

Article

Robust Non-Parametric Mortality and Fertility Modelling and Forecasting: Gaussian Process Regression Approaches

Ka Kin Lam  and Bo Wang ^{*} 

School of Mathematics and Actuarial Science, University of Leicester, Leicester LE1 7RH, UK; kkl21@le.ac.uk

^{*} Correspondence: bo.wang@le.ac.uk; Tel.: +44-116-252-2162

Abstract: A rapid decline in mortality and fertility has become major issues in many developed countries over the past few decades. An accurate model for forecasting demographic movements is important for decision making in social welfare policies and resource budgeting among the government and many industry sectors. This article introduces a novel non-parametric approach using Gaussian process regression with a natural cubic spline mean function and a spectral mixture covariance function for mortality and fertility modelling and forecasting. Unlike most of the existing approaches in demographic modelling literature, which rely on time parameters to determine the movements of the whole mortality or fertility curve shifting from one year to another over time, we consider the mortality and fertility curves from their components of all age-specific mortality and fertility rates and assume each of them following a Gaussian process over time to fit the whole curves in a discrete but intensive style. The proposed Gaussian process regression approach shows significant improvements in terms of forecast accuracy and robustness compared to other mainstream demographic modelling approaches in the short-, mid- and long-term forecasting using the mortality and fertility data of several developed countries in the numerical examples.

Keywords: demographic modelling; mortality forecasting; fertility forecasting; Gaussian process; non-parametric regression; Lee-Carter model



Citation: Lam, K.K.; Wang, B. Robust non-parametric mortality and fertility modelling and forecasting: Gaussian process regression approaches.

Forecasting **2021**, *3*, 207–227. <https://doi.org/10.3390/forecast3010013>

Received: 28 January 2021

Accepted: 6 March 2021

Published: 9 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There has been an increasing demand for demographic modelling and forecasting over the last few decades, driven by many developed countries are now suffering a rapid decline in mortality and fertility, leading to a significant increase in expenditures on health services for an ageing population and a shortage of future labour. Although many factors, including wealth, religion and migration, play important roles in demographic modelling and forecasting, mortality and fertility have been the principal determinants of the demographic models during the last century. Hence a better understanding of the mortality and fertility patterns and trends is of great importance for all stakeholders in a society as the mortality forecasts, for example, play a vital role for the insurance and pensions industries in pricing their insurance products. The fertility predictions are also of great interest to the government and education sectors in planing children's welfare and educational services.

Unlike the biological and the medical methods, statisticians have developed very different and purely mathematical methods to model the demographic patterns and trends which are well-documented by Preston et al. [1]. The history of demographic modelling with the mathematical approaches can be traced back to some deterministic models proposed in the mid-nineteenth century, see, for example, Gompertz [2] and Makeham [3]. The deterministic models are, however, restricted with few fixed factors and have no stochastic process considered owing to the lack of computing capability in that early period. With the advance of technology in computing, stochastic modelling has become a mainstream method of mortality and fertility curve fittings over the last three decades.

A significant milestone of the demographic modelling literature is the seminal work done by Lee and Carter [4], well-known as the Lee-Carter model, which can model and extrapolate a long-term mortality trend as a stochastic time series. It rapidly gained credit and popularity, given its simplicity and ability to capture most variations in demographic data evolved over time. For instance, Lee [5] applies it to fertility modelling and forecasting, and there has also been a series of extensions, variants and modifications proposed afterwards, see, for example, Bell [6], Lee and Miller [7], Booth et al. [8], Brouhns et al. [9], Renshaw and Haberman [10], Cairns et al. [11] and Hunt and Blake [12].

In the meantime, non-parametric or data-driven techniques, which do not need to meet certain assumptions on parametric forms in the model calibration, have also been introduced and developed in demographic modelling. This kind of techniques can be dated back to the classical graduation technique [13], which maintains an agnostic view of historical experience and solely focuses on removing random fluctuations in observed data then directly extrapolates the past trend to the future. In more recent years, Currie et al. [14] use P-splines to smooth the historical mortality surface across both age and year dimensions before fitting. Hyndman and Ullah [15] extend the Lee-Carter model to a functional data framework with a non-parametric smoothing method that allows for smooth functions of age and is more robust for demographic modelling. Some other developments in this area include Delwarde et al. [16], Debón et al. [17], Li et al. [18], Ludkovski et al. [19], Dokumentov et al. [20], Wu and Wang [21] and Alexopoulos et al. [22].

As an alternative to existing methods as well as having several desirable advantages over others, in this paper we propose a new non-parametric approach using Gaussian process regression with a natural cubic spline mean function and a spectral mixture covariance function for mortality and fertility modelling and forecasting. Unlike most of the existing approaches in demographic modelling, which depend on a few parameters to determine the movement of the whole mortality or fertility curve shifting from one year to another, we consider the mortality and fertility curves from their components of all age-specific mortality and fertility rates and assume each of them following a Gaussian process over time to fit the whole curves in a discrete but dense style.

More will be discussed in detail in the paper, and the rest of this paper is organised as follows. We briefly outline the theoretical background of Gaussian process regression in Section 2. In Section 3, we describe the framework of the proposed Gaussian process regression model for mortality and fertility modelling and forecasting. We then illustrate the proposed model with applications to the empirical mortality and fertility data, followed by comparisons to other existing approaches in terms of the systematic differences and forecasting performances using the observed mortality and fertility data of ten developed countries in Section 4. We lastly conclude this paper with discussions and remarks in Section 5.

2. Theoretical Background of Gaussian Process Regression

Gaussian process regression (also known as Kriging) is a regression method which belongs to a class of Bayesian non-parametric approaches to inferencing and modelling of an unknown latent function. This approach can be seen as conditioning of test data on training data in a joint Gaussian distribution. The theoretical basis for Gaussian process regression (GPR) was initially developed for estimating the most likely distribution of gold based on samples from a few boreholes in South Africa [23], and it was mainly used for spatial analysis and natural resources evaluation in geostatistics in the past, see, for example, Matheron [24], Journel and Huijbregts [25] and Cressie [26].

Over the last two decades, GPR has gained its popularity rapidly as different statistical tools among the data science community, such as, Engel et al. [27], Krause et al. [28] and Neal [29], due to its many desirable properties, such as the existence of explicit forms, the ease of obtaining and expressing uncertainty in predictions, the ability to capture a wide variety of patterns through different covariance functions, and a natural Bayesian interpretation. It has also been recognised as a powerful tool for regressions and forecast-

ing problems in the field of economics thanks to its abilities to quantify the uncertainties associated with the historical experience of the observed data and to generate the full stochastic trajectories for out-of-sample forecasts with prediction intervals under the Gaussian probabilistic framework. It is also capable of scaling to a large dataset with a minimum set of tunable (hyper-) parameters involved as well as capturing non-linear or periodic dynamics with a high degree of analytic tractability that extends the explorabilities and explainabilities of the classical linear regression model in more complicated scenarios. Some economic forecasting applications with GPR can be found in the literature. For instance, Alamaniotis et al. [30] perform a short-term load forecasting using an ensemble of GPR approaches, and Wu et al. [31] predict the tourism demand volume in Hong Kong using GPR. Since GPR is the fundamental technique applied in this article, we will give a brief review of GPR which will be used later. Readers can refer to MacKay [32] and Williams and Rasmussen [33] for more complete discussions of GPR.

Gaussian Process Regression

Let a training dataset with n pairs of observations of univariate covariates and responses be $\{t_i, y(t_i)\}_{i=1}^n$. We consider the following non-linear regression model

$$y(t_i) = f(t_i) + \epsilon(t_i), \quad \epsilon(t_i) \sim \mathcal{N}(0, \sigma^2),$$

where $f(\cdot)$ is the unknown function that needs to be estimated, and $\{\epsilon(t_i)\}_{i=1}^n$ are the independent and identically normally distributed error terms with mean 0 and constant variance σ^2 .

Following the Gaussian process paradigm, the unknown function $\{f(t) : t \in \mathcal{T}\}$ is assumed to have a Gaussian process prior with a specified mean function and a specified covariance function over the domain \mathcal{T} . It gives for all $t \in \mathcal{T}$

$$f(t) \sim \mathcal{GP}(\mu(t), K(t, t')), \quad (1)$$

where the mean function $\mu(t)$ and the covariance function $K(t, t')$ are defined as

$$\mu(t) = \mathbb{E}[f(t)],$$

$$K(t, t') = \text{Cov}(f(t), f(t')) = \mathbb{E}[(f(t) - \mu(t))(f(t') - \mu(t'))].$$

The specified mean function $\mu(\cdot)$ and the specified covariance function $K(\cdot, \cdot)$ in the Gaussian process prior reflect our prominent belief in the unknown function $f(\cdot)$, prior to any information about the observed data. A well-specified prior mean function has a profound impact on the forecast performance since the mean function will dominate the forecast results of the Gaussian process regression model in regions far beyond from the historical data. Meanwhile, the specified covariance function encodes the correlation between any pair of outputs $\{y(t), y(t')\}$, which determines the relativeness of one point to another, such as smoothness or periodicity.

A common covariance function example is the squared exponential covariance function which is used to model a smooth function. It has the form

$$K(t, t') = h^2 \exp\left(-\frac{(t - t')^2}{2l^2}\right),$$

where h is the response-scale amplitude determining the variation of function values, l is the characteristic length-scale which gives smooth variations in a covariate-scale and controls how far the observed data can be extrapolated.

Another common example is the periodic covariance function, which is used to model a periodic function. It gives

$$K(t, t') = h^2 \exp\left(-\frac{2\sin(\pi(t - t')/p)^2}{l^2}\right),$$

where p determines the distance between repetitions of a function, h and l are the same as in the squared exponential covariance function above.

A wide variety of covariance functions has been proposed, see, for example, Williams and Rasmussen [33], and this allows us great flexibilities in modelling among many different scenarios.

