A functional (lagged) time series regression model involves the regression of scalar response time series on a time series of regressors that consists of a sequence of random functions. In practice, the underlying regressor curve time series are not always directly accessible, but are latent processes observed (sampled) only at discrete measurement locations. In this article, we consider the so-called sparse observation scenario where only a relatively small number of measurement locations have been observed, possibly different for each curve. The measurements can be further contaminated by additive measurement error. A spectral approach to the estimation of the model dynamics is considered. The spectral density of the regressor time series and the cross-spectral density between the regressors and response time series are estimated by kernel smoothing methods from the sparse observations. The impulse response regression coefficients of the lagged regression model are then estimated by means of ridge regression (Tikhonov regularization) or principal component analysis (PCA) regression (spectral truncation). The latent functional time series are then recovered by means of prediction, conditioning on all the observed data. The performance and implementation of our methods are illustrated by means of a simulation study and the analysis of meteorological data.

Received 23 May 2019; Accepted 19 June 2020

Keywords: Autocovariance operator; lagged regression; functional data analysis; non-parametric regression; spectral density operator.

MOS subject classification: 62M10; 62M15; 60G10.

1. INTRODUCTION

A (lagged) time series regression model is perhaps the most basic – and certainly one of the most and longest studied (Kolmogoroff, 1941; Wiener, 1950) – forms of coupled analysis of two time series. In a general context, given two discrete-time stationary time series \( \{X_t\}_{t \in \mathbb{Z}} \) and \( \{Z_t\}_{t \in \mathbb{Z}} \), the input (or regressor) and output (or response), valued in some vector spaces \( H_1 \) and \( H_2 \), such a model postulates that

\[
Z_t = a + \sum_{k \in \mathbb{Z}} B_k X_{t-k} + e_t, \quad t \in \mathbb{Z},
\]

for some constant \( a \in H_2 \), a sequence of random disturbances \( \{e_t\}_{t \in \mathbb{Z}} \) valued in \( H_2 \) and a sequence of linear mappings \( B_k : H_1 \to H_2, k \in \mathbb{Z} \). This linear coupling would be the typical dependence model, for instance, if \( \{(X_t, Z_t)\}_{t \in \mathbb{Z}} \) were a jointly Gaussian stationary process in \( H_1 \times H_2 \), and is also known as a time-invariant linearly filtered time series model. The estimation problem is then to estimate the unknown transformations \( \{B_k\}_{k \in \mathbb{Z}} \) given the realization of a finite stretch of the joint series \( \{(Z_t, X_t)\}_{t \in \mathbb{Z}} \) (a problem also known as system identification, particularly in signal processing).

This problem is very well understood and has been extensively studied in the classical context where the spaces \( H_j, j = 1, 2 \), coincides with the Euclidean spaces of potentially different dimensions (Brillinger, 1981; Priestley,
Nevertheless, generalizing these results to the case where either space, and particularly the regressor space $H_1$, may be an infinite dimensional vector space is far from straightforward, and has been comparatively less studied. The difficulty in this case is that one needs to manipulate operations involving the inverses of compact or even trace-class operators, which fail to exist boundedly on the entire codomain. Consequently, analysis of such models requires drawing on tools from functional analysis, and developing novel methodology that incorporates suitable regularizing schemes. Such a setting has only recently been considered for functional time series regression models, for example by Hörmann et al. (2015b), who treat the problem of estimation of the filter coefficients by means of spectral truncation regularization (principal component analysis regression), and Pham and Panaretos (2018), who deduce convergence rates for the estimated coefficients when using Tikhonov regularization (ridge regression).

While this may seem to be an artificial abstraction at first sight, such infinite-dimensional time series are becoming increasingly prominent for applications. Indeed the abstraction of an infinite dimensional (Hilbert) space captures the scenario where the value of a series at each time is a square-integrable function, for example, a curve. Examples include time series of DNA minicircles evolving in solution (seen as a time series of closed curves in 3D indexed by discrete time Tavakoli and Panaretos, 2016) or the data constructed by dividing a continuously observed scalar time series into segments of an obvious periodicity, usually days. Examples of the latter form are particularly prominent in environmental applications, for example the analysis of particulate matter atmospheric pollution (Hörmann and Kokoszka, 2010; Hörmann et al., 2015, 2018; Aue et al., 2015), traffic data modelling (Klepšch et al., 2017) or financial applications of intra-day trading (Müller et al., 2011; Kokoszka et al., 2017). Another promising financial application of functional time series emerges in the yield curve modelling (Hays et al., 2012; Kowal et al., 2017b; Sen and Klüppelberg, 2019).

We focus here primarily on the case $H_2 = \mathbb{R}$, that is, when the response process is scalar. This is indeed the case that is most often studied in the literature, due in large part due to the many examples it covers, but also because it is the simplest version of the problem that captures the essence of higher generality: the difficulty in estimating the filter coefficients lies in the ill-possessedness of the spectral density operator inversion that requires regularization. For a scenario involving $H_2$ being infinite dimensional, that is, with a functional response, see Hörmann et al. (2015b) or the Supporting information of this article confirming this difficulty.

In practice, one can never observe the regressor time series $\{X_t\}_{t \in \mathbb{Z}}$ in its ‘Platonic’ continuum form. For example, if $H_1$ is a space of functions on $[0, 1]$ and each $X_t : [0, 1] \to \mathbb{R}$ is a curve, then one might be able to observe evaluations of $X_t(\cdot)$ at various locations of its domain. In some cases, the sampled locations are sufficiently dense, and the measurement instrumentation sufficiently precise to be free of any additional noise contamination, so that one can disregard the effects of sampling. This is essentially the approach taken in Hörmann et al. (2015b) and Pham and Panaretos (2018) where the regressors are treated as being fully observed as elements of $H_1 = L^2[0, 1]$. However, it may well happen that $\{X_t\}_{t \in \mathbb{Z}}$ is only measured at few and randomly positioned locations in its domain, indeed varying with time, and that the measurements are themselves contaminated by noise. That is, instead of observing $X_t(u)$ for all $u$ in $[0, 1]$, we instead observe

$$Y_t = X_t(x_j) + \varepsilon_t, \quad j = 1, \ldots, N_t, \quad t = 1, \ldots, T.$$ 

for a sequence of point processes $(x_1, \ldots, x_{N_t})$, $t = 1, \ldots, T$, independent in time, and a white noise measurement error sequence. This observation scheme is illustrated in Figure 1 and is exhibited, for example, when dealing with fair weather atmospheric electricity data (Tammet, 2009; Rubin and Panaretos, 2020).

Such a setting escapes both the methods and the theory developed at the level of continuum, and instead requires methodology that accounts for the observation scheme, as well as theory that incorporates the latent/emission processes explicitly. The purpose of this article is precisely to construct such a methodology and develop the associated asymptotic theory.

Our approach consists in estimating the complete space-time covariance structure of the data by estimating the spectral density operator of the regressor time series, and the cross-spectral density operator between the regressor and response time series. Our estimators are based on the kernel smoothing methods as discussed for example...
in Yao et al. (2005a,b), Hall et al. (2006) and Li and Hsing (2010), and extended to the time series case in Rubín and Panaretos (2020). Once these are estimated, one obtains estimating equations whose solution yields the estimators of the filter coefficients. The solution of the estimating equations, however, comprises an ill-posed inverse problem, hence regularization is required. We offer two regularization techniques, spectral truncation regularization (Hörmann et al., 2015b) and Tikhonov regularization (Pham and Panaretos, 2018). The forecasting of the response process is then implemented by first predicting the latent functional regressor data (using their estimated spectral characteristics) and then plugging-in these predictions into the estimated lagged regression model.

Sparsely observed functional time series have only recently received attention (Kowal et al. 2017a,b; Sen and Klüppelberg, 2019; Rubín and Panaretos, 2020), and our results appear to be the first in the context of the lagged regression model where the regressor process is functional. A related problem of dynamic function-on-scalar regression was studied by Kowal (2018) by means of Bayesian factor models.

Our presentation is organized as follows. Section 2 defines the framework of sparsely observed functional time series as well as the functional lagged regression model and its analysis in the spectral domain. The section explains the estimation methodology for the model components, the spectral and the cross-spectral densities, and the regularized estimator of the filter coefficients of the lagged regression. Furthermore, the forecasting algorithm is introduced. Section 3 presents the asymptotic results of the proposed method: consistency and convergence rates are established. Section 4 probes the finite sample properties of the proposed methodology on a simulation study. As a proof-of-concept illustration, we included an example featuring meteorological data in Section 5 of the article. The properties of the two considered regularization methods are summarized in Section 6. The proofs of the formal statements and some extensions of the proposed model are presented in the Supporting information of this article.

2. MODEL AND ESTIMATION METHODOLOGY

2.1. Functional Time Series Regression Model

The classical theory of functional data analysis views its data, random curves, as random elements in a suitable Hilbert space. Throughout our presentation we assume this space to be the Hilbert space of real square integrable function, \( H_1 = L^2([0, 1]) \), equipped with the inner product \( \langle \cdot, \cdot \rangle \) and the norm \( \| \cdot \| \). The functional time series is then a sequence of random elements in such space indexed by the integer variable \( t \in \mathbb{Z} \), interpreted as times (for example days), and denoted as \( \{ X_t \} \equiv \{ X_t(\cdot) \}_{t \in \mathbb{Z}} \). The argument of the individual functions, for the most
part of this article denoted as \( x \in [0, 1] \), can be interpreted as either the spatial location or intra-day time (for the sequence of intra-day data order day-by-day into a sequence). The basic assumption we impose is:

(A1) The functional time series \( \{X_t\} \) is assumed to be square-integrable, that is, \( \mathbb{E}[\|X_t\|^2] < \infty \), \( t \in \mathbb{Z} \), and second-order stationary in the variable \( t \). Furthermore, the sample paths (trajectories) of \( \{X_t\} \) are smooth.

