



Adaptive LASSO estimation for functional hidden dynamic geostatistical models

Paolo Maranzano^{1,2} · Philipp Otto³ · Alessandro Fassò⁴

Accepted: 1 May 2023

© The Author(s) 2023

Abstract

We propose a novel model selection algorithm based on a penalized maximum likelihood estimator (PMLE) for functional hidden dynamic geostatistical models (f-HDGM). These models employ a classic mixed-effect regression structure with embedded spatiotemporal dynamics to model georeferenced data observed in a functional domain. Thus, the regression coefficients are functions. The algorithm simultaneously selects the relevant spline basis functions and regressors that are used to model the fixed effects. In this way, it automatically shrinks to zero irrelevant parts of the functional coefficients or the entire function for an irrelevant regressor. The algorithm is based on an adaptive LASSO penalty function, with weights obtained by the unpenalised f-HDGM maximum likelihood estimators. The computational burden of maximisation is drastically reduced by a local quadratic approximation of the log-likelihood. A Monte Carlo simulation study provides insight in prediction ability and parameter estimate precision, considering increasing spatiotemporal dependence and cross-correlations among predictors. Further, the algorithm behaviour is investigated when modelling air quality functional data with several weather and land cover covariates. Within this application, we also explore some scalability properties of our algorithm. Both simulations and empirical results show that the prediction ability of the penalised estimates are equivalent to those provided by the maximum likelihood estimates. However, adopting the so-called one-standard-error rule, we obtain estimates closer to the real ones, as well as simpler and more interpretable models.

Keywords Functional HDGM · Adaptive LASSO · Model selection · Penalized maximum likelihood · Geostatistical models · Air quality Lombardy

1 Introduction

In recent years, we experienced a significant increase in the availability and size of geo-referenced datasets, especially in air quality monitoring (Vitolo et al. 2016), agriculture and livestock farming (Fassò, A., Rodeschini, J., Moro, A.F., Shaboviq, Q., Maranzano, P., Cameletti, M., Otto, P. 2023; Brown et al. 2023), and climate (Cruz-Alonso et al. 2023; Czernecki et al. 2020). As a result, geostatistical applications need efficient algorithms for variable selection, that is for selecting relevant predictors from a large set of candidates.

In addition, geostatistical data often defined on a functional domain because of their characteristics, e.g., vertical atmospheric profiles in climate studies (Fassò et al. 2018) or off-shore coastal profile measurements for beach monitoring (Otto et al. 2021). A functional data approach may also be used to reduce the dimensionality of high-frequency temporal observations. For example, Ignaccolo

✉ Paolo Maranzano
paolo.maranzano@unimib.it

Philipp Otto
philipp.otto@ikg.uni-hannover.de

Alessandro Fassò
alessandro.fasso@unibg.it

¹ Department of Economics, Management and Statistics (DEMS), University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milano, Italy

² Fondazione Eni Enrico Mattei (FEEM), Corso Magenta 63, 20123 Milano, Italy

³ Institute of Cartography and Geoinformatics (IKG), Leibniz University of Hannover, Appelstrasse 9a, 30167 Hannover, Lower Saxony, Germany

⁴ Department of Economics, University of Bergamo, Via dei Caniana 2, 24127 Bergamo, Italy

et al. (2008) considered the time series of air quality measurements at many stations as functional observations. Also, to understand the bike-sharing system, Piter et al. (2022) considered daily 5-min usage profiles of a bike-sharing system as daily functional observations. Due to the spatial nature of the underlying process, further applications can be found in environmetrics (e.g., Franco-Villoria and Ignaccolo 2017; Ignaccolo et al. 2013; Giraldo et al. 2011), medicine (e.g., Aristizabal et al. 2019), econometrics (e.g., Pineda-Ríos et al. 2019). As a result, the above mentioned need for efficient variable selection should cover also functional data models.

1.1 The *Big-N-problem*

In this context, statistical methods have to face the so-called *Big-N-problem*, in which the time complexity of estimation algorithms grows polynomially with an order greater than 2 when the number of locations is increasing and traditional methods are often not computationally feasible (cf. Katzfuss 2017a; Katzfuss and Cressie 2011). To reduce the complexity of such models, various approaches have been used, some of which are based on inducing sparsity in the spatial covariance matrix (Furrer et al. 2006; Kaufman et al. 2008; Stein 2013; Furrer et al. 2016). Some other approaches are related to the precision matrix, either using a graphical least absolute shrinkage and selection operator (LASSO) algorithm (Krock et al. 2021, 2021), or a sparse Cholesky factors approach based on the Vecchia approximations (Stein et al. 2004; Kang and Katzfuss 2021; Schäfer et al. 2021) and on multi-resolution approximations of Gaussian processes (Katzfuss 2017b; Jurek and Katzfuss 2021). In particular, Vecchia approximation can be efficiently used to perform high-dimensional spatiotemporal filtering (Jurek and Katzfuss 2022) and spatiotemporal smoothing (Jurek and Katzfuss 2022b). Low-rank covariance matrices have been also considered, including fixed-rank kriging and penalised methods (Banerjee et al. 2008; Cressie and Johannesson 2008; Chang et al. 2010a; Hsu et al. 2012a; Cressie et al. 2010). Eventually, combined approaches, like the so-called *full-scale approximation of the covariance matrix* have been proposed (Sang and Huang 2012).

1.2 Geostatistical variable selection

For large regression models, joint estimation and variable selection based on penalised estimation provides stable solutions compared to the classic backward and forward selection methods (Breiman 1996; Bondell et al. 2010). In this regard, we refer to the review paper of Müller and Welsh (2010).

Methods of selecting covariates have been developed based on penalised methods in spatial and spatiotemporal settings. For instance, Wang and Zhu (2009) suggested a penalised least squares approach for spatial regression models; Cai and Maiti (2020), for spatial autoregressive models; and Gonella et al. (2022), for conditional autoregressive models. For additive spatial models including potential nonlinear effects, Nandy et al. (2017) developed a weighted penalised least squares estimator. Alternatively, penalised maximum likelihood estimators (PMLE) are considered. For instance, Zhu et al. (2010) suggested PMLE for linear models with spatially correlated errors. Chu et al. (2011a) and Chu et al. (2011b) additionally reduced the model's complexity by combining covariance tapering and a PMLE for spatial and spatiotemporal settings, respectively.

It is important to note that geostatistical applications are prone to cross-correlated regressors due to their spatial nature. Cross-correlation among the regressors can be a critical issue in model selection. Indeed, when the covariates are correlated, as pointed out by Zhao and Yu (2006), the classic LASSO approach would generally not be selection-consistent. Similarly, Simon and Tibshirani (2012) showed that in the same situation, the group-LASSO estimator (cf. Yuan and Lin (2006)), which assumes orthonormal data within each group, would perform poorly in selecting the relevant features. Such issues motivate the choice of an adaptive LASSO penalty, which led to selection-consistent estimators in the case of cross-correlated regressors (see Zou 2006; Zou and Li 2008).

Following the least absolute shrinkage and selector operator (LASSO) methodology (Tibshirani 1996; Reyes et al. 2012) proposed a spatiotemporal adaptive LASSO algorithm for linear regression models with spatiotemporal neighbourhood structures. The estimation strategy involved both the penalised least squares and PMLE approaches. Other examples of penalised regression for spatiotemporal data are in Al-Sulami et al. (2019), in which an adaptive LASSO method was developed to simultaneously identify and estimate spatiotemporal lag interactions in the context of a data-driven semiparametric nonlinear model. Furthermore, Safikhani et al. (2020) considered LASSO methods for generalised spatiotemporal autoregressive models. The estimators are obtained by a modified version of the penalised least squares that accommodates hierarchical group LASSO-type penalties.

In general, spline basis functions are widely used tools in geostatistics for the spatiotemporal interpolation of environmental phenomena (see for group-LASSO approaches in this context Hofierka et al. 2002; Xiao et al. 2016; Chang et al. 2010b; Hsu et al. 2012b). Also, spatial and spatiotemporal model selection has been addressed in a Bayesian framework (see e.g., Katzfuss and Cressie 2012;

Carroll et al. 2016, 2016; Lawson et al. 2017; Carroll et al. 2018) but it will not be the focus of this paper.

Penalised methods are also commonly applied in the context of functional data analysis, such as penalised splines (see Silverman and Ramsay 2002; Claeskens et al. 2009). These methods usually regularise the smoothness of the estimated functions by penalising the integrated second derivatives. In this way, many basis functions can be used, thus avoiding the typical overfit resulting from unpenalised estimation methods.

In the context of functional data, some authors have proposed the use of penalty methods for selecting relevant predictors. As an example, Pannu and Billor (2017) proposed using group LASSO for selecting grouped variables (functional predictors) rather than individual variables; Ivanoff et al. (2016) proposed adaptive LASSO and group LASSO estimators for functional Poisson regression; Zhao et al. (2012) employed a wavelet-based LASSO approach for regressing scalars on functions aimed at identifying a relatively small number of non-zero wavelet coefficients; eventually, Centofanti et al. (2022) proposed a smooth LASSO able to locate regions where the coefficient function is zero, and to smoothly estimate non-zero values of the coefficient function. These contributions, although close to us in terms of purpose and methodology, do not consider the presence of spatiotemporal dependence and cross-correlation between covariates, which are the main focus of this paper.

1.3 Variable selection for spatiotemporal functional models

In this paper, we propose a penalised maximum likelihood estimate for the functional spatiotemporal model known as the *functional hidden dynamics geostatistical model* (f-HDGM) (Wang et al. 2021). More precisely, we develop an adaptive LASSO method able to estimate the relevant model coefficients and shrink to zero the irrelevant ones, while taking into account spatiotemporal correlation and cross-correlation among predictors.

In our functional setting, the LASSO selected coefficients are associated with the spline bases of the functional regressors. As a result we may have two cases. All the coefficients of a certain regressor are shrunk to zero, resulting in its drop out. The second and more subtle case, happens if only some of the coefficients of a single variable are shrunk to zero. In this case the functional coefficient may be zero in a subset of the functional domain. Similarly, when using periodic Fourier bases some frequencies for a specific variable may be neglected. For instance, Otto et al. (2021) showed that major storm floods have an effect only on specific parts of the coastal profiles, that is, those affected by high waves during a flood.

We test the performance of the algorithm through a Monte Carlo simulation study based on three settings with increasing level of complexity and representative of geostatistical applications. Furthermore, we apply the penalisation algorithm to an empirical example of air quality assessment. Within the application we study the computation time of the phases composing the penalty algorithm and its behaviour as the model complexity increases. Both simulations and applications are evaluated by highlighting the predictive ability of the penalised estimators, the interpretability of the estimates, the precision of parameter estimation, and the variable selection capability.

Our proposal extends the approach of Fassò et al. (2022), which considered a two stage variable selection algorithm approach for multivariate HDGM. In the first stage the classic trace LASSO is applied to the multivariate response variable without considering any spatiotemporal structure. This provides the so-called active sets associated to the penalty factors λ . In the second stage, for each active set the multivariate HDGM is estimated and tested using cross-validation. This provides the optimal active set and the corresponding selected variables according to an hybrid criterion. Here, considering functional response and predictors, we propose a penalized maximum likelihood method based on a second order approximation of the fHDGM likelihood.

The remainder of the paper is structured as follows. In Sect. 2, we briefly introduce the considered functional geostatistical model, namely the f-HDGM. In Sect. 3, we present the new penalised maximum likelihood approach using an adaptive LASSO penalty. In Sect. 4, we present the results of an extensive Monte Carlo simulation of three simulation settings. In Sect. 5, we illustrate an empirical application in which the penalisation algorithm is applied to daily air quality profiles in Lombardy, a region in Northern Italy. Finally, Sect. 6 concludes this paper and identifies potential topics for future research.

2 The functional model

In this section, we review the functional spatiotemporal model known as the functional hidden dynamics geostatistical model. The modelling rationale may be rooted in the state-space modelling approach (Ferreira et al. 2022; Jurek and Katzfuss 2021, 2022b). It is based on the classic idea so that the temporal dynamics is described by a fixed effect component plus an unobserved Markovian component with innovations spatially correlated. In this framework, the spatiotemporal covariance is assumed to be separable (see for a comparison of different spatiotemporal models Huang et al. 2007). The multivariate model and its maximum likelihood estimation are introduced by Calulli

et al. (2015). The procedure is based on maximising the likelihood function using an expectation-maximization (EM) algorithm, which is efficient thanks to the Kalman Filter algorithm (Rougier et al. 2023).

2.1 Model details

The f-HDGM is designed to handle functional data $\{y_{s,t}(h) : s \in D, t = 1, \dots, T\}$ defined on the interval $H = [h_1, h_2]$. That is, $y_{s,t}(h) : H \rightarrow \mathbb{R}$ can be observed at any $h \in H$ for any given location s in the spatial domain D and for any given discrete time t . Although the spatial domain D is continuous, we observe such data on n spatial points in an irregular grid $S = \{s_1, \dots, s_n\}$. Similarly, we observe the data for each function at a discrete set of points h_1, \dots, h_q , where both h_i and q may depend on s_i and t . These observations are denoted by vectors $\{y_{s,t} = (y_{s,t}(h_1), \dots, y_{s,t}(h_q))\}$. Overall, our data set is composed of $N = nT$ functional data.

To account for the spatial and temporal dependence, we model the process using a hidden dynamics geostatistical model that separates all regressive effects from the spatiotemporal interrelations. More precisely, the f-HDGM is defined by

$$y_{s,t}(h) = \mu_{s,t}(h) + \omega_{s,t}(h) + \varepsilon_{s,t}(h), \quad (1)$$

where the fixed effects component $\mu_{s,t}(h)$, the random effects component $\omega_{s,t}(h)$ and the modelling errors variance $\sigma^2(h) = \text{Var}(\varepsilon_{s,t}(h))$ are modelled using splines.

Let $B_{k,a}(h)$ be the k -th of the K_a basis functions of component $a \in \{\mu, \omega, \sigma\}$. In Eq. 1, the mean, or the fixed effects component, is a linear regression model in the functional domain. That is,

$$\mu_{s,t}(h) = \sum_{j=0}^p \sum_{k=1}^{K_\mu} X_{s,t,j}(h) B_{k,\mu}(h) \beta_{jk}, \quad (2)$$

where $X_{s,t,j}(h)$ denotes the functional observations of the j -th regressor, with $j = 0$ referring to the functional intercept, and $j = 1, \dots, p$ referring to the p functional covariates. For the generic j -th regressor, by multiplying of the spline basis matrix by the coefficients $\beta_j = (\beta_{j1}, \dots, \beta_{jK_\mu})'$, we obtain the functional coefficients shown in Fig. 1. In Sect. 3, we propose an adaptive LASSO procedure to penalize these regression coefficients. In a nutshell, whether all entries of the vector β_j are shrunk to zero or not, we can select the relevant regressors. That is if β_j contains only zeros, then, the j -th regressor is removed from the model. Moreover, if β_j is only partly shrunk to zero, we can select the relevant parts and knots of the j -th regressor in the functional domain.

