Spatiotemporal Prediction of COVID-19 Mortality and Risk Assessment

Antoni Torres-Signes

University of Málaga

M. Pilar Frías

University of Jaén <https://orcid.org/0000-0001-6886-0953>

María D.Ruiz-Medina (✉ mrui@ugr.es)

University of Granada <https://orcid.org/0000-0001-7445-7060>

Research Article

Keywords: Bayesian and Classical Functional Prediction, COVID-19 Mortality Risk, Machine Learning Forecasting, Nonlinear Parametric Regression

Posted Date: August 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-56955/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Spatiotemporal Prediction of COVID-19 Mortality and Risk Assessment

Antoni Torres-Signes · María P. Frías ·
María D. Ruiz-Medina

Received: date / Accepted: date

Abstract This paper presents a multivariate functional data statistical approach, for spatiotemporal prediction of COVID-19 mortality counts. Specifically, spatial heterogeneous nonlinear parametric functional regression trend model fitting is first implemented. Classical and Bayesian infinite-dimensional log-Gaussian linear residual correlation analysis is then applied. The nonlinear regression predictor of the mortality risk is combined with the plug-in predictor of the multiplicative error term. An empirical model ranking, based on random K -fold validation, is established for COVID-19 mortality risk forecasting and assessment, involving Machine Learning (ML) models, and the adopted Classical and Bayesian semilinear estimation approach. This empirical analysis also determines the ML models favored by the spatial multivariate Functional Data Analysis (FDA) framework. The results could be extrapolated to other countries.

Keywords Bayesian and Classical Functional Prediction, COVID-19 Mortality Risk, Machine Learning Forecasting, Nonlinear Parametric Regression.

Mathematics Subject Classification (2010) MSC code1 60G25 · 60G60 and 62J05 · MSC code2 62J10

A. Torres-Signes

Department of Statistics and Operation Research, Faculty of Sciences, University of Málaga, Spain

E-mail: atisignes@uma.es

M. P. Frías

Department of Statistics and Operation Research, Faculty of Sciences, University of Jaén, Spain

E-mail: mpfrias@ujaen.es

M.D. Ruiz-Medina

Department of Statistics and Operation Research, Faculty of Sciences, University of Granada, Spain

E-mail: mruiz@ugr.es

1 Introduction

Coronavirus disease 2019 (COVID-19) rapidly spreads around many other countries, since December 2019 when arises in China (see [44]; [51]). It has supposed a substantial damage in terms of both human lives and economic cost. There is no vaccine specifically designed for this virus. Thus, the absence of reliable data regarding this infectious disease transmission makes necessary cautious responses. A crucial problem is geo-localization of outbreaks, since it allows a more effective allocation of medical resources (see, e.g., [11]; [25]; [34]; [36], just to mention a few). Some mathematical models have been introduced in the literature, trying to describe the spatiotemporal dynamics of COVID-19. To generate and assess short-term forecasts of the cumulative number of reported cases, three models are presented in [37], which were validated with outbreaks of other diseases different from COVID-19. Alternative SEIR type models with little variations are formulated, involving, for instance, stochastic components (see [26]). In [19], a θ -SEIHRD model is introduced, well adapted to COVID-19, based on the Be-CoDiS model (see [20]). This model is able to estimate the number of cases, deaths, and needs of beds in hospitals. The most remarkable feature of the presented approach in [19] is the balance between complexity and indentifiability of model parameters.

Mathematics contributes in an essential way to the derivation of epidemiological models, and the subsequent analysis of the causes, dynamics, and spread of an epidemic/pandemic (see, e.g., [17] and [23] for an overview on mathematical models in Epidemiology). Particularly, ordinary differential equations have been extensively applied in the characterization of the time evolution of an infectious disease over uniform population distributions within a habitat. Most of those models are compartment epidemic models, such as the SIR (susceptible-infectious-recovered) model. The SIR model was first proposed in [24] to explain plague and cholera epidemics. It was later extended to the SIR-susceptible (SIRS) model to allow imperfect immunity (see, e.g., [12]). Indeed, there exists an extensive literature on SIR models, based on ordinary differential equations (see, e.g. [13]; [21]; [27]; [35]; [41]; [48]; [49]). Delay differential equations have also been applied in the SIR modeling context (see [5]; [31]; [40], among others). These models are limited, since any structure due to the spatial position is not reflected. Reaction-diffusion models provide a framework for their spatial extension, reflecting the infectious disease spread over a spatial region. Thus, they provide a simplest modeling alternative in Ecology and Epidemiology involving space and time. A diffusive predator-prey model is analyzed in [15]. A one dimensional SIR model, including constant diffusive movement of all individuals is derived in [46] with no-flux boundary conditions. SIR modeling based on hyperbolic partial differential equations is studied in [32], and, based on parabolic partial differential equation systems with no-flux boundary conditions in [28]. See also [39] and [38] on SEIRD (susceptible, exposed, infected, recovered, deceased) model, that incorporates the spatial spread of the disease with inhomogeneous diffusion terms.

The stochastic version of SIR-type models intends to cover several limitations detected regarding uncertainty, in the observations, and the hidden dynamical epidemic process. Specifically, the counts compartments are observed with error, and several factors introduce uncertainty in the hidden epidemic process, such as the presence of heterogeneous populations. The transmission, recovery and loss of immunity rate parameters, for example, are uncertain themselves. In a first stage, a probabilistic mechanism is adopted involving a Markov chain of SIR states (see [4]; [47]). Some more recent stochastic models involve complex networks (see [50]; [43]) or drug-resistant influenza (see [10]). As pointed out in [52] (see references therein), these models do not take into account the observation error in the counts, and obtain mass balance equation from the observed counts. The uncertainty in the parameter space is neither accounted for. The approach presented in [52] covers these gaps, adopting a Bayesian hierarchical statistical SIRS model framework. Specifically, the stochastic SIR-like model fitting is performed, assuming that the counts are observed with error and the dynamical evolution of the SIR process is stochastic. Parameter uncertainty is also modelled under this Bayesian statistical framework (see also [2]).

Many alternative models have been proposed to address the problem of uncertainty in the counts and the parameters, in the spatiotemporal statistical and Cox process context (see, e.g., [1]; [4]; [14], among others). Some of them go beyond the SIR modeling framework (see, e.g., [7] or [22], where the problem of correlation of rates of change between neighbouring sites or individuals is addressed). Survival Dynamical System are proposed in [45], paying attention to survival functions instead of population counts. We particularly refer to the approach presented in [33], where a spatially descriptive, temporally dynamic hierarchical model is formulated, to handle inference from small-area counts on the number of infected during successive, regular time intervals. The realizations of an underlying multivariate autoregressive process are involved in the infection relative risk modeling, incorporating the space-time dynamics, and providing an epidemic-field-based view. Markov chain Monte Carlo (MCMC) in a Bayesian statistical framework is applied to compute the posterior estimates of all parameters of interest.

Our paper extends some of the tools applied in [33] to the nonlinear trend model fitting, and infinite-dimensional statistical frameworks. The prediction approach presented involves nonlinear parametric functional regression, and spatial multivariate functional linear residual correlation analysis. It is well-known that Functional Data Analysis (FDA) techniques allow to make inference on the small-scale and large-scale sample properties, in our case, of the hidden stochastic mortality risk process generating the spatiotemporal counts (see, e.g., [42]). Note that the classical stochastic diffusion models offer a *particle* view rather than a field view (see, e.g., [30]). Thus, we incorporate large- and small-scale sample information in our statistical analysis framework. The first stage involves nonlinear parametric curve regression, from the observed realizations of the COVID-19 mortality log-risk process, at each one of the Spanish Communities. The second step performs a spatial multivariate

functional linear residual correlation analysis. Thus, the resulting semi-linear additive plug-in predictor approximates the behavior of the COVID-19 mortality log-risk process in space and time. Bayesian and Classical approaches are adopted in the implementation of the second step. Hence, our approach includes uncertainty in the observed counts, parameters and hidden COVID-19 mortality log-risk process. The presented estimation methodology is validated from the observed number of deaths registered at the Spanish Communities, during the period analyzed, since March, 8, 2020 until May, 13, 2020. Our results show a remarkable qualitative agreement with the reported epidemiological data.

Nowadays Machine learning models have established themselves as serious contenders to classical statistical models in the area of forecasting. Research started in the eighties with the development of the neural network model. Subsequently, research extended the concept to alternative models, such as support vector machines, decision trees, and others, that are collectively called machine learning models ([3]; [16]). This paper also aims to establish an empirical comparative study of these models and our approach. Applying random K -fold validation, the forecasting ability of ML models are tested into two categories, respectively corresponding to the original and projected functional data. A model ranking is established in both cases. In the second category, they are compared with our multivariate infinite-dimensional Bayesian and Classical semilinear approaches. From the COVID-19 mortality data set analyzed, the outperformance of Radial Basis Function Neural Network (RBF), and Gaussian Processes (GP) against the remaining forecasting models tested: Support Vector Regression (SVR), Bayesian Neural Networks (BNN), Classical and Bayesian multivariate infinite-dimensional semilinear estimation, Multilayer Perceptron (MLP), Generalized Regression Neural Network (GRNN) is observed. Note that, the best K -fold running for these superior RBF and GP methods is obtained under an FDA preprocessing of the data. Our approach presents a more stable and robust behavior against the possible random splitting of the functional sample into training, and validation or testing samples.

The outline of the paper is the following. The adopted modeling approach is introduced in Section 2. Section 3 derives the estimation methodology. This methodology is applied to the spatiotemporal statistical analysis of COVID-19 mortality in Spain in Section 4. The empirical comparative study with ML models is addressed in Section 5. An empirical model ranking is provided in Section 6 leading to some concluding remarks. In the Appendix additional outputs, obtained in our spatiotemporal estimation of COVID-19 mortality and risk assessment in Spain, are displayed. Finally, this Appendix also provides a brief introduction to our implementation of ML models in an FDA framework.

2 The Model

Let $(\Omega, \mathcal{A}, \mathcal{P})$ be the basic probability space. Consider H to be a real separable Hilbert space, with H^* denoting its dual. Let $A = \{A_{t,\mathbf{z}}, \mathbf{z} \in D \subset \mathbb{R}^d, t \in$

$\mathbb{R}_+, \}$ be a spatiotemporal risk process on $(\Omega, \mathcal{A}, \mathcal{P})$ such that

$$\mathcal{I}_{\mathcal{T}}(h) = \int_{\mathcal{T}} \exp \left(\int_D \ln(\lambda_{t,\mathbf{z}}) h(\mathbf{z}) d\mathbf{z} \right) dt < \infty, \forall h \in H, \mathcal{T} \in \mathcal{B}, \text{ a.s.}, \quad (1)$$

where $\mathcal{T}$ is a bounded Borel set in $\mathbb{R}_+$, and $\mathcal{B}$ denotes the Borel σ -algebra. Process Λ is assumed to satisfy $E[\|\ln(\Lambda_t(\cdot))\|_H^2] < \infty$, with $\Lambda_t(\mathbf{z}) = \Lambda_{t,\mathbf{z}}$, for every $\mathbf{z} \in D$ and $t \in \mathbb{R}_+$. Here,

$$\left\{ \exp \left(\int_D \ln(\lambda_{t,\mathbf{z}}) h(\mathbf{z}) d\mathbf{z} \right), t \in \mathbb{R}_+ \right\} = \{\mathcal{R}_t(h), t \in \mathbb{R}_+\}$$

defines a temporal realization of the generalized risk process evaluated at the test function $h \in H$. Note that, the weighted integration of the realizations of the log-risk process, in terms of test function $h \in H$, keeps the spatial information on this process. For certain families of compactly supported test functions, the local mean defined by this integral allows the localization of the values of the spatiotemporal risk process over a specific spatial region.

