Partly linear models on Riemannian manifolds

Wenceslao Gonzalez-Manteiga[a], Guillermo Henry[b] and Daniela Rodriguez[b]
[a] *Universidad de Santiago de Compostela, Spain*
[b] *Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina*

Abstract

In partly linear models, the dependence of the response y on $(\mathbf{x}^{\mathrm{T}}, t)$ is modeled through the relationship $y = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + g(t) + \varepsilon$ where ε is independent of $(\mathbf{x}^{\mathrm{T}}, t)$. We are interested in develop an estimation procedure that allows to combine the flexibility of the partly linear models, studied by several authors, but including some variables belong to a non Euclidean space. The motivating application for this paper deals to explain the atmospheric SO_2 pollution incidents using these models when the nature of some of predictive variables belong in a cylinder. In this paper, estimators of $\boldsymbol{\beta}$ and g are constructed when the explanatory variables t take values on a Riemannian manifold and asymptotic properties of the proposed estimators are obtained under suitable conditions. We illustrate the use of this estimation approach on the environment dataset and, we explored the performance of the estimators through a simulation study.

Key words and phrases: Environmental data, Hypothesis test, Nonparametric estimation, Partly linear models, Riemannian manifolds.

1 Introduction

The partly linear models were introduced by [7] to analyze the relationship between the electricity usage and average daily temperature. In recent years, this model has gained a lot of attention in order to explore the nature of complex nonlinear phenomena. Partly linear models were widely studied in the literature see for example [18], [5], [1] among others. The partly linear models allow modeling the response variable with a set of predictors that enter linearly in the model while one of them is considered in the model nonparametrically.

In this paper, we discuss the application of these models to the sulfur dioxide (SO_2) pollution problem. More specifically, we are interested in model the emission of SO_2 through of variables such as the temperature and the direction and speed of the wind. It is important to remark that the nature of the wind's variables have a common structure. Also, the circular structure of the wind direction allows to consider a cylinder as the space where the speed and direction take values. However, the partly linear models do not seem include this structure in the nature of the model. The case studied in this paper is only an example of predictive variables taking values on a Riemannian manifold rather than on Euclidean space. Some others examples of variables taking values in a non Euclidean space could be found in meteorology, astronomy, geology and other fields, that include naturally distributions on spheres, tangent bundles, Lie groups, etc. Research on the statistical analysis of variables with some one of these structures was studied by [4], [14] and more recently by [10], [16],[15] and [11].

The aim of this work is to study the partly linear models when the explanatory variable t takes values on a Riemannian manifold, i.e. when the variable to be modeled in a non-parametric way is in a manifold. We introduce an estimation procedure that include this structure of the variables.

This paper is organized as follows. In Section 2, we construct estimates for these models and give a review of the nonparametric estimation on Riemannian manifolds proposed in [15]. In Section 3, we present the asymptotic behavior of the proposed estimators under regular assumptions. In Section 4, we explored the performance of the estimators through a simulation study and we show an example using real data. Moreover, we review a cross validation procedure for partial linear models. The proofs of the theoretical results presented in Section 3 are given in the Appendix.

2 Estimators

2.1 Model and estimators

Let $(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i)$ be an i.i.d. random vectors valued in $I\!R^{p+1} \times M$ with identically distribution to $(y, \mathbf{x}^{\mathrm{T}}, t)$, where (M, ξ) is a Riemannian manifold of dimension d. The partly linear model assume that the relation between the response variable y_i and the covariates $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ can be represented as

$$y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + g(t_i) + \varepsilon_i \qquad 1 \leq i \leq n \;, \tag{1}$$

where the errors ε_i are independent and independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)^{\mathrm{T}}$, also $E(\varepsilon_i | \mathbf{x}_i, t_i) = 0$. In many situations, it seems reasonable to suppose that a relationship between the covariates $\mathbf{x}$ and t exists, so as in [18] and [1], we will assume that for $1 \leq j \leq p$

$$x_{ij} = \phi_j(t_i) + \eta_{ij} \qquad 1 \leq i \leq n \tag{2}$$

where the errors η_{ij} are independent. Denote $\phi_0(\tau) = E(y|t = \tau)$ and $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_p(t))$, then we have that $g(t) = \phi_0(t) - \boldsymbol{\phi}(t)^{\mathrm{T}} \boldsymbol{\beta}$ and hence, $y - \phi_0(t) = (\mathbf{x} - \boldsymbol{\phi}(t))^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon$. This equation suggest to estimate the unknown functions and parameters as follows. Let $\widehat{\phi}_j(t)$ be the nonparametric estimators of ϕ_j, $0 \leq j \leq p$. Regarding the estimation of the parameter $\boldsymbol{\beta}$, we note that using the nonparametric estimators of the functions ϕ_j, the regression parameter can be estimate considering the least square estimators obtained minimizing

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} [(y_i - \widehat{\phi}_0(t_i)) - (\mathbf{x}_i - \widehat{\boldsymbol{\phi}}(t_i))^{\mathrm{T}} \boldsymbol{\beta}]^2.$$

where $\widehat{\boldsymbol{\phi}}(t) = (\widehat{\phi}_1(t), \ldots, \widehat{\phi}_p(t))$. Then, the function g can be estimated as $\widehat{g}(t) = \widehat{\phi}_0(t) - \widehat{\boldsymbol{\phi}}(t)^{\mathrm{T}} \widehat{\boldsymbol{\beta}}$.

Note that the regression functions correspond to predictors taking values in a Riemannian manifold, nonparametric kernel type estimators adapted to this structure was considered in [15] and also studied in [12]. An overview of these estimators can be found in the following Subsection.

The proposed estimators are consistent with the respective estimators when the explanatory variable t take values on Euclidean spaces, i.e. in this case the proposed estimators correspond to the estimators introduced by [7].

2.2 Review of Nonparametric estimators on Riemannian manifolds

2.2.1 Preliminaries

Let (M, ξ) be a $d-$dimensional connected Riemannian manifold. We denote by d_ξ the distance in M induced by ξ. (M, ξ) is a complete Riemannian manifold if (M, d_ξ) is complete as a metric space. $\mathbb{R}^d$ endowed with the Euclidean metric ξ_e^d, the hyperbolic space (H^d, ξ_h), the d-dimensional torus (T^d, ξ) and the graph of a smooth function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ are examples of complete Riemannian manifolds.