With the assumption of the Gaussian distribution on a collection of all the discrete observations, $\mathbf{y} = [y(t_1), \dots, y(t_n)]^T$ follows an n -variate normal distribution, i.e.,

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}),$$

where $\boldsymbol{\mu} = [\mu(t_1), \dots, \mu(t_n)]^T$, and \mathbf{K} is the $n \times n$ covariance matrix with entries $K(t_i, t_j) + \sigma^2 \delta_{t_i, t_j}$. Here $\delta_{t_i, t_j} = 1$ if $t_i = t_j$ and 0 otherwise.

The log-likelihood function of the n -dimensional collection of the discrete observations \mathbf{y} for all the (hyper-) parameters in the specified mean function and the specified covariance function (denoted by a generic $\boldsymbol{\theta}$) and the noise parameter σ^2 is

$$\mathcal{L}(\boldsymbol{\theta}, \sigma^2) = -\frac{1}{2} \log |\mathbf{K}| - \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^T (\mathbf{K})^{-1} (\mathbf{y} - \boldsymbol{\mu}) - \frac{n}{2} \log(2\pi), \quad (2)$$

where $|\cdot|$ denotes the determinant of a matrix.

Standard gradient-based numerical optimisation techniques, such as Conjugate Gradient method, can be used to maximise the log-likelihood function $\mathcal{L}(\boldsymbol{\theta}, \sigma^2)$ in Equation (2) to obtain the estimates of the model parameters.

For the corresponding value $y(t^*)$ at any measurement time point t^* , we can denote the joint distribution of $[y(t_1), \dots, y(t_n), y(t^*)]^T$ as an $(n+1)$ -variate normal distribution with a mean vector $[\mu(t_1), \dots, \mu(t_n), \mu(t^*)]^T$ and a covariance matrix as

$$\begin{bmatrix} \mathbf{K} & \mathbf{K}^* \\ \mathbf{K}^{*T} & K(t^*, t^*) + \sigma^2 \end{bmatrix},$$

where $\mathbf{K}^* = [K(t^*, t_1), \dots, K(t^*, t_n)]^T$.

The conditional distribution of $y(t^*)$ given \mathbf{y} with the estimated (hyper-) parameters $\hat{\boldsymbol{\theta}}$ and noise variance $\hat{\sigma}^2$ through the optimisation of the log-likelihood function in Equation (2), is then $\mathcal{N}(\hat{y}(t^*), \hat{\sigma}^{*2})$, where

$$\hat{y}(t^*) = \mu(t^*) + \mathbf{K}^{*T} \mathbf{K}^{-1} (\mathbf{y} - \boldsymbol{\mu}),$$

$$\hat{\sigma}^{*2} = K(t^*, t^*) + \hat{\sigma}^2 - \mathbf{K}^{*T} \mathbf{K}^{-1} \mathbf{K}^*.$$

The Gaussian process regression approach for conditioning of an unknown value by some realisations of a stochastic process in a joint Gaussian distribution can be seen as expanding discrete data from a function space point of view over the same domain. This idea can also be applied to the task of forecasting. With the Gaussian probabilistic framework, it can quantify the uncertainty associated with the historical experience of the observed data and then generate the full stochastic trajectories for out-of-sample forecasts with prediction intervals. We will discuss this in detail with applications to mortality and fertility data in the next section.

3. Methodology

3.1. Gaussian Process Regression (GPR) Model for Mortality and Fertility Modelling and Forecasting

In this section, we introduce the proposed Gaussian process regression (GPR) model for mortality and fertility modelling and forecasting.

Let our discrete observed log mortality (or fertility) rates dataset be $\{t_i, y_x(t_i)\}$ for $i = 1, \dots, n$ and $x = 0, \dots, m$, where n is the total number of observed calendar years t , m is the maximum age of interest, and $y_x(t_i)$ is the log of mortality (or fertility) rates for a

given age x in a calendar year t_i . The proposed GPR model for mortality (or fertility) rates modelling and forecasting in a given age x is

$$y_x(t_i) = f_x(t_i) + \epsilon_x(t_i), \quad \epsilon_x(t_i) \sim \mathcal{N}(0, \sigma_x^2),$$

where $f_x(t_i)$ is the underlying function that needs to be estimated, $\{\epsilon_x(t_i)\}_{i=1}^n$ allow the observation errors varying based on the assumption of i.i.d. normally distributed random variables with mean 0 and constant variance σ_x^2 for a given age x .

Following the Gaussian process regression paradigm discussed in Section 2, the underlying function $\{f_x(t) : t \in \mathcal{T}\}$ for mortality (or fertility) modelling and forecasting of each age x is assumed to follow a Gaussian process with a specified prior mean function $\mu_x(t)$ and a specified prior covariance function $K_x(t, t')$ over the domain \mathcal{T} (note that mortality and fertility data are not directly of a functional nature, and here we assume that there are underlying functional time series which are observed with errors at discrete points). It gives for all $t \in \mathcal{T}$

$$f_x(t) \sim \mathcal{GP}(\mu_x(t), K_x(t, t')). \quad (3)$$

3.1.1. Specified Gaussian Process Prior Mean Function

Under the Gaussian process regression paradigm, we are allowed to choose a prior mean function that reflects our prominent belief in the unknown function $f_x(t)$ for the Gaussian process regression model. Due to incomplete information on the functional form among the time-series of age-specific mortality and fertility rates and a wide range of the patterns as exhibited in our numerical examples in Section 4 (see Figures 1a and 2a), we adopt a non-parametric mean function, namely the natural cubic spline function, in the proposed GPR model for mortality and fertility modelling and forecasting. The justifications for this choice here are that the natural cubic spline function can handle a broad range of complex and non-linear functions that exist in all the age-specific mortality and fertility rates as a general and unified approach. It behaves like a higher-order polynomial regression model for curve fitting with no presumption needed on the relationship between the covariate and response variables. Meanwhile, it can extrapolate the trend of historical data smoothly into the future in an appropriate direction without suffering the overfitting problem that occurs commonly in the higher-order polynomial regression model when it comes to forecasting. The information involved in the actual observations of the data is reflected in the coefficients involved which need to be estimated based on the data.

To explain further how it works mathematically, we consider the natural cubic spline function with strictly increasing K knots $\{\xi_k\}_{k=1}^K$ where $\xi_1 < \xi_2 < \dots < \xi_K$ over the observed time interval $[t_1, t_n]$ can be represented as the truncated power series representation for a cubic spline function. It gives

$$\mu_x(t) = \sum_{p=0}^3 \alpha_p t^p + \sum_{k=1}^K \beta_k (t - \xi_k)_+^3, \quad (4)$$

with additional constraints on the boundary conditions when $t < \xi_1$ and $t \geq \xi_K$, such that

$$\alpha_2 = 0, \alpha_3 = 0, \sum_{k=1}^K \beta_k = 0, \text{ and } \sum_{k=1}^K \beta_k \xi_k = 0, \quad (5)$$

where $(\cdot)_+$ denotes the positive part, $\{\alpha_p\}_{p=0}^3$ are the coefficients of the cubic polynomial and $\{\beta_k\}_{k=1}^K$ are the coefficients of the truncated power series for the cubic splines with K interior knots.

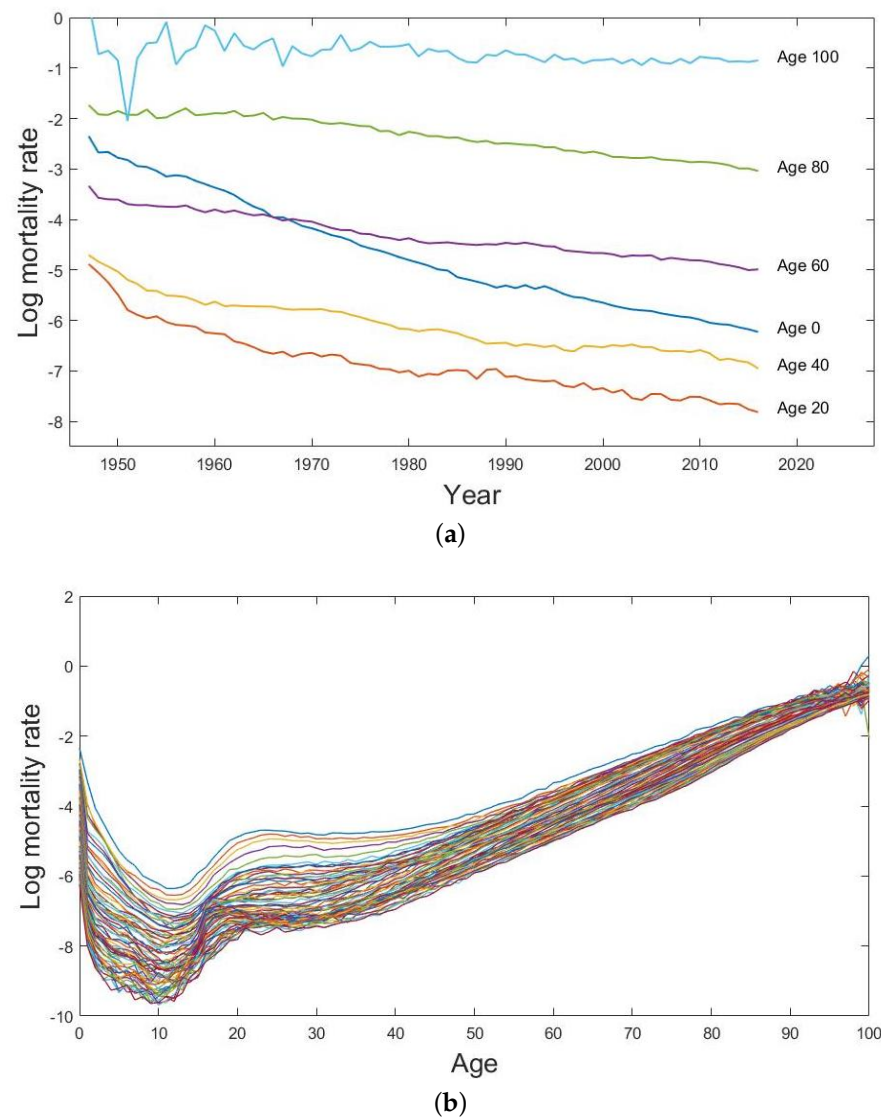


Figure 1. (a) Univariate time series of the log age-specific male mortality rates with 20-year age intervals and (b) the log male mortality curves from age 0 to age 100 from the year 1947 to the year 2016 in Japan.