For any bounded Borel set $\mathcal{T}$, let $N(\mathcal{T}) : (\Omega, \mathcal{A}, \mathcal{P}) \rightarrow \mathbb{N}$ be a measurable mapping, with values in the natural numbers $\mathbb{N}$. Given the observation of $\mathcal{I}_{\mathcal{T}}(h)$, the conditional probability distribution of the number of random events $N(\mathcal{T})$, that occur in the bounded interval $\mathcal{T} \in \mathcal{B}$, is a Poisson probability distribution with functional mean $\mathcal{I}_{\mathcal{T}}(h)$, $h \in H$. Thus, its conditional characteristic function is given by

$$\begin{aligned} f_{N(\mathcal{T})/\mathcal{I}_{\mathcal{T}}(h)}(\xi) &= E[\exp(i\xi N(\mathcal{T})) / \mathcal{I}_{\mathcal{T}}(h)] \\ &= \exp(\mathcal{I}_{\mathcal{T}}(h) (\exp(i\xi) - 1)), \quad \forall \xi, \quad h \in H. \end{aligned} \quad (2)$$

2.1 Data Model

Let us consider the following observation model for the generalized log-risk process: For $t = 1, \dots, T$,

$$\begin{aligned} \ln(\Lambda_t)(\psi_{p,\varpi_p}) &= g_t(\psi_{p,\varpi_p}, \boldsymbol{\theta}(p)) + \varepsilon_t(\psi_{p,\varpi_p}) \\ &= \langle g_t(\cdot, \boldsymbol{\theta}(p)), \psi_{p,\varpi_p}(\cdot) \rangle_H + \langle \varepsilon_t(\cdot), \psi_{p,\varpi_p}(\cdot) \rangle_H, \quad p = 1, \dots, P, \end{aligned} \quad (3)$$

where $\{\psi_{p,\varpi_p}, p = 1, \dots, P\} \subset H$ denotes a function family in H , whose elements have compact support defining the small-areas, where the counts are aggregated. Here, $\varpi_p, p = 1, \dots, P$, are the localization hyperparameters (in the next section, the center and bandwidth parameters), associated with those small-areas. For each $p \in \{1, \dots, P\}$, $\boldsymbol{\theta}(p) = (\theta^1(p), \dots, \theta^q(p)) \in \Theta$ denotes the unknown parameter vector to be estimated at the p th region, and Θ is the open set defining the parameter space, whose closure Θ^c is compact in $\mathbb{R}^q$. The parametric regression function g_t is continuous over $\boldsymbol{\theta}(\cdot) \in \Theta$, and $t \in \mathbb{R}_+$, with respect to H -norm. Furthermore, we assume that g_t is of the form

$$g_t(\boldsymbol{\theta}(\cdot)) = \sum_{k=1}^N (A_k(\cdot) \cos(\varphi_k(\cdot)t) + B_k(\cdot) \sin(\varphi_k(\cdot)t)), \quad t \in \mathbb{R}_+, \quad (4)$$

with

$$\begin{aligned}
\boldsymbol{\theta}(\cdot) &= (\theta_1(\cdot), \theta_2(\cdot), \theta_3(\cdot), \dots, \theta_{3N-2}(\cdot), \theta_{3N-1}(\cdot), \theta_{3N}(\cdot)) \\
&= (A_1(\cdot), B_1(\cdot), \varphi_1(\cdot), \dots, A_N(\cdot), B_N(\cdot), \varphi_N(\cdot)) \in H^{3N}, \\
C_k^2(\cdot) &= A_k^2(\cdot) + B_k^2(\cdot) > 0, \quad k = 1, \dots, N, \\
0 &\leq \underline{\varphi}(\cdot) < \varphi_1(\cdot) < \dots < \varphi_N(\cdot) < \overline{\varphi} < \infty.
\end{aligned} \tag{5}$$

In equation (3), for any $T \geq 2$, ε_t is assumed to be a temporal dynamical zero-mean process that satisfies the state equation

$$\varepsilon_t = \rho(\varepsilon_{t-1}) + \nu_t, \quad t = 1, \dots, T, \tag{6}$$

where $\{\nu_t, t = 1, \dots, T\}$ are independent H -valued zero-mean Gaussian random variables, with $E[\|\nu_t\|_H^2] = E[\|\nu_0\|_H^2] < \infty$, and $P[\nu_t \in H] = 1$, $t = 1, \dots, T$. The autocorrelation operator ρ is bounded linear operator. Thus, the observation noise ε_t obeys an Autoregressive Hilbertian equation of order one (ARH(1) equation), which admits an unique stationary solution under the condition $\|\rho\|_{\mathcal{L}(H)}^{k_0} < 1$, for certain $k_0 \in \mathbb{N}$ (see, e.g., [8]). Here, $\|\cdot\|_{\mathcal{L}(H)}$ denotes the bounded linear operator norm in H .

3 Trend Model Fitting and Residual Correlation Analysis

Let $L_1, \dots, L_P$ be the small-areas, where the counts are aggregated. Consider $\{\psi_{p, \rho_p}, p = 1, \dots, P\} \subset H$ to be a family of compactly supported radial functions in H , whose centers $c_p, p = 1, \dots, P$, are allocated at the regions $L_1, \dots, L_P$, respectively. The bandwidth parameters $\rho_1, \dots, \rho_P$ provide their associated windows with suitable size.

From (3), we first compute the least-squares estimator (LSE) $\widehat{\boldsymbol{\theta}}_T(p)$, in the Walker sense, of the parameter vector $\boldsymbol{\theta}(p)$, for $p = 1, \dots, P$ (see (4)–(5)). Thus, for $p = 1, \dots, P$, the LSE $\widehat{\boldsymbol{\theta}}_T(p)$ is obtained as the solution to the following optimization problem:

$$L_T(\widehat{\boldsymbol{\theta}}_T(p)) = \inf_{\boldsymbol{\tau}_p \in \Theta^c} L_T(\boldsymbol{\tau}_p) = \inf_{\boldsymbol{\tau}_p \in \Theta^c} \frac{1}{T} \sum_{t=1}^T |\ln(\Lambda_t)(\psi_{p, \rho_p}) - g_t(\boldsymbol{\tau}_p)|^2. \tag{7}$$

Hence,

$$\widehat{\ln(\Lambda_t)}(\psi_{p, \rho_p}) = g_t(\widehat{\boldsymbol{\theta}}_T(p)) \tag{8}$$

defines the parametric nonlinear regression predictor of the log-risk process at region p , for any time $t \in \mathbb{R}_+$, and for $p = 1, \dots, P$.

3.1 Classical and Bayesian componentwise estimation of the mortality residual log-risk process

Let $\mathbf{Y} = \{Y_t = \ln(\Lambda_t(\cdot)) - g_t(\widehat{\boldsymbol{\theta}}_T(\cdot)), t = 1, \dots, T\}$ be the functional residuals from the nonlinear parametric regression model fitting. The empirical autocovariance and cross-covariance operators, based on the residuals $Y_t(\cdot) = \ln(\Lambda_t(\cdot)) - g_t(\widehat{\boldsymbol{\theta}}_T(\cdot))$, $t = 1, \dots, T$, are respectively given by

$$\widehat{R}_{0,T}^{\mathbf{Y}}(h)(g) = \frac{1}{T} \sum_{t=1}^T Y_t(h)Y_t(g), \quad \widehat{R}_{1,T}^{\mathbf{Y}}(h)(g) = \frac{1}{T-1} \sum_{t=1}^{T-1} Y_t(h)Y_{t+1}(g), \quad (9)$$

for every $h, g \in H$. Operators $\widehat{R}_{0,T}^{\mathbf{Y}}$ and $\widehat{R}_{1,T}^{\mathbf{Y}}$ are also nuclear operators. In particular, we will consider the spectral decomposition

$$\widehat{R}_{0,T}^{\mathbf{Y}} = \sum_{k=1}^T \lambda_{k,T}(\widehat{R}_{0,T}^{\mathbf{Y}})[\phi_{k,T} \otimes \phi_{k,T}], \quad (10)$$

where $\{\lambda_{k,T}(\widehat{R}_{0,T}^{\mathbf{Y}}), k = 1, \dots, T\}$ and $\{\phi_{k,T}, k \geq 1\}$ denote the empirical eigenvalues and eigenvectors of $\widehat{R}_{0,T}^{\mathbf{Y}}$, respectively.

The following classical componentwise estimator of the autocorrelation operator is considered (see [8]): For every $h, g \in H$,

$$\widehat{\rho}_T(h)(g) = \sum_{k,l=1}^{K(T)} \frac{1}{\lambda_{k,T}(\widehat{R}_{0,T}^{\mathbf{Y}})} \langle h, \phi_{k,T} \rangle_H \widehat{R}_{1,T}^{\mathbf{Y}}(\phi_{k,T})(\phi_{l,T}) \langle g, \phi_{l,T} \rangle_H, \quad (11)$$

where $K(T)$ denotes the truncation parameter, that depends on the functional sample size T . An optimal choice is given by $K(T) = \ln(T)$. Under the conditions derived in Corollary 4.1, Theorem 4.8 and Theorem 8.7 in [8], $\widehat{\rho}_T$ is a strong-consistent componentwise estimator of the autocorrelation operator of the residuals, with respect to the norm in the space $\mathcal{L}(H)$ of bounded linear operators. Under the assumptions made in Theorem 1 in [18], the corresponding plug-in predictor

$$\widehat{Y}_t = \widehat{\rho}_T(Y_{t-1}), \quad t \geq 1, \quad (12)$$

is a weak-consistent estimator of ε_t in equation (6), with respect to H -norm. We refer to (12), as the classical plug-in predictor of the mortality residual log-risk process.

In the Bayesian componentwise estimation of ρ in (6), we assume that ρ is compact self-adjoint operator. An a-priori beta probability distribution with parameters $a_k > 0$ and $b_k > 0$ is considered to represent the uncertainty in the eigenvalue $\gamma_k(\rho)$ of ρ , for every $k \geq 1$. From the sample ε_t , $t = 1, \dots, T$,

of size T , since $\gamma_k(\rho) \sim \mathcal{B}(a_k, b_k)$, the likelihood function is computed as

$$\begin{aligned} \tilde{L}_k(\varepsilon_{1k}, \dots, \varepsilon_{Tk}, \gamma_k(\rho)) &= \frac{1}{(\sigma_k \sqrt{2\pi})^T} \exp \left(-\frac{1}{2\sigma_k^2} \sum_{t=1}^T (\varepsilon_t(\psi_k) - \gamma_k(\rho) \varepsilon_{t-1}(\psi_k))^2 \right) \\ &\quad \times [\gamma_k(\rho)]^{a_k-1} (1 - \gamma_k(\rho))^{b_k-1} \frac{\mathbb{I}_{\{0 < \gamma_k(\rho) < 1\}}}{\mathbb{B}(a_k, b_k)} \\ &= \frac{1}{(\sigma_k \sqrt{2\pi})^T} \exp \left(-\frac{1}{2\sigma_k^2} \sum_{t=1}^T [\nu_t(\psi_k)]^2 \right) \\ &\quad \times [\gamma_k(\rho)]^{a_k-1} (1 - \gamma_k(\rho))^{b_k-1} \frac{\mathbb{I}_{\{0 < \gamma_k(\rho) < 1\}}}{\mathbb{B}(a_k, b_k)}, \quad k \geq 1, \end{aligned} \quad (13)$$

where $\varepsilon_{tk} = \varepsilon_t(\psi_k) = \langle \varepsilon_t, \psi_k \rangle_H$, and $\sigma_k = \sqrt{E[\varepsilon_t(\psi_k)]^2}$, $t = 1, \dots, T$, with ψ_k being the k -th eigenvector of ρ , for each $k \geq 1$. Here, $\mathbb{I}$ is the indicator function on the interval $(0, 1)$, and $\mathbb{B}(a_k, b_k)$ denotes the beta function $\mathbb{B}(a_k, b_k) = \frac{\Gamma(a_k)\Gamma(b_k)}{\Gamma(a_k+b_k)}$. From (13), the Bayesian estimator $\widehat{\gamma_k(\rho)}$ of $\gamma_k(\rho)$ is given by

$$\begin{aligned} \widehat{\gamma_k(\rho)} &= \\ &= \frac{1}{2\zeta_k(T)} \left((\xi_k(T) + \zeta_k(T) \pm \sqrt{(\xi_k(T) - \zeta_k(T))^2 - 4\zeta_k(T)\sigma_k^2(2 - (a_k + b_k))}) \right), \end{aligned} \quad (14)$$

where, for each $k \geq 1$, $\xi_k(T) = \sum_{t=1}^T \varepsilon_{t-1}(\psi_k) \varepsilon_t(\psi_k)$, $\zeta_k(T) = \sum_{t=1}^T [\varepsilon_{t-1}(\psi_k)]^2$, and $\sigma_k^2 = E[\varepsilon_t(\psi_k)]^2$, for $t = 1, \dots, T$.

