A permutation approach to the analysis of spatio-temporal geochemical data in the presence of heteroscedasticity

Veronika Římalová[1], Alessandra Menafoglio[2], Alessia Pini[3],

Vilém Pechanec[4], Eva Fišerová[1]

[1] Department of Mathematical Analysis and Applications of
Mathematics, Faculty of Science, Palacky University Olomouc, Czech Republic,
(veronika.rimalova@upol.cz)

[2] MOX, Department of Mathematics, Politecnico di Milano, Italy

[3] Department of statistical sciences, Università Cattolica del Sacro Cuore, Italy

[4] Department of Geoinformatics, Faculty of Science, Palacky University Olomouc, Czech Republic

Abstract

This paper proposes a novel nonparametric approach to model and reveal differences in the geochemical properties of the soil, when these are described by space-time measurements collected in a spatial region naturally divided into two parts. The investigation is motivated by a real study on a space-time geochemical dataset, consisting of measurements of potassium chloride pH, water pH, and percentage of organic carbon collected during the growing season in the agricultural and forest areas of a site near Brno (Czech

1

Republic). These data are here modelled as spatially distributed functions of time. A permutation approach is introduced to test for the effect of covariates in a spatial functional regression model with heteroscedastic residuals. In this context, the proposed method accounts for the heterogeneous spatial structure of the data by grounding on a permutation scheme for estimated residuals of the functional model. Here, a weighted least-squares model is fitted to the observations, leading to asymptotically exchangeable, and thus, permutable residuals. An extensive simulation study shows that the proposed testing procedure outperforms the competitor approaches that neglect the spatial structure, both in terms of power and size. The results of modelling and testing on the case study are shown and discussed.

Keywords: Functional data, Geostatistics, Nonparametric inference, Functional regression, Edge effect on soil.

1 Introduction

Functional data analysis (FDA (Ramsay & Silverman, 2005); (Ramsay, Hooker, & Graves, 2009)) is a powerful set of methodologies for analysing complex data structures such as curves or images. Unlike multivariate statistics, FDA treats the curves as single data objects, *atoms* of the analysis, instead of a sequence of discrete observations (Ramsay & Silverman, 2005). FDA covers a lot of statistical areas, such as times series analysis (Fraiman, Justel, Liu, & Llop, 2014), change point detection (Aue, Gabrys, Horváth, & Kokoszka, 2009), or regression models ((Ramsay & Silverman, 2005); (Reiss, Huang, & Mennes, 2010)). Most FDA methods strongly rely on the assumption of

independence among observations. In the presence of spatial dependence, applying these methods is somewhat inappropriate and the analysis could fail because of consistency problems (Horváth & Kokoszka, 2012). The spatial dependence needs to be properly treated; for this purpose, the classical geostatistical methods can be extended to the functional framework ((Menafoglio, Secchi, & Dalla Rosa, 2013); (Menafoglio & Secchi, 2017); (Bernardi, Sangalli, Mazza, & Ramsay, 2017)).

In these days functional geostatistical theory is well-developed, methods like ordinary kriging (Giraldo, Delicado, & Mateu, 2011) or universal kriging ((Caballero, Giraldo, & Mateu, 2013); (Menafoglio, Grujic, & Caers, 2016)) have been successfully applied to functional data. To model the spatial trend in functional data, spatial coordinates and non-spatial predictors can be combined by kriging with external drift, as in (Ignaccolo, Mateu, & Giraldo, 2014). Methodologies for modelling spatial anisotropy have also been extended to the case of functional data. For instance, Bernardi, Carey, Ramsay and Sangalli (2018) proposed to model anisotropy by means of regression with partial differential regularization. Methods like clustering have also been adapted to the spatial functional setting; Romano, Cozza and Verde (2011) propose to cluster such data based on a kernel variogram estimator. However, other inferential settings – e.g. those for testing – are yet to be fully developed.

In general, testing for significance in the context of functional geostatistics can be handled via either parametric or nonparametric methods. The nonparametric approach appears particularly promising, as it allows for minimal assumptions on the data generative model. In this context, permutation

tests have been recently successfully developed, leading to flexible approaches for testing in FDA ((Pini & Vantini, 2017); (Abramowicz et al., 2018)).

The idea behind the permutation scheme, in the context of a two population test, is as follows (see, e.g., Pesarin & Salmaso, (2010)). Under the null hypothesis of equality in distribution of the populations, the distribution of the two samples is the same as the distribution of any two samples obtained upon permuting the populations. A test statistics computed for the original sample (e.g., the difference of the means) should thus have the same distribution as that calculated from any permutation of the data. Hence, the evidence against H_0 is provided by values of the test statistic under permutations that are statistically different from the one obtained on the original sample ((Pesarin & Salmaso, 2010); (Marozzi, 2002)). Due to the large number of all possible permutations (a factorial of the sample size) it would be hardly feasible to consider all possible rearrangements of data. Instead, the permutation distribution of the test statistic can be approximated with Monte Carlo methods, by randomly generating a subset of permutations ((Fortin & Jacquez, 2000); (Lindgren, 2010); (Pesarin & Salmaso, 2010)). The permutation approach requires only very general assumptions on the data's distribution (Marozzi, 2002); in the permutation framework the classical assumption of independence of observations is relaxed to their exchangeability under the null hypothesis ((Good, 2013); (Myllymäki, Mrkvička, Grabarnik, Seijo, & Hahn, 2017)), that is, the distribution of data is asked to be invariant under permutations (Kim, Fay, Feuer, & Midthune, 2000).

In the framework of linear models, it is not possible to directly observe exchangeable quantities under the null hypothesis, unless the model only con-

4

tains a single covariate. The permutation scheme can be adapted in different ways to obtain approximate exchangeability. Freedman and Lane (1983) proposed to permute the estimated residuals from the model under the null hypothesis (also referred to as the reduced model); ter Braak (1992) proposed to permute residuals of the full model; other authors ((Oja, 1987); (Manly, 2018)) propose to permute other quantities, such as the covariates' values, or the responses. Anderson and Robinson (2001) compared Freedman and Lane's methodology and other approaches by a simulation study, concluding that the Freedman and Lane permutation scheme gives the best empirical results in terms of the power and the size of the test. From a theoretical point of view, since the estimated residuals of the null model are asymptotically exchangeable, the obtained test is asymptotically exact. Winkler, Ridgway, Webster, Smith and Nichols (2014) have applied this approach into the multivariate case of testing for differences between two sets of images, Abramowicz et al. (2018) have adapted the aforementioned permutation scheme to testing in functional regression model for knee movement.

In this work we extend the approach of Abramowicz et al. (2018) to the space-time setting. The research is motivated by a real-world case of interest in geology and chemistry. The data consist of measurements of potassium chloride (KCl) pH, water pH and percentage of organic carbon in a site which is naturally divided into two parts, agricultural and forest, with different geochemical properties. The phenomenon known as the *edge effect* has been studied for decades, and is connected with margins of fields, meadows, or forests (Schröder & Fleig, 2017). It is accompanied by specific features, such as different variability in yield and utilization of cultures (Pechanec, Vávra,

Hovorková, Brus, & Kiliánová, 2014), a change of the water management regime, water-retention capacity, or different lighting conditions between the edge and the central part of the site. From the agricultural point of view, the edge effect is rather problematic since it has negative effect on the production due to the change of conditions at borders with other landscapes. On the contrary, from the ecological and landscaping point of view, the edge effect is a vitally important element. The edges of the site, left untreated, are an important part of the landscape in terms of the biological diversity.

In this setting, two main questions arise: Do the aforementioned two types of soil have different chemical properties? Do the measurements differ according to the type of soil? To face these challenging research questions, a permutation testing procedure based on a spatial regression model is here proposed. To deal with the nature of spatial observations, which depend on their spatial coordinates and are likely to be heterogeneous by virtue of their geographical closeness, we propose to permute the residuals of spatial regression model and not the observations themselves. For this purpose, the permutation scheme by Freedman and Lane is adapted here. In the paper, emphasis shall be given to the exchangeability of residuals, since this assumption can be easily violated whenever the data are heteroscedastic.

The paper is structured as follows: Section 2 provides a detailed description of geochemical properties of the analysed real-world dataset, as well as a mathematical formulation of the problem. In Section 3, spatio-temporal models under homoscedasticity and heteroscedasticity assumptions are defined. For each case, a permutation-based test for the effect of covariates in functional regression model with spatial covariates is introduced. In the last

part of Section 3, a model validation through a permutation-based test for homoscedasticity is described. Results of a simulation study for the empirical size and power of tests for the effect of covariates are reported in Section 4. In Section 5, all proposed methods are applied to the available geochemical data. The conclusions are discussed in Section 6.

2   Data description

This study is based on the data resulting from the analysis of soil samples, collected in the growing seasons (March-October) of the years 2015 and 2016 at a site located near Brno, Czech Republic. Its total area is 9131.4 m^2. Its surface is formed by brown earth and cambisol. The altitude of the site ranges from 524.3 m to 529.3 m (mean altitude 526.8 m). The slope ranges from 0.05° to 16.32° with mean 2.7°. The surface of the site is oriented to the south-west. Agricultural soil covers 60.5 % of the area, 16 % is covered by beech trees. The remaining part of the site is covered mainly by spruces and fixed coniferous forest.

