Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT
Journal of Multivariate Analysis 87 (2003) 133–158
Journal of Multivariate Analysis
http://www.elsevier.com/locate/jmva

Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes

Anestis Antoniadis[a,*] and Theofanis Sapatinas[b]

[a] *Laboratoire IMAG-LMC, University Joseph Fourier, 51 rue de Mathematiques, BP 53, 38041 Grenoble Cedex 9, France*
[b] *Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, CY 1678 Nicosia, Cyprus*

Received 10 September 2001

Abstract

We consider the prediction problem of a continuous-time stochastic process on an entire time-interval in terms of its recent past. The approach we adopt is based on the notion of autoregressive Hilbert processes that represent a generalization of the classical autoregressive processes to random variables with values in a Hilbert space. A careful analysis reveals, in particular, that this approach is related to the theory of function estimation in linear ill-posed inverse problems. In the deterministic literature, such problems are usually solved by suitable regularization techniques. We describe some recent approaches from the deterministic literature that can be adapted to obtain fast and feasible predictions. For large sample sizes, however, these approaches are not computationally efficient.

With this in mind, we propose three linear wavelet methods to efficiently address the aforementioned prediction problem. We present regularization techniques for the sample paths of the stochastic process and obtain consistency results of the resulting prediction estimators. We illustrate the performance of the proposed methods in finite sample situations by means of a real-life data example which concerns with the prediction of the entire annual cycle of climatological El Niño-Southern Oscillation time series 1 year ahead. We also compare the resulting predictions with those obtained by other methods available in the literature, in particular with a smoothing spline interpolation method and with a SARIMA model.

AMS 2000 subject classifications: 62M10; 62M20; 65F22

Keywords: Autoregressive Hilbert processes; Banach spaces; Besov spaces; Continuous-time prediction; El Niño-Southern Oscillation; Hilbert spaces; Ill-posed inverse problems; SARIMA

*Corresponding author.
E-mail addresses: anestis.antoniadis@imag.fr (A. Antoniadis), t.sapatinas@ucy.ac.cy (T. Sapatinas).

0047-259X/03/$ - see front matter
doi:10.1016/S0047-259X(03)00028-9

models; Singular value decomposition; Sobolev spaces; Smoothing splines; Tikhonov–Phillips regularization; Wavelets

1. Introduction

In many real life situations one seeks information on the evolution of a continuous-time stochastic process $X = (X(t);\ t \in \mathbb{R})$ in the future. Given a trajectory of X observed on the interval $[0, T]$, one would like to predict the behavior of X on the entire interval $[T, T+\delta]$, where $\delta > 0$, rather than at specific time-points. An appropriate approach to this problem is to divide the interval $[0, T]$ into subintervals $[i\delta, (i+1)\delta]$, $i = 0, 1, \dots, n-1$ with $\delta = T/n$, and to consider the process $Z = (Z_n;\ n \in \mathbb{Z})$ defined by

$$Z_n(t) = X(t + n\delta), \quad 0 \leqslant t \leqslant \delta, \ n \in \mathbb{Z}. \tag{1}$$

This representation is especially fruitful if X possesses a seasonal component with period δ. It can be also employed if the data are collected as curves indexed by time-intervals of equal lengths; these intervals may be *adjacent*, *disjoint* or even *overlapping* (see, for example, [39,40]).

To deal with the prediction problem, Bosq [13] introduced and studied the (zero-mean) Hilbert-valued (H-valued) *autoregressive* (*of order* 1) *processes* (ARH(1)) as the natural and infinite-dimensional extension of the classical $\mathbb{R}^d$-valued ($d \geqslant 1$) *autoregressive* (*of order* 1) *processes*. In this line of study, if Z in (1) is an ARH(1) process, the best prediction of Z_{n+1} given its past history $(Z_n, Z_{n-1}, \dots)$ is then obtained by

$$\begin{aligned} \tilde{Z}_{n+1} &= \mathbb{E}(Z_{n+1} \mid Z_n, Z_{n-1}, \dots) \\ &= \rho(Z_n), \quad n \in \mathbb{Z}, \end{aligned}$$

where ρ is a bounded linear operator associated with the ARH(1) process. In many practical situations, however, the stochastic process X is not centered and, therefore, is not weakly stationary. We will assume that its *mean* is a periodic, H-valued, function $a = (a_t;\ t \in \mathbb{R})$ with period δ and, hence, the centered stochastic process $Y = (Y_n = Z_n - a;\ n \in \mathbb{Z})$ is an ARH(1) process, implying that the best predictor of Z_{n+1} given $Z_n, Z_{n-1}, \dots$ is obtained by

$$\begin{aligned} \tilde{Z}_{n+1} &= \mathbb{E}(Z_{n+1} \mid Z_n, Z_{n-1}, \dots) \\ &= a + \rho(Z_n - a), \quad n \in \mathbb{Z}. \end{aligned} \tag{2}$$

If one is able to estimate the (unknown) periodic function a, say by $\hat{a}$, and the 'prediction' operator ρ, say by $\hat{\rho}$, given $Z_0, Z_1, \dots, Z_n$, then a statistical predictor of Z_{n+1} based on (2) is obtained by

$$\hat{\tilde{Z}}_{n+1} = \hat{a} + \hat{\rho}(Z_n - \hat{a}). \tag{3}$$

This is the approach that we consider in the following development. The article is organized as follows. In Section 2, we briefly state some material from ARH(1) processes that we shall need further. We also briefly discuss some existing methods in the literature to predict a continuous-time stochastic process on an entire time-interval in terms of its recent past, based on the notion of ARH(1) processes. Section 3 is devoted to a detailed analysis of this approach revealing, in particular, that it is related to the theory of function estimation in linear ill-posed inverse problems. In the deterministic literature, such problems are usually solved by suitable regularization techniques. We describe some recent approaches from the deterministic literature that can be adapted to obtain fast and feasible predictions. For large sample sizes, however, these approaches are not computationally efficient. With this in mind, in Section 4, we propose three linear wavelet methods to efficiently address the aforementioned prediction problem. We present regularization techniques for the sample paths of the stochastic process and obtain consistency results of the resulting prediction estimators. In Section 5, we illustrate the performance of the proposed methods in finite sample situations by means of a real-life data example which concerns with the prediction of the entire annual cycle of climatological El Niño–Southern Oscillation time series 1 year ahead. We also compare the resulting predictions with those obtained by other methods available in the literature, in particular with a smoothing spline interpolation method and with a SARIMA model. Concluding remarks and hints for possible extensions of the proposed linear wavelet methods are made in Section 6.

2. Preliminary results

In this section, we briefly state some material from ARH(1) processes that we shall need further. We also briefly discuss some existing methods in the literature to predict a continuous-time stochastic process on an entire time-interval in terms of its recent past, based on the notion of ARH(1) processes.

Let H be a (real) separable Hilbert space, endowed with the Hilbert inner product $\langle \cdot, \cdot \rangle$ and the Hilbert norm $|| \cdot ||$. Typically, H is chosen to be the space of squared-integrable functions on the interval $[a,b] \subseteq \mathbb{R}$ (i.e. $H = L^2[a,b]$) or the Sobolev space of s-smooth functions on the interval $[a,b] \subseteq \mathbb{R}$ (i.e. $H = W_2^s[a,b]$ with integer regularity index $s \geqslant 1$).

Let $\xi = (\xi_n;\ n \in \mathbb{Z})$ be a sequence of H-valued random variables defined on the same probability space $(\Omega, \mathscr{F}, \mathscr{P})$. We say that ξ is a (zero mean) ARH(1) process, if

$$\xi_n = \rho(\xi_{n-1}) + \varepsilon_n, \quad n \in \mathbb{Z}, \tag{4}$$

where $\rho : H \mapsto H$ is a bounded linear operator and $\varepsilon = (\varepsilon_n;\ n \in \mathbb{Z})$ is a H-valued strong white noise, i.e. a sequence of independent and identically distributed H-valued random variables such that $\mathbb{E}(\varepsilon_n) = 0$ and $\mathbb{E}(||\varepsilon_n||^2) < \infty$. Under some mild conditions, (4) has a unique solution which is a weakly stationary process with innovation ε (see, for example, [14, Chapter 3]).

Let H^* be the (topological) dual of H, i.e. the space of all bounded linear functionals on H. The covariance structure of ξ is related to two bounded linear operators from H^* to H, namely the *covariance* and *cross-covariance* (*of order* 1) *operators*. Since H^* may be identified with H (by Riesz representation), they are defined respectively by

$$f \in H \mapsto Cf = \mathbb{E}((\xi_0 \otimes \xi_0)(f))$$

and

$$f \in H \mapsto D^* f = \mathbb{E}((\xi_1 \otimes \xi_0)(f)),$$

where the tensor product $u \otimes v$ (for two fixed elements $u, v \in H$) is the bounded linear operator from H to H, defined by

$$x \in H \mapsto (u \otimes v)(x) = \langle u, x \rangle v.$$

If $\mathbb{E}||\xi_0||^2 < \infty$, the operator C is then symmetric, positive, nuclear and, therefore, Hilbert–Schmidt (see, for example, [14, Chapter 1]). The cross-covariance (of order 1) operator D (the adjoint of D^*) defined by

$$f \in H \mapsto Df = \mathbb{E}((\xi_0 \otimes \xi_1)(f))$$

is also nuclear and, therefore, Hilbert–Schmidt (see, for example, [14, Chapter 1]). The operators C, D and D^* can be (unbiasedly) estimated by their empirical counterparts C_n, D_n and D_n^* defined respectively by

$$C_n = \frac{1}{n+1} \sum_{k=0}^{n} \xi_k \otimes \xi_k,$$

$$D_n = \frac{1}{n} \sum_{k=0}^{n-1} \xi_k \otimes \xi_{k+1}$$

and

$$D_n^* = \frac{1}{n} \sum_{k=0}^{n-1} \xi_{k+1} \otimes \xi_k.$$

Since the ranges of C_n, D_n and D_n^* are finite-dimensional, it follows that they are nuclear and, therefore, Hilbert–Schmidt. Moreover, C_n is symmetric and positive. From now on, the eigenvalues of C and C_n will be respectively denoted (in decreasing order) by $\lambda_1 \geqslant \lambda_2 \geqslant \cdots$ and $\lambda_{1,n} \geqslant \lambda_{2,n} \geqslant \cdots$ with corresponding eigenfunctions respectively denoted by $e_1, e_2, \ldots$ and $e_{1,n}, e_{2,n} \ldots$. Consistency results for C_n, D_n, D_n^*, $\lambda_{i,n}$ and $e_{i,n}$ $(i = 1, 2, \ldots)$ can be found in, for example, [14, Chapter 4].

