Statistical and Numerical techniques for Spatial Functional Data Analysis

Laura M. SANGALLI
MOX - Dipartimento di Matematica, Politecnico di Milano
Starting grant project
SNAPLE, research team

Starting grant of Ministero dell’Istruzione dell’Università e della Ricerca

“Advanced statistical and numerical methods for the analysis of high dimensional functional data in life sciences and engineering”

SNAPLE Statistical and Numerical methods for the Analysis of Problems in Life sciences and Engineering

Laura Sangalli
PI
Statistics

PostDoc
2y

Simona Perotto
Numerical Analysis

PostDoc
2y

John A. D. Aston
Statistics

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Starting grant of Ministero dell’Istruzione dell’Università e della Ricerca

“Advanced statistical and numerical methods for the analysis of high dimensional functional data in life sciences and engineering”

Laura Sangalli
PI
Statistics

Bree Ettinger
Applied Math
2y, NOW at Emory U.

Simona Perotto
Numerical Analysis

Laura Azzimonti
Applied Math, Statistics
1y, NOW at MOX-off

John A. D. Aston
Statistics

PostDoc
1y

PhD student
3y
(PostDoc 2y)
Collaborators

Piercesare Secchi
Statistics
MOX – Politecnico di Milano

Jim Ramsay
Statistics
McGill University, Canada

Fabio Nobile
Numerical Analysis
École Polytechnique Fédérale de Lausanne, Switzerland
Spatial regression with differential regularization

- Problem: *surface estimation* and *spatial field estimation* (spatial regression)

- We interface statistical methodology and numerical analysis techniques and propose

  spatial regression models with partial differential regularization

→ estimation problem solved via *Finite Elements*

- Can handle data distributed over irregular domains
- Can comply with general conditions at domain boundaries

→ Sangalli, Ramsay, Ramsay, 2013, JRSSB
Spatial regression with differential regularization

► Can incorporate *priori knowledge* about phenomenon under study allowing for very flexible modelling of space variation (anisotropy and non-stationarity)

→ Azzimonti, Sangalli, Secchi, Nobile, Domanin, 2013, TechRep
→ Azzimonti, Nobile, Sangalli, Secchi, 2013, TechRep

► Can deal with data over bi-dimensional Riemannian manifolds

→ Ettinger, Perotto, Sangalli, 2012, TechRep
→ Dassi, Ettinger, Perotto, Sangalli, 2013, TechRep

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Spatial regression with differential regularization

- Data: \( \Omega \subset \mathbb{R}^2 \): a region of interest, bounded, with \( \partial \Omega \in C^2 \)

  for \( i = 1, \ldots, n \)

  - \( \mathbf{p}_i = (x_i, y_i) \in \Omega \)
  - \( z_i \): a real valued variable of interest observed \( \mathbf{p}_i \)
  - \( \mathbf{w}_i = (w_{i1}, \ldots, w_{iq})^t \): a \( q \)-vector of covariates associated to \( z_i \)

Buoy data (National Oceanic and Atmospheric Administration)

Estimate of water surface temperature
Spatial regression with differential regularization

- Data: \( \Omega \subset \mathbb{R}^2 \): a region of interest, bounded, with \( \partial \Omega \in C^2 \)
  
  for \( i = 1, \ldots, n \)

  - \( \mathbf{p}_i = (x_i, y_i) \in \Omega \)

  - \( z_i \): a real valued variable of interest observed \( \mathbf{p}_i \)

  - \( \mathbf{w}_i = (w_{i1}, \ldots, w_{iq})^t \): a \( q \)-vector of covariates associated to \( z_i \)

- Generalized Additive Model:
  
  \[ z_i = \mathbf{w}_i^t \beta + f(\mathbf{p}_i) + \epsilon_i \quad i = 1, \ldots, n \]

  - \( \epsilon_i, i = 1, \ldots, n \), i.i.d. mean 0 and variance \( \sigma^2 \)

  - \( \beta \in \mathbb{R}^q \)