In Eq. 1, the spatiotemporal dependence is modelled by the functional random effects $\omega_{s,t}(h)$, given by

$$\omega_{s,t}(h) = \sum_{k=1}^{K_\omega} B_{k,\omega}(h) z_{s,t,k}. \quad (3)$$

In Eq. 3, the latent component $z_{s,t} = (z_{s,t,1}, \dots, z_{s,t,K_\omega})$ follows a temporal Markovian process, i.e.,

$$z_{s,t} = G z_{s,t-1} + \eta_{s,t}, \quad (4)$$

where $\eta_{s,t}$ is a spatially correlated K_ω -dimensional zero-mean Gaussian process

$$\eta_{s,t} \sim N_{K_\omega}(0, \Gamma). \quad (5)$$

Let $\rho(d, \theta)$ be the exponential covariance function at distance d with spatial range θ . Then, the spatial covariance function Γ at location s and s' is given by

$$\Gamma = \text{diag}(v_1 \rho(s - s'; \theta_1), \dots, v_{K_\omega} \rho(s - s'; \theta_{K_\omega})), \quad (6)$$

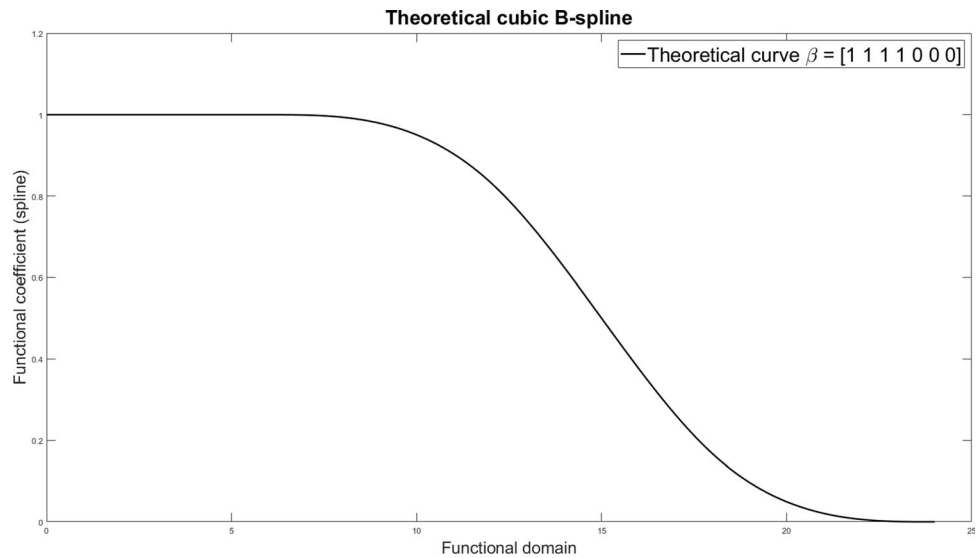
where v_i and θ_i are the variance and the range of the j -th component of $\eta_{s,t}$, respectively, with $i = 1, \dots, K_\omega$.

Eventually, the model errors $\varepsilon_{s,t}$ are assumed to be independent and identically normally distributed across space and time, but the error variance may vary across the functional domain as follows:

$$\sigma^2(h) = \sum_{k=1}^{K_\sigma} B_{k,\sigma}(h) \sigma_k^2. \quad (7)$$

Let $\beta = (\beta_1, \dots, \beta_p)'$ be the stacked vector of the spline coefficient vectors of the fixed effects model, let $\theta = (\theta_1, \dots, \theta_{K_\omega})'$ be the stacked vector of the spatial ranges, and let $v = (v_1, \dots, v_{K_\omega})'$ be the stacked vector of random effects variances. Also, let $\psi = \{G, V, \theta, v, \sigma^2\}$ be the set of all coefficients of the random effects, including the error term. Moreover, let H denote the Hessian matrix of the model's log-likelihood. The full set of parameters $\{\beta, \psi\}$ is estimated by maximising the log-likelihood using the EM algorithm. Let $\{\hat{\beta}_{MLE}, \hat{\psi}_{MLE}\}$ denote the maximum likelihood estimate of $\{\beta, \psi\}$. Moreover, let $H_{MLE} = H(\hat{\beta}_{MLE}, \hat{\psi}_{MLE})$ denote the Hessian matrix computed at the ML solutions $\{\hat{\beta}_{MLE}, \hat{\psi}_{MLE}\}$. The EM algorithm used for computation is implemented in the D-STEMv2 software (Finazzi and Fassò 2014; Wang et al. 2021) within the MATLAB environment. Starting with initial parameters $\{\beta^{(0)}, \psi^{(0)}\}$, the expectation of the complete data likelihood of the f-HDGM is iteratively maximised until the parameters converge or an upper bound of iterations is reached. Both, the convergence distance and the iteration limit, can be specified by the user. Moreover, the M-step involves the Kalman smoother estimating the random effects and possibly missing values. For details on the update steps, we

Fig. 1 Cubic B-splines functional coefficient with seven basis functions and coefficients equal $\beta = (1, 1, 1, 1, 0, 0, 0)'$ used in the simulations



refer the reader to Calculli et al. (2015) which is the basis of the algorithm implemented in D-STEMv2.

2.2 Approximations

Estimating the parameters of the f-HDGM can be computationally demanding. Following Wang et al. (2021), we will consider two approximations in order to reduce the computational time. In the first approximation, the variance-covariance matrix of the parameters is computed using an approximated approach. This task is performed by fixing a threshold for the overall improvement in the variance-covariance matrix computation (see Sect. 2.5 of Wang et al. 2021).

The second approximation concerns a spatial partitioning approach. According to Stein (2013), we divide the complete dataset into k groups (based on the geodesic distance) of size r , and assume that the data in the different groups are not correlated. This implies a factorised likelihood function and the possibility of computing the E-step in parallel. As a result, the computational complexity is reduced from $O(Tn^3b^3)$ to $O(Tkr^3b^3)$ (see Sect. 2.4 of Wang et al. 2021). Moreover, if the computing infrastructure can handle k parallel processes, the computing time may be further reduced to $O(Tr^3b^3)$.

3 Spatiotemporal adaptive LASSO estimation for functional coefficients

In this section, we suggest an adaptive LASSO approach to select (1) the relevant regressors, (2) the relevant sections of the functional coefficients and (3) the relevant knots of the fixed effects functional model $\mu_{s,t}(h)$. The emphasis is

on modelling the relationship between the covariates and the response variable. Therefore, the parameters of the random effects components are kept unpenalised. Moreover, for regularised regression approaches, the covariance matrix of the model errors is usually not penalised (e.g., Fan and Li 2001; Tibshirani 1996).

Spatiotemporal parameters could also be included in the penalised procedure (e.g., see for random effects shrinking in linear mixed models Bondell et al. 2010). However, in this case, the shrinkage target should be adjusted to the specific empirical case. Indeed, while the temporal dependence parameter matrix G could be shrunk to zero, i.e., in the case of temporal independence, a zero shrinkage target is not meaningful for the variance parameters and the range parameter of the spatial dependence θ .

We follow a penalised maximum likelihood estimation strategy for the fixed effects coefficients conditional on the random effects parameters, i.e.,

$$\hat{\beta}_{PMLE}(\lambda, \psi) = \arg \max_{\beta} \mathcal{L}(\beta, \psi) - \lambda f(\beta) \tag{8}$$

with the likelihood function \mathcal{L} and a penalty function f . To reduce the computational burden, we locally approximate the full model log-likelihood \mathcal{L} in (8) around the unpenalised and consistent ML estimates using a second-order local quadratic approximation (Jennrich and Sampson 1976; Longford 1987). That is, we obtain the approximated log-likelihood as follows:

$$\mathcal{L}(\beta) \cong \mathcal{L}(\hat{\beta}_{MLE}) + \frac{1}{2}(\beta - \hat{\beta}_{MLE})' H_{MLE}(\beta - \hat{\beta}_{MLE}), \tag{9}$$

where $H_{MLE} = \nabla^2 \mathcal{L}(\hat{\beta}_{MLE}, \hat{\psi}_{MLE})$.

Similar computational solutions involving local approximation of the likelihood have been proposed by Zou and Li (2008) for obtaining penalised estimates of the

parameters in Generalised Linear Models (GLM) via the one-step sparse estimator by Fan and Li (2001) for variable selection adopting nonconcave penalties; by McIlhagga (2016) for penalized GLM based on Fisher scoring algorithms; by Zhu et al. (2010) for adaptive spatial LASSO in lattice data; and by Reyes et al. (2012) for penalised likelihood problems in linear spatiotemporal contexts. This study extends the aforementioned studies by obtaining penalised estimates of the fixed effects coefficients of a linear mixed model for functional data, with the spatiotemporal dynamics modelled by a geostatistical random component.

We now consider the penalty function $f(\beta)$ in Equation (8). Motivated by the oracle properties of the adaptive LASSO estimates (Zou 2006; Bondell et al. 2010), we use an adaptive penalty for the likelihood of the functional HDGM. Because the observed data are supposed to be correlated in space and time, it is important that the algorithm be selection-consistent (Zhang 2010) even in the case of correlated observations. However, this may not often be the case for classic LASSO approaches (see among others, for conditions of selection consistency Zhao and Yu 2006). Thus, we suggest an adaptive LASSO penalty that has the desired property of selection consistency, as shown by Zou (2006); Zou and Li (2008); Huang et al. (2008).

We propose the following estimator with an adaptive LASSO penalty for the fixed effects coefficients of the f-HDGM:

$$\hat{\beta}_{PMLE}(\lambda; \hat{\psi}_{MLE}) = \arg \min_{\beta} -\frac{1}{2}(\beta - \hat{\beta}_{MLE})' H_{MLE}(\beta - \hat{\beta}_{MLE}) + N\lambda \|\omega \circ \beta\|_1 \tag{10}$$

where λ is the regularisation parameter, \circ is the element-wise product, and the penalty weights w are chosen as the inverse initial ML estimates, that is, $w = (w_i)_{i=1,\dots,p}$ with $w_i = \frac{1}{|\hat{\beta}_{MLE,i}|^\gamma}$, with $\gamma = 1$ and $\hat{\beta}_{MLE,i}$ being the i -th entry of $\hat{\beta}_{MLE}$. To increase or diminish the influence of the initial estimates, $\gamma \geq 0$ could also be chosen differently. Generally, to obtain the oracle properties, the penalty parameter λ should be of order \sqrt{n} (see Zou and Li 2008). In the next paragraph, we will go into further details on the selection of λ .

The algorithm used to solve minimisation in (10) is based on the BFGS quasi-Newton method over the non-zero coefficients, that is, the so called active set, with the initial values being $\hat{\beta}_{MLE}$. The algorithm requires limited computation effort, as the time consuming computation of the Hessian matrix H_{MLE} used in the approximation (9) is done only once, and the second-order derivatives can be computed numerically throughout the optimisation. Notice

that the dimension of problem (10) is much smaller compared to the full model MLE as it involves only β , while ψ is kept fixed at $\hat{\psi}_{MLE}$.

The penalised procedure shrinks irrelevant coefficients to zero. Because this applies to all basis functions separately, we do not impose that all coefficients associated with one regressor must be shrunk to zero simultaneously as for a block LASSO approach. It is then possible to select the relevant sections of a functional coefficient and exclude the irrelevant knots. However, the basis functions may overlap to some extent. This implies that the height of the functional coefficient at a given point in the functional domain (i.e., the sum of the weighted basis functions at a given point) is determined by several coefficients. If only some of such coefficients are zero, the functional coefficient is not zero. Hence, typically, smooth transitions shrunken to zero can be observed in the functional domain, depending on the number and location of the knots (i.e., the fewer knots there are, the smoother the estimated function is). This further encourages the use of an adaptive LASSO penalty, which leads to asymptotically unbiased estimates (see Zou 2006).

3.1 Cross-validation

The penalty parameter λ is determined by minimising a prediction error metric, say PE, obtained from a random K -fold cross-validation (CV) study. For this reason, let $\mathcal{D} = \{y_{s,t}(h), s \in \mathcal{S}, t = 1, \dots, T\}$ be the set of all available functional observations, and let $\mathcal{D}_1, \dots, \mathcal{D}_K$ be a random partition of \mathcal{D} , which is made by randomly assigning N/k observation to each group \mathcal{D}_i with $i = 1, \dots, K$.

For each subset \mathcal{D}_i , the penalised estimation is performed for a certain predefined sequence of λ , including $\lambda = 0$. In particular, the parameters are estimated according to (10) using data in $\mathcal{D} - \mathcal{D}_i = \{s \in \mathcal{D} : s \notin \mathcal{D}_i\}$. Then, the data in \mathcal{D}_i , are used to evaluate the out-of-sample prediction $\hat{y}_{s,t}(h, \lambda)$ for a fixed λ , and the corresponding prediction error metric $PE(\lambda)_i$. Eventually, for each λ , the overall performance measures $PE(\lambda)$ is obtained by averaging $PE(\lambda)_i$ for $i = 1, 2, \dots, K$. The optimal λ^* may be chosen either by minimising PE, namely

$$\lambda_{min,PE}^* = \operatorname{argmin}_{\lambda}(PE(\lambda))$$

In order to consider the random nature of sets \mathcal{D}_i , a popular solution is to adopt the so-called one-standard-error rule, namely to chose λ defined by

$$\lambda_{1\sigma,PE}^* = \inf_{\lambda} \{PE(\lambda) \geq PE(\lambda^*) + \sigma(PE)\}$$

where $\sigma(PE)$ is the standard error of the average of the PE_i . Of course any λ^* between the above two values would have the same statistical validity. The choice of the metric PE

may be based on the application at hand. Popular choices are the root mean square error and the mean absolute error which are discussed in the following Sect. 3.2.

