COVID-19 mortality analysis from soft-data multivariate curve regression and machine learning

Antoni Torres–Signes¹ · María P. Frías² · María D. Ruiz-Medina³

Accepted: 3 April 2021 / Published online: 19 April 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

A multiple objective space-time forecasting approach is presented involving cyclical curve log-regression, and multivariate time series spatial residual correlation analysis. Specifically, the mean quadratic loss function is minimized in the framework of trigonometric regression. While, in our subsequent spatial residual correlation analysis, maximization of the likelihood allows us to compute the posterior mode in a Bayesian multivariate time series soft-data framework. The presented approach is applied to the analysis of COVID-19 mortality in the first wave affecting the Spanish Communities, since March 8, 2020 until May 13, 2020. An empirical comparative study with Machine Learning (ML) regression, based on random k -fold cross-validation, and bootstrapping confidence interval and probability density estimation, is carried out. This empirical analysis also investigates the performance of ML regression models in a hard- and soft-data frameworks. The results could be extrapolated to other counts, countries, and posterior COVID-19 waves.

Keywords COVID-19 analysis · Curve regression · Hard-data · Machine learning · Multivariate time series · Soft-data

Mathematics Subject Classification 62F40 · 62F15 · 62F10 · 90B99

Introduction

Coronavirus disease 2019 (COVID-19) rapidly spreads around many other countries, since December 2019 when arises in China (see Sivakumar 2020; Wang et al. 2020; Zhou et al. 2020). The effective allocation of medical resources requires the derivation of predictive techniques, describing the spatiotemporal dynamics of COVID-19 (see, e.g., Du et al. 2020; Khan and Atangana 2020; Nishiura

et al. 2020; Remuzzi and Remuzzi 2020, just to mention a few). Epidemiological models can contribute to the analysis of the causes, dynamics, and spread of this pandemic (see, e.g. Huppert and Katriel 2013; Keeling and Rohani 2008; Laaroussi et al. 2018, and the references therein). Short-term forecasts can be obtained adopting the framework of compartmental SIR (susceptible-infectious-recovered) models, based on ordinary differential equations (see, e.g. Angulo et al. 2013; Elhia et al. 2014; Ji et al. 2012; Kermack and McKendrick 1927; Kuznetsov and Piccardi 1994; Milner and Zhao 2008; Pathak et al. 2010; Tornatore et al. 2005; Yu et al. 2009; Zhang et al. 2008). An extensive literature is available, including different versions of compartmental models, like SIR-susceptible (SIRS, Dushoff et al. 2004), and delay differential equations (see Beretta et al. 2001; McCluskey 2010; Sekiguchi and Ishiwata 2010). Spatial extensions, based on reaction-diffusion models, reflecting the infectious disease spread over a spatial region can be found, for instance, in Guin and Mandal (2014) and Webb (1981). SEIRD (susceptible, exposed, infected, recovered, deceased) models, incorporating the spatial spread of the disease with inhomogeneous

✉ María D. Ruiz-Medina
mruiz@ugr.es

Antoni Torres–Signes
atisignes@uma.es

María P. Frías
mpfrias@ujaen.es

¹ Department of Statistics and Operation Research, Faculty of Sciences, University of Málaga, Málaga, Spain

² Department of Statistics and Operation Research, Faculty of Sciences, University of Jaén, Jaén, Spain

³ Department of Statistics and Operation Research, Faculty of Sciences, University of Granada, Granada, Spain

diffusion terms are also analyzed (see Roques and Bonnfon 2016 and Roques et al. 2011). The stochastic version of SIR-type models intends to cover several limitations detected regarding uncertainty in the observations, and the hidden dynamical epidemic process. Markov chain SIR based modelling (see Anderson and Britton 2000; Xu et al. 2007), and some recent stochastic formulations involving complex networks (see Volz 2008; Zhou et al. 2006) or drug-resistant influenza (see Chao et al. 2012) constitute some alternatives. A Bayesian hierarchical statistical SIRS model framework is adopted in Aalen et al. (2008); Abboud et al. (2019); Anderson and Britton (2000); Fleming and Harrington (1991) taking into account the observation error in the counts, and uncertainty in the parameter space. Beyond SIR modeling, the multivariate survival analysis approach offers a suitable modelling framework, regarding infection, incubation and recovering random periods, affecting the containment of COVID-19 (see, e.g., Bolker and Grenfell 1996; Keeling et al. 1997; Pak et al. 2020; Wasiur et al. 2019).

In a first stage, most of the above referred models have been adapted and applied to approximate the space/time evolution of COVID-19 incidence and mortality. That is the case, for instance, of the three models presented in Roosa et al. (2020), which were validated with outbreaks of other diseases different from COVID-19. Alternative SEIR type models, involving stochastic components, are formulated in Kucharski et al. (2020). A revised SEIR model has also been proposed in Zhang et al. (2020) (see also He et al. 2020). A θ -SEIHRD model, able to estimate the number of cases, deaths, and needs of beds in hospitals, is introduced in Ivorra et al. (2020), adapted to COVID-19, based on the Be-CoDiS model (see Ivorra et al. 2015). Due to the low quality of the records available, and the hidden sample information, the most remarkable feature in this research area is the balance between complexity and indentifiability of model parameters. Recently, an attempt to simplify modelling strategies, applied to COVID-19 data analysis, is presented in Ramosa et al. (2020), in terms of θ -SEIHQRD model. Mitigation of undersampling is proposed in Langousis and Carsteanu (2020), based on re-scaling of summary statistics characterizing sample properties of the pandemic process, useful between countries with similar levels of health care.

Nowadays ML 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 this concept to alternative models, such as support vector machines, decision trees, and others (see, e.g., Alpaydin 2004; Blanquero et al. 2020; Hastie et al. 2001; Mohammady et al. 2021). In general, curve regression techniques based on a function basis, usually in the

space of square integrable functions with respect to a suitable probability measure, allow short- and long- term forecast. Thus, depending on our choice of the function basis, and the probability measure selected, *particle* and *field* views could be combined. Note that the classical stochastic diffusion models offer a *particle* rather than a *field* view (see, e.g., Malesios et al. 2016).

Linear regression, multilayer perceptron and vector autoregression methods have been applied in Sujath et al. (2020a, 2020b) to predicting COVID-19 spread, anticipating the potential patterns of COVID-19 effects (see also Section 2 of Sujath et al. (2020a), on related work). Early stage location of COVID-19 is addressed in Barstugan et al. (2020), applying machine learning strategies actualized on stomach Computed Tomography pictures. Chien and Chen (2020) evaluates association between meteorological factors and COVID-19 spread. They concluded that average temperature, minimum relative humidity, and precipitation were better predictors, displaying possible non-linear correlations with COVID-19 variables. These conclusions are crucial in the subsequent machine learning regression based analysis.

This paper presents a multiple objective space-time forecasting approach, where curve trigonometric log-regression is combined with multivariate time series spatial residual analysis. In our curve regression model fitting, we are interested on reflecting the cyclical behavior of COVID-19 mortality induced by the hardening or relaxation of the containment measures, adopted to mitigate the increase of infections and mortality. The trigonometric basis (sines and cosines) is then selected in our spatial heterogeneous curve log-regression model fitting. The ratio of the expected minimized empirical risk, and the corresponding expected value of the quadratic loss function at such a minimizer is considered for model selection (see, e.g., Chapelle et al. 2002). Note that this selection procedure provides an agreement between the expected minimum empirical risk, and the corresponding expected theoretical loss function value.

The penalized factor proposed in Chapelle et al. (2002), applied to our choice of the truncation parameter, leads to the dimension of the subspace where our curve regression estimator is approximated at any spatial location. This model selection procedure is asymptotically equivalent to Akaike correction factor. A robust modification of the Akaike information criterion can be found, for example, in Agostinelli (2001). As an alternative, one can consider cross-validation criterion for selecting the best subset of explanatory variables (see Takano and Miyashiro 2020, where a mixed-integer optimization approach is proposed in this context).

Beyond asymptotic analysis, model selection from finite sample sizes constitutes a challenging topic in our

approach. To address this problem, a bootstrap estimator of the ratio between the expected quadratic loss function and the expected training quadratic error, from different sets of explanatory variables, is implemented. Bootstrap confidence intervals are also provided for the spatial mean of the curve regression predictor, and for the expected training error of the curve regression, and of the multivariate time-series residual predictor. The bootstrap approximation of the probability distribution of these statistics is also computed.

In our multivariate time series analysis of the regression residuals, a classical and Bayesian componentwise estimation of the spatial linear correlation is achieved. The presented multiple objective forecasting approach is applied to the spatiotemporal analysis of COVID-19 mortality in the first wave affecting the Spanish Communities, since March, 8, 2020 until May, 13, 2020. Our results show a remarkable qualitative agreement with the reported epidemiological data.

