No effect and lack-of-fit permutation tests for functional regression

Hervé Cardot[1], Luboš Prchal[2,3], and Pascal Sarda[3]

[1] CESAER, 26, bd Docteur Petitjean, BP 87999, 21079 Dijon Cedex, France

[2] Univerzita Karlova v Praze, Katedra pravděpodobnosti a statistiky, Sokolovská 83, 18675 Praha 8, Czech Republic

[3] Université Paul Sabatier, Laboratoire de Statistique et Probabilités, 118, route de Narbonne, 31062 Toulouse Cedex, France

Summary

This paper proposes statistical procedures to check if a real-valued covariate X has an effect on a functional response $Y(t)$. A nonparametric kernel regression is considered to estimate the influence of X on $Y(t)$ and two test statistics based on residual sums of squares and smoothing residuals are proposed. Their acceptance levels are determined by means of permutations. The lack-of-fit test for a class of parametric models is then discussed as a consequence of the no effect procedure. Monte Carlo simulations provide an insight into the level and the power of the no effect tests. A study of atmospheric radiation illustrates the behavior of the proposed methods in practice.

Keywords: F-statistic, smoothing residuals, parametric models, atmospheric radiation

1 Introduction

Since the pioneer work of Deville (1974) on harmonic decomposition of random functions, functional data analysis, a new field of statistics combining modern probability theory with computer intensive implementations, has been developed. An overview of statistical problems, available methods and case studies in a functional data setting are given, among others, in monographs by Ramsay and Silverman (2002, 2005) and Ferraty and Vieu (2006).

Much work has been done in the regression setting when one or both, the response Y and the explanatory variable X, are valued in a function space. When a real response depends on a functional predictor, for instance, the functional linear regression is considered by Ramsay and Dalzell (1991) and Cardot *et al.* (1999), generalized linear models are discussed in Marx and Eilers (1999), Escabias *et al.* (2004), Cardot and Sarda (2005), Müller and Stadtmüller (2005), or James (2002) dealing with functional predictors observed in different time moments. A fully nonparametric approach is proposed by Ferraty and Vieu (2002). The case of both functional variables is outlined in Aguilera et *al.* (1999), Chiou *et al.* (2004) and Ramsay and Silverman (2005).

We consider the case of a regression setting where the variations of a functional response are explained by a real-valued predictor. This case has been firstly studied by Cardot (2006) which introduces kernel estimators in the context of the conditional functional principal components analysis. We use the same nonparametric estimator of the regression function, defined in Section 2. This regression problem has some interesting applications as discussed in Cardot (2006) and as illustrated in Section 6 below where we consider atmospheric radiation curves, the (scalar) predictor being time.

The main goal of our study is to test the significance of the covariate effect on the response. Testing procedures of a significant influence of the predictor on the response in the functional linear model are suggested by Cardot *et al.* (2003, 2004), the former based on asymptotic properties of the test statistics, the latter using computer intensive methods. A functional bootstrap, studied by Cuevas and Fraiman (2004), serves in a slightly different testing problem of functional ANOVA (Cuevas *et al.* 2004). In Section 3, two statistics for the no effect test procedure are suggested. At first, the classical Fisher's F-statistic is adapted for the functional setting in a way similar to that proposed by Bowman and Azzalini (1997) in the real-valued response case. An alternative test statistic based on smoothing residuals is also proposed. A permutation approach is applied in order to obtain p-values and the influence of the kernel smoothing parameter and the number of permutations are then discussed.

As a consequence of the no effect test, Section 4 presents the lack-of-fit test, a method enabling to decide whether a specified parametric class of regression functions is adequate for the considered regression problem. In Section 5, Monte Carlo simulations allow to check the good behavior of the two test procedures. A real data analysis concerning changes along time in atmospheric radiation illustrates the performance of the lack-of-fit test in Section 6.

2 Nonparametric regression with functional response

Let us consider a functional variable $Y(t)$ taking values in the separable Hilbert space of square integrable functions $L^2(\mathcal{T})$ defined on the compact interval $\mathcal{T} \subset \mathbb{R}$. The standard inner product $\langle \phi, \psi \rangle = \int_{\mathcal{T}} \phi(t)\psi(t)\,dt, \forall \phi, \psi \in L^2(\mathcal{T})$, and the standard norm $\|\phi\|^2 = \langle \phi, \phi \rangle, \forall \phi \in L^2(\mathcal{T})$, are considered throughout the paper. Let X be a real random variable with values in $\mathcal{X} \subset \mathbb{R}$ defined on the same probability space $(\Omega, \mathcal{A}, P)$ as $Y(t)$ and possibly containing some additional information on $Y(t)$.

Let us consider a relation between X and $Y(t)$ in the form of a regression function $m(t,x)$ expressed as

$$m(t,x) = \mathsf{E}\big[Y(t)|X = x\big], \qquad t \in \mathcal{T}, x \in \mathcal{X}. \tag{1}$$

Having a sample $\big(X_i, Y_i(t)\big), i = 1, \ldots, n$ of i.i.d. realizations of $\big(X, Y(t)\big)$, Cardot (2006) proposed a consistent kernel smoother estimator $\widehat{m}(t,x)$ of the conditional mean (1),

$$\widehat{m}(t,x) = \sum_{i=1}^{n} w_i(x,h) Y_i(t), \qquad t \in \mathcal{T}, x \in \mathcal{X}, \tag{2}$$

where the weights w_i are defined as

$$w_i(x,h) = \frac{K\big((X_i - x)/h\big)}{\sum_{i=1}^{n} K\big((X_i - x)/h\big)},$$

and the real kernel function $K(\cdot)$ is a positive bounded function, symmetric around zero, with compact support.

The bandwidth parameter $h > 0$ controls the smoothness of the estimator. Following the nonparametric regression ideas of Härdle and Marron (1985), we can perform a (data-driven) cross-validation procedure in order to obtain the "optimal" bandwidth h^*, i.e.

$$h^* = \operatorname*{arg\,min}_{h>0} \sum_{i=1}^{n} \big\| Y_i - \widehat{m}^{-i}(X_i) \big\|^2, \tag{3}$$

where $\widehat{m}^{-j}(t,x)$ denotes the kernel estimator (2) calculated from the sample $\big(X_i, Y_i(t)\big), i = 1, \ldots, n$, with the j-th pair $\big(X_j, Y_j(t)\big)$ excluded.

3 No effect test

The smoother (2) enables to quantify the covariate effect X on the variable of interest $Y(t)$, however, it does not provide any information about the statistical significance of the X-Y relation.

As discussed in Cardot (2006), working with the conditional moments and decomposition of $Y(t)$ may positively influence results of the analysis. On the other hand, the conditional models are more complicated from both the theoretical and the computational point of view. Thus, the non-negligible effect of the covariate should be justified before its inclusion into the analysis.

