**Spatial spline regression model**

Immagine che contiene testo

Descrizione generata automaticamente

SUMMARY:

We describe a model for the analysis of data distributed over irregularly shaped spatial domains with complex boundaries, strong concavities and interior holes. Adopting an approach that is typical of functional data analysis, we propose a spatial spline regression model that is computationally efficient, allows for spatially distributed covariate information and can impose various conditions over the boundaries of the domain. Accurate surface estimation is achieved by the use of piecewise linear and quadratic finite elements.

EXPLANATION:

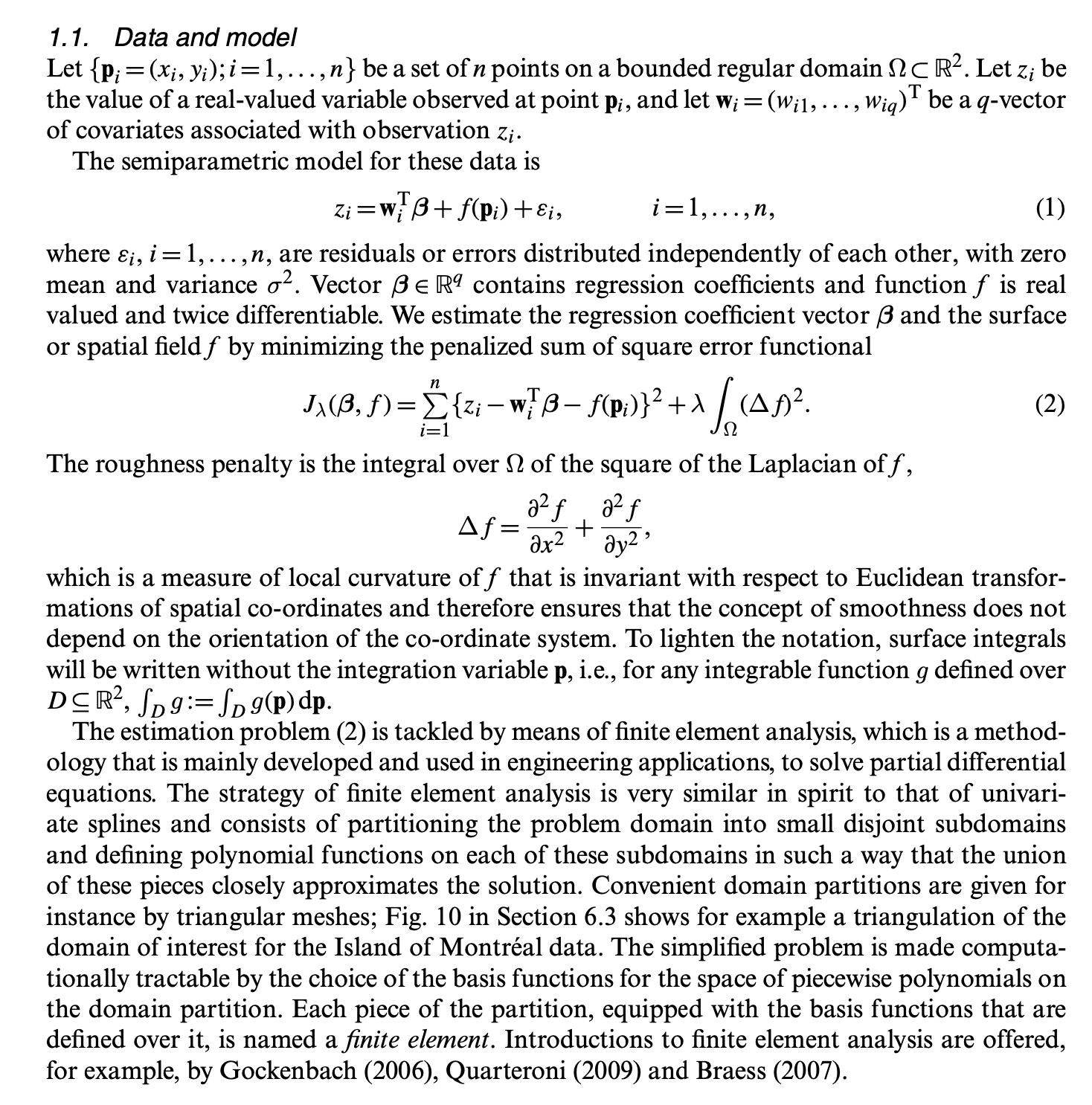
We propose a semiparametric model for the analysis of data distributed over spatial domains, including those with complex domain boundaries, strong concavities and interior holes.