By imposing the above boundary constraints in Equation (5), the natural cubic spline function can be further derived as a reduced basis when $t \geq \zeta_K$ for extrapolation as

$$\begin{aligned}
 \mu_x(t) &= \sum_{p=0}^3 \alpha_p t^p + \sum_{k=1}^K \beta_k (t - \zeta_k)^3 \\
 &= \sum_{p=0}^3 \alpha_p t^p + \sum_{k=1}^K \beta_k (t^3 - 3\zeta_k t^2 + 3\zeta_k^2 t - \zeta_k^3) \\
 &= \alpha_0 + \alpha_1 t + \underbrace{\alpha_2 t^2}_{=0} + \underbrace{\alpha_3 t^3}_{=0} + \sum_{k=1}^K \beta_k t^3 - \sum_{k=1}^K 3\beta_k \zeta_k t^2 + \sum_{k=1}^K 3\beta_k \zeta_k^2 t - \sum_{k=1}^K \beta_k \zeta_k^3 \\
 &= \alpha_0 + \alpha_1 t + \sum_{k=1}^K 3\beta_k \zeta_k^2 t - \sum_{k=1}^K \beta_k \zeta_k^3 \\
 &= (\alpha_0 - \sum_{k=1}^K \beta_k \zeta_k^3) + (\alpha_1 + \sum_{k=1}^K 3\beta_k \zeta_k^2) t.
 \end{aligned} \tag{6}$$

For simplicity of notation, we denote $c_0 = (\alpha_0 - \sum_{k=1}^K \beta_k \zeta_k^3)$ and $c_1 = (\alpha_1 + \sum_{k=1}^K 3\beta_k \zeta_k^2)$. Then

$$\mu_x(t) = c_0 + c_1 t,$$

thus $\mu_x(t)$ is a linear function with constants c_0 and c_1 when $t \geq \zeta_K$ for extrapolation.

Unlike the cubic spline function whose behaviour tends to be erratic near boundaries and the extrapolation can be unrealistic outside the observed time interval $[t_1, t_n]$, the natural cubic spline function forces its extrapolation outside regions beyond the observed data to be a linear function with zero second derivative that agrees with the spline in its intercept and slope values among its last knot [34]. With this distinctive feature, the natural cubic spline function can hence inherit the local information of recent data and gives a reasonable direction for extrapolation. In demographic modelling, it is often the case that more recent experience has greater relevance on future behaviour than those data from the distant past. On account of this, we specify the natural cubic spline function as the mean function for mortality and fertility modelling and forecasting in the proposed GPR model.

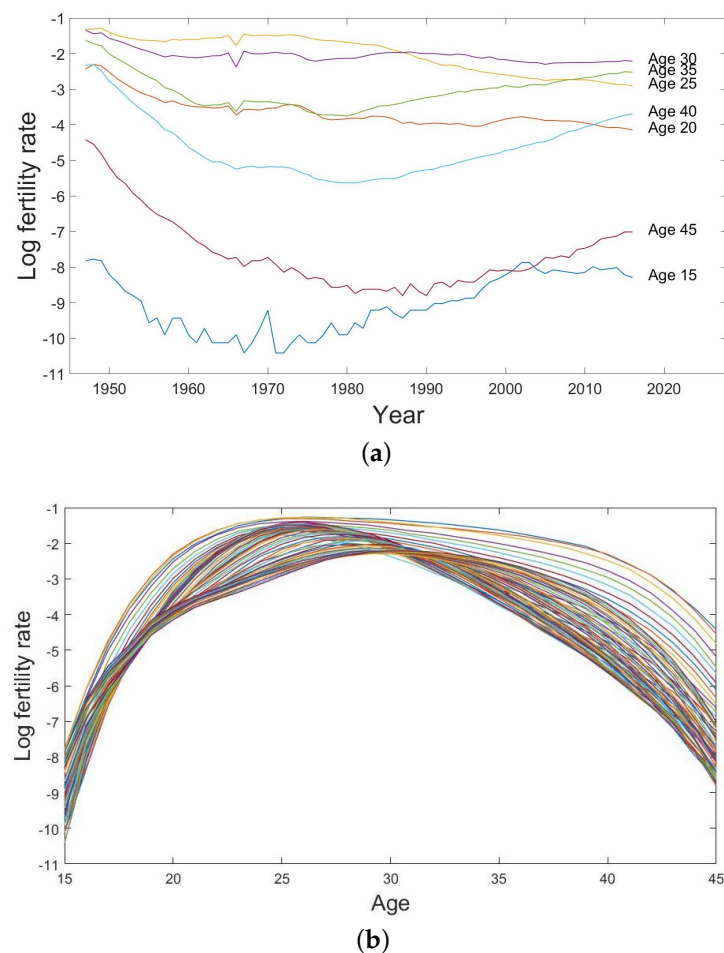


Figure 2. (a) Univariate time series of the log fertility rates with 5-year age intervals and (b) the log fertility curves from age 15 to age 45 from the year 1947 to the year 2016 in Japan.

3.1.2. Specified Gaussian Process Prior Covariance Function

In the proposed GPR model, the specified covariance function is used to discover and capture the similarities and various time-correlated structures among the historical demographic trends for different age groups, then project their patterns into the future.

Given that the choice of covariance functions in Gaussian process regression remains an ongoing research problem, in our study a range of commonly used covariance functions are tested. Apart from the squared exponential and the periodic covariance functions

mentioned in Section 2, we have also considered and tested four more common covariance functions, including the rational quadratic, Matérn 3/2 and Matérn 5/2 covariance functions documented in the Gaussian process literature [33], and the spectral mixture covariance function proposed by Wilson and Adams [35]. The spectral mixture covariance function is derived by modelling a spectral density of the Fourier transform of input-correlated structures with a Gaussian mixture, which can automatically discover patterns and extrapolate far beyond the available data; see Wilson and Adams [35] for detailed discussions.

From our empirical experiments, we have found that the spectral mixture covariance function is indeed relatively more successful in capturing different patterns of age-specific mortality and fertility rates. It also provides more accurate forecast results than the other covariance functions for mortality and fertility predictions. Wu and Wang [21] also discovered the similar results. The spectral mixture covariance function is thus selected as the primary covariance function in the proposed GPR model for all age-specific mortality and fertility rates as a unified approach.

Considering a Q number of Gaussian mixture components, in which the q -th component has mean λ_q and variance ν_q^2 , the spectral mixture covariance function is defined as

$$K_x(t, t') = \sum_{q=1}^Q w_q \exp\{-2\pi^2(t - t')^2 \nu_q^2\} \cos(2\pi(t - t')\lambda_q), \quad (7)$$

where w_q is the weight specifying the contribution of the q -th Gaussian mixture component.

3.1.3. Likelihood Function of the Proposed GPR Model

It yields that $\mathbf{y}_x = [y_x(t_1), \dots, y_x(t_n)]^T$ in each given age x follows an n -variate normal distribution, such that

$$\mathbf{y}_x \sim \mathcal{N}(\boldsymbol{\mu}_x, \mathbf{K}_x),$$

where $\boldsymbol{\mu}_x = [\mu_x(t_1), \dots, \mu_x(t_n)]^T$ and \mathbf{K}_x is the $n \times n$ specified covariance matrix with entries $K_x(t_i, t_j) + \sigma_x^2 \delta_{t_i, t_j}$. Here $\delta_{t_i, t_j} = 1$ if $t_i = t_j$ and 0 otherwise.

The log-likelihood function of the n -dimensional collection of the discrete observations \mathbf{y}_x in each age x for all the (hyper-) parameters (denoted by a generic $\boldsymbol{\theta}_x$) and the noise variance σ_x^2 is

$$\mathcal{L}(\boldsymbol{\theta}_x, \sigma_x^2) = -\frac{1}{2} \log |\mathbf{K}_x| - \frac{1}{2} (\mathbf{y}_x - \boldsymbol{\mu}_x)^T (\mathbf{K}_x)^{-1} (\mathbf{y}_x - \boldsymbol{\mu}_x) - \frac{n}{2} \log(2\pi), \quad (8)$$

where $|\cdot|$ is the determinant of a matrix.

Standard gradient-based numerical optimisation techniques can be used to maximise the log-likelihood function $\mathcal{L}(\boldsymbol{\theta}_x, \sigma_x^2)$ in Equation (8) to obtain the estimates of the model parameters.

3.1.4. Out-of-Sample Forecasts and Prediction Intervals of the Proposed GPR Model

Let t_h^* be the h -step ahead of the last observed calendar year where $t_n < t_1^* < \dots < t_h^*$ for all $t_i^* \in \mathcal{T}$. Then the joint distribution of $[y_x(t_1), \dots, y_x(t_n), y_x(t_1^*), \dots, y_x(t_h^*)]^T$ is an $(n + h)$ -variate normal distribution that comes with the mean vector $[\mu_x(t_1), \dots, \mu_x(t_n), \mu_x(t_1^*), \dots, \mu_x(t_h^*)]^T$ and the covariance matrix is

$$\begin{bmatrix} \mathbf{K}_x & \mathbf{K}_x^* \\ \mathbf{K}_x^{*T} & \mathbf{K}_x^{**} \end{bmatrix},$$

where \mathbf{K}_x^* is the size of $n \times h$ covariance matrix, and \mathbf{K}_x^{**} is the size of $h \times h$ covariance matrix with entries $K_x(t_i^*, t_j^*) + \sigma_x^2 \delta_{t_i^*, t_j^*}$. Here $\delta_{t_i^*, t_j^*} = 1$ if $t_i^* = t_j^*$ and 0 otherwise.

