Spatio-temporal random fields on meshed surfaces defined from advection-diffusion SPDEs: application to environmental data

Lucia Clarotto[1]
Joint work with
M. Pereira[2] and N. Desassis[2]

[1]MIA Paris-Saclay, AgroParisTech, INRAE
Université Paris Saclay
[2]Department of Geosciences and Geoengineering
Mines Paris – PSL University

Journée Thématique IA et Environnement
June 11, 2024

AgroParisTech

GEOSTATISTICAL SPATIO-TEMPORAL MODELING

Geostatistical paradigm: over the spatial domain $\mathcal{D}$ and a time segment $[0, T]$

Gaussian Random Field		Observed variable
$Z : \{Z(t, s) : s \in \mathcal{D},\ t \in [0, T]\}$	Realization $\longrightarrow$	$z : \{z(t, s) : s \in \mathcal{D},\ t \in [0, T]\}$
High correlation		High "similarity"

- Allows to model data which are not independent, identically distributed
- **Covariance function** C_{ST}:

$$C_{ST}(h, u) = \mathrm{Cov}[Z(s, t), Z(s', t')] = \mathrm{Cov}[Z(s, t), Z(s + h, t + u)]$$

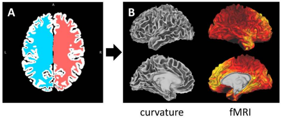
with $(s, t), (s', t') \in \mathbb{R}^d \times \mathbb{R}$ and $h = ||s - s'||$, $u = |t - t'|$

$\rightarrow$ used to model the spatio-temporal structure observed on the variable/data
$\rightarrow$ its matrix form must be inverted for estimation via maximum likelihood and for prediction

THREE CHALLENGES OF COVARIANCE-BASED APPROACH

Random fields on non-Euclidean domains

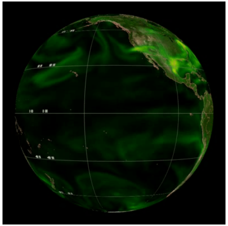
- Extensive literature for the sphere only: Marinucci and Peccati (2011), Lang et al. (2015), Lantuéjoul et al. (2019), Emery and Porcu (2019)



(Source: Mejia et al. (2020))

Physical phenomena and non-stationarity

- Focus on Euclidean domains and on the sphere: Porcu et al. (2021; 2018), Chen et al. (2021)



(Source: NASA)

Big "N" problem

- Need to restrict the choice of models to work with **sparse matrices**: Compactly-supported or tapered covariance functions (Gneiting, 2002, Furrer et al., 2006), Markovian models (Rue and Held, 2005)

A FIRST SOLUTION: THE SPDE APPROACH

Whittle (1954; 1963), Rozanov (1977)

Given the **spatial SPDE** ($s \in \mathbb{R}^d$)

$$(\kappa^2 - \Delta)^{\alpha/2} \mathcal{Z}(s) = \tau \mathcal{W}(s)$$

- $\mathcal{W}$ is a standard Gaussian white noise
- $\alpha > d/2, \ \tau > 0$

A FIRST SOLUTION: THE SPDE APPROACH

Whittle (1954; 1963), Rozanov (1977)

Given the **spatial SPDE** ($s \in \mathbb{R}^d$)

$$(\kappa^2 - \Delta)^{\alpha/2} \mathcal{Z}(s) = \tau \mathcal{W}(s)$$

- $\mathcal{W}$ is a standard Gaussian white noise
- $\alpha > d/2, \ \tau > 0$

the solution $\mathcal{Z}(s)$ is a Gaussian random field (GRF) with **Matérn covariance function**

$$C^{\mathcal{M}}_{\nu,a}(h) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{h}{a}\right)^{\nu} \mathcal{K}_\nu \left(\frac{h}{a}\right), \quad h = ||s - s'||$$

- smoothness $\nu = \alpha - d/2$
- scale parameter $a = 1/\kappa$
- variance $\sigma^2 = \tau^2 (4\pi)^{-d/2} \Gamma(\nu) \Gamma(\nu + d/2)^{-1} \kappa^{-2\nu}$
- $\mathcal{K}_\nu$ is the modified 2^{nd} order Bessel function

■ SOLUTIONS TO CHALLENGES

$$(\kappa^2 - \Delta)^{\alpha/2}\mathcal{Z} = \mathcal{W}$$

SPDE approach: Lindgren et al. (2011) use this last characterization

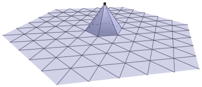
Problem	Solution proposed
Non-Euclidean domains, Non-stationarity	Define the SPDE over a more adequate domain than $\mathbb{R}^d$ (Riemannian manifold)

■ SOLUTIONS TO CHALLENGES

$$(\kappa^2 - \Delta)^{\alpha/2} \mathcal{Z} = \mathcal{W}$$

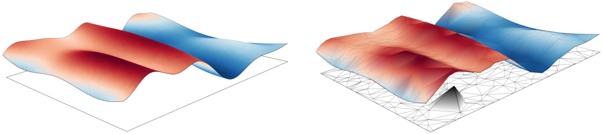
SPDE approach: Lindgren et al. (2011) use this last characterization

Problem	**Solution proposed**
Big “N” problem	Use the finite element method to solve the SPDE



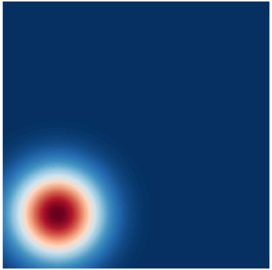
$\mathcal{D}$

$$\rightarrow \quad Z(\boldsymbol{p}) = \sum_{j=1}^{n} Z_j \psi_j(\boldsymbol{p}) \quad \rightarrow$$