Using (4), it is not difficult to see that we obtain the following relations:

$$D = \rho C, \tag{5}$$

$$D^* = C\rho^*, \tag{6}$$

where ρ^* denotes the adjoint operator of ρ. One could try to estimate the 'prediction' operator ρ by inverting the operator C in (5) and using the empirical estimates C_n and D_n of C and D, respectively. In other words, an estimator of ρ could be based on the relation

$$\rho = DC^{-1}. \tag{7}$$

In a finite-dimensional context, such a relation makes sense, provided the invertibility of C. In an infinite-dimensional context, however, (7) does not make sense because C^{-1} is an unbounded operator. Bosq [13] initiated research in this area and proposed to first project the data over a suitable finite dimensional subspace of H and then defined an appropriate estimator of ρ. Further work in this direction and extensions have been developed by Pumo [38], Mourid [35], Merlevéde [33] and Cardot [17]. In most applied contexts, however, the process Z defined in (1) is only observed at discrete time-points. Therefore, in order to use these results to deal with the prediction problem of a continuous-time stochastic process on an entire time-interval in terms of its recent past, as discussed in Section 1, one should first approximate the sample paths of Z and then derive appropriate estimators of ρ. Under some strict assumptions on the sample paths of Z (i.e. Hölder continuity of order s, $0<s\leqslant 1$), Pumo [38] used a linear interpolation of the discrete time-points and proceeded in approximating C and D by finite-rank estimators. The resulting class of estimators for ρ are called the *linear interpolation estimators*. Alternatively, by assuming that the predictable part of Z belong to a q-dimensional subspace H_q of smooth functions (i.e., the range of ρ is an s-smooth Sobolev space ($W_2^s[0,1]$) with integer regularity index $s\geqslant 1$), Besse and Cardot [11] proposed to *simultaneously* estimate the sample paths of Z and project the data using natural cubic splines (usually called smoothing splines in the nonparametric curve estimation literature). The resulting class of estimators for ρ are called the *smoothing spline interpolation estimators*.

On the other hand, since C is a compact operator, by the *closed graph theorem* (see, for example, [26, Theorem 5.20]) and the fact that the range of D^* is included in the domain of C^{-1}, the adjoint relation to (7) (using (6)) given by

$$\rho^* = C^{-1}D^* \tag{8}$$

is well-defined and bounded on the domain of C^{-1} which is dense in H. By standard results on linear operators (see, for example, [24, Theorem 3, p. 74]), ρ^* given in (8) can be extended by continuity to ρ. Furthermore, we have

$$\rho = \mathrm{Ext}(DC^{-1}) = (C^{-1}D^*)^* = (DC^{-1})^{**}, \tag{9}$$

where Ext denotes the extension to H of a bounded linear operator defined on a dense subspace of H. If one is, therefore, able to estimate ρ^* given in (8), then any theoretical result obtained on the estimate of ρ^* is applicable to the corresponding estimator of ρ. Moreover, for any $z\in H$, ρz can be estimated via estimates of relevant elements in the range of ρ^*, by first decomposing ρz on any basis in H and then using the adjoint property. The above discussion justifies why, in what follows, we merely concentrate on the estimation of ρ^*.

Recently, Mas [32, Chapter 3] considered two classes of estimators for ρ^*, namely the class of *projection estimators* and the class of *resolvent estimators*. The class of *projection estimators* is defined by

$$\rho_n^* = (\Pi^{k_n} C_n \Pi^{k_n})^{-1} D_n^* \Pi^{k_n}. \tag{10}$$

The (random) operator $(\Pi^{k_n} C_n \Pi^{k_n})^{-1}$ is, in fact, defined by inverting the operator $(\Pi^{k_n} C_n \Pi^{k_n})$ and completing it by the null operator on the subspace orthogonal to the space spanned by the first k_n eigenfunctions of C_n. The class of *resolvent estimators* is defined by

$$\rho^*_{n,p} = f_{n,p}(C_n) D^*_n, \tag{11}$$

where

$$f_{n,p}(C_n) = (C_n + b_n I)^{-(p+1)} C^p_n, \quad \text{for } p \geqslant 0, \ b_n > 0, \ n \geqslant 0.$$

The above expressions are defined in terms of powers of compact operators using their spectral measures (see, for example, [26, p. 356]). When $p \geqslant 1$, the resolvent operators $f_{n,p}(C_n)$ are compact. Moreover, contrary to the projection operators discussed above, these operators have a deterministic norm equal to b_n^{-1}. This allows one to control, through appropriate convergence rates towards 0 of the sequence b_n, the consistency properties of the corresponding resolvent estimators.

The classes (10) and (11) of estimators for ρ^* will be discussed further in Section 4 when we will construct two of the proposed linear wavelet methods to efficiently address the prediction problem (2).

3. Continuous-time prediction as a linear ill-posed inverse problem

In this section, we first show that the original estimator of Bosq [13] to predict a continuous-time stochastic process on an entire time-interval, based on the notion of ARH(1) processes, can be seen as a *projection-type estimator*. We then present a detailed analysis of his prediction setup, revealing, in particular, that the resulting projection-type estimator is related to the theory of function estimation in linear ill-posed inverse problems. In the deterministic literature, such problems are usually solved by suitable regularization techniques. We describe some recent approaches from the deterministic literature that can be adapted to obtain fast and feasible solutions to the prediction problem (2). For large sample sizes, however, these approaches are not computationally efficient.

Remark 3.1. Throughout this section, without loss of generality, we consider the prediction problem (2) with $a = 0$. The general case with $a \neq 0$ follows easily along the same lines with a replaced by its unbiased and mean-squared consistent estimator $\hat{a} = \bar{Z}$ (see [14, Theorem 3.7]).

Projection-type estimators. Let z be any generic vector in H. Using (6), one obtains the relation

$$D^* z = C \rho^* z. \tag{12}$$

Noting that C is a symmetric operator, (12) can be written as

$$C D^* z = C^* C \rho^* z. \tag{13}$$

Since we cannot invert C^*C, from standard literature on inverse problems (using the singular value decomposition of C), we have

$$\rho^* z = \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle e_i, D^* z \rangle e_i.$$

Using the m nonzero eigenvalues $\lambda_{i,n}$ and the corresponding eigenfunctions $e_{i,n}$ of C_n, and the estimator D_n^* of D^*, an estimator of $\rho^* z$ is then obtained by

$$\rho_{n,m}^* z = \sum_{i=1}^{m} \frac{1}{\lambda_{i,n}} \langle e_{i,n}, D_n^* z \rangle e_{i,n}. \tag{14}$$

The above expression (14) clearly represents a *projection-type* estimator of $\rho^* z$. Moreover, this estimator is nothing else than the regularized solution of (12) obtained by truncated singular value decomposition (where C and D are replaced by C_n and D_n, respectively) which corresponds to the original estimator obtained by Bosq [13].

The linear ill-posed inverse problem setting. On the other hand, it is easily seen that (12) can also be written as

$$D_n^* z + (D^* - D_n^*) z = C_n \rho^* z + (C - C_n) \rho^* z$$

which is equivalent to

$$\begin{aligned} D_n^* z &= C_n \rho^* z + (C - C_n) \rho^* z - (D^* - D_n^*) z \\ &= C_n \rho^* z + \eta_n. \end{aligned} \tag{15}$$

The above expression (15) clearly represents a linear *ill-posed* inverse problem that could be solved to estimate $\rho^* z$. The ill-posedness is caused by asymptotic instability due to lack of boundness of C^{-1}.

In the deterministic literature, such problems are usually solved by suitable regularization techniques. However, to adapt this approach in our stochastic setting in order to obtain an estimator of $\rho^* z$, one needs to control the error term $\eta_n = (C - C_n)\rho^* z - (D^* - D_n^*) z$ in (15). We first note that η_n is random because C_n and D_n are random. We are going to control the terms $(C - C_n)$ and $(D^* - D_n^*)$. It is easily seen that

$$\begin{aligned} ||\eta_n||^2 &\leqslant \left(||(C - C_n)\rho^* z|| + ||(D^* - D_n^*) z||\right)^2 \\ &\leqslant 2(||C - C_n||_{\mathscr{L}}^2 ||\rho^* z||^2 + ||D^* - D_n^*||_{\mathscr{L}}^2 ||z||^2), \end{aligned}$$

where $||\cdot||_{\mathscr{L}}$ stands for the supremum norm for bound linear operators from H to H.

Using the fact that the Hilbert–Schmidt norm is finer than the supremum norm and taking into account the asymptotic rates obtained by Bosq [13, Propositions 2.1 and 2.2], we then have

$$\mathbb{E}||\eta_n||^2 \leqslant \frac{2\kappa}{n}(||\rho^* z||^2 + ||z||^2),$$

where κ is a generic positive constant, leading to

$$E||\eta_n||^2 = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty.$$

The above expression ensures that the error term η_n in (15) is controlled. Furthermore, if in the definition of η_n we use the H-valued process Z instead of the generic vector $z \in H$, nothing changes since $\mathbb{E}||\rho^* Z||^2 < \infty$ and $\mathbb{E}||Z||^2 < \infty$.

We now describe some recent approaches from the deterministic literature that address suitable regularization algorithms which can be adapted to solve (15). Our analysis, enables regularization procedures to be 'lifted' from the deterministic literature, reducing thus to a minimum the considerable effort that it would be required to extend these methods to the *stochastic* setting.

3.1. Mair and Ruymgaart [29]

The preconditioning step in (13), for the original equation (12), turns out to be extremely expedient. The operator C^*C is Hermitian and strictly positive and, therefore, easier to deal with than C. A version of the spectral theorem due to Halmos [23] ensures that C^*C is unitary, equivalent to multiplication in a Hilbert space. More precisely, using the singular value decomposition for C, C^*C can be expressed as

$$C^*C = U^{-1} M_{h^2} U,$$

where U is a unitary operator $U : H \mapsto L^2(\mu)$, μ a σ-finite measure, $h \in L^\infty(\mu)$ a strictly positive function and M_{h^2} means multiplication by h^2. It is, therefore, more convenient to work in the *spectral domain*, $L^2(\mu)$, than in the *original domain*, H.