The spatiotemporal approach presented in this paper makes the fusion of generalized random field theory, and our multiple-objective space-time forecasting, based on nonlinear parametric regression, and bayesian analysis of the spatiotemporal correlation structure. Regarding the *site-specific* or *specificatory* knowledge bases (see Christakos et al. 2002), in our approach, several information sources can be incorporated in the description of the hidden epidemic process. Particularly, we distinguish here between hard-data or hard measurements providing a satisfactory level of accuracy for practical purposes, and soft-data displaying a non-negligible amount of uncertainty. That is, in this second data category, we include missing observations or imperfect observations, categorical data and fuzzy inputs (see also Christakos 2000, 2002; Christakos and Hristopulos 1998, and the references therein). In this paper, we consider hard-data sets given by numerical values of our count process at the Spanish Communities analyzed. Our soft-data sample complements hard measures, in terms of interpolated, smoothed, and spatial projected data. Particularly, spatial correlations between regions are incorporated in terms of soft-data. Additional information about the continuous functional nature of the underlying space-time COVID-19 mortality process is also reflected in our soft-data set. This information helps the implementation of the proposed estimation methodology in the framework of Functional Data Analysis (FDA) techniques.

As commented before, last advances in spatiotemporal mapping of epidemiological data incorporate ML regression models to improve and help the understanding of *general or core knowledge bases*. Thus, model fitting is achieved according to epidemiological systems laws, population dynamics, and theoretical space-time dependence models (see Christakos 2008, and the references

therein). See also Barstugan et al. 2020; Chien and Chen 2020 and Sujath et al. (2020a) in the hard-data context. It is well-known that the limited availability of hard-data affects space-time analysis. Hence, the incorporation of soft-data into ML regression models can help this analysis, providing a global view of the available sample information (see, e.g., Christakos et al. 2002). Particularly, in our empirical comparative analysis, involving ML regression models and our approach, input hard- and soft-data information is incorporated. Cross-validation, bootstrapping confidence intervals and probability density estimation support our comparative study. Specifically, random k -fold ($k = 5, 10$) cross-validation first evaluates the performance of the compared regression models from hard- and soft-data, in terms of Symmetric Mean Absolute Percentage Errors (SMAPEs). Bootstrap confidence intervals and probability density estimation of the spatially averaged SMAPEs approximate the distributional characteristics of the random k -fold cross-validation errors. Thus, a complete picture of SMAPEs supports our evaluation of the predictive ability of the regression models tested, from the analyzed hard- and soft-data sets.

From the empirical comparative analysis carried out, we can conclude that almost the best performance in both, hard- and soft-data categories, is displayed by Radial Basis Function Neural Network (RBF), and Gaussian Processes (GP). Both approaches are improved, when soft-data are incorporated into the regression analysis. Slightly differences are observed in the performance of Support Vector Regression (SVR) and Bayesian Neural Networks (BNN). Multilayer Perceptron (MLP) gets over GRNN, presenting better estimation results when hard-data are analyzed. The sample values and distributional characteristics of cross-validation SMAPEs, in Generalized Regression Neural Network (GRNN), are similar to the ones obtained in trigonometric curve regression, when spatial residual analysis is achieved in terms of empirical second-order moments. Note that, GRNN is also favored by the soft-data category. In this category, BNN and our approach show very similar performance, when trigonometric regression is combined with Bayesian multivariate time series residual prediction. Indeed, some slightly better bootstrapping distributional characteristics of our approach respect to BNN are observed in the soft-data category.

The outline of the paper is the following. The modeling approach is introduced in Sect. 2. Section 3 describes the multiple objective forecasting methodology. This methodology is applied to the spatiotemporal statistical analysis of COVID-19 mortality in Spain in Sect. 4. The empirical comparative study with ML regression models is given in Sect. 5. Conclusions about our data-driven model ranking can be found in Sect. 6. In the Supplementary Material, a brief introduction to our implementation of ML models

from hard- and soft-data is provided. Additional numerical estimation results, based on the complete sample, are also displayed. Particularly, the observed and predicted mortality cumulative cases, and log-risk curves are displayed.

Data model

Let $(\Omega, \mathcal{A}, \mathcal{P})$ be the basic probability space. Consider $H = L^2(\mathbb{R}^d)$, $d \geq 2$, the space of square-integrable functions on $\mathbb{R}^d$, to be the underlying real separable Hilbert space. In the following, we denote by $\mathcal{B}^d$ the Borel σ -algebra in $\mathbb{R}^d$, $d \geq 1$.

Let $X = \{X_t(\mathbf{z}), \mathbf{z} \in \mathbb{R}^d, t \in \mathbb{R}_+\}$ be our spatiotemporal input hard-data process on $(\Omega, \mathcal{A}, \mathcal{P})$, satisfying $E[\|X_t(\cdot)\|_H^2] < \infty$, for any time $t \in \mathbb{R}_+$. The input soft-data process over any spatial bounded set $D \in \mathcal{B}^d$ is then defined as

$$\left\{ X_t(h) = \int_D X_t(\mathbf{z})h(\mathbf{z})d\mathbf{z}, h \in C_0^\infty(D), t \in \mathbb{R}_+ \right\}, \quad (1)$$

where $C_0^\infty(D)$ denotes the space of infinite differentiable functions, with compact support contained in D . For each bounded set $D \in \mathcal{B}^d$, define

$$A = \{A_t(h) = \exp(X_t(h)), h \in C_0^\infty(D), t \in \mathbb{R}_+\}.$$

Assume that, for any finite positive interval $\mathcal{T} \in \mathcal{B}$, and bounded set $D \in \mathcal{B}^d$,

$$\mathcal{I}_{\mathcal{T}}(h) = \int_{\mathcal{T}} \exp(X_t(h))dt < \infty, \quad \forall h \in C_0^\infty(D), \quad (2)$$

where $\int_{\mathcal{T}} \exp(X_t(h))dt$ denotes the identity in the second-order moment sense. Let $\{N_h : (\Omega, \mathcal{A}, \mathcal{P}) \times \mathcal{B} \rightarrow \mathbb{N}, h \in H\}$ be a family of random counting measures. Given the observation $\{x_t(h), t \in \mathcal{T}\}$ at the finite temporal interval $\mathcal{T} \in \mathcal{B}$ of the input soft-data process over the spatial h -window in D , the conditional probability distribution of the number of random events $N_h(\mathcal{T})$ that occur in $\mathcal{T} \in \mathcal{B}$ is a Poisson probability distribution with mean $\int_{\mathcal{T}} \exp(x_t(h))dt$, for every $h \in C_0^\infty(D)$ and $D \in \mathcal{B}^d$. We refer to $\mathcal{I}_{\mathcal{T}}(h)$ as the generalized cumulative mortality risk random process over the interval $\mathcal{T}$. Hence, the input hard-data process $X = \{X_t(\mathbf{z}), \mathbf{z} \in \mathbb{R}^d, t \in \mathbb{R}_+\}$ defines the spatiotemporal mortality log-risk process.

From the sample values of our input soft-data process, the following observation model is considered in the curve regression model fitting

$$\begin{aligned} \ln(A_t)(\psi_{p,\varpi_p}) &= g_t(\psi_{p,\varpi_p}, \theta(p)) + \varepsilon_t(\psi_{p,\varpi_p}) \\ &= \left\langle g_t(\cdot, \theta(p)), \psi_{p,\varpi_p}(\cdot) \right\rangle_H \\ &\quad + \left\langle \varepsilon_t(\cdot), \psi_{p,\varpi_p}(\cdot) \right\rangle_H, \quad t \in \mathbb{R}_+, p = 1, \dots, P, \end{aligned} \quad (3)$$

where

$$\begin{aligned} g_t(\psi_{p,\varpi_p}, \theta(p)) &= \int_{\mathcal{D}_p} g_t(\mathbf{z}, \theta(p))\psi_{p,\varpi_p}(\mathbf{z})d\mathbf{z} \\ \langle f, g \rangle_H &= \int_{\mathbb{R}^d} f(\mathbf{z})g(\mathbf{z})d\mathbf{z}, \end{aligned} \quad (4)$$

with $\{\psi_{p,\varpi_p}, p = 1, \dots, P\} \subset H$ denoting a function family in H , whose elements have respective compact supports $\mathcal{D}_p$, $p = 1, \dots, P$, defining the p small-areas where the counts are aggregated, satisfying suitable regularity conditions. For each $p = 1, \dots, P$, the vector ϖ_p contains the center and bandwidth parameters, defining the window selected in the analysis of the small-area p . For each $p \in \{1, \dots, P\}$, $\theta(p) = (\theta^1(p), \dots, \theta^q(p)) \in \Theta$ represents the unknown parameter vector to be estimated at the p region, and Θ is the open set defining the parameter space, whose closure Θ^c is a compact set in $\mathbb{R}^q$. We assume that g_t is of the form (see, e.g., Ivanov et al. 2015)

$$\begin{aligned} g_t(\theta(p)) &= \sum_{k=1}^N (A_k(p) \cos(\varphi_k(p)t) + B_k(p) \sin(\varphi_k(p)t)), \\ p &= 1, \dots, P, \quad t \in \mathbb{R}_+, \end{aligned} \quad (5)$$