  - \( f : \Omega \rightarrow \mathbb{R} \)
Estimate $\beta$ and $f$ minimizing

$$J_{\lambda}(\beta, f) = \sum_{i=1}^{n} (z_i - w_i^t \beta - f(p_i))^2 + \lambda \int_\Omega (\Delta f)^2 d\Omega$$

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

Inclusion of simple Partial Differential Equations (PDE) in statistical models:

- Thin-plate splines (Wahba, 1990; Stone, 1988)
  $$\sum_{i=1}^{n} (z_i - f(p_i))^2 + \int_{\mathbb{R}^2} \left( \frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \frac{\partial^2 f}{\partial x \partial y} + \left( \frac{\partial^2 f}{\partial y^2} \right)^2$$

- Bivariate Splines (Guillas and Lai, 2010)

- FEL-splines (Ramsay, 2002), Soap-film smoothing (Wood et al., 2008): *irregular domains*

Irregularly shaped domains

Buoy data
(National Oceanic and Atmospheric Administration www.ndbc.noaa.gov)
→ Parnigoni Master thesis 2013

Fisheries data (NOAA)

Census Canada data

$Cov(Z(p_i), Z(p_j))$
stationarity, isotropy
few covariance models

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Boundary conditions

- Dirichlet
  \[ f|_{\partial \Omega} = g \]

- Neumann
  \[ \partial_{\nu} f|_{\partial \Omega} = g \]

- Robin (linear combination of the above)

- Mixed (different conditions in different parts of the boundary)

Fisheries data (NOAA)

Census Canada data

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Incorporating a priori information: using PDE to model space variation of the phenomenon

\[ J_\lambda(\beta, f) = \sum_{i=1}^{n} (z_i - w_i^T \beta - f(p_i))^2 + \lambda \int_{\Omega} (Lf - u)^2 d\Omega \]

Problem specific a priori information → (physics, mechanics, chemistry, morphology)

- **Diffusion tensor field**: non-stationary anisotropic diffusion
- **Transport vector field**: non-stationary directional smoothing
- **Reaction term**: non-stationary shrinking effect

PDEs are commonly used to describe complex phenomena behaviors in many fields of engineering and sciences. Model space variation spatially varying more complex partial differential operator (linear second order elliptic PDE)

\[ Lf = -\text{div}(K \nabla f) + b \cdot \nabla f + cf \]
A priori information

\[ J_\lambda(\beta, f) = \sum_{i=1}^{n} (z_i - w_i^T \beta - f(p_i))^2 + \lambda \int_{\Omega} (L f - u)^2 d\Omega \]

Buoy data

Prior (Gulf Stream)
Manifold domains

Object Oriented Data Analysis

\[ J_{\Gamma, \lambda}(\beta, f) = \sum_{i=1}^{n} \left( z_i - w_i^t \beta - f(x_i) \right)^2 + \lambda \int_{\Gamma} (\Delta_{\Gamma} f(x))^2 d\Gamma \]

Laplace-Beltrami operator associated to \( \Gamma \)

\( \Gamma \): surface embedded in \( \mathbb{R}^3 \)

\( x_i \in \Gamma \)

\( f : \Gamma \rightarrow \mathbb{R} \)

AneuRisk project http://mox.polimi.it/it/progetti/aneurisk/

http://mox.polimi.it/users/sangalli/firbSNAPLE.html

Ettinger et al. 2012, TechRep
Spatial regression with differential regularization

- Generalized Additive Model:

\[ z_i = w_i^t \beta + f(p_i) + \epsilon_i \quad i = 1, \ldots, n \]

\[ z = W \beta + f_n + \epsilon \]

\[ z := (z_1, \ldots, z_n)^t \]

\[ W := \begin{bmatrix} w_1^t \\ \vdots \\ w_n^t \end{bmatrix} \]

\[ H := W(W^tW)^{-1}W^t \quad Q := I - H \]

\[ f_n := (f(p_1), \ldots, f(p_n))^t, \text{ where } f \text{ is any function on } \Omega \]

- $H^m(\Omega)$: set of functions in $L^2(\Omega)$ having all weak derivatives up to order $m$ in $L^2(\Omega)$

