A Functional Wavelet-Kernel Approach for Continuous-time Prediction

Anestis Antoniadis, Efstathios Paparoditis, Theofanis Sapatinas

A Functional Wavelet-Kernel Approach for Continuous-time Prediction

Anestis ANTONIADIS,
Laboratoire IMAG-LMC,
University Joseph Fourier,
BP 53, 38041 Grenoble Cedex 9,
FRANCE.
Email: Anestis.Antoniadis@imag.fr

Efstathios PAPARODITIS
Department of Mathematics and Statistics,
University of Cyprus,
P.O. Box 20537,
CY 1678 Nicosia,
CYPRUS.
Email: stathisp@ucy.ac.cy

and

Theofanis SAPATINAS
Department of Mathematics and Statistics,
University of Cyprus,
P.O. Box 20537,
CY 1678 Nicosia,
CYPRUS.
Email: t.sapatinas@ucy.ac.cy

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 functional kernel nonparametric regression estimation techniques where observations are segments of the observed process considered as curves. These curves are assumed to lie within a space of possibly inhomogeneous functions, and the discretized times series dataset consists of a relatively small, compared to the number of segments, number of measurements made at regular times. We thus consider only the case where an asymptotically non-increasing number of measurements is available for each portion of the times series. We estimate conditional expectations using appropriate wavelet decompositions of the segmented sample paths. A notion of similarity, based on wavelet decompositions, is used in order to calibrate the prediction. Asymptotic properties when the number of segments grows to infinity are investigated under mild conditions, and a nonparametric resampling procedure is used to generate, in a flexible way, valid asymptotic pointwise confidence intervals for the predicted trajectories. We illustrate the usefulness of the proposed functional wavelet-kernel methodology in finite sample situations by means of three real-life datasets that were collected from different arenas.

Some key words: α-MIXING; BESOV SPACES; CONTINUOUS-TIME PREDICTION; FUNCTIONAL KERNEL REGRESSION; POINTWISE PREDICTION INTERVALS; RESAMPLING; SARIMA MODELS; SMOOTHING SPLINES; WAVELETS

1 INTRODUCTION

In many real life situations one seeks information on the evolution of a (real-valued) 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 $[l\delta, (l + 1)\delta]$, $l = 0, 1, \ldots, i - 1$ with $\delta = T/i$, and to consider the stochastic process $Z = (Z_i;\ i \in \mathbb{N}^+)$, where $\mathbb{N}^+ = \{1, 2, \ldots\}$, defined by

$$Z_i(t) = X(t + (i - 1)\delta), \quad i \in \mathbb{N}^+, \quad \forall\ t \in [0, \delta). \tag{1}$$

This representation is especially fruitful if X has a seasonal component with period δ and can be decomposed into locally stationary parts. 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, Ramsay & Silverman, 1997).

In the recent literature, practically all investigations to date for this prediction problem are for the case, where one assumes that an appropriately centered version of the stochastic process Z is a (zero-mean) Hilbert-valued *autoregressive (of order 1) processes* (ARH(1)); the best prediction of

Z_{n+1} given its past history $(Z_n, Z_{n-1}, \ldots, Z_1)$ is then given by

$$\begin{aligned} \tilde{Z}_{n+1} &= \mathbb{E}(Z_{n+1} \mid Z_n, Z_{n-1}, \ldots, Z_1) \\ &= \rho(Z_n), \quad n \in \mathbb{N}^+, \end{aligned}$$

where ρ is a bounded linear operator associated with the ARH(1) process. The adopted approaches mainly differ in the way of estimating the 'prediction' operator ρ, or its value $\rho(Z_n)$ given $Z_1, Z_2, \ldots, Z_n$ (see, e.g., Bosq, 1991; Besse & Cardot, 1996; Bosq, 2000; Antoniadis & Sapatinas, 2003).

In many practical situations, however, the stochastic process Z may not have smooth sample paths (lying in H) or may not be modelled with such an autoregressive structure. This is the case that we consider in the following development. In particular, we also assume that the (real-valued) continuous-time stochastic process $X = (X(t);\ t \in \mathbb{R})$ possesses a representation of the form (1) with short duration 'blocks' Z_i, for $i \in \mathbb{N}^+$. We then develop a version of prediction via functional regression analysis, in which both the predictor and response variables are functions of time, using a conditioning idea. Under mild assumptions on the observed time series, a one time-interval ahead prediction of the 'block' Z_{n+1} is obtained by kernel regression of the present 'block' Z_n on the past 'blocks' $\{Z_{n-1}, Z_{n-2}, \ldots, Z_1\}$. The resulting predictor will be seen as a weighted average of the past 'blocks', placing more weight on 'blocks' that are similar to the present one. Hence, the analysis is rooted in the ability to find 'similar blocks'. Considering that 'blocks' can be quite irregular curves, similarity matching is based on a distance metric on the wavelet coefficients of an appropriate wavelet decomposition of the 'blocks'. A resampling scheme, involving resampling of the original 'blocks' to form 'pseudo-blocks' of the same duration, is then used to calculate pointwise prediction intervals for the predicted 'block'.

The paper is organized as follows. In Section 2, we introduce basic notions for continuous-time prediction, relevant notions on wavelet-based orthogonal expansions of continuous-time stochastic processes, and describe the strictly stationarity and α-mixing assumptions that are going to be adopted for forecasting. In Section 3, we discuss the extension of the conditioning approach to the one time-interval ahead prediction. Resampling-based pointwise prediction intervals are presented in Section 4. In Section 5, we illustrate the usefulness of the proposed functional wavelet-kernel approach for continuous-time prediction by means of three real-life datasets that were collected from different arenas. We also compare the resulting predictions with those obtained by two other methods in the literature, in particular with a smoothing spline method and with the SARIMA model. Some concluding remarks are made in Section 6. Proofs and auxiliary results are compiled in the Appendix.

2 PRELIMINARIES AND NOTATIONS

2.1 Generalities

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, rich enough so that all random variables considered in the following development can be defined on this space, and let $X = (X(t);\ t \in \mathbb{R})$ be a (real-valued) continuous-time stochastic process defined on this space. Motivated by applications to prediction and forecasting, it is supposed that the time-interval on which the continuous-time stochastic process is observed is divided into intervals of constant-width $\delta > 0$ so that, from X, a functional-valued random variable sequence $(Z_i;\ i \in \mathbb{N}^+)$ is constructed according to the representation (1), i.e.,

$$Z_i(t) = X(t + (i-1)\delta), \quad i \in \mathbb{N}^+, \quad \forall\, t \in [0, \delta).$$

In the sequel, we regard the Z_i's as elements of a certain (semi-)normed functional linear space H equipped with (semi-)norm $\|\cdot\|$ and its Borel σ-field $\mathcal{F}_{\|\cdot\|}$. Recall that our aim is a one-ahead time interval prediction which, under the above notation, is reduced in studying a corresponding problem for the H-valued time series $Z = (Z_i;\ i \in \mathbb{N}^+)$. In what follows, we assume that the time series Z is strictly stationary with $\mathbb{E}(\|Z_i\|) < \infty$. If the time series Z is not stationary, it is assumed that it has been transformed to a stationary one by a preprocessing procedure, and the procedures that we are going to develop hold for the resulting stationary time series.

In practice, the random curves Z_i are only known at discretized equidistant time values, say $t_1, \ldots, t_P$. Thus, they must be approximated by some H-valued functions, which in our case will be realized by first expanding the Z_i's into a wavelet basis and then estimating consistently the coefficients from the observed discrete data. We may have used a fixed spline basis or a Fourier basis instead but there are some good reasons to prefer wavelet bases. When P is fixed, using spline interpolation could make sense if the sample paths exhibits a uniformly smooth temporal structure thus not requiring any smoothing to stabilize the variance. When P is large, a necessary smoothing step in the spline approximation would be necessary and since the smoothing is global this would not be appropriate when the sample paths are composed of different temporal structures. The same remarks apply for a Fourier basis. On the contrary, when P is fixed and the sample paths are continuous, one may choose an interpolation wavelet basis and the wavelets coefficients are computed directly from the sampled values instead of inner product integrals. When Z_i are observed either continuously or on a very fine discretization grid, then wavelets can be used successfully for compression of a continuous-time stochastic process, in the sense that the sample paths can be accurately reconstructed from a fraction of the full set of wavelet coefficients. Whatever the setting is, the wavelet decomposition of the sample paths will be a local one, so that if the information relevant to our prediction problem is contained in a particular part or parts of the sample path, as it is typically the case in many practical applications, this information will be carried by a very

small number of wavelet coefficients.

Below, we recall some background on interpolating wavelets and orthonormal wavelet expansions of continuous-time stochastic processes that we are going to use in the subsequent development.

2.2 Orthonormal wavelet expansions of continuous-time stochastic processes

The discrete wavelet transform (DWT), as formulated by Mallat (1989) and Daubechies (1992), is an increasingly popular tool for the statistical analysis of time series (see, e.g., Nason & von Sachs, 1999; Percival & Walden, 2000; Fryzlewicz, Van Bellegem & von Sachs, 2003). The DWT maps a time series into a set of wavelet coefficients, each one associated with a particular scale. Two distinct wavelet coefficients can be either 'within-scale' (i.e., both are associated with the same scale) or 'between-scale' (i.e., each has a distinct scale).

One reason for the popularity of the DWT in times series analysis is that measured data from most processes are inherently multiscale in nature, owing to contributions from events occurring at different locations and with different localization in time and frequency. Consequently, data analysis and modelling methods that represent the measured variables at multiple scales are better suited for extracting information from measured data than methods that represent the variables at a single scale.

Let $S(t) \equiv S(\omega, t)$ be a mean-square continuous-time stochastic process defined on $\Omega \times [0,1]$, i.e.

$$S \in L^2(\Omega \times [0,1]) = \left\{ X(t) : \Omega \to \mathbb{R},\ t \in [0,1] \ \middle|\ \mathbb{E}\left(\int_0^1 X^2(t)\, dt \right) < \infty \right\}.$$

Recall that (see, e.g., Neveu, 1968) $L^2(\Omega \times [0,1])$ is a separable Hilbert space equipped with inner product defined by

$$\langle X_1, X_2 \rangle = \mathbb{E}\left(\int_0^1 X_1(t) X_2(t)\, dt \right).$$

To develop a wavelet-based orthonormal expansion, we mimic the construction of a (regular) multiresolution analysis of $L^2([0,1])$ (see, e.g., Mallat, 1989). In other words, consider a ladder of closed subspaces

$$V_{j_0} \subset V_{j_0+1} \subset \cdots \subset L^2([0,1]),$$

with any fixed $j_0 \geq 0$, whose union is dense in $L^2([0,1])$, and where, for each j, V_j is spanned by 2^j orthonormal scaling functions $\{\phi_{j,k} :\ k = 0, \ldots, 2^j - 1\}$, such that $\mathrm{supp}(\phi_{j,k}) \subset [2^{-j}(k-c), 2^{-j}(k+c)]$, with c a constant not depending on j. At each resolution level j, the orthonormal complement W_j between V_j and V_{j+1} is generated by 2^j orthonormal wavelets $\{\psi_{j,k} :\ k = 0, \ldots, 2^j - 1\}$, such that $\mathrm{supp}(\psi_{j,k}) \subset [2^{-j}(k-c), 2^{-j}(k+c)]$. As a consequence, the family $\cup_{j \geq j_0} \{\psi_{j,k} :\ k = 0, \ldots, 2^j - 1\}$, completed with $\{\phi_{j_0,k} :\ k = 0, \ldots, 2^{j_0} - 1\}$, constitutes an orthonormal basis of $L^2([0,1])$.