For a fixed choice of PE and criterion for λ^* , the procedure is synthesised using the pseudo-code in Algorithm 1.

described in Sect. 4 is built under the Gaussianity assumption of Sect. 2. Instead, the air quality exercise of Sect. 5 does not fulfill such assumption. As discussed in Chai and Draxler (2014), the RMSE is appropriate when the error is Gaussian distributed. Whereas, when the error distribution is affected by skewness, outliers or is leptokurtic (Karunasingha 2022), MAE is preferred being

Algorithm 1 Adaptive LASSO estimation for functional hidden dynamic geostatistical models

- 1: **Phase 1:** EM algorithm for full model
 - 2: Estimate the f-HDGM model parameters $\hat{\beta}_{MLE}$ and $\hat{\psi}_{MLE}$ using the EM algorithm
 - 3: **Phase 2:** Hessian of full model
 - 4: Compute the Hessian matrix $H_{MLE} = H(\hat{\beta}_{MLE}, \hat{\psi}_{MLE})$
 - 5: **Phase 3:** Penalised estimation of functional fixed effects
 - 6: **for** each partition \mathcal{D}_i with $i = 1, \dots, K$ **do**
 - 7: **for** $\lambda = 0, \dots, \lambda_{max}$ **do**
 - 8: Compute $\hat{\beta}_{PMLE}(\lambda; \hat{\psi}_{MLE})$ using the BFGS quasi-Newton method in Equation (10) and data in $\mathcal{D} - \mathcal{D}_i$
 - 9: Predict out-of-sample values $\hat{y}_{s,t}(h, \lambda)$ using $\hat{\beta}_{PMLE}(\lambda; \hat{\psi}_{MLE})$
 - 10: Compute $PE_i(\lambda)$ for data in \mathcal{D}_i
 - 11: **end for**
 - 12: **end for**
 - 13: Compute overall $PE(\lambda)$ and find the optimal λ^*
-

3.2 Which prediction error metric?

In the simulations and application of the next sections, we will use the root-mean-square error (RMSE) and the mean absolute error (MAE) as prediction error metrics PE. For each random partition \mathcal{D}_i , they are defined as follows:

$$RMSE(\lambda)_i = \frac{1}{24\|\mathcal{D}_i\|} \sum_{y_{s,t} \in \mathcal{D}_i} \sum_{h \in H} (y_{s,t}(h) - \hat{y}_{s,t}(h, \lambda))^2 \quad (11)$$

$$MAE(\lambda)_i = \frac{1}{24\|\mathcal{D}_i\|} \sum_{y_{s,t} \in \mathcal{D}_i} \sum_{h \in H} |y_{s,t}(h) - \hat{y}_{s,t}(h, \lambda)| \quad (12)$$

where $\hat{y}_{s,t}(h, \lambda)$ is the prediction provided by the estimated model for a fixed λ , and $\|\mathcal{D} - \mathcal{D}_i\|$ denotes the cardinality of a the set \mathcal{D}_i . Although other alternatives are available (Chicco et al. 2021), RMSE and MAE are usual metrics used in the literature (Sammut and Webb 2010b, a).

In this paper, we employ both RMSE and MAE without stating a preference rule between the two as there is no universal consensus on the most appropriate metric for model errors (Cort and Kenji 2005). The simulation study

more robust (Willmott et al. 2009; Hodson 2022). Therefore, we develop the analysis using both metrics as suggested by Chai et al. (2009).

Using $\lambda_{min,PE}^*$ provides the better predictive capabilities for the data at hand. Instead, the one-standard-error rule leads to a parsimonious set of parameters and the selected model is simpler and more interpretable than previous one (Hastie et al. 2017, 2015). Therefore, in both simulations and the application we will compare the results coming from the four CV criteria: $\lambda_{min,RMSE}^*$, $\lambda_{min,MAE}^*$, $\lambda_{1\sigma,RMSE}^*$, and $\lambda_{1\sigma,MAE}^*$.

4 Monte carlo simulation study

In this section, we present and discuss an extensive Monte Carlo simulation study aimed at evaluating the performance of the algorithm in various contexts with increasing levels of complexity and representative of common application contexts.

Table 1 Specification of the simulation settings

Description	Setting I	Setting II	Setting III
	Uncorrelated response Uncorrelated predictors	Spatiotemporal correlation Uncorrelated predictors	Spatiotemporal correlation Correlated predictors
Spatial locations s	15	15	15
Days t	365	365	365
Functional domain	[0, 24]	[0, 24]	[0, 24]
Total observations	$15 \cdot 365 \cdot 24$	$15 \cdot 365 \cdot 24$	$15 \cdot 365 \cdot 24$
Covariates	$\mathbf{X} \sim N_3(\mathbf{0}, \Sigma_X)$	$\mathbf{X} \sim N_3(\mathbf{0}, \Sigma_X)$	$\mathbf{X} \sim N_3(\mathbf{0}, \Sigma_X)$
Var-cov matrix	$\Sigma_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\Sigma_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\Sigma_X = \begin{bmatrix} 1 & 0.9 & 0.7 \\ 0.9 & 1 & 0.5 \\ 0.7 & 0.5 & 1 \end{bmatrix}$
Spline type	B-spline	B-spline	B-spline
Interior knots	5	5	5
Spline order	3 (cubic)	3 (cubic)	3 (cubic)
Number of bases	7	7	7
β coefficients	[1 1 1 1 0 0 0]	[1 1 1 1 0 0 0]	[1 1 1 1 0 0 0]
θ	0 km	50 km	50 km
\mathbf{G}	diag(0,0,0)	diag(0.85,0.85,0.85)	diag(0.85,0.85,0.85)
Σ_η	diag(0,0,0)	diag(1,1,1)	diag(1,1,1)
Σ_ϵ	diag(1,1,1)	diag(1,1,1)	diag(1,1,1)

4.1 Experimental design

To evaluate the performance of the model selection algorithm, we perform a Monte Carlo simulation study based on three settings, labelled as *Setting I*, *Setting II*, and *Setting III*. The three schemes are summarised in Table 1. The settings represent the following situations of interest in geostatistical models. First, we regard a multiple regression model as a benchmark approach (i.e., all temporal and spatial dependence parameters are chosen such that the resulting process is independent in space and time). Second, we consider the case of a response variable that is correlated across space and time but with uncorrelated regressors. Third, in *Setting III*, we introduce cross-correlation among the regressors. In particular, we consider cross-correlation ranging from moderate (0.5) to strong (0.9). The latter setting represents the most challenging for model selection, but also the most realistic one.

The spatial dependence is exponentially decreasing with spatial range $\theta = 50$ km, implying a correlation is below 0.37 beyond a 50 km distance. The temporal autoregressive coefficients in the G matrix are all equal to 0.85, resulting in a pronounced temporal persistence, which is common in meteorological-related applications.

To represent a realistic spatial setting, we take the coordinates from the data that is used in the following empirical sections. More precisely, the coordinates of the spatial locations refer to the atmospheric monitoring sites belonging to the 84-stations network of *ARPA Lombardia*

(see the paper by Maranzano 2022). Regarding the temporal resolution, we consider that the data are observed over 365 days, with each day representing the functional domain.

For each of the three settings, 500 Monte Carlo replications are simulated using a random subset of 15 locations extracted from the full list of ARPA Lombardia monitoring network.

For the functional interpolation, we use a simple set-up of cubic B-spline basis functions (Ramsay 2005) with 7 knots, corresponding to the β coefficients given by $\beta = (1111000)$. This allows us to analyse the performance of the algorithm in selecting relevant parts across the functional domain, we considered functional regression coefficients, that is 1 at the start of a day and going smoothly to 0, as shown in Fig. 1. Note that B-splines used do not reproduce daily periodicity. This will be addressed using Fourier bases in the application.

The simulated values of the response variable are given by the sum of the random effect, the measurement error, the linear combination of the three covariates, and the functional intercept. Tibshirani (1996) suggest to standardise all the covariates and centering the dependent variable before applying the penalised regression. In our simulation design, all the covariates are simulated by a Gaussian distribution with zero mean and unit variance, thus no standardisation is needed. Since we are interested in the daily profile, we keep the response variable in its original scale and the penalty is only applied to the three

covariates, while the functional intercept is left unpenalised. In this way, even for high penalty values that set the spline coefficients to zero, the intercept can still capture the average intraday pattern.

The penalty term sequence was generated according to an exponentially decaying grid, starting from $\lambda_{min} = 10^{-5}$ up to $\lambda_{max} = 0.5$. As we will show in the simulation results, for a value of λ greater than 0.5, all the coefficients shrunk to 0. We added as the first value of the sequence $\lambda = 0$, corresponding to the unpenalised maximum likelihood solution. In total, we consider 101 different values. To identify the optimal value of λ , we perform a 10-fold random cross-validation across space and time.

Being a simulation experiment, we are able to determine the range of possible values in which the prediction error metrics used (i.e., RMSE and MAE) may occur. Based on the simulation setup described in Table 1, we simulate reference values for the minimum and maximum of each prediction error metric in each setting. Reference values are reported in Table 2. The reported reference values correspond to the average value of the minima and maxima of the three prediction error metrics calculated on $n = 100$ simulations of each of the three settings. As may be noticed, both the spatiotemporal dependence and the cross-correlation among the covariates increase both the minimum and the maximum. In addition to the mean value, under parenthesis we report the simulation standard errors.

Table 2 Simulated ($n = 100$ simulations) reference values for each setting. Reported values are the average across simulations of the minimum (*min*) and maximum (*max*) MSE, RMSE, and MAE. Values under parenthesis are the standard error of the mean computed across simulations

		MSE	RMSE	MAE
Setting I	min	0.9999 (0.0004)	1.0000 (0.0002)	0.7979 (0.0002)
	max	2.7257 (0.0013)	1.6510 (0.0004)	1.2724 (0.0003)
Setting II	min	0.9989 (0.0004)	0.9995 (0.0002)	0.7975 (0.0002)
	max	2.7238 (0.0013)	1.6504 (0.0004)	1.2719 (0.0003)
Setting III	min	0.9996 (0.0004)	0.9998 (0.0002)	0.7977 (0.0002)
	max	5.1389 (0.0022)	2.2669 (0.0005)	1.6956 (0.0003)

4.2 Simulation results

The common features of the three settings are numerous. First, we observe that both the presence of spatiotemporal dependence (Settings II and III) and cross-correlation between covariates (Setting III) increase the values of RMSE and MAE. This is reflected in the RMSE and MAE values estimated by penalisation (Fig. 2), whose minimum and maximum values are perfectly overlapping with the reference values (Table 2) for Setting I. However, not surprisingly, the estimated values of the metrics in Setting II and Setting III are slightly higher than the reference values. In fact, it is known that when spatiotemporal dependence occurs, part of the information is duplicated or redundant, thus the information embedded in the sample is lower compared to the case of independent data (Lee and Lund 2008; Griffith 2005). This results in an increase of the prediction errors variability (i.e., the non-predictable portion of the data).¹ Also, as expected, for all three settings, the penalised optimal estimates ($\lambda_{min,RMSE}^*$ and $\lambda_{min,MAE}^*$) are always different from the maximum likelihood estimates and are mutually consistent (i.e., in line with the Gaussian framework, the λ value minimising the two metrics is on average the same). However, from the purely predictive capability standpoint, penalised and unpenalised are equivalent (i.e., the confidence intervals for $\lambda = 0$, $\lambda_{min,RMSE}^*$ and $\lambda_{min,MAE}^*$ overlap). Last but not least, the one-standard-error metrics lead to coefficient estimates that are much closer to the true values compared to the MLE and minimum solutions.

Specifically for Setting III, the major insights are summarised as follows. Figure 2 shows the average RMSE and MAE across the Monte Carlo replications and the optimal λ values obtained using the four above mentioned criteria. The upper panels show MAEs (left) and RMSEs (right) for all the whole range of λ values considered, while the lower panels focus on the behaviour of MAE and RMSE near the optimum solutions. Moreover, we depict the CV variability with the error bars, computed as plus and minus one standard error of the CV prediction error. Both RMSE and MAE plots clearly show that λ minimising RMSE or MAE provides different performances from the MLE solution. Overall, both the RMSE and MAE show smooth patterns.

¹ Given the Markovian dynamics with respect to the time (Eq. 4) and the presence of the Gaussian Process with covariance matrix Γ (Eq. 5), it is easily demonstrated that the variance of the spatiotemporal effect $\omega_{st}(h)$ (Eq. 3) depends: (a) linearly on the marginal variances v_1, \dots, v_p ; (b) non-linearly (hyperbole) on the temporal autocorrelation coefficient g ; (3) non-linearly (exponential) on the range θ . Figure S3 in the Supplementary Information shows how the variability of $n = 10$ simulated random effects and the total variability evolves as the temporal autocorrelation, marginal variance and range change.

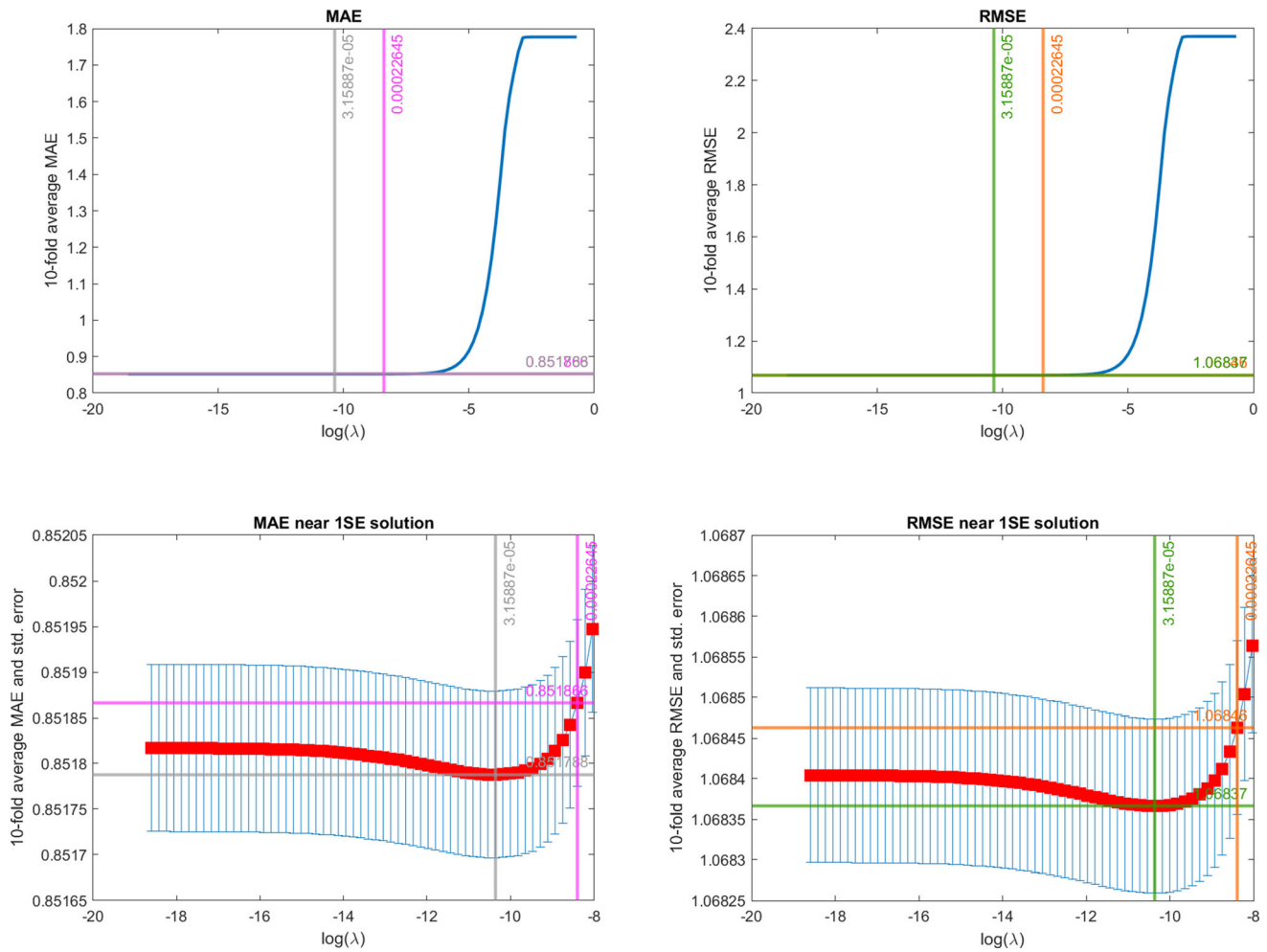


Fig. 2 RMSE and MAE for different values of λ in Setting III. Top panels: full λ range. Bottom panels: near-optimum λ range. Left panels: MAE. Right panels: RMSE. The vertical and horizontal lines

correspond to the considered selection rules (grey: $\lambda_{min,MAE}^*$; pink: $\lambda_{1\sigma,MAE}^*$; black: $\lambda_{min,RMSE}^*$; orange: $\lambda_{1\sigma,RMSE}^*$)