- $H^m_{n0}(\Omega)$: subset of $H^m(\Omega)$ consisting of functions whose normal deriv are 0 on the boundary of $\Omega$ ($H^m(\Omega) + \text{Neumann b.c.}$)
Proposition

The estimators $\hat{\beta} \in \mathbb{R}^q$ and $\hat{f} \in H^2_{n0}(\Omega)$ exist unique

(\star) $\hat{\beta} = (W^t W)^{-1} W^t (z - \hat{f}_n)$

(\star\star) $\hat{f}$ satisfies

$$u_n^t Q \hat{f}_n + \lambda \int_{\Omega} (\Delta u)(\Delta \hat{f}) = u_n^t Q z \quad \text{for all } u \in H^2_{n0}(\Omega)$$

Weak formulation: find $(\hat{f}, g) \in (H^1(\Omega) \cap C^0(\Omega)) \times H^1(\Omega)$ such that

$$u_n^t Q \hat{f}_n - \lambda \int_{\Omega} (\nabla u \cdot \nabla g) = u_n^t Q z$$

$$\int_{\Omega} v g - \int_{\Omega} (\nabla v \cdot \nabla \hat{f}) = 0$$

for all $(u, v) \in (H^1(\Omega) \cap C^0(\Omega)) \times H^1(\Omega)$.

Thanks to the regularity of the problem, $\hat{f}$ belongs to $H^2_{n0}(\Omega)$. 

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Finite element analysis has been mainly developed and used in engineering applications, to solve partial differential equations.

Finite element space: space of continuous piecewise-polynomial surfaces over a triangulation $\mathcal{T}$ of $\Omega$
\{\xi_1, \ldots, \xi_K\}: \text{ nodes of } \mathcal{T}

\Omega_{\mathcal{T}}: \text{ triangulated domain; } H^{1/2}_{\mathcal{T}}(\Omega): \text{ finite element space}

\psi = (\psi_1, \ldots, \psi_K)^t: \text{ finite element basis } \quad \Psi = \{\psi\}_{ij} := \psi_j(p_i)

\text{for any } g \text{ in the finite element space, } g = g^t\psi \quad \text{where } g := (g(\xi_1), \ldots, g(\xi_K))^t

R_0 := \int_{\Omega_{\mathcal{T}}} (\psi\psi^t) \quad R_1 := \int_{\Omega_{\mathcal{T}}} (\psi_x\psi_x^t + \psi_y\psi_y^t)

**Corollary.** The estimators \(\hat{\beta} \in \mathbb{R}^q\) and \(\hat{f} \in H^{1/2}_{\mathcal{T}}(\Omega)\), that solve the discrete counterpart of the estimation problem, exist unique

\(\hat{\beta} = (W^tW)^{-1}W^t(z - \hat{f}_n)\)

\(\hat{f} = \hat{f}^t\psi\), with \(\hat{f}\) satisfying

\[
\begin{bmatrix}
-\Psi^tQ\Psi & \lambda R_1 \\
\lambda R_1 & \lambda R_0
\end{bmatrix}
\begin{bmatrix}
f \\
g
\end{bmatrix}
= 
\begin{bmatrix}
-\Psi^tQz \\
0
\end{bmatrix}
\]
Spatial regression with differential regularization

\[ \hat{\beta} \text{ and } \hat{f} \text{ are linear in } z \rightarrow \text{linear estimators} \]

\[ \hat{f} \text{ has typical penalized regression form, being identified by} \]
\[ \hat{f}_n = (\Psi^t Q \Psi + \lambda P)^{-1} \Psi^t Q z \]

\[ \text{Classical inferential tools are readily derived} \]

\[ \triangleright \text{mean and variances of } \hat{\beta} \text{ and } \hat{f} \]
\[ \triangleright \text{confidence intervals for } \beta \]
\[ \triangleright \text{confidence bands for } f \]
\[ \triangleright \text{prediction intervals for new observations} \]
\[ \triangleright \text{estimate of error variance } \sigma^2 \]
\[ J_\lambda(\beta, f) = \sum_{i=1}^{n} (z_i - w_i^t \beta - f(p_i))^2 + \lambda \int_{\Omega} (\Delta f)^2 d\Omega \]

\[ \triangleright \text{selection of smoothing parameter } \lambda \text{ via generalized cross-validation} \]
Two sources of bias:

- \( \hat{f} \in H^1_f(\Omega) \) is affected by bias due to discretization:
  This bias disappears as \( n \to \infty \) with \( h \to 0 \)

- \( \hat{f} \in H^2_{n_0}(\Omega) \) and \( \hat{f} \in H^1_f(\Omega) \) are affected by bias due to regularization
  This bias disappears as \( n \to \infty \) with \( \lambda \to 0 \)
Simulation studies (Sangalli et al. 2013) show that the proposed models outperform thin-plate splines, filtered kriging and other state-of-the-art methods for spatially distributed data.
Simulation studies (Sangalli et al. 2013) show that the proposed models outperforms thin-plate splines, filtered kriging and other state-of-the-art methods for spatially distributed data.
Simulation studies (Sangalli et al. 2013) show that the proposed models outperform thin-plate splines, filtered kriging and other state-of-the-art methods for spatially distributed data.

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Illustrative example: Island of Montreal census data

- $p_i$: centroids of census tracts
- $z_i$: population density
  (1000 inhabitants per $km^2$)
- $w_i$: indicator of residential (1) commercial (0) census tract

- Mixed b.c. (homogeneous Neumann and homogeneous Dirichlet)
Illustrative example: Island of Montreal census data

\[
\hat{\beta} = 1.300
\]

approx. 95% confidence interval \([0.755; 1.845]\)
Motivating applied problem: AneuRisk project

Spatial regression over bi-dimensional Riemannian manifold domains

http://mox.polimi.it/it/progetti/aneurisk
Spatial regression over Riemannian manifolds

Heamodynamic simulations
Passerini et al. 2012

Object Oriented Data Analysis

- Nearest Neighbor Averaging
  Hagler, Saygin, Sereno, 2006, *NeuroImage*

- Heat Kernel Smoothing
  Chung et al., 2005, *NeuroImage*

- Methods for data over spheres, hyperspheres and other manifolds
  Lindgren, Rue, Lindstrom, 2011, *JRSSB*
  Gneiting, 2013, *Bernoulli*

\[ \Gamma \subset \mathbb{R}^3 \] - a non-planar surface domain
Artery wall

\[ \{ x_i = (x_{1i}, x_{2i}, x_{3i}) \in \Gamma \} \] - data locations

\[ z_i \in \mathbb{R} \] - variable of interest observed at \( x_i \)
Wall shear stress modulus at systolic peak

\[ w_i = (w_{i1}, \ldots, w_{iq}) \in \mathbb{R}^q \] - space varying covariates
Local curvature of vessel wall
Curvature of vessel
Local radius of the vessel

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Spatial regression over Riemannian manifolds

\[ z_i = w_i^{t} \beta + f(x_i) + \epsilon_i \quad i = 1, \ldots, n \]

\[ J_{\Gamma, \lambda}(\beta, f) = \sum_{i=1}^{n} (z_i - w_i^{t} \beta - f(x_i))^2 + \lambda \int_{\Gamma} (\Delta_{\Gamma} f(x))^2 dx \]
Spatial regression over Riemannian manifolds

- Estimators have typical penalized regression form
- Linear in observed data values
- Classical inferential tools

Conformal mapping
(fully encode information about complex 3D geometry)

Equivalent estimation problem on planar domain
→ Extend method for planar domains

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Spatial regression over Riemannian manifolds

\[ X : \Omega \rightarrow \Gamma \]

\[ \mathbf{u} = (u, v) \mapsto \mathbf{x} = (x_1, x_2, x_3) \]

(\(\Omega\) : open, convex, bounded set in \(\mathbb{R}^2\))

\(\Gamma\)

\(\mathbf{u} = (u, v)\)

