
SPATIO-TEMPORAL RANDOM FIELDS ON MESSED SURFACES DEFINED FROM ADVECTION-DIFFUSION SPDES

Mike PEREIRA, Nicolas DESASSIS and Lucia CLAROTTO

Department of Geosciences and Geoengineering

Mines Paris – PSL University, Fontainebleau, France.

`firstname.lastname@minesparis.psl.eu`

ABSTRACT

The aim of this work is to propose a statistical model for spatio-temporal data on meshed surfaces based on the SPDE modeling approach. To do so, we consider a class of advection-diffusion SPDEs defined on smooth compact orientable closed Riemannian manifolds of dimension 2, and their discretization using a Galerkin approach. We show how this approach allows to easily propose scalable algorithms for the simulation and prediction of Gaussian random fields that are solutions to the discretized SPDE.

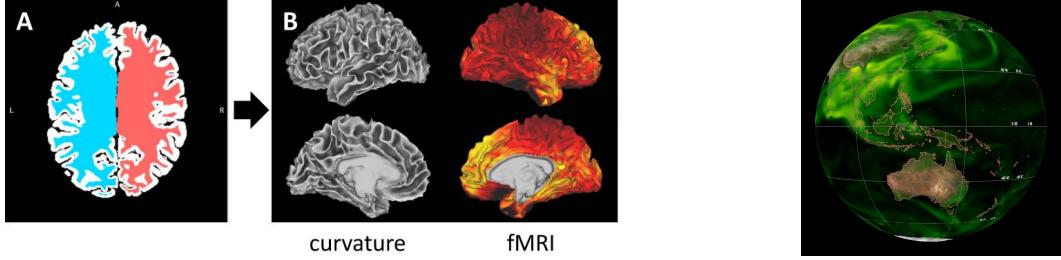
*Note: This document contains links to animated figures colored in dark red. All these animations are collected at:
<https://mike-pereira.github.io/STRF/>.*

1 Introduction

In Geostatistics, when modeling spatio-temporal data, the observed variable is seen as (a realization of) a space-time Gaussian random field (GRF), so that the mere characterization of its mean and covariance functions suffices to fully describe its statistical properties. In particular, these two functions are chosen to mimic the spatio-temporal variability and structure observed in the data [34]. This probabilistic framework has many advantages. On the one hand, it allows to perform simulations of random fields with the same spatio-temporal structure as the one observed in the data, and predictions of at unobserved locations. On the other hand, uncertainties can be quantified both on the variable behavior at unobserved locations (using so-called conditional simulations) or on the model parameters (through Bayesian approaches) [8].

Of particular interest in this work is the setting where the spatial domain on which the data lie is not Euclidean, but rather represents a meshed surface. For instance, this is the case when dealing fMRI data in neuroimaging applications, in which case the data lie on the cortical surface (i.e.. the surface of the brain): see eg. [20] and the illustration in Figure 1.1A. This is also the case when considering global data in environmental applications, for which the data lie on a sphere representing our planet (see eg. [29]), and on this surface transport phenomena (due to winds and currents for instance) can affect the structure of the data (see eg. Figure 1.1B). Hence the main motivations of this work: proposing models for spatio-temporal GRFs flexible enough to represent complex patterns of correlations in data lying on compact meshed surfaces, and but simple enough so that numerically efficient algorithms for their inference, simulation and prediction can be derived.

ACKNOWLEDGMENT The author would like to thank Nicolas Desassis and Lucia Clarotto for their insightful comments and advices.



(A) Volumetric (left) and surface (right) representations of fMRI data. In the right picture, both the curvature of the surface and the BOLD response from the fMRI volume are represented. (Source: [20])

(B) **Simulation of the presence of sulfate in the Earth atmosphere** (Source: NASA Global Modeling and Assimilation Office)

Figure 1.1: Examples of spatio-temporal data distributed on surfaces.

2 Context and state-of-the art

When it comes to defining and building spatio-temporal GRFs to model data, there exists two main approaches: either through the definition of valid covariance functions (which are then “fitted” on the data), or through dynamical models describing the evolution in space and time of the GRFs. Let us review the principle and limits of both approaches.

2.1 The covariance-based approach

Since we consider GRFs, tasks such as sampling from (un)-conditional distributions, predictions (through conditional expectations), and likelihood-based inference can all be performed by solving linear systems or adequately factorizing covariance matrices of the field [34]. Hence, the most straightforward (and classical) approach to spatio-temporal geostatistical modeling consists in fitting valid space-time covariance functions on the data, so that these covariance matrices may be built. Consequently, extensive literature on which covariance functions may be used to model spatio-temporal data, even with complex correlation patterns, is available (see [7, 27, 28] for recent reviews).

Nonetheless, the covariance-based approach has two main drawbacks. First, the matrix factorizations required in sampling, prediction and inference tasks have a complexity that scales as the cube of the number of observations and/or target points, thus making them unfeasible when this last number is large. To circumvent this, simplifying assumptions on the covariance model must be made, such as the separability of space and time dependencies or stationarity. This in turn may result in a lack of realism of the model. Secondly, since they rely on Euclidean or arc-length distances, most of the covariance models available in the literature are restricted to the setting where the spatial domain is either Euclidean or the sphere. Hence, they hardly generalize to other surfaces.

2.2 The dynamic approach

As foretold by its name, this approach relies on models of the dynamic evolution of the GRF in time and space. These models take the form of stochastic partial differential equations (SPDE), the solutions of which are GRFs. This “SPDE approach” to GRF modeling has been popularized by Lindgren et al. [17], and builds on a result from Whittle [35] which states that isotropic GRFs Z on \mathbb{R}^d ($d \in \mathbb{N}$) with a Matérn covariance function are stationary solutions of the SPDE given by

$$(\kappa^2 - \Delta)^{\alpha/2} Z = \tau \mathcal{W}, \quad (1)$$

where $\kappa > 0$, $\alpha > d/2$, $\tau > 0$, and $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudo-differential operator (defined as $(\kappa^2 - \Delta)^{\alpha/2}[\cdot] = \mathcal{F}^{-1}[\mathbf{w} \mapsto (\kappa^2 + \|\mathbf{w}\|^2)^{\alpha/2} \mathcal{F}[\cdot](\mathbf{w})]$), and \mathcal{W} is a Gaussian white noise on \mathbb{R}^d . Solving numerically this SPDE using stochastic finite elements allows to directly obtain an expression for the precision matrix (i.e. the inverse of the covariance matrix) of a GRF with Matérn covariance. Then, at the price of a minor approximation (called mass lumping), these expressions yield a Gaussian Markov random field representation of the GRF, characterized by a sparse precision matrix [18, 30]. This in turn results in significant computational gains since sparse matrix algorithms can be used to deal with the matrix factorizations and linear system solving involved when performing sampling, prediction and inference [13, 17].

The SPDE approach has been extensively used to model spatial data on Euclidean domains (see [18] for a recent review), and extended to model spatial data on surfaces (see e.g. [3, 11, 20]) and more generally on Riemannian manifolds (see e.g. [15, 16]) by replacing the Laplace operator $-\Delta$ in SPDE (1) by a Laplace–Beltrami operator.

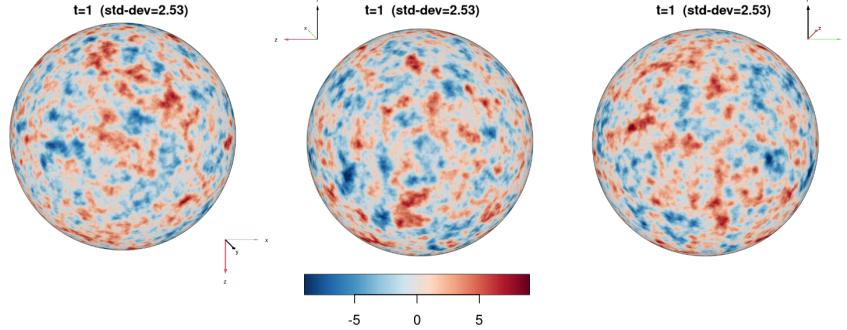


Figure 2.1: Simulation a spatio-temporal diffusion SPDE on the sphere represented from three different viewpoints on the surface (t represents the time step, std-dev the standard deviation of the field value across the surface).

Extensions of the SPDE approach to the spatio-temporal setting have also been proposed. Cameletti et al. [6] propose an approach where the spatial SPDE is coupled with an AR(1) process in time, thus yielding a separable model. Non-separable models based on a direct generalization of SPDE (1) have been proposed by Bakka et al. [1] and Rayner et al. [29], who consider the solutions of diffusion SPDE defined on Euclidean domains and on surfaces by

$$\frac{\partial Z}{\partial t} + (\kappa^2 - \Delta)^{\alpha/2} Z = \tau \mathcal{W}_T \otimes \mathcal{W}_S,$$

where \mathcal{W}_T denotes a temporal white noise and \mathcal{W}_S denotes either a white or colored noise in space. For reference, we provide in Figure 2.1 a simulation of a solution to this SPDE on the sphere. But these models result in random fields with even covariance functions, meaning that changing the sign of the spatial or temporal lag at which the covariance is evaluated does not change the value of the covariance. Consequently, these models are incapable of accounting for transport effects such as advection phenomena (which are intrinsically asymmetrical in time). Note however that in the Euclidean setting, extensions of the SPDE approach allowing to deal with asymmetries in the covariance structure have been proposed by Clarotto et al. [9], Liu et al. [19], Sigrist et al. [33]. However, to the best of our knowledge, the generalization of such models to more complex geometries is left open.

2.3 Proposed approach

The aim of this work is to propose new models for spatio-temporal data on meshed surfaces based on the SPDE modeling approach. To do so, we generalize the approach proposed by Clarotto et al. [9] to model spatio-temporal data on Euclidean domains using an advection-diffusion SPDE, to compact smooth orientable Riemannian manifolds of dimension 2. Then, following the framework proposed in [16], we define a counterpart to the advection-diffusion SPDE on a meshed triangulation of the manifold using Galerkin approximations of the differential operators. We show how this approach allows to easily propose scalable algorithms for the simulation and prediction of Gaussian random fields that are solutions to the resulting SPDE.

3 Advection-diffusion SPDE on a Riemannian manifold

We start by defining the advection-diffusion SPDE on a Riemannian manifold that will be considered on this work. Let (\mathcal{M}, g) denote a compact orientable smooth Riemannian manifold of dimension $d = 2$, without boundary. Let $T > 0$, we denote by $-\Delta_{\mathcal{M}}$ the Laplace–Beltrami operator of the surface. We consider the following advection-diffusion SPDE on the domain $[0, T] \times \mathcal{M}$:

$$\frac{\partial \mathcal{Z}}{\partial t} + \frac{1}{c} (P(-\Delta_{\mathcal{M}}) \mathcal{Z} + \operatorname{div}(\mathcal{Z} \gamma)) = \frac{1}{c} \mu_S + \frac{\tau}{\sqrt{c}} \mathcal{W}_T \otimes \mathcal{Y}_S, \quad (2)$$

where

- P is a polynomial such for all $\lambda \geq 0$, $P(\lambda) \geq \kappa^2$, for some fixed $\kappa > 0$,
- $s \in \mathcal{M} \mapsto \gamma(s)$ is a smooth vector field on $T\mathcal{M}$ (the tangent bundle of \mathcal{M}), such that

$$\kappa^2 + \frac{1}{2} \operatorname{div}(\gamma) \geq \eta_0 \quad (3)$$

for some fixed $\eta_0 > 0$,

- $\mu_S \in L^2(\mathcal{M})$ is a deterministic time-invariant source term,
- $\mathcal{W}_T \otimes \mathcal{Y}_S$ is a space-time separable stochastic forcing given as the product of a time-dependent Gaussian white noise \mathcal{W}_T and a space-dependent colored noise $\mathcal{Y}_S = f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S$ where $f_S : \mathbb{R}_+ \rightarrow \mathbb{R}$ is a bounded function (cf. Appendix A.1 for a definition of colored noise),
- $c > 0$ is a time-scaling parameter, and $\tau > 0$ is a variance-scaling parameter.

More precisely, the forcing term $\mathcal{W}_T \otimes \mathcal{Y}_S$ is defined as a generalized random field acting on functions of $L^2([0, T]) \times L^2(\mathcal{M})$. Let $\langle \cdot, \cdot \rangle$ (resp. $\langle \cdot, \cdot \rangle_T$) denote the usual inner product on $L^2(\mathcal{M})$ (resp. $L^2([0, T])$). Then for any $(\phi_T, \phi_S), (\varphi_T, \varphi_S) \in L^2([0, T]) \times L^2(\mathcal{M})$, $\mathcal{W}_T \otimes \mathcal{Y}_S(\phi_T, \phi_S)$ and $\mathcal{W}_T \otimes \mathcal{Y}_S(\varphi_T, \varphi_S)$ are centered Gaussian random variables, and

$$\text{Cov} \left[\mathcal{W}_T \otimes \mathcal{Y}_S(\phi_T, \phi_S), \mathcal{W}_T \otimes \mathcal{Y}_S(\varphi_T, \varphi_S) \right] = \langle \phi_T, \varphi_T \rangle_T \langle f_S(-\Delta_{\mathcal{M}})\phi_S, f_S(-\Delta_{\mathcal{M}})\varphi_S \rangle.$$

Hence, SPDE (2) can be rewritten in a perhaps more familiar form, for the readers used to stochastic differential equations (SDE) in infinite dimensions [12], as:

$$dZ = -\frac{1}{c}(P(-\Delta_{\mathcal{M}})Z + \text{div}(Z\gamma))dt + \frac{1}{c}\mu_S + \frac{\tau}{\sqrt{c}}d\widetilde{\mathcal{W}}_t. \quad (4)$$

where $\{\widetilde{\mathcal{W}}_t\}_{t \in [0, T]}$ denotes a cylindrical Wiener process (cf. Appendix B for more details). Note that Condition (3) ensures the coercivity of the bilinear form $a(u, v) = \langle P(-\Delta_{\mathcal{M}})u + \text{div}(u\gamma), v \rangle$ which in turn allows to ensure the existence and uniqueness of a weak solution to SPDE (4) [12, Theorem 5.4].

Remark 3.1. By linearity, a solution to (4) can be expressed as sum $Z = z_0 + \mathcal{X}$ where z_0 is a solution to the deterministic differential equation

$$dz_0 = -\frac{1}{c}(P(-\Delta_{\mathcal{M}})z_0 + \text{div}(z_0\gamma))dt + \frac{1}{c}\mu_S.$$

and \mathcal{X} is a solution to the SDE

$$d\mathcal{X} = -\frac{1}{c}(P(-\Delta_{\mathcal{M}})\mathcal{X} + \text{div}(\mathcal{X}\gamma))dt + \frac{\tau}{\sqrt{c}}d\widetilde{\mathcal{W}}_t.$$

where in particular, $\mathbb{E}[Z] = z_0$ and $\mathbb{E}[\mathcal{X}] = 0$.

We conclude this section with a few words about the vector field γ . A natural way to define (and parametrize) a smooth vector field γ on an arbitrary manifold \mathcal{M} is to assume that it is the gradient of a smooth scalar function $\xi : \mathcal{M} \rightarrow \mathbb{R}$, i.e. for any $s \in \mathcal{M}$,

$$\gamma(s) = \nabla\xi(s) \in T_s\mathcal{M}. \quad (5)$$

In this setting, the function ξ can be seen as a potential whose spatial variations locally define the direction of the advection.

Note that this decomposition is not general enough to parametrize all smooth vector fields. For instance, similarly to the Euclidean case, the curl of vector fields defined as in (5) will be zero. A complete characterization of smooth vector fields is given by Helmholtz-Hodge decomposition theorem, which states that vector fields may be uniquely decomposed as the sum of an irrotational component (whose curl is zero), a divergence-free component (whose div is zero) and a harmonic component [2]. In some specific cases, this decomposition can be easily parametrized. For instance, if $\mathcal{M} = \mathbb{S}^2$ is the 2-sphere, we can decompose any tangent vector field γ as

$$\gamma(s) = \nabla\xi(s) + \vec{n}(s) \times \nabla\chi(s) \in T_s\mathbb{S}^2, \quad (6)$$

for some scalar functions $\xi, \chi : \mathbb{S}^2 \rightarrow \mathbb{R}$, and with $\vec{n}(s)$ denoting the vector normal to \mathbb{S}^2 and pointing outwards [22].