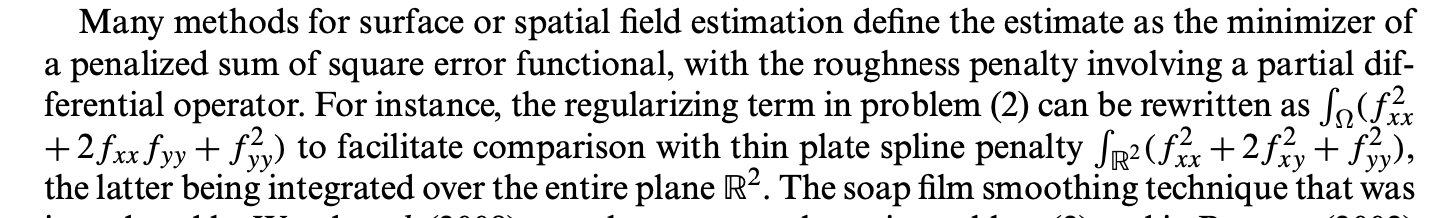
Well-known methods for spatial data analysis, such as kriging, kernel smoothing, wavelet-based smoothing, tensor product splines and thin plate splines, are not appropriate for these data, since they do not take into account the shape of the domain and also smooth across concave boundary regions;

We adopt a functional data analysis approach in proposing a *spatial spline regression* (SSR) model that overcomes these limitations, being able to deal efficiently with data distributed over irregularly shaped regions.

in this smoother, the roughness penalty consists of a Laplace operator that is integrated only over the region of interest thanks to a finite element formulation, that defines a system of local basis functions for continuous piecewise polynomial surfaces.

Our simulation studies show that SSR and soap film smoothing provide a large advantage over the other more classical techniques when dealing with data that are scattered over irregularly shaped domains.





**Lagrange triangular finite elements**

We consider a regular triangulation T of Ω, where adjacent triangles share either a vertex or a complete edge. Domain Ω is hence approximated by domain ΩT consisting of the union of all triangles, so that the boundary @Ω of Ω is approximated by a polygon (or more poly- gons, in the case for instance of domains with interior holes). It is assumed, therefore, that the number and density of triangles in T are sufficient to capture sharp features in @Ω as well as providing a basis for adequately describing the data.

Starting from set **V** of data locations has the advantage of providing triangulations that are naturally finer where there are more data points, and coarser where data points are fewer;

The surface to be constructed over ΩT is assumed to be a polynomial in *x* and *y* over any triangle and is continuous across edges and vertices. We consider the case of linear polynomial 🡪 linear polynomial defined over each triangle as linear combination of three basis functions, each having value 1 at a single vertex and 0 at the other two.

Each such piecewise polynomial basis function is called a *Lagrange finite element*.