\(\mathbf{x} = (x_1, x_2, x_3)\)

\(X^{-1}\) (Flattening map)
Spatial regression over Riemannian manifolds

\[ \frac{\partial X}{\partial u} (u), \frac{\partial X}{\partial v} (u): \text{column vectors} \]

\[ G(u) := \nabla X(u)'\nabla X(u) = \begin{pmatrix} \| \frac{\partial X}{\partial u} (u) \|^2 & \langle \frac{\partial X}{\partial u} (u), \frac{\partial X}{\partial v} (u) \rangle \\ \langle \frac{\partial X}{\partial u} (u), \frac{\partial X}{\partial v} (u) \rangle & \| \frac{\partial X}{\partial v} (u) \|^2 \end{pmatrix} \]

\[ W(u) := \sqrt{\det(G(u))}; \quad W(u) \, du = dx \]

\[ K(u) = W(u) G^{-1}(u) \]

For \( f \circ X \in C^2(\Omega) \),

\[ \nabla_{\Gamma} f(x) = \nabla X(u) G^{-1}(u)(\nabla f(X(u))) \]

\[ \Delta_{\Gamma} f(x) = \text{div}_{\Gamma}(\nabla_{\Gamma} f(X(u))) = \frac{1}{W(u)} \text{div}(K(u) \nabla f(X(u))) \]

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Spatial regression over Riemannian manifolds

- \( H_{n,0,K}^m(\Omega) = \{ h \in H^m(\Omega) : K\nabla h \cdot n = 0 \text{ on } \partial \Omega \} \subset H^m(\Omega) \)

**Equivalent estimation problem over the planar domain \( \Omega \)**

Find \( \beta \in \mathbb{R}^q \) and \( f \) with \( (f \circ X) \in H_{n,0,K}^2(\Omega) \) that minimizes

\[
J_{\Omega,\lambda}(\beta, f \circ X) = \sum_{i=1}^{n} (z_i - w_i'\beta - f(X(u_i)))^2 + \lambda \int_{\Omega} \frac{1}{W} \left( \text{div}(K\nabla (f \circ X)) \right)^2 d\Omega
\]

where \( u_i = X^{-1}(x_i) \)

For conformal maps, i.e. \( \| \frac{\partial X}{\partial u} (u) \|^2 = \| \frac{\partial X}{\partial v} (u) \|^2 \) and \( \langle \frac{\partial X}{\partial u} (u), \frac{\partial X}{\partial v} (u) \rangle = 0 \forall u \in \Omega \),

\[
J_{\Omega,\lambda}(\beta, f \circ X) = \sum_{i=1}^{n} (z_i - w_i'\beta - f(X(u_i)))^2 + \lambda \int_{\Omega} \left( \frac{1}{\sqrt{W(u)}} \Delta f(X(u)) \right)^2 d\Omega
\]
Spatial regression over Riemannian manifolds

\( H^m_{n_0,K}(\Omega) = \{ h \in H^m(\Omega) : K\nabla h \cdot n = 0 \text{ on } \partial\Omega \} \subset H^m(\Omega) \)

\( \Rightarrow \) Equivalent estimation problem over the planar domain \( \Omega \)

Find \( \beta \in \mathbb{R}^q \) and \( f \) with \( (f \circ X) \in H^2_{n_0,K}(\Omega) \) that minimizes

\[
J_{\Omega,\lambda}(\beta, f \circ X) = \sum_{i=1}^{n} (z_i - w'_i\beta - f(X(u_i)))^2 + \lambda \int_{\Omega} \frac{1}{W} \left( \text{div}(K\nabla(f \circ X)) \right)^2 d\Omega
\]

where \( u_i = X^{-1}(x_i) \)

\( \Rightarrow \) Extend method for planar domains
Proposition. The estimators $\hat{\beta} \in \mathbb{R}^q$ and $\hat{f} \in H^2_{n0,K}(\Omega)$ exist unique

\[(\star) \quad \hat{\beta} = (W^t W)^{-1} W^t (z - \hat{f}_n)\]

\[(\star\star) \quad \hat{f} \text{ satisfies } \mu^t_n Q\hat{f}_n + \lambda \int_\Omega \frac{1}{W} \left(\text{div}(K \nabla (\mu \circ X))\right)\left(\text{div}(K \nabla (\hat{f} \circ X))\right) d\Omega = \mu^t_n Qz\]

for any $\mu$ defined on $\Gamma$ such that $\mu \circ X \in H^2_{n0,K}(\Omega)$.

