Dote Ricercatore
RegioneLombardia - POLIMI

FIRB 2008
FUTURO IN RICERCA
SNAPLE

POLITECNICO DI MILANO

January 2014, Geilo, Norway

14th Winter School in eScience SINTEF

Big Data Challenges to Modern Statistics

Statistical and Numerical techniques for Spatial Functional Data Analysis

Laura M. SANGALLI

MOX - Dipartimento di Matematica, Politecnico di Milano

FIRB

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 **S**tatistical and **N**umerical methods for the **A**nalysis of **P**roblems in **L**ife sciences and **E**ngineering

Laura Sangalli

PI

Statistics

PostDoc

2y

Simona Perotto

Numerical Analysis

PostDoc

2y

John A. D. Aston

Statistics

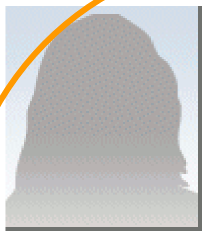
FIRB FUTURO IN RICERCA

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.



PostDoc
1y

Simona Perotto
Numerical Analysis

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

PhD student
3y
(PostDoc 2y)

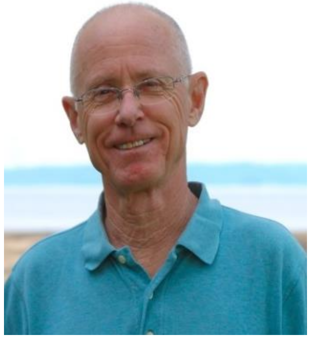
John A. D. Aston
Statistics

CRiSM THE UNIVERSITY OF WARWICK

Collaborators

Piercesare Secchi
Statistics
MOX – Politecnico di Milano

Fabio Nobile
Numerical Analysis
École Polytechnique Fédérale de Lausanne, Switzerland



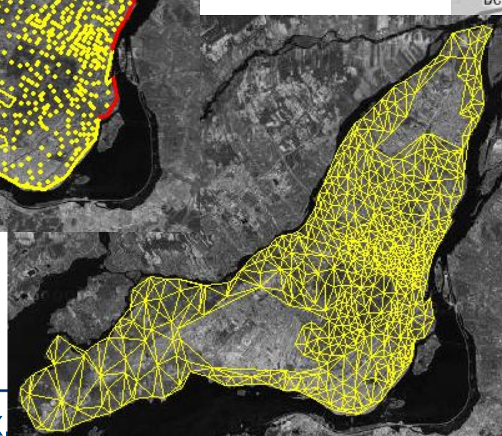
Jim Ramsay
Statistics
McGill University, Canada

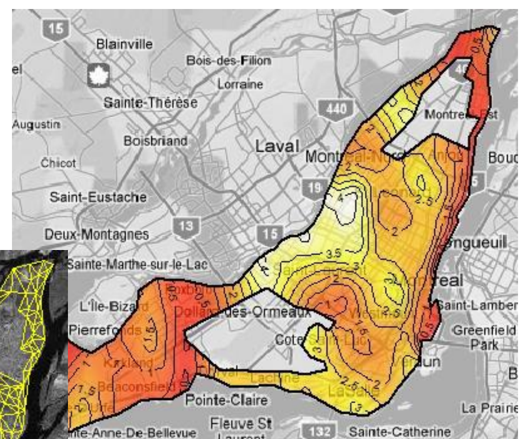
- Problem: *surface estimation* and *spatial field estimation* (spatial regression)

- 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 piori knowledge about phenomenon under study allowing for very flexible modelling of space variation (anysotropy and non-stationarity)

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

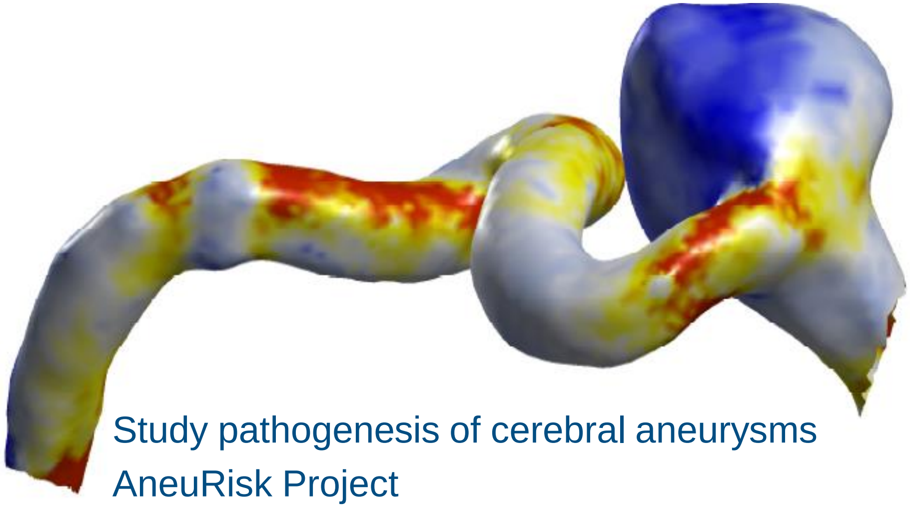
⟶ Azzimonti, Nobile, Sangalli, Secchi, 2013, TechRep

Study pathogenesis of atherosclerotic plaques
MAthematichs for CARotid ENdarterectomy @ MOX

► Can deal with data over bi-dimensional Riemannian manifolds

⟶ Ettinger, Perotto, Sangalli, 2012, TechRep

⟶ Dassi, Ettinger, Perotto, Sangalli, 2013, TechRep



Study pathogenesis of cerebral aneurysms
AneuRisk Project

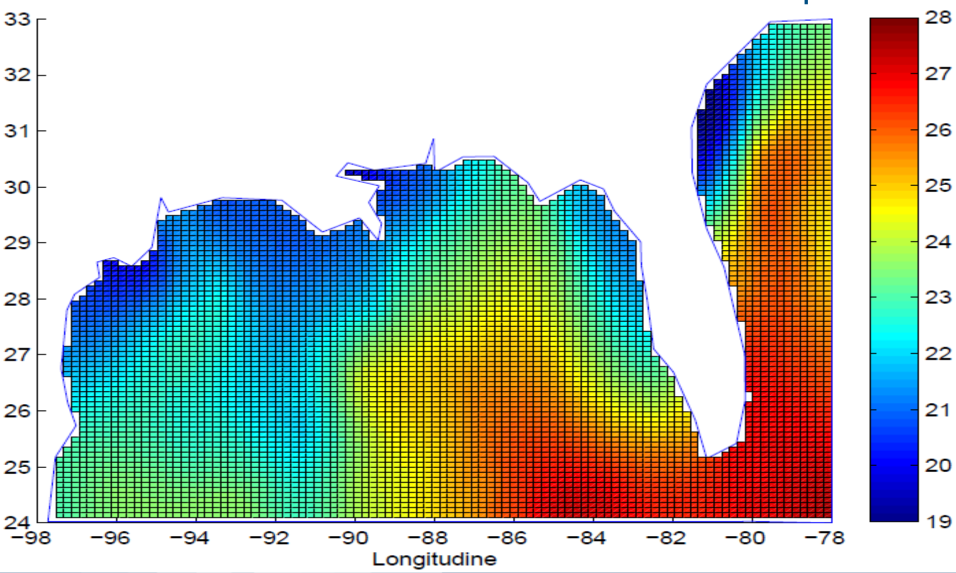
- Data: $\Omega \subset \mathbb{R}^2$: a region of interest, bounded, with $\partial\Omega \in \mathcal{C}^2$

  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



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

  for $i = 1, \dots, n$

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

- Generalized Additive Model:

$$z_i = \mathbf{w}_i^t \boldsymbol{\beta} + f(\mathbf{p}_i) + \epsilon_i \qquad i = 1, \dots, n$$

  - $\epsilon_i, i = 1, \dots, n$, i.i.d. mean 0 and variance σ^2
  - $\boldsymbol{\beta} \in \mathbb{R}^q$
  - $f : \Omega \to \mathbb{R}$

Sangalli et al. 2013, JRSSB

- Estimate $\boldsymbol{\beta}$ and f minimizing

$$J_\lambda(\boldsymbol{\beta}, f) = \sum_{i=1}^{n} \left(z_i - \mathbf{w}_i^t \boldsymbol{\beta} - f(\mathbf{p}_i)\right)^2 + \lambda \int_\Omega (\Delta f)^2 d\Omega$$

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

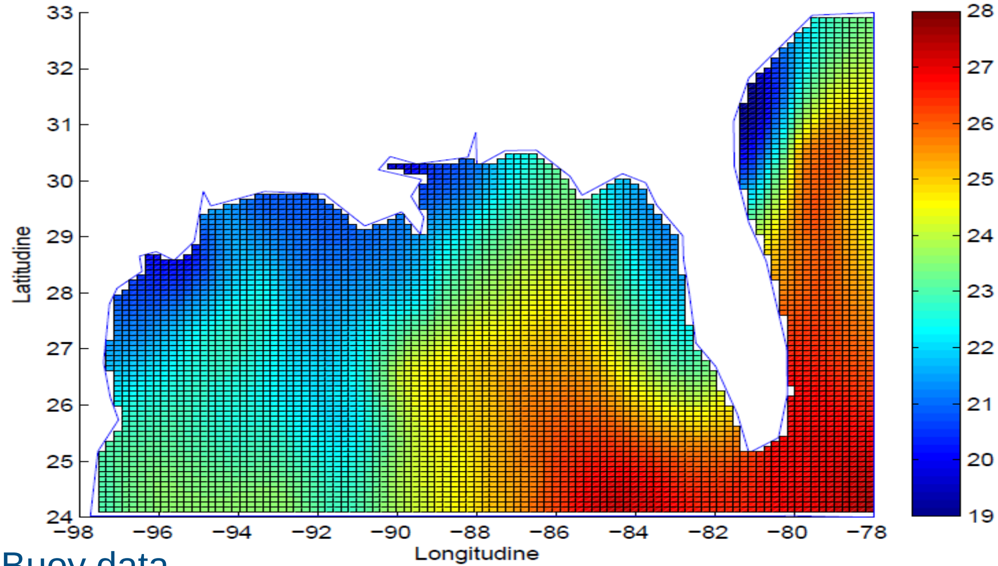
Inclusion of simple Partial Differential Equations (PDE) in statistical models:

- Thin-plate splines (Wahba,1990; Stone, 1988)

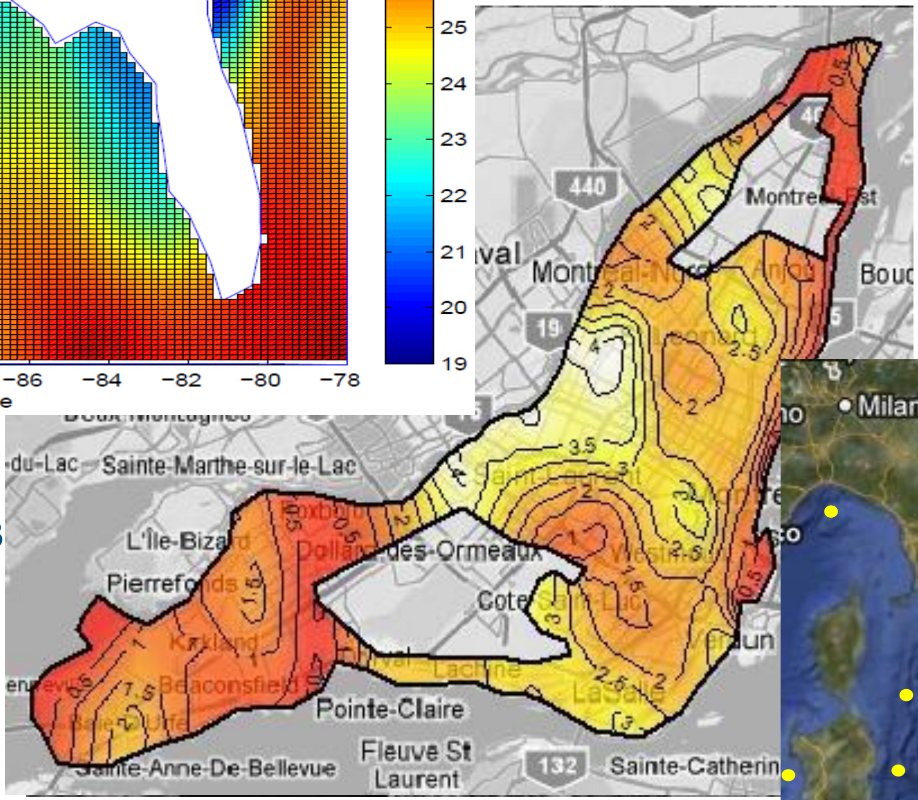
$$\sum_{i=1}^{n} \left(z_i - f(\mathbf{p}_i)\right)^2 + \int_{\mathbb{R}^2} \left(\frac{\partial^2 f}{\partial x^2}\right)^2 + 2\,\frac{\partial^2 f}{\partial x \partial y} + \left(\frac{\partial^2 f}{\partial y^2}\right)^2$$

- Bivariate Splines (Guillas and Lai, 2010)
- FEL-splines (Ramsay, 2002), Soap-film smoothing (Wood *et al.*, 2008): *irregular domains*
- Stochastic PDE: Lindgren *et al.* (2011), Bayesian inverse problems: Stuart (2010).

Irregularly shaped domains



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



Census Canada data

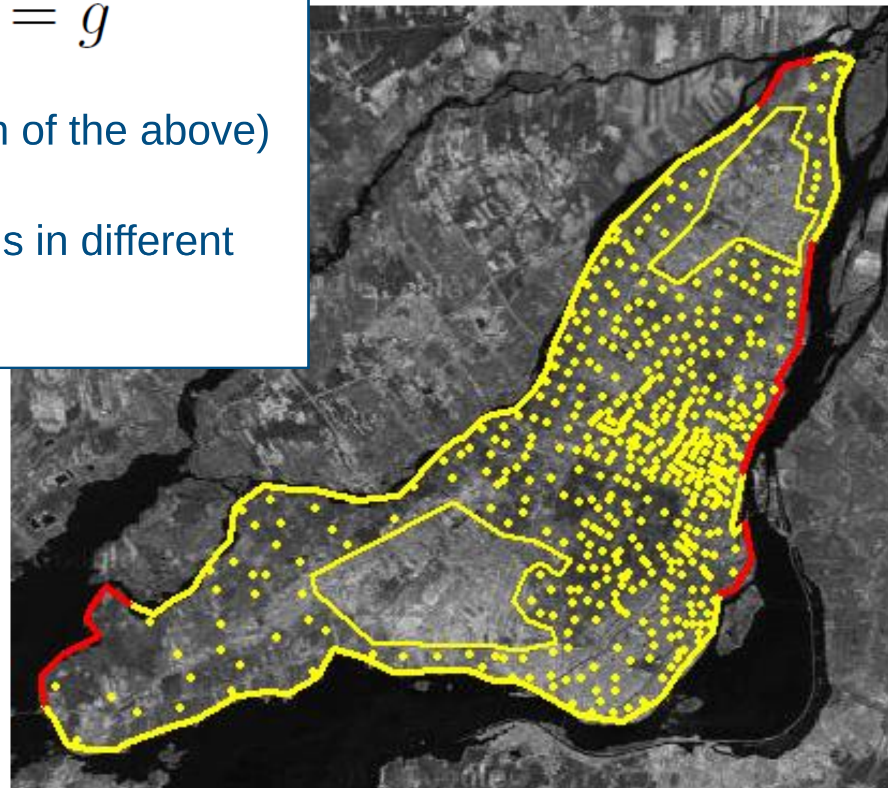
Fisheries data (NOAA)

$$Cov(Z(\mathbf{p}_i), Z(\mathbf{p}_j))$$

stationarity, isotropy
few covariance models

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)



Census Canada data

Fisheries data (NOAA)

A priori information

Azzimonti et al., 2013a, TechRep
Azzimonti et al., 2013b, TechRep

Incorporating a priori information:
using PDE to model space variation
of the phenomenon

log-likelihood
data fidelity

prior
model fidelity

$$J_\lambda(\boldsymbol{\beta}, f) = \sum_{i=1}^{n} \left(z_i - \mathbf{w}_i^t \boldsymbol{\beta} - f(\mathbf{p}_i)\right)^2 + \lambda \int_\Omega (Lf - u)^2 d\Omega$$

more complex partial differential operator
(linear second order elliptic PDE)

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

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

spatially varying

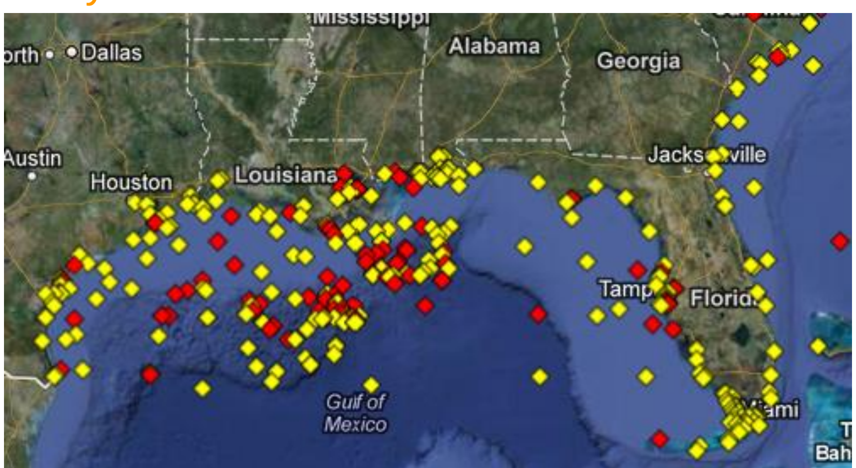
- ► PDEs are commonly used to describe *complex phenomena behaviors* in many fields of engineering and sciences
- ► model space variation

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

A priori information

$$J_\lambda(\boldsymbol{\beta}, f) = \sum_{i=1}^{n} \left(z_i - \mathbf{w}_i^t \boldsymbol{\beta} - f(\mathbf{p}_i)\right)^2 + \lambda \int_\Omega (Lf - u)^2 d\Omega$$