whose spatial-dependent parameters are given by the temporal scalings $(\varphi_1(\cdot), \dots, \varphi_N(\cdot))$, and the Fourier coefficients $(A_1(\cdot), B_1(\cdot), \dots, A_N(\cdot), B_N(\cdot))$. For simplifications purposes, we will consider that the scaling parameters φ_k , $k = 1, \dots, N$, are known, and fixed over the P spatial regions. Also, $C_k^2(\cdot) = A_k^2(\cdot) + B_k^2(\cdot) > 0$, for $k = 1, \dots, N$, where N denotes the truncation parameter, that will be selected according to the penalized factor proposed in Chapelle et al. (2002), as we explain in more detail in Sect. 3. Thus,

$$\theta(p) = (A_1(p), B_1(p), \dots, A_N(p), B_N(p)), \quad p = 1, \dots, P.$$

To analyze the spatial correlation between regions, a multivariate autoregressive model is considered for prediction of the regression residual term at each region $p \in \{1, \dots, P\}$. Particularly, for any $T \geq 2$, ε_t in equation (3) is assumed to satisfy the state equation, for $p = 1, \dots, P$,

$$\varepsilon_t(\psi_{p,\varpi_p}) = \sum_{q=1}^P \rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p})\varepsilon_{t-1}(\psi_{q,\varpi_q}) + v_t(\psi_{p,\varpi_p}), \quad (6)$$

where, for any $t \in \mathbb{R}_+$, and $p, q = 1, \dots, P$,

$$\begin{aligned} \varepsilon_t(\psi_{p,\varpi_p}) &= \int_{\mathcal{D}_p} \varepsilon_t(\mathbf{z})\psi_{p,\varpi_p}(\mathbf{z})d\mathbf{z} \\ v_t(\psi_{p,\varpi_p}) &= \int_{\mathcal{D}_p} v_t(\mathbf{z})\psi_{p,\varpi_p}(\mathbf{z})d\mathbf{z} \\ \rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p}) &= \int_{\mathcal{D}_p \times \mathcal{D}_q} \rho(\mathbf{z}, \mathbf{y})\psi_{p,\varpi_p}(\mathbf{z})\psi_{q,\varpi_q}(\mathbf{y})d\mathbf{y}d\mathbf{z}. \end{aligned}$$

Here, $(v_t(\psi_{p,\varpi_p}), p = 1, \dots, P), t \in \mathbb{R}_+$, are assumed to be independent zero-mean Gaussian P -dimensional vectors. For $p, q \in \{1, \dots, P\}$, the projection $\rho(\psi_{p,\varpi_p})(\psi_{q,\varpi_q})$ then keeps the temporal linear autocorrelation at each spatial region for $p = q$, and the temporal linear cross-correlation between regions for $p \neq q$ of the regression error $\{\varepsilon_t(\cdot), t \in \mathbb{R}_+\}$ (see, Bosq 2000).

Implementation of the curve regression model and spatial residual analysis

Let $\mathcal{D}_1, \dots, \mathcal{D}_P$ be the small-areas, where the counts are aggregated, and $\{\psi_{p,\varpi_p}, \varpi_p = (c_p, \rho_p), p = 1, \dots, P\} \subset H$ be the functions with respective compact supports $\mathcal{D}_1, \dots, \mathcal{D}_P$. Particularly, we denote by $c_p, p = 1, \dots, P$, the centers respectively allocated at the regions $\mathcal{D}_1, \dots, \mathcal{D}_P$, and by $\rho_1, \dots, \rho_P$, the bandwidth parameters providing the associated window sizes.

In practice, from the observation model (3), to find g_t in (5) minimizing the expected quadratic loss function, or expected risk, we look for the minimizer $\hat{\theta}_T(p)$ of the empirical regression risk

$$\begin{aligned} L_T(\hat{\theta}_T(p)) &= \inf_{\theta(p) \in \Theta^c} L_T(\theta(p)) \\ &= \inf_{\theta(p) \in \Theta^c} \frac{1}{T} \sum_{t=1}^T \left| \ln(A_t)(\psi_{p,\varpi_p}) - g_t(\theta(p)) \right|^2. \end{aligned} \quad (7)$$

Truncation parameter N is then selected to controlling the ratio between the expected quadratic loss function at $\hat{\theta}_T(p)$, and the expected value of the minimized empirical risk from the identity

$$\begin{aligned} E \left[\ln(A_t)(\psi_{p,\varpi_p}) - g_t(\psi_{p,\varpi_p}, \hat{\theta}_T(p)) \right]^2 \\ = E \left[L_T(\hat{\theta}_T(p)) \right] \left(1 - \frac{N}{T} \right)^{-1} \left(1 + \frac{\sum_{i=1}^N 1/\lambda_i}{T} \right), \end{aligned} \quad (8)$$

where, for $i = 1, \dots, N$, $1/\lambda_i$ denotes the inverse of the i th eigenvalue of the matrix $\Phi^T \Phi$, with Φ being a $T \times N$ matrix, whose elements are the values of the N trigonometric basis functions selected at the time points $t = 1, \dots, T$. Parameter N should be such that $N < T$. Note that, asymptotically, when $N \rightarrow \infty$, $\Phi^T \Phi$ goes to the identity matrix, and for $i = 1, \dots, N$, $1/\lambda_i \sim 1$. In equation (8), we have considered the minimized empirical risk

$$L_T(\hat{\theta}_T(p)) = \frac{1}{T} \tilde{\mathcal{R}}^T(p) \left(I_{T \times T} - \Phi(\Phi^T \Phi)^{-1} \Phi^T \right) \tilde{\mathcal{R}}(p), \quad (9)$$

for each spatial region $p = 1, \dots, P$, where

$$\begin{aligned} \tilde{\mathcal{R}}(p) &= \left(\sum_{k=N+1}^{\infty} (A_k(p) \cos(\varphi_k t) + B_k(p) \sin(\varphi_k t)) \right. \\ &\quad \left. + \varepsilon_t(\psi_{p,\varpi_p}), t = 1, \dots, T \right). \end{aligned}$$

Our regression predictor is then computed, for any $t \in \mathbb{R}_+$, from the identity

$$\ln(\widehat{A}_t)(\psi_{p,\varpi_p}) = g_t(\hat{\theta}_T(p)), \quad p = 1, \dots, P \quad (10)$$

(see Theorem 1 in Ivanov et al. (2015) about conditions for the weak-consistency of (10)).

The regression residuals

$$\mathbf{Y} = \left\{ Y_t(\psi_{p,\varpi_p}) = \ln(A_t)(\psi_{p,\varpi_p}) - g_t(\hat{\theta}_T(p)), t = 1, \dots, T, p = 1, \dots, P \right\},$$

and the empirical nuclear autocovariance and cross-covariance operators

$$\begin{aligned} \hat{R}_{0,T}^{\mathbf{Y}}(\psi_{p,\varpi_p})(\psi_{q,\varpi_q}) &= \frac{1}{T} \sum_{t=1}^T Y_t(\psi_{p,\varpi_p}) Y_t(\psi_{q,\varpi_q}) \\ \hat{R}_{1,T}^{\mathbf{Y}}(\psi_{p,\varpi_p})(\psi_{q,\varpi_q}) &= \frac{1}{T-1} \sum_{t=1}^{T-1} Y_t(\psi_{q,\varpi_q}) Y_{t+1}(\psi_{p,\varpi_p}), p, q = 1, \dots, P, \end{aligned} \quad (11)$$

will be considered in the estimation of the spatial linear residual correlation (see Bosq 2000). A truncation parameter $k(T)$ is also considered here to remove the ill-posed nature of this estimation problem. Particularly, $k(T)$ must satisfy $k(T) \rightarrow \infty$, $k(T)/T \rightarrow 0$, $T \rightarrow \infty$. A suitable choice of $k(T)$ also ensures strong-consistency of the estimator

$$\begin{aligned} \hat{\rho}_{k(T)}(\psi_{p,\varpi_p})(\psi_{q,\varpi_q}) \\ = \sum_{k,l=1}^{k(T)} \frac{\langle \psi_{p,\varpi_p}, \phi_{k,T} \rangle_H \langle \psi_{q,\varpi_q}, \phi_{l,T} \rangle_H}{\lambda_{k,T}(\hat{R}_{0,T}^Y)} \hat{R}_{1,T}^Y(\phi_{k,T})(\phi_{l,T}), \end{aligned} \quad (12)$$

for $p, q = 1, \dots, P$ (see Bosq 2000). Here,

$$\hat{R}_{0,T}^Y = \sum_{k=1}^T \lambda_{k,T}(\hat{R}_{0,T}^Y) [\phi_{k,T} \otimes \phi_{k,T}], \quad (13)$$

where $\{\lambda_{k,T}(\hat{R}_{0,T}^Y), k = 1, \dots, T\}$ and $\{\phi_{k,T}, k \geq 1\}$ denote the empirical eigenvalues and eigenvectors of $\hat{R}_{0,T}^Y$, respectively. Particularly, we consider $k(T) = \ln(T)$ (see Bosq 2000). The classical plug-in predictor is then computed, for each $p = 1, \dots, P$, as

$$\hat{Y}_t^{k(T)}(\psi_{p,\varpi_p}) = \sum_{q=1}^P \hat{\rho}_{k(T)}(\psi_{q,\varpi_q})(\psi_{p,\varpi_p}) Y_{t-1}(\psi_{q,\varpi_q}), \quad t \geq 1. \quad (14)$$

