Spatio-temporal prediction by stochastic partial differential equationsPrédiction spatio-temporelle par équations aux dérivées partielles stochastiques

Lucia Clarotto

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Mines Paris

Prédiction spatio-temporelle par Équations aux Dérivées Partielles Stochastiques

Spatio-temporal prediction by Stochastic Partial Differential Equations

Soutenue par
Lucia CLAROTTO
Le 12/06/2023

Co-dirigée par
Thomas ROMARY
Denis ALLARD

Co-encadrée par
Nicolas DESASSIS

École doctorale n°621
ISMME

Spécialité
Géostatistique et probabilités appliquées

Composition du jury :

Liliane BEL	Professeure, AgroParisTech	*Présidente du jury*
Emmanuel GOBET	Professeur, CMAP - École polytechnique	*Rapporteur*
Finn LINDGREN	Professeur, University of Edinburgh	*Rapporteur*
Geir-Arne FUGLSTAD	Associate Professor, NTNU	*Examinateur*
Carlo GAETAN	Full Professor, Università Ca' Foscari Venezia	*Examinateur*
Alessandra MENAFOGLIO	Associate Professor, MOX - Politecnico di Milano	*Examinatrice*
Thomas ROMARY	Maître Assistant, Centre de Géosciences Mines Paris	*Directeur de thèse*
Denis ALLARD	Directeur de Recherche, BioSP INRAE	*Co-directeur de thèse*

MINES PARIS | PSL

"*Although this may seem a paradox, all exact science is dominated by the idea of approximation. When a man tells you that he knows the exact truth about anything, you are safe in inferring that he is an inexact man.*"

Bertrand Russell

Acknowledgements

First of all, I would like to thank the committee members that accepted to evaluate my work. Thank you Professor Finn Lindgren, "the chief of SPDEs"; I'm really honored that you reviewed my thesis. Your scientific work has been a source of inspiration for these last years. Thank you Professor Emmanuel Gobet, you reviewed my work with care and interest, and your suggestions were insightful. Thank you Alessandra Menafoglio for being the person I began my journey in the scientific research with, and for being always so kind to me. Thank you Carlo Gaetan for always being interested in my research and for making me feel the Italian spirit among French researchers. Thank you Geir-Arne Fuglstad for motivating some of the research paths of this last year. Thank you Liliane Bel for being a fixed point in the French spatial statistics community.

These last three years have been a great journey. As always in my life, I rather travel in good company than alone. Here is my gratitude to all the people that walked by my side to get me where I am.

Merci Denis, parce que tu as vu une étincelle en moi et que tu ne l'as jamais laissé s'éteindre. Merci pour l'accueil chaleureux chaque fois que je passe par Avignon et pour m'avoir si bien comprise. Tu es un point fixe de ma carrière scientifique et j'espère que tu le seras encore longtemps.

Merci Thomas, pour ta présence constante. J'aime la précision et l'engagement que tu mets dans ton travail. Merci pour avoir créé les conditions de travail idéales et pour avoir été toujours de mon côté.

Merci Nico, pour le côté fou et en même temps génial de ton travail. J'adore l'énergie que tu mets dans les projets qui t'intéressent, et l'éclat qu'on voit dans tes yeux quand tu as une idée novatrice.

Merci les doctorants (et ex) du centre de géosciences, vous faites du centre un lieu vivant et joyeux. Alexis, pour nos visions si différentes du monde, mais en même temps très ouvertes à la découverte. Charlie et Ferdinand, pour faire vivre le bureau. Antoine, Nicolas, Fabien, Léo, Jean et les autres, pour tous les moments agréables ensemble. Je ne suis pas devenue une géologue, mais au moins je sais distinguer le basalte du grès. :D Thank you Martin, to light the team up this last spring.

Merci l'équipe Géostat, vous avez rendu heureux même mes aller-retours à Bleau. Nath, pour le "caféé" de 10h et pour tout ce que tu fais pour notre bien-être. Didier, pour tes blagues et pour ton faux R roulé. Lantu, pour ne jamais nous laisser seuls au labo et pour ton dévouement infini à la recherche. Mike, pour avoir rajeuni l'équipe (ooops) et pour m'aider à résoudre les problèmes mathématiques trop complexes pour moi. Fabien, pour toute l'aide informatique indispensable, et pour être toujours aussi gentil même à la n-ème fois que je te demande comment installer les logiciels. Xavier, Émilie, Jacques, Chantal pour les pauses café, moment essentiel de la vie de l'équipe.

Merci Paris, parce que tu m'as conquise avec les ruelles cachées aux touristes, les coins verts que je découvre encore comme une petite fille, la culture et la beauté qu'imbibent tes veines et la variété des gens qui t'habitent.

Merci la Rooft', l'immeuble le plus cool de Paris, parce que vous m'avez accueillie il y a trois ans et m'avez accompagnée tout au long de ce voyage. Étienne, pour la naturalité des moments qu'on passe ensemble, pour ton côté bordélique et désordonné qui me fait tant marrer, pour les mille inspirations de nos discussions. Val, pour l'air agréable que tu fais vivre dans l'appart, pour ton duo de personnalité que j'ai découvert au fur et à mesure et pour les folles soirées de la Driiing. Jenn, pour nos brunchs, nos séances de ciné et nos peintures, et pour ton sourire toujours sincère. Gwen, pour nos parties aux jeux de société et pour tes enseignements écologiques et engagés. Juliette, pour nos soirées ensemble, qu'elles soient disco ou tisane. Robin, pour Disneyland et pour tes couscous. Et tous les autres… sans oublier Edit, pour les chaussons en cadeau, les douces lettres et les pots des fleurs! :D

Merci au Buccrew, pour avoir accepté une fille dans votre bande de fous, pour perdre aux jeux de société avec moi, pour faire semblant que je joue suffisamment bien au tennis, pour les soirées incroyables dans la forêt.

Merci Béa, parce que tu me traites comme ta fille, tu me nourris avec du beurre et de l'amour, et tu voudrais que je ne parte jamais.

Merci Pierre-Louis, parce que tu m'apprends toujours plein de choses et tu me berces avec la beauté de ton français.

Merci Alexis, pour nos séances de ciné et nos discussions toujours enrichissantes. Je sais avoir trouvé un ami sincère.

Merci les Centraliens, parce que les weekends avec vous sont toujours inoubliables. Merci surtout à Jean, Tristan, Bastien, Apo.

Merci Marie et Victor, je vous ai retrouvés après deux ans de distance et je vous ai vus grandir et changer dans les rues parisiennes.

Grazie Angelino, Franco e Daniele, perché siete la mia famiglia italo-parigina, con tanto di pranzi della nonna e olio da giù in regalo.

Angelino, sei un essere incredibile, impossibile da descrivere a parole, e tutta Parigi e banlieu ormai lo sa.

Franco, adoro i tuoi pensieri costantemente in movimento, la tua sete di conoscenza, il tuo spirito libero ma riflessivo, la tua dolce follia.

Daniele, credo di non aver mai incontrato qualcuno tanto gentile quanto te, non perdere mai questo dono bellissimo.

Thank you Matthew, Jordan, Clément and Rama, for the great time we spent in Saudi. It's always a pleasure to share work and life interests with other people, and it's a blast to be able to meet you all over the world for conferences or easy time.

E poi il pensiero torna sempre a casa, la mia prima casa, che ha fatto nascere le persone più importanti della mia vita, che mi ha reso chi sono e che è sempre un piacere riscoprire a ogni ritorno: l'Italia.

Grazie mamma, perché non sempre mi dici quello che pensi, ma quando lo fai, mi fai capire che mi conosci più di chiunque altro. Sei la persona che più mi sprona a fare sempre le cose al meglio. Sei una forza della natura.

Grazie papà per tutto quello che fai per me. Sei l'esempio di come voglio vivere la mia vita: sempre appieno, dando tutto quello che ho, nella mente e nel cuore.

Grazie Marta di essere la donna più entusiasta che conosca. Ogni giorno mi stupisci. Sei una campionessa, nello sport e nella vita.

Grazie Chiara, che ci sei da sempre e per sempre. E a me basta questo.

Grazie Miri ed Eli, perché fate sempre il tifo per me.