4 Advection-diffusion SPDE on a meshed surface

4.1 Definition and discretization of the SPDE

Let \mathcal{M}_h be a discretization of the manifold \mathcal{M} into a polyhedral surface with mesh size $h > 0$ (by triangulation). In particular, following the surface finite element approach we assume the nodes of the polyhedral surface \mathcal{M}_h lie on the surface \mathcal{M} , and that \mathcal{M}_h is close enough to \mathcal{M} that there exists a smooth and invertible function $\ell : \mathcal{M}_h \rightarrow \mathcal{M}$ that maps any point of \mathcal{M}_h to a unique point of \mathcal{M} and vice-versa (i.e. \mathcal{M}_h can be “lifted” to \mathcal{M}).

Let $\{\psi_1, \dots, \psi_N\} \subset H^1(\mathcal{M}_h)$ be the *linear* finite element basis associated with \mathcal{M}_h , where N is the number of nodes of the triangulation. Let then $V_N = \text{span}\{\psi_k : 1 \leq k \leq N\}$ and $V_N^\ell = \text{span}\{\psi_k^\ell : 1 \leq k \leq N\} \subset H^1(\mathcal{M})$ be the set of lifted finite element functions : for any $1 \leq k \leq N$, $\psi_k^\ell = \psi_k \circ \ell^{-1}$.

In order to formulate an advection-diffusion SPDE on \mathcal{M}_h , we look for an approximation of the solution \mathcal{Z} of (2) that can be expressed as a V_N -valued random variable (at any time). We obtain it by approximating each term in the SPDE by operators or variables that “live” on the lifted finite element space V_N^ℓ . This is done in two steps.

Let $-\Delta_N$ and $\text{div}_N(\gamma \cdot)$ be the Galerkin approximation of $-\Delta_{\mathcal{M}}$ over V_N^ℓ (cf. Appendix A.2) and let $\{\lambda_k^{(N)}\}_{1 \leq k \leq N}$ denote its eigenvalues, and $\{e_k^{(N)}\}_{1 \leq k \leq N}$ be a set of associated eigenfunctions forming an orthonormal basis of V_N^ℓ . First, the colored noise \mathcal{Y}_S is approximated by a V_N^ℓ -valued random variable Y_S defined by

$$Y_S = f_S(-\Delta_N)W_S = \sum_{k=1}^n w_k f_S(\lambda_k^{(N)}) e_k^{(N)}, \quad (7)$$

where $\{w_k\}_{1 \leq k \leq N}$ a sequence of independent standard Gaussian variable. Note that this definition holds for any bounded f_S , and is equivalent to the definition of colored noise previously introduced, after replacing H by V_N^ℓ and the Laplace–Beltrami operator $-\Delta_{\mathcal{M}}$ by its Galerkin approximation $-\Delta_N$. Besides, when either of the decompositions (5)-(6) is used, we assume that the potential functions ξ and χ are also taken in V_N^ℓ , which yields a piecewise constant approximation of the vector field. Details about the computations related to such vector fields can be found in [26].

Then, we define an approximation $Z^\ell(t, \cdot) \in V_N^\ell$ of the field $\mathcal{Z}(t, \cdot)$ by considering SPDE (2) and replacing the operators $-\Delta_{\mathcal{M}}$ and $\text{div}(\gamma \cdot)$ by their Galerkin approximations, thus giving

$$\frac{\partial Z^\ell}{\partial t} + \frac{1}{c} \left(P(-\Delta_N)Z^\ell + \text{div}_N(\gamma Z^\ell) \right) = \frac{1}{c} M_S + \frac{\tau}{\sqrt{c}} \mathcal{W}_T \otimes Y_S, \quad t \in [0, T], \quad (8)$$

where M_S denotes the L^2 -orthogonal projection of μ_S onto V_N^ℓ , and $\mathcal{W}_T \otimes Y_S$ is defined in the same way as its counterpart $\mathcal{W}_T \otimes \mathcal{Y}_S$, i.e. as a generalized random field acting on functions of $L^2([0, T]) \times V_N^\ell$. Note in particular that the gradients and integrals now involved in the definition of (8) are now taken on the polyhedral surface.

Remark 4.1. Once again, we can identify $\mathcal{W}_T \otimes Y_S$ with a cylindrical Wiener process $\{\widetilde{W}_t\}_{t \in [0, T]}$ in V_N^ℓ , which can be decomposed as (40) where the eigenfunctions $\{e_j\}_{j \in \mathbb{N}}$ and eigenvalues $\{\lambda_j\}_{j \in \mathbb{N}}$ of $-\Delta_{\mathcal{M}}$, by the eigenfunctions $\{e_j^{(N)}\}_{1 \leq j \leq N}$ and eigenvalues $\{\lambda_j^{(N)}\}_{1 \leq j \leq N}$ of $-\Delta_N$.

We now discretize (8) in time, by applying an implicit Euler scheme with time step $\delta t > 0$. We start by discretizing the time interval $[0, T]$ into $K + 1$ regular time steps of size $\delta t = T/K$, and write $t_k = k\delta t$ for $k \in \{0, \dots, K\}$. Let then $Z^{(k)} = Z(t_k, \cdot)$ denote the approximation of the spatial trace of the solution to SPDE (8) at time t_k . We start from an initial condition $Z^{(0)} = Z(0, \cdot) \in V_N$, taken for instance as the projection onto V_N of the (random) initial condition $\mathcal{Z}(0, \cdot)$ of the original SPDE (2). Then, we have the recursion

$$Z^{(k+1)} - Z^{(k)} + \frac{\delta t}{c} \left(P(-\Delta_N)Z^{(k+1)} + \text{div}_N(\gamma Z^{(k+1)}) \right) = \frac{\delta t}{c} M_S + \tau \sqrt{\frac{\delta t}{c}} Y^{(k+1)}, \quad k \in \mathbb{N}_0, \quad (9)$$

where $\{Y^{(k)}\}_{k \in \mathbb{N}}$ is a sequence of independent samples of Y_S .

The next proposition presents explicit formulas to compute this recursion. First, let us introduce some standard finite element discretization matrices. We denote by \mathbf{C} , \mathbf{R} and \mathbf{B} the matrices whose entries are respectively given by

$$C_{ij} = \langle \psi_i^\ell, \psi_j^\ell \rangle, \quad R_{ij} = \langle \nabla \psi_i^\ell, \nabla \psi_j^\ell \rangle, \quad B_{ij} = \langle \psi_i^\ell, \text{div}(\gamma \psi_j^\ell) \rangle, \quad 1 \leq i, j \leq N, \quad (10)$$

and let $\sqrt{\mathbf{C}} \in \mathbb{R}^{N \times N}$ such that $\mathbf{C} = \sqrt{\mathbf{C}}(\sqrt{\mathbf{C}})^T$. We also introduce the scaled matrices $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{B}}$ defined by

$$\tilde{\mathbf{R}} = (\sqrt{\mathbf{C}})^{-1} \mathbf{R} (\sqrt{\mathbf{C}})^{-T}, \quad \tilde{\mathbf{B}} = (\sqrt{\mathbf{C}})^{-1} \mathbf{B} (\sqrt{\mathbf{C}})^{-T}.$$

Proposition 4.2. Let $\boldsymbol{\mu}^{(0)} = (\mu_1^{(0)}, \dots, \mu_N^{(0)})$, $\mathbf{m} = (m_1, \dots, m_N)$ and, for $0 \leq k \leq K$, let $\mathbf{z}^{(k)} = (z_1^{(k)}, \dots, z_N^{(k)})^T$ such that the time-discretization $\{Z^{(k)}\}_{0 \leq k \leq K}$ of SPDE (8) satisfy

$$M_S = \sum_{j=1}^N m_j \psi_j^\ell, \quad \mathbb{E}[Z^{(0)}] = \sum_{j=1}^N \mu_j^{(0)} \psi_j^\ell, \quad \text{and} \quad Z^{(k)} = \sum_{j=1}^N z_j^{(k)} \psi_j^\ell. \quad (11)$$

We also denote by Γ the matrix defined by

$$\Gamma = \mathbf{I} + \frac{\delta t}{c} (P(\tilde{\mathbf{R}}) + \tilde{\mathbf{B}}). \quad (12)$$

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

$$\begin{cases} \Gamma \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \frac{\delta t}{c} (\sqrt{\mathbf{C}})^T \mathbf{m} + f_{\delta t}(\tilde{\mathbf{R}}) \mathbf{w}^{(k+1)}, \\ \mathbf{z}^{(k+1)} = (\sqrt{\mathbf{C}})^{-T} \mathbf{x}^{(k+1)}, \end{cases} \quad (13)$$

where $\{\mathbf{w}^{(k)}\}_{k \in \mathbb{N}}$ is a sequence of independent centered Gaussian vectors with covariance matrix \mathbf{I} , and $f_{\delta t}$ is the function defined by

$$f_{\delta t}(\lambda) = \tau \sqrt{\frac{\delta t}{c}} f_S(\lambda), \quad \lambda \geq 0,$$

and the matrix function $f_{\delta t}(\tilde{\mathbf{R}})$ is defined in Appendix A.3.

Proof. We write for $k \in \mathbb{N}_0$, $\boldsymbol{\alpha}^{(k)} = (\alpha_1^{(k)}, \dots, \alpha_N^{(k)})^T \in \mathbb{R}^N$, $\mathbf{y}^{(k)} = (y_1^{(k)}, \dots, y_N^{(k)})^T$, where

$$P(-\Delta_N) Z^{(k)} = \sum_{j=1}^N \alpha_j^{(k)} \psi_j^\ell, \quad Y^{(k)} = \sum_{j=1}^N y_j^{(k)} \psi_j^\ell.$$

Besides, let $\boldsymbol{\xi}^{(k)}$ be the vector defined by $\boldsymbol{\xi}^{(k)} = (\langle Z^{(k)}, e_1^{(N)} \rangle, \dots, \langle Z^{(k)}, e_N^{(N)} \rangle)^T$.

Firstly, note that, following [16, Theorem 3.4], we can take

$$\mathbf{y}^{(k)} = (\sqrt{\mathbf{C}})^{-T} f_S(\tilde{\mathbf{R}}) \mathbf{w}^{(k)}, \quad k \in \mathbb{N},$$

where $\{\mathbf{w}^{(k)}\}_{k \in \mathbb{N}}$ is a sequence of independent centered Gaussian vectors with covariance matrix \mathbf{I} .

Then, by testing (9) against ψ_i (for $i \in \{1, \dots, N\}$), and injecting (11), we get the following linear system of equations

$$\mathbf{C} \mathbf{z}^{(k+1)} - \mathbf{C} \mathbf{z}^{(k)} + \frac{\delta t}{c} \left(\mathbf{C} \boldsymbol{\alpha}^{(k+1)} + \mathbf{B} \mathbf{z}^{(k+1)} \right) = \frac{\delta t}{c} \mathbf{C} \mathbf{m} + \tau \sqrt{\frac{\delta t}{c}} \mathbf{C} \mathbf{y}^{(k+1)} \quad (14)$$

On the one hand, following the definition of the map E in (38), we have

$$Z^{(k+1)} = E\left((\sqrt{\mathbf{C}})^T \mathbf{z}^{(k+1)}\right) \quad \text{and} \quad P(-\Delta_N) Z^{(k+1)} = E\left((\sqrt{\mathbf{C}})^T \boldsymbol{\alpha}^{(k+1)}\right) \quad (15)$$

On the other hand, following the definition of the basis $\{e_j^{(N)}\}_{1 \leq j \leq N}$,

$$Z^{(k+1)} = \sum_{j=1}^N \langle Z^{(k+1)}, e_j^{(N)} \rangle E(\mathbf{v}_j) = E\left(\sum_{j=1}^N \langle Z^{(k+1)}, e_j^{(N)} \rangle \mathbf{v}_j\right) = E(\mathbf{V} \boldsymbol{\xi}^{(k+1)})$$

Hence, since E is invertible, we have $(\sqrt{\mathbf{C}})^T \mathbf{z}^{(k+1)} = \mathbf{V} \boldsymbol{\xi}^{(k+1)}$.

Similarly, by definition of $P(-\Delta_N)$, we have

$$P(-\Delta_N) Z^{(k+1)} = \sum_{j=1}^N P(\lambda_j^{(N)}) \langle Z^{(k+1)}, e_j^{(N)} \rangle e_j^{(N)} = E(\mathbf{V} P(\boldsymbol{\Lambda}^{(N)}) \boldsymbol{\xi}^{(k+1)}) = E(P(\tilde{\mathbf{R}}) \mathbf{V} \boldsymbol{\xi}^{(k+1)})$$

Therefore, we have $P(-\Delta_N) Z^{(k+1)} = E(P(\tilde{\mathbf{R}})(\sqrt{\mathbf{C}})^T \mathbf{z}^{(k+1)})$, and using (15) and the fact that E is invertible, we can deduce that $\boldsymbol{\alpha}^{(k+1)} = (\sqrt{\mathbf{C}})^{-T} P(\tilde{\mathbf{R}})(\sqrt{\mathbf{C}})^T \mathbf{z}^{(k+1)}$.

In conclusion, we can now rewrite (14) as

$$\left(\mathbf{C} + \frac{\delta t}{c} (\sqrt{\mathbf{C}}) P(\tilde{\mathbf{R}}) (\sqrt{\mathbf{C}})^T + \frac{\delta t}{c} \mathbf{B} \right) \mathbf{z}^{(k+1)} = \mathbf{C} \mathbf{z}^{(k)} + \frac{\delta t}{c} \mathbf{C} \mathbf{m} + \tau \sqrt{\frac{\delta t}{c}} \mathbf{C} \mathbf{y}^{(k+1)}.$$

By then multiplying both sides of the equality by $(\sqrt{\mathbf{C}})^{-1}$ and introducing $\mathbf{x}^{(k)} = (\sqrt{\mathbf{C}})^T \mathbf{z}^{(k)}$, we retrieve the recursion (13). \square

Remark 4.3. In the case where there is no advection ($\gamma = 0$), an exact time integrator can be proposed for (8). This resulting numerical scheme can be applied to uneven time discretization steps, and relies solely on matrix functions of the scaled matrix $\tilde{\mathbf{R}}$. More details are provided in Appendix D.

Remark 4.4. If the initial condition is chosen so that $\mu^{(0)} = \mu$, then for any $k \in \{0, \dots, K\}$, $\mathbb{E}[\mathbf{z}^{(k)}] = \mu$, i.e. the expectation of the spatial trace of the field remains constant through time.

Finally, in order to define the solutions of advection-diffusion on a the polyhedral surface \mathcal{M}_h , we simply consider the recursion in Proposition 4.2, but replace the finite element matrices \mathbf{C} , \mathbf{R} and \mathbf{B} by their approximations on \mathcal{M}_h . These approximations are obtained by replacing the surface integrals on \mathcal{M} by integrals computed on its polyhedral approximation \mathcal{M}_h when computing the entries of \mathbf{C} , \mathbf{R} and \mathbf{B} , thus giving

$$C_{ij} = \langle \psi_i, \psi_j \rangle_{\mathcal{M}_h}, \quad R_{ij} = \langle \nabla \psi_i, \nabla \psi_j \rangle_{\mathcal{M}_h}, \quad B_{ij} = \langle \psi_i, \operatorname{div}(\gamma \psi_j) \rangle_{\mathcal{M}_h}, \quad 1 \leq i, j \leq N, \quad (16)$$

where $\langle \cdot, \cdot \rangle_{\mathcal{M}_h}$ denotes the inner product over $L^2(\mathcal{M}_h)$, and the gradient and div operator are defined along each (planar) face of \mathcal{M}_h . Such integrals can be easily computed on each face of \mathcal{M}_h and then summed over all the faces. In the remainder of the text, and unless specified otherwise, we assume that entries are computed according to this approach.