The Bayesian componentwise estimator $\tilde{\rho}_T$ of the autocorrelation operator ρ is then computed as

$$\tilde{\rho}_T = \sum_{k=1}^{K(T)} \widehat{\gamma_k(\rho)} [\psi_k \otimes \psi_k], \quad (15)$$

where $K(T)$ is the truncation parameter, usually selected as $K(T) = \ln(T)$.

The resulting Bayesian componentwise plug-in predictor is given by

$$\tilde{\varepsilon}_t = \tilde{\rho}_T(\varepsilon_{t-1}), \quad t \geq 1.$$

In practice, in the computation of the Bayesian componentwise estimator $\tilde{\rho}_T$, we replace in (13)–(14), ε_t by $\hat{\varepsilon}_t = Y_t = \ln(\Lambda_t(\cdot)) - g_t(\hat{\theta}_T(\cdot))$, for each $t = 1, \dots, T$. The eigenvectors $\{\psi_k, k \geq 1\}$ can also be replaced by their empirical counterpart. Hence,

$$\tilde{Y}_t = \tilde{\rho}_T(Y_{t-1}), \quad t \geq 1. \quad (16)$$

We refer to (16) as the Bayesian plug-in predictor of the residual log-risk process Y_t . From equations (12) and (16), the corresponding Classical and

Bayesian plug-in predictors implemented for spatiotemporal forecasting of COVID-19 mortality risk are given, for any t , by

$$\begin{aligned}\widehat{\mathbf{R}}_t(\psi_{p,\rho_p}) &= \exp\left(g_t(\widehat{\boldsymbol{\theta}}_T(p)) + \widehat{Y}_t(\psi_{p,\rho_p})\right), \quad p = 1, \dots, P \\ \widetilde{\mathbf{R}}_t(\psi_{p,\rho_p}) &= \exp\left(g_t(\widetilde{\boldsymbol{\theta}}_T(p)) + \widetilde{Y}_t(\psi_{p,\rho_p})\right), \quad p = 1, \dots, P.\end{aligned}\quad (17)$$

Classical and Bayesian estimation of the autocovariance operator of the innovation process ν_t in (6) is provided in Appendix.

4 Statistical Analysis of COVID-19 Mortality in Spain

The COVID-19 mortality analysis performed here is based on the daily records reported by the Spanish Statistical National Institute, since March, 8 to May, 13, 2020, at the 17 Spanish Communities. Interpolation at 265 temporal nodes, and cubic B-spline smoothing is applied to obtain our functional data set, in the implementation of the spatial heterogeneous nonlinear parametric curve regression model (3)–(4) (see Figures 9 and 10 in Appendix).

The spatiotemporal correlation structure of COVID-19 mortality residual log-risk process is estimated from equations (11) and (15). Thus, from (17), with $P = 17$,

$$\begin{aligned}\widehat{\ln(\Lambda_t)}(\psi_{p,\rho_p}) &= g_t(\widehat{\boldsymbol{\theta}}_T(p)) + \widehat{Y}_t(\psi_{p,\rho_p}), \quad t \in \mathbb{R}_+, \quad p = 1, \dots, 17 \\ \widetilde{\ln(\Lambda_t)}(\psi_{p,\rho_p}) &= g_t(\widetilde{\boldsymbol{\theta}}_T(p)) + \widetilde{Y}_t(\psi_{p,\rho_p}), \quad t \in \mathbb{R}_+, \quad p = 1, \dots, 17.\end{aligned}\quad (18)$$

Tables 9–10, in Appendix, display the parameter estimates $\widehat{A}_k(\cdot)$ and $\widehat{B}_k(\cdot)$, $k = 1, \dots, N$, in the nonlinear parametric trend model fitting. Here, we have considered $N = 6$. Thus, 12 parameters must be estimated, since $\varphi_k = \frac{2\pi}{66}$, for $k = 1, \dots, N = 6$, is assumed to be constant. In Tables below, the following Spanish Community codes have been considered: C1 for Andalucía; C2 for Aragón; C3 for Asturias; C4 for Islas Baleares; C5 for Canarias; C6 for Cantabria; C7 for Castilla La Mancha; C8 for Castilla y León; C9 for Cataluña; C10 for Comunidad Valenciana; C11 for Extremadura; C12 for Galicia; C13 for Comunidad de Madrid; C14 for Murcia; C15 for Navarra; C16 for Pas Vasco, and C17 for La Rioja.

In equation (14), for the computation of the Bayesian estimator (15) of the autocorrelation operator ρ , the prior $\gamma_k(\rho) \sim \beta(4, 5)$ has been tested, for every $k \geq 1$. Note that, although, $K(T) = \lfloor \ln(T) \rfloor^- = \lfloor \ln(265) \rfloor^- = \lfloor 5.57973 \rfloor^- = 5$ is usually considered, we have also tested the truncation parameter values $N = 8$ and $N = 9$. Table 1 displays the discrete approximation of $\|\cdot\|_{L^2([0,66])}$ -norm, multiplied by the scale factor $\mathcal{S} = 10^2$, of the empirical functional quadratic errors associated with these three truncation parameter values. We find that the sample size $T = 265$ leads to the threshold dimension $K(265) = 9$, providing the allowed model complexity, in our finite-dimensional implementation of the

Table 1: $\|\cdot\|_{L^2([0,66])}$ -norm, multiplied by the scale factor $S = 10^2$, of the empirical functional quadratic errors at the 17 Spanish Communities, after applying nonlinear trend model fitting, and Classical and Bayesian residual correlation analysis, for the truncation parameter values $K(T) = 5, 9, 8$ from the left to the right

SC/TP	B.5	C.5	B.9	C.9	B.8	C.8
C1	0.0069	0.0027	0.0067	0.0026	0.0066	0.0026
C2	0.0388	0.0186	0.0388	0.0185	0.0388	0.0185
C3	0.0048	0.0023	0.0042	0.0018	0.0044	0.0019
C4	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001
C5	0.0007	0.0003	0.0007	0.0003	0.0007	0.0003
C6	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
C7	0.0126	0.0067	0.0118	0.0060	0.0118	0.0060
C8	0.0018	0.0015	0.0008	0.0005	0.0008	0.0005
C9	0.0858	0.0409	0.0856	0.0405	0.0857	0.0405
C10	0.0016	0.0010	0.0011	0.0005	0.0011	0.0005
C11	0.0011	0.0009	0.0008	0.0006	0.0008	0.0006
C12	0.0405	0.0152	0.0408	0.0155	0.0408	0.0154
C13	0.0253	0.0162	0.0252	0.0161	0.0252	0.0161
C14	0.0002	0.0001	0.0002	0.0001	0.0002	0.0001
C15	0.0408	0.0174	0.0407	0.0174	0.0407	0.0173
C16	0.0033	0.0018	0.0029	0.0013	0.0029	0.0013
C17	0.0004	0.0003	0.0003	0.0002	0.0003	0.0002

COVID-19 mortality log-risk predictor, under both, Bayesian and Classical approaches.

Figures 1 and 2 respectively show the maps reflecting the original and estimated spatiotemporal evolution of COVID-19 mortality risk, and accumulated mortality counts, under the presented Bayesian and Classical semilinear estimation approaches, over the 17 Spanish Communities analyzed.

5 An Empirical Comparative Study

This section provides an empirical model ranking for spatiotemporal COVID-19 mortality risk assessment, based on random K -fold validation, involving the ML models introduced in Appendix, and the forecasting approach presented in this paper. The random K -fold ($K = 5, 10$) has been implemented from the records reported by the Spanish Statistical National Institute, since March, 8 to May, 13, 2020. The empirical ranking is referred to two categories, where we respectively perform forecasts from the observed pointwise, and from the generalized (projected) COVID-19 mortality log-risk process realizations. In the first category the ML models: Multilayer Perceptron (MLP), Radial Basis Function Neural Network (RBF), Bayesian Neural Network (BNN), Generalized Regression Neural Network (GRNN), Support Vector Regression (SVR), and Gaussian Processes (GP) are compared. In the

Fig. 1: COVID-19 mortality risk maps, since March, 8 to May, 13, 2020. Original (left) and estimated (right) maps, computed from the Bayesian (first line) and classical (second line) COVID-19 mortality predictors, for the truncation parameter value $K(T) = 8$

Fig. 2: COVID-19 mortality cumulative cases maps, since March, 8 to May, 13, 2020. Original (left) and estimated (right) maps, from the Bayesian (first line) and classical (second line) COVID-19 mortality count predictors, for the truncation parameter value $K(T) = 8$

second category, the performance of the presented approach is compared with these ML methodologies.

A brief introduction to the ML methodologies, here implemented in a classical and FDA frameworks, is provided in Appendix.

5.1 Results from the Empirical Comparative Study

We start from the functional data set obtained after the FDA preprocessing procedure has been applied, as described in the previous section. Specifically, data have been interpolated and cubic B -spline smoothed. The logarithmic transform and linear scaling are also applied. We held out the first ten points and the last three, for each log-risk curve, as an out of sample set. In the first category, ML models are implemented from this data set, establishing a first model ranking based on the obtained validation results. In the second category, COVID-19 mortality log-risk curves are projected onto the empirical eigenvectors of the autocovariance operator, considering the truncation parameter value $K(T) = 8$. Our choice of the truncation parameter $K(T) = K(265) = 8$ fits the threshold providing a balance between $K(T) = \lfloor \ln(T) \rfloor = \lfloor \ln(265) \rfloor = 5$, signing an agreement with the separation and velocity decay of the empirical eigenvalues of the autocovariance operator, and the parameter value $K(T) = K(265) = 9$, controlling model complexity according to the sample size $T = 265$. Based on the validation results obtained in this category, a second model ranking is established involving the ML models tested and the approach presented. The derived empirical model ranking is not affected by these two categories. However, it can be observed how each ML model performance is favored by one of the two categories (see Section 6). For instance, linear SVR is the best option for the first category, while better forecasting results are obtained when nonlinear SVR is implemented from the projected data. As we will see in Section 6, the particular features of the data at each one of the 17 Spanish Communities analyzed also affect the performance of the implemented prediction methodologies.