Grazie Bea, perché dentro di te convivono mille anime. Riesci a mettere in discussione la mia vita e, allo stesso tempo, a farmi ridere con i video trash. Sei un esempio per me e sono fiera della donna che stai diventando.

Grazie GLS per non scappare mai. Vi ritrovo sempre più maturi, belli e uniti ogni volta che torno a casa. Siete degli amici speciali.

Grazie Cami, per la FOMO che condividiamo e che ci fa correre in lungo e in largo per passare dei momenti incredibili insieme.

Grazie Leti, sai quanto la tua amicizia mi faccia bene; grazie per le folli serate in Rue de Courlis, per le notti nel mio lettone comodo, per i tuoi consigli sinceri.

Grazie Ele, perché casa tua è il nuovo ritrovo milanese per eccellenza, e quando vengo ritrovo le sensazioni di un tempo, la gioia spensierata degli anni dell'università.

Grazie Bea, perché ti godi la vita e anche io me la godo con te.

Grazie Nene, perché in questi anni ci hai insegnato che si può soffrire e ricominciare a sorridere, e il tuo sorriso è un dono per noi.

Grazie Marti, perché mi insegni che quando sai dove vuoi andare, basta mettersi in marcia.

Grazie Luca, perché la nostra sincerità mi fa del bene, e perché imparo sempre qualcosa di nuovo stando con te.

Grazie Gio, per i momenti stupendi a Grenoble et environs, per il tuo modo di vivere la vita, da cui prendo spunto appena posso.

Grazie Teo, per la tua presenza costante, che sento forte anche a 9000km di distanza. Sto bene a pensare che ti stai godendo "l'isola dei sogni".

Grazie Ruffo, grazie Giodaz, perché quando ci siete è festa per me.

Et pour finir, merci Daniel, pour grandir avec moi et illuminer mon visage avec ton amour. J'ai hâte de découvrir ensemble notre avenir.

Abstract

Prédiction spatio-temporelle par Équations aux Dérivées Partielles Stochastiques

Spatio-temporal prediction by Stochastic Partial Differential Equations

In the task of predicting spatio-temporal fields in environmental science using statistical methods, introducing statistical models inspired by the physics of the underlying phenomena that are numerically efficient is of growing interest. Large space-time datasets call for new numerical methods to efficiently process them. The Stochastic Partial Differential Equation (SPDE) approach has proven to be effective for the estimation and the prediction in a spatial context. We present here the unsteady advection-diffusion SPDE which defines a large class of nonseparable spatio-temporal models. A Gaussian Markov random field approximation of the solution to the SPDE is built by discretizing the temporal derivative with a finite difference method (implicit Euler) and by solving the spatial SPDE with a finite element method (continuous Galerkin) at each time step. The Streamline Diffusion stabilization technique is introduced when the advection term dominates the diffusion. Computationally efficient methods are proposed to estimate the parameters of the SPDE and to predict the spatio-temporal field by kriging, as well as to perform conditional simulations. The approach is applied to solar radiation and wind speed datasets. Its advantages and limitations are discussed, and new perspectives for future work are envisaged, especially involving a nonstationary extension of the approach. As a further contribution of the PhD, the nonseparable generalization of the Gneiting class of multivariate space-time covariance functions is presented. The main potential of the approach is the possibility to obtain entirely nonseparable models in a multivariate setting, and this advantage is shown on a weather trivariate dataset. Finally, a review of some methods for approximate estimation and prediction for spatial and spatio-temporal data is proposed, motivated by the objective of reaching a trade-off between statistical efficiency and computational complexity. These methods proved to be effective for parameter estimation and prediction in the context of the "Spatial Statistics Competition for Large Datasets" organized by the King Abdullah University of Science and Technology (KAUST) in 2021 and 2022. Lastly, possible further research directions are discussed.

Résumé

Prédiction spatio-temporelle par Equations aux Dérivées Partielles Stochastiques

L'introduction de modèles statistiques inspirés de la physique des phénomènes sous-jacents et numériquement efficaces est d'un intérêt croissant pour la prédiction de processus spatio-temporels en sciences environnementales. Les grands jeux de données spatio-temporelles nécessitent de nouvelles méthodes numériques efficaces. L'approche par Equations aux Dérivées Partielles Stochastiques (EDPS) s'est avérée efficace pour l'estimation et la prédiction dans un contexte spatial. Nous présentons ici une EDPS d'advection-diffusion avec une dérivée de premier ordre en temps qui définit une grande classe de modèles spatio-temporels non séparables. On construit une approximation de la solution de l'EDPS par un champ aléatoire Markovien Gaussien en discrétisant la dérivée temporelle par la méthode des différences finies (Euler implicite) et en résolvant l'EDPS spatiale par la méthode des éléments finis (Galerkin continu) à chaque pas de temps. La technique de stabilisation "Streamline Diffusion" est introduite lorsque le terme d'advection domine la diffusion. Des méthodes de calcul efficaces sont proposées pour estimer les paramètres de l'EDPS et pour prédire le champ spatio-temporel par krigeage, ainsi que pour effectuer des simulations conditionnelles. L'approche est appliquée à des jeux de données de rayonnement solaire et de vitesse du vent. Ses avantages et ses limites sont examinées, et de nouvelles perspectives de travail sont envisagées, notamment afin de proposer une extension dans un cadre non stationnaire. On présente également un travail portant sur la généralisation non séparable de la classe de Gneiting des fonctions de covariance spatio-temporelles multivariées. Le principal potentiel de l'approche est la possibilité d'obtenir des modèles entièrement non séparables dans un cadre multivarié, et les avantages sont illustrés sur un ensemble de données météorologiques trivariées. De plus, on propose une analyse de méthodes d'estimation et de prédiction approximées pour les données spatiales et spatio-temporelles, motivée par l'objectif de parvenir à un compromis entre l'efficacité statistique et la complexité computationnelle. Ces méthodes se sont avérées efficaces pour l'estimation des paramètres et la prédiction dans le contexte de la "Compétition de statistiques spatiales pour les grands jeux de données" organisée par la King Abdullah University of Science and Technology (KAUST) en 2021 et 2022. Enfin, d'autres pistes de recherche sont envisagées et examinées.

Contents

List of Figures

List of Tables

List of Abbreviations

PDE	**P**artial **D**ifferential **E**quation
ODE	**O**rdinary **D**ifferential **E**quation
SDE	**S**tochastic **D**ifferential **E**quation
SPDE	**S**tochastic **P**artial **D**ifferential **E**quation
GRF	**G**aussian **R**andom **F**ields
MRF	**M**arkov **R**andom **F**ields
GMRF	**G**aussian **M**arkov **R**andom **F**ields
PIML	**P**hysics-**I**nformed **M**achine **L**earning
INLA	**I**ntegrated **N**ested **L**aplace **A**pproximation
GeRF	**Ge**neralized **G**aussian **R**andom **F**ields
SCEU	**S**ufficient **C**ondition for **E**xistence and **U**niqueness
DEMF	**D**iffusion-based **E**xtension of the **M**atérn **F**ield
FDM	**F**inite **D**ifference **M**ethod
FEM	**F**inite **E**lement **M**ethod
FVM	**F**inite **V**olume **M**ethod
SD	**S**treamline **D**iffusion
GLS	**G**alerkin **L**east **S**quares
SUPG	**S**treamline **U**pwind **P**etrov **G**alerkin
ML	**M**aximum **L**ikelihood
CG	**C**onjugate **G**radient
GS	**G**auss **S**eidel
MSE	**M**ean **S**quare **E**rror
RMSE	**R**oot **M**ean **S**quare **E**rror
MAE	**M**ean **A**bsolute **E**rror
GHI	**G**lobal **H**orizontal **I**rradiance
RNN	**R**ecurrent **N**eural **N**etwork
ESN	**E**cho **S**tate **N**etwork
EOF	**E**mpirical **O**rthogonal **F**unction
PCA	**P**rincipal **C**omponent **A**nalysis
FRK	**F**ixed **R**ank **K**riging
PIT	**P**roportional **I**n **T**ime
PL	**P**airwise **L**ikelihood
MLOE	**M**ean **L**oss **O**f **E**fficiency
MMOM	**M**ean **M**isspecification **O**f the **M**ean Square Error