We present in Figure 4.1 an example of simulation to the advection-diffusion SPDE on a meshed sphere. We took P to be a polynomial of degree 1, $M_S = 0$ (which implies that $\mathbf{m} = \mathbf{0}$), $\mu^{(0)} = \mathbf{0}$, and f_S to be the inverse of a polynomial of degree 1. The vector field was parametrized thanks to two scalar functions as in (6). As seen in the simulation, the resulting random fields seems to be able to recreate complex convection and diffusion phenomena.

4.2 Mean and covariance of the field

The recursion in Proposition 4.2 can be used to derive, for any $K \in \mathbb{N}$, the expression of the mean and of the covariance matrix of the vectors $\mathbf{z}^{(0)}, \dots, \mathbf{z}^{(K)}$.

Proposition 4.5. Following the notations introduced in Proposition 4.2, let $\mu^{(k)} = \mathbb{E}[\mathbf{z}^{(k)}]$ and let $\Sigma^{(k)} = \operatorname{Var}[\mathbf{z}^{(k)}] = \mathbb{E}[(\mathbf{z}^{(k)} - \mu^{(k)})(\mathbf{z}^{(k)} - \mu^{(k)})^T]$ be respectively the mean and the covariance matrix of the vector $\mathbf{z}^{(k)}$ (for $k \in \mathbb{N}_0$). Then, on the one hand,

$$\mu^{(k)} = \mu + (\sqrt{\mathbf{C}})^{-T} \Gamma^{-k} (\sqrt{\mathbf{C}})^T (\mu^{(0)} - \mu), \quad (17)$$

where we write $\Gamma^{-k} = (\Gamma^{-1})^k$ and

$$\mu = (\sqrt{\mathbf{C}})^{-T} (P(\tilde{\mathbf{R}}) + \tilde{\mathbf{B}})^{-1} (\sqrt{\mathbf{C}})^T \mathbf{m}.$$

And on the other hand,

$$\Sigma^{(k)} = (\sqrt{\mathbf{C}})^{-T} \left(\Gamma^{-k} (\sqrt{\mathbf{C}})^T \Sigma^{(0)} (\sqrt{\mathbf{C}}) (\Gamma^{-k})^T + \sum_{i=1}^k \Gamma^{-i} f_{\delta t}^2(\tilde{\mathbf{R}}) (\Gamma^{-i})^T \right) (\sqrt{\mathbf{C}})^{-1} \quad (18)$$

Proof. On the one hand, by taking the expectation on both sides of the recursion (13), we find $\mu^{(0)} = \mathbb{E}[\mathbf{z}^{(0)}] = (\sqrt{\mathbf{C}})^{-T} \mathbb{E}[\mathbf{x}^{(0)}]$ and for $k \geq 0$,

$$\begin{cases} \Gamma \mathbb{E}[\mathbf{x}^{(k+1)}] = \mathbb{E}[\mathbf{x}^{(k)}] + \frac{\delta t}{c} \tilde{\mathbf{m}}, \\ \mu^{(k+1)} = \mathbb{E}[\mathbf{z}^{(k+1)}] = (\sqrt{\mathbf{C}})^{-T} \mathbb{E}[\mathbf{x}^{(k+1)}], \end{cases}$$

where we take $\tilde{\mathbf{m}} = (\sqrt{\mathbf{C}})^T \mathbf{m}$. Hence, we have for any $k \geq 0$,

$$\begin{aligned} \mathbb{E}[\mathbf{x}^{(k)}] - \frac{\delta t}{c} (\Gamma - \mathbf{I})^{-1} \tilde{\mathbf{m}} &= \Gamma \mathbb{E}[\mathbf{x}^{(k+1)}] - \frac{\delta t}{c} \tilde{\mathbf{m}} - \frac{\delta t}{c} (\Gamma - \mathbf{I})^{-1} \tilde{\mathbf{m}} = \Gamma \mathbb{E}[\mathbf{x}^{(k+1)}] - \frac{\delta t}{c} (\mathbf{I} + (\Gamma - \mathbf{I})^{-1}) \tilde{\mathbf{m}} \\ &= \Gamma \mathbb{E}[\mathbf{x}^{(k+1)}] - \frac{\delta t}{c} ((\Gamma - \mathbf{I}) + \mathbf{I}) (\Gamma - \mathbf{I})^{-1} \tilde{\mathbf{m}} = \Gamma \left(\mathbb{E}[\mathbf{x}^{(k+1)}] - \frac{\delta t}{c} (\Gamma - \mathbf{I})^{-1} \tilde{\mathbf{m}} \right). \end{aligned}$$

By iterating this last equation, we obtain

$$\mathbb{E}[\mathbf{x}^{(0)}] - \frac{\delta t}{c} (\Gamma - \mathbf{I})^{-1} \tilde{\mathbf{m}} = \Gamma^k \left(\mathbb{E}[\mathbf{x}^{(k)}] - \frac{\delta t}{c} (\Gamma - \mathbf{I})^{-1} \tilde{\mathbf{m}} \right).$$

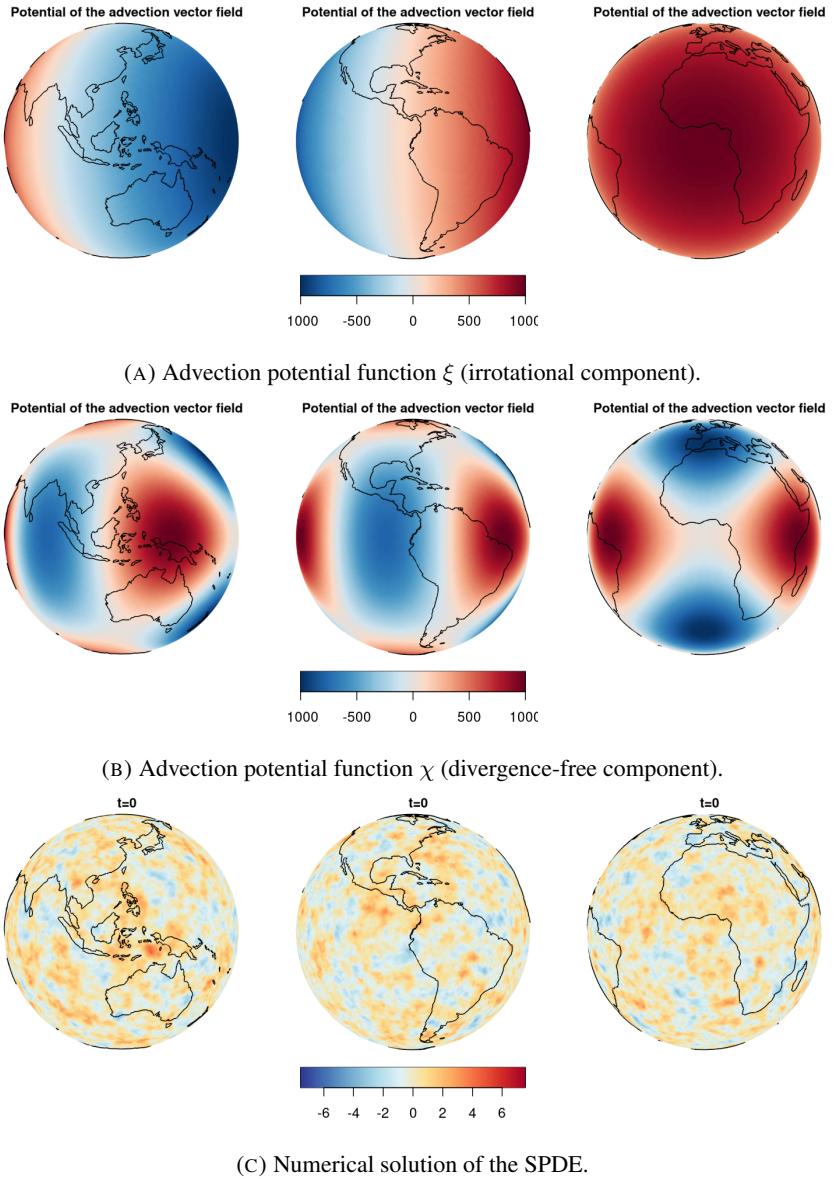


Figure 4.1: Simulation of a solution to the advection-diffusion SPDE on a meshed sphere, represented from three different viewpoints on the surface.

Finally, by multiplying both side of this equality by $(\sqrt{C})^{-T}$, we get

$$\begin{aligned}\boldsymbol{\mu}^{(0)} - \frac{\delta t}{c}(\sqrt{C})^{-T}(\boldsymbol{\Gamma} - \mathbf{I})^{-1}\widetilde{\mathbf{m}} &= \boldsymbol{\mu}^{(0)} - \boldsymbol{\mu} = (\sqrt{C})^{-T}\boldsymbol{\Gamma}^k(\sqrt{C})^T\left(\boldsymbol{\mu}^{(k)} - \frac{\delta t}{c}(\sqrt{C})^{-T}(\boldsymbol{\Gamma} - \mathbf{I})^{-1}\widetilde{\mathbf{m}}\right) \\ &= (\sqrt{C})^{-T}\boldsymbol{\Gamma}^k(\sqrt{C})^T\left(\boldsymbol{\mu}^{(k)} - \boldsymbol{\mu}\right)\end{aligned}$$

which in turn gives (17).

On the other hand, by taking the covariance on both sides of the recursion (13), we have, for any $k \geq 0$,

$$\begin{cases} \boldsymbol{\Gamma} \text{Var}[\mathbf{x}^{(k+1)}]\boldsymbol{\Gamma}^T = \text{Var}[\mathbf{x}^{(k)}] + f_{\delta t}^2(\widetilde{\mathbf{R}}), \\ \boldsymbol{\Sigma}^{(k+1)} = (\sqrt{C})^{-T}\text{Var}[\mathbf{x}^{(k+1)}](\sqrt{C})^{-1}, \end{cases} \quad (19)$$

Let us show by induction that for any $k \geq 0$,

$$\text{Var}[\mathbf{x}^{(k)}] = \boldsymbol{\Gamma}^{-k}\text{Var}[\mathbf{x}^{(0)}](\boldsymbol{\Gamma}^{-k})^T + \sum_{i=1}^k \boldsymbol{\Gamma}^{-i}f_{\delta t}^2(\widetilde{\mathbf{R}})(\boldsymbol{\Gamma}^{-i})^T \quad (20)$$

Formula (20) clearly holds for $k = 0$. Let us now assume that it holds for some fixed index $k \in \mathbb{N}_0$. Then, following (19), we have

$$\begin{aligned}\text{Var}[\mathbf{x}^{(k+1)}] &= \boldsymbol{\Gamma}^{-1}(\text{Var}[\mathbf{x}^{(k)}] + f_{\delta t}^2(\widetilde{\mathbf{R}}))\boldsymbol{\Gamma}^{-T} \\ &= \boldsymbol{\Gamma}^{-(k+1)}\text{Var}[\mathbf{x}^{(0)}](\boldsymbol{\Gamma}^{-(k+1)})^T + \sum_{i=1}^k \boldsymbol{\Gamma}^{-(i+1)}f_{\delta t}^2(\widetilde{\mathbf{R}})(\boldsymbol{\Gamma}^{-(i+1)})^T + \boldsymbol{\Gamma}^{-1}f_{\delta t}^2(\widetilde{\mathbf{R}})\boldsymbol{\Gamma}^{-T} \\ &= \boldsymbol{\Gamma}^{-(k+1)}\text{Var}[\mathbf{x}^{(0)}](\boldsymbol{\Gamma}^{-(k+1)})^T + \sum_{i=1}^{k+1} \boldsymbol{\Gamma}^{-i}f_{\delta t}^2(\widetilde{\mathbf{R}})(\boldsymbol{\Gamma}^{-i})^T\end{aligned}$$

Hence, Formula (20) also holds for the index $(k+1)$. In conclusion, by induction, Formula (20) holds for any $k \geq 0$. We then retrieve (18) as a consequence of (19). \square

We end this subsection with a result providing an alternative way to compute the mean and covariance vectors exposed in Proposition 4.8. Let us introduce some notations. For $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2 \in \mathbb{R}^{N \times N}$, let $\mathbf{L}(\boldsymbol{\Theta}_1)$, $\mathbf{D}(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2)$, $\mathbf{D}(\boldsymbol{\Theta}_1) \in \mathbb{R}^{(K+1)N \times (K+1)N}$ be the block matrices given by

$$\mathbf{L}(\boldsymbol{\Theta}_1) = \begin{pmatrix} \mathbf{I} & & & \\ -\mathbf{I} & \boldsymbol{\Theta}_1 & & \\ & \ddots & \ddots & \\ & & -\mathbf{I} & \boldsymbol{\Theta}_1 \end{pmatrix}, \quad \mathbf{D}(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) = \begin{pmatrix} \boldsymbol{\Theta}_1 & & & \\ & \boldsymbol{\Theta}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{\Theta}_2 \end{pmatrix}, \quad \text{and} \quad \mathbf{D}(\boldsymbol{\Theta}_1) = \mathbf{D}(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_1). \quad (21)$$

The next proposition gives the expression of the precision matrix of the coefficients $\mathbf{z}^{(0)}, \dots, \mathbf{z}^{(K)}$ obtained through the recursion (13).

Proposition 4.6. *Let us assume that the initial condition of SPDE (8) can be expressed as*

$$Z^{(0)} = \boldsymbol{\mu}^{(0)} + f_0(-\Delta_N)W_S$$

for some (deterministic) $\boldsymbol{\mu}^{(0)} \in V_N$ and some $f_0 : \mathbb{R}_+ \rightarrow \mathbb{R}$ that is bounded takes positive values (meaning in particular that $\mathbb{E}[Z^{(0)}] = \boldsymbol{\mu}^{(0)}$). Let then $\mathbf{Z}, \mathbf{M}_{\delta t} \in \mathbb{R}^{(K+1)N}$ be defined by

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}^{(0)} \\ \mathbf{z}^{(1)} \\ \vdots \\ \mathbf{z}^{(K)} \end{pmatrix} \quad \text{and} \quad \mathbf{M}_{\delta t} = \begin{pmatrix} \boldsymbol{\mu}^{(0)} \\ \frac{\delta t}{c}\mathbf{m} \\ \vdots \\ \frac{\delta t}{c}\mathbf{m} \end{pmatrix}$$

where the vectors the vectors $\mu^{(0)}$, \mathbf{m} , and $\{\mathbf{z}^{(k)}\}_{0 \leq k \leq K}$ are given in Proposition 4.2. Let μ_Z be the expectation and Q_Z be the precision matrix of Z . Then, we have

$$\mu_Z = \mathbb{E}[Z] = D((\sqrt{C})^{-T})L(\Gamma)^{-1}D((\sqrt{C})^T)M_{\delta t}, \quad (22)$$

and

$$Q_Z = \mathbb{E}[(Z - \mu_Z)(Z - \mu_Z)^T]^{-1} = D(\sqrt{C}) L(\Gamma)^T D(f_0^{-2}(\tilde{R}), f_{\delta t}^{-2}(\tilde{R})) L(\Gamma) D((\sqrt{C})^T). \quad (23)$$

Proof. Consider the vectors $\{\mathbf{w}^{(k)}, \mathbf{x}^{(k)}\}_{0 \leq k \leq K}$ defined in Proposition 4.2, and define the vectors $\mathbf{W} = ((\mathbf{w}^{(0)})^T, \dots, (\mathbf{w}^{(K)})^T)^T \in \mathbb{R}^{(K+1)N}$ and $\mathbf{X} = ((\mathbf{x}^{(0)})^T, \dots, (\mathbf{x}^{(K)})^T)^T \in \mathbb{R}^{(K+1)N}$. In particular, \mathbf{W} is a centered Gaussian vector with precision matrix $Q_W = I$. Besides, note that by definition of $Z(0, \cdot)$, we can write $\mathbf{z}^{(0)} = \mathbf{m}^{(0)} + (\sqrt{C})^{-T} f_0(\tilde{R}) \mathbf{w}^{(0)}$. On the one hand, we can rewrite the first equality of (13) in matrix form as

$$L(\Gamma) \mathbf{X} = D((\sqrt{C})^T) M_{\delta t} + D(f_0(\tilde{R}), f_{\delta t}(\tilde{R})) \mathbf{W}. \quad (24)$$

Taking the expectation on both sides of (24), we then get $\mathbb{E}[\mathbf{X}] = L(\Gamma)^{-1} D((\sqrt{C})^T) M_{\delta t}$.