One-step-ahead forecasting is selected in the implementation of the ML methodologies, and the proposed Classical and Bayesian semilinear approaches. Model fitting is evaluated in terms of the Symmetric Mean Absolute Percentage Errors (SMAPEs), given by, for $P = 17$, and $T = 265$,

$$\frac{1}{T} \sum_{t=1}^T \frac{\left| \widehat{\ln(A_t)}(\psi_{p,\rho_p}) - \ln(A_t)(\psi_{p,\rho_p}) \right|}{\left(\left| \ln(A_t)(\psi_{p,\rho_p}) \right| + \left| \widehat{\ln(A_t)}(\psi_{p,\rho_p}) \right| \right) / 2}, \quad p = 1, \dots, P. \quad (19)$$

The most widely used methodology for model selection in ML is K -fold validation approach. In our case, we have applied random K -fold validation. We have computed the mean of the SMAPEs obtained at each one of the K iterations of the random K -fold cross validation procedure implemented. This validation technique consists of random splitting the functional sample into a training and validation sample at each one of the K iterations. Model fitting is performed from the training sample, and the target outputs are defined from the validation or testing sample. In particular, we have selected $K = 5, 10$, in the K -fold validation implemented. By running each model ten times and averaging SMAPEs, we remove the fluctuations due to the random initial weights (for MLP and BNN models), and the differences in the parameter estimation

in all methods, due to the random specification of the sample splitting in the implementation of the random K -fold validation procedure.

The ten-running based random K -fold validation SMAPEs are displayed in Tables 2 ($K = 5$) and Table 3 ($K = 10$), for the six ML techniques tested, GRNN, MLP, SVR, BNN, RBF, and GP, from data in the first category. The ML model estimation results, based on the complete functional sample, are also displayed in Table 11 in Appendix (see also Figures 3 and 4). In Table 12 in Appendix, one can also find the corresponding estimation results from the projected data. (Note that, the same Spanish Community codes, as before, are used in all the tables referred in this section, that is, C1 for Andalucía; C2 for Aragón; C3 for Asturias; C4 for Islas Baleares; C5 for Canarias; C6 for Cantabria; C7 for Castilla La Mancha; C8 for Castilla y León; C9 for Cataluña; C10 for Comunidad Valenciana; C11 for Extremadura; C12 for Galicia; C13 for Comunidad de Madrid; C14 for Murcia; C15 for Navarra; C16 for Pas Vasco, and C17 for La Rioja).

Tables 4 ($K = 5$) and Table 5 ($K = 10$) provide the ten-running based random K -fold validation results, from the projected data category. The random K -fold validation results obtained after implementation of the Bayesian and Classical semilinear approaches, as well as the estimation results based on the global sample are displayed in Table 6. See also Figures 5 and 6, where original and estimated COVID-19 mortality log-risk and cumulative cases curves are respectively shown.

In the previous selection of ML model hyperparameters, we have also applied random K -fold validation ($K = 5, 10$). Our selection has been made from a suitable set of candidates. Specifically, the optimal numbers of hidden (NH) nodes in the implementation of MLP and BNN have been selected from the candidate sets $= [0, 1, 3, 5, 7, 9]$ and $[1, 3, 5, 7, 9]$, respectively. The cross-validation results in both cases, $K = 5, 10$, lead to the same choice of the NH optimal value. Namely, $NH = 1$ for MLP, and $NH = 5$ for BNN. The last one displays slight differences with respect to the values $NH = 3, 7$, in the 10-fold cross-validation implementation. In the same way, we have selected the spread and bandwidth parameters in the RBF and GRNN procedures. Thus, after applying random K -fold validation, with $K = 5, 10$, the optimal values $\beta = 2.5$, and $h = 0.05$ are obtained, from the candidate sets $[2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20]$ and $[0.05, 0.1, 0.2, 0.3, 0.5, 0.6, 0.7]$, respectively. As commented, in the first category, the best SVR performance corresponds to the linear case, finding the hyperparameters that minimize the five-fold cross-validation loss (by using automatic hyperparameter optimization from *fitrsvm* MatLab function). While, from the projected data category, the best option corresponds to the Gaussian kernel based nonlinear SVR model fitting (applying the same option of automatic hyperparameter optimization, in the argument of *fitrsvm* MatLab function). In the implementation of GP we follow the same tuning procedure for model selection. In this case, for both categories, we have opted by Bayesian cross-validation optimization (in the hyperparameter optimization argument of the *fitrgp* MatLab function).

In all the results displayed, the SMAPE–MEAN (M.) and SMAPE–TOTAL (T.) have been computed as rank-based performance measures, for comparing the ML models tested, and the presented Classical and Bayesian semilinear estimation approaches.

Table 2: **FIRST DATA CATEGORY. Average SMAPEs multiplied by the factor 10^2** (10 running of random K -fold validation, $K = 5$)

SC	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1962	0.0845	0.0890	0.0709	0.0635	0.0592
C2	0.6150	0.1531	0.0711	0.0805	0.0782	0.0710
C3	0.1541	0.0479	0.0438	0.0338	0.0371	0.0325
C4	0.0984	0.0388	0.0190	0.0226	0.0214	0.0226
C5	0.2065	0.0555	0.0378	0.0414	0.0429	0.0397
C6	0.1585	0.0423	0.0227	0.0257	0.0266	0.0263
C7	0.4957	0.0786	0.0771	0.0707	0.0712	0.0587
C8	0.0804	0.0352	0.0226	0.0215	0.0247	0.0189
C9	0.7280	0.2170	0.1061	0.0866	0.0421	0.0475
C10	0.2208	0.0724	0.0751	0.0577	0.0541	0.0494
C11	0.1273	0.0618	0.0373	0.0460	0.0451	0.0385
C12	0.5237	0.1706	0.1441	0.1449	0.1126	0.1065
C13	0.3637	0.0717	0.0701	0.0671	0.0605	0.0480
C14	0.1359	0.0388	0.0307	0.0235	0.0220	0.0213
C15	0.6105	0.1705	0.1592	0.1228	0.1101	0.1121
C16	0.2479	0.0931	0.0926	0.0753	0.0665	0.0632
C17	0.0667	0.0414	0.0195	0.0255	0.0241	0.0246
M.	0.2958	0.0867	0.0657	0.0598	0.0531	0.0494
T.	5.0293	1.4731	1.1177	1.0166	0.9026	0.8400

6 Concluding remarks

Model empirical selection results hold in a broad-based sense, as displayed in Section 5.1. Specifically, these results have been tested for $K = 5$ and $K = 10$ in the random K -fold validation, leading to almost the same model ranking in the two categories. See Table 7 and Figure 7 for the first category, and Table 8 and Figure 8 for the second category. (In Table 8 and Figure 8, C–ARH(1) and B–ARH(1) respectively denote the Classical and Bayesian semilinear estimation). The random K -fold validation results displayed for the presented Classical and Bayesian semilinear approaches locate these methodologies after SVR model. Thus, the fifth and sixth positions are respectively occupied in that empirical ranking. In the first category, RBF and GP are almost equal, while in the second category slightly larger differences can be observed. On the other hand, the projected data category favors the GRNN, RBF and GP performance, while better results are obtained from the original data in the implementation of MLP, SVR and BNN. Note that, the best models in the above rankings, RBF and GP, display a better performance in the second category. Thus, we can conclude that the FDA preprocessing produces a remarkable

Fig. 3: FIRST DATA CATEGORY. Observed and estimated COVID-19 mortality log-risk curves, from the implementation of Generalized Regression Neural Network (GRNN), Multilayer Perceptron (MLP), Support Vector Regression (SVR), Bayesian Neural Network (BNN), Radial Basis Function Neural Network (RBF), and Gaussian Processes (GP)

Fig. 4: FIRST DATA CATEGORY. Observed and estimated COVID-19 mortality cumulative cases curves from the implementation of Generalized Regression Neural Network (GRNN), Multilayer Perceptron (MLP), Support Vector Regression (SVR), Bayesian Neural Network (BNN), Radial Basis Function Neural Network (RBF), and Gaussian Processes (GP)

Table 3: **FIRST DATA CATEGORY. Average SMAPEs multiplied by the factor 10^2** (10 running of random K -fold validation, $K = 10$)

SC	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1957	0.0777	0.0700	0.0594	0.0543	0.0554
C2	0.6132	0.1490	0.0663	0.0738	0.0680	0.0654
C3	0.1556	0.0473	0.0350	0.0303	0.0331	0.0304
C4	0.0971	0.0342	0.0135	0.0200	0.0182	0.0211
C5	0.2049	0.0457	0.0318	0.0370	0.0369	0.0372
C6	0.1572	0.0368	0.0177	0.0234	0.0233	0.0247
C7	0.4898	0.0698	0.0644	0.0590	0.0616	0.0588
C8	0.0804	0.0340	0.0171	0.0191	0.0211	0.0177
C9	0.7258	0.1976	0.0979	0.0812	0.0326	0.0437
C10	0.2191	0.0704	0.0556	0.0482	0.0471	0.0463
C11	0.1262	0.0530	0.0310	0.0395	0.0375	0.0355
C12	0.5228	0.1578	0.1341	0.1282	0.0940	0.0993
C13	0.3594	0.0647	0.0576	0.0579	0.0533	0.0458
C14	0.1345	0.0366	0.0209	0.0204	0.0194	0.0207
C15	0.6080	0.1523	0.1411	0.1141	0.0982	0.1039
C16	0.2464	0.0889	0.0709	0.0622	0.0568	0.0594
C17	0.0660	0.0370	0.0148	0.0222	0.0203	0.0227
M.	0.2942	0.0796	0.0553	0.0527	0.0456	0.0463
T.	5.0022	1.3528	0.9397	0.8959	0.7757	0.7879

Table 4: **SECOND DATA CATEGORY. Average SMAPEs multiplied by the factor 10^2** (10 running of random K -fold, $K = 5$)

SC	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1569	0.0958	0.0695	0.0604	0.0240	0.0337
C2	0.1836	0.1835	0.0697	0.0880	0.0286	0.0329
C3	0.1029	0.1309	0.0474	0.0491	0.0273	0.0284
C4	0.0433	0.0299	0.0171	0.0181	0.0133	0.0130
C5	0.0609	0.0524	0.0282	0.0264	0.0153	0.0159
C6	0.0259	0.0239	0.0133	0.0148	0.0130	0.0129
C7	0.3783	0.2136	0.1018	0.0963	0.0309	0.0439
C8	0.0774	0.0467	0.0315	0.0320	0.0292	0.0207
C9	0.4968	0.2926	0.1458	0.1329	0.0254	0.0417
C10	0.1710	0.0935	0.0541	0.0489	0.0277	0.0316
C11	0.1556	0.0914	0.0456	0.0441	0.0221	0.0247
C12	0.3759	0.2235	0.1509	0.1396	0.0402	0.0560
C13	0.2894	0.1599	0.0903	0.0775	0.0259	0.0368
C14	0.0375	0.0250	0.0124	0.0153	0.0114	0.0109
C15	0.3646	0.2410	0.1378	0.1297	0.0334	0.0573
C16	0.1792	0.0828	0.0677	0.0578	0.0292	0.0344
C17	0.0900	0.0724	0.0216	0.0256	0.0129	0.0144
M.	0.1876	0.1211	0.0650	0.0622	0.0241	0.0299
T.	3.1893	2.0589	1.1047	1.0567	0.4100	0.5090

improvement, even in the communities, like, Navarra and Galicia, where the worst results are obtained with all the ML models tested. Particularly, this improvement is more significative at Murcia Community. In that sense, one can also observe that the spatial component of the data also affects the perfor-

Table 5: **SECOND DATA CATEGORY. Average SMAPEs multiplied by the factor 10^2** (from 10 running of random K -fold, $K = 10$)