The site contains 11 sampling points equally distributed on a straight line, perpendicular to the border between the field and the forest. The sampling points are located every 3 meters, the total length of the studied site is 30 meters. It is naturally divided in two parts by the central sampling point, the *ecotone*. The soil samples were taken 5 cm beneath the surface.

The analysis of samples was conducted in certified laboratories. Carbon samples were carried out through the method of oxidation of the sample chromium mixture and subsequent measurement on a spectrophotometer.

Following the methodology of the Central Institute for Supervising and Testing in Agriculture, Czech Republic (Zbíral et al., 2011), combustion tubes have been replaced by 100 ml glass flasks, heated on a heating plate instead of using a steam distiller. The measured values represent an amount of oxidisable carbon in weight percentage of dry matter. The active soil reaction was determined using a pH meter in soil water suspension with a glass ion-selective electrode after previous shaking of the suspension on a horizontal mechanical shaker for one hour followed by one hour rest. Just before the measurement itself, the suspension was briefly stirred with a glass bar (Zbíral & Honsa, 2010).

The main goal of this study is to compare the chemical properties (pH KCl, pH H_2O and percentage of organic carbon) of field (A) and forest (B). The dissimilarities between field and forest could be investigated from behaviour of their probability distributions. In particular, we aim at testing the hypothesis

$$H_0 : \mathcal{F}_{(A)} \sim \mathcal{F}_{(B)} \text{ against } H_1 : \mathcal{F}_{(A)} \not\sim \mathcal{F}_{(B)}, \tag{1}$$

where $\mathcal{F}_{(A)}, \mathcal{F}_{(B)}$ denote the distribution of pH KCl, pH H_2O, or percentage of organic carbon, in field and forest, respectively.

Using state-of-the-art permutation tests, the hypothesis would be tested by (i) randomly reassigning the data to the two groups and (ii) comparing the distributions of the random reassigned samples with the original ones. Under the null hypothesis, the distributions of the two groups should not differ when the data is permuted. However, such a permutation procedure

8

grounds on the key assumption that the data are exchangeable under the null hypothesis. This assumption may be clearly violated in the case of spatial data, since they are likely to depend on their coordinates. Instead of permuting the data directly, a spatial trend will be here fitted to the data and the estimated residuals from this model, assumed to be approximately exchangeable, will be permuted instead. The model and procedure is formally introduced in Section 3.

The measurements taken from each sampling locations shall be modelled as functions of time distributed over a one-dimensional spatial domain. Raw monthly measurements were turned into functional observations, through a projection over a cubic B-spline basis with knots at data points (i.e., 8 knots). The data were smoothed using the penalized residual sum of squares criterion with the smoothing parameter $\lambda = 10$ selected via the generalized cross-validation (Ramsay & Silverman, 2005).

The measurements taken at the ecotone were excluded from the dataset. Indeed, the ecotone is a sampling point having neither properties of forest, nor of field soil and, moreover, its influence disappears at around 3 meters away. The measurements from growing seasons in 2015 and 2016 do not differ from the geochemical point of view, therefore the data can be analysed together.

3   Problem formulation

In this section, a functional linear model for the spatial data described in Section 2 is introduced, and the testing procedure is proposed.

3.1 Model specification

Denote by $D \subset \mathbb{R}^d$ the spatial domain of interest. For the purpose of our study, we set $d = 1$. However, the procedure here detailed would also be valid in two- or three-dimensional study regions. Let s be a point in D and denote by $\mathcal{X}_s(t)$ the geochemical property under study (e.g., pH KCl) at location $s \in D$ at time $t \in T$, with $T = [3, 10]$ ($t = 3$ means measurement in March, $t = 10$ in October). In the following, we always assume that $\mathcal{X}_s(\cdot)$ is an element of the Hilbert space of squared-integrable function $L^2(T)$ (or L^2 for short). We call $\{\mathcal{X}_s(t), \ t \in T, s \in D \subset \mathbb{R}^d\}$ the functional random field of the time-varying geochemical property, defined on L^2. Following (Menafoglio et al., 2013), we assume for the random field the model

$$\mathcal{X}_s(t) = m_s(t) + \delta_s(t), \ s \in D, \ t \in T, \tag{2}$$

where $m_s(t)$ is a drift and $\delta_s(t)$ are residuals – realisations of a zero-mean random process. The drift captures a spatially non-constant mean variation. We here represent the drift by a linear model

$$m_s(t) = \sum_{l=0}^{L} \beta_l(t) f_l(s), \ s \in D, \ t \in T, \tag{3}$$

where $\beta_l(t)$, $l = 0, \ldots, L$, are unknown functional parameters independent of the spatial location and $f_l(s)$, $l = 0, \ldots, L$, are known functions of the spatial variable $s \in D$. Assuming the drift to be constant over the whole spatial domain leads to a stationary setting in the sense of Menafoglio et al. (2013).

Using a matrix notation, model (2) can be rewritten into the form

$$\boldsymbol{\mathcal{X}} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\delta}, \tag{4}$$

where $\boldsymbol{\mathcal{X}} = (\mathcal{X}_{s_1}(t), \ldots, \mathcal{X}_{s_n}(t))'$, $t \in T$, are functional observations at spatial points $s_1, \ldots, s_n \in D$, $\mathbf{F} = (f_l(s_i))$ is a known $n \times (L+1)$ design matrix, $\boldsymbol{\beta} = (\beta_0(t), \ldots, \beta_L(t))'$, $t \in T$, are unknown functional parameters, and $\boldsymbol{\delta} = (\delta_{s_1}(t), \ldots, \delta_{s_n}(t))'$, $t \in T$, are spatially correlated residuals with a global variance-covariance matrix $\boldsymbol{\Sigma}$, the symbol $'$ meaning transposition. Recall that, in the context of functional spatial processes, ((Delicado, Giraldo, Comas, & Mateu, 2010); (Menafoglio et al., 2013)), $\boldsymbol{\Sigma}$ can be defined as

$$\Sigma_{ij} = \mathrm{Cov}(\mathcal{X}_{s_i}, \mathcal{X}_{s_j}) = \mathrm{E}[\langle \mathcal{X}_{s_i} - m_{s_i}, \mathcal{X}_{s_j} - m_{s_j} \rangle], \ i, j = 1, \ldots, n. \tag{5}$$

Note that this definition of a variance-covariance matrix is associated with a well-defined *global* measure of spatial dependence (see, e.g., (Menafoglio et al., 2013); (Menafoglio & Petris, 2016)), that can be fully described through the trace variogram of the field $\{\mathcal{X}_s, s \in D\}$:

$$\gamma(h) = \frac{1}{2}\mathrm{Var}\left(\delta_{s_i}(t) - \delta_{s_j}(t)\right) = \frac{1}{2}\mathrm{E}[\|\delta_{s_i}(t) - \delta_{s_j}(t)\|^2], \ s_i, s_j \in D, \ h = \|s_i - s_j\|_D, t \in T, \tag{6}$$

where $\|.\|$ represents the L^2 norm (Abramowicz et al., 2018) and $\|.\|_D$ the norm over D. The representation of the spatial dependence through the trace-semivariogram allows for the global inference on the model coefficients proposed in this work, as we show in Section 3.2. A more local perspective

on the parameter inference (e.g., aimed to identify intervals within which the hypothesis is rejected, similarly as in (Pini & Vantini, 2017)) is outside the scope of this work, and would instead require a more local measure of spatial dependence, i.e., the full cross-covariance operator (see, e.g., (Menafoglio & Petris, 2016)).

In view of our case study, we assume that the residuals are heteroscedastic, and that their spatial variance-covariance matrix $\mathbf{\Sigma}$ is diagonal. Although the extension of the proposed procedure to non-diagonal matrices can be envisioned, it requires further technical developments.

In the following, we denote by $\mathbf{W}$ a matrix of weights set to $\mathbf{W} = \mathbf{\Sigma}^{-1}$. We further define $\boldsymbol{\varepsilon} = \mathbf{W}^{1/2}\boldsymbol{\delta}$, as

$$\boldsymbol{\varepsilon}_{s_i} = w_{ii}^{1/2}\,\delta_{s_i}, i = 1, \ldots, n. \tag{7}$$

Clearly, one has $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \mathbf{I}$, and $\{\varepsilon_{s_i}\}$ that are exchangeable. In principle, one may specify any model for the weights w_{ii}, without substantial modifications of the method here proposed. In the following, we shall assume that w_{ii} are constant within the groups, and possibly different between groups.