On the one hand, since by definition of $\mathbf{x}^{(k)}$ ($k \geq 0$) in Proposition 4.2, we have $\mathbf{X} = D((\sqrt{C})^T) Z$, we can conclude that

$$\mu_Z = D((\sqrt{C})^{-T}) L(\Gamma)^{-1} D((\sqrt{C})^T) M_{\delta t}.$$

On the other hand, we can also write

$$L(\Gamma)(\mathbf{X} - \mathbb{E}[\mathbf{X}]) = D(f_0(\tilde{R}), f_{\delta t}(\tilde{R})) \mathbf{W}. \quad (25)$$

Taking the covariance of both side of this equality, we get

$$L(\Gamma) Q_X^{-1} L(\Gamma)^T = D(f_0^2(\tilde{R}), f_{\delta t}^2(\tilde{R})),$$

where we used the fact that $D(f_0(\tilde{R}), f_{\delta t}(\tilde{R}))$ is block diagonal and that $Q_W = I$. By then inverting both sides, we obtain the following relation for the precision matrix Q_X of \mathbf{X} :

$$Q_X = L(\Gamma)^T D(f_0^{-2}(\tilde{R}), f_{\delta t}^{-2}(\tilde{R})) L(\Gamma).$$

Finally, recall that $\mathbf{X} = D((\sqrt{C})^T) Z$. By once again taking the covariance and then inverting both sides of this equality, we retrieve (23). \square

Remark 4.7. The explicit computation of the precision matrix Q_Z results in a block tri-diagonal matrix given by

$$Q_Z = \begin{pmatrix} \Gamma_0 & -\Gamma_2 & & \\ -\Gamma_2^T & \Gamma_1 & -\Gamma_2 & \\ & \ddots & \ddots & \ddots \\ & & -\Gamma_2^T & \Gamma_1 & -\Gamma_2 \\ & & & -\Gamma_2^T & \Gamma_3 \end{pmatrix},$$

where

$$\begin{aligned} \Gamma_0 &= (\sqrt{C})(f_0^{-2}(\tilde{R}) + f_{\delta t}^{-2}(\tilde{R}))(\sqrt{C})^T, & \Gamma_1 &= (\sqrt{C})(\Gamma^T f_{\delta t}^{-2}(\tilde{R}) \Gamma + f_{\delta t}^{-2}(\tilde{R}))(\sqrt{C})^T, \\ \Gamma_2 &= (\sqrt{C})f_{\delta t}^{-2}(\tilde{R})\Gamma(\sqrt{C})^T, & \Gamma_3 &= (\sqrt{C})\Gamma^T f_{\delta t}^{-2}(\tilde{R})\Gamma(\sqrt{C})^T. \end{aligned}$$

As an illustration, starting from the same model as the one simulated in Figure 4.1, we represent in Figure 4.2 the spatio-temporal evolution of the covariance between the value of the field at time $t = 0$ at three reference points (in blue), and the values of the field elsewhere and at later times. These covariance are computed using the formula in Proposition 4.6 for the spatio-temporal precision matrix of the field. As one can note in the animation, the zone of high-correlation “moves” along the advection direction, as expected in an advection problem.

The covariance matrix Q_Z of the Euler discretization of Z can be expressed as the product of block diagonal and block bi-diagonal matrices. This particular structure allows to propose scalable algorithms to perform products with vectors, solving linear systems and computing the log-determinant, all of which will prove useful when tackling inference and prediction problems (cf. Section 5.2).

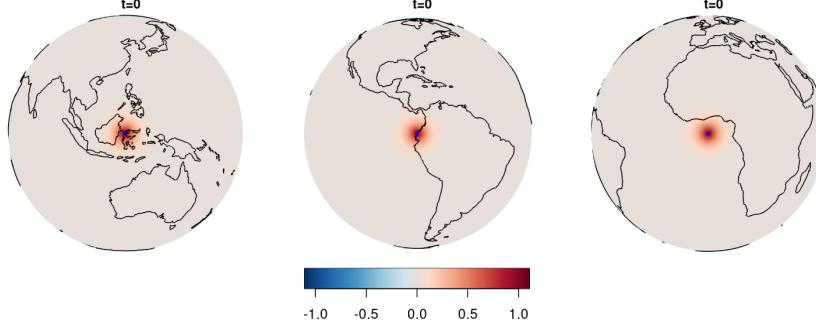


Figure 4.2: Spatio-temporal evolution of the covariance between a reference point (in blue) and the rest of the points in the domain, for the spatio-temporal model simulated in Figure 4.1. The color (red to blue) represent the value of the covariance.

4.3 Stability of the recursion

Similarly to the Euclidean case, the Galerkin approximation of advection-diffusion SPDEs on surfaces are subject to instabilities when the advection term dominates the diffusion term [21]. This phenomenon can be observed in Figure 4.3B, where a GRF on a cortical surface is simulated using the scheme described in Proposition 4.6: as one can notice in the animation, the values taken by the field quickly explode and tend to oscillate. These instabilities are due to the fact that in advection-dominated cases, the matrix Γ in (12) dominates and can cause the linear systems involving Γ to become ill-conditioned.

A natural question that now arises is whether or not the recursion in Proposition 4.2 is stable. By stability, we mean here that the mean and covariance matrices exposed in Proposition 4.8 stay at least bounded for any $k \in \mathbb{N}_0$. It turns out that under a mild condition on the (symmetric part of) advection matrix $\tilde{\mathbf{B}}$ we can ensure that the sequences $\{\mu^{(k)}\}_{k \in \mathbb{N}_0}$ and $\{\Sigma^{(k)}\}_{k \in \mathbb{N}_0}$ are not only bounded, but also are convergent.

Proposition 4.8. *Following the notations introduced in Proposition 4.2, let $\mu^{(k)} = \mathbb{E}[\mathbf{z}^{(k)}]$ and let $\Sigma^{(k)} = \text{Var}[\mathbf{z}^{(k)}]$ be respectively the mean and the covariance matrix of the vector $\mathbf{z}^{(k)}$, for any $k \in \mathbb{N}_0$.*

If the matrix $\tilde{\mathbf{G}} = \frac{1}{2}(\tilde{\mathbf{B}} + \tilde{\mathbf{B}}^T)$ satisfies $\kappa^2 + \lambda_{\min}(\tilde{\mathbf{G}}) > 0$, then the recursion in Proposition 4.2 is stable in the sense that the sequences $\{\mu^{(k)}\}_{k \in \mathbb{N}_0}$ and $\{\Sigma^{(k)}\}_{k \in \mathbb{N}_0}$ are bounded. Moreover, both sequences converge to the following limits

$$\lim_{k \rightarrow \infty} \mu^{(k)} = \mu, \quad \text{and} \quad \lim_{k \rightarrow \infty} \Sigma^{(k)} = (\sqrt{C})^{-T} \left(\sum_{i=1}^{\infty} \Gamma^{-i} f_{\delta t}^2(\tilde{\mathbf{R}}) (\Gamma^{-i})^T \right) (\sqrt{C})^{-1}. \quad (26)$$

Proof. Let $\mathbf{x} \in \mathbb{R}^n$. Note that $\mathbf{x}^T \Gamma \mathbf{x} = \mathbf{x}^T \Gamma^T \mathbf{x} = \frac{1}{2} \mathbf{x}^T (\Gamma + \Gamma^T) \mathbf{x}$. Hence,

$$\mathbf{x}^T \Gamma \mathbf{x} = \mathbf{x}^T \left(\frac{1}{2} (\Gamma + \Gamma^T) \right) \mathbf{x} = \mathbf{x}^T \left(\mathbf{I} + \frac{\delta t}{c} (P(\tilde{\mathbf{R}}) + \tilde{\mathbf{G}}) \right) \mathbf{x}$$

since the matrix $P(\tilde{\mathbf{R}})$ is symmetric. First, note that the min-max principle gives that $\mathbf{x}^T \tilde{\mathbf{G}} \mathbf{x} \geq \lambda_{\min}(\tilde{\mathbf{G}}) \|\mathbf{x}\|^2$. Note then that since the matrix $\tilde{\mathbf{R}}$ is positive semi-definite, its eigenvalues are lower-bounded by zero. Therefore, given that P is lower-bounded by κ^2 over \mathbb{R}_+ , we can deduce that $\lambda_{\min}(P(\tilde{\mathbf{R}})) \geq \kappa^2$. Finally, using the min-max principle, we can deduce that $\mathbf{x}^T P(\tilde{\mathbf{R}}) \mathbf{x} \geq \lambda_{\min}(P(\tilde{\mathbf{R}})) \|\mathbf{x}\|^2 \geq \kappa^2 \|\mathbf{x}\|^2$. In conclusion, we have for any $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^T \Gamma \mathbf{x} \geq \alpha \|\mathbf{x}\|^2$$

where $\alpha = \left(1 + \frac{\delta t}{c} (\kappa^2 + \lambda_{\min}(P(\tilde{\mathbf{G}}))) \right) > 1$.

Then, for any $\mathbf{x} \in \mathbb{R}^n$,

$$\alpha \|\Gamma^{-1} \mathbf{x}\|^2 \leq \mathbf{x}^T \Gamma^{-T} \Gamma \Gamma^{-1} \mathbf{x} = \mathbf{x}^T \Gamma^{-T} \mathbf{x} = \mathbf{x}^T \Gamma^{-1} \mathbf{x} \leq \|\mathbf{x}\| \|\Gamma^{-1} \mathbf{x}\|$$

which gives $\|\Gamma^{-1} \mathbf{x}\| \leq \alpha^{-1} \|\mathbf{x}\|$. Similarly, we have for any $\mathbf{x} \in \mathbb{R}^n$, $\alpha \|\Gamma^{-T} \mathbf{x}\|^2 \leq \mathbf{x}^T \Gamma^{-T} \mathbf{x} \leq \|\mathbf{x}\| \|\Gamma^{-T} \mathbf{x}\|$ which gives $\|\Gamma^{-T} \mathbf{x}\| \leq \alpha^{-1} \|\mathbf{x}\|$.

Consequently, we have for any $k \in \mathbb{N}_0$,

$$\begin{aligned}\|\boldsymbol{\mu}^{(k)} - \boldsymbol{\mu}\| &= \|(\sqrt{\mathbf{C}})^{-T} \boldsymbol{\Gamma}^{-k} (\sqrt{\mathbf{C}})^T (\boldsymbol{\mu}^{(0)} - \boldsymbol{\mu})\| \leq \|(\sqrt{\mathbf{C}})^{-T}\| \cdot \|\boldsymbol{\Gamma}^{-k} (\sqrt{\mathbf{C}})^T (\boldsymbol{\mu}^{(0)} - \boldsymbol{\mu})\| \\ &\leq \|(\sqrt{\mathbf{C}})^{-T}\| \cdot \alpha^{-k} \|(\sqrt{\mathbf{C}})^T (\boldsymbol{\mu}^{(0)} - \boldsymbol{\mu})\| \leq \alpha^{-k} \|(\sqrt{\mathbf{C}})^{-T}\| \cdot \|(\sqrt{\mathbf{C}})^T\| \cdot \|\boldsymbol{\mu}^{(0)} - \boldsymbol{\mu}\|\end{aligned}$$

Since $\alpha > 1$ we have $\alpha^{-k} \in (0, 1)$ for any $k \in \mathbb{N}_0$ and therefore the sequence $\{\boldsymbol{\mu}^{(k)}\}_{k \in \mathbb{N}_0}$ is bounded. Moreover, by taking the limit as $k \rightarrow \infty$ on both sides of the inequality, we get $\lim_{k \rightarrow \infty} \|\boldsymbol{\mu}^{(k)} - \boldsymbol{\mu}\| = 0$ and therefore $\lim_{k \rightarrow \infty} \boldsymbol{\mu}^{(k)} = \boldsymbol{\mu}$.

For $k \in \mathbb{N}$, let then $\mathbf{X}^{(k)} = \boldsymbol{\Gamma}^{-k} (\sqrt{\mathbf{C}})^T \boldsymbol{\Sigma}^{(0)} (\sqrt{\mathbf{C}}) (\boldsymbol{\Gamma}^{-k})^T$ and $\mathbf{Y}^{(k)} = \sum_{i=1}^k \boldsymbol{\Gamma}^{-i} f_{\delta t}^2(\tilde{\mathbf{R}}) (\boldsymbol{\Gamma}^{-i})^T$. In particular, we then have $\boldsymbol{\Sigma}^{(k)} = (\sqrt{\mathbf{C}})^{-T} (\mathbf{X}^{(k)} + \mathbf{Y}^{(k)}) (\sqrt{\mathbf{C}})^{-1}$. For any $\mathbf{x} \in \mathbb{R}^n$ we have

$$\begin{aligned}\|\mathbf{X}^{(k)} \mathbf{x}\| &= \|\boldsymbol{\Gamma}^{-k} (\sqrt{\mathbf{C}})^T \boldsymbol{\Sigma}^{(0)} (\sqrt{\mathbf{C}}) (\boldsymbol{\Gamma}^{-k})^T \mathbf{x}\| \leq \alpha^{-k} \|(\sqrt{\mathbf{C}})^T \boldsymbol{\Sigma}^{(0)} (\sqrt{\mathbf{C}}) (\boldsymbol{\Gamma}^{-k})^T \mathbf{x}\| \\ &\leq \alpha^{-k} \|(\sqrt{\mathbf{C}})^T \boldsymbol{\Sigma}^{(0)} (\sqrt{\mathbf{C}})\| \|(\boldsymbol{\Gamma}^{-k})^T \mathbf{x}\| \leq \alpha^{-2k} \|(\sqrt{\mathbf{C}})^T \boldsymbol{\Sigma}^{(0)} (\sqrt{\mathbf{C}})\| \|\mathbf{x}\|\end{aligned}$$

meaning in particular that $\|\mathbf{X}^{(k)}\| \leq \alpha^{-2k} \|(\sqrt{\mathbf{C}})^T \boldsymbol{\Sigma}^{(0)} (\sqrt{\mathbf{C}})\|$. Similarly, we have for any $\mathbf{x} \in \mathbb{R}^n$

$$\|\mathbf{Y}^{(k)} \mathbf{x}\| = \left\| \sum_{i=1}^k \boldsymbol{\Gamma}^{-i} f_{\delta t}^2(\tilde{\mathbf{R}}) (\boldsymbol{\Gamma}^{-i})^T \mathbf{x} \right\| \leq \sum_{i=1}^k \|\boldsymbol{\Gamma}^{-i} f_{\delta t}^2(\tilde{\mathbf{R}}) (\boldsymbol{\Gamma}^{-i})^T \mathbf{x}\| \leq \sum_{i=1}^k \alpha^{-2i} \|f_{\delta t}^2(\tilde{\mathbf{R}})\| \|\mathbf{x}\|$$

meaning in particular that

$$\|\mathbf{Y}^{(k)}\| \leq \|f_{\delta t}^2(\tilde{\mathbf{R}})\| \sum_{i=1}^k \alpha^{-2i} = \frac{1 - \alpha^{-2k}}{1 - \alpha^{-2}} \|f_{\delta t}^2(\tilde{\mathbf{R}})\|$$

And, similarly as before, we can conclude that the sequences $\{\mathbf{X}^{(k)}\}_{k \in \mathbb{N}_0}$ and $\{\mathbf{Y}^{(k)}\}_{k \in \mathbb{N}_0}$ are both bounded, and therefore that the sequence $\{\boldsymbol{\Sigma}^{(k)}\}_{k \in \mathbb{N}_0}$ is bounded. Besides by taking the limit as $k \rightarrow \infty$, we get that $\|\mathbf{X}^{(k)}\| \rightarrow 0$ and that the matrix series $\mathbf{Y}^{(k)}$ converges (absolutely). Hence, we end up with

$$\lim_{k \rightarrow \infty} \boldsymbol{\Sigma}^{(k)} = (\sqrt{\mathbf{C}})^{-T} \left(\lim_{k \rightarrow \infty} \mathbf{X}^{(k)} + \lim_{k \rightarrow \infty} \mathbf{Y}^{(k)} \right) (\sqrt{\mathbf{C}})^{-1} = (\sqrt{\mathbf{C}})^{-T} \left(\sum_{i=1}^{\infty} \boldsymbol{\Gamma}^{-i} f_{\delta t}^2(\tilde{\mathbf{R}}) (\boldsymbol{\Gamma}^{-i})^T \right) (\sqrt{\mathbf{C}})^{-1}.$$

□

Hence, whenever the smallest eigenvalue of the symmetric part of the matrix $\tilde{\mathbf{B}}$ is strictly lower-bounded by $-\kappa^2$, no instability occurs. It turns out that when the finite element matrices \mathbf{C} , \mathbf{R} and \mathbf{B} are computed according to (10) this last condition is in automatically satisfied whenever the divergence $\text{div}(\gamma)$ satisfies the condition (cf. Proposition E.1)

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

Note that this last inequality is one of the conditions we assumed for the well-posedness of the SPDE (2), meaning that within our assumptions the recursion is always stable.