SC	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1545	0.0983	0.0666	0.0573	0.0234	0.0312
C2	0.1844	0.1730	0.0660	0.0749	0.0277	0.0301
C3	0.1029	0.1192	0.0481	0.0452	0.0273	0.0274
C4	0.0432	0.0286	0.0165	0.0158	0.0124	0.0123
C5	0.0610	0.0476	0.0258	0.0248	0.0144	0.0149
C6	0.0260	0.0217	0.0133	0.0140	0.0124	0.0125
C7	0.3750	0.2026	0.1095	0.0924	0.0307	0.0399
C8	0.0764	0.0482	0.0305	0.0300	0.0262	0.0187
C9	0.4894	0.3198	0.1753	0.1212	0.0229	0.0372
C10	0.1680	0.0815	0.0521	0.0462	0.0252	0.0290
C11	0.1537	0.0839	0.0436	0.0397	0.0199	0.0219
C12	0.3689	0.2558	0.1505	0.1249	0.0401	0.0490
C13	0.2848	0.1582	0.0968	0.0792	0.0240	0.0320
C14	0.0367	0.0226	0.0120	0.0143	0.0106	0.0104
C15	0.3618	0.2264	0.1201	0.1227	0.0317	0.0522
C16	0.1773	0.0835	0.0651	0.0545	0.0264	0.0318
C17	0.0884	0.0623	0.0210	0.0231	0.0125	0.0136
M.	0.1854	0.1196	0.0655	0.0577	0.0228	0.0273
T.	3.1524	2.0333	1.1129	0.9801	0.3877	0.4642

Table 6: **SECOND DATA CATEGORY. Average SMAPEs multiplied by the factor 10^2** (10 running of random K -fold, $K = 5, 10$)

SC	B. k5	C. k5	B. k10	C. k10	B. est.	C. est.
C1	0.0781	0.0645	0.0750	0.0623	0.0726	0.0611
C2	0.1227	0.0952	0.1188	0.0944	0.1152	0.0940
C3	0.0797	0.0598	0.0767	0.0595	0.0736	0.0591
C4	0.0221	0.0145	0.0209	0.0145	0.0197	0.0144
C5	0.0302	0.0243	0.0291	0.0234	0.0282	0.0227
C6	0.0146	0.0125	0.0142	0.0123	0.0138	0.0120
C7	0.1966	0.1089	0.1850	0.1075	0.1736	0.1065
C8	0.0524	0.0349	0.0494	0.0346	0.0465	0.0347
C9	0.3947	0.2228	0.3709	0.2236	0.3479	0.2257
C10	0.0512	0.0364	0.0490	0.0348	0.0470	0.0335
C11	0.0637	0.0346	0.0600	0.0337	0.0565	0.0330
C12	0.2389	0.1883	0.2317	0.1866	0.2231	0.1866
C13	0.1591	0.0938	0.1488	0.0896	0.1388	0.0840
C14	0.0155	0.0133	0.0150	0.0131	0.0146	0.0127
C15	0.1624	0.1279	0.1555	0.1232	0.1504	0.1204
C16	0.0666	0.0482	0.0638	0.0460	0.0612	0.0441
C17	0.0323	0.0203	0.0304	0.0202	0.0287	0.0201
M.	0.1047	0.0706	0.0997	0.0694	0.0948	0.0685
T.	1.7807	1.2002	1.6942	1.1792	1.6115	1.1647

mance of the ML models tested through the two categories distinguished (e.g., the best performance overall ML models tested is obtained at Islas Baleares, Castilla-León and La Rioja, and the worst at Navarra and Galicia). Remark

Fig. 5: SECOND DATA CATEGORY. Observed and estimated COVID-19 mortality log-risk curves, from the implementation of Generalized Regression Neural Network (GRNN), Multilayer Perceptron (MLP), Support Vector Regression (SVR), Bayesian Neural Network (BNN), Radial Basis Function Neural Network (RBF), Gaussian Processes (GP), and Classical and Bayesian semilinear estimation

Fig. 6: SECOND DATA CATEGORY. Observed and estimated COVID-19 mortality cumulative cases curves, from the implementation of Generalized Regression Neural Network (GRNN), Multilayer Perceptron (MLP), Support Vector Regression (SVR), Bayesian Neural Network (BNN), Radial Basis Function Neural Network (RBF), Gaussian Processes (GP), and Classical and Bayesian semilinear estimation

also the outperformance of the Classical and Bayesian semilinear approaches at the Islas Baleares and Murcia Communities.

Note also that, the rank-based performance measure applied only refers to estimation accuracy, and other features of the probability distribution of the estimators have not been investigated yet. Indeed, the presented approach offers more stable, and robust estimations against the random selection of the training and validation samples. Thus, an improved model ranking position could be occupied by our approach if variability in that sense is penalized. Another subject to be addressed in a near future is the optimal selection of the test functions $\{\psi_{p,\rho_p}, p = 1, \dots, P\}$ involved in the formulation of the generalized least-squares loss function (7), and in the COVID-19 mortality risk assessment from equation (17). Our final goal would be to obtain an optimal design, from a set of candidates, of the test function family, providing the spatial sample information required, for spatiotemporal ML, and Classical and Bayesian semilinear forecasting, in a multivariate functional data framework (see, e.g., [6] regarding optimal selection of time intervals, or, instants, in the implementation of SVR from a multivariate functional data set).

Table 7: **MODEL RANKING FOR THE FIRST CATEGORY (10-FOLD CV)**

O.D.C.	T.-10-fold-SMAPE
RBF	0.7757(10^{-2})
GP	0.7879(10^{-2})
BNN	0.8959(10^{-2})
SVR	0.9397(10^{-2})
MLP	1.3528(10^{-2})
GRNN	5.0022(10^{-2})

Table 8: **MODEL RANKING FOR THE SECOND CATEGORY (10-FOLD CV)**

P.D.C.	T.-10-fold-SMAPE
RBF	0.3877(10^{-2})
GP	0.4642(10^{-2})
BNN	0.9801(10^{-2})
SVR	1.1129(10^{-2})
C-ARH(1)	1.1792(10^{-2})
B-ARH(1)	1.6942(10^{-2})
MLP	2.0333(10^{-2})
GRNN	3.1524(10^{-2})

Fig. 7: **ML MODEL RANKING.** T.-10-fold-SMAPE in green line for the first category (O.D.C. category), and red line for the second category (P.D.C category), and T.-5-fold-SMAPE in black line for the first category (O.D.C category), and blue line for the second category (P.D.C category)

Fig. 8: **JOINT MODEL RANKING.** T.-10-fold-SMAPE (red line) and T.-5-fold-SMAPE (blue line) in the second category (P.D.C category)

Acknowledgements

This work has been supported in part by project PGC2018-099549-B-I00 of the Ministerio de Ciencia, Innovación y Universidades, Spain (co-funded with FEDER funds).

We also thank *Unidad de Transferencia del IMUS*, and the organizer, Professor Emilio Carrizosa, by the seminar *Matemáticas y la COVID*, that constitutes the germinating seed in the development of the present work.

A statement potential conflicts

There are not conflicts of interest in the developed methodology, and the derived results in this paper, on Statistical Analysis of COVID-19 Mortality in Spain.

Appendix

Bayesian and Classical estimation of the autocovariance operator of the innovations defining the mortality residual log-risk process

In the Bayesian componentwise estimation of the autocovariance operator of the residual log-risk innovation process, ν_t , we can apply the approach presented in [9]. Specifically, for each $k \geq 1$, consider $\tau_k = \frac{1}{2\lambda_k(R_0^\nu)}$, with $\lambda_k(R_0^\nu)$ being the k th eigenvalue of $R_0^\nu = E[\nu_t \otimes \nu_t] = E[\nu_0 \otimes \nu_0]$, for every $t = 1, \dots, T$, $T \geq 2$. When the eigenvectors $\{\phi_{k,\nu}, k \geq 1\}$ of R_0^ν are known, we define the loss function, given by the likelihood, as

$$\begin{aligned} L(\varepsilon_{1k}, \dots, \varepsilon_{Tk}, \tau_k) &= C\tau_k^{T/2} \exp\left(-\tau_k \sum_{i=1}^T [\varepsilon_t(\phi_{k,\nu}) - \rho(\varepsilon_{t-1})(\phi_{k,\nu})]^2\right) \\ &= C\tau_k^{T/2} \exp\left(-\tau_k \sum_{i=1}^T [\nu_t(\phi_{k,\nu})]^2\right), \quad k \geq 1, \end{aligned} \quad (20)$$

where $\{\nu_t(\phi_{k,\nu}), t = 1, \dots, T\}$ are independent and identically distributed zero-mean Gaussian random variables with variance $\frac{1}{2\tau_k}$, and $\varepsilon_{tk} = \varepsilon_t(\phi_{k,\nu}) = \langle \varepsilon_t, \phi_{k,\nu} \rangle_H$, for each $k \geq 1$.

In equation (20), as conjugate prior for parameter $\tau_k = \frac{1}{2\lambda_k(R_0^\nu)}$, let us consider $\tau_k \sim \Gamma(g_k, \beta_k)$ (with β_k denoting the rate parameter), for $g_k > 2$, $k \geq 1$. Hence, $E\left[\frac{1}{\tau_k}\right] = \frac{\beta_k}{g_k - 1}$. For $k \geq 1$, the posterior density of τ_k , given $x_{ik} = \nu_i(\phi_{k,\nu})$, $i = 1, \dots, T$, is

$$\mathcal{C} \left(\sum_{i=1}^T x_{ik}^2 \right) \tau_k^{T/2 + g_k - 1} \exp\left(-\tau_k \left(\beta_k + \sum_{i=1}^T x_{ik}^2 \right)\right).$$

That is, the posterior is a Gamma probability distribution with parameters $T/2 + g_k$ and $(\beta_k + \sum_{i=1}^T x_{ik}^2)$.

For the loss function defined in terms of the squared error, the Bayes estimator of $2\lambda_k(R_0^\nu) = 1/\tau_k$ is the posterior expectation of $1/\tau_k$, given by $(\beta_k + \sum_{i=1}^T x_{ik}^2)/(T/2 + g_k - 1)$, for each $k \geq 1$. Thus, the Bayes estimator $\widetilde{\lambda}_{kT}(R_0^\nu)$ of $\lambda_k(R_0^\nu) = 1/2\tau_k$ is

$$\widetilde{\lambda}_{kT}(R_0^\nu) = \frac{\beta_k + \sum_{i=1}^T x_{ik}^2}{T + 2g_k - 2}, \quad k \geq 1 \quad (21)$$

(see [29], p.236). Hence, from (21), the Bayesian componentwise estimator of R_0^ν is computed as follows:

$$\widetilde{R}_{0,T}^\nu = \sum_{k=1}^{\infty} \widetilde{\lambda}_{kT}(R_0^\nu) [\phi_{k,\nu} \otimes \phi_{k,\nu}]. \quad (22)$$

In the classical approach, an unbiased estimator of $\lambda_k(R_0^\nu)$ is obtained from the identity

$$\bar{\lambda}_{kT}(R_0^\nu) = \frac{1}{T} \sum_{i=1}^T x_{ik}^2 = \frac{1}{T} \sum_{i=1}^T [\nu_i(\phi_{k,\nu})]^2, \quad k \geq 1. \quad (23)$$

Thus, the classical componentwise estimator of R_0^ν is defined as

$$\bar{R}_{0,T}^\nu = \sum_{k=1}^{\infty} \bar{\lambda}_{kT}(R_0^\nu) [\phi_{k,\nu} \otimes \phi_{k,\nu}]. \quad (24)$$

In practice, in the computation of equations (22) and (24), when the eigenvectors $\{\phi_{k,\nu}, k \geq 1\}$ are unknown, one can consider their consistent approximation in terms of the empirical eigenvectors $\{\phi_{k,T,\nu}, k \geq 1\}$, under the conditions assumed in Corollary 4.1, Lemma 4.3, and Theorem 4.8 in [8]. Indeed, under the conditions in Theorem 1 of [18], and the referred setting of conditions in [8], for each $k \geq 1$, $Y_t(\phi_{k,\nu}) - \widehat{\rho}(Y_{t-1})(\phi_{k,\nu})$ and $Y_t(\phi_{k,T,\nu}) - \widehat{\rho}(Y_{t-1})(\phi_{k,T,\nu})$, consistently approximate $x_{tk} = \nu_t(\phi_{k,\nu})$, for $t = 1, \dots, T$, in the computation of equations (21) and (23), for the cases where the eigenvectors $\{\phi_{k,\nu}, k \geq 1\}$ are known and unknown, respectively.