The functional time series \( \{X_t\} \) is assumed to be only latent and therefore not accessible directly. Denote \( N_t \), the number of measurements available for the random curve at time \( t = 1, \ldots, T \) where \( T \in \mathbb{N} \) is the time of the latest curve with at least one measurement. We denote \( Y_{ij} \) the \( j \)th measurement on the \( r \)th curve at spatial position \( x_i \in [0, 1] \) where \( j = 1, \ldots, N_t \) and \( t = 1, \ldots, T \). The measurements are assumed to be contaminated by additive measurement errors \( \epsilon_{ij} \), \( j = 1, \ldots, N_t \), \( t = 1, \ldots, T \) which constitute an independent identically distributed zero-mean array of scalar random variables with variance \( \sigma^2 > 0 \). The number of measurement locations \( \{N_t\} \), the measurement errors \( \{\epsilon_{ij}\} \), and the underlying functional time series \( \{X_t\} \) are assumed to be independent. The above outlined sparse observation regime leads to the model equation

\[
Y_{ij} = X_i(x_j) + \epsilon_{ij}, \quad j = 1, \ldots, N_t, \quad t = 1, \ldots, T.
\]

The assumption (A1) allows us to define the mean function as \( \mu(x) = \mathbb{E}X_0(x), \ x \in [0, 1] \), and the lag-\( h \) autocovariance kernel \( R_h^X(x, y) = \mathbb{E}[\{X_0(x) - \mu(x)\}\{X_0(y) - \mu(y)\}], \ x, y \in [0, 1] \), for \( h \in \mathbb{Z} \). The integral operator \( R_h^X \) associated with this kernel, the lag-\( h \) autocovariance operator, is defined by the right integration \( (R_h^X f)(x) = \int_0^1 R_h^X(x, y)f(y)dy, f \in L^2([0,1]) \). The autocovariance kernels and operators are assumed to be summable in the supremum norm (denoted \( \| \cdot \|_\infty \)) and the nuclear norm (denoted \( \| \cdot \|_1 \)) respectively,

(A2)

\[
\sum_{h \in \mathbb{Z}} \|R_h^X\|_\infty = \sum_{h \in \mathbb{Z}} \sup_{x, y \in [0, 1]} |R_h^X(x, y)| < \infty, \quad \sum_{h \in \mathbb{Z}} \|R_h^X\|_1 = \sum_{h \in \mathbb{Z}} \text{trace} \left\{ \sqrt{(R_h^X)^*(R_h^X)} \right\} < \infty.
\]

Throughout the article we assume that the response time series \( \{Z_t\} \) is scalar. We make this assumption to facilitate the presentation while keeping in mind the fact that the scalar case already demonstrates the difficulty of the ill-posed estimation problem and the requirement of regularization. The extension to a functional response time series is presented in the Supporting information of the article.

The functional lagged regression model with a scalar response time series \( \{Z_t\} \) is given by the equation

\[
Z_t = a + \sum_{k \in \mathbb{Z}} B_k X_{t-k} + e_t, \quad t \in \mathbb{Z},
\]

where \( a \in \mathbb{R} \) is a constant, called the intercept; \( B_k : \mathcal{H} \to \mathbb{R}, k \in \mathbb{Z} \), are fixed continuous linear functionals called the filter coefficients, with Riesz-representers \( \{b_k\}_{k \in \mathbb{Z}} \) (i.e. \( B_kf = \langle f, b_k \rangle \) for \( f \in \mathcal{H} \)); and \( \{e_t\}_{t \in \mathbb{Z}} \) is a sequence of independent identically distributed zero-mean real random variables with variance \( \tau^2 > 0 \). The time series \( \{X_t\} \), the sampling counts \( \{N_t\} \), the observation locations \( \{x_i\} \), and the errors \( \{e_{ij}\} \) and \( \{e_t\} \) are assumed to be independent. The filter coefficients are assumed to be summable in the vector norm and the supremum norm

(A3)

\[
\sum_{k \in \mathbb{Z}} \|B_k\| < \infty, \quad \sum_{k \in \mathbb{Z}} \|B_k\|_\infty = \sum_{k \in \mathbb{Z}} \sup_{x \in [0, 1]} |b_k(x)| < \infty.
\]

where given any continuous linear functional \( \mathcal{G} : \mathcal{H} \to \mathbb{R} \) the notation \( \|\mathcal{G}\| \) will be used interchangeably with \( \|g\| \) for \( g \) the Riesz-representer of \( \mathcal{G} \), that is, \( \langle \mathcal{G}, f \rangle \).

We make here a few simplifying assumptions concerning the first order structure of the data, to keep focus on the more challenging second-order structure. First, we assume that the regressor time series \( \{X_t\} \) is centred, that
is, \( \mu = 0 \). If that was not the case, the mean function \( \mu(\cdot) \) could be estimated by the local-linear kernel smoother (Yao et al., 2005a; Rubín and Panaretos, 2020) and all the sparse observations centred. Second, we assume that the intercept is null, that is, \( a = 0 \). Otherwise it could be estimated using the relation \( a = \mathbb{E} [Z_0] - \sum_{k \in \mathbb{Z}} B_k \mu \) where \( \mathbb{E} [Z_0] \) is estimated by the sample mean and the coefficients \( B_k \) using the methods of this article. As a consequence of these assumptions, the response time series is also centred, that is, \( \mathbb{E} [X_t] = 0, \mathbb{E} [Z_t] = 0 \).

Finally, we define the lag-\( h \) cross-covariance kernel and the lag-\( h \) cross-covariance operator between the scalar time series \( \{Z_t\} \) and the functional time series \( \{X_t\} \), for \( h \in \mathbb{Z} \), as

\[
R^{ZX}_{h}(x) = \mathbb{E} [Z_h X_0(x)], \quad x \in [0, 1], \quad \mathcal{R}^{ZX}_{h} = \mathbb{E} [Z_h (X_h \cdot)] : \mathcal{H} \to \mathbb{R}.
\]

### 2.2. Spectral Analysis of Functional Lagged Regression

Under the assumptions (A1) and (A2), the spectral density kernel and the spectral density operator (Panaretos and Tavakoli, 2013) at frequency \( \omega \in [\pi, \pi] \) are defined by

\[
f^{X}_\omega (\cdot, \cdot) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} R^{X}_{h}(\cdot, \cdot)e^{-i\omega h}, \quad \mathcal{F}^{X}_\omega = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \mathcal{R}^{X}_{h} e^{-i\omega h}.
\]

The sums (3) converge in the supremum and the trace-class norm respectively and define the spectral density kernel and the spectral density operator at frequency \( \omega \in [\pi, \pi] \) respectively. The spectral density operator \( \mathcal{F}^{X}_\omega \) is a non-negative self-adjoint trace-class operator for all \( \omega \in [-\pi, \pi] \). Hence it admits the spectral representation

\[
\mathcal{F}^{X}_\omega = \sum_{m=1}^{\infty} \lambda^{m}_\omega \varphi^{m}_\omega \otimes \varphi^{m}_\omega = \sum_{m=1}^{\infty} \varphi^{m}_\omega \langle \varphi^{m}_\omega \cdot \rangle \varphi^{m}_\omega
\]

where \( \otimes \) is the tensor product in \( \mathcal{H} \), defined by the inner product formula on the right-hand side of (4). The elements of the sequence \( \lambda^{m}_\omega \geq \lambda^{2}_\omega \geq \cdots \geq 0 \) are called the harmonic eigenvalues and the corresponding functions \( \{\varphi^{m}_\omega\}_{m=1}^{\infty} \) are called the harmonic eigenfunctions.

Furthermore, the lagged autocovariance kernels and operators can be recovered by the inversion formula (Panaretos and Tavakoli, 2013) that holds in the supremum and the nuclear norm respectively:

\[
R^{X}_{h}(\cdot, \cdot) = \int_{-\pi}^{\pi} f^{X}_\omega (\cdot, \cdot)e^{i\omega h} d\omega, \quad \mathcal{F}^{X}_h = \int_{-\pi}^{\pi} \mathcal{F}^{X}_\omega e^{i\omega h} d\omega, \quad h \in \mathbb{Z}.
\]

Hörmann et al. (2015b) defined the cross-spectral density operator between \( \{Z_t\}_{t \in \mathbb{Z}} \) and \( \{X_t\}_{t \in \mathbb{Z}} \) by the formula

\[
\mathcal{F}^{ZX}_{\omega} = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \mathcal{R}^{ZX}_{h} e^{-i\omega h}, \quad \omega \in [-\pi, \pi],
\]

and showed that this sum converges in the Hilbert–Schmidt norm, provided the autocovariance function \( \mathcal{R}^{X}_{h} \) is summable in the Hilbert–Schmidt norm and the sequence of filter coefficient operators are also summable in the operator norm. Since our assumptions (A1)–(A3) are stronger and we consider only scalar response, (6) is well defined in the setting of our article.
Likewise, one defines the cross-spectral density kernel between \( \{Z_t\}_{t \in \mathbb{Z}} \) and \( \{X_t\}_{t \in \mathbb{Z}} \) by
\[
\hat{f}_{XZ}(\omega) = \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} R^{XZ}_{\hat{h}}(\omega) e^{-i\omega t}, \quad \omega \in [-\pi, \pi].
\] (7)

The sum on the right-hand side of (7) converges in the supremum norm.

Furthermore, Hörmann et al. (2015b) introduced the frequency response operator
\[
\mathcal{B}_\omega = \sum_{h \in \mathbb{Z}} B_h e^{-i\omega h}, \quad \omega \in [-\pi, \pi],
\]
and obtained the relation between the spectral density operators, cross-spectral density operators and the frequency response operators
\[
\mathcal{F}_{XZ} = \mathcal{B}_\omega \mathcal{F}_X, \quad \omega \in [-\pi, \pi],
\] (8)

which provides the basis for the estimation of the filter coefficients introduced in the next section. In our case of scalar response, \( \mathcal{B}_\omega \) is in fact a functional, that is, \( \mathcal{B}_\omega : \mathcal{H} \rightarrow \mathbb{R} \). The filter coefficients \( \{B_h\}_{h \in \mathbb{Z}} \) can be recovered by the formula
\[
B_h = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{B}_\omega e^{i\omega} d\omega, \quad k \in \mathbb{Z}.
\] (9)

2.3. Non-Parametric Estimation in the Lagged Regression

The non-parametric estimation in functional models given sparse noisy measurements has been frequently based on local-polynomial kernel smoothers (Yao et al., 2005a; Hall et al., 2006; Li and Hsing, 2010; Rubín and Panaretos, 2020). Throughout the article we use the Epanechnikov kernel \( K(v) = \frac{3}{4}(1 - v^2) \) for \( v \in [-1, 1] \), and 0 otherwise, because it is known to minimize the mean square error in the classical local-polynomial smoothing problem (Fan and Gijbels, 1996, Section 3.2.6). Moreover, the Epanechnikov kernel’s bounded support speeds up the numerical calculations. Nevertheless, the choice of the kernel function is not crucial for the statistical performance of the smoothing estimators (Fan and Gijbels, 1996, Section 3.2.6) and any other commonly used kernel function could be adopted.