However, recall that in practice we replace the surface integrals on \mathcal{M} by integrals computed on its polyhedral approximation \mathcal{M}_h when computing the finite element matrices \mathbf{C} , \mathbf{R} and \mathbf{B} (according to (16)). When doing so, differences of order $\mathcal{O}(h^2)$ (where h is the size of the mesh) are to be expected when comparing both types of integrals. Consequently, the stability of the recursion is no longer guaranteed. However, by decreasing size of the mesh, we reduce the discrepancy between the two types of integrals and therefore we can retrieve stability (which is guaranteed at the limit $h \rightarrow 0$). In practice though, we only need to decrease the mesh size until the matrix $\tilde{\mathbf{G}}$ in Proposition 4.8 satisfies the inequality $\kappa^2 + \lambda_{\min}(\tilde{\mathbf{G}}) > 0$.

Another approach to deal with instabilities consists in introducing a stabilization term to advection-diffusion SPDEs [10]. For instance, following the approach of Clarotto et al. [9], we can use the so-called Streamline Diffusion stabilization method [5] for which we add an additional term in the variational formulation of SPDE (8), namely the bilinear form defined on V_N by

$$a_S(v_1, v_2) = h \langle g(\gamma, \nabla v_1), \frac{1}{\sqrt{g(\gamma, \gamma)}} g(\gamma, \nabla v_2) \rangle, \quad v_1, v_2 \in V_N,$$

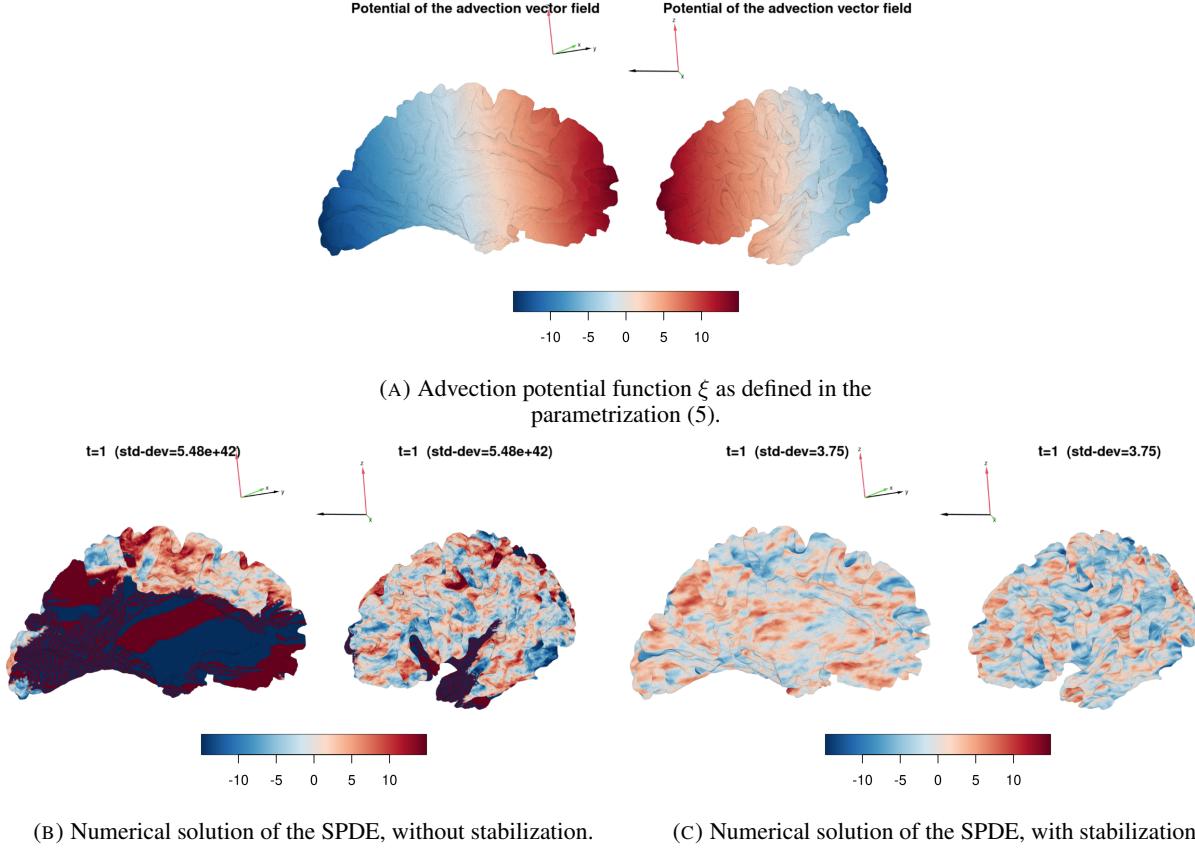


Figure 4.3: Simulation of the spatio-temporal advection-diffusion SPDE on a cortical surface, represented from three different viewpoints on the surface (t represents the time step, std-dev the standard deviation of the field values across the surface).

where h denotes the size of the triangulation mesh. This term effectively adds a small diffusion along the local direction of the advection γ . Note that scaling this term by h allows to minimize its impact on the SPDE solution: for instance, in the Euclidean setting, the resulting errors due to the introduction of this term are of order $\mathcal{O}(h)$.

In practice, when adding the stabilization term SPDE (8), the derivation of the recursion remains the same. The only change is the introduction of an additional matrix in the definition (12) of Γ , which now reads:

$$\Gamma = \mathbf{I} + \frac{\delta t}{c} (P(\tilde{\mathbf{R}}) + \tilde{\mathbf{B}} + \tilde{\mathbf{S}}), \quad (27)$$

where $\tilde{\mathbf{S}} = (\sqrt{C})^{-1} \mathbf{S} (\sqrt{C})^{-T}$ is the scaled stabilization matrix, and the entries of \mathbf{S} are given by

$$S_{ij} = h \langle g(\gamma, \nabla \psi_i), \frac{1}{\sqrt{g(\gamma, \gamma)}} g(\gamma, \nabla \psi_j) \rangle, \quad 1 \leq i, j \leq N.$$

As seen Figure 4.3C, this method allowed us to get rid of the instabilities in the simulation.

5 Prediction and inference from data

Let $T > 0$. We assume that the time interval $[0, T]$ is discretized into $K + 1$ regular time steps of size $\delta t = T/K$. We hence write $t_k = k\delta t$ for $k \in \{0, \dots, K\}$.

We consider a statistical model with fixed and random effects to model observations of some variable u in the spatiotemporal domain $[0, T] \times \mathcal{M}_h$. The fixed effects correspond to a regression of a set of q covariates, and the random effects are modeled as the solution ψ , to which an independent measurement noise is added. More precisely, for each $k \in \{0, \dots, K\}$, we assume that we have n_k observations of u at some locations $s_1^{(k)}, \dots, s_{n_k}^{(k)} \in \mathcal{M}_h$. We then denote

by $\mathbf{u}^{(k)} \in \mathbb{R}^{n_k}$ the vector containing these observations, i.e. $\mathbf{u}^{(k)} = (u(t_k, s_1^{(k)}), \dots, u(t_k, s_{n_k}^{(k)}))^T$. Finally, we denote by $\mathbf{U} \in \mathbb{R}^{N_o}$ the vector containing all the $N_o = n_0 + \dots + n_K$ observations, i.e. $\mathbf{U} = (\mathbf{u}^{(0)})^T, \dots, (\mathbf{u}^{(K)})^T$. Let $\boldsymbol{\varepsilon}$ be a vector of N_o independent standard Gaussian variables and $\sigma > 0$. Then the statistical model for the observations takes the form

$$\mathbf{U} = \boldsymbol{\eta}\mathbf{b} + \mathbf{A}^T \mathbf{Z} + \sigma\boldsymbol{\varepsilon}, \quad (28)$$

where $\mathbf{b} \in \mathbb{R}^q$ is the vector of q fixed effects, $\boldsymbol{\eta} \in \mathbb{R}^{N_o \times q}$ is a matrix of covariates, \mathbf{Z} is the vector containing the weights defining the solution of SPDE (8) as in Proposition 4.6, and $\mathbf{A} \in \mathbb{R}^{(K+1)N \times N_o}$ is the block diagonal matrix whose k -th block ($k \in \{0, \dots, K\}$) $\mathbf{A}^{(k)} \in \mathbb{R}^{N \times n_k}$ is defined by

$$[\mathbf{A}^{(k)}]_{ij} = \psi_i(s_j^{(k)}), \quad i \in \{1, \dots, N\}, j \in \{1, \dots, n_k\}.$$

5.1 Prediction by kriging

Predictions of the spatio-temporal field are tackled using conditional expectations (and variances), following the approach outlined by Clarotto et al. [9], Pereira et al. [25]. The next proposition provides explicit formulas for the computation of the conditional expectation and variance of the field \mathbf{Z} given the observations \mathbf{U} .

Proposition 5.1. *The conditional expectation (also called kriging predictor) $\mathbb{E}[\mathbf{Z}|\mathbf{U}]$ of \mathbf{Z} given \mathbf{U} is given*

$$\mathbb{E}[\mathbf{Z}|\mathbf{U}] = \boldsymbol{\mu}_{\mathbf{Z}} + (\mathbf{Q}_{\mathbf{Z}} + \sigma^{-2} \mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A}(\mathbf{U} - \mathbf{A}^T \boldsymbol{\mu}_{\mathbf{Z}} - \boldsymbol{\eta}\mathbf{b}),$$

where the expressions of $\boldsymbol{\mu}_{\mathbf{Z}}$ and $\mathbf{Q}_{\mathbf{Z}}$ are given in Proposition 4.6. Besides, the conditional variance $\text{Var}[\mathbf{Z}|\mathbf{U}]$ is given by

$$\text{Var}[\mathbf{Z}|\mathbf{U}] = \mathbb{E}[(\mathbf{Z} - \mathbb{E}[\mathbf{Z}|\mathbf{U}])(\mathbf{Z} - \mathbb{E}[\mathbf{Z}|\mathbf{U}])^T | \mathbf{U}] = (\mathbf{Q}_{\mathbf{Z}} + \sigma^{-2} \mathbf{A} \mathbf{A}^T)^{-1}.$$

Proof. This result is a direct consequence of the fact that the vector $(\mathbf{Z}^T, \mathbf{U}^T)^T$ is Gaussian, and a complete proof is given in [25, Proposition 3.1]. \square

For $k \in \{0, \dots, K\}$, the spatial prediction $Z^*(t_k, p)$ of the field $Z(t_k, \cdot)$ at any location $p \in \mathcal{M}_h$ is deduced from the conditional expectation $\mathbb{E}[\mathbf{Z}|\mathbf{U}]$ by leveraging the linearity of the (conditional) expectation, thus giving:

$$Z^*(t_k, p) = \mathbb{E}[Z(t_k, p)|\mathbf{U}] = \begin{pmatrix} \psi_1(p) \\ \vdots \\ \psi_N(p) \end{pmatrix}^T \mathbb{E}[\mathbf{Z}|\mathbf{U}]. \quad (29)$$

Time extrapolation at times t_k , $k > K$, can be tackled by similarly. Indeed, by taking the conditional expectation $\mathbb{E}[\cdot | \mathbf{U}]$ on both sides of the recursion (13) we get

$$\mathbb{E}[\mathbf{z}^{(k+1)}|\mathbf{U}] = (\sqrt{\mathbf{C}})^{-T} \boldsymbol{\Gamma}^{-1} (\sqrt{\mathbf{C}})^T \left(\mathbb{E}[\mathbf{z}^{(k)}|\mathbf{U}] + \frac{\delta t}{c} \mathbf{m} \right), \quad k \geq K, \quad (30)$$

where we recall that $\mathbf{z}^{(k)}$ is the weight vector defining the solution Z at time t_k (cf. Proposition 4.2), and $\mathbb{E}[\mathbf{z}^{(K)}|\mathbf{U}]$ corresponds to the N last rows of $\mathbb{E}[\mathbf{Z}|\mathbf{U}]$. Then spatial predictions at any locations can one again be obtained using (29).

From now on, for any $s > 0$ and $\mathbf{Q} \in \mathbb{R}^{(K+1)N \times (K+1)N}$ a positive definite matrix, let us denote by $\mathcal{K}(\cdot | \mathbf{Q}, s^2)$ the linear map given by

$$\mathcal{K}(\mathbf{v} | \mathbf{Q}, s^2) = s^{-2}(\mathbf{Q} + s^{-2} \mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{v}, \quad \mathbf{v} \in \mathbb{R}^{N_o}. \quad (31)$$

As seen in Proposition 5.1, we have

$$\mathbb{E}[\mathbf{Z}|\mathbf{U}] = \boldsymbol{\mu}_{\mathbf{Z}} + \mathcal{K}(\mathbf{U} - \mathbf{A}^T \boldsymbol{\mu}_{\mathbf{Z}} - \boldsymbol{\eta}\mathbf{b} | \mathbf{Q}_{\mathbf{Z}}, \sigma^2).$$

Therefore, efficient algorithms to evaluate $\mathcal{K}(\cdot | \mathbf{Q}, s^2)$ are fundamental to perform predictions. A first approach to compute (31) consists in factorizing the matrix $\boldsymbol{\Psi} = (\mathbf{Q} + s^{-2} \mathbf{A} \mathbf{A}^T)$ (using for instance a Cholesky decomposition), and then using the factorization to efficiently solve the linear system in (31). However, since the matrix $\boldsymbol{\Psi}$ has size $N(K+1) \times N(K+1)$, building, storing and factorizing can become computationally prohibitive in some applications where either N or K (or both) are large.

An alternative approach to evaluate $\mathcal{K}(\cdot | \mathbf{Q}, s^2)$ consists in solving the linear system in (31) using a matrix-free iterative algorithm [31]. Such algorithms yield an approximate solution of the linear system using through an iterative process

which requires at each iterations products between Ψ and vectors. Such products can in turn be evaluated without having to explicitly build the matrix Ψ , but using instead the diagonal block structure of A . In this setting, only the “spatial” matrices Γ , \tilde{R} and \sqrt{C} and $A^{(k)}$ ($0 \leq k \leq K$), which are sparse and of size at most $N \times N$, are stored.