Therefore, we are interested in testing the null hypothesis of *no effect* of X on the conditional mean $m(t, x)$, i.e.

$$\mathsf{H}_0 : \forall x \in \mathcal{X} \quad m(t, x) = \mu(t), \qquad t \in \mathcal{T}, \tag{4}$$

where $\mathsf{E}Y(t) \equiv \mu(t) \in L^2(\mathcal{T})$ is an unknown function. We test against the alternative that the conditional mean $m(t, x)$ is a general function of x, i.e.

$$\mathsf{A}_0 : \exists x \in \mathcal{X} \quad m(t, x) \neq \mu(t), \qquad t \in \mathcal{T}. \tag{5}$$

In the following, we propose two test statistics for the no effect hypothesis: an adaptation of Fisher's F-statistic and an approach based on smoothing residuals.

3.1 F-statistic

In a classical linear model setting, ratio of residual sums of squares (F-statistic) is widely used to test a submodel. Adapting this idea for functional regression we propose a test statistic in the form

$$F = \frac{\mathsf{RSS}_0 - \mathsf{RSS}_1}{\mathsf{RSS}_1}, \tag{6}$$

where residual sums of squares RSS_0 and RSS_1 are taken with respect to the hypothesis and the alternative, respectively.

Introducing an estimator $\widehat{\mu}(t)$ of the mean function $\mu(t)$

$$\widehat{\mu}(t) = \frac{1}{n} \sum_{j=1}^{n} Y_j(t), \qquad t \in \mathcal{T}, \tag{7}$$

and estimating the conditional expectation $m(t, x)$ by the nonparametric smoother (2), the corresponding residual sums of squares take the forms

$$\begin{aligned}\mathsf{RSS}_0 &= \sum_{i=1}^{n} \left\| Y_i - \widehat{\mu} \right\|^2 = \sum_{i=1}^{n} \int_{\mathcal{T}} \big(Y_i(t) - \widehat{\mu}(t)\big)^2 \, dt, \\ \mathsf{RSS}_1 &= \sum_{i=1}^{n} \left\| Y_i - \widehat{m}(X_i) \right\|^2 = \sum_{i=1}^{n} \int_{\mathcal{T}} \big(Y_i(t) - \widehat{m}(t, X_i)\big)^2 \, dt.\end{aligned}$$

The idea is to reject the hypothesis H_0 at a prescribed level $\alpha \in (0,1)$ for "large" values of F, i.e. when the functional regression estimator profits from the knowledge of X and fits the data significantly better than the unconditional sample mean function.

3.2 Smoothing residuals

Let us define the empirical residuals

$$\widehat{e}_i(t) = Y_i(t) - \widehat{\mu}(t), \qquad t \in \mathcal{T}, i = 1, \dots, n.$$

If the null hypothesis H_0 is true, the residuals $\widehat{e}_i(t) \in L^2(\mathcal{T})$ should form a sample of centered uncorrelated functional variables, i.e. $\mathsf{E}\widehat{e}_i(t) = 0, \forall t \in \mathcal{T}, \ \forall i$, and $\mathsf{E}(\widehat{e}_i(s)\widehat{e}_j(t)) = 0, \forall (s,t) \in \mathcal{T} \times \mathcal{T}, \ \forall i \neq j$. Denoting by $\widetilde{m}_e(t,x)$ the kernel smoother (2) applied to the residuals $\widehat{e}_i(t)$, i.e.

$$\widetilde{m}_e(t,x) = \sum_{i=1}^{n} w_i(x,h)\widehat{e}_i(t), \qquad t \in \mathcal{T}, x \in \mathcal{X}, \tag{8}$$

this estimator, under H_0, is close to 0 (the zero function). Thus, being motivated by a real-valued regression case discussed by Hart (1997), we suggest a *smoothing residuals* test statistic

$$R = \sum_{i=1}^{n} \left\| \widetilde{m}_e(X_i) \right\|^2 = \sum_{i=1}^{n} \int_{\mathcal{T}} \left[\widetilde{m}_e(t, X_i)\right]^2 dt. \tag{9}$$

As well as for F-statistic, we reject the no effect null hypothesis, if the value of R is too large.

Remark. The norm of the smoother $\widetilde{m}_e(t,x)$ in the test statistic (9) should, at the first glance, be taken with respect to the variability of residuals. As discussed in the following, such a "standardization" is not necessary with a permutation approach.

3.3 The permutation principle

We have presented two approaches that allow to build test statistics for the null hypothesis (4), however, finding their exact, or at least asymptotic, distributions in order to get critical values seems to be quite a hard task. Therefore, we propose to adopt computationally intensive permutation procedures discussed by *e.g.* Raz (1990), Bowman and Azzalini (1997), or Cardot *et al.* (2004).

These approaches rely on the fact that the pairing of any particular X and $Y(t)$ in the observed sample is, under H_0, entirely random. Hence, the distribution of the proposed statistics can be obtained by calculating their values based on all permutated samples $\big(X_{\pi_k(i)}, Y_i(t)\big), k = 1, \ldots, n!$, where π_k is the k-th permutation of $\{1, \ldots, n\}$. Note that the permutation principle is the way to obtain, at least theoretically, a distribution-free unbiased test, see Lehmann (1959).

Since performing all the $n!$ permutations is almost impossible, we approximate in practice the distribution by K values of the test statistic based on randomly chosen K permutations. The empirical p-value pv is then defined as the proportion of simulated test statistics F_k or R_k which are greater than the one F_{obs}, R_{obs} respectively, observed from the original data set,

$$\mathsf{pv}_F = \frac{1}{K}\sum_{k=1}^{K} \mathbb{I}(F_k > F_{\mathsf{obs}}), \qquad \mathsf{pv}_R = \frac{1}{K}\sum_{k=1}^{K} \mathbb{I}(R_k > R_{\mathsf{obs}}), \tag{10}$$

where $\mathbb{I}$ denotes the indicator function. The null hypothesis is rejected if the empirical p-value pv is smaller then the prescribed level of the test α.

3.4 Bandwidth choice and number of permutations

The estimated p-value (10) depends on the chosen bandwidth parameter h which controls the smoothness of the kernel estimator (2) as well as on the number of permutations K performed to simulate the distribution of the test statistic.

We have already mentioned the cross-validation criterion as a method to obtain "optimal" bandwidth h^*. However, this procedure "optimizes" a value of h with respect to a predictive error which is not always optimal with respect to the testing problem. Therefore, we recommend to perform the permutation test for different values of h in some interval around h^* and base the definitive decision on the "optimal" p-value pv^* and a proportion of a significance trace lying under the level of the test. An illustration of a significance trace encouraging the rejection of the null hypothesis is shown on Figure 1.

To decide how many permutations are sufficient to obtain a reasonable p-value, one can plot a diagram similar to Figure 2. It represents the relation between the number of performed permutations K and 10 estimated p-values for a simulated data set with $n = 40$. One can see that different estimated

p-values become very close for $K > 5000$ with p-values varying about 0.075 (see Table 1 for numerical summary). In other words, a mean of 10 simulated p-values for 5 000 permutations or, equivalently, performing 50 000 permutations seems to give accurate enough estimates of the p-value for this particular data set.