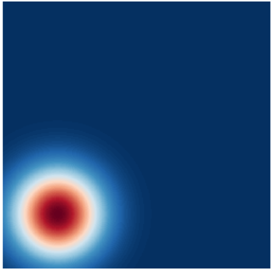
Lindgren et al. (2011) show that the weights $\boldsymbol{Z} = (Z_1, \dots, Z_n)$ form a Gaussian vector with **sparse precision matrix**

■ PHYSICAL PHENOMENA



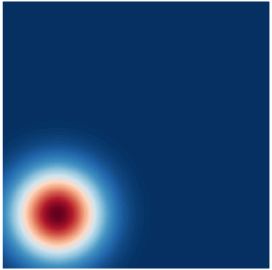
Advection

$$\frac{\partial z}{\partial t} + \gamma \cdot \nabla z = 0$$



Diffusion

$$\frac{\partial z}{\partial t} - \Delta z = 0$$



Advection + Diffusion

$$\frac{\partial z}{\partial t} + \gamma \cdot \nabla z - \Delta z = 0$$

SPATIO-TEMPORAL SPDE APPROACH

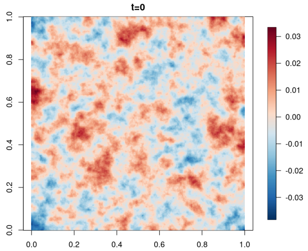
On **Euclidean domains**, consider the **advection-diffusion** SPDE (Liu et al., 2022, Sigrist et al., 2015, Clarotto et al., 2022)

$$\frac{\partial \mathcal{Z}}{\partial t} + \frac{1}{c}\left((\kappa^2 - \Delta)^\alpha \mathcal{Z} + \gamma \cdot \nabla \mathcal{Z}\right) = \frac{\tau}{\sqrt{c}} \mathcal{W}_T \otimes \mathcal{Y}_S,$$

where

- γ is a vector defining the advection direction
- $\mathcal{W}_T \otimes \mathcal{Y}_S$ is a space-time separable stochastic forcing (noise white in time and colored in space)

$$(\kappa^2 - \Delta)^{\alpha_S/2} \mathcal{Y}_S = \mathcal{W}_S$$

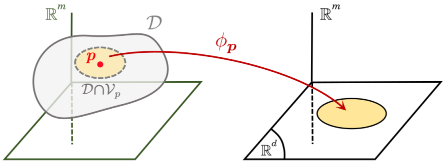


DEFINITION: RIEMANNIAN MANIFOLDS

Let $m \geq 1$ and $1 \leq d \leq m$

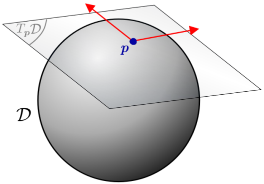
$(\mathcal{M}, g)$ is a **compact Riemannian (sub)manifold of dimension** 2

- $\mathcal{M} \subset \mathbb{R}^m$ is a **smooth (sub)manifold**
  - Locally Euclidean of dimension d
  - Can be entirely mapped by a set of smoothly *compatible* charts



Ex: Euclidean domains, smooth surfaces (eg. sphere, torus,...)

- $\mathcal{M}$ is equipped with a **Riemannian metric** g
  - $g_{\boldsymbol{p}}$: inner product on the tangent space of $\mathcal{M}$ at $\boldsymbol{p} \in \mathcal{D}\mathcal{M}$
  - $g : \boldsymbol{p} \mapsto g_{\boldsymbol{p}}$ is "smooth"



Lengths and angles of tangent vectors $\boldsymbol{u}, \boldsymbol{v}$:

$$\|\boldsymbol{u}\|_{\boldsymbol{p}} = \sqrt{g_{\boldsymbol{p}}(\boldsymbol{u}, \boldsymbol{u})}$$

$$\cos\left(\theta(\boldsymbol{u}, \boldsymbol{v})\right) = \frac{g_{\boldsymbol{p}}(\boldsymbol{u}, \boldsymbol{v})}{\|\boldsymbol{u}\|_{\boldsymbol{p}}\|\boldsymbol{v}\|_{\boldsymbol{p}}}$$

ADVECTION-DIFFUSION SPDE ON RIEMANNIAN MANIFOLD

On a compact smooth orientable Riemannian manifold $(\mathcal{M}, g)$ of dimension 2

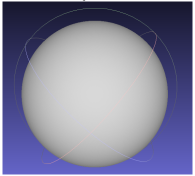
$$\frac{\partial \mathcal{Z}}{\partial t} + \frac{1}{c}\left((\kappa^2 - \Delta_{\mathcal{M}})^{\alpha}\mathcal{Z} + \mathrm{div}_{\mathcal{M}}(\gamma \mathcal{Z})\right) = \frac{\tau}{\sqrt{c}}\mathcal{W}_T \otimes \mathcal{Y}_S,$$

where

- $-\Delta_{\mathcal{M}}$ is the Laplace–Beltrami operator and $\mathrm{div}_{\mathcal{M}}$ the divergence operator on $(\mathcal{M}, g)$
- $\mathcal{W}_T \otimes \mathcal{Y}_S$ is a noise white in time, colored in space
- $s \in \mathcal{M} \mapsto \gamma(s)$ is a smooth tangent vector field