Note that linear systems of the form of (31) are classically encountered in regularized least-square problems, and are known to be ill-conditioned when s^2 is small. This can make the convergence of classical iterative algorithms (e.g Conjugate gradient) very slow. To circumvent this problem, appropriate preconditioning should be applied to the system. For instance, Clarotto et al. [9] propose to use the Gauss–Siedel preconditioner, which in our case takes the form of lower-triangular block matrix. An alternative method consists in noting that the solution x of the linear system $(Q + s^{-2}AA^T)x = Av$ satisfies is (part of) the solution of the augmented system

$$\begin{pmatrix} Q & A \\ A^T & -s^2 I \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} v \\ 0 \end{pmatrix}, \quad (32)$$

which in turn is equivalent (after applying a block diagonal left and right preconditioners) to the system

$$\begin{pmatrix} I & (L_Q^{-1}A) \\ (L_Q^{-1}A)^T & -s^2 I \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} = \begin{pmatrix} L_Q^{-1}\hat{v} \\ 0 \end{pmatrix}, \quad (33)$$

where $L_Q = D(\sqrt{C}, \sqrt{C})L(\Gamma, I)^T D(f_0^{-1}(\tilde{R}), f_{\delta t}^{-1}(\tilde{R}))$, $\hat{x} = L_Q^T x$, and $\hat{y} = y$. Systems of the form (32)–(33) are known as saddle-point systems and specific algorithms have been devised to tackle them, such as the TriMR algorithm [23]. In the numerical applications presented in this work, the evaluations of the map \mathcal{K} are done using the Julia implementation of the TriMR algorithm in the Krylov.jl to solve the system (33).

As an illustration, we present in Figure 5.1 the results of a kriging prediction based on data sampled from a simulation of the solution to the advection-diffusion model. For the SPDE, we use the same model as the one in Figure 4.1. The observations are assumed to be at the same spatial locations at each time step, and a measurement noise of standard-deviation $\sigma = 0.1$ is considered. When comparing the kriging predictor to the actual simulation, we can notice that the predictor enforces the transport phenomena modeled by the SPDE.

Finally, note that the conditional expectation can also be used to generate conditional simulations at time steps t_k , $0 \leq k \leq K$ of the field Z , by leveraging the fact that the conditional variance does not depend explicitly on the conditioning data U . This approach, presented in more details in [9, Section 3.3], is recalled below:

1. Compute a non-conditional simulation Z_{NC} by running the recursion in Proposition 4.2.
2. Generate new observations by computing $U_{NC} = \eta b + A^T Z_{NC} + \sigma \epsilon_{NC}$.
3. Compute the residuals $r_{NC} = Z_{NC} - \mathbb{E}[Z_{NC}|U_{NC}]$.
4. Return the conditional simulation $Z_C = \mathbb{E}[Z|U] + r_{NC}$.

Conditional simulations at further time steps are then obtained using once again the recursion in Proposition 4.2.

5.2 A few words about inference

Let θ be the vector containing the parameters of (8), and let $\nu = (\theta^T, \mu^{(0)}, m^T, b^T, \sigma^2)^T$ be the vector containing all the parameters of the statistical model. Let us first assume that we observe the field Z exactly, i.e. $U = Z$. In that case, the log-likelihood of the parameters is simply given by

$$\mathcal{L}(\nu) = -\frac{N}{2} \log 2\pi + \frac{1}{2} \log |\mathbf{Q}_Z(\theta)| - \frac{1}{2} (Z - \mu_Z(\theta))^T \mathbf{Q}_Z(\theta) (Z - \mu_Z(\theta)), \quad (34)$$

where the expressions of μ_Z and \mathbf{Q}_Z are given in Proposition 4.6. Evaluating this log-likelihood can be done as follows. The log-determinant \mathbf{Q}_Z can be deduced from the log-determinants of the diagonal block entries of the matrices $L(\Gamma)$, $D(f_0^{-2}(\tilde{R}), f_{\delta t}^{-2}(\tilde{R}))$ and $D(\sqrt{C})$. We obtain in particular the relations

$$\begin{aligned} \log |\mathbf{Q}_Z| &= 2(K+1) \log |\sqrt{C}| + K \log |\Gamma^T \Gamma| + \log |f_0^{-2}(\tilde{R})| + K \log |f_{\delta t}^{-2}(\tilde{R})|, \\ &= 2(K+1) \log |\sqrt{C}| + K \log |\Gamma^T \Gamma| + 2 \log |f_0^{-1}(\tilde{R})| + 2K \log |f_{\delta t}^{-1}(\tilde{R})| \end{aligned}$$

Hence, in order to compute its log-determinant, we do not need to build and store the matrix \mathbf{Q}_Z , but rather only need to store the matrices \tilde{R} , Γ and \sqrt{C} . Indeed, the log-determinant $\log |\Gamma^T \Gamma|$ can be deduced from the LU decomposition of

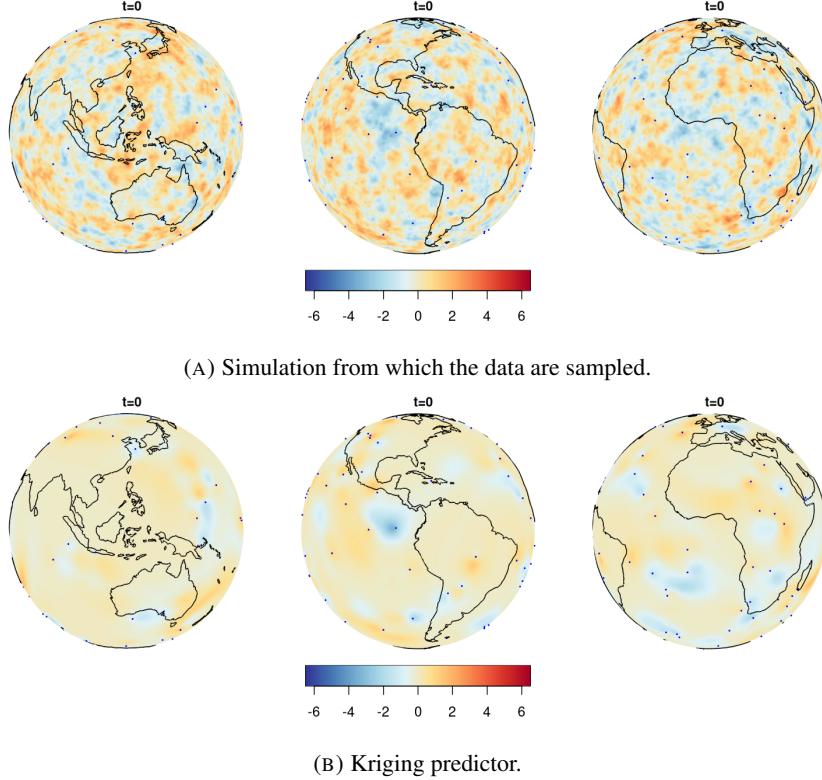


Figure 5.1: Spatio-temporal kriging predictor from data sampled from a simulation of the SPDE model and with a measurement noise of standard-deviation $\sigma = 0.1$. The blue points locate the places where the observations are taken.

Γ , and the log-determinants $\log |f_0^{-1}(\tilde{\mathbf{R}})| = -\log |f_0(\tilde{\mathbf{R}})|$ and $\log |f_{\delta t}^{-1}(\tilde{\mathbf{R}})| = -\log |f_{\delta t}(\tilde{\mathbf{R}})|$ can be approximated using the Hutchinson trick (cf. [24]) or through Cholesky decompositions (when f_0 , f_0^{-1} , $f_{\delta t}$, and/or $f_{\delta t}^{-1}$ are polynomials with low degrees). This ensures that the storage needs for that computation remain the same as the number of time steps considered K increases (which is not the case when using directly the matrix $\mathbf{Q}_Z \in \mathbb{R}^{(K+1)N \times (K+1)N}$). As for the matrix-vector product involving \mathbf{Q}_Z , it can be computed iteratively while requiring only products between the matrices $\tilde{\mathbf{R}}$, Γ and \sqrt{C} and vectors (cf. Algorithm 1).

Remark 5.2. *The linear system $\mathbf{Q}_Z \mathbf{X} = \mathbf{Y}$ can be also solved iteratively, by substitution and leveraging the bi-diagonal structure of the matrix $\mathbf{L}(\Gamma)$. This iterative approach only requires to compute products between functions of the matrix $\tilde{\mathbf{R}}$ and vectors, and to be able to solve linear systems involving Γ (cf. Algorithm 4).*

The next proposition, proven in [9, Section 3.1] gives the log-likelihood in the more general setting of observations given by (28).

Proposition 5.3. *The log-likelihood function of the vector of observations \mathbf{U} is given by*

$$\begin{aligned} \mathcal{L}(\nu) &= -\frac{N_o}{2} \log 2\pi + \frac{1}{2} \log |\mathbf{Q}_U(\nu)| - \frac{\sigma^{-2}}{2} \|\mathbf{U} - \mathbf{A}^T \boldsymbol{\mu}_Z - \boldsymbol{\eta}_b\|_2^2 \\ &\quad + \frac{\sigma^{-2}}{2} (\mathbf{U} - \mathbf{A}^T \boldsymbol{\mu}_Z - \boldsymbol{\eta}_b)^T \mathbf{A}^T \mathcal{K}(\mathbf{U} - \mathbf{A}^T \boldsymbol{\mu}_Z - \boldsymbol{\eta}_b \mid \mathbf{Q}_Z(\theta), \sigma^2), \end{aligned} \tag{35}$$

where $\mathbf{Q}_U(\nu) = (\mathbf{A}^T \mathbf{Q}_Z(\theta)^{-1} \mathbf{A} + \sigma^2 \mathbf{I})^{-1}$, and

$$\log |\mathbf{Q}_U(\nu)| = -N_o \log \sigma^2 + \log |\mathbf{Q}_Z(\theta)| - \log |\mathbf{Q}_Z(\theta)| + \sigma^{-2} \mathbf{A} \mathbf{A}^T.$$

The evaluation of the log-likelihood (35) can be tackled using the approach described in [25, Section 4] and [9, Section 3.1]. Indeed, the map \mathcal{K} can be evaluated using either a preconditioned Conjugate gradient algorithm or a saddle point algorithm (cf. Section 5.1). And the log-determinant terms can be evaluated using either Cholesky decompositions of the matrices (when their size allow it) or using matrix-free approaches based on Hutchinson trace estimators [14]. The maximization of the log-likelihood can then be tackled using gradient-based optimization (based on finite-difference approximations of the gradient) or the Nelder-Mead algorithm (which only requires evaluations of the cost function).

References

- [1] H. Bakka, E. Krainski, D. Bolin, H. Rue, and F. Lindgren. The diffusion-based extension of the Matérn field to space-time. 2020.
- [2] H. Bhatia, G. Norgard, V. Pascucci, and P.-T. Bremer. The helmholtz-hodge decomposition—a survey. *IEEE Transactions on visualization and computer graphics*, 19(8):1386–1404, 2012.
- [3] A. Bonito, D. Guignard, and W. Lei. Numerical approximation of Gaussian random fields on closed surfaces. *arXiv preprint arXiv:2211.13739*, 2022.
- [4] C.-E. Bréhier. A short introduction to stochastic pdes. 2014.
- [5] A. N. Brooks and T. J. Hughes. Streamline upwind/petrov-galerkin formulations for convection dominated flows with particular emphasis on the incompressible navier-stokes equations. *Computer methods in applied mechanics and engineering*, 32(1-3):199–259, 1982.
- [6] M. Cameletti, F. Lindgren, D. Simpson, and H. Rue. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, 97:109–131, 2013.
- [7] W. Chen, M. G. Genton, and Y. Sun. Space-time covariance structures and models. *Annual Review of Statistics and Its Application*, 8:191–215, 2021.
- [8] J.-P. Chiles and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*, volume 497. John Wiley & Sons, 2009.
- [9] L. Clarotto, D. Allard, T. Romary, and N. Desassis. The spde approach for spatio-temporal datasets with advection and diffusion. *arXiv preprint arXiv:2208.14015*, 2022.
- [10] R. Codina. On stabilized finite element methods for linear systems of convection–diffusion-reaction equations. *Computer Methods in Applied Mechanics and Engineering*, 188(1-3):61–82, 2000.
- [11] S. Coveney, C. Corrado, C. H. Roney, R. D. Wilkinson, J. E. Oakley, F. Lindgren, S. E. Williams, M. D. O’Neill, S. A. Niederer, and R. H. Clayton. Probabilistic interpolation of uncertain local activation times on human atrial manifolds. *IEEE Transactions on Biomedical Engineering*, 67(1):99–109, 2019.
- [12] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge university press, 2014.
- [13] T. A. Davis. *Direct Methods for Sparse Linear Systems*, volume 2. SIAM, 2006.
- [14] I. Han, D. Malioutov, and J. Shin. Large-scale log-determinant computation through stochastic chebyshev expansions. In *International Conference on Machine Learning*, pages 908–917. PMLR, 2015.
- [15] L. Herrmann, K. Kirchner, and C. Schwab. Multilevel approximation of Gaussian random fields: fast simulation. *Mathematical Models and Methods in Applied Sciences*, 30(01):181–223, 2020.
- [16] A. Lang and M. Pereira. Galerkin–Chebyshev approximation of Gaussian random fields on compact Riemannian manifolds. *arXiv preprint arXiv:2107.02667*, 2021.
- [17] F. Lindgren, H. Rue, and J. Lindström. 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, 2011.
- [18] F. Lindgren, D. Bolin, and H. Rue. The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running. *Spatial Statistics*, 50:100599, 2022.
- [19] X. Liu, K. Yeo, and S. Lu. Statistical modeling for spatio-temporal data from stochastic convection-diffusion processes. *Journal of the American Statistical Association*, 117(539):1482–1499, 2022.
- [20] A. F. Mejia, Y. Yue, D. Bolin, F. Lindgren, and M. A. Lindquist. A Bayesian general linear modeling approach to cortical surface fMRI data analysis. *Journal of the American Statistical Association*, 115(530):501–520, 2020.
- [21] G. T. Mekuria, J. A. Rao, et al. Adaptive finite element method for steady convection-diffusion equation. *American Journal of Computational Mathematics*, 6(03):275, 2016.
- [22] J. Molina and R. M. Slevinsky. A rapid and well-conditioned algorithm for the helmholtz–hodge decomposition of vector fields on the sphere. *arXiv preprint arXiv:1809.04555*, 2018.
- [23] A. Montoison and D. Orban. Trigc and trimr: Two iterative methods for symmetric quasi-definite systems. *SIAM Journal on Scientific Computing*, 43(4):A2502–A2525, 2021.
- [24] M. Pereira, N. Desassis, and D. Allard. Geostatistics for large datasets on Riemannian manifolds: A matrix-free approach. *Journal of Data Science*, 20(4):512–532, 2022.
- [25] M. Pereira, N. Desassis, and D. Allard. Geostatistics for large datasets on riemannian manifolds: A matrix-free approach. *Journal of Data Science*, 20(4):512–532, 2022. ISSN 1680-743X. doi:10.6339/22-JDS1075.

- [26] K. Poelke and K. Polthier. Boundary-aware hodge decompositions for piecewise constant vector fields. *Computer-Aided Design*, 78:126–136, 2016.
- [27] E. Porcu, A. Alegria, and R. Furrer. Modeling temporally evolving and spatially globally dependent data. *International Statistical Review*, 86(2):344–377, 2018.
- [28] E. Porcu, R. Furrer, and D. Nychka. 30 years of space–time covariance functions. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(2):e1512, 2021.
- [29] N. A. Rayner, R. Auchmann, J. Bessembinder, S. Brönnimann, Y. Brugnara, F. Capponi, L. Carrea, E. M. Dodd, D. Ghent, E. Good, et al. The EUSTACE project: delivering global, daily information on surface air temperature. *Bulletin of the American Meteorological Society*, 101(11):E1924–E1947, 2020.
- [30] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*. CRC press, 2005.
- [31] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.
- [32] S. Särkkä and A. Solin. *Applied stochastic differential equations*, volume 10. Cambridge University Press, 2019.
- [33] F. Sigrist, H. R. Künsch, and W. A. Stahel. Stochastic partial differential equation based modelling of large space-time data sets. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 3–33, 2015.
- [34] H. Wackernagel. *Multivariate Geostatistics: An Introduction with Applications*. Springer Science & Business Media, 2003.
- [35] P. Whittle. On stationary processes in the plane. *Biometrika*, pages 434–449, 1954.