As shown in (Menafoglio et al., 2013), if the design matrix is of full column rank, i.e. $rank(\mathbf{F}) = L + 1 \leq n$, and the variance-covariance matrix is known and of full rank, i.e. $rank(\mathbf{\Sigma}) = n$, the functional parameters $\boldsymbol{\beta}$ can be estimated through the weighted least-squares (WLS) estimator

$$\widehat{\boldsymbol{\beta}}_{\mathrm{WLS}} = (\mathbf{F}'\mathbf{W}\mathbf{F})^{-1}\mathbf{F}'\mathbf{W}\boldsymbol{\mathcal{X}}. \tag{8}$$

The WLS estimator is unbiased and associated with the variance-covariance matrix $\mathrm{Var}(\widehat{\boldsymbol{\beta}}_{\mathrm{WLS}}) = (\mathbf{F}'\mathbf{W}\mathbf{F})^{-1}$.

In view of the application, the residuals $\boldsymbol{\delta}$ will be considered either homoscedastic and uncorrelated, i.e. $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix and σ^2 is an unknown parameter, or heteroscedastic and uncorrelated, i.e. $\boldsymbol{\Sigma} = \mathbf{W}^{-1}$, where $\mathbf{W}$ is an unknown diagonal matrix.

In case $\boldsymbol{\Sigma} = \mathbf{W}^{-1}$, the diagonal matrix $\mathbf{W}$ is estimated from residuals and a two-step procedure to estimate the functional parameters can be applied. In particular, the OLS model is fitted to the observations, obtaining the estimated residuals $\widehat{\boldsymbol{\delta}}$, by which the weight matrix $\mathbf{W}$ is estimated. Then, the functional parameters $\boldsymbol{\beta}$ are estimated by the WLS estimator using $\widehat{\mathbf{W}}$ in formula (8). The estimation of the variance-covariance matrix $\boldsymbol{\Sigma}$ will be discussed in Section 3.3.

3.2   Tests for the effect of covariates

The hypothesis (1) can be reinterpreted in the context of the functional regression model introduced in Section 3.1, by using group indicators as covariates and testing for the significance of the associated (functional) parameters. In this section we thus discuss the problem of testing for the significance of functional parameters in model (3), i.e.,

$$H_0 : \beta_1(t) = \ldots = \beta_L(t) = 0 \; \forall t \in T, \; \text{against}$$

$$H_1 : \beta_l(t) \neq 0 \; \text{for some } l \in \{1, \ldots, L\} \text{ and some } t \in T.$$

This hypothesis can be equivalently written in a matrix form as

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0} \ \forall t \in T \text{ against } H_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{0} \text{ for some } t \in T, \tag{9}$$

where $\mathbf{C}$ is an $L \times (L+1)$ block matrix of the form $\mathbf{C} = (\mathbf{0}|\mathbf{I})$, where $\mathbf{0}$ is an L-dimensional vector of zeros and $\mathbf{I}$ is an $(L \times L)$ identity matrix.

When the variance-covariance matrix is of the form $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$, one can use the test statistic

$$\text{T}_{\text{OLS}} = \int_T \frac{1}{\widehat{\sigma}^2} (\mathbf{C}\widehat{\boldsymbol{\beta}}_{\text{OLS}})' \left[\mathbf{C}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{C}'\right]^{-1} (\mathbf{C}\widehat{\boldsymbol{\beta}}_{\text{OLS}}) dt, \tag{10}$$

where

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \|\delta_{s_i}(t)\|^2 \tag{11}$$

A decision about the hypothesis (9) can be formulated by means of a permutation test, the global p-value of the T_{OLS} test being computed via the adaptation of the Freedman and Lane scheme (1983) to functional data (Abramowicz et al., 2018). In this setting, the T_{OLS} distribution under permutations is estimated with a Monte Carlo technique, by evaluating it over a high number of permuted datasets, obtained by randomly permuting the residuals $\widehat{\boldsymbol{\delta}}$ estimated from the model under the null hypothesis. If H_0 is true, the distribution of T_{OLS} is *asymptotically* invariant under permutations since permutations are applied to the estimated residuals, which are only asymptotically exchangeable. The resulting global p-value is obtained as the proportion of permutations leading to a value of the T_{OLS} statistic that is higher than or equal to the one of observed data.

For a general diagonal variance-covariance matrix $\mathbf{\Sigma}$, the testing procedure needs to be modified to account for the spatial structure of the residuals, which are no longer asymptotically exchangeable. In this case, a test statistic accounting for the covariance structure of the residuals can be defined as

$$\mathrm{T_{WLS}} = \int_T \left(\mathbf{C}\widehat{\boldsymbol{\beta}}_{\mathrm{WLS}}\right)' \left[\mathbf{C}(\mathbf{F}'\widehat{\mathbf{W}}\mathbf{F})^{-1}\mathbf{C}'\right]^{-1} \left(\mathbf{C}\widehat{\boldsymbol{\beta}}_{\mathrm{WLS}}\right) dt, \qquad (12)$$

where $\widehat{\boldsymbol{\beta}}_{\mathrm{WLS}}$ is the weighted least-squares estimator of functional parameters $\boldsymbol{\beta}$ obtained by plugging in $\widehat{\mathbf{W}} = \widehat{\mathbf{\Sigma}}^{-1}$ into formula (8). In this work, we propose to consider for the permutation scheme the standardized residuals $\boldsymbol{\varepsilon}$, which can be obtained from the estimated residuals $\widehat{\boldsymbol{\delta}}$ as $\widehat{\boldsymbol{\varepsilon}} = \widehat{\mathbf{W}}^{1/2}\widehat{\boldsymbol{\delta}}$. Assuming the homoscedasticity within groups and heteroscedasticity between groups, the variance $\sigma^2_{(j)}$ in the jth group can be estimated via the sample variance within this group, similarly as in equation (11). The estimated standardized residuals $\widehat{\boldsymbol{\varepsilon}}$ are then randomly permuted and the corresponding permuted responses are used similarly as in the previous case.

The permutation scheme is described in detail in Algorithm 1. This algorithm is a new proposal for the case of heteroscedasticity among observations. Note that the testing procedure based on the $\mathrm{T_{WLS}}$ statistic could be considered for more general forms of the variance-covariance matrix $\mathbf{\Sigma}$, provided that this is properly estimated from the data. The algorithm for homoscedastic functional data can be found in the supplementary material.

Algorithm 1 Freedman and Lane permutation scheme in case $\Sigma = \mathbf{W}^{-1}$.

1. Estimate the residuals $\delta_{s_i}(t)$ of the model under H_0;
   $\mathcal{X}_{s_i}(t) = \beta_0(t) + \delta_{s_i}(t)$. Obtain $\widehat{\beta}_0(t), \widehat{\delta}_{s_i}(t), i = 1, \ldots, n$.

2. Estimate the exchangeable residuals $\boldsymbol{\varepsilon}$ from $\widehat{\boldsymbol{\delta}}$ by $\widehat{\boldsymbol{\varepsilon}} = \widehat{\mathbf{W}}^{1/2}\widehat{\boldsymbol{\delta}}$.

3. Permute the estimated exchangeable residuals $\widehat{\varepsilon}_{s_1}(t), \ldots, \widehat{\varepsilon}_{s_n}(t)$, obtaining $\widehat{\varepsilon}_{s_i}^*(t), i = 1, \ldots, n$.

4. Compute the permuted responses $\mathcal{X}^*$ through the fitted reduced model and the permuted estimated residuals $\widehat{\boldsymbol{\delta}}^* = \widehat{\mathbf{W}}^{-1/2}\widehat{\boldsymbol{\varepsilon}}^*$, as $\mathcal{X}_{s_i}^*(t) = \widehat{\beta}_0(t) + \widehat{\delta}_{s_i}^*(t), i = 1, \ldots, n$.

5. Estimate the functional parameters $\boldsymbol{\beta}$ of the full model from the permuted responses $\mathcal{X}^*$ by WLS by plugging-in $\widehat{\mathbf{W}}$. Obtain $\widehat{\boldsymbol{\beta}}_{\mathrm{WLS}}^*$.

6. Compute the test statistic $\mathrm{T}_{\mathrm{WLS}}^*$ as
   $\mathrm{T}_{\mathrm{WLS}}^* = \int_T \left(\mathbf{C}\widehat{\boldsymbol{\beta}}_{\mathrm{WLS}}^*\right)' \left[\mathbf{C}(\mathbf{F}'\widehat{\mathbf{W}}\mathbf{F})^{-1}\mathbf{C}'\right]^{-1} \left(\mathbf{C}\widehat{\boldsymbol{\beta}}_{\mathrm{WLS}}^*\right) dt$.

3.3 Assessment of model assumptions

We here discuss a possible approach to the assessment of the model assumptions introduced in Sections 3.1 - 3.2, particularly those related with the spatial covariance structure of the residuals. The spatial dependence among residuals can be measured using the covariance function (also known as a trace-covariogram (Menafoglio & Secchi, 2017)), $\mathrm{Cov}(\delta_{s_i}(t), \delta_{s_j}(t))$, or the semivariogram defined in equation (6). Its empirical counterpart ((Menafoglio & Secchi, 2017); (Menafoglio et al., 2013)) is given by

$$\widehat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} \|\delta_{s_i}(t) - \delta_{s_j}(t)\|^2, \tag{13}$$

where $N(h)$ is the set of all pairs of observations at a distance approximately h and $|N(h)|$ is its cardinality. More detailed information regarding the variography can be found in ((Cressie, 1993); (Menafoglio & Secchi, 2017) or (Menafoglio et al., 2013)).