Regarding the variance of the proposed binomially distributed estimators (10), Good (1997) claims that relatively small numbers $K \sim 2000 - 3000$ of permutations ensure a negligible variance of pv_F and pv_R. However, one has to be very careful specially when the p-value is close to the requested level.

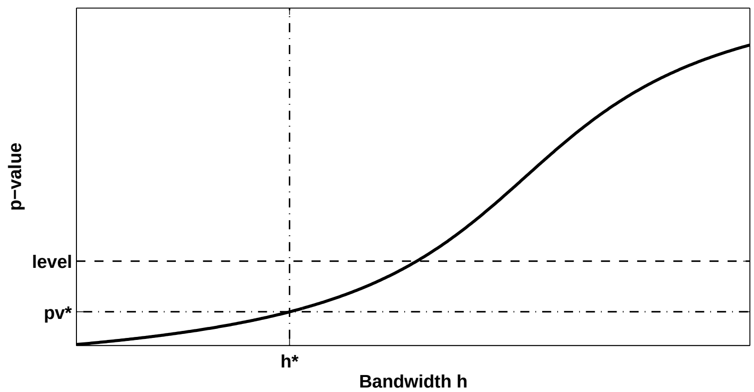

Figure 1: *A significance trace. A final decision whether to reject the hypothesis is recommended to be taken with respect to the optimal p-value pv^* and to the proportion of the significance trace lying under the level of the test.*

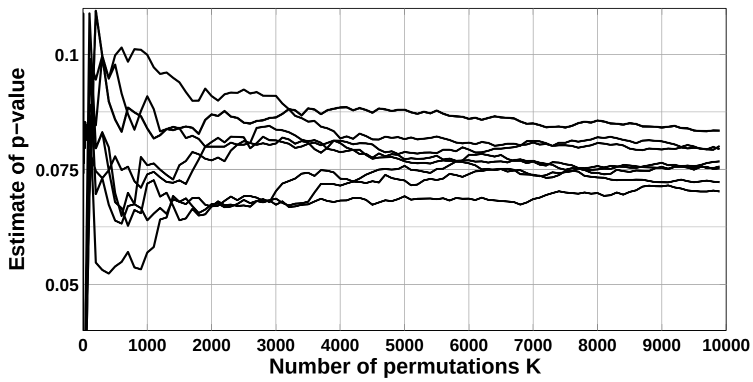

Figure 2: *As the beginning of the simulation may significantly influence the estimate, a huge amount (50 000 in this case) of permutations is recommended to obtain a reliable p-value.*

$K/10^3$	1	2	3	4	5	6	7	8	9	10
mean pv $\times 10^2$	7.79	7.86	7.96	7.96	7.90	7.89	7.85	7.85	7.83	7.80
std pv $\times 10^2$	1.24	0.88	0.82	0.69	0.61	0.54	0.54	0.56	0.48	0.47

Table 1: *An illustration of the mean values and standard deviations of 10 estimated p-values for different numbers of permutations K.*

As

$$\mathsf{pv}_F^{k+1} = \begin{cases} \frac{k}{k+1}\mathsf{pv}_F^k & \text{if } F_{k+1} \leq F_{\mathsf{obs}}, \\ \frac{k}{k+1}\mathsf{pv}_F^k + \frac{1}{k+1} & \text{if } F_{k+1} > F_{\mathsf{obs}}, \end{cases}$$

and $k/(k+1) \sim 1$ with increasing k, one recognizes, that the value of pv is significantly influenced by the beginning of the simulation. The one-step change in pv is almost negligible as K increases, and thus a huge amount of permutations (10^5) is requested to obtain a reliable estimate. Even if Good's advice may work in practical analysis, we recommend to simulate as many permutations as possible from the time-consumption point of view.

4 Lack-of-fit test

If the no effect hypothesis H_0 is rejected, a natural question appears: Can we find an appropriate (parametric) model explaining the $Y(t)$ on X regression?

Let us denote by $r(t, x; \boldsymbol{\theta})$ a parametric function supposed to be a reasonable model of the unknown mean function $m(t, x)$. Denoting by $\mathcal{C}_{\boldsymbol{\Theta}}$ a class of functions $\mathcal{C}_{\mathbf{\Theta}} = \left\{r(\cdot,\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \mathbf{\Theta} \subseteq \mathbb{R}^q\right\}$, we want to test the null hypothesis

$$\mathsf{H}_1 : \forall x \in \mathcal{X} \quad m(t,x) \in \mathcal{C}_{\mathbf{\Theta}}, \qquad t \in \mathcal{T}, \tag{11}$$

against the general alternative that H_1 does not hold.

Let $\widehat{\boldsymbol{\theta}}$ be a consistent estimator of $\boldsymbol{\theta}$ assuming the null hypothesis is true. Then, if H_1 holds, residuals $\widehat{e}_i(t)$ defined as

$$\widehat{e}_i(t) = Y_i(t) - r(t, X_i; \widehat{\boldsymbol{\theta}}), \qquad t \in \mathcal{T}, i = 1, \dots, n,$$

have to satisfy the no effect hypothesis H_0 discussed in the previous section. Performing the lack-of-fit test thus simply consists in estimating the unknown parameter $\boldsymbol{\theta}$ and applying the no effect test on residuals described in previous section.

5 A simulation study

In this section, we propose to compare the performances of the two test procedures described before thanks to a Monte Carlo study.

For sake of simplicity, the interval $\mathcal{T}$ is taken to be $\mathcal{T} = [0,1]$ and all functional variables and mean functions are discretized in $p = 100$ equispaced points $0 = t_1 < t_2 < \cdots < t_p = 1$.

5.1 The level of the test

Let us consider an additive model with a fixed mean function $\mu(t)$ and a random error term $\varepsilon(t)$, i.e.

$$Y(t) = \mu(t) + \varepsilon(t).$$

The following mean functions are taken:

$$\mu_1(t) = 0, \tag{12}$$
$$\mu_2(t) = 3(t - 1/2)^2, \tag{13}$$
$$\mu_3(t) = \log(50t + 1) + \cos(2\pi t), \tag{14}$$
$$\mu_4(t) = \sin(\pi t)\cos(3\pi t). \tag{15}$$

The explanatory variables X_i are drawn independently on ε from the standard normal distribution, the uniform distribution on $[0,1]$, and the $\mathtt{t}$ distribution with 2 degrees of freedom, respectively. Random terms $\varepsilon(t_j)$, $j = 1 \dots, p$, were simulated either as a zero mean Brownian motion or i.i.d. zero mean random variables following:

- the normal distribution $\mathsf{N}(0, \sigma^2)$, $\sigma^2 = 0.1, 0.25, 1, 4$;
- the uniform distribution on $[-1/2, 1/2]$;
- $\mathtt{t}$ distribution with 2 degrees of freedom transformed to have a zero mean.

Samples $\big(X_i, Y_i(t)\big)$ of different sample sizes $n = 20, 50$, and 100 are generated for each simulation. For each sample, K permutations are performed to obtain an appropriate p-value of the test statistic. The number of permutations K has been chosen using the *ad hoc* procedure based on diagrams similar to those presented in Figure 2. We decided to perform $K = 5000$ permutations when the sample size is $n = 20$, $K = 10000$ when the sample size is $n = 50$, and $K = 20000$ for $n = 100$.