List of Symbols

i	imaginary unit	
$\mathbb{R}^d$	d-dimensional Euclidean space	
$\mathbb{S}^d$	d-dimensional sphere	
$\mathbf{s}$	spatial location	$\in \mathbb{R}^d$
$\mathbf{s}_i$	spatial location of observation i	$\in \mathbb{R}^d$
$\mathbf{h} = \mathbf{s} - \mathbf{s}'$	spatial lag between $\mathbf{s}$ and $\mathbf{s}'$	$\in \mathbb{R}^d$
$\|\cdot\|$	Euclidean distance	$\mathbb{R}^d \to \mathbb{R}$
t	temporal location	$\in \mathbb{R}$
t_i	temporal location of observation i	$\in \mathbb{R}$
$u = \|t - t'\|$	temporal distance	$\in \mathbb{R}$
$\mathbb{E}(\cdot)$	expectation	
$\mathrm{Cov}(\cdot, \cdot)$	covariance	
$C_{ST}(\cdot, \cdot)$	stationary spatio-temporal covariance function	$(\mathbb{R}^d \times \mathbb{R}) \times (\mathbb{R}^d \times \mathbb{R}) \to \mathbb{R}$
$C_S(\cdot, \cdot)$	stationary spatial covariance function	$\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$
$C_T(\cdot, \cdot)$	stationary temporal covariance function	$\mathbb{R} \times \mathbb{R} \to \mathbb{R}$
N_S	number of spatial mesh points	$\in \mathbb{R}$
N_T	number of time steps	$\in \mathbb{R}$
n	number of observations	$\in \mathbb{R}$
n_S	number of spatial observations	$\in \mathbb{R}$
$X(\cdot, \cdot)$	spatio-temporal GRF with zero-mean	$\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$
$Y(\cdot, \cdot)$	spatio-temporal stochastic process	$\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$
$\mu(\cdot, \cdot)$	spatio-temporal trend	$\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$
$W(\cdot, \cdot)$	spatio-temporal Gaussian white noise	$\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$
$W_S(\cdot)$	spatial Gaussian white noise	$\mathbb{R}^d \to \mathbb{R}$
$W_T(\cdot)$	temporal Gaussian white noise	$\mathbb{R} \to \mathbb{R}$
$\mathcal{L}$	spatio-temporal differential operator	
$\mathcal{L}_S$	spatial differential operator	
$\mathcal{L}_T$	temporal differential operator	
$\boldsymbol{\xi}$	spatial frequency	$\in \mathbb{R}^d$
ω	temporal frequency	$\in \mathbb{R}$
$\mathcal{F}$	Fourier transform	
$\mathcal{F}_S$	spatial Fourier transform	
$\mathcal{F}_T$	temporal Fourier transform	
$g(\cdot)$	spatial symbol function	$\mathbb{R}^d \to \mathbb{C}$

$S(\cdot, \cdot)$	spatio-temporal spectral density	$\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$
$\mathbf{\Sigma}$	variance-covariance matrix	
$\mathcal{T}$	spatial triangulation	
$\mathcal{T}'$	spatio-temporal triangulation	
$\mathbf{b}$	vector of q fixed effects	$\in \mathbb{R}^q$
$\boldsymbol{\eta}$	matrix of covariates	$\in \mathbb{R}^n \times \mathbb{R}^q$
$\mathbf{A}$	projection matrix between $\mathcal{T}'$ and data	$\in \mathbb{R}^{N_S N_T} \times \mathbb{R}^n$
$\boldsymbol{\varepsilon}$	error term	$\in \mathbb{R}^{N_S N_T}$
σ_0^2	variance of the error term	$\in \mathbb{R}^+$
κ	scale parameter of Matérn covariance	$\in \mathbb{R}$
ν	smoothness parameter of Matérn covariance	$\in \mathbb{R}$
$\mathcal{K}_\nu(\cdot)$	modified 2$^{\text{nd}}$ order Bessel function	$\mathbb{R}^d \to \mathbb{R}$
r	practical range of a Matérn process	$\in \mathbb{R}^+$
$C_\nu^M(\cdot)$	unit variance and scale Matérn covariance function	$\mathbb{R}^d \to \mathbb{R}$
$\boldsymbol{\gamma}$	velocity vector	$\in \mathbb{R}^d$
τ	standard deviation factor	$\in \mathbb{R}^+$
$\boldsymbol{\theta} = [\kappa, \gamma_x, \gamma_y, c, \tau]^\top$	parameters of the SPDE	
$\boldsymbol{\psi} = [\boldsymbol{\theta}^\top, \mathbf{b}^\top, \sigma_0]^\top$	parameters of the statistical model	

To teacher Katia, Prof. Molteni and Prof. Scappatura,
who made me fall in love with maths

Chapter 1

Introduction and background

Résumé

Dans ce chapitre, nous introduisons le cadre de la thèse, en détaillant les sujets fondamentaux : la statistique spatio-temporelle, les notions de nonséparabilité et d'asymétrie, l'introduction de la physique dans les modèles statistiques, le problème de la grande dimension des jeux de données. Nous présentons ensuite la méthode des Équations aux Dérivées Partielles Stochastiques (EDPS) et son intérêt dans la modélisation, la simulation, l'inférence et la prédiction des données spatio-temporelles. Nous précisons les objectifs de ce travail consistant en des approches précises et computationnellement efficaces pour la simulation, l'inférence et la prédiction de données environnementales spatio-temporelles. Nous présentons aussi d'autres contributions de la thèse, notamment la généralisation entièrement non séparable de la classe des fonctions de covariance spatio-temporelles multivariées de Gneiting, et de nouvelles techniques pour l'estimation et la prédiction approchées pour les grands jeux de données spatiales et spatio-temporelles, qui se sont avérées très performantes dans la participation à la "Compétition de statistique spatiale pour les grands jeux de données" proposée par la King Abdullah University of Science and Technology (KAUST) en 2021 et 2022.

1.1 Spatio-temporal statistics

Many areas of environmental science seek to predict a space-time variable of interest from observations at scattered points in the space cross time domain of study, e.g., among other possible applications, wind prediction (Lenzi and Genton, 2020; Huang et al., 2022), precipitation forecasting (Sigrist et al., 2011), urban air quality inference (Paciorek et al., 2009), climate models inference (Genton and Kleiber, 2015; Edwards et al., 2019), meteorological predictions (Bourotte et al., 2016; Allard et al., 2022), data assimilation in oceanography (Bertino et al., 2003). A more detailed list of references is available in Porcu et al. (2021).

On the one hand, when data are dense in time and sparse in space, one can work within the framework of multivariate time series. For example, this is the case of high frequency measurements arising from a small set of sensors. On the other hand, when data are dense in space and sparse in time, the analysis can be done in a

multivariate geostatistical setting, where repetitions in time are treated as different variables. When data are dense both in time and space or when they show complex spatio-temporal correlations, these approaches fail due to the inability to build very large multivariate temporal or spatial processes. A much more flexible and richer approach is to consider spatio-temporal models.

Even if data for environmental processes are often taken at fixed locations and times, the processes themselves are generally defined on continuous sets in space and time. Thus, the models are defined on $\mathbb{R}^d \times \mathbb{R}$, where d is the spatial dimension of the process. When dealing with a spatial domain defined on a sphere, $\mathbb{R}^d$ must be replaced by $\mathbb{S}^d$, the d-dimensional sphere. In this work we will consider only spatio-temporal models whose spatial domain is the d-dimensional Euclidean space $\mathbb{R}^d$.

We denote $Y(\mathbf{s}, t)$ the value of the spatio-temporal random field at location $\mathbf{s} \in \mathbb{R}^d$ and time $t \in \mathbb{R}$. Mathematically, one may consider $Y(\mathbf{s}, t)$ as a random function in $\mathbb{R}^{d+1}$ and apply geostatistics techniques as usual. This approach is mathematically correct, but it misses a key point, which is the fact that time flows only in one direction. Good statistical models should take this asymmetry into account.

