PREDICTION OF SPATIO-TEMPORAL GAUSSIAN PROCESSES BY ADVECTION-DIFFUSION STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS

UQSay

Lucia Clarotto[1] (joint work with M. Pereira[2], N. Desassis[2], T. Romary[2], D. Allard[3])

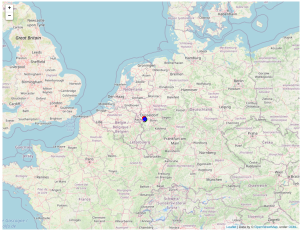
April 9, 2026

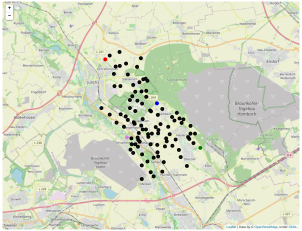
[1]MIA Paris-Saclay, AgroParisTech, INRAE, Université Paris Saclay, France
[2]STIM, Mines Paris, Université PSL, France
[3]BioSP, INRAE, Avignon, France

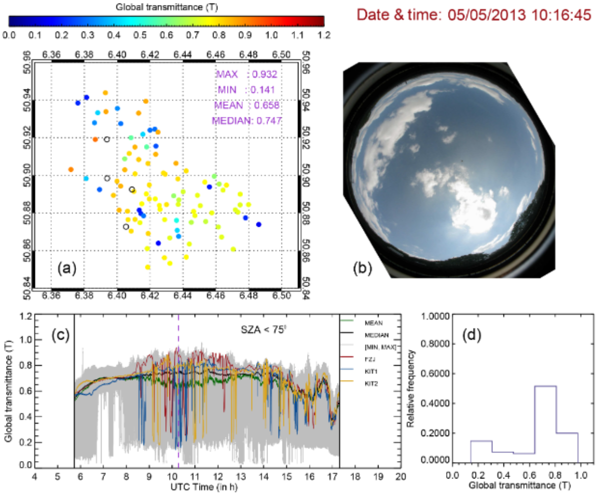
DATA AND CONTEXT

SPATIO-TEMPORAL PROCESS IN ENVIRONMENTAL SCIENCE

SOLAR RADIATION DATASET





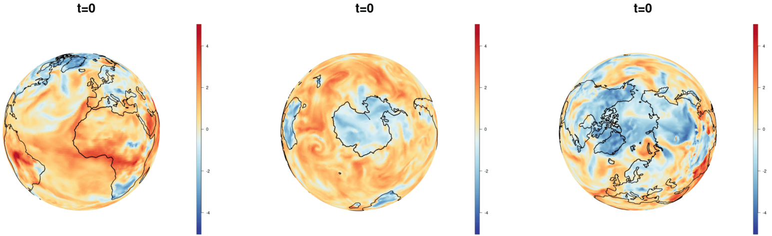


Aerosols are tiny solid or liquid particles suspended in the atmosphere

- Sources: dust, sea salt, volcanic ash, wildfire smoke, pollution
- Effects: cool or warm surface, influence cloud formation, affect health

AOD (Aerosol Optical Depth) at 550nm, every 3 hours, 01/12/2024 - 03/12/2024 (ECMWF Atmospheric Composition Reanalysis 4)

- Spatial resolution of $0.75° \times 0.75°$ $\Rightarrow$ 21 time steps and 163842 spatial locations
- Measures how aerosols affect light (higher AOD $\Rightarrow$ hazier skies)



SPATIO-TEMPORAL MODEL

On a spatial domain Ω and a time interval $[0, T]$, a space-time (ST) variable $Y(\mathbf{s}, t)$ is modeled as

$$Y(\mathbf{s}, t) = \underbrace{\mu(\mathbf{s}, t)}_{\text{fixed effects}} + \underbrace{Z(\mathbf{s}, t)}_{\text{random effects}} + \underbrace{\epsilon}_{\text{residual}}$$

where $Z(\mathbf{s}, t)$ is a second order Gaussian **random field**

- Models data that are not independent, identically distributed
- High **correlation** in the Gaussian field $\Rightarrow$ High **similarity** in the observed variable
- Parametric **covariance function** C_{ST} (in the stationary case):

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

  with $(\mathbf{s}, t), (\mathbf{s}', t') \in \mathbb{R}^d \times \mathbb{R}$ and $\mathbf{h} = ||\mathbf{s} - \mathbf{s}'||$, $u = |t - t'|$
  $\rightarrow$ used to model the observed spatio-temporal structure
  $\rightarrow$ its matrix form Σ defines the covariance between all observed points: big and dense matrix
  $\rightarrow$ Σ must be inverted for parameter estimation and prediction

SPATIO-TEMPORAL MODEL

On a spatial domain Ω and a time interval $[0, T]$, a space-time (ST) variable $Y(\mathbf{s}, t)$ is modeled as

$$Y(\mathbf{s}, t) = \underbrace{\mu(\mathbf{s}, t)}_{\text{fixed effects}} + \underbrace{Z(\mathbf{s}, t)}_{\text{random effects}} + \underbrace{\epsilon}_{\text{residual}}$$

where $Z(\mathbf{s}, t)$ is a second order Gaussian **random field**

- Models data that are not independent, identically distributed
- High **correlation** in the Gaussian field $\Rightarrow$ High **similarity** in the observed variable
- Parametric **covariance function** C_{ST} (in the stationary case):

