Journal of Forecasting
J. Forecast. (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/for.1169

Functional Methods for Time Series Prediction: A Nonparametric Approach

GERMÁN ANEIROS-PÉREZ,* RICARDO CAO AND JUAN M. VILAR-FERNÁNDEZ
Departamento de Matemáticas, Facultad de Informática, Universidade da Coruña, Spain

ABSTRACT

The problem of prediction in time series using nonparametric functional techniques is considered. An extension of the local linear method to regression with functional explanatory variable is proposed. This forecasting method is compared with the functional Nadaraya–Watson method and with finite-dimensional nonparametric predictors for several real-time series. Prediction intervals based on the bootstrap and conditional distribution estimation for those nonparametric methods are also compared. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS time series forecasting; functional data; nonparametric regression; bootstrap

INTRODUCTION

Prediction of future observations is an important problem in time series. Given an observed series $Z_1, Z_2, \ldots, Z_n$, the aim is to predict a future value Z_{n+l}, for some integer $l \geq 1$. A useful approach for the prediction problem is to consider that the series follows an autoregressive process of order q:

$$Z_t = m(Z_{t-1}, Z_{t-2}, \ldots, Z_{t-q}) + \varepsilon_t$$

where ε_t is the error process, assumed to be independent of the past of Z_t, i.e., $Z_{t-1}, Z_{t-2}, \ldots$. It is clear then that the first task is to estimate the function $m(\cdot)$.

A classical approach to this problem consists in assuming that $m(\cdot)$ belongs to a class of functions, only depending on a finite number of parameters to be estimated. Examples of such classes are the well-known ARIMA models, widely studied in the literature (see, among many others, the books by Box and Jenkins, 1976; Brockwell and Davis, 1987; Makridakis *et al.*, 1998). This problem can

*Correspondence to: Germán Aneiros-Pérez, Departamento de Matemáticas, Facultad de Informática, Universidade da Coruña, Campus de Elviña, s/n. 15071 A Coruña, Spain. E-mail: ganeiros@udc.es

also be addressed via nonparametric methods. These methods do not assume any functional form on $m(\cdot)$, but only impose regularity conditions on it.

Nonparametric regression estimation under dependence is a useful tool for forecasting in time series. Some relevant works in this field include Györfi *et al.* (1989), Härdle and Vieu (1992), Hart (1991, 1996), Masry and Tjostheim (1995), Härdle *et al.* (1997, 1998) and Bosq (1998). Other papers more specifically focused on prediction using nonparametric techniques are Carbon and Delecroix (1993), Nottingham and Cook (2001), Matzner-Lober *et al.* (1998) and Vilar-Fernández and Cao (2007).

The literature on methods for time series prediction in the context of functional data is much more limited. The books by Bosq (2000) and Ferraty and Vieu (2006) are comprehensive references for parametric (linear) and nonparametric functional data analysis, respectively. Applications of the FAR models (a functional version of the classical AR models) can be seen in Besse *et al.* (2000). Masry (2005) has proven asymptotic normality of the kernel regression functional estimator under dependence, while Antoniadis *et al.* (2006) proposed a functional wavelet-kernel approach for time series prediction. Aneiros-Pérez and Vieu (2008) deal with the problem of nonparametric time series prediction using a semi-functional partial linear model.

In this paper we adopt a nonparametric view for the problem of time series prediction using functional data techniques. Specifically, a local-linear regression estimator for this problem is been proposed. This estimator is compared with the Nadaraya–Watson kernel estimator for the regression functional, as well as with the classical finite-dimensional versions of the Nadaraya–Watson and the local-linear regression estimator for the problem of time series prediction. These four methods are applied to three real-time series concerning electricity consumption, ozone concentration and air temperature.

The rest of the paper is organized as follows. The mathematical formulation of the nonparametric prediction problem is presented in the next section. The third section contains details of the local-linear regression estimator for functional data and how to use it to construct point forecasts and nonparametric prediction intervals. A comparative empirical study of the new method and other nonparametric approaches is included in the fourth section, where some conclusions are drawn.

FORMULATION OF THE PROBLEM

Let us consider a continuous-time stochastic process, $\{Z(t)\}_{t\in\mathbb{R}}$, observed for $t \in$ [a, b), and suppose we are interested in predicting $Z(b + r)$, for some $r \geq 0$. Let us assume that $Z(t)$ is (or may be) stational, with seasonal length τ and $b = a + (n' + 1)\tau$. In other words, we assume that the interval [a, b) consists of $n' + 1$ seasonal periods of length τ of the stochastic process $\{Z(t)\}_{t\in\mathbb{R}}$. For simplicity we will assume the following Markov property

$$Z(b+r)|_{\{Z(t),t\in[a,b)\}} \stackrel{d}{=} Z(b+r)|_{\{Z(t),t\in[b-\tau,b)\}}$$

By defining the functional data $\{(\mathbf{X}_i, Y_i)\}_{i=1}^{n'}$, where $\mathbf{X}_i(t) = Z(a + (i - 1)\tau + t)$ with $t \in C = [0, \tau)$, and $Y_i = Z(a + i\tau + r)$ with $r \in C$, we may look at the problem of predicting $Z(b + r)$ by computing nonparametric estimations, $\hat{m}(\mathbf{x}_{n'+1})$, of the autoregression functional:

DOI: 10.1002/for

$$m(\mathbf{x}_{n'+1}) = E\left(Y|_{\mathbf{X}=\mathbf{x}_{n'+1}}\right) \tag{1}$$

with functional explanatory variable, $\mathbf{X}$, and scalar response, Y.

In practice, we typically only observe a discrete version of the functional data in s equispaced instants ($s \in \mathbb{N}$). More specifically, we only observe $\mathbf{X}_i(t)$ for $t = \frac{j}{s}\tau$, with $j = 0, 1, \ldots, s-1$. In such a case, defining $n = (n'+1)s$, we may formulate the prediction problem in terms of a discrete-time process. Given the observed sample from the time series:

$$Z_1 = Z(a), Z_2 = Z\left(a + \frac{1}{s}\tau\right), \ldots,$$

$$Z_n = Z\left(a + \frac{n's + s - 1}{s}\tau\right) = Z\left(b - \frac{1}{s}\tau\right)$$

our aim is to predict

$$Z_{n+l} = Z(b + r) = Z\left(a + \frac{n's + s - 1 + l}{s}\tau\right)$$

with $r = \frac{l-1}{s}\tau$, for some fixed $l = 1, 2, \ldots, s$.