Buoy data



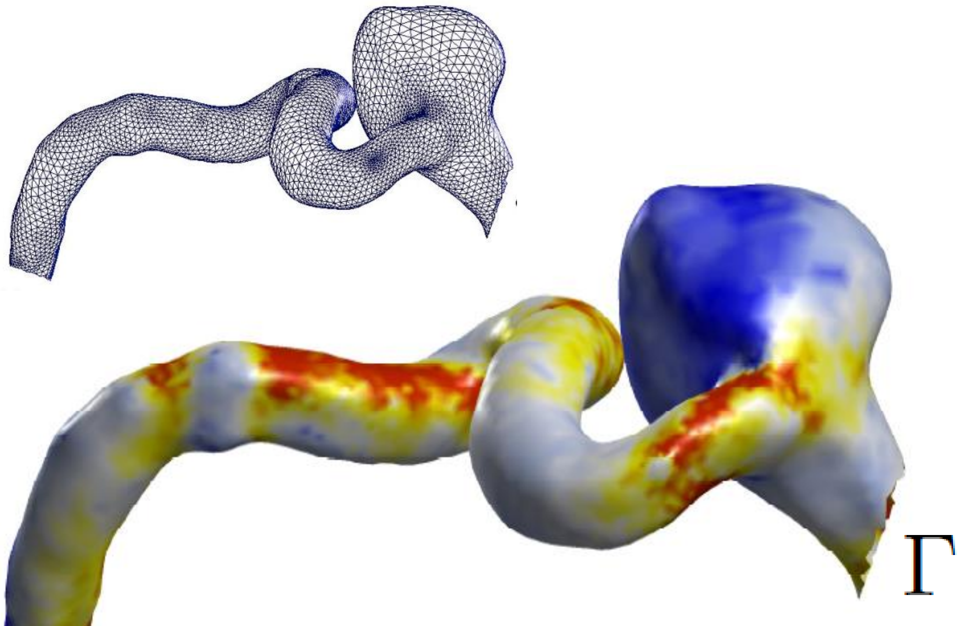
Prior (Gulf Stream)

Ettinger et al. 2012, TechRep

Object Oriented Data Analysis

$$J_{\Gamma,\lambda}(\boldsymbol{\beta}, f) = \sum_{i=1}^{n} \left(z_i - \mathbf{w}_i^t \boldsymbol{\beta} - f(\mathbf{x}_i)\right)^2 + \lambda \int_{\Gamma} (\Delta_\Gamma f(\mathbf{x}))^2 d\mathbf{x}$$

Laplace-Beltrami operator associated to Γ



Γ: surface embedded in $\mathbb{R}^3$

$\mathbf{x}_i \in \Gamma$

$f : \Gamma \rightarrow \mathbb{R}$

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

- Generalized Additive Model:

$$z_i = \mathbf{w}_i^t \boldsymbol{\beta} + f(\mathbf{p}_i) + \epsilon_i \qquad\qquad i = 1, \dots, n$$

$$\mathbf{z} = W\,\boldsymbol{\beta} + \mathbf{f}_n + \boldsymbol{\epsilon}$$

▷ $\mathbf{z} := (z_1, \dots, z_n)^t$

▷ $W := \begin{bmatrix} \mathbf{w}_1^t \\ \vdots \\ \mathbf{w}_n^t \end{bmatrix} \qquad H := W(W^tW)^{-1}W^t \qquad Q := I - H$

▷ $\mathbf{f}_n := (f(\boldsymbol{p}_1), \dots, f(\boldsymbol{p}_n))^t$, where f is any function on Ω

▷ $H^m(\Omega)$: 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 Ω ($H^m(\Omega)$ + Neumann b.c.)

$$J_\lambda(\boldsymbol{\beta}, f) = \sum_{i=1}^{n} \left(z_i - \mathbf{w}_i^t \boldsymbol{\beta} - f(\mathbf{p}_i)\right)^2 + \lambda \int_\Omega (\Delta f)^2 d\Omega$$

Proposition

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

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

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

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

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

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

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

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

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

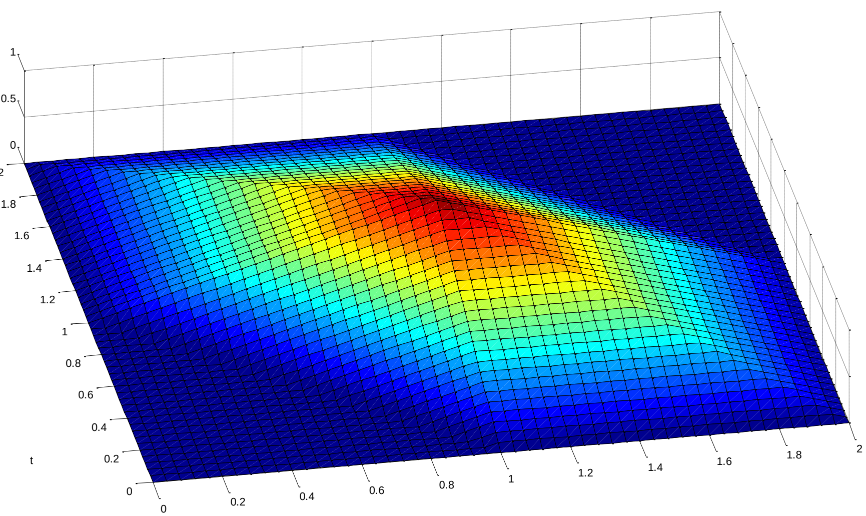
Infinite-dimensional problem

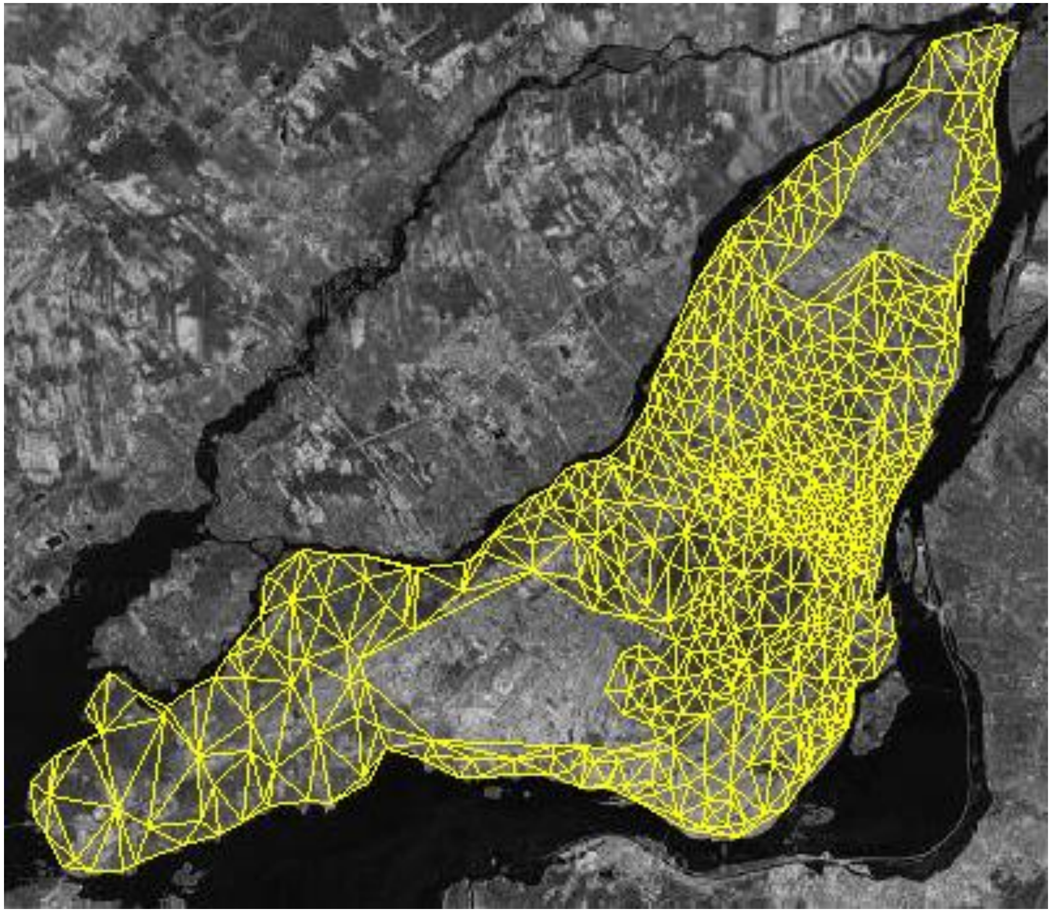
Basis expansion
Finite Elements

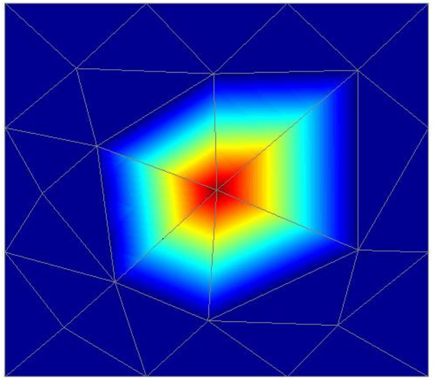
Finite-dimensional problem

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 Ω







- $\{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_K\}$: nodes of $\mathcal{T}$
- $\Omega_{\mathcal{T}}$: triangulated domain; $H^1_{\mathcal{T}}(\Omega)$: finite element space
- $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_K)^t$: finite element basis $\qquad \Psi = \{\Psi\}_{ij} := \psi_j(\mathbf{p}_i)$
- for any g in the finite element space, $g = \mathbf{g}^t\boldsymbol{\psi}$ where $\mathbf{g} := (g(\boldsymbol{\xi}_1), \ldots, g(\boldsymbol{\xi}_K))^t$
- $R_0 := \int_{\Omega_{\mathcal{T}}}(\boldsymbol{\psi}\boldsymbol{\psi}^t) \qquad R_1 := \int_{\Omega_{\mathcal{T}}}(\boldsymbol{\psi}_x\boldsymbol{\psi}_x^t + \boldsymbol{\psi}_y\boldsymbol{\psi}_y^t)$

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

- $\hat{\boldsymbol{\beta}} = (W^tW)^{-1}W^t(\boldsymbol{z} - \hat{\mathbf{f}}_n)$
- $\hat{f} = \hat{\mathbf{f}}^t\boldsymbol{\psi}$, with $\hat{\mathbf{f}}$ satisfying

$$\begin{bmatrix} -\Psi^t Q \Psi & \lambda R_1 \\ \lambda R_1 & \lambda R_0 \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \boldsymbol{g} \end{bmatrix} = \begin{bmatrix} -\Psi^t Q\, \boldsymbol{z} \\ \mathbf{0} \end{bmatrix}$$

- $\hat{\boldsymbol{\beta}}$ and $\hat{f}$ are linear in $\mathbf{z}$ $\rightarrow$ *linear estimators*

  $\hat{f}$ has typical penalized regression form, being identified by

  $$\hat{\mathbf{f}}_n = (\Psi^t Q \Psi + \lambda P)^{-1} \Psi^t Q \, \mathbf{z}$$

- Classical inferential tools are readily derived
  - mean and variances of $\hat{\boldsymbol{\beta}}$ and $\hat{f}$
  - confidence intervals for $\boldsymbol{\beta}$
  - confidence bands for f
  - *prediction intervals for new observations*
  - estimate of error variance σ^2
  - selection of smoothing parameter λ via generalized cross-validation

$$J_\lambda(\boldsymbol{\beta}, f) = \sum_{i=1}^{n} \left(z_i - \mathbf{w}_i^t \boldsymbol{\beta} - f(\mathbf{p}_i)\right)^2 + \lambda \int_\Omega (\Delta f)^2 d\Omega$$

Azzimonti et al., 2013b, TechRep

Two sources of bias:

- $\hat{f} \in H^1_{\mathcal{T}}(\Omega)$ is affected by bias due to discretization:

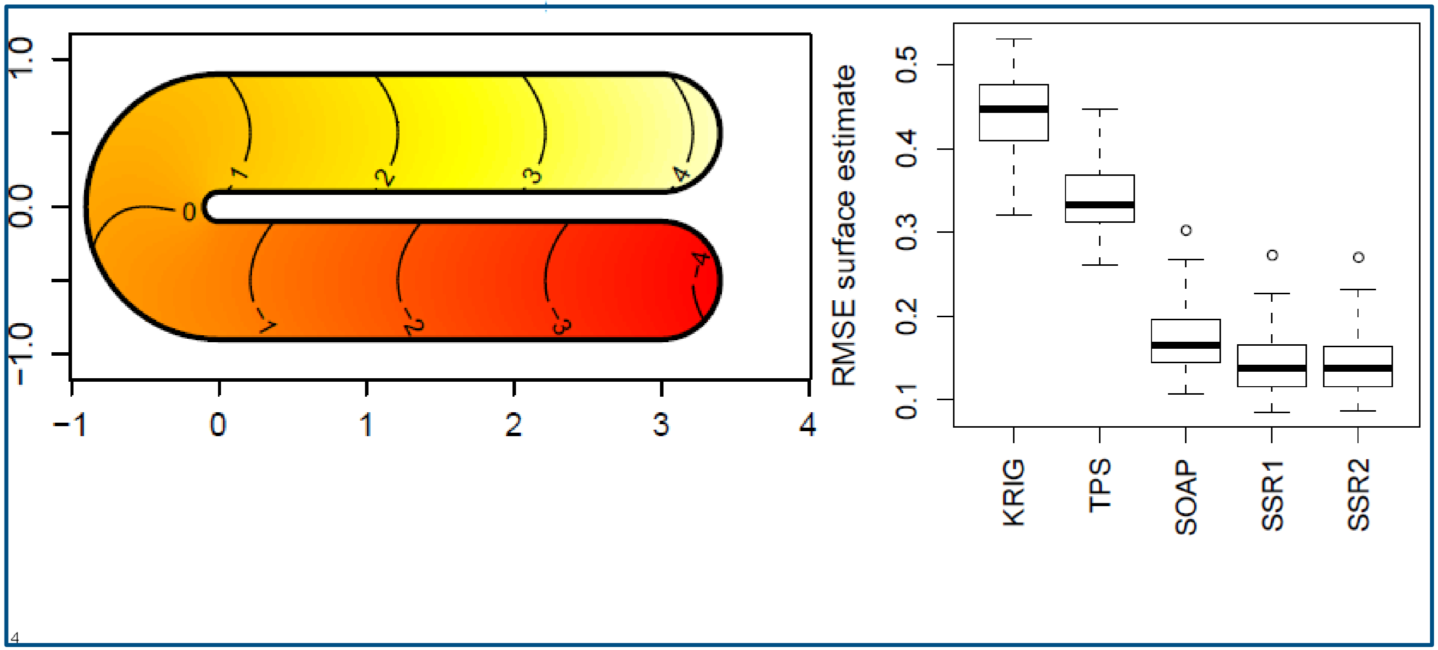
  This bias disappears as $n \to \infty$ with $h \to 0$

- $\hat{f} \in H^2_{n0}(\Omega)$ and $\hat{f} \in H^1_{\mathcal{T}}(\Omega)$ are affected by bias due to regularization

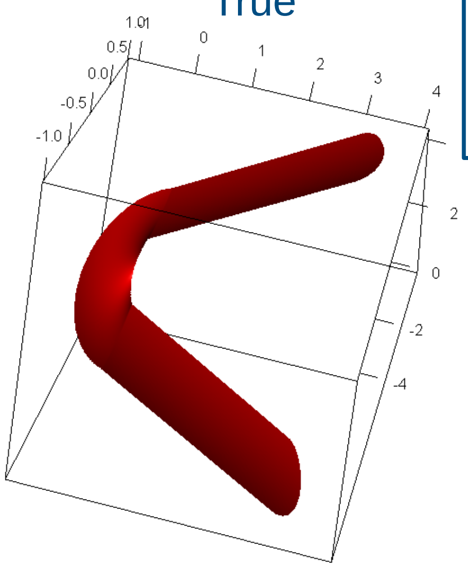
  This bias disappears as $n \to \infty$ with $\lambda \to 0$

Simulation studies

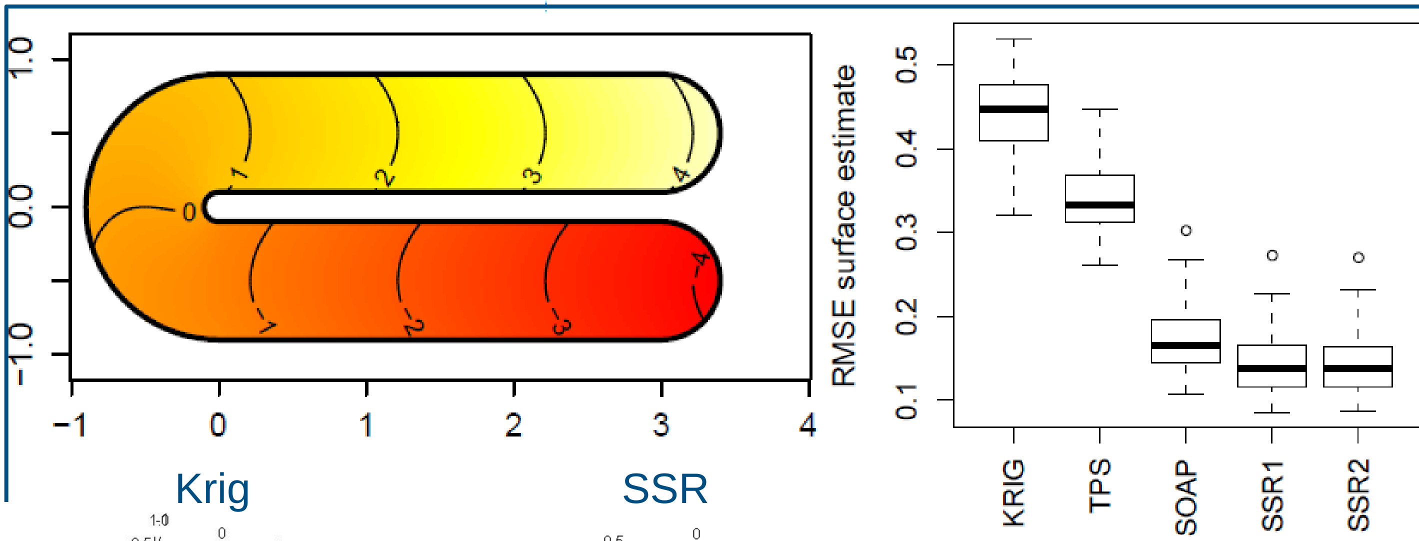
Simulation studies (Sangalli et al. 2013) show that the proposed models outperforms thin-plate splines, filtered kriging and other state-of-the-art methods for spatially distributed data.