Weak formulation:

Find $(\hat{f} \circ X, \gamma \circ X) \in (H^1_{n0,K}(\Omega) \cap C^0(\bar{\Omega})) \times H^1(\Omega)$ such that

\[
\mu^t_n Q\hat{f}_n - \lambda \int_\Omega K \nabla (\mu \circ X) \cdot \nabla (\gamma \circ X) d\Omega = \mu^t_n Qz
\]

\[
\int_\Omega (\xi \circ X)(\gamma \circ X) W d\Omega + \int_\Omega \nabla (\xi \circ X) K \nabla (\hat{f} \circ X) d\Omega = 0
\]

for any $(\mu \circ X, \xi \circ X) \in (H^1_{n0,K}(\Omega) \cap C^0(\bar{\Omega})) \times H^1(\Omega)$.

Thanks to the regularity of the problem, $\hat{f} \circ X$ still belongs to $H^2_{n0,K}(\Omega)$.
Conformal parametrization

\[ \begin{align*}
-\Delta_{\Sigma} u &= 0 \text{ on } \Sigma \\
u &= 0 \text{ on } \sigma_0 \\
u &= 1 \text{ on } \sigma_1
\end{align*} \]

\[ E_D(u) = \frac{1}{2} \int_{\Gamma} \| \nabla_{\Gamma} u \|^2 d\Gamma \]

\[ \begin{align*}
-\Delta_{\Sigma} v &= 0 \text{ on } \Sigma \\
v(\zeta) &= \int_{\zeta_0}^{\zeta} \frac{\partial u}{\partial \nu} \, ds \text{ on } B
\end{align*} \]

\[ E_D(v) = \frac{1}{2} \int_{\Gamma} \| \nabla_{\Gamma} v \|^2 d\Gamma \]


http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Spatial regression over Riemannian manifolds

Conformal parametrization

Spatial regression over Riemannian manifolds

- \{\xi_1, ..., \xi_K\}: nodes of planar triangulation \( \mathcal{T} \)

- \( \Omega_{\mathcal{T}} \): planar triangulated domain; \( H^1_{\mathcal{T}}(\Omega) \): finite element space

- \( \psi = (\psi_1, \ldots, \psi_K)^t \): finite element basis \( \Psi = \{\Psi\}_{ij} := \psi_j(p_i) \)

- for any \( h \) in the finite element space, \( h = h^t \psi \) where \( h := (h(\xi_1), \ldots, h(\xi_K))^t \)

- \( R_0 := \int_{\Omega_{\mathcal{T}}} (\psi \psi^t) \mathcal{W} \quad R_1 := \int_{\Omega_{\mathcal{T}}} \nabla \psi' K \nabla \psi \)

**Corollary.** The estimators \( \hat{\beta} \in \mathbb{R}^q \) and \( \hat{f} \in H^1_{\mathcal{T}}(\Omega) \), that solve the discrete counterpart of the estimation problem, exist unique

- \( \hat{\beta} = (W^t W)^{-1} W^t (z - \hat{f}_n) \)

- \( \hat{f} = \hat{f}^t \psi \), with \( \hat{f} \) satisfying

\[
\begin{bmatrix}
-\Psi^t Q \Psi & \lambda R_1 \\
\lambda R_1 & \lambda R_0
\end{bmatrix}
\begin{bmatrix}
f \\
g
\end{bmatrix}
= \begin{bmatrix}
-\Psi^t Q z \\
0
\end{bmatrix}
\]
Simulation (without covariates)

TRUE

TRUE + NOISE

ESTIMATE

RMSE

50 simulation replicates

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Spatial regression over Riemannian manifolds

