The SPDE approach for spatio-temporal datasets with advection and diffusion

Lucia Clarotto, Denis Allard, Thomas Romary, Nicolas Desassis

The SPDE approach for spatio-temporal datasets with advection and diffusion

Lucia Clarotto[1*], Denis Allard[2†], Thomas Romary[1†] and Nicolas Desassis[1†]

[1*]Centre for geosciences and geoengineering, Mines Paris, PSL University, Fontainebleau, 77300, France.
[2]Biostatistiques et Processus Spatiaux (BioSP), INRAE, Avignon, 84914, France.

*Corresponding author(s). E-mail(s): lucia.clarotto@minesparis.psl.eu;
†These authors contributed equally to this work.

Abstract

In the task of predicting spatio-temporal fields in environmental science, introducing models inspired by the physics of the underlying phenomena that are numerically efficient is of growing interest in spatial statistics. The size of space-time datasets calls for new numerical methods to efficiently process them. The SPDE (Stochastic Partial Differential Equation) approach has proven to be effective for the estimation and the prediction in a spatial context. We present here the advection-diffusion SPDE with first order derivative in time to enlarge the SPDE family to the space-time context. By varying the coefficients of the differential operators, the approach allows to define a large class of non-separable spatio-temporal models. A Gaussian Markov random field approximation of the solution of the SPDE is built by discretizing the temporal derivative with a finite difference method (implicit Euler) and by solving the purely spatial SPDE with a finite element method (continuous Galerkin) at each time step. The "Streamline Diffusion" stabilization technique is introduced when the advection term dominates the diffusion term. Computationally efficient methods are proposed to estimate the parameters of the SPDE and to predict the spatio-temporal field by kriging, as well as to perform conditional simulations. The approach is applied to a solar radiation dataset. Its advantages and limitations are discussed.

1 Introduction

Many areas of environmental science seek to predict a space-time variable of interest from observations at scattered points in the space-time field of study. Among modern techniques proposing efficient methods for estimation and prediction in a spatio-temporal framework, there is a clear distinction between two possible ways of constructing and treating spatio-temporal models (Wikle and Hooten, 2010): either one follows the traditional geostatistical paradigm, using joint space-time covariance functions (see for example Cressie and Huang (1999), Gneiting (2002), Stein (2005)), or one uses dynamical models, by combining time series and spatial statistics (see for example Wikle and Cressie (1999), Sigrist et al. (2012) and Martínez-Hernández and Genton (2022)).

While the theoretical aspects of spatio-temporal geostatistics are well developed (Cressie and Wikle, 2011), implementations lack behind. The geostatistical paradigm can be computationally expensive for large spatio-temporal datasets, due to the factorization of dense covariance matrices, whose complexity scales with the cube of the number of observation. Banerjee et al. (2014) called this issue the "big n problem". Moreover, it is hard to define complex space-time covariance functions. For this reason, separable space-time covariance functions have often been applied to spatio-temporal models to take advantage of their computational convenience, even when they are not realistic in describing the processes due to the impossibility of allowing space-time interaction in the covariance. Recent studies have focused on constructing non-separable models, which are physically more realistic, albeit computationally more expensive. Non-separable space-time covariance models can be constructed from Fourier transforms of permissible spectral densities, mixtures of separable models, and partial differential equations (PDEs) representing physical laws (Chen et al., 2021; Lindgren et al., 2022). They can be fully symmetric or asymmetric, stationary or non-stationary, univariate or multivariate, and in the Euclidean space or on the sphere. See Porcu et al. (2021) for a recent comprehensive review.

In this paper, we follow the dynamic approach that makes use of physical laws and study models which are defined through Stochastic Partial Differential Equations (SPDEs), where the stochasticity is obtained by adding a random noise as a forcing term. The SPDE approach relies on the representation of a continuously indexed Gaussian Field (GF) as a discretely indexed random process, i.e. a Gaussian Markov Random Field (GMRF, see Rue and Held (2005)). Passing from a GF to a GMRF, the covariance function and the dense covariance matrix are substituted respectively by a neighborhood structure and a sparse precision matrix. The advantage of using GMRFs is that the use of sparse precision matrices implies computationally efficient numerical methods, especially for matrix factorization. The link between GF and GMRFs in the purely spatial case has been pioneered by Lindgren et al. (2011), who proposed to construct a GMRF representation of the spatial Matérn field on a triangulated mesh of the domain through the discretization of a diffusion SPDE with a Finite Element Method (FEM). We refer to Bakka (2018) for a

simple explanation of FEM applied to the spatial SPDE or to Section 2.3 for a detailed generalization to spatio-temporal SPDE.

In the spatial framework, major mathematical and algorithmic advances in the SPDE approach have been made (Fuglstad et al., 2015; Pereira and Desassis, 2019; Pereira et al., 2022), making it possible to efficiently process very large datasets, even in the presence of non-stationarities and varying local anisotropy. The development of SPDE-based approaches to Gaussian processes has led to several practical solutions, among which we find the R package for approximate Bayesian inference R-INLA (Rue et al., 2009; Lindgren and Rue, 2015) that uses SPDEs to sample from spatial and spatio-temporal models.

When generalizing to the spatio-temporal framework, a direct space-time formulation of the SPDE approach was first suggested in Lindgren et al. (2011), without any precise detail on estimation and prediction methods. The SPDE approach was coupled with the Bayesian framework by Cameletti et al. (2011) to provide a separable space-time model. Non-separable spatio-temporal models have been elaborated in Krainski et al. (2018) and Lindgren et al. (2020) as a spatio-temporal generalization of the diffusion-Matérn model of Lindgren et al. (2011).

In all the approaches overviewed above, the space-time processes are symmetrical in the sense that the spatio-temporal covariance does not change when the sign of the space and/or time lag changes. However, atmospheric and geophysical processes are often asymmetric due to transport effects, such as air and water flows. Carrizo-Vergara et al. (2022) defined new spatio-temporal models incorporating the physical processes linked to the studied phenomena (advection, diffusion, etc.). Problems relating to the estimation of the parameters and the conditioning to the observed data remained however open. Sigrist et al. (2015) built non-symmetrical and non-separable space-time Gaussian as a solution to an advection-diffusion SPDE with computationally efficient algorithms for statistical estimation using fast Fourier transforms and Kalman filters. Liu et al. (2020) extended the approach to spatially-varying advection-diffusion and non-zero mean source-sink, leading to a space-time covariance which is non-stationary in space. The applicability of this approach remains difficult however, especially with scattered data, as it relies on the Fourier transform of the data.

In this work, we propose an alternative approach for dealing with spatio-temporal SPDEs that includes both a diffusion and an advection terms. In contrast to Sigrist et al. (2015), we make use of the sparse formulation of the spatio-temporal field which is the approximate solution of the SPDE obtained by a combination of FEM and finite differences. This sparse formulation allows to get fast algorithms for parameter estimation and spatio-temporal prediction. We also treat the case of an advection-dominated SPDE, by introducing the Streamline Diffusion stabilization term in the SPDE (Hughes and Brooks, 1981).

The paper is organized as follows: Section 2 first presents the background of the spatio-temporal SPDE approach, then defines the spatio-temporal

advection-diffusion model developed in the paper and its discretization. Moreover, the stabilization of advection-dominated SPDEs is introduced. Section 3 explores fast and scalable estimation methods, kriging formula for prediction and conditional simulations. Section 4 presents an application of the proposed spatio-temporal SPDE approach to a solar radiation dataset. Section 5 discusses the advantages and the limitations of the approach and opens the way to further works on the subject.

2 The spatio-temporal advection-diffusion SPDE and its discretization

2.1 Background

In the SPDE representation, Gaussian random fields on $\mathbb{R}^d$ are viewed as solutions to specific stochastic partial differential equations (Whittle, 1954, 1963). In particular, Gaussian Whittle-Matérn fields, analyzed in details in Lindgren et al. (2011) and reviewed in Lindgren et al. (2022), are solutions to

$$(\kappa^2 - \Delta)^{\alpha/2} X(\mathbf{s}) = \tau W(\cdot), \tag{1}$$

with $\alpha > d/2$ and $\tau > 0$. $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$ is the Laplacian operator and $W(\cdot)$ is a standard spatial Gaussian white noise, whose definition is briefly recalled: $W(\cdot)$ is as a Generalized GRF such that $\mathbb{E}[W(A)] = 0$ and $\mathbb{C}\text{ov}(W(A), W(B)) = |A \cap B|$ for any two Borelians of $\mathbb{R}^d$ and where $|A|$ is the Lebesgue measure of A over $\mathbb{R}^d$. In principle, $W(\cdot)$ has no elementwise definition, but for an easier reading, we will allow ourselves to write $W(\mathbf{s})$.

The covariance function of the Gaussian Whittle-Matérn field solution to Equation (1) is the well known Matérn covariance function

$$C_\nu^M(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa h)^\nu \mathcal{K}_\nu (\kappa h), \tag{2}$$

with smoothness parameter $\nu = \alpha - d/2 > 0$, scale parameter κ and 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$^{\text{nd}}$ order Bessel function and $h = \|\mathbf{s} - \mathbf{s}'\|$ is the Euclidean distance between the two locations $\mathbf{s}$ and $\mathbf{s}'$ in $\mathbb{R}^d$. In particular, when $\nu = 1/2$ we get the exponential covariance function and when $\nu \to +\infty$, after proper renormalization, (2) tends to the Gaussian covariance function.

In Lindgren et al. (2011), the smoothness parameter ν considered in the Matérn covariance function corresponds to integer values of α. When non-integer values of ν are introduced in the modeling, the SPDE is said to be fractional. Recent reviews of results and applications of the fractional SPDE approach are available in Xiong et al. (2022); Bolin and Kirchner (2020); Roques et al. (2022), but this case will not be treated further in this work.

When generalizing to spatio-temporal processes $X(\mathbf{s}, t)$, we consider the framework proposed in Carrizo-Vergara et al. (2022) for extending the SPDE approach to a wide class of linear spatio-temporal SPDEs. Let us denote $\boldsymbol{\xi} \in \mathbb{R}^d$ a spatial frequency and $\omega \in \mathbb{R}$ a temporal frequency. The space-time white noise with unit variance, denoted $W(\mathbf{s}, t)$, is characterized by its spectral measure $d\mu_W(\boldsymbol{\xi}, \omega) = (2\pi)^{-(d+1)} d\boldsymbol{\xi}\, d\omega$. New spatio-temporal models were obtained from known PDEs describing physical processes, such as diffusion, advection, and oscillations with stochastic forcing terms. In particular, Carrizo-Vergara et al. (2022) provides sufficient conditions to the existence and uniqueness of stationary solutions to

$$\left[\frac{\partial^\beta}{\partial t^\beta} + \mathcal{L}_g\right] X(\mathbf{s}, t) = W_S(\mathbf{s}) \otimes W_T(t). \tag{3}$$

In (3), the spatial operator $\mathcal{L}_g$ is defined using the spatial Fourier transform on $\mathbb{R}^d$, denoted $\mathcal{F}_S$,

$$\mathcal{L}_g(\cdot) = \mathcal{F}_S^{-1}(g\mathcal{F}_S(\cdot)),$$

where $g : \mathbb{R}^d \to \mathbb{C}$ is a sufficiently regular and Hermitian-symmetric function called the *symbol function* of the operator $\mathcal{L}_g$. The temporal operator $\frac{\partial^\beta}{\partial t^\beta}$ is

$$\frac{\partial^\beta}{\partial t^\beta}(\cdot) = \mathcal{F}_T^{-1}((i\omega)^\beta \mathcal{F}_T(\cdot)),$$

where $\mathcal{F}_T$ is the temporal Fourier transform on $\mathbb{R}$ and where we have used the symbol function over $\mathbb{R}$

$$\omega \mapsto (i\omega)^\beta = |\omega|^\beta e^{i\,\mathrm{sgn}(\omega)\beta\pi/2}.$$

The spatio-temporal symbol function of the operator involved in (3) is thus

$$(\boldsymbol{\xi}, \omega) \mapsto (i\omega)^\beta + g(\boldsymbol{\xi}) = |\omega|^\beta \cos\left(\frac{\beta\pi}{2}\right) + g_R(\boldsymbol{\xi}) + i\left(\mathrm{sgn}(\omega)|\omega|^\beta \sin\left(\frac{\beta\pi}{2}\right) + g_I(\boldsymbol{\xi})\right)$$

where g_R and g_I are the real and imaginary part of the spatial symbol function $g(\boldsymbol{\xi})$. Theorem 1 and Proposition 3 in Carrizo-Vergara et al. (2022) state that (3) admits a unique stationary solution for every arbitrary g_I function if $|g_R|$ is inferiorly bounded by the inverse of a strictly positive polynomial and $g_R \cos\left(\frac{\beta\pi}{2}\right) \geq 0$.

2.2 The spatio-temporal advection-diffusion SPDE

The advection-diffusion equation is a Partial Differential Equation (PDE) that describes physical phenomena where particles, energy, or other physical quantities evolve inside a physical system due to two processes: diffusion and advection. Advection represents the mass transport due to the average velocity

of all particles, and diffusion represents the mass transport due to the instantaneously varying velocity of individual particles. The advection-diffusion SPDE studied in this work writes

$$\left[\frac{\partial}{\partial t} + \frac{1}{c}(\kappa^2 - \nabla \cdot \mathbf{H}\nabla)^\alpha + \frac{1}{c}\boldsymbol{\gamma}\cdot\nabla\right] X(\mathbf{s},t) = \frac{\tau}{\sqrt{c}} Z(\mathbf{s},t), \tag{4}$$

where

- the operator $\nabla \cdot \mathbf{H}\nabla$ is a *diffusion* term that can incorporate *anisotropy* in the matrix $\mathbf{H}$. When the field is isotropic, i.e. when $\mathbf{H} = \lambda\mathbf{I}$, this term reduces to the Laplacian operator $\lambda\Delta$;
- the operator $\boldsymbol{\gamma}\cdot\nabla$ models the *advection*, $\boldsymbol{\gamma} \in \mathbb{R}^d$ being a velocity vector;
- $\alpha \geq 0$ relates to the smoothness of $X(\cdot,t)$, $\kappa^2 > 0$ accounts for *damping* and c is a positive time-scale parameter;
- $\tau \geq 0$ is a standard deviation factor and Z is a stochastic forcing term. From now on, we will assume a Gaussian distribution for Z.