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

with $(\mathbf{s}, t), (\mathbf{s}', t') \in \mathbb{R}^d \times \mathbb{R}$ and $\mathbf{h} = \|\mathbf{s} - \mathbf{s}'\|$, $u = |t - t'|$
$\rightarrow$ used to model the observed spatio-temporal structure
$\rightarrow$ its matrix form Σ defines the covariance between all observed points: big and dense matrix
$\rightarrow$ Σ must be inverted for parameter estimation and prediction

EXAMPLES OF SPATIAL DEPENDENCE STRUCTURE

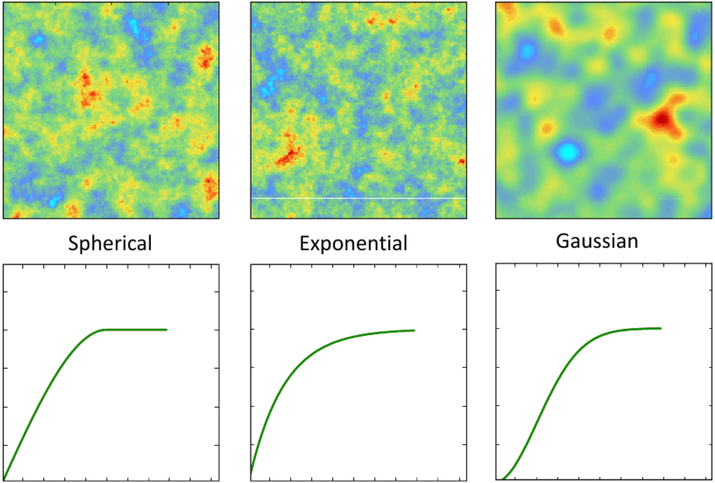


Figure: Spatial Gaussian random field spatial simulated with different covariance functions

Separability

$C_{ST}(\mathbf{h}, u) = C(\mathbf{0}, 0)^{-1} C_S(\mathbf{h}) C_T(u)$

Covariance matrix $\Sigma = \Sigma_T \otimes \Sigma_S$

- Easy to understand, less memory and easy coding
- **Under Gaussian hypothesis**, it's equivalent to conditional independence :

$$\left(Z(\mathbf{s}, t) \perp Z(\mathbf{s}', t')\right) | Z(\mathbf{s}, t')$$

- No complex spatio-temporal interaction
- Not realistic in many environmental applications

$\Rightarrow$ We choose **nonseparable** models to describe **more complex** phenomena

Separability

$C_{ST}(\mathbf{h}, u) = C(\mathbf{0}, 0)^{-1} C_S(\mathbf{h}) C_T(u)$

Covariance matrix $\Sigma = \Sigma_T \otimes \Sigma_S$

- Easy to understand, less memory and easy coding
- **Under Gaussian hypothesis**, it's equivalent to conditional independence :

$$\left(Z(\mathbf{s}, t) \perp Z(\mathbf{s}', t')\right) | Z(\mathbf{s}, t')$$

- No complex spatio-temporal interaction
- Not realistic in many environmental applications

$\Rightarrow$ We choose **nonseparable** models to describe **more complex** phenomena

Symmetry

$$C_{ST}(\mathbf{h}, u) = C_{ST}(-\mathbf{h}, u) = C_{ST}(\mathbf{h}, -u) = C_{ST}(-\mathbf{h}, -u)$$

- Symmetric models cannot distinguish differing effects wrt past or future time

$\Rightarrow$ We choose **asymmetric** models, because full symmetry is incompatible with environmental processes under the **influence of air or water flows**

Symmetry

$$C_{ST}(\mathbf{h}, u) = C_{ST}(-\mathbf{h}, u) = C_{ST}(\mathbf{h}, -u) = C_{ST}(-\mathbf{h}, -u)$$

- Symmetric models cannot distinguish differing effects wrt past or future time

$\Rightarrow$ We choose **asymmetric** models, because full symmetry is incompatible with environmental processes under the **influence of air or water flows**

Stationarity

$$C_{ST}(\mathbf{h}, u) = \mathrm{Cov}[Z(\mathbf{s}, t), Z(\mathbf{s}+\mathbf{h}, t+u)], \quad (\mathbf{s}, t), (\mathbf{s}+\mathbf{h}, t+u) \in \mathbb{R}^d \times \mathbb{R}$$

- Stationary models cannot distinguish different dependence structures in different locations of the domain

$\Rightarrow$ We choose **nonstationary** models to capture **complex dependence structures**

Stationarity

$$C_{ST}(\mathbf{h}, u) = \mathrm{Cov}[Z(\mathbf{s}, t), Z(\mathbf{s}+\mathbf{h}, t+u)], \quad (\mathbf{s}, t), (\mathbf{s}+\mathbf{h}, t+u) \in \mathbb{R}^d \times \mathbb{R}$$

- Stationary models cannot distinguish different dependence structures in different locations of the domain

$\Rightarrow$ We choose **nonstationary** models to capture **complex dependence structures**

Goal

Predict a **space-time** variable of interest from scattered observations

Challenges

- Introduce **physical phenomena** in a statistical spatio-temporal model
- Relax the hypothesis of **stationarity** to model more complex phenomena
- Treat data on **non Euclidean surfaces**
- Manage the **abundance of data** $(N_T \times N_S)$ and design fast and scalable methods for parameter **estimation** and **prediction**

$\Rightarrow$ Gaussian random fields defined from advection-diffusion SPDEs on Riemannian manifolds