Under the Gaussian distribution of v_t , in the Bayesian estimation of ρ , from (6), the likelihood function, defining the objective function, is given by, for each $p = 1, \dots, P$,

$$\begin{aligned} \tilde{L}_p(\varepsilon_{1p}, \dots, \varepsilon_{Tp}, \varepsilon_{0q}, \dots, \varepsilon_{(T-1)q}) \\ \rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p}), q = 1, \dots, P \\ \exp\left(-\frac{1}{2\sigma_p^2} \sum_{t=1}^T \left(\varepsilon_t(\psi_{p,\varpi_p}) - \sum_{q=1}^P \varepsilon_{t-1}(\psi_{q,\varpi_q}) \rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p})\right)^2\right) \\ = \frac{1}{(\sigma_p \sqrt{2\pi})^T} \\ \times \prod_{q=1}^P [\rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p})]^{a_{pq}-1} \\ (1 - \rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p}))^{b_{pq}-1} \\ \times \frac{\mathbb{I}_{\{0 < \rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p}) < 1\}}}{\mathbb{B}(a_{pq}, b_{pq})} \\ = \frac{1}{(\sigma_p \sqrt{2\pi})^T} \exp\left(-\frac{1}{2\sigma_p^2} \sum_{t=1}^T [\varepsilon_t(\psi_{p,\varpi_p})]^2\right) \\ \times \prod_{q=1}^P [\rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p})]^{a_{pq}-1} \\ (1 - \rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p}))^{b_{pq}-1} \\ \times \frac{\mathbb{I}_{\{0 < \rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p}) < 1\}}}{\mathbb{B}(a_{pq}, b_{pq})}, \end{aligned} \quad (15)$$

where, for each $p = 1, \dots, P$, the beta probability distributions with shape parameters a_{pq} and b_{pq} , $q = 1, \dots, P$, respectively define the prior probability distributions of the independent random variables $\{\rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p}), q = 1, \dots, P\}$. Here, for each $p = 1, \dots, P$, $\varepsilon_{tp} = \varepsilon_t(\psi_{p,\varpi_p}) = \langle \varepsilon_t, \psi_{p,\varpi_p} \rangle_H$, and $\sigma_p = \sqrt{E[\varepsilon_t(\psi_{p,\varpi_p})]^2}$, for $t = 0, \dots, T$. As

before, ψ_{p,ϖ_p} weights the spatial sample information about the p small-area, for $p = 1, \dots, P$. As usual, $\mathbb{I}_{0 < \cdot < 1}$ denotes the indicator function on the interval $(0, 1)$, and $\mathbb{B}(a_{pq}, b_{pq})$ is the beta function,

$$\mathbb{B}(a_{pq}, b_{pq}) = \frac{\Gamma(a_{pq})\Gamma(b_{pq})}{\Gamma(a_{pq} + b_{pq})}.$$

From (15), the Bayesian predictor is obtained, for $p = 1, \dots, P$, as

$$\tilde{\varepsilon}_t(\psi_{p,\varpi_p}) = \sum_{q=1}^P \tilde{\rho}(\psi_{q,\varpi_q})(\psi_{p,\varpi_p}) \varepsilon_{t-1}(\psi_{q,\varpi_q}), \quad t \geq 1, \quad (16)$$

with $(\tilde{\rho}(\psi_{1,\varpi_1})(\psi_{p,\varpi_p}), \dots, \tilde{\rho}(\psi_{P,\varpi_P})(\psi_{p,\varpi_p}))$ being computed by maximizing (15), to find the posterior mode (see Bosq and Ruiz-Medina 2014, where Bayesian estimation is introduced in an infinite-dimensional framework). We refer to (16) as the Bayesian plug-in predictor of the residual mortality log-risk process at the p small area, for $p = 1, \dots, P$. In practice, equation (15) is approximated from the computed values of the regression residual process.

Statistical analysis of COVID-19 mortality

Our analysis is based on daily records of COVID-19 mortality reported by the Carlos III Health Institute, since March, 8 to May, 13, 2020, at the 17 Spanish Communities. We first describe the main steps of the proposed estimation algorithm, referring to the inputs and outputs at different stages.

Step 1 Daily records of COVID-19 mortality are accumulated over the entire period at every Spanish Community. The resulting step cumulative curves are interpolated at 265 temporal nodes, and cubic B-spline smoothed. Their derivatives and logarithmic transforms are then computed.

Step 2 Our soft-input-data process is obtained from the spatial projection of the outputs in Step 1 onto the compactly supported basis $\{\psi_{p,\varpi_p}, p = 1, \dots, P = 17\}$. We choose the tensorial product of Daubechies wavelet bases. Here, for $p = 1, \dots, 17$, $\varpi_p = (N(p), j(p), \mathbf{k}(p))$, whose components respectively provide the order of the Daubechies wavelet functions, the resolution level, and the vector of spatial displacements, according to the area occupied by each Spanish community (see, e.g., Daubechies 1988).

Table 1 Regression parameter estimates $\hat{A}_k(\cdot)$, $k = 1, \dots, 6$, at the 17 Spanish Communities

SC/PE	$\hat{A}_1(\cdot)$	$\hat{A}_2(\cdot)$	$\hat{A}_3(\cdot)$	$\hat{A}_4(\cdot)$	$\hat{A}_5(\cdot)$	$\hat{A}_6(\cdot)$
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

Table 2 Regression parameter estimates $\hat{B}_k(\cdot)$, $k = 1, \dots, 6$, at the 17 Spanish Communities

SC/PE	$\hat{B}_1(\cdot)$	$\hat{B}_2(\cdot)$	$\hat{B}_3(\cdot)$	$\hat{B}_4(\cdot)$	$\hat{B}_5(\cdot)$	$\hat{B}_6(\cdot)$
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

Step 3 The choice $N = 6$ in (8) corresponds to 1.1304 value of the ratio between the mean quadratic loss function and expected minimized empirical risk. Hence 12 coefficients should be estimated. Note that the eigenvalues in (8) are computed from the trigonometric basis.

Step 4 Under $N = 6$ in Step 3, the least-squares estimates of the 12 Fourier coefficients are

computed from (7), in terms of the soft-input-data process obtained as output in Step 2.

Step 5 The regression residuals are then calculated from Step 4.

Step 6 The auto- and cross-covariance operators in (11) are computed from the outputs of Step 5. The residual spatial linear correlation matrix is then obtained from (12). The truncation scheme $k(T) = \ln(T)$ has been adopted, with $T = 265$.

Fig. 1 At the top, COVID-19 mortality mean cumulative curve in Spain, since March, 8, 2020 to May, 13, 2020 (continuous red line, 265 temporal nodes), and bootstrap curve confidence intervals, at the left-hand-side, $\mathcal{I}_1$ (dashed blue lines) and $\mathcal{I}_2$ (dashed magenta lines), and at the right-hand-side, $\mathcal{I}_3$ (dashed green lines) and $\mathcal{I}_4$ (dashed yellow lines). Plots at the center and bottom reflect the same information respectively referred to the mean intensity (spatial averaged COVID-19 mortality risk curve), and log-intensity (spatial averaged COVID-19 mortality log-risk curve) curves in Spain. All the confidence bootstrap intervals are computed at confidence level $1 - \alpha = 0.95$, from 1000 bootstrap samples

Table 3 Computed values $L_{265}(\hat{\theta}_{265}(p))$, $p = 1, \dots, 17$

$L_{265}(\hat{\theta}_{265}(p))$	$p =$	1...17			
0.0155	0.0259	0.0668	0.0408	0.0927	
0.0623	0.1642	0.0883	0.2174	0.0313	
0.0559	0.1904	0.0054	0.1602	0.1640	
0.0003	0.1238				

Step 7 The residual predictor (14) is computed from Step 6.

Step 8 100 bootstrap samples are generated from the empirical autocorrelation projections. The bootstrap prior fitted suggests us to consider a scaled beta probability density with shape parameters 14 and 13.

