Constrained Kriging for smoothing and forecasting mortality rates

Zied Chaieb
Laboratoire de recherche QUANTLABS
zied.chaieb@quantlabs.fr

Djibril Gueye
Laboratoire de recherche QUANTLABS
djibril.gueye@quantlabs.fr

Abstract

Mortality surface is a function of age and year with the main characteristic of being increasing in age direction from a given age. One of the major challenges of its construction is to take this last specificity into account. In this paper, we propose to use constrained Kriging for such construction. Our approach is based on the finite-dimensional approximation of the Gaussian process. We first show the ability of the constrained Kriging to construct mortality surfaces and then compare its performance against classical Kriging models with trend functions such as those used in [LRZ18]. Our empirical study based on mortality data from three countries (France, Italy, and Germany) showed the need to add a constraint of convexity in age direction and illustrated through an RMSE criterion that the constrained Kriging provided better results in terms of out-of-sample forecasting.

1 Introduction

Mortality surface is at the heart of risk pricing in life insurance. It allows insurers to evaluate the technical premiums of life insurance contracts (savings, retirement, individual and collective pension arrangements) and to know their commitments towards the insured in the Solvency 2 regulatory context. For a life insurer, being able to predict the distortion of the risk of death shortly remains a central issue. This risk is evaluated based on mortality tables. The latter represents, as a function of age and calendar year, the empirical death rates estimated over annual periods. The construction of these surfaces is a classic and fundamental subject in life insurance. The classical methods essentially revolve around those of Lee and Carter [LC92] and its extensions (e.g., [BDV02], [Cur13])..

Machine learning methods, in particular those of classical Kriging, have recently been used in this context. Kriging is a geostatistical method used for spatial interpolation by the regression of a Gaussian process (GP) ([WR06]). The pertinence and efficiency of Kriging for estimating mortality rates have been shown in [WW18] and in [LRZ18], especially through its ability to quantify the uncertainty on the variable of interest.

However, in [LRZ18], the authors did not take into account some specificities that must be respected for the reconstruction and the projection of these surfaces, in particular the monotony on the ages. Indeed it is admitted that from a certain age, the death rates are increasing in the direction of the ages. In addition to this shape constraint, there is high variability in the death rates as we reach the highest ages due to the low number of observations in these areas. This is why it becomes important to be able to project the mortality surface despite the small set of data used.

It becomes then necessary to provide a technique that smoothes out the mortality surface and takes into account their specificities. To respect the shape constraint on the surface, the usual techniques are essentially based on constrained *splines* ([Woo94; Cam+12]). However, these techniques, being deterministic, do not quantify the uncertainty of the variable to be estimated.

In this work, we propose a constrained Kriging method which not only makes it possible to smooth the surface by taking the shape constraint on the ages into account but also makes it possible to quantify the uncertainty on the variable of interest by constructing confidence intervals on areas of low observations. Given this shape constraint which will be interpreted as a finite number of linear inequalities, the method consists in finding the posterior surface as being a Gaussian process truncated from a prior Gaussian process Y.

Indeed, conditionally to this form constraint, the posterior process is no longer Gaussian, unlike the classical Kriging case. We, therefore, adopt the finite-dimensional approximation approach $Y^N(.) = \sum_j^N \phi_j(.)\xi_j$ de Y ([CMR16]), where $\phi_1, ..., \phi_N$ are triangular-type based functions and $\xi_1, ..., \xi_N$ are Gaussian random variables.

The advantage of this approximation comes from the fact that the estimation of the posterior process is based essentially on the estimation of the Gaussian variables ξ_j truncated on a convex domain of linear inequalities. The estimation of the most probable coefficient vector will be made through a quadratic optimization criterion of the empirically observed data from the prospective mortality tables. We then use the Hamiltonian Monte Carlo method (HMM) to estimate the posterior mortality surfaces.

Our approach meets the limitations of the classical Kriging method used and mentioned in [Lud18]. The authors did not take into account the shape constraint on the ages and adjusted the GP on their whole data set. This method can have a good performance on the *in-sample RMSE* of the GP but cannot guarantee sufficient performance on the *out-of-sample RMSE* of the latter one.

To overcome this, we used a low-dimensional randomized sampling from the initial data set. The rest is considered as test data (or even missing data) on the learning grid and can be used later in the backtesting process. This method gives better projections since if we look at the angle of [TP15], the forecast data can be considered as missing data.

This work is structured as follows: first, we recall the construction process of the mortality surface from prospective mortality tables. Then we describe the framework of classical Kriging and the transition to constrained Kriging. Finally, we illustrate our approach through numerical simulations on data of three countries as France, Italy, and Germany.

2 Problem and methodology

Kriging has been widely used in the field of actuarial sciences. Among others applications in this domain, it has been employed for learning mortality rate and mortality improvement (see [LRZ18], [WW18]). To the best of our knowledge, it has never been used for imposing age-monotonicity constraints for the mortality surface.

Adding inequality constraint to the Kriging method (what is called contained Kriging) is not an easy task but allows to solve a wide range of problems in actuarial sciences.

Constrained Kriging has first been adapted in quantitative finance by Cousin, Maatouk, and Rullière [CMR16] who showed the extension of classical spline interpolation by constrained Kriging techniques developed in [MB14] to ensure non-arbitrable and error-controlled yield-curve and CDS curve interpolation. This technique is then applied for imposing non-arbitrage conditions in volatility surface construction (see [Cha+21]).