As in the scalar case (Cressie, 1993), the shape of the trace-variogram can be used to determine the viability of the assumption that Σ is diagonal. Indeed, uncorrelated residuals δ_{s_i} are associated with a pure nugget model, i.e., a constant trace-variogram function.

In the presence of uncorrelated residuals one has also to check the assumption of their homoscedasticity, as this determines their exchangeability. The test of homoscedasticity can be formulated as test (1), where under the alternative hypothesis the two distributions differ because of a difference in variance. For this reason, we employ a test based on the same permutations as before, but with a test statistic specifically aiming at detecting differences in variance. Note that, using multiple comparison principles, the test can be extended to the case of more populations.

Let $\delta_{s_i(j)}(t)$, $i = 1, \ldots, n_j$, $j = A, B$, be the two groups of residual functions from model (2)-(3), with i being the unit index and j the population index. Let $\sigma^2_{(A)}$ and $\sigma^2_{(B)}$ be the global variances (constant over T) of these two populations, i.e., $\sigma^2_{(j)} = \mathrm{E}[\|\delta_{s_i(j)}\|^2]$. The latter can be estimated as

$$\widehat{\sigma}^2_{(j)} = \frac{1}{n_j} \sum_{i=1}^{n_j} \|\widehat{\delta}_{s_i(j)}(t)\|^2, \ s_i \in D, \ i = 1, \ldots, n_j, \ j = A, B, \ t \in T. \quad (14)$$

Note that, by construction, the overall residual sample mean is zero; further, the residual sample mean within groups is zero if dummy variables are in-

cluded among covariates (i.e., in a functional ANCOVA setting). The goal is now to test the hypothesis

$$H_0 : \sigma^2_{(A)} = \sigma^2_{(B)} \text{ against } H_1 : \sigma^2_{(A)} \neq \sigma^2_{(B)}. \tag{15}$$

For this purpose, we propose as a test statistic the absolute value of the log-proportion of variances:

$$\mathrm{T_{Var}} = \left| \log\left(\frac{\widehat{\sigma}^2_{(A)}}{\widehat{\sigma}^2_{(B)}} \right) \right|. \tag{16}$$

To perform the test, we consider a permutation scheme similar to the one discussed in Section 3.2. Under H_0, residuals are approximately exchangeable, and the permutation procedure of Abramowicz et al. (2018) can be applied. The global p-value of test (15) is computed as the proportion of permutations leading to a value of $\mathrm{T_{Var}}$ higher than the one observed on the data. The testing process is described in Algorithm 2. Its results can be used to establish which permutation strategy to use to test the significance of the functional parameters among those detailed in Section 3.2. Note that the proposed statistical inference concerning homoscedasticity is only approximate due to its definition on the estimated residuals.

4   A simulation study

In this section, we report the results of a simulation study intended to evaluate the empirical size and power of the proposed permutation-based global tests for significance of the effect of covariates in the spatial regression model

---
Algorithm 2 The permutation scheme for the homoscedasticity test (15).
---

1. Sort function's indices as $\underbrace{1, 2, \ldots, n_A,}_{\text{group } A} \underbrace{n_A + 1, n_A + 2, \ldots, n_A + n_B}_{\text{group } B}.$

2. Permute the estimated residuals $\widehat{\delta}_{s_i(j)}(t)$, $j = A, B$, $i = 1, \ldots, n$, from both groups together. Obtain

$$\underbrace{\widehat{\delta}^*_{s_1}(t), \ldots, \widehat{\delta}^*_{s_{n_A}}(t),}_{\text{group } A} \underbrace{\widehat{\delta}^*_{s_{n_A+1}}(t), \ldots, \widehat{\delta}^*_{s_{n_A+n_B}}(t)}_{\text{group } B}$$

3. Using the rearranged functions $\widehat{\delta}^*_{s_i(j)}(t)$, $j = A, B$, estimate the group variances $\widehat{\sigma}^{*2}_{(A)} = \frac{1}{n_A} \sum_{i=1}^{n_A} \|\widehat{\delta}^*_{s_i(A)}(t)\|^2$ and $\widehat{\sigma}^{*2}_{(B)} = \frac{1}{n_B} \sum_{i=n_A+1}^{n_A+n_B} \|\widehat{\delta}^*_{s_i(B)}(t)\|^2.$

4. Compute the test statistic $T^*_{\text{Var}} = \left| \log\left(\frac{\widehat{\sigma}^{*2}_{(A)}}{\widehat{\sigma}^{*2}_{(B)}} \right) \right|.$

---

discussed in Section 3. In the following, we consider two simulation scenarios. First, the empirical size and power of T_{OLS} and T_{WLS} are assessed for a simple functional linear model with one covariate. Second, the empirical size of both tests is evaluated for a model with more covariates. The studies are designed based on the real-world applications presented in Section 5.

4.1 A simple model with an indicator

The first simulation study is based on a functional linear model with one covariate. To evaluate the empirical size and power of tests T_{OLS} and T_{WLS}, a hypothesis of no effect of the covariate is tested. The nominal size of the test is set to $\alpha = 0.05$. We consider sample sizes $n = 20, 40, 200$ and 500. The data are divided into two groups, A with a variance $\sigma^2_{(A)}$ and B with

$\sigma^2_{(B)} = k\sigma^2_{(A)}$, $k = 3, 9$. The sample sizes of A and B are chosen as follows: unbalanced design I, where $n_B = 3n_A$, balanced design II with $n_A = n_B$, and unbalanced design III, such that $n_A = 3n_B$. For each scenario, 2000 simulations are performed. In each case, the model is built as

$$\mathcal{X}_{s(j)}(t) = \beta_0(t) + r \cdot \beta_1(t)soil(s) + \delta_{s(j)}(t), \ \ s \in D, \ t \in [3, 10], \tag{17}$$

where $j = A, B$,

$$soil(s) = \begin{cases} 0 & \text{for group } A, \\ 1 & \text{for group } B, \end{cases} \tag{18}$$

and $r \in [0, 5]$ controls the severity of the deviation between the means of groups A and B.
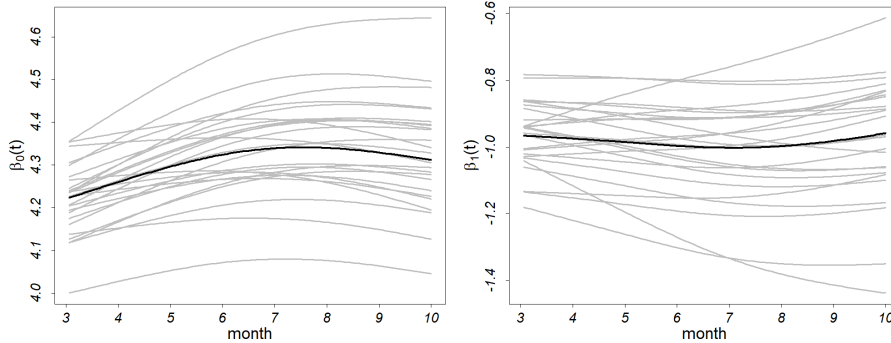

Figure 1: The true $\beta_0(t)$ and $\beta_1(t)$ curves (black) and their estimates (grey) for the simple model with an indicator (17). For the sake of clarity, only 25 estimates are shown.

The residuals' generation is inspired by the geochemical data described in Section 2 and further analysed in Section 5. Denote by $\{\phi_q, q = 1, \dots, 10\}$ the B-spline basis used to represent the data, and by $\boldsymbol{c} = (c_1, \dots, c_{10})'$ the corresponding coefficient vector (see Section 2). Throughout the simulations,

the residuals of the i-th simulation are built on the same B-spline basis expansion $\{\phi_q, \ q = 1, \ldots, 10\}$ as

$$\delta_s(t) = \sum_{q=0}^{10} c_q^{*i} \phi_q(t), \quad t \in T, \tag{19}$$

where the basis coefficients c_q^{*i} are drawn from a multivariate normal distribution with zero-mean and a variance-covariance matrix S_c. For the group A, the matrix S_c was set to the sample variance-covariance matrix of $\boldsymbol{c}$. To ensure heteroscedasticity, the variance-covariance matrix for the group B was multiplied by a scalar a, i.e., $S_c^{(B)} = a \cdot S_c^{(A)}$, $a = 3, 9$. Next, the simulated residuals were added to the drift model – whose true functional parameters $\beta_0(t), \beta_1(t)$ are displayed in Figure 1 – eventually obtaining the functional observations $\mathcal{X}_{s(j)}(t), \ j = A, B$.

