

Modeling prior knowledge on complex phenomena behaviors via partial differential equations

Come descrivere informazione a priori sul comportamento di fenomeni complessi mediante equazioni alle derivate parziali

Laura Azzimonti, Laura M. Sangalli and Piercesare Secchi

Abstract In this talk we shall describe an innovative method for the analysis of spatially distributed data, when a prior knowledge on the phenomenon under study is available and can be formalized in terms of a partial differential model. The prior knowledge may for instance derive from physics, physiology, morphology, chemistry or mechanics of the problem at hand, and may as well concern conditions characterizing the phenomenon at the boundary of the problem domain. The proposed models exploit advanced numerical techniques and specifically make use of the Finite Element method.

Abstract *In questo seminario presenteremo un metodo innovativo per l'analisi di dati spazialmente distribuiti, quando una conoscenza a priori del fenomeno sotto studio è disponibile e può essere formalizzata in termini di un modello alle derivate parziali. Tale informazione a priori può per esempio derivare dalla fisica, fisiologia, morfologia, chimica o meccanica del problema in esame, e può altresì riguardare condizioni che caratterizzano il fenomeno ai bordi del dominio di interesse del problema. I modelli proposti sfruttano tecniche numeriche avanzate ed in particolare fanno uso del metodo degli elementi finiti.*

Key words: functional data analysis, spatial data analysis, penalized regression.

1 Introduction

We present a class of regression models for spatially distributed data, where the estimated entity both adhere to the data and incorporate problem-specific a priori knowledge on the phenomenon under study, defined by a Partial Differential Equation (PDE). PDEs are indeed commonly used to describe complex phenome-

Laura Azzimonti, Laura M. Sangalli and Piercesare Secchi
MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano
e-mail: laura.azzimonti@polimi.it, laura.sangalli@polimi.it, piercesare.secchi@polimi.it

na behaviors in many fields of engineering and sciences, including bio-sciences, geo-sciences and physical-sciences.

The inclusion of simple forms of PDEs in statistical models has been recently explored by other researchers. For instance, [7] proposes Bayesian spatial models that link Gaussian fields and Gaussian Markov random fields via a stochastic PDE inducing a Matern covariance. [11] considers more general Bayesian inverse problems. Other recent proposals, such as [9, 12, 6, 4, 10], assume a penalized regression form, with the roughness term involving the partial differential operator.

The proposed class of models, Spatial Regression with PDE regularization (SR-PDE), generalizes the method in [10] and considers more complex PDEs with respect to the methods mentioned above. See [3, 2, 1]. An important novelty of the proposed models is that the PDE is here used to model the space variation of the phenomenon, using problem-specific a priori information. Moreover SR-PDE allows for important modeling flexibility in this respect, accounting also for space anisotropy and non-stationarity in a straightforward way, as well as unidirectional smoothing effects.

2 Data and model

Let $\{\mathbf{p}_i = (x_i, y_i); i = 1 \dots, n\}$ be a set of n points on a bounded domain $\Omega \subset \mathbb{R}^2$ with boundary $\partial\Omega \in C^2$. Let z_i be the value of a real-valued variable of interest observed at location $\mathbf{p}_i$. Assume the model

$$z_i = f(\mathbf{p}_i) + \varepsilon_i \quad (1)$$

where $\varepsilon_i, i = 1, \dots, n$, are independent errors with zero mean and constant variance σ^2 , and $f : \Omega \rightarrow \mathbb{R}$ is the surface or spatial field to be estimated. Assume that problem specific prior information is available, that can be described in terms of a PDE, $Lf = u$, modeling to some extent the phenomenon under study. Moreover, the prior knowledge can also concern possible conditions that f has to satisfy at the boundary $\partial\Omega$ of the problem domain. Generalizing the models in [9] and [10], we propose to estimate f by minimizing the penalized sum-of-square-error functional