Therefore, the h -step ahead out-of-sample forecast and its variance of the age-specific mortality (or fertility) rates can be found from the conditional distribution of

$\mathbf{y}_x^* = [y_x(t_1^*), \dots, y_x(t_h^*)]^T$ given \mathbf{y}_x with the estimated (hyper-) parameters $\hat{\boldsymbol{\theta}}_x$ and noise variance $\hat{\sigma}_x^2$ through the optimisation of the log-likelihood function in Equation (8), is then $\mathcal{N}(\hat{\mathbf{y}}_x^*, \text{Var}(\mathbf{y}_x^*))$, and let $\boldsymbol{\mu}_x^* = [\mu_x(t_1^*), \dots, \mu_x(t_h^*)]^T$, where

$$\hat{\mathbf{y}}_x^* = \boldsymbol{\mu}_x^* + \mathbf{K}_x^{*T} \mathbf{K}_x^{-1} (\mathbf{y}_x - \boldsymbol{\mu}_x), \quad (9)$$

$$\text{Var}(\mathbf{y}_x^*) = \mathbf{K}_x^{**} - \mathbf{K}_x^{*T} \mathbf{K}_x^{-1} \mathbf{K}_x^* + \hat{\sigma}_x^2 \mathbf{I}. \quad (10)$$

With the normality assumptions on the model error and the known $\text{Var}(\mathbf{y}_x^*)$, a $100(1 - \alpha)\%$ prediction interval for \mathbf{y}_x^* can be calculated as $\mathbf{y}_x^* \pm z_\alpha \sqrt{\text{Var}(\mathbf{y}_x^*)}$, where z_α is the $(1 - \alpha/2)$ quantile of the standard normal distribution.

4. Applications

In this section, we demonstrate the usefulness of the proposed GPR model for modelling and forecasting two different sets of demographic data—mortality data and fertility data. We first apply the proposed model to the male mortality data and the fertility data of Japan for illustration purposes. We then compare and evaluate the quality of the fitted mortality and fertility curves by the proposed method with some other existing approaches using the mortality and fertility data of ten different developed countries.

4.1. Male Mortality Data

The male mortality data of Japan are available from the year 1947 to the year 2016 from the Human Mortality Database [36]. The age-specific male mortality rates are defined as the number of deaths in males during a calendar year, proportional to the male resident population of the same age during the same calendar year. The database consists of the age-specific male mortality rates by a single calendar year of age up to 110 years old. We restrict our experimental mortality data up to age 100 to avoid any potential problem associated with the erratic mortality rates above age 100.

The observed male mortality data are presented in Figure 1a as separate univariate time series of the log age-specific male mortality rates with 20-year age intervals from age 0 to age 100 from the year 1947 to the year 2016. We can see that there is a general decrease in male mortality rates among all the selected age groups during the examined period. The decline in mortality rates at higher ages seems to change more slowly than those at younger ages. Figure 1b presents the log male mortality curves, which give us information about the general trends and the variations of the age-specific male mortality rates over the observed period. There is an apparent hump around 18 to 25 years old which usually relates to reckless behaviour in teenage ages and a general drop for all population in mortality rates over time, due primarily to the advances in medical technology.

4.2. Mortality Modelling and Forecasting

In the demonstration of the proposed model, we aim to display 10-years-ahead out-of-sample forecasts of mortality rates. We first split the dataset into a training dataset with the observed mortality rates from the year 1947 to the year 2006 and a test dataset with the remaining observed mortality data from the year 2007 to the last sample data in the year 2016. We use our GPR model in Equation (3) to fit the male mortality rates for each given age in the training dataset. We equally place the interior knots ξ inside the quantile position of the training time interval ($K = 4$) in the natural cubic spline mean function in Equation (4), and use two Gaussian mixture components ($Q = 2$) in the spectral mixture covariance function in Equation (7). These settings are the minimal requirements for the number of parameters needed for the Gaussian process mean function and covariance function. Further tests on the numbers of knots K and Gaussian mixture components Q have been made, and no significant difference in our experiment results was found when $K > 4$ and $Q > 2$. The parameters $\boldsymbol{\theta}_x$ including the parameters of the mean function $\boldsymbol{\theta}_{\mu_x}$ and the hyperparameters of the covariance function $\boldsymbol{\theta}_{K_x} = \{(w_q, m_q, \nu_q)\}_{q=1}^Q$ and the noise variance σ_x^2 can be estimated by maximising the log likelihood function in Equation (8)

for each given age x . However, estimating all unknown parameters involved both in the mean and the covariance structures for all age groups can be time demanding with high computational costs in practice. Following the suggestions of Shi et al. [37], we remedy this problem by computing the parameters of the mean function through the ordinary least squares approach and then estimating the parameters of the covariance function by maximising the log likelihood function in Equation (8). This estimating procedure is analogous to de-trending the age-specific mortality rates by the empirical mean function then modelling the residuals by a zero-mean Gaussian process. It indeed improved the stability and the speed of the estimation procedure for all (hyper-) parameters in our numerical examples.

The predictive mean and variance for forecasting age-specific mortality rates can be obtained by Equations (9) and (10) respectively. As demonstration Figure 3 shows the predicted results of some age-specific mortality rates by the proposed GPR model. We can see that the proposed model can capture the varying patterns in mortality rates among different age groups properly. The predicted mortality curve for all ages for any specified year can then be constructed by extracting the predicted age-specific mortality rates from age 0 to age 100 in that specified year. Figure 4 gives an example of the 10-years-ahead out-of-sample forecast results of the male mortality curve with the 95% prediction intervals using the proposed GPR model for the year 2016 based on the observations from the year 1947 to the year 2006.

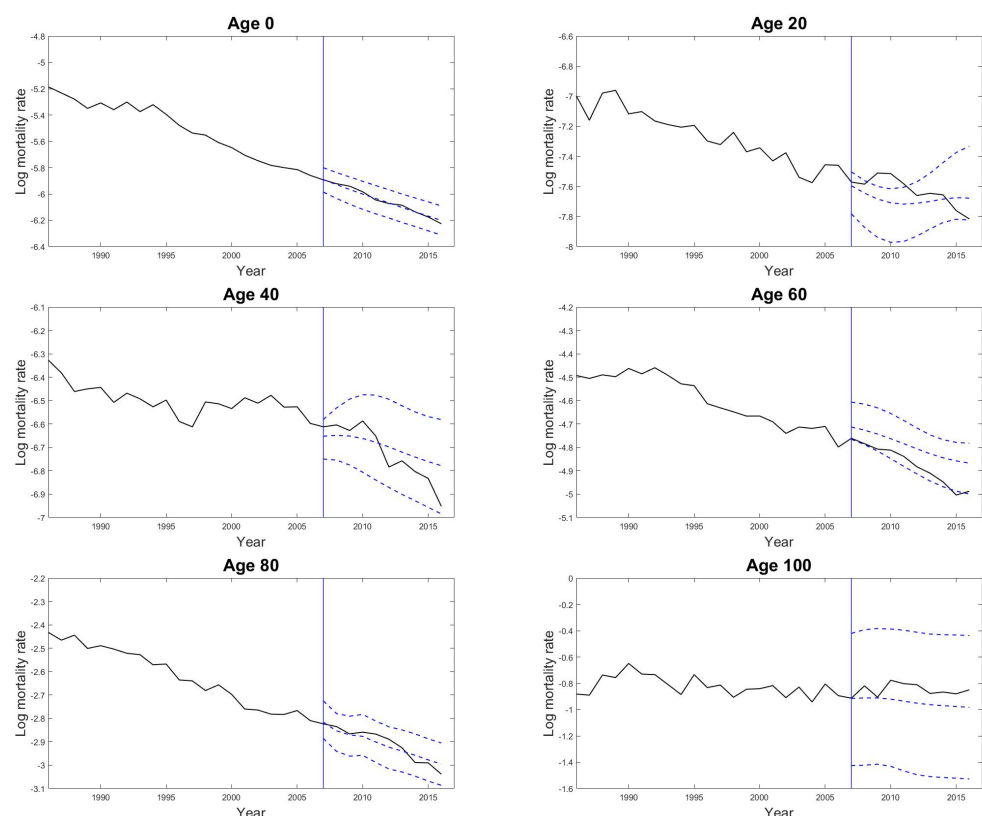


Figure 3. Predicted male mortality rates of the selected age groups from age 0 to age 100 with 20-year age intervals using the proposed GPR model from the year 2007 to the year 2016 based on the observations from the year 1947 to the year 2006 in Japan. The solid lines are the observed values, and the dashed lines are the predictions and the 95% prediction intervals. The vertical line indicates the starting point of the predictions.

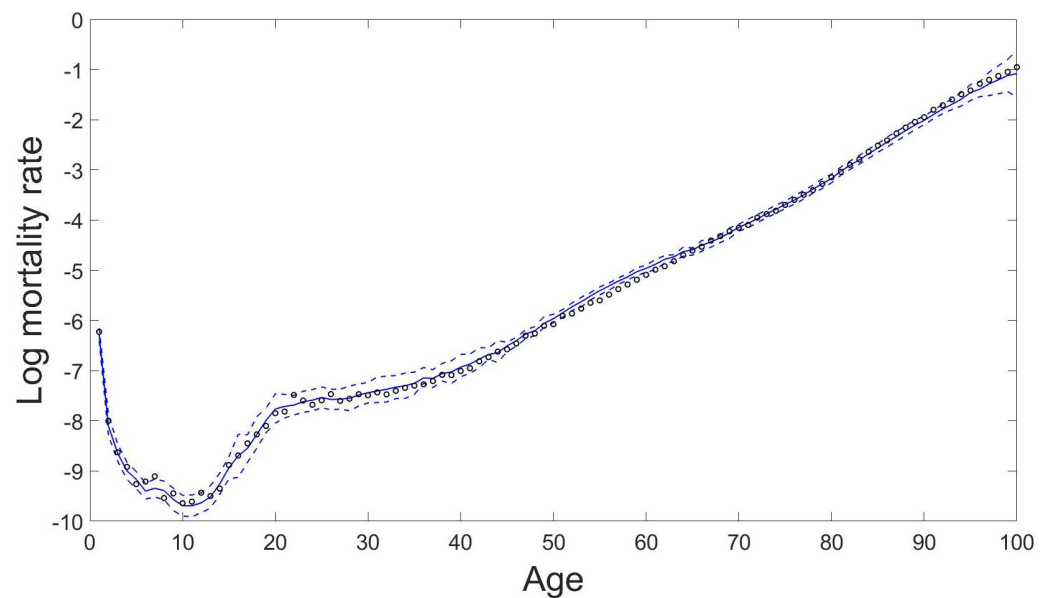


Figure 4. Predicted male mortality curve from age 0 to age 100 with the 95% prediction intervals using the proposed GPR model for the year 2016 based on the observations from the year 1947 to the year 2006 in Japan. The circles are the true log mortality rates, the solid line is the prediction, and the dashed lines are the 95% prediction intervals.