Covariates:
- Local curvature of vessel wall → Negative association
- Curvature of vessel → Positive association
- Local radius of vessel → Negative association
Future work

Variability across patients

(data registration)
Facing big data challenges:

- iterative algorithms
- mesh simplification algorithms
Spatial Regression models over Riemannian manifolds

\[ c(e, v^*) := \alpha c_{\text{geo}}(e, v^*) + (1 - \alpha) c_{\text{data}}(e, v^*), \quad 0 \leq \alpha \leq 1 \]
Spatial regression over Riemannian manifolds

\[ c_{\text{equi}}(e, v^*) := \frac{1}{\#(T_{\text{cont}})} \left( \sum_{T \in T_{\text{cont}}} (N_T - \bar{N})^2 \right) \]

\[ N_T := n_{\text{faces}} + \frac{1}{2} n_{\text{edges}} + \frac{1}{\#(T_{v_1})} n_{v_1} + \frac{1}{\#(T_{v_2})} n_{v_2} + \frac{1}{\#(T_{v_3})} n_{v_3} \]
Spatial regression over Riemannian manifolds

TRUE

TRUE + NOISE

ESTIMATE

http://mox.polimi.it/users/sangalli/firbSNAPLE.html

![Brain images showing true, true + noise, and estimate](image)
Motivating applied problem: MACAREN@MOX

MACAREN@MOX Project: MAtematics for CARotid ENdarterectomy @ MOX

Aim: Study the pathogenesis of atherosclerotic plaques

Statistics
Computer science
Numerical analysis
Vascular surgery

Magnetic Resonance Imaging (MRI)
Vessel morphology

Echo-Color Doppler (ECD)
Blood fluid-dynamics

Statistics
Computer science
Numerical analysis
Vascular surgery

External carotid artery
Internal carotid artery
Common carotid artery

Magnetic Resonance Imaging (MRI)
Vessel morphology

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Motivating applied problem: MACAREN@MOX

Longitudinal section of the carotid
Motivating applied problem: MACAREN@MOX

Elliptic beam

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Motivating applied problem: MACAREN@MOX
Motivating applied problem: MACAREN@MOX

Systolic peak
Motivating applied problem: MACAREN@MOX

- Systolic peak
- Mean velocity

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
FIRST GOAL:
estimate blood-flow velocity field over the carotid section

For each patient, 7 ECD measurements over the carotid section located 2 cm before the carotid bifurcation

Mean measurements over 7 beams (systolic peak)
Spatial Spline Regression

\[ J(f) = \sum_{i=1}^{n} (f(p_i) - z_i)^2 + \lambda \int_{\Omega} (\Delta f)^2 \]

Boundary conditions

- Dirichlet
  \[ f|_{\partial \Omega} = 0 \]

Mean measurements over 7 beams (systolic peak)

**Physiological boundary conditions:** velocity=0 near the arterial wall
Motivating applied problem: MACAREN@MOX

Spatial Spline Regression

\[ J(f) = \sum_{i=1}^{n} (f(p_i) - z_i)^2 + \lambda \int (\Delta f)^2 \]

Mean measurements over 7 beams (systolic peak)

- **Non-physiological velocity field:** Squared isolines caused by cross-shaped pattern of observations

- **Prior information:** Theoretical solution for velocity field in perfectly straight pipe without turbulence has parabolic profile

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Spatial regression models with differential regularization

\[ \bar{z}_i = \frac{1}{|D_i|} \int_{D_i} f_0 + \eta_i \]

Areal data
(subdomain \( D_i \): i-th beam)

\[ \bar{J}(f) = \sum_{i=1}^{N} \frac{1}{|D_i|} \left( \int_{D_i} (f - \bar{z}_i) \right)^2 + \lambda \int_{\Omega} (Lf - u)^2 \]

→ Weighted least-square-error term for areal mean over subdomains \( D_i \)
→ Roughness term penalizing misfit with respect to more complex PDE known to model to some extent the phenomenon under study

general second order elliptic operator

\[ Lf = -div(K \nabla f) + b \cdot \nabla s + cs \]

forcing term

\[ u \in L^2(\Omega) \]