GOAL AND CHALLENGES

Goal

Predict a **space-time** variable of interest from scattered observations

Challenges

- Introduce **physical phenomena** in a statistical spatio-temporal model
- Relax the hypothesis of **stationarity** to model more complex phenomena
- Treat data on **non Euclidean surfaces**
- Manage the **abundance of data** $(N_T \times N_S)$ and design fast and scalable methods for parameter **estimation** and **prediction**

$\Rightarrow$ Gaussian random fields defined from advection-diffusion SPDEs on Riemannian manifolds

GOAL AND CHALLENGES

Goal

Predict a **space-time** variable of interest from scattered observations

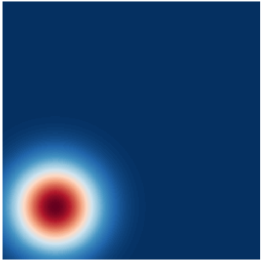
Challenges

- Introduce **physical phenomena** in a statistical spatio-temporal model
- Relax the hypothesis of **stationarity** to model more complex phenomena
- Treat data on **non Euclidean surfaces**
- Manage the **abundance of data** $(N_T \times N_S)$ and design fast and scalable methods for parameter **estimation** and **prediction**

$\Rightarrow$ **Gaussian random fields defined from advection-diffusion SPDEs on Riemannian manifolds**

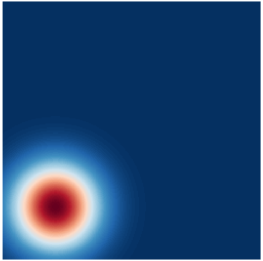
RANDOM FIELDS VIA ADVECTION-DIFFUSION SPDES

PHYSICAL PHENOMENA



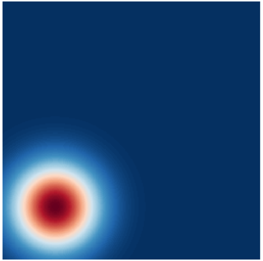
Diffusion

$$\frac{\partial x}{\partial t} - \Delta x = f$$



Advection

$$\frac{\partial x}{\partial t} + \text{div}(x\gamma) = f$$



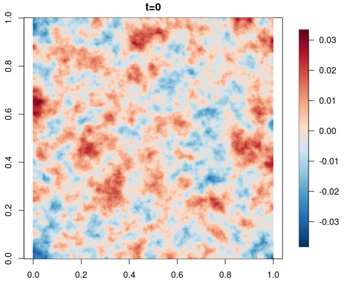
Advection-diffusion

$$\frac{\partial x}{\partial t} - \Delta x + \text{div}(x\gamma) = f$$

On **Euclidean domains**, consider the **advection-diffusion** SPDE (Clarotto et al., 2024)

$$\frac{\partial \mathcal{Z}}{\partial t} + \frac{1}{c}\Big(\underbrace{(\kappa^2 - \Delta)\mathcal{Z}}_{\text{diffusion with damping}} + \underbrace{c_{adv}\,\text{div}(\mathcal{Z}\gamma)}_{\text{advection}} \Big) = \frac{\tau}{\sqrt{c}} \underbrace{\mathcal{W}_T \otimes \mathcal{Y}_S}_{\text{stochastic forcing}} ,$$

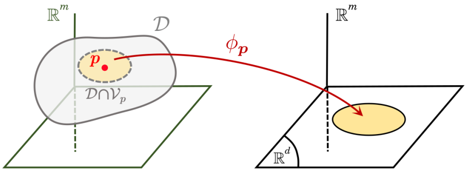
- γ is a vector defining the advection and c_{adv} is a scaling factor
- $\mathcal{W}_T \otimes \mathcal{Y}_S$ is a spatio-temporal noise, white in time and colored in space, such that $(\kappa_S^2 - \Delta)^{\alpha_S/2}\mathcal{Y}_S = \mathcal{W}_S$



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

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

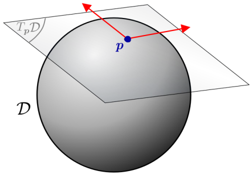
- locally Euclidian of dimension d
- can be entirely mapped by a set of smoothly compatible charts



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

$\mathcal{M}$ is equipped with a **Riemannian metric** g

- g_p: inner product on the tangent space of $\mathcal{M}$ at $p \in \mathcal{DM}$
- $g : p \mapsto g_p$ is smooth



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

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

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

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

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

- $-\Delta_{\mathcal{M}}$ 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 and colored in space
- $s \in \mathcal{M} \mapsto \gamma(s)$ is a smooth **tangent vector field**

Example of tangent vector field $\gamma(s)$ on the 2-sphere $\mathbb{S}^2$
$\Rightarrow$ Helmholtz-Hodge decomposition: curl-free component + divergence-free component

$$\gamma(s) = \nabla\xi(s) + \vec{n}(s) \times \nabla\chi(s) \in T_s\mathbb{S}^2,$$

where $\xi, \chi : \mathcal{M} \to \mathbb{R}$ smooth functions, and $\vec{n}(s)$ outward normal at $s \in \mathcal{M}$

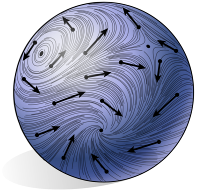


Figure: Tangent vector field

ADVECTION-DIFFUSION SPDE ON RIEMANNIAN MANIFOLD

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

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

- $-\Delta_{\mathcal{M}}$ 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 and colored in space
- $s \in \mathcal{M} \mapsto \gamma(s)$ is a smooth **tangent vector field**