This equation was mentioned in Lindgren et al. (2011), Carrizo-Vergara et al. (2022) and Lindgren et al. (2020), and was analyzed using spectral approaches in Sigrist et al. (2015) and Liu et al. (2020). The stochastic forcing term $Z(\mathbf{s},t)$ is assumed separable with

$$Z(\mathbf{s},t) = W_T(t) \otimes Z_S(\mathbf{s}),$$

where Z_S is a spatial Generalized Random Field and W_T is a temporal white noise. Z_S is often chosen to be a spatial white noise, denoted W_S in this case. To ensure a sufficient regularity for Z, Z_S can alternatively be a *colored noise*, such as for example the solution to the spatial Whittle-Matérn SPDE (Lindgren et al., 2011)

$$(\kappa^2 - \nabla \cdot \mathbf{H}\nabla)^{\alpha_S/2} Z_S(\mathbf{s}) = W_S(\mathbf{s}), \tag{5}$$

where W_S is a Gaussian white noise. Notice that the parameter κ^2 in the forcing term has been set identical to that in the diffusion term in the left-hand-side of (4) to make sure that the spatial marginalization of the process is a Matérn field, as detailed below.

When $\alpha > 0$, $X(\mathbf{s},t)$ is a non-separable spatio-temporal field. The advection-diffusion equation (4) is a particular first order evolution model as in Equation (3) with $\beta = 1$. Its spatial symbol function

$$g(\boldsymbol{\xi}) = \frac{1}{c}\left[(\kappa^2 - \boldsymbol{\xi}^\top \mathbf{H}\boldsymbol{\xi})^\alpha + i\boldsymbol{\gamma}^\top\boldsymbol{\xi}\right],$$

verifies the sufficient condition for existence and uniqueness of a stationary solution recalled at the end of Section 2.1. It can be shown that the advection term does not affect the spatial trace of the solution, as stated in the following Proposition 1.

Proposition 1 *Let $Z(\mathbf{s},t)$ be a spatio-temporal noise colored in space (such that $Z_S(\mathbf{s})$ satisfies (5)), and let define $\alpha_{tot} = \alpha + \alpha_S$. If $\alpha_{tot} > d/2$, the spatial trace of the stationary solution $X(\mathbf{s},t)$ of the SPDE (4) is a Gaussian Matérn field with covariance (2) with $\nu = \alpha_{tot} - d/2$ and marginal variance σ^2 equal to*

$$\sigma^2 = \frac{\tau^2 \Gamma(\alpha_{tot} - d/2)}{\Gamma(\alpha_{tot}) 2 (4\pi)^{d/2} \kappa^{2(\alpha_{tot}-d/2)} |\mathbf{H}|^{1/2}}. \tag{6}$$

Proposition 1 is adapted from Proposition 1 in Lindgren et al. (2020). For completeness, a proof including an anisotropy matrix $\mathbf{H}$ is reported in Appendix A.

Corollary 2 *Let the coefficients of the SPDE (4) be such that $\alpha = 0$ and $\boldsymbol{\gamma} = \mathbf{0}$; the spatial operator applied to the spatio-temporal field $X(\mathbf{s},t)$ is then the constant value c^{-1}. Let $Z(\mathbf{s},t)$ be a spatio-temporal noise colored in space, with $Z_S(\mathbf{s})$ satisfying (5). If $\alpha_S > d/2$, the stationary solution of the SPDE is a separable spatio-temporal field whose covariance is the product of an exponential temporal covariance (with scale parameter equal to c) and a Matérn spatial covariance (2) with scale parameter equal to κ, smoothness parameter $\nu = \alpha_S - d/2$ and marginal variance equal to*

$$\sigma^2 = \frac{\tau^2 \Gamma(\alpha_S - d/2)}{2\Gamma(\alpha_S)(4\pi)^{d/2} \kappa^{2(\alpha_S - d/2)} |\mathbf{H}|^{1/2}}.$$

The corollary is a particular case of Proposition 1.

2.3 Discretization

The advection-diffusion SPDE in (4) is discretized in time and space, using a Finite Difference Method (FDM) and a Finite Element Method (FEM), respectively. The temporal domain $[1, T]$ is discretized in N_T regular time steps of length T/N_T. Since implicit solvers are usually less sensitive to numerical instability than explicit solvers, the implicit Euler scheme is chosen for the temporal discretization. This choice implies stability, hence convergence towards the stationary solution. The FEM method for the spatial discretization is the continuous Galerkin method with Neumann Boundary Conditions as detailed in Lindgren et al. (2011).

The solution in two dimensions is now detailed. The solution in three dimensions involve geometrical technicalities, but is otherwise very similar. Let $\Omega \subset \mathbb{R}^2$ be a compact and connected domain of $\mathbb{R}^2$. Ω is meshed using a triangulation $\mathcal{T}$ with N_S vertices $\{\mathbf{s}_1, \ldots, \mathbf{s}_{N_S}\} \subset \Omega$. Let $h := \max_{\mathrm{Tr} \in \mathcal{T}} h_{\mathrm{Tr}}$, where h_{Tr} is the diameter of the triangle $\mathrm{Tr} \in \mathcal{T}$. A first order finite element representation X_h of the solution to the spatial SPDE is a linear combination $X_h = \sum_{i=1}^{N_S} x_i \psi_i$ of piecewise linear basis functions $\{\psi_i\}_{i=1}^{N_S}$, each ψ_i being equal to 1 at the vertex $\mathbf{s}_i$ and 0 at all the other vertices. The weights $\{x_i\}_{i=1}^{N_S}$ define uniquely the values of the field at the vertices, while the values in the interior of the triangles are determined by linear interpolation. The continuous Galerkin solution is then obtained by finding the weights that fulfill the

weak formulation of Equation (4) for test functions belonging to the space $\mathcal{V}$ spanned by $\{\psi_i\}_{i=1}^{N_S}$.

Proposition 3 *Let $X(\mathbf{s},t)$ be the spatio-temporal process solution to Equation (4) with $\mathbf{H}=\mathbf{I}$, $\alpha \in \{0,1\}$ and spatio-temporal white noise, i.e. $Z(\mathbf{s},t)=W(\mathbf{s},t)=W_T(t)\otimes W_S(\mathbf{s})$. Let $\mathcal{T}$ be a triangulation of Ω and $\{\psi_i\}_{i=1}^{N_S}$ be the piecewise linear basis functions defined over $\mathcal{T}$. Let us define the mass matrix $\mathbf{M}=[M_{ij}]_{i,j=1}^{N_S}$, the stiffness matrix $\mathbf{G}=[G_{ij}]_{i,j=1}^{N_S}$, the advection matrix $\mathbf{B}=[B_{ij}]_{i,j=1}^{N_S}$ and the matrix $\mathbf{K}=[K_{ij}]_{i,j=1}^{N_S}$ as follows:*

$$
\begin{aligned}
M_{ij} &= \int_\Omega \psi_i(\mathbf{s})\psi_j(\mathbf{s})\,\mathrm{d}\,\mathbf{s}\,,\\
G_{ij} &= \int_\Omega \nabla\psi_i(\mathbf{s})\cdot\nabla\psi_j(\mathbf{s})\,\mathrm{d}\,\mathbf{s}\,,\\
B_{ij} &= \int_\Omega \boldsymbol{\gamma}\cdot\nabla\psi_i(\mathbf{s})\psi_j(\mathbf{s})\,\mathrm{d}\,\mathbf{s}\,,\\
(K_{\kappa^2})_{ij} &= (\kappa^2 M_{ij}+G_{ij})^\alpha.
\end{aligned}
$$

Then, at each time step, the continuous Galerkin finite element solution vector $\mathbf{x}_{t+dt}=\{x_{t+dt,i}\}_{i=1}^{N_S}$ satisfies

$$
\left(\mathbf{M}+\frac{dt}{c}(\mathbf{K}+\mathbf{B})\right)\mathbf{x}_{t+dt}=\mathbf{M}\,\mathbf{x}_t+\frac{\tau\sqrt{dt}}{\sqrt{c}}\,\mathbf{M}^{1/2}\,\mathbf{z}_{t+dt}, \tag{7}
$$

where $\mathbf{z}_{t+dt}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_{N_S})$, $\mathbf{M}^{1/2}$ is any matrix such that $\mathbf{M}^{1/2}\,\mathbf{M}^{1/2}=\mathbf{M}$ and $dt=T/N_T$. When the noise on the right-hand side is colored in space, i.e. $Z(\mathbf{s},t)=W_T(t)\otimes Z_S(\mathbf{s})$, the discretization reads

$$
\left(\mathbf{M}+\frac{dt}{c}(\mathbf{K}+\mathbf{B})\right)\mathbf{x}_{t+dt}=\mathbf{M}\,\mathbf{x}_t+\frac{\tau\sqrt{dt}}{\sqrt{c}}\,\mathbf{M}\,\mathbf{L}_S^{\top}\,\mathbf{z}_{t+dt},
$$

where $\mathbf{L}_S$ is the Cholesky decomposition of $\mathbf{Q}_S^{-1}$, the covariance matrix of the discretized solution of the spatial SPDE (5), obtained with the continuous Galerkin FEM (Lindgren et al., 2011).

Proof The proof is available in Appendix B. □

Remark that the elements of the matrices $\mathbf{M}$, $\mathbf{G}$, $\mathbf{B}$ and $\mathbf{K}$ are non-zero only for pairs of basis functions which share common triangles. This implies that the matrix $(\mathbf{M}+\frac{dt}{c}(\mathbf{K}+\mathbf{B}))$ is sparse and that Equation (7) can be solved by Cholesky decomposition in an efficient way.

2.4 Stabilization of advection-dominated SPDE

When the advection term is too strong with respect to the diffusion term, advection-domination occurs. In the framework outlined above, when $\alpha=1$, the non-symmetric matrix $\left[\mathbf{M}+\frac{dt}{c}(\mathbf{K}+\mathbf{B})\right]$ becomes ill-conditioned, which induces oscillations and unstable solutions for the continuous Galerkin approximation. Specifically, the advection-domination occurs when the Péclet number

$\text{Pe}^h = \frac{\|\boldsymbol{\gamma}\|h}{2\lambda} > 1$, where λ is the coefficient of the isotropic Laplacian operator (see for example Mekuria and Rao (2016) or Quarteroni (2008, Chapter 5)).

One possible solution is to decrease the diameter h, i.e., to refine the triangulation, until the advection no longer dominates on the element-level, with $\text{Pe}^h < 1$. However, in many cases this is not a feasible solution because it would increase the number of vertices beyond computation limits. Another solution, adopted here, is to introduce a stabilization term. Many stabilization approaches are possible, some being more accurate than others (Quarteroni, 2008, Chapter 5). In our case, we opt for the Streamline Diffusion (SD) stabilization approach (Hughes and Brooks, 1981), considered as a good trade-off between accuracy and computational complexity. Essentially, the SD approach consists in stabilizing the advection by introducing an artificial diffusion term along the advection direction. A detailed presentation of the stabilization approach is reported in Appendix C. The following proposition presents the stabilized solution to (4).

Proposition 4 *Assume the same hypotheses as in Proposition 3 with $\alpha = 1$. The solution to Equation (4) in presence of Streamline Diffusion stabilization is*

$$\left(\mathbf{M} + \frac{dt}{c}(\mathbf{K} + \mathbf{B} + \mathbf{S})\right)\mathbf{x}_{t+dt} = \mathbf{M}\,\mathbf{x}_t + \frac{\tilde{\tau}\sqrt{dt}}{\sqrt{c}}\,\mathbf{M}^{1/2}\,\mathbf{z}_{t+dt}, \tag{8}$$

where $\mathbf{S} = [S_{ij}]_{i,j=1}^{N_S}$ is the matrix of the Streamline Diffusion stabilization operator $\mathcal{S}$, such that $S_{ij} = \mathcal{S}(\psi_i, \psi_j) = h|\boldsymbol{\gamma}|^{-1}\int_\Omega(\boldsymbol{\gamma}\cdot\nabla\psi_i)(\boldsymbol{\gamma}\cdot\nabla\psi_j)\,\mathrm{d}\,\mathbf{s}$, and $\tilde{\tau} = \tau\left(|\mathbf{H} + h|\boldsymbol{\gamma}|^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top|\right)^{-1/4}(|\mathbf{H}|)^{1/4}$. When the noise on the right-hand side of Equation (4) is colored in space, the discretization becomes

$$\left(\mathbf{M} + \frac{dt}{c}(\mathbf{K} + \mathbf{B} + \mathbf{S})\right)\mathbf{x}_{t+dt} = \mathbf{M}\,\mathbf{x}_t + \frac{\tilde{\tau}\sqrt{dt}}{\sqrt{c}}\,\mathbf{M}\,\mathbf{L}_S^\top\,\mathbf{z}_{t+dt},$$

where $\mathbf{L}_S$ is as in Proposition 3.