Step 9 Assuming a Gaussian scenario for our log-regression residuals, our constrained nonlinear

Fig. 2 1000 bootstrap samples have been generated of the spatially averaged minimum empirical regression risk (SAMERR). The corresponding sample values are displayed at the top. The bootstrap histogram can be found at the bottom-left-hand side. The bootstrap probability density is plotted at the bottom-right-hand-side

Table 4 Bootstrap confidence intervals for $\overline{L}_{265}$ (confidence level $1 - \alpha = 0.95$)

CIS	1000	10000
$\mathcal{I}_1$	[0.0593, 0.1222]	[0.0594, 0.1236]
$\mathcal{I}_2$	[0.0564, 0.1196]	[0.0567, 0.1207]
$\mathcal{I}_3$	[0.0584, 0.1215]	[0.0579, 0.1217]
$\mathcal{I}_4$	[0.0592, 0.1233]	[0.0581, 0.1208]
$\mathcal{I}_5$	[0.0484, 0.1281]	[0.0494, 0.1215]

multivariate objective function (15) is computed from the prior proposed in Step 8.

Step 10 To maximize the objective function computed in Step 9, we implement an hybrid genetic algorithm, constructed from 'gaoptimset' MaLab function, implemented with the 'HybridFcn' option that handles to a function to continuing optimization after the genetic algorithm terminates. This last function applies quasi-Newton methodology in the

optimization procedure, involving an inverse Hessian matrix estimate.

Step 11 The soft-data based bayesian predictor (16) of the residual COVID-19 mortality log-risk is finally computed from the outputs in Step 10.

Step 12 Our multiple objective space-time predictor is obtained from Steps 4 and 11, by addition the regression and residual predictors, applying inverse spatial wavelet transform.

Tables 1–2 below display the parameter estimates $\hat{A}_k(\cdot)$ and $\hat{B}_k(\cdot)$, $k = 1, \dots, 6$, where $\varphi_k = \frac{2\pi}{265}$ has been considered, for $k = 1, \dots, N = 6$. In these tables and below, the following Spanish Community (SC) codes appear: 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 País Vasco, and C17 for La Rioja.

Bootstrap curve confidence intervals at confidence level $1 - \alpha = 0.95$, based on 1000 bootstrap samples, are

Fig. 3 At the left-hand side, empirical projections of the autocorrelation operator ρ , reflecting temporal autocorrelation and cross-correlation between the 17 Spanish Communities analyzed. At the right-hand side, the considered prior probability density (red squares)

computed for the spatial mean, over the 17 Spanish Communities, of the curve regression predictors. Their construction is based on the bias corrected and accelerated percentile method ($\mathcal{I}_1$); Normal approximated interval with bootstrapped bias and standard error ($\mathcal{I}_2$); basic percentile method ($\mathcal{I}_3$), and bias corrected percentile method ($\mathcal{I}_4$) (see Fig. 1). The minimized regression empirical risk values $L_{265}(\hat{\theta}_{265}(p))$, $p = 1, \dots, 17$, are displayed in Table 3.

Figure 2 at the top displays the 1000 bootstrap sample values

$$\overline{L_{265}}(\omega_i) = \frac{1}{P} \sum_{p=1}^P L_{265}(\omega_i, \hat{\theta}_{265}(p)), \quad \omega_i \in \Omega, \quad i = 1, \dots, 1000,$$

of the spatial averaged minimized empirical quadratic risk in the trigonometric regression. Note that the sample mean of these values is $\bar{L} = 0.0262$, showing a good performance of the least-squares regression predictor, according to the value $T(265, 12) = 1.1304$ obtained. The bootstrap histogram and the corresponding approximation of the probability density function, computed from $\overline{L_{265}}(\omega_i)$, $i = 1, \dots, 1000$, are also plotted at the bottom of Fig. 2.

Bootstrap confidence intervals for $\overline{L_{265}}$ have also been computed at level $1 - \alpha = 0.95$, from 1000 and 10000 bootstrap samples. Table 4 displays these intervals respectively based on the bias corrected and accelerated percentile method ($\mathcal{I}_1$); Normal approximated interval with bootstrapped bias and standard error ($\mathcal{I}_2$); basic percentile method ($\mathcal{I}_3$); bias corrected percentile method ($\mathcal{I}_4$), and Student-based confidence interval ($\mathcal{I}_5$).

The classical and Bayesian plug-in predictors of the residual COVID-19 mortality log-risk process at each one

of a scaled, by factor 1/3, Beta distributed random variable with shape parameters 14 and 13 is compared with the bootstrap fitting of an empirical prior (blue squares)

of the Spanish Communities are respectively computed from equations (14) and (16) for $P = 17$.

Given the empirical spectral characteristics observed in the regularized approximation $\hat{\rho}_{k(T)}$ of ρ in (12), from the singular value decomposition of the empirical operators in (11), our choice of the prior for the projections of ρ has been a scaled, by factor 1/3, Beta prior with hyper-parameters $a_{pq} = 14$, and $b_{pq} = 13$, for $p, q = 1, \dots, 17$. The suitability of this data-driven choice, regarding localization of the mode, and the tails thickness, is illustrated in Fig. 3. Specifically, at the right plot in Fig. 3, both, the scaled Beta probability density, with shape parameters 14 and 13 (red-square line), and the fitted probability density (blue-square line), from the generated bootstrap samples, based on the empirical projections of ρ , are displayed. Note that the observed range of the empirical projections of ρ is well fitted, as one can see from the left plot in Fig. 3.

Bootstrap confidence intervals $\mathcal{I}_1, \dots, \mathcal{I}_5$ at level $1 - \alpha = 0.95$, for the expected training standard error of the multivariate time series classical and Bayesian residual COVID-19 mortality log-risk predictors, based on 1000 bootstrap samples, are displayed in Table 5:

Maps plotted in Fig. 4 show the observed spatiotemporal evolution of COVID-19 mortality risk, and its prediction, from the fitted curve trigonometric regression model, and the subsequent classical and Bayesian time series analysis.

An empirical comparative study

The ML regression models introduced in the Supplementary Material are applied to COVID-19 mortality analysis, and compared, via random k -fold cross-validation and

Fig. 4 COVID-19 mortality risk maps, since March, 8 to May, 13, 2020. Observed (left-hand-side) and estimated (right-hand side) maps, computed from trigonometric regression, combined with classical (first line) and Bayesian (second line) residual predictors

Table 5 Bootstrap confidence intervals for the expected training standard error of the classical and Bayesian residual COVID-19 mortality log-risk predictors ($1 - \alpha = 0.95$)

CI/S	Classical	Bayesian
$\mathcal{I}_1$	[0.0474, 0.0597]	[0.0173, 0.0228]
$\mathcal{I}_2$	[0.0455, 0.0578]	[0.0167, 0.0220]
$\mathcal{I}_3$	[0.0463, 0.0588]	[0.0169, 0.0225]
$\mathcal{I}_4$	[0.0460, 0.0586]	[0.0172, 0.0226]
$\mathcal{I}_5$	[0.0421, 0.0563]	[0.0158, 0.0215]

bootstrap estimators, with the multiple objective space-time forecasting approach presented. We distinguish two categories respectively referred to the strong-sense (hard-data) and weak-sense (soft-data) definition of our data set. Random k -fold ($k = 5, 10$) cross-validation, in terms of Symmetric Mean Absolute Percentage Errors (SMAPEs), evaluates the performance of the compared regression models, from hard- and soft-data. Bootstrap confidence intervals, and probability density estimates of the spatially averaged SMAPEs are also computed. Section 6 provides a data-driven model classification, based on SMAPEs, in the two categories analyzed, from random k -fold cross-validation, and the bootstrap estimation procedures applied.

Results from random k -fold cross-validation

After interpolation and cubic B -spline smoothing of our original data set, the logarithmic transform and linear

scaling are applied. We held out the first ten points and the last three, for each COVID-19 mortality log-risk curve, as an out of sample set. Our approach is implemented in the second-category from soft-data. In this implementation, we consider $N = 6$, adopting the model selection criterion given in equation (8) (see Chapelle et al. 2002). In the multivariate time series classical and Bayesian prediction, our choice of $k(T) = k(265) = 8$ provides a balance between $k(T) = [\ln(T)]^- = [\ln(265)]^- = 5$, signing an agreement with the separation and velocity decay of the empirical eigenvalues of the autocovariance operator, and the parameter value $k(T) = 9$, controlling model complexity according to the sample size $T = 265$. The random fluctuations observed at the $k(T)$ empirical projections of the spatial autocorrelation matrix ρ are also well-fitted by our choice of the shape hyperparameters, characterizing the prior Beta probability density.

Model fitting is evaluated in terms of the Symmetric Mean Absolute Percentage Errors (SMAPEs), given by, for $P = 17$, and $T = 265$,

$$\frac{1}{T} \sum_{t=1}^T \frac{\left| \ln(\widehat{A}_t)(\psi_{p,\rho_p}) - \ln(A_t)(\psi_{p,\rho_p}) \right|}{\left(\left| \ln(A_t)(\psi_{p,\rho_p}) \right| + \left| \ln(\widehat{A}_t)(\psi_{p,\rho_p}) \right| \right) / 2}, \quad p = 1, \dots, P. \quad (17)$$

We have computed the mean of the SMAPEs obtained at each one of the k iterations of the random k -fold cross-validation procedure. This validation technique consists of

Fig. 5 Hard-data category. From 1000 bootstrap samples, spatially averaged SMAPEs histograms and probability densities are plotted, for GRNN (top), MLP (center), and linear SVR (bottom)

random splitting the functional sample into a training and validation samples 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.