Similarly, we define a sequence of approximating spaces of $L^2(\Omega \times [0,1])$ by

$$V_j(\Omega \times [0,1]) = \left\{ X \in L^2(\Omega \times [0,1]) \;\middle|\; X(t) = \sum_{k=0}^{2^j-1} \xi_{j,k}\phi_{j,k}(t),\ \sum_{k=0}^{2^j-1} \mathbb{E}(\xi_{j,k})^2 < \infty \right\},$$

where $\{\xi_{j,k} :\ k = 0, \ldots, 2^j - 1\}$ is a sequence of random variables and $\phi_{j,k}$ is the scaling basis of V_j. Note that since $L^2(\Omega \times [0,1])$ is isomorphic to the Hilbert tensor product $L^2(\Omega) \otimes L^2([0,1])$, the stochastic approximating spaces $V_j(\Omega \times [0,1])$ are closed subspaces of $L^2(\Omega \times [0,1])$. Note also that for every $X \in V_j(\Omega \times [0,1])$, one has $\mathbb{E}(X) \in L^2([0,1])$ since

$$\int_0^1 [\mathbb{E}(X(t))]^2\, dt = \int_0^1 \left[\mathbb{E}\left(\sum_{k=0}^{2^j-1} \xi_{j,k}\phi_{j,k}(t) \right) \right]^2 dt = \sum_{k=0}^{2^j-1} [\mathbb{E}(\xi_{j,k})]^2 \leq \sum_{k=0}^{2^j-1} \mathbb{E}(\xi_{j,k})^2,$$

by orthonormality of the scaling functions. Following Cohen & D'Ales (1997), it is easy to see that $\{V_j(\Omega \times [0,1]) :\ j \in \mathbb{N}_0\}$ is a multiresolution analysis of $L^2(\Omega \times [0,1])$. Moreover if $W_j(\Omega \times [0,1])$ denotes the orthonormal complement of $V_j(\Omega \times [0,1])$ in $V_{j+1}(\Omega \times [0,1])$, then one naturally has the following stochastic wavelet orthonormal expansion

$$X \in L^2(\Omega \times [0,1]) \leftrightarrow X(t) = \sum_{k=0}^{2^{j_0}-1} \xi_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \eta_{j,k}\psi_{j,k}(t),$$

where $\xi_{j,k} = \int_0^1 \phi_{j,k}(t)X(t)\, dt$ and $\eta_{j,k} = \int_0^1 \psi_{j,k}(t)X(t)\, dt$. The above remarks clearly show that the wavelet-based orthonormal expansion is a fundamental tool for viewing the continuous-time stochastic process in both time and scale domains.

In order to allow for inhomogeneous sample paths, the notion of the Besov space ($B^s_{p,q}$) comes naturally into the picture. Without getting into mathematical details, we just point out that Besov spaces are known to have exceptional expressive power: for particular choices of the parameters s, p and q, they include, e.g., the traditional Hölder ($p = q = \infty$) and Sobolev ($p = q = 2$) classes of smooth functions, and the inhomogeneous functions of bounded variation sandwiched between $B^1_{1,\infty}$ and $B^1_{1,1}$. The parameter p can be viewed as a degree of function's inhomogeneity while s is a measure of its smoothness. Roughly speaking, the (not necessarily integer) parameter s indicates the number of function's (fractional) derivatives, where their existence is required in an L^p-sense, while the additional parameter q provides a further finer gradation (see, e.g.,. Meyer, 1992). Therefore, from an approximation perspective, if the sample paths of the continuous-time stochastic process X belong to an inhomogeneous Besov space of regularity $s > 0$, and if one uses regular enough scaling functions, one may approximate in $L^2(\Omega \times [0,1])$ any sample path of the process X by its projection onto V_J at a rate of the order $\mathcal{O}(2^{-sJ})$ (see Cohen & D'Ales, 1997, Theorem 2.1). This is, in fact, a simple rephrasing, in the stochastic framework, of the deterministic results on the multiresolution approximation of functions in Besov spaces when the

error is measured in the $L^2([0,1])$-norm. This result has the advantage that dimension reduction by basis truncation in the wavelet domain is controlled more precisely.

Consider now the case where the stochastic signal is sampled over a finite number P of equidistant points and assume that $P = 2^J$. One then may use an interpolating wavelet basis, as the one constructed by Donoho (1992), to interpolate the observed values. The interpolating scaling function ϕ corresponds to the autocorrelation function of an orthogonal, regular enough, scaling function and the projection onto the scaling approximating space V_J is then given by

$$\mathcal{P}_J(Z)(t) = \sum_{k=0}^{2^J-1} Z(t_k)\phi(2^J t - k).$$

While this projector has no orthogonality property, it retains however the good approximations properties of projection operators derived from orthogonal multiresolution analyses (see Mallat, 1999, Theorem 7.21). When P is not anymore a power of two, one may still use the above scheme by adapting the interpolating wavelet basis to the sampling grid using an appropriate subdivision scheme for interpolation (see Cohen, Dyn & Matei, 2003).

We conclude this section by recalling the strictly stationarity and α-mixing concepts that we are going to adopt for developing the proposed functional wavelet-kernel prediction methodology.

2.3 Strictly stationarity and α-mixing

One of our main assumption in predicting the times series Z will be its strictly stationarity. We therefore recall some results on strictly stationarity from the above stochastic multiresolution analysis perspective. Recall that, for all $X \in V_J(\Omega \times [0,1])$, we have

$$X(t) = \sum_{k=0}^{2^J-1} \xi_{J,k}\phi_{J,k}(t).$$

Therefore,

$$\begin{aligned} X(t+s) &= \sum_{k=0}^{2^J-1} \xi_{J,k}\phi_{J,k}(t+s) = \sum_{k=0}^{2^J-1} \xi_{J,k}2^{J/2}\phi(2^J(t+s)-k) \\ &= \sum_{k=0}^{2^J-1} \xi_{J,k}2^{J/2}\phi(2^J t - (k-2^J s)) = \sum_{k=0}^{2^J-1} \xi_{J,k+2^J s}\phi_{J,k}(t). \end{aligned}$$

¿From the above, it is easy to see that if X is a strictly stationary process then, at any resolution level j, the vector of its scaling coefficients, is also strictly stationary. As shown in Cheng & Tong (1996), the converse is also true. It follows that strictly stationarity of the discrete-time series Z implies strictly stationarity in the above sense of the sequence of the scaling coefficients vectors at any resolution. Note, moreover, that if the strictly stationarity assumption is too strong, one

could calibrate the non-stationarity by considering only J-stationarity (see Cheng & Tong, 1998), that is, strictly stationarity of the scaling coefficients up to (finest) resolution J, with eventually a different distribution at each resolution level j.

Our theoretical results will be derived under α-mixing assumptions on the time series $Z = (Z_i;\ i \in \mathbb{N}^+)$. Recall that for a strictly stationary series $Z = (Z_i;\ i \in \mathbb{N}^+)$, the α-mixing coefficient (see Rosenblatt, 1956) is defined by

$$\alpha_Z(m) = \sup_{A \in \mathcal{D}_l, B \in \mathcal{D}_{l+m}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

where $\mathcal{D}_l = \sigma(Z_i, i \leq l)$ and $\mathcal{D}_{l+m} = \sigma(Z_i, i \geq l+m)$ are the σ-fields generated by $(Z_i;\ i \leq l)$ and $(Z_i;\ i \geq l+m)$ respectively, for any $m \geq 1$. The stationary sequence $Z = (Z_i;\ i \in \mathbb{N}^+)$ is said to be α-mixing if $\alpha_Z(m) \to 0$ as $m \to \infty$. Among various mixing conditions used in the literature, α-mixing is reasonably weak (see, e.g., Doukhan, 1994).

Since in the subsequent development we are dealing with wavelet decompositions, for each $i \in \mathbb{N}^+$, denote by $\Xi_i = \{\xi_i^{(J,k)} :\ k = 0, 1, \ldots, 2^J - 1\}$ the set of scaling coefficients up to resolution level J of the i-th segment Z_i. Note that because $Z = (Z_i;\ i \in \mathbb{N}^+)$ is a strictly stationary stochastic process, the same is also true for the 2^J-dimensional stochastic process $\{\Xi_i;\ i \in \mathbb{N}^+\}$. Furthermore, denote by $\mathcal{A}_{J,l} = \sigma(\xi_i^{J,k}, i \leq l)$ and $\mathcal{A}_{J,l+m} = \sigma(\xi_i^{J,k}, i \geq l+m)$ the σ-fields generated by $(\xi_i^{(J,k)};\ i \leq l)$ and $(\xi_i^{(J,k)};\ i \geq l+m)$ respectively. Because $\sigma(\xi_i^{(J,k)}, i \in I) \subseteq \sigma(Z_i, i \in I)$ for any $I \subset \mathbb{N}^+$, we get

$$\begin{aligned} \alpha_{J,k}(m) &= \sup_{A \in \mathcal{A}_{J,l}, B \in \mathcal{A}_{J,l+m}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \\ &\leq \sup_{A \in \mathcal{D}_{J,l}, B \in \mathcal{D}_{J,l+m}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \\ &= \alpha_Z(m). \end{aligned}$$

Note that the above observation is also true when dealing with sample paths discretized over a fixed equidistant grid on $[0, \delta]$ of size P, since then $\xi_i^{(J,k)} = Z_i(t_k)$ for all $k = 1, 2, \ldots, P$.

3 A Functional Wavelet-kernel prediction

3.1 Finite-dimensional Kernel Prediction

Consider the nonparametric prediction of a (real-valued) stationary discrete-time stochastic process. Let $X_{n,(r)} = (X_n, X_{n-1}, \ldots, X_{n-r+1}) \in \mathbb{R}^r$ be the vector of lagged variables, and let s be the forecast horizon. It is well-known that the autoregression function plays a predominant forecasting role in the above time series context. Recall that the autoregression function f is defined by

$$f(\mathbf{x}) = \mathbb{E}(X_{n+s} \mid X_{n,(r)} = \mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^r.$$

It is clear that the first task is to estimate f. The classical approach to this problem is to find some parametric estimate of f. More specifically, it is assumed that f belongs to a class of functions, only depending on a finite number of parameters to be estimated. This is the case of the very well-known ARIMA models, widely studied in the literature (see, e.g., Box & Jenkins, 1976; Brockwell & Davis, 1991).