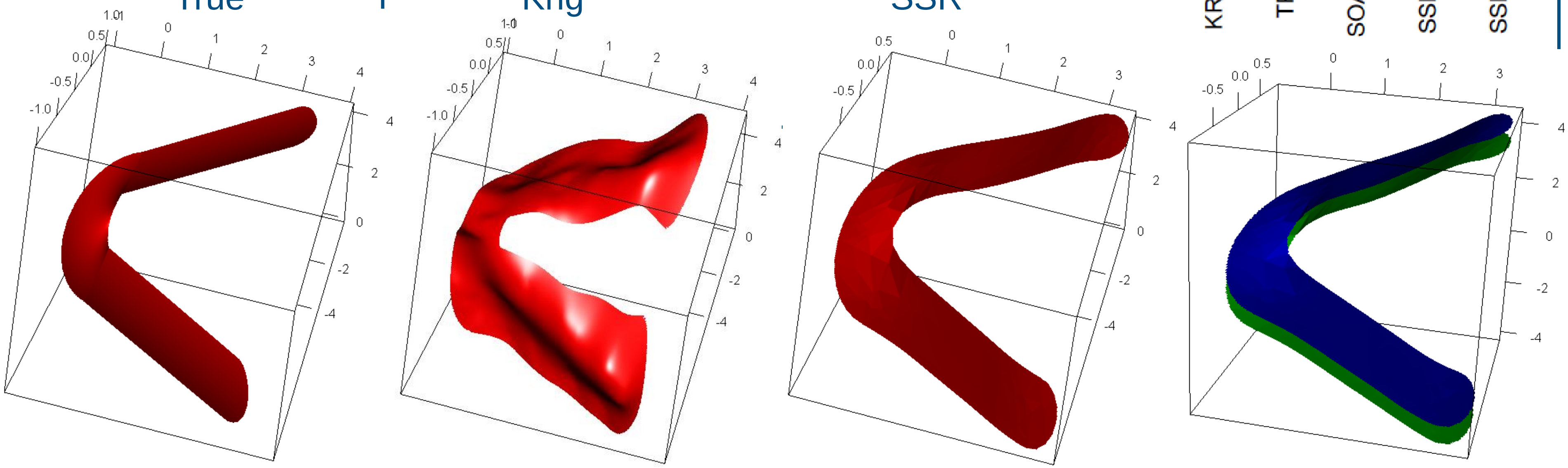
True



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.

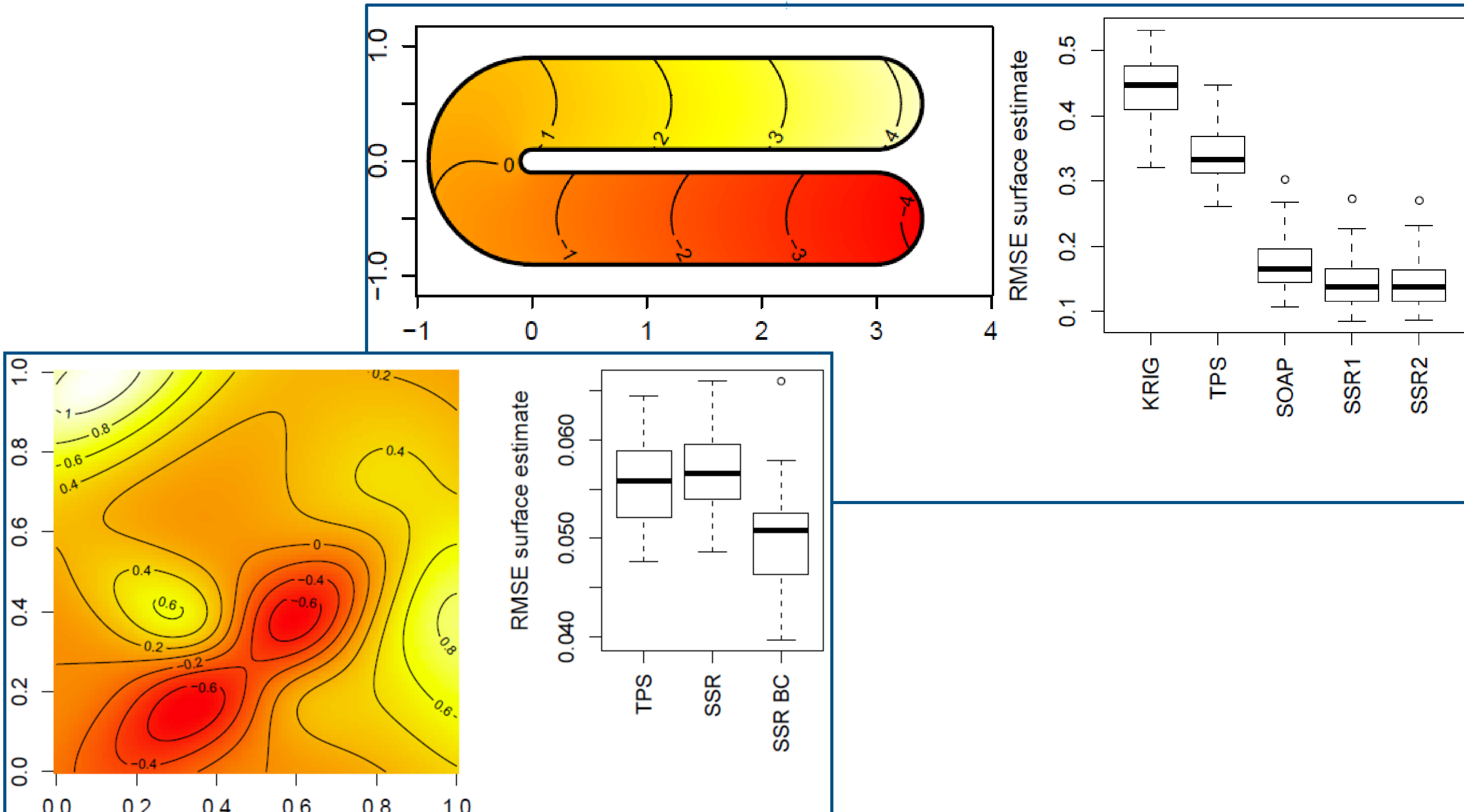


True Krig SSR



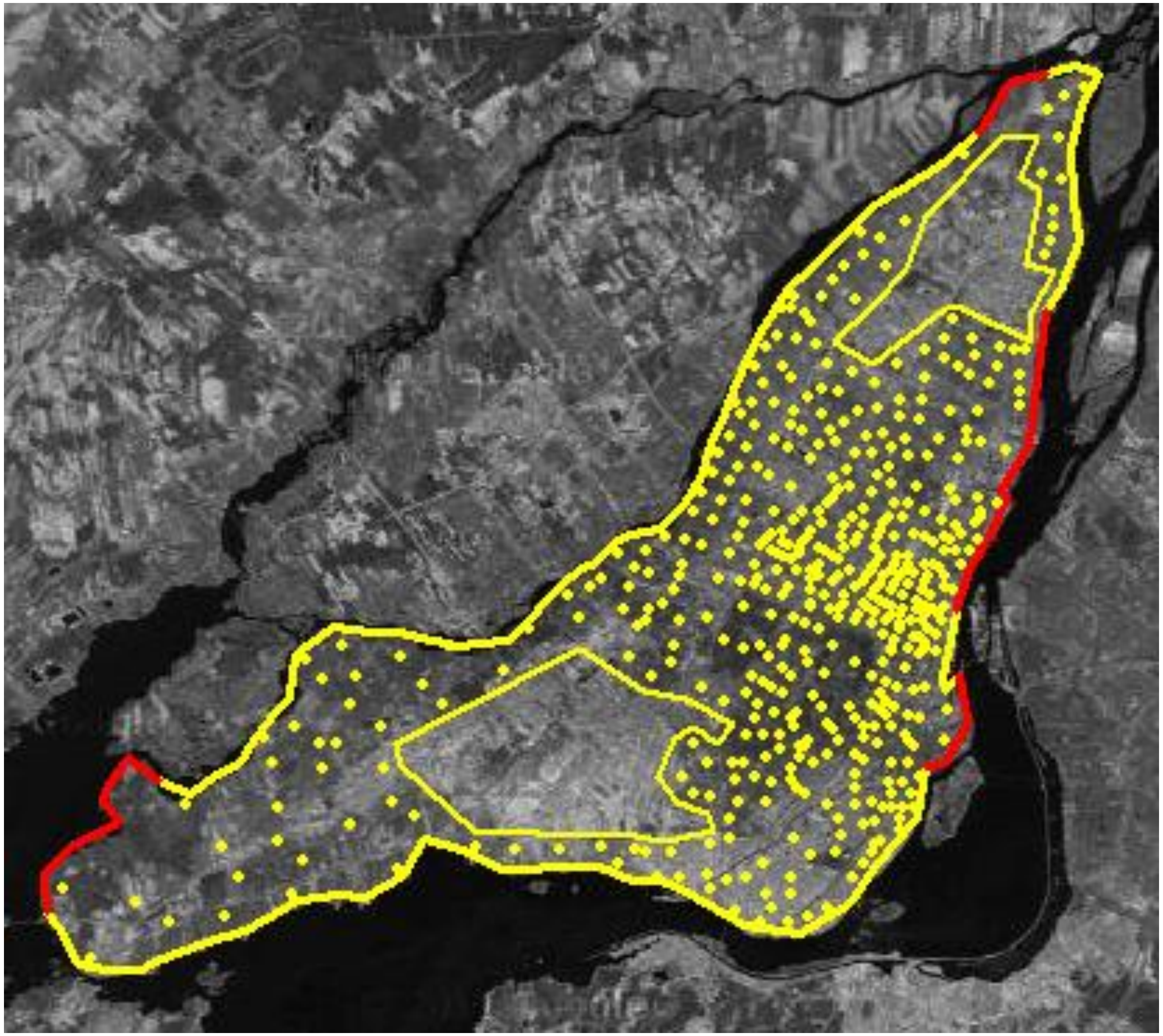
Simulation studies

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.



Illustrative example: Island of Montreal census data

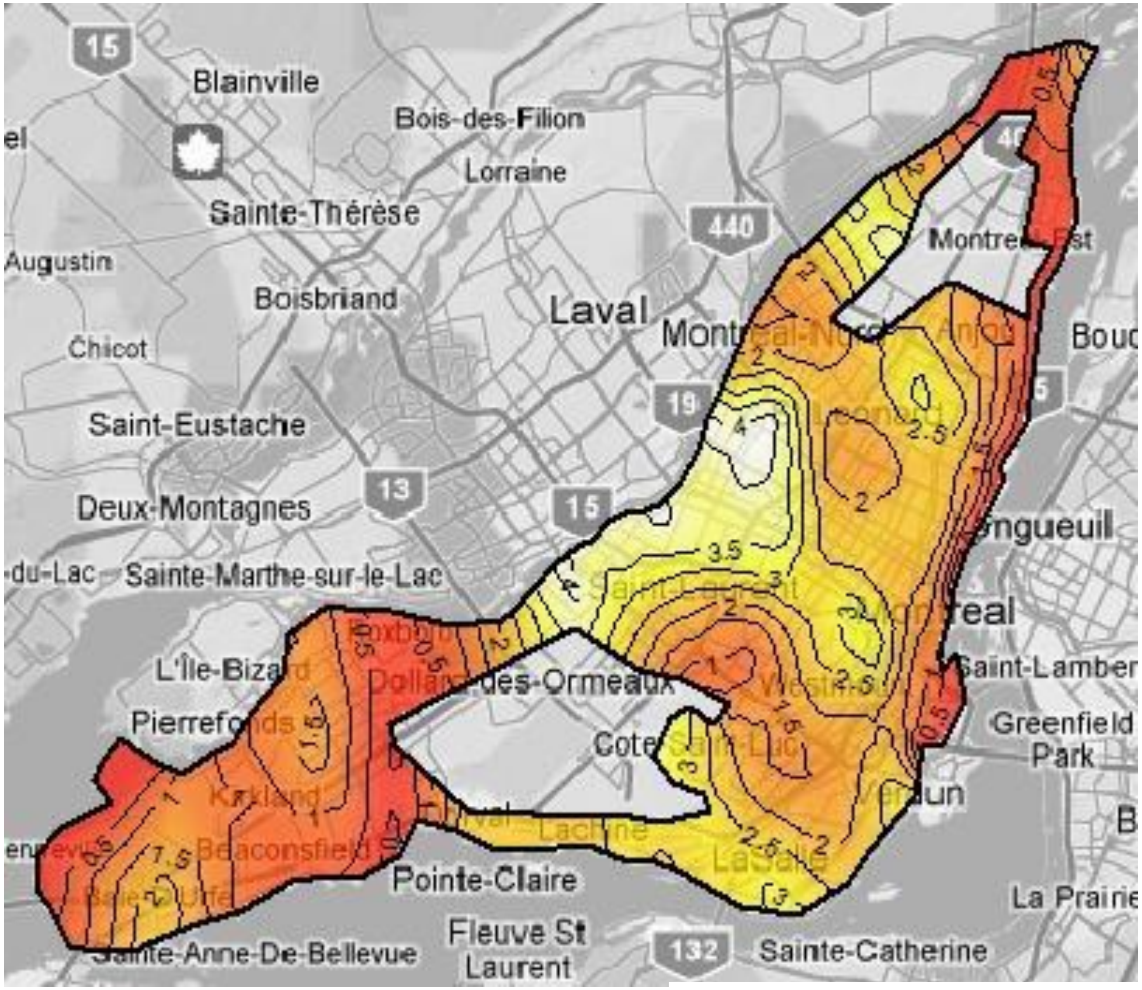
SNAPLE



- ▷ $\mathbf{p}_i$: centroids of census tracts
- ▷ z_i: population density (1000 inhabitants per km^2)
- ▷ $\mathbf{w}_i$: indicator of residential (1) commercial (0) census tract

► Mixed b.c. (homogeneous Neumann and homogeneous Dirichlet)

Illustrative example: Island of Montreal census data

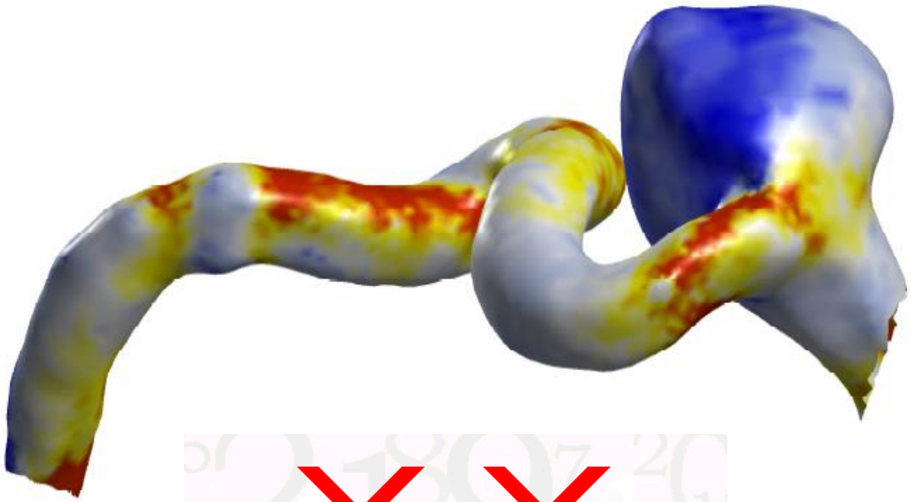


$$\hat{\beta} = 1.300$$

approx. 95% confidence interval $[0.755; 1.845]$

Motivating applied problem: AneuRisk project

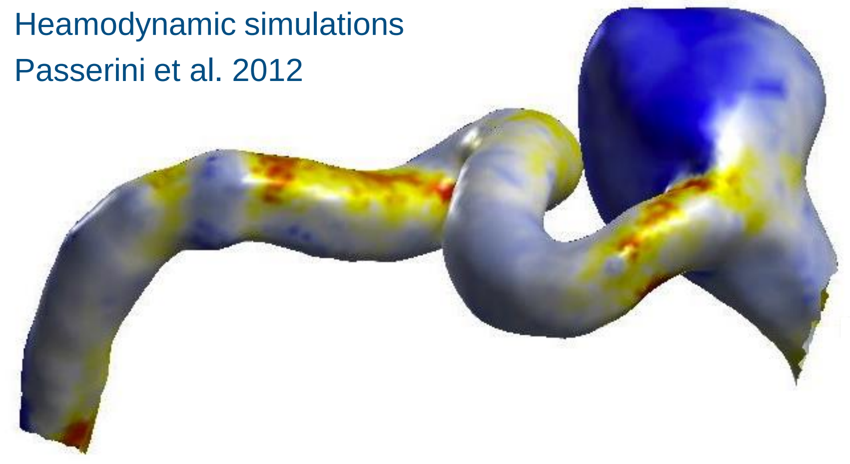
Spatial regression over bi-dimensional Riemannian manifold domains



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

Ettinger et al., 2012, TechRep

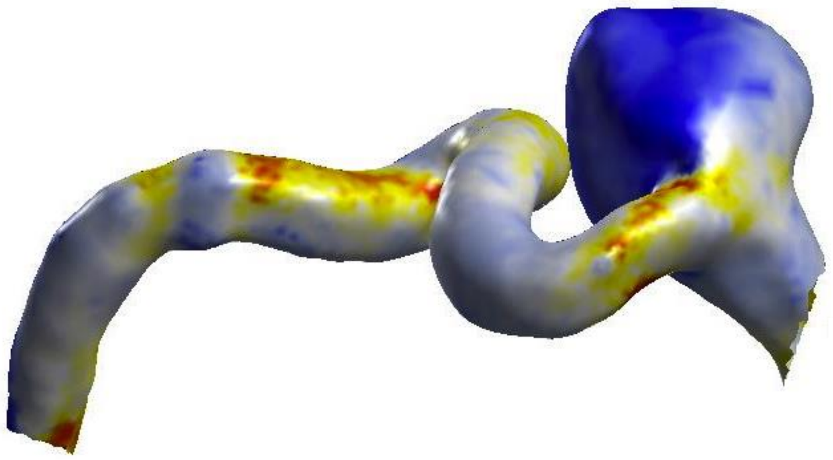
Heamodynamic simulations
Passerini et al. 2012



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
  - Baramidze, Lai, Shum, 2006, *SIAM J.S.C.*
  - Wahba,1981, *SIAM J.S.C.*
  - Lindgren, Rue, Lindstrom, 2011, *JRSSB*
  - Gneiting, 2013, *Bernoulli*