LOCAL-LINEAR FUNCTIONAL PREDICTION

In this section we present the Nadaraya–Watson regression estimator for functional data and extend it to the local-linear estimator in the context of the functional explanatory variable. These two nonparametric estimators are useful techniques for point forecasting based on the autoregression functional.

Two methods are also introduced to compute prediction intervals based on the two previous nonparametric forecasts. One is based on a bootstrap resampling of the residuals and the other uses the conditional prediction distribution.

There are plenty of papers in the statistical literature that are concerned with the use of bootstrap methods for time series prediction. Among them we mention the works by Thombs and Schucany (1990), Breidt *et al.* (1995), García-Jurado *et al.* (1995) and Zagdański (2001).

Nadaraya–Watson estimator for the regression functional

Given the functional sample $\{(\mathbf{X}_i, Y_i)\}_{i=1}^{n'}$, the Nadaraya–Watson (NW) kernel estimator evaluated at a given function $\mathbf{u}$, $\hat{m}_{h,NW}(\mathbf{u})$, is of the form

$$\hat{m}_{h,\mathrm{NW}}(\mathbf{u}) = \sum_{j=1}^{n'} W_{h,\mathrm{NW}}(\mathbf{u}, \mathbf{X}_j) Y_j \tag{2}$$

DOI: 10.1002/for

where

$$W_{h,\mathrm{NW}}(\mathbf{u}, \mathbf{X}_j) = \frac{K_h(\|\mathbf{X}_j - \mathbf{u}\|)}{\sum_{i=1}^{n'} K_h(\|\mathbf{X}_i - \mathbf{u}\|)}$$

and

$$K_h(t) = \frac{1}{h} K\left(\frac{t}{h}\right)$$

is the rescaled kernel function with bandwidth $h > 0$ and $\|\ \|$ is a suitable seminorm in the functional space

$$\mathcal{F} = \{f : C \to \mathbb{R}; f \in L_2\}$$

and $\mathbf{u} \in \mathcal{F}$ (see Ferraty and Vieu, 2006, pp. 55–56, p. 223, for details on both the estimator (2) and the crucial role of the seminorm, respectively). The kernel K is a non-negative real-valued function such that $\int_0^\infty K(t)\,\mathrm{d}t = 1$.

Local-linear estimator for the regression functional

In this section we propose a local-linear (LL) functional regression estimator for (1) at $\mathbf{u}$. We extend the ideas in Fan and Gijbels (1996) to the case of functional data. First of all, we use a linear approximation of m in a neighbourhood of a given function $\mathbf{u}$:

$$m(\mathbf{x}) \simeq \beta_0 + \int_C \beta(t)(x(t) - \mathbf{u}(t))\,\mathrm{d}t$$

for some $\beta_0 \in \mathbb{R}$ and $\beta \in \mathcal{F}$. The constant β_0 plays the role of $m(\mathbf{u})$, while the function β is the gradient of m at the 'point' $\mathbf{u}$.

Given the sample $\{(\mathbf{X}_i, Y_i)\}_{i=1}^{n'}$, defined above, we construct the LL estimator of β_0 and β as the minimizers of

$$\Psi(\beta_0, \beta) = \sum_{i=1}^{n'} \left[Y_i - \left(\beta_0 + \int_C \beta(t)(\mathbf{X}_i(t) - \mathbf{u}(t))\,\mathrm{d}t \right) \right]^2 K_{h,i} \tag{3}$$

with $K_{h,i} = K_h(\|\mathbf{X}_i - \mathbf{u}\|)$, for a suitable seminorm $\|\ \|$.

In order to minimize (3) with respect to β_0 and β, we impose that the partial derivative with respect to β_0 is zero:

$$\frac{\partial \Psi}{\partial \beta_0} = 0 \Leftrightarrow \beta_0 \sum_{i=1}^{n'} K_{h,i} + \sum_{i=1}^{n'} K_{h,i} \int_C \beta(t)(\mathbf{X}_i(t) - \mathbf{u}(t))\,\mathrm{d}t = \sum_{i=1}^{n'} K_{h,i} Y_i \tag{4}$$

DOI: 10.1002/for

and that the directional derivative of Ψ in the direction of any $\mathbf{v} \in \mathcal{F}$ is also zero:

$$\begin{aligned}\left.\frac{\partial \Psi(\beta_0, \beta + \varepsilon \mathbf{v})}{\partial \varepsilon}\right|_{\varepsilon=0} = 0 \Leftrightarrow \beta_0 \sum_{i=1}^{n'} K_{h,i} \int_C \mathbf{v}(t)(\mathbf{X}_i(t) - \mathbf{u}(t))\,\mathrm{d}t \\ + \sum_{i=1}^{n'} K_{h,i} \left(\int_C \mathbf{v}(t)(\mathbf{X}_i(t) - \mathbf{u}(t))\,\mathrm{d}t\right)\left(\int_C \beta(t)(\mathbf{X}_i(t) - \mathbf{u}(t))\,\mathrm{d}t\right) \\ = \sum_{i=1}^{n'} K_{h,i} Y_i \left(\int_C \mathbf{v}(t)(\mathbf{X}_i(t) - \mathbf{u}(t))\,\mathrm{d}t\right)\end{aligned} \tag{5}$$

In order to solve in β_0 and β the system of functional equations (4) and (5) for all $\mathbf{v} \in \mathcal{F}$, we first write the unknown function β in terms of a basis, $\{\mathbf{e}_j(\cdot)\}_{j \in \mathbb{N}}$, of $\mathcal{F}$, $\beta(t) = \sum_{j=1}^{\infty} \lambda_j \mathbf{e}_j(t)$, and apply equation (5) for $\mathbf{v} = \mathbf{e}_k$, $k \in \mathbb{N}$. Thus we have the following system of infinitely many linear equations:

$$\beta_0 \sum_{i=1}^{n'} K_{h,i} + \sum_{j=1}^{\infty} \lambda_j \sum_{i=1}^{n'} K_{h,i} a_{ij} = \sum_{i=1}^{n'} K_{h,i} Y_i \tag{6}$$

$$\beta_0 \sum_{i=1}^{n'} K_{h,i} a_{ik} + \sum_{j=1}^{\infty} \lambda_j \sum_{i=1}^{n'} K_{h,i} a_{ij} a_{ik} = \sum_{i=1}^{n'} K_{h,i} Y_i a_{ik},\ k = 1, 2, \ldots \tag{7}$$

where

$$a_{ij} = \int_C (\mathbf{X}_i(t) - \mathbf{u}(t)) \mathbf{e}_j(t)\,\mathrm{d}t$$