The proof of the discretized equation follows the same reasoning as that of Proposition 3 with the addition of the matrix $\mathbf{S}$. The Streamline Diffusion approach can be seen as a perturbation of the original SPDE (Bank et al., 1990). Indeed, by making the classical hypothesis of Neumann boundary condition on Ω and by using the Green's first identity, we get

$$\int_\Omega(\boldsymbol{\gamma}\cdot\nabla x)(\boldsymbol{\gamma}\cdot\nabla v)\,\mathrm{d}\,\mathbf{s} = -\int_\Omega \nabla\cdot(\boldsymbol{\gamma}\boldsymbol{\gamma}^\top)\nabla x v\,\mathrm{d}\,\mathbf{s}\,.$$

As a consequence, the original SPDE (4) can be rewritten with an additional diffusion term as

$$\left[\frac{\partial}{\partial t} + \frac{1}{c}\left(\kappa^2 - \nabla\cdot\left(\mathbf{H} + h|\boldsymbol{\gamma}|^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top\right)\nabla + \boldsymbol{\gamma}\cdot\nabla\right)\right]X(\mathbf{s},t) = \frac{\tilde{\tau}}{\sqrt{c}}Z(\mathbf{s},t). \tag{9}$$

The term $(h|\boldsymbol{\gamma}|^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top)$ acts as an anisotropic "diffusion" matrix that is added to the anisotropy (or identity) matrix $\mathbf{H}$ of the original diffusion. This extra diffusion stabilizes the advection directed along the direction $\boldsymbol{\gamma}$. By following the proof of Proposition 1, we find that the variance of the field $X(\mathbf{s},t)$ of Equation (9) is equal to

$$\sigma^2 = \frac{\tau^2\Gamma(\alpha_{tot}-d/2)}{\Gamma(\alpha_{tot})2(4\pi)^{d/2}\kappa^{2(\alpha_{tot}-d/2)}|\mathbf{H}+h|\boldsymbol{\gamma}|^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top|^{1/2}}.$$

For the variance to be equal to the variance in Proposition 1, τ must be replaced by $\tilde{\tau} = \tau\left(|\mathbf{H}+h|\boldsymbol{\gamma}|^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top|\right)^{-1/4}(|\mathbf{H}|)^{1/4}$.

2.5 Spatio-temporal Gaussian Markov Random Field approximation

Proposition 5 *In presence of an advection-dominated flow and a spatio-temporal white noise on the right-hand side of Equation (4), the discretized vector* $\mathbf{x}_{t+dt}$ *on the mesh* $\mathcal{T}$ *at each time step is the solution of the following equation:*

$$\begin{aligned} \mathbf{x}_1 &\sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}),\\ \mathbf{x}_{t+dt} &= \mathbf{D}\,\mathbf{x}_t + \mathbf{E}\,\mathbf{z}_{t+dt}, \quad \forall t>1 \end{aligned} \tag{10}$$

where

$$\begin{aligned} \mathbf{D} &= \left(\mathbf{M}+\frac{dt}{c}(\mathbf{K}+\mathbf{B}+\mathbf{S})\right)^{-1}\mathbf{M},\\ \mathbf{E} &= \frac{\tilde{\tau}\sqrt{dt}}{\sqrt{c}}\left(\mathbf{M}+\frac{dt}{c}(\mathbf{K}+\mathbf{B}+\mathbf{S})\right)^{-1}\mathbf{M}^{1/2}, \end{aligned} \tag{11}$$

and $\mathbf{z}_{t+dt} \sim \mathcal{N}(\mathbf{0},\mathbf{I}_{N_S})$ *is independent of* $\mathbf{x}_1,\ldots,\mathbf{x}_{t+dt}$.

In presence of a spatio-temporal noise colored in space on the right-hand side of Equation (4), the matrix $\mathbf{E}$ *reads*

$$\mathbf{E} = \frac{\tilde{\tau}\sqrt{dt}}{\sqrt{c}}\left(\mathbf{M}+\frac{dt}{c}(\mathbf{K}+\mathbf{B}+\mathbf{S})\right)^{-1}\mathbf{M}\,\mathbf{L}_S^\top,$$

where $\mathbf{L}_S$ *is defined in Proposition 3.*

Proof Starting from Equation (8), which represents the numerical scheme for the advection-diffusion spatio-temporal SPDE with stabilization, it is straightforward to obtain (10). □

When the SPDE is not advection-dominated, which implies that no stabilization term is needed, Equation (11) is replaced by the similar equation where the matrix $\mathbf{S}$ is deleted and where $\tilde{\tau}$ is replaced by τ.

The covariance matrix $\boldsymbol{\Sigma}$ can be taken to be equal to any admissible positive definite matrix. The speed of convergence to the stationary solution of the equation will depend on the proximity of the covariance of the spatial trace of $X(\mathbf{s},t)$ to $\boldsymbol{\Sigma}$. When the noise is colored in space, an efficient option is to choose

$\mathbf{\Sigma}$ so as to correspond to the Matérn covariance of the spatial trace defined in Proposition 1. For this reason κ is identical in both hand-sides of Equation (4).

To obtain a GMRF representation of the spatio-temporal discretized solution $\mathbf{x}_{1:N_T} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_T}]^\top$, its precision matrix must be sparse. For this reason $\mathbf{M}$ is replaced by the diagonal matrix $\widetilde{\mathbf{M}}$, where $\widetilde{\mathbf{M}}_{ii} = \langle \psi_i, 1 \rangle$ (Lindgren et al., 2011). This technique is called mass lumping and is common practice in FEM (Quarteroni, 2008, Chapter 5). From now on, we always use the diagonal matrix $\widetilde{\mathbf{M}}$, but for ease of reading, it will still be denoted $\mathbf{M}$.

Proposition 6 *Let $\mathbf{x}_{1:N_T} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_T}]^\top$ be the vector containing all spatial solutions until time step N_T of Equation (10). The global precision matrix $\mathbf{Q}$ of the vector $\mathbf{x}_{1:N_T}$ of size $(N_S N_T, N_S N_T)$ reads*

$$\mathbf{Q} = \begin{pmatrix} \mathbf{\Sigma}^{-1} + \mathbf{D}^\top \mathbf{F}^{-1} \mathbf{D} & -\mathbf{D}^\top \mathbf{F}^{-1} & 0 & \ldots & 0 \\ -\mathbf{F}^{-1} \mathbf{D} & \mathbf{F}^{-1} + \mathbf{D}^\top \mathbf{F}^{-1} \mathbf{D} & -\mathbf{D}^\top \mathbf{F}^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -\mathbf{F}^{-1} \mathbf{D} & \mathbf{F}^{-1} + \mathbf{D}^\top \mathbf{F}^{-1} \mathbf{D} & -\mathbf{D}^\top \mathbf{F}^{-1} \\ 0 & \ldots & 0 & -\mathbf{F}^{-1} \mathbf{D} & \mathbf{F}^{-1} \end{pmatrix}, \tag{12}$$

where $\mathbf{F} = \mathbf{E}\,\mathbf{E}^\top$.

Proof The proof is available in Appendix D. □

3 Estimation, prediction and simulation

This section presents how the estimation of the parameters and the spatio-temporal prediction can be efficiently implemented within the spatio-temporal SPDE framework described in Section 2. We will consider the advection-diffusion SPDE (4) with $d = 2$, $\alpha = 1$, $\mathbf{H} = \mathbf{I}_{N_S}$ (isotropic diffusion) and colored noise in space with $\alpha_S = 2$. Similar computations can be generalized to other values of α_S such that $\alpha_S/2$ is integer or to anisotropic diffusion.

We consider n spatio-temporal data $\mathbf{y}_{1:N_T} = [\mathbf{y}_1, \ldots, \mathbf{y}_{N_T}]^\top$ scattered in the spatio-temporal domain $\Omega \times [1, T]$, discretized in space with a triangulation $\mathcal{T}$ with N_S nodes and discretized in time by means of N_T regular time steps. We denote this space-time discretization $\mathcal{T}' = \mathcal{T} \times \{T/N_T, 2T/N_T, \ldots, T\}$. We consider a statistical model with fixed and random effects. The fixed effect is a regression on a set of covariates and the random effect is modeled as the FEM discretization of a random field described by the SPDE (4) with the addition of random noise:

$$\mathbf{y}_{1:N_T} = \boldsymbol{\eta}\,\mathbf{b} + \mathbf{A}^\top \mathbf{x}_{1:N_T} + \sigma_0\,\boldsymbol{\varepsilon}, \tag{13}$$

where $\mathbf{b}$ is the vector of q fixed effects and $\boldsymbol{\eta}$ is a (n, q) matrix of covariates with $[\boldsymbol{\eta}]_{jk} = \eta_k(\mathbf{s}_j, t_j)$, $j = 1 \ldots, n$ and $k = 1, \ldots, q$. The matrix $\mathbf{A}$ is the $(N_S N_T, n)$ projection matrix between the points in $\mathcal{T}'$ and the data, and $\boldsymbol{\varepsilon}$ is a standard Gaussian random vector with independent components. When the locations

of the observation stations do not change during the time window, $\mathbf{A}\,\mathbf{A}^\top$ is a $(N_S N_T, N_S N_T)$ block-diagonal matrix with all (N_S, N_S) equal blocks.

3.1 Estimation of the parameters

The parameters of the SPDE are estimated using Maximum Likelihood. We collect the parameters of the SPDE in the vector $\boldsymbol{\theta}^\top = (\kappa, \gamma_x, \gamma_y, c, \tau)$, while the parameters of the statistical model are collected in $\boldsymbol{\psi}^\top = (\boldsymbol{\theta}^\top, \mathbf{b}^\top, \sigma_0)$. Following (13), $\mathbf{y}_{1:N_T}$ is a Gaussian vector with expectation $\boldsymbol{\eta}\,\mathbf{b}$ and covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{y}_{1:N_T}} = \mathbf{A}^\top\,\mathbf{Q}^{-1}(\boldsymbol{\theta})\,\mathbf{A} + \sigma_0^2\,\mathbf{I}_n,$$

where $\mathbf{Q}(\boldsymbol{\theta})$ is a precision matrix of size $(N_S N_T, N_S N_T)$ depending on the parameters $\boldsymbol{\theta}$. For ease of notation, we use $\mathbf{Q}$ instead of $\mathbf{Q}(\boldsymbol{\theta})$. The log-likelihood is equal to

$$\mathcal{L}(\boldsymbol{\psi}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{\mathbf{y}_{1:N_T}}(\boldsymbol{\psi})| - \frac{1}{2}(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\,\mathbf{b})^\top\,\boldsymbol{\Sigma}^{-1}_{\mathbf{y}_{1:N_T}}(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\,\mathbf{b}). \tag{14}$$

We use the Broyden, Fletcher, Goldfarb, and Shanno optimization algorithm (Nocedal and Wright, 2006), that makes use of the second-order derivative of the objective function. The gradients of the log-likelihood function (14) with respect to the different parameters included in $\boldsymbol{\psi}$ are approximately computed with a Finite Difference Method. We now propose a computationally efficient formulation of each term of the log-likelihood (14).

Proposition 7 *In the framework outlined above, we have*

$$\log|\boldsymbol{\Sigma}_{\mathbf{y}_{1:N_T}}| = n\log\sigma_0^2 - \log|\mathbf{Q}| + \log|\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top|. \tag{15}$$

Proof To compute $\log|\boldsymbol{\Sigma}_{\mathbf{y}_{1:N_T}}|$, let us consider the augmented matrix

$$\boldsymbol{\Sigma}_c = \begin{pmatrix} \mathbf{Q}^{-1} & \mathbf{Q}^{-1}\,\mathbf{A} \\ \mathbf{A}^\top\,\mathbf{Q}^{-1} & \boldsymbol{\Sigma}_{\mathbf{y}_{1:N_T}} \end{pmatrix}. \tag{16}$$

Hence,

$$\mathbf{Q}_c = \boldsymbol{\Sigma}_c^{-1} = \begin{pmatrix} \mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top & -\sigma_0^{-2}\,\mathbf{A} \\ -\sigma_0^{-2}\,\mathbf{A}^\top & \sigma_0^{-2}\,\mathbf{I}_n \end{pmatrix}. \tag{17}$$

By using block formulas, we have

$$\log|\boldsymbol{\Sigma}_c| = -\log|\mathbf{Q}_c| = -\log|\mathbf{Q}| + n\log\sigma_0^2,$$

and

$$\begin{aligned} \log|\boldsymbol{\Sigma}_c| &= \log|\boldsymbol{\Sigma}_{\mathbf{y}_{1:N_T}}| + \log|\mathbf{Q}^{-1} - \mathbf{Q}^{-1}\,\mathbf{A}\,\boldsymbol{\Sigma}^{-1}_{\mathbf{y}_{1:N_T}}\,\mathbf{A}^\top\,\mathbf{Q}^{-1}| \\ &= \log|\boldsymbol{\Sigma}_{\mathbf{y}_{1:N_T}}| - \log|\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top|, \end{aligned}$$

where the last equality is a consequence of the Woodbury identity. This leads to the result. □

Proposition 8 *The term* $\log|\mathbf{Q}|$ *in Equation* (15) *can be computed with the computationally cheaper formula*

$$\log|\mathbf{Q}| = \log|\mathbf{\Sigma}^{-1}| + (N_T - 1)\log|\mathbf{F}^{-1}|, \tag{18}$$

with

$$\mathbf{F}^{-1} = \frac{c}{\tilde{\tau}^2 dt}(\mathbf{M} + \frac{dt}{c}(\mathbf{K} + \mathbf{B} + \mathbf{S}))^\top \mathbf{M}^{-1}(\mathbf{M} + \frac{dt}{c}(\mathbf{K} + \mathbf{B} + \mathbf{S})).$$

When the noise is colored in space, $\mathbf{M}^{-1}$ *must be replaced by* $\mathbf{M}^{-1}\,\mathbf{Q}_S\,\mathbf{M}^{-1}$.