4.3. Fertility Data

We move to the second case study of modelling and forecasting the age-specific fertility rates. The age-specific fertility rates are defined as the number of births during a calendar year, based on the age of the mother proportional to the total number of the female resident population. The fertility data of Japan are available from the year 1947 to the year 2016 from the Human Fertility Database [38]. The database consists of the age-specific fertility rates by calendar year from 12 to 55 years old. In our study, we choose the age-specific fertility rates from age 15 to age 45, because the absence of the age-specific fertility data below age 15 and above age 45 is prevalent in the datasets from the Human Fertility Database. As we want to test the forecast performance of the proposed model for each age and for a range of countries without having to deal with the missing data issues we focus on the age-specific fertility rates from age 15 to age 45 only for which the data are available for all the countries in the study.

We present the historical fertility data of Japan as separate univariate time series of the log age-specific fertility rates with 5-year age intervals in Figure 2a and as the log fertility curve from age 15 to age 45 from the year 1947 to the year 2016 in Figure 2b. They reflect the patterns of fertility change caused by social conditions in different periods in Japan. For example, there was a ‘baby boom’ in all age groups after the end of World War II in 1945 and a rapid decrease in the birth rate during the ‘Japanese economic miracle period’ in the early 1980s due to a delay in child-bearing while the economy was establishing rapidly at that time. In more recent years, there was an increasing trend in fertility at higher ages, because the females devoted more time in their educations and careers. In general, females from age 20 to age 40 have relatively higher fertility rates compared to the age groups 15 and 45. The bunch of fertility curves in Figure 2b display a concave shape. It shows that the fertility rates climb from age 15 and reach their peak at around age 30 then decline. There exist some sparse patterns in the later part of the fertility curves, which reflects that the variations in birth rates above age 35 are obvious across the observed years.

4.4. Fertility Modelling and Forecasting

For the task of fertility forecasting, we again attempt to make 10-years-ahead out-of-sample forecasts of fertility rates for demonstration in this section. We maintain the same settings in the GPR model as for the mortality case, including the split of dataset, the number of interior knots in the mean structure and the Gaussian mixture components in the covariance structure with the same estimation procedures for all (hyper-) parameters involved. As examples Figure 5 presents the forecast results of some age-specific fertility rates by the GPR model. We can see that the GPR method can catch the varying patterns of fertility reasonably well, except the cases where the test data do not follow the historical trends determined by the training data. Figure 6 shows the predicted fertility curve from age 15 to age 45 with the 95% prediction intervals using the proposed GPR model for the year 2016 based on the observations from the year 1947 to the year 2006.

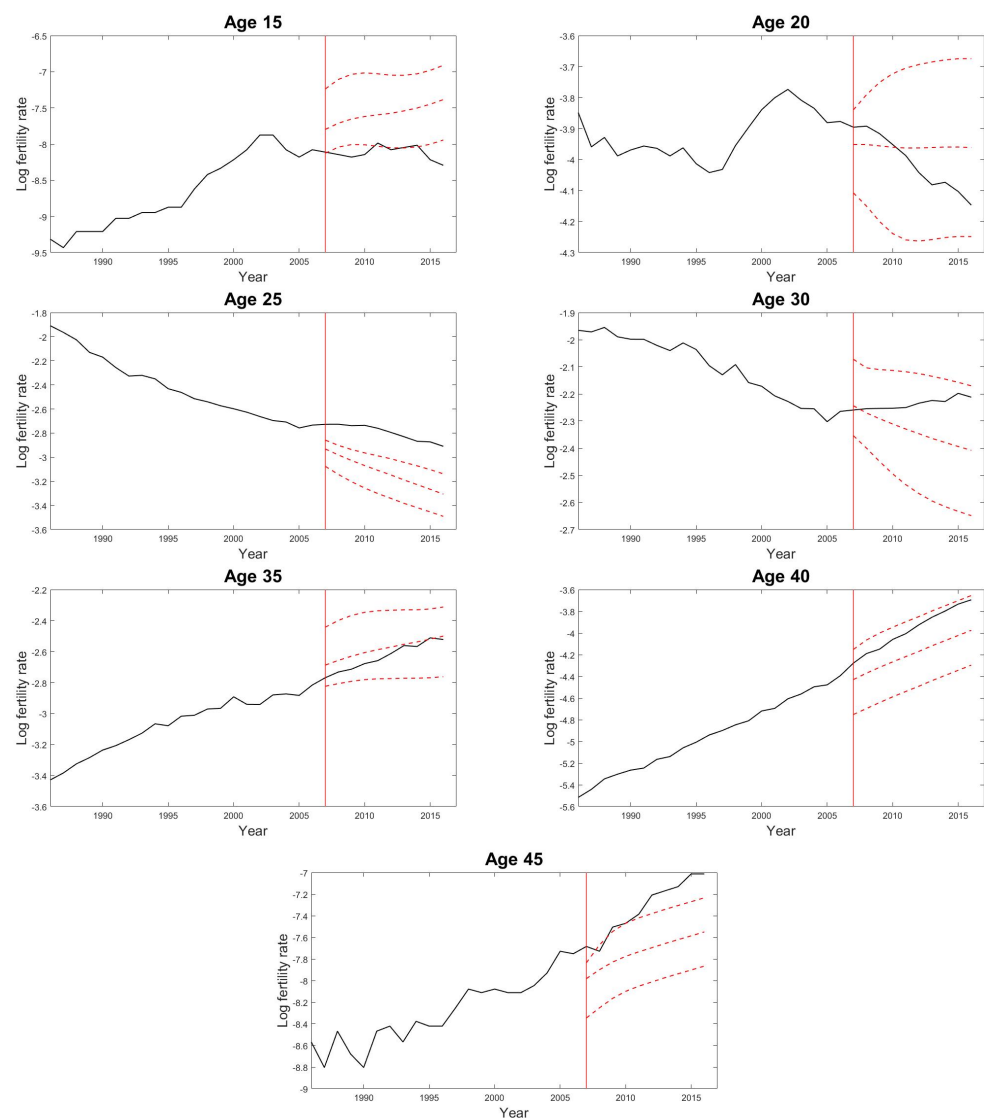


Figure 5. Predicted fertility rates of the selected age groups from age 15 to age 45 with 5-year age intervals using the GPR model from the year 2007 to the year 2016 based on the observations from the year 1947 to the year 2006 in Japan. The solid lines are the observed values, and the dashed lines are the predictions and the 95% prediction intervals. The vertical line indicates the starting point of the predictions.

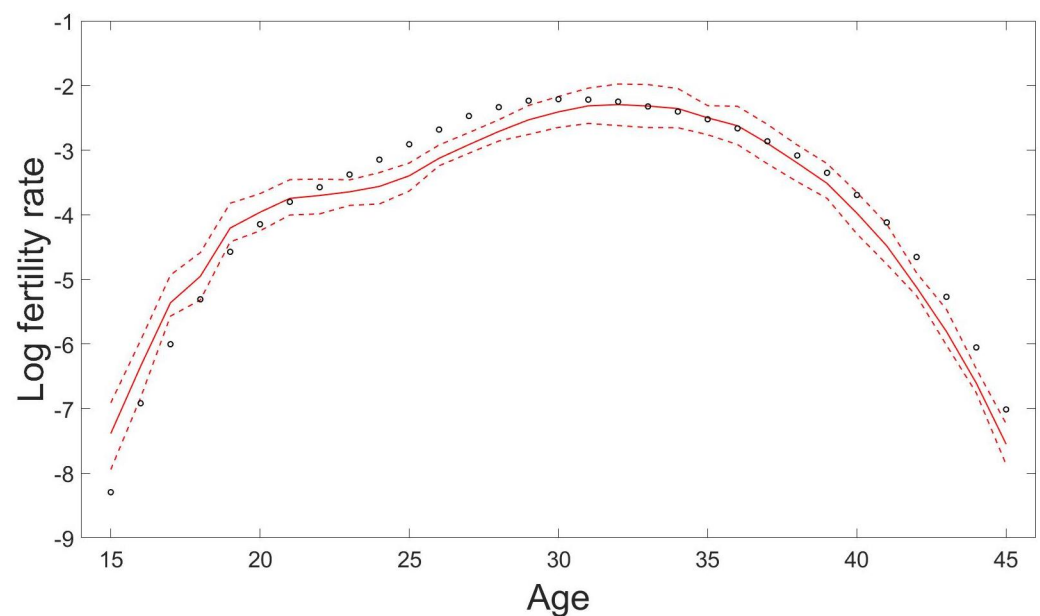


Figure 6. Predicted fertility curve from age 15 to age 45 with the 95% prediction intervals using the GPR model for the year 2016 based on the observations from the year 1947 to the year 2006 in Japan. The circles are the true log fertility rates, the solid line is the prediction, and the dashed lines are the 95% prediction intervals.

4.5. Comparisons and Forecast Accuracy Evaluations with Existing Models

We now compare and evaluate the forecast performance and accuracy of the proposed GPR model with four other mainstream approaches in the demographic modelling literature, namely, the Lee-Carter (LC) model [4], the Lee-Miller (LM) model [7], the Booth-Maindonald-Smith (BMS) model [8] and the Hyndman-Ullah (HU) model [15].