The system (6)–(7) can be written in a simpler way after introducing some new notation:

$$\lambda_0 = \beta_0,\ b_{kj} = \sum_{i=1}^{n'} K_{h,i} a_{ik} a_{ij},\ \text{for } k,\ j \in \mathbb{N}$$

$$b_{0j} = \sum_{i=1}^{n'} K_{h,i} a_{ij},\ \text{for } j \in \mathbb{N},\ b_{k0} = \sum_{i=1}^{n'} K_{h,i} a_{ik},\ \text{for } k \in \mathbb{N}$$

$$b_{00} = \sum_{i=1}^{n'} K_{h,i},\ d_k = \sum_{i=1}^{n'} K_{h,i} Y_i a_{ik},\ \text{for } k \in \mathbb{N},\ d_0 = \sum_{i=1}^{n'} K_{h,i} Y_i$$

This gives the following linear system:

$$\sum_{j=0}^{\infty} b_{kj} \lambda_j = d_k,\ k = 0, 1, \ldots \tag{8}$$

DOI: 10.1002/for

The solution of the previous system would give estimators of both m and the gradient of m at $\mathbf{u}$: $\hat{m}_{h,LL}(\mathbf{u}) = \hat{\beta}_0 = \hat{\lambda}_0$ and $\hat{\beta}_{h,\mathrm{LL}}(t) = \sum_{j=1}^{\infty} \hat{\lambda}_j \mathbf{e}_j(t)$, respectively. Of course, in general it is not possible to find an explicit solution of this infinite linear system and truncation ideas can be applied to solve a finite system approximating (8). Fix some $N \in \mathbb{N}$ and consider just the first $N + 1$ equations in (8). This requires finding the solution $\tilde{\lambda}_k$, $k = 0, 1, \ldots, N$ of

$$\sum_{j=0}^{N} b_{kj}\lambda_j = d_k, k = 0, 1, \ldots, N$$

Finally we obtain $\tilde{m}_{h,N,LL}(\mathbf{u}) = \tilde{\beta}_0 = \tilde{\lambda}_0$ and $\hat{\beta}_{h,N,\mathrm{LL}}(t) = \sum_{j=1}^{N} \tilde{\lambda}_j \mathbf{e}_j(t)$. These are approximated solutions of (8), using only the first N functions in the basis: $\mathbf{e}_1(\cdot)$, $\mathbf{e}_2(\cdot)$, $\ldots$, $\mathbf{e}_N(\cdot)$.

Residual-based bootstrap prediction intervals (RBB)

In this subsection and the next one, we follow the lines of Vilar-Fernández and Cao (2007). For this reason we omit the details. The first bootstrap method for interval prediction is based on resampling the residuals. A sketch of the algorithm follows.

1. Compute the residuals, $\hat{\varepsilon}_j = Y_j - \hat{m}(\mathbf{X}_j)$, using some global cross-validation bandwidth, where $\hat{m}(\mathbf{u})$ is either the NW or the LL estimator for functional data.
2. Use s_ε, the standard deviation of the $\hat{\varepsilon}_j$, to compute the smoothing parameter $g = \left(\frac{4}{3n'}\right)^{1/5} s_\varepsilon$.
3. Draw smoothed bootstrap residuals:

$$\hat{\varepsilon}_i^* = \hat{\varepsilon}_{I_i} + g\xi_i, \quad i =, \ldots, B$$

where $I_i \overset{d}{=} U(\{1, \ldots, n'\})$, $\xi_i \overset{d}{=} N(0,1)$ and B is the number of bootstrap replications.

4. Sort the bootstrap residuals: $\{\hat{\varepsilon}_{(i)}^* : i = 1, \ldots, B\}$ and compute the $1 - \alpha$ prediction interval:

$$\left(\hat{m}(\mathbf{X}_{n'+1}) + \hat{\varepsilon}^*_{[(\alpha/2)B]}, \hat{m}(\mathbf{X}_{n'+1}) + \hat{\varepsilon}^*_{[(1-(\alpha/2))B]}\right)$$

Prediction intervals based on the conditional distribution (CD)

This method is based on estimating the conditional distribution function of $Y|_{\mathbf{X}=\mathbf{x}}$. The basic ideas can be found in Cao (1999). For a given real value y, the conditional cumulative distribution function can be viewed as a regression function

$$F(y|_{\mathbf{x}}) = E\left(1_{\{Y \le y\}}|_{\mathbf{X}=\mathbf{x}}\right)$$

Consequently, the functional NW or LL estimators can be used to estimate this conditional distribution function. The prediction algorithm proceeds as follows.

DOI: 10.1002/for

1. Use the sample $\{(\mathbf{X}_j, Y_j): 1 \le j \le n'\}$ to compute $\hat{F}(y|_{\mathbf{x}})$ by means of the NW or the LL method. This function is smoothed again in the variable y.
2. The $1-\alpha$ prediction interval is (L, U), where

$$\hat{F}\left(L|_{\mathbf{x}_{n'+1}}\right)=\frac{\alpha}{2} \quad \text{and} \quad \hat{F}\left(U|_{\mathbf{x}_{n'+1}}\right)=1-\frac{\alpha}{2}$$

EMPIRICAL STUDY

The two functional data nonparametric point forecast and prediction intervals are compared with their finite-dimensional counterparts. Specifically, these prediction methods are applied to three real-time series concerning electricity consumption, ozone concentration and air temperature. As we will show later, these time series are seasonal and, for each one, we have 12 observations taken in equi-spaced instants within each seasonal period. In this sense, we can consider that the length of the seasonal periods is $\tau = 12$ (in some units).

For the finite-dimensional nonparametric approach we follow the procedure in Vilar-Fernández and Cao (2007). An important question in the finite-dimensional setting is how to select the autoregressor variables, $(Z_{t-i_1}, Z_{t-i_2}, \ldots, Z_{t-i_p})$, for predicting Z_{t+l}. We adopt the approach by Tjostheim and Auestad (1994). It consists in minimizing a nonparametric estimation of the final prediction error.

Selection of the tuning parameters

In the functional data setup (as well as in the finite-dimensional case) there are several tuning parameters that need to be selected. We briefly mention now some procedures to do this.

The Epanechnikov kernel is been used for the NW and the LL estimators. Cross-validation methods (see Rachdi and Vieu, 2007; Benhenni *et al.*, 2007), expressed in terms of k-nearest neighbours, are used for smoothing parameter selection. Global cross-validation is used for constructing the residuals corresponding to the RBB prediction intervals, and local cross-validation for computing the nonparametric point forecast and the estimation of the conditional distribution function.