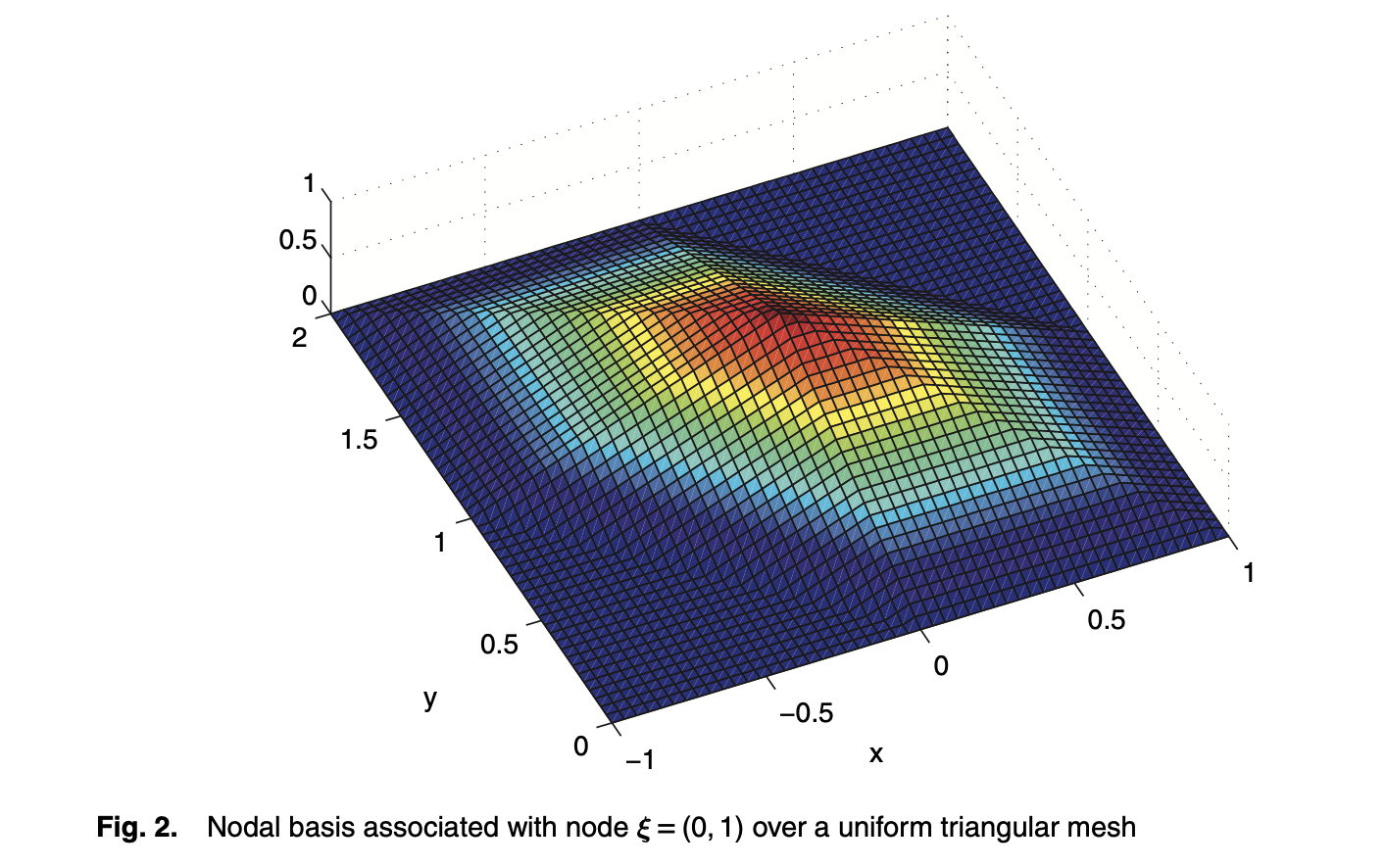


Immagine che contiene testo

Descrizione generata automaticamente

**Modeling spatially dependent functional data via regression with differential regularization**

In our case, instead of the PDE in the regularization term in space, we simply had the second order derivative of f with respect to space (Laplacian)

Immagine che contiene testo

Descrizione generata automaticamente

Abstract

We propose a method for modeling spatially dependent functional data, based on regres- sion with differential regularization. The regularizing term enables to include problem- specific information about the spatio-temporal variation of the phenomenon under study, formalized in terms of a time-dependent partial differential equation. The method is implemented using a discretization based on finite elements in space and finite differences in time. This non-tensor product basis allows to handle efficiently data distributed over complex domains and where the shape of the domain influences the phenomenon’s behavior. Moreover, the method can comply with specific conditions at the boundary of the domain of interest.