The results of the simulations for the size of tests T_{OLS} and T_{WLS} are presented in Table 1; the power functions are shown in Figure 2. One can see that, for the balanced design II, the T_{OLS} and T_{WLS} tests show similar behaviour regarding the size and the power of the tests; here, the empirical size of both tests is close to the nominal one $\alpha = 5\,\%$. In case of the unbalanced design I ($n_B = 3n_A$ and $\sigma_{(B)}^2 = k\sigma_{(A)}^2, k = 3, 9$), the T_{OLS} test is very conservative for each considered sample size. Its empirical size varies from 0.2 % to 1.3 % (95% confidence intervals: (0.1,0.5), (0.9,1.9), respectively). In contrast, for the unbalanced design III ($n_A = 3n_B$ and $\sigma_{(B)}^2 = k\sigma_{(A)}^2, k = 3, 9$), the T_{OLS} test is, for each sample size n, very liberal. The empirical size ranges from 12.4 % to 23.9 % (95% confidence intervals: (11.0,13.9), (22.1,25.8), respectively). In general, under heteroscedasticity, T_{WLS} performs better than

T_{OLS} whenever the design is unbalanced. The empirical size of T_{WLS} is close to the nominal size $\alpha = 5\%$ in the majority of cases. However, for design III and small sample size $n = 20$ or 40, the empirical size of T_{WLS} is slightly higher than the nominal one, maximal empirical size of the T_{WLS} test being 9.6% (95% confidence interval: $(8.4, 11.0)$). This effect relates to the design III where it is necessary to estimate high variance ($k\sigma^2_{(A)}, k = 3, 9$) from relatively very small number of observations ($n_B = n/4$).

Focusing on the size of the difference among group variances ($\sigma^2_{(B)}$ being 3 or 9 times higher than $\sigma^2_{(A)}$), we can see that, as k increases, the empirical power of both T_{OLS} and T_{WLS} converges to 1 for bigger difference among group means (compare the power functions for $k = 3$ and $k = 9$ in Figure 2). In case of design III, the T_{OLS} test may be mistakenly considered as more powerful than T_{WLS}. However, this is only the consequence of the highly liberal behaviour of the T_{OLS}. As one could naturally expect, the power of both tests increases with higher sample size n, as can be seen in every plot in Figure 2.

Table 1 here.

4.2   A more complex model

In this simulation scenario, the tests T_{OLS} and T_{WLS} are evaluated through a functional linear model with three covariates – an indicator function, similarly as in Section 4.1, a linear function of distance, and an interaction term of these two. In total, 2000 simulations were run for a sample size $n = 200$.
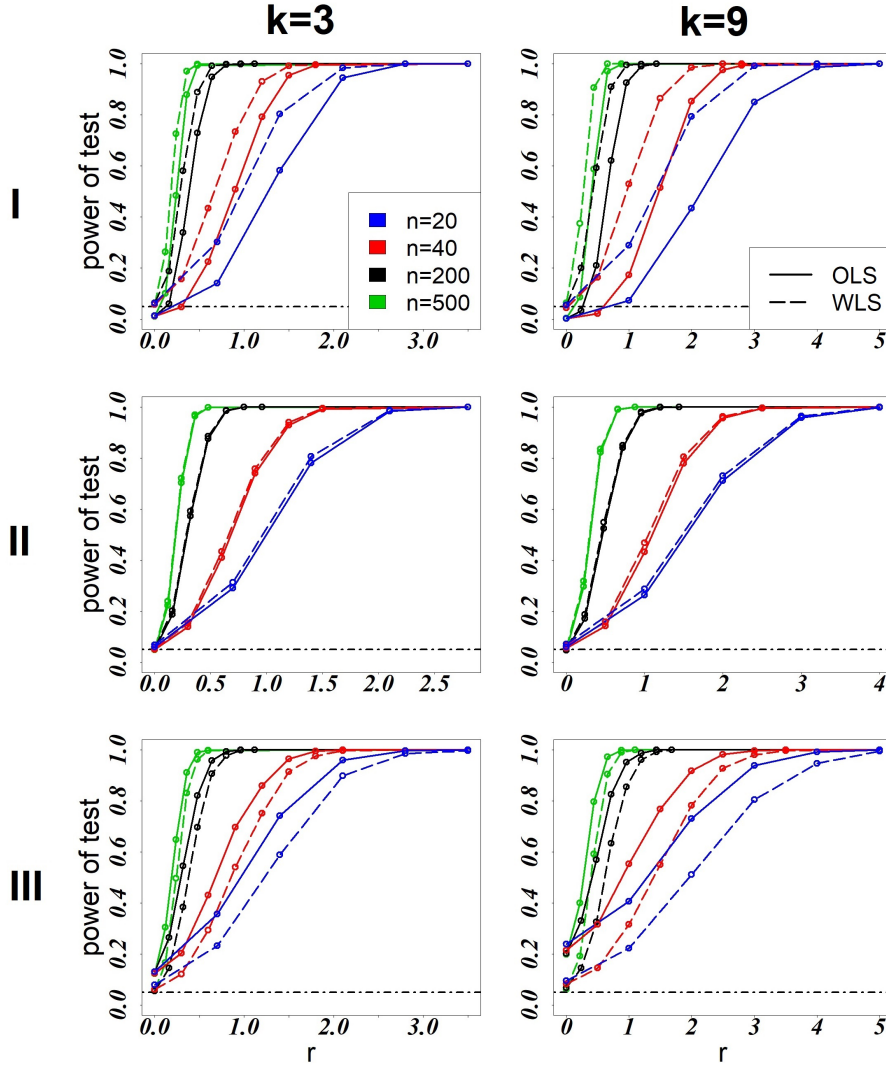
Figure 2: The empirical power of tests T_{OLS} (solid line) and T_{WLS} (dashed line) under different $k = \sigma^2_{(B)}/\sigma^2_{(A)}$ and design. The mean difference among groups is modelled as $r \cdot \beta_1(t)$. The dot-dashed horizontal line represents the nominal size of the test $- \alpha = 0.05$.

Again, inspired by the case study from Section 5, the model is built as

$$\mathcal{X}_{s(j)}(t) = \beta_0(t) + r\cdot\beta_1(t)soil(s) + \beta_2(t)dist(s) + r\cdot\beta_3(t)soil(s)\cdot dist(s) + \delta_{s(j)}(t),$$

(20)

where $s \in D$, $t \in [3, 10]$. The term $dist(s)$ denotes the distance from the *ecotone*, such that $dist(s) = 3, 6, \ldots, 3n_j$, $j = A, B$. The true functional parameters $\beta_0(t), \ldots, \beta_3(t)$ are presented in Figure 3. The simulation study is performed for the same setting as in the previous case. Moreover, we here also consider the situation when $\sigma^2_{(A)} = \sigma^2_{(B)}$.

We aim at examining the empirical size of tests T_{OLS} and T_{WLS} for the global null hypothesis $H_0 : \beta_1(t) = \beta_2(t) = \beta_3(t) = 0$ and comparing it to the empirical size for the indicator model (17). The results (Table 2) suggest that the tests behave similarly for both simple model with an indicator (17) and the more complex model (20). In the presence of heteroscedasticity, the T_{OLS} approach in model (20) is, compared to the results for model (17), less conservative for design I, and even more liberal for design III. Under homoscedasticity, the empirical size of both tests is close to the nominal one $\alpha = 5\,\%$. The T_{WLS}, despite being applied to homoscedastic data, gives results comparable to the (proper) T_{OLS} approach. Some discussion about the precision of the OLS and WLS functional parameter estimates can be found in the supplementary material.

The results from the simulation studies from Sections 4.1 and 4.2 suggest that the T_{WLS} approach introduced in this paper is capable of handling complex functional regression model and heteroscedasticity at the same time. Concerning the either liberal or conservative behaviour of T_{OLS}, our results are consistent with those obtained, in the scalar case, by Huang, Xu, Calian and Hsu (2006).
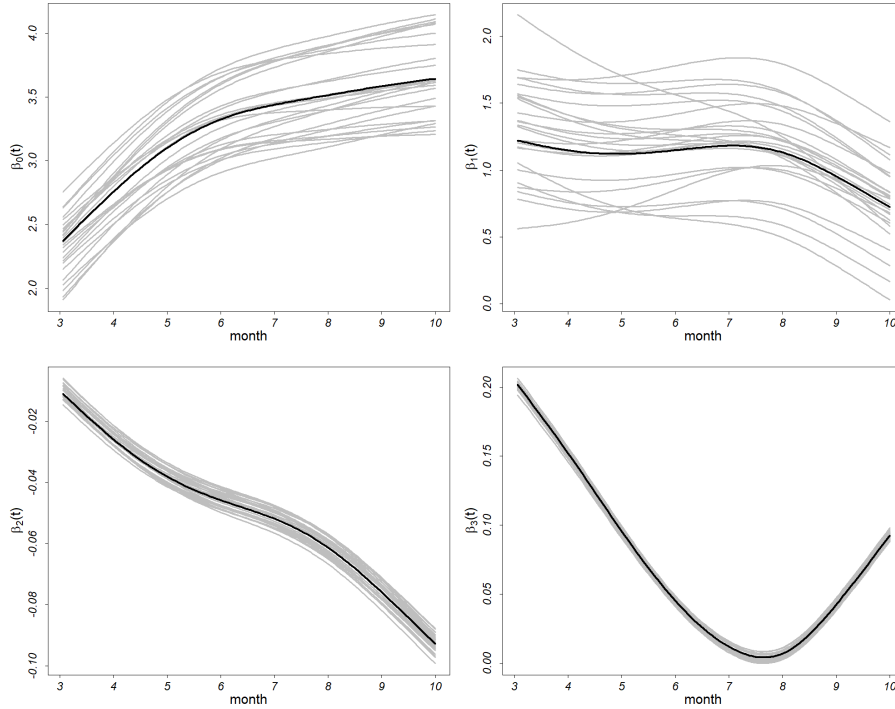
<div align="center">Table 2 here.</div>

Figure 3: The estimates $\widehat{\beta}_l(t), l = 0, 1, 2, 3$, (grey lines) for the complex model (20), where $n = 200$, design I $(n_B = 3n_A)$ and $\sigma^2_{(B)} = 3\sigma^2_{(A)}$. Black lines represent the true functional parameters $\beta_l(t)$. For the sake of clarity, only 25 estimates are shown.