Note also that, under the conditions assumed in [9], in particular, when, for every $k \geq 1$, $\frac{\beta_k}{2(g_k-1)+1} \leq \lambda_k(R_0^\nu)$, we obtain the following equivalent asymptotic behavior in the Hilbert-Schmidt operator norm $\|\cdot\|_{\mathcal{S}(H)}$, for both estimation approaches:

$$\lim_{T \rightarrow \infty} TE \left\| \widetilde{R}_{0,T}^\nu - R_0^\nu \right\|_{\mathcal{S}(H)}^2 = \lim_{T \rightarrow \infty} TE \left\| \bar{R}_{0,T}^\nu - R_0^\nu \right\|_{\mathcal{S}(H)}^2 = 2 \sum_{k=1}^{\infty} [\lambda_k(R_0^\nu)]^2.$$

In the practical implementation of the Bayesian componentwise estimator (22) in Sections 4–5, we have considered the prior probability distribution

$\tau_k \sim \Gamma(g_k, \beta_k)$ with $g_k = 6$, and $\beta_k = (2g_k - 1)\lambda_{K(T),T}(\hat{R}_{0,T}^\nu)/10$ for each $k \geq 1$, where $\lambda_{K(T),T}(\hat{R}_{0,T}^\nu)$ denotes the $K(T)$ th empirical eigenvalue of the autocovariance operator of the innovation process defining the residual log-risk process, and $K(T)$ represents the truncation parameter value (see [9]).

Additional Outputs in the Spatiotemporal Log-Gaussian Analysis of COVID-19 Mortality Risk

The original COVID-19 mortality cumulative cases curves, and its interpolation and B -spline smoothing are shown in Figure 9. The corresponding COVID-19 mortality risk curves are also displayed in Figure 10.

Fig. 9: Observed, and Interpolated and cubic B-spline smoothed COVID-19 mortality cumulative cases curves at the 17 Spanish Autonomous Communities, since March, 8 to May, 13, 2020

The least-squares parameter estimates computed in the spatial heterogeneous nonlinear parametric trend model fitting are shown in Tables 9–10.

Finally, the COVID-19 mortality log-risk estimations, based on spatial heterogeneous linear and nonlinear trend model fitting, are compared in Figure 11, under a Classical (C) and Bayesina (B) approximation of the functional residuals, considering the truncation parameter value $K(265) = 9$.

Fig. 10: Observed COVID-19 mortality risk curves at the 17 Spanish Autonomous Communities, since March, 8 to May, 13, 2020

Table 9: NONLINEAR TREND MODEL FITTING. Parameter estimates $\hat{A}_k(\cdot)$, $k = 1, \dots, 6$, at the 17 Spanish Communities

SC/AE	$\hat{A}_1(\cdot)$	$\hat{A}_2(\cdot)$	$\hat{A}_3(\cdot)$	$\hat{A}_4(\cdot)$	$\hat{A}_5(\cdot)$	$\hat{A}_6(\cdot)$
C1	3.6343	-0.4814	-0.0075	-0.0258	0.0189	0.0193
C2	3.4345	-0.3923	0.0416	0.0265	-0.0709	-0.0572
C3	3.2031	-0.1364	-0.0088	0.0221	0.0430	0.0289
C4	3.1445	-0.1118	0.0041	0.0337	0.0062	0.0072
C5	3.1015	-0.0693	-0.0345	0.0352	-0.0112	0.0003
C6	3.1347	-0.1397	0.0020	0.0300	-0.0061	-0.0002
C7	4.0591	-0.5487	-0.0907	0.0951	0.0992	0.0842
C8	3.8032	-0.5500	-0.1007	0.0633	0.0139	0.0277
C9	4.5095	-0.7435	-0.1134	0.1809	0.2231	0.2026
C10	3.6321	-0.4685	-0.0540	0.0384	-0.0152	0.0011
C11	3.2967	-0.2274	-0.0083	0.0553	0.0250	0.0240
C12	3.3454	-0.2122	-0.0927	-0.0330	0.0724	0.0679
C13	4.8419	-0.6790	-0.2455	0.0311	0.0554	0.0667
C14	3.0941	-0.1037	0.0210	0.0141	-0.0016	0.0041
C15	3.2877	-0.2598	-0.0524	0.0842	-0.0423	-0.0348
C16	3.6870	-0.4302	-0.0086	0.0078	-0.0027	-0.0017
C17	3.2197	-0.2071	0.0162	0.0079	0.0206	0.0110

A Brief Introduction to our FDA Implementation of ML models

The implemented ML models are introduced here. Their reformulation in the multivariate FDA framework adopted is also provided.

Multilayer Perceptron (MLP)

MLP shares the philosophy of nonlinear regression, in terms of a link function g , the hidden node output, defining the following approximation of the

Table 10: NONLINEAR TREND MODEL FITTING. Parameter estimates $\hat{B}_k(\cdot)$, $k = 1, \dots, 6$, at the 17 Spanish Communities

SC/BE	B_1	B_2	B_3	B_4	B_5	B_6
C1	0	-0.0052	-0.1330	-0.0123	0.0064	-0.0195
C2	0	-0.0367	-0.0998	-0.0462	-0.0343	-0.0107
C3	0	-0.0531	-0.0074	-0.0142	-0.0003	0.0020
C4	0	-0.0074	-0.0284	-0.0151	-0.0092	0.0012
C5	0	0.0433	-0.0438	-0.0116	-0.0118	0.0046
C6	0	0.0018	-0.0174	-0.0068	-0.0089	0.0000
C7	0	-0.0365	-0.2451	-0.1791	-0.0820	0.0026
C8	0	0.0953	-0.2389	-0.0431	-0.0313	-0.0045
C9	0	-0.1587	-0.4054	-0.2269	-0.1010	0.0047
C10	0	0.1118	-0.1579	-0.0458	-0.0418	-0.0220
C11	0	0.0754	-0.1138	-0.0166	-0.0048	0.0072
C12	0	-0.1104	-0.1338	0.1330	0.0761	-0.0017
C13	0	0.4654	-0.1302	-0.1602	-0.1061	-0.0038
C14	0	0.0355	-0.0560	0.0119	0.0025	-0.0044
C15	0	-0.0187	-0.0021	-0.0897	-0.0562	0.0134
C16	0	0.0025	-0.0707	-0.0638	-0.0439	-0.0267
C17	0	0.0389	-0.0270	-0.0174	-0.0006	0.0019

Fig. 11: Observed and estimated COVID-19 mortality log-risk curves, at the 17 Spanish Communities, since March, 8 to May, 13, 2020, under a spatial heterogeneous linear and nonlinear trend model fitting, applying a classical (C) and Bayesian (B) componentwise estimation of the residual correlation, under the truncation parameter value $K(T) = 9$

response:

$$\hat{\mathbf{y}} = \eta_0 + \sum_{k=1}^{NH} \eta_k g(\beta_k^T \mathbf{x}), \quad (25)$$

from the *input* vector $\mathbf{x}^T = (\mathbf{1}, \mathbf{x})$ augmented with 1, and the weight vector β_k associated with the k th hidden node, $k = 1, \dots, NH$, defining $\beta =$

$(\beta_1, \dots, \beta_{NH})^T$. The hidden node outputs are also weighted by the components of $(\eta_0, \dots, \eta_{NH})$. In this context, $\hat{\mathbf{y}}$ is usually referred as the network output. MLP allows the approximation of any given continuous function on a compact set as close as arbitrarily, from a given network with a finite number of hidden nodes. Optimization algorithms are applied to obtain the weights from the least-squares loss function. In our implementation of MLP, in an FDA framework, the input vector is defined by

$$\mathbf{x}^T = (1_{\mathcal{D}}(\cdot), \ln(\lambda_t)(\phi_1), \dots, \ln(\lambda_t)(\phi_{K(T)}), \dots, \ln(\lambda_{t-p})(\phi_{K(T)})) ,$$

with $1_{\mathcal{D}}(\cdot)$ denoting the indicator function of the domain $\mathcal{D} \subset \mathbb{R}^2$, the spatial support of the functions in H . Here, $\ln(\lambda_{t-j})(\phi_l)$, $l = 1, \dots, K(T)$, $j = 0, \dots, p$, denote the observed projected log-risk values, onto the autocovariance operator eigenvectors $\phi_1, \dots, \phi_{K(T)}$, at times $t, \dots, t-p$. Parameter $K(T)$ denotes the truncation achieved in our finite-dimensional approximation, and T the number of temporal nodes, or training data points. In practice, the theoretical eigenvectors are usually replaced by the empirical eigenvectors. Parameter $p+1$ provides the number of temporal lags incorporated in the prediction. In our case, since we have considered one-step-ahead forecasting, $p = 1$, the response

$$\mathbf{y}_{t+1} = (\ln(\lambda_{t+1})(\phi_1), \dots, \ln(\lambda_{t+1})(\phi_{K(T)})) \quad (26)$$

at time $t+1$, is approximated from the input vector

$$(\ln(\lambda_t)(\phi_1), \dots, \ln(\lambda_t)(\phi_{K(T)})) . \quad (27)$$

Hence, equation (25) can be rewritten as

$$\hat{\mathbf{y}}_{t+1} = \eta_0 + \sum_{k=1}^{NH} \eta_k g(\langle \beta_k, \mathbf{x} \rangle_H), \quad \beta_k, \mathbf{x} \in H, \quad k = 1, \dots, NH. \quad (28)$$

Usually, the logistic function $g(u) = \frac{1}{1+\exp(-u)}$ is considered.

Radial Basis Function Neural Network (RBF)

RBF works with node functions, depending on a center and scale parameters, fitting the local smoothness of the response. Specifically, from an initial blank network, the nodes are sequentially added, around the training pattern, until an acceptable error is reached. All the output layer weights are then recomputed using the least squares formula. Gaussian functions have been widely selected as node functions. Equation (27) defines the input vector in our implementation of RBF in an infinite-dimensional setting given by the formula

$$\mathbf{y} = \sum_{j=1}^{NH} \eta_j \exp\left(-\frac{\|\mathbf{x} - c_j\|_H^2}{\rho^2}\right), \quad (29)$$

where, as usual, the weight parameters η_j , $j = 1, \dots, NH$, define the linear combination of radial basis functions. Here, the scale parameter ρ , and the

functional location parameters $c_j \in H$, $j = 1, \dots, NH$, respectively provide the width, and the functional center of the node functions. In practice, a $K(T)$ -dimensional implementation is considered based on the truncation parameter $K(T)$, as given in (27).

Support Vector Regression (SVR)

SVR implementation involves a loss function leading to a balance between model complexity and precision (accurate prediction). A bias parameter b is also considered in its formulation as reflected in the following equation:

$$y = f(\mathbf{x}) = \beta^T \mathbf{x} + b, \quad (30)$$

where the loss function

$$\mathcal{L} = \frac{1}{2} \|\beta\|^2 + C \sum_{m=1}^M |y_m - f(x_m)|_\epsilon, \quad (31)$$

is considered. Here, x_m and y_m respectively denote the m th training input vector and the target output, for $m = 1, \dots, M$, and

$$|y_m - f(x_m)|_\epsilon = \max \{0, |y_m - f(x_m)| - \epsilon\}.$$

Thus, the errors below ϵ are not penalized. The solution to the optimization problem associated with the loss function (31) is given by

$$f(\mathbf{x}) = (\alpha_m^* - \alpha_m) \mathbf{x}^T \mathbf{x} + b, \quad (32)$$

where α_m^* and α_m are Lagrange multipliers. The support vectors are the training vectors giving nonzero Lagrange multipliers. A nonlinear version is obtained when the inner product $\mathbf{x}^T \mathbf{x}$ in (32) is replaced by $\mathcal{K}(\mathbf{x}^T \mathbf{x})$, with $\mathcal{K}$ being a kernel. A usual kernel choice is the Gaussian kernel.