The inverse $(C^*C)^{-1}$ can be now be represented by multiplication by $1/h^2$, where defined. By regularization, we can replace $1/h^2$, which may be unbounded, with a function close to it but still bounded. Certain types of regularized estimators in Hilbert (particularly Sobolev) spaces, with applications in deconvolution and indirect nonparametric regression, have been considered by Mair and Ruymgaart [29]. It is shown that the resulting estimators attain the asymptotic *minimax* rate.

3.2. Plato and Vainikko [37], Maass and Rieder [28] and Rieder [41]

If we consider a Tikhonov–Phillips regularization method then, because the operator C^{-1} is unbounded, (12) is replaced by the following variational problem:

$$\min_{\rho^* z} \{||C\rho^* z - D^* z||^2 + \lambda||\rho^* z||^2;\ \rho^* z \in H\},$$

where $\lambda > 0$ is the *regularization parameter*. Theoretically, the minimum is given by the *normal equations*

$$(C^*C + \lambda I)(\rho^* z)_\lambda = C^* D^* z. \tag{16}$$

Remark 3.2. Note that the above solution (16), when C and D are replaced by C_n and D_n, respectively, is a special case of the resolvent estimator (11) (for $p = 0$) obtained by Mas [32, Chapter 3].

For a computable approximation one has to project the normal equations given in (16) to a finite-dimensional subspace of H. In other words, we can replace (16) by

$$(C_l^* C_l + \lambda I)(\rho^* z)_l = C_l^* D^* z, \tag{17}$$

where $C_l = CP_l$, $P_l : H \mapsto V_l$, V_l is a finite-dimensional subspace of H with $V_l \subset V_{l+1}$, $\overline{\bigcup_l V_l} = H$. With these properties, the solution of (17) converges to the minimal solution of (16) as $l \to \infty$, provided λ is chosen appropriately (see [37]).

The original problem (16), or its computational approximation (17), may be efficiently solved via a multilevel solver. The general idea of multilevel methods is to approximate the original problem by a sequence of related auxiliary problems at smaller scales which can be solved very cheaply and efficiently. Maass and Rieder [28] and Rieder [41] constructed a numerical algorithm for implementing such multilevel iterations for Tikhonov–Phillips regularization by employing wavelet techniques. More specifically, let the sequences $\{V_j;\ j \in \mathbb{Z}\}$ and $\{W_j;\ j \in \mathbb{Z}\}$ be respectively the *approximated* and the *detailed* spaces associated with a multi-resolution analysis of H (see [30]). By writing

$$V_l = V_{l_{\min}} \oplus \bigoplus_{j=l_{\min}}^{l-1} W_j, \quad l_{\min} \leqslant l-1$$

and denoting with Q_j the orthogonal projection of H onto W_j, everything will now depend (speed to the right solution) on the decay rate of the quantity

$$\gamma_l = ||C^*C - C_l^* C_l|| = ||C^* C(I - P_l)|| \quad \text{as } l \to \infty.$$

Using Lemma 1 of [28] and the fact that the operator C^*C is compact, we have

$$||C^* C Q_l|| \leqslant \gamma_l \to 0 \quad \text{as } l \to \infty.$$

Under these conditions, the solution may be obtained by applying an additive Schwarz relaxation iterative solver, similar to the one described in [28]. In practice, however, the operators C and D in the above expressions are unknown, and one could replace them respectively by C_n and D_n in the above computations. Provided n is large enough, and since C_n and D_n converge to C and D respectively (see, for example, [13, Propositions 2.1 and 2.2]), the convergence analysis of the numerical iterative solver of [28] carries over when C_n is replaced by $C_{n,l} = C_n P_l$.

Moreover, instead of replacing C_n by $C_{n,l} = C_n P_l$, one could also approximate C_n by any approximation operator $C_{n,h}$ such that the error of approximation is controlled. A particular way to do that is to compute the (truncated) singular value decomposition of C_n and to take $C_{n,h}$ as

$$C_{n,h}(\cdot) = \sum_{k \leqslant m(h)} \lambda_k \langle \cdot, e_k \rangle e_k,$$

where $|\lambda_k| \leqslant h$ for all $k \geqslant m(h)$. Replacing the normal equations (16) by the approximating version

$$(C_{n,h}^* C_{n,h} + \lambda I)(\rho^* z)_h = C_{n,h}^* D_n^* z, \tag{18}$$

the regularized solution of (18) is then explicitly given by

$$(\rho_n^* z)_{\lambda,h} = \sum_{k \leqslant m(h)} \frac{\lambda_k}{\lambda_k^2 + \lambda} \langle e_k, D_n^* z \rangle e_k. \tag{19}$$

Remark 3.3. (i) By taking $\lambda = 0$ in (19), one obtains a prediction estimator similar to the projection-type estimator (14) that appeared as our interpretation of the prediction estimator obtained by Bosq [13]. If instead of using the weights $\lambda_k/(\lambda_k^2 + \lambda)$ we consider the weights $\lambda_k^p/(\lambda_k^2 + \lambda)^{p+1}$ for some $p \geqslant 0$, and take $h = 0$, one then obtains the resolvent estimator (11) obtained by Mas [32, Chapter 3].

(ii) Similar results to the ones presented above can also be obtained by a more general approximation to the normal equations given in (16). In other words, one can consider the generalized Tikhonov method (see, for example, [37]) which consists in finding the solution of

$$((C^* C)^{q+1} + \lambda I)(\rho^* z)_\lambda = (C^* C)^q C^* D^* z, \tag{20}$$

where $q \geqslant -1/2$.

The above discussion, and the analysis of Bosq [13] and Mas [32, Chapter 3] approaches, allows one to numerically obtain fast and feasible solutions to the prediction problem (2) using some ad hoc wavelet techniques. However, if n is large, the resulting systems are then too large to be efficiently implemented. With this in mind, we propose in the following section some specific-designed linear wavelet methods that are well-suited and theoretically sound for continuous-time prediction.

4. Linear wavelet methods for continuous-time prediction

In this section, we propose three linear wavelet methods to efficiently address the prediction problem (2). We present regularization techniques for the sample paths of the stochastic process and obtain consistency results of the resulting estimators. It is worth pointing out that the proposed linear wavelet methods allow one to obtain asymptotic rates under much weaker assumptions on the sample paths than second order differentiability. We shall assume that the sample paths of Z belongs to the Sobolev space $H = W_2^s[0,1]$ with noninteger regularity index $s > 1/2$. This space consists of relatively smooth functions, but not as smooth as the usual Sobolev space used in *smoothing spline* approaches, i.e. $H = W_2^s[0,1]$ with integer regularity index $s \geqslant 1$ (see, for example, [8]).

Hereafter, we assume that we are working within an orthonormal basis generated by dilations and translations of a compactly supported scaling function, ϕ, and a compactly supported mother wavelet, ψ, associated with an r-regular ($r \geqslant 0$) multiresolution analysis of $L^2[0,1]$ (see, for example, [18, Chapter 5]). For simplicity in exposition, we work with periodic wavelet bases on $[0,1]$ (see, for example, [31, Section 7.5.1]), letting

$$\phi^{\mathrm{p}}_{jk}(t) = \sum_{l \in \mathbb{Z}} \phi_{jk}(t-l) \quad \text{and} \quad \psi^{\mathrm{p}}_{jk}(t) = \sum_{l \in \mathbb{Z}} \psi_{jk}(t-l), \quad \text{for } t \in [0,1],$$

where

$$\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k), \quad \psi_{jk}(t) = 2^{j/2}\psi(2^j t - k).$$

For any $j_0 \geqslant 0$, the collection

$$\{\phi^{\mathrm{p}}_{j_0 k},\ k = 0,1,\dots,2^{j_0}-1;\ \psi^{\mathrm{p}}_{jk},\ j \geqslant j_0 \geqslant 0,\ k = 0,1,\dots,2^j - 1\}$$

is then an orthonormal basis of $L^2[0,1]$. The superscript "p" will be suppressed from the notation for convenience. Despite the poor behavior of periodic wavelets near the boundaries, where they create high amplitude wavelet coefficients, they are commonly used because the numerical implementation is particular simple.

For detailed expositions of the mathematical aspects of wavelets we refer, for example, to [18,31,34]. For comprehensive expositions and reviews on wavelets in statistical settings we refer, for example, to [1,5,6,43].

4.1. Regularized wavelet-vaguelette estimators

It is easily seen that the best predictor of Y_1 given Y_0, leading to the best predictor of Z_1 given Z_0 (see (2) and (3)), can be expressed as

$$\tilde{Y}_1 = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j - 1} \langle \rho Y_0, \psi_{jk} \rangle \psi_{jk}. \tag{21}$$

Our purpose is first to compute the coefficients $\langle \rho Y_0, \psi_{jk} \rangle = \langle Y_0, \rho^* \psi_{jk} \rangle$. These are similar to the ones obtained by wavelet-vaguelette decompositions of a homogeneous operator (see [21]). The wavelet-vaguelette method for regularizing linear ill-posed problems has also been considered by Dicken and Maass [20]; the approach presented in this section is closely related to the one used in [20, Section 4].

Recall from (6) that $D^* = C\rho^*$ and, therefore, $D^*\psi_{jk} = C\rho^*\psi_{jk}$. One way to solve the latter equation and get $\rho^*\psi_{jk}$ is by regularization, since the operator C is not invertible. By regularization (see Section 3.2), we estimate $\rho^*\psi_{jk}$ by

$$(\widehat{\rho^*\psi_{jk}})_l = (C^*_{n,l} C_{n,l} + \lambda I)^{-1} C^*_{n,l} D^*_n \psi_{jk},$$

where $C_{n,l} = C_n P_l$, $P_l : H \mapsto V_l$ and V_l is the scaling space corresponding to a periodic multiresolution analysis associated with the wavelet ψ. Therefore, we estimate $\langle \rho Y_0, \psi_{jk} \rangle$ by $\langle \widehat{\rho Y_0, \psi_{jk}} \rangle = \langle Y_0, (\widehat{\rho^*\psi_{jk}})_l \rangle$ and $D^*\psi_{jk}$ by $D^*_n \psi_{jk} = \frac{1}{n}\sum_{i=0}^{n-1} \langle Y_{i+1}, \psi_{jk} \rangle Y_i$.