2.1 Kriging concept and mortality modeling

We consider the problem of learning an unknown mapping function f from observations of input and output couples $((\mathbf{x}_i, \mathbf{z}_i), i = 1, \ldots, n)$ that represents the evolution of a reference quantity $\mathbf{z}$ as a function of some selected factors or explanatory variables $\mathbf{x}$ with some possible noises:

$$\mathbf{z} = f(\mathbf{x}) + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{\varepsilon}$ is considered to be a zero-mean Gaussian vector with homoscedasticity variance ς^2, Kriging consists in making a Gaussian prior Y on f, with mean function $\mu(\mathbf{x})$ and covariance function γ. Here, we consider the d-dimensional isotropic covariance kernel given, for any $\mathbf{x} = (x^1, \ldots, x^d)$ and $\mathbf{x}' = (x'^1, \ldots, x'^d)$ as

$$\gamma(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{i=1}^{d} R_i(x^i - x'^i, \theta_{x^i})$$

where $\theta = (\theta_{x^1}, \ldots, \theta_{x^d}) \in \mathbb{R}^d$ and σ are respectively the length scale and the variance hyperparameters of the kernel function γ and the functions $(R_i)_i$ are kernel correlation functions. These hyperparameters together with the noise variance ς^2 can be estimated using the Maximum Likelihood estimator (MLE).

In this setting, the Best Linear Unbiased Estimator (BLUE) of f is

$$\eta(\mathbf{x}) = \mu(\mathbf{x}) + \mathbf{c}(\mathbf{x})^\top (C + \varsigma^2 I_n)^{-1} (z - \mu(\mathbf{x})) \tag{2}$$

with $\mathbf{c}(\mathbf{x}) = [\gamma(\mathbf{x}, (\mathbf{x}_1)), \ldots, \gamma(\mathbf{x}, (\mathbf{x}_n))]^\top$ and the conditional covariance function

$$\gamma^\star(\mathbf{x}, \mathbf{x}') = \gamma(\mathbf{x}, \mathbf{x}') - \mathbf{c}(\mathbf{x})^\top (C + \varsigma^2 I_n)^{-1} \mathbf{c}(\mathbf{x}') \tag{3}$$

can then be used to obtain confidence bounds around the predicted function f. The previous result holds from the fact that conditionally to (1), the posterior process is still Gaussian with mean function (2) and covariance function (3). This technique is known as classical Kriging.

In the framework of mortality surface construction, the unknown function f represents the log of mortality rate and the inputs $\mathbf{x}$ are the pair (age, year). In most cases, a functional form of the trend function $\mu(\mathbf{x})$ is chosen to get better results in terms of forecasting. In [Lud18] several forms of $\mu(\mathbf{x})$ have been tested, Wu and Wang [WW18] proposed to use a linear trend in the year direction for a fixed age because of capturing the decreasing effect of mortality rate function with respect to the years for most ages.
Classical Kriging is shown to be a suitable tool for mortality surface construction. However, when monotonicity in the age direction is considered, this method is not robust (see [CG21]). Therefore, in the following subsection, we are going to discuss how to handle monotonicity in the age direction for mortality construction using Kriging.

2.2 Imposing non-decreasing effect in the age direction using Kriging

Mortality data set $D = (\mathbf{x}, \mathbf{z})$ consists in a set of n inputs $\mathbf{x} = ((x_1, t_1), \ldots, (x_n, t_n))$ and n outputs $\mathbf{z} = (z_1, \ldots, z_n)$, where the $(x_i)_i$ represent the ages, $(t_i)_i$ designate the years and $(z_i)_i$ are the log of mortality rates.
Given a mortality data D, we consider the following model:

$$z_i = m(t_i) + f(x_i, t_i) + \varepsilon_i \tag{4}$$

where m is a trend function which is assumed to be linear in the year direction t and constant in the age direction x and ε_i is a Gaussian random variable with variance ς_i^2. We assume that f is a realization of a zero-mean Gaussian process Y with covariance function γ. It is clear that the response variable is non-decreasing in the age direction x if and only if the function f is non-decreasing in x. Therefore, in order

to impose the monotonicity in x, we consider the finite-dimensional approximation Y^N of the process Y. Dealing with this approach is equivalent to dealing with the Gaussian model in [CG21] with the fact that the observed outputs are considered to be $z_i - m(t_i)$, for $i = 1, \ldots, n$. The parameters of the linear function m together with the hyperparameters and the variance of the noise will be estimated using MLE.
In this contrast, we aim at estimating Y given (4) and conditionally to the fact that Y is non-decreasing in the age direction x.
For this purpose, we consider a discretized version of the input domain $\Omega = [x_1, x_2] \times [t_1, t_2]$ with a $N = (N_t + 1) \times (N_x + 1)$ regular grid containing the set of knots $(u_i, v_j) \in \Omega$, $i = 1, \ldots, N_t$, $j = 1, \ldots, N_x$ where $u_i = \frac{i}{N_t}$ and $v_j = \frac{j}{N_x}$. The finite-dimensional approximation of Y is defined by

$$Y^N(t,x) = \sum_{i=0}^{N_t} \sum_{j=0}^{N_x} \xi_{ij} \phi_{i,j}(t,x), \quad \text{for all } (t,x) \in \Omega, \tag{5}$$

where $\xi_{ij} = Y(u_i, v_j)$ and $(\phi_{i,j})_{i,j}$ are triangular (hat) basis functions defined as the following tensor product