Note that $|\mathbf{F}^{-1}|$ is now the determinant of a (N_S, N_S) sparse, symmetric and positive definite matrix. The computation of its determinant can be obtained by Cholesky decomposition of $\mathbf{F}^{-1}$.

Proof Following Powell (2011), let $\mathbf{N}_N = [\mathbf{N}_{i\,j}]_{i,j=1}^N$ be an (nN, nN) matrix, which is partitioned into N blocks, each of size (n, n). Then the determinant of $\mathbf{N}_N$ is

$$|\mathbf{N}_N| = \prod_{k=1}^{N} |\alpha_{kk}^{(N-k)}|,$$

where the $\alpha^{(k)}$ are defined by

$$\begin{aligned} \alpha_{ij}^{(0)} &= \mathbf{N}_{ij} \\ \alpha_{ij}^{(k+1)} &= \alpha_{ij}^{(k)} - \alpha_{i,N-k}^{(k)}(\alpha_{N-k,N-k}^{(k)})^{-1}\alpha_{N-k,j}^{(k)}, \quad k \geq 1. \end{aligned}$$

$\mathbf{Q}$ is a block-matrix organized as $\mathbf{N}_N$. Hence, the formula for $|\mathbf{Q}|$ is

$$|\mathbf{Q}| = |\mathbf{\Sigma}^{-1}||\mathbf{F}^{-1}|^{N-1}. \tag{19}$$

Applying the logarithm, we obtain Equation (18). □

The term $\log|\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top|$ requires a detailed analysis. The term $\sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top$ is an $(N_S N_T, N_S N_T)$ diagonal block matrix, whose (N_S, N_S) blocks are sparse. The computation of $\log|\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top|$ is not as straightforward as in the case of $\log|\mathbf{Q}|$, because there is no way of reducing the computation to purely spatial matrices. Depending on the size $N_S N_T$, we can either apply a Cholesky decomposition of the $(N_S N_T, N_S N_T)$ matrix $(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)$ or the matrix-free approach proposed in Pereira et al. (2022). For the sake of completeness, the approach is briefly sketched. The logarithm function is first approximated by a Chebyschev polynomial $P(\cdot)$, then the Hutchinson's estimator (Hutchinson, 1990) is used to obtain a stochastic estimate of $\mathrm{tr}[P(|\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top|)]$. The method is detailed in Algorithm 5 in Pereira et al. (2022).

Concerning the computation of the quadratic term of the log-likelihood, we can work with the more convenient expression obtained thanks to the Woodbury formula

$$\mathbf{\Sigma}^{-1}_{\mathbf{y}_{1:N_T}} = \sigma_0^{-2}\,\mathbf{I}_n - \sigma^{-4}\,\mathbf{A}^\top(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)^{-1}\,\mathbf{A}\,.$$

Hence

$$(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\,\mathbf{b})^\top\,\boldsymbol{\Sigma}^{-1}_{\mathbf{y}_{1:N_T}}(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\,\mathbf{b}) = \sigma_0^{-2}(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\,\mathbf{b})^\top\,\mathbf{I}_n(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\,\mathbf{b}) - \sigma_0^{-4}(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\,\mathbf{b})^\top\,\mathbf{A}^\top(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)^{-1}\,\mathbf{A}(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\,\mathbf{b}).$$

The second term can be computed either by Cholesky decomposition or using the Conjugate Gradient method. This latter method solves $\mathbf{N}\,\mathbf{v} = \mathbf{w}$ with respect to $\mathbf{v}$ and computes $\mathbf{v}_{sol} = \mathbf{w}^\top\,\mathbf{v}$, with $\mathbf{N} = (\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)$ and $\mathbf{w} = \mathbf{A}(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\,\mathbf{b})$. In this case, it is useful to find a good preconditioner for the matrix $(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)$ to ensure fast convergence of the conjugate gradient method. We found that a temporal block Gauss-Seidel preconditioner was a good choice in this case. A detailed presentation of the Conjugate Gradient method and the Gauss-Seidel preconditioner is available in Appendix E.

3.2 Prediction by Kriging

Under a Gaussian assumption, optimal prediction is the conditional expectation, also known in the geostatistics literature as kriging. We here detail two prediction settings: space-time interpolation and temporal extrapolation.

In the space-time interpolation setting, the spatio-temporal vector $\mathbf{x}_{1:N_T}$ is predicted on the entire spatial mesh during the time window $[1, T]$, i.e. on $\mathcal{T}'$, using the data $\mathbf{y}_{1:N_T}$ defined in Equation (13). The kriging predictor is directly read from Equation (17):

$$\mathbf{x}^\star_{1:N_T} = \mathbb{E}(\mathbf{x}_{1:N_T} \mid \mathbf{y}_{1:N_T}) = \sigma_0^{-2}(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)^{-1}\,\mathbf{A}(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\widehat{\mathbf{b}}). \tag{20}$$

The computation of (20) requires the inversion of $(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)$, see Section 3.1. The conditional variance, also called kriging variance, is

$$\mathrm{Var}(\mathbf{x}_{1:N_T}|\mathbf{y}_{1:N_T}) = (\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)^{-1}.$$

The computation of the diagonal of an inverse matrix is not straightforward when only the Cholesky decomposition of the matrix is available. Among the existing methods there is the Takahashi recursive algorithm described in Takahashi et al. (1973) and Erisman and Tinney (1975). Another way of computing the kriging variance is through conditional simulations, as detailed in Section 3.3.

In the temporal extrapolation setting, the vector $\mathbf{x}_{N_T+1}$ is predicted at time step $(N_T + 1)$ on $\mathcal{T}$ using all the data available until time T, i.e. from $\mathbf{y}_{1:N_T}$. Following Equation (10), we have

$$\mathbf{x}_{N_T+1} = \mathbf{D}\,\mathbf{x}_{N_T} + \mathbf{E}\,\mathbf{z}_{N_T+1}, \tag{21}$$

where $\mathbf{z}_{N_T+1}$ is a standardized Gaussian vector.

The kriging predictor $\mathbf{x}^{\star}_{N_T+1}$ is

$$\mathbf{x}^{\star}_{N_T+1} = \mathbb{E}(\mathbf{x}_{N_T+1} \mid \mathbf{y}_{1:N_T}) = \mathbf{D}\,\mathbb{E}(\mathbf{x}_{N_T} \mid \mathbf{y}_{1:N_T}) = \mathbf{D}\,\mathbf{x}^{\star}_{N_T}, \tag{22}$$

where $\mathbf{x}^{\star}_{N_T}$ is extracted from $\mathbf{x}^{\star}_{1:N_T}$. The same procedure can be iterated to predict $\mathbf{x}$ at further time steps.

3.3 Conditional simulations

To perform a conditional simulation, we use the conditional kriging paradigm presented below. This approach relies on the fact that kriging predictors and kriging residuals are uncorrelated (independent under Gaussian assumption), see Chilès and Delfiner (2012, Chapter 7). First, a non-conditional simulation $\mathbf{x}^{(NC)}_{1:N_T}$ is performed on the spatio-temporal grid $\mathcal{T}'$. From this simulation, kriging residuals

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

are computed over the entire spatio-temporal grid $\mathcal{T}'$. The conditional expectation is computed using the method presented in the previous section. In a second step, these independently generated residuals are added to the usual kriging of the data to get the conditional simulation

$$\mathbf{x}^{(C)}_{1:N_T} = \mathbf{x}^{*}_{1:N_T} + \mathbf{r}_{1:N_T}\,.$$

Conditional simulations at further time steps are obtained by iteratively computing $\mathbf{x}^{(C)}_{N_T+k}$ using the propagation equation (21) with $k \geq 1$.

3.4 Simulation study

We here report some results of estimation of the parameters $\boldsymbol{\theta}^{\top} = (\kappa, \gamma_x, \gamma_y, c, \tau)$ for a spatio-temporal model simulated with the SPDE (4). We set $\mathbf{H} = \mathbf{I}_{N_S}$, $\alpha = 1$ and $\alpha_S = 2$. The spatial domain is a $[100, 100]$ square with a grid triangulation of $N_S = 900$ spatial points. The time window is $[1, 10]$ with unit $N_T = 10$ time steps. The $n_S = 100$ observations are randomly located into the spatial domain and their position do not change during the N_T time steps (hence $n = 1000$). Since the size of both the dataset and the spatio-temporal mesh is quite small, we report only the estimations computed with the Cholesky decomposition.

As initial values we used estimated values obtained from the spatial and temporal traces of the process. Specifically, the initial value for κ is the estimated scale parameter of a Matérn covariance function with smoothness parameter $\nu = \alpha + \alpha_S - 1 = 2$ considering independent temporal repetitions, the initial value for c is deduced from the estimated parameter of n_S independent repetitions of AR(1) processes of length N_T and τ^2 is computed from Equation (6) with σ^2 being the empirical variance computed on the data.

Finally, the initial value for $\boldsymbol{\gamma}$ is the null vector. The results are depicted in Table 1.

κ	$\hat{\kappa}$	γ_x	$\hat{\gamma}_x$	γ_y	$\hat{\gamma}_y$	c	$\hat{c}$	τ	$\hat{\tau}$	average time (s)
0.2	0.203 (0.051)	-2	-2.043 (0.262)	3	2.979 (0.351)	1	1.037 (0.093)	1	1.036 (0.092)	120
0.33	0.328 (0.059)	-1	-1.008 (0.134)	1	1.018 (0.143)	0.5	0.546 (0.041)	1.2	1.221 (0.037)	124

Table 1 Mean (and standard deviation) of ML estimates $\hat{\boldsymbol{\theta}}^\top = (\hat{\kappa}, \hat{\gamma}_x, \hat{\gamma}_y, \hat{c}, \hat{\tau})$ over 10 simulations for two different subsets of advection-diffusion model parameters

4 Application to a solar radiation dataset

The approach detailed in the previous sections is now applied to a solar radiation dataset for which experts agree on the presence of advection due to Western prevailing winds transporting the clouds from one side of the domain to the other. The HOPE campaign (Macke et al., 2017) recorded Global Horizontal Irradiance (GHI) (or SSI, Surface Solar Irradiance) over a $10 \times 16\ km^2$ region in West Germany near the city of Jülich from the 2nd of April, to the second of July, 2013. The sensors were located at 99 stations located as pictured in Figure 1 and GHI was recorded every 15 seconds. A detailed description of the campaign can be found in Macke et al. (2017).

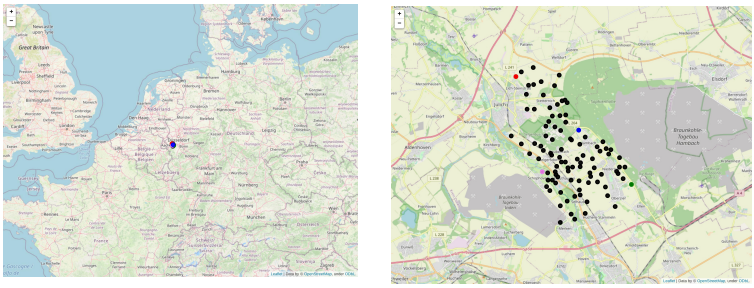

Fig. 1 Stations over the spatial domain

The dataset was cleaned for outlying values and non-operating sensors, and the temporal resolution was reduced from 15 seconds to 1 minute. Figure 2 (left panel) shows GHI as a function of time (in minute, during a full day – the 28th of May 2013) at 4 different stations. These stations, represented in color in Figure 1, are located at the border of the domain, far from each other. The GHI starts close to 0, increases after sunrise, peaks at midday and tends to 0 at sunset. The maximal theoretical amount of irradiance reaching the sensor follows an ideal concave curve. The divergence between the measured irradiance and the optimal curve can be slight or important, depending on the presence of clouds. One can see on this example that the evolution among the

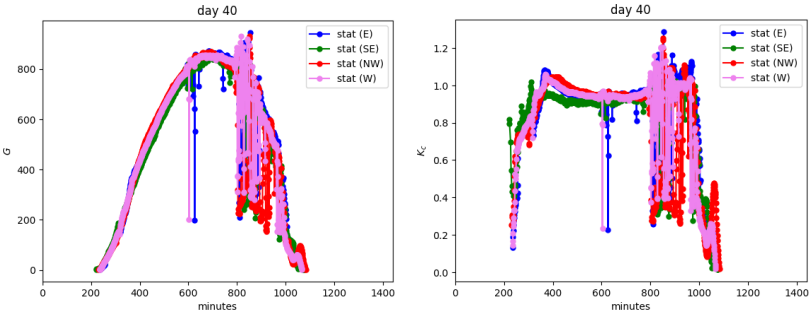

Fig. 2 GHI G and Clear Sky Index K_c for 4 different stations on the 28th of May 2013

4 stations is similar, with variations accounting for spatio-temporal variations of the clouds.

