# Supplementary material for
# Spatial regression models over two-dimensional manifolds

Published on *Biometrika* (2016), 103 (1), 71-88.

BY B. ETTINGER

*Mathematics and Computer Science Department, Emory University, Atlanta, GA 30322, U.S.A.*
betting@emory.edu

S. PEROTTO AND L. M. SANGALLI

*Laboratory for Modeling and Scientific Computing MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy*
simona.perotto@polimi.it laura.sangalli@polimi.it

## 7. PROOF OF PROPOSITION 1

For the proof of Proposition 1, we need the following characterization theorem from Chapter 2 of Braess (2007).

*Lemma.* Let $\mathcal{A}$ be a symmetric coercive bilinear form on a vector space $V$ and $\mathcal{B}$ a linear form on $V$. Then, there exists in $V$ a minimizer $\zeta$ of the form $\mathcal{A}(y,y) - 2\mathcal{B}(y)$, for $y \in V$, if and only if $\mathcal{A}(\zeta, y) = \mathcal{B}(y)$, for any $y \in V$. Moreover, the minimizer $\zeta$ is unique. □

Now, we write the functional in (5) as

$$J_{\Omega,\lambda}(f \circ X) = z^T z - 2\mathrm{f}_N^T z + \mathrm{f}_N^T \mathrm{f}_N + \lambda \int_\Omega \frac{1}{\mathcal{D}} \Big[ \mathrm{div}\big\{ K \nabla (f \circ X) \big\} \Big]^2 d\Omega.$$

Since we are solving the optimization problem with respect to $f \circ X$, we can ignore the terms that are constant with respect to $f \circ X$ and look for a solution $f \circ X \in H^2_{n0,K}(\Omega)$ that minimizes

$$\tilde{J}_{\Omega,\lambda}(f \circ X) = \left( \mathrm{f}_N^T \mathrm{f}_N + \lambda \int_\Omega \frac{1}{\mathcal{D}} \Big[ \mathrm{div}\big\{ K \nabla (f \circ X) \big\} \Big]^2 d\Omega \right) - 2\mathrm{f}_N^T z.$$

To minimize $\tilde{J}_{\Omega,\lambda}(f \circ X)$, we apply Lemma 1 with

$$\mathcal{A}(f \circ X, y \circ X) = \mathrm{y}_N^T \mathrm{f}_N + \lambda \int_\Omega \frac{1}{\mathcal{D}} \mathrm{div}\big\{ K \nabla (f \circ X) \big\} \mathrm{div}\big\{ K \nabla (y \circ X) \big\} d\Omega,$$

$\mathcal{B}(y) = \mathrm{y}_N^T z$, and $V = H^2_{n0,K}(\Omega)$. To show that the bilinear form $\mathcal{A}$ is coercive, we suppose that $\mathcal{A}(f \circ X, f \circ X) = 0$ for some $f \circ X \in H^2_{n0,K}(\Omega)$. Then, we have $\mathrm{f}_N^T \mathrm{f}_N = 0$ and $\int_\Omega \mathcal{D}^{-1} [\mathrm{div}\big\{ K \nabla (f \circ X) \big\}]^2 d\Omega = 0$, where $\mathcal{D}$ is positive and the matrix $K$ is positive definite. The boundary conditions imposed on the co-normal derivatives in $H^2_{n0,K}(\Omega)$ force $f \circ X$ to be a constant on $\Omega$. Moreover, the condition $\mathrm{f}_N^T \mathrm{f}_N = 0$ implies that $f \circ X$ is the constant null function on $\Omega$. Thus $\mathcal{A}$ is coercive on $H^2_{n0,K}(\Omega)$ and via Lemma 1, the function $\hat{f} \circ X$ is the unique minimizer of (5) in $H^2_{n0,K}(\Omega)$ if and only if $\hat{f} \circ X$ satisfies (8).