Example of tangent vector field $\gamma(s)$ on the 2-sphere $\mathbb{S}^2$
$\Rightarrow$ Helmholtz-Hodge decomposition: curl-free component + divergence-free component

$$\gamma(s) = \nabla\xi(s) + \vec{n}(s) \times \nabla\chi(s) \in T_s\mathbb{S}^2,$$

where $\xi, \chi : \mathcal{M} \to \mathbb{R}$ smooth functions, and $\vec{n}(s)$ outward normal at $s \in \mathcal{M}$

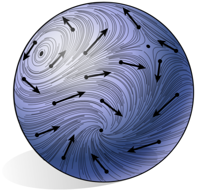


Figure: Tangent vector field

SIMULATION OF A RANDOM FIELD ON THE SPHERE

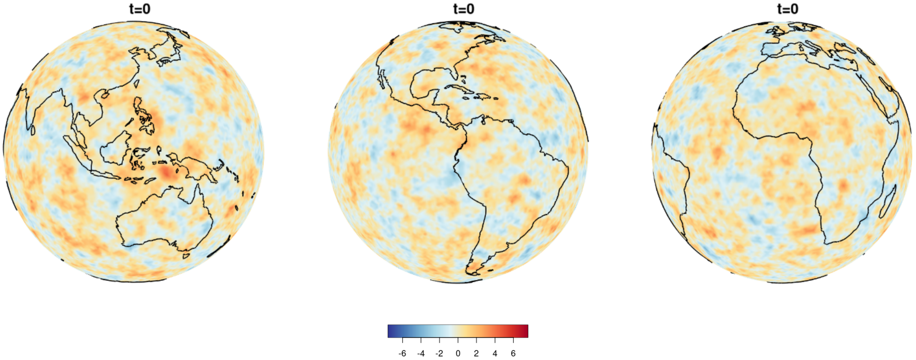


Figure: Simulation of an advection-diffusion SPDE model on a Riemannian manifold

COVARIANCE ON THE SPHERE

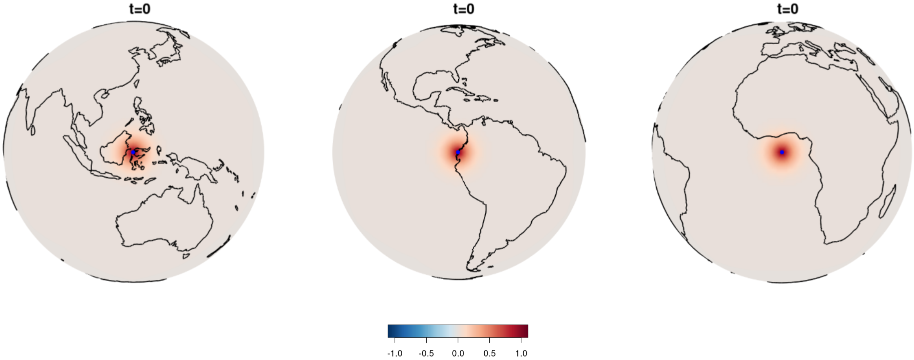


Figure: ST evolution of covariance between reference point and other points on the domain

ADVECTION MODELING FOR AOD

The advection vector field γ is chosen to be parameterized by a covariate

$\Rightarrow$ 10m u-component and v-component of wind

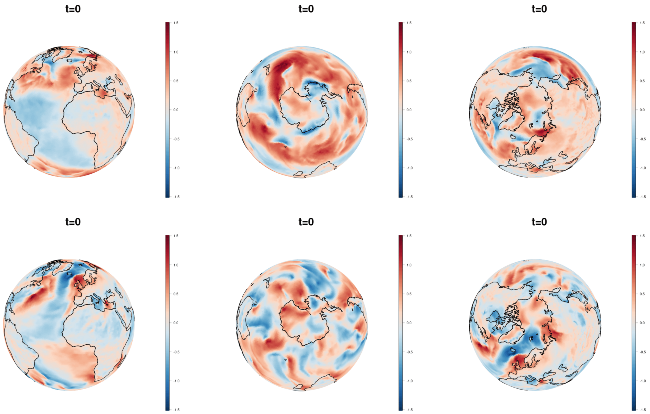


Figure: 10m u-component and v-component of wind

For the SPDE to be stable, a sufficient condition is

$$\inf_{\mathcal{M}} \left(\kappa^2 + \frac{1}{2} \text{div}(\gamma) \right) > 0$$

If we only take the divergence-free term of the Helmholtz-Hodge decomposition of wind field, we are sure to respect the condition.
$\Rightarrow$ This implies forgetting about small scales and tropical patterns in wind field.

Transport modeling

The modeling choices for $c_{adv}\text{div}(\mathcal{Z}\gamma)$ imply the following properties:

- Temporally varying vector field $\gamma(t, \cdot)$
- A **global scaling factor** c_{adv} to be estimated

For the SPDE to be stable, a sufficient condition is

$$\inf_{\mathcal{M}} \left(\kappa^2 + \frac{1}{2}\text{div}(\gamma) \right) > 0$$

If we only take the divergence-free term of the Helmholtz-Hodge decomposition of wind field, we are sure to respect the condition.
$\Rightarrow$ This implies forgetting about small scales and tropical patterns in wind field.