We start by defining the ‘raw’ covariances of \( \{X_t\} \) as \( G^X_{h,j} (x_{t+h}, x_{t+j}) = Y_{t+h} Y_{t+j} \) for \( h = -T, \ldots, T, \) \( j = 1, \ldots, N_t \). The lag-0 covariance kernel \( R^X_0(\cdot, \cdot) \) is estimated by the local-linear surface smoother, setting \( \hat{R}^X_0(x, y) = c_{0}^{(1)} \) where
\[
\left( c_{0}^{(1)}, c_{1}^{(1)}, c_{2}^{(1)} \right) = \arg \min_{c_0, c_1, c_2} \sum_{t=1}^{T} \sum_{j=1}^{N_t} \sum_{k=1}^{N_t} \left( \frac{X_{t+j} - X}{B_R} \right) K \left( \frac{X_{t+k} - X}{B_R} \right) \times \left\{ G^X_{0,0} (x_{t+k}, x_{t+j}) - c_{0}^{(1)} - c_{1}^{(1)} (x - x_{t+k}) - c_{2}^{(1)} (y - x_{t+j}) \right\}^2
\]
for \( x, y \in [0, 1] \) and where \( B_R > 0 \) is the bandwidth parameter for the surface smoother.

In the next step we aim to estimate the measurement error variance \( \sigma^2 \) by the approach suggested by Yao et al. (2003,2005a) for which we need the two following ingredients: the estimator of the diagonal of the lag-0 covariance kernel of \( \{X_t\} \) with and without the measurement noise contamination. First, we estimate the diagonal of \( R^X_0 \) without the measurement noise contamination by the local-quadratic smoother along the direction perpendicular...
where \( \Delta(x_j, x_k) \) is the distance of the point \((x_j, x_k)\) from the diagonal equipped with the positive sign if the point \((x_j, x_k)\) is above the diagonal, and with the negative sign if below. Formally

\[
\Delta(x_j, x_k) = \text{sign}(x_k - x_j) \sqrt{(P(x_j, x_k) - x_j)^2 + (P(x_j, x_k) - x_k)^2},
\]

where \( \text{sign}(\cdot) \in \{-1, 0, 1\} \) is the sign function and \( P(x_j, x_k) \) is the first coordinate of the point \((x_j, x_k)\) projected onto the diagonal of \([0, 1]^2\).

Second, we estimate the function \( x \mapsto R^X_0(x, x) + \sigma^2 \), \( x \in [0, 1] \), that is, the noise contaminated diagonal of the lag-0 covariance operator. For \( x \in [0, 1] \) and a bandwidth parameter \( B_V > 0 \), we use the local-linear line smoother and set \( \hat{X}^X(x) = c_0^{(0)} \), where

\[
\begin{align*}
\left( c_0^{(0)}, c_1^{(0)} \right) &= \arg \min_{c_0^{(0)}, c_1^{(0)}} \sum_{t=1}^{T} \sum_{j=1}^{N_t} K \left( \frac{x_{ij} - x}{B_V} \right) \left( x_{ij} - c_0^{(0)} - c_1^{(0)} f_{ij} \right)^2, \\
\left( c_0^{(3)}, c_1^{(3)} \right) &= \arg \min_{c_0^{(3)}, c_1^{(3)}} \sum_{t=1}^{T} \sum_{j=1}^{N_t} K \left( \frac{x_{ij} - x}{B_V} \right) \left( c_{0,j}^{X}(x_{ij}, x_{jk}) - c_0^{(3)}(x - x_{ij}) \right)^2.
\end{align*}
\]

Having the estimates \( \hat{R}^X_0(\cdot) \) and \( \hat{X}^X(\cdot) \), the measurement error variance \( \sigma^2 \) is estimated by integrating the difference

\[
\sigma^2 = \int_0^1 \left( \hat{X}^X(x) - \hat{R}^X_0(x) \right) \, dx.
\]

In case the right-hand side of (10) is negative, it is recommended to replace it with a small positive number (Yao et al., 2005a).

The spectral density kernels \( \left\{ R^X_{0,j} \right\}_{j \in \mathbb{Z}} \) are estimated by Bartlett’s approach (Hörmann et al., 2015a; Rubín and Panaretos, 2020), weighing down higher-order lags using Bartlett’s (triangular) weights defined as \( W_h = (1 - |h|/L) \) for \(|h| < L \) and 0 otherwise. The parameter \( L \in \mathbb{N} \) is called Bartlett’s span parameter and controls the amount of regularization involved in the spectral density estimation. For a fixed \( \omega \in [-\pi, \pi] \), the spectral density kernel at frequency \( \omega \) and \((x, y) \in [0, 1]^2\) is estimated as

\[
\hat{f}^X_0(x, y) = \frac{L}{2\pi} \hat{c}_0^{(4)} (x, y) \in \mathbb{C},
\]

where \( \hat{c}_0^{(4)} \in \mathbb{C} \) is obtained by minimizing the following weighted sum of squares

\[
\begin{align*}
\left( c_0^{(4)}, c_1^{(4)}, c_2^{(4)} \right) &= \arg \min_{c_0^{(4)}, c_1^{(4)}, c_2^{(4)}} \sum_{h = -L}^{L} \frac{1}{\mathcal{N}_h} \sum_{|h| = \max(1, |h|)}^{\min(T, T-h)} \sum_{j=0}^{N_{ij}} \sum_{k=1}^{N_i} G_{wh}^X(x_{ij+h, j}, x_{ik}) e^{-ih\omega} \\
& \quad \left( c_0^{(4)} - c_0^{(4)}(x_{ij+h, j} - x) - c_2^{(4)}(x_{ij+h, j} - y) \right)^2 W_h \frac{1}{B^2_R} K \left( \frac{x_{ij+h, j} - x}{B_R} \right) K \left( \frac{x_{ij+h, j} - y}{B_R} \right),
\end{align*}
\]

where \( \mathcal{N}_h = (T - |h|)(\bar{N})^2 \) for \( h \neq 0 \), \( \mathcal{N}_0 = T(\bar{N}^2 - \bar{N}) \), and where \( \bar{N} = (1/T) \sum_{t=1}^{T} N_t \), and \( \bar{N}^2 = (1/T) \sum_{t=1}^{T} N_t^2 \).

wileyonlinelibrary.com/journal/jtsa  © 2020 John Wiley & Sons Ltd J. Time Ser. Anal. 41: 858–882 (2020)

DOI: 10.1111/jtsa.12551
The spectral density estimate \( \hat{R}_h^X(\cdot,\cdot) \) and its operator counterpart \( \hat{F}_h^X \) enable us to estimate all autocovariance kernels and operators, thus to estimate the complete second-order structure of \( \{X_t\} \), by the inversion formula

\[
\hat{R}_h^X(\cdot,\cdot) = \int_{-\pi}^{\pi} \hat{R}_h^X(\cdot)e^{i\omega h_0}d\omega, \quad \hat{F}_h^X = \int_{-\pi}^{\pi} \hat{F}_h^X e^{i\omega h_0}d\omega.
\]

(13)

In the following paragraphs we extend these kernel smoothing techniques to estimate the cross-spectral density between the response time series \( \{Z_t\}_{t\in\mathbb{Z}} \) and the regressor time series \( \{X_t\}_{t\in\mathbb{Z}} \).

Define the ‘raw’ lag-\(h\) cross covariances \( G_{h,j}(x_{\omega}) = Z_{t+h}Y_{t} \) where \( h = -T + 1, \ldots, T - 1, t = \max(1,T-h), \ldots, \min(T,T-h) \) and \( j = 1, \ldots, N_t \). Using the local linear kernel smoothing techniques, we estimate the cross-spectral density at frequency \( \omega \in [-\pi,\pi] \) and at \( x \in [0,1] \) as

\[
f^{XY}(x) = \frac{L}{2\pi} c_0^{(s)} (\in C),
\]

where \( c_0^{(s)} \) is realized as the minimizer of the following weighted sum of squares

\[
\left( c_0^{(s)}, c_1^{(s)} \right) = \arg\min \left( \sum_{h=-L}^{L} \sum_{j=1}^{N_t} W_h B_C K \left( \frac{x_{\omega} - x}{B_C} \right) \right) \times \left| G_{h,j}(x_{\omega})e^{-i\omega h} - c_0^{(s)} - c_1^{(s)}(x_{\omega}) \right| ^2,
\]

(15)

where \( B_C > 0 \) is a bandwidth parameter.

The solutions to all above least square optimization problems can be found explicitly by using a standard argument in local-polynomial regression (Fan and Gijbels, 1996, Section 3.1) or (Rubin and Panaretos, 2020, Section B.2). Moreover, the solutions to the spectral density estimator (12) and (15) depend on a handful of terms independent of the frequency \( \omega \in [-\pi,\pi] \), that can be precalculated, and multiplication by complex exponentials. This allows a computationally feasible evaluation even on a fine grid of frequencies.

Once the estimates of the spectral density kernels \( \{ \hat{F}_h^X \}_{h \in [-\pi,\pi]} \) and the cross-spectral density \( \{ \hat{G}^{\omega}_h \}_{\omega \in [-\pi,\pi]} \) have been constructed, we focus attention on the estimation of the frequency response operators \( \{ \hat{B}_\omega \}_{\omega \in [-\pi,\pi]} \).

Heuristically, from relation (8), we would like to write \( \hat{B}_\omega = \hat{F}^X \hat{F}^{-1}_\omega \), \( \omega \in [-\pi,\pi] \). This formula is indeed only heuristic because the operator \( \hat{F}^{-1}_\omega \), being trace class, is not boundedly invertible. The same issue is present also for its empirical counterpart \( \hat{B}^{\omega}_h \) \( \hat{F}^{-1}_\omega \). Therefore, to achieve consistent estimation, a regularization of the inverse \( \hat{F}^{-1}_\omega \) is required.

Being a self-adjoint trace class operator, \( \hat{F}^X \) admits the spectral representation

\[
\hat{F}^X = \sum_{j=1}^{\infty} \lambda_j^\omega \hat{x}_j^\omega \otimes \hat{x}_j^\omega, \quad \omega \in [-\pi,\pi],
\]

which can be viewed as the empirical version of (4). The difficulty in inverting \( \hat{F}^{-1}_\omega \) can be seen from the fact that \( \lambda_j^\omega = \text{trace}(\hat{F}_j^X) < \infty \), implying that \( \lambda_j^\omega \) decays at least as fast as \( j^{-1+\delta} \), \( \delta > 0 \). It is the small values of \( \lambda_j^\omega \) that cause problems and there are two classical strategies to overcome the issue: spectral truncation and the Tikhonov regularization.

1. **Spectral truncation.** The inverse \( \hat{F}^{-1}_\omega \) is replaced by

\[
\sum_{j=1}^{K_\omega} \frac{1}{\lambda_j^\omega} \hat{x}_j^\omega \otimes \hat{x}_j^\omega, \quad \omega \in [-\pi,\pi],
\]

\( J. \; Time \; Ser. \; Anal. \; 41: \; 858–882 \; (2020) \) © 2020 John Wiley & Sons Ltd wileyonlinelibrary.com/journal/jtsa
where \( K_\omega \in \mathbb{N} \), \( \omega \in [-\pi, \pi] \), is the spectral truncation parameter that needs to grow to infinity sufficiently slowly to allow for the consistency. It may or may not depend on the frequency \( \omega \in [-\pi, \pi] \).