EXPLANATION

We consider the problem of modeling functional data with complex dependencies, such as spatially dependent curve data or time dependent surface data.

Here we extend spatial regression with differential regularization [5,38,41] to spatially dependent functional data.

the estimation problems can be discretized in space by means of the finite element method, similarly to [7,38,41], and in time by means of the finite difference method. The finite element basis used in space allows to handle efficiently data distributed over irregularly shaped domains. This is crucial when the shape of the domain influences the phenomenon under study, as in the applied problem that has stimulated this research.

First of all, the shape of the domain, the carotid cross section, influences the spatio-temporal blood flow velocity field, and hence must be explicitly considered during the estimation process. Unfortunately, almost all the available techniques naturally work over rectangular or tensorized domain.

We can here profit of a detailed problem-specific information, that can be formalized in terms of a time- dependent PDE, modeling the spatio-temporal behavior of the phenomenon under study. Using the proposed approach, this problem-specific information, thus formalized, can be profitably included in the estimation process, to define an anisotropic and spatially non-stationary estimator that yields physiological estimates

**MODEL**

describes the proposed spatio-temporal regression with time-dependent PDE regularization, under the simplifying assumption that the functional data are available continuously over time;

discretization of the estimation problem by means of the finite element method in space and the finite difference method in time.

COMPARISONS OF METHODS

The other methods we consider are based on differential regularization with two roughness terms that account separately for the regularity of the field in space and in time. All these methods use a tensor product approach.

The first method, denoted by the acronym TPS, adopts a thin plate spline basis in space and a cubic B-spline basis in time; the spatial penalty is the thin plate spline energy and the temporal penalty is the L2 norm of the second derivative in time. The second method, denoted by the acronym SOAP and proposed by [3,29], uses soap film smoothing in space [46] and cubic B-splines in time; the penalization is composed by the L2 norm of the Laplacian in space and the L2 norm of the second derivative in time. Both TPS and SOAP are implemented using the function gam of the R package mgcv [45]. The last model we consider, denoted by the acronym ST-PDE and proposed in [7], employs finite elements in space and cubic B-Splines in time; this method penalizes the L2 norm of the Laplacian in space and the L2 norm of the second derivative in time. Both the proposed ST-tPDE method and the ST-PDE method by [7] are implemented in R and C++, based on the R package fdaPDE.

MODEL SELECTION

The values of the smoothing parameters for the four methods are selected, for each replicate and method, by generalized cross-validation.

Immagine che contiene testo

Descrizione generata automaticamente

PAPER MARA  
**A PENALIZED REGRESSION MODEL FOR SPATIAL FUNCTIONAL DATA WITH APPLICATION TO THE ANALYSIS OF THE PRODUCTION OF WASTE IN VENICE PROVINCE**

ABSTRACT

We propose a method for the analysis of functional data with complex dependencies, such as spa- tially dependent curves or time dependent surfaces, over highly textured domains. The models are based on the idea of regression with partial differential regularizations. In particular, we consider here two roughness penalties that account separately for the regularity of the field in space and in time. Among the various modelling features, the proposed method is able to deal with spatial domains fea- turing peninsulas, islands and other complex geometries. Space-time varying covariate information is included in the model via a semi-parametric framework. The proposed method is compared via simulation studies to other spatio- temporal techniques and it is applied to the analysis of the annual production of waste in the towns of Venice province.