■ ADVECTION-DIFFUSION ON MESHED SURFACE

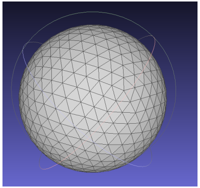
$\mathcal{M}$



$$\frac{\partial \mathcal{Z}}{\partial t} + \frac{1}{c}\left((\kappa^2 - \Delta_{\mathcal{M}})^{\alpha}\mathcal{Z} + \mathrm{div}_{\mathcal{M}}(\gamma \mathcal{Z})\right) = \frac{\tau}{\sqrt{c}}\mathcal{W}_T \otimes \mathcal{Y}_S$$

Triangulation of the surface (mesh size h)
+ Galerkin approximation
+ Implicit Euler

$\mathcal{M}_h$

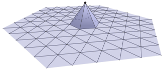


$\boldsymbol{C}$ (diagonal mass lumped matrix), $\boldsymbol{R}$ (stiffness matrix), $\boldsymbol{B}$ (advection matrix)

$$C_{ij} = \langle \psi_i, 1 \rangle \delta_{ij}, \quad R_{ij} = \langle \nabla \psi_i, \nabla \psi_j \rangle, \quad B_{ij} = \langle \psi_i, \mathrm{div}(\gamma \psi_j) \rangle, \quad 1 \leq i, j \leq N_h$$

For $k \in \mathbb{N}$, we have the recursion

$$\left(\boldsymbol{C} + \boldsymbol{C}^{1/2}(\kappa^2 \boldsymbol{I} + \widetilde{\boldsymbol{R}})\boldsymbol{C}^{1/2} + \boldsymbol{B}\right)\boldsymbol{z}^{(k+1)} = \boldsymbol{C}\left(\boldsymbol{z}^{(k)} + \tau\sqrt{\frac{\delta t}{c}}\boldsymbol{y}_S^{(k+1)}\right)$$



with $\widetilde{\boldsymbol{R}} = \boldsymbol{C}^{-1/2}\boldsymbol{R}\boldsymbol{C}^{-1/2}$ and $\{\boldsymbol{y}_S^{(k)}\}_{k \in \mathbb{N}}$ i.i.d. realizations of the spatial noise $\mathcal{Y}_S$

PRECISION MATRIX OF THE FIELD

$$\boldsymbol{\Gamma} \boldsymbol{z}^{(k+1)} = \boldsymbol{C} \left(\boldsymbol{z}^{(k)} + \tau \sqrt{\frac{\delta t}{c}} \boldsymbol{y}_S^{(k+1)} \right), \quad \text{with} \quad \boldsymbol{\Gamma} = \left(\boldsymbol{C} + \boldsymbol{C}^{1/2} (\kappa^2 \boldsymbol{I} + \widetilde{\boldsymbol{R}}) \boldsymbol{C}^{1/2} + \boldsymbol{B} \right)$$

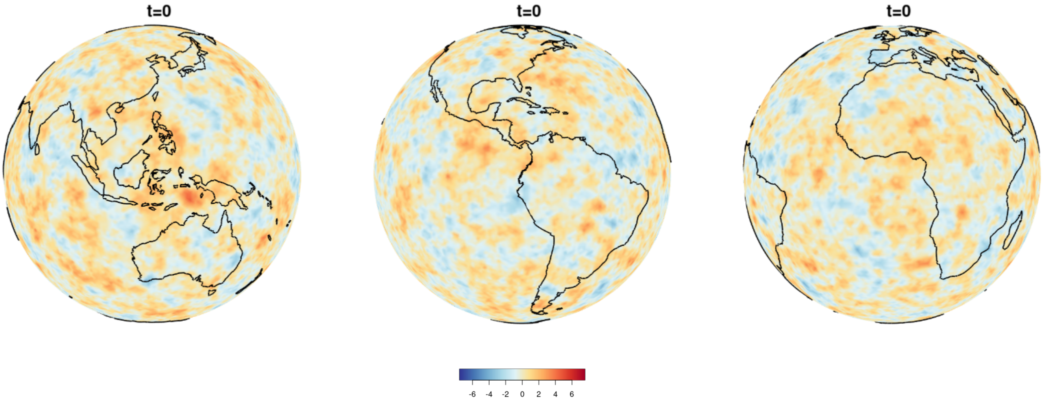
Then the precision matrix of $\boldsymbol{Z} = (\boldsymbol{z}^{(0)}, \ldots, \boldsymbol{z}^{(K)})$ is given by

$$\boldsymbol{Q}_{\boldsymbol{Z}} = \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{C} & & \\ & \boldsymbol{\Gamma}^T & \ddots & \\ & & \ddots & -\boldsymbol{C} \\ & & & \boldsymbol{\Gamma}^T \end{pmatrix} \begin{pmatrix} \boldsymbol{Q}_0 & & & \\ & \boldsymbol{Q}_S & & \\ & & \ddots & \\ & & & \boldsymbol{Q}_S \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & & & \\ -\boldsymbol{C} & \boldsymbol{\Gamma} & & \\ & \ddots & \ddots & \\ & & -\boldsymbol{C} & \boldsymbol{\Gamma} \end{pmatrix}$$

where $\boldsymbol{Q}_S$ (resp. $\boldsymbol{Q}_0$) is the precision matrix of colored noise vector $\boldsymbol{y}_S$ (resp. initial condition $\boldsymbol{z}^{(0)}$)