In our implementation of this methodology in an FDA framework from equations (26)–(27), we will consider

$$\begin{aligned} f(\mathbf{x}) &= \langle \beta, \mathbf{x} \rangle_H + b, \quad \beta, b, \mathbf{x} \in H, \\ \mathcal{L} &= \frac{1}{2} \|\beta\|_H^2 + C \sum_{t=1}^T \sum_{l=1}^{K(T)} |\ln(\lambda_{t+1})(\phi_l) - f(\ln(\lambda_t)(\phi_l))|_{\epsilon_l}, \end{aligned} \quad (33)$$

where T denotes, as before, the number of temporal nodes, and for $l \geq 1$,

$$\begin{aligned} &|\ln(\lambda_{t+1})(\phi_l) - f(\ln(\lambda_t)(\phi_l))|_{\epsilon_l} \\ &= \max \{0, |\ln(\lambda_{t+1})(\phi_l) - f(\ln(\lambda_t)(\phi_l))| - \epsilon_l\}. \end{aligned} \quad (34)$$

Bayesian Neural Network (BNN)

The design of BNN involves Bayesian estimation, and the concept of regularization. The network parameters or weights are considered random variables, following a prior probability distribution. Smooth fits are usually favored in the selection of the prior probability of the weights to reduce model complexity. The posterior probability distribution of the weights is obtained after data are observed. The network prediction is then computed. Specifically, the optimal prediction is obtained by minimizing the following expression:

$$J = \nu E_{\mathbf{O}} + (1 - \nu) E_{\mathbf{W}}, \quad (35)$$

where $E_{\mathbf{O}}$ is the sum of the square errors in the network output, based on the posterior distribution of the parameters, $E_{\mathbf{W}}$ represents the sum of the squares of the weights or the network parameters, and $\nu \in (0, 1)$ denotes the regularization parameter. Let L be the number of parameters. An L -dimensional Gaussian prior probability distribution is usually assumed for the network parameters with zero-mean and variance-covariance matrix $\frac{1}{2(1-\nu)} I_{L \times L}$, with $I_{L \times L}$ denoting the $L \times L$ identity matrix. Thus,

$$p(\mathbf{w}) = \left[\frac{1 - \nu}{\pi} \right]^{L/2} \exp(-(1 - \nu) E_W(\mathbf{w})). \quad (36)$$

This prior puts more weight onto small network parameter values close to zero. The posterior probability density, given the observed data $\mathbf{O} = \mathbf{o}$, and the value ν of the regularization parameter, is defined as

$$p(\mathbf{w}/\mathbf{o}, \nu) = \frac{p(\mathbf{o}/\mathbf{w}, \nu) p(\mathbf{w}/\nu)}{p(\mathbf{o}/\nu)}. \quad (37)$$

Considering that the errors are also Gaussian distributed the conditional probability of the data $\mathbf{O}$ given the parameters ν and $\mathbf{w}$, is obtained as

$$p(\mathbf{o}/\mathbf{w}, \nu) = \left(\frac{\nu}{\pi} \right)^{M/2} \exp(-\nu E_{\mathbf{O}}(\mathbf{o})), \quad (38)$$

where M denotes the number of training data points. From equations (36)–(38),

$$p(\mathbf{w}/\mathbf{o}, \nu) = c \exp(-J(\mathbf{w}, \mathbf{o}, \nu)), \quad (39)$$

where c denotes the normalizing constant. The conditional probability of the parameter ν given the observed data $\mathbf{O} = \mathbf{o}$ is also computed under a Bayesian framework as

$$p(\nu/\mathbf{o}) = \frac{P(\mathbf{o}/\nu) p(\nu)}{p(\mathbf{o})}. \quad (40)$$

Equations (39) and (40) are maximized to obtain the optimal weights and the regularization parameter ν , respectively.

In our FDA implementation from (26)–(27), the corresponding optimization problem is formulated by conditioning to the projected data. Note that,

in the selected Gaussian prior probability framework, the error projections are also Gaussian, and our choice of the function basis, diagonalizing the empirical autocovariance operator of the errors, leads to a projected error vector with independent Gaussian components, which is also normalized, in terms of the empirical eigenvalues as described in the preprocessing procedure in Section 5.1. Hence, optimization from equations (39) and (40) can be implemented in a similar way.

Generalized Regression Neural Network (GRNN)

GRNN is based on kernel regression. The kernel estimator is computed from the weighted sum of the observed responses, or target outputs associated with the training data points in a neighborhood of the objective data point x , where prediction must be computed. The training data points are selected in the vicinity of the given objective point x . Specifically, the following formula is applied in the approximation of the response value at the point x :

$$\hat{y}(x) = \sum_{j=1}^M \frac{\mathcal{K}\left(\frac{\|x-x_j\|}{h}\right)}{\sum_{l=1}^M \mathcal{K}\left(\frac{\|x-x_l\|}{h}\right)} y_j, \quad (41)$$

where y_j is the target output for training data point x_j , for $j = 1, \dots, M$, and $\mathcal{K}$ is the kernel function. Usually an isotropic rapidly decreasing kernel function is considered, e.g., the Gaussian kernel $\mathcal{K}(u) = \exp(-u^2/2) / \sqrt{2\pi}$ constitutes a common choice. The bandwidth parameter h defines the smoothness of the fit. Thus, h controls the size of the smoothing region. Hence, large values of h correspond to a stronger smoothing than the smallest values allowing a larger degree of local variation. Our FDA formulation, from (26)–(27), follows straightforward by considering the H -norm $\|\cdot\|_H$ in (41), instead of the Euclidean norm $\|\cdot\|$.

Gaussian Processes (GP)

A good performance is usually observed in the implementation of GP regression, based on the multivariate normal probability distribution assumption, characterizing the observed responses at the different training data points. Specifically, we consider the observation model

$$\mathbf{Y} = \mathbf{Z} + \boldsymbol{\epsilon}, \quad (42)$$

where the additive noise vector $\boldsymbol{\epsilon}$ has independent components. A multivariate normal distribution of the random vector $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, with covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}}$ is assumed. This matrix provides the variances and covariances between the output vector components in the training set. Applying Bayes rule we obtain the solution to the inverse problem (42). Thus, the estimation of $\mathbf{Z}$, for a given input vector $\mathbf{x}_*$, is obtained as

$$\hat{\mathbf{Z}}_{\mathbf{x}_*} = E[\mathbf{Z}/\mathbf{Y} = \mathbf{x}_*] = \boldsymbol{\Sigma}_{\mathbf{Y},\mathbf{Z}} [\boldsymbol{\Sigma}_{\mathbf{Z}} + \sigma_{\boldsymbol{\epsilon}}^2 \mathbf{I}]^{-1} \mathbf{Y}. \quad (43)$$

Here, we consider an alternative multivariate FDA interpretation, in terms of a training test function set (instead of a training data point set, based on temporal nodes). Hence, $\Sigma_{\mathbf{Z}}$ is replaced by the matrix autocovariance operator $R_{\mathbf{Z}}$ of the output, and $\Sigma_{\mathbf{Y},\mathbf{Z}}$ is replaced by the matrix cross-covariance operator $R_{\mathbf{Y},\mathbf{Z}}$, both evaluated at the training test function set. In this case, σ_{ϵ}^2 denotes the functional variance or trace of the autocovariance operator, characterizing the Gaussian infinite-dimensional distribution of the $\mathcal{H}$ -valued independent components of the zero-mean observation noise vector ϵ . Hence, the identity matrix $\mathbf{I}$ is replaced by the identity operator $\mathbf{I}_{\mathcal{H}^P}$ on $\mathcal{H}^P$, with P denoting the number of training test functions. That is,

$$\hat{\mathbf{Z}}_{\star} = E \left[\hat{\mathbf{Z}}_{\star} / \mathbf{Y} = \mathbf{x}_{\star} \right] = R_{\mathbf{Y},\mathbf{Z}} \left[R_{\mathbf{Z}} + \sigma_{\epsilon}^2 \mathbf{I}_{\mathcal{H}^P} \right]^{-1} \mathbf{Y}. \quad (44)$$

ML Estimation Results in the Two Categories

The estimation results obtained from the implementation of the above ML models in the first category are displayed in Table 11. The SMAPEs from the projected data are shown in Table 12.

Table 11: **FIRST CATEGORY. SMAPEs multiplied by factor 10^2 .**

AC	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1964	0.0611	0.0665	0.0535	0.0483	0.0490
C2	0.6117	0.1172	0.0588	0.0678	0.0609	0.0567
C3	0.1565	0.0375	0.0328	0.0284	0.0300	0.0273
C4	0.0969	0.0314	0.0129	0.0191	0.0167	0.0185
C5	0.2044	0.0427	0.0290	0.0356	0.0334	0.0328
C6	0.1571	0.0319	0.0161	0.0230	0.0217	0.0218
C7	0.4889	0.0545	0.0619	0.0583	0.0569	0.0503
C8	0.0808	0.0273	0.0161	0.0180	0.0189	0.0153
C9	0.7231	0.1976	0.0850	0.1102	0.0318	0.0352
C10	0.2185	0.0526	0.0532	0.0446	0.0428	0.0415
C11	0.1255	0.0487	0.0298	0.0350	0.0338	0.0316
C12	0.5216	0.1666	0.1210	0.1022	0.0870	0.0887
C13	0.3584	0.0592	0.0548	0.0613	0.0490	0.0412
C14	0.1342	0.0286	0.0201	0.0202	0.0182	0.0180
C15	0.6064	0.1470	0.1307	0.0931	0.0864	0.0927
C16	0.2456	0.0681	0.0674	0.0573	0.0514	0.0532
C17	0.0655	0.0356	0.0147	0.0207	0.0181	0.0200
M.	0.2936	0.0710	0.0512	0.0499	0.0415	0.0408
T.	4.9916	1.2078	0.8707	0.8480	0.7053	0.6936

References

1. O. O. Aalen, O. Borgan and H. K. Gjessing (2008). *Survival and event history analysis: a process point of view*. Springer Science & Business Media, New-York.