MODEL

In this work we deal with spatio-temporal data distributed over a spatial domain which presents complex geometries. That is, the irregular shape of the domain influences the phenomenon under study and there are important

geographical elements within the boundary such as islands and peninsulas that impact the distribution of the data. We refer to such domains as textured.

**NB: our assumption 🡪 we shall make a strong simplification of the nature of these data, and consider them in the framework of geostatistical functional data where the datum is observable in principle in any point of the domain, instead of in the framework of functional areal data.**

Immagine che contiene testo, quotidiano, screenshot

Descrizione generata automaticamente

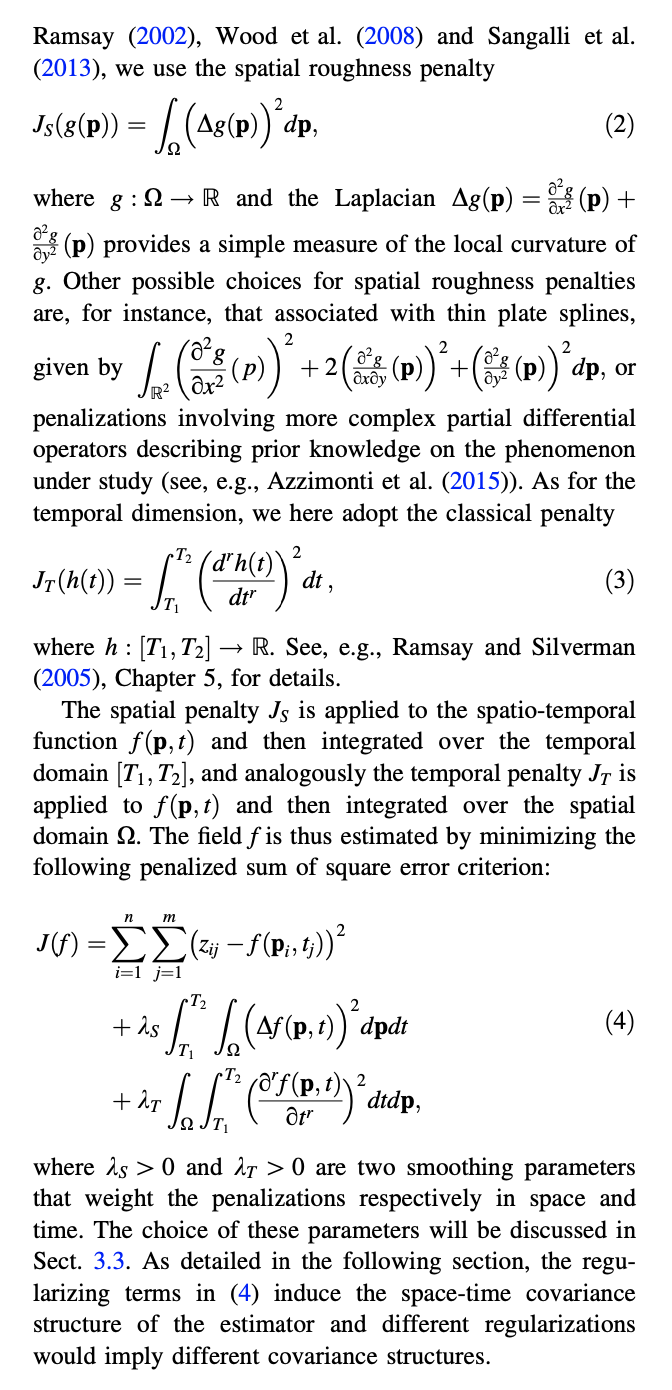


Immagine che contiene testo

Descrizione generata automaticamente

Immagine che contiene testo

Descrizione generata automaticamente

**Basis system in space-time**

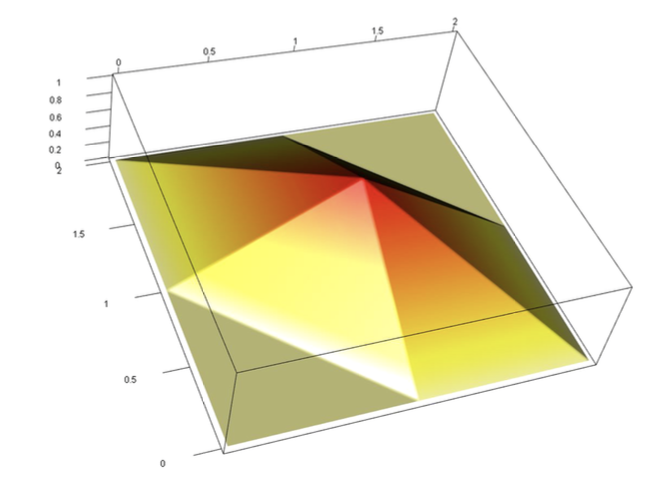
In this work, we use in space a finite element basis on a triangulation Xs of the spatial domain X of interest. This choice leads to an effi- cient discretization of the functional J and allows an accurate account of the shape of the spatial domain.

A triangulation of the resulting simplified domain is then obtained using the R package fdaPDE (Lila et al. 2016). In particular, we start from a Delaunay triangulation, con- strained within the simplified boundary, where each of the town locations and each point defining the simplified boundary become a triangle vertex. A more regular mesh is then obtained with additional vertices, imposing a maxi- mum value to the triangle areas.

instead of using as coordinates the latitude and longitude, we employ the UTM coordinate system, which allows to compute the distance between two points on the Earth’s surface by means of the Euclidean distance instead of the geodesic distance.

The finite element basis is composed by globally con- tinuous functions that coincides with a polynomial of a certain degree on each element of the domain triangulation. In particular we use here linear finite element basis, that are

piecewise linear functions. The dimension of the spatial basis is strictly related to the triangulation of the spatial domain: there is one basis function for each knot of the triangulation; for linear finite elements, each basis is associated to a vertex of the triangulation and has value 1 at that vertex and 0 at all other vertices. Figure 7 shows an example of linear basis function.