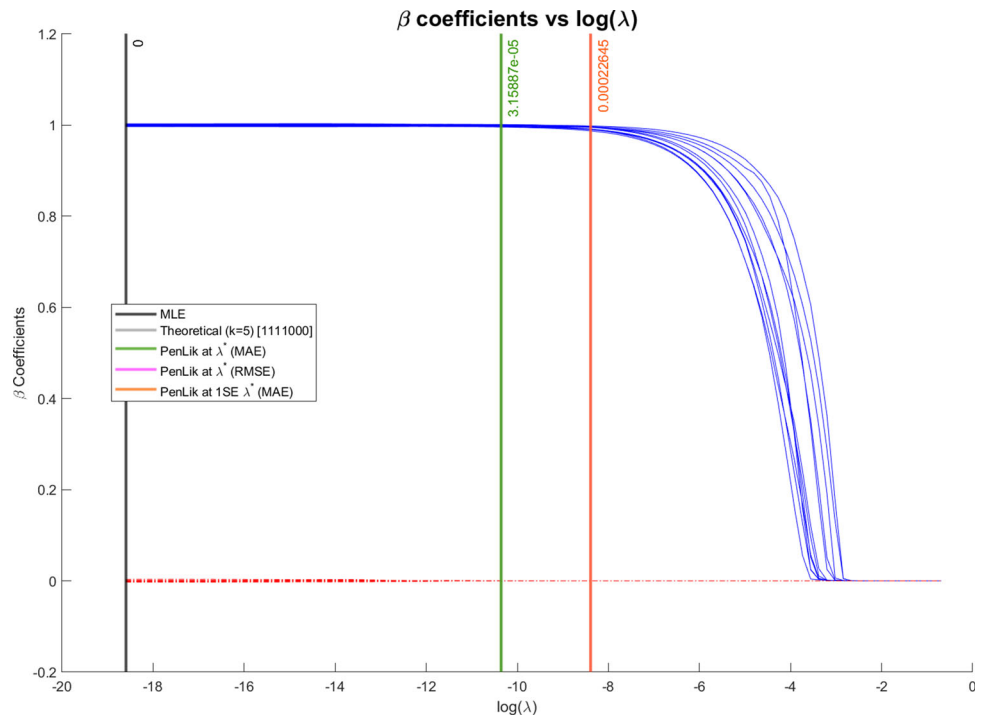
Consistently with the Gaussian simulation framework, the λ minimising the metrics (grey and green vertical lines) coincide, as well as the penalties associated with the one-standard-error rule (pink and orange vertical lines). As previously stated, in terms of predictive capacity, each model with λ lying between the MLE and the one-standard-error solutions are equivalent as the corresponding intervals overlap.

In Fig. 3 we can observe that the coefficients averages are smoothly shrunk towards 0. For values of the penalty term λ greater than 0.03 (i.e., $\log(\lambda) > -3.50$), all the coefficients are shrunk to 0 and the RMSE and MAE stabilise around 2.35 and 1.75, respectively.

In Fig. 4, we plot the empirical distribution (i.e., the box-plot) across the simulations of each fixed effects coefficient for $\lambda = \lambda_{min,RMSE}^*$ (upper panel) and $\lambda = \lambda_{1\sigma,RMSE}^*$ (lower panel).

In both cases, the following is very noticeable: (1) the coefficients are estimated very close to their actual value, which indicates that the penalised estimators are approximately unbiased; (2) the variability for null coefficients is considerably smaller than for coefficients equal to 1; and (3) for increasing values of penalisation (i.e. moving from $\lambda_{min,RMSE}^*$ to $\lambda_{1\sigma,RMSE}^*$ solutions) the coefficients variability decreases, especially for the true zero coefficients. The two last considerations are strengthened by Table S5 in the Supplementary Information. The findings suggest that the one-standard-error estimates (for both MAE and RMSE) are quite different from the minimum and MLE solutions. In fact, it is easy to see that the coefficients associated with bases 5, 6 and 7 of each covariate (all null in the simulative setup) tend to the true value 0 as the penalty increases, while the first four coefficients remain around the true value 1. In the case of the null parameters, comparing the distances between the estimators and the true coefficients (“RMSE” columns), it is noticeable tht the distance

Fig. 3 Average estimated coefficients for different values of λ in Setting III. The positive coefficients are drawn in blue, and the zero coefficients are depicted by the red dashed lines



decreases as one moves from the MLE estimators to the penalized ones (reaching exactly zero with the standard-error estimators). In contrast, high penalty values introduce a bias in the one-unit coefficients, leading to an increase in their variability. Thus, we can state that the one-standard-error rule enjoy an oracle property in the sense that for the null coefficients we detect the true zero values. Also, from an inferential standpoint, the sparsity induced by the one-standard-error estimates leads to models less complex and easier to interpret than the MLE estimates.

5 Application to air quality in Lombardy

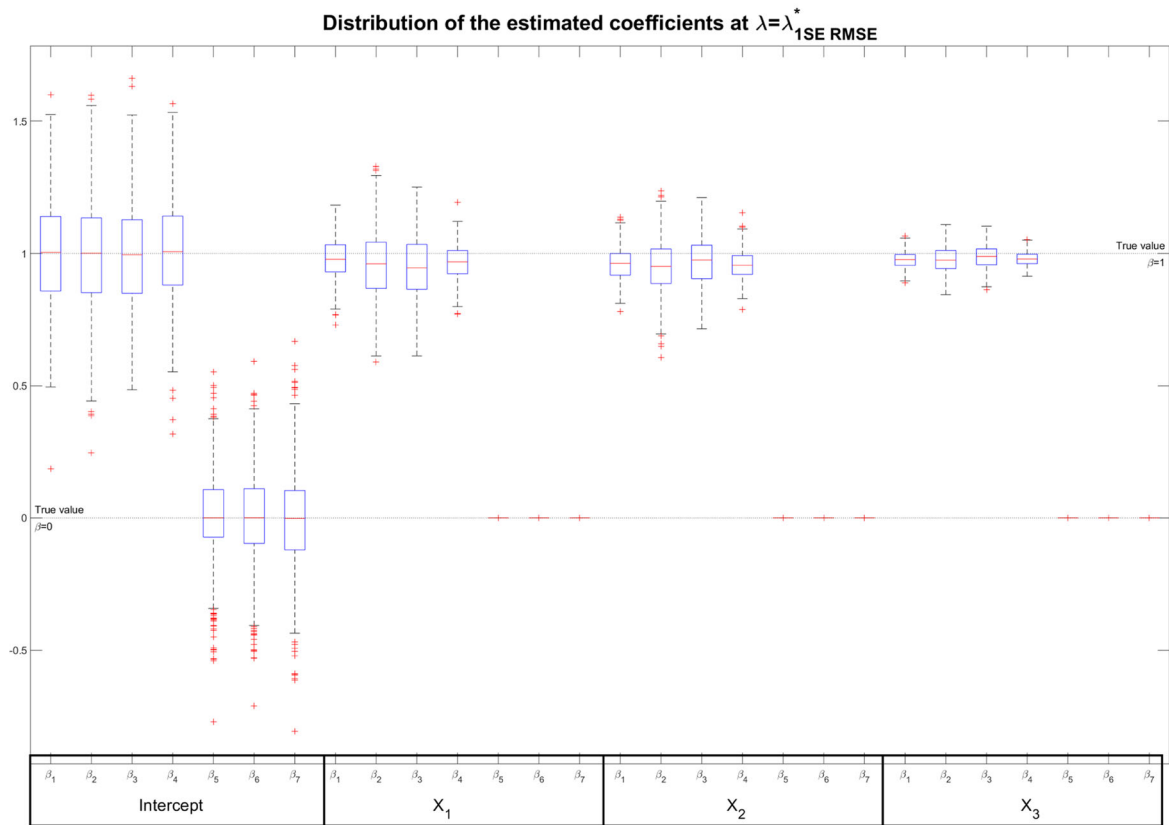
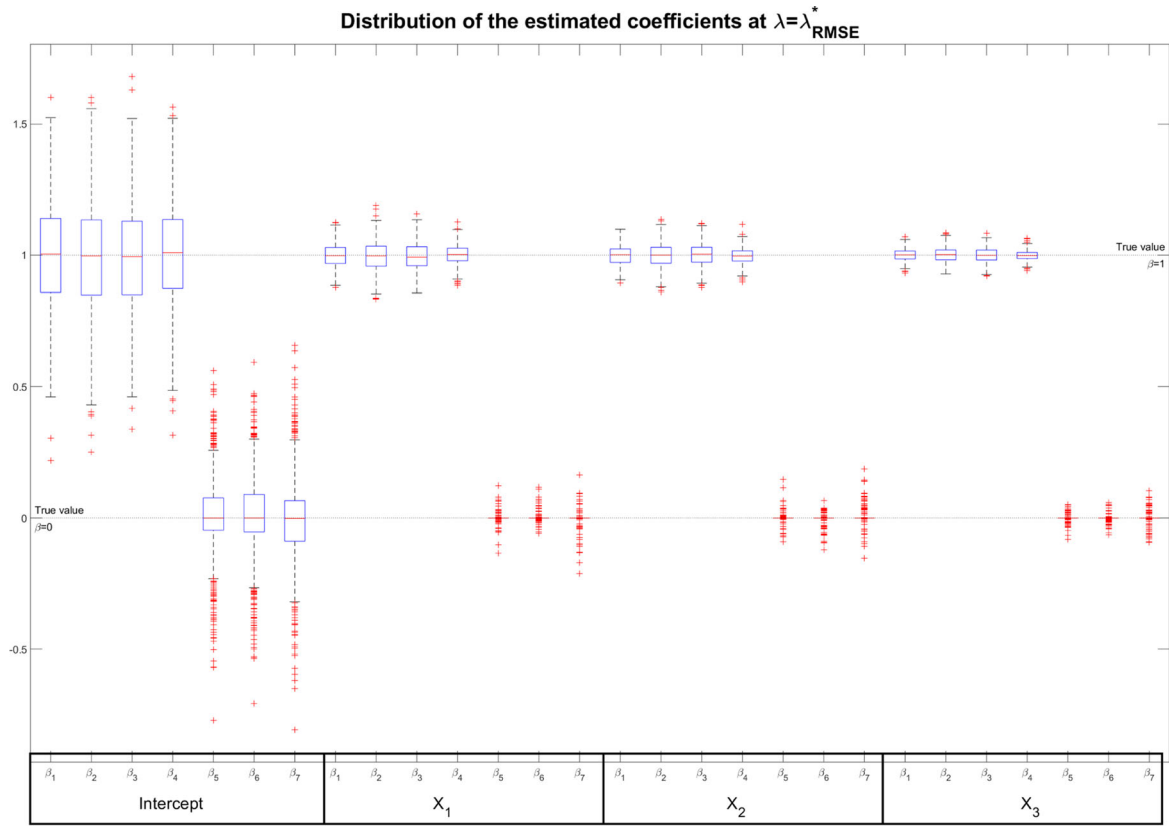
In this section, we present an empirical application of the penalisation algorithm to an air quality assessment case study. The application is structured in such a way to study some scalability properties of the procedure, that is the computation time of the phases composing Algorithm 1, and its behaviour as the model complexity increases.

The application refers to air quality data recorded during the COVID-19 pandemic in Lombardy region, Italy (see the upper panel of Fig. 5). Airborne pollution in Lombardy has attracted considerable research interest for many years. With the COVID-19 emergence, many researchers have become increasingly interested in the short-term effects of lockdowns on air quality (Cameletti 2020; Collivignarelli et al. 2020; Lovarelli et al. 2020; Fassò et al. 2022).

We model hourly nitrogen dioxide (NO_2) concentrations obtained by the $n = 84$ stations depicted in Fig. 5 (bottom

panel) from 1 March, 2020, to 31 May, 2020, that is $T = 92$ days. We consider the concentration throughout the day as a functional observation. For a first insight on the intraday variations, in Fig. 6 we show the regional intradaily evolution of the NO_2 concentrations through a functional box plot computed on the full sample. The 24-hour profile clearly shows the intra-day dynamics of the average (blue curve) and median (black) concentrations, as well as of their variability (the red line represents the functional standard deviation). In particular, there are strong differences between the NO_2 concentrations at night and day. They are in accordance with anthropic activities – that is, very high concentrations are seen during peak hours, between 7am and 9am and between 5pm and 11pm. Also, the standard deviation's curve shows a pattern very close to the average, with two local maxima at peak hours and the minimum value during the afternoon. On average, the standard deviation is bounded between $8.50\mu\text{g}/\text{m}^3$ (3pm) and $16.62\mu\text{g}/\text{m}^3$ (9pm), with mean $12.76\mu\text{g}/\text{m}^3$.

To explain the airborne pollutant concentrations, we consider a set of nine meteorological and land cover variables: temperature ($^{\circ}\text{C}$), precipitation (mm), relative humidity (%), atmospheric pressure (Pa), eastward and northward component of the wind (m/s), geopotential height (m^2/s^2) and high and low vegetation covering (measured as one-half of the total green leaf area per unit horizontal ground surface area, cf. Sabater 2019). Since the variables present different scales and ranges, we standardise both the response variable and the covariates with respect to their overall 24-hour mean and standard



◀ **Fig. 4** Box plot of the estimated coefficients across 500 simulations at $\lambda_{min, RMSE}^*$ (upper panel) and at $\lambda_{1\sigma, RMSE}^*$ (lower panel) for Setting III

deviation. Following Tibshirani (1996), having standardised both the response variable and the covariates, penalisation is applied to predictors only, while the functional intercept is left unpenalised. The total number of observations was $n \times T = 185472$ for each variable.

To account for the natural daily cycle, in Equation (1), we set t as the day, whereas h is the time across the day. Hence, periodic Fourier basis functions with b bases and a support between 0 and 24 were used.² Recall that by construction, Fourier splines require an odd number of bases, and their interpretation depends on the frequency. In fact, except for the first basis, the other basis pairs were calculated at increasing frequency. For example, if the number of bases was $b = 5$, the pair formed by the fourth and fifth bases would have twice the frequency of the second and third pair. The use of Fourier bases ensures the continuity of the last hour of a day to the first hour of the consecutive day. According to the number of basis functions, the total number of parameters to estimate is equal to $b \times 14$. In particular, $b \times 10$ parameters are associated with the covariates and the functional intercept; $b \times 3$ with spatiotemporal dynamics, and b with residual component variances.