A first preprocessing was made in order to stationarize the phenomenon. Oumbe et al. (2014) showed that the solar irradiance at ground level, GHI (denoted G for short from now on), computed by a radiative transfer model can be approximated by the product of the irradiance under clear atmosphere (called Clear Sky GHI, or G_c) and a modification factor due to cloud properties and ground albedo only (Clear Sky Index, or K_c, Beyer et al. (1996)):

$$G \simeq G_c K_c. \tag{23}$$

The error made in using this approximation depends mostly on the solar zenith angle, the ground albedo and the cloud optical depth. In most cases, the maximum errors (95th percentile) on global and direct surface irradiances are less than 15 Wm^{-2} and less than 2- to 5 % in relative value, as recommended by the World Meteorological Organization for high-quality measurements of the solar irradiance (Oumbe et al., 2014). Practically, it means that a model for fast calculation of surface solar irradiance may be separated into two distinct and independent models: i) a deterministic model for G, under clear-sky conditions, as computed according to Gschwind et al. (2019), considered as known in this study; ii) a model for K_c which accounts for cloud influence on the downwelling radiation and is expected to change in time and space. K_c is modeled as a random spatio-temporal process and will be the subject of our analysis. Figure 2 (right panel) shows the variable K_c at the same 4 stations and on the same day of left panel. In general, K_c lies between 0 and 1, but in rare occasions, values above 1 can be observed. This phenomenon is called *overshooting* (Schade et al., 2007).

A time window of 20 minutes around 4 p.m. on the 28th of May 2013 is extracted with observations every minute at the 73 stations with well recorded values. Parameters will be estimated on this 20-minute window using the method described in Section 3.1. The spatio-temporal grid contains $N_T = 20$ one-minute time steps, from $t = 1$ to $t = 20$ and $N_S = 900$ spatial mesh points.

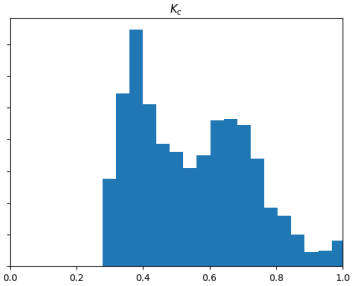

Fig. 3 Histogram of GHI over 20 time steps

4.1 Estimation and prediction

Six different models are fitted to the data and used for prediction: 3 models with advection (called "adv-diff") and 3 models without advection (called "diff") obtained by setting $\boldsymbol{\gamma} = \mathbf{0}$. Both groups contain the three following sub-models: (i) a separable model with a Matérn spatial trace with $\nu = 1$; (ii) a non-separable model whose spatial trace has no known closed form expression for the covariance function; (iii) a non-separable model with a Matérn spatial trace with $\nu = 2$. In the general model of Equation (4) they correspond respectively to $(\alpha, \alpha_S) = (0, 2), (1, 0), (1, 2)$. The parameters of the SPDE are estimated for each model separately. The results are reported in Table 2. The log-likelihood of the models that include advection is at least 34 units larger than those with diffusion only. As a point of comparison, if all spatio-temporal dependencies were ignored, the BIC penalization to the log-likelihood of the model with advection would be $2\ln(1460) \simeq 14.5$. These results indicate strong evidence in favor of the model with advection.

We then perform prediction with two different validation settings containing 80% of conditioning data and 20% of validation data. In the first case (called "Random") the validation locations are uniformly randomly selected. In the second case (called "South-East") the validation locations are located downwind (i.e. South-East) with respect to the estimated advection direction. See Figure 4 for a representation of the validation settings.

Recall that a time window containing 20 time-steps, from $T = 1$ to $T = 20$, was selected. For each validation setting and from $T = 11$ to $T = 20$, three prediction configurations using conditioning data from time $(T - 9)$ to T are computed and compared to the real values, allowing us to compute a Root Mean Square Error (RMSE) validation score. First, the kriging is performed spatially only (hereafter referred to as "S" kriging). Second, a temporal extrapolation is computed at the conditioning locations at time horizons $(T + 1)$, $(T + 2)$ and $(T + 3)$ ("T1", "T2", "T3" kriging). Third, the spatio-temporal prediction is computed at the validation locations at time horizons $(T + 1)$, $(T + 2)$ and $(T + 3)$ ("ST1", "ST2", "ST3" kriging). We thus have a total of 6 models $\times$ 2 validation settings $\times$ 3 prediction configurations. RMSEs are averaged on the 10 repetitions. Results are shown in Table 3. It is clear that

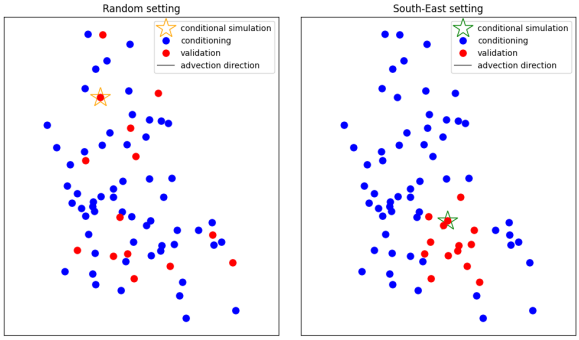

Fig. 4 Validation settings: Random (left) and South-East (right)

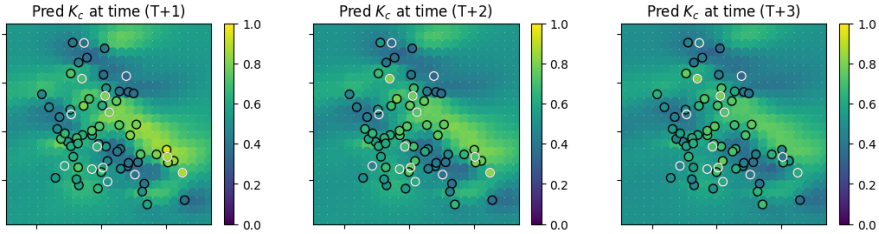

Fig. 5 Predictions of K_c at $(T+1)$, $(T+2)$ and $(T+3)$ with model (3) ("adv-diff" with $\alpha = 1$ and $\alpha_S = 2$). The black contoured dots are the conditioning locations and the white contoured dots are the validation locations for the Random setting

non-separable advection-diffusion models ((2) and (3)) outperform the others in both temporal and spatio-temporal settings.

An example of prediction maps on $\mathcal{T}$ at time horizons $(T+1)$, $(T+2)$ and $(T+3)$ is reported in Figure 5, along with the observed values (black contoured dots). The white contoured dots are the locations used for validation in the Random setting.

	Model	α	α_S	log-likelihood	$\hat{\kappa}$	$\hat{\gamma}_x$	$\hat{\gamma}_y$	$\hat{c}$	$\hat{\tau}$	$\hat{\sigma}_0$	$\hat{b}$
(1)	adv-diff	0	2	2587	1.477	9.642	-5.382	11.659	2.254	0.052	0.570
(2)	adv-diff	1	0	2577	0.237	4.718	-0.928	9.315	0.458	0.045	0.598
(3)	adv-diff	1	2	2579	1.299	17.325	-8.442	41.017	3.072	0.058	0.574
(4)	diff	0	2	2507	1.240	0	0	12.558	1.081	0.059	0.569
(5)	diff	1	0	2545	0.246	0	0	6.594	0.436	0.047	0.577
(6)	diff	1	2	2512	0.940	0	0	34.607	1.248	0.064	0.580

Table 2 Estimated parameters and log-likelihood for 6 different models from all data on a 20-minute window

				Random			South-East		
	Model	α	α_S	S (min,max)			S (min,max)		
(1)	adv-diff	0	2	**0.088 (0.052,0.127)**			**0.103 (0.051,0.138)**		
(2)	adv-diff	1	0	0.103 (0.064,0.142)			0.105 (0.045,0.158)		
(3)	adv-diff	1	2	0.102 (0.062,0.134)			0.109 (0.054,0.149)		
(4)	diff	0	2	0.119 (0.074,0.140)			0.134 (0.092,0.181)		
(5)	diff	1	0	0.094 (0.060,0.131)			0.136 (0.067,0.187)		
(6)	diff	1	2	0.110 (0.066,0.132)			0.140 (0.085,0.192)		
	Model	α	α_S	T1 (min,max)	T2 (min,max)	T3 (min,max)	T1 (min,max)	T2 (min,max)	T3 (min,max)
(1)	adv-diff	0	2	0.095 (0.067,0.120)	0.146 (0.111,0.186)	0.181 (0.131,0.236)	0.095 (0.065,0.122)	0.142 (0.106,0.185)	0.172 (0.121,0.228)
(2)	adv-diff	1	0	**0.071 (0.046,0.090)**	**0.093 (0.054,0.124)**	**0.102 (0.060,0.143)**	0.074 (0.045,0.099)	0.097 (0.065,0.128)	0.109 (0.069,0.148)
(3)	adv-diff	1	2	0.072 (0.046,0.093)	0.095 (0.054,0.127)	0.104 (0.055,0.144)	**0.074 (0.049,0.097)**	**0.096 (0.062,0.122)**	**0.106 (0.061,0.139)**
(4)	diff	0	2	0.094 (0.058,0.123)	0.137 (0.091,0.181)	0.166 (0.102,0.231)	0.090 (0.063,0.116)	0.128 (0.097,0.167)	0.154 (0.113,0.209)
(5)	diff	1	0	0.079 (0.058,0.098)	0.108 (0.082,0.135)	0.124 (0.099,0.158)	0.081 (0.057,0.105)	0.111 (0.090,0.147)	0.128 (0.100,0.169)
(6)	diff	1	2	0.083 (0.054,0.108)	0.110 (0.077,0.149)	0.125 (0.085,0.180)	0.084 (0.056,0.109)	0.109 (0.086,0.152)	0.123 (0.091,0.176)
	Model	α	α_S	ST1 (min,max)	ST2 (min,max)	ST3 (min,max)	ST1 (min,max)	ST2 (min,max)	ST3 (min,max)
(1)	adv-diff	0	2	0.105 (0.067,0.144)	0.147 (0.106,0.193)	0.179 (0.140,0.231)	0.102 (0.081,0.121)	0.158 (0.130,0.180)	0.210 (0.160,0.241)
(2)	adv-diff	1	0	0.091 (0.052,0.131)	0.103 (0.062,0.161)	0.110 (0.071,0.165)	**0.079 (0.044,0.119)**	**0.075 (0.039,0.127)**	**0.070 (0.038,0.132)**
(3)	adv-diff	1	2	**0.085 (0.050,0.127)**	**0.094 (0.058,0.142)**	**0.100 (0.061,0.157)**	0.090 (0.052,0.121)	0.099 (0.056,0.149)	0.109 (0.057,0.172)
(4)	diff	0	2	0.123 (0.072,0.186)	0.150 (0.081,0.237)	0.170 (0.116,0.257)	0.128 (0.092,0.157)	0.165 (0.112,0.195)	0.199 (0.116,0.236)
(5)	diff	1	0	0.104 (0.070,0.153)	0.122 (0.095,0.181)	0.131 (0.095,0.187)	0.107 (0.050,0.201)	0.100 (0.042,0.223)	0.094 (0.0380,0.217)
(6)	diff	1	2	0.108 (0.073,0.153)	0.126 (0.085,0.188)	0.134 (0.082,0.199)	0.114 (0.060,0.204)	0.111 (0.058,0.229)	0.108 (0.059,0.223)

Table 3 Averaged RMSE computed at 10 successive time steps for 6 different models, 2 validation settings (Random and South-East) and 3 prediction configurations (S, T and ST); see text for details. In each case, the best score among the models is in bold font

4.2 Conditional simulations

Figure 6 shows 100 conditional simulations of K_c computed at time $T = 11$ and horizons $(T+1), (T+2), \ldots, (T+6)$ with the advection-diffusion model (3). Two validation stations have been selected: one selected in the "Random" setting, in the North-West part of the domain (the orange star in the left panel of Figure 4) and one selected in the "South-East" setting (the green star in the right panel of Figure 4). Given that there is an advection from NW to SE, it is therefore expected that the advection-diffusion model should be able to transport some information. The mean of the 100 simulations and the envelopes corresponding to twice the pointwise standard deviation have also been represented, along with the true values. As expected, most of the conditional simulations lie within the envelopes in both cases and at all time horizons. The remarkable result is that the variance of the conditional simulations at the green station is smaller than that at the orange one at every time step, especially when the time horizon increases. This is due to the advection term in model (3), able to propagate information from North-West to South-East.

5 Conclusion

The spatio-temporal SPDE approach based on advection-diffusion equations proposed in this work combines elements of physics, numerical analysis and statistics. It can be seen as a first step toward *physics informed geostatistics*, which introduces physical dynamics into a statistical model, accounting for possible hidden structures governing the evolution of the spatio-temporal phenomenon. The different terms of the SPDE (advection, diffusion) directly influence the spatio-temporal dependencies of the process, by controlling its variability in space and time. Compared to spatio-temporal models built on covariance functions such as the Gneiting class (Gneiting, 2002), we gain in

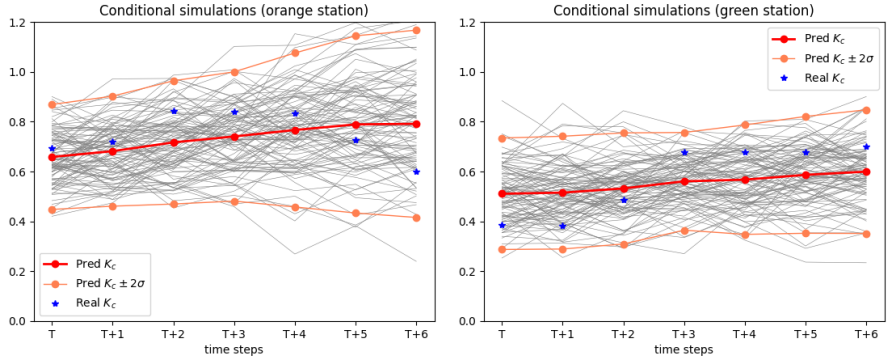

Fig. 6 Real K_c, mean of conditional simulations of K_c and $\pm 2\sigma$ envelope at time horizons $T, (T+1), (T+2), \ldots, (T+6)$. Left: orange station with Random setting. Right: green station with South-East setting

interpretability since the parameters of the model can be directly linked to the physical coefficients of SPDEs.