We assume that the spatio-temporal field Y can be decomposed as a sum of fixed and random effects

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + X(\mathbf{s}, t), \tag{1.1}$$

where $\mu(\mathbf{s}, t) = \mathbb{E}[Y(\mathbf{s}, t)]$ is the spatio-temporal trend, i.e., the deterministic part. In this work, we will always assume, unless stated explicitly otherwise, that $X(\mathbf{s}, t)$ is a zero-mean second-order spatio-temporal random field, i.e., that its covariance function is

$$C_{ST}(\mathbf{h}, u) = \mathrm{Cov}[X(\mathbf{s}, t), X(\mathbf{s}+\mathbf{h}, t+u)] = \mathbb{E}[X(\mathbf{s}, t)X(\mathbf{s}+\mathbf{h}, t+u)], \tag{1.2}$$

with $(\mathbf{s}, t)$ and $(\mathbf{s}+\mathbf{h}, t+u) \in \mathbb{R}^d \times \mathbb{R}$. This statement assumes that all the non-stationarities of the process are accounted for in the deterministic part, which is a quite difficult hypothesis to verify. Under the assumption of a Gaussian process X, the mean and covariance functions completely characterize its distribution (see e.g. Chilès and Delfiner (1999)).

The spatio-temporal field $Y(\mathbf{s}, t)$ is sampled at locations and times $(\mathbf{s}_i, t_i)$, $i = 1, \ldots, n$ where n is the number of space-time data. Ideally, measurements are available for all times and locations, but very often data are missing for some times at some locations. Predicting the spatio-temporal field $Y(\mathbf{s}, t)$ at an unsampled spatio-temporal coordinate $(s_0, t_0) \in \mathbb{R}^d \times \mathbb{R}$ from available measurements $\{Y(\mathbf{s}_i, t_i)\}_{i=1,\ldots,n}$ requires a model for the trend $\mu(\mathbf{s}, t)$ and a covariance model for the random component $X(\mathbf{s}, t)$. Several possibilities exist for modeling the deterministic part, including parametric or non-parametric, linear or non-linear models, with or without covariates. A major step in fitting a Gaussian model to data is the estimation of the covariance function of $X(\mathbf{s}, t)$. Covariance functions are a measure of the linear dependence between the random variables X at the space–time locations $(\mathbf{s}, t)$ and $(\mathbf{s}+\mathbf{h}, t+u)$.

They allow computing the covariance between any couple of points $X(\mathbf{s}_i, t_i)$ and $X(\mathbf{s}_j, t_j)$, hence they must be positive definite functions. Conversely, any positive definite function is the covariance function of a Gaussian process. This necessary and sufficient condition limits the choice of available models and makes the construction of models with realistic features difficult. The definition of spatio-temporal covariance functions that appropriately capture complex interactions between space and time has become a very active field of research. Important developments in this field are detailed in Section 1.2.

If X is a spatio-temporal real stationary Random Field (RF) with covariance function C_{ST}, then, for every $t \in \mathbb{R}$, the spatial RF $X(\cdot, t)$ is a stationary spatial RF, called the *spatial trace* of X. All spatial traces of X are stationary with same stationary spatial covariance function given by $C_S(\mathbf{h}) = C_{ST}(\mathbf{h}, 0)$ for all $\mathbf{h} \in \mathbb{R}^d$. Analogously, for every $\mathbf{s} \in \mathbb{R}^2$, the temporal RF $X(\mathbf{s}, \cdot)$ is called the *temporal trace* of X, and all the temporal traces of X are stationary temporal RF with same covariance function given by $C_T(u) = C_{ST}(0, u)$ for all $u \in \mathbb{R}$.

Among modern techniques proposing efficient methods for estimation and prediction in a spatio-temporal framework, there is a distinction between two possible ways of constructing and treating spatio-temporal models (Wikle and Hooten, 2010): either one follows the traditional geostatistical paradigm, using joint space-time covariance functions (see for example Cressie and Huang (1999), Gneiting (2002), Stein (2005), as well as the recent reviews Porcu et al. (2021), Chen et al. (2021)), or one uses dynamical models, including functional time series of surfaces or solutions to Stochastic Partial Differential Equations (SPDEs), see for example Wikle and Cressie (1999), Sigrist et al. (2012) and Martínez-Hernández and Genton (2022). In this work, we will propose two approaches to construct spatio-temporal models: one, the unsteady advection-diffusion SPDE, belongs to the class of dynamical models; the other, the nonseparable Gneiting class, belongs to the geostatistical class of models.

1.2 Nonseparability and asymmetry

Since it is often difficult to scrutinize spatial and temporal variations simultaneously, it is tempting to focus on $C_{ST}(\mathbf{0}, \cdot)$, i.e., how the covariances at a single location vary across time, and $C_{ST}(\cdot, 0)$, i.e., how the covariances at a single time vary across space. If these were the only variations that mattered, then separable models would suffice. A space-time covariance function is space-time *separable* if it is the product of a spatial covariance function with a temporal covariance function, i.e. if $C_{ST}(\mathbf{h}, u) = C_{ST}(\mathbf{0}, 0)^{-1} C_S(\mathbf{h}) C_T(u)$. This is a valid covariance function, since when C_S is positive definite on $\mathbb{R}^d$ and C_T is positive definite on $\mathbb{R}$, then the product $C_S C_T$ is positive definite on $\mathbb{R}^d \times \mathbb{R}$. The advantage of separability lies in the fact that it leads to a reduced number of parameters, allows separate estimation of the parameters relating to space and time and provides faster prediction techniques

for a constant spatial design, thanks to fast computations of the inverse and of the determinant of the matrix involved in estimation and kriging.

However, using only products of spatial covariances and temporal covariances to define the spatio-temporal covariances is a severe restriction. Separable space-time covariance functions have often been used to take advantage of their computational convenience, even when they are not realistic in describing the processes due to the absence of space-time interaction. Separability implies that small changes in the locations of observations can lead to large changes in the correlations between linear combinations of observations (Stein, 1999). The source of this behavior is the fact that separable covariance functions can be rougher away from the origin than at the origin (Stein, 1999). For this reason, it is necessary to seek space-time covariance functions that are smooth everywhere except possibly at the origin, i.e., when $\mathbf{h} = \mathbf{0}$ and $u = 0$. Moreover, it can be interesting to find models that allow different degrees of smoothness across space than across time. These models are called nonseparable.

A space-time covariance C_{ST} is said to be positively (resp. negatively) space-time *nonseparable* if

$$s(\boldsymbol{h}, u) = C_{ST}(\boldsymbol{h}, u)C_{ST}(\mathbf{0}, 0) - C_S(\boldsymbol{h})\, C_T(u) \geq 0 \quad (\text{resp. } \leq 0). \tag{1.3}$$

If such inequalities hold for all $(\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$, then C_{ST} is called uniformly positively (resp. negatively) nonseparable (Rodrigues and Diggle, 2010; De Iaco and Posa, 2013). The function $s(\boldsymbol{h}, u)$ will be referred to as *separability function*. Their empirical versions can be used as exploratory tools to characterize the type of nonseparability in a dataset. In most applications, nonseparable models show better predictions than separable models, since they are physically more realistic, albeit computationally more expensive.

Nevertheless, it is difficult to write down directly nonseparable space-time covariance functions. Recent studies have focused on constructing nonseparable models, see Gneiting (2002), De Iaco et al. (2003), Porcu et al. (2006), Salvaña and Genton (2021), as well as Bourotte et al. (2016) and Allard et al. (2022) in a multivariate context. Nonseparable space-time covariance models can be constructed from Fourier transforms of permissible spectral densities (Stein, 2005), mixtures of separable models (De Iaco et al., 2003), and partial differential equations (PDEs) representing physical laws (Carrizo-Vergara et al., 2022; Lindgren et al., 2020). They can be stationary or nonstationary, univariate or multivariate, in the Euclidean space or on the sphere. See Porcu et al. (2021) and Chen et al. (2021) for recent comprehensive reviews.

A notion closely related to separability is that of full symmetry. A spatio-temporal covariance function is *fully symmetric* (Gneiting et al., 2006) if

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

for all $(\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$. A fully symmetric model is unable to distinguish possibly differing effects as time moves forward or backward. Isotropy in space and symmetry

in time imply full symmetry. Separable covariance functions are also fully symmetric, but not vice versa. Product-sum models (De Iaco et al., 2003) and the Gneiting class (Gneiting, 2002) are fully symmetric, nonseparable (uniformly negatively nonseparable and uniformly positively nonseparable, respectively) covariance functions (Rodrigues and Diggle, 2010).