In order now to compute $\langle \rho Y_0, \psi_{jk} \rangle$, we note that it is enough to compute $\langle \rho Y_0, \phi_{Jk} \rangle$, $J = \log_2(n)$, $k = 0, 1, \ldots, 2^j - 1$. This is so, because ρ is linear and we can use the pyramid algorithm (the *discrete wavelet transform*) to compute the wavelet coefficients at coarser scales (see [30]). Indeed, we have

$$\phi_{j\ell}(x) = \sum_k h_{\ell-2k}\phi_{j-1,k}(x) + \sum_k g_{\ell-2k}\psi_{j-1,k}(x), \quad x \in L^2[0,1],$$

where h and g are the quadrature mirror filters associated with the pyramid algorithm, implying that

$$\begin{aligned}\langle \rho Y_0, \phi_{j\ell} \rangle &= \left\langle \rho Y_0, \sum_k h_{\ell-2k}\phi_{j-1,k} \right\rangle + \left\langle \rho Y_0, \sum_k g_{\ell-2k}\psi_{j-1,k} \right\rangle \\ &= \sum_k h_{\ell-2k} \langle \rho Y_0, \phi_{j-1,k} \rangle + \sum_k g_{\ell-2k} \langle \rho Y_0, \psi_{j-1,k} \rangle. \end{aligned} \tag{22}$$

(Note that a similar remark for such a fast evaluation is given in [20, Proposition 5.2]). In order to estimate $\langle \rho Y_0, \phi_{Jk} \rangle = \langle Y_0, \rho^* \phi_{Jk} \rangle$, we need to estimate $\rho^* \phi_{Jk}$. To do that, we recall again from (6) that $D^* = C\rho^*$ and, therefore, $D^* \phi_{Jk} = C\rho^* \phi_{Jk}$. Again, by regularization, we estimate $\rho^* \phi_{Jk}$ by

$$(\widehat{\rho^* \phi_{Jk}})_l = (C_{n,l}^* C_{n,l} + \lambda I)^{-1} C_{n,l}^* D_n^* \phi_{Jk},$$

where $C_{n,l}$ is defined above, and $D_n^* \phi_{Jk} = \frac{1}{n}\sum_{i=0}^{n-1} \langle Y_{i+1}, \phi_{Jk} \rangle Y_i$. If ϕ is now a Coiflet of order $L > [s] + 1$, where $[x]$ is the integer part of x (see [18, p. 258]), then one can get the approximation (see, [3]) $\langle Y_{i+1}, \phi_{Jk} \rangle \simeq 2^{-J/2} Y_{i+1}(k/2^J)$. The resulting approximation $\widehat{D_n^* \phi_{Jk}} = \frac{2^{-J/2}}{n}\sum_{i=0}^{n-1} Y_{i+1}(k/2^J) Y_i$, while presenting a very small bias, leads to a highly oscillatory function (see [3]). In order to further regularize it, we associate with the discretization grid size m (hereafter, the discretization grid size m depends on n, the number of simple paths of an ARH(1) process, i.e. $m := m(n)$, but for notation simplicity we will omit this dependency and denote it by m), a resolution level $J(m) < J$, and estimate $D_n^* \phi_{Jk}$ by $\widetilde{D_n^* \phi_{J(m)k}}$, its orthogonal projection onto the scaling space $V_{J(m)}$ associated with the multiresolution analysis of H.

To control the previous approximations, we need to bound the *mean-squared approximation error* of $\mathbb{E}||\widehat{\rho f} - \rho f||^2$, for any $f \in H$, where

$$\widehat{\rho f} = \sum_{k=0}^{2^{J(m)}-1} \langle f, \widehat{\rho^* \phi_{J(m)k}} \rangle \phi_{J(m)k}$$

and

$$\widehat{\rho^* \phi_{J(m)k}} = f_{n,0}(C_n) \widetilde{D_n^* \phi_{J(m)k}}$$

with $f_{n,0} = (C_n + b_n I)^{-1}$ (see (11)). The error of approximating $D_n^* \phi_{Jk}$ by $\widetilde{D_n^* \phi_{J(m)k}}$ will be called the *interpolation error*. It is bounded above by the interpolation error of approximating $D_n^* \phi_{Jk}$ by $\widehat{D_n^* \phi_{Jk}}$. If ϕ is now a Coiflet of order $L > [s] + 1$, using

Lemma 3.1 in [3], we then have

$$\mathbb{E}||D_n^*\phi_{Jk} - \widehat{D_n^*\phi_{Jk}}||^2 = O(2^{-J(2s+1)})\mathbb{E}||\bar{Y}_n||^2 \quad \text{as } n \to \infty,$$

where $\bar{Y}_n = \frac{1}{n+1}\sum_{l=0}^{n} Y_l$. If $\mathbb{E}||Y_0||^2 < \infty$, then $E||\bar{Y}_n||^2 = O(\frac{1}{n})$, as $n \to \infty$ (see [14, Theorem 3.7]). Therefore,

$$\mathbb{E}||D_n^*\phi_{Jk} - \widehat{D_n^*\phi_{Jk}}||^2 = O(2^{-2J(s+1)}) \quad \text{as } n \to \infty,$$

implying that

$$\begin{aligned}\mathbb{E}||D_n^*\phi_{Jk} - \widetilde{D_n^*\phi_{J(m)k}}||^2 &= O(2^{-2J(s+1)}2^{2(J-J(m))}) \\ &= O(2^{-2(Js+J(m))}) \quad \text{as } n \to \infty.\end{aligned}$$

Moreover, for any $f \in H$, we have

$$||\widehat{\rho f} - \rho f||^2 = \sum_{k=0}^{2^{J(m)}-1} \langle f, \rho^*\phi_{J(m)k} - \widehat{\rho^*\phi_{J(m)k}} \rangle^2 + ||(I - P_{V_{J(m)}})\rho f||^2,$$

where $P_{V_{J(m)}}$ denotes the orthogonal projection operator of H onto $V_{J(m)}$. The assumption that the sample paths belong to $H = W_2^s[0,1]$, for $s > 1/2$, implies that $\rho f \in H = W_2^s[0,1]$, $s > 1/2$. Therefore, by using standard wavelet approximation results (see, for example, [18] or [34], for any $f \in H$, we have that

$$||(I - P_{V_{J(m)}})\rho f||^2 = O(2^{-2sJ(m)}) \quad \text{as } n \to \infty.$$

Using the Cauchy–Schwarz inequality and Proposition 3 in [32] (including the arguments in his proof), we get

$$\begin{aligned}&\sum_{k=0}^{2^{J(m)}-1} \langle f, \rho^*\phi_{J(m)k} - \widehat{\rho^*\phi_{J(m)k}} \rangle^2 \\ &\leqslant ||f||^2 \sum_{k=0}^{2^{J(m)}-1} ||\rho^*\phi_{J(m)k} - \widehat{\rho^*\phi_{J(m)k}}||^2 \\ &= |f||^2 \sum_{k=0}^{2^{J(m)}-1} ||\rho^*\phi_{J(m)k} - \widehat{\widehat{\rho^*\phi_{J(m)k}}} + \widehat{\widehat{\rho^*\phi_{J(m)k}}} - \widehat{\rho^*\phi_{J(m)k}}||^2,\end{aligned}$$

where $\widehat{\widehat{\rho^*\phi_{J(m)k}}}$ is the kth discrete wavelet coefficient at scale $J(m)$ of $f_{n,0}(C_n)D_n^*\phi_{J(m)k}$. Hence, by the triangular inequality, we get

$$\begin{aligned}&\mathbb{E}\left[\sum_{k=0}^{2^{J(m)}-1} \langle f, \rho^*\phi_{J(m)k} - \widehat{\rho^*\phi_{J(m)k}} \rangle^2\right] \\ &\leqslant 2\,||f||^2 \left[O(2^{J(m)}2^{-2(Js+J(m))}) + O\left(\frac{2^{J(m)-J}}{b_n}\right)\right] \\ &= 2||f||^2 \left[O(2^{-(2Js+J(m))}) + O\left(\frac{2^{J(m)-J}}{b_n}\right)\right],\end{aligned}$$

implying that, for any $f \in H$

$$\mathbb{E}||\widehat{\rho f} - \rho f||^2 = O(2^{-(2Js+J(m))}) + O\left(\frac{2^{J(m)-J}}{b_n}\right) + O(2^{-2sJ(m)}) \quad \text{as } n \to \infty, \qquad (23)$$

where the first term in (23) depends on $f \in H$ (not uniformly).

In order that the bias (the last term in (23)) tends to zero, we need $J(m) \to \infty$ as $m \to \infty$. But $J(m)$ has to go to ∞ at a slower speed than J, in other words, $J - J(m) \to \infty$. For example, if one takes $J(m) = \frac{\log_2(n)}{(1+2s)}$ and $b_n = \frac{1}{\log(n)}$, then the mean-squared approximation error is, for any $f \in H$

$$\mathbb{E}||\widehat{\rho f} - \rho f||^2 = O(\log(n)\ n^{-\frac{2s}{2s+1}}), \quad \text{as } n \to \infty,$$

implying the (mean-squared) consistency of our prediction estimator at the optimal (up to a logarithmic factor) rate.

Remark 4.1. To obtain the above approximation results, we have used a linear wavelet method, by assuming that the sample paths belong to a Sobolev space $H = W_2^s[0,1]$ with non-integer regularity index $s > 1/2$. If, instead, we assume that the sample paths belong to a Besov space on $[0,1]$ (see, for example, [34]) then, given the results in [20], we conjecture that a nonlinear wavelet thresholding estimator for correlated data (see, for example, [25]), computed in the same spirit as above, would define an adaptive consistent prediction estimator of ρf, $f \in H$. For the asymptotic rates of such estimators see, for example, [19]).

4.2. Regularized wavelet interpolation estimators

As mentioned in Section 2, one way to regularize the sample paths and to get a reasonable estimator for the prediction problem (2) is to use smoothing splines (see [11,17]). However, as mentioned at the beginning of Section 4, by using wavelet-based regularization techniques, we can reach stochastic processes with less smooth sample paths.