$$J(f) = \sum_{i=1}^n (f(\mathbf{p}_i) - z_i)^2 + \lambda \int_{\Omega} (Lf(\mathbf{p}) - u(\mathbf{p}))^2 d\mathbf{p} \quad (2)$$

with respect to $f \in V$, where V is the space of functions in $L^2(\Omega)$ with first and second derivatives in $L^2(\Omega)$, that satisfy the required boundary conditions. The penalized error functional hence trades off a data fitting criterion, the sum-of-square-error, and a model fitting criterion, that penalizes departures from a PDE problem-specific description of the phenomenon. This trade off is controlled by the positive smoothing parameter λ : the smaller this smoothing parameter the more the estimate will adhere to the observed data, the higher this smoothing parameter the higher the

strength given the a priori information on f described by the differential regularization. Although not itself formalized in a Bayesian framework, the proposed model can be seen as a regularized least square analogous to the Bayesian inverse problems presented, e.g., in [11]. In particular, the least square term in $J(f)$ corresponds to a log-likelihood for Gaussian errors, while the regularizing term translates the prior knowledge on the spatial field f , with λ controlling the strength or concentration of the prior. With respect to [11], besides the different model framework, we are able to deal with a larger class of operators, including also non-stationary anisotropic diffusion, transport and reaction terms.

In particular, we consider here phenomena that are well described in terms of linear second order elliptic operators L ,

$$Lf = -\operatorname{div}(\mathbf{K}\nabla f) + \mathbf{b} \cdot \nabla f + cf, \quad (3)$$

where the symmetric and positive definite matrix $\mathbf{K} \in R^{2 \times 2}$ is the diffusion tensor, $\mathbf{b} \in R^2$ is the transport vector and $c \geq 0$ is the reaction term. These terms may be space varying, and they induce non-stationary and anisotropic smoothing effects, non-stationary unidirectional smoothing effects and non-stationary shrinking effects, thus allowing for a very flexible modeling of the space variation. The forcing term $u \in L^2(\Omega)$ can be either $u = 0$ (homogeneous case) or $u \neq 0$ (non-homogeneous case), allowing for further model flexibility.

The prior information can moreover concern conditions characterizing the phenomenon at the boundary $\partial\Omega$ of the problem domain. In many applicative contexts, such as the one described in Section 3, accurately complying with this prior knowledge is crucial to obtain meaningful estimates. SR-PDE models can include this prior information. In particular, the boundary conditions may involve the evaluation of the function and/or its normal derivative at the boundary, allowing for a very complex modeling of the behavior of the spatial field at the boundary of the domain of interest.

In [3] we show that, under suitable regularity conditions, the estimation problem (1) is well posed. The spatial field estimator is obtained solving a coupled system of PDEs. This infinite dimensional problem is thus reduced to a finite dimensional one using finite elements, that provide a space of piecewise polynomial surfaces. Finite element analysis (see, e.g., [8]) is mainly developed and used in engineering applications, to solve PDEs. The strategy of finite element analysis is very similar in spirit to that of univariate splines and consists of partitioning the problem domain into small disjoint subdomains and defining polynomial functions on each of these subdomains in such a way that the union of these pieces closely approximates the solution to the estimation problem. The introduction of the finite element space reduces the estimation problem to solving a linear system. The resulting estimators are linear in the observed data values and have a penalized regression form. The classical inferential tools are thus derived. Moreover, the use of finite elements makes the method computationally highly efficient. The good numerical properties of the estimators are analyzed in [2]. The models can also be extended to include space-

varying covariate information, according to the semi-parametric approach described in [10].

3 Illustrative problem

The motivating applied problem concerns the estimation of the blood-flow velocity field on a cross-section of the common carotid artery, using data provided by echo-color doppler acquisitions. This applied problem arises within the research project *MAThematichs for CARotid ENdarterectomy@MOX* (*MACAREN@MOX*), which aims at studying atherosclerosis pathogenesis. Carotid Echo-Color Doppler (ECD) is a medical imaging procedure that uses reflected ultrasound waves to create images of an artery and to measure the velocity of blood cells in some locations within the artery. Specifically, the ECD data measure the velocity of blood within beams located on the considered artery cross-section. For this study, during the ECD scan 7 beams are considered, located in a cross-shaped pattern.