$$\phi_{i,j}(t,x) := \max(1 - N_t|t - u_i|, 0) \max(1 - N_x|x - v_j|, 0) .$$

Note that the finite-dimensional process Y^N uniformly converges to Y on Ω as $N_X \to \infty$ and $N_T \to \infty$, almost surely and the vector $\boldsymbol{\xi} = [\xi_{0,0}, \ldots, \xi_{i,j} \ldots, \xi_{N_T,N_X}]^\top$ is a zero-mean Gaussian vector with $N \times N$ covariance matrix Γ^N with (ι_1, ι_2)-th entry (corresponding to (i_1, j_1) and (i_2, j_2) respectively) given by $\Gamma^N_{\iota_1,\iota_2} = \gamma((u_{i_1}, v_{j_1}), (u_{i_2}, v_{j_2}))$ (see [Maa17]).
The main advantage of this approximation lies in the fact that incorporating some inequality constraints on Y^N is equivalent to handling these constraints in the vector of Gaussian coefficient $\boldsymbol{\phi}(t,x)$ which corresponds to a finite number of inequality constraints. Hence the process Y^N is non-decreasing in the age direction x if and only if $\xi_{i,j} \le \xi_{i,j+1}, \forall (i,j)$.
Therefore estimating Y^N given (4) and conditionally to Y^N non-decreasing is equivalent to estimating ξ conditionally to

$$\begin{cases} z_i = m(t_i) + \boldsymbol{\phi}(t_i, x_i) \cdot \xi + \varepsilon_i, \text{ for } i = 1, \ldots, n \\ \xi_{i,j} \le \xi_{i,j+1}, \forall (i,j), \end{cases} \tag{6}$$

where $\boldsymbol{\phi}(t_i, x_i) = [\phi_{0,0}(t,x), \ldots, \phi_{k,l}(t,x), \ldots, \phi_{N_T,N_X}(t,x)]$.

In a matrix reformulation, our problem consists in sampling $\boldsymbol{\xi}$ given

$$\begin{cases} \boldsymbol{z} = \mathbf{m}(\mathbf{t}) + \boldsymbol{\Phi}(\mathbf{t},\mathbf{x}) \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} \\ A \cdot \boldsymbol{\xi} \ge 0 \end{cases} \tag{7}$$

where we set $\mathbf{m}(\mathbf{t}) = [m(t_1), \ldots, m(t_n)]^T$ and we denote by $\boldsymbol{\Phi}(\mathbf{t},\mathbf{x})$ the $n \times N$ matrix of basis function in which, each row l corresponds to the vector $\boldsymbol{\phi}(t_l, x_l)$. The matrix A characterizes the linear inequality constraints on $\boldsymbol{\xi}$ and is defined as

$$A = \begin{bmatrix} 1 & -1 & 0 & \ldots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & 1 & -1 \end{bmatrix}_{N \times N}$$

The conditional distribution of $\boldsymbol{\xi}$ given $\boldsymbol{z} = \mathbf{m}(\mathbf{t}) + \boldsymbol{\Phi}(\mathbf{t},\mathbf{x}) \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon}$ is Gaussian with mean function and covariance function

$$\boldsymbol{\eta}_{\mathbf{z}}(\mathbf{t},\mathbf{x}) = \Gamma^N \boldsymbol{\Phi}(\mathbf{t},\mathbf{x})^\top (\boldsymbol{\Phi}(\mathbf{t},\mathbf{x}) \Gamma^N \boldsymbol{\Phi}(\mathbf{t},\mathbf{x})^\top + \varsigma^2 I_n)^{-1} (\mathbf{z} - \mathbf{m}(\mathbf{t})) \tag{8}$$

$$\gamma_{\mathbf{z}}(\mathbf{t},\mathbf{x}) = \Gamma^N \boldsymbol{\Phi}(\mathbf{t},\mathbf{x})^\top \left(\boldsymbol{\Phi}(\mathbf{t},\mathbf{x}) \Gamma^N \boldsymbol{\Phi}(\mathbf{t},\mathbf{x})^\top + \varsigma^2 I_n \right)^{-1} \boldsymbol{\Phi}(\mathbf{t},\mathbf{x}) \Gamma^N . \tag{9}$$

Hence, the simulation of the Gaussian coefficient vector ξ will be done using Hamiltonian Monte Carlo (HMC) method (see [PP14]). Since in the exact HMC sampling method the initial vector must satisfy the inequality constrained, a wise choice of this vector could be the Maximum a Posterior (MAP) of ξ (as it is done in [LL+18; Cha+21]). We recall that the MAP $\widehat{\mathbf{c}}$ of ξ is a solution to the following quadratic problem:

$$\widehat{\mathbf{c}} = \underset{\mathbf{m}(\mathbf{t})+\mathbf{\Phi}(\mathbf{t},\mathbf{x})\cdot c+\varepsilon=z,\, A\cdot c\geq 0}{\arg\min} \left(c^\top (\Gamma^N)^{-1} c\right) \tag{10}$$

and satisfies the linear inequality constraint in the whole domain. Once we compute the MAP $\widehat{\mathbf{c}}$ of ξ, the one $\widehat{Y}^N$ of Y^N can be derived as $\widehat{Y}^N(\mathbf{t},\mathbf{x}) = \mathbf{\Phi}(\mathbf{t},\mathbf{x})\cdot\widehat{\mathbf{c}}$.
In this contrast, the most probable response mortality surface is then given by:

$$\exp\left(\mathbf{m}(\mathbf{t}) + \widehat{Y}^N(\mathbf{t},\mathbf{x})\right). \tag{11}$$