Atmospheric or environmental processes are often under the influence of air or water flows, which are incompatible with full symmetry (Gneiting et al., 2006). The Lagrangian reference framework, here detailed, is useful in this case. Transport effects can be modeled with the help of a stationary spatial Gaussian random field and a d-dimensional velocity vector $\boldsymbol{\gamma}$, possibly random. The resulting space–time covariance, which is stationary, but not fully symmetric, is defined as

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

This idea was applied to precipitation fields by Cox and Isham (1988) and Benoit et al. (2018a), to wave heights by Ailliot et al. (2011) and to wind by Salvaña and Genton (2021). It will also be discussed in Section 2.3.3. Covariance functions that are not fully symmetric can also be constructed based on advection-diffusion equations or SPDEs, as first discussed in Gneiting et al. (2006) and later in Lindgren et al. (2011); Sigrist et al. (2015); Carrizo-Vergara et al. (2022). This construction will be used in Chapter 2.

1.3 Physics-informed statistics

Understanding real-world dynamical phenomena is a challenging task. Across various scientific disciplines, machine learning has been used to analyze big data and dynamical systems, and it has become a powerful tool to make data-driven decisions based on the input data alone. However, real-world data are stochastic, noisy, and incomplete, and contain uncertainty. Statistical methods and machine learning prediction approaches ignore the fundamental laws of physics and often, although they perform well at fitting observations, they fail to make plausible predictions.

In the last years, researchers have been trying to integrate data, statistical models and physics to improve the predictive potential of traditional methods. The idea is to introduce the prior knowledge about the dynamical models into the statistical approach, in order to respect underlying physical principles. The Physics-Informed Machine Learning (PIML) models (Raissi et al., 2019; Karniadakis et al., 2021) often perform well for both interpolation and extrapolation, even for a small amount of noisy, missing or sparse data. The physical prior information can be derived either by Partial Differential Equations (PDEs), Ordinary Differential Equations (ODEs) and Stochastic Differential Equations (SDEs) or by intuitive physics, and it can be incorporated into different parts of machine learning models such as data, model architecture, loss function, optimizer and inference algorithm (Linka et al., 2022).

In this specific work, we will not make use of the most common PIML object, which is the Physics-Informed Neural Network (PINN), that is designed to be trained to satisfy the given data as well as the imposed governing equations. Instead, we will apply the Physics-Informed framework in a more general meaning, by introducing physical information into a statistical model. In this particular case, we will define a spatio-temporal random field $X(\mathbf{s}, t)$ as the solution of a Stochastic Partial Differential Equation (SPDE) and use this interpretation to model the physical phenomena underlying the spatio-temporal process $Y(\mathbf{s}, t)$ under study (see Chapter 2).

1.4 Big n problem

While the theoretical aspects of spatio-temporal geostatistics are well-developed (Cressie and Wikle, 2011), their implementation still faces difficulties. The geostatistical paradigm is computationally expensive for large spatio-temporal datasets, due to the factorization of dense covariance matrices, whose complexity scales with the cube of the number of observations. This well known problem is often referred to as the *"big n problem"* (Banerjee et al., 2014).

By definition, the spatio-temporal covariance matrix contains $(N_S N_T)^2$ covariance values, where N_S and N_T are the number of spatial and temporal locations, respectively. These locations can be the grid points where we want to compute a simulation or the locations of data. This means that $(N_S N_T)^2$ values should be computed and stored, implying heavy computational and storage needs when this number becomes very large. Using separable covariance functions can significantly reduce the dimension of the covariance matrices that need to be inverted, alleviating the computational demand. This is due to the fact that, in this particular case, the large covariance matrix is defined as a Kronecker product of smaller matrices and only those smaller matrices must be inverted.

However, as detailed in Section 1.2, separable models are not always appropriate for environmental data due to their lack of flexibility, and even when they are, the "big n problem" can arise from the spatial covariance matrix only. Numerous solutions have been proposed in the stationary case to tackle this problem (see Sun et al. (2012) for a comprehensive review). Some options are the use of compactly supported covariance functions (Gneiting, 2002; Bevilacqua et al., 2019), the covariance tapering (Furrer et al., 2006; Kaufman et al., 2008), the Vecchia approximations and their generalizations (Vecchia, 1988; Stein et al., 2004; Guinness, 2018; Katzfuss and Guinness, 2021), among others. These methods limit the number of non-zero entries in the covariance matrix.

Another way of dealing with the "big n problem" is to impose that the considered Gaussian Random Field (GRF) is Markovian, i.e., a Gaussian Markov Random Field (GMRF) (Besag, 1974; Rue and Held, 2005). This ensures that the resulting *precision matrix*, the inverse of the covariance matrix, has a limited number of non-zero entries (Rue and Held, 2005), leading to a reformulation of the problems using the

sparse precision matrix instead of the covariance matrix. The computational gains from GMRFs have been exploited to provide fast and accurate Bayesian inference for latent Gaussian field models through the Integrated Nested Laplace Approximation (INLA), proposed in Rue et al. (2009). However, GMRF approximations were initially criticized, because they don't allow a direct modeling of some specific correlation behaviors and because it is not possible to write down the distribution at a selected location, but only determine a conditional distribution given a set of locations. Moreover, the GMRF approximation was restricted to GRF observed over a regular lattice. Those criticisms were overcome by Lindgren et al. (2011) who introduced the GMRF approximation of a GRF with Matérn covariance function as a solution of a Stochastic Partial Differential Equation (SPDE).

1.5 SPDE approach

The use of SPDEs for modeling spatial or spatio-temporal Gaussian processes is an approach that has been used in the statistical community for decades, starting with Whittle (1954), Heine (1955) and Whittle (1963). Later works include Jones and Zhang (1997) and Brown et al. (2000). Since the spectral density can be directly obtained from the SPDE, spectral methods can be considered to analyze data, as in Fuentes (2007) for the spatial case and in Sigrist et al. (2015) and Liu et al. (2020) for the space-time case. Rozanov (1977) proved that the solution of a specific class of linear SPDEs is a Markov Random Field (MRF). This occurs when the field's spectral density is the inverse of a polynomial function of the frequencies. Lindgren et al. (2011) used this property to show how the approximated solution of a certain class of spatial SPDEs, obtained with the Finite Element Method (FEM), can be considered as a GMRF. The SPDE framework developed in Lindgren et al. (2011) was largely adopted in spatial statistics because of two main properties. First, Lindgren et al. (2011) linked a specific linear SPDE to the Matérn covariance function, which is probably the most common covariance function in geostatistics. Second, the FEM that was considered to build the GMRF representation was made available for practitioners in the `R-INLA` and `inlabru` packages (Lindgren and Rue, 2015; Krainski et al., 2018; Bakka et al., 2018). The SPDE's parameters are related to the local properties of the spatial or spatio-temporal model and are translated into the conditional distributions in the GMRF representation. One can interpret the physical parameters of the SPDE, build the prior distributions for the model parameters and obtain the corresponding spatial or spatio-temporal field. A complete review of the developments of the SPDE approach in the last decade, both from the methodological and the applied point of view, is available in Lindgren et al. (2022).

1.6 Thesis statement and main contributions

This PhD dissertation starts from the considerations detailed in the sections above to propose methods for spatio-temporal inference and prediction that are physically justified (see Section 1.3), computationally efficient (see Section 1.4) and, especially, capable to model complex spatio-temporal dependence structures through nonseparable and asymmetric models (see Section 1.2).

The first contribution is the development of an SPDE approach for nonseparable asymmetric spatio-temporal models. A Spatio-Temporal (ST) advection-diffusion SPDE is proposed; its coefficients, which can be also spatially and/or temporally varying, directly influence the spatio-temporal dependencies of the process, by explaining its variability in space and time. Nonseparable and separable spatio-temporal fields can be obtained as sub-models by choosing appropriate parameters. The advection-dominated case is also discussed and dealt with by introducing the "Streamline Diffusion" stabilization technique to reduce the convergence problems related to an advection-dominated flow.