The estimator of the spectral transfer function becomes

\[
\hat{\mathcal{R}}_{\omega}^{\text{trunc}} = \sum_{j=1}^{K_\omega} \frac{1}{\lambda_j^\omega} \langle \hat{\phi}_j^\omega, \cdot \rangle \hat{\mathcal{F}} X \hat{\phi}_j^\omega, \quad \omega \in [-\pi, \pi].
\] (16)

We opt to implement the spectral truncation by relying on eigenvalue thresholding approach (Hörmann et al., 2015b) where we implement the eigenvalue threshold selection by cross-validation (more in Section 4).

2. Tikhonov regularization. Here, the inverse of \( \hat{\mathcal{F}} X \) is replaced by

\[
(\hat{\mathcal{F}} X + \rho I)^{-1} = \sum_{j=1}^{\infty} \frac{1}{\lambda_j^\omega + \rho} \hat{\phi}_j^\omega \otimes \hat{\phi}_j^\omega, \quad \omega \in [-\pi, \pi],
\]

where \( I \) is the identity operator on \( H \) and the Tikhonov regularization parameter \( \rho > 0 \) tends to zero as \( T \to \infty \) slowly enough to allow for consistency. Even though the parameter \( \rho \) may, in general, depend on \( \omega \) we carry out further analysis with the frequency independent parameter. We do so because in the implementation (more in Section 4) we select the tuning parameter \( \rho \) using the cross-validation where it is feasible to optimize over a single (frequency independent) tuning parameter.

The estimator of the spectral transfer function becomes

\[
\hat{\mathcal{R}}_{\omega}^{\text{Tikh}} = \hat{\mathcal{F}} X (\hat{\mathcal{F}} X + \rho I)^{-1} = \sum_{j=1}^{\infty} \frac{1}{\lambda_j^\omega + \rho} \langle \hat{\phi}_j^\omega, \cdot \rangle \hat{\mathcal{F}} X \hat{\phi}_j^\omega, \quad \omega \in [-\pi, \pi].
\] (17)

This form of regularization was adopted and studied by Pham and Panaretos (2018).

Once the estimators of the spectral transfer operator \( \hat{\mathcal{R}}_{\omega} \) have been constructed by either of the above regularization techniques, the filter coefficients are estimated by

\[
\hat{B}_{trunc}^{\text{trunc}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\mathcal{R}}_{\omega}^{\text{trunc}} e^{i\omega k} d\omega, \quad k \in \mathbb{Z},
\] (18)

\[
\hat{B}_{trunc}^{\text{Tikh}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\mathcal{R}}_{\omega}^{\text{Tikh}} e^{i\omega k} d\omega, \quad k \in \mathbb{Z}.
\] (19)

2.4. Forecasting the Response Process

In Section 2 we assumed the data to be available up to time \( T \) for both the regressor time series \( \{X_t\} \) as well as the response time series \( \{Z_t\} \). It may very well happen, though, that the measurement of the response variable is terminated at an earlier time \( S \) where \( 1 < S < T \) but the measurements of the regressor time series \( \{X_t\} \) are available until time \( T \). In this case it is of interest to forecast the unobserved values of \( Z_{S+1}, \ldots, Z_T \). It turns out that the forecast of \( Z_{S+1}, \ldots, Z_T \) is surprisingly straightforward, once the model dynamics have been estimated. We present our forecasting method first by assuming the model dynamics are known, and then we plug-in its estimates.

Since the filter coefficients satisfy the assumption (A3), their norms converge to zero as \( |k| \to \infty \). Choose a constant \( M \in \mathbb{N} \) such that \( \|B_k\| \) is negligible for \( |k| > M \). Denote \( \mathcal{X} = [X_{M+1}, \ldots, X_{T+M}] \in \mathbb{H}^{T+2M} \) the random element representing ‘stacked’ curves of the latent regressors time series.

The vector \( \mathcal{Y} = (Y_{1i}, \ldots, Y_{N_1i}, \ldots, Y_{1N}, \ldots, Y_{TN}) \in \mathbb{R}^{N_1} \) consists of all observed data of the regressor time series where \( N_1 = \sum_{i=1}^{T} N_i \) is the total number of observations up to time \( T \). The measurement
errors \( \{ e_t \} \) are stacked into a vector denoted \( \mathcal{E} \in \mathbb{R}^{T \times T} \). Further define the evaluation operators \( H_t : \mathcal{H} \rightarrow \mathbb{R}^N, \xi \mapsto (\xi(x_{t1}), \ldots, \xi(x_{tN})) \) for each \( t = 1, \ldots, T \) and the ‘joint’ evaluation operator \( \mathcal{H} : \mathcal{H}^{T+2M} \rightarrow \mathbb{R}^{N \times T}, [\xi_{-M+1}, \ldots, \xi_{T+M}] \mapsto (H_{1T} \xi_1, \ldots, H_{1T} \xi_T) \). Hence the observation scheme (1) becomes \( \mathcal{Y} = \mathcal{H} \mathcal{X} + \mathcal{E} \).

By the symbol \( \Pi(\cdot | \mathcal{Y}) \) we denote the best linear unbiased predictor of the term represented by the dot, given the data \( \mathcal{Y} \). A key observation that simplifies the forecasting is the fact that we may predict the latent functional data first, and then plug them into the filter coefficients. Proposition 1 summarizes this assertion formally.

**Proposition 1.** The best linear unbiased predictor of \( Z_t \) given \( \mathcal{Y} \), denoted as \( \Pi(Z_t | \mathcal{Y}) \), is equivalent to constructing the best linear unbiased predictors of \( X_t \) given \( \mathcal{Y} \), denoted as \( \Pi(X_t | \mathcal{Y}) \), for all \( t \in \mathbb{Z} \) and then applying the filter coefficients \( \{ B_k \}_{k \in \mathbb{Z}} \) to these predictions:

\[
\Pi(X_t | \mathcal{Y}) = \sum_{k \in \mathbb{Z}} B_k \Pi(X_{t-k} | \mathcal{Y}) , \quad s \in \mathbb{Z} .
\]

In the following paragraph we explain how to construct the predictors \( \Pi(X_t | \mathcal{Y}) \), \( t = -M + 1, \ldots, T + M \), following Rubín and Panaretos (2020). First, we comment on the distributional properties of the above defined terms. The random element \( \mathcal{X} \) inherits the second-order structure of \( \{ X_t \} \), thus \( \mathbb{E} \mathcal{X} = 0 \) and

\[
\text{Var}(\mathcal{X}) \equiv \mathsf{S} = \begin{bmatrix}
\mathcal{R}_0^X & (\mathcal{R}_1^X)^* & \cdots & (\mathcal{R}_{T+2M-2}^X)^* \\
\mathcal{R}_1^X & \mathcal{R}_0^X & \cdots & (\mathcal{R}_{T+2M-1}^X)^* \\
\vdots & \vdots & \ddots & \vdots \\
\mathcal{R}_{T+2M-2}^X & \mathcal{R}_{T+2M-3}^X & \cdots & \mathcal{R}_0^X \\
\mathcal{R}_{T+2M-1}^X & \mathcal{R}_{T+2M-2}^X & \cdots & \mathcal{R}_0^X
\end{bmatrix} (20)
\]

understood as a linear operator acting on \( \mathcal{H}^{T+2M} \) and where * denotes the adjoint operator. Moreover, due to independence of the measurement errors, \( \text{Var}(\mathcal{E}) = \sigma^2 \mathcal{I}_{N^T \times N^T} \) where \( \mathcal{I}_{N^T} \) is the identity matrix of size \( N^T \times N^T \).

The best linear unbiased predictor of \( \mathcal{X} \) given \( \mathcal{Y} \) is then given by

\[
\Pi(\mathcal{X} | \mathcal{Y}) = \mathcal{H}^+ \left( \mathcal{H} \mathcal{S} \mathcal{H}^+ + \sigma^2 \mathcal{I}_{N^T} \right)^{-1} \mathcal{Y} \in \mathcal{H}^{T+2M} , (21)
\]

where the term \( \mathcal{H} \mathcal{S} \mathcal{H}^+ + \sigma^2 \mathcal{I}_{N^T} \) is a matrix of size \( N^T \times N^T \) and its inverse is well conditioned since \( \sigma^2 > 0 \). The best linear predictor of each functional datum \( X_t \), \( t = -M + 1, \ldots, T + M \) given the observed data \( \mathcal{Y} \), denoted as \( \Pi(X_t | \mathcal{Y}) \), is then given by the projection \( \Pi(X_t | \mathcal{Y}) = P_t \Pi(X_t | \mathcal{Y}) \) where \( P_t : \mathcal{H}^{T+2M} \rightarrow \mathcal{H}, [\xi_{-M+1}, \ldots, \xi_{T+M}] \mapsto \xi_t \) is the projection operator for \( t = -M + 1, \ldots, T + M \).

The predictor (21) requires the knowledge of the unknown dynamics of the regressor time series through the autocovariance operators (20) as well as the measurement error variance \( \sigma^2 \). Instead of these parameters we plug-in the estimated counterparts (13) and (10) and denote these estimated predictors as \( \hat{\Pi}(\cdot | \mathcal{Y}) \).

In summary, the forecasting algorithm consists of the following steps:

1. From the measurements \( \mathcal{Y} \) realized on the regressor time series \( \{ X_t \} \) estimate the spectral density \( \{ \hat{\mathcal{R}}_t^X \}_{t \in [-T, T]} \) and the measurement error variance \( \sigma^2 \). Using the formula (13), integrate the estimated spectral density to obtain the complete space time covariance \( \{ \hat{\mathcal{R}}_k^X \}_{k \in \mathbb{Z}} \) of the regressor time series \( \{ X_t \} \).
2. From the measurements \( \mathcal{Y} \) and the observed response times series \( Z_1, \ldots, Z_T \), estimate the cross-spectral density \( \{ \hat{\mathcal{R}}_t^Z \}_{t \in [-T, T]} \). Using either truncation regularization (16) or Tikhonov regularization (17), estimate the spectral transfer function \( \{ \hat{\mathcal{R}}_r \}_{r \in [-T, T]} \). By means of formula (18) or (19), depending on the regularization scheme, integrate the spectral transfer function to obtain the filter coefficients \( \{ \hat{B}_k^{\text{trunc}} \}_{k \in \mathbb{Z}} \) or \( \{ \hat{B}_k^{\text{bayes}} \}_{k \in \mathbb{Z}} \).
3. Choose $M$ such that the estimated filter coefficients $\hat{B}^{\text{trunc}}$ (or $\tilde{B}_k^{(d)}$) are negligible for $|k| > M$. Using the methodology explained at the beginning of this section, construct the prediction of the latent functional data $\hat{f}(X_{-M+1} | \mathcal{V}), \ldots, \hat{f}(X_{T+M} | \mathcal{V})$.