Concerning the choice of the bandwidth parameter h, the automatic cross-validation criterion (3) has been used throughout the simulation study. However, before the simulation started we have carefully checked the significance traces for several samples and all simulation settings. The obtained p-values have been stable in the neighborhood of the cross-validated h^*, and thus the automatic procedure producing the p-value pv^* can be considered reliable.

The results of the simulations are similar for all combinations of the mean functions and distributions of X and $\varepsilon(t)$ and both considered test statistic. Tables 2, 3, 4, and 5 give the empirical levels of the test statistics for i.i.d. gaussian error terms and Brownian motion error terms, respectively, uniformly distributed X and two theoretical levels $\alpha = 5\%$ and $\alpha = 10\%$. We first remark that simulated levels are close to the theoretical ones for both test statistics. It seems that the test procedure performs similarly for both considered error terms, i.e. the pointwise correlation between $\varepsilon_i(t)$ and $\varepsilon_i(s)$ does not influence the test behavior. As expected, the different mean functions do not play a significant role either.

	$\mu_1(t) = 0$			$\mu_2(t) = 3(t - 1/2)^2$		
	$n = 20$	$n = 50$	$n = 100$	$n = 20$	$n = 50$	$n = 100$
$\sigma^2 = 0.01$	4.9 %	5.2 %	5 %	4.7 %	5.1 %	5.2 %
$\sigma^2 = 0.25$	5.2 %	5.2 %	4.5 %	5.5 %	5 %	4.8 %
$\sigma^2 = 1$	4.7 %	4.9 %	4.8 %	5.2 %	5.3 %	5 %
$\sigma^2 = 4$	4.9 %	5.1 %	5.1 %	6 %	4.9 %	4.9 %

	$\mu_3(t) = \log(50t + 1) + \cos(2\pi t)$			$\mu_4(t) = \sin(\pi t)\cos(3\pi t)$		
	$n = 20$	$n = 50$	$n = 100$	$n = 20$	$n = 50$	$n = 100$
$\sigma^2 = 0.01$	6 %	4.5 %	5 %	4.2 %	4.5 %	5 %
$\sigma^2 = 0.25$	5 %	4.5 %	5.1 %	6.2 %	4.9 %	5 %
$\sigma^2 = 1$	5.5 %	5.2 %	5 %	5.1 %	5.2 %	5.1 %
$\sigma^2 = 4$	5 %	4.9 %	5 %	4.5 %	4.8 %	5 %

Table 2: *Simulated levels of the no effect test using F statistic for uniformly distributed X and error terms $\varepsilon(t_j)$ having the normal distribution $\mathsf{N}(0, \sigma^2)$. The theoretical level $\alpha = 5\%$.*

	$\mu_1(t) = 0$			$\mu_2(t) = 3(t - 1/2)^2$		
	$n = 20$	$n = 50$	$n = 100$	$n = 20$	$n = 50$	$n = 100$
$\sigma^2 = 0.01$	4.8 %	5 %	4.6 %	4.8 %	5.1 %	5 %
$\sigma^2 = 0.25$	5.3 %	5.3 %	4.5 %	5.1 %	5.2 %	4.9 %
$\sigma^2 = 1$	6 %	4.8 %	5.2 %	5.2 %	5.3 %	5 %
$\sigma^2 = 4$	6 %	5.1 %	5.8 %	5 %	4.8 %	5.2 %

	$\mu_3(t) = \log(50t + 1) + \cos(2\pi t)$			$\mu_4(t) = \sin(\pi t)\cos(3\pi t)$		
	$n = 20$	$n = 50$	$n = 100$	$n = 20$	$n = 50$	$n = 100$
$\sigma^2 = 0.01$	6 %	4.7 %	5 %	4.5 %	4.8 %	5 %
$\sigma^2 = 0.25$	5.2 %	4.8 %	5.1 %	6 %	4.9 %	4.9 %
$\sigma^2 = 1$	4.9 %	5.3 %	4.9 %	5.2 %	5.1 %	5.1 %
$\sigma^2 = 4$	5.5 %	4.9 %	5 %	4.7 %	4.8 %	5 %

Table 3: *Simulated levels of the no effect test using R statistic for uniformly distributed X and error terms $\varepsilon(t_j)$ having the normal distribution $\mathsf{N}(0, \sigma^2)$. The theoretical level $\alpha = 5\%$.*

	$\mu_1(t) = 0$			$\mu_2(t) = 3(t-1/2)^2$		
	$n = 20$	$n = 50$	$n = 100$	$n = 20$	$n = 50$	$n = 100$
$\sigma^2 = 0.01$	9.5 %	8.5 %	10.5 %	11.5 %	11 %	10.2 %
$\sigma^2 = 0.25$	9 %	9.8 %	9.8 %	10 %	10.2 %	10.3 %
$\sigma^2 = 1$	8.5 %	9 %	10.5 %	12 %	9.5 %	9.8 %
$\sigma^2 = 4$	12.2 %	12 %	10.2 %	9 %	9.9 %	10.5 %

	$\mu_3(t) = \log(50t+1) + \cos(2\pi t)$			$\mu_4(t) = \sin(\pi t)\cos(3\pi t)$		
	$n = 20$	$n = 50$	$n = 100$	$n = 20$	$n = 50$	$n = 100$
$\sigma^2 = 0.01$	10 %	9.5 %	10.2 %	9 %	9.5 %	9.8 %
$\sigma^2 = 0.25$	9.5 %	9.5 %	9.7 %	10 %	10.1 %	11.1 %
$\sigma^2 = 1$	9 %	10.2 %	11.3 %	9.6 %	9.9 %	10.3 %
$\sigma^2 = 4$	10.5 %	10 %	10.5 %	9 %	10.3 %	9.7 %

Table 4: *Simulated levels of the no effect test using F statistic for uniformly distributed X and Brownian motion error terms. The theoretical level $\alpha =$ 10%.*

	$\mu_1(t) = 0$			$\mu_2(t) = 3(t-1/2)^2$		
	$n = 20$	$n = 50$	$n = 100$	$n = 20$	$n = 50$	$n = 100$
$\sigma^2 = 0.01$	10.1 %	9.5 %	10.3 %	9.6 %	11 %	10 %
$\sigma^2 = 0.25$	10.2 %	9.8 %	9.8 %	10.5 %	10.2 %	10.1 %
$\sigma^2 = 1$	9.8 %	10.3 %	9.6 %	9.7 %	9.5 %	9.6 %
$\sigma^2 = 4$	9.9 %	10 %	10.2 %	10.2 %	9.9 %	10.5 %