## 8. WEAK FORMULATION OF THE ESTIMATION PROBLEM

To obtain an equivalent formulation for (8) suited for a finite element approximation, we introduce an auxiliary function $g$ defined on $\Gamma$. Then, the problem of finding $\hat{f}$ defined on $\Gamma$ such that $\hat{f} \circ X \in H^2_{n0,K}(\Omega)$ satisfies (8) for any $q$ on $\Gamma$ where $q \circ X \in H^2_{n0,K}(\Omega)$, can be rewritten as the problem of finding a pair of functions $\hat{f}$ and $g$ such that $(\hat{f} \circ X, g \circ X) \in H^2_{n0,K}(\Omega) \times L^2(\Omega)$ and satisfies

$$\begin{aligned} &\mathrm{q}_n^T \hat{\mathrm{f}}_n + \lambda \int_\Omega (g \circ X)\ \mathrm{div}\big\{K \nabla (q \circ X)\big\} d\Omega = \mathrm{q}_n^T z \\ &\int_\Omega (g \circ X)(\zeta \circ X) \mathcal{D} d\Omega - \int_\Omega \mathrm{div}\big\{K \nabla (\hat{f} \circ X)\big\}(\zeta \circ X) d\Omega = 0 \end{aligned} \tag{22}$$

for any $(q \circ X, \zeta \circ X) \in H^2_{n0,K}(\Omega) \times L^2(\Omega)$. If the pair $(\hat{f} \circ X, g \circ X) \in H^2_{n0,K}(\Omega) \times L^2(\Omega)$ satisfies (22) for any $(q \circ X, \zeta \circ X) \in H^2_{n0,K}(\Omega) \times L^2(\Omega)$, then $\hat{f} \circ X$ also satisfies (8). Of course, if $\hat{f} \circ X$ satisfies (8), then the pair $[\hat{f} \circ X, \mathrm{div}\{K \nabla (f \circ X)\}]$ satisfies (22). Now, we ask for higher regularity of the auxiliary function $g$ and the test function $\zeta$, i.e., $g \circ X, \zeta \circ X \in H^1(\Omega)$. With the added regularity and by exploiting Green's Theorem, problem (22) can be reformulated as finding $(\hat{f} \circ X, g \circ X) \in \{H^1_{n0,K}(\Omega) \cap C^0(\bar{\Omega})\} \times H^1(\Omega)$ such that (12) is verified for any $(q \circ X, \zeta \circ X) \in \{H^1_{n0,K}(\Omega) \cap C^0(\bar{\Omega})\} \times H^1(\Omega)$. Moreover, the elliptic regularity property ensures that $\hat{f} \circ X$ still belongs to $H^2_{n0}(\Omega)$ (see, e.g., Lions & Magenes, 1973, Chapter 8).

## 9. COMPUTATIONAL ALGORITHM FOR THE FLATTENING MAP

We approximate the conformal coordinates $u_{[1]}$ and $u_{[2]}$ in (9)-(10) in the space $H^1_{\mathcal{T}}(\Gamma)$, i.e., with functions that are globally continuous and linear over each triangle of $\Gamma_{\mathcal{T}}$. Care must be used in the choice of the three-dimensional mesh $\Gamma_{\mathcal{T}}$, because degenerate triangles can be generated by flattening a triangle with all the vertices on the boundary. Below, we outline the finite element procedure adopted to approximate the flattening map.

1. An approximation of $u_{[1]}$ is found by minimizing $\mathcal{E}(u_{[1]})$ in (11) over $H^1_{\mathcal{T}}(\Gamma)$. The energy functional is invariant with respect to conformal changes of the domain metric (see, e.g. Pinkall & Polthier, 1993). This fact yields a convenient cotangent formula for the stiffness matrix $D$. In particular, the entries of $D$ are computed by gathering terms associated with the same edge. Specifically, if $x_j$ and $x_l$ are connected by an edge of the triangular mesh $\Gamma_{\mathcal{T}}$, then $D_{jl} = -\frac{1}{2}(\cot \alpha_j + \cot \beta_j)$ where $\alpha_j$ and $\beta_j$ are the angles opposite to the edge identified by $x_l$ and $x_j$. If $x_j$ and $x_l$ are not connected by an edge, then $D_{jl} = 0$. The diagonal entries of $D$ are such that $D_{jj} = -\sum_{l \neq j} D_{jl}$. For the boundary conditions stated in (9), each interior vertex $x_j \in \Gamma_{\mathcal{T}}$ satisfies
$$\sum_{x_l \in \Gamma_{\mathcal{T}}} D_{jl} u_{1l} = -\sum_{x_l \in b_1} D_{jl}. \tag{23}$$
Solving the system above approximates the conformal parameter $u_{[1]}$.
2. Cut $\Gamma_{\mathcal{T}}$ following the gradient of $u_{[1]}$. The maximum principle ensures that the solution to a Laplace equation reaches its maximum on the boundary (Lions & Magenes, 1973). This implies that there always exists a vertex adjacent to the current vertex with a larger value. We

use this fact to find the cut $C$ on the surface. Start by picking a vertex on $b_0$ to be the starting point, call it $\varrho_0$. Search the adjacent vertices and move to the vertex with a larger value of $u_{[1]}$. Continue to search the vertices adjacent to current vertex driven by the same criterion, i.e., always moving to the near vertex with a larger value of $u_{[1]}$. Once a vertex on $b_1$ is reached, then the cut is completed.