Transport modeling

The modeling choices for $c_{adv}\text{div}(\mathcal{Z}\gamma)$ imply the following properties:

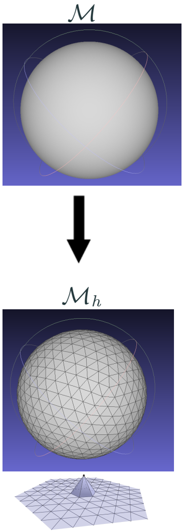
- Temporally varying vector field $\gamma(t, \cdot)$
- A **global scaling factor** c_{adv} to be estimated

INFERENCE AND PREDICTION

From AOD data, we would like to:

(1) Infer the parameters of the SPDE-based random field
(2) Make spatio-temporal prediction from scattered data
(3) Make conditional simulations of the SPDE-based random field

For all these tasks, we need to work with the covariance (or precision) matrix and make it as sparse as possible



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

Triangulation of the surface

+ Galerkin approximation

+ Implicit Euler

$\boldsymbol{C}$ (mass matrix), $\widetilde{\boldsymbol{R}}$ (scaled stiffness matrix), $\widetilde{\boldsymbol{B}}^{(k)}$ (k-th scaled advection matrix)

Let $\boldsymbol{x}^{(0)} = (\sqrt{\boldsymbol{C}})^T \boldsymbol{z}^{(0)}$. Then we have the following recursion for $0 \leq k < K$:

$$\begin{cases} \boldsymbol{\Gamma}^{(k)} \boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \frac{\delta t}{c}(\sqrt{\boldsymbol{C}})^T \boldsymbol{m} + f_{\delta t}(\widetilde{\boldsymbol{R}})\boldsymbol{w}^{(k)}, \\ \boldsymbol{z}^{(k+1)} = (\sqrt{\boldsymbol{C}})^{-T} \boldsymbol{x}^{(k+1)}, \end{cases}$$

with $\boldsymbol{\Gamma}^{(k)} = \boldsymbol{I} + \frac{\delta t}{c}\left(\kappa^2 \boldsymbol{I} + \widetilde{\boldsymbol{R}}) + \widetilde{\boldsymbol{B}}^{(k)}\right)$, $\quad f_{\delta t}(\lambda) = \tau\sqrt{\frac{\delta t}{c}}(\kappa_S^2 + \lambda)^{-1}$ and $\{\boldsymbol{w}^{(k)}\}_{k=0,\dots,K-1}$ i.i.d. realizations of standard Gaussian spatial random vectors

PRECISION MATRIX OF THE FIELD

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

$$\boldsymbol{Q}_{\boldsymbol{Z}} = \begin{pmatrix} \boldsymbol{\Gamma}_1^{(0)} & -\boldsymbol{\Gamma}_2^{(0)} & & & \\ -(\boldsymbol{\Gamma}_2^{(0)})^T & \boldsymbol{\Gamma}_1^{(1)} & -\boldsymbol{\Gamma}_2^{(1)} & & \\ & \ddots & \ddots & \ddots & \\ & & -(\boldsymbol{\Gamma}_2^{(K-2)})^T & \boldsymbol{\Gamma}_1^{(K-1)} & -\boldsymbol{\Gamma}_2^{(K-1)} \\ & & & -(\boldsymbol{\Gamma}_2^{(K-1)})^T & \boldsymbol{\Gamma}_1^{(K)} \end{pmatrix}$$

$$\boldsymbol{\Gamma}_1^{(k)} = \begin{cases} (\sqrt{\boldsymbol{C}})\big(f_0^{-2}(\widetilde{\boldsymbol{R}}) + f_{\delta t}^{-2}(\widetilde{\boldsymbol{R}})\big)(\sqrt{\boldsymbol{C}})^T & \text{if } k = 0 \\ (\sqrt{\boldsymbol{C}})\left((\boldsymbol{\Gamma}^{(k-1)})^T f_{\delta t}^{-2}(\widetilde{\boldsymbol{R}})\,\boldsymbol{\Gamma}^{(k-1)} + f_{\delta t}^{-2}(\widetilde{\boldsymbol{R}})\right)(\sqrt{\boldsymbol{C}})^T & \text{if } 1 \leq k \leq K-1 \\ (\sqrt{\boldsymbol{C}})(\boldsymbol{\Gamma}^{(K-1)})^T f_{\delta t}^{-2}(\widetilde{\boldsymbol{R}})\,\boldsymbol{\Gamma}^{(K-1)}(\sqrt{\boldsymbol{C}})^T & \text{if } k = K \end{cases}$$

$$\boldsymbol{\Gamma}_2^{(k)} = (\sqrt{\boldsymbol{C}})f_{\delta t}^{-2}(\widetilde{\boldsymbol{R}})\boldsymbol{\Gamma}^{(k)}(\sqrt{\boldsymbol{C}})^T \qquad \text{for } 0 \leq k \leq K-1$$

$\Rightarrow$ $\boldsymbol{Q}_{\boldsymbol{Z}}$ is block tri-diagonal in time and sparse in each spatial block

$\Rightarrow$ $\boldsymbol{Z}$ is a Gaussian Markov random field

Data $U(t_k, s_i^{(k)})$, observations at time step $t_k \in [0, T]$ and locations $s_1^{(k)}, \dots, s_{n_k}^{(k)} \in \mathcal{M}_h$

Model

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

with $\varepsilon_i \sim \mathcal{N}(0,1)$, $\boldsymbol{A}^{(k)}$ spatial observation matrix and Z approx. solution of the adv-diff SPDE
Equivalently, $\boldsymbol{U} = \boldsymbol{A}^\top \boldsymbol{Z} + \sigma \boldsymbol{\varepsilon}$

Goal Estimation of parameters $\boldsymbol{\nu} = (\kappa, \kappa_S, c_{adv}, c, \tau, \sigma)$