Denote by $\mathbf{Z}_l = Z_l(t_i)$, $i = 1, \dots, m$, $l = 0, 1, \dots, n$, $t_i = \frac{i}{m} = \frac{i}{2^J}$, the m-dimensional vector of sampled-values for the lth block of an ARH(1) process. As in [4], we shall first approximate the discrete sample paths of Z by a continuous process

$$I_m \mathbf{Z}_l = 2^{-J/2} \sum_{k=0}^{2^J-1} Z_l\left(\frac{k}{2^J}\right) \phi_{Jk} \qquad (24)$$

which is an approximation of the projection of $\mathbf{Z}_l$ onto the space V_J, associated with a multiresolution analysis of H. Using again Coiflets of order $L > [s] + 1$, the following (uniformly) bounds hold (as consequences of Lemma 3.1 in [4])

$$\sup_{t \in [0,1]} \left\{ 2^{-\frac{J}{2}} \sum_{k=0}^{2^J-1} \left| 2^{\frac{J}{2}} \langle \mathbf{Z}_l, \phi_{Jk} \rangle - Z_l\left(\frac{k}{2^J}\right) \right| \ |\phi_{Jk}| \right\} \leqslant O(2^{-Js}) \quad \text{a.s.} \quad \text{as } n \to \infty,$$

justifying our approximation $I_m\mathbf{Z}_l$ given in (24). For a given primary resolution level $j_0 \geqslant 0$, (24) can be expanded as

$$I_m\mathbf{Z}_l = \sum_{k=0}^{2^{j_0}-1} c^l_{j_0k}\phi_{j_0k} + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} d^l_{jk}\psi_{jk},$$

where, for any $l = 0, 1, \ldots, n$,

$$c^l_{j_0k} = \langle I_m\mathbf{Z}_l, \phi_{j_0k}\rangle, \quad k = 0, 1, \ldots, 2^{j_0} - 1,$$

$$d^l_{jk} = \langle I_m\mathbf{Z}_l, \psi_{jk}\rangle, \quad j = j_0, \ldots, J-1, \;\; k = 0, 1, \ldots, 2^j - 1.$$

We are going to use wavelet-based regularization to obtain smooth estimates of the sample paths. More precisely, one could solve the following variational problem:

$$\inf_{f \in H} \{||I_m\mathbf{Z}_l - f||^2_{L^2[0,1]} + \lambda ||P_{V^{\perp}_{j_0}} f||^2;\; f \in H\}, \tag{25}$$

where $\lambda > 0$ is the regularization parameter and $P_{V^{\perp}_{j_0}}$ denotes the orthogonal projection operator of H onto the orthogonal complement of V_{j_0}, associated with a multiresolution analysis of H. Here, for any $l = 0, 1, \ldots, n$,

$$f = \sum_{k=0}^{2^{j_0}-1} \alpha^l_{j_0k}\phi_{j_0k} + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta^l_{jk}\psi_{jk},$$

and

$$\alpha^l_{j_0k} = \langle f, \phi_{j_0k}\rangle, \quad k = 0, 1, \ldots, 2^{j_0} - 1,$$

$$\beta^l_{jk} = \langle f, \psi_{jk}\rangle, \quad j \geqslant j_0, \;\; k = 0, 1, \ldots, 2^j - 1.$$

Using the equivalent sequence norms for $H = W^s_2[0,1]$, $s > 1/2$, minimizing functional (25) is equivalent to minimizing the following expression (see [4]):

$$\sum_{k=0}^{2^{j_0}-1} (c^l_{j_0k} - \alpha_{j_0k})^2 + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} \left\{ (d^l_{jk} - \beta_{jk})^2 + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \lambda 2^{2js}\beta^2_{jk} \right\}. \tag{26}$$

One minimizes (26) by minimizing each term separately, and the solution is then given by

$$\widehat{\alpha^l_{j_0k}} = c^l_{j_0k}, \quad k = 0, 1, \ldots, 2^{j_0} - 1,$$

$$\widehat{\beta^l_{j_0k}} = \frac{d^l_{jk}}{(1 + \lambda 2^{2sj})}, \quad j = j_0, \ldots, J-1, \;\; k = 0, 1, \ldots, 2^j - 1,$$

$$\widehat{\beta^l_{jk}} = 0, \quad j \geqslant J, \;\; k = 0, 1, \ldots, 2^j - 1,$$

leading to the following smoothed sample paths

$$\tilde{Z}_{l,\lambda} = \sum_{k=0}^{2^{j_0}-1} \widehat{\alpha^l_{j_0k}}\phi_{j_0k} + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} \widehat{\beta^l_{jk}}\psi_{jk}.$$

Let now the empirical mean of the smoothed sample paths be denoted by

$$\tilde{a}_{n,\lambda} = \frac{1}{n+1} \sum_{l=0}^{n} \tilde{Z}_{l,\lambda}, \tag{27}$$

and let

$$\tilde{Y}_{l,\lambda} = \tilde{Z}_{l,\lambda} - \tilde{a}_{n,\lambda}, \quad l = 0, 1, \ldots, n$$

be the centered smoothed sample paths. We define the *smoothed covariance* and *smooth cross-covariance* (of order 1) operators, respectively, by

$$\tilde{C}_{n,\lambda} = \frac{1}{n+1} \sum_{l=0}^{n} \tilde{Y}_{l,\lambda} \otimes \tilde{Y}_{l,\lambda},$$

$$\tilde{D}_{n,\lambda} = \frac{1}{n} \sum_{l=0}^{n-1} \tilde{Y}_{l,\lambda} \otimes \tilde{Y}_{l+1,\lambda}$$

and

$$\tilde{D}^{*}_{n,\lambda} = \frac{1}{n} \sum_{l=0}^{n-1} \tilde{Y}_{l+1,\lambda} \otimes \tilde{Y}_{l,\lambda}.$$

Note that all smoothed sample paths $\tilde{Y}_{l,\lambda}$ belong to V_J. Using our previous notation (see (10)), let $\tilde{\Pi}^{k_n}_{\lambda}$ be the projection operator onto the space spanned by the first k_n eigenfunctions of $\tilde{C}_{n,\lambda}$. Then, we define the *regularized wavelet projection estimator* of ρ^* by

$$\widetilde{\rho^*}_{n,\lambda} = (\tilde{\Pi}^{k_n}_{\lambda} \tilde{C}_{n,\lambda} \tilde{\Pi}^{k_n}_{\lambda})^{-1} \tilde{D}^{*}_{n,\lambda} \tilde{\Pi}^{k_n}_{\lambda}. \tag{28}$$

Similarly, using our previous notation (see (11)), we define the *regularized wavelet resolvent estimator* by

$$\widetilde{\rho^*}_{n,\lambda,p} = f_{n,p}(\tilde{C}_{n,\lambda}) \tilde{D}^{*}_{n,\lambda}, \tag{29}$$

where

$$f_{n,p}(\tilde{C}_{n,\lambda}) = (\tilde{C}_{n,\lambda} + b_n I)^{-(p+1)} \tilde{C}^{p}_{n,\lambda}, \quad p \geqslant 0, \ b_n > 0, \ n \geqslant 0.$$

Remark 4.2. (i) The advantage of calculating (29) is that one does not need to compute the singular value decomposition of $\tilde{C}_{n,\lambda}$ to construct an estimator for ρ^* (meaning that not assumptions are needed to control the behavior of the eigenvalues of C). Large positive powers p produce smoother solutions. However, the larger the value of p the slower the convergence speed is in proving consistency results (see [32, Chapter 3]).

(ii) The resolvent estimator $\widetilde{\rho^*}_{n,\lambda,0}$ may be seen as a ridge regression estimator with smoothing parameter b_n.

4.2.1. Consistency results

Hereafter, we assume that

$$(\mathrm{H}_1):\quad \mathbb{E}(||Y_0||^4) < \infty \quad \text{and} \quad ||\rho^{j_0}||_{\mathscr{L}} < 1 \quad \text{for some } j_0 \geqslant 1.$$

This ensures that $Y = (Y_n;\ n \in \mathbb{Z})$ is a weakly stationary stochastic process with innovation ε (see, for example, [14, Theorem 3.1]). We shall also assume that

$$(\mathrm{H}_2):\quad C \quad \text{is one-to-one,}$$

otherwise ρ cannot be defined uniquely (see [32, Chapter 3]). It is also assumed that

$$(\mathrm{H}_3):\quad \mathscr{P}(\liminf \mathscr{E}_n) = 1, \quad \text{where } \mathscr{E}_n = \{\omega \in \Omega\text{: } \dim R(\Pi^{k_n} C_n \Pi^{k_n}) = k_n\}$$

with $R(A)$ denoting the range of the operator A. This assumption guarantees that the operator $\tilde{\Pi}_\lambda^{k_n} \tilde{C}_{n,\lambda} \tilde{\Pi}_\lambda^{k_n}$ in (28) is almost surely invertible after some $n \in \mathbb{Z}$ (see [32, Chapter 3]). Moreover, we assume that

$$(\mathrm{H}_4):\quad n\lambda_{k_n}^4 \to \infty \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^{k_n} \frac{g_k}{\lambda_k^2} \to 0, \quad \text{as } n \to \infty,$$

where $g_k = \max\{(\lambda_{k-1} - \lambda_k)^{-1}, (\lambda_k - \lambda_{k+1})^{-1}\}$ (see [32, Chapter 3]).

We are going first to show that $\tilde{a}_{n,\lambda}$ in (27) is a consistent estimator of the mean a.

Lemma 4.1. *The following holds*:

$$\mathbb{E}||a - \tilde{a}_{n,\lambda}||^2 = O\left(\frac{1}{n}\right) + O(\lambda + 2^{-2Js}) \quad \textit{as } n \to \infty.$$

Proof. Let $\tilde{a}_\lambda = \mathbb{E}(\tilde{a}_{n,\lambda})$. It is easily seen that

$$\mathbb{E}||a - \tilde{a}_{n,\lambda}||^2 \leqslant ||a - \tilde{a}_\lambda||^2 + \mathbb{E}||\tilde{a}_\lambda - \tilde{a}_{n,\lambda}||^2.$$

Due to the fact that $(a - \tilde{a}_\lambda) \in V_J^\perp$ and $(\tilde{a}_\lambda - \tilde{a}_{n,\lambda}) \in V_J$, we have that

$$\begin{aligned}\mathbb{E}||\tilde{a}_\lambda - \tilde{a}_{n,\lambda}||^2 &\leqslant 3\left\{\mathbb{E}||\tilde{a}_\lambda - a||^2 + \mathbb{E}||\frac{1}{(n+1)}\sum_{l=0}^{n}(Z_l - a)||^2.\right. \\ &\quad \left. + \mathbb{E}\left|\left|\frac{1}{(n+1)}\sum_{l=0}^{n}(\tilde{Z}_{l,\lambda} - Z_l)\right|\right|^2\right\} \\ &= 3(A_1 + A_2 + A_3), \quad \text{as } n \to \infty.\end{aligned}$$

We have that $A_2 = O(\frac{1}{n})$, as $n \to \infty$ (see [14, Theorem 3.7]), and that $A_1 = O\big(\lambda + 2^{-2sJ}\big)$, $A_3 = O\big(\lambda + 2^{-2sJ}\big)$, as $n \to \infty$ (see [4, Theorem 3.1]), completing thus the proof of the lemma. $\square$

By Lemma 4.1, we assume without loss of generality, that the process is centered and, therefore, we can work with the centered process $Y_l = Z_l - a$.