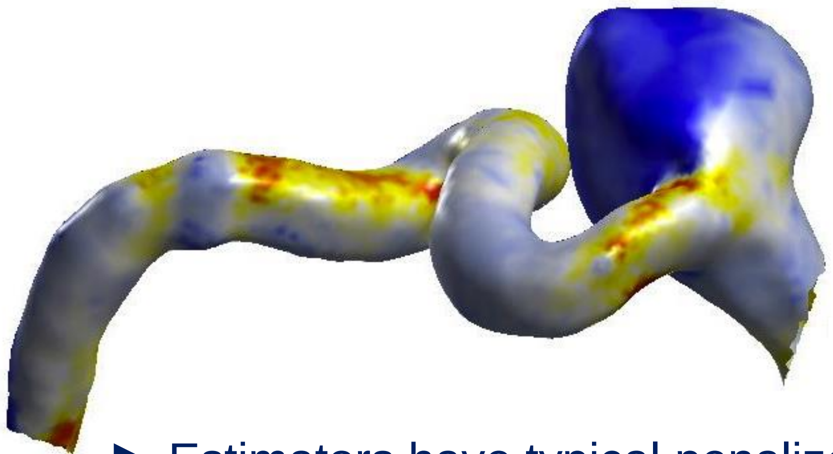
- ▷ $\Gamma \subset \mathbb{R}^3$ - a non-planar surface domain
  Artery wall
- ▷ $\{\mathbf{x}_i = (x_{1i}, x_{2i}, x_{3i}) \in \Gamma\}$ - data locations
- ▷ $z_i \in \mathbb{R}$ - variable of interest observed at $\mathbf{x}_i$
  Wall shear stress modulus at systolic peak
- ▷ $\mathbf{w}_i = (w_{i1}, \ldots, w_{iq}) \in \mathbb{R}^q$ - space varying covariates
  Local curvature of vessel wall
  Curvature of vessel
  Local radius of the vessel



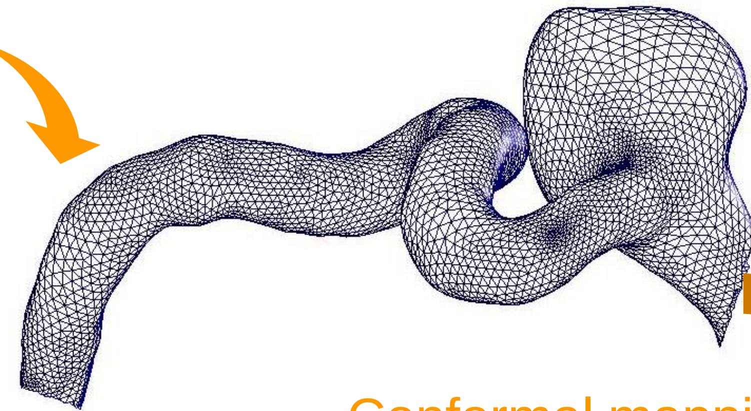
$$z_i = \mathbf{w}_i^t \boldsymbol{\beta} + f(\mathbf{x}_i) + \epsilon_i \qquad\qquad i = 1, \ldots, n$$

$$J_{\Gamma,\lambda}(\boldsymbol{\beta}, f) = \sum_{i=1}^{n} \left(z_i - \mathbf{w}_i^t \boldsymbol{\beta} - f(\mathbf{x}_i)\right)^2 + \lambda \int_\Gamma (\Delta_\Gamma f(\mathbf{x}))^2 d\mathbf{x}$$

Spatial regression over Riemannian manifolds

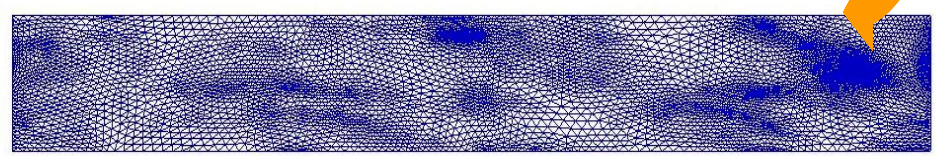


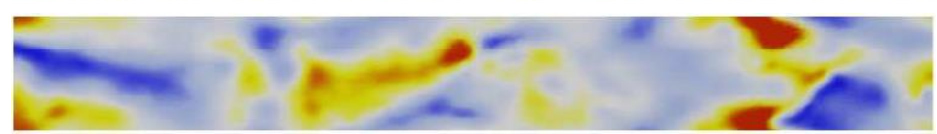
- ► Estimators have typical penalized regression form
- ► Linear in observed data values
- ► Classical inferential tools



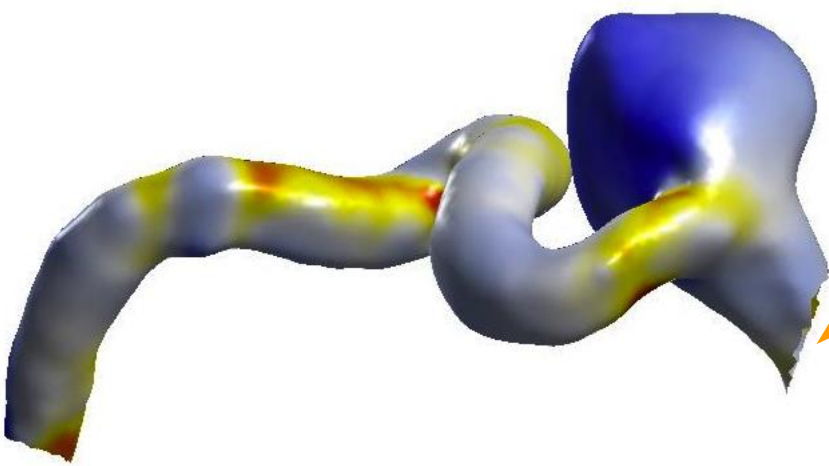
Conformal mapping

(fully encode information about complex 3D geometry)





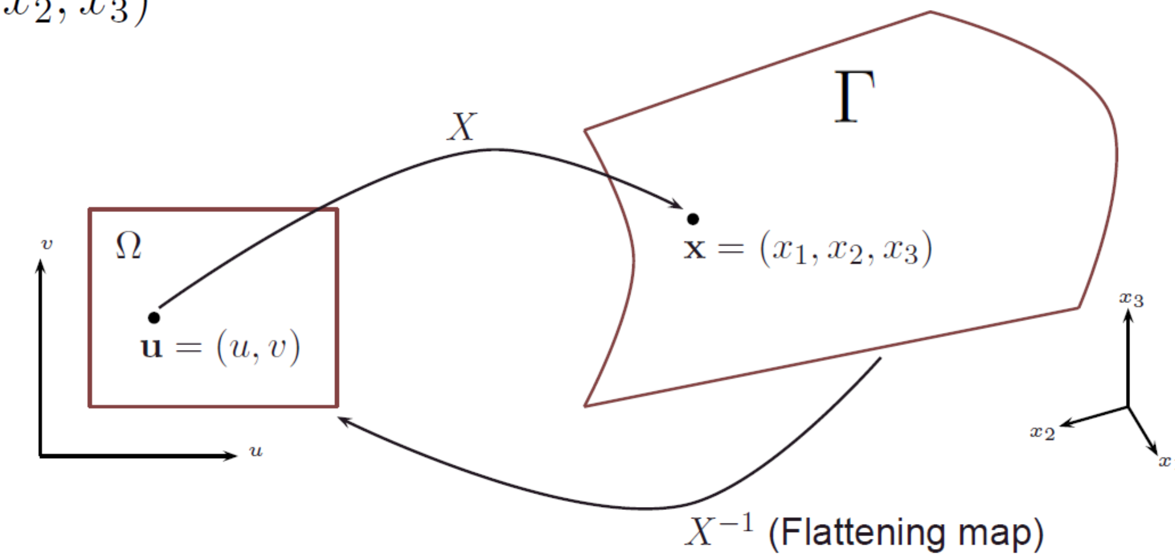
Equivalent estimation problem on planar domain



⟶ Extend method for planar domains

▷ $X : \Omega \to \Gamma$ (Ω : open, convex, bounded set in $\mathbb{R}^2$)

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



- $\frac{\partial X}{\partial u}(\mathbf{u})$, $\frac{\partial X}{\partial v}(\mathbf{u})$: column vectors

- space varying metric tensor:

$$G(\mathbf{u}) \;:=\; \nabla X(\mathbf{u})' \nabla X(\mathbf{u}) = \begin{pmatrix} \left\|\frac{\partial X}{\partial u}(\mathbf{u})\right\|^2 & \langle \frac{\partial X}{\partial u}(\mathbf{u}), \frac{\partial X}{\partial v}(\mathbf{u})\rangle \\ \langle \frac{\partial X}{\partial u}(\mathbf{u}), \frac{\partial X}{\partial v}(\mathbf{u})\rangle & \left\|\frac{\partial X}{\partial v}(\mathbf{u})\right\|^2 \end{pmatrix}$$

- $\mathcal{W}(\mathbf{u}) := \sqrt{\det(G(\mathbf{u}))}\,;$ $\quad \boxed{\mathcal{W}(\mathbf{u})d\mathbf{u} = d\mathbf{x}}$

- $\mathbf{K}(\mathbf{u}) = \mathcal{W}(\mathbf{u})\, G^{-1}(\mathbf{u})$

- For $f \circ X \in \mathcal{C}^2(\Omega)$, $\quad \mathbf{u} = X^{-1}(\mathbf{x})$

$$\nabla_\Gamma f(\mathbf{x}) \;=\; \nabla X(\mathbf{u})\; G^{-1}(\mathbf{u})(\nabla f(X(\mathbf{u})))$$

$$\boxed{\Delta_\Gamma f(\mathbf{x})} \;=\; \mathrm{div}_\Gamma(\nabla_\Gamma f(X(\mathbf{u}))) = \boxed{\frac{1}{\mathcal{W}(\mathbf{u})}\mathrm{div}\big(\mathbf{K}(\mathbf{u})\nabla f(X(\mathbf{u}))\big)}$$

▷ $H^m_{n0,\mathbf{K}}(\Omega) = \{h \in H^m(\Omega) \;:\; \mathbf{K}\nabla h \cdot n = 0 \text{ on } \partial\Omega\} \subset H^m(\Omega)$

▷ **Equivalent estimation problem over the planar domain Ω**

Find $\boldsymbol{\beta} \in \mathbb{R}^q$ and f with $(f \circ X) \in H^2_{n0,\mathbf{K}}(\Omega)$ that minimizes

$$J_{\Omega,\lambda}(\boldsymbol{\beta}, f \circ X) = \sum_{i=1}^{n} \left(z_i - \mathbf{w}_i'\boldsymbol{\beta} - f(X(\mathbf{u}_i))\right)^2 + \lambda \int_\Omega \frac{1}{\mathcal{W}} \Big(\operatorname{div}(\mathbf{K}\nabla(f \circ X))\Big)^2 d\Omega$$

where $\mathbf{u}_i = X^{-1}(\mathbf{x}_i)$

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

$$J_{\Omega,\lambda}(\boldsymbol{\beta}, f \circ X) = \sum_{i=1}^{n} \left(z_i - \mathbf{w}_i'\boldsymbol{\beta} - f(X(\mathbf{u}_i))\right)^2 + \lambda \int_\Omega \left(\frac{1}{\sqrt{\mathcal{W}(\mathbf{u})}} \Delta f(X(\mathbf{u}))\right)^2 d\Omega$$

▷ $H^m_{n0,\mathbf{K}}(\Omega) = \{h \in H^m(\Omega) \;:\; \mathbf{K}\nabla h \cdot n = 0 \text{ on } \partial\Omega\} \subset H^m(\Omega)$

▷ **Equivalent estimation problem over the planar domain Ω**

Find $\boldsymbol{\beta} \in \mathbb{R}^q$ and f with $(f \circ X) \in H^2_{n0,\mathbf{K}}(\Omega)$ that minimizes

$$J_{\Omega,\lambda}(\boldsymbol{\beta}, f \circ X) = \sum_{i=1}^{n} \left(z_i - \mathbf{w}_i'\boldsymbol{\beta} - f(X(\mathbf{u}_i))\right)^2 + \lambda \int_\Omega \frac{1}{W} \Big(\operatorname{div}(\mathbf{K}\nabla(f \circ X))\Big)^2 d\Omega$$

where $\mathbf{u}_i = X^{-1}(\mathbf{x}_i)$

→ Extend method for planar domains

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

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

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

$$\boldsymbol{\mu}_n^t \mathbf{Q}\hat{\mathbf{f}}_n + \lambda \int_\Omega \frac{1}{\mathcal{W}} \Big(\mathrm{div}(\mathbf{K}\nabla(\mu \circ X))\Big)\Big(\mathrm{div}\Big(\mathbf{K}\nabla(\hat{f} \circ X)\Big)\Big) d\Omega = \boldsymbol{\mu}_n^t \mathbf{Q}\mathbf{z}$$

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

- Weak formulation:

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

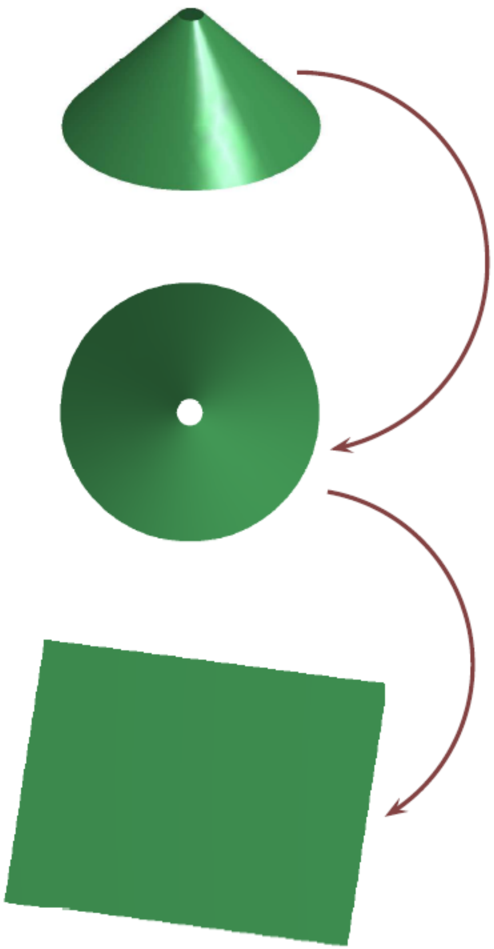
$$\boldsymbol{\mu}_n^t \mathbf{Q}\hat{\mathbf{f}}_n - \lambda \int_\Omega \mathbf{K}\nabla(\mu \circ X) \cdot \nabla(\gamma \circ X) d\Omega = \boldsymbol{\mu}_n^t \mathbf{Q}\mathbf{z}$$

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

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

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

Conformal parametrization



$$\begin{cases} -\Delta_{\Sigma} u = 0 \text{ on } \Sigma \\ u = 0 \text{ on } \sigma_0 \\ u = 1 \text{ on } \sigma_1 \end{cases}$$

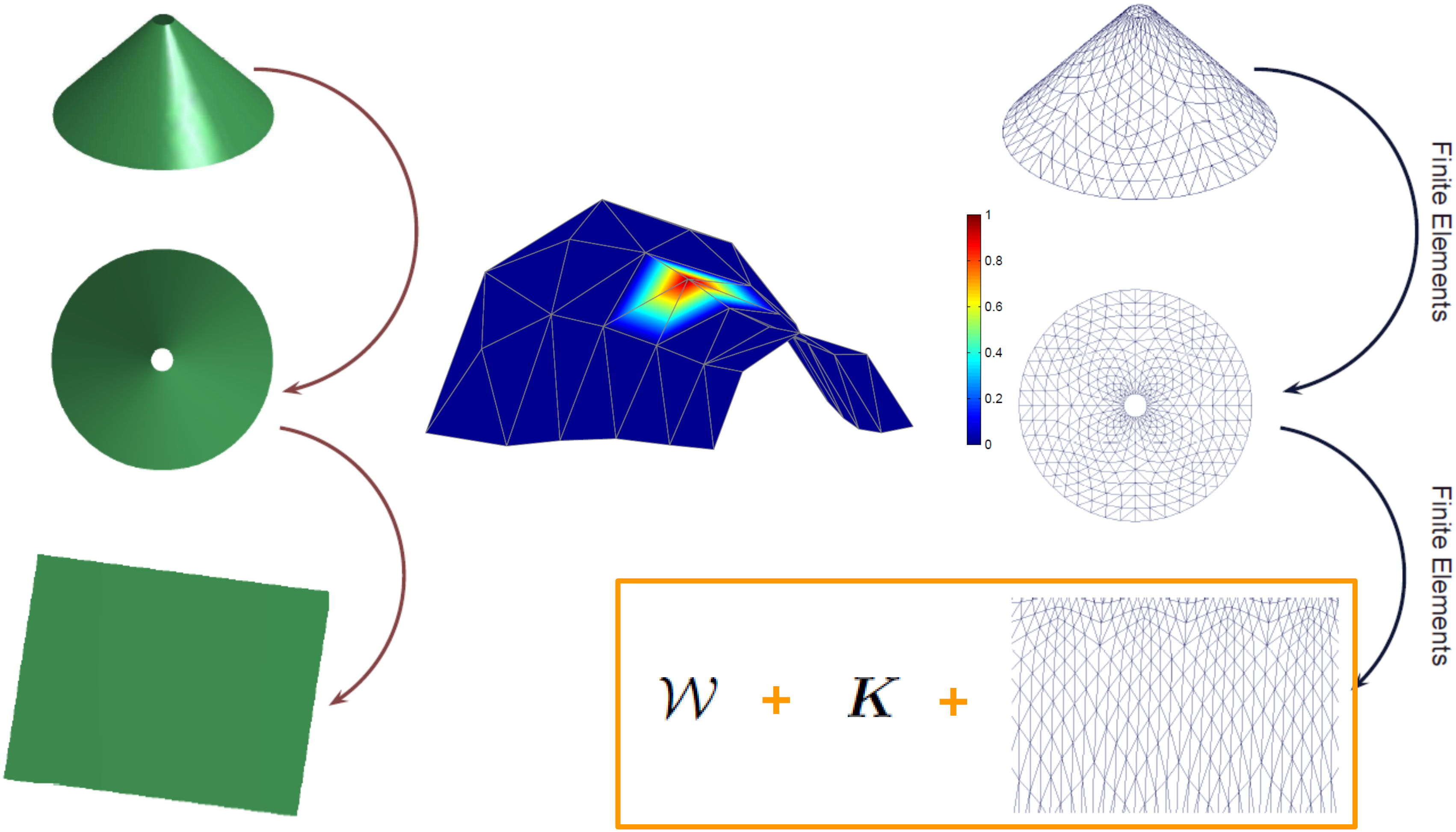
$$E_D(u) = \frac{1}{2} \int_{\Gamma} \|\nabla_{\Gamma} u\|^2 d\Gamma$$

$$\begin{cases} -\Delta_{\Sigma} v = 0 \text{ on } \Sigma \\ v(\zeta) = \int_{\zeta_0}^{\zeta} \frac{\partial u}{\partial \nu} \, ds \ \text{ on } B \end{cases}$$