	$\mu_3(t) = \log(50t+1) + \cos(2\pi t)$			$\mu_4(t) = \sin(\pi t)\cos(3\pi t)$		
	$n = 20$	$n = 50$	$n = 100$	$n = 20$	$n = 50$	$n = 100$
$\sigma^2 = 0.01$	10.2 %	10 %	10 %	10.9 %	9.5 %	9.3 %
$\sigma^2 = 0.25$	10.7 %	9.8 %	9.5 %	10 %	10.5 %	10.1 %
$\sigma^2 = 1$	9.8 %	10.1 %	10.1 %	9.6 %	10.2 %	10.3 %
$\sigma^2 = 4$	9.9 %	10.5 %	9.5 %	9.8 %	9.8 %	10 %

Table 5: *Simulated levels of the no effect test using R statistic for uniformly distributed X and Brownian motion error terms. The theoretical level $\alpha =$ 10%.*

5.2 Power of the test

To illustrate the power of the no effect test and to compare the test statistics under the alternative we have simulated samples $\big(X_i, Y_i(t)\big)$, $i = 1, \ldots, 80$, as follows:

$$Y_i(t) = \begin{cases} \sin\left(\frac{2\pi}{\omega_i}(t-\gamma_i)\right), & \text{if } X_i < 25, \\ \sin\left(\frac{2\pi}{\omega_i}(t-\gamma_i)\right) + \beta(X_i - 25), & \text{otherwise,} \end{cases} \tag{16}$$

where X_i is drawn from the uniform distribution on $[0, 80]$, ω_i is drawn from the uniform distribution on $[0.95, 1.05]$, γ_i from the uniform distribution on $[-0.05, 0.05]$. The coefficient β is a fixed coefficient defining the alternative to the no effect hypothesis and allows us to check the ability of the test procedures to detect the alternative when the mean function is not too far from the null hypothesis. Indeed, when β is equal to zero, the null hypothesis is true.

The simulation scheme was motivated by the real data example discussed in Section 6. Although it is written in an unusual form for the regression setting, $Y_i(t)$ may also be expressed as $Y_i(t) = m(t, x) + \varepsilon_i(t)$, where

$$\varepsilon_i(t) = \sin\left(\frac{2\pi}{\omega}(t-\gamma)\right) - \mathsf{E}\sin\left(\frac{2\pi}{\omega}(t-\gamma)\right), \qquad i = 1, \ldots, 80.$$

We have made 100 replications of sample $\big(X_i, Y_i(t)\big)_{i=1}^{80}$. For each replication, we have performed 10 000 permutations to obtain the corresponding p-values. The power of the test procedures is then estimated as a fraction of p-values less than a prescribed level of the test. The results are summarized in Table 6 and shown in Figure 3. As expected, the power increases when the absolute value of β increases, as the alternative becomes more easy to detect and when the level increases since the null hypothesis is rejected more often.

	F statistic				R statistic			
$\beta \times 10^4$	$\alpha = 1\%$	2.5%	5%	10%	1%	2.5%	5%	10%
5	4 %	10 %	12 %	22 %	2 %	8 %	14 %	18 %
7.5	10 %	18 %	24 %	34 %	0 %	4 %	16 %	30 %
10	4 %	20 %	26 %	46 %	4 %	12 %	26 %	44 %
12.5	16 %	24 %	40 %	64 %	10 %	24 %	43 %	64 %
15	8 %	26 %	60 %	96 %	6 %	26 %	66 %	96 %
15.5	16 %	34 %	60 %	100 %	16 %	38 %	70 %	100 %
16	16 %	36 %	70 %	100 %	10 %	42 %	78 %	100 %
16.5	18 %	48 %	72 %	100 %	18 %	48 %	82 %	100 %
17	26 %	48 %	82 %	100 %	18 %	54 %	92 %	100 %
17.5	18 %	66 %	94 %	100 %	24 %	66 %	98 %	100 %
20	56 %	92 %	100 %	100 %	58 %	98 %	100 %	100 %
22.5	84 %	100 %	100 %	100 %	94 %	100 %	100 %	100 %

Table 6: *Illustration of the power of the tests based on simulated samples according to (16) for different alternatives β and the test levels α.*

It seems, from Table 6 and plots in Figure 3, that both considered test statistics behave well and similarly. The F statistic performs slightly better for small values of β, i.e. in the cases when the alternative is close to the hypothesis, R seems to be more powerful for the more distant alternatives.

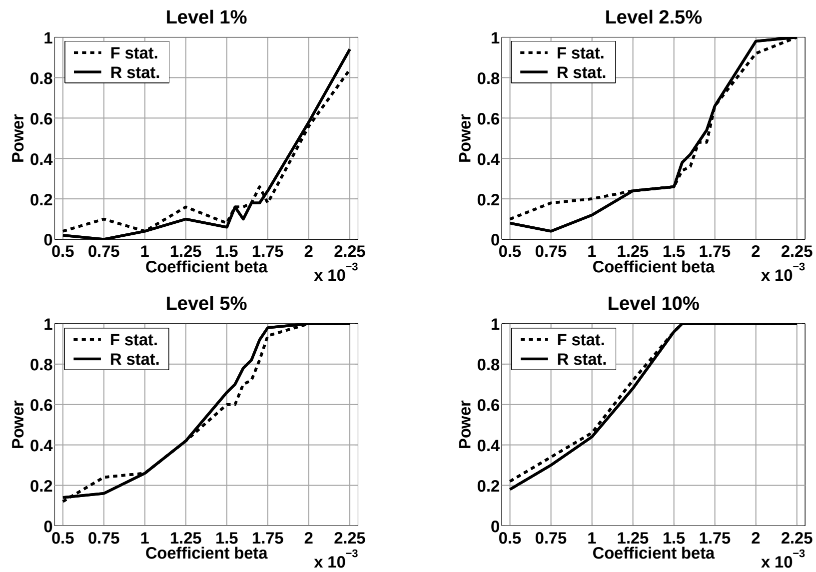

Figure 3: *The simulated power of the no effect test for both considered test statistics, different test levels and the "alternative" coefficient* $\beta \times 10^3$ *varying in* $[0.5, 2.25]$.

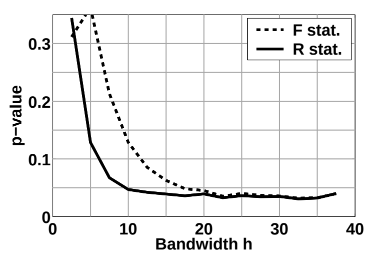

Figure 4: *The significance trace for a simulated sample according to (16) with* $\beta = 1.6 \times 10^{-3}$.

However, it turns out from the performed simulations that the R statistic is more stable than the F one in the sense of the significance trace. A typical example for one data set simulated according to (16) with $\beta = 1.6 \times 10^{-3}$ is given on Figure 4. Although both statistics reach p-values slightly smaller than 5% for the "optimal" cross-validated $h^* = 20$, the F statistic, in contrast to R, indicates to accept the hypothesis for $h < 20$. As we use data-driven method to choose the optimal bandwidth, the comparison of the significance traces of both test statistics is recommended to take into account the stability of the procedures before the final acceptance/rejection is done.