The algorithm is initialised by estimating both the fixed and random effects of the f-HDGM using the unpenalised MLE. After estimating the full model, we apply the penalised likelihood model selection algorithm using an exponentially decaying grid of penalty coefficients λ that ranged from $\lambda_{min} = 10^{-4}$ to $\lambda_{max} = 0.50$. We also includes a value of $\lambda = 0$ for the unpenalised estimates.

5.1 Scalability of the algorithm

Using the air quality data introduced above, we study the computing time of Phases 1, 2 and 3 of Algorithm 1 and the algorithm's behaviour for increasing model complexity. To do this, we consider an increasing number of Fourier bases b for each covariate and an increasing number of spatial partitions. We also consider the impact of approximation methods for the fixed-effect coefficients (i.e., $\hat{\beta}_{MLE}$ and the Hessian matrix (i.e., H_{MLE}) of the initial unpenalised f-HDGM introduced in Sect. 2.2.

Moreover, we examine the algorithm's ability to select only the relevant frequencies of the Fourier bases by shrinking irrelevant frequencies towards 0. Thus, we test

² Following Ramsay (2005) (Section 3.4), a set of periodic Fourier basis functions is defined as $B_0(h) = 1$, $B_{2K_a-1} = \sin(hK_a \frac{2\pi}{24})$, and $B_{2K_a} = \cos(hK_a \frac{2\pi}{24})$

the following occurrences as the complexity increases: (1) the increase in the number of fixed zero coefficients, and (2) the higher concentration of zeroes for the coefficients associated with high Fourier frequencies. The latter is consistent with observed NO₂ concentrations (Fig. 6), whose intra-day behaviour is fairly smooth and shows two peaks, which mean that from a modelling perspective, a small number of seasonal frequencies (low complexity) is expected.

A total of 17 models were evaluated. For each model, we considered the computation time for the three main phases of the algorithm – that is, the initial model estimation with the EM algorithm, the computation of the variance-covariance matrix of the parameters, and the penalised likelihood algorithm.

For model complexity we consider three scenarios: $b = 5$, $b = 7$ and $b = 9$. In the first case, the total number of the model's parameters is $b \times 14 = 70$; in the second case, is 98, and in the third case is 126. Therefore, the numbers of fixed effect parameters is 50, 70, and 90, respectively. Moreover, for spatial partitioning we consider groups varying from $k = 1$ (no spatial partitioning) to $k = 5$. Eventually, the threshold for the overall improvement in the variance-covariance matrix computation is fixed to 0.01.

The results are summarised in Figs. 7 and 8, and full details are given in Table S1 of the Supplementary Information.

In particular, Fig. 7 shows the computational cost of the penalised likelihood algorithm as a function of the model complexity, the spatial partitioning and the adoption of an approximation for the variance-covariance matrix. The main results are summarised as follows:

- Variance-covariance matrix approximation (upper left panel of Fig. 7): compared to the exact solution, the approximation reduces the computation time of 66% for Phase 2, and 25% for the overall computation. This holds independently from the model complexity. Of course, the penalisation algorithm is not affected;
- Spatial partitioning (upper right panel of Fig. 7): the application of a spatial partitioning reduces the initial computation time by 30% to 50% and the penalisation phase 3 by up to 38%. Moreover, it reduces the overall time of more than 30%. The variance-covariance matrix computation was not affected. The time gain is negligible when the number of groups increase (i.e. $k \geq 4$);
- Model complexity (left panels of Fig. 7): when the number of basis functions b is reduced from 9 to 7, the computation time of all phases significantly decreases, independent of the approximation of the variance-covariance matrix. In particular, the penalised

Fig. 5 Physical map of Po Valley (upper left panel) and Lombardy (upper right panel) and the ARPA Lombardia air quality monitoring network by type of station (lower left panel) and type of area (lower right panel)

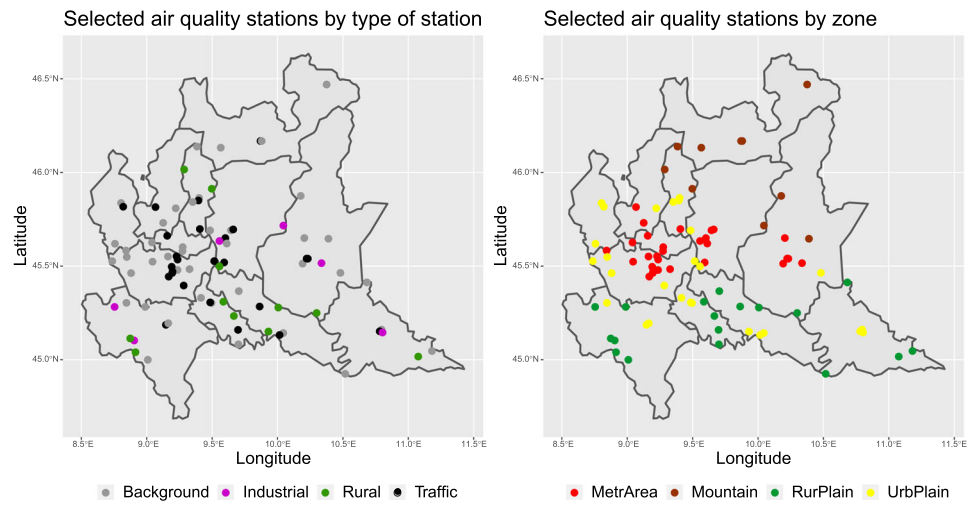
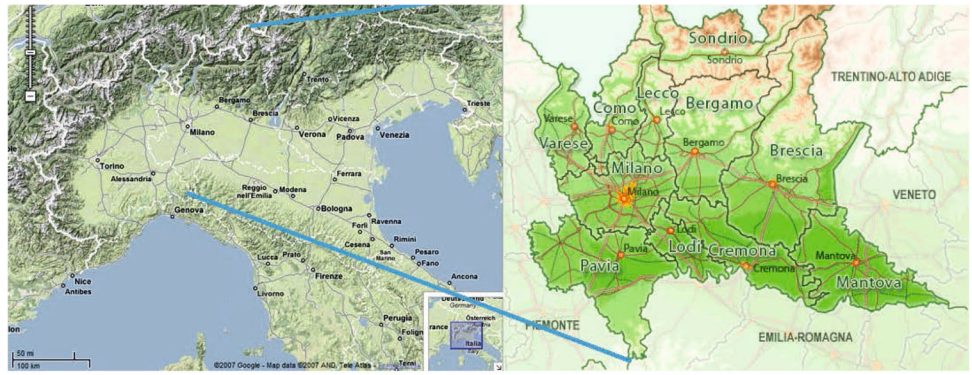
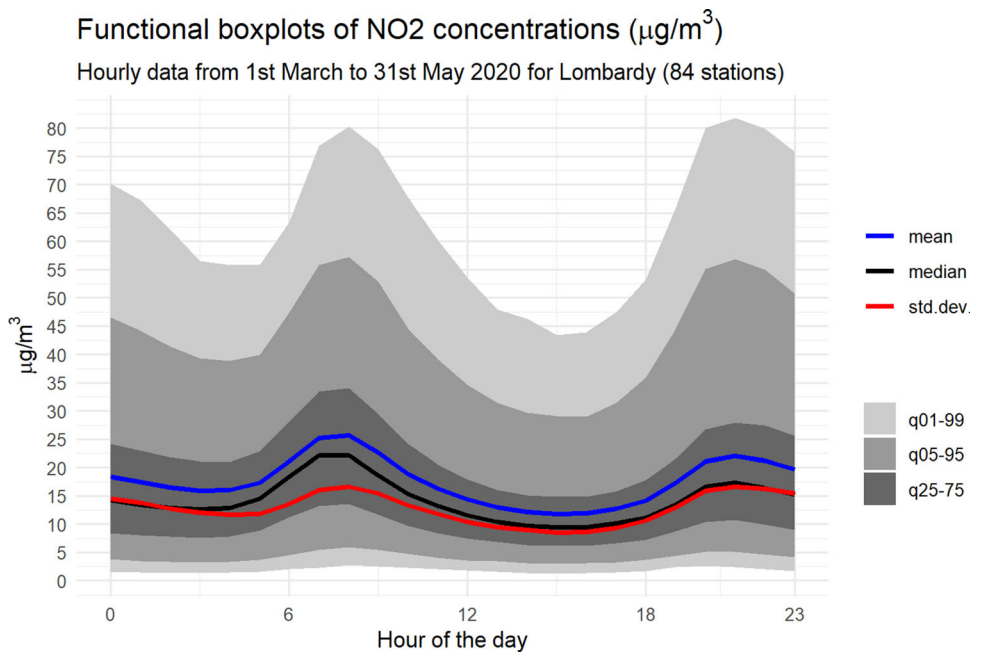


Fig. 6 Intraday box-plot of NO₂ concentrations ($\mu\text{g}/\text{m}^3$) observed between the 1 March 2020 and 31 May 2020



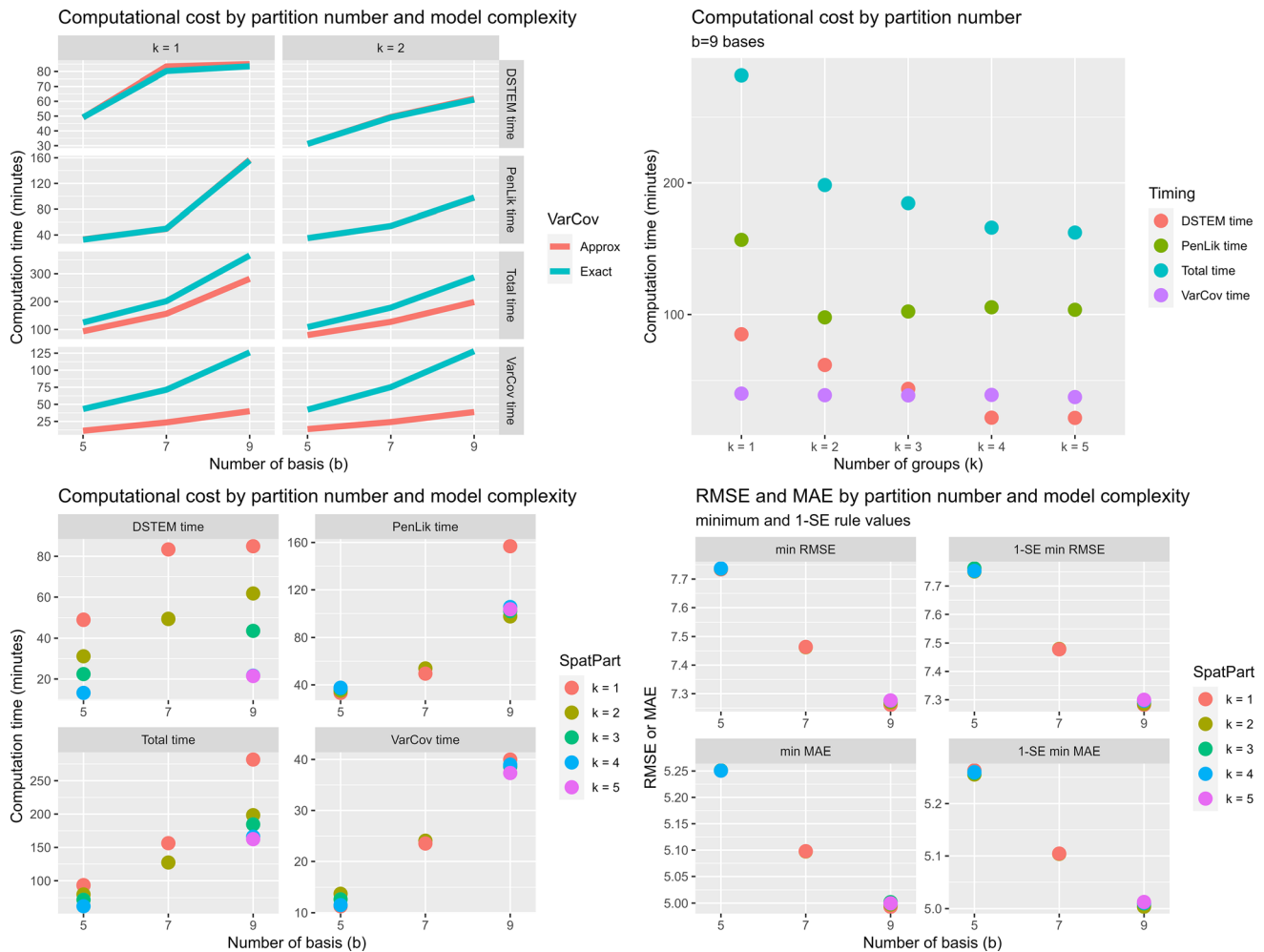


Fig. 7 Computation time and cross-validation errors across the models. Computation time of each phase by model complexity with and without spatial partitioning (upper left panel); computation time of each phase by increasing level of spatial partitioning (upper right

panel); computation time of each phase by increasing level of spatial partitioning and model complexity (lower left panel); RMSE and MAE by model complexity and increasing spatial partitioning (lower right panel)

likelihood estimation decreases by up to 68% and the overall computation by 45%. When b further decreases to 5, the gain is less pronounced;

- Concerning the cross-validated model error, both MAE and RMSE are affected only by the model complexity (lower right panel of Fig. 7). Indeed, independent of the approximation of the covariance matrix or of the spatial partitioning, both MAE and RMSE decreased as the number of basis functions increased for all four criteria used to define the optimum $\lambda_{min,PE}^*$.

In Fig. 8 we present the relationship between model complexity and the share of regressors removed with the adaptive LASSO. The plot clearly shows that when the number of basis functions is large, the overall proportion of zero coefficients increases up to 25% of the total. The graph on the right examines the excluded coefficients in

detail and shows that the highest frequencies (corresponding to $\beta_6, \beta_7, \beta_8$ and β_9) are the most frequently removed by the algorithm. This is consistent with the NO_2 concentrations shown in Fig. 6. Indeed, since the response variable exhibits a very smooth intra-day pattern, the number of frequencies required to model the relationship with the covariates is low.

This result allows us to state that the adaptive LASSO algorithm proposed in this study can be a useful tool for identifying the most relevant frequencies as it is precise in its selection and implementable even in contexts with large data sets. If time computing time is not an issue, using a higher number of frequencies (e.g., $b = 9$ in our case) provides better forecasting performance (i.e., lower RMSE and MAE) while avoiding an excessive number of non-zero coefficients.