4. For each $s = S + 1, \ldots, T$, construct the forecast $\hat{f}(Z_s | \mathcal{V}) = \sum_{k=-M}^{M} \hat{B}_k^{\text{trunc}} \hat{f}(X_{s-k} | \mathcal{V})$, in the case of spectral truncation or $\hat{f}(Z_s | \mathcal{V}) = \sum_{k=-M}^{M} \tilde{R}_k^{(d)} \hat{f}(X_{s-k} | \mathcal{V})$, in the case of Tikhonov regularization.

### 3. ASYMPTOTIC RESULTS

We turn to establishing consistency of the proposed estimators of the spectral density (11) and the cross-spectral density (14), and the regularized estimators of the filter coefficients (18) and (19). A key requirement for consistency is to control the temporal dependence of $X_t$. Here we employ cumulant mixing conditions which have been successfully used in functional time series before (Panaretos and Tavakoli, 2013; Rubín and Panaretos, 2020). Alternatives to the cumulant assumptions are $L^p$-$m$-approximability (Hörmann and Kokoszka, 2010; Hörmann et al., 2015a,b) and strong mixing conditions (Rubín and Panaretos, 2020).

Bellow we list the assumptions that we will make use of to study the asymptotics of the spectral density estimator $\{f^X_0 \}_{t \in [-x,x]}$ and the cross-spectral density estimator $\{f^Z_0 \}_{t \in [-x,x]}$.

- **(B1)** The number of measurements $N_t$ at time $t$ are independent random variables, identically distributed with a random variable $N$ satisfying $N \geq 0$, $\mathbb{E}[N] < \infty$ and $\Pr(N > 1) > 0$.

- **(B2)** The measurement locations $x_j, j = 1, \ldots, N_t, t = 1, \ldots, T$ are independent random variables generated from the density $g(\cdot)$ and are independent of the number of measurements $\{N_t\}_{t = 1, \ldots, T}$. The density $g(\cdot)$ is assumed to be twice continuously differentiable and strictly positive on $[0,1]$.

- **(B3)** The autocovariance kernels, $R_k(\cdot, \cdot)$, are twice continuously differentiable on $[0,1]^2$ for each $h \in \mathbb{Z}$.

Moreover, the terms

$$
\sup_{y,x \in [0,1]} \left| \frac{\partial^2}{\partial y \partial x} R^X_h(y,x) \right|
$$

are uniformly bounded in $h$ for all combinations of $\alpha_1, \alpha_2 \in \mathbb{N}_0$ satisfying $\alpha_1 + \alpha_2 = 2$.

- **(B4)** The fourth order cumulant kernel of $\{X_t\}$ is summable in the supremum norm

$$
\sum_{h_1, h_2, h_3, h_4 = -\infty}^{\infty} \sup_{x_1, x_2, x_3, x_4 \in [0,1]} \left| \text{cum}(X_{h_1}, X_{h_2}, X_{h_3}, X_{h_4})(x_1, x_2, x_3, x_4) \right| < \infty.
$$

where

$$
\text{cum}(X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4})(x_1, x_2, x_3, x_4) = \mathbb{E} \left[ X_{t_1}(x_1) X_{t_2}(x_2) X_{t_3}(x_3) X_{t_4}(x_4) \right] - \mathbb{E} \left[ X_{t_1}(x_1) X_{t_3}(x_3) \right] \mathbb{E} \left[ X_{t_2}(x_2) X_{t_4}(x_4) \right] - \mathbb{E} \left[ X_{t_1}(x_1) X_{t_2}(x_2) \right] \mathbb{E} \left[ X_{t_3}(x_3) X_{t_4}(x_4) \right] - \mathbb{E} \left[ X_{t_1}(x_1) \right] \mathbb{E} \left[ X_{t_2}(x_2) X_{t_3}(x_3) X_{t_4}(x_4) \right] - \mathbb{E} \left[ X_{t_2}(x_2) \right] \mathbb{E} \left[ X_{t_1}(x_1) X_{t_3}(x_3) X_{t_4}(x_4) \right] - \mathbb{E} \left[ X_{t_3}(x_3) \right] \mathbb{E} \left[ X_{t_1}(x_1) X_{t_2}(x_2) X_{t_4}(x_4) \right] - \mathbb{E} \left[ X_{t_4}(x_4) \right] \mathbb{E} \left[ X_{t_1}(x_1) X_{t_2}(x_2) X_{t_3}(x_3) \right]
$$

for $t_1, t_2, t_3, t_4 \in \mathbb{N}$ and $x_1, x_2, x_3, x_4 \in [0,1]$.

- **(B5)** The sequence $\{R^X_h \}_{h \in \mathbb{Z}}$ satisfies the weak dependence condition

$$
\sum_{h=-\infty}^{\infty} \left| h \right| \sup_{x,y \in [0,1]} \left| R^X_h(x,y) \right| < \infty.
$$

Finally, we state assumptions on the bandwidth parameters $B_h, B_y, B_C$ and the Bartlett span $L$:
(B6) \[ B_R \to 0, \quad TB_R^6 \to \infty, \]

(B7) \[ B_V \to 0, \quad TB_V^4 \to \infty, \]

(B8) \[ B_C \to 0, \quad TB_C^4 \to \infty, \]

(B9) \[ L \to \infty, \quad L = o(\sqrt{T B_R^2}), \quad L = o(\sqrt{T B_C}). \]

Under the assumptions (A1)–(A3), (B1)–(B6), (B9) the spectral density estimator \( \hat{f}_X \omega \) is consistent uniformly in the supremum norm (Rubín and Panaretos, 2020). Moreover, an upper bound for the convergence rate was established. Assuming further (B7) ensures the consistency of the measurement error variance estimator \( \hat{\sigma}^2 \) given by formula (10).

Likewise, the estimator of the cross-spectral density \( \hat{f}_Z \omega \) is consistent by Proposition 2.

**Proposition 2.** Under the conditions (A1)–(A3), (B1)–(B3), (B8), (B9), the cross-spectral density is estimated consistently:

\[
\sup_{\omega \in [-\pi, \pi]} \sup_{x \in [0, 1]} | \hat{f}_Z \omega (x) - f_Z \omega (x) | = o_p(1) \quad \text{as} \quad T \to \infty.
\]

Assuming further the condition (B5), we obtain the convergence rate:

\[
\sup_{\omega \in [-\pi, \pi]} \sup_{x \in [0, 1]} | \hat{f}_Z \omega (x) - f_Z \omega (x) | = O_p \left( L \frac{1}{\sqrt{T B_C}} \right) \quad \text{as} \quad T \to \infty.
\]  

The estimators of the spectral density and the cross-spectral density are essential building blocks for the estimation of the filter coefficients. In the following paragraphs we list the conditions for the consistency of the filter coefficients obtained via Tikhonov regularization (19) and via truncation regularization (18).

First, the following condition is required for the regression model (2) to be identifiable, regardless of the regularization method used.

(C) For all \( \omega \in [-\pi, \pi] \) the operators \( \mathcal{X}_\omega : \mathcal{H} \to \mathcal{H} \) satisfy ker \( \mathcal{X}_\omega \) = 0.

To ensure the consistency of the filter coefficients estimator by the Tikhonov method we only need to guarantee that the regularization parameter vanishes slowly.

(D) The Tikhonov regularization parameter satisfies

\[
\frac{1}{\rho} L \frac{1}{\sqrt{T B_C}} \to 0, \quad \text{as} \quad T \to \infty.
\]

\[
\frac{1}{\rho^2} L \frac{1}{\sqrt{T B_R^2}} \to 0, \quad \text{as} \quad T \to \infty.
\]

Theorem 1 establishes the consistency of the Tikhonov filter coefficient estimators.
Theorem 1. Under the conditions (A1)–(A3), (B1)–(B6), (B8), (B9), (C), (D), the filter coefficient estimators (19) constructed by means of Tikhonov regularization are consistent in the sense that:

\[
\sup_{k \in \mathbb{Z}} \left\| \hat{B}_{T_{k}}^{\text{Tikh}} - B_{k} \right\| = o_{p}(1) \quad \text{as} \quad T \to \infty.
\]

We now turn to the truncation estimator (18) of the filter coefficients, whose consistency requires more technical assumptions. We use the result by Hörmann et al. (2015b, Theorem 1) relies on having consistent estimators of the spectral density and cross-spectral density operators with a known rate of convergence, on a condition on the eigenvalue spacing, and on an assumption that the spectral truncation parameter \(K_{\omega}\) grows sufficiently slowly. In what follows, we review their conditions and adapt them to the setting when the spectral density kernels and the cross-spectral density are estimated by the kernel smoothing methods from sparse noisy observations.

Recall the eigendecomposition of the spectral frequency operator (4) and that its harmonic eigenvalues and harmonic eigenfunction are denoted \(\{\lambda_{\omega}^{m}\}_{m=1}^{\infty}\) and \(\{q_{\omega}^{m}\}_{m=1}^{\infty}\) respectively. Define

\[
\Lambda_{1}^{\omega} = \lambda_{1}^{\omega} - \lambda_{2}^{\omega},
\]

\[
\Lambda_{k}^{\omega} = \min \left\{ \lambda_{k}^{\omega} - \lambda_{k+1}^{\omega}, \lambda_{k-1}^{\omega} - \lambda_{k}^{\omega} \right\}, \quad k \geq 2.
\]

The following condition guarantees that the eigenspaces belonging to each of the eigenvalues \(\{\lambda_{\omega}^{m}\}_{m=1}^{\infty}\) are one-dimensional, hence the eigenfunctions \(\{q_{\omega}^{m}\}_{m=1}^{\infty}\) can be identified (up to multiplication by a complex number with modulus 1).

(E1) For all \(k \geq 1\) we assume \(\inf_{\omega \in [-\pi, \pi]} \Lambda_{k}^{\omega} > 0\).

Furthermore we need to assume that the truncation parameter \(K_{\omega}\) needs to grow sufficiently slowly.