5 Application to geochemical data

In this section, we present the application of the models and methods introduced in Section 3 to the smoothed data of the potassium chloride pH data, H_2O pH, and the percentage of organic carbon, illustrated in Section 2.

5.1 Modelling potassium chloride and water pH

The potassium chloride (KCl) dataset contains 20 functional observations, 10 for the field and 10 for the forest soil. For the KCl pH data we consider

the model

$$\mathcal{X}_s^{\mathrm{KCl}}(t) = \beta_0(t) + \beta_1(t) \cdot soil(s) + \delta_s(t), \ \ s \in D, \ t \in [3, 10], \qquad (21)$$

where the interval $T = [3, 10]$ denotes months from March to October, D is a set of sample spaces and $soil(s)$ is an indicator function for the type of soil, taking value 0 for observations from the field and 1 from the forest soil, i.e.

$$soil(s) = \begin{cases} 0 & \text{for field sample spaces}, \\ 1 & \text{for forest sample spaces}. \end{cases} \qquad (22)$$

The functional parameter $\beta_0(t)$, for each t, represents the mean pH KCl for the field observations, whereas $\beta_1(t)$ can be interpreted, for each t, as the difference between pH KCl from field and forest part of the site.

The fitted model (21) for the drift using the OLS estimation procedure, together with the variogram of residuals, are displayed in Figure 4. The residuals show a pure nugget structure, and thus, they can be considered spatially uncorrelated (Cressie, 1993).
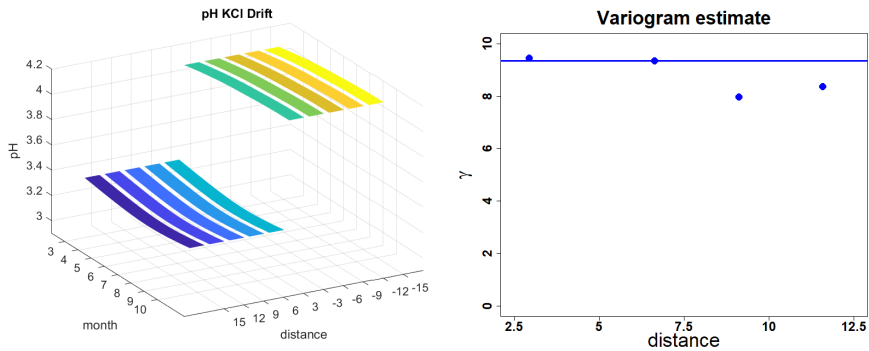

Figure 4: Left: the fitted model (21) for the drift; right: the variogram of the residuals from the model (21).

As the next step, the homoscedasticity of residuals from the field and forest is verified. Let $j = A, B$ denote the field and forest part of the site, respectively. The hypothesis of homoscedasticity is tested using the permutation procedure from Section 3.3. The global p-value of the test – estimated from 1000 random permutations – is equal to 0.001 and, consequently, the hypothesis of homoscedasticity among residuals is rejected. In fact, although the residuals are spatially uncorrelated, the spatial position influences residuals in terms of their variance. Hence, such residuals are not exchangeable over the spatial domain D and the inferential procedure based on T_{OLS} would not be correct.

We thus consider the heteroscedastic setting, and decompose the residuals $\delta_{s(j)}(t)$ as $\sigma_{(j)}\varepsilon_s(t)$, $j = A, B$, where $\sigma_{(j)}$ is the standard deviation for the j-th type of soil, and $\varepsilon_s(t)$ are approximately exchangeable over D. We also assume that the standard deviations are constant on the corresponding part of the spatial domain.

The variances $\sigma^2_{(j)}$, $j = A, B$, are estimated as in (14). Their estimates are $\widehat{\sigma}^2_{(A)} = 14.36$ and $\widehat{\sigma}^2_{(B)} = 3.32$. Notice that the variance of the field soil is more than 4 times higher than that of the forest soil.

Estimated the data variances under heteroscedasticity, model (21) is fitted to the observations through the weighted least-squares method. The quality of the fit can be assessed by the (normalized) root mean squared error RMSE (NRMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\mathcal{X}_{s_i}(t) - \widehat{\mathcal{X}}_{s_i}(t)\|^2}, \ \text{NRMSE} = \frac{\text{RMSE}}{\|\bar{\mathcal{X}}(t)\|}, \qquad (23)$$

and by the R-squared

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n} \|\mathcal{X}_{s_i}(t) - \widehat{\mathcal{X}}_{s_i}(t)\|^2}{\sum_{i=1}^{n} \|\mathcal{X}_{s_i}(t) - \bar{\mathcal{X}}(t)\|^2}, \tag{24}$$

where $\widehat{\mathcal{X}}_{s_i}(t)$ are the fitted values and $\bar{\mathcal{X}}(t)$ is the functional mean. For the model (21), RMSE = 2.973 (NRMSE = 0.074) and $\mathrm{R}^2 = 0.707$.

The effect of the type of soil on KCl pH is tested using the permutation scheme for spatial heteroscedastic data described in Section 3.1, based on 1000 permutations, obtaining a p-value equal to 0. The type of soil has indeed a significant influence on the potassium chloride pH, which tends to be more acidic in the forest part than in the field part of the site (the mean KCl pH ranging from 3.23 to 3.29 in the forest soil, and from 4.03 to 4.12 in the field soil, see Figure 5).

For the H_2O pH measurements, 20 functional observations were given, evenly distributed at the field and forest part of the site. We consider the same model as for the KCl pH. A preliminary test on the variances in the two groups shows that the setting is homoscedastic in this case (p-value = 0.122). The behaviour of H_2O pH appears however similar to that of KCl pH, as shown by the test based on $\mathrm{T_{OLS}}$. The latter allows to conclude that the mean H_2O pH is significantly different at field (range 4.93-5.00) and forest (range 3.65-3.83) part of the site (p-value = 0). The detailed analysis of this dataset is given in the supplementary material.
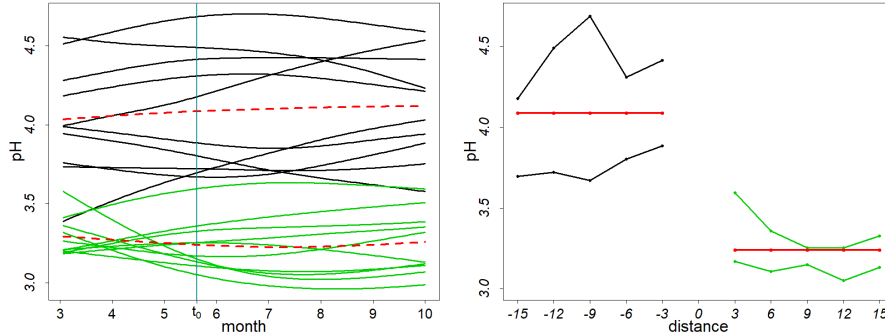
Figure 5: The KCl pH data. Left: original functional observations (black: field, green: forest) and the fitted model (21) (red). Right: A view of data in space at fixed t_0 (specified by a vertical line in the figure on the left). The points indicate the section of the curves at t_0, for clarity connected by lines.

5.2 Modelling the percentage of carbon

Unlike in previous datasets, the carbon measurements are unbalanced – 15 functional observations, from which 10 were measured in the field and 5 in the forest soil. We model the percentage of carbon as

$$\mathcal{X}_s^{\mathrm{C}}(t) = \beta_0(t) + \beta_1(t) \cdot soil(s) + \beta_2(t) \cdot dist(s) + \delta_s(t), \ s \in D, \ t \in [3, 10], \quad (25)$$

where $dist(s)$ denotes the distance of the sampling point from the ecotone, such that $dist(s) = 3, 6, 9, 12, 15$. Unlike the case of the pH observations, the soil type indicator itself is here insufficient in capturing the spatial trend in the data. The variography results again suggest that the data are not spatially correlated (not shown). The test for equality of variances for field and forest part of the site leads to the global p-value of the test (16) equal to 0.055, thus we cannot reject the hypothesis of equality of variances. However, in the light of the simulation results, we opt for considering a heteroscedastic mod-

29

elling setting, the latter being more robust than that based on homoscedastic assumptions (see Section 4). We thus fit model (25) by WLS (RMSE = 2.945, NRMSE = 0.083 and R^2 = 0.857). Further, we consider a permutation test based on the standardized errors $\varepsilon_s(t) = \delta_{s(j)}(t)/\sigma(j)$, with $\widehat{\sigma}^2_{(A)} = 10.85$ and $\widehat{\sigma}^2_{(B)} = 4.06$.