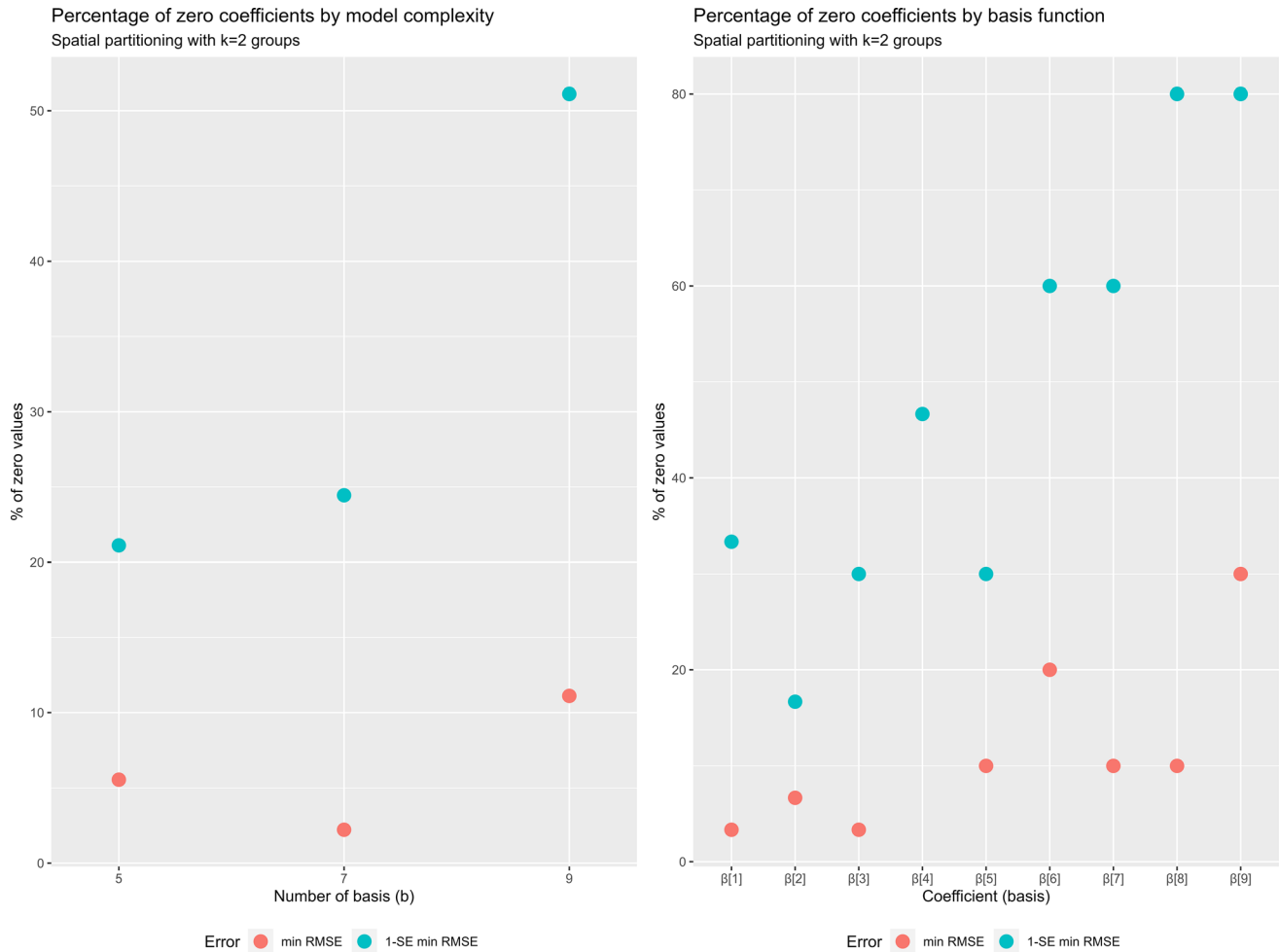


Fig. 8 Percentage of zero coefficients across the models. Percentage of zero coefficients by model complexity when applying a spatial partitioning with $k = 2$ groups (left panel); percentage of zero

coefficients by basis function (coefficients) when applying a spatial partitioning with $k = 2$ groups

5.2 Penalised estimates

The above estimates show very limited variability across the models and exhibit weak sensitivity to the spatial partitioning and Hessian matrix approximations. As shown in the preceding section, the impact of our adaptive LASSO procedure is more pronounced as the number of fixed effects coefficients increases. Thus, we report and comment on the empirical results obtained considering the model with the lowest prediction error among those using $b = 9$ basis functions. Specifically, we consider the case with an approximate variance-covariance matrix and without spatial partitioning.

In Fig. 9, we depict the behavior of the average cross-validation MAE (left panels) and RMSE (right panels) for increasing values of the penalty term λ . The computed RMSE falls between $7.26 \mu\text{g}/\text{m}^3$ (for small values of λ) and $7.75 \mu\text{g}/\text{m}^3$ (for large values of λ). These values are in the same scale as the observed standard deviation (shown in

Fig. 6), but still below the minimum variability observed in the data (about $8.50 \mu\text{g}/\text{m}^3$). By computing the ratio of the minimum MSE to the average intra-day variance we obtain the average proportion of variance unexplained by the model, equivalent to 32.34%. This percentage is consistent with the plot of the variance of measurement errors (Figure S1 in the Supplementary Information), whose average value is 28.60%. For large values of the penalty term (i.e., $\log(\lambda) > -4$) all the covariates are drop out. However, both one-standard-error-rules provide more parsimonious models with prediction errors not significantly different from the optimal ones. The estimated coefficients that correspond to the unrestricted model and the models associated with the four penalty terms considered are shown in Fig. 10.

In Fig. 11, we show the 24-hour estimated functional coefficients for each variable. The black lines correspond to the unpenalised MLE solution; the green lines to the optimal λ w.r.t RMSE; the grey lines, to the optimal λ w.r.t

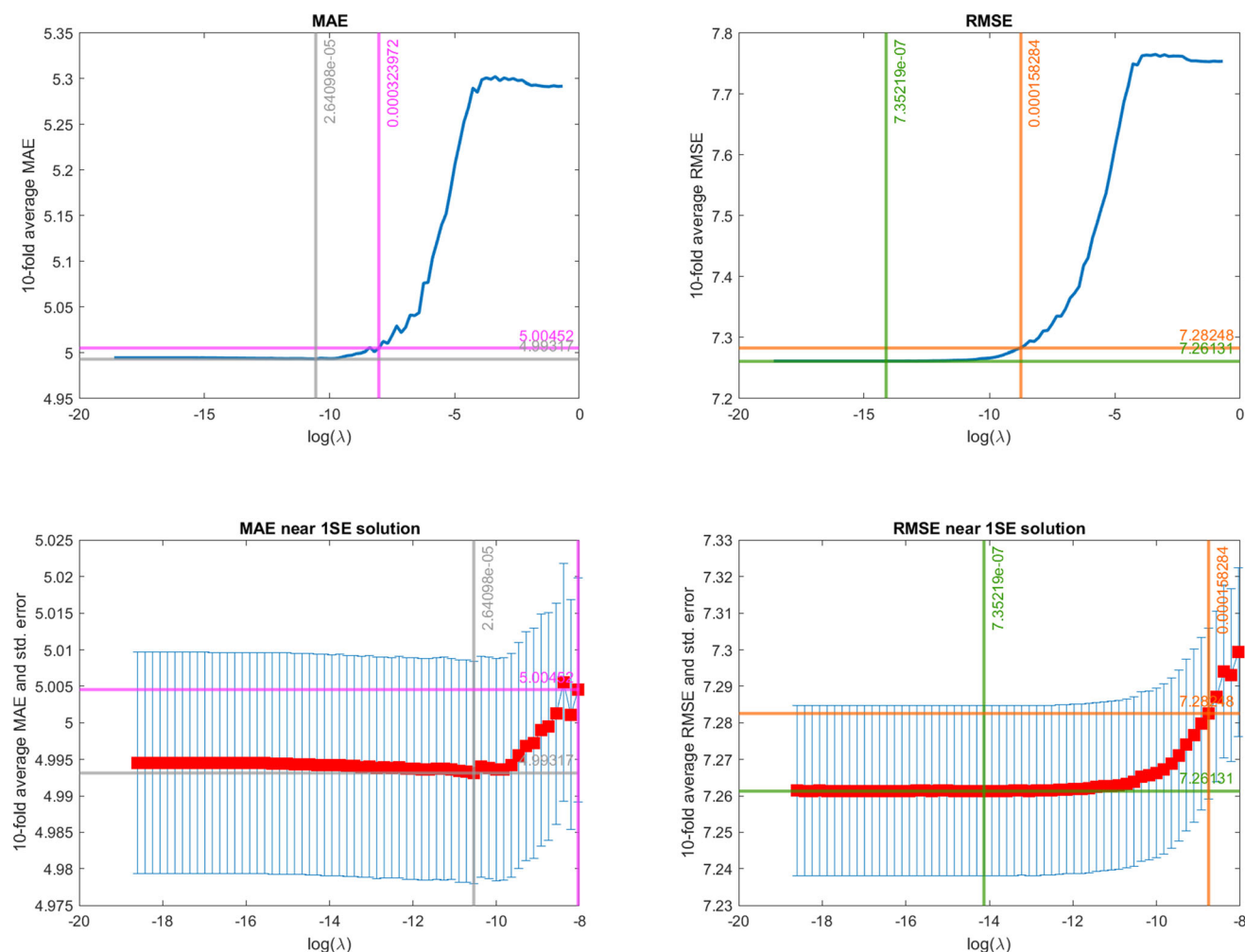


Fig. 9 MAE against the logarithm of the penalty term λ (left panels) and RMSE against the logarithm of the penalty term λ (right panels). The horizontal lines represent the values of MAE and RMSE for key values of $\log(\lambda)$, the optimal ($\lambda_{min,RMSE}^*$ and $\lambda_{min,MAE}^*$), 1-SE optimal

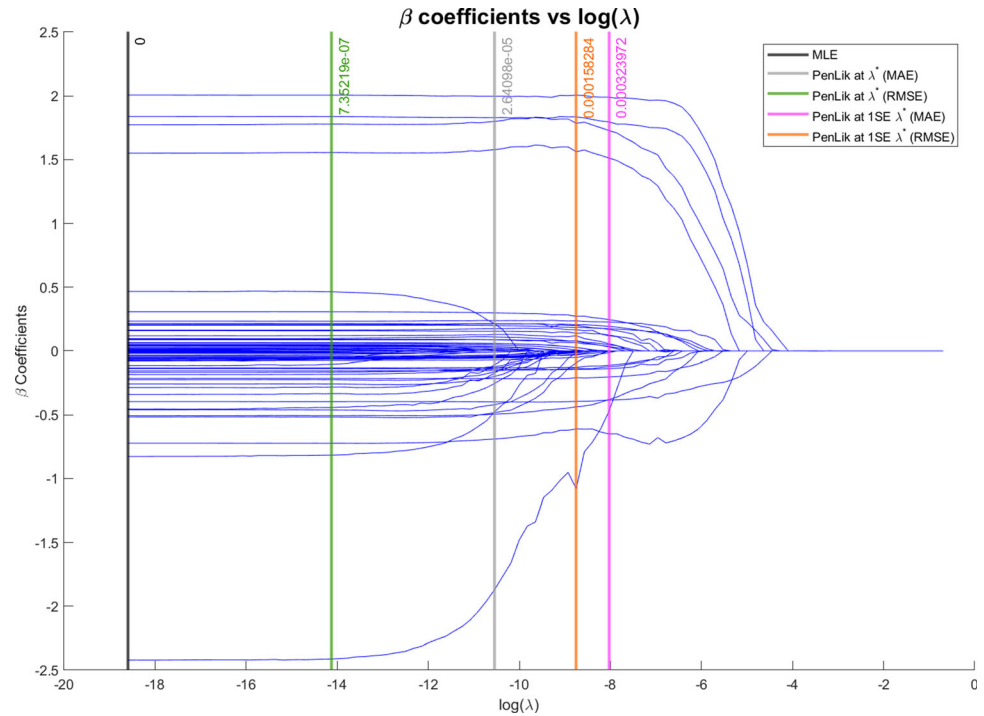
values ($\lambda_{1SE,RMSE}^*$ and $\lambda_{1SE,MAE}^*$). The bottom panels are details near the optimum

MAE; the orange lines, to the 1SE optimal λ w.r.t RMSE; and the pink lines, to the 1SE optimal λ w.r.t MAE. The estimated coefficients associated with temperature always exhibits negative values, particularly in the late afternoon and evening hours. The patterns obtained for different values of λ did not show large discrepancies and tended to overlap throughout the day, leaving the overall dynamics unchanged during the day. However, the penalty seemed to mitigate the temperature effect at peak hours (10 a.m. and 8 p.m.). Rainfall shows a negative effect on the NO_2 concentrations, especially in the evening and just before dawn. In both moments, the effect reached the minimum peaks. Unlike the temperature, whose daily pattern varied slightly as the penalty increased, for higher values of λ , the precipitation diminished its effect and tended to flatten slightly towards 0. Considering the one-standard-error rule, between 7 a.m. and 5 p.m., the curve flattened to a constant

negative value without being exactly 0. For the same λ values, the two negative peak periods are greatly mitigated. These elements confirm the important role of temperature and rainfall in mitigating NO_2 concentrations, which is highlighted in literature (e.g., Fassò et al. 2022). Relative humidity presents some very interesting findings. First, its effect is null at around midnight, slightly negative at night before dawn and strongly positive in the daylight hours. Moreover, the penalisation appeared to produce no effect on the intra-day behaviour. This is consistent with the fact that whereas temperature and rainfall showed more complex patterns during the day, relative humidity already exhibits a simple pattern and did not need further smoothing.

Both the effects of atmospheric pressure and geopotential height (used as a proxy of elevation) depend on the moment of the day that was being considered. In both

Fig. 10 Functional coefficients against the logarithm of the penalty term λ . Vertical lines represent key values of $\log(\lambda)$, i.e. unpenalized ($\lambda = 0$), optimal (λ_{RMSE}^* and λ_{MAE}^*), 1-SE rule ($\lambda_{1\sigma, RMSE}^*$ and $\lambda_{1\sigma, MAE}^*$)



cases, the estimates show a positive effect at the start and at the end of the day and a negative effect in the afternoon. However, in the case of elevation, the functional coefficient in the early and late hours is very close to zero, and in the central hours, it deviates significantly from 0 regardless of the penalty used. For both variables, penalisation does not play a significant role, as the difference between penalised and non-penalised curves approaches 0 and the infra-daily dynamics appears to be stable.

Also, the U (eastward) and V (northward) components of wind show a time-varying behaviour across the day. In both cases, the effect on the NO_2 concentrations is positively estimated during the early stage of the day, especially between 5 a.m. and 10 a.m., and it strongly weakened in the afternoon and at night, reaching values very close to 0 between 3 p.m. and 8 p.m. However, the cleaning effect is limited to the early part of the day. The shrinkage effect induced by the penalisation algorithm is more pronounced in the eastward component than in the northward component. In fact, we noticed that the effect of the eastward component was strongly smoothed in the morning hours, and the coefficient was cancelled during the afternoon. The northward component, although also smoothed, showed a significantly positive effect in the early hours of the day.