Following the recommendations of Ferraty and Vieu (2006, p. 223), for choosing the seminorm in practical situations we base our choice on the smoothness or roughness of the explanatory curves. Specifically, when the curves are smooth we use the L_2 norm of the qth derivative of the curve, $\|\,\|_q^{\text{derivative}}$, while for rough curves the seminorm is based on principal component analysis, $\|\,\|_q^{PCA}$ (q being the number of principal components). For the definition of this class of seminorms see Ferraty and Vieu (2006, p. 28–30). The Fourier basis is used in the LL functional estimator. The parameter q in the seminorm and the number of functions in the Fourier basis, N, are also selected by cross-validation. The parameter N is selected within $\{3, 5, 7\}$, while q was selected in the set $\{1, \ldots, 12\}$, when the seminorm is based on principal components, and within $\{0, 1, 2\}$ for the seminorm based on the qth derivative.

When the prediction horizon is larger than one, point forecasts are carried out in two different ways. The first one is the direct method and consists of the approach mentioned in the previous section. The second alternative is the recursive method. It computes a one-ahead forecast and

DOI: 10.1002/for

includes it in the sample to perform again a one-lag prediction, as many times as needed. These two methods are compared in the empirical study.

Methods and error criteria

Four nonparametric point forecasts are computed: (a) a finite-dimensional NW forecast, (b) a finite-dimensional LL forecast, (c) a functional NW forecast and (d) a functional LL forecast. These forecasts are performed using either the direct method or the recursive one.

Four types of prediction intervals (only using the direct method) are computed. These are the four combinations for the nonparametric method used for point forecast (NW or LL) and the basic procedure for constructing prediction intervals (residual-based bootstrap and conditional distribution). These four approaches are used for both finite-dimensional and functional autoregression estimation.

The nominal level for the prediction intervals is 95%. The number of bootstrap replications is set to $B = 1000$. A maximum horizon of $s = 12$ is considered.

The performance of the point forecasts is evaluated by excluding the last seasonal period (last 12 observations) from the data, computing the point forecasts for these values and comparing the predicted values, $\hat{z}_n(l)$, with the real ones, z_{n+l}. Several error measures are considered. The root mean squared error:

$$\mathrm{RMSE} = \left[\frac{1}{s} \sum_{l=1}^{s} (\hat{z}_n(l) - z_{n+l})^2 \right]^{1/2}$$

the mean absolute error:

$$\mathrm{MAE} = \frac{1}{s} \sum_{l=1}^{s} |\hat{z}_n(l) - z_{n+l}|$$

and the relative error:

$$\mathrm{RE} = \frac{1}{s} \sum_{l=1}^{s} \frac{|\hat{z}_n(l) - z_{n+1}|}{\hat{\sigma}_l}$$

where $\hat{\sigma}_l^2$ is the quasi-variance of $\{z_{(j-1)\tau+l}\}_{j=1}^{n'}$.

Electricity consumption data

The first dataset analysed consists of monthly electricity consumption in the USA during the period January 1972–January 2005 (397 months). The source of the data was the US Government (Department of Energy), and they are available at the website http://www.economagic.com. The data are transformed using logarithms and then differentiated to eliminate the trend. The seasonal period for this time series is one year. This gives 33 curves (see Ferraty and Vieu, 2006, pp. 17–20, for details about this dataset). From Figure 1 we can observe that the functional data are quite rough curves. Thus we use the class of seminorms $\{\|\|_q^{\mathrm{PCA}}\}_{q=1}^{12}$.