The above prediction problem can also be undertaken with a nonparametric view, without any assumption on the functional form of f. This is a much more flexible approach that only imposes regularity conditions on f. Nonparametric methods for forecasting in time series can be viewed, up to a certain extent, as a particular case of nonparametric regression estimation under dependence (see, e.g., Bosq, 1991; Hart, 1991; Härdle & Vieu, 1992; Hart, 1996). A popular nonparametric method for such task is to use the kernel smoothing ideas because they have good properties in real-valued regression problems both from a theoretical and a computational point of view. The kernel estimator $\hat{f}_n$ (based on $X_1, \ldots, X_n$) of f is defined by

$$\hat{f}_n(\mathbf{x}) = \frac{\sum_{t=r}^{n-s} \mathbb{K}((\mathbf{x} - X_{n,(r)})/h_n) X_{t+s}}{\sum_{t=r}^{n-s} \mathbb{K}((\mathbf{x} - X_{n,(r)})/h_n)},$$

or 0 if the denominator is null. In our development, for simplicity, we consider a product kernel, i.e., for each $\mathbf{x} = (x_1, \ldots, x_r)$, $\mathbb{K}(\mathbf{x}) = \prod_{i=1}^{r} K(x_i)$; also h_n is a sequence of positive numbers (the bandwidths). The s-ahead prediction is then simply given by $X_{n+s\,|\,n} = \hat{f}_n(X_{n,(r)})$. Theoretical results show that the detailed choice of the kernel function does not influence strongly the asymptotic behavior of prediction but the choice of the bandwidth values are crucial for prediction accuracy (see, e.g., Bosq, 1998).

As it is readily seen, the prediction is expressed as a locally weighted mean of past values, where the weights measure the similarity between $(X_{t,(r)};\ t = r, \ldots, n-s)$ and $X_{n,(r)}$, taking into account the process history. Let now $\|\cdot\|$ be a generic notation for a Euclidean norm. If the kernel values decrease to zero as $\|\mathbf{x}\|$ increases, the smoothing weights have high values when the $(X_{t,(r)})$ is close to $X_{n,(r)}$, and is close to zero otherwise. In other words, the prediction $X_{n+s\,|\,n}$ is obtained as a locally weighted average of future blocks of horizon s in all blocks of length r in the past, weighted by similarity coefficients $w_{n,t}$ of these blocks with the current block, where

$$w_{n,t}(\mathbf{x}) = \frac{\mathbb{K}((\mathbf{x} - X_{t,(r)})/h_n)}{\sum_{m=r}^{n-s} \mathbb{K}((\mathbf{x} - X_{m,(r)})/h_n)}.$$

3.2 Functional Wavelet-Kernel Prediction

Recall that, in our setting, the strictly stationary time series $Z = (Z_i;\ i \in \mathbb{N}^+)$ is functional-valued rather than $\mathbb{R}$-valued, i.e., each Z_i is a random curve. In this functional setup, and to simplify notation, we address, without loss of generality, the prediction problem for a horizon $s = 1$. We

could mimic the above kernel regression ideas, and use the following estimate

$$Z_{n+1\,|\,n}(\cdot) = \sum_{m=1}^{n-1} w_{n,m} Z_{m+1}(\cdot), \tag{2}$$

where the triangular-array of local weights $\{w_{n,m} : \ m = 1, 2, \ldots, n-1; n \in \mathbb{N}^+\}$ increases with the closeness or similarity of the last observed path Z_n and the paths Z_m in the past, in a (semi-)norm sense; this is made more precise in (3) below. The literature on this infinite-dimension kernel regression related topic is relatively limited, to our knowledge. Bosq & Delecroix (1985) dealt with general kernel predictors for Hilbert-valued stationary Markovian processes. A similar idea was applied by Besse, Cardot & Stephenson (1999) for ARH(1) processes in the special case of a Sobolev space. Extending and justifying these kernel regression techniques to infinite-dimensional stochastic processes with no specific structures (e.g., ARH(1) or more general Markovian processes), will require using measure-theoretic assumptions on infinite-dimensional spaces (e.g., a probability density function with respect to an invariant measure) thus restricting the analysis and applicability to a small class of stochastic processes (e.g., diffusion processes). Such kind of assumptions are more natural in finite-dimensional spaces such as those obtained through orthonormal wavelet decompositions, especially when the discretized sample paths of the observed process are quite coarse. Taking advantage of these latter remarks, our forecasting methodology relies upon a wavelet decomposition of the observed curves, and uses the concepts of strictly stationarity and α-mixing within the stochastic multiresolution analysis framework discussed in Section 2. Moreover, note that using distributional assumptions such as those given in the appendix on the wavelet coefficients is much less restrictive than using similar assumptions on the original process Z.

To summarize, the proposed forecasting methodology is decomposed into two phases

Ph1: find within the past paths the ones that are 'similar' to the last observed path (this determines the weights);

Ph2: use the weights and the stochastic multiresoltion analysis to forecast by a locally weighted averaging process as the one described by (2).

Since we are dealing with a wavelet decomposition, it is worth to isolate the first phase (Ph1) by discussing possible ways to measure the similarity of two curves, by means of their wavelet approximation, and then to proceed to the second phase (Ph2), using again this wavelet approximation. The analysis of the proposed kernel-based functional prediction method is based on finding similar paths. Similarity is now defined in terms of a distance metric related to the functional space in which the sample paths lie. When the space is a Besov space, it is well-known that its norm is characterized by a weighted ℓ_p-norm of the wavelet coefficients of its elements (see, e.g., Meyer, 1992). It is therefore natural to address the similarity issue on the wavelet decomposition

of the observed sample paths. The wavelet transform is applied to the observed sample paths, and due to the approximation properties of the wavelet transform, only a few coefficients of the transformed data will be used; a kind of contractive property of the wavelet transform.

Applying the DWT to each path, decomposes the temporal information of the time series into discrete wavelet coefficients associated with both time and scale. Discarding scales in the DWT that are associated with high-frequency oscillations, provides a straightforward data reduction step and decreases the computational burden. We want to use the distributional properties of the wavelet coefficients of the observed series. Imagine first that we are given 2 observed series, and let $\theta_{j,k}^{(i)}$, $i = 1, 2$, be the discrete wavelet coefficient of the DWT of each signal at scale j ($j = j_0, \ldots, J-1$) and location k ($k = 0, 1, \ldots, 2^j - 1$). At each scale $j \geq j_0$, define a measure of discrepancy in terms of a distance

$$d_j\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}\right) = \left\{ \sum_{k=0}^{2^j-1} \left(\theta_{jk}^{(1)} - \theta_{j,k}^{(2)}\right)^2 \right\}^{1/2},$$

which measures how effectively the two signals match at scale j. To combine all scales, we then use

$$D\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}\right) = \sum_{j=j_0}^{J-1} 2^{-j} d_j\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}\right).$$

Such a measure of discrepancy is natural and is often used to test the equality of two regression functions in the wavelet domain (see, e.g., Spokoiny, 1996; Abramovich, Antoniadis, Sapatinas & Vidakovic, 2004).

As for the second phase (Ph2), recall that, for each $i \in \mathbb{N}^+$, $\Xi_i = \{\xi_i^{(J,k)} : \ k = 0, 1, \ldots, 2^J - 1\}$ denotes the set of scaling coefficients up to resolution level J of the i-th segment Z_i. The kernel prediction of the scaling coefficients at time $n+1$, $\Xi_{n+1\,|\,n}$, is given by

$$\Xi_{n+1\,|\,n} = \frac{\sum_{m=1}^{n-1} K(D(\mathcal{C}(\Xi_n), \mathcal{C}(\Xi_m))/h_n)\Xi_{m+1}}{1/n + \sum_{m=1}^{n-1} K(D(\mathcal{C}(\Xi_n), \mathcal{C}(\Xi_m))/h_n)}, \tag{3}$$

where the $1/n$ factor in the denominator allows expression (3) to be properly defined and does not affect its convergence rate. Here, for simplicity, we use the notation $D(x, y)/h_n = D(x/h_n, y/h_n)$, and $\mathcal{C}(\Xi_k)$ is the set of wavelet coefficients obtained by applying the "pyramid algorithm" (see Mallat, 1989) on the set of (finest level) scaling coefficients Ξ_k, for $k = 1, 2, \ldots, n$. This, leads to the time-domain prediction at time $n+1$,

$$Z_{n+1\,|\,n}^J(t) = \sum_{k=0}^{2^J-1} \xi_{n+1\,|\,n}^{(J,k)} \phi_{J,k}(t), \quad \forall\, t \in [0, \delta),$$

of $\mathbb{E}(Z_{n+1}(\cdot)\,|\,Z_n(\cdot))$, where $\xi_{n+1\,|\,n}^{(J,k)}$ are the components of the predicted scaling coefficients $\Xi_{n+1\,|\,n}$. The following theorem shows its consistency property.

Theorem 3.1 *Suppose that the Assumptions (A1)-(A7), given in the Appendix, are true.*

(i) If $t \in \{0, 1, \ldots, 2^J - 1\}$ *and if* $h_n = \left(\frac{\log n}{n}\right)^{1/(2+2^J)}$, *then, as* $n \to \infty$, *we have*

$$\left| Z^J_{n+1\,|\,n}(t) \;-\; \mathbb{E}(Z_{n+1}(t)\,|\,Z_n) \right| = O\left(2^{J/2} \left(\frac{\log n}{n}\right)^{1/(2+2^J)} \right), \quad \textit{almost surely.}$$

(ii) If the sample paths are sampled on a grid of size 2^J *and if* $h_n = \left(\frac{\log n}{n}\right)^{1/(2+2^J)}$, *then, as* $n \to \infty$, *we have*

$$\left| Z^J_{n+1\,|\,n}(t) \;-\; \mathbb{E}(Z_{n+1}(t)\,|\,Z_n) \right| = O\left(2^{J/2} \left(\frac{\log n}{n}\right)^{1/(2+2^J)} + 2^{-sJ} \right), \quad \textit{almost surely.}$$

Remark 3.1 Note that in both assertions of Theorem 3.1, the size of the sampling grid over each segment affects the convergence rate of our predictor. In the first case, which is the most usual in practice, the rate becomes slower as the dimension P increases but we still have consistency as the number of segments increases to infinity. In the second case, however, an extra term is given by the wavelet approximation of the sample paths at resolution J and getting a larger J affects considerably the rate of the estimator which is the well-known problem of the "curse of dimensionality". One possible way to deal with this problem would be to look at the regressor in an infinite-dimensional space but, as already noted above, this is a difficult problem, since one would need some concentration assumption about the distribution of the functional-valued time series $Z = (Z_i;\ i \in \mathbb{N}^+)$ without referring to any particular probability density function.