We showed that it is possible to build an accurate space-time approximations of the process driven by the SPDE using a combination of FEM in space and implicit Euler scheme in time. It leads to sparse structured linear systems. We obtained promising results for the estimation and for the prediction of processes both in terms of precision and speed. When the size of the dataset is moderate, direct matrix implementation is possible. We showed how matrix-free methods can be implemented in order to obtain scalable computations even for very large datasets.

The application to the solar radiation dataset showed that the non-separable advection-diffusion model exhibited the best prediction performances on a phenomenon that is certainly governed by advection and diffusion processes.

Further work would be necessary to better assess the prediction accuracy and the computational complexity. Applications to larger and more complex datasets should be considered and comparison to models expressing the advection in a Lagrangian framework (Ailliot et al., 2011; Salvaña and Genton, 2021) should be performed.

We implemented a maximum likelihood approach for parameter estimation. It would be interesting to investigate a possible Bayesian estimation procedure through the INLA approach (Rue et al., 2009).

One of the main advantages of the SPDE formulation is that it is easy to generalize to non-stationary settings. Non-stationary fields can be defined by letting the parameters $(\kappa(\mathbf{s}, t), \boldsymbol{\gamma}(\mathbf{s}, t))$ be space-time-dependent. This generalization implies only minimal changes to the method used in the stationary case concerning the simulation, but needs more work for estimation and prediction, since the maximum likelihood approach becomes much more expensive. We can also incorporate models of spatially varying anisotropy by modifying the general operator $\nabla \cdot \mathbf{H}(\mathbf{s}, t) \nabla X(\mathbf{s}, t)$ with a non-stationary anisotropic matrix

$\mathbf{H}(\mathbf{s}, t)$. The introduction of non-stationarities could allow to better describe phenomena where local variations are clearly present. The generalization of the approaches by Fuglstad et al. (2015) and Pereira et al. (2022) should be investigated and generalized to the spatio-temporal framework.

Another interesting consequence of defining the models through local stochastic partial differential equations is that the SPDEs still make sense when $\mathbb{R}^d$ is replaced by a space that is only locally flat. We can define non-stationary Gaussian fields on manifolds, and still obtain a GMRF representation. Important improvements were obtained in the spatial case (Pereira et al., 2022). The generalization to space-time processes could be explored further.

Possible generalization to spatio-temporal SPDEs with a fractional exponent in the diffusion term could also be considered. A development of the methods proposed by Bolin and Kirchner (2020) and Vabishchevich (2015) should be explored.

Acknowledgments. We are grateful to the O.I.E. center of Mines Paris – ARMINES, especially to Yves-Marie Saint-Drenan, Philippe Blanc and Hadrien Verbois, for providing the data and for inspiring discussions about renewable resources evaluation. In addition, we thank the Mines Paris / INRAE chair "Geolearning" for the constant support. We are thankful to the STSDS department of King Abdullah University of Science and Technology (KAUST), and especially to Professor Marc Genton, for the insightful work carried out during a stay in Saudi Arabia.

References

Ailliot, P., Baxevani, A., Cuzol, A., Monbet, V., and Raillard, N. (2011). Space–time models for moving fields with an application to significant wave height fields. *Environmetrics*, 22(3):354–369.

Bakka, H. (2018). How to solve the stochastic partial differential equation that gives a matérn random field using the finite element method.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data (2nd ed.)*. Chapman & Hall/CRC Press. Boca Raton, FL.

Bank, R. E., Bürgler, J. F., Fichtner, W., and Smith, R. K. (1990). Some upwinding techniques for finite element approximations of convection-diffusion equations. *Numerische Mathematik*, 58(1):185–202.

Beyer, H. G., Costanzo, C., and Heinemann, D. (1996). Modifications of the heliosat procedure for irradiance estimates from satellite images. *Solar Energy*, 56(3):207–212.

Bolin, D. and Kirchner, K. (2020). The rational spde approach for gaussian random fields with general smoothness. *Journal of Computational and Graphical Statistics*, 29(2):274–285.

Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. (2011). Spatio-temporal modeling of particulate matter concentration through the spde approach. *AStA Advances in Statistical Analysis*, 97.

Carrizo-Vergara, R., Allard, D., and Desassis, N. (2022). A general framework for spde-based stationary random fields. *Bernoulli*, 28(1):1–32.

Chen, W., Genton, M. G., and Sun, Y. (2021). Space-time covariance structures and models. *Annual Review of Statistics and Its Application*, 8(1):191–215.

Chilès, J.-P. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty, Second Edition*. John Wiley & Sons.

Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1340.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley.

Erisman, A. M. and Tinney, W. F. (1975). On computing certain elements of the inverse of a sparse matrix. *Commun. ACM*, 18(3):177–179.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2015). Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics*, 14(Part C):505–531.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97(458):590–600.

Gschwind, B., Wald, L., Blanc, P., Lefèvre, M., Schroedter-Homscheidt, M., and Arola, A. (2019). Improving the mcclear model estimating the downwelling solar radiation at ground level in cloud-free conditions - mcclear-v3. *Meteorologische Zeitschrift*, 28(2):147–163.

Hughes, T. and Brooks, A. (1981). A multidimensional upwind scheme with no crosswind diffusion. *Analytical and Numerical Approaches to Asymptotic Problems in Analysis*, pages 99–116.

Hutchinson, M. (1990). A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics - Simulation and Computation*, 19(2):433–450.

Krainski, E., Gómez Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. Chapman & Hall/CRC Press. Boca Raton, FL.

Lindgren, F., Bakka, H., Bolin, D., Krainski, E., and Rue, H. (2020). A diffusion-based spatio-temporal extension of gaussian matérn fields.

Lindgren, F., Bolin, D., and Rue, H. (2022). The spde approach for gaussian and non-gaussian fields: 10 years and still running. *Spatial Statistics*, 50:100599. Special Issue: The Impact of Spatial Statistics.

Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with r-inla. *Journal of Statistical Software*, 63(19):1–25.

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. (2020). Statistical modeling for spatio-temporal data from stochastic convection-diffusion processes. *Journal of the American Statistical Association*, pages 1–37.

Macke, A., Seifert, P., Baars, H., Barthlott, C., Beekmans, C., Behrendt, A., Bohn, B., Brueck, M., Bühl, J., Crewell, S., Damian, T., Deneke, H., Düsing, S., Foth, A., Di Girolamo, P., Hammann, E., Heinze, R., Hirsikko, A., Kalisch, J., Kalthoff, N., Kinne, S., Kohler, M., Löhnert, U., Madhavan, B. L., Maurer, V., Muppa, S. K., Schween, J., Serikov, I., Siebert, H., Simmer, C., Späth, F., Steinke, S., Träumner, K., Trömel, S., Wehner, B., Wieser, A., Wulfmeyer, V., and Xie, X. (2017). The hd(cp)^2 observational prototype experiment (hope) – an overview. *Atmospheric Chemistry and Physics*, 17(7):4887–4914.

Martínez-Hernández, I. and Genton, M. G. (2022). Surface time series models for large spatio-temporal datasets. *Spatial Statistics*, page 100718.

Mekuria, G. and Rao, J. (2016). Adaptive finite element method for steady convection-diffusion equation. *American Journal of Computational Mathematics*, 06:275–285.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, New York, NY, USA, 2e edition.

Oumbe, A., Qu, Z., Blanc, P., Lefèvre, M., Wald, L., and Cros, S. (2014). Decoupling the effects of clear atmosphere and clouds to simplify calculations of the broadband solar irradiance at ground level. *Geoscientific Model Development*, 7(4):1661–1669.

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

Pereira, M., Desassis, N., and Allard, D. (2022). Geostatistics for large datasets on riemannian manifolds: A matrix-free approach. *Journal of Data Science*, 20(4):512–532.

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

Powell, P. (2011). Calculating determinants of block matrices.

Quarteroni, A. (2008). *Modellistica Numerica per Problemi Differenziali*. Springer Milano.

Roques, L., Allard, D., and Soubeyrand, S. (2022). Spatial statistics and stochastic partial differential equations: A mechanistic viewpoint. *Spatial Statistics*, 50:100591.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Monographs on statistics and applied probability 104. Chapman & Hall/CRC, 1 edition.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Salvaña, M. L. O. and Genton, M. G. (2021). *Lagrangian Spatio-Temporal Nonstationary Covariance Functions*, pages 427–447. Springer International Publishing, Cham.

Schade, N. H., Macke, A., Sandmann, H., and Stick, C. (2007). Enhanced solar global irradiance during cloudy sky conditions. *Meteorologische Zeitschrift*, 16:295–303.

Sigrist, F., Künsch, H. R., and Stahel, W. A. (2012). A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. *The Annals of Applied Statistics*, 6(4).

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)*, 77(1):3–33.

Stein, M. L. (2005). Space-time covariance functions. *Journal of the American Statistical Association*, 100(469):310–321.

Takahashi, K., Fagan, J., and Chin, M. (1973). Formation of a sparse bus impedance matrix and its application to short circuit study. 8th PICA Conf. Proc. June 4-6, Minneapolis, Minn.

Vabishchevich, P. N. (2015). Numerically solving an equation for fractional powers of elliptic operators. *Journal of Computational Physics*, 282:289–302.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41(3-4):434–449.

Whittle, P. (1963). Stochastic processes in several dimensions. *Bull. Inst. Int. Statist.*, 40:974–994.

Wikle, C. and Hooten, M. (2010). A general science-based framework for dynamical spatio-temporal models. *TEST*, 19:417–451.

Wikle, C. K. and Cressie, N. (1999). A dimension-reduced approach to space-time kalman filtering. *Biometrika*, 86(4):815–829.

Xiong, Z., Simas, A. B., and Bolin, D. (2022). Covariance-based rational approximations of fractional spdes for computationally efficient bayesian inference.

Appendix A Proof of Proposition 1

Proof The spatial trace is the covariance function between $X(\mathbf{s},t)$ and $X(\mathbf{s}',t)$ for a spatial lag $\mathbf{h}=\mathbf{s}-\mathbf{s}'$

$$\begin{aligned}\mathrm{Cov}(\mathbf{h},0) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}} \exp(i\,\mathbf{h}\,\boldsymbol{\xi}) S(\boldsymbol{\xi},\omega)\,\mathrm{d}\omega\,\mathrm{d}\,\boldsymbol{\xi} \\ &= \int_{\mathbb{R}^d} \exp(i\,\mathbf{h}\,\boldsymbol{\xi}) \left[\int_{\mathbb{R}} S(\boldsymbol{\xi},\omega)\,\mathrm{d}\omega\right]\mathrm{d}\,\boldsymbol{\xi} \\ &= \int_{\mathbb{R}^d} \exp(i\,\mathbf{h}\,\boldsymbol{\xi}) S_S(\boldsymbol{\xi}), \end{aligned} \tag{A1}$$

where $S_S(\boldsymbol{\xi})$ is the spatial spectral density defined as

$$\begin{aligned} S_S(\boldsymbol{\xi}) &= \frac{\tau^2}{(2\pi)^d c(\kappa^2+\boldsymbol{\xi}^\top\mathbf{H}\,\boldsymbol{\xi})^{\alpha_S}} \int_{\mathbb{R}} \frac{1}{2\pi\left[\omega^2+c^{-2}(\kappa^2+\boldsymbol{\xi}^\top\mathbf{H}\,\boldsymbol{\xi})^{2\alpha}\right]}\,\mathrm{d}\omega \\ &= \frac{\tau^2}{(2\pi)^d c(\kappa^2+\boldsymbol{\xi}^\top\mathbf{H}\,\boldsymbol{\xi})^{\alpha_S}} \frac{1}{2\left[c^{-2}(\kappa^2+\boldsymbol{\xi}\,\mathbf{H}\,\boldsymbol{\xi})^{2\alpha}\right]^{1/2}} \\ &= \frac{\tau^2}{2(2\pi)^d(\kappa^2+\boldsymbol{\xi}^\top\mathbf{H}\,\boldsymbol{\xi})^{\alpha_{tot}}}. \end{aligned} \tag{A2}$$

By using the change of variable $\boldsymbol{\xi}=\kappa\,\mathbf{H}^{-1/2}\,\mathbf{w}$ and by plugging Equation (A2) into (A1), we obtain

$$\begin{aligned}\mathrm{Cov}(\mathbf{h},0) &= \frac{\tau^2}{2}\int_{\mathbb{R}^d} \frac{e^{i\,\mathbf{h}\,\boldsymbol{\xi}}}{(2\pi)^d(\kappa^2+\kappa^2\,\boldsymbol{\xi}^\top\mathbf{H}\,\boldsymbol{\xi})^{\alpha_{tot}}}\,\mathrm{d}\,\boldsymbol{\xi} \\ &= \frac{\tau^2}{2}\int_{\mathbb{R}^d} \frac{e^{i\,\mathbf{h}\,\kappa\,\mathbf{H}^{-1/2}\,\mathbf{w}}|\kappa\,\mathbf{H}^{-1/2}|}{(2\pi)^d(\kappa^2+\kappa^2\,\mathbf{w}^\top\,\mathbf{w})^{\alpha_{tot}}}\,\mathrm{d}\,\mathbf{w} \\ &= \frac{\tau^2}{2\kappa^{2(\alpha_{tot}-d/2)}|\mathbf{H}|^{1/2}}\int_{\mathbb{R}^d} \frac{e^{i\,\mathbf{h}\,\kappa\,\mathbf{H}^{-1/2}\,\mathbf{w}}}{(2\pi)^d(1+\mathbf{w}^\top\,\mathbf{w})^{\alpha_{tot}}}\,\mathrm{d}\,\mathbf{w} \\ &= \frac{\tau^2\Gamma(\alpha_{tot}-d/2)}{2\Gamma(\alpha_{tot})(4\pi)^{d/2}\kappa^{2(\alpha_{tot}-d/2)}|\mathbf{H}|^{1/2}} C^M_{\alpha_{tot}-d/2}\left(\kappa\|\mathbf{h}\|\right). \end{aligned}$$