Let $t \in (M, \xi)$ and T_tM the tangent fiber of M at t. Let $\exp_t : T_tM \longrightarrow M$ be the exponential map induced by the metric ξ , see for instance [6]. Recall that $\exp_t(0_t) = t$, where 0_t is the null vector of T_tM, and there exists a neighborhood W of 0_t that satisfies $\exp_t |_W : W \longrightarrow \exp_t(W)$ is a diffeomorphism. We say that a ball $B_r(t) = \{q \in M : d_\xi(t, q) < r\}$ is *normal* if there exists a ball $B_r(0_t)$ such that

$$\exp_t |_{B_r(0_t)} : B_r(0_t) \longrightarrow B_r(p)$$

is a diffeomorphism.

The injectivity radius of (M, ξ) is defined by

$$inj_\xi(M) := \inf_{r>0}\{B_r(t) \text{ is a normal ball}\}.$$

In the case of the Euclidean or the Hyperbolic space is easy to see that $inj_{\xi_e^d}(\mathbb{R}^d) = inj_{\xi_h}(H^d) = \infty$. The surface $M := \{x^2 + y^2 - e^{-z} = 0\} \subseteq \mathbb{R}^3$ endowed with the metric induced by ξ_e^3 is a complete Riemannian manifold without boundary with injectivity radius equals zero. If (S^d, ξ_0^d) is the d-dimensional sphere with its canonical metric, then $inj_{\xi_0^d}(S^d) = \pi$. Every compact Riemannian manifold has positive injectivity radius.

In this paper we are going to consider only complete Riemannian manifolds without boundary and positive injectivity radius. The assumption about the positive injectivity radius, it will be clear in the next subsection when we introduce the estimator proposed by Pelletier.

Let $t \in (M, \xi)$ and $\{v_1, \ldots, v_d\}$ be an orthonormal basis of T_tM and let $B_r(t)$ be a normal ball. The exponential map induced a coordinate system (U, ψ) in (M, ξ) as follows: Let $\gamma : T_tM \longrightarrow \mathbb{R}^d$ defined by

$$\gamma(v) = (z_1, \ldots, z_d) \text{ if } v = \sum_{i=1}^{d} z_i v_i,$$

then $U = B_r(t)$ and $\psi : B_r(t) \longrightarrow I\!R^d$ is given by

$$\psi(s) = \gamma \circ (\exp_t |_{B_r(0_t)})^{-1}(s).$$

The coordinate system (U, ψ) is called a *normal charts centered at* t.

If $s \in U$ let us denote by $\{\partial/\partial\psi_1|_s, \ldots, \partial/\partial\psi_d|_s\}$ the basis of T_sM induced by (U, ψ), i.e. let $q \in B_r(0_t)$ such that $\exp_t(q) = s$, then $\partial/\partial\psi_i|_s$ is the velocity at time 0 of the curve $\alpha_i(a) = exp_t(q + a.v_i)$.

For any $s \in M$ the metric ξ restricted to the tangent fiber of M at s is a symmetric positive defined bilinear form $\xi_s : T_sM \times T_sM \longrightarrow I\!R$. If (U, ξ) is a coordinate system in M and $s \in U$ let the smooth matricial map $V_{(U,\xi)} : U \longrightarrow I\!R^{d\times d}$ given by

$$(V_{(U,\xi)}(s))_{ij} = \xi_s(\partial/\partial\psi_i|_s, \partial/\partial\psi_j|_s).$$

Let $t \in M$ and (U, ξ) be a normal chart centered at t. Consider the function $\theta_t : U \longrightarrow I\!R$

$$\theta_t(s) = (det(V_{(U,\xi)}(s)))^{\frac{1}{2}}$$

which is not other than the volume of the parallelepiped $\{\partial/\partial\psi_1|_s, \ldots, \partial/\partial\psi_d|_s\}$. This function is called *the volume density function.* It is not difficult to see that θ_t does not depend on the election of the normal chart and that $\theta_t(s) = \theta_s(t)$, see [11] for instance.

For the Euclidean space $(I\!R^d, \xi_e^d)$ and for the cylinder $(S^1 \times I\!R, \xi_0^1 + \xi_e^1)$, $\theta_t(s) = 1$ for all s, t where the function is well defined. In [11], we calculate the volume density on the sphere, in this case,

$$\theta_s(t) = \frac{|\sin(d_\xi(s,t))|}{d_\xi(s,t)} \text{ for } t \neq s, -s \text{ and } \theta_s(s) = 1.$$

2.2.2 The nonparametric estimators

Let $(y_1, t_1), \ldots, (y_n, t_n)$ be i.i.d random objects that take values on $I\!R \times M$. In order to estimate $r(\tau) = E(y|t = \tau)$, Pelletier [15] proposed a nonparametric kernel type estimators. The proposal introduced by Pelletier was build an analogue to the kernel type estimator on (M, ξ) considering the distance d_ξ on M and the volume density function of (M, ξ) in order to take into account the curvature of the manifold. More precisely, the nonparametric estimator can be defined as,

$$r_n(t) = \sum_{i=1}^{n} w_{n,h}(t, t_i) y_i \tag{3}$$

with $w_{n,h}(t, t_i) = \theta_t^{-1}(t_i)K(d_\xi(t, t_i)/h)/[\sum_{k=1}^n \theta_t^{-1}(t_k)K(d_\xi(t, t_k)/h)]^{-1}$ where $K : I\!R \to I\!R$ is a non-negative function, $\theta_t(s)$ the volume density function on (M, ξ) and the bandwidth h is a sequence of real positive numbers such that $\lim_{n\to\infty} h = 0$ and $h < inj_\xi M$, for all n. This last requirement on the bandwidth guarantees that (3) is well defined for all $t \in M$. Pelletier (2006) studied some properties of this estimators such as the asymptotic pointwise mean squared error. On the other hand, Henry and Rodriguez [11] proposed a robust version that generalized these estimators and studied some asymptotic properties.

3 Asymptotic behavior

We begin with the following assumptions required to derive the large sample properties of the proposed estimators. In this section, we study the asymptotic behavior of the regression parameter estimator and the nonparametric component of the model under the following conditions.

*H*1. Let M_0 be a compact set on M such that: f is a bounded function such that $\inf_{t\in M_0} f(t) = A > 0$ and $\inf_{t,s\in M_0} \theta_t(s) = B > 0$.