We conclude this section by pointing out that, as in any nonparametric smoothing approach, the choice of the smoothing parameter h_n (the bandwidth) is of great importance. Once h_n is specified, only time segments that lie within a similarity distance from the segment Z_n within h_n will be used to estimate the prediction at time $n+1$. Intuitively, a large value of h_n will lead to an estimator that incurs large bias, while a small value, might reduce the bias but the variability of the predicted curve could be large since only few segments are used in the estimation. A good choice of h_n should balance the bias-variance trade off. In our implementation, we have used the leave-one out cross-validation for times-series data as suggested by Hart (1996). The principle of the cross-validation criterion is to select the bandwidth which, for our given prediction horizon $s = 1$, minimizes the mean squared prediction errors of the $(i+1)$-th segment using all segments in the past except the i-th, i.e., the value of h_n that minimizes

$$CV(h) = \frac{1}{n-1} \sum_{i=1}^{n-1} \left\| Z_{i+1} - Z^{(i)}_{i+1\,|\,i} \right\|^2,$$

where $Z^{(i)}_{i+1\,|\,i}$ is the kernel regression estimate with bandwidth h obtained using the series without its i-th segment. This is the method for choosing the bandwidth adopted in the numerical results presented in Section 5.

4 Resampling-based Pointwise Prediction Intervals

Apart from the prediction $Z^J_{n+1\,|\,n}(t)$ discussed in Section 3, we also construct resampling-based pointwise prediction intervals for $Z_{n+1}(t)$. In particular, suppose that $Z_n(t)$ is observed at the set $0 \le t_1 < t_2 < \ldots < t_P \le \delta$ of discrete points on the interval $[0, \delta]$. A pointwise prediction interval for $Z_{n+1}(t)$ is defined to be a set of lower and upper points $L_{n+1,\alpha}(t_i)$ and $U_{n+1,\alpha}(t_i)$ respectively, such that for every $t_i, i = 1, 2, \ldots, P$, and every $\alpha \in (0, 1)$,

$$\mathbb{P}\left(L_{n+1,\alpha}(t_i) \le Z_{n+1}(t_i) \le U_{n+1,\alpha}(t_i)\right) \ge 1 - 2\alpha.$$

Note that since we are looking at the one step prediction of $Z_{n+1}(t)$ given Z_n, we are in fact interested in the conditional distribution of $Z_{n+1}(t)$ given Z_n, i.e., $L_{n+1,\alpha}(t_i)$ and $U_{n+1,\alpha}(t_i)$ are the lower and upper α-percentage points of the conditional distribution of $Z_{n+1}(t_i)$ given Z_n.

To construct such a prediction interval the following simple resampling procedure is proposed. Given Z_n, i.e., given $C(\Xi_n)$, define the weights

$$w_{n,m} = \frac{K(D(\mathcal{C}(\Xi_n), \mathcal{C}(\Xi_m))/h_n)}{n^{-1} + \sum_{m=1}^{n-1} K(D(\mathcal{C}(\Xi_n), \mathcal{C}(\Xi_m))/h_n)} + \frac{1}{1 + n\sum_{m=1}^{n-1} K(D(\mathcal{C}(\Xi_n), \mathcal{C}(\Xi_m))/h_n))}.$$

Note that the weights have been selected appropriately so that

$$0 \le w_{n,m} \le 1 \quad \text{and} \quad \sum_{m=1}^{n-1} w_{n,m} = 1.$$

Now, given Z_n, generate pseudo-realizations $Z^*_{n+1}(t)$ such that for $m = 1, 2, \ldots, n-1$,

$$\mathbb{P}\left(Z^*_{n+1}(t) = Z_{m+1}(t)\,|\,Z_n\right) = w_{n,m},$$

i.e., $Z^*_{n+1}(t)$ is generated by choosing randomly a segment from the whole set of observed segments $Z_{m+1}(t)$, where the probability that the $(m+1)$-th segment is chosen depends on how 'similar' is the preceding segment Z_m to Z_n. This 'similarity' is measured by the resampling probability $w_{n,m}$.

Given pseudo-replicates $Z^*_{n+1}(t)$, calculate $R^*_{n+1}(t_i) = Z^*_{n+1}(t_i) - Z^J_{n+1|\,n}(t_i)$, where $Z^J_{n+1|\,n}(t_i)$ is our time-domain conditional mean predictor. Let $R^*_{n+1,\alpha}(t_i)$ and $R^*_{n+1,1-\alpha}(t_i)$ be the lower and upper α-percentage points of $R^*_{n+1}(t_i)$. Note that these percentage points can be consistently estimated by the corresponding empirical percentage points over B realizations $Z^{*(b)}_{n+1}(t_i)$, $b = 1, 2, \ldots, B$, of $Z^*_{n+1}(t_i)$. A $(1-\alpha)100\%$ pointwise prediction interval for $Z_{n+1}(t_i)$ is then obtained by

$$\{[\ L^*_{n+1,\alpha}(t_i),\ U^*_{n+1,\alpha}(t_i)\ ],\quad i = 1, 2, \ldots, P\},$$

where $L^*_{n+1,\alpha}(t_i) = R^*_{n+1,\alpha}(t_i) + Z^J_{n+1|\,n}(t_i)$ and $U^*_{n+1,\alpha}(t_i) = R^*_{n+1,1-\alpha}(t_i) + Z^J_{n+1|\,n}(t_i)$.

The following theorem shows that the proposed method is asymptotically valid, i.e., the so-constructed resampling-based prediction interval achieves the desired pointwise coverage probability.

Theorem 4.1 *Suppose that the Assumptions (A1)-(A7), given in the Appendix, are true. Then, for every* $i = 1, 2, \ldots, P$, *and every* $\alpha \in (0, 1)$

$$\lim_{n\to\infty} \mathbb{P}\left(L^*_{n+1,\alpha}(t_i) \leq Z_{n+1}(t_i) \leq U^*_{n+1,\alpha}(t_i) \,|\, Z_1, \ldots, Z_n\right) \geq 1 - 2\alpha.$$

5 Applications to real-life datasets

We now illustrate the usefulness of the proposed functional wavelet-kernel (W-K) approach for continuous-time prediction in finite sample situations by means of three real-life datasets that were collected from different arenas, in particular with (*a*) the prediction of the entire annual cycle of climatological El Niño-Southern Oscillation time series one-year ahead from monthly recordings, (*b*) the one-week ahead prediction of Paris electrical load consumption from half-hour daily recordings, and (*c*) the one-year ahead prediction of the Nottingham temperature data from monthly recordings.

For the W-K approach, the interpolating wavelet transform of Donoho (1992) based on *Symmlet 6* (see Daubechies, 1992, p. 195) was used. Preliminary simulations show that the analysis is robust with respect to the wavelet filter, e.g., using *Coiflet 3* (see Daubechies, 1992, p. 258). In the case where the number of time points (P) in each segment is not a power of 2, each segment is extended by periodicity at the right to a length closest to the nearest power of 2. The Gaussian kernel (K) was adopted in our analysis. Again, preliminary simulations show that our analysis is robust with respect to kernels with unbounded support (e.g., Laplace). The bandwidth (h_n) was chosen by the leave-one out cross-validation for times-series data as suggested by Hart (1996). For the associated 95% resampling-based pointwise prediction intervals, the number of resampling samples (B) was taken equal to 500.

We compare the resulting predictions with those obtained by some well-established methods in the literature, in particular with a smoothing spline (SS) method and with the classical SARIMA model. The SS method, introduced by Besse & Cardot (1996), assumes an ARH(1) structure for the time series $Z = (Z_i;\ i \in \mathbb{N}^+)$ and handles the discretization problem of the observed curves by simultaneously estimating the sample paths and projecting the data on a q-dimensional subspace (that the predictable part of Z assumed to belong) using smoothing splines (by solving an appropriate variational problem). The corresponding smoothing parameter (λ) and dimensionality (q) are chosen by a cross-validation criterion. Following the Box-Jenkins methodology (see Box

& Jenkins, 1976, Chapter 9), a suitable SARIMA model is also adjusted to the times series $Z = (Z_i;\ i \in \mathbb{N}^+)$.

The quality of the prediction methods was measured by the *relative mean-absolute error* (RMAE) defined by

$$\text{RMAE} = \frac{1}{P}\sum_{t=1}^{P}\frac{|\hat{Z}_{n_0}(t_i) - Z_{n_0}(t_i)|}{|Z_{n_0}(t_i)|}, \tag{4}$$

where Z_{n_0} is the n_0-th element of the time series Z and $\hat{Z}_{n_0}$ is the prediction of Z_{n_0} given the past.

The computational algorithms related to wavelet analysis were performed using Version 8.02 of the freeware WaveLab software. The entire numerical 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 extra-tropical variability over both Hemispheres, is related to ENSO. For a detailed review of ENSO the reader is referred, for example, to Philander (1990).

An useful index of El Niño variability is provided by the sea surface temperatures averaged over the Niño-3 domain (5^oS $-$ 5^oN, 150^oW $-$ 90^oW). Monthly mean values have been obtained from January 1950 to December 1996 from gridded analyses made at the U.S. National Centers for Environmental Prediction (see Smith, Reynolds, Livezey & Stokes, 1996). The time series of this EN index is depicted in Figure 5.1, and shows marked inter-annual variations superimposed on a strong seasonal component. It has been analyzed by many authors (see, for example, Besse, Cardot & Stephenson, 2000; Antoniadis & Sapatinas, 2003).

The bandwidth (h_n) for the W-K method was chosen by a cross-validation criterion and found equal to 0.11. We have compared our results with those obtained by Besse, Cardot & Stephenson (2000), using the SS method, with smoothing parameter (λ) and dimensionality (q) chosen optimally by a cross-validation criterion and found equal to 1.6×10^{-5} and 4, respectively. 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, and the most parsimonious SARIMA model, validated through a portmanteau test for serial correlation of the fitted residuals, was selected.

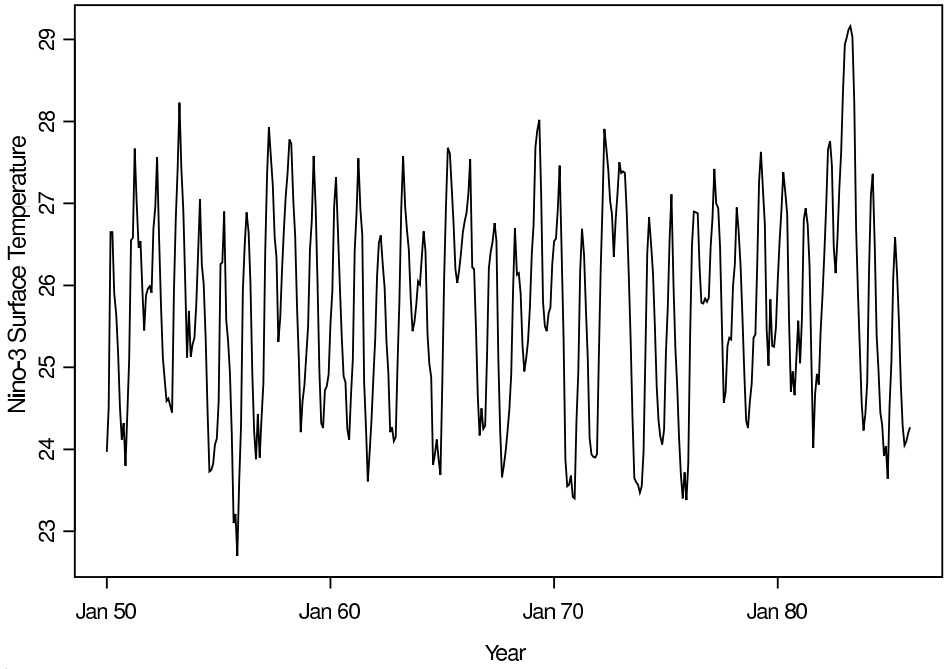

Figure 5.1: The monthly mean Niño-3 surface temperature index in (deg C) which provides a contracted description of ENSO.