DOI: 10.1002/for

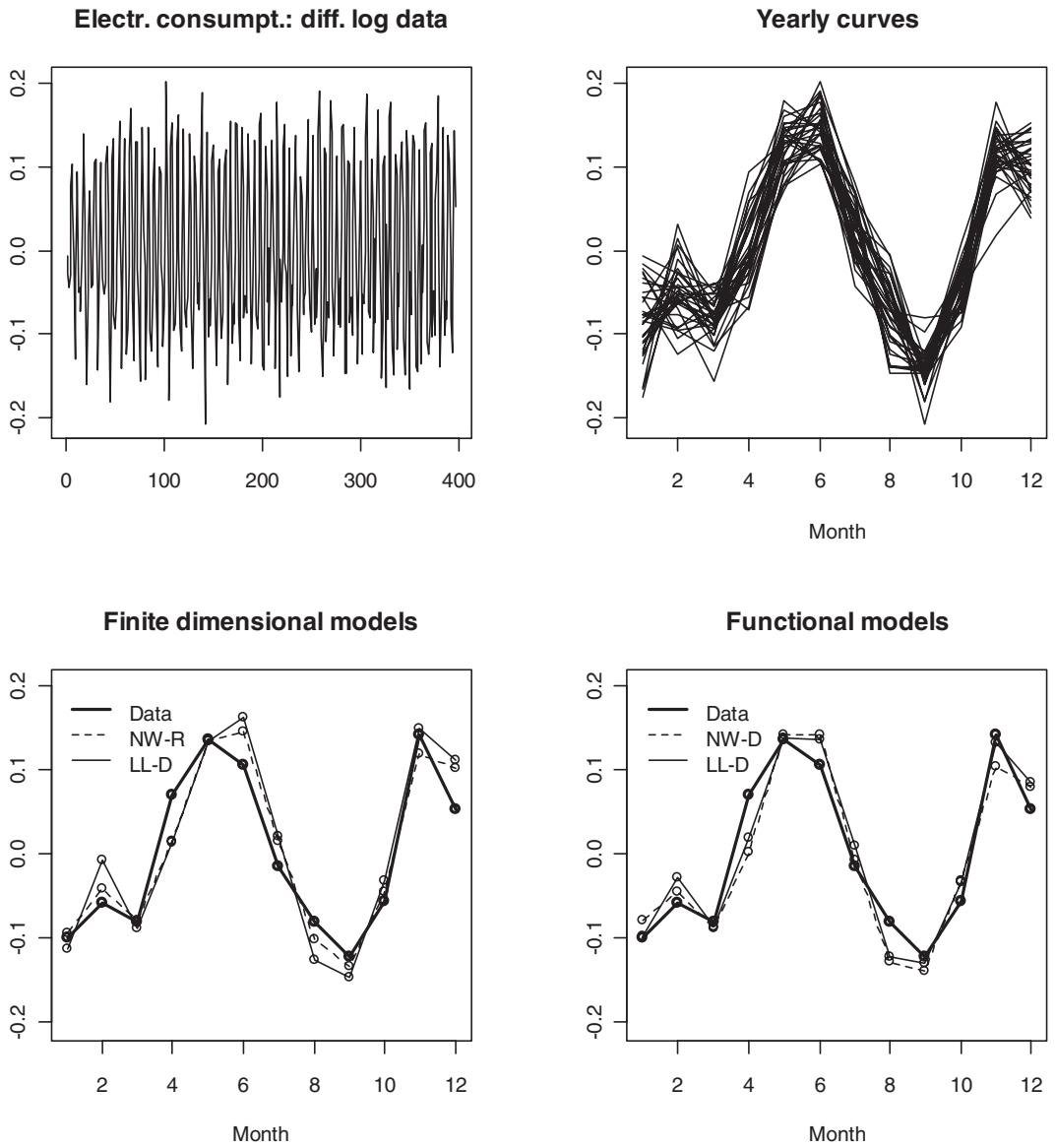

Figure 1. Time series and functional data (upper panels) together with the best forecast (differenced log) electricity consumption for each type of model and estimator (lower panels). Both the direct (D) and the recursive (R) method are combined with the Nadaraya–Watson (NW) and local-linear (LL) estimator

Table I collects the point forecasting errors. Figure 1 shows some plots of the time series along time, the functional data and the best forecasts, for every type of model and estimator. In other words, for each kind of model (finite-dimensional or functional) and each kind of estimator (NW or LL), only the results corresponding to the best (direct or recursive) forecast are shown. Figure 2 reports

DOI: 10.1002/for

Table I. Error criteria for the finite dimensional and functional models using Nadaraya–Watson (NW) and the local-linear (LL) forecasts with the direct (D) and recursive (R) approach for electricity consumption data

Estimator	Error criteria		
	RMSE	MAE	RE
Finite-dimensional			
NW-D	0.0277	0.0225	0.7326
NW-R	0.0281	0.0224	0.7208
LL-D	0.0381	0.0321	1.0313
LL-R	0.0426	0.0357	1.1538
Functional			
NW-D	0.0315	0.0260	0.8241
NW-R	0.0343	0.0287	0.9377
LL-D	0.0269	0.0218	0.6863
LL-R	0.0764	0.0470	1.7574

the prediction intervals (only using the direct method) for the four nonparametric forecasts (NW and LL either finite-dimensional or functional) with the two possible methods for interval construction (RBB and CD).

From a graphical point of view, Figure 1 suggests that the nonparametric forecasts have a good behaviour, in the sense that the predictions follow the trend of the data. To compare quantitatively the different prediction methods used in this paper, we need to consider the information contained in Table I. On the one hand, this table shows that the LL forecast for functional data (direct version) beats the other nonparametric methods (either finite-dimensional or functional) for the analysed series. On the other hand, we should mention the problems in the use of the recursive method, because a poor prediction in a specific instant causes even worse predictions for future instants. In fact, the bad performance of the recursive LL functional predictor suggests that at any instant a poor one-ahead forecast is obtained.

Figure 2 reports results on the prediction intervals. From this figure, and focusing on each class of model (finite-dimensional or functional models), there are no large differences between the intervals constructed by NW estimators and those constructed using LL. Nevertheless, prediction intervals using the functional data approach are, generally speaking, more narrow than those using finite-dimensional models. In addition, functional LL–RBB prediction intervals are most of the time more accurate and narrow than the others.

Ozone concentration data

The second series collects ozone concentrations every second hour from 18 May 2005 to 29 June 2005 (516 items of data) recorded in Getafe (Madrid, Spain). These data, published by the Autonomous Community of Madrid (Environmental Department), can be found at the website http://gestiona.madrid.org/aireinternet. There exist a clear daily seasonality in this series, which gives 43