(E2)

\[
K_{\omega} = \min \{K^{(i)}, 1 \leq i \leq 4\},
\]

where

\[
K^{(1)} = \max \left\{ k \geq 1 : \inf_{\omega \in [-\pi, \pi]} \hat{\lambda}_{k} \geq 2LT^{-1/2}B_{R}^{-2} \right\},
\]

\[
K^{(2)} = \max \left\{ k \geq 1 : LT^{-1/2}B_{C}^{-1} \int_{-\pi}^{\pi} W^{K(\omega)}(\omega) \, d\omega \leq 1 \right\},
\]

\[
K^{(3)} = \max \left\{ k \geq 1 : \int_{-\pi}^{\pi} \left( W^{K(\omega)}(\omega) \right)^{2} \, d\omega \leq L^{-1/2}T^{-1/4}B_{R} \right\},
\]

\[
K^{(4)} = \max \left\{ k \geq 1 : \int_{-\pi}^{\pi} \left( W^{K(\omega)}(\omega) \right)^{2} \, d\omega \leq L^{-1/2}T^{-1/4}B_{R} \right\}
\]

and where we further define

\[
W^{K(\omega)}(\omega) = \left( \sum_{m=1}^{k} \frac{1}{\hat{\lambda}_{m}^{\omega}} \right)^{1/2}, \quad W_{\Lambda}(\omega) = \left( \sum_{m=1}^{k} \frac{1}{\hat{\Lambda}_{m}^{\omega}} \right)^{1/2}
\]

and \(\{\hat{\Lambda}_{m}^{\omega}\}\) are the empirical counterparts of \(\{\Lambda_{m}^{\omega}\}\) where the estimates \(\hat{\lambda}_{m}^{\omega}\) are plugged-in.

Under the above stated assumptions, the filter coefficient estimator (18) obtained by means of truncation regularization is consistent.
Theorem 2. Under the conditions (A1)–(A3), (B1)–(B6), (B8), (B9), (C), (E1), (E2), the filter coefficients estimates (18) constructed by the spectral truncation regularization are consistent:

$$\sup_{k \in \mathbb{Z}} \left\| \hat{B}_k^{\text{true}} - B_k \right\| = o_p(1) \quad \text{as} \quad T \to \infty.$$ 

4. NUMERICAL EXPERIMENTS

4.1. Simulation Setting

In this simulation study we assess the performance of the proposed methodology on the basis of two criteria: the estimation error of the filter coefficients estimator (19), and the prediction error of the forecasts of the response process (Section 2.4). We also compare the performance of the two regularization techniques, and corroborate that neither dominates the other. To illustrate this, we introduce two different filter coefficient function for the lagged process (Section 2.4). We also compare the performance of the two regularization techniques, and corroborate that neither dominates the other. To illustrate this, we introduce two different filter coefficient function for the lagged process, see (26) and (27), in which one technique is expected to perform better than the other, and vice versa.

The MATLAB code and the results of the simulation are openly available on GitHub (Rubín and Panaretos, 2019).

We simulate the functional regressor series \( \{X_t\}_{t \in \mathbb{Z}} \) as functional linear processes: either a functional autoregressive process or a functional moving average process. We define \( \{E_t\}_{t \in \mathbb{Z}} \) to be the stochastic innovation term for these linear processes. It is assumed to be a sequence of i.i.d. zero-mean Gaussian random variables in \( H = L^2([0, 1]) \) with the covariance kernel \( K(x, y) \) given by

\[
K(x, y) = \sin(2\pi x) \sin(2\pi y) + 0.6 \cos(2\pi x) \cos(2\pi y) + 0.3 \sin(4\pi x) \sin(4\pi y) + 0.1 \cos(4\pi x) \cos(4\pi y) + 0.1 \sin(6\pi x) \sin(6\pi y) + 0.1 \cos(6\pi x) \cos(6\pi y) + 0.05 \sin(8\pi x) \sin(8\pi y) + 0.05 \cos(8\pi x) \cos(8\pi y) + 0.05 \sin(10\pi x) \sin(10\pi y) + 0.05 \cos(10\pi x) \cos(10\pi y), \quad x, y \in [0, 1].
\]

We simulate realizations of the functional autoregressive process of order 1 and the functional moving average process of order 4 defined in the following two settings:

(FAR(1)) The process \( \{X_t\}_{t \in \mathbb{Z}} \) is a functional autoregressive process of order 1 (Bosq, 2012) defined by the iteration

\[
X_{t+1} = AX_t + E_t, \quad t \in \mathbb{Z}. \tag{24}
\]

The operator \( A \) is assumed to be a Hilbert–Schmidt operator and we define its kernel as \( A(x, y) = \kappa \sin(x - y), x, y \in [0, 1] \), where \( \kappa > 0 \) is chosen such that \( \|A\|_{L^2(H)} = 0.7 \) and \( \|\cdot\|_{L^2(H)} \) is the operator norm in the space of linear operators on \( H \).

(FMA(4)) The process \( \{X_t\}_{t \in \mathbb{Z}} \) is considered to be the functional moving average process of order 4 defined by

\[
X_t = E_t + M_1E_{t-1} + M_2E_{t-2} + M_3E_{t-3} + M_4E_{t-4}, \quad t \in \mathbb{Z}. \tag{25}
\]

The operators \( M_1, \ldots, M_4 \) are assumed to be Hilbert–Schmidt and given by their kernels \( M_1(x, y) = \kappa_1 \sin(x + y), M_2(x, y) = \kappa_2 \sin(1 - x - y), M_3(x, y) = \kappa_3 \sin(1 + x - y), M_4(x, y) = \kappa_4 \sin(2 - x - y), \)
The filter coefficients are stationary and Gaussian (Bosq, 2012).

The proposed methodology requires the selection of the tuning parameters. We implemented the choice of the bandwidths $\sigma_B$ and $B_C$ for the estimation of $(\mathcal{F}_x, \omega)_{\omega \in [-\pi, \pi]}$ and $\sigma^2$ by means of K-fold cross validation, as explained in Section 5 of our previous work (Rubín and Panaretos, 2020).

The functional autoregressive process $(\text{FAR}(1))$, defined uniquely by (24), and the functional moving average process $(\text{FMA}(4))$ are stationary and Gaussian (Bosq, 2012).

4.2. Estimation Procedure

The estimation procedure involves several steps:

1. **Filter Coefficients Selection**: The filter coefficients $b_0, b_1, b_2, b_3, b_4, b_5$ are non-zero but with decaying magnitude. They are set to either $(\beta_A, 0.9\beta_A, 0.7\beta_A, 0.5\beta_A, 0.3\beta_A, 0.1\beta_A)$ or $(\beta_B, 0.9\beta_B, 0.7\beta_B, 0.5\beta_B, 0.3\beta_B, 0.1\beta_B)$ depending on the chosen shape of the filter coefficients.

2. **Measurement Error**

   The variance of the measurement error $\sigma^2$ is chosen so that the signal-to-noise ratio is $\tau^2 = 0.001$.

3. **Approximation Procedure**

   For each combination of the settings, that is, each of the 2 linear processes of $\{X_t\}_{t \in \mathbb{Z}}$, each of 4 length parameters $T \in \{300, 600, 900, 1200\}$, each of 4 sampling density parameters $N_{\max} \in \{10, 20, 40, 60\}$, each of 3 regression schemata, and 2 shapes of the filter coefficients, we run 90 independent runs. The simulations have to be obviously performed in a finite dimension. We approximate the infinite dimensional dynamics of the process $\{X_t\}_{t \in \mathbb{Z}}$ by the B-Spline basis of dimension 21. This follows the same scheme as Rubín and Panaretos (2020), where further details on the definition of the basis as well as the approximation procedure can be found. Moreover, we consider also the regime of complete functional observations in the setting of Hörmann et al. (2015b) in order to compare how much information is lost due to sparse sampling.

4.3. Results

The results are presented in Figure 3 and Table 1. The performance of the proposed methodology is evaluated in terms of the estimation accuracy of the filter coefficients and the measurement error. The methodology achieves accurate estimation for all considered scenarios, with a low bias and variance in the estimated coefficients.

5. Conclusion

The proposed methodology provides a robust framework for the estimation of functional autoregressive and moving average processes. It is particularly suitable for sparse sampling scenarios, where the loss of information is a critical issue. Future work will focus on extending the methodology to more complex settings and incorporating additional functional features.
Figure 2. Top row: The covariance kernel of the stochastic innovation $K_{(23)}$, the autoregressive kernel $A_{(24)}$, and the filter coefficients $\beta_A_{(26)}$ and $\beta_B_{(27)}$ respectively. Bottom row: The kernels of the moving average process $M_1, \ldots, M_4_{(25)}$ [Color figure can be viewed at wileyonlinelibrary.com]

In detail in Rubín and Panaretos (2020). The Bartlett span parameter for the estimation of the spectral density is set to $L = \lfloor \frac{2T}{3} \rfloor$.

In order to select the regularization parameters for either of the two proposed regularization methods we resort to holdout cross-validation. We split the response time series $Z_1, \ldots, Z_T$ into the training set $Z_1, \ldots, Z_S$ and the test set $Z_{S+1}, \ldots, Z_T$. The split is set to be 80:20 in favour of the training set, that is, $S = 0.8T$. The cross-spectral density $\{\mathcal{F}_{ZX}\}_{\omega \in [-\pi, \pi]}$ is estimated from the data in the training set $Z_1, \ldots, Z_S$ and the bandwidth parameter $B_C$ is selected by K-fold cross-validation within the training set.

The truncation estimator (16) is constructed by means of eigenvalue thresholding (Hörmann et al., 2015b). Specifically, we set $K_{0}(v) = \max_{m \geq 1} \{ \hat{\lambda}_m > v \}$ where $v > 0$ is a parameter to be chosen by holdout cross-validation in the following way. Having estimated $\{\mathcal{F}_{X}^{\infty}\}_{\omega \in [-\pi, \pi]}$ and $\sigma^2$ from the sparsely observed regressor time series $\{X_t\}_{t=1}^{T}$ and the cross-spectral density $\{\mathcal{F}_{XZ}\}_{\omega \in [-\pi, \pi]}$ from $\{X_t\}_{t=1}^{T}$ and the training partition of the response $Z_1, \ldots, Z_S$, the spectral transfer function is estimated by the formula (16) using a candidate value of $K_0(v)$. The forecasts $\hat{Z}_{S+1}, \ldots, \hat{Z}_T$ are produced by the methodology outlined in Section 2.4. Comparing the forecasts with the true values of $Z_{S+1}, \ldots, Z_T$ yields a mean square forecast error on the holdout partition which we minimize with respect to $v$. The Tikhonov estimator (17) involves the selection of the parameter $\rho$. In the same way as for the truncation regularization, we chose $\rho$ by holdout cross-validation based on the mean square forecast error on $Z_{S+1}, \ldots, Z_T$.