Figure 5.2 displays the observed data of the 37th year (1986) and its predictions obtained by the W-K, SS and SARIMA methods. The RMAE of each prediction method are displayed in Table 1 (we have taken $n_0 = 37$ and $P = 12$). As observed in both the figure and the table, the W-K and SS estimators give almost similar predictions, both visually and in terms of RMAE. The prediction obtained by the SARIMA model is strongly and uniformly biased, failing thus to produce an adequate prediction (see Besse, Cardot & Stephenson, 2000, for an explanation). Note that, after May, all predictions are not very close to the true points and this difficulty in prediction is captured in Figure 5.3 which displays the corresponding 95% resampling-based pointwise prediction interval for the Niño-3 surface temperature during 1986, based on the corresponding prediction obtained by the W-K method. As observed in the figure, it becomes clear that as one moves from May onwards, this interval gets larger.

5.2 Paris Electrical Load Consumption

This application concerns with the one-week ahead prediction of Paris electrical load consumption from half-hour daily recordings. The short-term predictions are based on data sampled over 30 minutes, obtained after eliminating certain components linked to weather conditions, calendar

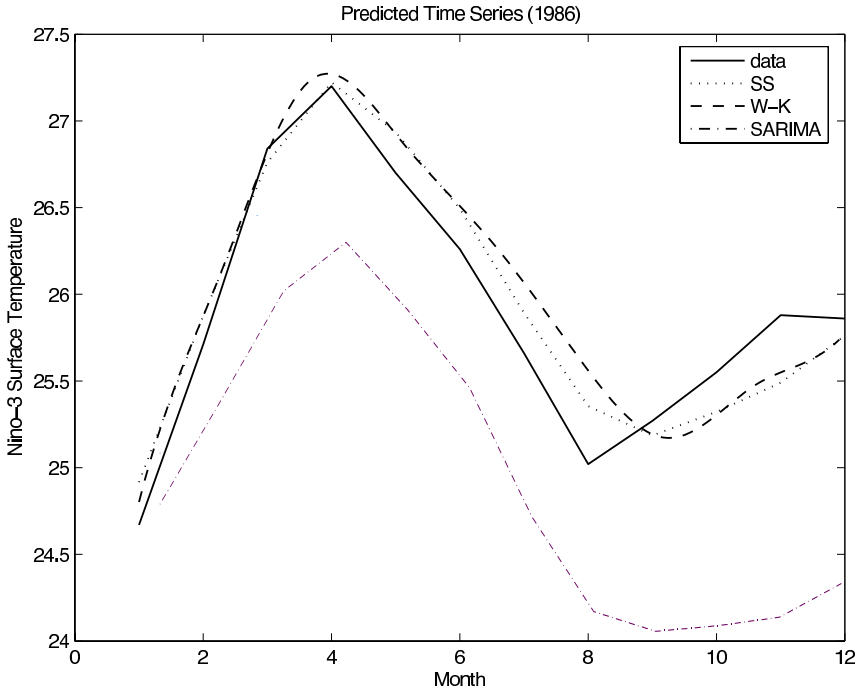

Figure 5.2: The Niño-3 surface temperature during 1986 (—) and its various predictions using the W-K (- - -), SS ($\cdots$), and SARIMA (- $\cdot\cdot$ -) methods.

Prediction Method	**RMAE**
W-K	0.86 %
SS	0.76 %
SARIMA	3.72 %

Table 1: RMAE for the prediction of Niño-3 surface temperatures during 1986 based on the W-K, SS and SARIMA methods.

effects, outliers and known external actions. The dataset analyzed is part of a larger series recorded from the French national electricity company (EDF) during the period running from the 1st of August 1985 to the 4th of July 1992. The time period that we have analyzed runs 35 days, starting from the 24th of July 1991 to the 27th of August 1991, and it is displayed in Figure 5.4. One may note quite a regularity in this time series and a marked periodicity of 7 days (linked to economic rhythms) together with a pseudo-daily periodicity. However, daily consumption patterns due to holidays, weekends and discounts in electricity charges (e.g., relay-switched water heaters to benefit from special night rates), make the use of SARIMA modelling for forecasting problematic for about

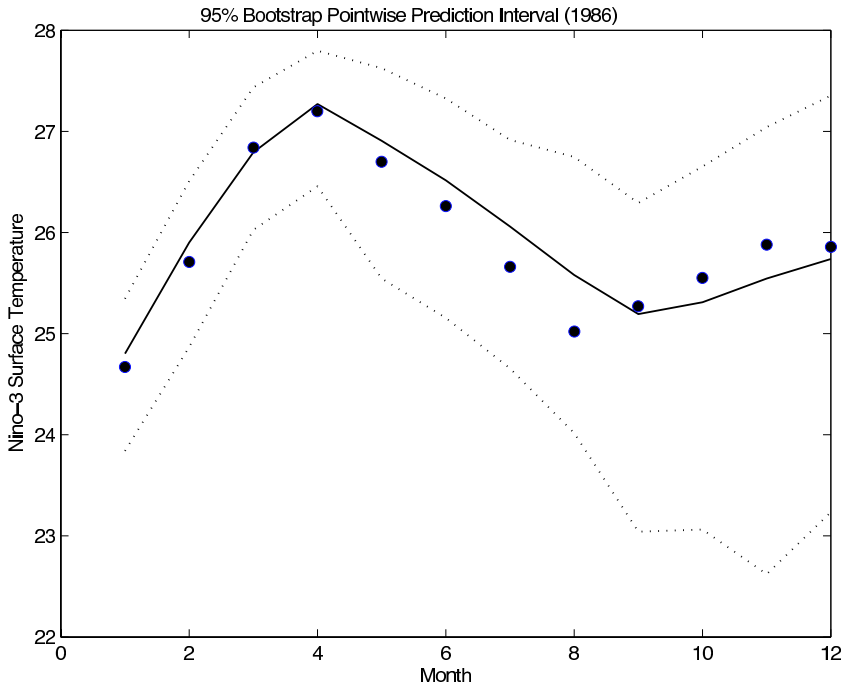

Figure 5.3: 95% resampling-based pointwise prediction interval ($\cdots$) for the Niño-3 surface temperature during 1986, based on the corresponding prediction obtained by the W-K (—) method. The true points ($\bullet$) are also displayed.

10% of the days when working with half-hour data (see Misiti, Misiti, Oppenheim & Poggi, 1994).

The bandwidth (h_n) for the W-K method was chosen by cross-validation and found equal to 0.01. We have compared our results with those obtained using the SS method, with smoothing parameter (λ) and dimensionality (q) chosen by cross-validation and found equal to 10^{-4} and 4, respectively. Figure 5.5 displays the observed data of the 2th August 1991 and its predictions obtained only by the W-K and SS methods. The RMAE of both prediction methods are displayed in Table 2 (we have taken $n_0 = 35$ and $P = 48$). As observed in both the figure and the table, the prediction obtained by the W-K method is reasonably close to the true points, while the prediction made by the SS method falls far off from them. This example, clearly illustrates the impact of the proposed wavelet-based approach since the trajectory to be predicted seems not regular with some peculiar peaks. On the other hand, the smoothing spline-based approach relies upon more stringent smoothness assumptions and the corresponding prediction is therefore largely biased, falling well off the true points. Figure 5.6 displays the 95% resampling-based pointwise prediction interval for the half-hour electricity load consumption in Paris during 27th of August 1991, based on the corresponding prediction obtained by the W-K method.

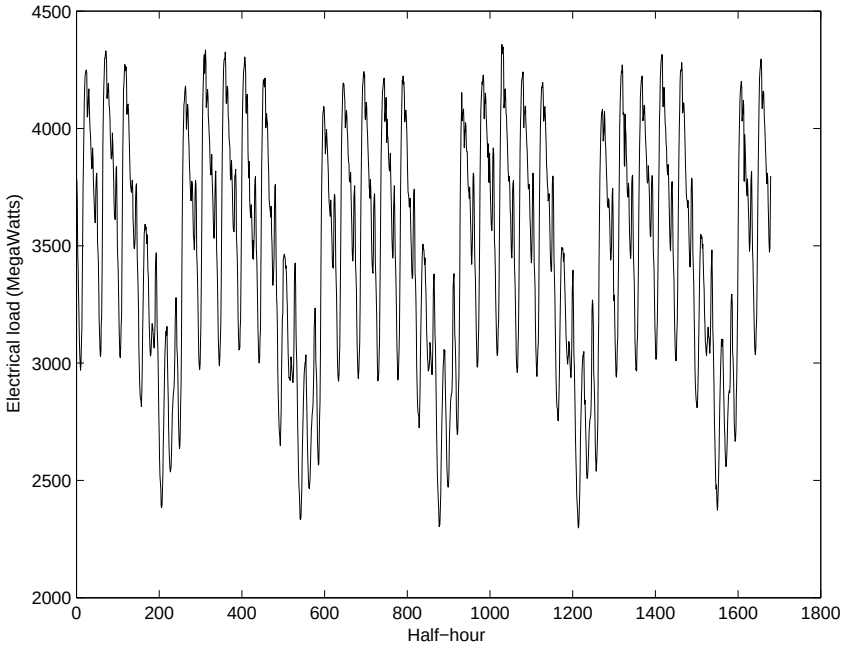

Figure 5.4: The half-hour electricity load consumption in Paris from the 24th of July 1991 to the 27th of August 1991.

Prediction Method	**RMAE**
W-K	12 %
SS	36 %

Table 2: RMAE for the prediction of half-hour electricity load consumption in Paris during 27th of August 1991 based on the W-K and SS methods.

