Physics-informed spatial and functional data analysis over non-Euclidean domains

by Laura Sangalli, MOX, Politecnico di Milano

Discussant: Ilaria Prosdocimi

Ca' Foscari University of Venice

Università Ca'Foscari Venezia

A busy landscape

Laura’s work has similarities with other approaches, among others:

A busy landscape

Laura's work has similarities with other approaches, among others:

- Generalised Additive Models and extensions:
    - Simon Wood's work implemented in the `mgcv` R package
    - The Bayesian perspective of structured additive distributional regression models implemented in the `BayesX` and `bamlss` software
    - Many more approaches in this landscape (boosting, GAMLSS, etc.)

A busy landscape

Laura's work has similarities with other approaches, among others:

- Generalised Additive Models and extensions:
    - Simon Wood's work implemented in the `mgcv` R package
    - The Bayesian perspective of structured additive distributional regression models implemented in the `BayesX` and `bamlss` software
    - Many more approaches in this landscape (boosting, GAMLSS, etc.)
- Latent Gaussian models represented as Gaussian Markov Random Field (GMRF) made popular in applications by the `R-INLA` package by Håvard Rue

A busy landscape

Laura's work has similarities with other approaches, among others:

- Generalised Additive Models and extensions:
    - Simon Wood's work implemented in the `mgcv` R package
    - The Bayesian perspective of structured additive distributional regression models implemented in the `BayesX` and `bamlss` software
    - Many more approaches in this landscape (boosting, GAMLSS, etc.)
- Latent Gaussian models represented as Gaussian Markov Random Field (GMRF) made popular in applications by the `R-INLA` package by Håvard Rue

Some common traits: smart use of numerical methods; very stable and usable software

What is a penalty? and why?

"Penalised" estimation: maximise

$$l(\theta; y) - \lambda \text{pen}(\theta)$$

Why would you do that?

What is a penalty? and why?

"Penalised" estimation: maximise

$$l(\theta; y) - \lambda \text{pen}(\theta)$$

Why would you do that?

- Regularising estimation: in regression multi-collinearity, $n < p$, variable selection, . . .

What is a penalty? and why?

"Penalised" estimation: maximise

$$l(\theta; y) - \lambda \text{pen}(\theta)$$

Why would you do that?

- Regularising estimation: in regression multi-collinearity, $n < p$, variable selection, ...

  Ridge Regression (Tikhonov regularization), Lasso, Elastic Net, ...

What is a penalty? and why?

"Penalised" estimation: maximise

$$l(\theta; y) - \lambda \mathrm{pen}(\theta)$$

Why would you do that?

- Regularising estimation: in regression multi-collinearity, $n < p$, variable selection, ...

  Ridge Regression (Tikhonov regularization), Lasso, Elastic Net, ...
- Setting constraints on some of the parameters: eg smoothness of functions in GAMs

What is a penalty? and why?

"Penalised" estimation: maximise

$$l(\theta; y) - \lambda \text{pen}(\theta)$$

Why would you do that?

- Regularising estimation: in regression multi-collinearity, $n < p$, variable selection, ...

  Ridge Regression (Tikhonov regularization), Lasso, Elastic Net, ...
- Setting constraints on some of the parameters: eg smoothness of functions in GAMs
- Setting a prior on the parameters: eg Bayesian GAMs

What is a penalty? and why?

"Penalised" estimation: maximise

$$l(\theta; y) - \lambda \text{pen}(\theta)$$

Why would you do that?

- Regularising estimation: in regression multi-collinearity, $n < p$, variable selection, ...

  Ridge Regression (Tikhonov regularization), Lasso, Elastic Net, ...
- Setting constraints on some of the parameters: eg smoothness of functions in GAMs
- Setting a prior on the parameters: eg Bayesian GAMs
- Include physics of the problem as in SR-PDE

What is a penalty? and why?

"Penalised" estimation: maximise

$$l(\theta; y) - \lambda \text{pen}(\theta)$$

Why would you do that?

- Regularising estimation: in regression multi-collinearity, $n < p$, variable selection, ...