The LC model applies the principal component analysis to decompose the age-time matrix of the log mortality (or fertility) rates y_{x,t_i} into a linear combination of age and time parameters from the first-order principal component, i.e.

$$y_{x,t_i} = a_x + b_x k_{t_i} + \epsilon_{x,t_i}, \quad \epsilon_{x,t_i} \sim \mathcal{N}(0, \sigma^2), \quad (11)$$

where a_x is the averaged log mortality (fertility) rates at age x across all calendar years, b_x reflects the relative change in the log mortality (fertility) rates at age x , k_{t_i} measures the general time trend of the log mortality (fertility) rates, and $\{\epsilon_{x,t_i}\}_{i=1}^n$ are the i.i.d. normally distributed error terms with zero mean and constant variance σ^2 . The LC model relies on the h -step ahead extrapolated time parameter \hat{k}_{t_i+h} by some time series models, such as ARIMA model, to produce a h -year ahead forecast of the mortality (or fertility) curve across all age groups.

The LM model, the BMS model and the HU model can be thought of as the extensions and variants using the similar framework of the LC model in Equation (11). Their ways to make forecasts of the mortality (or fertility) curves also depend on the extrapolation of the time parameters derived from the age-time matrix of the log mortality (or fertility) rates [39]. More specifically, the LM model is a modification to the LC model where the coefficient series is adjusted so that the fitted life expectancy is equal to the observed life expectancy in each year in an attempt to reduce forecast bias. The BMS model modifies the LC model to adjust the coefficients using the age-at-death distribution and determines the optimal fitting period beforehand to address the non-linearity problem in the time component. The HU model extends the LC model by adapting the functional data paradigm using non-parametric smoothing to reduce outliers in observed data and more principal components for robust forecasts.

In contrast to the LC method and its numerous variants and extensions which rely on the time parameters to determine the movement of the whole mortality (or fertility) curve across all ages groups shifting from one year to another, our approach provides a novel point of view in demographic modelling and forecasting. The proposed GPR method treats demographic data in each age group as time series data and assumes each of them following a Gaussian process to achieve the same tasks in a discrete but intensive fashion. The natural cubic spline mean function in the proposed GPR method can automatically discover and extract more recent information from the observed data and then extrapolate its future trend in an appropriate direction. The spectral mixture covariance function in the GPR model, on the other hand, addresses the problems associated with the non-linearity and periodicity in the demographic data.

Figure 7 presents an example of the 10-years-ahead out-of-sample forecast results of the Japanese male mortality curves from age 0 to age 100 for the year 2016 using the LC model (with RMSE = 0.2172), the LM model (with RMSE = 0.1660), the BMS model (with RMSE = 0.1109), the HU model (with RMSE = 0.2342) and the GPR model (with RMSE = 0.0895) based on the observations from the year 1947 to the year 2006. The root mean square error (RMSE) measures the standard deviation of the average squared prediction error regardless of the positive or negative sign, and is defined here as

$$\text{RMSE} = \sqrt{\frac{1}{101} \sum_{x=0}^{100} \left(y_{x,t_{2016}} - \hat{y}_{x,t_{2016}} \right)^2},$$

where $y_{x,t_{2016}}$ is the log male mortality rates aged x in the year 2016.

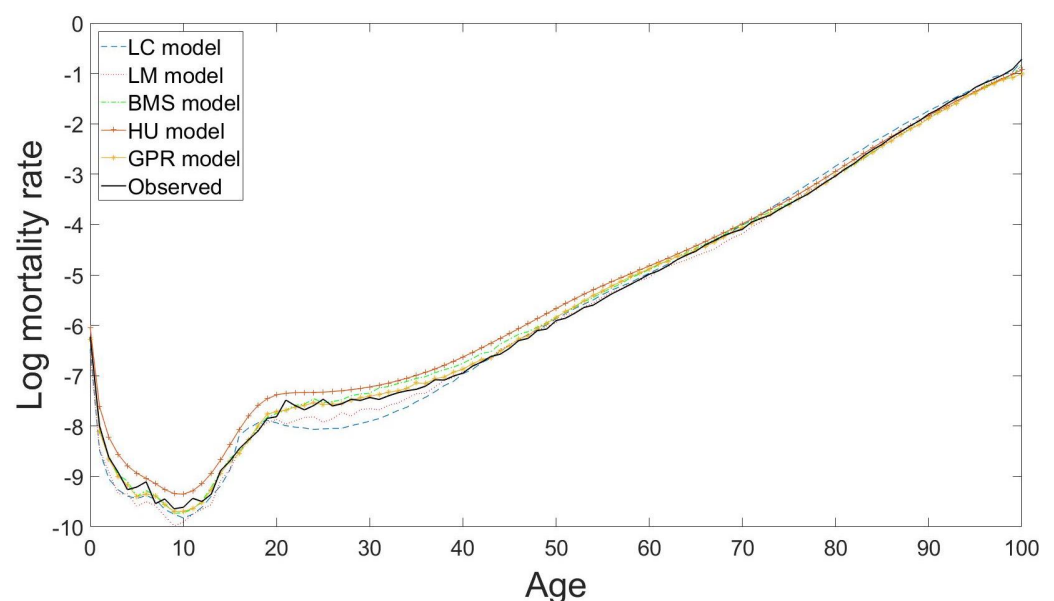


Figure 7. Predicted male mortality curves from age 0 to age 100 for the year 2016 using the LC model (with RMSE = 0.2172), the LM model (with RMSE = 0.1660), the BMS model (with RMSE = 0.1109), the HU model (with RMSE = 0.2342) and the GPR model (with RMSE = 0.0895) based on the observations from the year 1947 to the year 2006 in Japan.

Figure 8 shows another example of the 10-years-ahead out-of-sample forecast results of the Japanese fertility curves from age 15 to age 45 for the year 2016 using the LC model (with RMSE = 0.9401), the LM model (with RMSE = 0.7490), the BMS model (with RMSE = 1.2198),

the HU model (with RMSE = 0.4205) and the GPR model (with RMSE = 0.3764) based on the observations from the year 1947 to the year 2006. The RMSE here is

$$\text{RMSE} = \sqrt{\frac{1}{31} \sum_{x=15}^{45} \left(y_{x,t_{2016}} - \hat{y}_{x,t_{2016}} \right)^2},$$

where $y_{x,t_{2016}}$ is the log fertility rates aged x in the year 2016.

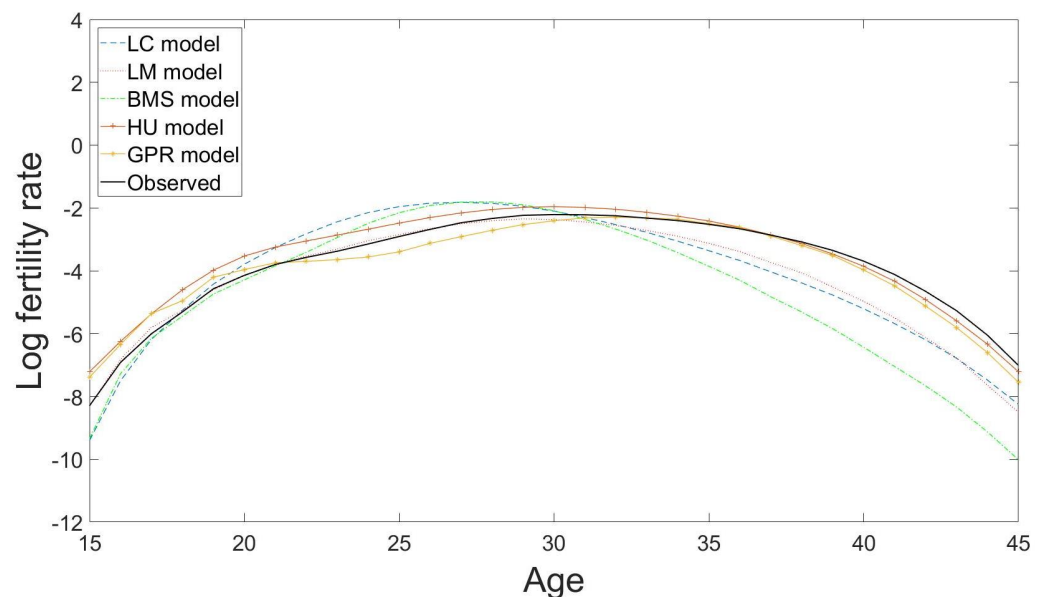


Figure 8. Predicted fertility curves from age 15 to age 45 for the year 2016 using the LC model (with RMSE = 0.9401), the LM model (with RMSE = 0.7490), the BMS model (with RMSE = 1.2198), the HU model (with RMSE = 0.4205) and the GPR model (with RMSE = 0.3764) based on the observations from the year 1947 to the year 2006 in Japan.

Forecast Accuracy Evaluations Using Rolling-Window Analysis

For evaluating the forecast accuracy, we consider ten developed countries for which mortality and fertility data are also available in the Human Mortality Database [36] and the Human Fertility Database [38]. We focus the data periods of all selected countries commencing in the year 1947 up to the year 2016 (70 years in total) for a unified purpose. Although we intent to use the same set of ten countries for both experiments in mortality and fertility, we are restricted by the availability of either mortality or fertility data in some countries. Hence, Belgium and the Netherlands are used for the mortality experiments only while Germany and Italy are selected for the fertility experiments. The remaining eight other countries are the same in both mortality and fertility experiments. We implement the four existing models mentioned above using the R package ‘*demography*’ with the instructions of Booth et al. [40] on how the parameters should be set in these models. Rolling-window analysis is used for assessing the consistency of the forecasting ability of a model by rolling a fixed size prediction interval (window) throughout the observed period [41]. We hold the sample data from the initial year up to the year t_m , where $t_m < t_n$, as holdout samples and produce the forecast for the t_{m+h} year where h is the forecast horizon. The forecasts errors are then determined by comparing the out-of-sample forecast results with the actual data. We increase one rolling-window (1 year ahead) in year t_{m+1} to make the same procedure again for the year t_{m+1+h} until the rolling-window analysis covers all available data in year t_n . We consider four different forecast horizons ($h = 5, 10, 15$ and 20) with ten sets of rolling-window to examine the short-term, the mid-term, and the

long-term forecast abilities of the models. In the mortality rolling-window experiments, the RMSE formula is

$$\text{RMSE}_c(h) = \sqrt{\frac{1}{10 \times 101} \sum_{w=0}^9 \sum_{x=0}^{100} \left(y_{x,t_m+w+h} - \hat{y}_{x,t_m+w+h} \right)^2},$$

where c is the selected country, w is the index of the rolling-window sets and x is the age from age 0 to age 100. And the RMSE in the fertility rolling-window experiments is

$$\text{RMSE}_c(h) = \sqrt{\frac{1}{10 \times 31} \sum_{w=0}^9 \sum_{x=15}^{45} \left(y_{x,t_m+w+h} - \hat{y}_{x,t_m+w+h} \right)^2},$$

where x is the age from age 15 to age 45.