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

Table 6 Hard-data category. Averaged SMAPEs, based on 10 running of random 10-fold cross-validation

SC($\times 10^{-2}$)	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 7 Soft-data category. Averaged SMAPEs, based on 10 running of random 10-fold cross-validation

SC($\times 10^{-2}$)	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 8 Our approach. Averaged SMAPEs, based on 10 running of random 10-fold cross-validation, for testing trigonometric regression combined with Classical (C.) and Bayesian (B.) residual analysis

SC	C. k10	B. k10
C1	0.0024	$0.7106(10)^{-3}$
C2	0.0019	$0.4003(10)^{-3}$
C3	0.0016	$0.6797(10)^{-3}$
C4	0.0017	$0.4367(10)^{-3}$
C5	0.0023	$0.6530(10)^{-3}$
C6	0.0018	$0.5854(10)^{-3}$
C7	0.0017	$0.6341(10)^{-3}$
C8	0.0016	$0.6593(10)^{-3}$
C9	0.0013	$0.5979(10)^{-3}$
C10	0.0019	$0.6954(10)^{-3}$
C11	0.0017	$0.5444(10)^{-3}$
C12	0.0016	$0.5016(10)^{-3}$
C13	0.0020	$0.4832(10)^{-3}$
C14	0.0026	$0.6544(10)^{-3}$
C15	0.0023	$0.6616(10)^{-3}$
C16	0.0015	$0.7134(10)^{-3}$
C17	0.0022	$0.6781(10)^{-3}$
M.	0.0019	$0.60524(10)^{-3}$
T.	0.0321	0.0103

specification of the sample splitting in the random k -fold cross-validation procedure.

The ten-running based random 10-fold cross-validation SMAPEs are displayed in Table 6, for the six ML techniques tested, GRNN, MLP, SVR, BNN, RBF, and GP, when hard-data are considered (see also Table 3 of the Supplementary Material on random 5-fold cross-validation results). Table 7 provides the ten-running based random 10-fold cross-validation results, from soft-data category (see also Table 4 of the Supplementary Material on random 5-fold cross validation results). The corresponding cross-validation results of the presented approach from soft-data are displayed in Table 8.

ML model hyperparameter selection has been achieved by applying random k -fold cross-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 random 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 random 10-fold cross-validation implementation. In the same way, we have selected the respective spread β and bandwidth h parameters in the RBF and GRNN procedures. Thus, after applying random

Table 9 Hard-data category. Bootstrap confidence intervals ($1 - \alpha = 0.95$) for the spatially averaged SMAPEs from 1000 bootstrap samples ($T = 265$, $P = 17$)

CI/ML	GRNN	MLP
$\mathcal{I}_1$	$[2.1(10)^{-3}, 4.1(10)^{-3}]$	$[0.5(10)^{-3}, 1(10)^{-3}]$
$\mathcal{I}_2$	$[2(10)^{-3}, 3.9(10)^{-3}]$	$[0.4776(10)^{-3}, 0.9483(10)^{-3}]$
$\mathcal{I}_3$	$[2(10)^{-3}, 4(10)^{-3}]$	$[0.4746(10)^{-3}, 0.9713(10)^{-3}]$
$\mathcal{I}_4$	$[2(10)^{-3}, 4(10)^{-3}]$	$[0.5118(10)^{-3}, 0.9878(10)^{-3}]$
$\mathcal{I}_5$	$[1.7(10)^{-3}, 3.9(10)^{-3}]$	$[0.2780(10)^{-3}, 0.9244(10)^{-3}]$

CI/ML	SVR	BNN
$\mathcal{I}_1$	$[0.3682(10)^{-3}, 0.7219(10)^{-3}]$	$[0.3720(10)^{-3}, 0.6659(10)^{-3}]$
$\mathcal{I}_2$	$[0.3516(10)^{-3}, 0.6763(10)^{-3}]$	$[0.3587(10)^{-3}, 0.6379(10)^{-3}]$
$\mathcal{I}_3$	$[0.3493(10)^{-3}, 0.6770(10)^{-3}]$	$[0.3668(10)^{-3}, 0.6509(10)^{-3}]$
$\mathcal{I}_4$	$[0.3508(10)^{-3}, 0.6865(10)^{-3}]$	$[0.3654(10)^{-3}, 0.6379(10)^{-3}]$
$\mathcal{I}_5$	$[0.3050(10)^{-3}, 0.6661(10)^{-3}]$	$[0.3099(10)^{-3}, 0.6335(10)^{-3}]$

CI/ML	RBF	GP
$\mathcal{I}_1$	$[0.3260(10)^{-3}, 0.5310(10)^{-3}]$	$[0.3243(10)^{-3}, 0.5350(10)^{-3}]$
$\mathcal{I}_2$	$[0.3155(10)^{-3}, 0.5159(10)^{-3}]$	$[0.3065(10)^{-3}, 0.5126(10)^{-3}]$
$\mathcal{I}_3$	$[0.3140(10)^{-3}, 0.5270(10)^{-3}]$	$[0.3095(10)^{-3}, 0.5188(10)^{-3}]$
$\mathcal{I}_4$	$[0.3247(10)^{-3}, 0.5338(10)^{-3}]$	$[0.3152(10)^{-3}, 0.5222(10)^{-3}]$
$\mathcal{I}_5$	$[0.2677(10)^{-3}, 0.5141(10)^{-3}]$	$[0.2505(10)^{-3}, 0.5046(10)^{-3}]$

k -fold cross-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 (see Supplementary Material). Better performance from hard-data is observed in linear SVR. In its implementation, automatic hyperparameter optimization from *ftrsvm* MatLab function is applied. While, from the soft-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 *ftrsvm* MatLab function). In the implementation of GP, we follow the same tuning procedure for model selection. In this case, for both categories, we have selected Bayesian cross-validation optimization (in the hyperparameter optimization argument of the *ftrgp* MatLab function).

In all the results displayed, the SMAPE-MEAN (M.) and SMAPE-TOTAL (T.) have been computed as

Table 10 Soft-data category. Bootstrap confidence intervals ($1 - \alpha = 0.95$) for the spatially averaged SMAPEs from 1000 bootstrap samples ($T = 265$, $P = 17$)

CI/ML	GRNN	MLP
$\mathcal{I}_1$	$[1.3(10)^{-3}, 2.6(10)^{-3}]$	$[0.6(10)^{-3}, 1.3(10)^{-3}]$
$\mathcal{I}_2$	$[1.3(10)^{-3}, 2.6(10)^{-3}]$	$[0.6(10)^{-3}, 1.3(10)^{-3}]$
$\mathcal{I}_3$	$[1.3(10)^{-3}, 2.7(10)^{-3}]$	$[0.6(10)^{-3}, 1.3(10)^{-3}]$
$\mathcal{I}_4$	$[1.3(10)^{-3}, 2.6(10)^{-3}]$	$[0.7(10)^{-3}, 1.3(10)^{-3}]$
$\mathcal{I}_5$	$[1(10)^{-3}, 2.7(10)^{-3}]$	$[0.5(10)^{-3}, 1.3(10)^{-3}]$

CI/ML	SVR	BNN
$\mathcal{I}_1$	$[0.4096(10)^{-3}, 0.8221(10)^{-3}]$	$[0.3588(10)^{-3}, 0.6177(10)^{-3}]$
$\mathcal{I}_2$	$[0.3764(10)^{-3}, 0.7763(10)^{-3}]$	$[0.3433(10)^{-3}, 0.6053(10)^{-3}]$
$\mathcal{I}_3$	$[0.3900(10)^{-3}, 0.7889(10)^{-3}]$	$[0.3454(10)^{-3}, 0.6037(10)^{-3}]$
$\mathcal{I}_4$	$[0.4108(10)^{-3}, 0.7805(10)^{-3}]$	$[0.3559(10)^{-3}, 0.5988(10)^{-3}]$
$\mathcal{I}_5$	$[0.3105(10)^{-3}, 0.7818(10)^{-3}]$	$[0.3003(10)^{-3}, 0.6129(10)^{-3}]$

CI/ML	RBF	GP
$\mathcal{I}_1$	$[0.1794(10)^{-3}, 0.2478(10)^{-3}]$	$[0.2095(10)^{-3}, 0.3248(10)^{-3}]$
$\mathcal{I}_2$	$[0.1754(10)^{-3}, 0.2474(10)^{-3}]$	$[0.2065(10)^{-3}, 0.3215(10)^{-3}]$
$\mathcal{I}_3$	$[0.1785(10)^{-3}, 0.2485(10)^{-3}]$	$[0.2079(10)^{-3}, 0.3262(10)^{-3}]$
$\mathcal{I}_4$	$[0.1743(10)^{-3}, 0.2494(10)^{-3}]$	$[0.2091(10)^{-3}, 0.3258(10)^{-3}]$
$\mathcal{I}_5$	$[0.1616(10)^{-3}, 0.2542(10)^{-3}]$	$[0.1941(10)^{-3}, 0.3232(10)^{-3}]$