2.3 Main steps of our methodology

We aim to generate a set of sampled mortality surfaces compatible with some input noisy mortality rates for different pairs (year, age) and which respect the monotonicity (non-decreasing) constraint in the age direction. We make a Gaussian prior Y in the log of mortality rate and use the finite-dimensional approximation (Y^N) of this original Gaussian process prior Y to handle monotonicity constraints on the entire domain. The prior process is then characterized by the Gaussian vector ξ with the same number of elements that knots in the basis function grid.
The construction is as follows:

1. Import input observed mortality data,
2. Construct the grid of basis functions non necessarily a rectangular grid (i.e., with constant step for each direction):
   a. Define the input domain Ω on which we want to study the phenomenon,
   b. Construct the grid on that domain.
3. Construct the constraint matrices A (matrices that characterize the inequality constraints on the finite-dimensional approximation of the Gaussian process Y^N),
4. Construct the equality constraint matrix Φ,
5. Maximize the log marginal likelihood using multi-start or global optimization starting from a range of admissible parameters,
6. Construct the maximum a posterior (MAP) $\widehat{Y}^N$ for Y^N,
7. Sample the desired number of paths of the truncated Gaussian process using the exact HMC method,
8. Compute the $[0.05 - 0.95]$ quantiles for mortality surfaces.

3 Numerical illustrations

The main aim of our empirical study is to examine the performance of Kriging for smoothing and forecasting mortality surfaces that respect the monotonicity constraint in the age direction. In particular, we first show that constrained Kriging is a useful tool for constrained mortality surface construction. Then we compare our model to the one studied in [LRZ18] who used classical Kriging for constructing mortality surface by testing several trend functions of the GP.

Results are illustrated on the mortality data of three countries as Germany, France, and Italy. Here, we only show the ones obtained with french male data.

Our Data come from the Human Mortality Database via the R package demography ([Tic+12]) and include mortality rates of males which covers the period 2000-2015 for individuals aged between 45 and 90. Figure 1 shows the 2-d scatter plot of french male mortality data where the training points are in blue and the test ones in red. We observe the increasing effect in the age direction for the observed mortality rates. In addition, the mortality rate seems to be convex in the age direction. This last may require adding a convexity constraint in that direction regarding the best fit of the surface as well as better performances for projection in age direction.

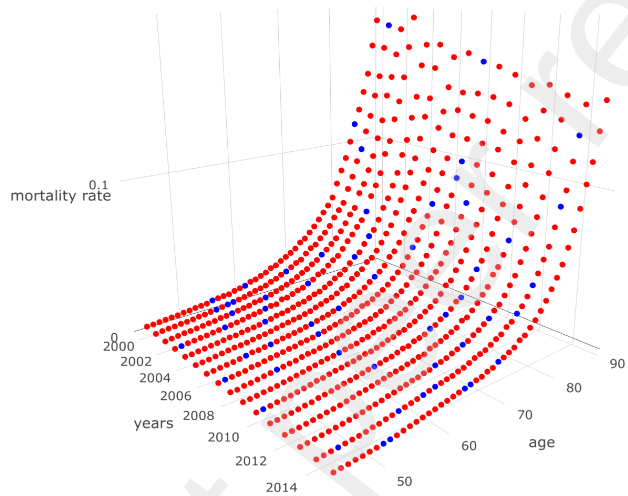

Figure 1: *Observed mortality rates.*

We first make a log transformation of the mortality rate for the sake of stabilizing the noise variance. Then we use only 10 percent of the data to construct our training set, the rest constitutes the test and validation set. To show that constrained Kriging is adapted for this construction, we fit the Gaussian process Y without the linear trend in our training set by using a Matèrn 5/2 kernel. The parameters of the model are estimated using MLE implemented in the DiceKriging R package (Roustant, Ginsbourger, and Deville [RGD12]). These parameters are such that the length scale $\widehat{\theta}_x = 46.2858$ in the age direction, the length scale $\widehat{\theta}_t = 9.8961$ in the year direction, the variance hyperparameter $\widehat{\sigma}^2 = 4.388$ and the noise variance $\widehat{\varsigma} = 0.000451$. Note that the noise variance is too small and together with the fact that the observed mortality rates respect the monotonicity constraint, we expect a narrow confidence interval.

To construct the finite-dimensional approximation Y^N of the Gaussian process Y, we use 40 basis functions in the year direction and 30 basis functions in the age direction which correspond to the total of $N = 40 \times 30$ grids. The R function "quadprog" in the package "pracma" is used for solving the quadratic problem (10).

We generate 500 sample surfaces of the constrained Gaussian process using the Hamiltonian Monte Carlo algorithm. Figure 2 shows the MAP of the constrained GP and one of the observation noises which represents the most probable distance between the observed mortality rate and the monotonicity constrained Gaussian field with the specified kernel (i.e, matèrn 5/2).