Table 12: **SECOND CATEGORY. SMAPEs multiplied by factor 10^2 .**

CA	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1526	0.0857	0.0592	0.0501	0.0223	0.0305
C2	0.1831	0.1393	0.0619	0.0606	0.0248	0.0285
C3	0.1024	0.0804	0.0457	0.0374	0.0268	0.0278
C4	0.0431	0.0212	0.0157	0.0154	0.0117	0.0120
C5	0.0610	0.0362	0.0242	0.0240	0.0133	0.0138
C6	0.0260	0.0167	0.0125	0.0135	0.0118	0.0121
C7	0.3734	0.1301	0.0914	0.0737	0.0298	0.0398
C8	0.0762	0.0435	0.0290	0.0238	0.0244	0.0180
C9	0.4850	0.2467	0.1470	0.0818	0.0199	0.0360
C10	0.1663	0.0607	0.0460	0.0421	0.0236	0.0276
C11	0.1533	0.0540	0.0394	0.0383	0.0180	0.0209
C12	0.3641	0.2325	0.1320	0.0887	0.0369	0.0480
C13	0.2827	0.1350	0.0707	0.0699	0.0215	0.0304
C14	0.0361	0.0197	0.0117	0.0121	0.0102	0.0104
C15	0.3590	0.1843	0.1226	0.1049	0.0290	0.0506
C16	0.1759	0.0754	0.0566	0.0489	0.0243	0.0302
C17	0.0878	0.0352	0.0190	0.0219	0.0122	0.0134
N.	0.1840	0.0939	0.0579	0.0475	0.0212	0.0265
T.	3.1282	1.5966	0.9845	0.8070	0.3605	0.4497

2. C. Abboud, O. Bonnefon, E. Parent and S. Soubeyrand (2019). Dating and localizing an invasion from post-introduction data and a coupled reaction–diffusion–absorption model. *Journal of Mathematical Biology* **79**, 765–789.
3. E. Alpaydin (2004). *Introduction to Machine Learning*. MIT Press, Cambridge, MA.
4. H. Anderson and T. Britton (2000.). *Stochastic epidemic models and their statistical analysis*. Springer–Verlag, New–York.
5. E. E. Beretta, T. Hara, W. Ma and Y. Takeuchi (2001). Global asymptotically stability of an SIR epidemic model with distributed time delay. *Nonlinear Anal Theory Methods Appl* **47**, 4107–4115.
6. R. Blanquero, E. Carrizosa, A. Jiménez–Cordero and B. Martín–Barragán (2020). Selection of time instants and intervals with support vector regression for multivariate functional data <https://www.researchgate.net/publication/327552293>.
7. B. M. Bolker and B. Grenfell (1996). Impact of vaccination on the spatial correlation and persistence of measles dynamics. *Proceedings of the National Academy of Sciences* **93**, 12648–12653.
8. D. Bosq (2000). *Linear processes in function spaces*. Lecture notes in statistics **149**. Springer, New–York.
9. D. Bosq and M. D. Ruiz–Medina (2014). Bayesian estimation in a high dimensional parameter framework. *Electron J Statist* **8**, 1604–1640.
10. D. L. Chao, J. D. Bloom, B. F. Kochin, R. Antia and Jr, I. M. Longini (2012). The global spread of drug-resistant influenza. *Journal of the Royal Society Interface* **9**, 648–656.
11. Z. Du, X. Xu, Y. Wu, L. Wang, L. A. Cowling and B. J. Meyers (2020). Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg. Infect. Dis.* **26**(6).
12. J. Dushoff, J. Plotkin, S. Levin and D. Earn (2004). Dynamical resonance can account for seasonality of influenza epidemics. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16915–16916.
13. M. Elhia, A. Laaroussi, M. Rachik, Z. Rachik and E. Labriji (2014). Global stability of a susceptible–infected–recovered (SIR) epidemic model with two infectious stages and treatment. *Int J Sci Res* **3**, 114–121.
14. T. R. Fleming and D. P. Harrington (1991). *Counting processes and survival analysis*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New–York.

15. L. N. Guin and P. K. Mandal (2014). Spatiotemporal dynamics of reaction–diffusion models of interacting populations. *Appl Math Model* **38**, 4417–4427.
16. T. Hastie, R. Tibshirani and J. Friedman (2001). *The elements of statistical learning*. Springer Series in Statistics. Springer–Verlag, New–York.
17. A. Huppert and G. Katriel (2013). Mathematical modelling and prediction in infectious disease epidemiology. *Clin Microbiol Infect* **19**, 999–1005.
18. A.V. Ivanov, N.N. Leonenko, M.D. Ruiz Medina, and B.M. Zhurakovsky (2015). Estimation of harmonic component in regression with cyclically dependent errors. *Stastics: A Journal of Theoretical and Applied Statistics* **49**, 156–186.
19. B. Ivorra, M. R. Ferrández, M. Vela-Pérez and A. M. Ramos (2020). Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Commun Nonlinear Sci Numer Simulat* **88**, 105–303.
20. B. Ivorra, A. M. Ramos and D. Ngom (2015). Be-CoDiS: A mathematical model to predict the risk of human diseases spread between countries. Validation and application to the 2014 ebola virus disease epidemic. *Bull Math Biol* **77**, 1668–1704.
21. C. Ji, D. Jiang and N. Shi (2012). The behavior of an SIR epidemic model with stochastic perturbation. *Stoch Anal Appl* **30**, 755–773.
22. M. J. Keeling, D. A. Rand and A. J. Morris (1997). Correlation models for childhood epidemics. *Proceedings of the Royal Society of London* **264**, 1149–1156.
23. M. J. Keeling and P. Rohani (2008). *Modeling infectious diseases in humans and animals*. Princeton University Press, Princeton.
24. W. Kermack and A. McKendrick (1927). Contributions to the mathematical theory of epidemics - I. *Proceedings of the Royal Society of Edinburgh A* **115**, 700–721.
25. M. A. Khan and A. Atangana (2020). Modeling the dynamics of novel coronavirus (2019-nCov) with fractional derivative. *Alex. Eng. J.* doi.org/10.1016/j.aej.2020.02.033.
26. A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds and S. Funk et al. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis*. doi.org/10.1016/S1473-3099(20)30144-4.
27. Y. A. Kuznetsov and C. Piccardi (1994). Bifurcation analysis of periodic SEIR and SIR epidemic models. *J Math Biol* **32**, 109–121.
28. A. E. Laaroussi, M. Rachik and M. Elhia (2018). An optimal control problem for a spatiotemporal SIR model *Int. J. Dynam. Control* **6**, 384–397.
29. E. L. Lehmann and G. Casella (1998). *Theory of point estimation*. Springer (2th Rdition), New York.
30. C. Malesios, N. Demiris, P. Kostoulas, K. Dadousis, T. Koutroumanidis and Z. Abas (2016). Spatio-temporal modelling of foot-and-mouth disease outbreaks. *Epidemiol. Infect.* **144**, 2485–2493.
31. C. C. McCluskey (2010). Complete global stability for an SIR epidemic model with delay distributed or discrete. *Nonlinear Anal Real World Appl* **11**, 55–59.
32. F. A. Milner and R. Zhao (2008). SIR model with directed spatial diffusion. *Math Popul Stud* **15**, 160–181.
33. A. S. Mugglin, N. Cressie and I. Gemmell (2002). Hierarchical statistical modelling of influenza epidemic dynamics in space and time. *Statist. Med.* **21**, 2703–2721.
34. H. Nishiura, N. M. Linton and A. R. Akhmetzhanov (2020). Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **93**, 284–286.
35. S. Pathak, A. Maiti and G. Samanta (2010). Rich dynamics of an SIR epidemic model. *Nonlinear Anal Model Control* **15**, 71–81.
36. A. Remuzzi and G. Remuzzi (2020). COVID-19 and Italy: what next? The Lancet. doi.org/10.1016/S0140-6736(20)30690-5
37. K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. Hyman et al. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th. *Infect Dis Modell* **5**, 256–263.
38. L. Roques and O. Bonnefon (2016). Modelling population dynamics in realistic landscapes with linear elements: A mechanistic-statistical reaction-diffusion approach. *PloS One* **11**(3):e0151217.
39. L. Roques, S. Soubeyrand and J. Rousselet (2011). A statistical-reaction-diffusion approach for analyzing expansion processes. *J Theor Biol.* **274**, 43–51.

40. M. Sekiguchi and E. Ishiwata (2010). Global dynamics of a discretized SIRS epidemic model with time delay. *J Math Anal Appl* **371**, 195–202.
41. E. Tornatore, S. M. Buccellato and P. Vetro (2005). Stability of a stochastic SIR system. *Phys A Stat Mech Its Appl* **354**, 111–126.
42. A. Torres, M. P. Frías and M. D. Ruiz-Medina (2016). Log-Gaussian Cox processes in infinite-dimensional spaces. *Theor Prob Math Stat* **95**, 157–177.
43. E. Volz (2008). SIR dynamics in random networks with heterogeneous connectivity. *Journal of Mathematical Biology* **56**, 293–310.
44. C. Wang, P. W. Horby, F. Hayden and G. F. Gao (2020). A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470–473.
45. R. K. Wasiur, B. Choiy, E. Kenahz and G. A. Rempa (2019). Survival dynamical systems for the population-level analysis of epidemics. arXiv.1901.00405.
46. G. Webb (1981). A reaction-diffusion model for a deterministic diffusive epidemic. *J Math Anal Appl* **84**, 150–161.
47. Y. Xu, L. Allena and A. Perelson (2007). Stochastic model of an influenza epidemic with drug resistance. *Journal of Theoretical Biology* **248**, 179–193.
48. J. Yu, D. Jiang and N. Shi (2009). Global stability of two-group SIR model with random perturbation. *J Math Anal Appl.* **360** 235–244.
49. F. Zhang, Z. Li and F. Zhang (2008). Global stability of an SIR epidemic model with constant infectious period. *Appl Math Comput.* **199**, 285–291.
50. T. Zhou, Z. Fu and B. Wang (2006). Epidemic dynamics on complex networks. *Progress in Natural Science* **16**, 452–457.
51. F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, et al (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*. doi.org/10.1016/S0140-6736(20)30566-3.
52. L. Zhuang and N. Cressie (2014). Bayesian hierarchical statistical SIRS models. *Statistical Methods & Applications* **23**, 601–646.

Figures

Figure 1

COVID–19 mortality risk maps, since March, 8 to May, 13, 2020. Original (left) and estimated (right) maps, computed from the Bayesian (first line) and classical (second line) COVID–19 mortality predictors, for the truncation parameter value $K(T)=8$

Figure 2

COVID–19 mortality cumulative cases maps, since March, 8 to May, 13, 2020. Original (left) and estimated (right) maps, from the Bayesian (first line) and classical (second line) COVID–19 mortality count predictors, for the truncation parameter value $K(T)=8$

Figure 3

FIRST DATA CATEGORY. Observed and estimated COVID-19 mortality log-risk curves, from the implementation of Generalized Regression Neural Network (GRNN), Multilayer Perceptron (MLP), Support Vector Regression (SVR), Bayesian Neural Network (BNN), Radial Basis Function Neural Network (RBF), and Gaussian Processes (GP)

Figure 4

FIRST DATA CATEGORY. Observed and estimated COVID–19 mortality cumulative cases curves from the implementation of Generalized Regression Neural Network (GRNN), Multilayer Perceptron (MLP), Support Vector Regression (SVR), Bayesian Neural Network (BNN), Radial Basis Function Neural Network (RBF), and Gaussian Processes (GP)

Figure 5

SECOND DATA CATEGORY. Observed and estimated COVID-19 mortality log-risk curves, from the implementation of Generalized Regression Neural Network (GRNN), Multilayer Perceptron (MLP), Support Vector Regression (SVR), Bayesian Neural Network (BNN), Radial Basis Function Neural Network (RBF), Gaussian Processes (GP), and Classical and Bayesian semilinear estimation

Figure 6

SECOND DATA CATEGORY. Observed and estimated COVID–19 mortality cu- mulative cases curves, from the implementation of Generalized Regression Neural Network (GRNN), Multilayer Perceptron (MLP), Support Vector Regression (SVR), Bayesian Neural Network (BNN), Radial Basis Function Neural Network (RBF), Gaussian Processes (GP), and Classical and Bayesian semilinear estimation

Figure 7

ML MODEL RANKING. T.-10-fold-SMAPE in green line for the first category (O.D.C. category), and red line for the second category (P.D.C category), and T.-5-fold-SMAPE in black line for the first category (O.D.C category), and blue line for the second category (P.D.C category)

Figure 8

JOINT MODEL RANKING. T.-10-fold-SMAPE (red line) and T.-5-fold-SMAPE (blue line) in the second category (P.D.C category)

Figure 9

Observed, and Interpolated and cubic B–spline smoothed COVID–19 mortality cumulative cases curves at the 17 Spanish Autonomous Communities, since March, 8 to May, 13, 2020

Figure 10

Observed COVID–19 mortality risk curves at the 17 Spanish Autonomous Communities, since March, 8 to May, 13, 2020

Figure 11

Observed and estimated COVID–19 mortality log–risk curves, at the 17 Spanish Communities, since March, 8 to May, 13, 2020, under a spatial heterogeneous linear and nonlinear trend model fitting, applying a classical (C) and Bayesian (B) componentwise estimation of the residual correlation, under the truncation parameter value $K(T)=9$