APPENDIX

A Mathematical tools

A.1 Functions of the Laplacian and colored noise

Let $\{\lambda_k\}_{k \in \mathbb{N}}$ denote the set of eigenvalues of the Laplace–Beltrami operator $-\Delta_{\mathcal{M}}$ on (\mathcal{M}, g) , and $\{e_k\}_{k \in \mathbb{N}}$ denote the associated eigenfunctions. In particular, $\{e_k\}_{k \in \mathbb{N}}$ form a basis of the space $L^2(\mathcal{M})$. For $f : \mathbb{R}_+ \rightarrow \mathbb{R}$, we define $\mathcal{D}_f \subset L^2(\mathcal{M})$ as

$$\mathcal{D}_f = \{\phi \in L^2(\mathcal{M}) : \sum_{j \in \mathbb{N}} f(\lambda_j)^2 \langle \phi, e_j \rangle^2 < \infty\}.$$

Note that if f is bounded, then $\mathcal{D}_f = L^2(\mathcal{M})$. Then, the *function of the Laplacian* $f(-\Delta_{\mathcal{M}})$ is the operator $f(-\Delta_{\mathcal{M}}) : \mathcal{D}_f \rightarrow L^2(\mathcal{M})$ defined by

$$f(-\Delta_{\mathcal{M}})\phi = \sum_{j \in \mathbb{N}} f(\lambda_j) \langle \phi, e_j \rangle e_j, \quad \phi \in L^2(\mathcal{M}).$$

Let then $f_S : \mathbb{R}_+ \rightarrow \mathbb{R}$ be a bounded function, and $\{w_j\}_{j \in \mathbb{N}}$ be a sequence of independent standard Gaussian variable. We call *colored noise* the linear functional $f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S$ defined by

$$f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S : \phi \in H \mapsto \langle f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S, \phi \rangle = \sum_{j \in \mathbb{N}} w_j f_S(\lambda_j) \langle \phi, e_j \rangle.$$

Note that for any $\phi \in H$, the series $\langle f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S, \phi \rangle$ converges in quadratic mean since $\mathbb{E}[\langle f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S, \phi \rangle] = 0$, and by independence of the variables w_j ,

$$\mathbb{E}[|\langle f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S, \phi \rangle|^2] = \sum_{j \in \mathbb{N}} |f_S(\lambda_j)|^2 |\langle \phi, e_j \rangle|^2 \leq (\sup_{\mathbb{R}_+} |f_S|)^2 \|\phi\|_S^2 < \infty.$$

Besides, using the same arguments, we have for any $\phi_1, \phi_2 \in H$,

$$\text{Cov}[\langle f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S, \phi_1 \rangle, \langle f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S, \phi_2 \rangle] = \sum_{j \in \mathbb{N}} |f_S(\lambda_j)|^2 \langle \phi_1, e_j \rangle \langle \phi_2, e_j \rangle = \langle f_S(-\Delta_{\mathcal{M}})\phi_1, f_S(-\Delta_{\mathcal{M}})\phi_2 \rangle,$$

Hence, in the case where $f_S(\lambda) = 1$ for any $\lambda \geq 0$, $f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S = \mathcal{W}_S$ corresponds to the definition of the spatial Gaussian white noise on $L^2(\mathcal{M})$. Also, whenever f_S satisfies $f_S(\lambda) = \mathcal{O}_{\lambda \rightarrow \infty}(\lambda^{-\alpha})$ with $\alpha > d/4$, $f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S$ can be identified with a square-integrable H -valued random variable, and decomposed as [16, Proposition 2.7]:

$$f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S = \sum_{j \in \mathbb{N}} w_j f_S(\lambda_j) e_j.$$

For instance, if $f_S(\lambda) = |\kappa^2 + \lambda|^{-\alpha}$, then $\mathcal{Y}_S = f_S(-\Delta_{\mathcal{M}})\mathcal{W}_S$ can be seen as a solution of the Whittle-Matérn SPDE

$$(\kappa^2 - \Delta_{\mathcal{M}})^{\alpha} \mathcal{Y}_S = \mathcal{W}_S \tag{36}$$

and can thus be seen as a Whittle-Matérn random field on \mathcal{M} .

A.2 Galerkin approximation

For $N \in \mathbb{N}$, let $V_N = \text{span}\{\psi_k : 1 \leq k \leq N\}$ where $\psi_1, \dots, \psi_N \in H^1(\mathcal{M})$ are linearly independent functions. The Galerkin approximation of $-\Delta_{\mathcal{M}}$ over V_N is the linear operator $-\Delta_N : V_N \rightarrow V_N$ which maps any $\phi \in V_N$ to the element $-\Delta_N \phi \in V_N$ satisfying

$$\langle -\Delta_N \phi, v \rangle = \langle \nabla \phi, \nabla v \rangle, \quad \text{for any } v \in V_N.$$

Similarly, we can define the Galerkin approximation of the operator $f \in H^1(\mathcal{M}) \mapsto \text{div}(\gamma f)$ as the linear operator $\text{div}_N(\gamma \cdot) : V_N \rightarrow V_N$ which maps any $\phi \in V_N$ to the element $\text{div}_N(\gamma \phi) \in V_N$ satisfying, for any $v \in V_N$,

$$\langle \text{div}_N(\gamma \phi), v \rangle = \langle \text{div}(\gamma \phi), v \rangle, \quad \text{for any } v \in V_N.$$

As defined, $-\Delta_N$ is a symmetric endomorphism, and as such is diagonalizable. Let $\{\lambda_k^{(N)}\}_{1 \leq k \leq N}$ denote its eigenvalues, and let $\{e_k^{(N)}\}_{1 \leq k \leq N}$ be a set of associated eigenfunctions forming an orthonormal basis of V_N .

Let $\mathbf{C}, \mathbf{R} \in \mathbb{R}^{N \times N}$ be the matrices whose entries are respectively given by

$$C_{ij} = \langle \psi_i, \psi_j \rangle, \quad R_{ij} = \langle \nabla \psi_i, \nabla \psi_j \rangle, \quad 1 \leq i, j \leq N. \tag{37}$$

Note that, following [16, Corollary 3.2], $\{\lambda_k^{(N)}\}_{1 \leq k \leq N}$ are also the eigenvalues of the matrix $\tilde{\mathbf{R}}$. Besides, the map

$$E : \mathbf{v} \in \mathbb{R}^N \mapsto \sum_{k=1}^N \left[(\sqrt{\mathbf{C}})^{-T} \mathbf{v} \right]_k \psi_k \in V_N \tag{38}$$

is an isomorphism that maps the eigenvectors of $\tilde{\mathbf{R}}$ to the eigenfunctions of $-\Delta_N$, and an isometry between \mathbb{R}^N (equipped with the Euclidean metric) and V_N (equipped with the metric induced by the inner product $\langle \cdot, \cdot \rangle$).

A.3 Matrix functions

Let $\mathbf{S} \in \mathbb{R}^{N \times N}$ be a real symmetric matrix and let $f : \mathbb{R} \rightarrow \mathbb{R}$. In particular let us denote by $\lambda_1, \dots, \lambda_N$ the eigenvalues of \mathbf{S} and let $\mathbf{V} \in \mathbb{R}^{N \times N}$ be an orthogonal matrix such that

$$\mathbf{S} = \mathbf{V} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{pmatrix} \mathbf{V}^T. \tag{39}$$

Then, the matrix function $f(\mathbf{S}) \in \mathbb{R}^{N \times N}$ is the matrix defined by

$$f(\mathbf{S}) = \mathbf{V} \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_N) \end{pmatrix} \mathbf{V}^T.$$

Note in particular that this definition is independent of the choice of matrix \mathbf{V} in (39), and that when f is a polynomial, $f(\mathbf{S})$ coincides with the usual notion of matrix polynomial.

B Analogy between SPDE formulations

As defined, the forcing term $\mathcal{W}_T \otimes \mathcal{Y}_S$ in SPDE (2) can be identified with a cylindrical Wiener process $\{\widetilde{\mathcal{W}}_t\}_{t \in [0, T]}$ in $L^2(\mathcal{M})$ through

$$\widetilde{\mathcal{W}}_t(\phi_S) = (\mathcal{W}_T \otimes \mathcal{Y}_S)(\mathbb{1}_{[0, t]}, \phi_S), \quad \phi_S \in L^2(\mathcal{M}), \quad t \in [0, T],$$

where $\mathbb{1}_{[0, t]}$ denotes the indicator function of the segment $[0, t]$ [4]. As such, we have (almost-surely) the following decomposition of $\widetilde{\mathcal{W}}_t$

$$\widetilde{\mathcal{W}}_t = \sum_{j \in \mathbb{N}} f_S(\lambda_j) \beta_j(t) e_j, \quad t \in [0, T], \tag{40}$$

where $\{e_j\}_{j \in \mathbb{N}}$ denotes an orthonormal basis of $L^2(\mathcal{M})$ composed eigenfunctions of the Laplace–Beltrami operator $-\Delta_{\mathcal{M}}$, and $\{\lambda_j\}_{j \in \mathbb{N}}$ their associated eigenvalues. This identification allows in turn to write

$$\mathcal{W}_T \otimes \mathcal{Y}_S(\phi_T, \phi_S) = \left\langle \int_0^T \phi_T d\widetilde{\mathcal{W}}_t, \phi_S \right\rangle, \quad (\phi_T, \phi_S) \in L^2([0, T]) \times L^2(\mathcal{M}).$$

where the integral term is given by

$$\int_0^T \phi_T d\widetilde{\mathcal{W}}_t = \sum_{j \in \mathbb{N}} f_S(\lambda_j) \left(\int_0^T \phi_T d\beta_j(t) \right) e_j.$$

Hence, we can interpret the forcing term $\mathcal{W}_T \otimes \mathcal{Y}_S$ as the (time) derivative of the cylindrical Wiener process $\{\widetilde{\mathcal{W}}_t\}_{t \in [0, T]}$. This analogy allows in particular to rewrite SPDE (2) as in (4).

C Algorithms

We expose in this section a few algorithms that are necessary to perform matrix-free predictions.

Algorithm 1 Matrix-vector product by \mathbf{Q}_Z

Depends on: Matrices $(\sqrt{\mathbf{C}}), \tilde{\mathbf{R}}, \Gamma$ in (23).
Input: Vector $\mathbf{X} = ((\mathbf{x}^{(0)})^T, \dots, (\mathbf{x}^{(K)})^T)^T \in \mathbb{R}^{(K+1)N}$.
Output: Vector $\mathbf{Y} = ((\mathbf{y}^{(0)})^T, \dots, (\mathbf{y}^{(K)})^T)^T = \mathbf{Q}_Z \mathbf{X}$.

```

1: for  $k = 0$  to  $K$  do
2:   Initialize  $\mathbf{y}^{(k)} = (\sqrt{\mathbf{C}})^T \mathbf{x}^{(k)}$ 
3:   Initialize  $\mathbf{z}^{(k)} = \mathbf{0}$ 
4: end for
5: Set  $\mathbf{z}^{(0)} \leftarrow \mathbf{y}^{(0)}$ 
6: for  $k = 1$  to  $K$  do
7:   Set  $\mathbf{z}^{(k)} \leftarrow \Gamma \mathbf{y}^{(k)} - \mathbf{y}^{(k-1)}$ 
8: end for
9: Set  $\mathbf{z}^{(0)} \leftarrow f_0^{-2}(\tilde{\mathbf{R}}) \mathbf{z}^{(0)}$ 
10: for  $k = 1$  to  $K$  do
11:   Set  $\mathbf{z}^{(k)} \leftarrow f_{\delta t}^{-2}(\tilde{\mathbf{R}}) \mathbf{z}^{(k)}$ 
12: end for
13: Set  $\mathbf{y}^{(K)} \leftarrow \Gamma^T \mathbf{z}^{(K)}$ 
14: for  $k = K-1$  to  $1$  do
15:   Set  $\mathbf{y}^{(k)} \leftarrow \Gamma^T \mathbf{z}^{(k)} - \mathbf{z}^{(k+1)}$ 
16: end for
17: Set  $\mathbf{y}^{(0)} \leftarrow \mathbf{z}^{(0)} - \mathbf{z}^{(1)}$ 
18: for  $k = 0$  to  $K$  do
19:   Set  $\mathbf{y}^{(k)} \leftarrow (\sqrt{\mathbf{C}}) \mathbf{y}^{(k)}$ 
20: end for
21: return  $\mathbf{Y} = ((\mathbf{y}^{(0)})^T, \dots, (\mathbf{y}^{(K)})^T)^T$ .

```

Algorithm 2 Solve a linear system defined by $\mathbf{L}(\Theta_1, \Theta_2)$

Depends on: Matrices Θ_1, Θ_2 in (21).
Input: Vector $\mathbf{X} = ((\mathbf{x}^{(0)})^T, \dots, (\mathbf{x}^{(K)})^T)^T \in \mathbb{R}^{(K+1)N}$.
Output: Vector $\mathbf{Y} = ((\mathbf{y}^{(0)})^T, \dots, (\mathbf{y}^{(K)})^T)^T = \mathbf{L}(\Theta_1, \Theta_2)^{-1} \mathbf{X}$.

```

1: for  $k = 0$  to  $K$  do
2:   Initialize  $\mathbf{y}^{(k)} = \mathbf{0}$ 
3: end for
4: Set  $\mathbf{y}^{(0)} \leftarrow \mathbf{x}^{(0)}$ 
5: for  $k = 1$  to  $K$  do
6:   Set  $\mathbf{y}^{(k)} \leftarrow \Theta_1^{-1}(\Theta_2 \mathbf{y}^{(k-1)} + \mathbf{x}^{(k)})$ 
7: end for
8: return  $\mathbf{Y} = ((\mathbf{y}^{(0)})^T, \dots, (\mathbf{y}^{(K)})^T)^T$ .

```

Algorithm 3 Solve a linear system defined by $\mathbf{L}(\Theta_1, \Theta_2)^T$

Depends on: Matrices Θ_1, Θ_2 in (21).
Input: Vector $\mathbf{X} = ((\mathbf{x}^{(0)})^T, \dots, (\mathbf{x}^{(K)})^T)^T \in \mathbb{R}^{(K+1)N}$.
Output: Vector $\mathbf{Y} = ((\mathbf{y}^{(0)})^T, \dots, (\mathbf{y}^{(K)})^T)^T = \mathbf{L}(\Theta_1, \Theta_2)^{-T} \mathbf{X}$.

```

1: for  $k = 0$  to  $K$  do
2:   Initialize  $\mathbf{y}^{(k)} = \mathbf{0}$ 
3: end for
4: Set  $\mathbf{y}^{(K)} \leftarrow \Theta_1^{-T} \mathbf{x}^{(K)}$ 
5: for  $k = K - 1$  to  $1$  do
6:   Set  $\mathbf{y}^{(k)} = \Theta_1^{-T} (\Theta_2^T \mathbf{y}^{(k+1)} + \mathbf{x}^{(k)})$ 
7: end for
8: Set  $\mathbf{y}^{(0)} \leftarrow \Theta_2^T \mathbf{y}^{(1)} + \mathbf{x}^{(0)}$ 
9: return  $\mathbf{Y} = ((\mathbf{y}^{(0)})^T, \dots, (\mathbf{y}^{(K)})^T)^T$ .