Lemma 4.2. *Using the notation of Section* 4.2, *the following hold*:

$$\mathbb{E}||\tilde{C}_{n,\lambda} - C||^2_{\mathrm{HS}} = O\left(\frac{1}{n}\right) + O(\lambda + 2^{-2sJ}) \quad as\ n \to \infty,$$

$$\mathbb{E}||\tilde{D}_{n,\lambda} - D||^2_{\mathrm{HS}} = O\left(\frac{1}{n}\right) + O(\lambda + 2^{-2sJ}) \quad as\ n \to \infty,$$

$$\mathbb{E}||\tilde{D}^*_{n,\lambda} - D^*||^2_{\mathrm{HS}} = O\left(\frac{1}{n}\right) + O(\lambda + 2^{-2sJ}) \quad as\ n \to \infty,$$

where $||\cdot||_{\mathrm{HS}}$ *stands for the Hilbert–Schmidt norm for Hilbert–Schmidt operators from* H *to* H.

Moreover, if $J > \frac{1+\gamma}{2s}\log_2(n)$ *for some* $\gamma > 0$ *and* $\lambda < n^{-(1+\delta)}$, $\delta > 0$, *the following hold*

$$||\tilde{C}_{n,\lambda} - C||_{\mathrm{HS}} \to 0 \quad \text{a.s.} \quad as\ n \to \infty,$$

$$||\tilde{D}_{n,\lambda} - D||_{\mathrm{HS}} \to 0 \quad \text{a.s.} \quad as\ n \to \infty,$$

$$||\tilde{D}^*_{n,\lambda} - D^*||_{\mathrm{HS}} \to 0 \quad \text{a.s.} \quad as\ n \to \infty.$$

Proof. It is easily seen that

$$\begin{aligned}\mathbb{E}||\tilde{C}_{n,\lambda} - C||^2_{\mathrm{HS}} &\leqslant 2\{\mathbb{E}||\tilde{C}_{n,\lambda} - C_n||^2_{\mathrm{HS}} + \mathbb{E}||C_n - C||^2_{\mathrm{HS}}\} \\ &= 2(B_1 + B_2).\end{aligned}$$

We have that $B_2 = O(\frac{1}{n})$, as $n \to \infty$ (see [14, Theorem 3.7]). Also, using similar arguments as in Lemma 4.1, B_1 can be expressed as

$$\begin{aligned}B_1 &= \mathbb{E}\left|\left|\frac{1}{(n+1)}\sum_{l=0}^{n}(\tilde{Y}_{l,\lambda} \otimes \tilde{Y}_{l,\lambda} - Y_l \otimes Y_l)\right|\right|^2_{\mathrm{HS}} \\ &\leqslant \frac{1}{(n+1)^2}\mathbb{E}\left|\left|\sum_{l=0}^{n}\{(Y_l - \tilde{Y}_{l,\lambda}) \otimes Y_l - \tilde{Y}_{l,\lambda} \otimes (Y_l - \tilde{Y}_{l,\lambda})\}\right|\right|^2_{\mathrm{HS}} \\ &\leqslant \frac{4}{(n+1)^2}\mathbb{E}\left(\sum_{l=0}^{n}||Y_l||_{L^2}||Y_l - \tilde{Y}_{l,\lambda}||_{L^2}\right)^2 \\ &\leqslant 4\mathbb{E}||Y||^2_{L^2}\mathbb{E}||Y - \tilde{Y}_\lambda||^2_{L^2} \\ &= O(\lambda + 2^{-2sJ}) \quad \text{as } n \to \infty.\end{aligned}$$

Similarly, the same result is true for $\mathbb{E}||\tilde{D}_{n,\lambda} - D||^2_{\mathrm{HS}}$ and $\mathbb{E}||\tilde{D}^*_{n,\lambda} - D^*||^2_{\mathrm{HS}}$.

The assumptions $J > \frac{1+\gamma}{2s}\log_2(n)$ for some $\gamma > 0$ and $\lambda < n^{-(1+\delta)}$, $\delta > 0$, suffices to ensure the convergence of the series $P(||\tilde{C}_{n,\lambda} - C||^2_{\mathrm{HS}} > \varepsilon)$, for some $\varepsilon > 0$, and the Borel–Cantelli lemma provides the almost surely convergence of $\tilde{C}_{n,\lambda}$. The same is true for $\tilde{D}_{n,\lambda}$ and $\tilde{D}^*_{n,\lambda}$, completing thus the proof of the lemma. $\square$

Lemma 4.3. *Let $\tilde{\lambda}_{k,n,\lambda}$ and $\tilde{e}_{k,n,\lambda}$ $(k = 1, \ldots, k_n)$ be respectively the first k_n eigenvalues and eigenfunctions of the operator $\tilde{C}_{n,\lambda}$. Then, the following hold*:

$$\mathbb{E}|\tilde{\lambda}_{k,n,\lambda} - \lambda_k|^2 = O\left(\frac{1}{n}\right) + O(\lambda + 2^{-2sJ}) \quad (\textit{for } k = 1, \ldots, k_n), \quad \textit{as } n \to \infty,$$

$$\mathbb{E}||\tilde{e}_{k,n,\lambda} - e_k||^2 = a_k^2 \left\{ O\left(\frac{1}{n}\right) + O(\lambda + 2^{-2sJ}) \right\} \quad (\textit{for } k = 1, \ldots, k_n), \quad \textit{as } n \to \infty,$$

where

$$a_1 = \frac{2\sqrt{2}}{(\lambda_1 - \lambda_2)} \quad \textit{and} \quad a_k = \frac{2\sqrt{2}}{\min(\lambda_{k-1} - \lambda_k, \lambda_k - \lambda_{k+1})}, \quad k = 2, \ldots, k_n.$$

Remark 4.3. Since the operator $\tilde{\Pi}_\lambda^{k_n} \tilde{C}_{n,\lambda} \tilde{\Pi}_\lambda^{k_n}$ is assumed to be almost surely invertible after some $n \in \mathbb{Z}$, the a_k $(k = 1, \ldots, k_n)$ are all different than zeros.

Proof. Using Lemma 3.1 in [13], we have that

$$|\tilde{\lambda}_{k,n,\lambda} - \lambda_k| \leqslant ||\tilde{C}_{n,\lambda} - C||_{\mathrm{HS}},$$

$$||\tilde{e}_{k,n,\lambda} - e_k|| \leqslant a_k \; ||\tilde{C}_{n,\lambda} - C||_{\mathrm{HS}}.$$

The rates of convergence follow now from Lemma 4.2, completing thus the proof of the lemma. □

We are now in the position to give consistency results for the regularized wavelet interpolation estimators $\widetilde{\rho^*}_{n,\lambda}$ given in (28) and $\widetilde{\rho^*}_{n,\lambda,p}$ given in (29) of ρ^*. In the following theorems, $\xrightarrow{p}$ stands for convergence *in probability*.

Theorem 4.1. *Under the assumptions* (H_1), (H_2), (H_3) *and* (H_4), *if* $\lambda_k = ar^k$, $a > 0$, $r \in (0, 1)$ *and if* $k_n = o(\min(\log(n), -\log(\lambda + 2^{-2sJ})))$ *as* $n \to \infty$ *and* $\lambda \to 0$, *then*

$$||\widetilde{\rho^*}_{n,\lambda}(\tilde{Y}_{n,\lambda}) - \rho^*(Y_n)|| \xrightarrow{p} 0, \quad \textit{as } n \to \infty.$$

Proof. The proof follows along the same lines of Proposition 4.6 in [13] by using Lemmas 4.1, 4.2 and 4.3. □

Theorem 4.2. *Under the assumptions* (H_1) *and* (H_2), *if* $J = O(\log_2(n))$, $\lambda = O(n^{-(1+\delta)})$ *for some* $\delta > 0$, *as* $n \to \infty$, *and if* $b_n \to 0$, $b_n^{p+2}\sqrt{n} \to \infty$ *for some* $p \geqslant 0$, *as* $n \to \infty$, *then*

$$||\widetilde{\rho^*}_{n,\lambda,p}(\tilde{Y}_{n,\lambda}) - \rho^*(Y_n)|| \xrightarrow{p} 0, \quad \textit{as } n \to \infty.$$

Proof. The proof follows along the same lines of Proposition 3 in [32, Chapter 3] by using Lemmas 4.1 and 4.2. □

Remark 4.4. Theorems 4.1 and 4.2 are concerned with the estimation of $\rho^*(Y_n)$. However, by continuous extension (see Section 2), they also apply to the best prediction $\rho(Y_n)$.

5. Applications and comparisons

The purpose of this section is to illustrate the performance of the proposed linear wavelet methods in finite sample situations by means of a real-life data example. We also compare our prediction results with those obtained by other methods available in the literature, in particular with a smoothing spline interpolation method and with a SARIMA model.

The computational algorithms related to wavelet analysis were performed using Version 8 of the WaveLab toolbox for MATLAB [16] that is freely available from `http://www-stat.stanford.edu/software/software.html`. The entire study was carried out using the MATLAB programming environment.

5.1. El Niño-Southern Oscillation

This application concerns with the prediction of a climatological times series describing El Niño-Southern Oscillation (ENSO) during the 12-month period of 1986, from monthly observations during the 1950–1985 period. ENSO is a natural phenomenon arising from coupled interactions between the atmosphere and the ocean in the tropical Pacific Ocean. El Niño (EN) is the ocean component of ENSO while Southern Oscillation (SO) is the atmospheric counterpart of ENSO. Most of the year-to-year variability in the tropics, as well as a part of the extratropical variability over both Hemispheres, is related to ENSO. For a detailed review of ENSO the reader is referred, for example, to [36].