5.3 The case of noisy discretized curves

In practice, one can not observe the whole curves $Y_i(t)$, $t \in \mathcal{T}$, but has discretized and noisy data

$$y_{ij} = Y_i(t_{ij}) + \epsilon_{ij}, \qquad j = 1, \ldots, p_i,\ i = 1, \ldots, n,$$

at design points $t_{i1} < t_{i2} < \cdots < t_{ip_i}$, which may vary from one trajectory to another, with ϵ_{ij} staying for a zero mean white noise. Some pre-smoothing steps, discussed by e.g. Benko et *al.* (2006), Besse et *al.* (1997), Staniswallis and Lee (1998), or Yao et *al.* (2005), must generally be performed to transform the data to the same design grid. In the following paragraphs we give some remarks on the influence of different design point settings on the test procedure.

Beside the *equidistant* fixed design points common for all variables (D1), we have considered random designs at $p = 100$ points (common for all variables) following

(D2) the uniform distribution, i.e. $y_{ij} = Y_i(T_j)$, $T_j \sim \mathsf{U}[0, 1]$;

(D3) the $\mathsf{Beta}(1.5, 2.75)$ distribution, i.e. $y_{ij} = Y_i(T_j)$, $T_j \sim \mathsf{Beta}(1.5, 2.75)$.

Finally, an *individual* random design (D4) according to the $\mathsf{Beta}(1.5, 2.75)$ distribution, i.e. $y_{ij} = Y_i(T_{ij})$, $T_{ij} \sim \mathsf{Beta}(1.5, 2.75)$. In the latter case (D4), each simulated curve $Y_i(t)$ has been "replaced" with its "classical" kernel estimator $\widehat{Y}_i(t)$ obtained by regressing the couples (y_{ij}, t_{ij}). The leave-one-out cross-validation criterion has been used to tune the bandwidth parameter. This pre-smoothing step enables to re-discretized the curves $\widehat{Y}_i(t)$ into the equidistant fixed design and consequently to use quadrature rules in order to approximate the integrals in (3), (6), and (9). More details on estimating $m(t, x)$ from (noisy) non-equidistant discrete data can be be found in Cardot (2006).

Table 7 illustrates the power of the test for the simulating scheme (16) with $\beta = 12.5 \times 10^{-4}$. Results are similar for all considered designs, it seems that pre-smoothing of the response variables increases a little bit the power mainly for small significance levels α. The obtained results are not surprising as the effect of the covariate X in (16) is additive and thus influences the response $Y(t)$ homogenously in $t \in \mathcal{T}$.

	Power $\alpha = 5\%$				Power $\alpha = 10\%$			
	(D1)	(D2)	(D3)	(D4)	(D1)	(D2)	(D3)	(D4)
F	40 %	41 %	50 %	56 %	64 %	67 %	71%	67%
R	43 %	45 %	51 %	52 %	64 %	65 %	70 %	67%

Table 7: *The influence of different design settings on the power of the test procedure – the additive simulation model (16).*

Design points play more important role when the covariate X effects the functional variable only in a part of its support $\mathcal{T}$. To illustrate the behavior, we have considered a simple multiplicative setting

$$Y_i(t) = a(2\pi(t - \gamma_i), X_i) \sin\big(2\pi(t - \gamma_i)\big), \tag{17}$$

where the amplitude function $a(t, X_i)$ is constant in the first half of the sine period and depends on the uniformly distributed explicative X_i in the second half, i.e.

$$a(t, X) = \begin{cases} 1, & \text{if } 2k\pi < t \leq (2k+1)\pi, k \in \mathbb{Z}, \\ 0.15X, & \text{if } (2k+1)\pi < t \leq 2k\pi, k \in \mathbb{Z}, \end{cases}$$

and where γ_i follows the uniform distribution on $[-0.05, 0.05]$.

From the results presented in Table 8 one can see, that the asymmetric Beta design (D3) with the discretization points concentrated in the first unaffected half of the sine function is considerably less powerful than the equidistant (D1) or the uniform (D2) case. For this particular simulation it turns out that pre-smoothing applied to the Beta design remarkably increases the power of the test.

	Power $\alpha = 5\%$				Power $\alpha = 10\%$			
	(D1)	(D2)	(D3)	(D4)	(D1)	(D2)	(D3)	(D4)
F	53 %	60 %	15 %	98 %	92 %	85 %	33%	100%
R	52 %	61 %	14 %	99 %	94 %	85 %	32 %	100%

Table 8: *The influence of different design settings on the power of the test procedure – the multiplicative simulation model (17).*

To conclude, if the discretization design is not homogenous, pre-smoothing steps are worth applying. The final decision is the to be taken with respect to the test performance for both, the original and the smoothed, data sets.

6 Application to vertical atmospheric radiation profiles

Since 1994, vertical profiles of atmospheric radioactivity have been measured at the upper air meteorological station of the Czech Hydrometeorological Institute in Prague-Libuš. Four times a year (monthly during the years 1994–2000) a meteorological balloon with a radioactivity sonde system ascends from the Earth's surface up to approximately 35 km and detects short current pulses coming from the interaction between the radiation and the sonde material. Each launch of a balloon thus provides one observation of the *vertical atmospheric radiation profile* – giving a measure of dependence of the radiation intensity on the altitude.

About 100 profiles, overall, have been measured since 1994. We have carefully chosen 69 observations that were measured (at least) to the altitude of 35 km and do not suffer from any evident measurement errors due to technical problems with signal transmitting.

Let us denote by $Y_i(a), i = 1, \ldots, n = 69$, a radiation profile measured at a random time X_i (in days since the first observation) with $a \in [0, 35]$ representing the altitude. For the purpose of this analysis, we assume $Y_i(a)$ to be random parametric functions

$$\begin{aligned} Y_i(a) = r(a; \boldsymbol{\vartheta}^i) &= \vartheta_1^i \vartheta_3^i \left(1 + (\vartheta_2^i - 1)\mathrm{e}^{-\vartheta_3^i(a-\vartheta_4^i)}\right)^{\vartheta_2^i/(1-\vartheta_2^i)} \\ &\quad + \vartheta_5^i \left(1 + \mathrm{e}^{-\vartheta_6^i(a-\vartheta_7^i)}\right) + \vartheta_8^i. \end{aligned} \tag{18}$$

Reasons to consider atmospheric profiles in the form (18) additively combining the derivative of the Richards growth curve with the logistic curve are explained in details in Antoch et *al.* (2006) as well as an estimation procedure of the parameters $\boldsymbol{\vartheta}^i \in \mathbb{R}^8$.

Regarding the available sample of radiation profiles $Y_i(a)$ (Figure 5) through meteorologists' eyes, two questions appear:

a) Do radiation profiles change along time ?
b) If there is a change, can we find an appropriate model taking time into account ?

To answer the questions we first apply the no effect permutation test for

$$m_{\mathsf{rad}}(a, x) = \mathsf{E}\big[Y(a)|X = x\big], \qquad a \in [0, 35], x \in \mathbb{R}^+, \tag{19}$$

As Hlubinka and Prchal (2006) suggest parametric estimators of $m_{\mathsf{rad}}(a, x)$, we then show their adequacy using the lack-of-fit test.