In the case of complete functional observations we again use holdout cross-validation for the selection of the eigenvalue thresholding parameter as well as the Tikhonov parameter.
10 20 40 60 inf
N_{max}
0 0.2 0.4 0.6 0.8 1
MSE, filters
T=300, shape=A
trunc Tikh
10 20 40 60 inf
N_{max}
0 0.2 0.4 0.6 0.8 1
MSE, filters
T=600, shape=A
trunc Tikh
10 20 40 60 inf
N_{max}
0 0.2 0.4 0.6 0.8 1
MSE, filters
T=900, shape=A
trunc Tikh
10 20 40 60 inf
N_{max}
0 0.2 0.4 0.6 0.8 1
MSE, filters
T=1200, shape=A
trunc Tikh

Figure 3. The median mean square error $\delta^B$ of the filter coefficient estimates (28) for the truncation regularization ('trunc') and Tikhonov regularization method ('Tikh'), displayed as a function of the time series length $T \in \{300, 600, 900, 1200\}$ and the maximum number of the observation locations $N_{max} \in \{10, 20, 40, 60, \text{inf}\}$ where 'inf' stands for the fully observed functional data. The top and the bottom row show the results for the filter coefficients of the shapes (26) and (27) respectively.

4.3. Evaluation Criteria

We assess the estimation error of the filter coefficients by the mean square error criterion:

$$\delta^B = \sum_{k \in \mathbb{Z}} \left\| \hat{B}_k - B_k \right\|^2.$$  (28)

Next we want to assess the forecasting performance of the proposed methodology. Because the entire sample was used for fitting the model dynamics, we simulate an independent copy of the regressor time series, denoted as $\{X^{copy}_t\}_{t=1}^T$, and the response process $\{Z^{copy}_t\}_{t=1}^T$. Using the estimates of the model dynamics from the original data, we produce the predictions $\{\hat{Z}^{copy}_t\}_{t=1}^T$ from the sparsely observed $\{X^{copy}_t\}_{t=1}^T$ and compare with the true values $\{Z^{copy}_t\}_{t=1}^T$. The prediction relative mean square error is then defined

$$\delta^{pred} = \frac{1}{T} \sum_{t=1}^T \frac{(Z^{copy}_t - Z^{copy}_t)^2}{\text{Var}(Z_0)}.$$  (29)

Moreover, we include the prediction error of the oracle estimator that assumes that both the dynamics of the regressor time series $\{X_t\}_{t \in \mathbb{Z}}$ and the filter coefficients $\{B_k\}_{k \in \mathbb{Z}}$ are known. The oracle estimator completes the steps 3 and 4 of the algorithm of Section 2.4 where the estimates $\{\hat{R}^i_h(\cdot,\cdot)\}_{i \in \mathbb{Z}}$ and $\{\hat{B}_k^{\text{trunc}}\}_{k \in \mathbb{Z}}$ (or $\{\hat{B}_k^{\text{Tikh}}\}_{k \in \mathbb{Z}}$) are replaced by the true values of $\{R^i_h(\cdot,\cdot)\}_{i \in \mathbb{Z}}$ and $\{B_k\}_{k \in \mathbb{Z}}$.

4.4. Results of Numerical Experiments

Due to the large number of simulation settings considered, we display the results in an aggregated form. Figures 3 and 4 present the results for the filter coefficient estimation and the prediction performance as a function of the
Figure 4. The median mean square prediction error $\delta_{\text{pred}}$ (29) for the truncation regularization (‘trunc’), Tikhonov regularization method (‘Tikh’) and the oracle estimator, displayed as a function of the time series length $T \in \{300, 600, 900, 1200\}$ and the maximum number of the observation locations $N_{\text{max}} \in \{10, 20, 40, 60, \text{inf}\}$ where ‘inf’ stands for the fully observed functional data. The top and the bottom row show the results for the filter coefficients of the shapes (26) and (27) respectively.

Figure 5. The median mean square error $\delta_{\text{filt}}$ of the filter coefficient estimates (28) for the truncation regularization (‘trunc’) and Tikhonov regularization method (‘Tikh’) with respect to the simulated dynamics of $\{X_t\}$ and the regression scheme. The results are aggregated over all sparse observation setups $T \in \{300, 600, 900, 1200\}$ and $N_{\text{max}} \in \{10, 20, 40, 60\}$. The left and the right figures show the results for the filter coefficients of the shapes (26) and (27) respectively.

An inspection of Figures 3 and 4 reveals that there is no clear winner between the truncation and the Tikhonov methods. The numeric experiments with the shape (A) defined by (26) show that the Tikhonov method dominates sample size parameters $T$ and $N_{\text{max}}$. The results are aggregated over both types of simulated dynamics of $\{X_t\}$, that is, the functional autoregressive process (FAR(1)) and the functional moving average process (FMA(4)), and over all three regression schemes (reg1), (reg2), and (reg3). Figures 5 and 6, on the other hand, present the results for different the simulated dynamics of the process and the different considered regression schemes separately, aggregated over all time series length parameters $T \in \{300, 600, 900, 1200\}$ and all sparse observation regimes $N_{\text{max}} \in \{10, 20, 40, 60\}$.

An inspection of Figures 3 and 4 reveals that there is no clear winner between the truncation and the Tikhonov methods. The numerical experiments with the shape (A) defined by (26) show that the Tikhonov method dominates.
the truncation method in all considered settings for the estimation of the filter coefficients. The simulations with the shape (B) defined by (27) yield the opposite behaviour: the truncation regularization prevails. We attribute this dichotomy to the following reasons:

- The shape (B) corresponds to the leading eigenfunction of the covariance kernel (23). Even though the functional regression is performed in the spectral domain, the spectral transfer function $\mathcal{B}_\omega$ is still well aligned with the first eigenfunction of $\mathcal{F}_\omega^A$ and therefore is enough to cutoff after the first eigenvalue, thus favoring truncation regularization. See Figure 7 to visualize the alignment in the spectral domain.

- The shape (A) corresponds to the fourth eigenfunction of the covariance kernel (23). Moreover the fourth eigenvalue is tied with the fifth and the sixth one. This structure is preserved also in the spectral domain, c.f. Figure 7. Since the Tikhonov regularization does not discard the eigenspace corresponding to any of these eigenvalues, it achieves lower estimation error in this non-aligned case. Moreover, the Tikhonov regularization enjoys the advantage of being stable to spectral eigenvalue ties (Hall and Horowitz, 2007; Pham and Panaretos, 2018).

- Note that the shape (A) is generally more difficult to estimate than shape (B). This is not surprising because the signal in the fourth eigenfunction is much weaker than in the first one.

There is no notable difference in the difficulty of estimation between the two simulated dynamics of the regressor time series $\{X_t\}$, that is, the functional autoregressive process (FAR(1)) and the functional moving average process (FMA(4)). The considered regression schemes do not reveal any surprises: the longer the lagged dependence is, the more difficult estimation becomes. Therefore the (reg3) scheme produces the largest errors while (reg1) is the lowest.

The prediction error of the response process $\delta_{\text{pred}}$, which is presented in Figures 4 and 6, follows the same conclusions as the estimation of the filter coefficients. The shape (A) is more challenging to predict and the Tikhonov regularization is seen to be preferable not only for estimation, but for prediction too. The shape (B) is easier to predict using the truncation regularization. The predictions by either of the two techniques, feature twice to thrice greater prediction error $\delta_{\text{pred}}$ than the oracle estimator, that is, the prediction assuming the model of the data to be known and the uncertainty coming only from sparse noisy sampling regime.

5. DATA ANALYSIS

We illustrate the proposed methodology on measurements recorded (Tammet, 2009) at the scientific observatory located at mount Wank located in southern Germany. We remark that this analysis should be seen primarily as
an example of the type of data that fall in our framework, rather than a complete data analysis, since there are presumably further important covariates that otherwise should be included. The considered period is January 1, 1977–December 31, 1979 consisting of $T = 1095$ days. In particular, we analyse the interdependence of three time series:

* Atmospheric electricity. The ionization processes in the atmosphere cause the air to be conductive and the conductivity can be measured in terms of electric potential difference per distance, expressed in volts per meter (V/m). The atmospheric electricity is an important indicator for climate research (Tammet, 2009) and air pollution (Israelsson and Tammet, 2001). However, the atmospheric electricity can be reliably measured only under fair-weather conditions, otherwise the atmospheric ionization processes are changed and a different quantity is recorded. The standard meteorological methodologies (Xu et al., 2013; Israelsson and Tammet, 2001) suggest to discard the observations under unfair conditions and analyse only those observations recorded under fair weather. Given these guidelines, we take into account only those hourly observations of atmospheric electricity where the wind speed was below 20 km/hour and the atmospheric electricity $E$ itself satisfies $0 < E < 250$ V/m. The meteorological community also advocate discarding the data based on cloud coverage (Xu et al., 2013; Israelsson and Tammet, 2001), but, unfortunately, the mount Wank dataset (Tammet, 2009) does not contain cloud coverage information.

Based on the above criteria, we eventually retain an unevenly sampled scalar time series which we consequently decatenate into individual days. This technique is useful (Aue et al., 2015; Hörmann et al., 2015a, 2018; Hörmann and Kokoszka, 2010) in separating the intra-day variability and the temporal dependence across days. Thus we arrive at a sparsely observed functional time series, where the latent functional data are interpreted as the ‘atmospheric electricity had the weather been fair’.

In what follows, we denote the fair weather electricity time series as $\{X_t^{(E)}\}_{t=1}^T$. The time series features total of 18,326 measurements or 16.7 measurements per day on average.
• **Temperature.** The temperature was recorded hourly at Wank over the considered period. First, we remove the yearly periodicity in the data, then divide the time domain into individual days, and finally convert the hourly observations into functional data using B-splines. The produced fully observed functional time series is denoted as \( \{ X_t^{(E)} \}_{t=1}^T \). The time series includes 21 missing days which we treat as missing completely at random.

• **Recorded visibility.** The reported visibility was recorded hourly at a range of locations. We define the response time series \( \{ Z_t \}_{t=1}^T \) as the average visibility on the given day. The time series includes 42 missing values which we treat as missing completely at random.

Since the goal of our analysis is to illustrate the lagged regression methodology and compare the Tikhonov and the truncation regularization, we split the response time series into two parts, the training component \( Z_1, \ldots, Z_{822} \) and the test component \( Z_{823}, \ldots, Z_{1095} \) consisting of roughly 75% and 25% of the observations respectively. We fit three models on the entire time span of \( \{ X_t^{(E)} \}_{t=1}^T \) and/or \( \{ X_t^{(T)} \}_{t=1}^T \) and the training set \( \{ Z_t \}_{t=1}^{822} \):

• **Atmospheric electricity model (E).** In this model we use the sparsely observed functional time series \( \{ X_t^{(E)} \}_{t=1}^T \) as the regressor time series for the response \( \{ Z_t \}_{t=1}^T \) exploiting the methodology outlined in this article.

• **Temperature model (T).** This model handles the fully observed functional time series \( \{ X_t^{(T)} \}_{t=1}^T \) as the regressor for the response \( \{ Z_t \}_{t=1}^T \) using the methodology developed by Hörmann et al. (2015b).