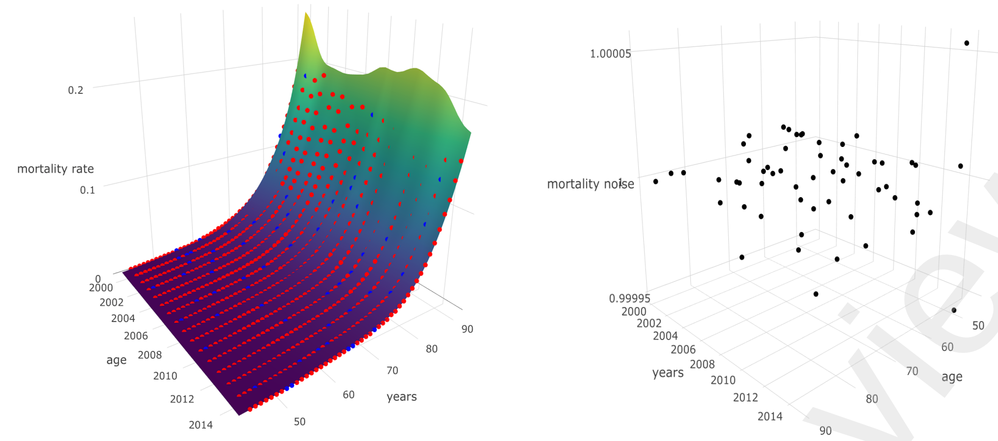

Figure 2: Most probable surface (left) vs most probable noise values (right).

Using quantiles instead of the full support of the simulations for GP uncertainty quantification, our MAP would lie outside the uncertainty bounds. Figure 3 shows some different slices in the age direction of the most likely surface (i.e., the MAP) together with the ones of the 95% credible bounds. These surfaces verify the non-decreasing effect in the age direction. Note from the year 2000 that the posterior distribution seems to be skewed for some age in between 75 and 80 since the GP MAP is close to the lower boundary of the posterior paths sampling. In addition, the in-sample RMSE (i.e., 1.40737×10^{-05}) and the out-sample RMSE (i.e., 4.713633×10^{-02}) show good performances for the fitting and validation of the GP. This shows that Constrained Kriging obtains good training and testing RMSE and quantifies the uncertainty interestingly. These results are almost similar for all countries (see appendix section 4 for more plots).

Comparison with classical Kriging adding a trend function: To test the projection performance of our model, we compare it with the models developed in [LRZ18] where authors add different trend functions. In this paper, we consider the two following trends for the classical Kriging:

- **Linear Trend Model**: $m(t,x) := \alpha_0 + \alpha_1^x\, x + \alpha_1^t\, t, \forall (t,x)$
- **Quadratic Trend Model**: $m(t,x) := \alpha_0 + \alpha_1^x\, x + \alpha_2^x\, x^2 + \alpha_1^t\, t, \forall (t,x)$.

Concerning the constrained Kriging model we use a linear trend in year t, i.e.,

$$m(t,x) := \alpha_0 + \alpha_1^t\, t.$$

We proceed as follows: fitting the three models in the domain [AGE_min_train, AGE_max_train] × [YEAR_min_train, YEAR_max_train] and evaluating their accuracies in [AGE_max_train + 1, AGE_end_data] × [YEAR_min_train, YEAR_max_train] by computing the Root Mean Square Error (RMSE) between the observed outputs and the predicted ones.

Results: We present an example where AGE_min_train = 45, AGE_max_train = 65, YEAR_min_train = 2001, YEAR_max_train = 2012.

Fig 4 shows the estimated MAP from the three different models. In both cases, we fit the GP with a matèrn 5/2 kernel function. The hyperparameters together with the trend parameters estimated using MLE are summarized in table 1.

We generally observe a good fit between the three models for ages in the training sample set. However, The estimated paths of these models start at age 65 (which separates the two data sets) to deviate from the

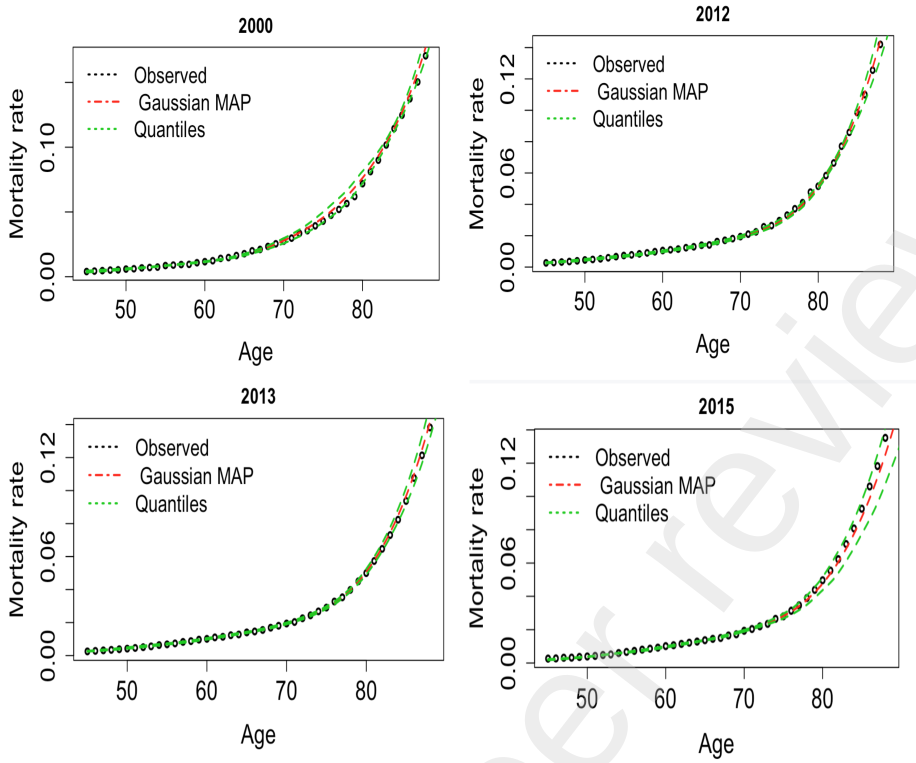

Figure 3: Different slices of the constrained GP with respect to age direction.