*H*2. The sequence h is such that $nh^4 \to 0$ and $nh_n^d/\log n \to \infty$ as $n \to \infty$.

*H*3. $K : I\!R \to I\!R$ is a bounded nonnegative Lipschitz function of order one, with compact support $[0,1]$ satisfying: $\int_{I\!R^d} K(\|\mathbf{u}\|)d\mathbf{u} = 1$, $\int_{I\!R^d} \mathbf{u}K(\|\mathbf{u}\|)d\mathbf{u} = \mathbf{0}$ and $0 < \int_{I\!R^d} \|\mathbf{u}\|^2 K(\|\mathbf{u}\|)d\mathbf{u} < \infty$.

*H*4. For any open set U_0 of M such that $M_0 \subset U_0$, the functions g, ϕ_j for $1 \le j \le p$ are of class C^2 on U_0.

*H*5. The errors ε_i and η_{ij} for $1 \le i \le n$ and $1 \le j \le p$ are independent and $E|\varepsilon_1|^r + \sum_{j=1}^p E|\eta_{1j}|^r < \infty$ for $r \ge 3$, $\sigma_\varepsilon^2 = \text{var}(\varepsilon_1) > 0$ and $\Sigma = E(\eta_1^{\mathrm{T}}\eta_1)$ is a positive definite matrix.

We are now ready to establish the large sample properties of the estimators. The theorem below provides the asymptotic normality of the $\widehat{\boldsymbol{\beta}}$ and the rate of convergence of the nonparametric estimator $\widehat{g}(t)$.

Theorem 3.1. Under $H1$ to $H5$

i) $\sqrt{n}\,(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N(0, \sigma_\varepsilon^2 \Sigma^{-1})$.

ii) $\sup_{t\in M_0} |\widehat{g}(t) - g(t)| = O(h^2) + O(\sqrt{\log n/nh^d})$.

Note that this theorem is consistent with the corresponding results in the Euclidean case.

Remark 3.1. The fact that $\theta_t(t) = 1$ for all $t \in M$ guarantees that the bound of θ in $H1$ holds. The assumptions $H2$ and $H3$ are standard assumptions when kernel estimators are considered.

In many practical problems it is interesting to make inference on the regression parameter, such as construction of confidence regions or hypothesis test. The obtained asymptotic distribution can be used to construct a Wald–type statistic, more precisely, to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. It seems natural to test H_0 through the Wald–type statistic

$$T_n = \frac{n}{\widehat{\sigma}_\varepsilon^2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}}\widehat{\Sigma}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

where $\widehat{\sigma}_\varepsilon^2$ and $\widehat{\Sigma}$ are estimates of σ_ε^2 and Σ, respectively. For example,

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \widehat{g}(t_i))^2 \quad \text{and} \quad \widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \widehat{\boldsymbol{\phi}}(t))^{\mathrm{T}}(\mathbf{x}_i - \widehat{\boldsymbol{\phi}}(t))$$

may be the considered estimators.

Lemma A.2 and Remmark A.5 in the Appendix, show that $\widehat{\Sigma}$ and $\widehat{\sigma}_\varepsilon^2$ are consistent of Σ and σ_ε^2, respectively. Therefore, under the null hypothesis $T_n \xrightarrow{\mathcal{D}} \chi_p^2$. Thus, if we test H_0 at a significance level α, the Wald–test rejects H_0 when $T_n > \chi_{p,\alpha}^2$ where $\chi_{p,\alpha}^2$ is the corresponding $1-\alpha$ quantile of the χ_p^2.

4 Case studies

4.1 Selection of the smoothing parameter

An important issue in any smoothing procedure is the choice of the smoothing parameter. Under a nonparametric regression model with carriers in an Euclidean space, i.e., when (M,ξ) is $(I\!R^d, \xi_e^d)$, two commonly used approaches are L^2 cross–validation and plug–in methods. In this section, we included a cross-validation method for the choice of the bandwidth in the case of partly linear models. The asymptotic properties of data–driven estimators require further careful investigation and are beyond the scope of this paper.

The cross-validation method constructs an asymptotically optimal data-driven bandwidth, and thus adaptive data-driven estimators, by minimizing

$$CV(h) = \sum_{i=1}^{n}[(y_i - \widehat{\phi}_{0,-i,h}(t_i)) - (\mathbf{x}_i - \widehat{\boldsymbol{\phi}}_{-i,h}(t_i))^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}]^2,$$

where $\widehat{\phi}_{0,-i,h}(t)$ and $\widehat{\boldsymbol{\phi}}_{-i,h}(t) = (\widehat{\phi}_{1,-i,h}(t),\ldots,\widehat{\phi}_{p,-i,h}(t))$ denote the nonparametric estimators computed with bandwidth h using all the data expect the $i-$th observation and $\widetilde{\boldsymbol{\beta}}$ minimize $\sum_{i=1}^{n}[(y_i - \widehat{\phi}_{0,-i,h}(t_i)) - (\mathbf{x}_i - \widehat{\boldsymbol{\phi}}_{-i,h}(t_i))^{\mathrm{T}}\boldsymbol{\beta}]^2$ in $\boldsymbol{\beta}$.

4.2 Simulation study

To evaluate the performance of the estimation procedure, we conduct a simulation study. We consider two models in two different Riemannian manifolds, the sphere and the cylinder endowed with the metric induced by the canonical metric of $I\!R^3$. We performed 1000 replications of independent samples of size $n = 50, 100, 150, 200$ according to the following models:

Sphere case: The variables (y_i, x_i, t_i) for $1 \le i \le n$ were generated as

$$y_i = \beta\, x_i + \exp\{-(t_{i1} + 2t_{i2} + t_{i3})^2\} + \varepsilon_i \quad \text{and} \quad x_i = t_{i1} + t_{i2} + t_{i3} + \eta_i$$

where $t_i = (\cos(\theta_i)\cos(\gamma_i), \sin(\theta_i)\cos(\gamma_i), \sin(\gamma_i))$ with θ_i and γ_i follow a von Mises distribution with means 0 and π and concentration parameters 3 and 5, respectively. In this case, the functions $g(t)$ and $\boldsymbol{\phi}(t)$ are equal to $\exp\{[(1,1,1)^{\mathrm{T}}t]^2\}$ and $(1,1,1)^{\mathrm{T}}t$, respectively.