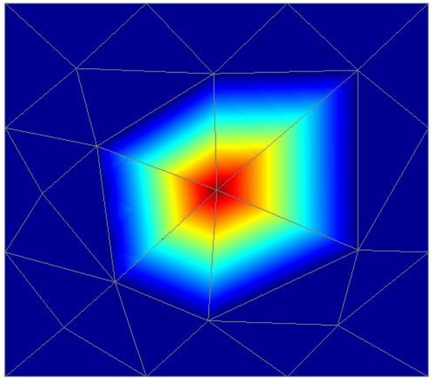
$$E_D(v) = \frac{1}{2} \int_{\Gamma} \|\nabla_{\Gamma} v\|^2 d\Gamma$$

Haker *et al*, 2000, *IEEE Trans. Med. Imag*

Conformal parametrization



Haker *et al*, 2000, *IEEE Trans. Med. Imag*



- $\{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_K\}$: nodes of planar triangulation $\mathcal{T}$
- $\Omega_{\mathcal{T}}$: planar triangulated domain; $H^1_{\mathcal{T}}(\Omega)$: finite element space
- $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_K)^t$: finite element basis $\quad \Psi = \{\Psi\}_{ij} := \psi_j(\mathbf{p}_i)$
- for any h in the finite element space, $h = \mathbf{h}^t \boldsymbol{\psi}$ where $\mathbf{h} := (h(\boldsymbol{\xi}_1), \ldots, h(\boldsymbol{\xi}_K))^t$
- $R_0 := \int_{\Omega_{\mathcal{T}}} (\boldsymbol{\psi}\boldsymbol{\psi}^t) \mathcal{W} \qquad R_1 := \int_{\Omega_{\mathcal{T}}} \nabla \boldsymbol{\psi}' \boldsymbol{K} \nabla \boldsymbol{\psi}$

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

- $\hat{\boldsymbol{\beta}} = (W^t W)^{-1} W^t (\boldsymbol{z} - \hat{\mathbf{f}}_n)$
- $\hat{f} = \hat{\mathbf{f}}^t \boldsymbol{\psi}$, with $\hat{\mathbf{f}}$ satisfying

$$\begin{bmatrix} -\Psi^t Q \Psi & \lambda R_1 \\ \lambda R_1 & \lambda R_0 \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \boldsymbol{g} \end{bmatrix} = \begin{bmatrix} -\Psi^t Q\, \boldsymbol{z} \\ \mathbf{0} \end{bmatrix}$$

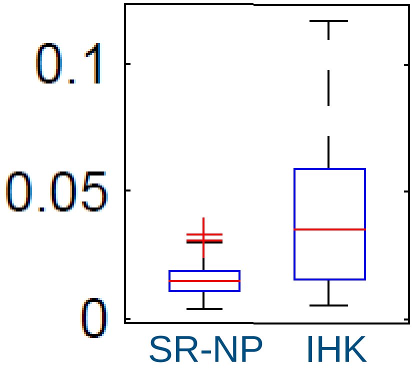
Simulation (without covariates)

TRUE

TRUE + NOISE

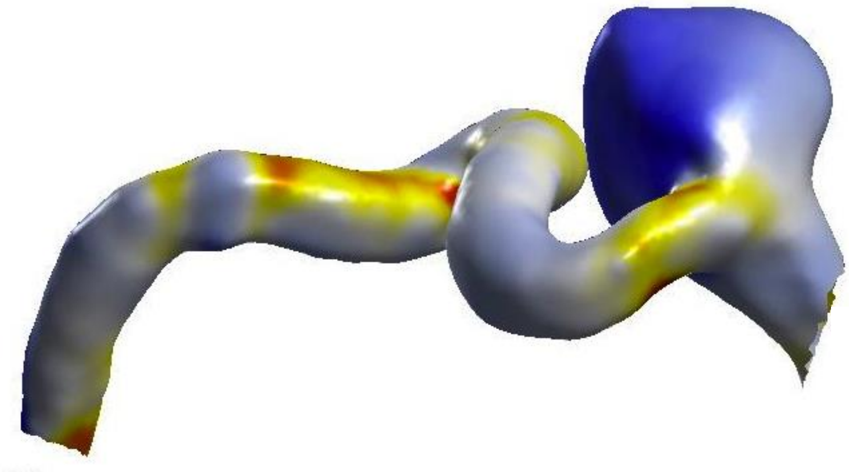
ESTIMATE

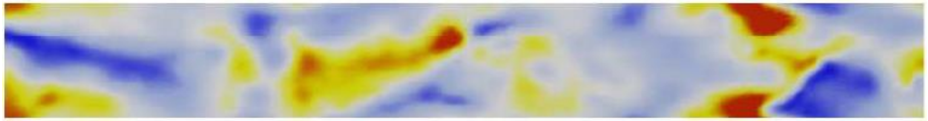
RMSE



50 simulation replicates

Spatial regression over Riemannian manifolds



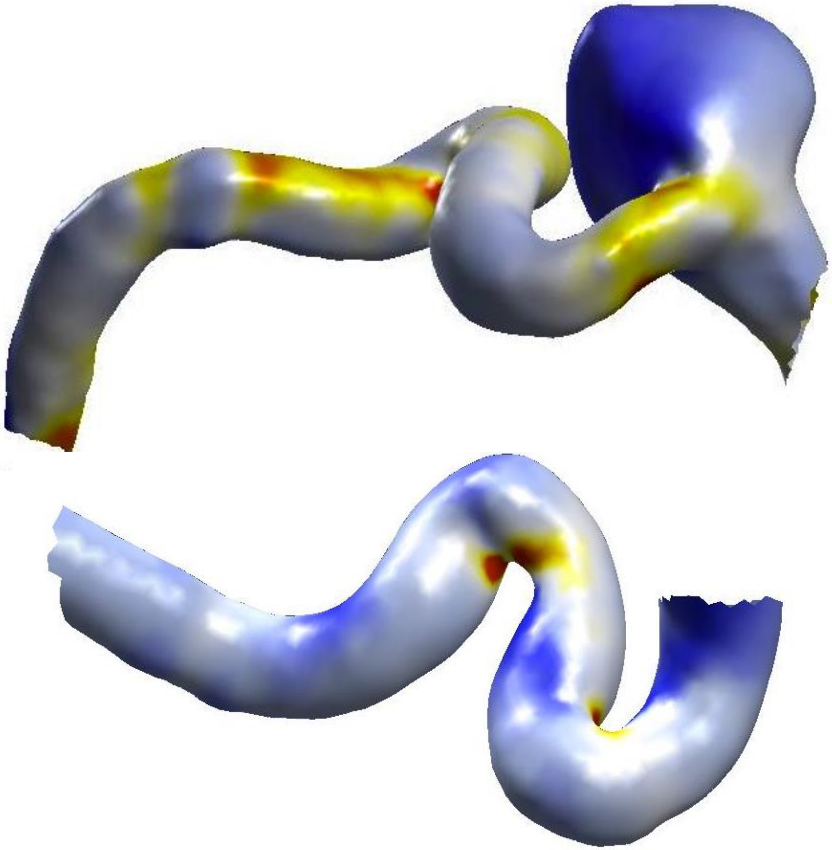


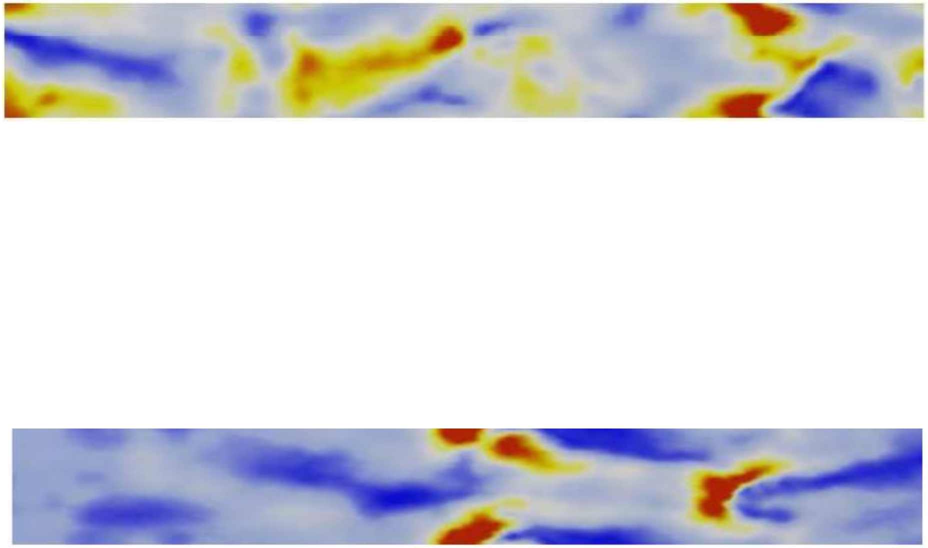
Covariates:

- Local curvature of vessel wall → Negative association

- Curvature of vessel → Positive association

- Local radius of vessel → Negative association





Variability across patients

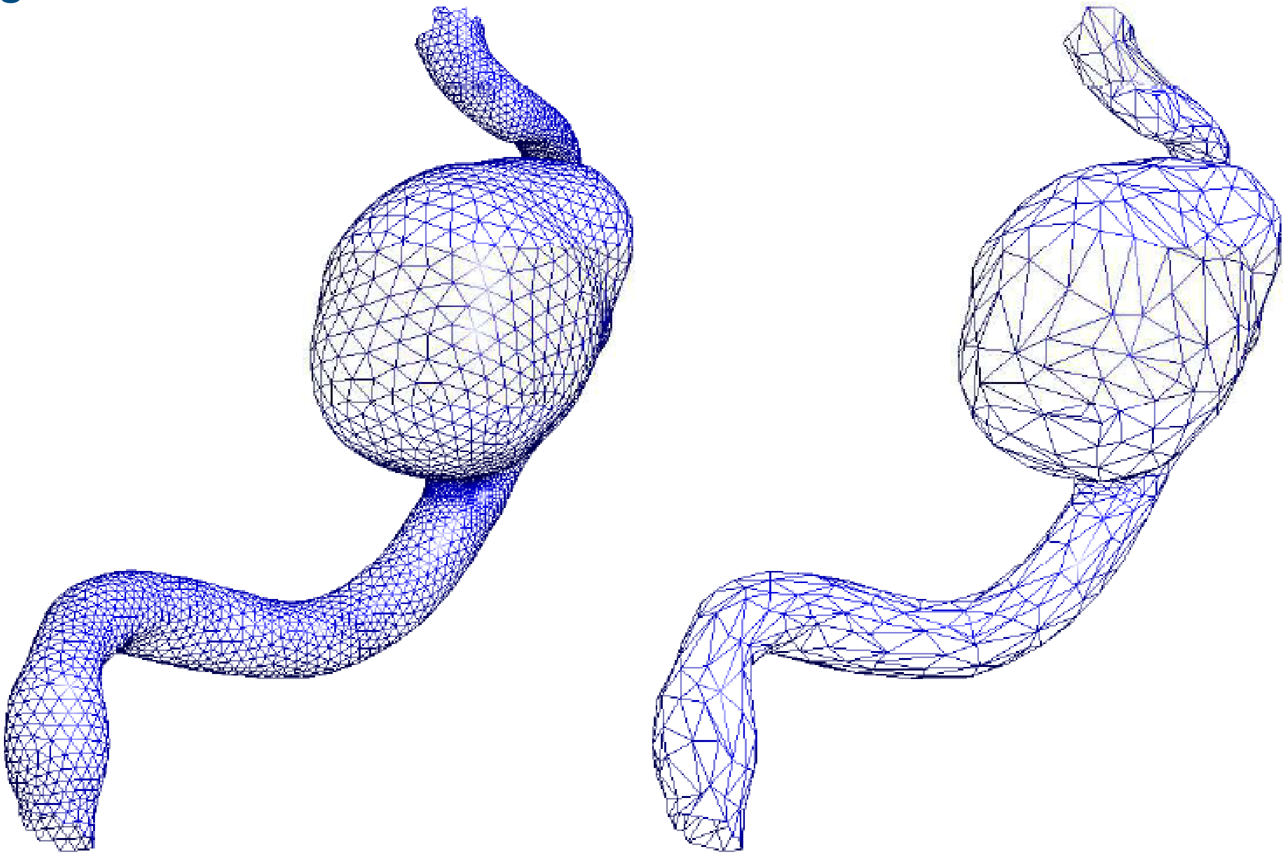
...

(data registration)

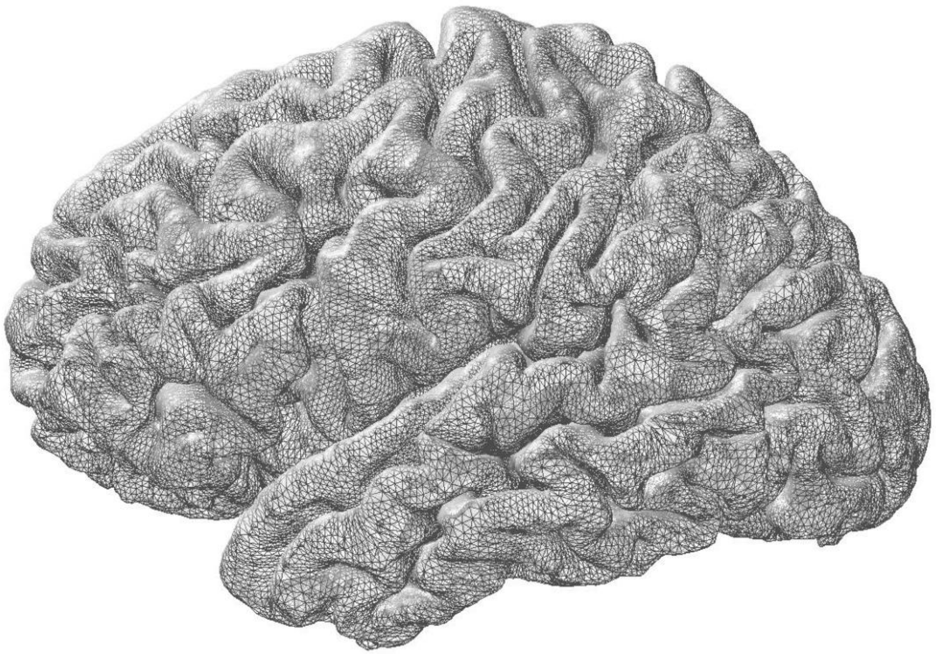
Facing big data challenges:

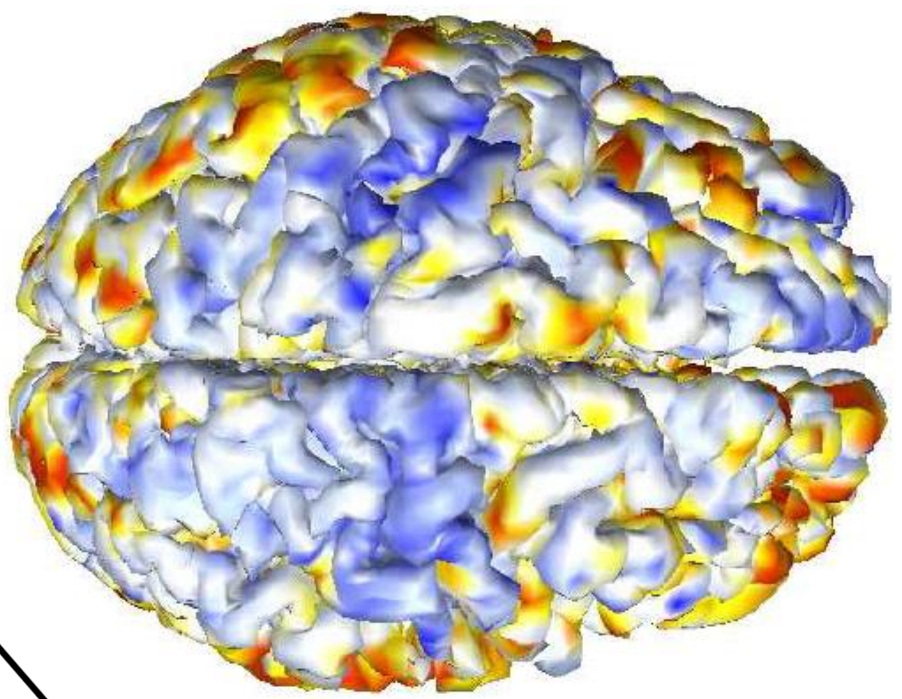
- iterative algorithms

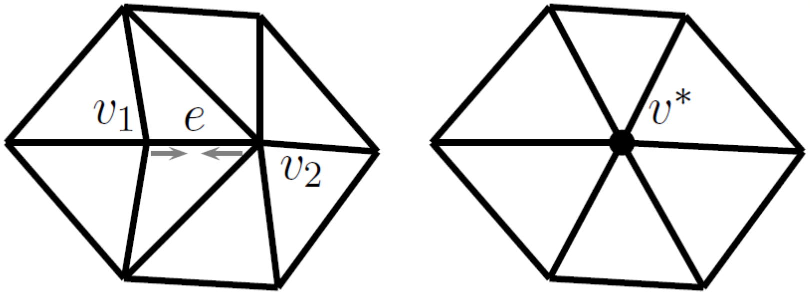
- mesh simplification algorithms



Dassi et al., 2013, TechRep



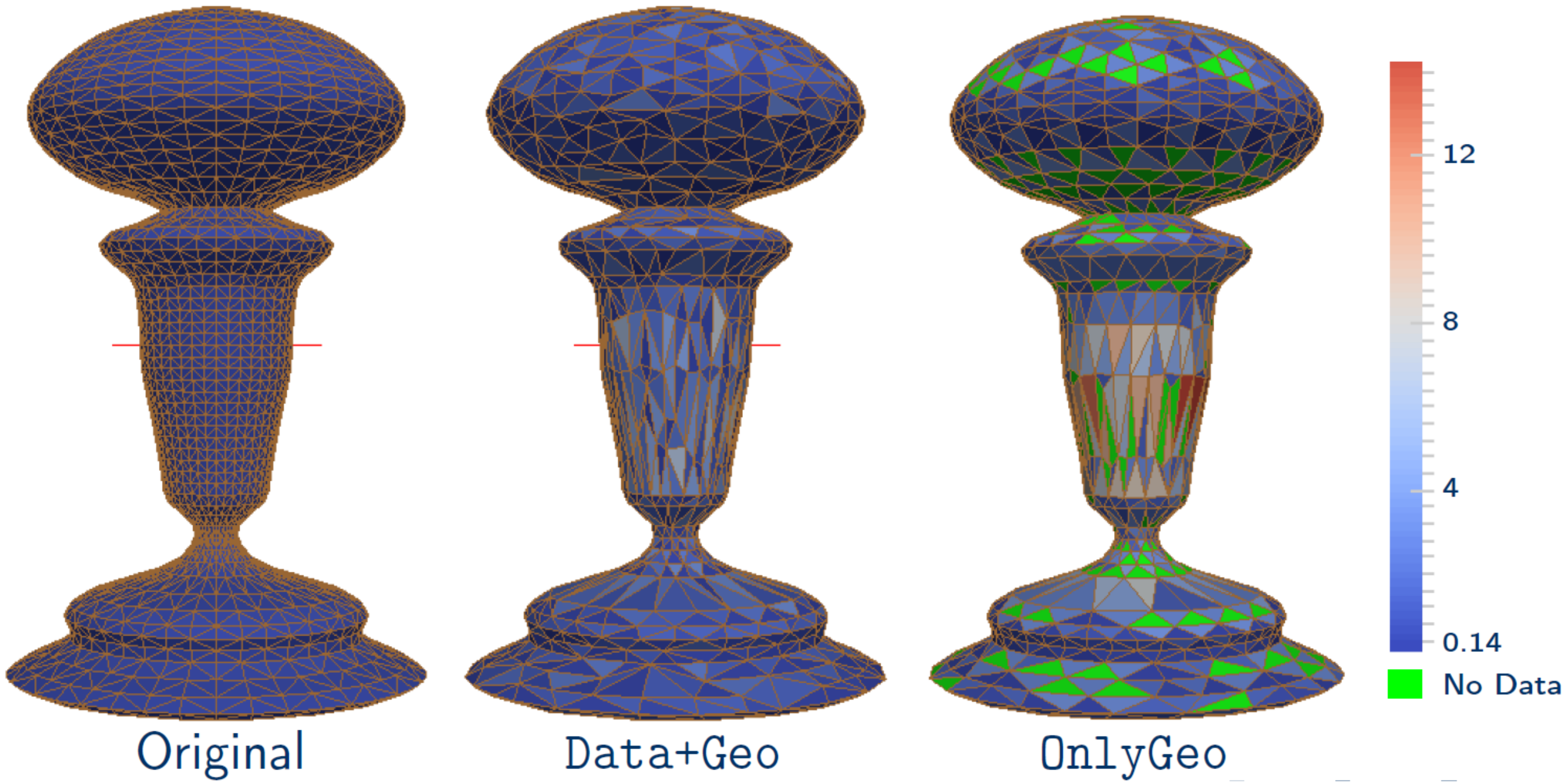




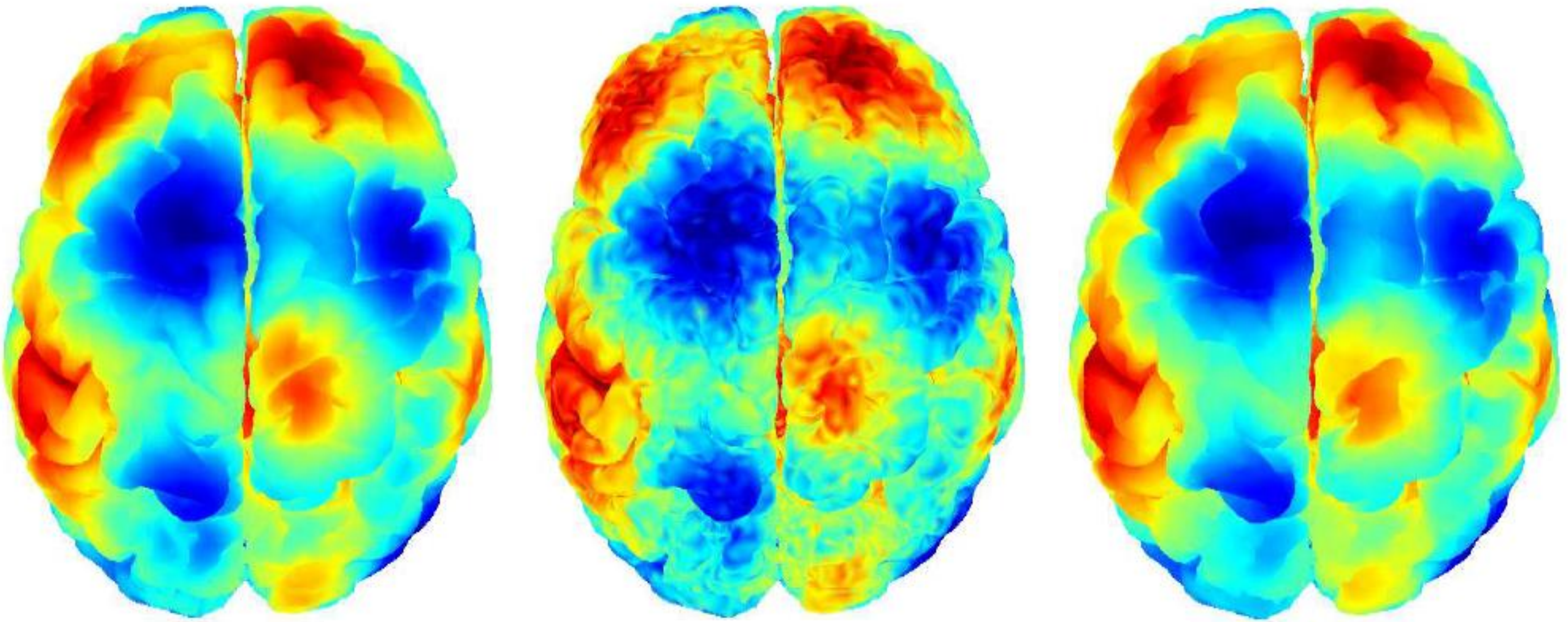
$$c(e, v^*) := \alpha c_{\text{geo}}(e, v^*) + (1 - \alpha) c_{\text{data}}(e, v^*), \quad 0 \leq \alpha \leq 1$$

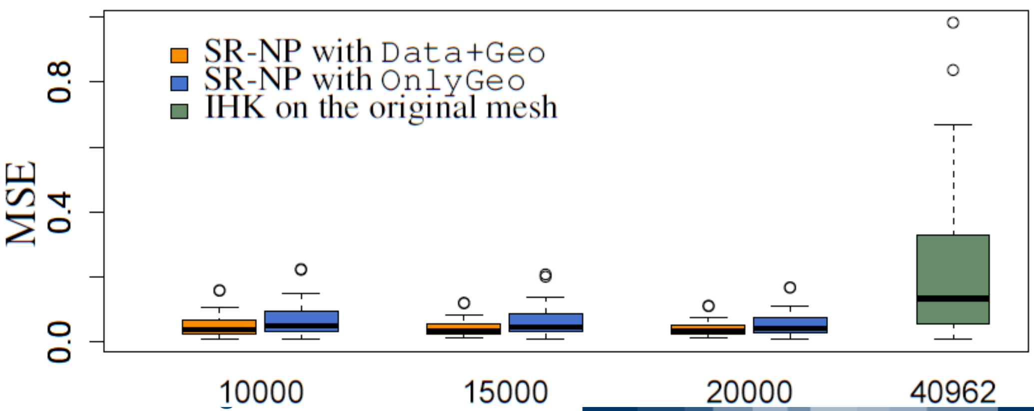
$$c_{\text{equi}}(e,\, v^*) := \frac{1}{\#(\mathcal{T}_{\text{cont}})}\left(\sum_{T\in\mathcal{T}_{\text{cont}}} (N_T - \overline{N})^2\right)$$

$$N_T := n_{\text{faces}} + \frac{1}{2}n_{\text{edges}} + \frac{1}{\#(\mathcal{T}_{v_1})}n_{v_1} + \frac{1}{\#(\mathcal{T}_{v_2})}n_{v_2} + \frac{1}{\#(\mathcal{T}_{v_3})}n_{v_3}$$



TRUE | TRUE + NOISE | ESTIMATE





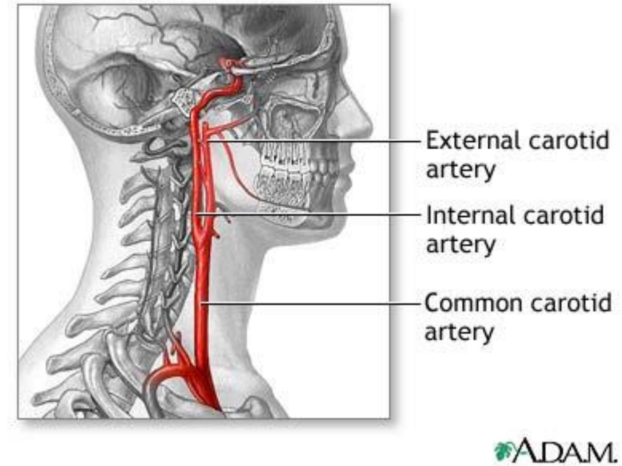
MACAREN@MOX Project: MAthematics for CARotid ENdarterectomy @ MOX

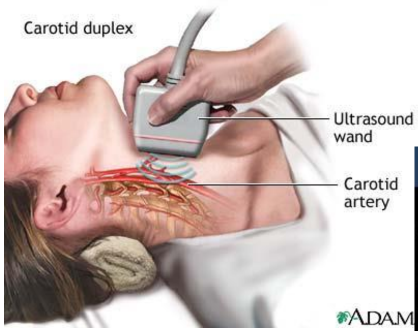
Aim: Study the pathogenesis of atherosclerotic plaques

MODELLISTICA E CALCOLO SCIENTIFICO MOX MODELING AND SCIENTIFIC COMPUTING — Statistics, Computer science

UNIVERSITA' DEGLI STUDI DI BERGAMO — Numerical analysis

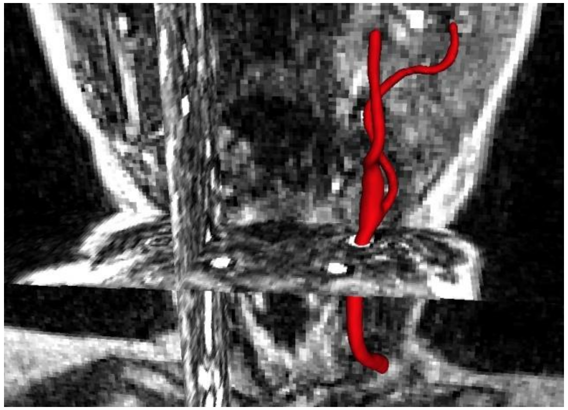
OSPEDALE MAGGIORE POLICLINICO, MANGIAGALLI E REGINA ELENA FONDAZIONE IRCCS DI NATURA PUBBLICA — Vascular surgery





Echo-Color Doppler (ECD)

Blood fluid-dynamics



Magnetic Resonance Imaging (MRI)