It will be shown that it is possible to build an accurate space-time approximation of the process driven by the advection-diffusion SPDE using a combination of a Finite Element Method in space and implicit Euler scheme in time. This approximation leads to sparse structured linear systems. We will obtain promising results, both in terms of precision and speed, for estimation and prediction of spatio-temporal processes based on the SPDE approach from information at scattered spatio-temporal locations. When the size of the spatio-temporal mesh is moderate, direct matrix implementation is possible through the Cholesky decomposition of sparse matrices. We will show how "matrix-free" methods (conjugate gradient method, preconditioners, etc.) can be implemented in order to obtain scalable computations even for very large meshes. The method will be applied to two different spatio-temporal environmental datasets, about solar radiation and wind speed, to show the potential of the approach. Finally, a possible generalization to nonstationary fields is proposed and some preliminary results are illustrated.

The second contribution is a generalization of the well-known Gneiting class of multivariate space-time covariance functions (Gneiting et al., 2010; Bourotte et al., 2016). The proposal is a very general parametric class of fully nonseparable direct and cross-covariance functions for multivariate random fields, where each component has a spatial covariance function from the Matérn family with its own smoothness and scale parameters and, unlike all currently available models, its own correlation function in time (Allard et al., 2022). The application of the proposed model will be illustrated on a weather trivariate dataset. It will be shown that this new model yields better fitting and better predictive scores compared to a more parsimonious model with common temporal correlation function. In this case, the main potential of the proposed approach is the possibility to obtain entirely nonseparable models in a multivariate setting.

The third contribution is a discussion about methods for approximate estimation and prediction for spatial and spatio-temporal data, motivated by the objective of reaching a trade-off between statistical efficiency and computational complexity. This trade-off is required by the massive spatial and spatio-temporal datasets available nowadays. Those techniques will be discussed in terms of predictive capabilities based on the results of the "Spatial Statistics Competition for Large Datasets" proposed by the King Abdullah University of Science and Technology (KAUST) in 2021 and 2022. We will present the techniques that led the team RESSTE, composed of 4 researchers from Mines Paris and INRAE Avignon (Denis Allard, Thomas Romary, Thomas Opitz and myself), to rank first in some of the sub-competitions in both editions.

1.7 Outline

The dissertation aims at introducing and applying the SPDE approach for nonseparable advection-diffusion spatio-temporal models. In particular, we derive methods for simulation, estimation and prediction of such models. In Chapter 2 we start by setting up the mathematical framework and the main notions necessary to work with spatio-temporal SPDEs; we then present the classes of spatio-temporal SPDE models developed in the literature and the model on which this work is based: the unsteady advection-diffusion SPDE model. Chapter 3 focuses on the algorithms designed to simulate from the models via a numerical approximation of the SPDE, by first reviewing the methods that are currently used and then introducing the approach chosen in this specific work. This approach results in a convenient sparsity pattern of the global spatio-temporal precision matrix, which leads to reduced computational cost for all statistical methods (estimation, prediction, conditional simulations). Chapter 4 is devoted to the non-conditional simulation of stationary advection-diffusion SPDE models. Chapter 5 tackles the estimation and the prediction of a stationary advection-diffusion SPDE model from its partial and noisy observations. We suggest solving this problem using an approach inspired by kriging theory. In both cases, we propose practical algorithms based on the Cholesky decomposition of sparse matrices and on "matrix-free" methods. Chapter 6 presents an application of the approach for the prediction of a spatio-temporal field from its partial and noisy observations, both for solar radiation and wind speed datasets. Finally, Chapter 7 is devoted to the nonstationary generalization of the SPDE approach and to some related results, who are however still preliminary.

Moreover, the dissertation presents other contributions made during the PhD (Chapter 8). First, a different point of view on nonseparable spatio-temporal models is presented, through more classical models and their generalizations. In Section 8.1 we enlarge the Gneiting class of space-time covariance functions by introducing a very general parametric class of fully nonseparable direct and cross-covariance functions

for multivariate random fields. Then, Section 8.2 describes the contribution to the "Spatial Competition for Large Datasets" proposed by KAUST in 2021 and 2022.

Finally, Chapter 9 presents the conclusions and the perspectives of the PhD. Several possible research directions are explored.

Chapter 2

Spatio-temporal SPDE models

Résumé

Dans ce chapitre, nous introduisons un cadre mathématique minimal permettant d'étudier les modèles spatio-temporels qui sont solutions d'Equations aux Dérivées Partielles Stochastiques (EDPS). Nous commençons par introduire les principales classes d'EDPS spatio-temporelles dévéloppées dans la littérature statistique. Nous présentons ensuite la classe d'EDPS d'advection-diffusion en étudiant ses caractéristiques statistiques et en la comparant aux EDPS proposées précedemment.

We consider a spatio-temporal model $Y(\mathbf{s}, t)$ observed through some measurements, with no direct measurements of the underlying stochastic process $X(\mathbf{s}, t)$, as presented in Equation (1.1), and reported here:

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + X(\mathbf{s}, t).$$

$Y(\mathbf{s}, t)$ is the sum of fixed and random effects. The fixed effects' term can include a temporal trend, a spatial trend, and time-varying or space-varying regressions. We usually have little information about the covariance structure of the spatio-temporal random model component $X(\mathbf{s}, t)$, except that it should mimic the dependency structure in models of physical processes. We would like not to impose any hypothesis of separability on X, in order to be able to capture complex spatio-temporal dependencies (see Section 1.2). We use the SPDE framework, presented in Section 1.5 and detailed in the following sections. From a theoretical viewpoint, in contrast to statistically oriented constructions of nonseparable spatio-temporal models such as the product-sum models and the Gneiting class (De Iaco et al., 2003; Gneiting, 2002; Porcu et al., 2006), the SPDE approach proposes a physically grounded construction, for which the parameters carry physical interpretation such as diffusion, reaction and transport. It allows the construction of models $X(\mathbf{s}, t)$ with interesting nonseparability and asymmetry properties, and where nonstationarity can also easily be accounted for.

We first give a definition of *Generalized Random Field* (GeRF) (Itô, 1954; Rozanov et al., 1982), which is the stochastic analogous to the generalization of functions presented in Schwartz's theory of distributions (Schwartz, 1966).

Definition 2.1. *A generalized random field (GeRF) $\mathcal{X}$ on $\mathbb{R}^d$ is a linear and continuous functional that associates to any smooth function with compact support $\phi \in \mathcal{C}_0^\infty(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$ a random variable $\mathcal{X}(\phi) \in \mathbb{R}$. $\mathcal{X}$ is characterized by its probability distribution, which is the set of all joint distributions $F_{\phi_1,\dots,\phi_m}$ defined by*

$$F_{\phi_1,\dots,\phi_m} : (a_1, \dots, a_m) \in \mathbb{R}^m \mapsto \mathbb{P}[\mathcal{X}(\phi_1) \leq a_1, \dots, \mathcal{X}(\phi_m) \leq a_m]$$

for any $m \geq 1$ and $\phi_1, \dots, \phi_m \in \mathcal{C}_0^\infty(\mathbb{R}^d)$.

The previous proposition implies that GeRFs have only meaning when applied to test functions in some particular functional space, and not necessarily when evaluated in points of the space.

We define a *white noise* as the functional W defined over $\mathbb{R}^d$ that satisfies

$$\mathbb{E}[W(\phi)] = 0, \quad \forall \phi \in L^2(\mathbb{R}^d)$$

and

$$\mathrm{Cov}[W(\phi_1), W(\phi_2)] = \int_{\mathbb{R}^d} \phi_1 \phi_2 \, \mathrm{d}\,\mathbf{s}\,, \quad \forall \phi_1, \phi_2 \in L^2(\mathbb{R}^d). \tag{2.1}$$

If, moreover, for any $m \geq 1$ and any linearly independent $\phi_1, \dots, \phi_m \in L^2(\mathbb{R}^d)$, the random vector $[W(\phi_1), \dots, W(\phi_m)]^\top$ is a Gaussian vector, then W is called *Gaussian white noise*.