Cylinder case: The variables (y_i, x_i, t_i) for $1 \leq i \leq n$ were generated as

$$y_i = \beta\, x_i + s_i^2 + \sin(\theta_i) + \varepsilon_i \quad \text{and} \quad x_i = \exp(\theta_i) + \eta_i$$

where $t_i = (\cos(\theta_i), \sin(\theta_i), s_i)$ with the variables θ_i follow a von Mises distribution with mean π and concentration parameter 3 and the variables s_i are uniform in $(-2, 2)$, i.e. t_i have support in the cylinder with radius 1 and height between $(-2, 2)$. Note that in this model, $g(t) = (e_3^{\mathrm{T}}t)^2 + \sin(\arctan(e_2^{\mathrm{T}}t/e_1^{\mathrm{T}}t))$ and $\boldsymbol{\phi}(t) = \exp(\arctan(e_2^{\mathrm{T}}t/e_1^{\mathrm{T}}t))$ where e_i for $i = 1, 2, 3$ are the canonical vectors of $I\!R^3$.

In all cases, the regression parameter $\boldsymbol{\beta}$ was taken equal 5 and the errors ε_i and η_i are i.i.d. normal with mean 0 and standard deviation 1. In the smoothing procedure, the kernel was taken as $K(u) = 30\, u^2(1-u)^2 I_{(0,1)}(u)$. Respect to the selection of the smoothing parameter, we apply the cross validation procedure described in Section 4.1. Furthermore, to analyzed the effect of the bandwidth in the estimation procedure, we computed the estimators on a grid of bandwidths. We consider an equispaced grid of length 10 between 0.5 and π in the sphere case, and between 1 and 2π in the cylinder case. The distance d_ξ for these manifolds can be found in [12] and [11] and the volume density function in Section 2.2.1. Table 4.2.1 and Table 4.2.3 give the mean, standard deviations (sd), mean square error (MSE) for the regression estimates of β and the mean of the mean square error of the regression function g over the 1000 replications when we consider the cross validation procedure. Table 4.2.2 and Table 4.2.4 report the mean square error (MSE) for the regression estimates of β and the mean of the mean square error of the regression function g over the 1000 replications for each bandwidth considered.

	mean($\widehat{\boldsymbol{\beta}}$)	sd($\widehat{\boldsymbol{\beta}}$)	MSE($\widehat{\boldsymbol{\beta}}$)	MSE($\widehat{g}$)
$n = 50$	5.1103	0.1411	0.0321	0.2102
$n = 100$	5.1147	0.0993	0.0230	0.1846
$n = 150$	5.1023	0.0907	0.0187	0.1709
$n = 200$	5.1024	0.0807	0.0170	0.1644

Table 4.2.1: Performance of $\widehat{\boldsymbol{\beta}}$ and $\widehat{g}$ in the sphere case using cross-validation.

	bandwidth									
	0.5	0.793	1.087	1.380	1.674	1.967	2.261	2.554	2.848	3.141
					$n = 50$					
MSE($\widehat{\boldsymbol{\beta}}$)	0.0311	0.0248	0.0218	0.0207	0.0253	0.0256	0.0293	0.0300	0.0325	0.0337
MSE($\widehat{g}$)	0.2715	0.1698	0.1620	0.1772	0.1929	0.2038	0.2102	0.2126	0.2112	0.2130
					$n = 100$					
MSE($\widehat{\boldsymbol{\beta}}$)	0.0119	0.0116	0.0102	0.0118	0.0146	0.0164	0.0191	0.0209	0.0211	0.0233
MSE($\widehat{g}$)	0.1557	0.1134	0.1256	0.1527	0.1747	0.1861	0.1916	0.1936	0.1918	0.1891
					$n = 150$					
MSE($\widehat{\boldsymbol{\beta}}$)	0.0082	0.0071	0.0069	0.0079	0.0112	0.0143	0.0170	0.0172	0.0192	0.0202
MSE($\widehat{g}$)	0.1142	0.0932	0.1150	0.1437	0.1672	0.1792	0.1889	0.1866	0.1863	0.1832
					$n = 200$					
MSE($\widehat{\boldsymbol{\beta}}$)	0.0056	0.0052	0.0052	0.0072	0.0099	0.0127	0.0155	0.0166	0.0183	0.0196
MSE($\widehat{g}$)	0.0931	0.0820	0.1092	0.1388	0.1640	0.1768	0.1835	0.1856	0.1833	0.1818

Table 4.2.2: Performance of $\widehat{\boldsymbol{\beta}}$ and $\widehat{g}$ in the sphere case for different bandwidths.

	mean($\widehat{\boldsymbol{\beta}}$)	sd($\widehat{\boldsymbol{\beta}}$)	MSE($\widehat{\boldsymbol{\beta}}$)	MSE($\widehat{g}$)
$n = 50$	4.9836	0.0166	0.0005	0.1756
$n = 100$	4.9866	0.0119	0.0003	0.1545
$n = 150$	4.9873	0.0097	0.0003	0.1473
$n = 200$	4.9877	0.0086	0.0002	0.1397

Table 4.2.3: Performance of $\widehat{\boldsymbol{\beta}}$ and $\widehat{g}$ in the cylinder case using cross-validation.

	bandwidth									
	1	1.454	1.908	2.362	2.816	3.270	3.724	4.178	4.632	5.086
					$n = 50$					
MSE($\widehat{\boldsymbol{\beta}}$)	0.0006	0.0005	0.0005	0.0006	0.0007	0.0007	0.0007	0.0008	0.0008	0.0007
MSE($\widehat{g}$)	0.9471	0.5392	0.2428	0.1936	0.2413	0.2786	0.3049	0.3122	0.3158	0.3089
					$n = 50$					
MSE($\widehat{\boldsymbol{\beta}}$)	0.0003	0.0003	0.0004	0.0004	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
MSE($\widehat{g}$)	0.5914	0.4312	0.2269	0.1797	0.2283	0.2622	0.2855	0.2930	0.2935	0.2987
					$n = 50$					
MSE($\widehat{\boldsymbol{\beta}}$)	0.0002	0.0002	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
MSE($\widehat{g}$)	0.4983	0.3567	0.2214	0.1740	0.2232	0.2592	0.2822	0.2891	0.2903	0.2917
					$n = 50$					
MSE($\widehat{\boldsymbol{\beta}}$)	0.0002	0.0002	0.0003	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
MSE($\widehat{g}$)	0.4476	0.3222	0.2191	0.1713	0.2210	0.2573	0.2786	0.2858	0.2857	0.2882

Table 4.2.4: Performance of $\widehat{\boldsymbol{\beta}}$ and $\widehat{g}$ in the cylinder case for different bandwidths.