5.3 Nottingham Temperature Data

This application concerns with the one-year ahead prediction of the Nottingham temperature data from monthly recordings. The dataset analyzed are mean monthly air temperatures (°F) at Nottingham castle from January 1920 to December 1939, from 'Meteorology of Nottingham', in *City Engineer Surveyor*. Since February 1929 was an exceptionally cold month in England, and since the seasonal pattern is fairly stable over time, the original dataset has been 'corrected'. In other words, since this 'outlier' will distort the fitting process, the value of February 1929 was altered it to a low value for February of 35°F (see Venables & Ripley, 1999, Chapter 13). This 'corrected' dataset, which is the series `nottem` in the `MASS` library of S-PLUS, is the one analyzed

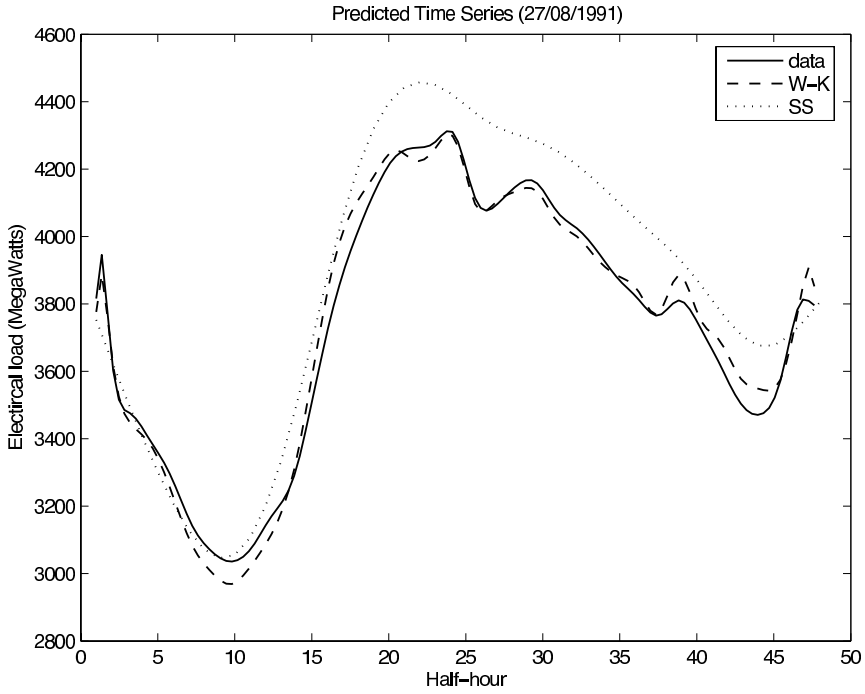

Figure 5.5: The half-hour electricity load consumption in Paris during 27th of August 1991 (—) and its various predictions using the W-K (- - -) and SS ($\cdots$) methods.

below.

The bandwidth (h_n) for the W-K method was chosen by cross-validation and found equal to 2.5. We have compared our results with those obtained using the SS method, with smoothing parameter (λ) and dimensionality (q) chosen by cross-validation and found equal to 10^{-4} and 1, respectively. To complete the comparison, a suitable ARIMA model, including 12 month seasonality, has also been adjusted to the times series from January 1920 to December 1938, and the most parsimonious SARIMA model, validated through a portmanteau test for serial correlation of the fitted residuals, was selected. Figure 5.8 displays the observed data of the year 1939 and its various predictions obtained by the W-K, SS and SARIMA methods. It is well-known in the literature (see, e.g., Venables & Ripley, 1999, Chapter 13) that this time series can be adequately predicted by a parametric model (SARIMA), but it is evident that both nonparametric models (W-K and SS) perform equally-well. The RMAE of each prediction method is displayed in Table 3 (we have taken $n_0 = 20$ and $P = 12$). As observed in both the figure and the table, before March and after October, all predictions are largely biased and hence not very close to the true points. This difficulty in prediction is captured in Figure 5.6 which displays the corresponding 95% resampling-based pointwise prediction interval for the mean monthly air temperatures (^{o}F) at Nottingham castle

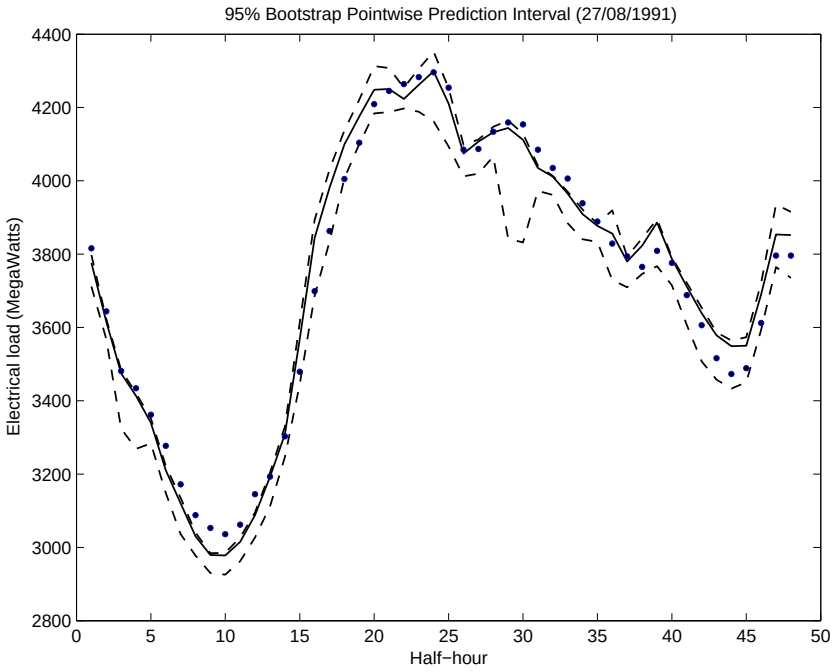

Figure 5.6: 95% resampling-based pointwise prediction interval (- - -) for the half-hour electricity load consumption in Paris during 27th of August 1991, based on the corresponding prediction obtained by the W-K (—) method. The true points (•) are also displayed.

Prediction Method	**RMAE**
W-K	30 %
SS	28 %
SARIMA	31 %

Table 3: RMAE for the prediction of half-hour electricity load consumption in Paris during 27th of August 1991 based on the W-K and SS methods.

during 1939, based on the corresponding prediction obtained by the W-K method. As observed in the figure, it becomes clear that as one moves from March backwards and from October onwards, this interval gets larger.

6 Conclusions

The functional wavelet-kernel prediction methodology, of a continuous-time stochastic process on an entire time-interval in terms of its recent past, developed in this paper exhibits very good

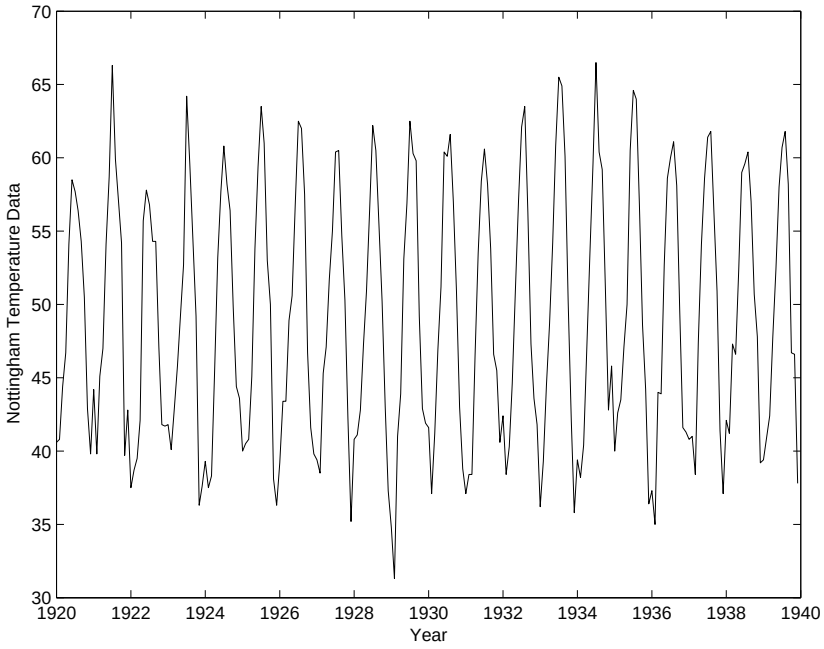

Figure 5.7: The mean monthly air temperatures (°F) at Nottingham castle from January 1920 to December 1939.

performance with respect to other well-known parametric and nonparametric techniques. As it is demonstrated in the real-life datasets analyzed, the proposed wavelet-based prediction methodology outperforms the smoothing spline-based prediction methodology for stochastic processes with inhomogeneous sample paths, and performs equally-well for stochastic processes with quite regular sample paths. Moreover, it performs reasonably well in situations where the classical seasonal parametric model exhibits very good predictions. We have, however, noted than when the number of sampling points within each time-segment is large, the curse of dimensionality leads to inefficiency of the proposed nonparametric prediction method unless the number of segments is really large.

Acknowledgements

Anestis Antoniadis was supported by the 'IAP Research Network P5/24' and the 'Cyprus-France CY-FR/0204/04 Zenon Program'. Efstathios Paparoditis and Theofanis Sapatinas were supported by the 'Cyprus-France CY-FR/0204/04 Zenon Program'. We would like to thank Jean-Michel Poggi (Universite Paris-Sud, France) for providing us with the Paris electrical load consumption data.

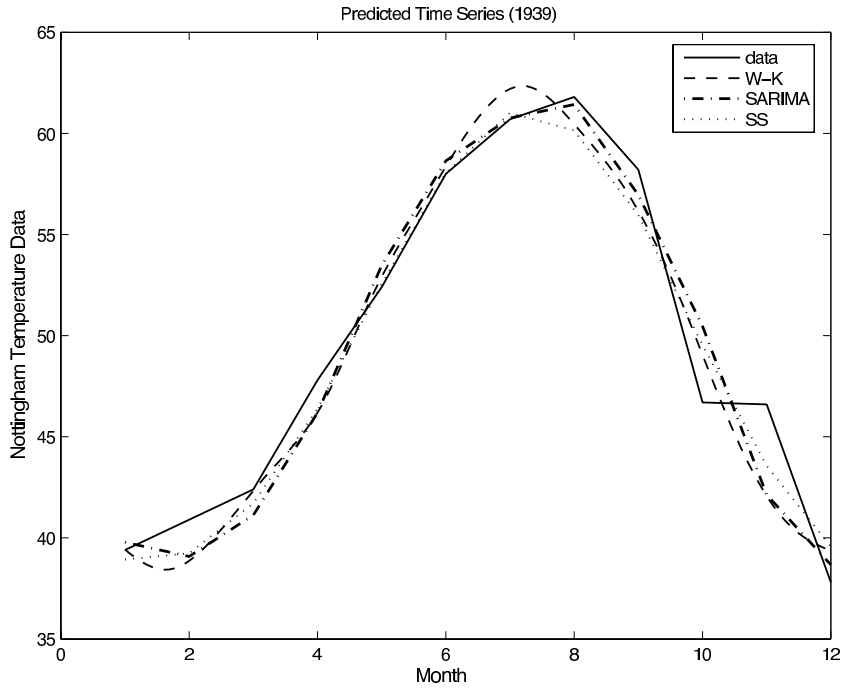

Figure 5.8: The mean monthly air temperatures (°F) at Nottingham castle during 1939 (—) and its various predictions using the W-K (- - -), SS ($\cdots$), and SARIMA (- $\cdot$ -) methods.