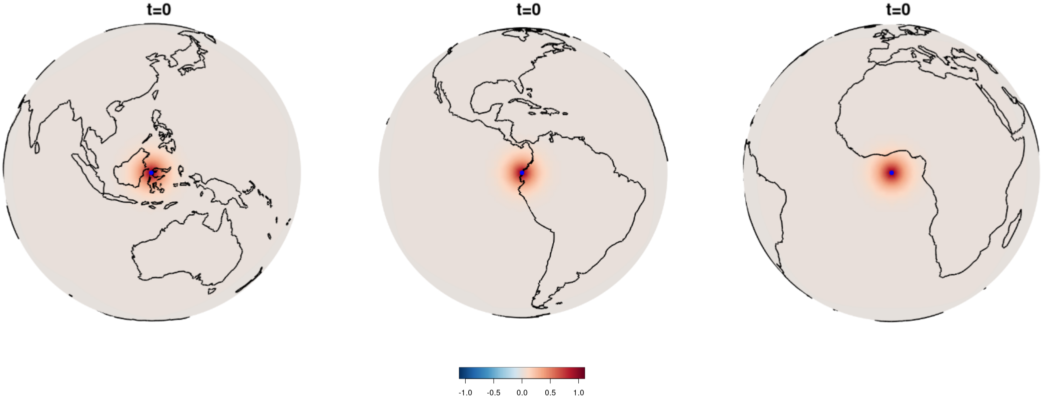
$\rightarrow$ (Time) scalable algorithms to compute products by $\boldsymbol{Q}_{\boldsymbol{Z}}$ and $\boldsymbol{Q}_{\boldsymbol{Z}}^{-1}$ because of **sparsity**

■ SIMULATION ON THE SPHERE



Numerical solution of the SPDE

COVARIANCE ON THE SPHERE



Spatio-temporal evolution of the covariance between three reference points (blue) and the other points of the domain

PREDICTION BY KRIGING

Data At each time step t_k of the time interval $[0, T]$, we have n_k observations $Y(t_k, s_i^{(k)})$ of a variable Y at some points $s_1^{(k)}, \dots, s_{n_k}^{(k)} \in \mathcal{M}_h$

Model We assume that

$$Y(t_k, s_i^{(k)}) = \boldsymbol{A}^{(k)} Z(t_k, s_i^{(k)}) + \sigma\varepsilon_i, \quad 1 \leq i \leq n_k$$

where $\varepsilon_i \sim \mathcal{N}(0, 1)$, $\boldsymbol{A}^{(k)}$ is the spatial observation matrix, and Z is a solution of the advection-diffusion SPDE on $\mathcal{M}_h$
Equivalently, $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z} + \sigma\boldsymbol{\varepsilon}$

Goal Predictions $Z^*(t_k, \cdot)$ of the value of Z at each time step

PREDICTION BY KRIGING

Data At each time step t_k of the time interval $[0, T]$, we have n_k observations $Y(t_k, s_i^{(k)})$ of a variable Y at some points $s_1^{(k)}, \dots, s_{n_k}^{(k)} \in \mathcal{M}_h$

Model We assume that

$$Y(t_k, s_i^{(k)}) = \boldsymbol{A}^{(k)} Z(t_k, s_i^{(k)}) + \sigma\varepsilon_i, \quad 1 \leq i \leq n_k$$

where $\varepsilon_i \sim \mathcal{N}(0,1)$, $\boldsymbol{A}^{(k)}$ is the spatial observation matrix, and Z is a solution of the advection-diffusion SPDE on $\mathcal{M}_h$

Equivalently, $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z} + \sigma\boldsymbol{\varepsilon}$

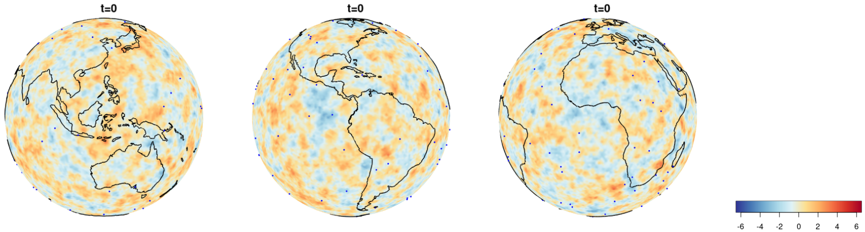
Goal Predictions $Z^*(t_k, \cdot)$ of the value of Z at each time step

Answer: Conditional expectation (a.k.a. Kriging prediction)

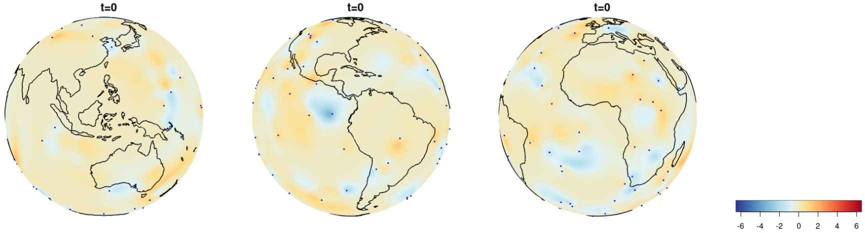
$$\boldsymbol{Z}^* = \mathbb{E}[\boldsymbol{Z} | \boldsymbol{Y}] = \sigma^{-2}(\boldsymbol{Q_Z} + \sigma^{-2}\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{Y}$$

$\rightarrow$ Solved using generalized least-square or saddle point algorithms

■ EXAMPLE: PREDICTION ON THE SPHERE



Simulation from which data are sampled



Kriging predictions

INFERENCE

Model Vector of observations $\boldsymbol{Y}$ modeled as

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z} + \sigma\boldsymbol{\varepsilon}$$

Goal Estimation of parameters of the SPDE $\boldsymbol{\theta} = (\kappa, \tau, c, \gamma)$