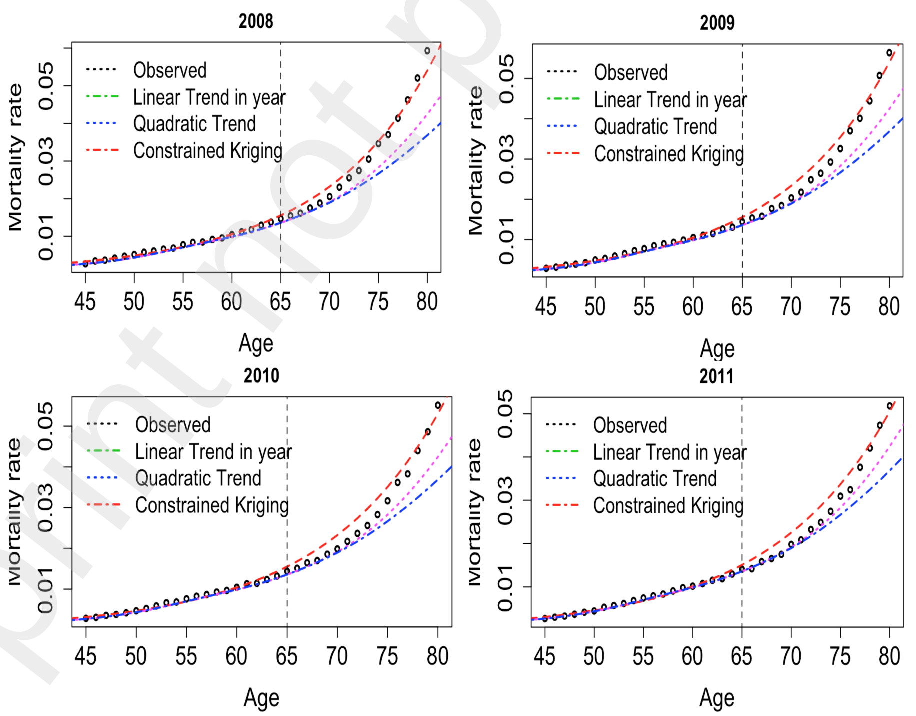

Figure 4: Slices of the MAP estimate of the three models.

	Trend parameters				Covariance matrix parameters			
	α_0	α_1^x	α_2^x	α_1^t	θ_x	θ_t	σ	ς
CK	90.63	-	-	-0.04	26.85	9.67	1.003161	0.0005030339
LT	32.23	0.08	-	-0.02	4.99	4.28	0.004761	0.0005503786
QT	30.81	0.1433	-0.0006	-0.02	4.62	3.85	0.003222	0.0005500401

Table 1: Parameters estimated of Constrained Kriging (CK), Linear Trend (LT) and Quadratic Trend (QT) models

predicted points. This deviation is in almost all cases much more narrowed for the constrained Kriging and a little bit more lager in the case of the quadratic model. Regarding these plots and their corresponding RMSE presented in table 2 one can conclude that adding constraints such as monotonicity and convexity in age direction can outperform the quadratic and linear model. In addition, the benefit of adding convexity constraint in the GP allows obtaining good fit and better forecast.

RMSE (Mortality rate RMSE)	Constrained Kriging	Linear model	Quadratic model
2008	0.0015121017	0.004605289	0.006104437
2009	0.0017016039	0.003870105	0.005373691
2010	0.0018920603	0.003323485	0.004827971
2011	0.0018228994	0.002501947	0.004005669

Table 2: RMSE of the three models for different years.

Recall that the main attempt of adding a mean function to the GP is to obtain better performances when predicting out of sample, somehow by minimizing the distance between the estimated path and the chosen points in the out of sample set. A conclusion from [LRZ18] is the fact that the choice of this mean function is no more clear. This can be seen between linear and quadratic trend models which one gives better results sometimes and not better the other time.

4 Conclusion

The objective of this paper was on the one hand to show that constrained Kriging is well suited to construct mortality surfaces and on the other hand to test its performance against the models developed in [LRZ18]. Our approach is carried out on the finite-dimensional approximation of the Gaussian process with the addition of a linear trend in the year direction.

The originality of our method is mainly based on the use of data for the construction of such surfaces. We used a low-dimensional random sample from the initial data set. The rest is considered as test data (see missing data) on the learning grid and can be used, subsequently, in the backtesting process. Illustrations on different mortality data sets from three countries have shown the latter's ability to obtain surfaces facing the monotonic-type shape constraint on the direction of ages.

The comparison with the models developed in [LRZ18] based on the addition of trend functions such as linear and quadratic motivated us to add a convexity constraint in the age direction. Indeed, it was shown in [LRZ18] that the quadratic tendency, which gives them better results in their study, adds a constraint of

convexity on the ages. Thanks to the RMSE criterion, this last specificity is taken into account in constrained Kriging makes it possible to obtain better performance in terms of out-of-sample forecasting.
The constrained Kriging allowed us to highlight differences between the lines, but it would be possible to continue the exploration of the finite-dimensional approximation by adding a quadratic trend on the ages.