  Ridge Regression (Tikhonov regularization), Lasso, Elastic Net, ...
- Setting constraints on some of the parameters: eg smoothness of functions in GAMs
- Setting a prior on the parameters: eg Bayesian GAMs
- Include physics of the problem as in SR-PDE

(Are these really different things?)

Not all penalties are equal

In the SR-PDE model:

$$\text{pen}(\theta) = \int_{\mathcal{D}} (Lf - u)^2 d\mathbf{p}$$

with

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

The K, $\mathbf{b}$ and c term encapsulate some understanding of the physics of the problem.

Not all penalties are equal

In the SR-PDE model:

$$\text{pen}(\theta) = \int_{\mathcal{D}} (Lf - u)^2 d\mathbf{p}$$

with

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

The K, $\mathbf{b}$ and c term encapsulate some understanding of the physics of the problem.

Could we turn this *problem-specific prior information* into a Bayesian prior?

Not all penalties are equal

In the SR-PDE model:

$$\text{pen}(\theta) = \int_{\mathcal{D}} (Lf - u)^2 d\mathbf{p}$$

with

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

The K, $\mathbf{b}$ and c term encapsulate some understanding of the physics of the problem.

Could we turn this *problem-specific prior information* into a Bayesian prior?

If so, we could specify strong or weak prior depending on our understanding?

Not all penalties are equal

In the SR-PDE model:

$$\text{pen}(\theta) = \int_{\mathcal{D}} (Lf - u)^2 d\mathbf{p}$$

with

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

The K, $\mathbf{b}$ and c term encapsulate some understanding of the physics of the problem.

Could we turn this *problem-specific prior information* into a Bayesian prior?

If so, we could specify strong or weak prior depending on our understanding?

How hard is it to derive/elicit the components of the penalty term?

Model selection

For some special cases of K, $\mathbf{b}$ and c we find back other well known penalities/priors/models.

Beside the isotropic iid case, can we retrieve some known spatial models as special cases of the SR-PDE penalty?

Model selection

For some special cases of K, $\mathbf{b}$ and c we find back other well known penalities/priors/models.

Beside the isotropic iid case, can we retrieve some known spatial models as special cases of the SR-PDE penalty?

Furthermore: Can SR-PDE perform a *model selection* step?

Model selection

For some special cases of K, $\mathbf{b}$ and c we find back other well known penalities/priors/models.

Beside the isotropic iid case, can we retrieve some known spatial models as special cases of the SR-PDE penalty?

Furthermore: Can SR-PDE perform a *model selection* step?

Can we use it to assess the (relative) importance of say, anisotropy, in a system?

Model selection

For some special cases of K, $\mathbf{b}$ and c we find back other well known penalities/priors/models.

Beside the isotropic iid case, can we retrieve some known spatial models as special cases of the SR-PDE penalty?

Furthermore: Can SR-PDE perform a *model selection* step?

Can we use it to assess the (relative) importance of say, anisotropy, in a system?

Or at least quantify what we loose when the physics-informed term is not included.

Model selection and uncertainty

Model selection is related to uncertainty in the estimates. There results on the variance of SR-PDE estimators: it depends on λ and the FEM bases $\psi_j(\mathbf{p})$.

In the spirit of Wood Pya and Säfken (2016, JASA): can we enhance the variance estimation to take into account the uncertainty in λ and $\psi_j(\mathbf{p})$?

In general: can we think of general approaches to assess uncertainties derived from modelling choices?

The more the merrier

SR-PDE is an excellent addition to a growing class of flexible models built with applications in mind: co-development with the end user who can encode understanding of the system.

The more the merrier

SR-PDE is an excellent addition to a growing class of flexible models built with applications in mind: co-development with the end user who can encode understanding of the system.

Statistics as one of the methods to solve real world problems.

The more the merrier

SR-PDE is an excellent addition to a growing class of flexible models built with applications in mind: co-development with the end user who can encode understanding of the system.

Statistics as one of the methods to solve real world problems.

Statistics employing methods developed in other fields: we can all learn from cross-contamination.