An useful index of El Niño variability is provided by the sea surface temperatures averaged over the Niño-3 domain (5°S–5°N, 150°W–90°W). Monthly mean values have been obtained from January 1950 to December 1996 from gridded analyses made at the US National Centers for Environmental Prediction (see [42]). The time series of this EN index is depicted in Fig. 1, and shows marked interannual variations superimposed on a strong seasonal component. It has been analyzed by many authors (see, for example, [12]) and it is freely available from `http://www.cpc.ncep.noaa.gov/data/indices`.

Assuming first that an ARH(1) process can model these data, we have fitted this model with the wavelet methods proposed in Section 4. For these methods, the first 36 years, from 1950 to 1985, were considered as a learning sample. The discretization grid size for each sample path, which is equal to $m = 12$, is relatively small and not even a power of 2. In order to apply our wavelet methods, we have used the binned wavelet transform of [9], which is suited when the sample sizes are not a power of 2. The idea is to approximate directly the scaling function, ϕ, and the mother wavelet, ψ, by some simpler functions which can be more easily evaluated at a given point.

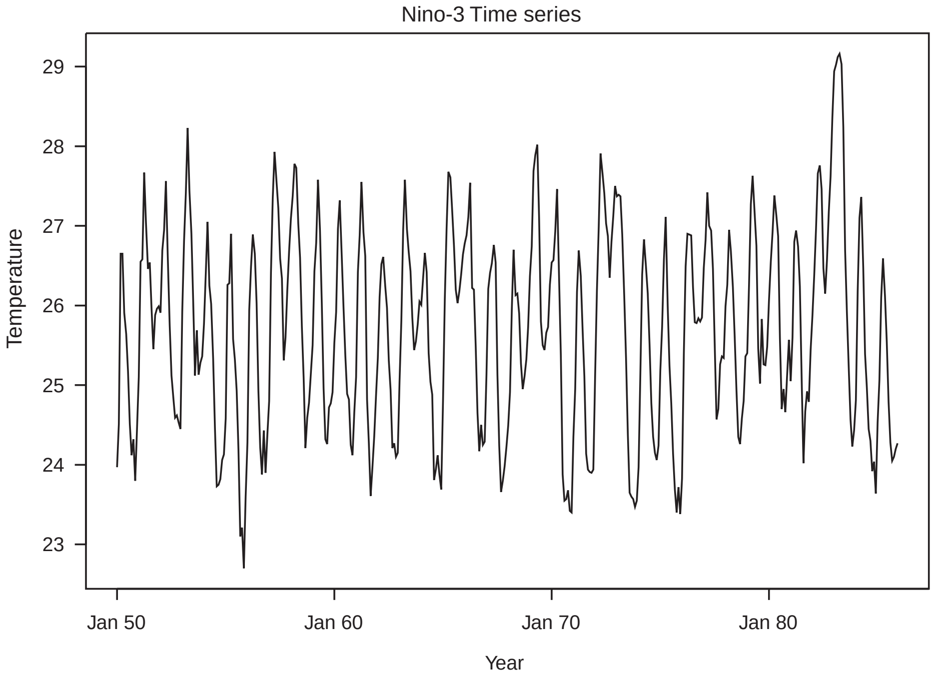

Fig. 1. The monthly mean Niño-3 surface temperature index in (deg C) which provides a contracted description of ENSO.

The resulting wavelet transform procedure, called BINWAV, is able to deal with random designs and sample sizes that are not power of 2. We refer to [9] for the algorithmic details.

As with any other nonparametric smoothing method, the proposed linear wavelet methods depend on tuning parameters. It is, therefore, desirable to select such parameters automatically. The problem of selecting the optimal resolution level, $J(m)$, is rather easier than the smoothing parameter, λ, and the dimensionality, q, for smoothing spline interpolation estimators. This is so because the optimal resolution level $J(m)$ is essentially reduced to being such that $J(m) < \frac{1}{2}\log_2(m)$. A commonly used selection rule adapted to our setting is to choose $J(m)$ as the minimizer of the cross-validation function

$$CV(J(m)) = m^{-1} \sum_{i=1}^{m} (X_i - \hat{X}_{(i)}(t_i))^2,$$

where $\hat{X}_{(i)}(t)$ is the leave-one-out predictor obtained by evaluating $\hat{X}$ (as a function of $J(m)$ and t) with the ith data point removed. It is found that this criterion produces reasonable results. In practice, for sample sizes between 10 and 100, we have found that it suffices to examine only $J(m) = 1$, 2 and 3. For this example, the data are pretty regular and the optimal value of the resolution level $J(m)$ for the regularized wavelet-vaguelette estimator (Wavelet I) discussed in Section 4.1 is small and was set equal to 1. For the regularized wavelet projection estimator (Wavelet II) discussed in Section 4.2, the smoothing parameter k_n was set equal to 2. For the regularized wavelet resolvent estimator (Wavelet III) discussed in Section 4.2, the smoothing parameters p and b_n were chosen to be equal to 1 and 0.25 respectively. We noted that similar predictions were reached for a range of b_n from 0.1 to 1.5, and values of p between 1 and 2. For the Wavelet I, II and III estimators, Daubechies nearly symmetric wavelets of order 6, *Symmlet 6*, were used (see [18, p. 195]), while the primary resolution level j_0 for the Wavelet II and III estimators was set equal to 2.

We have compared our results with those obtained by Besse et al. [12], using the smoothing spline interpolation estimator (Splines), with smoothing parameter λ and dimensionality (q) chosen optimally by a cross-validation criterion. In this case, λ was found equal to 1.6e-0.5 while q was found equal to 4. To complete the comparison, a suitable ARIMA model, including 12 month seasonality, has also been adjusted to the times series from January 1950 to December 1985. Following the Box–Jenkins methodology (see [15, Chap. 9]), the most parsimonious SARIMA model, validated through a portmanteau test for serial correlation of the fitted residuals, was driven by the parameters $(0,1,1) \times (1,0,1)_{12}$, i.e.

$$(1-\Theta_{12}B^{12})(1-B)X(t) = (1-\Theta_1 B)(1-\Theta^*_{12}B^{12})\varepsilon(t), \tag{30}$$

where B denotes the backshift operator and Θ_1, Θ_{12} and $\Theta^\star_{12}$ are appropriate coefficients. The left-hand side of this model corresponds to differencing at lags 1 and 12, and the right-hand side is a multiplicative moving average at lags 1 and 12.

The quality of the prediction is measured by two criteria: the *mean-squared error* (MSE) defined by

$$\mathrm{MSE} = \frac{1}{m}\sum_{t=1}^{m} (\hat{X}_{T+t} - x_{T+t})^2, \tag{31}$$

and the *relative mean-absolute error* (RMAE) defined by

$$\mathrm{RMAE} = \frac{1}{m}\sum_{t=1}^{m} \frac{|\hat{X}_{T+t} - x_{T+t}|}{x_{T+t}}, \tag{32}$$

where $\{T+t;\ t=1,\dots,m\}$ are the months of the year to be forecasted. In this application, we have taken $T = 1985$ and $m = 12$.

Fig. 2 displays the observed data of the 37th year (1986) and its predictions obtained by the various methods. One can notice that the Wavelet II estimator gives an almost similar prediction as the Splines estimator, both visually and in terms of the prediction criteria (31) and (32). Moreover, the Wavelet II estimator has a much less computational effort since it is 20 times faster than the Splines methods in CPU time. Note also the almost perfect prediction of the Wavelet III estimator for the first 8 months of 1986. This is a very nice property of this estimation, considering that present day forecasts of ENSO only show skill for leads of less than six months (see,

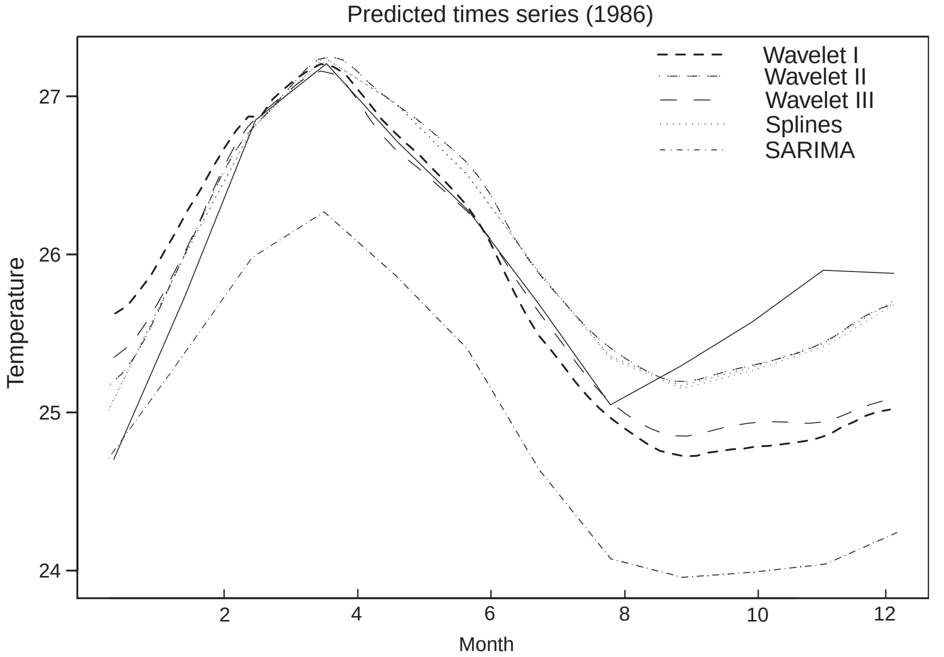

Fig. 2. The Niño-3 surface temperature during 1986 and its various predictions.

Table 1
MSE and RMAE for the prediction of Niño-3 surface temperatures during 1986 based on various methods

Prediction method	MSE	RAME (%)
Wavelet II	0.0630	0.89
Splines	0.0655	0.89
Wavelet III	0.1913	1.20
Wavelet I	0.3052	1.63
SARIMA	1.4567	3.72

for example, [27]). Note also the failure of the SARIMA estimator to produce an adequate prediction in this example (see [12], for an explanation). Finally the MSE and RAME of each prediction method are displayed in Table 1. The results have been presented in descending order, starting from the best method.