DOI: 10.1002/for

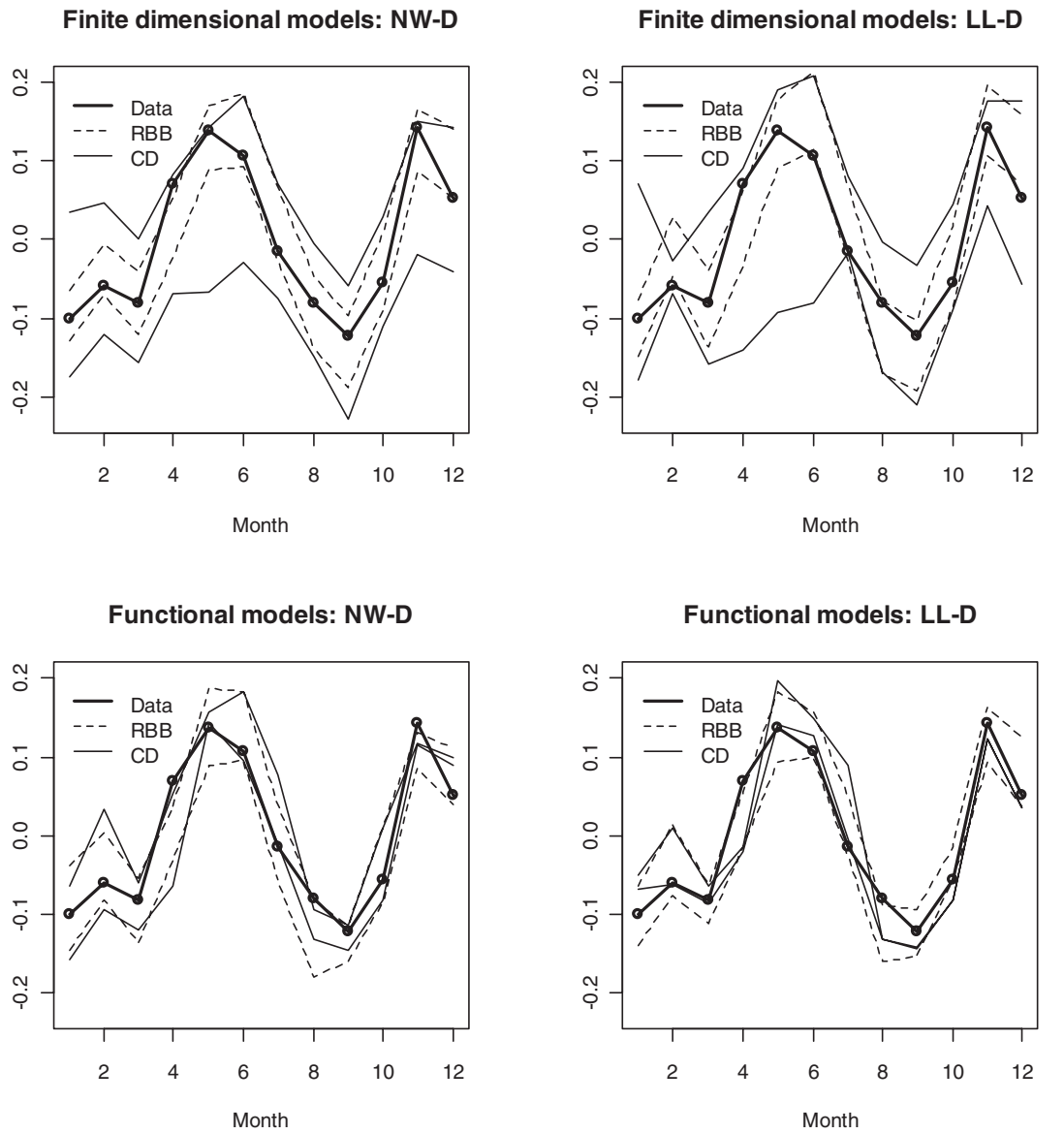

Figure 2. Prediction intervals corresponding to the (differenced log) electricity consumption data. They are based on both finite-dimensional (upper panels) and functional (lower panels) models. The Nadaraya–Watson (NW) and local-linear (LL) estimator are used for constructing the residual-based bootstrap (RBB) prediction intervals and prediction intervals based on the conditional distribution (CD). Only the direct (D) method is considered

DOI: 10.1002/for

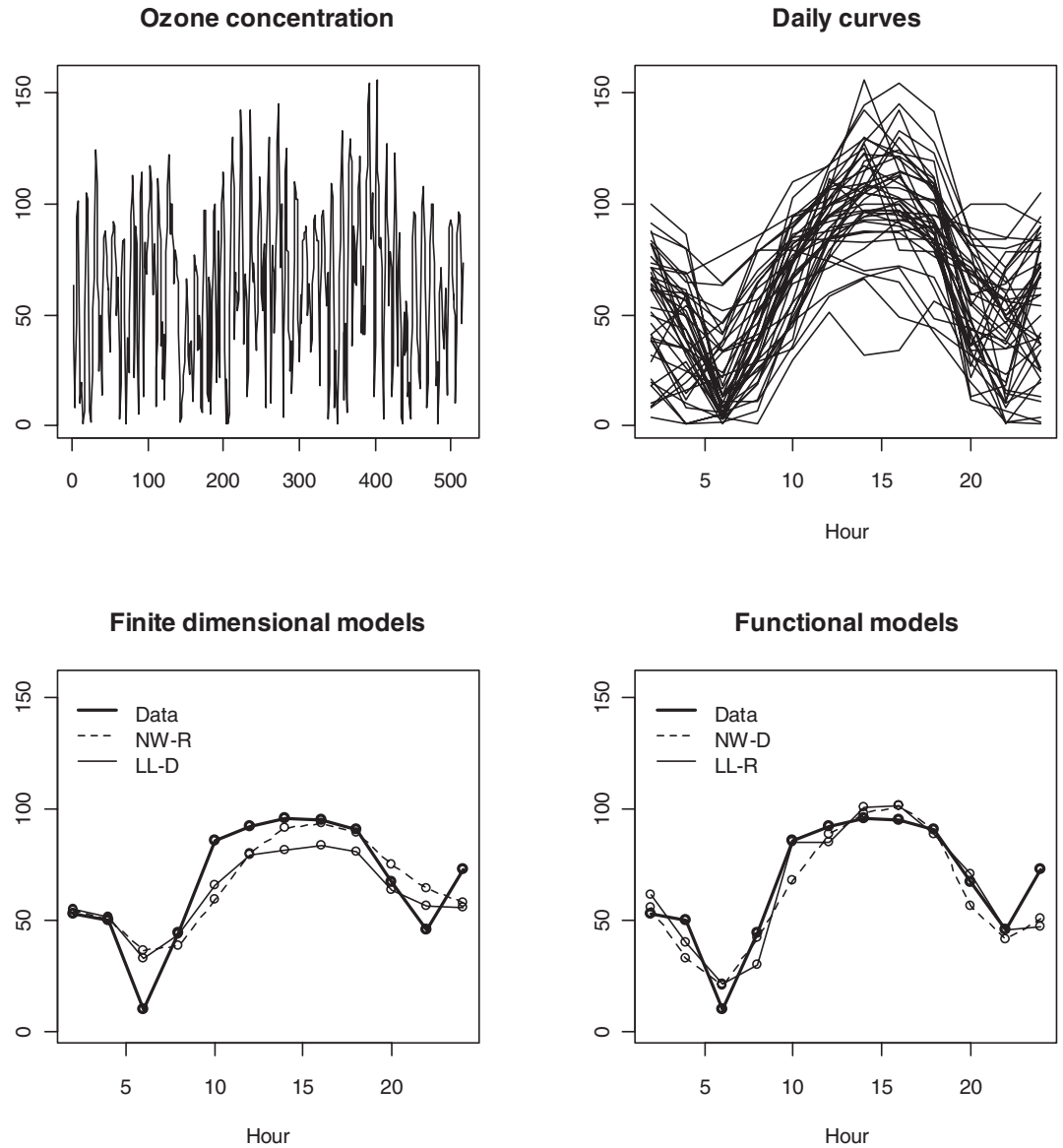

Figure 3. Time series and functional data (upper panels) together with the best forecast ozone concentration for each type of model and estimator (lower panels). Both the direct (D) and recursive (R) method are combined with the Nadaraya–Watson (NW) and local-linear (LL) estimator

curves (see Aneiros-Pérez and Vieu, 2008, for more information about this series). The smooth shape of the curves (see Figure 3) suggests use of the class of seminorms $\{\|\,\|_q^{\text{derivative}}\}_{q=0}^{2}$.

The point forecasting errors can be seen in Table II. Figure 3 collects some plots of the time series along time, the functional data and the best forecasts, for every type of model and estimator.