Answer: Let define $\boldsymbol{\Sigma}_{\boldsymbol{Y}}(\boldsymbol{\theta}) = \boldsymbol{A}\boldsymbol{Q}_{\boldsymbol{Z}}^{-1}(\boldsymbol{\theta})\boldsymbol{A}^T + \sigma^2\boldsymbol{I}$

Then, $\boldsymbol{Q}_{\boldsymbol{Y}} = \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} = \sigma^{-2}\boldsymbol{I} - \sigma^{-4}\boldsymbol{A}(\boldsymbol{Q}_{\boldsymbol{Z}} + \sigma^{-2}\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$

Log-likelihood

$$\mathcal{L}(\boldsymbol{\theta}) = \log|\boldsymbol{Q}_{\boldsymbol{Y}}(\boldsymbol{\theta})| - \boldsymbol{Y}^T\boldsymbol{Q}_{\boldsymbol{Y}}(\boldsymbol{\theta})\boldsymbol{Y} + \text{Constant}$$

INFERENCE

Model Vector of observations $\boldsymbol{Y}$ modeled as

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z} + \sigma\varepsilon$$

Goal Estimation of parameters of the SPDE $\boldsymbol{\theta} = (\kappa, \tau, c, \gamma)$

Answer: Let define $\boldsymbol{\Sigma}_{\boldsymbol{Y}}(\boldsymbol{\theta}) = \boldsymbol{A}\boldsymbol{Q}_{\boldsymbol{Z}}^{-1}(\boldsymbol{\theta})\boldsymbol{A}^T + \sigma^2\boldsymbol{I}$

Then, $\boldsymbol{Q}_{\boldsymbol{Y}} = \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} = \sigma^{-2}\boldsymbol{I} - \sigma^{-4}\boldsymbol{A}(\boldsymbol{Q}_{\boldsymbol{Z}} + \sigma^{-2}\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$

Log-likelihood

$$\mathcal{L}(\boldsymbol{\theta}) = \log|\boldsymbol{Q}_{\boldsymbol{Y}}(\boldsymbol{\theta})| - \boldsymbol{Y}^T\boldsymbol{Q}_{\boldsymbol{Y}}(\boldsymbol{\theta})\boldsymbol{Y} + \text{Constant}$$

$\rightarrow$ Solved again by matrix-free approach (with saddle point algorithms)

INFERENCE

Model Vector of observations $\boldsymbol{Y}$ modeled as

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z} + \sigma\varepsilon$$

Goal Estimation of parameters of the SPDE $\boldsymbol{\theta} = (\kappa, \tau, c, \gamma)$

Answer: Let define $\boldsymbol{\Sigma}_{\boldsymbol{Y}}(\boldsymbol{\theta}) = \boldsymbol{A}\boldsymbol{Q}_{\boldsymbol{Z}}^{-1}(\boldsymbol{\theta})\boldsymbol{A}^T + \sigma^2\boldsymbol{I}$

Then, $\boldsymbol{Q}_{\boldsymbol{Y}} = \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} = \sigma^{-2}\boldsymbol{I} - \sigma^{-4}\boldsymbol{A}(\boldsymbol{Q}_{\boldsymbol{Z}} + \sigma^{-2}\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$

Log-likelihood

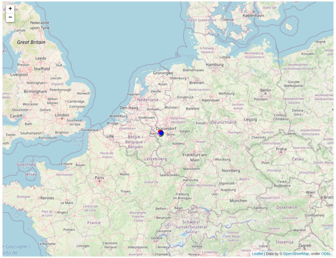
$$\mathcal{L}(\boldsymbol{\theta}) = \log|\boldsymbol{Q}_{\boldsymbol{Y}}(\boldsymbol{\theta})| - \boldsymbol{Y}^T\boldsymbol{Q}_{\boldsymbol{Y}}(\boldsymbol{\theta})\boldsymbol{Y} + \text{Constant}$$

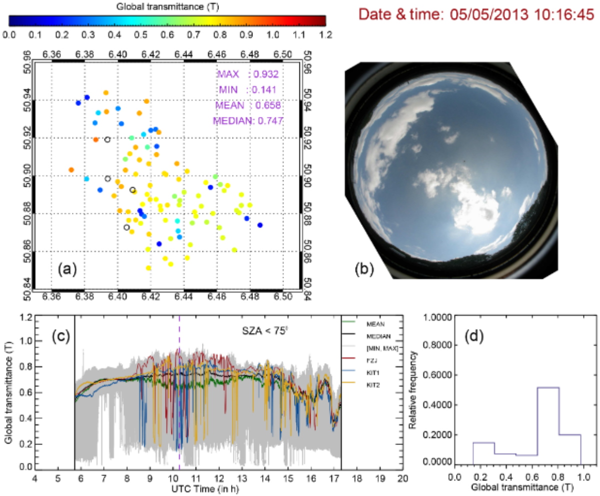
$$\log|\boldsymbol{Q}_{\boldsymbol{Y}}(\boldsymbol{\theta})| = -n\log\sigma^2 + \log|\boldsymbol{Q}_{\boldsymbol{Z}}(\boldsymbol{\theta})| - \log|\boldsymbol{Q}_{\boldsymbol{Z}}(\boldsymbol{\theta}) + \sigma^{-2}\boldsymbol{A}^T\boldsymbol{A}|$$

$\rightarrow$ Hutchinson estimator (Hutchinson, 1989)

$$\log|h(\boldsymbol{B})| = \mathrm{Trace}(\log h(\boldsymbol{B})) = \mathbb{E}[\boldsymbol{W}^T\log h(\boldsymbol{B})\boldsymbol{W}], \quad \boldsymbol{W} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