In Tables 4.2.1 and 4.2.3 we can see a good behavior of the estimators in the two considered schemes. In both cases the mean of the mean square error of the parametric and nonparametric estimators is small and reflect a good performance of the proposed estimators. As we expected, it is interesting to note that when the sample size increases the performance of the mean square errors of $\boldsymbol{\beta}$ and g is even better. This effect is also summarized in Tables 4.2.2 and 4.2.4. In these Tables we can see that the behavior of the proposed estimator is stable across of the bandwidths considered, except, as we expected, when the bandwidths and the samples size are small.

4.3 Application to real data

In this Subsection, we applied a partly linear model to an environment dataset in order to study the atmospheric SO_2 pollution incidents. The variables included in the study are the direction and the speed of the wind, the temperature and the SO_2 concentration in the meteorologic station at Villalba (Lugo in Galicia, Spain). In [9] and [17], the authors applied models to the prediction of atmospheric SO_2 pollution incidents in the vicinity of the coal/oil-fired power station at As Pontes, A Coruña, Galicia, Spain. They observed that the prediction of the SO_2 time series is challenging, because they consist in near-zero values interrupted between a few days and several weeks by episodes lasting a few hours at random in which values rise to high levels and then fall back to zero.

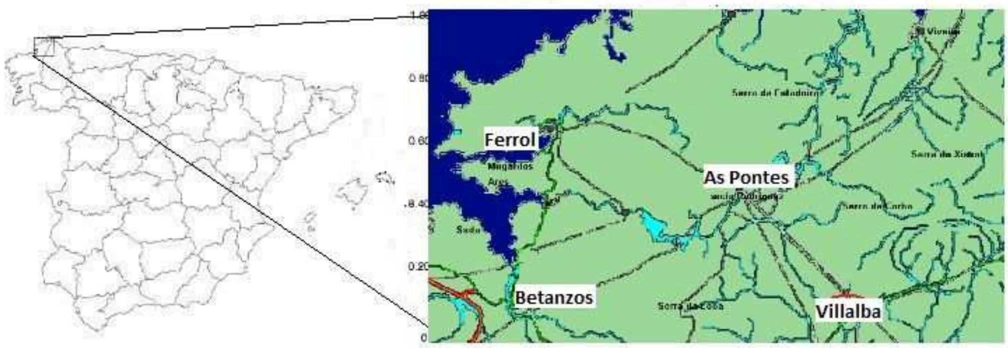


Figure 4.3.1: Location of Villalba, Galicia Spain.

Therefore, in order for predictions to be based on data representing a reasonably large number of incidents, we reproduced the construction of the historical matrix introduced in [17]. The authors took as samples 1000 rows of the historical matrix that was constructed and updated as follows.

First, they determined the range of 2-hour means observed during the previous 2 years. Then, they divided the non-near-zero region of this range into 10 strata containing approximately equal numbers of values, randomly selected 100 values y_l from each stratum, and associated each with the corresponding predictor values x_l and t_l. The 1000 $(1+p+d)$–tuplets so formed made up the 1000 rows of the seed of the historical matrix. Thereafter, during on-line processing, the historical matrix was updated whenever a non-near-zero y_k occurred by identifying the stratum to which y_k belonged and substituting x_k and t_k for the oldest row of the historical matrix a time $k-1$ belonging to this stratum.

The data was recorded daily in each minute during the year 2009 and we was considered a 1500–row historical matrix. The variables that we considered in the model were

y_i	SO_2 emission is measured in $\mu g/m^3$
x_{1i}	SO_2 emission in the instant $i-30$
x_{2i}	SO_2 emission difference between the instant $i-35$ and $i-30$
x_{3i}	the temperature in °C
t_{1i}	wind direction in radians from the north
t_{2i}	wind speed in m/s

Table 4.3.1: Environment variables considered in the model.

Note that the variables $t_i = (t_{1i}, t_{2i})$ have support in the cylinder. The maximum of the wind speed in this cases is 7.7 then we consider that the variable t belongs in the cylinder of high between 0 and 10. Therefore, we modeled the response variable using the following model $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + g(t_i) + \varepsilon_i$.

In the smoothing procedure, we considered the same kernel that we use in the simulation study and we choose the bandwidth using a cross validation procedure. Because of the computational burden of the cross-validation method, and because there is really no need to use this method with a sample as large as 1500, we also determined h by the split sample method, i.e. by dividing the historical matrix into a 750-member training set with odd index and a 750-member validation set with even index, and taking for h the value minimizing

$$SV(h) = \sum_{i=1}^{[n/2]} [(y_{2i} - \widehat{\phi}_{0,E,h}(t_{2i})) - (\mathbf{x}_{2i} - \widehat{\boldsymbol{\phi}}_{E,h}(t_{2i}))^{\mathrm{T}} \widetilde{\boldsymbol{\beta}}]^2.$$

where $\widehat{\boldsymbol{\phi}}_{E,h}(t) = (\widehat{\phi}_{1,E,h}(t), \ldots, \widehat{\phi}_{p,E,h}(t))$ and $\widehat{\phi}_{0,E,h}(t)$ denote the nonparametric estimators computed with bandwidth h using the data with odd index and $\widetilde{\boldsymbol{\beta}}$ minimize $\sum_{i=1}^{[n/2]}[(y_{2i} - \widehat{\phi}_{0,E,h}(t_{2i})) - (\mathbf{x}_{2i} - \widehat{\boldsymbol{\phi}}_{E,h}(t_{2i}))^{\mathrm{T}} \boldsymbol{\beta}]^2$ in $\boldsymbol{\beta}$. In this case the selected bandwidth was $h_{sv} = 2.1$. Table 4.3.2 reports the estimates values of the regression parameters. Figure 4.3.2. and Figure 4.3.3. show the estimate of the regression function over a grid of 1200 points in the cylinder. The graphics are quite similar. The differences arise only in the form that we plot the function. Figure 4.3.4. shows the residuals vs. order in each case of the nonparametric model considered. In particular, Figure a) shows the residuals $x_{1i} - \widehat{\phi}_1(t_i)$ against i, b) and c) the same plots related with x_{2i} and x_{3i}, respectively and d) $y_i - \widehat{\phi}_0(t_i)$ against i. These plots allow to verify the hypothesis of independence assumed over the errors.