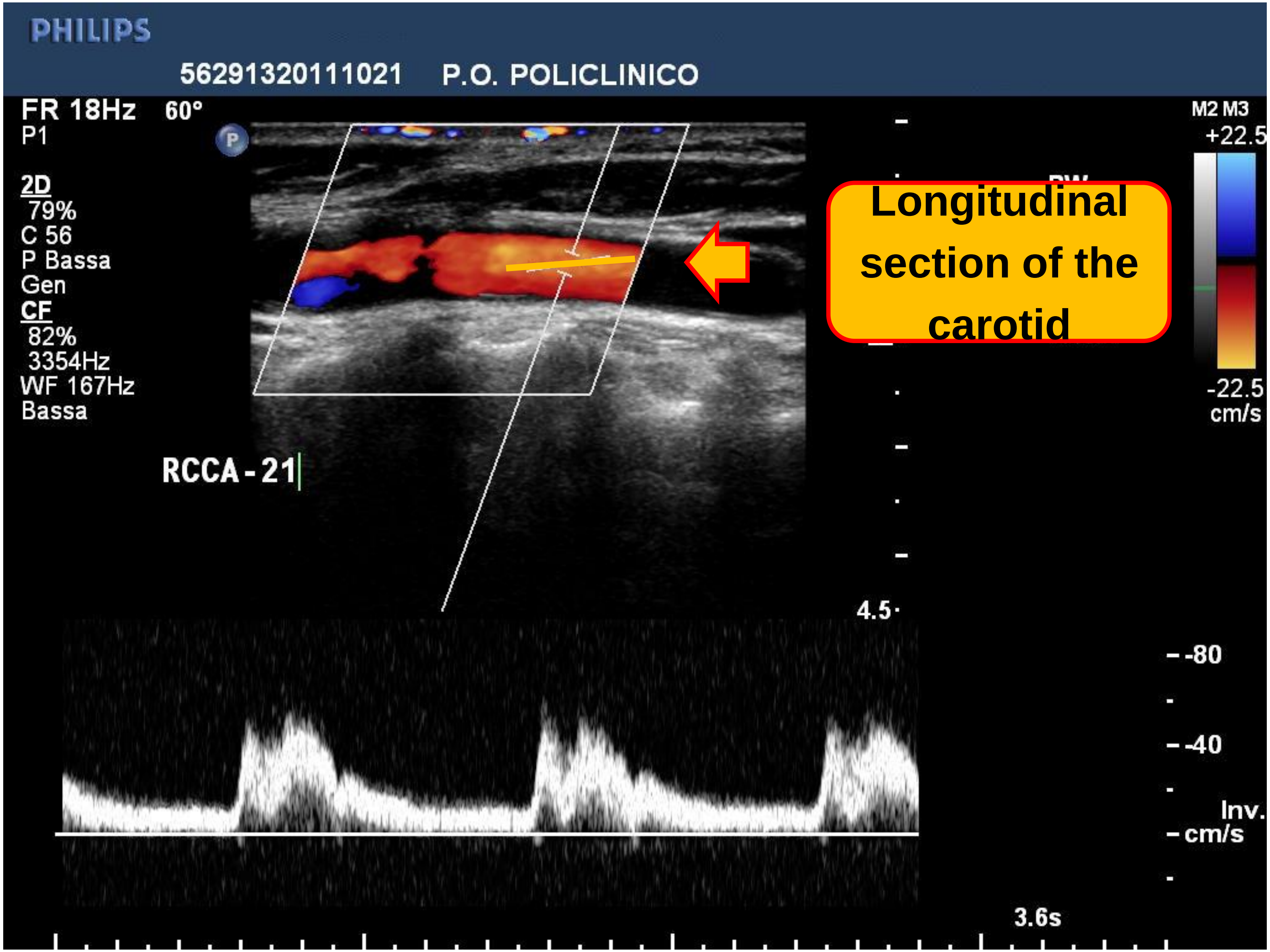
Vessel morphology

Motivating applied problem: *MACAREN@MOX*

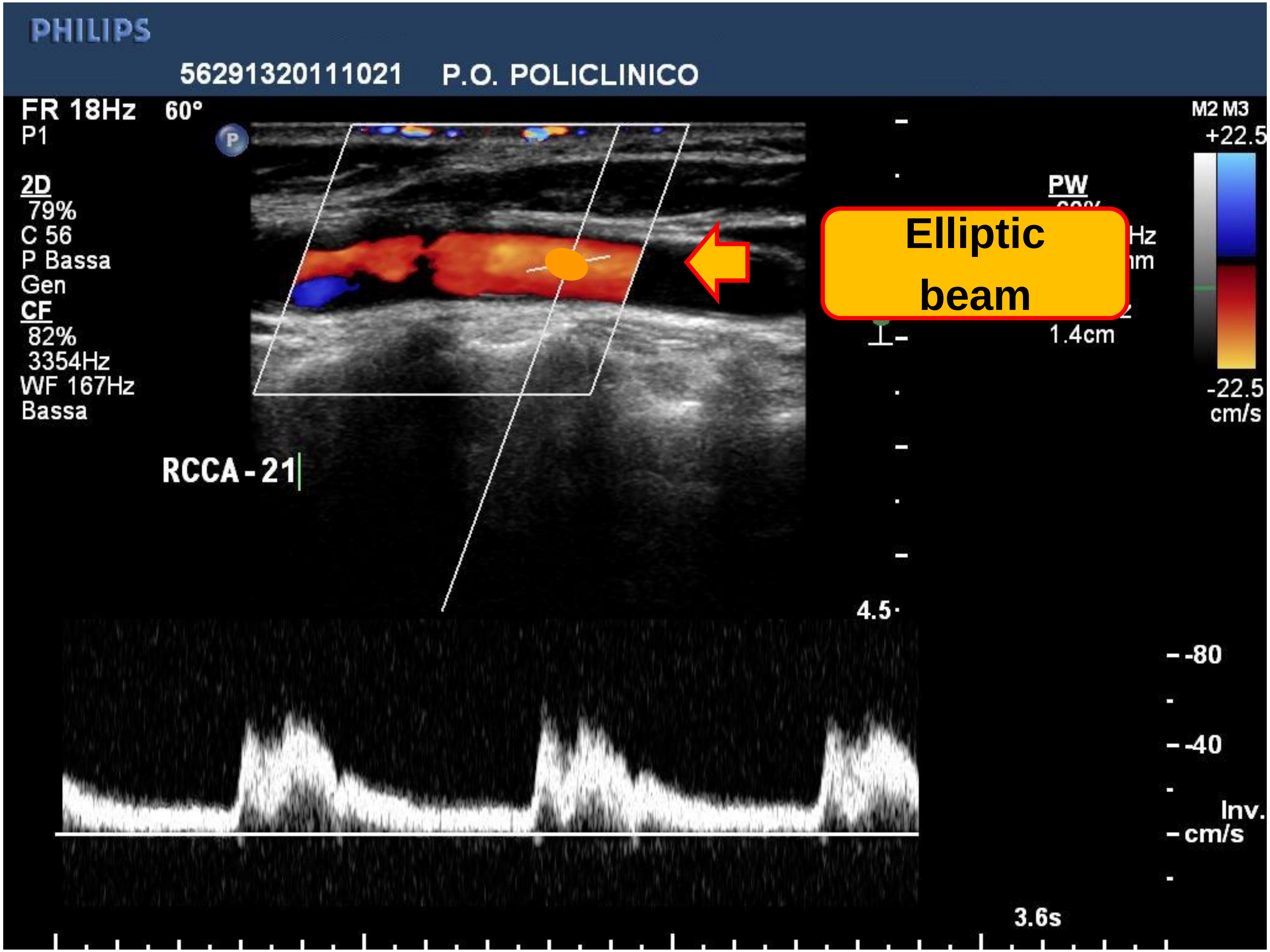
POLITECNICO DI MILANO 150°

FUTURO IN RICERCA

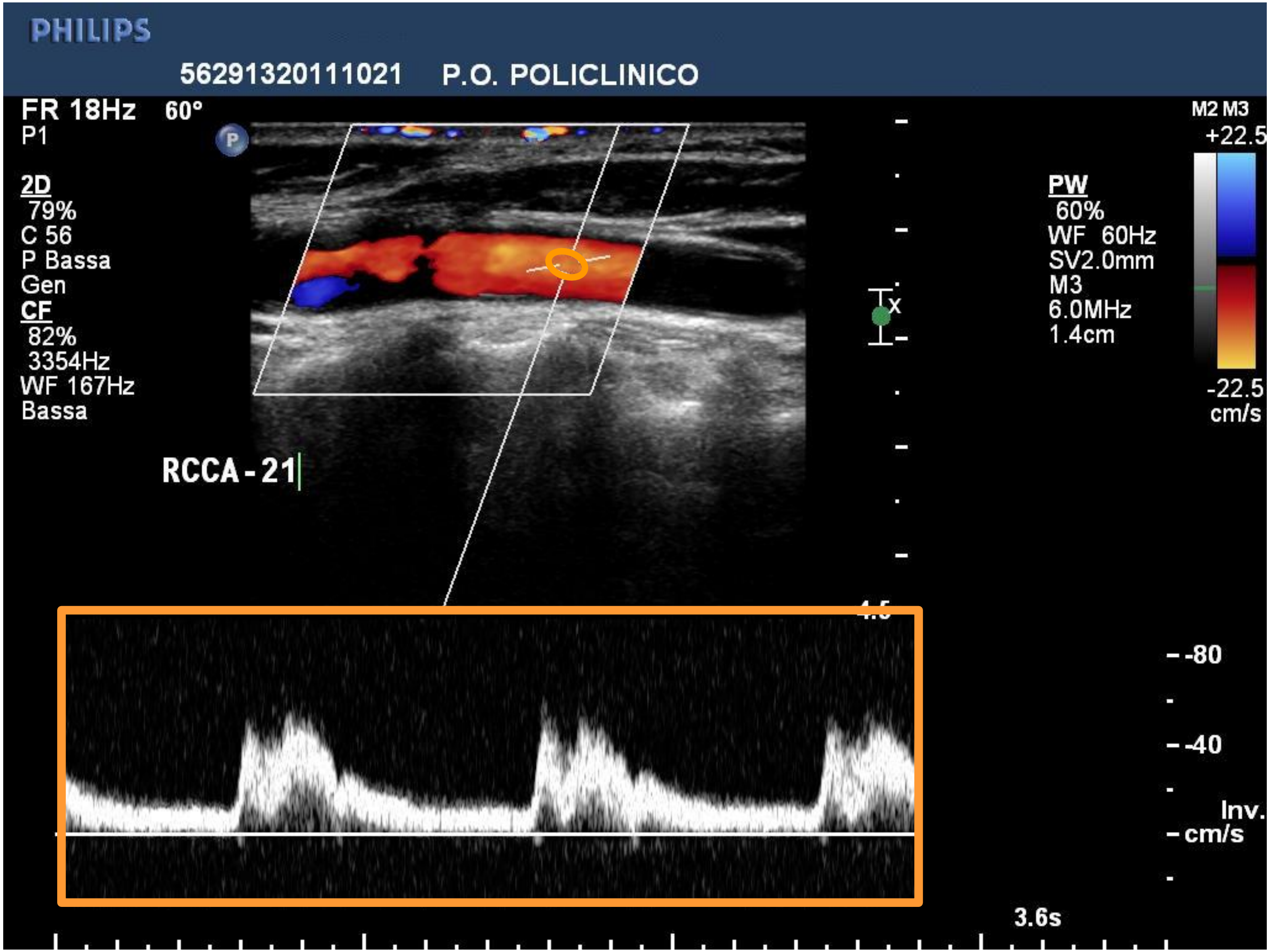
SNAPLE

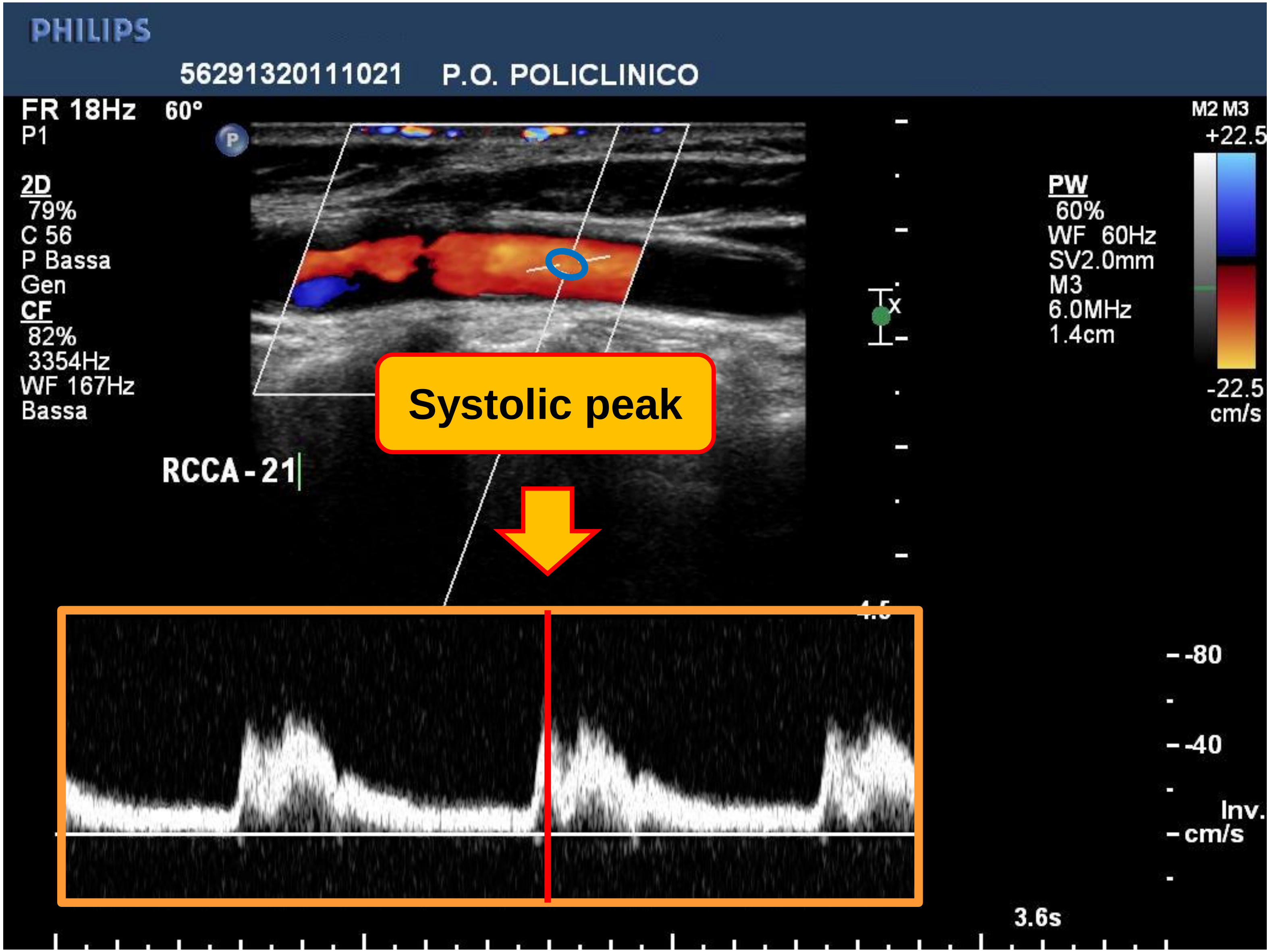


Motivating applied problem: *MACAREN@MOX*



Motivating applied problem: *MACAREN@MOX*



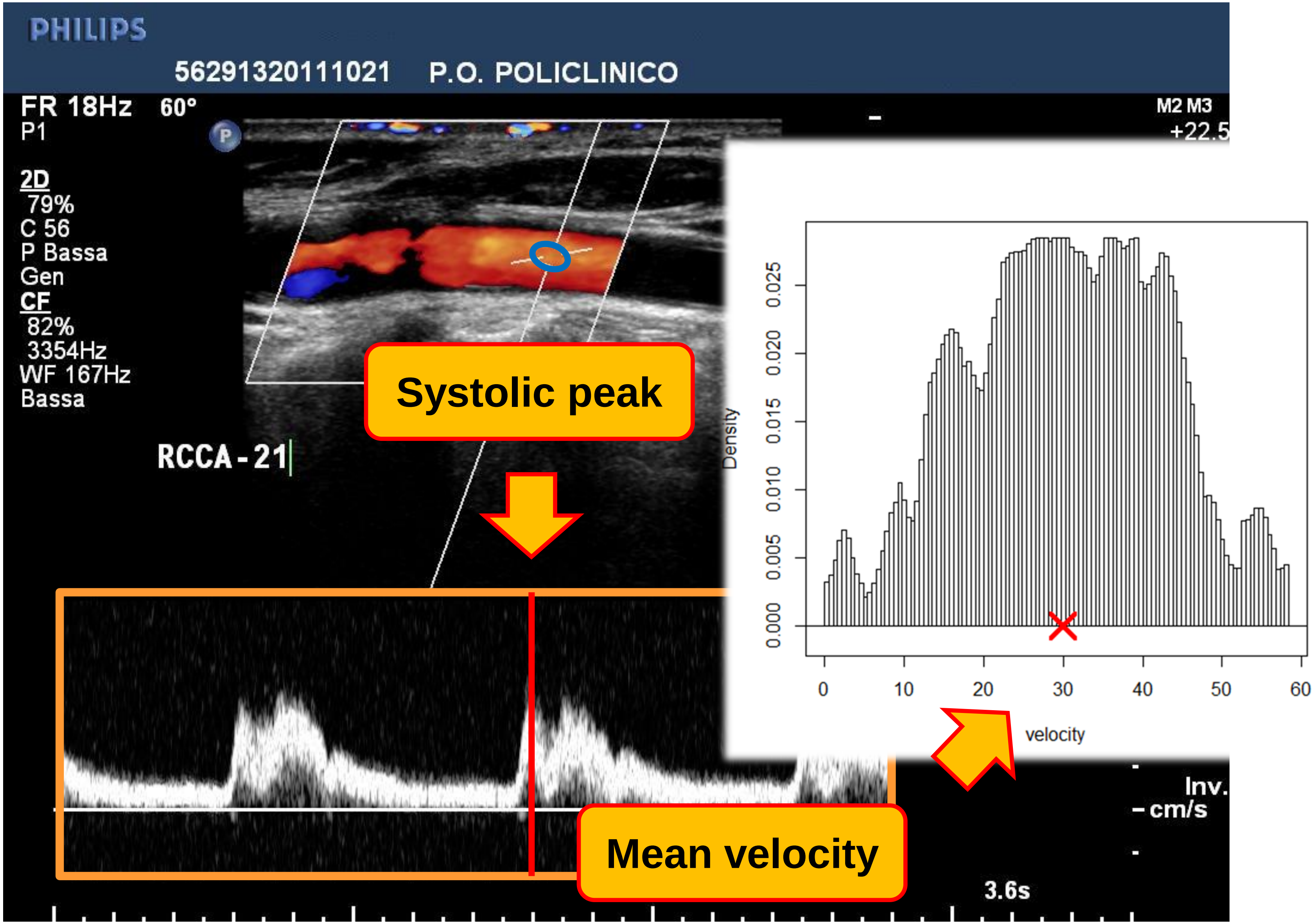

PHILIPS
56291320111021 P.O. POLICLINICO
FR 18Hz 60°
P1
2D
79%
C 56
P Bassa
Gen
CF
82%
3354Hz
WF 167Hz
Bassa
RCCA-21
PW
60%
WF 60Hz
SV2.0mm
M3
6.0MHz
1.4cm
M2 M3
+22.5
-22.5
cm/s
Systolic peak
-80
-40
Inv.
cm/s
3.6s

Motivating applied problem: *MACAREN@MOX*

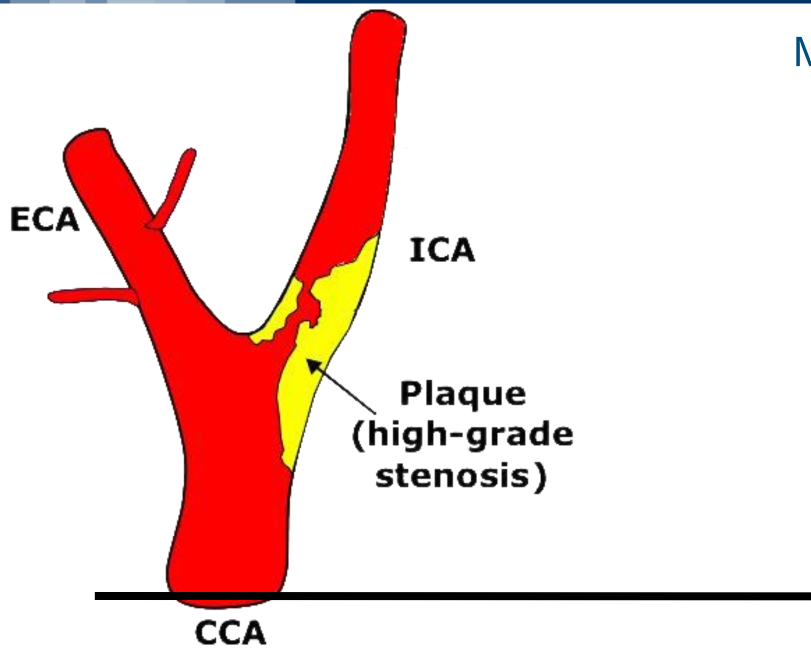
POLITECNICO DI MILANO 150°

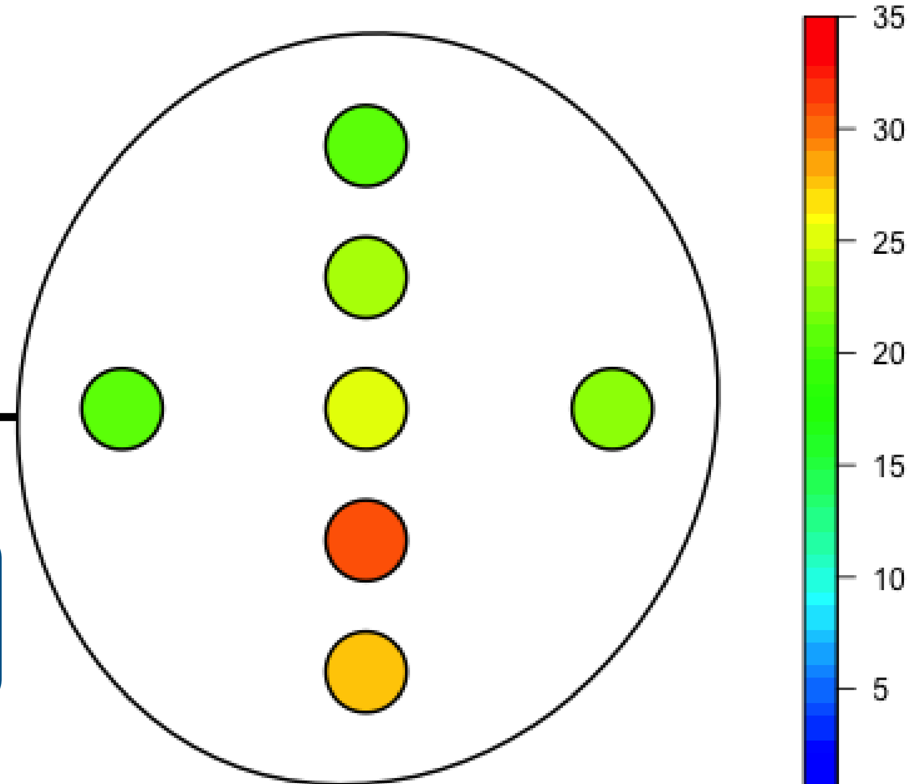
FUTURO IN RICERCA

SNAPLE



Mean measurements over 7 beams (systolic peak)





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

FIRST GOAL:
estimate blood-flow velocity field over the carotid section

Motivating applied problem: *MACAREN@MOX*

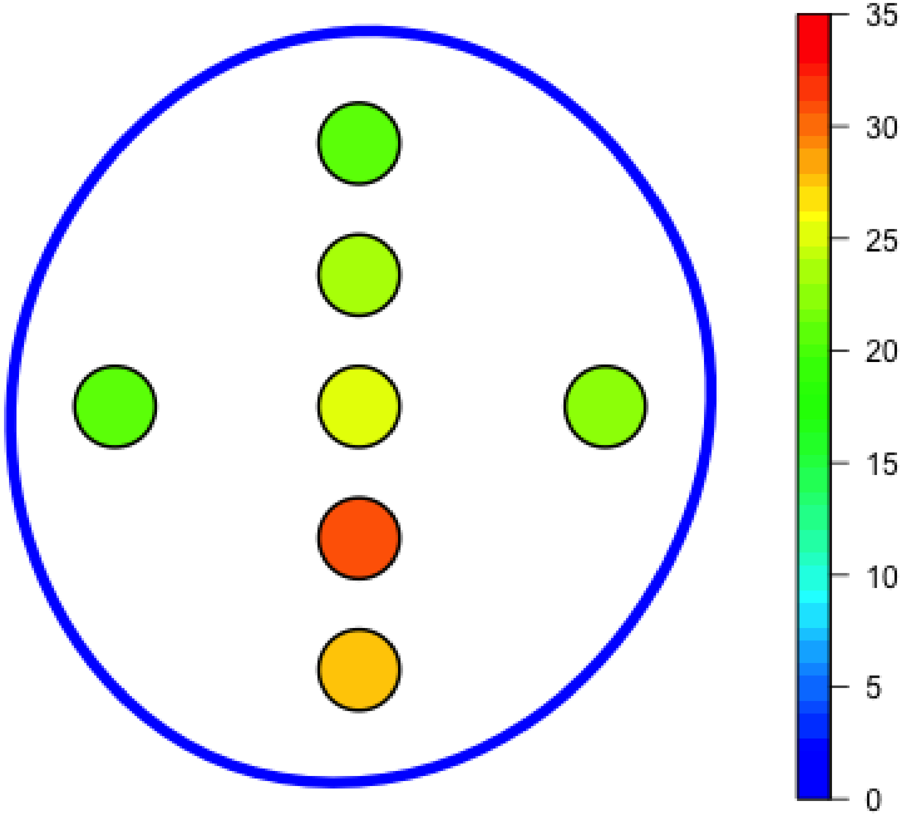
Spatial Spline Regression

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

Boundary conditions

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

Mean measurements over 7 beams (systolic peak)