3. Create the oriented boundary $B$. Let $B$ start from the vertex $\varrho_0$ identified in the previous step. Then, let $B$ run from $\varrho_0$ around $b_0$ back to $\varrho_0$, and then up the cut $C$ and around $b_1$ and back down $C$ in the opposite direction back to $\varrho_0$ creating a closed curve. Notice that $B$ must run around $b_0$ and $b_1$ in a way that preserves the orientation of the surface. See Figure 2, top center. Moreover, the vertices along $C$ need to be repeated twice since they will end up on opposite sides of the rectangle.
4. Generate the boundary values for $u_{[2]}$ in (10) by integrating $u_{[1]}$ along $B$, as $u_{[2]}(\varrho) = \int_{\varrho_0}^{\varrho} \frac{\partial u_{[1]}}{\partial \nu} ds$, where $ds$ is the arc-length element along $B$. Since $u_{[1]}$ is harmonic, the divergence theorem yields $\oint_B \frac{\partial u_{[1]}}{\partial \nu} ds = 0$, where $\oint_B$ is the line integral over the closed boundary $B$. The cut $C$ follows the gradient of $u_{[1]}$, thus $\frac{\partial u_{[1]}}{\partial \nu} = 0$ along $C$. Hence, $u_{[2]}$ is constant along $C$. Note that the height of the cylinder must be scaled properly. The height of the cylinder becomes the width of the rectangle which is forced to have length equal to one. Hence the height of the rectangle will be the circumference of the cylinder divided by the height of the cylinder. If the proportions of the rectangle are not scaled properly, then the map will not be conformal.
5. Set up and solve the system for $u_{[2]}$ as in Step 1., adjusting the right-hand side of (23) to take into account the boundary values for $u_{[2]}$ in (10).

## 10. ASYMPTOTIC PROPERTIES OF THE ESTIMATORS

The following results quantify the bias of the estimators that is induced by the roughness penalty.

PROPOSITION 5. *Let $f$ be the true function in model (1), $X$ be the map in (3) and $\hat{f} \circ X$ be the estimator satisfying (8). Then the following inequality holds:*

$$\big|\big|E(\hat{f} \circ X) - f \circ X\big|\big|^2_{J_\Omega} \leq 4\lambda \big|\big|\mathcal{D}(u)^{-1/2} \text{div}\big[K(u)\nabla f\{X(u)\}\big]\big|\big|^2_{L^2(\Omega)}$$

*where $||\cdot||_{J_\Omega}$ denotes the norm on $H^2(\Omega)$ induced by the functional $J_{\Omega,\lambda}$ in (5), i.e.*

$$||h \circ X||^2_{J_\Omega} = \sum_{i=1}^{n} \big[h\{X(u_i)\}\big]^2 + \lambda \int_\Omega \Big(\mathcal{D}(u)^{-1/2} \text{div}\big[K(u)\nabla h\{X(u)\}\big]\Big)^2 d\Omega$$

*and $||\cdot||_{L^2(\Omega)}$ is the standard $L^2$ norm over $\Omega$.*

This result follows from Lemma 1 in Azzimonti et al. (2014a), which considers the simpler case of the estimators defined over planar domains, owing to the regularity of the map $X$. Exploiting the inverse map $X^{-1}$, it is then possible to derive the following analogous result directly for the estimator $\hat{f}$ defined over the manifold domain $\Gamma$.

COROLLARY 1. *Let $f$ be the true function in model (1) and let $\hat{f}$ be the minimizer of (2). Then the following inequality holds:*

$$\big|\big|E(\hat{f}) - f\big|\big|^2_{J_\Gamma} \leq 4\lambda \big|\big|\Delta_\Gamma f\big|\big|^2_{L^2(\Gamma)}$$

*where* $||\cdot||_{J_\Gamma}$ *is the norm on* $H^2(\Gamma)$ *induced by the functional* $J_{\Gamma,\lambda}$ *in (2), i.e.*

$$||h||^2_{J_\Gamma} = \sum_{i=1}^{n} \left\{h(x_i)\right\}^2 + \lambda \int_\Gamma \left\{\Delta_\Gamma f(x)\right\}^2 \, d\Gamma$$

*and* $||\cdot||_{L^2(\Gamma)}$ *is the* $L^2$ *norm over* $\Gamma$.