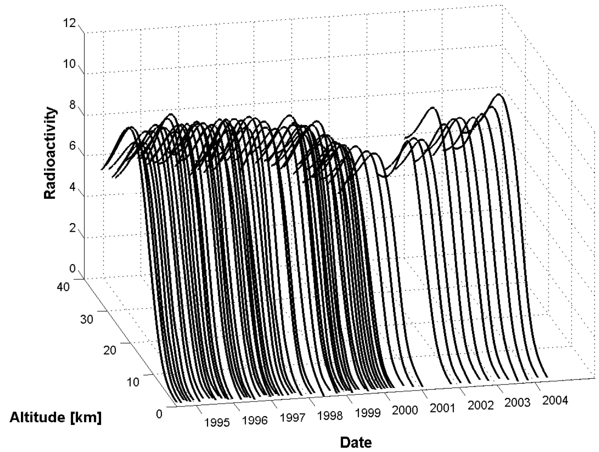

Figure 5: *The sample of 69 radiation profiles measured in Prague within the years 1994 – 2004.*

Remark. Meteorologists suppose that there exists a natural radiation background in the atmosphere possibly varying in a long-time period connected with the 11-years solar cycle. In a short-time (days) period, the radiation in low altitude levels (up to 10km) may be effected by local weather (storms, rainfalls, etc.). On the other hand, the permanent turbulent changes in the upper (stratospheric) layers are not supposed to influence the natural radiation. As we observe the profiles monthly and four times per year, respectively, the short-time dependence does not have to be taken into account. Hence, we can suppose the couples $\big(X_i, Y_i(a)\big)$ are an i.i.d. sample and we can apply regression techniques rather than "time-series" tools.

6.1 No effect test

It seems from the available data sample that there was a decrease of radiation in the period 2000 – 2002, at least a decrease in altitudes where the radiation reaches its maximal intensity (25th km).

To get a better idea on the time dependence, Hlubinka and Prchal (2006) apply the nonparametric kernel estimator (2) and discuss in details different choices of the smoothing parameter h. As an illustration, the estimate for a moderate value $h = 265$ is given in Figure 6.

In order to statistically prove "evident" changes in radiation we have performed the no effect permutation tests of the hypothesis

$$\mathsf{H}^0_{\mathsf{rad}} : \forall x \in \mathbb{R}^+ \quad m_{\mathsf{rad}}(a, x) = \mu_{\mathsf{rad}}(a), \qquad a \in [0, 35]. \tag{20}$$

We have simulated 20 000 permutations of the both discussed test statistics resulting into practically zero p-values ($\ll 10^{-5}$) for all reasonable values of the smoothing parameter $h \in [50, 3000]$. We thus, as expected, reject the null hypothesis $\mathsf{H}^0_{\mathsf{rad}}$ for all common test levels $\alpha = 1, 5, 10\%$.

6.2 Lack-of-fit test

Aside nonparametric kernel smoothing, Hlubinka and Prchal (2006) suggest (non-periodic and periodic) parametric estimators of (19). As we have accepted a significant dependence of the radiation on time, the lack-of-fit test serves to decide, whether their parametric estimators are appropriate descriptions of radiation time changes.

To illustrate the lack-of-fit test performance we have chosen a non-periodic model

$$\widehat{m}_{\mathsf{rad}}(a, x) = r(a; \boldsymbol{\vartheta}^0) \exp\Big\{\beta(a) \exp\left\{-(x-\gamma)^2/\delta\right\}\Big\}, \tag{21}$$

where $r(a)$ of the form (18) stays for a baseline radiation with eight real parameters $\boldsymbol{\vartheta}^0$ to be estimated, $\gamma, \delta \in \mathbb{R}$ are unknown time parameters, and the functional parameter $\beta(a)$ stays for the amplitude. We distinguish two cases – the *basic model* with a constant amplitude parameter $\beta(a) \equiv \alpha \in \mathbb{R}$ and the *adaptive model* with an altitude dependent amplitude $\beta(a)$ to be estimated.

The least squares estimates of the parameters are summarized in Table 9. Figure 6 then shows comparison of the kernel and parametric estimates – see Hlubinka and Prchal (2006) for further details.

We have performed 20 000 permutations in order to obtain p-values of the lack-of-fit tests. Figure 7 presents the reached p-values for different bandwidth parameters varying around the cross-validated optimal one. Table 10 numerically summarizes the figure. We see that both statistics performs similarly and that employing the adaptive parameter $\beta(a)$ significantly improves the estimator, as the p-values increase from 0.07 up to almost 0.3.

Although the results are omitted here, let us remark, that the periodic proposals of Hlubinka and Prchal (2006) behaves even better than the non-periodic ones. Both adaptive models thus may be considered as (statistically) adequate descriptions of the radiation time changes and are to be proposed to the meteorological community to discuss their physical meanings and consequences.

Basic model			
ϑ_1^0	108	ϑ_7^0	7.92
ϑ_2^0	0.84	ϑ_8^0	0.18
ϑ_3^0	0.14	α	-0.23
ϑ_4^0	18.8	γ	2273
ϑ_5^0	4.36	δ	3×10^5
ϑ_6^0	0.59		

Adaptive model			
ϑ_1^0	110	ϑ_7^0	7.9
ϑ_2^0	0.81	ϑ_8^0	0.18
ϑ_3^0	0.13	α	—
ϑ_4^0	18.9	γ	2428
ϑ_5^0	4.33	δ	3×10^5
ϑ_6^0	0.6		

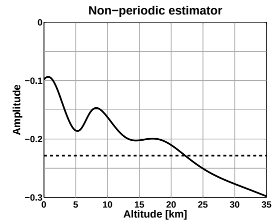

Table 9: *The least squares estimates of the parametric model (21). The plot shows the estimated parameter $\beta(a)$.*

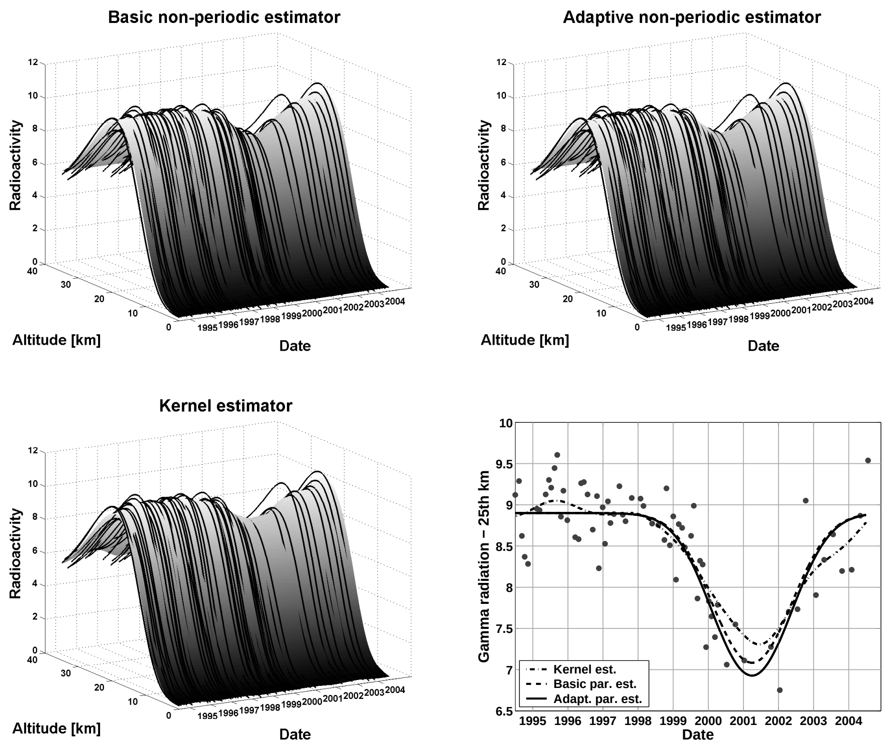

Figure 6: *The kernel and parametric estimates of the radiation profiles and their comparison at the altitude of 25 km.*