$\widehat{\boldsymbol{\beta}}_1$	$\widehat{\boldsymbol{\beta}}_2$	$\widehat{\boldsymbol{\beta}}_3$
0.7988	0.5766	-0.0013

Table 4.3.2: Estimates of regression parameter.

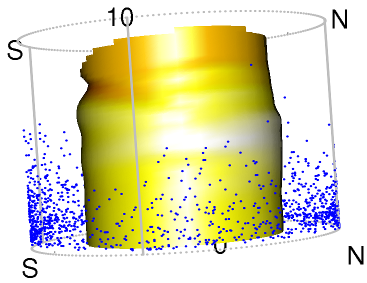


Figure 4.3.2: Estimates of the regression function over the cylinder.

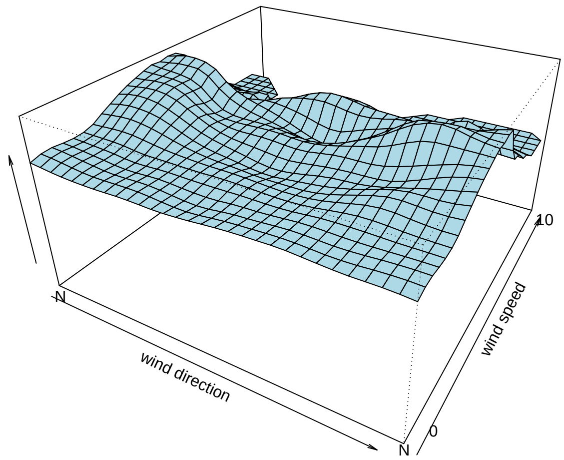
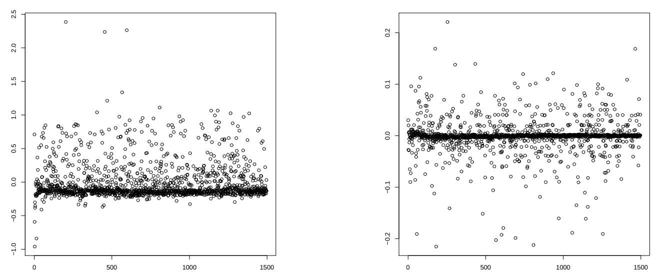


Figure 4.3.3: Estimates of the regression function projected in the plane.

$a)$ $b)$



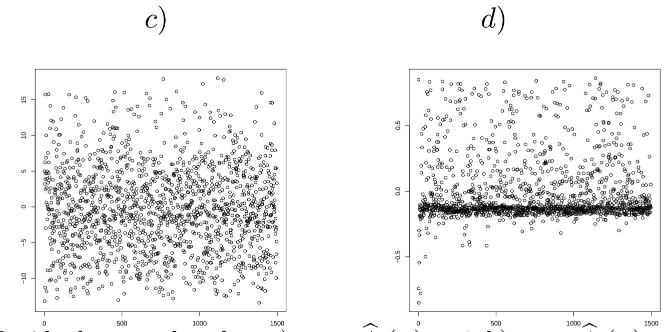


Figure 4.3.4: Residuals vs. order plot. a) $x_{1i} - \widehat{\phi}_1(t_i)$ vs i, b) $x_{2i} - \widehat{\phi}_2(t_i)$ vs i, c) $x_{2i} - \widehat{\phi}_2(t_i)$ vs i and d) $y_i - \widehat{\phi}_0(t_i)$ against i.

As we mentioned in Section 3, we made a Wald–type test for the hypothesis H_0 : $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. If we consider the hypothesisH_0 : $\boldsymbol{\beta} = (0.8, 0.6, 0)$, the test statistic and the p–value were $T_n = 1.59$ and p–value$= 0.6594$, respectively. Therefore, we do not reject the null hypothesis. Thus, we can conclude that the temperature seems not have significance impact factor in the pollution incidents.

To evaluate the performance of the partial linear model, we consider a full nonparametric model to explain y_i based on the variables x_{1i}, x_{2i} and x_{3i} through an unknown function η. In this case we use the Naradaya-Watson estimator with quadratic kernel. In the smoothing procedure as matrix of the bandwidth, we consider a multiple of the identity matrix. We compare the prediction error for both models computing, in the case of the full nonparametric model, $EP(h) = \sum_{i=1}^{[n/2]}[(y_{2i} - \widehat{\eta}(x_{1,2i}, x_{2,2i}, x_{3,2i}))]^2$ for a grid of 100 equispaces bandwidth between 0.1 and 4. For the partial linear model we compute the $SV(h)$ for the same grid of bandwidth. As we can see in Figure 4.3.2. a), the partial linear model has a better level of predictive than the full nonparametric model. The comparison of the errors in the figure seems to be in different scale and the proposed method seems to be unsensitive with respect to the bandwidth. The Figure 4.3.2. b) shows only the prediction error of the proposed estimator. Here, we can see that the behavior for small values of bandwidth seems more unstable but better than in the full nonparametric model.

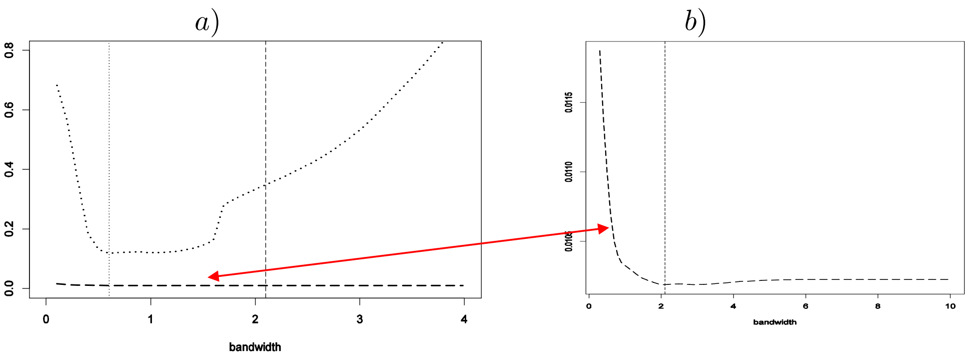


Figure 4.3.2: Comparative of the errors: the dotted line corresponds to the full nonparametric model and the dashed line to the partly linear model. The vertical lines correspond to the optimal bandwidths in each models.