For the temporal dimension, we use here a cubic B-spline basis with penalization of the second derivative, with knots coinciding with the sampling time instants of the dataset.

Finally, the values of the smoothing parameters lambdaS and lambdaT may be chosen via Generalized Cross-Validation (GCV).

Immagine che contiene testo

Descrizione generata automaticamente

Immagine che contiene testo, dispositivo, calibro

Descrizione generata automaticamente

**SIMULATION STUDIES**

We then consider two space-time models presented in Augustin et al. (2013) and Marra et al. (2012). One model adopt a thin plate spline basis in space and a cubic spline basis in time, and minimizes a functional analogous to (16), where the spatial penalty is replaced by the thin plate spline energy recalled in Sect. 2. The other model uses the soap film smoothing described in Wood et al. (2008) in space and a cubic spline basis in time, and minimizes the same functional as in (16). The two latter methods are imple- mented using the function gam of the R package mgcv (Wood 2006). Finally, for these two methods, as well as for the model here proposed, the values of the smoothing parameters lambda\_S, lambda\_T are chosen via GCV.

We present here five simulation studies: in the first two cases we consider different sampling schemes, with fewer or more observations in space and in time; in the third case we include covariates; in the fourth case we consider cor- related noise; finally, in the fifth case we start from areal data and consider the approximation consisting in assign- ing each datum to the area centroid.

Simulation with areal data assigned to area centroid:

In any case, we show that also in this approximated and simplified data setting, the model proposed outperforms the competitor methods.

SMOOTH.FEM.TIME() FUNCTION

Bestlambda: If GCV is TRUE, a 2-elements vector with the indices of smoothing parameters returnig the lowest GCV

PARAMETERS INFERENCE

Immagine che contiene testo

Descrizione generata automaticamente

***Sign-flip score test for SR-PDE***

Immagine che contiene testo

Descrizione generata automaticamente

***Eigen sign-flip score test for SR-PDE***

We now propose a modification of the sign-flip test introduced in Section 4.1, which is aimed to avoid the issues related to the lack of invariance under sign-flips of the covariance of the statistics *T*⇧. We name this new test Eigen sing-flip test.