Finally, we notice that penalisation generates a remarkable influence on the two land cover variables, that is the high vegetation and low vegetation indices. Both variables are heavily squeezed towards 0 even for the contained values of λ until reaching zero for the values

associated with the one-standard-error rule. Similar results are presented in Fassò et al. (2022), in which the effect of the same covariates on the NO_2 , PM_{10} and $\text{PM}_{2.5}$ concentrations in Lombardy is estimated to be close to 0, and thus not statistically significant, with the exception of the most urbanised areas.

To sum up, as expected, from Fig. 9 we see that from the predictive ability standpoint, penalised and non-penalised estimators are equivalent. However, Fig. 11 suggests that for some covariates the non-penalised intraday functional dynamics may be different from the penalised ones. In fact, for most of the covariates the smoothing increases moving from the MLE to the one-standard-error solution. In this regard, the functional coefficient of the rainfall shrinks close to zero in the middle of the day, whereas the vegetation variables lose much of their importance shrinking close to zero for the large part of the day. So, compared to the MLE and λ_{min}^* solutions, the one-standard-error estimate leads to a simpler and more interpretable model. This holds both for the RMSE and the MAE.

6 Conclusions and future developments

In this paper, we introduced an adaptive LASSO estimator for the so called functional hidden dynamic geostatistical models (f-HDGM). This new estimation approach based on penalised maximum-likelihood estimation can be used to efficiently estimate the relevant model coefficients and

β coefficients over the 24 hours

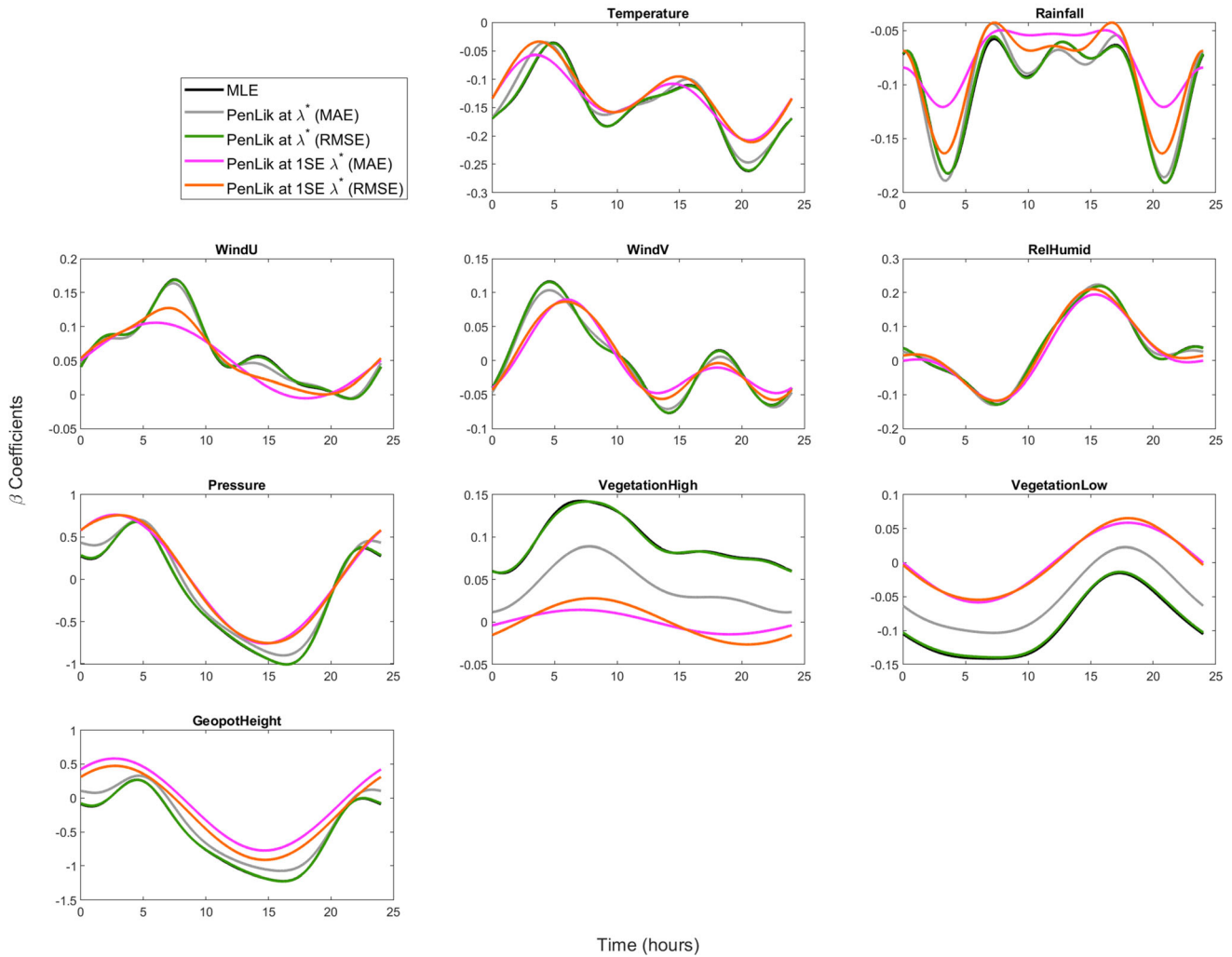


Fig. 11 Estimated functional β coefficients for differently selected optimal penalty parameters

shrink to zero the irrelevant ones, while taking into account spatiotemporal correlation and cross-correlation among predictors. In addition to the case in which all the coefficients of the splines basis associated with a certain covariate are set to zero, it is possible that only some of them are shrunk to zero. Indeed, the algorithm can be successfully applied to identify the relevant part of a functional coefficient across its functional domain.

From a computational perspective, we showed that the estimation can be efficiently implemented as a local quadratic approximation around the maximum of the log-likelihood function. To find this maximum, the EM algorithm can be used (see Wang et al. 2021). Then, a BFGS quasi-Newton iterative method can be used to optimise the penalised function.

We analysed the performance of this estimation procedure through a Monte Carlo simulation study based on three settings with increasing level of complexity and

representative of common applied contexts. To be precise, we considered settings where only parts of the functional coefficients had zero effects and where the regressors were cross-correlated and driven by spatiotemporal dynamics, as is often observed in geostatistical applications. The results of the simulations show that from the pure prediction ability perspective, penalised estimates are equivalent to maximum likelihood estimates. However, from an inferential standpoint, the one-standard-error estimates lead to models much more parsimonious and easier to interpret than the MLE estimates. As expected, having simulated in a Gaussian context, MAE and RMSE give almost equivalent results. Finally, the estimates produced by the one-standard-error rule enjoy an oracle property in the sense that for the null coefficients we detect the true zero value.

Furthermore, we applied have the penalisation algorithm to an empirical example of air quality assessment using hourly NO_2 concentrations observed in Lombardy, Italy.

Like the simulations, the empirical results reveal that from the pure prediction capability perspective, the penalised estimates are equivalent to the maximum likelihood estimates. While the solutions of the one-standard-error rule for RMSE and MAE are very similar, the minimum solution of the MAE suggests a higher degree of penalisation for the coefficients. In terms of inference, the one-standard-error estimates of the coefficients associated with different weather variables are very smoothed and in some cases almost zero. We also provided an extended study of the scalability of the algorithm when applied to real world data. In particular, we showed that even with high model complexity, the computation time (both of the penalised likelihood and overall) can greatly benefit from approximations in model estimation, leaving performance essentially unaffected.

This paper focused on model selection in functional data contexts, performed using an adaptive LASSO penalisation algorithm. However, further extensions can be pursued. Our proposal may be useful in several environmental policy assessment contexts, such as agricultural policies (Fassò, A., Rodeschini, J., Moro, A.F., Shaboviq, Q., Maranzano, P., Cameletti, M., Otto, P. 2023), air quality assessment (Fassò et al. 2022), and energy policies (Yuan et al. 2018). Indeed, the smoothness of the estimated functional coefficients can also be of high interest in such applications because too large a number of spline bases leads to over-fitting and non-smooth estimated effects. From the methodological perspective, a further penalty term based on the integrated second derivatives could counter these effects. Thus, an elastic net structure that includes the smoothness penalty and an adaptive LASSO penalty is a very interesting topic for future research. Eventually, using the results of Simon and Tibshirani (2012), the standardised group-LASSO estimator could be extended to spatiotemporal functional models by optimising a penalized likelihood function with quadratic approximation and by assuming that the spline basis functions associated with each covariate are a group.

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement. The authors declare that this research was funded by Fondazione Cariplo under the grant 2020- 4066 “AgrImOnIA: the impact of agriculture on air quality and the COVID-19 pandemic” from the “Data Science for science and society” program.

Data and codes availability All the results presented in this paper can be reproduced using MATLAB and R software. Data and codes for both the simulation and application results are available at the following GitHub web page: https://github.com/PaoloMaranzano/PM_PO_AF_AdaptiveLASSO_SERRA2023.git

Declarations

Conflict of interest The authors declare that they do not have any competing interest.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00477-023-02466-5>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al-Sulami D, Jiang Z, Lu Z, Zhu J (2019) On a semi-parametric data-driven nonlinear model with penalized spatio-temporal lag interactions. *J Time Ser Anal* 40(3):327–342. <https://doi.org/10.1111/jtsa.12442>
- Aristizabal J-P, Giraldo R, Mateu J (2019) Analysis of variance for spatially correlated functional data: application to brain data. *Spatial Statistics* 32:100381
- Banerjee S, Gelfand AE, Finley AO, Sang H (2008) Gaussian predictive process models for large spatial data sets. *J Royal Statistical Soc: Series B (Statistical Methodology)* 70(4):825–848
- Bondell HD, Krishna A, Ghosh SK (2010) Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*. <https://doi.org/10.1111/j.1541-0420.2010.01391.x>
- Breiman L (1996) Heuristics of instability and stabilization in model selection. *Ann Statistics* 24(6):2350–2383. <https://doi.org/10.1214/aos/1032181158>
- Brown D, de Sousa K, van Etten J (2023) ag5tools: an r package for downloading and extracting agrometeorological data from the agera5 database. *SoftwareX* 21:101267 Retrieved from <https://www.sciencedirect.com/science/article/pii/S2352711022001856><https://doi.org/10.1016/j.softx.2022.101267>
- Cai L, Maiti T (2020) Variable selection and estimation for high-dimensional spatial autoregressive models [Journal Article]. *Scand J Stat* 47(2):587–607. <https://doi.org/10.1111/sjos.12452>
- Calulli C, Fassò A, Finazzi F, Pollice A, Turnone A (2015) Maximum likelihood estimation of the multivariate hidden dynamic geostatistical model with application to air quality in Apulia Italy. *Environmetrics* 26(6):406–417
- Cameletti M (2020) The effect of corona virus lockdown on air pollution: Evidence from the city of Brescia in Lombardia region (Italy) [Journal Article]. *Atmospheric Environment* 239:117794 Retrieved from <http://www.sciencedirect.com/science/article/pii/S1352231020305288><https://doi.org/10.1016/j.atmosenv.2020.117794>
- Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K (2016) Bayesian model selection methods in modeling small area colon cancer incidence [Journal Article]. *Annals of Epidemiology* 26(1):43–49 Retrieved from <https://www.elsevier.com/locate/annepi>

- [sciencedirect.com/science/article/pii/S1047279715004524](https://doi.org/10.1016/j.annepidem.2015.10.011)<https://doi.org/10.1016/j.annepidem.2015.10.011>
- Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K (2016) Spatio-temporal bayesian model selection for disease mapping. *Environmetrics* 27(8):466–478. <https://doi.org/10.1002/env.2410>
- Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K (2018) Spatially-dependent bayesian model selection for disease mapping. *Stat Methods Med Res* 27(1):250–268. <https://doi.org/10.1177/0962280215627298>
- Centofanti F, Fontana M, Lepore A, Vantini S (2022) Smooth lasso estimator for the function-on-function linear regression model. *Computational Statistics & Data Analysis* 176:107556 Retrieved from <https://www.sciencedirect.com/science/article/pii/S0167947322001360><https://doi.org/10.1016/j.csda.2022.107556>
- Chai T, Carmichael GR, Tang Y, Sandu A, Heckel A, Richter A, Burrows JP (2009) Regional nox emission inversion through a four-dimensional variational approach using sciamachy tropospheric no2 column observations. *Atmospheric Environment* 43(32):5046–5055 Retrieved from <https://www.sciencedirect.com/science/article/pii/S1352231009005597><https://doi.org/10.1016/j.atmosenv.2009.06.052>
- Chai T, Draxler RR (2014) Root mean square error (rmse) or mean absolute error (mae)? - arguments against avoiding rmse in the literature [Journal Article]. *Geosci Model Dev* 7(3):1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chang Y-M, Hsu N-J, Huang H-C (2010a) Semiparametric estimation and selection for nonstationary spatial covariance functions. *J Comput Graph Stat* 19(1):117–139
- Chang Y-M, Hsu N-J, Huang H-C (2010b) Semiparametric estimation and selection for nonstationary spatial covariance functions. *J Comput Graph Stat* 19(1):117–139. <https://doi.org/10.1198/jcgs.2010.07157>
- Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation [Journal Article]. *PeerJ Computer Sci* 7:e623. <https://doi.org/10.7717/peerj-cs.623>
- Chu T, Zhu J, Wang H (2011) Penalized maximum likelihood estimation and variable selection in geostatistics. *Ann Statistics* 39(5):2607–2625. <https://doi.org/10.1214/11-AOS919>
- Chu T, Zhu J, Wang H (2011b) Penalized maximum likelihood estimation and variable selection in geostatistics. *Ann Stat* 39(5):2607–2625
- Claeskens G, Krivobokova T, Opsomer JD (2009) Asymptotic properties of penalized spline estimators. *Biometrika* 96(3):529–544
- Collivignarelli MC, Abbà A, Bertanza G, Pedrazzani R, Ricciardi P, Carnevale Miino M (2020) Lockdown for COVID-2019 in milan: What are the effects on air quality? [Journal Article]. *Science of The Total Environment* 732:139280 Retrieved from <http://www.sciencedirect.com/science/article/pii/S0048969720327972><https://doi.org/10.1016/j.scitotenv.2020.139280>
- Cort JW, Kenji M (2005) Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance [Journal Article]. *Climate Research* 30(1):79–82. Retrieved from <https://www.int-res.com/abstracts/cr/v30/n1/p79-82>
- Cressie N, Johannesson G (2008) Fixed rank kriging for very large spatial data sets. *J Royal Statistical Soc: Series B (Statistical Methodology)* 70(1):209–226
- Cressie N, Shi T, Kang EL (2010) Fixed rank filtering for spatio-temporal data. *J Comput Graph Stat* 19(3):724–745
- Cruz-Alonso V, Pucher C, Ratcliffe S, Ruiz-Benito P, Astigarraga J, Neumann M, Rodríguez-Sánchez F (2023) The easyclimate r package: Easy access to high-resolution daily climate data for europe [Journal Article]. *Environmental Modelling & Software* 105627. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1364815223000130><https://doi.org/10.1016/j.envsoft.2023.105627>
- Czernacki B, G logowski A, Nowosad J (2020) Climate: An r package to access free in-situ meteorological and hydrological datasets for environmental assessment [Journal Article]. *Sustainability* 12(1). <https://doi.org/10.3390/su12010394>
- Fan J, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc* 96(456):1348–1360
- Fassò A, Finazzi F, Madonna F (2018) Statistical issues in radiosonde observation of atmospheric temperature and humidity profiles. *Statistics & Probab Lett* 136:97–100
- Fassò A, Maranzano P, Otto P (2022) Spatiotemporal variable selection and air quality impact assessment of covid-19 lockdown. *Spatial Statistics* 49:100549. Retrieved from <https://www.sciencedirect.com/science/article/pii/S2211675321000592><https://doi.org/10.1016/j.spasta.2021.100549>
- Fass`o, A., Rodeschini, J., Moro, A.F., Shaboviq, Q., Maranzano, P., Cameletti, M., Otto, P. (2023) Agrimonia: a dataset on livestock, meteorology and air quality in the lombardy region, italy [Journal Article]. *Scientific Data* 10(1):143. <https://doi.org/10.1038/s41597-023-02034-0>
- Ferreira G, Mateu J, Porcu E (2022) Multivariate Kalman filtering for spatio-temporal processes. Retrieved from, *Stochastic Environmental Research and Risk Assessment*. <https://doi.org/10.1007/s00477-022-02266-3>
- Finazzi F, Fassò A (2014) D-STEM: a software for the analysis and mapping of environmental space-time variables. *J Stat Softw* 62(6):1–29
- Franco-Villoria M, Ignaccolo R (2017) Bootstrap based uncertainty bands for prediction in functional kriging. *Spatial Statistics* 21:130–148
- Furrer R, Bachoc F, Du J (2016) Asymptotic properties of multivariate tapering for estimation and prediction. *J Multivar Anal* 149:177–191
- Furrer R, Genton MG, Nychka D (2006) Covariance tapering for interpolation of large spatial datasets. *J Comput Graph Stat* 15(3):502–523
- Giraldo R, Delicado P, Mateu J (2011) Ordinary kriging for functionvalued spatial data. *Environ Ecol Stat* 18(3):411–426
- Gonella R, Bourel M, Bel L (2022) Facing spatial massive data in science and society: variable selection for spatial models. *Spatial Statistics* 1:100627. <https://doi.org/10.1016/j.spasta.2022.100627>
- Griffith DA (2005) Effective geographic sample size in the presence of spatial autocorrelation. *Ann Assoc Am Geograph* 95(4):740–760
- Hastie T, Tibshirani R, Friedman J (2017) *The elements of statistical learning: Data mining, inference, and prediction (springer series in statistics)* [Book]. Springer, New York
- Hastie T, Tibshirani R, Wainwright M (2015) *Statistical learning with sparsity: the lasso and generalizations* [Book]. CRC Press
- Hodson TO (2022) Root-mean-square error (rmse) or mean absolute error (mae): when to use them or not. *Geosci Model Dev* 15(14):5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- Hofierka J, Parajka J, Mitasova H, Mitas L (2002) Multivariate interpolation of precipitation using regularized spline with tension [Journal Article]. *Trans GIS* 6(2):135–150. <https://doi.org/10.1111/1467-9671.00101>