References

[BDV02] Natacha Brouhns, Michel Denuit, and Jeroen K Vermunt. "A Poisson log-bilinear regression approach to the construction of projected lifetables". In: *Insurance: Mathematics and economics* 31.3 (2002), pp. 373–393.

[Cam+12] Carlo G Camarda et al. "MortalitySmooth: An R package for smoothing Poisson counts with P-splines". In: *Journal of Statistical Software* 50.1 (2012), pp. 1–24.

[Cha+21] Marc Chataigner, Areski Cousin, Stéphane Crépey, Matthew Dixon, and Djibril Gueye. "Beyond Surrogate Modeling: Learning the Local Volatility via Shape Constraints". In: *SIAM Journal on Financial Mathematics* 12.3 (2021), SC58–SC69.

[CG21] Areski Cousin and Djibril Gueye. "KRIGING FOR IMPLIED VOLATILITY SURFACE". In: (2021).

[CMR16] Areski Cousin, Hassan Maatouk, and Didier Rullière. "Kriging of financial term-structures". In: *European J. Oper. Res.* 255.2 (2016), pp. 631–648. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2016.05.057. URL: https://doi.org/10.1016/j.ejor.2016.05.057.

[Cur13] Iain D Currie. "Smoothing constrained generalized linear models with an application to the Lee-Carter model". In: *Statistical Modelling* 13.1 (2013), pp. 69–93.

[LC92] RD Lee and Lawrence Carter. "Modelling and forcasting the times series of US mortality". In: *Journal of the American Statistical Association* 87.419 (1992).

[LL+18] Andrés F López-Lopera, François Bachoc, Nicolas Durrande, and Olivier Roustant. "Finite-dimensional Gaussian approximation with linear inequality constraints". In: *SIAM/ASA Journal on Uncertainty Quantification* 6.3 (2018), pp. 1224–1255.

[Lud18] Michael Ludkovski. "Kriging metamodels and experimental design for Bermudan option pricing". In: (2018).

[LRZ18] Mike Ludkovski, Jimmy Risk, and Howard Zail. "Gaussian process models for mortality rates and improvement factors". In: *ASTIN Bulletin: The Journal of the IAA* 48.3 (2018), pp. 1307–1347.

[Maa17] Hassan Maatouk. "Finite-dimensional approximation of Gaussian processes with inequality constraints and errors measurements". In: (2017).

[MB14] Hassan Maatouk and Xavier Bay. *Gaussian Process Emulators for Computer Experiments with Inequality Constraints. hal-01096751*. 2014.

[PP14] Ari Pakman and Liam Paninski. "Exact hamiltonian monte carlo for truncated multivariate gaussians". In: *Journal of Computational and Graphical Statistics* 23.2 (2014), pp. 518–542.

[RGD12] Olivier Roustant, David Ginsbourger, and Yves Deville. "DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization". In: (2012).

[TP15] Chon Sern Tan and Ah Hin Pooi. "Prediction of mortality rates in the presence of missing values". In: *AIP Conference Proceedings*. Vol. 1691. 1. AIP Publishing LLC. 2015, p. 050027.

[Tic+12] John Maindonald Tickle, Simon Wood, R Core Team, and Maintainer Rob J Hyndman. "Package 'demography'". In: (2012).

[WR06] Christopher K Williams and Carl Edward Rasmussen. "Gaussian processes for machine learning". In: *the MIT Press* 2.3 (2006), p. 4.

[Woo94] Simon N Wood. "Monotonic smoothing splines fitted by cross validation". In: *SIAM Journal on Scientific Computing* 15.5 (1994), pp. 1126–1133.

[WW18] Ruhao Wu and Bo Wang. "Gaussian process regression method for forecasting of mortality rates". In: *Neurocomputing* 316 (2018), pp. 232–239.

A Results of the comparison of the three models using Italy and Germany mortality data

A.1 Parameter estimated of the three models using Germany male mortality data

	Trend parameters				Covariance matrix parameters			
	α_0	α_1^x	α_2^x	α_1^t	θ_x	θ_t	σ	ς
CK	25.19		-	-0.015	20.60	16.67	0.7924864	0.000397417
LT	35.48	0.09	-	-0.02	8.48	5.52	0.0052787	0.000390033
QT	34.97	0.1535	-0.0005	-0.02	3.85	4.47	0.0036205	0.000386172

Table 3: Parameters estimated of Constrained Kriging (CK), Linear Trend (LT) and Quadratic Trend (QT) models for Germany male mortality data