Appendix

Our asymptotic results will be based on the following set of assumptions, which we detail below before proceeding to the proofs.

Main Assumptions

We first impose an assumption on the sample paths of the underlying stochastic process.

Assumption (A1): The sample paths of the strictly stationary process $Z = (Z_i;\ i \in \mathbb{N}^+)$ are assumed to lie within a Besov space $B^s_{p,q}$, where $0 < s < r$, $1 \leq p, q \leq \infty$, and r is the regularity of the scaling functions associated with the regular multiresolution analysis discussed in Section 2.3.

Assumption (A2): When we only observe a fixed number P of samples values from each sample path, we assume that the sampled paths of the strictly stationary process $Z = (Z_i;\ i \in \mathbb{N}^+)$ are continuous on $[0, \delta)$, and that the interpolating scaling function ϕ of the wavelet interpolating

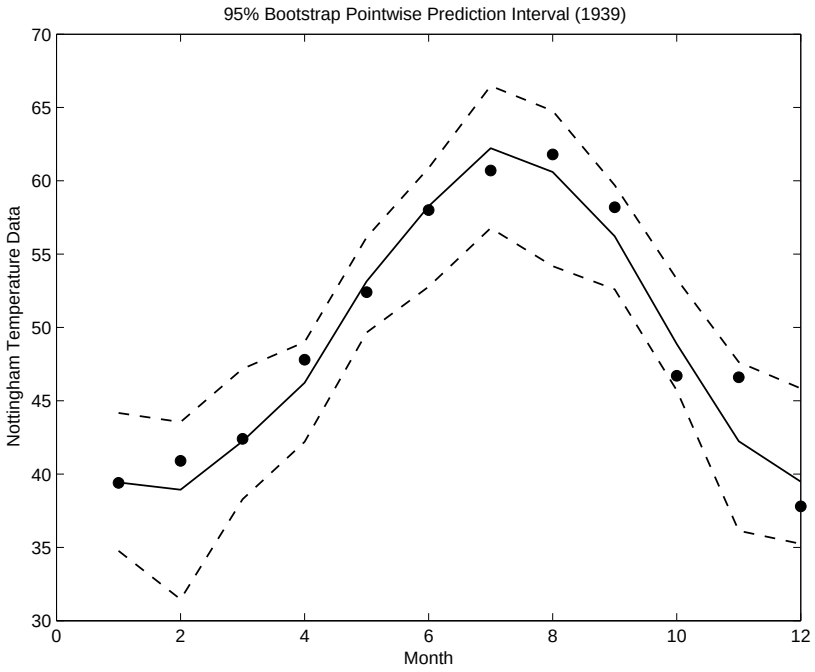

Figure 5.9: 95% resampling-based pointwise prediction interval (- - -) for the mean monthly air temperatures (°F) at Nottingham castle during 1939, based on the corresponding prediction obtained by the W-K (—) method. The true points (•) are also displayed.

basis has an exponential decay.

Assumption (A3): The α_Z-mixing coefficient of the strictly stationary process $Z = (Z_i;\ i \in \mathbb{N}^+)$ satisfies

$$\sum_{m=N}^{\infty} \alpha_Z(m)^{1-2/l} = O(N^{-1}) \quad \text{for some} \quad l > 4. \tag{5}$$

(Note that the scaling coefficients of Z inherit the above mixing property, according to the relevant discussion in Section 2.3.)

We next impose some assumptions on the joint and conditional probability density functions of the scaling coefficients $\xi_i^{(J,k)}$.

Assumption (A4): $E|\xi_i^{(J,k)}|^l < \infty$, for $l > 4$ and every $k = 0, 1, \ldots, 2^J - 1$.

Assumption (A5): The joint probability density function $f_{\xi_{i+1}^{(J,k)}, \xi_i^{(J,k)}}$ of $(\xi_{i+1}^{(J,k)}, \xi_i^{(J,k)})$ exist, it is absolute continuous with respect to Lebesgue measure, and it satisfies the following conditions

(*i*) $f_{\xi_{i+1}^{(j,k)},\xi_i^{(j,k)}}$ is Lipschitz continuous, i.e.,

$$\left|f_{\xi_{i+1}^{(J,k)},\xi_i^{(J,k)}}(x_1,x_2)-f_{\xi_{i+1}^{(J,k)},\xi_i^{(J,k)}}(y_1,y_2)\right|\leq C\,\|(x_1,x_2)-(y_1,y_2)\|.$$

(*ii*) The random vector of the scaling coefficients at scale J admits a compactly supported probability density function f (with support S) which is strictly positive and twice continuously differentiable.

(*iii*) The conditional probability density function of $\xi_{i+1}^{(J,k)}$ given $\xi_i^{(J,k)}$ is bounded, i.e., $f_{\xi_{i+1}^{(J,k)}\,|\,\xi_i^{(J,k)}}(\cdot\mid x)\leq C<\infty$.

We also impose some conditions on the kernel function and the bandwidth associated with it.

Assumption (A6): The (univariate) kernel K is a bounded symmetric density on $\mathbb{R}$ satisfying $|K(x)-K(y)|\leq C\,|x-y|$ for all $x,y\in\mathbb{R}$. Furthermore, $\int xK(x)dx=0$ and $\int x^2K(x)dx<\infty$.

Assumption (A7): The bandwidth h satisfies $h\to 0$ and $nh^{2^J}\to\infty$ as $n\to\infty$.

Let us now explain the meaning of the above assumptions. Assumptions (A1) (or (A2)) and (A3) are quite common in times series prediction (see Bosq, 1998). Assumptions (A4)-(A5) are essentially made on the distributional behavior of the scaling coefficients and, therefore, are less restrictive. They are moreover natural in nonparametric regression. Assumption (A5)-(ii) is natural as far as it concerns the scaling coefficients since the decay of the scaling coefficients is ensured by the approximation properties of the corresponding transform. Moreover, it is needed for obtained uniform consistency results. The conditions (A6)-(A7) are classical for kernel regression estimation.

Proof of Theorem 3.1. The difference in the two assertions of the theorem is due to the nature of the observations. In case (*i*), each observed segment is a time series with fixed, finite length, and we use an interpolating wavelet transform at the appropriate resolution J that makes interpolation error negligible, i.e.,

$$\mathbb{E}(\mathcal{P}_J(Z_{n+1})\,|\,Z_n)=\mathbb{E}(Z_{n+1})\,|\,Z_n).$$

In case (*ii*), the observed segments are continuous-time stochastic processes and one can not neglect anymore the approximation error due to the projection of the observed segment onto the scaling space V_J. However, under our assumptions, and according to the results recalled in Section 2, this error is uniformly bounded by a constant times 2^{-sJ}, where s denotes the Besov smoothness index s, i.e.,

$$||\mathbb{E}(\mathcal{P}_J(Z_{n+1})\,|\,Z_n)-\mathbb{E}(Z_{n+1})\,|\,Z_n)||=O\left(2^{-sJ}\right),\tag{6}$$

resulting in a different rate for the second case. Hence, in both cases, we proceed by deriving the appropriate rates for

$$||Z^J_{n+1|n} - \mathbb{E}(\mathcal{P}_J(Z_{n+1}) \,|\, Z_n)||.$$

We first show that, as $n \to \infty$,

$$||\Xi_{n+1|\,n} \;-\; \mathbb{E}(\Xi_{n+1} \,|\, \Xi_n)|| \to 0, \quad \textit{almost surely.} \tag{7}$$

For this, it suffices to show that, for every $k = 0, 1, \ldots, 2^J - 1$, as $n \to \infty$,

$$\xi^{(J,k)}_{n+1|\,n} \to \mathbb{E}(\xi^{(J,k)}_{n+1} \,|\, \xi^{(J,k)}_n), \quad \textit{almost surely.}$$

Let $x \in \mathbb{R}^{2^J}$, let $\Xi_{n+1|\,n}(x)$ be the value of $\Xi_{n+1|\,n}$ in (3) for $\Xi_n = x$, and denote by $\xi^{(J,k)}_{n+1|\,n}(x)$ the k-th component of $\Xi_{n+1|\,n}(x)$. Consider the 2^J-dimensional random variable $W_l = \mathcal{C}(\Xi_l)$, and denote by $f_{\xi^{(J,k)}_{l+1},W_l}$ and f_{W_l} the joint and marginal densities of $(\xi^{(J,k)}_{l+1}, W_l)$ and W_l, respectively. Notice that because of (A4), and the fact that W_l is a linear transformation of Ξ_l, $f_{\xi^{(J,k)}_{l+1}W_l}$ and f_{W_l} exist with respect to Lebesgue measure for every $k = 0, 1, \ldots, 2^J - 1$. Let

$$\widehat{f}_{W_l}(x) = (nh_n^{2^J})^{-1} \sum_{m=1}^{n-1} K(D(x, W_m)/h_n)$$

and notice that $\widehat{f}_{W_l}(x)$ is the kernel estimator of the 2^J-dimensional density $f_{W_l}(x)$. For notational convenience, in what follows, let $\Phi_{n,k}(x) = \mathbb{E}(\xi^{(J,k)}_{n+1} \,|\, \xi^{(J,k)}_n = x)$ and $\hat{g}_{n,k}(x) = (nh_n^{2^J})^{-1} \sum_{m=1}^{n-1} K(D(x, W_m)/h_n)\xi^{(J,k)}_{m+1}$.

We then have

$$\Big(\xi^{(J,k)}_{n+1|\,n}(x) - \mathbb{E}(\xi^{(J,k)}_{n+1} \,|\, \xi^{(J,k)}_n = x)\Big) = \frac{1}{\widehat{f}_{W_l}(x)}\Big\{\hat{g}_{n,k}(x) - \Phi_{n,k}(x) f_{W_l}(x)\Big\} - \frac{\Phi_{n,k}(x)}{\widehat{f}_{W_l}(x)}\Big\{\widehat{f}_{W_l}(x) - f_{W_l}(x)\Big\}. \tag{8}$$

Using now the assumptions of the theorem, the above decomposition, Lemma 2.1 and Theorem 3.2 of Bosq (1998), it follows that, as $n \to \infty$,

$$\sup_{x \in S} \Big|\xi^{(J,k)}_{n+1|\,n}(x) - \mathbb{E}(\xi^{(J,k)}_{n+1} \,|\, \xi^{(J,k)}_n = x)\Big| = O\left(\left(\frac{\log n}{n}\right)^{1/(2+2^J)}\right), \quad \textit{almost surely.} \tag{9}$$

Recalling now that our estimator is defined as

$$Z^J_{n+1|\,n}(t) = \sum_{k=0}^{2^J-1} \xi^{J,k}_{n+1|\,n} \phi_{J,k},$$

using the rate given in expression (9), and the fact that we have used a regular multiresolution analysis, we have, as $n \to \infty$,

$$\begin{aligned} \sup_t |Z^J_{n+1,n}(t) - \mathbb{E}(P_J(Z_{n+1}(t)) \,|\, Z_n)| &= 2^{J/2} \max_k \left|\xi^{J,k}_{n+1|n} - \mathbb{E}(\xi^{(J,k)}_{n+1} \,|\, \xi^{(J,k)}_n)\right| \sup_t \sum_{k=0}^{2^J-1} |\phi(2^J t - k)| \\ &= O\left(2^{J/2}\left(\frac{\log n}{n}\right)^{1/(2+2^J)}\right), \quad \textit{almost surely.} \end{aligned} \tag{10}$$