If $\{Z_i\}_{i \in \mathbb{N}}$ is a sequence of independent, standard Gaussian variables, then the function W defined over $L^2(\mathbb{R}^d)$ by

$$W(\phi) = \sum_{j \in \mathbb{N}} Z_j \int_{\mathbb{R}^d} \phi e_j \, \mathrm{d}\,\mathbf{s}\,, \quad \forall \phi \in L^2(\mathbb{R}^d),$$

where $\{e_j\}_{j \in \mathbb{N}}$ denotes an orthonormal basis of $L^2(\mathbb{R}^d)$, is a Gaussian white noise on $\mathbb{R}^d$.

Let us denote $\boldsymbol{\xi} \in \mathbb{R}^d$ a spatial frequency and $\omega \in \mathbb{R}$ a temporal frequency. The *Fourier transform* of an integrable function $\phi(\mathbf{x})$, with $\mathbf{x} \in \mathbb{R}^d$ and $\phi \in L^2(\mathbb{R}^d)$, is

$$\mathcal{F}(\phi)(\boldsymbol{\xi}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\,\mathbf{x}\boldsymbol{\xi}} \phi(\mathbf{x}) \, \mathrm{d}\,\mathbf{x}\,, \quad \boldsymbol{\xi} \in \mathbb{R}^d\,. \tag{2.2}$$

Its inverse is called *inverse Fourier transform* and is defined as

$$\mathcal{F}^{-1}(\phi)(\mathbf{x}) = \int_{\mathbb{R}^d} e^{i\,\mathbf{x}\boldsymbol{\xi}} \phi(\boldsymbol{\xi}) \, \mathrm{d}\,\boldsymbol{\xi}\,, \quad \mathbf{x} \in \mathbb{R}^d\,.$$

In this case, we necessarily have that $\mathcal{F}^{-1}(\mathcal{F}(\phi)) = \phi$.

An important result for the following sections is that the covariance structure of a stationary second-order random field X can be equivalently characterized either by its covariance function C_X or by its spectral density S_X. This is due to Bochner's

theorem (see e.g. Donoghue (1969, Chapter 37)), that states that any continuous covariance function C_X is the Fourier transform of a positive and finite measure. If we want C_X to be real, the measure should be also even. In our case, the measure is absolutely continuous with respect to the Lebesgue measure. Hence, we can suppose that there exists a function S_X, positive and integrable, called *spectral density*, such that (in the spatial case)

$$C_X(\mathbf{h}) = \mathcal{F}(S(\cdot))(\mathbf{h}) = \int_{\mathbb{R}^d} e^{-i\,\mathbf{h}\,\boldsymbol{\xi}} S(\boldsymbol{\xi})\,\mathrm{d}\,\boldsymbol{\xi}\,,$$

where $\mathcal{F}$ denotes the Fourier transform on $\mathbb{R}^d$ and $\int_{\mathbb{R}^d} S(\boldsymbol{\xi})\,\mathrm{d}\,\boldsymbol{\xi} = C_X(\mathbf{0}) < \infty$. The same result holds also for temporal and spatio-temporal covariances.

In the following sections, after presenting the spatial SPDE framework by Lindgren et al. (2011) (Section 2.1), we will review several classes of separable and nonseparable spatio-temporal SPDE models proposed in the last years (Section 2.2) and we will introduce the model on which this work is based: the unsteady advection-diffusion SPDE model (Section 2.3).

2.1 Spatial SPDE model

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

$$(\kappa^2 - \Delta)^{\alpha/2} X(\mathbf{s}) = \tau W(\mathbf{s}), \tag{2.3}$$

with $\alpha > d/2$ and $\tau > 0$. $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$ is the Laplacian operator defining a diffusion term (see Section 2.3.1 for more physical details), and $W(\mathbf{s})$ is a standard spatial Gaussian white noise. The equality is understood in a weak sense in L^2. The spectral density of W is $S_W(\boldsymbol{\xi}) = (2\pi)^{-d}$.

The spectral density of $X(\mathbf{s})$, solution to (2.3), is

$$S(\boldsymbol{\xi}) = \frac{1}{(2\pi)^d(\kappa^2 + \|\boldsymbol{\xi}\|^2)^\alpha}. \tag{2.4}$$

When α is an integer, this density is the reciprocal of a strictly positive and isotropic polynomial. Rozanov's Theorem (Rozanov et al., 1982, Section 3.2.3) allows us to state that this model is an isotropic stationary Markov Rrandom Field (MRF). In Rozanov's theory, a MRF is a RF such that for every domain of $\mathbb{R}^d$, evaluations of the RF on the interior of the domain are independent of evaluations on the interior of the complement of the domain, conditionally on the behavior of the RF on a neighborhood of the boundary of the domain. An evaluation is the inner product between the field and a test function whose support is included in the interior of the corresponding set.

A generalization to an anisotropic model can be obtained by applying an anisotropy matrix $\mathbf{H}$ to $\boldsymbol{\xi}$ in Equation (2.4) and by replacing the Laplacian operator Δ with its anisotropic form $\nabla \cdot \mathbf{H} \nabla$. When α is not an integer, the SPDE (2.3) is said to be fractional and the field X is no longer a MRF. This limitation on integer α can be avoided by considering the approximation to the best Markov approximation, suggested in the reply to Lindgren et al. (2011) and proposed in Bolin and Kirchner (2017) as a combination of the FEM discretization in space with a rational approximation. Other possible ways to tackle fractional SPDEs are proposed in Xiong et al. (2022); Bolin and Kirchner (2020); Roques et al. (2022), but we will not treat this subject further in this work.

The covariance function of the Gaussian field solution to Equation (2.3) is the well known Matérn covariance function (Matern, 1986)

$$C(\mathbf{h}) = \sigma^2 C_\nu^{\mathcal{M}}(\kappa\|\mathbf{h}\|) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa\|\mathbf{h}\|)^\nu \mathcal{K}_\nu (\kappa\|\mathbf{h}\|), \tag{2.5}$$

with smoothness parameter $\nu = \alpha - d/2 > 0$, scale parameter κ and variance $\sigma^2 = \tau^2(4\pi)^{-d/2}\Gamma(\nu)\Gamma(\nu + d/2)^{-1}\kappa^{-2\nu}$. $\mathcal{K}_\nu$ is the modified 2$^{\text{nd}}$ order Bessel function and $\mathbf{h} = \mathbf{s} - \mathbf{s}'$ is the spatial lag between two locations $\mathbf{s}$ and $\mathbf{s}'$ in $\mathbb{R}^d$. In particular, when $\nu = 1/2$, we get the exponential covariance function and when $\nu \to +\infty$, after proper renormalization, (2.5) tends to the Gaussian covariance function.

The spatial SPDE framework has been used for estimation and prediction of spatial processes in several scientific domains. We refer to Lindgren et al. (2022) for a complete review of applications of the spatial SPDE approach in the last years.

2.2 Spatio-temporal SPDE models: state of the art

To generalize the SPDE approach to linear spatio-temporal processes $X(\mathbf{s}, t)$, we consider the framework proposed in Carrizo-Vergara et al. (2022).

Carrizo-Vergara et al. (2022) provided sufficient conditions to the existence and uniqueness of stationary solutions to the spatio-temporal SPDE

$$\left[\frac{\partial^\beta}{\partial t^\beta} + \mathcal{L}_g\right] X(\mathbf{s}, t) = Z(\mathbf{s}, t), \tag{2.6}$$

with $\beta > 0$. $\mathcal{L} = \left[\frac{\partial^\beta}{\partial t^\beta} + \mathcal{L}_g\right]$ is a spatio-temporal operator. The spatial operator $\mathcal{L}_g$ is defined using the spatial Fourier transform on $\mathbb{R}^d$, denoted $\mathcal{F}_S$,

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

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

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

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

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

These operators are called *pseudo-differential* operators (Lang and Potthoff, 2011).

In contrast to the spatial SPDE model detailed in Section 2.1, we remark the presence of a temporal differential operator and of a spatio-temporal stochastic forcing term $Z(\mathbf{s}, t)$. Unless otherwise stated, we will force the stochastic source term to be a spatio-temporal white noise with unit variance $Z(\mathbf{s}, t) = W(\mathbf{s}, t) = W_S(\mathbf{s}) \otimes W_T(t)$, but it was proved that solutions of SPDEs with white noise as source term are the starting point of more general solutions, when the source term can be any stationary GeRF (Carrizo-Vergara et al., 2022). Carrizo-Vergara et al. (2022) established that, under mild conditions, the covariance of the stationary solution to a given SPDE with a general random source term with covariance C_Z is the convolution between the covariance of the same SPDE with white noise source term and C_Z. This result is a powerful tool for easily characterizing solutions of very general SPDEs. The space-time white noise $W(\mathbf{s}, t)$ is characterized by its spectral density $S_W(\boldsymbol{\xi}, \omega) = (2\pi)^{-(d+1)}$.