The parameters can be space-varying
PRIOR information described by a partial differential model

\[ L f = -\text{div}(K \nabla f) + \mathbf{b} \cdot \nabla s + c s \]

Diffusion tensor field: *anisotropic non-stationary diffusion* that smooths the observations along concentric circles

Forcing term:
\[ u = 0 \]

Reaction term: *shrinking effect*
\[ c = 0 \]

Transport vector field: *directional smoothing* that smooths the observations along the radial direction
Spatial regression models with differential regularization

\[ \mathbf{z} = (z_1, \ldots, z_n)^t \]

\[ \{\xi_1, \ldots, \xi_K\}: \text{nodes of } T \]

\[ \mathbf{\psi} = (\psi_1, \ldots, \psi_K)^t: \text{finite element basis} \]

\[ \Psi = \{\Psi\}_{ij} := \psi_j(p_i) \]

\[ \text{for any } g \text{ in the finite element space, } g = g^t \mathbf{\psi} \text{ where } g := (g(\xi_1), \ldots, g(\xi_K))^t \]

\[ R = \{R\}_{jk} := \int_{\Omega_T} (\psi_j \psi_k)^t \quad A = \{A\}_{jk} := \int_{\Omega_T} (K \nabla \psi_j \cdot \nabla \psi_k + b \cdot \nabla \psi_j \psi_k + c \psi_j \psi_k) \]

**Corollary.** The finite element estimator \( \hat{f} \) that solves the discrete counterpart of the estimation problem, exist unique and is given by \( \hat{f} = f^t \mathbf{\psi} \) where \( f \) satisfies

\[
\begin{bmatrix}
-\Psi^t \Psi & \lambda A \\
\lambda A & \lambda R
\end{bmatrix}
\begin{bmatrix}
f \\
g
\end{bmatrix}
= \begin{bmatrix}
-\Psi^t \mathbf{z} \\
0
\end{bmatrix}
\]

(Here for simplicity: pointwise case, \( u \equiv 0 \), homogeneous Neumann b.c.)
\( \hat{f} \) is linear in \( z \) and has typical penalized regression form:

\[
f_n = (\Psi^t \Psi + \lambda P)^{-1} \Psi^t z
\]

\( P = A^t R^{-1} A \) is discretization of penalty

- Classical inferential tools are readily derived
  - mean and variance of \( \hat{f} \)
  - confidence bands for \( f \)
  - prediction intervals for new observations
  - estimate of error variance \( \sigma^2 \)
  - selection of smoothing parameter \( \lambda \) via generalized cross validation

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Spatial regression models with differential regularization

\[ \int_\Omega (Lf - g)^2 \]

Physiological velocity field
Asymmetry due to curvature of carotid artery and to carotid bifurcation

Relevant features: eccentricity, reversion of fluxes

Patient-specific inflow conditions for Computation Fluid-Dynamics

Nobile, Pozzoli, Vergara

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Main references on spatial regression with PDE regularization

Irregularly shaped domains and boundary conditions

Incorporating a priori knowledge

Manifold domains
- Gneiting, T. (2013), Strictly and non-strictly positive definite functions on spheres, Bernoulli, 19, 4, 1087-1500

http://mox.polimi.it/users/sangalli/firbSNAPLE.html
Acknowledgments:

**FIRB2008 Futuro in Ricerca** 2010 - 2013 Starting grant from Ministero dell’Istruzione dell’Università e della Ricerca “SNAPLE: Statistical and Numerical methods for the Analysis of Problems in Life sciences and Engineering”. http://mox.polimi.it/users/sangalli/firbSNAPLE.html


**McGill University**: Visiting research grant (Apr.- July 2010), Department of Mathematics and Statistics, McGill University, Montreal, Canada.

**SAMSI**: Visiting research grant (Jan.- Feb. 2011), SAMSI Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, North Carolina, US

http://mox.polimi.it/users/sangalli