■ APPLICATION TO SOLAR RADIATION DATA





■ APPLICATION TO SOLAR RADIATION DATA

Data

99 photodiodes measuring **Solar Radiation** G every minute in a region in Germany (HOPE campaign, Macke et al., 2017)

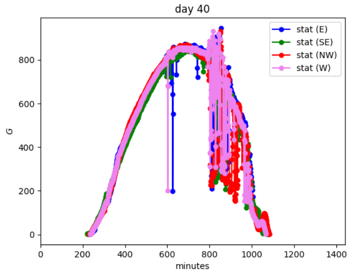
Goal

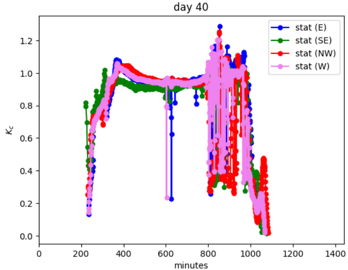
Create a complete spatio-temporal cartography over a fine spatial domain for short-term prediction

$$G = G_c \, K_c$$

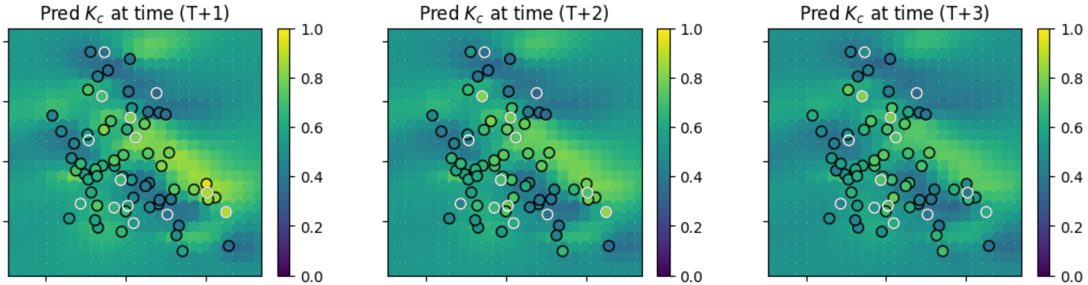
irradiance under clear sky — physics

clear sky index — **statistics**





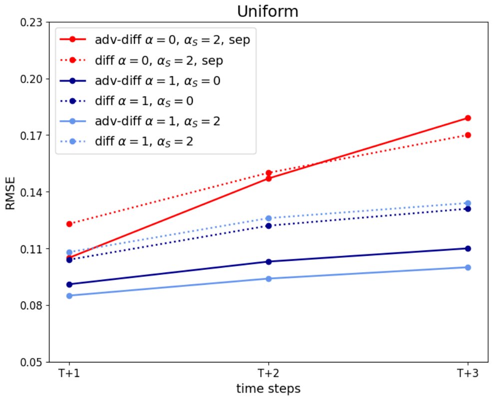
SPATIO-TEMPORAL PREDICTION



Predictions of K_c at $(T+1)$, $(T+2)$ and $(T+3)$ with "adv-diff" model with $\alpha = 1$ and $\alpha_S = 2$

Black dots are conditioning points

■ PREDICTION SCORES

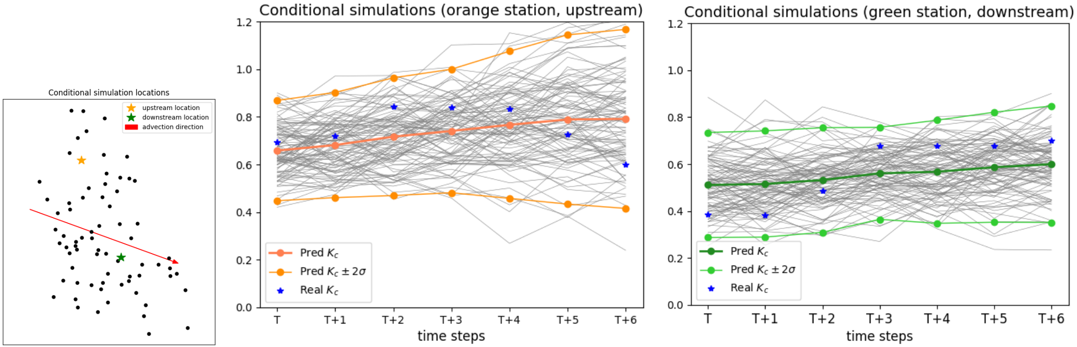


Prediction RMSE at $(T+1)$, $(T+2)$ and $(T+3)$ for 6 different models

Result

- **Advection-diffusion** models have **smaller errors** than purely diffusive models with same α and α_S
- **Non-separable** models have **smaller errors** than **separable** models

■ CONDITIONAL SIMULATIONS



Real K_c, mean of conditional simulations of K_c and $\pm 2\sigma$-envelope for temporal horizons $T, (T+1), \ldots, (T+6)$

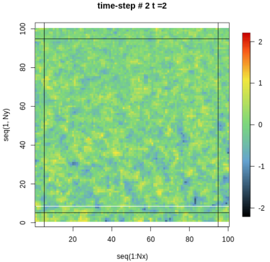
Result

- Conditional simulations have **lower uncertainty** at the **"downstream"** location
- The advection field moves information from **NW** to **SE**

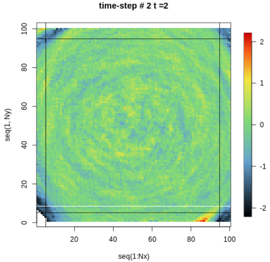
NON-STATIONARITY: SIMULATION

Non-stationary dependence structure

- We introduce nonstationarities by making $\kappa(s,t)$ or $\boldsymbol{\gamma}(s,t)$ **vary** in space and/or time
- The **discretization** method is equivalent (Implicit Euler + Finite Elements) $\rightarrow$ very fast!