These results state that the estimators are asymptotically unbiased, in the norm induced by the estimation functional, either if the true unknown function $f$ is such that it annihilates the penalty term, or if $\lambda \to 0$ for $n \to +\infty$. Letting $\lambda$ decrease with $n$ appears natural, since having more observations decreases the need to impose a regularization.

Convergence to zero of the bias due to the discretization follows from Theorem 3 in Azzimonti et al. (2014b) for the estimators in the planar domain, combined with classical convergence results for a finite element discretization (Quarteroni, 2014) to guarantee the convergence of the approximated conformal map to the exact one.

A more complex issue, and still an open problem, concerns the convergence of the variance term. This topic is studied in the classical setting of smoothing splines (see, e.g., Cox, 1983; Heckman, 1986), thin-plate splines, or multidimensional smoothing splines (see, e.g., Cox, 1984; Györfi et al., 2002; Cucker & Zhou, 2007, and references therein), that show consistency of these estimators when the smoothing parameter $\lambda$ goes to zero with a proper rate with respect to $n$. Unfortunately these results cannot be directly extended to our setting and a different approach needs to be developed to show the consistency of our models. We are currently trying to prove this result, looking for the appropriate rate of decrease of $\lambda$ with respect to $n$ that makes the variance vanish, starting with the simpler planar setting considered in Sangalli et al. (2013); Azzimonti et al. (2014a,b). Finally, Section 11·2 reports a simulation study in support of the pointwise asymptotic normality of $\hat{f}$ and the asymptotic normality of $\hat{\beta}$.

## 11. SIMULATION STUDIES

### 11·1. *Comparison with a volume smoother*

We extend the comparative study in Section 5 to a method that is not constrained to the manifold domain, but works in the full volume embedding the manifold. In particular, we consider a tensor product of univariate cubic smoothing splines, with penalization of the $L^2$-norms of the second order derivatives. This method is implemented via the R package mgcv (Wood, 2006). The smoothing parameter is selected at each simulation replicate and for each domain by generalized cross validation.

With respect to Figure 5, Figure 6 includes the additional box plots of the mean square errors over the fifty simulation replicates for this technique, compared to the methods already considered in Section 5. Smoothing with respect to the three space coordinates, regardless of the domain of the data, leads to good results only for Geometry 2, that has large portions approximately parallel to the planes formed by the main axes. This method performs instead very poorly for the other geometries. With this tensor product technique it is also possible to use different smoothing parameters in the three dimensions, thus allowing for anisotropy in the direction of the three coordinate axes. This leads to an improvement of the estimates for Geometry 1, whose centerline is linear and parallel to one of the main axes, but does not yield any appreciable improvement for Geometry 3 and 4.

These results illustrate that a volume smoother does not efficiently handle data spatially distributed over bidimensional manifolds.

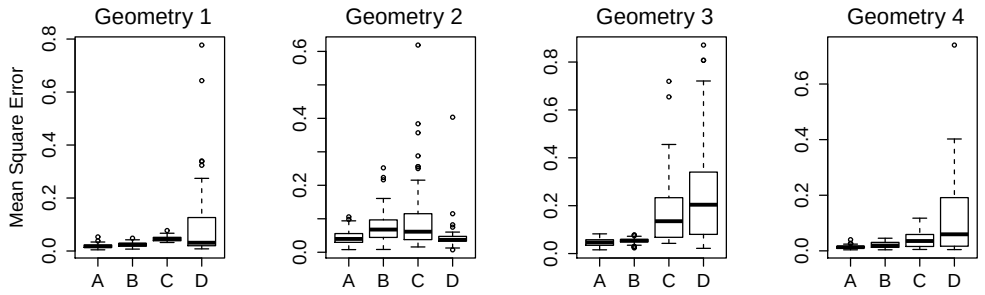


Fig. 6. Box plots of the mean square errors for the four estimators over fifty simulations. A: spatial regression over non-planar domains; B: spatial regression over planar domains, combined with cylindrical flattening; C: iterative heat kernel smoothing; D: tensor product of univariate smoothing splines.