□

Appendix B Discretization of spatio-temporal advection-diffusion SPDE

Here, we detail the discretization scheme of the advection-diffusion spatio-temporal SPDE (4). For the sake of a clearer exposition, we have set $\mathbf{H}=\mathbf{I}_{N_S}$, $\alpha=1$ and we consider a spatio-temporal white noise $Z(\mathbf{s},t)=W(\mathbf{s},t)$. The proof for the general case follows exactly the same lines as the proof below. The considered SPDE is

$$\left[\frac{\partial}{\partial t}+\frac{1}{c}(\kappa^2-\Delta)+\frac{1}{c}\,\boldsymbol{\gamma}\cdot\nabla\right]X(\mathbf{s},t)=\frac{\tau}{\sqrt{c}}W(\mathbf{s},t). \tag{B3}$$

For the discretization of the temporal derivative in Equation (B3), we opt for the implicit Euler scheme, which considers the differential equation

$$\frac{\partial X}{\partial t} = f(t, X),$$

with initial value $X(t_0) = X_0$, where both the function f and the initial data t_0 and X_0 are known; the function X depends on the real variable t and is unknown. The method produces a sequence $X_0, X_1, X_2, \ldots$, such that X_k approximates $X(t_0 + kdt)$, where dt is the time step size. The approximation reads

$$X_{k+1} = X_k + dtf(t_{k+1}, X_{k+1}).$$

In the specific case of Equation (B3), the implicit Euler discretization step reads

$$X(\mathbf{s}, t+dt) - X(\mathbf{s}, t) + dt\left[\frac{1}{c}(\kappa^2 - \Delta) + \frac{1}{c}\boldsymbol{\gamma}\cdot\nabla\right] X(\mathbf{s}, t+dt) = \frac{\sqrt{dt}\tau}{\sqrt{c}} W_{S,t+dt}(\mathbf{s}), \tag{B4}$$

where $W_{S,t+dt}(\mathbf{s})$ is a spatial white noise. For ease of notation, we denote $x = X(\mathbf{s}, t + dt)$ the unknown variable, defined with respect to $x_t = X(\mathbf{s}, t)$.

At each time step of the temporal discretization, a spatial Finite Element Method method is applied. In our case, we use the continuous Galerkin with Neumann boundary condition. The weak form of Equation (B4) is

$$\begin{aligned}\int_\Omega xv\,\mathrm{d}\,\mathbf{s} + \frac{dt}{c}\left(\int_\Omega \kappa^2 xv\,\mathrm{d}\,\mathbf{s} - \int_\Omega \Delta xv\,\mathrm{d}\,\mathbf{s} + \int_\Omega \boldsymbol{\gamma}\cdot\nabla xv\,\mathrm{d}\,\mathbf{s}\right)\\ = \int_\Omega x_t v\,\mathrm{d}\,\mathbf{s} + \frac{\sqrt{dt}\tau}{\sqrt{c}}\int_\Omega vW(\mathbf{s})\,\mathrm{d}\,\mathbf{s}, \quad \forall v \in \mathcal{V},\end{aligned}$$

where $\mathcal{V}$ is the space of the solutions.

By applying Green's first identity, i.e., by writing $\int_\Omega \Delta xv\,\mathrm{d}\,\mathbf{s} = -\int_\Omega \nabla x \cdot \nabla v\,\mathrm{d}\,\mathbf{s} + \int_{\partial\Omega} v\cdot(\nabla x\cdot\hat{\mathbf{n}})d\sigma$, with $\hat{\mathbf{n}}$ being the normal vector on the boundary, and by simplifying the second term thanks to the Neumann boundary condition, we obtain

$$\begin{aligned}\underbrace{\int_\Omega xv\,\mathrm{d}\,\mathbf{s} + \frac{dt}{c}\left(\int_\Omega \kappa^2 xv\,\mathrm{d}\,\mathbf{s} + \int_\Omega \nabla x\cdot\nabla v\,\mathrm{d}\,\mathbf{s} + \int_\Omega \boldsymbol{\gamma}\cdot\nabla xv\,\mathrm{d}\,\mathbf{s}\right)}_{\mathcal{A}(x,v)}\\ = \underbrace{\int_\Omega x_t v\,\mathrm{d}\,\mathbf{s}}_{\mathcal{C}(x_t,v)} + \underbrace{\frac{\sqrt{dt}\tau}{\sqrt{c}}\int_\Omega vW(\mathbf{s})\,\mathrm{d}\,\mathbf{s}}_{\mathcal{E}(v)}, \quad \forall v \in \mathcal{V}.\end{aligned}$$

Let $\mathcal{V}_h$ be the space of finite element solutions spanned by the basis functions $\{\psi_i\}_{i=1}^{N_S}$. The generalized Galerkin method allows us to find an

approximated solution $x_h \in \mathcal{V}_h \subseteq \mathcal{V}$ to the SPDE, such that

$$\mathcal{A}(x_h, v_h) = \mathcal{C}(x_{t,h}, v_h) + \mathcal{E}(v_h) \quad \forall v_h \in \mathcal{V}_h.$$

The functions x_h, $x_{t,h}$ and v_h are linear combinations of the basis functions, with

$$x_h = \sum_{i=1}^{N_S} x_{h,i}\psi_i; \quad x_{t,h} = \sum_{i=1}^{N_S} x_{t,h,i}\psi_i; \quad v_h = \sum_{i=1}^{N_S} v_{h,i}\psi_i,$$

Because of the linearity in the first argument of $\mathcal{A}(\cdot,\cdot)$ and $\mathcal{C}(\cdot,\cdot)$, we get

$$\sum_{i=1}^{N_S} \mathcal{A}(\psi_i, v_h)x_{h,i} = \sum_{i=1}^{N_S} \mathcal{C}(\psi_i, v_h)x_{t,h,i} + \mathcal{E}(v_h), \quad \forall v_h \in \mathcal{V}_h, \tag{B5}$$

where

$$\mathcal{A}(\psi_i, v_h) = \mathcal{M}(\psi_i, v_h) + \frac{dt}{c}\left(\mathcal{K}(\psi_i, v_h) + \mathcal{B}(\psi_i, v_h)\right)$$
$$\mathcal{C}(\psi_i, v_h) = \mathcal{M}(\psi_i, v_h),$$

with $\mathcal{K}(\psi_i, v_h) = \kappa^2\mathcal{M}(\psi_i, v_h) + \mathcal{G}(\psi_i, v_h)$. Here, $\mathcal{M}$ and $\mathcal{G}$ are the mass and stiffness operators, respectively $\mathcal{M}(v, w) = \int_\Omega vw \,\mathrm{d}\,\mathbf{s}$ and $\mathcal{G}(v, w) = \int_\Omega \nabla v \cdot \nabla w \,\mathrm{d}\,\mathbf{s}$. $\mathcal{B}$ is the advection operator, i.e., $\mathcal{B}(v, w) = \int_\Omega \boldsymbol{\gamma} \cdot \nabla v w \,\mathrm{d}\,\mathbf{s}$. Finally, $\mathcal{E}$ is the operator of the form $\mathcal{E}(v) = \frac{\sqrt{dt}\tau}{\sqrt{c}} \int_\Omega vW(\mathbf{s}) \,\mathrm{d}\,\mathbf{s}$.

Since any v_h can be written as a linear combination of basis functions, the formulation (B5) is equivalent to

$$\sum_{i=1}^{N_S} \mathcal{A}(\psi_i, \psi_j)x_{h,i} = \sum_{i=1}^{N_S} \mathcal{C}(\psi_i, \psi_j)x_{t,h,i} + \mathcal{E}(\psi_j), \quad \forall j \tag{B6}$$

We define $\mathbf{M} = [M_{ij}]_{i,j=1}^{N_S} = [\mathcal{M}(\psi_i, \psi_j)]_{i,j=1}^{N_S}$, $\mathbf{G} = [G_{ij}]_{i,j=1}^{N_S} = [\mathcal{G}(\psi_i, \psi_j)]_{i,j=1}^{N_S}$, $\mathbf{B} = [B_{ij}]_{i,j=1}^{N_S} = [\mathcal{B}(\psi_i, \psi_j)]_{i,j=1}^{N_S}$ the mass, stiffness and advection matrices, respectively.

$\mathcal{E}(\psi_j)$ is a random variable with expectation 0 and covariance equal to

$$\begin{aligned}\mathrm{Cov}(\mathcal{E}(\psi_i), \mathcal{E}(\psi_j)) &= \frac{dt\tau^2}{c}\mathrm{Cov}\left(\int_\Omega \psi_i W(\mathbf{s}) \,\mathrm{d}\,\mathbf{s}, \int_\Omega \psi_j W(\mathbf{s}) \,\mathrm{d}\,\mathbf{s}\right) \\ &= \frac{dt\tau^2}{c}\int_\Omega \psi_i\psi_j \,\mathrm{d}\,\mathbf{s} = \frac{dt\tau^2}{c}M_{ij}.\end{aligned}$$

If $\mathbf{z}_{t+dt}$ is a (N_S)-Gaussian vector such that $\mathbf{z}_{t+dt} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_S})$, $\mathbf{x}_{t+dt}$ is the vector containing the values $\{x_{h,i}\}_{i=1}^{N_S}$ and $\mathbf{x}_t$ is the vector containing the

values $\{x_{t,h,i}\}_{i=1}^{N_S}$, then the sparse linear system corresponding to Equation (B6) reads

$$\mathbf{M}\,\mathbf{x}_{t+dt} + \frac{dt}{c}(\mathbf{K}+\mathbf{B})\mathbf{x}_{t+dt} = \mathbf{M}\,\mathbf{x}_t + \frac{\sqrt{dt}\tau}{\sqrt{c}}\,\mathbf{M}^{1/2}\,\mathbf{z}_{t+dt}, \tag{B7}$$

where $\mathbf{K} = \kappa^2\,\mathbf{M} + \mathbf{G}$.

When the spatial noise is colored, i.e. $z_S = Z_S(\mathbf{s})$, the right-hand side operator $\mathcal{E}_S(v)$ becomes

$$\mathcal{E}_S(z_S, v) = \frac{\sqrt{dt}\tau}{\sqrt{c}} \int_\Omega z_S v \,\mathrm{d}\,\mathbf{s}$$

and it satisfies

$$\mathcal{E}_S(z_{S,h}, v_h) = \sum_{i=1}^{N_S} \mathcal{E}_S(\psi_i, v_h) z_{S,h,i}.$$

Hence,

$$\sum_{i=1}^{N_S} \mathcal{A}(\psi_i, \psi_j) x_{h,i} = \sum_{i=1}^{N_S} \mathcal{C}(\psi_i, \psi_j) x_{t,h,i} + \sum_{i=1}^{N_S} \mathcal{M}(\psi_i, \psi_j) z_{S,h,i}, \quad \forall j,$$

where $z_{S,h}$ has precision matrix equal to $\mathbf{Q}_S$. The sparse linear system thus reads

$$\mathbf{M}\,\mathbf{x}_{t+dt} + \frac{dt}{c}(\mathbf{K}+\mathbf{B})\mathbf{x}_{t+dt} = \mathbf{M}\,\mathbf{x}_t + \frac{\sqrt{dt}\tau}{\sqrt{c}}\,\mathbf{M}\,\mathbf{L}_S^\top\,\mathbf{z}_{t+dt}, \tag{B8}$$

where $\mathbf{L}_S$ is the Cholesky decomposition of $\mathbf{Q}_S^{-1}$.

Appendix C Advection-dominated SPDE

The stabilization of advection-dominated SPDEs is made through the introduction of a stabilization term for eliminating, or at least reducing, the numerical oscillations produced by the Galerkin method when the mesh is not fine enough. This term must vanish as $h \to 0$ to ensure consistency.

In the simplified case of a one-dimensional PDE with diffusion and advection terms

$$-\nabla \cdot \lambda \nabla u + \gamma \cdot \nabla u = f,$$

a way of stabilizing the advection operator $\gamma \cdot \nabla$ is to replace the diffusion coefficient λ with $\tilde{\lambda} = \lambda(1 + \phi(\mathrm{Pe}_h))$, where Pe_h is the Péclet number defined in Section 2.4 and $\lim_{h\to 0} \phi(\mathrm{Pe}_h) = 0$. A clear explanation of this method is detailed in Quarteroni (2008, Chapter 5). The method is called *upwind* (U) in the simplified case where $\phi(\mathrm{Pe}_h) = \mathrm{Pe}_h$. The idea behind the stabilization method is to add an artificial diffusion term equal to $\nabla \cdot \lambda \phi(\mathrm{Pe}_h) \nabla u$ that

depends on the size of the discretization mesh h and on the Péclet number. In this way, the equation with the additional stabilization term reads

$$-\nabla \cdot \lambda(1 + \phi(\mathrm{Pe}_h))\nabla u + \boldsymbol{\gamma} \cdot \nabla u = f$$

and it is no more advection-dominated since its Péclet number $\widetilde{\mathrm{Pe}}_h$ is now equal to

$$\widetilde{\mathrm{Pe}}_h = \frac{\mathrm{Pe}_h}{1 + \phi(\mathrm{Pe}_h)}$$

and always satisfies $\widetilde{\mathrm{Pe}}_h < 1$.