Acknowledgments

We wish to thank three anonymous referees and the Editor for valuable comments which led to an improved version of the original paper. We would like to thank María Leyenda Rodríguez for the preparation of the dataset. This work was done when the second and third authors were visiting the University of Santiago de Compostela, they are very grateful to all statistics group for their kind hospitality. This research was partly supported by Grants X-018 from Universidad de Buenos Aires, PIP 1122008010216 from CONICET and PICT - 00821 from ANPCYT, Argentina, and also by Spanish Grants MTM2008-03010/MTM of the Ministerio Español de Ciencia e Innovación and XUGA Grants PGIDIT07PXIB207031PR.

A Appendix

Lemma A.1: Let $\widetilde{\phi}_j(t) = \phi_j(t) - \sum_{i=1}^n w_{n,h}(t,t_i)x_{ij}$ *for* $1 \le j \le p$ *and* $\widetilde{\phi}_0(t) = \phi_0(t) - \sum_{i=1}^n w_{n,h}(t,t_i)y_i$*. Under* $H1$ *to* $H4$ *we have that*

$$\sup_{t\in M_0} |\widetilde{\gamma}(t)| = O(h^2) + O\left(\sqrt{\log n/nh^d}\right) \quad a.s$$

where $\widetilde{\gamma} \in \{\widetilde{\phi}_j; \quad 0 \le j \le p\}$.

Proof of Lemma A.1: Let $\gamma \in \{\phi_j; \quad 0 \le j \le p\}$. To fix ideas, we consider $\gamma = \phi_j(t)$ for $1 \le j \le p$, the case $j = 0$ is similar, then

$$\begin{aligned} \widetilde{\gamma}(t) &= \phi_j(t) - \sum_{i=1}^n w_{n,h}(t,t_i)x_{ij} \\ &= \frac{\frac{1}{nh^d}\sum_{i=1}^n \frac{1}{\theta_t(t_i)} K(d_\xi(t,t_i)/h)(\phi_j(t) - x_{ij})}{\frac{1}{nh^d}\sum_{i=1}^n \frac{1}{\theta_t(t_i)} K(d_\xi(t,t_i)/h)} \end{aligned}$$

By $H1$, $\inf_{t\in M} \frac{1}{h^d} E\left(\frac{1}{\theta_t(t_1)}K(d_\xi(t,t_1)/h)\right) \ge A > 0$ and the strong uniform consistency of $\widehat{f}_n(t) = (nh^d)^{-1}\sum_{k=1}^n \theta_t^{-1}(t_k)K(d_\xi(t,t_k)/h)$ obtained in [12], we can concentrate only the numerator of $\widetilde{\gamma}(t)$.

Using the results obtained in [15] and [12] we have that,

$$\sup_{t\in M_0} \left| E\left(\frac{1}{nh^d}\sum_{i=1}^n \frac{1}{\theta_t(t_i)} K(d_\xi(t,t_i)/h)(\phi_j(t) - x_{ij})\right)\right| = O(h^2) \tag{4}$$

$$\sup_{t\in M_0} \left| \text{var}\left(\frac{1}{nh^d}\sum_{i=1}^n \frac{1}{\theta_t(t_i)} K(d_\xi(t,t_i)/h)x_{ij}\right)\right| = O(nh^d) \tag{5}$$

Finally, the proof follows in analogous way that the proof of Lemma 3.1 in [8].□

Lemma A.2: Under $H1$ *to* $H4$ *we have that* $n^{-1}\sum_{i=1}^n \widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_i \xrightarrow{p} \Sigma$ *where* $\widetilde{\mathbf{x}}_i = \mathbf{x}_i - \widehat{\boldsymbol{\phi}}(t_i)$.

Proof of Lemma A.2: The element l,s of $n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_i$ can be written as

$$(n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_i)_{ls} = n^{-1}\left(\sum_{i=1}^{n}\eta_{il}\eta_{is} + \sum_{i=1}^{n}\widetilde{\phi}_l(t_i)\eta_{is} + \sum_{i=1}^{n}\widetilde{\phi}_s(t_i)\eta_{il} + \sum_{i=1}^{n}\widetilde{\phi}_l(t_i)\widetilde{\phi}_s(t_i)\right)$$

where $\widetilde{\phi}_j(t) = \phi_j(t) - \widehat{\phi}_j(t)$. We need to show that all terms except the first term converge to zero. Applying the strong law of large numbers we get that $n^{-1}\sum_{i=1}^{n}\eta_{il}\eta_{is} \xrightarrow{p} \Sigma_{ls}$.

Since Lemma A.1, the fact that $n^{-1}\sum_{i=1}^{n}\eta_{il}^2 \xrightarrow{p} \Sigma_{ll}$ and using the Cauchy-Schwartz inequality we get the result.□

Lemma A.3: Under $H1$ to $H3$, we have that $\sup_{t\in M}\max_{1\leq j\leq n}|w_{n,h}(t,t_j)| = O((nh^d)^{-1})$.

Proof of Lemma A.3:

Note that

$$w_{n,h}(t,t_j) = \frac{\frac{1}{nh^d}\frac{1}{\theta_t(t_j)}K(d_\xi(t,t_j)/h)\left[\frac{1}{h^d}E\left(\frac{1}{\theta_t(t_1)}K(d_\xi(t,t_1)/h)\right)\right]^{-1}}{\frac{1}{nh^d}\sum_{i=1}^{n}\frac{1}{\theta_t(t_i)}K(d_\xi(t,t_i)/h)\left[\frac{1}{h^d}E\left(\frac{1}{\theta_t(t_1)}K(d_\xi(t,t_1)/h)\right)\right]^{-1}}$$

According to [12], we have that

$$\sup_{t\in M}\left|\frac{1}{nh^d}\sum_{i=1}^{n}\frac{1}{\theta_t(t_i)}K(d_\xi(t,t_i)/h)\left[\frac{1}{h^d}E\left(\frac{1}{\theta_t(t_1)}K(d_\xi(t,t_1)/h)\right)\right]^{-1} - 1\right| = o(1) \quad \text{a.s.} \qquad (6)$$

$$\inf_{t\in M}\frac{1}{h^d}E\left(\frac{1}{\theta_t(t_1)}K(d_\xi(t,t_1)/h)\right) \geq A > 0. \qquad (7)$$

Then by (6) and (7) and the boundedness of K and θ_t, the lemma holds.□

Remark A.4: Note that by Lemmas A.1 and A.3 and using Lemma A.1 in [13]; we have that

$$\max_{1\leq i\leq n}|\gamma(t_i) - \sum_{k=1}^{n}w_{n,h}(t_i,t_k)\gamma(t_k)| = O(h^2) + O\left(\sqrt{\log n/nh^d}\right) \qquad \text{a.s.}$$

for any $\gamma \in \{\phi_j;\quad 0\leq j\leq p\}$.