### 11·2. *Simulation study with covariates*

We briefly show a simulation study where neither of the competitor models considered in Section 5, namely the planar spatial regression method with the cylindrical flattening and the Iterative Heat Kernel smoothing, can be implemented. The simulation uses the real inner carotid artery geometry in Figure 1 (b), also considered in Section 6, and includes space-varying covariates. As mentioned in the Introduction, for the carotid artery application, it is in fact of great interest to account for covariates concerning the local morphology of the artery. In this case, the planar spatial regression method with the cylindrical flattening cannot be implemented as the cylindrical map cannot handle the large aneurysmal sac, while Iterative Heat Kernel smoothing cannot be implemented as it is not currently designed to account for space-varying covariates.

We generate fifty test functions of the same form as (21) and consider two covariates; the local curvature $c$ of the vessel wall, as computed in Section 6, and another covariate $w$ that is randomly generated from a $N(0.5, 1)$ distribution. We create noisy data values at each data location $x_i$ via $z_i = 0.5w_i + 0.2c_i + f(x_i) + (1/3)\epsilon_i$ where the errors $\epsilon_i$ are independent with distribution t-student with five degrees of freedom. Across the fifty simulation repetitions, we obtain a median $\hat{\beta}_1 = 0.5020$ with an IQR of $0.0107$, and a median $\hat{\beta}_2 = 0.2004$ with an IQR of $0.0095$, thus illustrating that the proposed technique efficiently handles space-varying covariates. The distribution of $\hat{\beta}$ is well approximated by a Gaussian distribution, as verified by Shapiro-Wilk normality test (with p-value of 0.28 for the test on $\hat{\beta}$, and p-values of 0.30 and 0.76 for the tests on $\hat{\beta}_1$ and $\hat{\beta}_2$ respectively). The same can be said for the pointwise evaluations of $\hat{f}$ (if performing a Shapiro-Wilk normality test at each of the 4089 data locations $x_i$, then 94% of the p-values are larger than 0.05; all of the p-values are larger than 0.29 when corrected by false discovery rate). This simulation supports the asymptotic Gaussianity of the estimators, that can be used to construct approximate confidence intervals for $\beta$ and pointwise approximate confidence bands for $f$. For the first simulation replicate, we obtain for instance an approximate $95\%$ confidence interval of $(0.495, 0.521)$ for $\beta_1$ and an approximate $95\%$ confidence interval of $(0.195, 0.201)$ for $\beta_2$.

## REFERENCES

AZZIMONTI, L., NOBILE, F., SANGALLI, L. M. & SECCHI, P. (2014a). Mixed finite elements for spatial regression with PDE penalization. *SIAM/ASA J. Uncertain. Quantif.* **2**, 305–335.

AZZIMONTI, L., SANGALLI, L. M., SECCHI, P., DOMANIN, M. & NOBILE, F. (2014b). Blood flow velocity field estimation via spatial regression with PDE penalization. *J. Amer. Statist. Assoc.*

DOI:10.1080/01621459.2014.946036.

BRAESS, D. (2007). *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. New York, NY: Cambridge University Press, 2nd ed.

COX, D. D. (1983). Asymptotics for $M$-type smoothing splines. *Ann. Statist.* **11**, 530–551.

COX, D. D. (1984). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21**, 789–813.

CUCKER, F. & ZHOU, D.-X. (2007). *Learning theory: an approximation theory viewpoint*, vol. 24 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge. With a foreword by Stephen Smale.

GYÖRFI, L., KOHLER, M., KRZYŻAK, A. & WALK, H. (2002). *A distribution-free theory of nonparametric regression*. Springer Series in Statistics. Springer-Verlag, New York.

HECKMAN, N. E. (1986). Spline smoothing in a partly linear model. *J. Roy. Statist. Soc. Ser. B* **48**, 244–248.

LIONS, J.-L. & MAGENES, E. (1973). *Non-Homogeneous Boundary Value Problems and Applications*, vol. III. New York: Springer-Verlag.

PINKALL, U. & POLTHIER, K. (1993). Computing discrete minimal surfaces and their conjugates. *Experimental Mathematics* **2**, 15–36.

QUARTERONI, A. (2014). *Numerical models for differential problems*, vol. 8 of *Modeling, Simulation and Applications*. Springer, Milan, 2nd ed.

SANGALLI, L. M., RAMSAY, J. O. & RAMSAY, T. O. (2013). Spatial spline regression models. *J. R. Statist. Soc. Ser. B* **75**, 681–703.

WOOD, S. N. (2006). *Generalized additive models*. Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL. An introduction with ß$fR$.

[*Received April* 2012. *Revised September* 2012]