CI/OA	Classical	Bayesian
$\mathcal{I}_1$	$[2.2(10)^{-3}, 3.7(10)^{-3}]$	$[0.2943(10)^{-3}, 0.5177(10)^{-3}]$
$\mathcal{I}_2$	$[2.1(10)^{-3}, 3.3(10)^{-3}]$	$[0.2802(10)^{-3}, 0.4854(10)^{-3}]$
$\mathcal{I}_3$	$[2.1(10)^{-3}, 3.3(10)^{-3}]$	$[0.2833(10)^{-3}, 0.4884(10)^{-3}]$
$\mathcal{I}_4$	$[2.2(10)^{-3}, 3.4(10)^{-3}]$	$[0.2900(10)^{-3}, 0.5124(10)^{-3}]$
$\mathcal{I}_5$	$[1.8(10)^{-3}, 3.2(10)^{-3}]$	$[0.2418(10)^{-3}, 0.4664(10)^{-3}]$

performance measures, for comparing the ML models tested, and our approach.

Bootstrap based classification results

For the ML regression models tested, in the hard- and soft-data categories, bootstrap confidence intervals ($1 - \alpha = 0.95$ confidence level) for the spatially averaged SMAPEs, based on 1000 bootstrap samples, are constructed. Our approach requires the soft-data information to be incorporated. As before, the computed bootstrap

Fig. 6 Hard-data category. From 1000 bootstrap samples, spatially averaged SMAPEs histograms and probability densities are plotted, for BNN (top), RBF (center), and GP (bottom)

confidence intervals $\mathcal{I}_i$, $i = 1, \dots, 5$, are respectively based on the bias corrected and accelerated percentile method ($\mathcal{I}_1$); Normal approximated interval with bootstrapped bias

and standard error ($\mathcal{I}_2$); basic percentile method ($\mathcal{I}_3$); bias corrected percentile method ($\mathcal{I}_4$), and Student-based confidence interval ($\mathcal{I}_5$) (see Tables 9 and 10). The bootstrap

Fig. 7 Soft-data category. From 1000 bootstrap samples, spatially averaged SMAPEs histograms and probability densities are plotted, for GRNN (top), MLP (center) and non-linear SVR (bottom)

histogram, and probability density of the spatially averaged SMAPEs are displayed in Figs. 5 and 6, for the hard-data category, and in Figs. 7, 8 and 9, for the soft-data category.

The data-driven performance-based model classification results obtained are discussed in Sect. 6.

Fig. 8 Soft-data category. From 1000 bootstrap samples, spatially averaged SMAPEs histograms and probability densities are plotted, for BNN (top), RBF (center) and GP (bottom)

Fig. 9 Soft-data category. From 1000 bootstrap samples, spatially averaged SMAPEs histograms and probability densities are plotted, for trigonometric regression, combined with empirical-moment based classical (top), and Bayesian (bottom) residual prediction

Final comments

One can observe the agreement between the respective performance-based model classification results, obtained from random k -fold cross-validation, and bootstrap estimation in Sects. 5.1 and 5.2. In the hard-data category, the best performance is displayed by RBF and GP. Similar bootstrapping characteristics are observed for BNN and SVR, with slightly larger values of spatially averaged SMAPEs, reflected in the location of the mode, in the histograms and probability densities displayed in Figs. 5 and 6. These four regression methodologies show a similar degree of variability, regarding the spatially averaged SMAPEs sample values. A higher variability than RBF, GP, BNN and SVR is displayed by the bootstrap sample values of spatially averaged SMAPEs in MLP validation. MLP bootstrapped mode is also slightly shifted to the right. The worst performance corresponds to GRNN (see also Table 6). In the soft-data category, where our approach is incorporated to the empirical comparative study, almost the same empirical ML model ranking holds. Some differences

are found in the bootstrap confidence intervals, and histogram and probability densities computed. For instance, GRNN seems to be favored by soft-data category, while MLP displays worse performance in this category. Hence, smaller differences between GRNN and MPL are displayed in the soft-data category. A slightly improvement in the soft-data category of BNN relative to SVR is observed, preserving almost the same performance. RBF and GP display better performance in the soft-data category, being RBF a bit superior to GP in this category (see Table 10 and Fig. 8). The trigonometric regression, and multivariate time series residual prediction approach based on the empirical moments displays similar results to GRNN, with slightly better performance of GRNN, observed in the bootstrap intervals and histogram/probability density (see Figs. 7 and 9). However, as given in Figs. 8 and 9, the trigonometric regression and Bayesian residual prediction presents almost the same 'performance as BNN, with some slightly better probability distribution features of our approach respect to BNN (see also bootstrap intervals). Our approach is less affected by the random splitting of the sample, in the

implementation of the random k -fold cross validation procedure, since a dynamical spatial residual model is fitted in a second (objective) step. Thus, the proposed multivariate time series classical and Bayesian regression residual modeling fits the short-term spatial linear correlations displayed by the soft-data category. However, the price we pay for increasing model complexity is reflected in the resulting SMAPEs based random k -fold and bootstrap model classification results obtained.

The spatial component effect is reflected in Tables 6 (hard-data), where spatial heterogeneities displayed by random 10-fold cross-validation SMAPEs errors are observed (see also Table 3 in the Supplementary Material). While Table 7 (see also Table 4 in the Supplementary Material) reveals the benefits obtained in some of the ML regression models tested from soft-data information. Particularly, in this category, possible spatial linear correlations are incorporated to the analysis, in terms of soft-data.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00477-021-02021-0>.

Acknowledgements This work has been supported in part by projects PGC2018–099549-B-I00 of the Ministerio de Ciencia, Innovación y Universidades, Spain (co-funded with FEDER funds), and by grant A-FQM-345-UGR18 cofinanced by ERDF Operational Programme 2014–2020, and the Economy and Knowledge Council of the Regional Government of Andalusia, Spain.

The subject of this paper was originally developed, in a first stage, under the seminars hold in the *Unidad de Transferencia del IMUS* about *Matemáticas y la COVID*. We also thank the organizer, Professor Emilio Carrizosa.

Author Contributions Not applicable

Funding This work has been supported in part by projects PGC2018–099549-B-I00 of the Ministerio de Ciencia, Innovación y Universidades, Spain (co-funded with FEDER funds), and by grant A-FQM-345-UGR18 cofinanced by ERDF Operational Programme 2014–2020, and the Economy and Knowledge Council of the Regional Government of Andalusia, Spain. (In section Acknowledgements below, all sources of funding are also declared).

Data availability The Supplementary Material uploaded contains the plots of the observed COVID-19 mortality cumulative cases curves, as well as of the corresponding COVID-19 mortality log-risk curves. The rest is not applicable.

Declarations

Conflict of interest The author(s) declare(s) that they have no competing interests.