Table 1 presents the average RMSE results of ten sets of rolling-window analysis across ten countries in four different forecast horizons for the mortality experiments. The proposed GPR model performs consistently the most desirable for mortality forecasting. It occupies the major positions of having the least forecast errors among the ten selected countries in the four different prediction horizons. The method is shown to be capable of capturing various mortality curve patterns across different periods and age groups. We can see that the GPR model has significant improvements in the prediction accuracy in comparison with the four other tested models, particularly for the mortality data of Japan. It may mainly be thanks to the intensive treatment of mortality curve fitting by the GPR approach, which enhances the forecast accuracy and robustness, especially for countries with high mortality fluctuations in certain age groups due possibly to natural disasters. It is also worth noting that the forecasting performances of the LM model in the short term and the mid-short term are remarkably well with the UK and the USA male mortality data. The consistency and small variabilities of mortality curves in the UK and the USA over age and time may contribute to the superiority of the LM model compared to the other tested models in our experiments. The forecast performances of the LC model, the BMS model and the HU model are reasonably similar with no particularly outstanding point in our experiments.

Table 2 shows the average RMSE results of ten sets of rolling-window analysis across the selected ten countries in the four different prediction horizons in the fertility experiments. The proposed GPR model continues to maintain the dominant positions of having the smallest forecast errors in the short term and the mid-short term forecast horizons and is comparable with the HU model in the fifteen-year forecast horizon. As for the long term prediction of the twenty-year forecast horizon, it appears that the HU model is more suitable to capture the fertility patterns with smaller forecasts errors than the GPR model. The HU model also fits the French fertility data better than the other models. It may be because the fertility data contain more outliers or measurement errors than the mortality data, and the HU model includes the smoothing techniques, which can improve the model forecasting performances if the observed fertility data are smoothed in advance.

Table 1. The average RMSEs of ten sets of rolling-window analysis on the predicted male mortality curves using the LC model, the LM model, the BMS model, the HU model and the GPR model. The lowest forecast errors among all the models are highlighted.

Country	LC Model	LM Model	BMS Model	HU Model	GPR Model
<i>h</i> = 5					
Austria	0.3122	0.3191	0.2596	0.2591	0.2581
Belgium	0.2560	0.2783	0.2429	0.2458	0.2359
Canada	0.1850	0.1574	0.1762	0.1519	0.1583
France	0.1879	0.1260	0.1365	0.1477	0.1457
Japan	0.1903	0.1365	0.1564	0.1533	0.1164
Netherlands	0.2679	0.2284	0.2046	0.2265	0.2233
Sweden	0.2873	0.3178	0.2695	0.2633	0.2566
Switzerland	0.3346	0.3597	0.3077	0.3109	0.2774
UK	0.1805	0.1301	0.1382	0.1584	0.1504
USA	0.1266	0.0875	0.1324	0.1183	0.1258
Average	0.2328	0.2141	0.2024	0.2035	0.1948
<i>h</i> = 10					
Austria	0.3504	0.3289	0.2793	0.2809	0.2783
Belgium	0.2902	0.2945	0.2820	0.3044	0.2634
Canada	0.2197	0.1837	0.2078	0.1862	0.2017
France	0.2443	0.1871	0.2074	0.2391	0.1985
Japan	0.2630	0.2187	0.2824	0.2382	0.1208
Netherlands	0.3294	0.2706	0.2629	0.2861	0.2678
Sweden	0.3153	0.3311	0.2869	0.2813	0.2740
Switzerland	0.3888	0.4087	0.3942	0.4030	0.3842
UK	0.2271	0.1730	0.1987	0.2133	0.1899
USA	0.1514	0.1231	0.1706	0.1474	0.1724
Average	0.2780	0.2519	0.2572	0.2580	0.2351
<i>h</i> = 15					
Austria	0.4022	0.3707	0.3119	0.3582	0.3130
Belgium	0.3251	0.3148	0.3050	0.3558	0.2906
Canada	0.2659	0.2413	0.2340	0.2220	0.2595
France	0.2992	0.2699	0.2937	0.3431	0.2924
Japan	0.3631	0.3366	0.3551	0.2688	0.2063
Netherlands	0.3787	0.3316	0.3145	0.3335	0.2992
Sweden	0.3677	0.3637	0.3194	0.3331	0.2986
Switzerland	0.4574	0.4724	0.5270	0.5347	0.5279
UK	0.2836	0.2292	0.2373	0.2604	0.2310
USA	0.1855	0.1825	0.2109	0.1798	0.2172
Average	0.3328	0.3113	0.3109	0.3189	0.2936
<i>h</i> = 20					
Austria	0.4600	0.4289	0.3833	0.4275	0.3813
Belgium	0.3666	0.3402	0.3330	0.3740	0.3200
Canada	0.3113	0.2747	0.2394	0.2594	0.2613
France	0.3491	0.3218	0.3236	0.4051	0.3105
Japan	0.5193	0.5120	0.4385	0.3333	0.2828
Netherlands	0.4218	0.3828	0.3968	0.3739	0.3321
Sweden	0.4214	0.3943	0.3794	0.4154	0.3270
Switzerland	0.4910	0.5259	0.5930	0.6774	0.6169
UK	0.3529	0.2894	0.2750	0.3774	0.2717
USA	0.2017	0.2089	0.1856	0.2159	0.2122
Average	0.3895	0.3679	0.3548	0.3859	0.3316

Table 2. The average RMSEs of ten sets of rolling-window analysis on the predicted fertility curves using the LC model, the LM model, the BMS model, the HU model and the GPR model. The lowest forecast errors among all the models are highlighted.

Country	LC Model	LM Model	BMS Model	HU Model	GPR Model
<i>h</i> = 5					
Austria	0.6112	0.2509	0.6203	0.2287	0.1801
Canada	0.6044	0.2481	0.1330	0.3236	0.2432
France	0.4927	0.1565	0.2334	0.0981	0.2462
Germany	0.6495	0.2420	0.5870	0.1608	0.1489
Italy	0.5414	0.2635	0.3719	0.4402	0.2805
Japan	0.6533	0.3876	0.6721	0.5302	0.2769
Sweden	0.5970	0.1905	0.1562	0.2768	0.1553
Switzerland	0.6734	0.2822	0.4553	0.2536	0.2059
UK	0.4280	0.2392	0.2554	0.2970	0.2920
USA	0.4499	0.1757	0.2170	0.3554	0.2431
Average	0.5701	0.2436	0.3702	0.2964	0.2272
<i>h</i> = 10					
Austria	0.7328	0.5081	0.7915	0.3370	0.3276
Canada	0.7476	0.5054	0.2581	0.3630	0.2881
France	0.6026	0.3428	0.4888	0.2043	0.3742
Germany	0.7531	0.4990	0.7365	0.2909	0.2008
Italy	0.7064	0.5449	0.8189	0.5346	0.4802
Japan	0.8998	0.7846	1.2119	0.6443	0.3424
Sweden	0.7524	0.4169	0.2699	0.3857	0.1789
Switzerland	0.7926	0.5536	0.4835	0.4489	0.3274
UK	0.5928	0.4780	0.3448	0.3594	0.3294
USA	0.5058	0.3172	0.3092	0.4458	0.4241
Average	0.7086	0.4950	0.5713	0.4014	0.3273
<i>h</i> = 15					
Austria	0.8863	0.7721	0.9997	0.4972	0.6215
Canada	0.9300	0.8080	0.5972	0.4347	0.4894
France	0.7871	0.5908	0.8119	0.4960	0.5973
Germany	0.9231	0.7684	0.9732	0.4319	0.2961
Italy	0.9339	0.8319	1.3051	0.7208	0.9365
Japan	1.2392	1.2202	1.5252	0.6989	0.5641
Sweden	0.8930	0.6031	0.5837	0.5307	0.3582
Switzerland	0.9837	0.8390	0.9742	0.6711	0.6488
UK	0.7850	0.7286	0.4473	0.5277	0.4505
USA	0.6125	0.5309	0.4280	0.5312	0.5539
Average	0.8986	0.7690	0.8495	0.5603	0.5439
<i>h</i> = 20					
Austria	1.0445	1.0165	1.2876	0.6219	1.0984
Canada	1.2030	1.1651	1.2181	0.5261	0.9405
France	0.9948	0.8514	1.1597	0.8039	0.9926
Germany	1.1324	1.0523	1.3186	0.6835	0.6710
Italy	1.1945	1.1354	1.6735	0.9316	1.6623
Japan	1.6871	1.7571	1.6005	0.7553	0.8688
Sweden	1.0161	0.7615	0.9025	0.6986	0.6580
Switzerland	1.2332	1.1462	1.4133	0.7295	1.2452
UK	1.0005	0.9784	0.7820	0.8654	0.6823
USA	0.7946	0.7948	0.7017	0.6077	0.6676
Average	1.1301	1.0659	1.2057	0.7224	0.9487

5. Discussion and Conclusion Remarks

Following the theoretical framework of Gaussian process regression discussed in Section 2, in this paper, we have introduced a new design of the Gaussian process regression model equipped with the natural cubic spline mean function and the spectral mixture covariance function as a new approach for mortality and fertility modelling and forecasting. The use of the natural cubic spline mean function in the proposed GPR model can exploit

the local information of the recent data and force its extrapolation beyond the observed data as a linear function to provide smooth and robust forecasts. The spectral mixture covariance function, on the other hand, detects and handles any non-linearity and periodicity of the demographic data unexplained by the fitted mean function.