DOI: 10.1002/for

Table II. Error criteria for the finite-dimensional and functional models using the Nadaraya–Watson (NW) and local-linear (LL) forecasts with the direct (D) and recursive (R) approach for ozone concentration data

Estimator	Error criteria		
	RMSE	MAE	RE
Finite-dimensional			
NW-D	13.9557	11.9029	0.5454
NW-R	13.5441	10.0406	0.4669
LL-D	12.7884	10.6460	0.4884
LL-R	15.0158	11.9409	0.5550
Functional			
NW-D	10.9547	8.2943	0.3631
NW-R	14.3201	11.7257	0.5531
LL-D	14.3737	11.8001	0.5000
LL-R	10.3891	7.8917	0.3395

The discussion given in the previous example on the point forecast applies essentially for the ozone concentration data, as can be seen in Table II and Figure 3. The only difference is based on the fact that the recursive version of the LL functional predictor shows now a better behaviour than the direct version. This suggests that all the one-ahead forecasts involved in the recursive approach have a good performance.

For the sake of brevity, we omit the results for the prediction intervals. The conclusions are the same as those obtained from Figure 2.

Air temperature data

The Mabegondo data comprise the third time series we analyse. It is available at the website http://www.meteogalicia.es (source: Xunta of Galicia). Air temperature was recorded every 2 hours at Mabegondo meteorological station (Mabegondo, Galicia, Spain) over the period 1 January–30 March 2008. The seasonal period is one day (1092 items of data and 91 curves). As in the case of the ozone concentration data, we are in a situation in which the curves are smooth (see Figure 4). Thus we use the class of seminorms $\{\|\|_q^{\text{derivative}}\}_{q=0}^{2}$.

Table III reports the point forecasting errors. Figure 4 shows some plots of the time series along time, the functional data and the best forecasts, for every type of model and estimator. Both Table III and Figure 4 show poor behaviour of the finite-dimensional predictors. The good performance of the functional forecasts remains here (especially in the case of the LL forecasts). In addition, we observe in Table III similar values for the error criteria when the direct method or the recursive one is used in the LL functional forecasts. As in the previous subsection, we do not report results on the prediction intervals. In fact, the conclusions for the prediction intervals are similar to those presented for the electricity consumption data.

In summary, it is worth mentioning that the curves corresponding to the electricity data are rough, while those corresponding to both ozone concentration and air temperature data are smooth. On the other hand, we note that the curves corresponding to the last two datasets are more sparse than those corresponding to the electricity data. Thus the empirical study covers different situations that are common in practice.

DOI: 10.1002/for

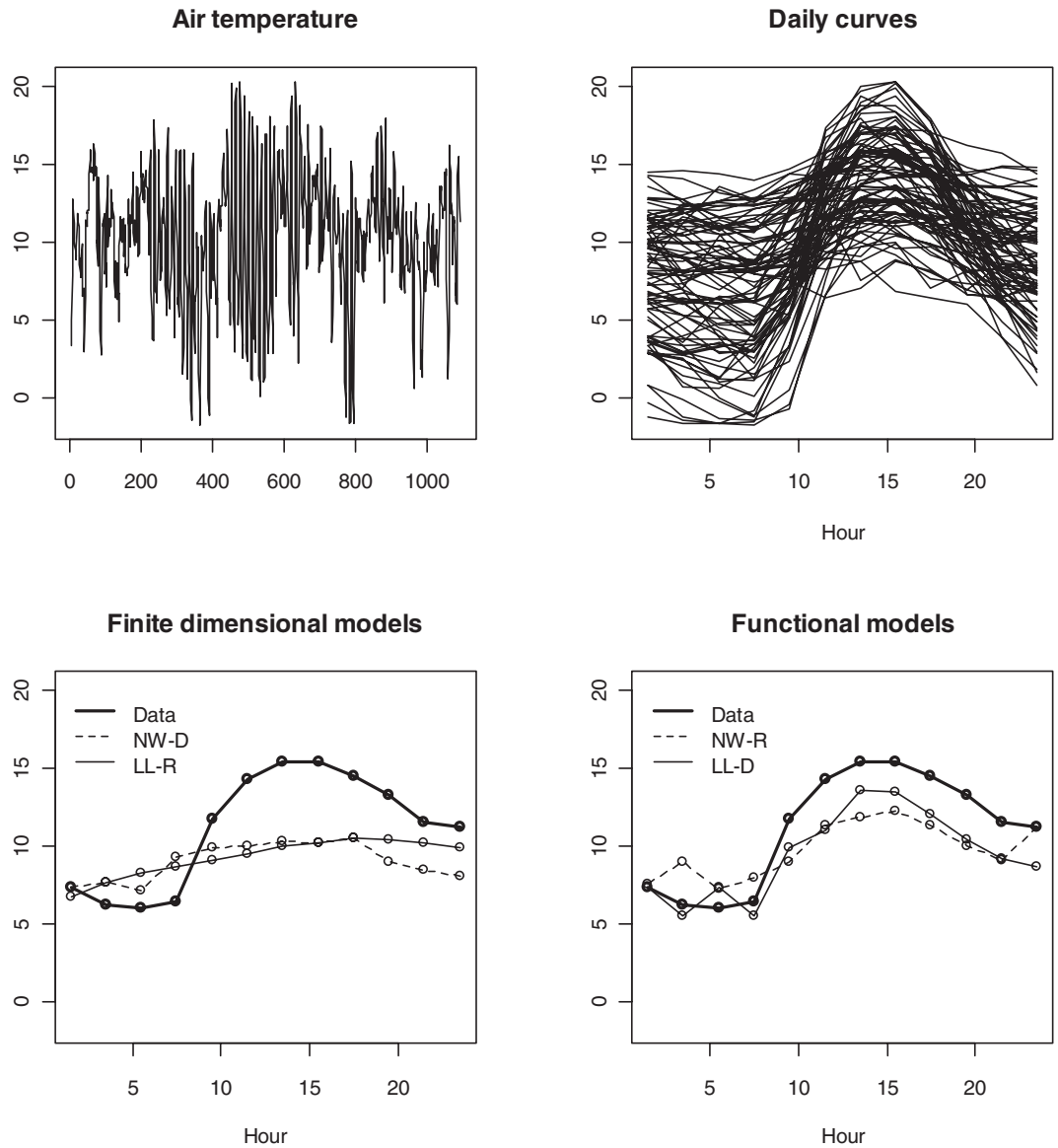

Figure 4. Time series and functional data (upper panels) together with the best forecast air temperature for each type of model and estimator (lower panels). Both the direct (D) and recursive (R) method are combined with the Nadaraya–Watson (NW) and local-linear (LL) estimator

DOI: 10.1002/for

Table III. Error criteria for the finite-dimensional and functional models using the Nadaraya–Watson (NW) and local-linear (LL) forecasts with the direct (D) and recursive (R) approach for air temperature data

Estimator	Error criteria		
	RMSE	MAE	RE
Finite-dimensional			
NW-D	3.4198	3.0401	1.0730
NW-R	3.8345	3.3167	1.1674
LL-D	3.3392	2.9590	1.0516
LL-R	3.2507	2.8530	0.9827
Functional			
NW-D	2.8224	2.5496	0.8886
NW-R	2.5482	2.2725	0.7891
LL-D	2.0530	1.8415	0.6556
LL-R	2.4964	1.9118	0.6658

ACKNOWLEDGEMENTS

We are grateful to Hervé Cardot, who kindly provided us some R code for functional data analysis. This research was partially supported by Grants PGIDIT07PXIB105259PR and 07SIN012105PR from Xunta de Galicia (Spain), by Grant number MTM2008-00166 from Ministerio de Ciencia e Innovación (Spain), and by the research group MODES.