Log-likelihood optimization by Nelder-Mead

$$\mathcal{L}(\boldsymbol{\nu}) = -\frac{N_o}{2}\log 2\pi + \frac{1}{2}\log|\boldsymbol{Q}_U| - \frac{\sigma^{-2}}{2}\|\boldsymbol{U}\|_2^2 + \frac{\sigma^{-2}}{2}\boldsymbol{U}^T\boldsymbol{A}^T\sigma^{-2}(\boldsymbol{Q}_Z + \sigma^{-2}\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\boldsymbol{U},$$

where $\boldsymbol{Q}_U = (\boldsymbol{A}^T\boldsymbol{Q}_Z^{-1}\boldsymbol{A} + \sigma^2\boldsymbol{I})^{-1}$ and $\log|\boldsymbol{Q}_U| = -N_o\log\sigma^2 + \log|\boldsymbol{Q}_Z| - \log|\boldsymbol{Q}_Z + \sigma^{-2}\boldsymbol{A}\boldsymbol{A}^T|$

$\Rightarrow$ We compute the quadratic form by only solving sparse systems by GMRES preconditioned with block symmetric Gauss-Seidel

$\Rightarrow$ We use scalable stochastic approximation based on Lanczos tridiagonalization to compute logdets of matrices (Dong et al, 2017)

Data $U(t_k, s_i^{(k)})$, observations at time step $t_k \in [0, T]$ and locations $s_1^{(k)}, \dots, s_{n_k}^{(k)} \in \mathcal{M}_h$

Model

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

with $\varepsilon_i \sim \mathcal{N}(0,1)$, $\boldsymbol{A}^{(k)}$ spatial observation matrix and Z approx. solution of the adv-diff SPDE
Equivalently, $\boldsymbol{U} = \boldsymbol{A}^\top \boldsymbol{Z} + \sigma\boldsymbol{\varepsilon}$

Goal Estimation of parameters $\boldsymbol{\nu} = (\kappa, \kappa_S, c_{adv}, c, \tau, \sigma)$

Log-likelihood optimization by Nelder-Mead

$$\mathcal{L}(\boldsymbol{\nu}) = -\frac{N_o}{2}\log 2\pi + \frac{1}{2}\log|\boldsymbol{Q}_U| - \frac{\sigma^{-2}}{2}\|\boldsymbol{U}\|_2^2 + \frac{\sigma^{-2}}{2}\boldsymbol{U}^T\boldsymbol{A}^T\sigma^{-2}(\boldsymbol{Q}_Z + \sigma^{-2}\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\boldsymbol{U},$$

where $\boldsymbol{Q}_U = (\boldsymbol{A}^T\boldsymbol{Q}_Z^{-1}\boldsymbol{A} + \sigma^2\boldsymbol{I})^{-1}$ and $\log|\boldsymbol{Q}_U| = -N_o\log\sigma^2 + \log|\boldsymbol{Q}_Z| - \log|\boldsymbol{Q}_Z + \sigma^{-2}\boldsymbol{A}\boldsymbol{A}^T|$

$\Rightarrow$ We compute the quadratic form by only solving sparse systems by GMRES preconditioned with block symmetric Gauss-Seidel

$\Rightarrow$ We use scalable stochastic approximation based on Lanczos tridiagonalization to compute logdets of matrices (Dong et al, 2017)

Data $U(t_k, s_i^{(k)})$, observations at time step $t_k \in [0, T]$ and locations $s_1^{(k)}, \dots, s_{n_k}^{(k)} \in \mathcal{M}_h$

Model

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

with $\varepsilon_i \sim \mathcal{N}(0,1)$, $\boldsymbol{A}^{(k)}$ spatial observation matrix and Z approx. solution of the adv-diff SPDE
Equivalently, $\boldsymbol{U} = \boldsymbol{A}^\top \boldsymbol{Z} + \sigma\boldsymbol{\varepsilon}$

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

Kriging

The conditional expectation (also called **kriging** predictor) $\mathbb{E}[\boldsymbol{Z}|\boldsymbol{U}]$ of $\boldsymbol{Z}$ given $\boldsymbol{U}$ is given by

$$\boldsymbol{Z}^* = \mathbb{E}[\boldsymbol{Z}|\boldsymbol{U}] = \sigma^{-2}(\boldsymbol{Q}_Z + \sigma^{-2}\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{A}\boldsymbol{U},$$

$\Rightarrow$ We solve the system by GMRES preconditioned with block symmetric Gauss-Seidel

Data $U(t_k, s_i^{(k)})$, observations at time step $t_k \in [0, T]$ and locations $s_1^{(k)}, \dots, s_{n_k}^{(k)} \in \mathcal{M}_h$

Model

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

with $\varepsilon_i \sim \mathcal{N}(0,1)$, $\boldsymbol{A}^{(k)}$ spatial observation matrix and Z approx. solution of the adv-diff SPDE
Equivalently, $\boldsymbol{U} = \boldsymbol{A}^\top \boldsymbol{Z} + \sigma \boldsymbol{\varepsilon}$

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

Kriging

The conditional expectation (also called **kriging** predictor) $\mathbb{E}[\boldsymbol{Z}|\boldsymbol{U}]$ of $\boldsymbol{Z}$ given $\boldsymbol{U}$ is given by

$$\boldsymbol{Z}^* = \mathbb{E}[\boldsymbol{Z}|\boldsymbol{U}] = \sigma^{-2}(\boldsymbol{Q}_Z + \sigma^{-2}\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{A}\boldsymbol{U},$$

$\Rightarrow$ We solve the system by GMRES preconditioned with block symmetric Gauss-Seidel