- Hsu N-J, Chang Y-M, Huang H-C (2012a) A group lasso approach for non-stationary spatial-temporal covariance estimation. *Environmetrics* 23(1):12–23
- Hsu N-J, Chang Y-M, Huang H-C (2012b) A group lasso approach for non-stationary spatial-temporal covariance estimation [Journal Article]. *Environmetrics* 23(1):12–23. <https://doi.org/10.1002/env.1130>
- Huang H-C, Martinez F, Mateu J, Montes F (2007) Model comparison and selection for stationary space-time models. *Comput Statistics & Data Anal* 51(9):4577–4596
- Huang J, Horowitz JL, Ma S (2008) Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Ann Statistics* 36(2):587–613. <https://doi.org/10.1214/009053607000000875>
- Ignaccolo R, Ghigo S, Bande S (2013) Functional zoning for air quality. *Environ Ecol Stat* 20(1):109–127
- Ignaccolo R, Ghigo S, Giovenali E (2008) Analysis of air quality monitoring networks by functional clustering. *Environmetrics* 19(7):672–686
- Ivanoff S, Picard F, Rivoirard V (2016) Adaptive lasso and group-lasso for functional poisson regression. *J Mach Learn Res* 17(1):1903–1948
- Jennrich RI, Sampson PF (1976) Newton-raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics* 18(1):11–17. <https://doi.org/10.1080/00401706.1976.10489395>
- Jurek M, Katzfuss M (2021) Multi-resolution filters for massive spatio-temporal data [Journal Article]. *J Computational and Graphic Statistics* 30(4):1095–1110. <https://doi.org/10.1080/10618600.2021.1886938>
- Jurek M, Katzfuss M (2022) Hierarchical sparse cholesky decomposition with applications to high-dimensional spatio-temporal filtering [Journal Article]. *Stat Comput* 32(1):15. <https://doi.org/10.1007/s11222-021-10077-9>
- Jurek M, Katzfuss M (2022b) Scalable spatio-temporal smoothing via hierarchical sparse cholesky decomposition. *Environmetrics*, [arXiv:2207.09384](https://arxiv.org/abs/2207.09384)
- Kang M, Katzfuss M (2021) Correlation-based sparse inverse Cholesky factorization for fast Gaussian-process inference. *arXiv preprint* [arXiv:2112.14591](https://arxiv.org/abs/2112.14591)
- Karunasingha DSK (2022) Root mean square error or mean absolute error? use their ratio as well. *Inform Sci* 585:609–629. <https://doi.org/10.1016/j.ins.2021.11.036>
- Katzfuss M (2017a) A multi-resolution approximation for massive spatial datasets. *J Am Stat Assoc* 112(517):201–214
- Katzfuss M (2017b) A multi-resolution approximation for massive spatial datasets [Journal Article]. *J Am Stat Assoc* 112(517):201–214. <https://doi.org/10.1080/01621459.2015.1123632>
- Katzfuss M, Cressie N (2011) Spatio-temporal smoothing and estimation for massive remote-sensing data sets. *J Time Ser Anal* 32(4):430–446
- Katzfuss M, Cressie N (2012) Bayesian hierarchical spatio-temporal smoothing for very large datasets. *Environmetrics* 23(1):94–107
- Kaufman CG, Schervish MJ, Nychka DW (2008) Covariance tapering for likelihood-based estimation in large spatial data sets. *J Am Stat Assoc* 103(484):1545–1555
- Krock M, Kleiber W, Becker S (2021) Nonstationary modeling with sparsity for spatial data via the basis graphical lasso. *J Comput Graph Stat* 30(2):375–389
- Krock M, Kleiber W, Hammerling D, Becker S (2021) Modeling massive highly-multivariate nonstationary spatial data with the basis graphical lasso. *arXiv preprint* [arXiv:2101.02404](https://arxiv.org/abs/2101.02404)
- Lawson AB, Carroll R, Faes C, Kirby RS, Aregay M, Watjou K (2017) Spatiotemporal multivariate mixture models for bayesian model selection in disease mapping. *Environmetrics* 28(8):e2465. <https://doi.org/10.1002/env.2465>
- Lee J, Lund R (2008) Equivalent sample sizes in time series regressions. *J Stat Comput Simul* 78(4):285–297. <https://doi.org/10.1080/10629360600758484>
- Longford NT (1987) A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects [Journal Article]. *Biometrika* 74(4):817–827. <https://doi.org/10.1093/biomet/74.4.817>
- Lovarelli D, Conti C, Finzi A, Bacchetti J, Guarino M (2020) Describing the trend of ammonia, particulate matter and nitrogen oxides: The role of livestock activities in Northern Italy during COVID-19 quarantine. *Environmental Research* 191:110048. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0013935120309452><https://doi.org/10.1016/j.envres.2020.110048>
- Maranzano P (2022) Air quality in Lombardy, Italy: an overview of the environmental monitoring system of ARPA Lombardia. *Earth* 3(1):172–203
- McIlhagga W (2016) penalized: A matlab toolbox for fitting generalized linear models with penalties. 2016, 72(6):21. Retrieved from <https://www.jstatsoft.org/v072/i06><https://doi.org/10.18637/jss.v072.i06>
- Müller S, Welsh AH (2010) On model selection curves. *Int Stat Rev* 78(2):240–256. <https://doi.org/10.1111/j.1751-5823.2010.00108.x>
- Nandy S, Lim CY, Maiti T (2017) Additive model building for spatial regression. *J Royal Statistical Soc: Series B (Statistical Methodology)* 79(3):779–800. <https://doi.org/10.1111/rssb.12195>
- Otto P, Piter A, Gijsman R (2021) Statistical analysis of beach profiles - a spatiotemporal functional approach. *Coastal Engineering* 170:103999. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0378383921001526><https://doi.org/10.1016/j.coastaleng.2021.103999>
- Pannu J, Billor N (2017) Robust group-lasso for functional regression model [Journal Article]. *Commun Statistics - Simulation and Comput* 46(5):3356–3374. <https://doi.org/10.1080/03610918.2015.1096375>
- Pineda-Ríos W, Giraldo R, Porcu E (2019) Functional sar models: With application to spatial econometrics. *Spatial statistics* 29:145–159
- Piter A, Otto P, Alkhatib H (2022) The Helsinki bike-sharing system - insights gained from a spatiotemporal functional model. *Journal of the Royal Statistical Society Series A (online first)*
- Ramsay SB, JO (2005) *Functional data analysis* [Book]. Springer, New York NY. <https://doi.org/10.1007/b98888>
- Reyes PE, Zhu J, Aukema BH (2012) Selection of spatial-temporal lattice models: assessing the impact of climate conditions on a mountain pine beetle outbreak [Journal Article]. *J Agric Biol Environ Stat* 17(3):508–525. <https://doi.org/10.1007/s13253-012-0103-0>
- Rougier J, Brady A, Bamber J, Chuter S, Royston S, Vishwakarma BD, Ziegler Y (2023) The scope of the kalman filter for spatio-temporal applications in environmental science [Journal Article]. *Environmetrics* 34(1):e2773. <https://doi.org/10.1002/env.2773>
- Sabater M (2019) Era5-land hourly data from to present. copernicus climate change service (c3s) climate data store (cds). (accessed on 31- 01-2022) <https://doi.org/10.24381/cds.e2161bac> [Journal Article]
- Safikhani A, Kamba C, Mudigonda S, Faghih SS, Moghimi B (2020) Spatio-temporal modeling of yellow taxi demands in New York City using generalized STAR models. *Int J Forecasting* 36(3), 1138–1148. <https://doi.org/10.1016/j.ijforecast.2018.10.001>
- Sammut C, Webb GI (2010a) Mean absolute error [Journal Article]. *Encyclopedia of Machine Learning*, 652

- Sammut C, Webb GI (2010b) Mean squared error. *Encyclopedia of Mach Learn*, 653
- Sang H, Huang JZ (2012) A full scale approximation of covariance functions for large spatial data sets. *J Royal Statistical Soc: Series B (Statistical Methodology)* 74(1):111–132
- Schäfer F, Katzfuss M, Owahdi H (2021) Sparse Cholesky factorization by Kullback-Leibler minimization. *SIAM J Sci Comput* 43(3):A2019–A2046
- Silverman B, Ramsay J (2002) *Applied functional data analysis: methods and case studies* [Book]. Springer, New York, NY. <https://doi.org/10.1007/b98886>
- Simon N, Tibshirani R (2012) Standardization and the group lasso penalty [Journal Article]. *Stat Sin* 22(3):983
- Stein ML (2013) Statistical properties of covariance tapers [Journal Article]. *J Comput Graph Stat* 22(4):866–885. <https://doi.org/10.1080/10618600.2012.719844>
- Stein ML, Chi Z, Welty LJ (2004) Approximating likelihoods for large spatial data sets. *J Royal Statistical Soc: Series B (Statistical Methodology)* 66(2):275–296
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Roy Stat Soc: Ser B (Methodol)* 58(1):267–288
- Vitolo C, Russell A, Tucker A (2016) rdefra: Interact with the uk air pollution database from defra [Journal Article]. *J Open Source Softw* 1(4):51
- Wang H, Zhu J (2009) Variable selection in spatial regression via penalized least squares [Journal Article]. *Can J Statistics* 37(4):607–624. <https://doi.org/10.1002/cjs.10032>
- Wang Y, Finazzi F, Fassò A (2021) D-stem v2: a software for modeling functional spatio-temporal data. *J Statistical Softw* 99(10):1–29
- Willmott CJ, Matsuura K, Robeson SM (2009) Ambiguities inherent in sums-of-squares-based error statistics. *Atmos Environ* 43(3):749–752
- Xiao Y, Gu X, Yin S, Shao J, Cui Y, Zhang Q, Niu Y (2016) Geostatistical interpolation model selection based on arcgis and spatiotemporal variability analysis of groundwater level in piedmont plains, northwest China [Journal Article]. *Springerplus* 5(1):425. <https://doi.org/10.1186/s40064-016-2073-0>
- Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables [Journal Article]. *J Royal Statistical Soc: Series B (Statistical Methodology)* 68(1):49–67. <https://doi.org/10.1111/j.1467-9868.2005.00532.x>
- Yuan Y, Wu S, Yu Y, Tong G, Mo L, Yan D, Li F (2018) Spatiotemporal interaction between ecosystem services and urbanization: case study of Nanjing City, China. *Ecol Indic* 95:917–929. <https://doi.org/10.1016/j.ecolind.2018.07.056>
- Zhang C-H (2010) Nearly unbiased variable selection under minimax concave penalty. *Ann Statistics* 38(2):894–942. <https://doi.org/10.1214/09-AOS729>
- Zhao P, Yu B (2006) On model selection consistency of lasso. *J Mach Learn Res* 7:2541–2563
- Zhao Y, Ogden RT, Reiss PT (2012) Wavelet-based lasso in functional linear regression [Journal Article]. *J Comput Graph Stat* 21(3):600–617. <https://doi.org/10.1080/10618600.2012.679241>
- Zhu J, Huang H-C, Reyes PE (2010) On selection of spatial linear models for lattice data. *J Royal Statistical Soc: Series B (Statistical Methodology)* 72(3):389–402. <https://doi.org/10.1111/j.1467-9868.2010.00739.x>
- Zou H (2006) The adaptive lasso and its oracle properties. *J Am Stat Assoc* 101(476):1418–1429
- Zou H, Li R (2008) One-step sparse estimates in nonconcave penalized likelihood models. *Ann Stat* 36(4):1509

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