REFERENCES

Aneiros-Pérez G, Vieu P. 2008. Nonparametric time series prediction: a semi-functional partial linear modeling. *Journal of Multivariate Analysis* **99**: 834–857.

Antoniadis A, Paparoditis E, Sapatinas T. 2006. A functional wavelet-kernel approach for time series prediction. *Journal of the Royal Statistical Society, Series B* **68**: 837–857.

Benhenni K, Ferraty F, Rachdi M, Vieu P. 2007. Local smoothing regression with functional data. *Computational Statistics* **22**: 353–369.

Besse PC, Cardot H, Stephenson DB. 2000. Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics* **27**: 673–687.

Bosq D. 1998. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Lecture Notes in Statistics 110. Springer: Berlin.

Bosq D. 2000. *Linear Processes in Function Spaces: Theory and Applications*. Lecture Notes in Statistics 149. Springer: Berlin.

Box GEP, Jenkins GM. 1976. *Time Series Analysis: Forecasting and Control*. Holden-Day: San Francisco, CA.

Breidt FJ, Davis RA, Dunsmuir WTM. 1995. Improved bootstrap prediction intervals for autoregressions. *Journal of Time Series Analysis* **16**: 177–200.

DOI: 10.1002/for

Brockwell PJ, Davis RA. 1987. *Time series: Theory and methods*. Springer: Berlin.
Cao R. 1999. An overview of bootstrap methods for estimating and predicting in time series. *Test* **8**: 95–116.
Carbon M, Delecroix M. 1993. Non-parametric vs parametric forecasting in time series: a computational point of view. *Applied Stochastic Models and Data Analysis* **9**: 215–229.
Fan J, Gijbels I. 1996. *Local Polynomial Modelling and its Applications*. Chapman & Hall: London.
Ferraty F, Vieu P. 2006. *Nonparametric Functional Data Analysis*. Springer Series in Statistics. Springer: New York.
García-Jurado I, González-Manteiga W, Prada-Sánchez JM, Febrero-Bande M, Cao R. 1995. Predicting using Box–Jenkins, nonparametric, and bootstrap techniques. *Technometrics* **37**: 303–310.
Györfi L, Härdle W, Sarda P, Vieu P. 1989. *Nonparametric Curve Estimation from Time Series*. Lecture Notes in Statistics 60. Springer: Berlin.
Härdle W, Vieu P. 1992. Kernel regression smoothing of time series. *Journal of the Time Series Analysis* **13**: 209–232.
Härdle W, Lütkepohl H, Chen R. 1997. A review of nonparametric time series analysis. *International Statistical Review* **65**: 49–72.
Härdle W, Tsybakov A, Yang L. 1998. Nonparametric vector autoregression. *Journal of Statistical Planning and Inference* **68**: 22–245.
Hart JD. 1991. Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society, Series B* **53**: 173–187.
Hart JD. 1996. Some automated methods of smoothing time-dependent data. *Journal of Nonparametric Statistics* **6**: 115–142.
Makridakis S, Wheelwright SC, Hyndman RJ. 1998. *Forecasting: Methods and Applications*. Wiley: Chichester.
Masry E. 2005. Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Processes and their Applications* **115**: 155–177.
Masry E, Tjostheim D. 1995. Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* **11**: 258–289.
Matzner-Lober E, Gannoun A, De Gooijer JG. 1998. Nonparametric forecasting: a comparison of three kernel based methods. *Communications in Statistics: Theory and Methods* **27**: 1593–1617.
Nottingham QJ, Cook DC. 2001. Local linear regression for estimating time series data. *Computational Statistics and Data Analysis* **37**: 209–217.
Rachdi M, Vieu P. 2007. Nonparametric regression for functional data: automatic smoothing parameter selection. *Journal of Statistical Planning and Inference* **137**: 2784–2801.
Thombs LA, Schucany WR. 1990. Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association* **85**: 486–492.
Tjostheim D, Auestad BH. 1994. Nonparametric identification of nonlinear time series: selecting significant lags. *Journal of the American Statistical Association* **89**: 1410–1419.
Vilar-Fernández JM, Cao R. 2007. Nonparametric forecasting in time series: a comparative study. *Communications in Statistics: Simulation and Computation* **36**: 311–334.
Zagdański A. 2001. Prediction intervals for stationary time series using the sieve bootstrap method. *Demonstratio Mathematica* **34**: 469–481.

Authors' biographies:

Germán Aneiros-Pérez is Associate Professor of Statistics at University of A Coruña (Spain) since 2003. He was awarded a Ph.D. in Statistics from University of Santiago de Compostela (Spain) in 2001. He has been working on several research topics over the past few years. This includes nonparametric methods for time series forecasting and, more recently, research in the setting of nonparametric methods for dependent functional data.

Ricardo Cao is Professor of Statistics at University of A Coruña (Spain) since 1999. He is Editor in Chief of the scientific journal *Test*, and Associate Editor of *Computational Statistics* and *Journal of Nonparametric Statistics*. He has published over sixty papers in international scientific journals in the topics of nonparametric curve estimation, bootstrap methods, survival analysis and time series, some of them with applications to the environment, material science and forensic science.

DOI: 10.1002/for

Juan M. Vilar-Fernández is Professor of Statistics at University of A Coruña (Spain) since 1994. His recent research deals with nonparametric methods under dependence conditions, including nonparametric forecasting in time series, local polynomial regression, time series clustering, comparison tests for regression curves and conditional variance estimation.

Authors' addresses:
Germán Aneiros-Pérez, **Ricardo Cao** and **Juan M. Vilar-Fernández**, Departamento de Matemáticas, Facultad de Informática, Universidade da Coruña, Campus de Elviña, s/n. 15071 A Coruña, Spain.

DOI: 10.1002/for