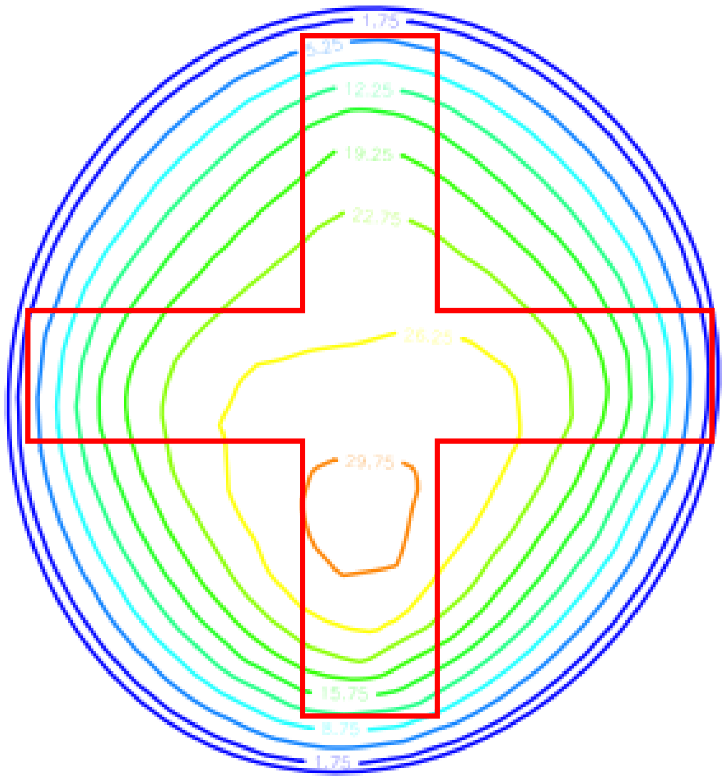


► *Physiological boundary conditions:*
velocity=0 near the arterial wall

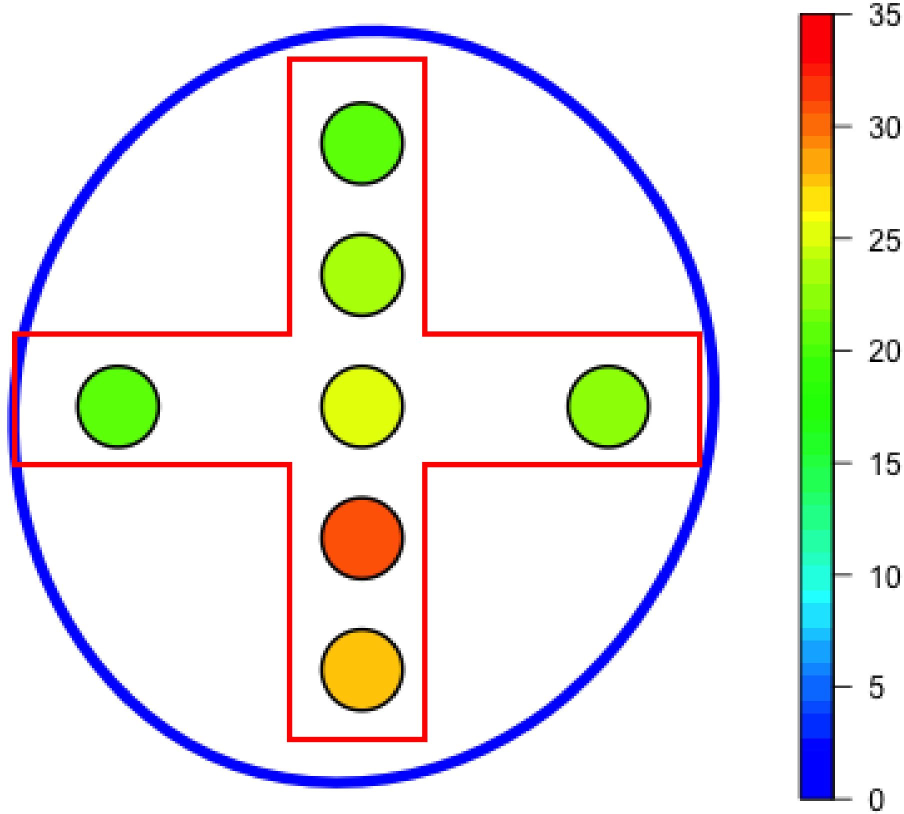
Motivating applied problem: *MACAREN@MOX*

Spatial Spline Regression

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



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 turbolence has parabolic profile

Azzimonti et al., 2013a, TechRep

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

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

$$Lf = -div(K \nabla f) + \mathbf{b} \cdot \nabla s + cs$$

forcing term

$$u \in L^2(\Omega)$$

The parameters can be space-varying

PRIOR information described by a partial differential model

Forcing term

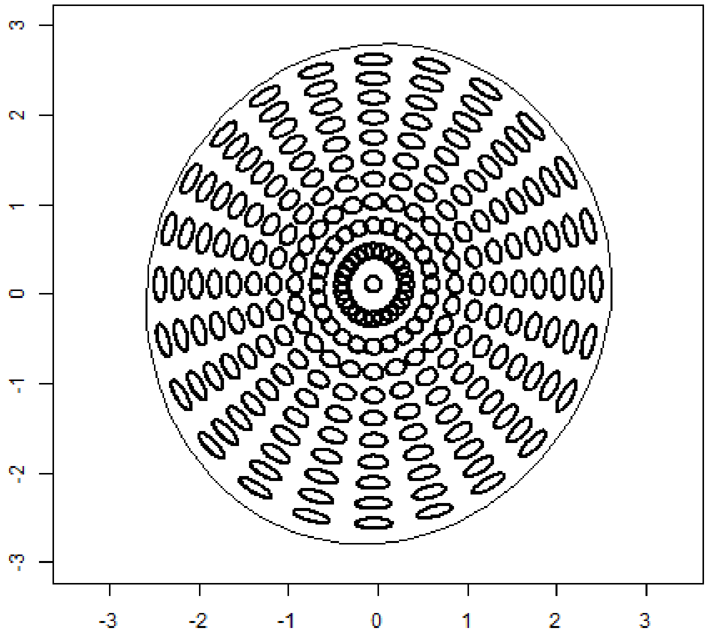
$$u \equiv 0$$

$$Lf = -div(K\nabla f) + \mathbf{b}\cdot\nabla s + cs$$

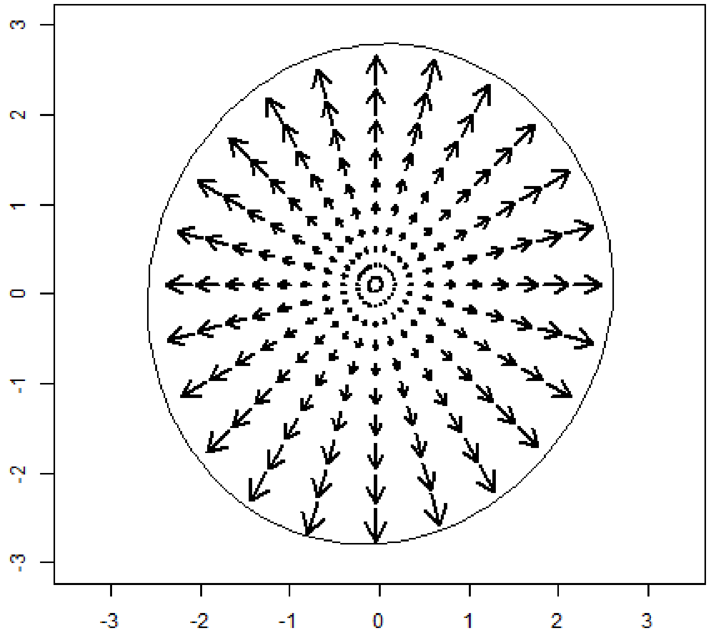
Reaction term: *shrinking effect*

$$c = 0$$

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



Transport vector field: *directional smoothing* that smooths the observations along the radial direction



- $\boldsymbol{z} = (z_1, \ldots, z_n)^t$
- $\{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_K\}$: nodes of $\mathcal{T}$
- $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_K)^t$: finite element basis $\quad \Psi = \{\Psi\}_{ij} := \psi_j(\mathbf{p}_i)$
- for any g in the finite element space, $g = \mathbf{g}^t \boldsymbol{\psi}$ where $\mathbf{g} := (g(\boldsymbol{\xi}_1), \ldots, g(\boldsymbol{\xi}_K))^t$
- $R = \{R\}_{jk} := \int_{\Omega_\mathcal{T}} (\psi_j {\psi_k}^t) \quad A = \{A\}_{jk} := \int_{\Omega_\mathcal{T}} (\mathbf{K}\nabla\psi_j \cdot \nabla\psi_k + \mathbf{b} \cdot \nabla\psi_j \psi_k + c\,\psi_j\psi_k)$

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

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

(Here for simplicity: pointwise case, $u \equiv 0$, homogeneous Neumann b.c.)

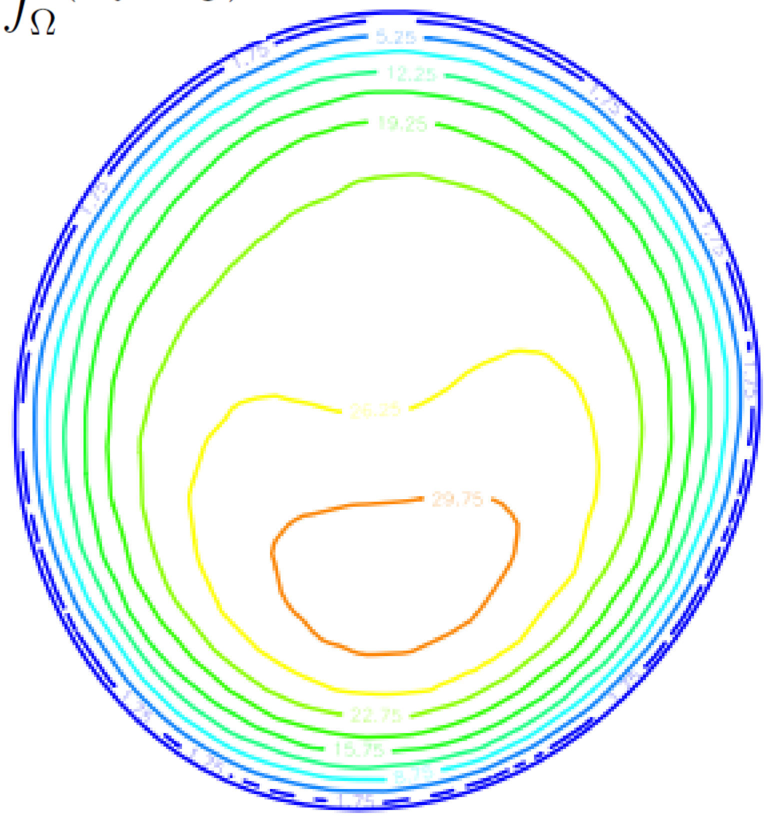
$\hat{f}$ is *linear* in $\mathbf{z}$ and has typical penalized regression form:

$$\mathbf{f}_n = (\Psi^t \Psi + \lambda P)^{-1} \Psi^t \mathbf{z}$$

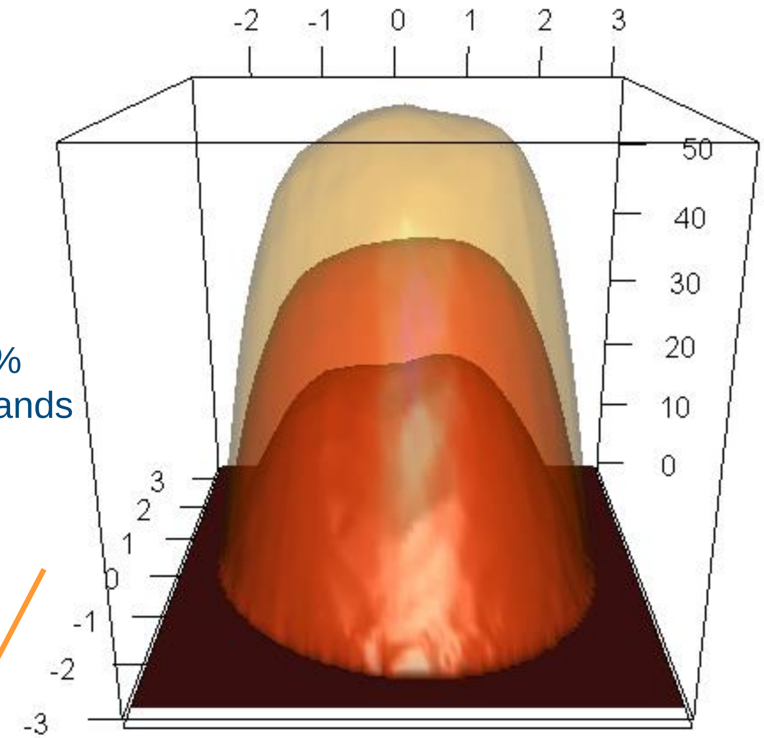
$$P = A^t \, R^{-1} \, A \qquad \text{is discretization of penalty}$$

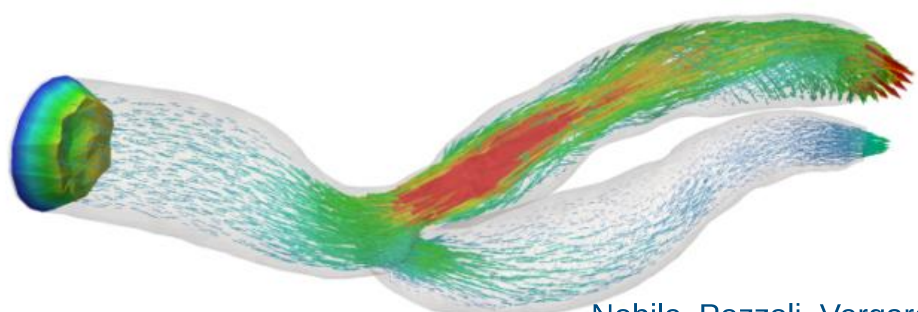
- Classical inferential tools are readily derived
  - mean and variance of $\hat{f}$
  - confidence bands for f
  - prediction intervals for new observations
  - estimate of error variance σ^2
  - selection of smoothing parameter λ via generalized cross validation

$$\int_{\Omega}(Lf-g)^2$$



pointwise 95% confidence bands





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

Relevant features: eccentricity, reversion of fluxes

Nobile, Pozzoli, Vergara
Patient-specific inflow conditions for Computation Fluid-Dynamics

Irregularly shaped domains and boundary conditions

- Sangalli L.M., Ramsay J.O., Ramsay T.O. (2013), "Spatial splines regression models", *Journal of the Royal Statistical Society Ser. B, Statistical Methodology*, 75, 4, 681-703.
- Ramsay J.O., Ramsay T.O., Sangalli L.M. (2011), "Spatial Functional Data Analysis", in Recent Advances in Functional Data Analysis and Related Topics, Springer Series Contribution to Statistics, pp. 269--276.

Incorporating a priori knowledge

- Azzimonti L., "Blood flow velocity field estimation via spatial regression with PDE penalization", PhD Thesis, Politecnico di Milano, 2013.
- Azzimonti L., Sangalli L.M., Secchi P., Domanin M. and Nobile F. (2013), "Blood flow velocity field estimation via spatial regression with PDE penalization", Tech.rep N. 19/2013, MOX, Dip. di Matematica, Politecnico di Milano.
- Azzimonti L., Nobile F., Sangalli L.M., Secchi P. (2013), "Mixed Finite Elements for spatial regression with PDE penalization", Tech.rep N. 20/2013, MOX, Dip. di Matematica, Politecnico di Milano.

Manifold domains

- Ettinger B., Perotto S., Sangalli L.M. (2012), "Regression models over two-dimensional manifolds", Tech.rep N. 54/2012, MOX, Dip. di Matematica, Politecnico di Milano.
- Ettinger B., Passerini T., Perotto S., Sangalli L.M. (2013), "Spatial smoothing for data distributed over non-planar domains", in Complex Models and Computational Methods in Statistics, Springer Series Contribution to Statistics, pp. 123-136.
- Dassi F., Ettinger B., Perotto S., Sangalli L.M. (2013), " Mesh simplification for spatial regression of cortical surface data", Tech.rep N. 31/2013, MOX, Dip. di Matematica, Politecnico di Milano.

Other references

- Baramidze, V., Lai,M. J., and Shum, C. K. (2006), "Spherical splines for data interpolation and fitting," *SIAM J. Sci. Comput.*, 28, 241–259.
- Chung, M. K., Robbins, S. M., Dalton, K. M., Davidson, R. J., Alexander, A. L., and Evans, A. C. (2005), "Cortical thickness analysis in autism with heat kernel smoothing," *NeuroImage*, 25, 1256–1265.
- Gneiting, T. (2013), Strictly and non-strictly positive definite functions on spheres, Bernoulli, 19, 4, 1087-1500
- Guillas, S. and Lai, M. (2010), "Bivariate splines for spatial functional regression models," *J. Nonparametric Stat*, 22, 477–497.
- Hagler, Jr., D. J., Saygin, A. P., and Sereno, M. I. (2006), "Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data," *NeuroImage*, 33, 1093–1103.
- Haker, S., Angenent, S., Tannenbaum, A., and Kikinis, R. (2000), "Nondistorting flattening maps and the 3-D visualization of colon CT images," *IEEE Trans. Med. Imag.*, 19, 665–670.
- Lindgren, F., Rue, H., and Lindstr ¨om, J. (2011), "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73, 423–498.
- Passerini, T., Sangalli, L. M., Vantini, S., Piccinelli, M., Bacigaluppi, S., Antiga, L., Boccardi, E., Secchi, P., and Veneziani, A. (2012), "An integrated CFD-statistical investigation of parent vasculature of cerebral aneurysms," *Cardio. Eng. and Tech.*, 3, 26–40.
- Ramsay. Spline smoothing over difficult regions. J. R. Stat. Soc. Ser. B Stat. Methodol., 64, 307–319, 2002.
- Stone, G. (1988). Bivariate splines. PhD thesis, University of Bath, Bath.
- Stuart, A. (2010), "Inverse problems: a Bayesian perspective," *Acta Numerica*, 19, 451–559.
- Wahba, G. (1990). Spline models for observational data. Society for Industrial and Applied Mathematics.
- Wahba, G. (1981), "Spline interpolation and smoothing on the sphere," *SIAM J. Sci. Stat. Comput.*, 2, 5–16.
- Wood, S., Bravington, M., and Hedley, S. (2008), "Soap film smoothing," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 70, 931–955.

Acknowledgments:

FUTURO IN RICERCA SNAPLE

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

Dote Ricercatore Politecnico di Milano – Regione Lombardia: Research Project (2010 - 2013) "Functional data analysis for life sciences".

RegioneLombardia

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