• **Joint model (E+T).** Finally this model includes the information from both the sparsely observed functional time series \( \{ X_t^{(E)} \}_{t=1}^T \) and the fully observed functional time series \( \{ X_t^{(T)} \}_{t=1}^T \) to predict the response process \( \{ Z_t \}_{t=1}^T \) outlined by the equation

\[
Z_t = a + \sum_{k \in \mathbb{Z}} B_k^{(E)} X_{t-k}^{(E)} + \sum_{k \in \mathbb{Z}} B_k^{(T)} X_{t-k}^{(T)} + e_t,
\]

which extends the model (2).

The estimation and the prediction in a model that includes multiple functional time series, some under sparse and others under full observation, is explained in detail in the Supporting information. Figure 8 displays a schematic visualization of the prediction for a specific day.

We estimated the filter coefficients in all three models. Figure 9 displays the estimated filter coefficients in the joint model (E+T). The filter coefficients estimates in the marginal models (E) and (T) (not presented here) are very similar to the corresponding filters estimated in (E+T) thanks to the fact that the time series \( \{ X_t^{(E)} \}_{t=1}^T \) and \( \{ X_t^{(T)} \}_{t=1}^T \) are essentially uncorrelated.

A first look at Figure 9 reveals that truncation-based estimates (depicted via their Riesz-representers) feature more spikes. This, combined with a worse prediction performance commented upon later, suggests that the Tikhonov regularization provides a better fit. Therefore we comment only on the interpretation of the filter coefficients estimated by the Tikhonov regularization. The filter coefficient \( B_0^{(E)} \) is negative, especially in the morning hours, suggesting an obvious interpretation: high atmospheric electricity (which is linked to pollution (Israelsson and Tammet, 2001)) implies a reduction of visibility. The filter coefficients \( B_k^{(E)} \) corresponding to the temperature time series reveal an opposite effect. The filters \( B_0^{(T)} \) and \( B_k^{(T)} \) are positive, therefore high temperatures on the same day and the next day predict a higher visibility today. We recall that our model is not causal, therefore the filter coefficient \( B_k^{(E)} \) indeed predicts an effect backwards in time.

Table I presents the prediction performance of the considered models. We calculate the mean square error (MSE) on the test partition \( \{ Z_t \}_{t=1}^{1095} \) and calculate the \( R^2 \) coefficient of determination. The table reveals that the Tikhonov regularization delivers a better prediction performance for all considered models.

6. **DISCUSSION OF THE CONSIDERED REGULARIZATION TECHNIQUES**

This article presents the methodology for the functional lagged regression problem where the regressor time series is observed sparsely and with noise contamination. We have shown how to estimate the (cross-)spectral density
Figure 8. A schema demonstrating prediction of the response time series at time $t = 848$ in the joint model (E+T). Recall that because the response $\{Z_t\}$ is scalar, the filter coefficients $B^E_k$ and $B^{(\tau)}_k$ are functionals and thus can be viewed as inner products with fixed functions which are visualized here. Top part of the figure: the contribution of the atmospheric electricity, bottom part: the contribution of the temperature.
Figure 9. The estimated filter coefficients for lags $k \in \{-3, -2, -1, 0, 1, 2, 3\}$ for the joint model (E+T). Solid line: the estimates by Tikhonov regularization, dashed line: the estimates by truncation regularization, dotted line: the reference line for zero. Top row: the filter coefficients $B^{(E)}_k$ for the atmospheric electricity, bottom row: the filter coefficients $B^{(T)}_k$ for the temperature time series.

Table I. The mean square prediction error (MSE) and $R^2$ coefficients of determination of each of the model with either truncation or Tikhonov regularization based estimation of the filter coefficients. Both the MSE’s and the $R^2$ coefficients were determined on the test partition of the response $\{Z_t\}_{t=823}^{1095}$ using surface smoothers. The estimation of the spectral transfer function and consequently the filter coefficients using the estimated (cross)-spectral density are ill-posed problems and therefore require regularization. We considered two regularization strategies, namely spectral truncation and Tikhonov regularization, and compared them on a simulation study and the analysis of a data set. In the following we summarize some observations on the differences, strengths and weaknesses of the two approaches.

The simulation study presented in Section 4 illustrates that neither of the two regularization method can dominate the other. In one of the considered scenarios the spectral transfer function $\mathcal{B}_\omega$ is well-aligned with the leading eigenfunction of the spectral density operator $\mathcal{F}_\omega$ thus being estimated better by means of truncation. In the other considered setting, $\mathcal{B}_\omega$ is explained by the fourth, the fifth and the sixth leading eigenvalue of $\mathcal{F}_\omega$, which are moreover nearly tied, resulting in a more challenging estimation task where the stability of Tikhonov regularization to ties leads to better results than truncation.

The data analysis illustration given in Section 5 analysed the dependence of visibility on atmospheric electricity and temperature. The comparative analysis of the two regularization methods revealed that the estimates obtained by the Tikhonov regularization feature better predictive performance. Moreover we found that the filter coefficients...
estimated by the Tikhonov regularization were easier to interpret. At least in this application setting, the spectral transfer function $\mathcal{B}_\omega$ does not seem to be well-aligned with the spectral density operator $\mathcal{F}_{XX}$.

We conclude that both of the regularization techniques should belong to the statistician’s repertoire as neither can dominate the other. However, if we were to choose only one to broadly recommend, this would be the Tikhonov approach, as it seems more robust to ‘spectral misalignment’ and eigenvalue ties, and its theoretical treatment requires fewer assumptions.

ACKNOWLEDGEMENTS
The authors thank the editors and reviewers for their thoughtful and constructive comments.

DATA AVAILABILITY STATEMENT
The MATLAB code and the results that support the findings of this study are openly available on GitHub at DOI: 10.5281/zenodo.3190740, reference Rubín and Panaretos (2019).

SUPPLEMENTARY MATERIAL
Additional supplementary material available on-line include: (i) extensions of the functional lagged regression methodology to the models with a functional response, (ii) extensions to models with multiple functional time series as regressor inputs, and (iii) the proofs of the formal statements presented in the article.

REFERENCES
Aue A, Norinho DD, Hörmann S. 2015. On the prediction of stationary functional time series. Journal of the American Statistical Association 110: 378–392.
Bosq D. 2012. Linear Processes in Function Spaces: Theory and Applications, Vol. 149: Springer Science & Business Media.
Brillinger DR. 1981. Time Series: Data Analysis and Theory, Vol. 36: SIAM.
Fan J, Gijbels I. 1996. Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability, Vol. 66: CRC Press.
Hall P, Horowitz JL. 2007. Methodology and convergence rates for functional linear regression. The Annals of Statistics 35: 70–91.
Hall P, Müller H-G, Wang J-L. 2006. Properties of principal component methods for functional and longitudinal data analysis. The Annals of Statistics: 1493–1517.
Hays S., Shen H., Huang J. Z. 2012. Functional dynamic factor models with application to yield curve forecasting. The Annals of Applied Statistics 6: 870–894.
Hörmann S, Kokoszka P. 2010. Weakly dependent functional data. The Annals of Statistics 38: 1845–1884.
Hörmann S, Kidziński L, Hallin M. 2015a. Dynamic functional principal components. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77: 319–348.
Hörmann S, Kidziński L, Kokoszka P. 2015b. Estimation in functional lagged regression. Journal of Time Series Analysis 36: 541–561.
Hörmann S., Kokoszka P., Nisol G. 2018. Testing for periodicity in functional time series, Vol. 46.
Israelsson S, Tammert H. 2001. Variation of fair weather atmospheric electricity at Marsta observatory, Sweden, 1993–1998. Journal of Atmospheric and Solar-Terrestrial Physics 63: 1693–1703.
Klepisch J, Klüppelberg C, Wei T. 2017. Prediction of functional ARMA processes with an application to traffic data. Econometrics and Statistics 1: 128–149.
Kokoszka P, Miao H, Reimherr M, Taoufik B. 2017. Dynamic functional regression with application to the cross-section of returns. Journal of Financial Econometrics 16: 461–485.
Kolmogoroff A. 1941. Interpolation und extrapolation von stationaren zufälligen folgen. Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya 5: 3–14.
Kowal D. R. 2018. Dynamic function-on-scalars regression, arXiv preprint arXiv:1806.01460.
Kowal DR, Matteson DS, Ruppert D. 2017a. A Bayesian multivariate functional dynamic linear model. Journal of the American Statistical Association 112: 733–744.
Kowal DR, Matteson DS, Ruppert D. 2017b. Functional autoregression for sparsely sampled data. Journal of Business & Economic Statistics: 1–13.
Li Y, Hsing T. 2010. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* **38**: 3321–3351.

Müller H-G, Sen R, Stadtmüller U. 2011. Functional data analysis for volatility. *Journal of Econometrics* **165**: 233–245.

Panaretos VM, Tavakoli S. 2013. Fourier analysis of stationary time series in function space. *The Annals of Statistics* **41**: 568–603.

Pham T., Panaretos V. 2018. Methodology and convergence rates for functional time series regression. *Statistica Sinica* **28**: 2521–2539. (Special Issue in Memory of Peter Hall).

Priestley MB. 1981. *Spectral Analysis and Time Series*: Academic Press.

Rubín T, Panaretos VM. 2020. Sparsely observed functional time series: estimation and prediction. *Electronic Journal of Statistics* **14**: 1137–1210.

Rubín T., Panaretos V. M. 2019. Github repository: functional lagged regression with sparse noisy observations.

Sen R, Klüppelberg C. 2019. Time series of functional data with application to yield curves. *Applied Stochastic Models in Business and Industry*: 1–16.

Shumway RH, Stoffer DS. 2000. Time series analysis and its applications. *Studies in Informatics and Control* **9**: 375–376.

Tammet H. 2009. A joint dataset of fair-weather atmospheric electricity. *Atmospheric Research* **91**: 194–200.

Tavakoli S, Panaretos VM. 2016. Detecting and localizing differences in functional time series dynamics: a case study in molecular biophysics. *Journal of the American Statistical Association* **111**: 1020–1035.

Wiener N. 1950. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*: Technology Press.

Xu B, Zou D, Chen BY, Zhang JY, Xu GW. 2013. Periodic variations of atmospheric electric field on fair weather conditions at YBJ, Tibet. *Journal of Atmospheric and Solar-Terrestrial Physics* **97**: 85–90.

Yao F, Müller H-G, Clifford AJ, Duerer SR, Follett J, Lin Y, Buchholz BA, Vogel JS. 2003. Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics* **59**: 676–685.

Yao F, Müller H-G, Wang J-L. 2005a. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**: 577–590.

Yao F, Müller H-G, Wang J-L. 2005b. Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33**: 2873–2903.