A.2 Plots of MAP estimate from Germany data

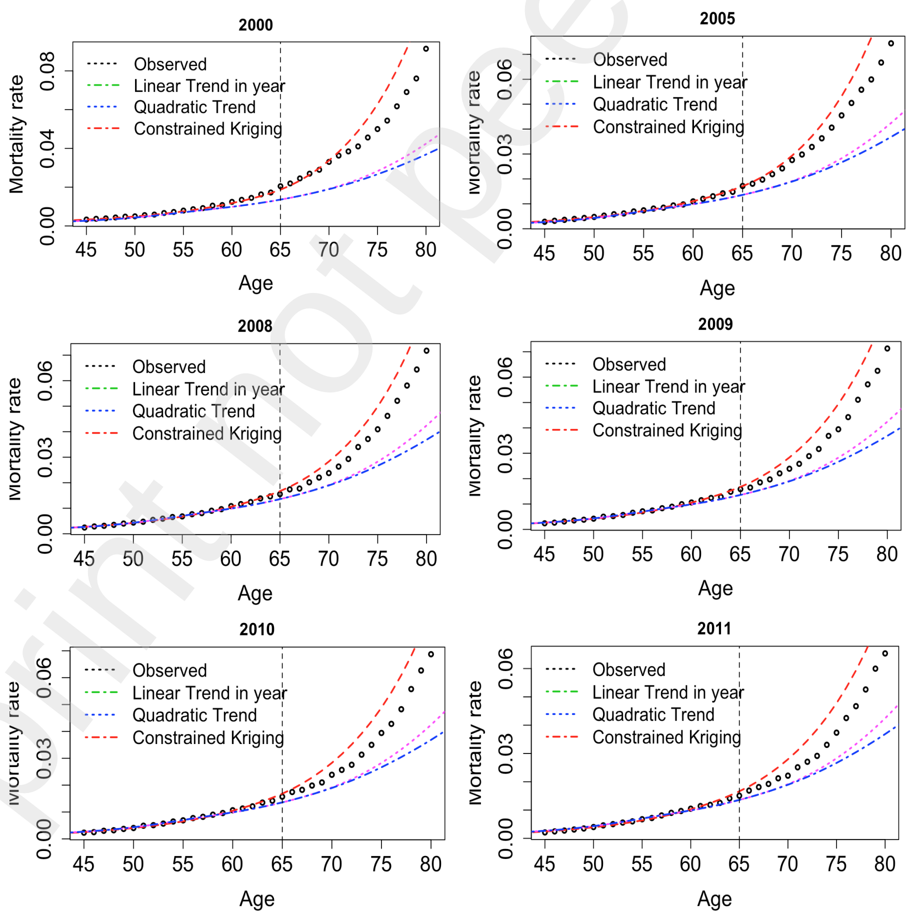

Figure 5: Slices of the MAP estimate of the three models from Germany data.

A.3 Parameter estimated of the three models using Italy male mortality data

	Trend parameters				Covariance matrix parameters			
	α_0	α_1^x	α_2^x	α_1^t	θ_x	θ_t	σ	ς
CK	98.48		-	-0.0516	29.97	21.21	1.77004900	0.000511804
LT	43.66	0.098	-	-0.0270	3.11	2.06	0.00045189	0.0004306641
QT	43.36	0.1067	-0.0001	-0.0271	3.03	2.03	0.00043478	0.0004297515

Table 4: Parameters estimated of Constrained Kriging (CK), Linear Trend (LT), and Quadratic Trend (QT) models for Italy male mortality data

A.4 Plots of MAP estimate from Italy data

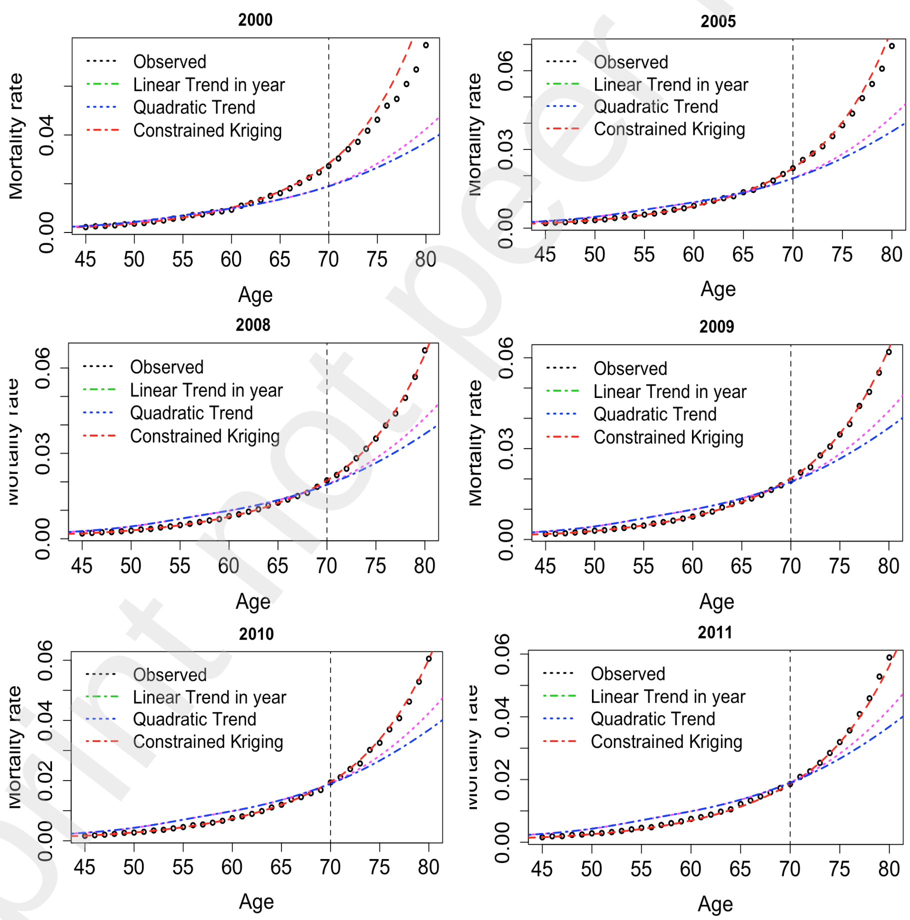

Figure 6: Slices of the MAP estimate of the three models from Italy data.