Based on the heteroscedastic permutation test for the significance of covariates, we conclude that the percentage of carbon is significantly affected by the covariates (p-value = 0), again displaying a significant difference between soil and forest groups: the mean percentage of carbon is approximately 1.4 times higher in the forest compared to the field and it decreases approximately 0.04 % per 3 m with increasing distance from the ecotone at both parts of the site. Figure 6 reports the results for the estimated model (25).
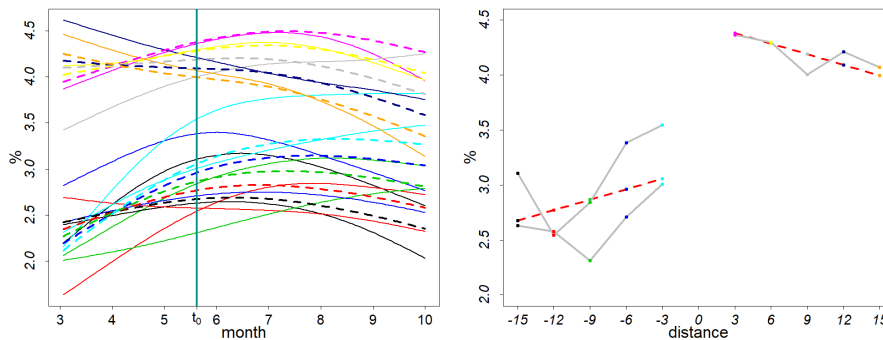

Figure 6: The percentage of carbon data, WLS model (25). Left: original functional observations (solid lines) and fitted model (dashed lines). Each colour represents the observation from a specific sample point. Right: a view of data in space at fixed t_0 (specified by a vertical line in the figure on the left). The points indicate the section of the curves at t_0, for clarity connected by grey lines. The fitted values are for clarity connected by a dashed red line.

6   Discussion and conclusion

In the framework of permutation tests for spatio-temporal functional mod-
elling, a testing procedure suitable for spatially heteroscedastic functional
data is introduced. We propose to fit a functional regression model to the
data, where the residuals are allowed to be characterized by heterogeneous
variances across groups. We thus extended the permutations schemes of
(Freedman & Lane, 1983) and (Abramowicz et al., 2018), to the standard-
ized errors of the model, assumed to be approximately exchangeable. With
the purpose of assessing the homoscedasticity assumption, we also proposed
a permutation-based test for equality of variances in two groups of data. The
proposed model and testing procedures can be easily extended to any number
$g \geq 2$ of groups. In this case, the mean term would be associated with $g - 1$
dummy variables, and the testing procedure could be performed (jointly) on
the associated coefficients, similarly as detailed in this work. In this more
general situation, heteroscedasticity among groups may be expected; tests
for heteroscedasticity could be developed by using an ANOVA-like setting,
or multiple pairwise tests with appropriate level corrections.

The proposed model allows to account for more general drift terms than
that used in the case study. This methodology could be also considered for
larger areas, provided that one is able to distinguish groups in data (e.g. by
type of soil, or possibly identied via a clustering procedure), or to quantify the
landuse types through one or more effective covariates. As usual in statistical
practice, an effective model formulation is here seen as a crucial point for a
successful use of the proposed methodology.

The testing procedure developed in this work is applicable to larger datasets as well, although it may clearly become computationally demanding for large datasets. In the literature, permutation schemes of similar structure have been successfully applied to datasets of relatively large size. For instance, in (Campos-Sánchez, Cremona, Pini, Chiaromonte, & Makova, 2016) several functional datasets of size larger than 1000 were considered. In our context, computations were relatively fast. For example, for the datasets of size $n = 500$ considered in the simulation study the testing procedure in Algorithm 1 run, on average, in 10 seconds, on a standard laptop (Windows 10 Home machine equipped with AMD Ryzen 7 2700U, 2.2GHz, 8GB RAM).

The simulation studies showed that the introduced test performs well concerning its size and power, especially in the case of unbalanced designs. In contrast, if the non-constant variance in the data is neglected and an ordinary least-squares approach is applied, the test is either too liberal or too conservative, whenever the two populations in data have unequal sample sizes. Based on simulations, the empirical size of the OLS test (10) could be more than 10 times smaller, or almost 5 times greater than the nominal one. However, for the balanced design, the OLS test works well even in the presence of heteroscedasticity. Our results are consistent with previous studies pointing out that the effect of heteroscedasticity is stronger for unbalanced designs ((Box et al., 1954); (Moder, 2010)). In fact, one should pay close attention to the specific spatial structure of the data during the model fitting, especially in the unbalanced design.

From the application viewpoint, the set of testing methodologies introduced was used to study geochemical functional data on KCl pH, H_2O pH,

and on the percentage of carbon. A balanced heteroscedastic, balanced homoscedastic and unbalanced heteroscedastic model, respectively, were fit to the data. Significance testing on of the functional regression parameters enables to conclude that (i) KCl pH is affected by the type of soil and the variability is different with respect to the soil type; (ii) values of H_2O pH depend on the type of soil, but the variability is not significantly different for field and forest part of the site; (iii) the percentage of carbon is associated with a more complex spatial model, depending not only on the soil type, but also on the distance from the ecotone, with increasing values when getting closer to the ecotone.

The results of the analysis are supported from the geochemical point of view. Indeed, a lower mean pH for both KCl and H_2O in the forest area may be due to a presence of fallen needles and their acidic character. As a consequence of its cumulation on the surface, it leaks into lower layers and the soil pH decreases. On contrary, in the agricultural area, a higher soil pH is indication of the agrotechnology in general, fertilization, a method of harvesting, etc. The behaviour of the organic carbon may be a confirmation of the *edge effect*. The amount of carbon is proportional to the water-retention of the soil. It is supposed to be highest in places with optimal hydro-pedological conditions, such as the ecotone. The results of the statistical analysis of the percentage of carbon from the Section 5.2 are in agreement with this claim. As we move to agricultural soil, less water is retained in the soil. One of the factors could be a bare surface which, indeed, is sensitive to the changes of temperature and to the evaporation during the day, month, or season. Another cause may be the human activity which extracts carbon from the

soil and has a negative impact on soil microorganisms, which contribute to the binding and prediction of soil carbon. Towards the forest, the behaviour of organic carbon is linked to the type and age of the forest. Moreover, due to the shadow and low pH caused by fallen needles, the soil microorganisms become less active. As we compare field and forest parts of the site, the different amount of carbon in general may be caused by a different degree of natural processes and cultivation. According to the results, the amount of carbon in forest soil has more variability than in agricultural soil (estimated forest variance is more than twice as high), presumably due to greater diversity of forest soils and the occurrence of humus.

Additional information and supporting material for this article is available online at the journal's website.

Data Availability Statement

The data and the computer code that support the findings of this study are openly available in Figshare.

Acknowledgement

References

Abramowicz, K., Häger, C. K., Pini, A., Schelin, L., Sjöstedt de Luna, S., & Vantini, S. (2018). Nonparametric inference for functional-on-scalar linear models applied to knee kinematic hop data after injury of the anterior cruciate ligament. *Scandinavian Journal of Statistics*, *45*(4), 1036–1061.

Anderson, M. J., & Robinson, J. (2001). Permutation tests for linear models. *Australian & New Zealand Journal of Statistics*, *43*(1), 75–88.

Aue, A., Gabrys, R., Horváth, L., & Kokoszka, P. (2009). Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, *100*(10), 2254–2269.

Bernardi, M. S., Carey, M., Ramsay, J. O., & Sangalli, L. M. (2018). Modeling spatial anisotropy via regression with partial differential regularization. *Journal of Multivariate Analysis*, *167*, 15–30.

Bernardi, M. S., Sangalli, L. M., Mazza, G., & Ramsay, J. O. (2017). A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province. *Stochastic environmental research and risk assessment*, *31*(1), 23–38.

Box, G. E., et al. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The annals of mathematical statistics*, *25*(2), 290–302.

Caballero, W., Giraldo, R., & Mateu, J. (2013). A universal kriging approach for spatial functional data. *Stochastic environmental research and risk*

assessment, *27*(7), 1553–1563.

Campos-Sánchez, R., Cremona, M. A., Pini, A., Chiaromonte, F., & Makova, K. D. (2016). Integration and fixation preferences of human and mouse endogenous retroviruses uncovered with functional data analysis. *PLoS computational biology*, *12*(6), e1004956.

Cressie, N. A. (1993). Statistics for spatial data. John Wiley and Sons. *Inc., New York*.

Delicado, P., Giraldo, R., Comas, C., & Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics: The official journal of the International Environmetrics Society*, *21*(3-4), 224–239.

Fortin, M.-J., & Jacquez, G. M. (2000). Randomization tests and spatially auto-correlated data. *Bulletin of the Ecological Society of America*, *81*(3), 201–205.