The idea behind the eigen decomposition is inspired by similar approaches proposed by [24, 25] for the classical linear model. In the context of classical linear models, the covariance matrix of the residuals is not diagonal, thus making the observations not exchangeables. To ensure a safe application of the permutation principle, the authors thus introduce the premultiplication of residuals and predictors by the eigenvectors of the residual projection matrix. Since the residual projection matrix is positive semi-definite and idempotent, its eigenvalues are only ones and zeros. This makes the resulting transformed residuals exchangeable (and reduced in number, equal to the rank of the residual projection matrix). In our case, the smoothing matrix is not a projection matrix; therefore, the eigenvalues do not take values in {0, 1}. Because of this, the rescaled residuals, on which the test statistic in (14) is based, are not homoscedastic, and the permutation approach is hence inapplicable. For this reason our proposal, unlike the method of [25], leverages instead on the sign-flip procedure. The proposed approach, together with the decomposition of the smoothing matrix, leads to a preservation of the covariance structure with respect to random sign-flips, as shown in Proposition 4.

As highlighted by the simulations in Section 5, the Eigen sign-flip score test here presented, with respect to the sign-flip score test in Section 4.1, ensures an higher power, while keeping a good control of the Type-I error.

The eigen sign-flip (solid red lines), on the contrary, it is not a↵ected by such problem and it is able to preserve the covariance structure of the test statistic also in the finite sample scenario, even when the covariate exhibits a strong spatial dependence. This leads, even in the more complicated cases, to a very good control of Type-I error, accompanied by an high power.

Result of test

The test returns a *p*-value of 0.326; this concludes quite clearly that the elevation does indeed not have a significant impact on the rainfall. This is probably due to the fact that the effect of elevation on rainfall is not linear, so that this effect is not captured by a model where elevation is included as a linear regressor. Rather, the distribution of rainfall is the result of complex interaction between the geomorphology of the region and the atmospheric circulation.

**POSSIBLE IMPROVEMENTS**

**Immagine che contiene testo

Descrizione generata automaticamente**

**Immagine che contiene testo

Descrizione generata automaticamente**

TRANSFORM THE COVARIATES IN UTM

Extending the work of Azzimonti et al. (2015) it is also possible to include a priori information available on the phenomenon under study, using more complex differential regularizations modelling the space and/or time behavior of the phenomenon. This also allows to account for non-sta- tionarities and anisotropies in space and/or time. Along the same lines, if a priori information about the interaction between space and time was available, then it would make sense to consider a unique space/time regularizing term based on a time-dependent PDE that governs the phe- nomenon behavior.

Loss correlate a tanti fattori, pero varroa sembra essere quello piu influente in inverno; ma le api non muoiono di inverno perche fa freddo, paradossalmente negli stati piu freddi le loss sono piu basse. Allora se non è temperatura, proviamo a vedere gli stressori, in particolare il varroa, è il piu frequente. Vediamo che il varroa (con cluster di Gege), è divisibile in gruppi; in alcuni (est), la loss è piu alta. Modello bayesiano clusterizza su persistenza.

GAM 🡪 spiega che a ogni aumento percentuale di 1% del Varroa, porta a 0.11% in piu di colony loss (più impattante tra tutti gli stressori, coerente che ci siano fattori che non conosciamo).

Modello con prezzo mostra che la perdita in valore economico (legata alla colony loss in percentuale) è più alta negli stati continentali, in particolare a est, coerente con il cluster fatto sul varroa 🡪 stati non stanno prendendo abbastanza misure contro varroa o per altre motivazioni ambientali il varroa resiste maggiormente.