In this applied problem we have a prior knowledge on the phenomenon under study that could be exploited to derive accurate physiological estimates. There is in fact a vast literature devoted to the study of fluid dynamics and hemodynamics, see for example [5] and references therein. This prior information concerns both the shape of the field, which can be conveniently described via a linear second order PDE, and the conditions at the boundary of the problem domain, i.e., specifically, at the wall of the carotid cross-section. The SR-PDE method efficiently uses the a priori information on the phenomenon under study and returns a realistic and physiological estimate of the dynamic blood flow, which is not affected by the pattern of the observations. See [3]. Moreover SR-PDE estimates accurately highlight important features of the blood flow, such as eccentricity, asymmetry and reversion of the fluxes, that are of interest to the medical doctors, in order to understand how the local hemodynamics influences atherosclerosis pathogenesis.

Acknowledgements Funding by the research program Dote Ricercatore Politecnico di Milano - Regione Lombardia, project “Functional data analysis for life sciences”, and by the starting grant *FIRB Futuro in Ricerca*, MIUR Ministero dell’Istruzione dell’Università e della Ricerca, research project “Advanced statistical and numerical methods for the analysis of high dimensional functional data in life sciences and engineering”(http://mox.polimi.it/users/sangalli/firbSNAPLE.html).

References

1. Azzimonti L.: Blood flow velocity field estimation via spatial regression with PDE penalization, PhD Thesis, Politecnico di Milano (2013)
2. Azzimonti, L., Nobile, F., Sangalli, L. M., Secchi, P.: Mixed Finite Elements for spatial regression with PDE penalization. TechRep. MOX 20/2013, Dipartimento di Matematica, Politecnico di Milano (2013)

3. Azzimonti, L., Sangalli, L. M., Secchi, P., Domanin, M., Nobile, F.: Blood flow velocity field estimation via spatial regression with PDE penalization. TechRep. MOX 19/2013, Dipartimento di Matematica, Politecnico di Milano (2013)
4. D’Elia, M., Perego, M., and Veneziani, A.: A Variational Data Assimilation Procedure for the Incompressible Navier-Stokes Equations in Hemodynamics. *SIAM Journal of Scientific Computing*, Vol. 52, Part 4, pp. 340–359. 2012, Volume 52, Issue 2, pp 340-359 (2012)
5. Formaggia, L., Quarteroni, A., and Veneziani, A.: Cardiovascular mathematics: modeling and simulation of the circulatory system, Springer (2009)
6. Guillas, S. and Lai, M.: Bivariate splines for spatial functional regression models. *Journal of Nonparametric Statistics*, Vol. 22, Part 4, pp. 477–497 (2010)
7. Lindgren, F., Rue, H., and Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Ser. B, Statistical Methodology*, Vol. 73, pp. 423–498, with discussions and a reply by the authors (2011)
8. Quarteroni, A. Numerical Models for Differential Problems. *Springer* (2013)
9. Ramsay, T.: Spline Smoothing over Difficult Regions. *Journal of the Royal Statistical Society Ser. B, Statistical Methodology*, Vol. 64, pp. 307–319 (2002)
10. Sangalli, L. M., Ramsay, J. O., Ramsay, T. O.: Spatial spline regression models. *Journal of the Royal Statistical Society Ser. B, Statistical Methodology*, Vol. 75, Part 4, pp. 681–703 (2013)
11. Stuart, A.: Inverse problems: a Bayesian perspective. *Acta Numerica*, Vol. 19, pp. 451–559 (2010)
12. Wood, S. N., Bravington, M. V., and Hedley, S. L.: Soap film smoothing. *Journal of the Royal Statistical Society Ser. B, Statistical Methodology*, Vol. 70, pp. 931–955 (2008)