Estimator	Optimal bandwidth h	F stat. p-value	R stat. p-value
Basic	1230	0.0784	0.0827
Adaptive	3000	0.2424	0.2748

Table 10: *The optimal bandwidths and the corresponding p-values of the lack-of-fit test performed for both parametric models.*

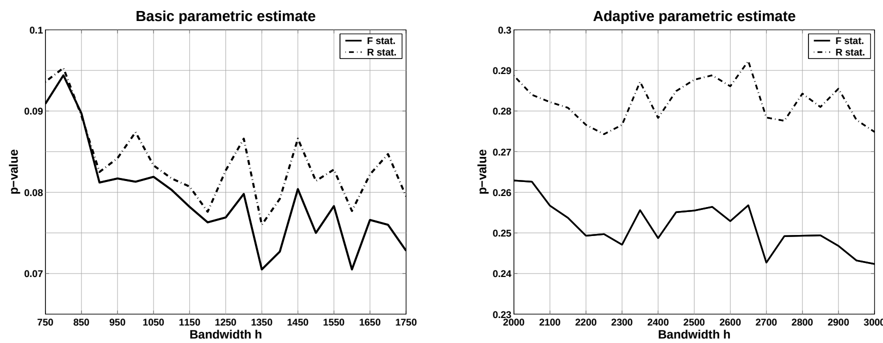

Figure 7: *Estimated p-values of the lack-of-fit test for both basic and modified parametric models.*

7 Conclusion

We have proposed a permutation approach to check if a real covariate has a significant effect on a functional response in a regression setting. The principles of both suggested test statistics, the adapted F-statistic and the smoothing residuals one, are easy to implement in practice. Even if a huge amount of permutations is usually required to obtain a reliable estimate of the p-value, the time consumption of the procedures does not seem to be a limiting factor nowadays.

The Monte Carlo simulation study confirms that the behavior of both test statistics corresponds to their theoretical unbiasedness. The R statistics then seems to be more stable with respect to different values of the smoothing parameter h. For the practical purposes, however, we recommend to perform the test for both test statistics and base the final decision on their significance plots.

The real-data analysis of the atmospheric radiation profiles shows that the proposed lack-of-fit test performs in the expected way and may serve as a tool in a decision problem of testing if some parametric class of models is adequate to describe the evolution along time of the observed curves.

Acknowledgements. The authors would like to express their thanks to Prof. Jaromír Antoch and all the members of the STAPH working group in Toulouse for their generous support, valuable comments and help. The research of L. Prchal was partially supported by the Grant No. 201/05/H007 of the Grant Agency of the Czech Republic and by the Research Project No. MSM 0021620839 of the Ministry of Education of the Czech Republic.

References

Aguilera, A. M., Ocaa, F. A., and Valderrama, M. J. (1999). Forecasting with unequally spaced data by a functional principal component approach. *Test*, **8**, 233 – 253.

Antoch, J., Hlubinka, D., and Prchal, L. (2006). Statistical methods for analysis of meteorological measurements, *preprint*.

Benko, M., Härdle, W., and Kneip, A. (2006). Common functional principal components. *SFB649 Economic Risk Discussion Paper 2006–10*, Humboldt University, Berlin.

Besse, P., Cardot, H., and Ferraty, F. (1997). Simultaneous non-parametric regressions of unbalanced longitudinal data. *Comput. Statist. Data Anal.*, **24**, 255 – 270.

Bowman, A. W., and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis*. Clarendon Press.

Cardot, H. (2006). Conditional functional principal components analysis. *Scand. J. Statist.*, to appear.

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional Linear Model. *Statist. & Prob. Letters*, **45**, 11 – 22.

Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Statist.*, **30**, 241 – 255.

Cardot, H., Goia A., and Sarda, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. *Comm. Statist. Simulation Comput.,* **33**, 179 – 199.

Cardot, H., and Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *J. Multivariate Anal.*, **92**, 24 – 41.

Chiou, J. M., Müller, H. G., and Wang, J. L. (2004). Functional response models. *Statist. Sinica*, **14**, 659 – 677.

Cuevas, A., and Fraiman, R. (2004). On the bootstrap methodology for functional data. *In: Antoch, J. (ed.), Proceedings in Computational Statistics, COMPSTAT 2004*. Physica-Verlag, 127 – 135.

Cuevas, A., Febrero, M., and Fraiman, R. (2004). An anova test for functional data. *Comput. Statist. Data Anal.*, **47**, 111 – 122.

Deville, J. C. (1974). Méthodes statistiques et numériques de l'analyse harmonique. *Ann. Insee,* **15**, 3 – 104.

Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2004). Principal component estimation of functional logistic regression: discussion of two different approaches. *J. Nonparametr. Stat.*, **16**, 365 – 384.

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

Ferraty, F., and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Springer.

Good, P. (2000). *Permutation Tests. A Practital Guide to Resampling Methods for Testing Hypotheses*. 2nd edition. Springer.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer.

Härdle, W., and Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.*, **13**, 1465 – 1481.

Hlubinka, D., and Prchal, L. (2006). Changes in atmospheric radiation from the statistical point of view. *Comput. Statist. Data Anal.*, to appear.

James, G. M. (2002). Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **64**, 411 – 432.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley.

Marx, B. D., and Eilers, P. H. (1999). Generalized linear regression on sampled signals and curves: a p-spline approach. *Technometrics*, **41**, 1 – 13.

Müller, H. G., and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.*, **33**, 774 – 805.

Raz, J. (1990). Testing for no effect when estimating a smooth function by nonparametric regression: a randomization approach. *J. Amer. Stat. Assoc.,* **85**, 132 – 138.

Ramsay, J. O., and Dalzell, C. J. (1991). Some tools for functional data analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **52**, 539 – 572.

Ramsay, J. O., and Silverman, B. W. (2005). *Functional Data Analysis*. Springer-Verlag, 2nd edition.

Ramsay, J. O., and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer-Verlag.

Staniswalis, J. G., and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *J. Amer. Stat. Assoc.*, **93**, 1403 – 1418.

Yao, F., Müller, H. G., and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Stat. Assoc.*, **100**, 577 – 590.