Proof of Theorem 3.1:

i) We can write $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_i)^{-1}n^{-1/2}\left[A_{1n} - A_{2n} + A_{3n}\right]$ where

$$A_{1n} = \sum_{i=1}^{n}\widetilde{\mathbf{x}}_i g^*(t_i) \quad A_{2n} = \sum_{i=1}^{n}\widetilde{\mathbf{x}}_i\left(\sum_{i=1}^{n}w_{n,h}(t_i,t_j)\varepsilon_j\right) \quad A_{3n} = \sum_{i=1}^{n}\widetilde{\mathbf{x}}_i\varepsilon_i$$

and $g^*(t) = g(t) - \sum_{i=1}^{n}w_{n,h}(t,t_i)g(t_i)$. Using Lemmas A.1 to A.3, the asymptotic behavior of A_{1n}, A_{2n} and A_{3n} can be obtained in the same way that in [2].

Specifically, considering the assumptions imposed on h, we can obtain

$$\begin{aligned}
A_{1n} &= O(nh^4 + h^{-d}\log^2 n) + O(n^{1/2}h^2\log n + h^{-d/2}\log^2 n) + O(n^{1/2}h^2h^{-d/2}\log n) \\
&+ O(h^{-d}\log^2 n) = o(n^{1/2}), \\
A_{2n} &= O(n^{1/2}h^2h^{-d/2}\log n + h^{-d}\log^2 n) + O(h^{-d/2}\log^2 n) + O(h^{-d}\log^2 n) = o(n^{1/2})
\end{aligned}$$

and

$$A_{3n} = O(n^{1/2}h^2\log n) + O(h^{-d/2}\log^2 n) + \sum_{i=1}^{n}\eta_i\varepsilon_i + O(h^{-d/2}\log^2 n) = \sum_{i=1}^{n}\eta_i\varepsilon_i + o(n^{1/2}).$$

Finally, the central limit theorem gives the desired result.

ii) Note that $\widehat{g}(t) - g(t) = \widehat{\phi}_0(t) - \phi_0(t) + (\widehat{\boldsymbol{\phi}}(t) - \boldsymbol{\phi}(t))^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \boldsymbol{\phi}(t)^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Therefore Lemma A.1 and the part i) of this theorem allow to complete the proof. □

Remark A.5: Finally, we note that Lemma A.1 and Theorem 3.1 state the consistency of the estimator $\widehat{\sigma}_\varepsilon^2$ *of* σ_ε^2 *defined in Section 3. More precisely, note that*

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i + \mathbf{x}_i^{\mathrm{T}}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + g(t_i) - \widehat{g}(t_i))^2.$$

Thus, if we distribute the terms, all of them converge to zero by Lemma A.1 and Theorem 3.1 except $\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2$ *since by the strong law of large numbers converges to* σ_ε^2.

References

[1] *Aneiros-Pérez, G. and Quintela del Río, G. Plug-in bandwidth choice in partial linear regression models with autoregressive errors. J. Statist. Planning and Inference.* **57**, 23-48. (2002).

[2] Aneiros-Pérez, G. and Vieu, F. Semi-Functional partial linear regression. *Statistics & Probability Letters.* **76**, 1102-1110. (2006).

[3] Besse, A. Manifolds all of whose Geodesics are Closed. *Springer-Verlag.* (1978).

[4] Bhattacharya, R. and Patrangenaru, V. Nonparametric estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference.* **108,** 23-35. (2002).

[5] Chen, H. Convergence rates for parametric components in a partly linear model. *Ann. Statist.* **16,** 136-146. (1988).

[6] Do Carmo, M. Riemannian Geometry. *Projeto Euclides, IMPA.* (2005).

[7] Engle, R.; Granger, C.; Rice, J. and Weiss, A. Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81,** 310-320. (1986).

[8] Ferraty, F. and Vieu, F. Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *Journal of Nonparametric Statistics.* **16,** 111-125. (2004).

[9] García-Jurado, I.; Gonzalez-Manteiga, W.; Prada-Sanchez, J.M., Febrero-Bande, M. and Cao, R. (1995). Predicting using Box-Jenkins, nonparametric and bootstrap techniques. *Technometrics.* **37,** 303-310.

[10] Hendriks, H. and Landsman, Z. Asymptotic data analysis on manifolds. *Ann. Statist.* **35, 1,** 109-131. (2007).

[11] Henry, G. and Rodriguez, D. Robust Nonparametric Regression on Riemannian Manifolds. *Journal of Nonparametric Statistics.* **21, 5,**611-628. (2009).

[12] Henry, G. and Rodriguez, D. (2009). Kernel Density Estimation on Riemannian Manifolds: Asymptotic Results. *Journal Math. Imaging Vis.* **43,** 235-639.

[13] Liang, H. Asymptotic normality of nonparametric part in partially linear models with measurement error in the nonparametric part. *Journal of statistical planning and inference.* **86,** 51-62. (2000).

[14] Mardia, K. Statistics of Directional Data. *Academic Press, London.* (1972).

[15] Pelletier, B. Nonparametric regression estimation on closed Riemannian manifolds. *Journal of Nonparametric Statistics.* **18,** 57-67. (2006).

[16] Pennec, X. Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal Math. Imaging Vis.* **25,** 127-154. (2006).

[17] Prada-Sanchez, J.M.; Febrero, M.; Cotos-Yañez, T.; Gonzalez-Manteiga, W.; Bermudez-Cela, J. and Lucas-Dominguez, T. Prediction of SO_2 pollution incidents near a power station using partially linear models and an historical matrix of predictor-response vectors. *Environmetrics.* **11,** 209-225. (2000).

[18] Speckman, P. Kernel smoothing in partial linear models. *J. Roy. Statist. Soc.* Ser. B. **50**, 413-436. (1988).