PREDICTION OF AOD

- Observations $\boldsymbol{U}$ at 5000 (3%) randomly distributed spatial locations at 21 time steps
- Inference from data at the first 11 time steps
- Predictions at all mesh locations at all 21 time steps

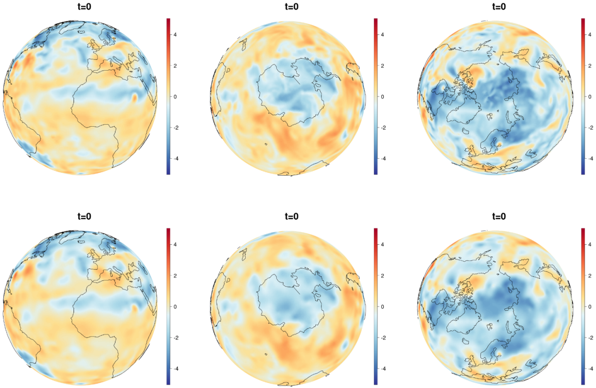
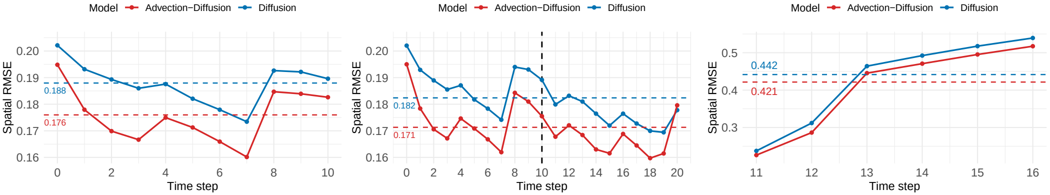


Figure: Up: AOD data. Down: Predictions by kriging $\mathbb{E}[\boldsymbol{Z}|\boldsymbol{U}]$

We assess prediction performance with **RMSE** between the kriging predictors and the observed data, averaged over all spatial locations on the sphere, at each time step



(a) Spatial RMSE at inference time steps (0-10)

(b) Spatial RMSE at inference time steps (0-10) and subsequent time steps (11-20)

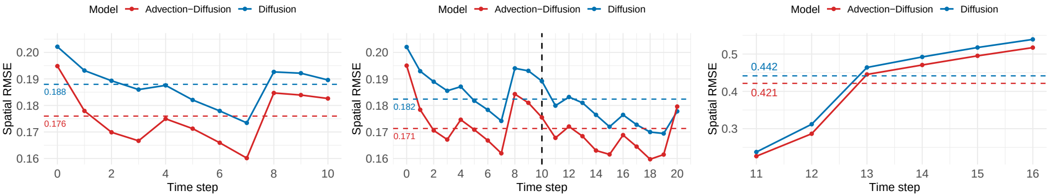
(c) Extrapolation spatial RMSE at subsequent time steps (11-16)

Figure: Spatial RMSE of the kriging predictors, at each time step

$\Rightarrow$ The advection–diffusion model outperforms the diffusion model, with an improvement of 6%

$\Rightarrow$ It is not necessary to perform inference using the full set of observations

We assess prediction performance with **RMSE** between the kriging predictors and the observed data, averaged over all spatial locations on the sphere, at each time step



(a) Spatial RMSE at inference time steps (0-10)

(b) Spatial RMSE at inference time steps (0-10) and subsequent time steps (11-20)

(c) Extrapolation spatial RMSE at subsequent time steps (11-16)

Figure: Spatial RMSE of the kriging predictors, at each time step

$\Rightarrow$ The advection–diffusion model outperforms the diffusion model, with an improvement of 6%

$\Rightarrow$ It is not necessary to perform inference using the full set of observations

Conditioning by kriging

(1) **Non-conditional** simulation $\boldsymbol{Z}^{(NC)}$ is performed on $\mathcal{M}_h \times [0, T]$ by FEM approximation of the SPDE

(2) **Kriging residuals** are computed as

$$\mathbf{r} = \boldsymbol{Z}^{(NC)} - \mathbb{E}\left(\boldsymbol{Z} \mid \boldsymbol{A}^\top \boldsymbol{Z}^{(NC)}_{1:N_T}\right)$$

(3) **Conditional** simulation is computed as sum of kriging prediction and kriging residuals

$$\boldsymbol{Z}^{(C)} = \boldsymbol{Z}^* + \mathbf{r}$$

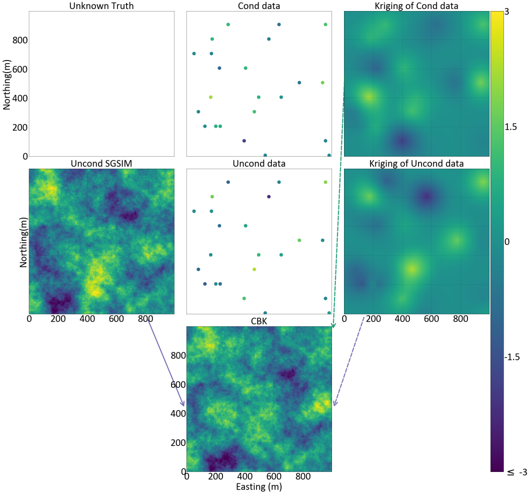
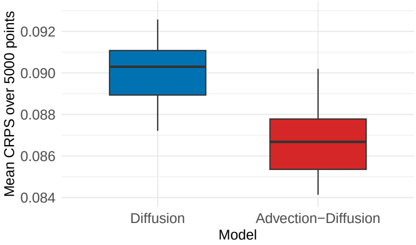