Diffusion SPDE
$\kappa(s)$ spatially varying (evolution on y-axis)
$\kappa(x,y) = y/200$

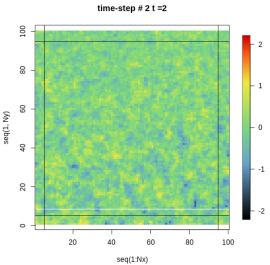


Advection-diffusion SPDE, $\kappa = 0.33$
$\boldsymbol{\gamma}(s)$ spatially varying (rotation around (50,50))
$\boldsymbol{\gamma}(x,y) = [-0.5(x-50), 0.5(y-50)]^\top$

■ NON-STATIONARITY: SIMULATION

Non-stationary dependence structure

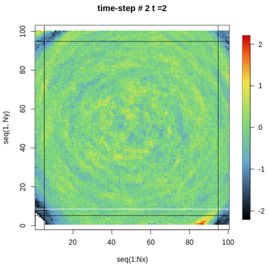
- We introduce nonstationarities by making $\kappa(s,t)$ or $\boldsymbol{\gamma}(s,t)$ **vary** in space and/or time
- The **discretization** method is equivalent (Implicit Euler + Finite Elements) $\rightarrow$ very fast!



Diffusion SPDE

$\kappa(s)$ spatially varying (evolution on y-axis)

$\kappa(x,y) = y/200$



Advection-diffusion SPDE, $\kappa = 0.33$

$\boldsymbol{\gamma}(s)$ spatially varying (rotation around (50,50))

$\boldsymbol{\gamma}(x,y) = [-0.5(x-50), 0.5(y-50)]^\top$

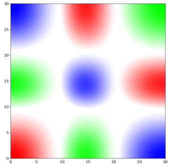
$\rightarrow$ **Estimation is not straightforward!**

■ PARAMETERIZATION BASED ON B-SPLINES

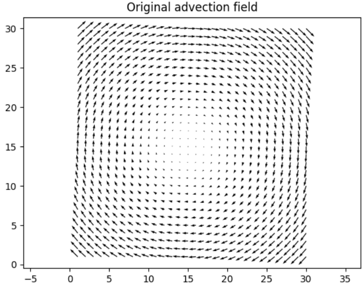
Parameterization based on basis functions (Fuglstad et al., 2015)

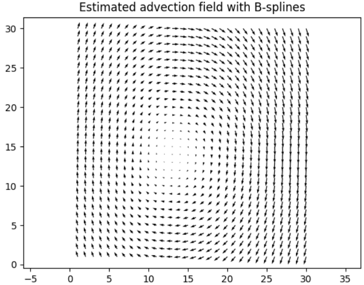
$$g(s) = \underbrace{\boldsymbol{f}(s)}_{\text{basis}}^{\top} \underbrace{\boldsymbol{a}_g}_{\text{coefficients}}$$

In 2D, the **bi-dimensional basis** is $f_{ij}(s) = B_{x,i}(x)B_{y,j}(y), \ \forall i,j \in \{1,\ldots,m\}$ where $B_{x,i}$ (and $B_{y,j}$) is the i-th (j-th) **clamped B-spline**

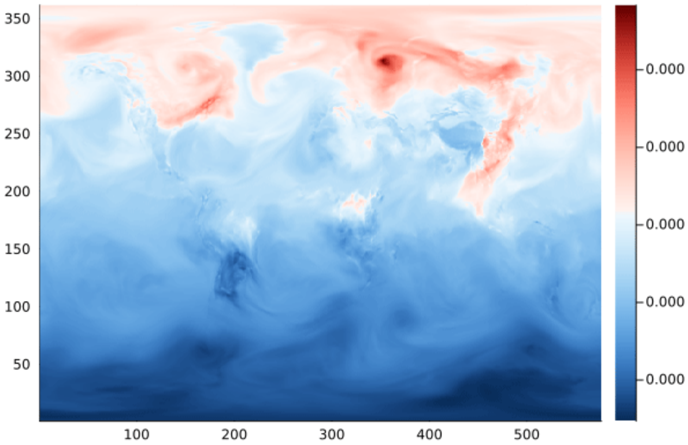


B-splines in 2D





■ APPLICATION TO CO_2 DATA



CO_2 concentration

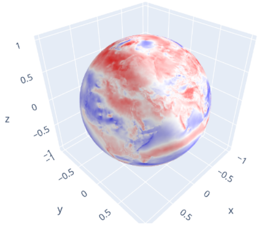
■ MODEL FOR CO_2

$$\boldsymbol{U} = \boldsymbol{\eta}\boldsymbol{b} + \boldsymbol{A}\boldsymbol{Z} + \sigma\boldsymbol{\varepsilon}$$

- $\boldsymbol{b} \in \mathbb{R}^d$ vector of q fixed effects
- $\boldsymbol{\eta} \in \mathbb{R}^{N\times q}$ matrix of covariates
- $\boldsymbol{Z}$ weights of the solution Z to the advection-diffusion SPDE on meshed manifold
- $\boldsymbol{A}$ observation matrix
- $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{I})$

Transport $\gamma(s)$ is a vector field $\Rightarrow$ covariance function is **non-stationary**