```

Algorithm 4 Solve a linear system defined by \mathbf{Q}_Z

Depends on: Matrices $(\sqrt{\mathbf{C}}), \tilde{\mathbf{R}}, \Gamma$ in (23).
Input: Vector $\mathbf{X} = ((\mathbf{x}^{(0)})^T, \dots, (\mathbf{x}^{(K)})^T)^T \in \mathbb{R}^{(K+1)N}$.
Output: Vector $\mathbf{Y} = ((\mathbf{y}^{(0)})^T, \dots, (\mathbf{y}^{(K)})^T)^T = \mathbf{Q}_Z^{-1} \mathbf{X}$.

```

1: for  $k = 0$  to  $K$  do
2:   Initialize  $\mathbf{y}^{(k)} = (\sqrt{\mathbf{C}})^{-1} \mathbf{x}^{(k)}$ 
3: end for
4: Set  $\mathbf{Y} \leftarrow \mathbf{L}(\Gamma, \mathbf{I})^{-T} \mathbf{Y}$  using Algorithm 3
5: Set  $\mathbf{y}^{(0)} \leftarrow f_0^2(\tilde{\mathbf{R}}) \mathbf{y}^{(0)}$ 
6: for  $k = 1$  to  $K$  do
7:   Set  $\mathbf{y}^{(k)} \leftarrow f_{\delta t}^2(\tilde{\mathbf{R}}) \mathbf{y}^{(k)}$ 
8: end for
9: Set  $\mathbf{Y} \leftarrow \mathbf{L}(\Gamma, \mathbf{I})^{-1} \mathbf{Y}$  using Algorithm 2
10: for  $k = 0$  to  $K$  do
11:   Set  $\mathbf{y}^{(k)} \leftarrow (\sqrt{\mathbf{C}})^{-T} \mathbf{x}^{(k)}$ 
12: end for
13: return  $\mathbf{Y} = ((\mathbf{y}^{(0)})^T, \dots, (\mathbf{y}^{(K)})^T)^T$ .

```

D Diffusion-only case

In the absence of advection term (i.e. when $\gamma = 0$), an alternative time discretization method for the SPDE (8) can be proposed. Indeed, note that in this case, SPDE (8) takes the form

$$\frac{\partial Z}{\partial t}(t, \cdot) + \frac{1}{c} P(-\Delta_N) Z(t, \cdot) = \frac{\tau}{\sqrt{c}} \mathcal{W}_T \otimes Y_S, \quad t \in [0, T], \quad (41)$$

Let us write for any $j \in \{1, \dots, N\}$ and $t \in [0, T]$, $\xi_j(t) = \langle Z(t, \cdot), e_j^{(N)} \rangle$. By testing (41) against $e_j^{(N)}$, and following the link between the stochastic forcing term $\mathcal{W}_T \otimes Y_S$ and cylindrical Wiener processes, we get that ξ_j satisfies the following SDE:

$$d\xi_j(t) = -\frac{1}{c} P(\lambda_j^{(N)}) \xi_j(t) dt + \frac{\tau}{\sqrt{c}} f_S(\lambda_j^{(N)}) d\beta_j(t), \quad 1 \leq j \leq N, \quad t \in [0, T], \quad (42)$$

where $\{\beta_j\}_{j \in \mathbb{N}}$ denotes a sequence of independent Brownian motions. The SDE (42) models an Ornstein–Uhlenbeck process, whose analytical solution is known [32]. In particular, ξ_j is a Gaussian process and for any $t_0 \geq 0$ and $h > 0$, the conditional distribution $\pi(\xi_j(t_0 + h) | x_j(t_0))$ of $\xi_j(t_0 + h)$ given $\xi_j(t_0)$ is Gaussian and given by

$$\pi(\xi_j(t_0 + h) | \xi_j(t_0)) = \mathcal{N}\left(m_h(\lambda_j^{(N)}) \xi_j(t_0), \sigma_h(\lambda_j^{(N)})^2\right),$$

where m_h and σ_h are the functions defined by

$$m_h(\lambda) = e^{-(h/c)P(\lambda)}, \quad \lambda \geq 0,$$

and

$$\sigma_h(\lambda) = \tau \frac{f_S(\lambda)}{\sqrt{2P(\lambda)}} \sqrt{1 - e^{-2(h/c)P(\lambda)}}, \quad \lambda \geq 0.$$

For $t \in [0, T]$, let $\xi(t) = (\xi_1(t), \dots, \xi_N(t))^T$ be the vector containing the N Ornstein–Uhlenbeck processes. Using the fact that the entries of ξ are independent (since the Brownian motions β_j are independent), we have

$$\pi(\xi(t_0 + h) | \xi(t_0)) = \mathcal{N}\left(m_h(\Lambda^{(N)})\xi(t_0), \sigma_h^2(\Lambda^{(N)})\right).$$

In turn, we introduce for $t \in [0, T]$, the vector $\mathbf{z}(t) = (z_1(t), \dots, z_N(t))^T$ such that

$$Z(t, \cdot) = \sum_{j=1}^N z_j(t) \psi_j \tag{43}$$

and let $\mathbf{x}(t) = (\sqrt{C})^T \mathbf{z}(t)$. Note that using the same arguments as the ones used in proof of Proposition 4.2, we can deduce the relation

$$\mathbf{x}(t) = (\sqrt{C})^T \mathbf{z}(t) = \mathbf{V}\xi(t), \quad t \in [0, T].$$

Hence, we have

$$\pi(\mathbf{x}(t_0 + h) | \mathbf{x}(t_0)) = \mathcal{N}\left(\mathbf{V}m_h(\Lambda^{(N)})\mathbf{V}^T \mathbf{x}(t_0), \mathbf{V}\sigma_h^2(\Lambda^{(N)})\mathbf{V}^T\right).$$

which gives, by definition of the matrix functions,

$$\pi(\mathbf{x}(t_0 + h) | \mathbf{x}(t_0)) = \mathcal{N}\left(m_h(\tilde{\mathbf{R}})\mathbf{x}(t_0), \sigma_h^2(\tilde{\mathbf{R}})\right). \tag{44}$$

This last property can be used to derive the joint distribution of observations of the vector $\mathbf{x}(t)$ (and therefore of the vector $\mathbf{z}(t)$) at irregular time steps, and for instance sample them. Indeed, let $0 \leq t_0 \leq \dots \leq t_K = T$, then using the Markov property of the process $\mathbf{x}(t)$ (inherited from the process $\xi(t)$), we have

$$\pi(\mathbf{x}(t_0), \dots, \mathbf{x}(t_K)) = \pi(\mathbf{x}(t_K) | \mathbf{x}(t_{K-1})) \cdots \pi(\mathbf{x}(t_1) | \mathbf{x}(t_0))\pi(\mathbf{x}(t_0)).$$

This decomposition implies that (jointly) sampling the vectors $(\mathbf{x}(t_0), \dots, \mathbf{x}(t_K))$ can be done by first sampling $\mathbf{x}(t_0)$ and then, for $k \geq 0$, by drawing $\mathbf{x}(t_{k+1})$ from the distribution $\pi(\mathbf{x}(t_{k+1}) | \mathbf{x}(t_k))$ given in (44). hence the following proposition.

Proposition D.1. *Let $K \in \mathbb{N}$, and $0 \leq t_0 \leq \dots \leq t_K = T$. For $k \in \{0, \dots, K\}$, let $\mathbf{z}(t_k)$ be the vector defining the solution of the diffusion (41) at time t_k as in (43) and let $\delta t_k = t_{k+1} - t_k$.*

Let $\mathbf{x}(t_0) = (\sqrt{C})^T \mathbf{z}(t_0)$. Then we have the following recursion for $k \geq 0$,

$$\begin{cases} \mathbf{x}(t_{k+1}) = m_{\delta t_k}(\tilde{\mathbf{R}})\mathbf{x}(t_k) + \sigma_{\delta t_k}(\tilde{\mathbf{R}})\mathbf{w}^{(k+1)}, \\ \mathbf{z}(t_{k+1}) = (\sqrt{C})^{-T}\mathbf{x}(t_{k+1}), \end{cases} \tag{45}$$

where $\{\mathbf{w}^{(k)}\}_{1 \leq k \leq K}$ is a sequence of independent centered Gaussian vectors with covariance matrix \mathbf{I} .

Contrary to the implicit Euler scheme introduced in Proposition 4.2, the present scheme is exact since it results from the exact solution (in distribution) to the Ornstein–Uhlenbeck processes that define the SPDE solution. Once again, the precision matrix of the vectors can be derived, using the same approach as for Proposition 4.6.

Proposition D.2. *Let us assume that the initial condition of SPDE (8) can be expressed as*

$$Z(t_0, \cdot) = f_0(-\Delta_N)Y_N,$$

for some function $f_0 : \mathbb{R}_+ \rightarrow \mathbb{R}$ that is bounded takes positive values. Let then $\mathbf{Z} = ((\mathbf{z}^{(0)})^T, \dots, (\mathbf{z}^{(K)})^T)^T$ be the vector obtained by concatenating the vectors $\mathbf{z}^{(0)}, \dots, \mathbf{z}^{(K)}$ defined in Proposition D.1, and let \mathbf{Q}_Z be the precision matrix of \mathbf{Z} . Then, we have

$$\mathbf{Q}_Z = \mathbf{D}(\sqrt{C}) \mathbf{L}_m^T \mathbf{D}_\sigma \mathbf{L}_m \mathbf{D}((\sqrt{C})^T),$$

where

$$\mathbf{L}_m = \begin{pmatrix} \mathbf{I} & & & \\ -m_{\delta t_0}(\tilde{\mathbf{R}}) & \mathbf{I} & & \\ & \ddots & \ddots & & \\ & & -m_{\delta t_{K-1}}(\tilde{\mathbf{R}}) & \mathbf{I} \end{pmatrix}, \quad \text{and} \quad \mathbf{D}_\sigma = \begin{pmatrix} f_0^{-2}(\tilde{\mathbf{R}}) & & & \\ & \sigma_{\delta t_0}^{-2}(\tilde{\mathbf{R}}) & & \\ & & \ddots & & \\ & & & & \sigma_{\delta t_{K-1}}^{-2}(\tilde{\mathbf{R}}) \end{pmatrix}.$$

Note that once again, products with vectors, solving linear systems and computing the log-determinant can be done without requiring to build the matrix \mathbf{Q}_Z . Indeed, the log-determinants are now given by

$$\begin{cases} \log |\mathbf{Q}_X| = \log |f_0^{-2}(\tilde{\mathbf{R}})| + \sum_{k=0}^{K-1} \log |\sigma_{\delta t_k}^{-2}(\tilde{\mathbf{R}})| = \log |f_0^2(\tilde{\mathbf{R}})| + \sum_{k=0}^{K-1} \log |\sigma_{\delta t_k}^{-2}(\tilde{\mathbf{R}})|, \\ \log |\mathbf{Q}_Z| = 2(K+1) \log |\sqrt{\mathbf{C}}| + \log |\mathbf{Q}_X|, \end{cases}$$

and Algorithms 1 and 4 used earlier to compute matrix-vector products and to solve linear systems can be straightforwardly adapted to the new expression of \mathbf{Q}_Z .

E Stability of the recursion

Proposition E.1. *Assume that the finite element matrices \mathbf{C} , \mathbf{R} and \mathbf{B} are defined according to (10) and let $\mathbf{G} = \frac{1}{2}(\mathbf{B} + \mathbf{B}^T)$. Then we have,*

$$G_{ij} = \left\langle \frac{1}{2} \operatorname{div}(\gamma) \psi_i^\ell, \psi_j^\ell \right\rangle, \quad 1 \leq i, j \leq n.$$

In particular, we have $\kappa^2 + \lambda_{\min}(\tilde{\mathbf{G}}) \geq \inf_{\mathcal{M}} \left(\kappa^2 + \frac{1}{2} \operatorname{div}(\gamma) \right)$ meaning that the recursion in Proposition 4.2 is stable whenever

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

Proof. Let $1 \leq i, j \leq n$. On the one hand, we have using the Leibnitz rule,

$$\langle \psi_j^\ell, \operatorname{div}(\gamma \psi_i^\ell) \rangle = \langle \psi_j^\ell, \operatorname{div}(\gamma) \psi_i^\ell \rangle + \langle \psi_j^\ell, g(\gamma, \nabla \psi_i^\ell) \rangle$$

On the other hand, the integration by parts formula gives

$$\langle \psi_i^\ell, \operatorname{div}(\gamma \psi_j^\ell) \rangle = -\langle \nabla \psi_i^\ell, \gamma \psi_j^\ell \rangle = -\langle g(\nabla \psi_i^\ell, \gamma), \psi_j^\ell \rangle$$

Hence, we can write

$$2G_{ij} = \langle \psi_i^\ell, \operatorname{div}(\gamma \psi_j^\ell) \rangle + \langle \psi_j^\ell, \operatorname{div}(\gamma \psi_i^\ell) \rangle = -\langle g(\nabla \psi_i^\ell, \gamma), \psi_j^\ell \rangle + \langle \psi_j^\ell, \operatorname{div}(\gamma) \psi_i^\ell \rangle + \langle \psi_j^\ell, g(\gamma, \nabla \psi_i^\ell) \rangle$$

which gives

$$G_{ij} = \frac{1}{2} \langle \psi_j^\ell, \operatorname{div}(\gamma) \psi_i^\ell \rangle$$

Let then $\mathbf{x} \in \mathbb{R}^n$ be fixed but arbitrary. Let us take $\tilde{\mathbf{x}} = (\sqrt{\mathbf{C}})^{-T} \mathbf{x}$ and $\varphi = \sum_{i=1}^n \tilde{x}_i \psi_i^\ell \in V_N$. Then,

$$\mathbf{x}^T (\kappa^2 \mathbf{I} + \tilde{\mathbf{G}}) \mathbf{x} = \tilde{\mathbf{x}}^T (\kappa^2 \mathbf{C} + \mathbf{G}) \tilde{\mathbf{x}} = \kappa^2 \tilde{\mathbf{x}}^T \mathbf{C} \tilde{\mathbf{x}} + \tilde{\mathbf{x}}^T \mathbf{G} \tilde{\mathbf{x}}$$

where in particular $\tilde{\mathbf{x}}^T \mathbf{C} \tilde{\mathbf{x}} = \sum_{1 \leq i, j \leq n} \tilde{x}_i \langle \psi_i, \psi_j \rangle \tilde{x}_j = \langle \varphi, \varphi \rangle$ and $\tilde{\mathbf{x}}^T \mathbf{G} \tilde{\mathbf{x}} = \sum_{1 \leq i, j \leq n} \tilde{x}_i \langle \frac{1}{2} \operatorname{div}(\gamma) \psi_i, \psi_j \rangle \tilde{x}_j = \langle \frac{1}{2} \operatorname{div}(\gamma) \varphi, \varphi \rangle$. Hence, we have

$$\mathbf{x}^T (\kappa^2 \mathbf{I} + \tilde{\mathbf{G}}) \mathbf{x} = \left\langle \left(\kappa^2 + \frac{1}{2} \operatorname{div}(\gamma) \right) \varphi, \varphi \right\rangle \geq \inf_{\mathcal{M}} \left(\kappa^2 + \frac{1}{2} \operatorname{div}(\gamma) \right) \langle \varphi, \varphi \rangle = \inf_{\mathcal{M}} \left(\kappa^2 + \frac{1}{2} \operatorname{div}(\gamma) \right) \tilde{\mathbf{x}}^T \mathbf{C} \tilde{\mathbf{x}}$$

Finally, note that by definition of $\tilde{\mathbf{x}}$, we have $\tilde{\mathbf{x}}^T \mathbf{C} \tilde{\mathbf{x}} = \mathbf{x}^T \mathbf{x}$, thus giving that, for any $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^T (\kappa^2 \mathbf{I} + \tilde{\mathbf{G}}) \mathbf{x} \geq \inf_{\mathcal{M}} \left(\kappa^2 + \frac{1}{2} \operatorname{div}(\gamma) \right) \|\mathbf{x}\|^2$$

Consequently, we have by the min-max principle, $\kappa^2 + \lambda_{\min}(\tilde{\mathbf{G}}) = \lambda_{\min}(\kappa^2 \mathbf{I} + \tilde{\mathbf{G}}) \geq \inf_{\mathcal{M}} \left(\kappa^2 + \frac{1}{2} \operatorname{div}(\gamma) \right)$. The stability result is then a direct consequence of Proposition 4.8. \square