The extension of the 1D upwind stabilization model to dimension $d = 2$ is obtained by adding to the bilinear form $\mathcal{A}$ defined in Appendix B the stabilization term $\mathcal{S}_U$ such that

$$\mathcal{S}_U(u_h, v_h) = Qh \int_\Omega \nabla u_h \cdot \nabla v_h \, \mathrm{d}\,\mathbf{s}\,, \quad Q > 0. \tag{C9}$$

This stabilization term can be considered as an additional artificial diffusion equal to $Qh\Delta X(\mathbf{s}, t)$ in SPDE (4). This diffusion is not only on the direction of the transport, where we aim to reduce the oscillations, but also on the orthogonal direction, where there is no problem of convergence. For this reason, we use a different stabilization method, called *Streamline Diffusion* method (SD) (Hughes and Brooks, 1981), that considers only an artificial diffusion along the advection direction by adding in the left-hand side of SPDE (4) the term

$$\mathcal{S}_{SD}(u_h, v_h) = \frac{h}{|\boldsymbol{\gamma}|} \int_\Omega (\boldsymbol{\gamma} \cdot \nabla u_h)(\boldsymbol{\gamma} \cdot \nabla v_h) \, \mathrm{d}\,\mathbf{s}\,. \tag{C10}$$

It is worth emphasizing that in both the stabilization terms (C9) and (C10) the scaling coefficient h has been introduced to recover consistency. Both methods are only weakly consistent and provide an error that is $\mathcal{O}(h)$ if finite elements are used (first order convergent). In our work, we opt for the Streamline Diffusion method and define $\mathcal{S} = \mathcal{S}_{SD}$ for ease of notation.

Appendix D Global precision matrix

We present here the proof of Proposition 6. Let us denote $\mathbf{x}_{1:N_T} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_T}]^\top$ the vector containing all spatial solutions until time step N_T. Then,

$$\mathbf{x}_{1:N_T} = \mathbf{R} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{z}_{2:N_T} \end{pmatrix},$$

with $\mathbf{z}_{2:N_T} = [\mathbf{z}_2, \dots, \mathbf{z}_{N_T}]^\top$ and

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}_{N_S} & 0 & 0 & 0 & \dots & 0 \\ \mathbf{D} & \mathbf{E} & 0 & 0 & \dots & 0 \\ \mathbf{D}^2 & \mathbf{D}\,\mathbf{E} & \mathbf{E} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \mathbf{D}^2 & \mathbf{D} & \mathbf{E} \end{pmatrix}.$$

$\mathbf{R}$ has a block structure which allows easy computation of its inverse

$$\mathbf{R}^{-1} = \begin{pmatrix} \mathbf{I}_{N_S} & 0 & 0 & 0 & \dots & 0 \\ -\mathbf{E}^{-1}\mathbf{D} & \mathbf{E}^{-1} & 0 & 0 & \dots & 0 \\ 0 & -\mathbf{E}^{-1}\mathbf{D} & \mathbf{E}^{-1} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & -\mathbf{E}^{-1}\mathbf{D} & \mathbf{E}^{-1} \end{pmatrix}.$$

The precision matrix of $\mathbf{x}_{1:N_T}$ is thus

$$\mathbf{Q} = {\mathbf{R}^{-1}}^\top \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & 0 & \dots & 0 \\ 0 & \mathbf{I}_{N_S} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{I}_{N_S} \end{pmatrix} \mathbf{R}^{-1}.$$

By denoting $\mathbf{F} = \mathbf{E}\,\mathbf{E}^\top$, the global precision matrix reads

$$\mathbf{Q} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1} + \mathbf{D}^\top\mathbf{F}^{-1}\mathbf{D} & -\mathbf{D}^\top\mathbf{F}^{-1} & 0 & \dots & 0 \\ -\mathbf{F}^{-1}\mathbf{D} & \mathbf{F}^{-1} + \mathbf{D}^\top\mathbf{F}^{-1}\mathbf{D} & -\mathbf{D}^\top\mathbf{F}^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -\mathbf{F}^{-1}\mathbf{D} & \mathbf{F}^{-1} + \mathbf{D}^\top\mathbf{F}^{-1}\mathbf{D} & -\mathbf{D}^\top\mathbf{F}^{-1} \\ 0 & \dots & 0 & -\mathbf{F}^{-1}\mathbf{D} & \mathbf{F}^{-1} \end{pmatrix}.$$

By replacing the values of $\mathbf{D}$ and $\mathbf{F}$ and by defining $\mathbf{J} = \left[\mathbf{M} + \frac{dt}{c}(\mathbf{K} + \mathbf{B})\right]$, we obtain

$$\mathbf{Q} = \frac{c}{\tau^2 dt} \begin{pmatrix} \boldsymbol{\Sigma}^{-1} + \mathbf{Q}_S & -\mathbf{M}^{-1}\mathbf{J}\,\mathbf{Q}_S & 0 & \dots & 0 \\ -\mathbf{M}^{-1}\mathbf{J}^\top\mathbf{Q}_S & \mathbf{J}^\top\mathbf{M}^{-1}\mathbf{Q}_S\mathbf{M}^{-1}\mathbf{J} + \mathbf{Q}_S & -\mathbf{M}^{-1}\mathbf{J}\,\mathbf{Q}_S & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -\mathbf{M}^{-1}\mathbf{J}^\top\mathbf{Q}_S & \mathbf{J}^\top\mathbf{M}^{-1}\mathbf{Q}_S\mathbf{M}^{-1}\mathbf{J} + \mathbf{Q}_S & -\mathbf{M}^{-1}\mathbf{J}\,\mathbf{Q}_S \\ 0 & \dots & 0 & -\mathbf{M}^{-1}\mathbf{J}^\top\mathbf{Q}_S & \mathbf{J}^\top\mathbf{M}^{-1}\mathbf{Q}_S\mathbf{M}^{-1}\mathbf{J} \end{pmatrix}.$$

Appendix E Matrix-free approach for solving systems

In computational mathematics, a matrix-free method is an algorithm for solving a linear system of equations that does not store the coefficient matrix explicitly, but accesses the matrix by evaluating matrix-vector products. Such methods can be preferable when the matrix is so big that storing and manipulating it would cost a lot of memory and computing time, even with the use of methods for sparse matrices. Many iterative methods allow for a matrix-free implementation, including the Conjugate Gradient method.

E.1 Conjugate Gradient method

Suppose we want to solve the system of linear equations

$$\mathbf{A}\,\mathbf{x} = \mathbf{b}$$

for the vector $\mathbf{x}$, where the known (n, n) matrix $\mathbf{A}$ is symmetric, positive-definite, and real, and $\mathbf{b}$ is known as well. We denote the unique solution of this system by $\mathbf{x}^\star$. The Conjugate Gradient (CG) method is an iterative method that allows us to approximately solve systems where n is so large that the direct method would take too much time.

We denote the initial guess for $\mathbf{x}^\star$ by $\mathbf{x}_0$ (without loss of generality $\mathbf{x}_0 = 0$). Starting with $\mathbf{x}_0$, at each iteration we need a metric to tell us whether we are closer to the solution $\mathbf{x}^\star$. This metric comes from the fact that the solution $\mathbf{x}^\star$ is also the unique minimizer of the following quadratic function

$$f(\mathbf{x}) = \frac{1}{2}\,\mathbf{x}^\top\,\mathbf{A}\,\mathbf{x} - \mathbf{x}^\top\,\mathbf{b}, \quad \mathbf{x} \in \mathbf{R}^n,$$

The existence of a unique minimizer is guaranteed by the fact that the Hessian matrix of f is symmetric positive-definite $\mathbf{H}(f(\mathbf{x})) = \mathbf{A}$, and that the minimizer solves the initial problem, since $\nabla f(\mathbf{x}) = \mathbf{A}\,\mathbf{x} - \mathbf{b}$.

We take the first basis vector $\mathbf{p}_0$ to be the negative of the gradient of f at $\mathbf{x} = \mathbf{x}_0$, leading to $\mathbf{p}_0 = \mathbf{b} - \mathbf{A}\,\mathbf{x}_0$. The other vectors in the basis will be conjugate to the gradient. Note that $\mathbf{p}_0$ is also the residual $\mathbf{r}_0$ provided by this initial step of the algorithm. In fact, $\mathbf{r}_k$ is such that $\mathbf{r}_k = \mathbf{b} - \mathbf{A}\,\mathbf{x}_k$. The directions $\mathbf{p}_k$ has to be conjugate to each other. To enforce this condition, we require the next search direction to be built out of the current residual and all previous search directions. The conjugation constraint is an orthonormal-type constraint, which makes the algorithm an example of Gram-Schmidt orthonormalization. This gives the expression

$$\mathbf{p}_k = \mathbf{r}_k - \sum_{i<k} \frac{\mathbf{p}_i^\top\,\mathbf{A}\,\mathbf{r}_k}{\mathbf{p}_i^\top\,\mathbf{A}\,\mathbf{p}_i}\,\mathbf{p}_i\,.$$

Following this direction, the next optimal location is given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \, \mathbf{p}_k,$$

with

$$\alpha_k = \frac{\mathbf{p}_k^\top (\mathbf{b} - \mathbf{A}\,\mathbf{x}_k)}{\mathbf{p}_k^\top \, \mathbf{A}\,\mathbf{p}_k} = \frac{\mathbf{p}_k^\top \, \mathbf{r}_k}{\mathbf{p}_k^\top \, \mathbf{A}\,\mathbf{p}_k}.$$

The expression for α_k can be derived if one substitutes the expression for $\mathbf{x}_{k+1}$ into f and minimizes it w.r.t. α_k:

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k + \alpha_k \, \mathbf{p}_k) = g(\alpha_k)$$

$$g'(\alpha_k) \neq 0 \Rightarrow \alpha_k = \frac{\mathbf{p}_k^\top (\mathbf{b} - \mathbf{A}\,\mathbf{x}_k)}{\mathbf{p}_k^\top \, \mathbf{A}\,\mathbf{p}_k}.$$

The algorithm seems to require storage of all previous searching directions and residual vectors, as well as many matrix-vector multiplications, leading to expensive computations. However, a closer analysis of the algorithm shows that $\mathbf{r}_i$ is orthogonal to $\mathbf{r}_j$, i.e. $\mathbf{r}_i^\top \, \mathbf{r}_j = 0$ for $i \neq j$, and $\mathbf{p}_i$ is $\mathbf{A}$-orthogonal to $\mathbf{p}_j$, i.e. $\mathbf{p}_i^\top \, \mathbf{A}\,\mathbf{p}_j = 0$ for $i \neq j$. This means that, as the algorithm progresses, $\mathbf{p}_i$ and $\mathbf{r}_i$ span the same Krylov subspace. $\{\mathbf{r}_i\}$ form the orthogonal basis with respect to the standard inner product, and $\{\mathbf{p}_i\}$ form the orthogonal basis with respect to the inner product induced by $\mathbf{A}$. Therefore, $\mathbf{x}_k$ can be regarded as the projection of $\mathbf{x}$ on the Krylov subspace.

E.2 Gauss-Seidel preconditioner

The general concept behind a preconditioner is the following: given a linear system $\mathbf{A}\,\mathbf{x} = \mathbf{b}$, we want to find the matrix $\mathbf{P}_R$ and/or $\mathbf{P}_L$ such that the condition number of $\mathbf{A}\,\mathbf{P}_R^{-1}$ (right preconditioner) or $\mathbf{P}_L^{-1}\,\mathbf{A}$ (left preconditioner) or $\mathbf{P}_L^{-1}\,\mathbf{A}\,\mathbf{P}_R^{-1}$ is better than for $\mathbf{A}$ and that we can easily solve $\mathbf{P}_L\,\mathbf{y} = \mathbf{g}$ or $\mathbf{P}_R\,\mathbf{y} = \mathbf{g}$ for any $\mathbf{g}$. Then we solve for $\mathbf{A}\,\mathbf{P}_R^{-1}\,\mathbf{y} = \mathbf{b}$ (right preconditioner), $\mathbf{P}_R\,\mathbf{x} = \mathbf{y}$ or $\mathbf{P}_L^{-1}\,\mathbf{A}\,\mathbf{x} = \mathbf{P}_L^{-1}\,\mathbf{b}$ (left preconditioner) or even both $\mathbf{P}_L^{-1}\,\mathbf{A}\,\mathbf{P}_R^{-1}\,\mathbf{y} = \mathbf{P}_L^{-1}\,\mathbf{b}$, $\mathbf{P}_R\,\mathbf{x} = \mathbf{y}$.

The best choice is of course $\mathbf{P} = \mathbf{A}$, but this does not make life easier. One of the ideas is to use other iterative methods as preconditioners, such as the Jacobi method, the Gauss-Seidel method or the SOR(ω) method (Successive over-relaxation).

A well-known method is Gauss-Seidel (GS) method, whose matrix form is here detailed. Given the matrix $\mathbf{A}$, we have

$$\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{L}^\star,$$

where $\mathbf{D}$ is the diagonal of $\mathbf{A}$, $\mathbf{L}$ is lower-triangular part with zero on the diagonal. One iteration of the GS method reads

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{L} + \mathbf{D})^{-1}(\mathbf{A}\,\mathbf{x}_k - \mathbf{b})$$

and we refer to the preconditioner $\mathbf{P} = \mathbf{L} + \mathbf{D}$ as Gauss-Seidel preconditioner. A good property of this preconditioner is that $\rho(\mathbf{I} - (\mathbf{L} + \mathbf{D})^{-1}\,\mathbf{A}) < 1$, where ρ is the spectral radius, i.e., for a positive definite matrix GS-method always converges.