Fraiman, R., Justel, A., Liu, R., & Llop, P. (2014). Detecting trends in time series of functional data: A study of Antarctic climate change. *Canadian Journal of Statistics*, *42*(4), 597–609.

Freedman, D., & Lane, D. (1983). A nonstochastic interpretation of reported significance levels. *Journal of Business & Economic Statistics*, *1*(4), 292–298.

Giraldo, R., Delicado, P., & Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, *18*(3), 411–426.

Good, P. (2013). *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media.

Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications* (Vol. 200). Springer Science & Business Media.

Huang, Y., Xu, H., Calian, V., & Hsu, J. C. (2006). To permute or not to permute. *Bioinformatics*, *22*(18), 2244–2248.

Ignaccolo, R., Mateu, J., & Giraldo, R. (2014). Kriging with external drift for functional data for air quality monitoring. *Stochastic environmental research and risk assessment*, *28*(5), 1171–1186.

Kim, H.-J., Fay, M. P., Feuer, E. J., & Midthune, D. N. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in medicine*, *19*(3), 335–351.

Lindgren, K.-O. (2010). Dyadic regression in the presence of heteroscedasticity - An assessment of alternative approaches. *Social networks*, *32*(4), 279–289.

Manly, B. F. (2018). *Randomization, bootstrap and Monte Carlo methods in biology.* Chapman and Hall/CRC.

Marozzi, M. (2002, 02). Some notes on nonparametric inferences and permutation tests. *Metron - International Journal of Statistics*, *LX*, 139-151.

Menafoglio, A., Grujic, O., & Caers, J. (2016). Universal kriging of functional data: Trace-variography vs cross-variography? application to gas forecasting in unconventional shales. *Spatial Statistics*, *15*, 39–55.

Menafoglio, A., & Petris, G. (2016). Kriging for Hilbert-space valued random fields: The operatorial point of view. *Journal of Multivariate Analysis*, *146*, 84–94.

Menafoglio, A., & Secchi, P. (2017). Statistical analysis of complex and spatially dependent data: a review of object oriented spatial statistics.

European journal of operational research, *258*(2), 401–410.

Menafoglio, A., Secchi, P., & Dalla Rosa, M. (2013). A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space. *Electronic Journal of Statistics*, *7*, 2209–2240.

Moder, K. (2010). Alternatives to F-test in one way ANOVA in case of heterogeneity of variances (a simulation study). *Psychological Test and Assessment Modeling*, *52*(4), 343–353.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., & Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *79*(2), 381–404.

Oja, H. (1987). On permutation tests in multiple regression and analysis of covariance problems. *Australian Journal of Statistics*, *29*(1), 91–100.

Pechanec, V., Vávra, A., Hovorková, M., Brus, J., & Kiliánová, H. (2014). Analyses of moisture parameters and biomass of vegetation cover in southeast Moravia. *International Journal of Remote Sensing*, *35*(3), 967–987.

Pesarin, F., & Salmaso, L. (2010). *Permutation tests for complex data: theory, applications and software.* John Wiley & Sons.

Pini, A., & Vantini, S. (2017). Interval-wise testing for functional data. *Journal of Nonparametric Statistics*, *29*(2), 407–424.

Ramsay, J., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and MATLAB.* Springer-Verlag New York. doi: 10.1007/978-0-387-98185-7

Ramsay, J., & Silverman, B. W. (2005). *Functional data analysis.* Springer-Verlag New York. doi: 10.1007/978-1-4757-7107-7

Reiss, P. T., Huang, L., & Mennes, M. (2010). Fast function-on-scalar regression with penalized basis expansions. *The international journal of biostatistics*, *6*(1).

Romano, E., Verde, R., & Cozza, V. (2011). Clustering Spatial Functional Data: A Method Based on a Nonparametric Variogram Estimation. In *New perspectives in statistical modeling and data analysis* (pp. 339–346). Springer.

Schröder, T., & Fleig, F. D. (2017). Spatial patterns and edge effects on soil organic matter and nutrients in a forest fragment of southern Brazil. *Soil research*, *55*(7), 649–656.

Ter Braak, C. J. (1992). Permutation versus bootstrap significance tests in multiple regression and ANOVA. In *Bootstrapping and related techniques* (pp. 79–85). Springer.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage*, *92*, 381–397.

Zbíral, J., & Honsa, I. (2010). *Unified working procedures. Soil analysis I.* Central Institute for Supervising and Testing in Agriculture.

Zbíral, J., Tieffová, P., Plhalová, Š., Urbánková, E., Niedobová, E., Srnková, J., & Střížová, I. (2011). *Unified working procedures. Soil analysis II.* Central Institute for Supervising and Testing in Agriculture.

	design	I		II		III	
n	k	3	9	3	9	3	9
20	OLS	1.3	0.3	6.0	5.9	13.1	23.9
		(0.9,1.9)	(0.1,0.7)	(5.0,7.1)	(5.0,7.0)	(11.7,14.7)	(22.1,25.8)
	WLS	6.2	5.5	6.8	7.0	8.0	9.6
		(5.2,7.3)	(4.6,6.6)	(5.8,8.0)	(6.0,8.2)	(6.9,9.3)	(8.4,11.0)
40	OLS	1.2	0.2	4.9	5.4	12.5	21.5
		(0.8,1.8)	(0.1,0.5)	(4.0,5.9)	(4.5,6.5)	(11.1,14.0)	(19.8,23.4)
	WLS	5.7	4.5	5.5	6.5	6.1	8.2
		(4.8,6.8)	(3.7,5.5)	(4.6,6.6)	(5.5,7.7)	(5.1,7.2)	(7.1,9.5)
200	OLS	1.2	0.3	5.2	4.7	13.1	20.3
		(0.8,1.8)	(0.1,0.7)	(4.3,6.3)	(3.9,5.7)	(11.7,14.7)	(18.7,22.1)
	WLS	6.5	5.4	5.8	5.4	5.5	6.9
		(5.5,7.7)	(4.5,6.5)	(4.9,6.9)	(4.5,6.5)	(4.6,6.6)	(5.9,8.1)
500	OLS	1.1	0.5	4.9	4.6	12.4	19.9
		(0.7,1.6)	(0.2,0.9)	(4.0,5.9)	(3.8,5.6)	(11.0,13.9)	(18.2,21.7)
	WLS	5.8	6.5	5.4	5.3	5.6	6.2
		(4.9,6.9)	(5.5,7.6)	(4.5,6.4)	(4.4,6.3)	(4.7,6.7)	(5.2,7.3)

Table 1: The empirical sizes (in percentages) together with the 95% confidence intervals of tests T_{OLS} and T_{WLS} for model (17) under different conditions. The nominal size of tests was set to $\alpha = 5\%$. The sample size is denoted by n, the data are divided in two groups A and B, such that: unbalanced design I: $n_B = 3n_A$, balanced design II: $n_A = n_B$ and unbalanced design III: $n_A = 3n_B$. The difference among group variances is achieved through $k = 3, 9$, such that $\sigma^2_{(B)} = k\sigma^2_{(A)}$.

model	k	1	3	9
design I				
(17)	OLS	5.2 (4.3,6.3)	1.2 (0.8,1.8)	0.3 (0.1,0.7)
	WLS	5.7 (4.8,6.8)	6.5 (5.5,7.7)	5.4 (4.5,6.5)
(20)	OLS	6.2 (5.2,7.3)	2.5 (1.9,3.3)	2 (1.5,2.7)
	WLS	6.5 (5.5,7.7)	5.4 (4.5,6.5)	5.6 (4.7,6.7)
design II				
(17)	OLS	5.2 (4.3,6.3)	5.2 (4.3,6.3)	4.7 (3.9,5.7)
	WLS	5.4 (4.5,6.5)	5.8 (4.9,6.9)	5.4 (4.5,6.5)
(20)	OLS	4.3 (3.5,5.3)	6.3 (5.3,7.5)	7.3 (6.2,8.5)
	WLS	4.1 (3.3,5.1)	5.7 (4.8,6.8)	6.4 (5.4,7.6)
design III				
(17)	OLS	5.6 (4.7,6.7)	13.1 (11.7,14.7)	20.3 (18.6,22.1)
	WLS	6.0 (5.0,7.1)	5.5 (4.6,6.6)	6.9 (5.9,8.1)
(20)	OLS	4.3 (3.5,5.3)	15.6 (14.1,17.3)	24.7 (22.9,26.6)
	WLS	5.0 (4.1,6.0)	6.5 (5.5,7.7)	5.9 (5.0,7.0)

Table 2: The empirical sizes (in percentages) together with the 95% confidence intervals of tests T_{OLS} and T_{WLS} for models (17) and (20) and sample size $n = 200$. The data are divided in two groups A and B, such that: unbalanced design I: $n_B = 3n_A$, balanced design II: $n_A = n_B$ and unbalanced design III: $n_A = 3n_B$. The relationship among group variances is modelled via $k = 1, 3, 9$, such that $\sigma^2_{(B)} = k\sigma^2_{(A)}$.