We conclude this section by pointing out that, apart from having much faster implementations, the improvement of the proposed linear wavelet methods over the smoothing spline interpolation method is not so obvious in this example. We have also applied the proposed linear wavelet methods to another data set which concern with the prediction of the degree of hotel occupation in Granada (Spain) during the 12-month period of 1994 from monthly observations during the 1974–1993 period. This data set has also been used by Aguilera et al. [2] to illustrate continuous-time prediction by an approximated principal component method. It has been found that the proposed linear wavelet methods produce the best predictions, both visually and in terms of the prediction criteria (31) and (32). Due to lack of space, however, we do not include this example here but we refer instead to [10].

6. Concluding remarks

We have proposed some linear wavelet methods to predict a continuous-time stochastic process on an entire time-interval in terms of its recent past, based on the notion of ARH(1) processes.

One of the most important advantages of the proposed linear wavelet methods is that they have a much faster implementation. While these methods require the specification of a number of tuning parameters for optimal prediction, the range of these parameters is much more restricted than the smoothing spline interpolation method. As shown in the application that we have used to illustrate the proposed linear wavelet methods, there is a quite simple recipe for the choice of these parameters.

Another advantage of the proposed linear wavelet methods is that they provide an excellent setup for continuous-time prediction of stochastic processes with much less smooth sample paths than other smoothing prediction methods available in the literature. In particular, the regularized wavelet interpolation estimators discussed in Section 4.2, were derived by penalizing the approximation squared error $||I_m\mathbf{Z}_l - f||^2_{L^2[0,1]}$ with a penalty function that penalizes the wavelet coefficients (β_{jk}) of $f \in H$, leading to a linear wavelet method of estimation which is optimal when $H = W_2^s$ with noninteger regularity index $s > 1/2$ (for smoothing spline methods this is only true for integer regularity index $s \geqslant 1$).

Assuming that the sample paths of $\mathbf{Z}_l$ belong to such a Sobolev space is, however, a strong assumption on the integral kernel associated with the covariance operator C. In order to make weaker assumptions, one could work along the lines suggested below extending thus the proposed linear wavelet methods to deal with less regular sample paths. A natural setup, dealing in particular with inhomogeneous sample paths, is to assume that they belong to specific Besov spaces (for example, $B^s_{p,p}$, $p \geqslant 2$ or $B^s_{p,2}$, $1 \leqslant p \leqslant 2$). In such cases, it is known that linear wavelet methods are not optimal (see, for example, [22]) and that one should seek nonlinear wavelet methods (see, for example, [1,5,6,43]). Nonlinear wavelet methods can be derived through penalization of functionals of the form

$$||I_m\mathbf{Z}_l - f||^2 + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} r_\lambda(|\beta_{jk}|)$$

over $f \in B^s_{p,p}$ $(p \geqslant 2)$ or $f \in B^s_{p,2}$ $(1 \leqslant p \leqslant 2)$. For appropriate forms of the penalties r_λ, see [7].

The classical autoregressive Hilbert model could be extended to the above mentioned Besov spaces by using the results in [14, Chapter 6] for Banach-valued autoregressive processes. We, therefore, conjecture that nonlinear wavelet approaches would lead to optimal consistency rates for the resulting prediction estimators. However, a deeper analysis of the arguments to support such a conjecture go beyond the intent of this paper, but provides an interesting topic for future research.

Acknowledgments

A. Antoniadis was supported by 'Project IDOPT, INRIA-CNRS-IMAG', 'Project ADÈMO, Rhône-Alpes' and 'Project AMOA, IMAG'. Theofanis Sapatinas was supported by a 'Royal Society Research Grant' and a 'Highly Structured Stochastic Systems Grant'. Theofanis Sapatinas would like to thank Anestis Antoniadis for financial support and excellent hospitality while visiting Grenoble to carry out this work. Helpful comments made by two anonymous referees are gratefully acknowledged. We also thank one of the referees for bringing into our attention the article by Dicken and Maass [20].

References

[1] F. Abramovich, T.C. Bailey, T. Sapatinas, Wavelet analysis and its statistical applications, Statistician 49 (2000) 1–29.
[2] A.M. Aguilera, F.A. Ocaña, M.J. Valderrama, An approximated principal component prediction model for continuous-time stochastic processes, Appl. Stochast. Models Data Anal. 13 (1997) 61–72.
[3] A. Antoniadis, Smoothing noisy data with coiflets, Statist. Sinica 4 (1994) 651–678.
[4] A. Antoniadis, Smoothing noisy data with tapered coiflets series, Scand. J. Statist. 23 (1996) 313–330.
[5] A. Antoniadis, Wavelets in statistics: a review (with discussion), J. Ital. Statist. Soc. 6 (1997) 97–144.
[6] A. Antoniadis, J. Bigot, T. Sapatinas, Wavelet estimators in nonparametric regression: a comparative simulation study, J. Statist. Software 6 (6) (2001) 1–83.
[7] A. Antoniadis, J. Fan, Regularization of wavelets approximations (with discussion), J. Amer. Statist. Assoc. 96 (2001) 939–967.
[8] A. Antoniadis, G. Grégoire, I.W. McKeague, Wavelet methods for curve estimation, J. Amer. Statist. Assoc. 89 (1994) 1340–1353.
[9] A. Antoniadis, D.T. Pham, Wavelet regression for random or irregular design, Comput. Statist. Data Anal. 28 (1998) 353–370.
[10] A. Antoniadis, T. Sapatinas, Wavelet methods for continuous-time prediction using representations of autoregressive processes in Hilbert spaces, Rapport de Recherchie, RR-1042-S, Laboratoire IMAG-LMC, University Joseph Fourier, France, 2001.
[11] P.C. Besse, H. Cardot, Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1, Canad. J. Statist. 24 (1996) 467–487.
[12] P.C. Besse, H. Cardot, D.B. Stephenson, Autoregressive forecasting of some functional climatic variations, Scand. J. Statist. 27 (2000) 673–687.
[13] D. Bosq, Modelization, nonparametric estimation and prediction for continuous time processes, in: G. Roussas (Ed.), Nonparametric Functional Estimation and Related Topics, Nato ASI Series C, Vol. 335, Kluwer Academic Publishers, Dortrecht, 1991, pp. 509–529.

[14] D. Bosq, Linear Processes in Function Spaces, in: Lecture Notes in Statistics, Vol. 149, Springer, New York, 2000.
[15] G.E. Box, G.M. Jenkins, Time Series Analysis, Holden Day, San Francisco, 1976.
[16] J.B. Buckheit, S. Chen, D.L. Donoho, I.M. Johnstone, J. Scargle, About WaveLab, Technical Report, Department of Statistics, Stanford University, USA, 1995.
[17] H. Cardot, Contibution à l'estimation et à la prévision statistique de données fonctionnelles, Ph.D. Thesis, University of Toulouse 3, France, 1997.
[18] I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.
[19] B. Delyon, A. Juditsky, On minimax wavelet estimators, Appl. Comput. Harm. Anal. 3 (1996) 215–228.
[20] V. Dicken, P. Maass, Wavelet-Galerkin methods for ill-posed problems, J. Inverse Ill-Posed Probl. 4 (1996) 203–222.
[21] D.L. Donoho, Nonlinear solutions of linear inverse problems by wavelet-vaguelette decompositions, Appl. Comput. Harm. Anal. 2 (1995) 101–126.
[22] D.L. Donoho, I.M. Johnstone, Minimax estimation via wavelet shrinkage, Ann. Statist. 26 (1998) 879–921.
[23] P.R. Halmos, What does the spectral theorem say?, Amer. Math. Monthly 70 (1963) 241–247.
[24] G. Helmerg, Introduction to Spectral Theory in Hilbert Space, in: North-Holland Series in Applied Mathematics and Mechanics, Vol. 6, North-Holland, Amsterdam, 1969.
[25] I.M. Johnstone, B.W. Silverman, Wavelet threshold estimators for data with correlated noise, J. R. Statist. Soc. Ser. B 59 (1997) 319–351.
[26] T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, Berlin, 1976.
[27] M. Latif, T. Barnett, M. Cane, M. Flugel, N. Graham, J. Xu, S. Zebiak, A review of ENSO prediction studies, Climate Dynamics 9 (1994) 167–179.
[28] P. Maass, A. Rieder, Wavelet-accelerated Tikhonov-Phillips regularization with applications, in: H.W. Engl, A.K. Louis, W. Rundell (Eds.), Inverse Problems in Medical Imaging and Nondestructive Testing, Springer-Verlag, Wien, 1997, pp. 134–159.
[29] B.A. Mair, F.H. Ruymgaart, Statistical inverse estimation in Hilbert scales, SIAM J. Appl. Math. 56 (1996) 1424–1444.
[30] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Machine Intell. 11 (1989) 674–693.
[31] S.G. Mallat, A Wavelet Tour of Signal Processing, 2nd Edition, Academic Press, San Diego, 1999.
[32] A. Mas, Estimation d'opérateurs de corrélation de processus fonctionnels: lois limites, tests, déviations modérées, Ph.D. Thesis, University of Paris 6, France, 2000.
[33] F. Merlevéde, Processus linéaires Hilbertiens: inversibilité, theorémes limites, Estimation et prévision, Ph.D. Thesis, University of Paris 6, France, 1996.
[34] Y. Meyer, Wavelets and Operators, Cambridge University Press, Cambridge, 1992.
[35] T. Mourid, Contribution à la statistique des processus autorégressifs à temps continu, Ph.D. Thesis, University of Paris 6, France, 1995.
[36] S. Philander, El Niño, La Niña and the Southern Oscillation, Academic Press, San Diego, 1990.
[37] R. Plato, G. Vainikko, On the regularization of projection methods for solving ill-posed problems, Numer. Math. 57 (1990) 63–79.
[38] B. Pumo, Estimation et prévision de processus autorégressifs fonctionnels. Application aux processus á temps continu, Ph.D. Thesis, University of Paris 6, France, 1992.
[39] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, Springer-Verlag, New York, 1997.
[40] J.O. Ramsay, B.W. Silverman, Applied Functional Data Analysis, Springer-Verlag, New York, 2002.
[41] A. Rieder, A wavelet multilevel method for ill-posed problems stabilized by Tikhonov regularization, Numer. Math. 75 (1997) 501–522.
[42] T.M. Smith, R. Reynolds, R. Livezey, D. Stokes, Reconstruction of historical sea surface temperatures using empirical orthogonal functions, J. Climate 9 (1996) 1403–1420.
[43] B. Vidakovic, Statistical Modeling by Wavelets, John Wiley & Sons, New York, 1999.