References

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

- Abboud C, Bonnefon O, Parent E, Soubeyrand S (2019) Dating and localizing an invasion from post-introduction data and a coupled reaction-diffusion-absorption model. *J Math Biol* 79:765–789
- Agostinelli C (2001) Robust model selection in regression via weighted likelihood methodology. *Stat Probab Lett* 56:289–300
- Alpaydin E (2004) Introduction to machine learning. MIT Press, Cambridge, MA
- Anderson H, Britton T (2000) Stochastic epidemic models and their statistical analysis. Springer-Verlag, New-York
- Angulo J, Yu H-L, Langousis A, Kolovos A, Wang J, Madrid AE, Christakos G (2013) Spatiotemporal infectious disease modeling: a BME-SIR approach. *PLoS One* 8(9):e72168
- Barstugan M, Ozkaya U, Ozturk S (2020). Coronavirus (COVID–19) classification using ct images by machine learning methods. arXiv preprint [arXiv:2003.09424](https://arxiv.org/abs/2003.09424)
- Beretta EE, Hara T, Ma W, Takeuchi Y (2001) Global asymptotically stability of an SIR epidemic model with distributed time delay. *Nonlinear Anal Theor Methods Appl* 47:4107–4115
- Blanquero R, Carrizosa E, Jiménez-Cordero MA, Martín-Barragán B (2020) Selection of time instants and intervals with support vector regression for multivariate functional data. *Comput Oper Res*. <https://doi.org/10.1016/j.cor.2020.105050>
- Bolker BM, Grenfell B (1996) Impact of vaccination on the spatial correlation and persistence of measles dynamics. *Proc Natl Acad Sci* 93:12648–12653
- Bosq D (2000) Linear processes in function spaces. Lecture notes in statistics **149**. Springer, New-York
- Bosq D, Ruiz-Medina MD (2014) Bayesian estimation in a high dimensional parameter framework. *Electron J Statist* 8:1604–1640
- Chao DL, Bloom JD, Kochin BF, Antia R, Longini IM (2012) The global spread of drug-resistant influenza. *J Royal Soc Interf* 9:648–656
- Chapelle O, Vapnik V, Bengio Y (2002) Model selection for small sample regression. *Mach Learn* 48:9–23
- Chien L-Ch, Chen L-W (2020) Meteorological impacts on the incidence of COVID-19 in the U.S. *Stoch Environ Res Risk Assess* 34:1675–1680
- Christakos G (2000) Modern spatiotemporal geostatistics. Oxford University Press, New-York, NY
- Christakos G (2002) On assimilation of uncertain physical knowledge bases: Bayesian and non-Bayesian techniques. *Adv. Water Resour* 25:1257–1274
- Christakos G, Bogaert P, Serre ML (2002) Advanced functions of temporal GIS. Springer-Verlag, New York, N.Y
- Christakos G, Hristopulos DT (1998) Spatiotemporal environmental health modelling: a Tractatus Stochasticus. Kluwer, Boston
- Christakos G (2008) Bayesian maximum entropy advanced mapping of environmental data: geostatistics, machine learning, and Bayesian maximum entropy. Wiley, New York, NY, pp 247–306
- Daubechies I (1988) Orthonormal basis of compactly supported wavelets. *Comm Pure Appl Math* 41:909–9996
- Du Z, Xu X, Wu Y, Wang L, Cowling LA, Meyers BJ (2020) Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg Infect Dis* 26(6)
- Dushoff J, Plotkin J, Levin S, Earn D (2004) Dynamical resonance can account for seasonality of influenza epidemics. *Proc Natl Acad Sci USA* 101:16915–16916
- Elhia M, Laaroussi A, Rachik M, Rachik Z, Labriji E (2014) Global stability of a susceptible-infected-recovered (SIR) epidemic model with two infectious stages and treatment. *Int J Sci Res* 3:114–121
- Fleming TR, Harrington DP (1991) Counting processes and survival analysis. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New-York

- Guin LN, Mandal PK (2014) Spatiotemporal dynamics of reaction-diffusion models of interacting populations. *Appl Math Model* 38:4417–4427
- Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning. Springer Series in Statistics. Springer-Verlag, New-York
- He J, Chen G, Jiang Y, Jin R, Shortridge A, Agusti S, Hea M, Wua J, Duarte CM, Christakos G (2020) Comparative infection modeling and control of COVID-19 transmission patterns in China, South Korea, Italy and Iran. *Science of the Total Environment* 747
- Huppert A, Katriel G (2013) Mathematical modelling and prediction in infectious disease epidemiology. *Clin Microbiol Infect* 19:999–1005
- Ivanov AV, Leonenko NN, Ruiz Medina MD, Zhurakovsky BM (2015) Estimation of harmonic component in regression with cyclically dependent errors. *Stat J Theor Appl Stat* 49:156–186
- Ivorra B, Ferrández MR, Vela-Pérez M, Ramos AM (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
- Ivorra B, Ramos AM, Ngom D (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
- Ji C, Jiang D, Shi N (2012) The behavior of an SIR epidemic model with stochastic perturbation. *Stoch Anal Appl* 30:755–773
- Keeling MJ, Rand DA, Morris AJ (1997) Correlation models for childhood epidemics. *Proc Royal Soc London* 264:1149–1156
- Keeling MJ, Rohani P (2008) Modeling infectious diseases in humans and animals. Princeton University Press, Princeton
- Kermack W, McKendrick A (1927) Contributions to the mathematical theory of epidemics - I. *Proc Royal Soc Edinburgh A* 115:700–721
- Khan MA, Atangana A (2020) Modeling the dynamics of novel coronavirus (2019-nCov) with fractional derivative. *J Alex Eng.* <https://doi.org/10.1016/j.aej.2020.02.033>
- Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S et al (2020) Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis.* [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4)
- Kuznetsov YA, Piccardi C (1994) Bifurcation analysis of periodic SEIR and SIR epidemic models. *J Math Biol* 32:109–121
- Laaroussi AE, Rachik M, Elhia M (2018) An optimal control problem for a spatiotemporal SIR model. *Int J Dynam Control* 6:384–397
- Langousis A, Carsteau AA (2020) Undersampling in action and at scale: application to the COVID-19 pandemic. *Stoch Environ Res Risk Assess* 34:1281–1283
- Malesios C, Demiris N, Kostoulas P, Dadousis K, Koutroumanidis T, Abas Z (2016) Spatio-temporal modelling of foot-and-mouth disease outbreaks. *Epidemiol Infect* 144:2485–2493
- McCluskey CC (2010) Complete global stability for an SIR epidemic model with delay distributed or discrete. *Nonlinear Anal Real World Appl* 11:55–59
- Milner FA, Zhao R (2008) SIR model with directed spatial diffusion. *Math Popul Stud* 15:160–181
- Mohammady M, Reza Pourghasemi H, Amiri M, Tiefenbacher JP (2021). Spatial modelling of susceptibility to subsidence using machine learning techniques. <https://doi.org/10.1007/s00477-020-01967-x>
- Nishiura H, Linton NM, Akhmetzhanov AR (2020) Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis* 93:284–286
- Pak D, Langohr K, Ning J, Cortés Martínez J, Gómez-Melis G, Shen Y (2020) Modeling the coronavirus disease 2019 incubation period: impact on quarantine policy <https://doi.org/10.1101/2020.06.27.20141002>
- Pathak S, Maiti A, Samanta G (2010) Rich dynamics of an SIR epidemic model. *Nonlinear Anal Model Control* 15:71–81
- Ramosa AM, Ferrández MR, Vela-Pérez M, Ivorra B (2020) A simple but complex enough θ -SIR type model to be used with COVID-19 real data. Application to the case of Italy <https://doi.org/10.13140/RG.2.2.32466.17601>
- Remuzzi A, Remuzzi G (2020) COVID-19 and Italy: What next? *The Lancet.* [https://doi.org/10.1016/S0140-6736\(20\)30690-5](https://doi.org/10.1016/S0140-6736(20)30690-5)
- Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J 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
- Roques L, Bonnefon O (2016) Modelling population dynamics in realistic landscapes with linear elements: a mechanistic-statistical reaction-diffusion approach. *PloS One* 11(3):e0151217
- Roques L, Soubeyrand S, Rousselet J (2011) A statistical-reaction-diffusion approach for analyzing expansion processes. *J Theor Biol* 274:43–51
- Sekiguchi M, Ishiwata E (2010) Global dynamics of a discretized SIRS epidemic model with time delay. *J Math Anal Appl* 371:195–202
- Sivakumar B (2020) COVID-19 and water. *Stoch Environ Res Risk Assess.* <https://doi.org/10.1007/s00477-020-01837-6>
- Sujath R, Chatterjee JM, Hassanien AE (2020) A machine learning forecasting model for COVID-19 pandemic in India. *Stoch Environ Res Risk Assess.* 34:959–972
- Sujath R, Chatterjee JM, Hassanien AE (2020) Correction to: a machine learning forecasting model for COVID-19 pandemic in India. *Stoch Environ Res Risk Assess.* <https://doi.org/10.1007/s00477-020-01843-8>
- Takano Y, Miyashiro R (2020) Best subset selection via cross-validation criterion. *TOP* 28:475–488
- Tornatore E, Buccellato SM, Vetro P (2005) Stability of a stochastic SIR system. *Phys A Stat Mech Its Appl* 354:111–126
- Volz E (2008) SIR dynamics in random networks with heterogeneous connectivity. *J Math Biol* 56:293–310
- Wang C, Horby PW, Hayden F, Gao GF (2020) A novel coronavirus outbreak of global health concern. *Lancet* 395:470–473
- Wasiur RK, Choij B, Kenahz E, Rempa GA (2019) Survival dynamical systems for the population-level analysis of epidemics. [arXiv.1901.00405](https://arxiv.org/abs/1901.00405)
- Webb G (1981) A reaction-diffusion model for a deterministic diffusive epidemic. *J Math Anal Appl* 84:150–161
- Xu Y, Allena L, Perelson A (2007) Stochastic model of an influenza epidemic with drug resistance. *J Theor Biol* 248:179–193
- Yu J, Jiang D, Shi N (2009) Global stability of two-group SIR model with random perturbation. *J Math Anal Appl* 360:235–244
- Zhang Wb, Ge Y, Liu M et al (2020) Risk assessment of the step-by-step return-to-work policy in Beijing following the COVID-19 epidemic peak. *Stoch Environ Res Risk Assess.* <https://doi.org/10.1007/s00477-020-01929-3>
- Zhang F, Li Z, Zhang F (2008) Global stability of an SIR epidemic model with constant infectious period. *Appl Math Comput* 199:285–291
- Zhou T, Fu Z, Wang B (2006) Epidemic dynamics on complex networks. *Prog Nat Sci* 16:452–457
- Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X 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.* [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.