The above bound (10), together with inequality (6), ensures the validity of both the assertions (i) and (ii). This completes the proof of Theorem 3.1. □

Proof of Theorem 4.1. For every $t_i \in \{t_1, t_2, \ldots, t_P\}$, note that $Z_{n+1}(t_i) = \xi_{n+1}^{(J,i)}$. Since

$$\begin{aligned} L^*_{n+1,\alpha}(t_i) &= R^*_{n+1,\alpha}(t_i) + Z^J_{n+1\,|\,n}(t_i) \\ &= Z^*_{n+1}(t_i) - E(Z_{n+1}(t_i)\,|\,Z_n) + (Z^J_{n+1\,|\,n}(t_i) - E(Z_{n+1}(t_i)\,|\,Z_n)), \end{aligned}$$

and, as $n \to \infty$, $|E(Z_{n+1}(t_i)\,|\,Z_n) - Z^J_{n+1\,|\,n}(t_i)| \to 0$ *in probability*, it suffices to show that the distribution of $Z^*_{n+1}(t_i) - E(Z_{n+1}(t_i)\,|\,Z_n)$ approximates correctly the conditional distribution of $Z_{n+1}(t_i) - E(Z_{n+1}(t_i)\,|\,Z_n)$ given Z_n.

Now, given $Z_n = x$, i.e., given $\Xi_n = \widetilde{x}$, we have

$$\begin{aligned} \mathbb{P}(Z^*_{n+1}(t_i) - E(Z_{n+1}(t_i)\,|\,Z_n = x) \le y) &= \sum_{m=1}^{n-1} 1_{(-\infty,y]}(Z_{m+1}(t_i) - E(Z_{n+1}(t_i)\,|\,Z_n = x))w_{n,m} \\ &= \sum_{m=1}^{n-1} 1_{(-\infty,\widetilde{y}]}(Z_{m+1}(t_i))w_{n,m} \\ &= \sum_{m=1}^{n-1} 1_{(-\infty,\widetilde{y}]}(\xi_{m+1}^{(J,i)})w_{n,m} \\ &= \frac{\sum_{m=1}^{n-1} \mathbf{1}_{(-\infty,\widetilde{y}]}(\xi_{m+1}^{(J,i)})\, K(D(\mathcal{C}(\widetilde{x}), \mathcal{C}(\Xi_m))/h_n)}{n^{-1} + \sum_{m=1}^{n-1} K(D(\mathcal{C}(\widetilde{x}), \mathcal{C}(\Xi_m))/h_n)} \\ &\quad + O(n^{-1}), \end{aligned} \tag{11}$$

where $\widetilde{y} = y + E(Z_{n+1}(t_i)\,|\,Z_n = x)$. Note that (11) is a kernel estimator of the conditional mean $E(\mathbf{1}_{(-\infty,\widetilde{y}]}(\xi_{n+1}^{(J,i)})\,|\,\Xi_n = \widetilde{x}) = \mathbb{P}(\xi_{n+1}^{(J,i)} \le \widetilde{y}\,|\,\Xi_n = \widetilde{x})$, i.e., of the conditional distribution of $\xi_{n+1}^{(J,i)}$ given that $\Xi_n = \widetilde{x}$. Denote now the conditional distribution of $\xi_{n+1}^{(J,i)}$ given Ξ_n by $F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(\cdot\,|\,\Xi_n)$ and its kernel estimator given in (11) by $\widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(\cdot\,|\,\Xi_n)$. Then by the same arguments as in Theorem 3.1 we get that, for every $y \in \mathbb{R}$, as $n \to \infty$,

$$\sup_{x \in S} \left| \widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y\,|\,x) - F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y\,|\,x) \right| \to 0, \quad \textit{in probability.}$$

It remains to show that the above convergence is also uniformly over y. Fix now a x in the support S of Ξ_n, and let $\epsilon > 0$ arbitrary. Since $F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y\,|\,x)$ is continuous we have that, for every $k \in \mathbb{N}$, points $-\infty = y_0 < y_1 < \ldots < y_{k-1} < y_k = \infty$ exist such that $F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x) = i/k$. For $y_{i-1} \le y \le y_i$, and using the monotonicity of $\widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}$ and $\widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}$, we have

$$\begin{aligned} \widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_{i-1}\,|\,x) - F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x) &\le \widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y\,|\,x) - F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y\,|\,x) \\ &\le \widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x) - F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x). \end{aligned}$$

¿From this, we get

$$\left|\widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y\,|\,x) - F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y\,|\,x)\right| \leq \sup_i \left|\widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x) - F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x)\right| + \frac{1}{k},$$

and, therefore,

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y\,|\,x) - F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y\,|\,x)\right| > \epsilon\right) &\leq \mathbb{P}\left(\sup_i \left|\widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x) - F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x)\right| + k^{-1} > \epsilon\right) \\ &\leq \mathbb{P}\left(\sup_i \sup_x \left|\widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x) - F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x)\right| + k^{-1} > \epsilon\right).\end{aligned}$$

Now, chose k large enough such that $1/k < \epsilon/2$. For such a fixed k, and because, for every $y \in \mathbb{R}$, as $n \to \infty$,

$$\sup_x \left|\widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y\,|\,x) - F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y\,|\,x)\right| \to 0, \quad \textit{in probability},$$

we can choose n large enough such that

$$\mathbb{P}\left(\sup_{1\leq i\leq k} \sup_x \left|\widehat{F}_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x) - F_{\xi_{n+1}^{(J,i)}\,|\,\Xi_n}(y_i\,|\,x)\right| > \epsilon/2\right) < \tau,$$

for any desired τ. Since τ is independent on y and x, the desired convergence follows. This completes the proof of Theorem 4.1. □

References

[1] Abramovich, F., Antoniadis, A., Sapatinas, T. & Vidakovic, B. (2004). Optimal testing in a fixed-effects functional analysis of variance model. *Int. J. Wavelets Multiresolut. Inf. Process.*, **2**, 323–349.

[2] Antoniadis, A. & Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.*, **87**, 133–158.

[3] Besse, P.C. & Cardot, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canad. J. Statist.*, **24**, 467–487.

[4] Besse, P.C., Cardot, H. & Stephenson, D.B. (2000). Autoregressive forecasting of some functional climatic variations. *Scand. J. Statist.*, **27**, 673–687.

[5] Bosq, D. & Delecroix, M. (1985). Nonparametric prediction of a Hilbert-space valued random variable. *Stochastic Process. Appl.* **19**, 271–280.

[6] Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In *Nonparametric Functional Estimation and Related Topics*, Ed. G. Roussas, pp. 509–529, Nato ASI Series C, Vol. **335**, Dortrecht: Kluwer Academic Publishers.

[7] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*. Lecture Notes in Statistics, Vol. **110**, New York: Springer-Verlag.

[8] Bosq, D. (2000). *Linear Processes in Function Spaces*. Lecture Notes in Statistics, Vol. **149**, New York: Springer-Verlag.

[9] Box, G.E. & Jenkins, G.M. (1976). *Time Series Analysis*. San Francisco: Holden Day.

[10] Brockwell, P.J. & Davis, R.A. (1991). *Time Series: Theory and Methods*. 2nd Edition. New York: Springer-Verlag.

[11] Cardot, H., Ferraty, F. & Sarda, P. (1999). Functional linear model. *Statist. Probab. Lett.*, **45**, 11–22.

[12] Cheng, B. & Tong, H. (1996). A theory of wavelet representation and decomposition for a general stochastic process. In *Lect. Notes Statist.*, **115**, pp. 115–129, Springer Verlag: New York.

[13] Cheng, B. & Tong, H. (1998). k-stationarity and wavelets. *J. Statist. Plann. Inference*, **68**, 129–144.

[14] Cohen, A. & D'Ales, J.P. (1997). Nonlinear approximation of random functions. *SIAM J. Appl. Math.*, **57**, 518–540.

[15] Cohen, A., Dyn, N. & Matei, B. (2003). Quasilinear subdivision schemes with applications to ENO interpolation. *Appl. Comp. Harm. Anal.*, **15**, 89–116.

[16] Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.

[17] Donoho, D.L. (1992). Interpolating wavelet transforms. *Technical Report*, **408**, Department of Statistics, Stanford University, USA.

[18] Doukhan, P. (1994). *Mixing: Properties and Examples*. New York: Springer-Verlag.

[19] Ferraty, F. & Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Computat. Statist.*, **17**, 545–564.

[20] Fryzlewicz, P., Van Bellegem & von Sachs, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Ann. Inst. Statist. Math.*, **55**, 737–764.

[21] Härdle, W. & Vieu, P. (1992). Kernel regression smoothing of time series. *J. Time Ser. Anal.*, **13**, 209–232.

[22] Hart, J.D. (1996). Some automated methods of smoothing time-dependent data. *J. Nonparametr. Statist.* **6**, 115–142.

[23] Latif, M., Barnett, T., Cane, M., Flugel, M., Graham, N., Xu, J. & Zebiak, S. (1994). A review of ENSO prediction studies. *Climate Dynamics*, **9**, 167–179.

[24] Mallat, S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pat. Anal. Mach. Intel.*, **11**, 674–693.

[25] Mallat, S.G. (1999). *A Wavelet Tour of Signal Processing*. 2nd Edition, San Diego: Academic Press.

[26] Meyer, Y. (1992). *Wavelets and Operators*. Cambridge: Cambridge University Press.

[27] Misiti, M., Misiti, Y., Oppenheim, G. & Poggi, J.-M. (1994), Décomposition en ondelettes et méthodes comparatives: étude d'une courbe de charge électrique. *Rev. Statist. Appl.*, **XLII**, 57–77.

[28] Nason, G.P. & von Sachs, R. (1999). Wavelets in time-series analysis. *R. Soc. Lond. Philos. Trans.*, Series A, **357**, 2511–2526.

[29] Neveu, J. (1968). Processus aleatoirés gaussiens. *Technical Report*, Presses de l'Université de Montréal, Canada.

[30] Percival, D.B. & Walden, A.T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press.

[31] Philander, S. (1990). *El Niño, La Niña and the Southern Oscillation*. San Diego: Academic Press.

[32] Ramsay, J.O. & Silverman, B.W. (1997). *Functional Data Analysis*. New York: Springer-Verlag.

[33] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Ac. Sc. USA*, **42**, 43–47.

[34] Smith, T.M., Reynolds, R., Livezey, R. & Stokes, D. (1996). Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, 1403–1420.

[35] Venables, W.N. & Ripley, B.D. (1999). *Modern Applied Statistics with S-Plus*. New York: Springer-Verlag.