- parameterization based on spherical harmonics
  $\Rightarrow$ reduce number of coefficients to estimate
- parameterization based on covariate (wind speed)
  $\Rightarrow$ estimate the scaling factor only



Eastward component of wind speed

RESULTS

In a nutshell:

- Formulation of an advection-diffusion SPDE on a meshed surface
- Introduction of spatially varying parameters (advection field) resulting in a non-stationary covariance structure
- Explicit formula for the precision matrix that depends on sparse matrices
- Time scalable algorithms for sampling and kriging-based predictions (even for millions of data)
- Good results on application to solar radiation data

CONCLUSION AND OUTLOOKS

Still a lot to do:

- Speed up the inference (for now, Nelder Mead for optimization of the log-likelihood)

  $\Rightarrow$ Introduce automatic differantiation for the computation of the gradient of the log-likelihood

  or

  $\Rightarrow$ Use a variational formulation (variational EM)

  – Introduce variational approximation $q(\boldsymbol{Z};\boldsymbol{\psi}) \approx p(\boldsymbol{Z}|\,\boldsymbol{Y};\boldsymbol{\theta})$
  – Optimize ELBO jointly $(\boldsymbol{\theta},\boldsymbol{\psi})$
  – Variational approximation only used as an auxiliary quantity to learn $\boldsymbol{\theta}$
  – Linear computional cost $\mathcal{O}(n)$

CONCLUSION AND OUTLOOKS

- Application to $C0_2$ concentration in the atmosphere around the globe: inference and spatio-temporal prediction

  $\Rightarrow$ Spatio-temporal non-stationary model from advection-diffusion SPDE
  - Advection field as scaled wind speed field
  - Forcing term as sum of stochastic term and term based on built-up land cover

THANK YOU FOR YOUR ATTENTION!

For more on this subject

Pereira, M., **Clarotto, L.**, Desassis, N. (2023). A note on spatio-temporal random fields on meshed surfaces defined from advection-diffusion SPDEs. *hal-04132148*.

Clarotto, L., Allard, D., Romary, T., Desassis, N. (2022). The SPDE approach for spatio-temporal datasets with advection and diffusion *arXiv preprint arXiv:2208.14015*.

Lang, A. and Pereira, M. (2021). Galerkin–Chebyshev approximation of Gaussian random fields on compact Riemannian manifolds. *arXiv preprint arXiv:2107.02667*.

Pereira, M., Desassis, N., Allard D. (2022). Geostatistics for Large Datasets on Riemannian Manifolds: A Matrix-Free Approach, *Journal of Data Science, 20(4), 512-532*.

Pereira, M. and Desassis, N. (2019). Efficient simulation of Gaussian Markov random fields by Chebyshev polynomial approximation. *Spatial Statistics, 31:100359*.

REFERENCES

Chen, W., Genton, M. G., and Sun, Y. (2021). Space-time covariance structures and models. *Annual Review of Statistics and Its Application*, 8:191–215.

Clarotto, L., Allard, D., Romary, T., and Desassis, N. (2022). The spde approach for spatio-temporal datasets with advection and diffusion. *arXiv:2208.14015*.

Emery, X. and Porcu, E. (2019). Simulating isotropic vector-valued gaussian random fields on the sphere through finite harmonics approximations. *Stochastic Environmental Research and Risk Assessment*.

Fuglstad, G.-A., Lindgren, F., Simpson, D., and Rue, H. (2015). Exploring a new class of non-stationary spatial gaussian random fields with varying local anisotropy. *Statistica Sinica*, pages 115–133.

Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523.

Gneiting, T. (2002). Compactly supported correlation functions. *Journal of Multivariate Analysis*, 83(2):493–508.

■ REFERENCES

Hutchinson, M. F. (1989). A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 18(3):1059–1076.

Lang, A., Schwab, C., et al. (2015). Isotropic gaussian random fields on the sphere: regularity, fast simulation and stochastic partial differential equations. *The Annals of Applied Probability*, 25(6):3047–3094.

Lantuéjoul, C., Freulon, X., and Renard, D. (2019). Spectral simulation of isotropic gaussian random fields on a sphere. *Mathematical Geosciences*.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Liu, X., Yeo, K., and Lu, S. (2022). Statistical modeling for spatio-temporal data from stochastic convection-diffusion processes. *Journal of the American Statistical Association*, 117(539):1482–1499.

Marinucci, D. and Peccati, G. (2011). *Random fields on the sphere: representation, limit theorems and cosmological applications*, volume 389. Cambridge University Press.

REFERENCES

Mejia, A. F., Yue, Y., Bolin, D., Lindgren, F., and Lindquist, M. A. (2020). A Bayesian general linear modeling approach to cortical surface fMRI data analysis. *Journal of the American Statistical Association*, 115(530):501–520.

Porcu, E., Alegria, A., and Furrer, R. (2018). Modeling temporally evolving and spatially globally dependent data. *International Statistical Review*, 86(2):344–377.

Porcu, E., Furrer, R., and Nychka, D. (2021). 30 years of space–time covariance functions. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(2):e1512.

Rozanov, J. A. (1977). Markov random fields and stochastic partial differential equations. *Mathematics of the USSR-Sbornik*, 32(4):515.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. CRC press.

Sigrist, F., Künsch, H. R., and Stahel, W. A. (2015). Stochastic partial differential equation based modelling of large space-time data sets. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 3–33.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, pages 434–449.

Whittle, P. (1963). Stochastic-processes in several dimensions. *Bulletin of the International Statistical Institute*, 40(2):974–994.