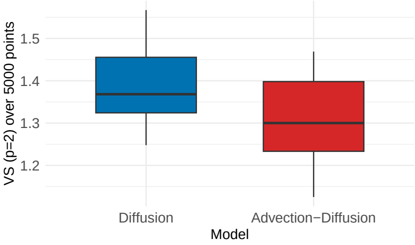
Figure: Conditioning by Kriging

We assess the probabilistic prediction performance of the advection-diffusion and diffusion models from 50 **conditional simulations**

- The **CRPS** evaluates the predictive performance of marginal distributions
- The **Variogram Score** assesses the quality of dependence structure across pairs of locations



(a) Average CRPS



(b) Variogram Score

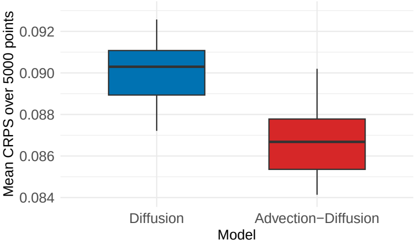
Figure: Scores computed for the two models

$\Rightarrow$ The advection–diffusion model outperforms the diffusion model, with an improvement of 4% in CRPS and 6% in VS

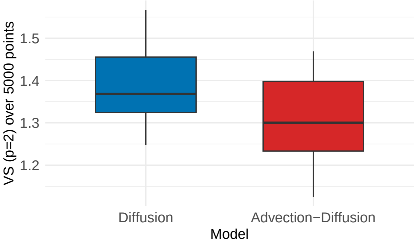
PROBABILISTIC SCORING RULES

We assess the probabilistic prediction performance of the advection-diffusion and diffusion models from 50 **conditional simulations**

- The **CRPS** evaluates the predictive performance of marginal distributions
- The **Variogram Score** assesses the quality of dependence structure across pairs of locations



(a) Average CRPS

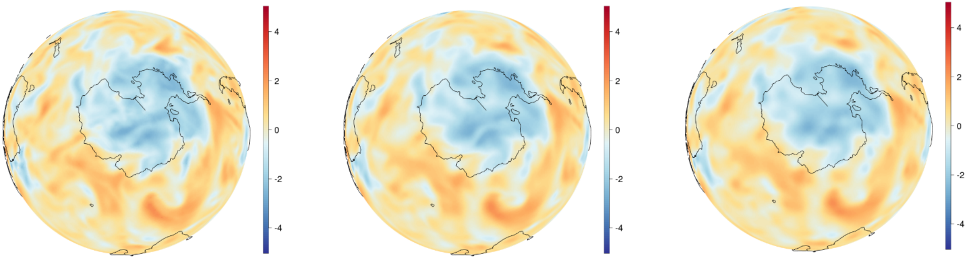


(b) Variogram Score

Figure: Scores computed for the two models

$\Rightarrow$ The advection–diffusion model outperforms the diffusion model, with an improvement of 4% in CRPS and 6% in VS

IMPROVEMENT IN LOCAL DETAILS



(a) AOD data (b) Advection-diffusion model (c) Diffusion model

Figure: AOD data and predictions with advection-diffusion and diffusion models at a given time step

$\Rightarrow$ The advection–diffusion model clearly captures the non-stationary channel-type behavior of AOD over central Antarctica

PERSPECTIVES

Still a lot to do

- Speed up the inference (for now, Nelder Mead for optimization of the log-likelihood) through different possibilities:
  - ⇒ Introduce **automatic differentiation** for the computation of gradients of the log-likelihood (is it really feasible?)
  - ⇒ Use **simulation-based inference** for parameter estimation (Alexandre Loret's PhD)
- More complex parameterization of the advection field when no covariate is available
  - ⇒ How to estimate an entire field?

Thank you for your attention!

lucia.clarotto@agroparistech.fr

Pereira, M., **Clarotto, L.**, Desassis, N. (2025). Prediction of spatio-temporal data on meshed surfaces using advection-diffusion SPDEs. hal-04132148.

Clarotto, L., Allard, D., Romary, T., Desassis, N. (2024). The SPDE approach for spatio-temporal datasets with advection and diffusion. Spatial Statistics, 62 :100847.

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.

PROBABILISTIC SCORING RULES

Used to assess the quality of predictive distributions, both in calibration (how well predicted probabilities match observations) and sharpness (the concentration of the predictive distribution).

- The **Continuous Ranked Probability Score (CRPS)** (Gneiting et al., 2007) is the most popular univariate scoring rule and is defined as

$$\text{CRPS}(F, y) = \mathbb{E}_F|X - y| - \frac{1}{2}\mathbb{E}_F|X - X'|,$$

where $y \in \mathbb{R}$ and X and X' are i.i.d. random variables following F, with a finite first moment.

$\Rightarrow$ The CRPS evaluates the accuracy of marginal predictive distributions at single locations.

- The **Variogram Score (VS)** (Scheuerer et al., 2015) is a multivariate score designed to assess the dependence structure. The VS of order p is defined as

$$\text{VS}_p(F, \boldsymbol{y}) = \sum_{i=1}^{d}\sum_{j=1}^{d} w_{i,j}\left(\mathbb{E}_F[|y_i - y_j|^p] - \mathbb{E}_F[|X_i - X_j|^p]\right)^2$$

where X_i is the i-component of the random vector X following F, $w_{i,j}$ are nonnegative weights, and $p > 0$ is the order of the variogram score.

$\Rightarrow$ The VS captures the quality of the forecasts' spatial dependence.