We have demonstrated the usefulness and flexibility of the proposed GPR model through two empirical data applications: one is to forecast the male mortality data, and the other is to predict the fertility data. Our experiments have proved the forecasting ability of the proposed GPR model. In these two experiments, the accuracies of the predicted mortality and fertility curves are improved by the proposed GPR model in terms of the average forecast errors for the ten tested countries in the different forecasting horizons when compared to the other four mainstream approaches. The prediction performances of these four methods rely barely on the extrapolation of the time parameters as they determine how the movement of the whole mortality (or fertility) curve should shift over time across all age groups. The entire fitted curve can go far away from its expected location when the fitted time parameters in these four methods are not well predicted. This problem can be seen in our demonstration for the predicted Japanese fertility curve in Figure 8 as the forecast results of the fertility curves by the mainstream approaches, such as the LC model and the BMS model, deviate from the observation noticeably in some parts of the curve. In contrast, the proposed method is more robust to this issue, because it treats each age group as time-series data and assumes each of them following a Gaussian process to achieve the same tasks in a discrete but very intensive fashion to avoid this issue.

Furthermore, the proposed GPR model is in time series, non-linear and non-parametric structures and is not restricted to mortality and fertility modelling and forecasting only. It can be used for other factors that impact population forecasting, such as migration flows, provided that the age-specific data are available. It can also support a wider range of potential applications in many domains of applied science and engineering, such as signal processing or weather forecasting. Its particular features enable more flexibility than the four existing models considered in our study. Although Wu and Wang [21] proposed a similar approach using Gaussian process regression with a weighted linear mean function, the weighted linear mean structure relies on the assumption that more recent data tend to have more weights on the future mortality, which requires on subjective judgements on how the weights are assigned to reflect the impacts of different periods. It can affect the accuracy of forecasts remarkably when different subjective choices of weights are adopted, or the weight parameters are inappropriately specified [21]. In contrast, our method exploits the natural cubic spline function as a non-parametric mean structure without any assumption made on the relationship between the mortality rates and the time, and it lets the natural cubic spline mean function automatically identify the local information from the observed data and then project the future mortality in an appropriate direction.

The main limitations of the proposed GPR models also relate to the characteristics of the classes of independent time series and simple non-parametric single population extrapolative methods of which it belongs to. Although it can capture the trends of the historical demographic data robustly, at the same time, it lacks the ability to incorporate and model other related information, such as the changes in medical technology, environment and social-economy. One may also expect that the mortality or fertility rates for different ages may be closely correlated, especially among the neighbouring ages. Our GPR model would be more desirable if it could model the dependence on different age groups simultaneously while taking their heterogeneity into account. We, therefore, aim to extend the current GPR model to a spatial or a multi-output GPR model for connecting the relationships between different age groups and time periods altogether. Another limitation is that the models do not incorporate the natural upper limits of mortality and fertility rates explicitly—the forecast is based on the historical data only, although the lower limits are removed by considering the logarithm of the rates. It is also noted that our models do not take account of the cohort effect which refers to the phenomenon that the mortality experience of the

same cohort group of individuals who were born in the same calendar year is highly correlated. All of these will be left for our future works to achieve.

Author Contributions: Conceptualization, K.K.L. and B.W.; methodology, K.K.L.; software, K.K.L.; writing—original draft preparation, K.K.L.; writing—review and editing, K.K.L. and B.W.; supervision, B.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: The authors thank the editor and the reviewers for their very helpful and constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Preston, S.; Heuveline, P.; Guillot, M. *Demography: Measuring and Modeling Population Processes*; Blackwell Publishers: Malden, MA, USA, 2000.
2. Gompertz, B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos. Trans. R. Soc. Lond.* **1825**, *115*, 513–583.
3. Makeham, W.M. On the law of mortality and the construction of annuity tables. *J. Inst. Actuar.* **1860**, *8*, 301–310. [[CrossRef](#)]
4. Lee, R.D.; Carter, L.R. Modeling and forecasting US mortality. *J. Am. Stat. Assoc.* **1992**, *87*, 659–671.
5. Lee, R.D. Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. *Int. J. Forecast.* **1993**, *9*, 187–202. [[CrossRef](#)]
6. Bell, W.R. Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *J. Off. Stat.* **1997**, *13*, 279–303.
7. Lee, R.; Miller, T. Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography* **2001**, *38*, 537–549. [[CrossRef](#)]
8. Booth, H.; Maindonald, J.; Smith, L. Applying Lee-Carter under conditions of variable mortality decline. *Popul. Stud.* **2002**, *56*, 325–336. [[CrossRef](#)] [[PubMed](#)]
9. Brouhns, N.; Denuit, M.; Vermunt, J.K. A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insur. Math. Econ.* **2002**, *31*, 373–393. [[CrossRef](#)]
10. Renshaw, A.; Haberman, S. Lee-Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2003**, *52*, 119–137. [[CrossRef](#)]
11. Cairns, A.J.; Blake, D.; Dowd, K. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *J. Risk Insur.* **2006**, *73*, 687–718. [[CrossRef](#)]
12. Hunt, A.; Blake, D. A general procedure for constructing mortality models. *N. Am. Actuar. J.* **2014**, *18*, 116–138. [[CrossRef](#)]
13. Whittaker, E.T. On a new method of graduation. *Proc. Edinb. Math. Soc.* **1922**, *41*, 63–75. [[CrossRef](#)]
14. Currie, I.D.; Durban, M.; Eilers, P.H. Smoothing and forecasting mortality rates. *Stat. Model.* **2004**, *4*, 279–298. [[CrossRef](#)]
15. Hyndman, R.J.; Ullah, M.S. Robust forecasting of mortality and fertility rates: A functional data approach. *Comput. Stat. Data Anal.* **2007**, *51*, 4942–4956. [[CrossRef](#)]
16. Delwarde, A.; Denuit, M.; Eilers, P. Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach. *Stat. Model.* **2007**, *7*, 29–48. [[CrossRef](#)]
17. Debón, A.; Martínez-Ruiz, F.; Montes, F. A geostatistical approach for dynamic life tables: The effect of mortality on remaining lifetime and annuities. *Insur. Math. Econ.* **2010**, *47*, 327–336. [[CrossRef](#)]
18. Li, H.; O'Hare, C.; Zhang, X. A semiparametric panel approach to mortality modeling. *Insur. Math. Econ.* **2015**, *61*, 264–270. [[CrossRef](#)]
19. Ludkovski, M.; Risk, J.; Zail, H. Gaussian process models for mortality rates and improvement factors. *ASTIN Bull. J. IAA* **2018**, *48*, 1307–1347. [[CrossRef](#)]
20. Dokumentov, A.; Hyndman, R.J.; Tickle, L. Bivariate smoothing of mortality surfaces with cohort and period ridges. *Stat* **2018**, *7*, e199. [[CrossRef](#)]
21. Wu, R.; Wang, B. Gaussian process regression method for forecasting of mortality rates. *Neurocomputing* **2018**, *316*, 232–239. [[CrossRef](#)]
22. Alexopoulos, A.; Dellaportas, P.; Forster, J.J. Bayesian forecasting of mortality rates by using latent Gaussian models. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2019**, *182*, 689–711. [[CrossRef](#)]
23. Krige, D.G. A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. S. Afr. Inst. Min. Metall.* **1951**, *52*, 119–139.

24. Matheron, G. Principles of Geostatistics. *Econ. Geol.* **1963**, *58*, 1246–1266. [[CrossRef](#)]
25. Journel, A.G.; Huijbregts, C.J. *Mining Geostatistics*; Academic Press: Cambridge, MA, USA, 1978.
26. Cressie, N. Geostatistics. *Am. Stat.* **1989**, *43*, 197–202.
27. Engel, Y.; Mannor, S.; Meir, R. Reinforcement learning with Gaussian processes. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 201–208.
28. Krause, A.; Singh, A.; Guestrin, C. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *J. Mach. Learn. Res.* **2008**, *9*, 235–284.
29. Neal, R.M. *Bayesian Learning for Neural Networks*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 118.
30. Alamaniotis, M.; Ikononopoulos, A.; Tsoukalas, L.H. Evolutionary multiobjective optimization of kernel-based very-short-term load forecasting. *IEEE Trans. Power Syst.* **2012**, *27*, 1477–1484. [[CrossRef](#)]
31. Wu, Q.; Law, R.; Xu, X. A sparse Gaussian process regression model for tourism demand forecasting in Hong Kong. *Expert Syst. Appl.* **2012**, *39*, 4769–4774. [[CrossRef](#)]
32. MacKay, D.J. Introduction to Gaussian processes. *NATO ASI Ser. F Comput. Syst. Sci.* **1998**, *168*, 133–166.
33. Williams, C.K.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; Volume 2.
34. Friedman, J.; Hastie, T.; Tibshirani, R. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer: New York, NY, USA, 2001; Volume 1.
35. Wilson, A.; Adams, R. Gaussian process kernels for pattern discovery and extrapolation. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 1067–1075.
36. Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available online: <https://www.mortality.org/> (accessed on 21 March 2020).
37. Shi, J.; Wang, B.; Murray-Smith, R.; Titterton, D. Gaussian process functional regression modeling for batch data. *Biometrics* **2007**, *63*, 714–723. [[CrossRef](#)]
38. Human Fertility Database. University of California, Berkeley (USA), Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). Available online: <https://www.humanfertility.org> (accessed on 2 April 2020).
39. Booth, H.; Tickle, L. Mortality modelling and forecasting: A review of methods. *Ann. Actuar. Sci.* **2008**, *3*, 3–43. [[CrossRef](#)]
40. Booth, H.; Hyndman, R.J.; Tickle, L. Prospective life tables. In *Computational Actuarial Science with R*; Taylor & Francis: Boca Raton, FL, USA, 2014; pp. 319–344.
41. Zivot, E.; Wang, J. *Modeling Financial Time Series with S-Plus®*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2007; Volume 191.