The spatio-temporal symbol function of the operator involved in (2.6) is

$$(\boldsymbol{\xi}, \omega) \mapsto (i\omega)^\beta + g(\boldsymbol{\xi}) = |\omega|^\beta \cos\left(\frac{\beta\pi}{2}\right) + g_R(\boldsymbol{\xi}) + i\left(\mathrm{sgn}(\omega)|\omega|^\beta \sin\left(\frac{\beta\pi}{2}\right) + g_I(\boldsymbol{\xi})\right) \tag{2.7}$$

where g_R and g_I are the real and imaginary part of the spatial symbol function g.

If $|g_R|$ is inferiorly bounded by the inverse of a strictly positive polynomial and $g_R \cos\left(\frac{\beta\pi}{2}\right) \geq 0$, Theorem 1 and Proposition 3 in Carrizo-Vergara et al. (2022) state that (2.6) admits a unique stationary solution for every arbitrary g_I function. For the rest of the Chapter, we will call this property the Sufficient Condition for Existence and Uniqueness (SCEU).

Remark 2.2.1. When the closed set $g^{-1}(\{0\}) = \{\boldsymbol{\xi} \in \mathbb{R}^d \,|g(\boldsymbol{\xi}) = 0\}$ is non-empty, the non-uniqueness is due to the existence of stationary solutions to the homogeneous problem $\mathcal{L}_g X(\mathbf{s}, t) = 0$.

In the following sections, we will review several classes of spatio-temporal SPDE models in the framework detailed above: first we introduce separable models, then we present nonseparable models in the chronological order in which they appeared in the literature, by showing their similarities and differences. We consider the real positive parameters τ, γ_s, γ_t, s, κ, α, β, α_1, α_2, α_S and δ. τ is a standard deviation parameter, γ_t is a time correlation parameter, γ_s is a spatial range parameter, s and κ are inverse range parameters, α, β, α_1, α_2, α_S and δ are smoothness parameters. We only consider stationary fields, where the parameters of the SPDE are constant values. Nevertheless, we note that nonstationarity can be obtained naturally by making the operator $\mathcal{L}$ spatially and/or temporally varying $\mathcal{L} = \mathcal{L}(\mathbf{s}, t)$, as discussed by Lindgren et al. (2011) and in Chapter 7.

2.2.1 Separable spatio-temporal models

The easiest way to extend the spatial SPDE model of Lindgren et al. (2011) to space-time is to use the spatial SPDE as the spatial component in a spatio-temporal separable model. Jones and Zhang (1997) discussed how separable covariance functions can be understood through differential operators, written as $\mathcal{L} = \mathcal{L}_S\mathcal{L}_T$, where $\mathcal{L}_S$ is a purely spatial operator and $\mathcal{L}_T$ is a purely temporal operator.

A separable space-time model can be defined as a SPDE model for the spatial trace and an autoregressive model of order 1, i.e., AR(1), for the time dimension. The space-time precision matrix of the separable model is defined by the Kronecker product between the precision matrices of the spatial and temporal random effects (Cameletti et al., 2011).

The same model is obtained in Krainski (2018) as a one dimensional SPDE describing the temporal evolution and another SPDE modeling the driving noise as spatially correlated. This model writes

$$\begin{cases} \left(\frac{\partial}{\partial t} + a\right) X(\mathbf{s}, t) = Z(\mathbf{s}, t) \\ (\kappa^2 - \Delta)^{\alpha/2} Z(\mathbf{s}, t) = \tau W(\mathbf{s}, t) \end{cases}$$

However, as discussed in Section 1.2, separability is a strict hypothesis, rarely satisfied in the real world. For this reason, researchers have been proposing spatio-temporal nonseparable models as solutions of SPDEs.

2.2.2 Jones and Zhang (Jones and Zhang, 1997)

Jones and Zhang (1997) were the first in the spatio-temporal statistics community to define a nonseparable spatio-temporal random field as solution to the SPDE

$$\left[\gamma_t \frac{\partial}{\partial t} + (\kappa^2 - \Delta)^{\alpha/2}\right] X(\mathbf{s}, t) = \tau W(\mathbf{s}, t). \tag{2.8}$$

The symbol function of the spatio-temporal operator is

$$(\boldsymbol{\xi}, \omega) \to [(\gamma_t i\omega) + (\kappa^2 + \|\boldsymbol{\xi}\|^2)^{\alpha/2}] \tag{2.9}$$

and the spectral density of the stationary solution X is

$$S(\boldsymbol{\xi}, \omega) = \frac{\tau^2}{(2\pi)^{d+1}[\gamma_t^2\omega^2 + (\kappa^2 + \|\boldsymbol{\xi}\|^2)^{\alpha}]}. \tag{2.10}$$

Equation (2.8) verifies the SCEU and it can be proved that, when $\alpha > d$, the spatial trace has a Matérn covariance with smoothness equal to $\nu = (\alpha - d)/2$ and scale parameter κ.

2.2.3 Stein (Stein, 2005)

Stein (2005) defined a nonseparable spatio-temporal model with spectral density

$$S(\boldsymbol{\xi}, \omega) = \frac{1}{(2\pi)^{d+1}[b(s^2 + |\omega|^2)^{\alpha_1} + a(\kappa^2 + \|\boldsymbol{\xi}\|^2)^{\alpha_2}]^{\delta}}, \tag{2.11}$$

where $a > 0$ and $b > 0$ are scaling parameters, s and κ are scale parameters and α_1, α_2 and δ are smoothness parameters.

Starting from this density, Krainski (2018), Carrizo-Vergara et al. (2022) and Lindgren et al. (2020) derived the symbol function

$$(\boldsymbol{\xi}, \omega) \to [b(s^2 + \omega^2)^{\beta} + a(\kappa^2 + \|\boldsymbol{\xi}\|^2)^{\alpha}]^{\delta/2} \tag{2.12}$$

and defined the SPDE whose solution is the spatio-temporal field with spectral density equal to (2.11)

$$\left[b\left(s^2 - \frac{\partial^2}{\partial t^2}\right)^{\alpha_1} + a(\kappa^2 - \Delta)^{\alpha_2}\right]^{\delta/2} X(\mathbf{s}, t) = W(\mathbf{s}, t). \tag{2.13}$$

The model verifies the SCEU. Moreover, to obtain an integrable marginal spatial covariance, it is needed $s^2 + \kappa^2 > 0$ and a minimum smoothing condition stated by $d/\alpha_2 + 1/\alpha_1 < 2\delta$. In this case, the marginal spatial covariance is a Matérn with scale κ and smoothness $\nu = \alpha_2(\delta - 1/(2\alpha_1)) - 1$ and the marginal temporal covariance ($\kappa = 0$) is a Matérn with scale s and smoothness $\nu_T = \alpha_1(\delta - 1/(2\alpha_2)) - 1/2$.

The interesting property of the Stein model is that, without being a separable model, the temporal and spatial smoothness of the paths of the random function can be controlled separately thanks to the parameters α_1 and α_2. There are only a few special cases of spectral densities of the form (2.11) for which explicit expressions for the Fourier transform are available.

Having α_1 and α_2 integers and δ half-integer gives the reciprocal of Equation (2.11) a polynomial in $\boldsymbol{\xi}$ and ω and thus, the associated process is a MRF. Hence, there is an explicit GMRF representations associated, as in Lindgren et al. (2011).

2.2.4 Stochastic Heat Equation

The *Stochastic Heat Equation* was proposed by both Krainski (2018) and Carrizo-Vergara et al. (2022) and can be deduced from the Jones and Zhang model (2.8) with the absence of the damping parameter ($\kappa = 0$) and a spatial smoothness parameter equal to $\alpha = 2$. It comes from the physical equation of heat, well known in the PDE literature (see e.g. Quarteroni (2008)). It writes

$$\left[\gamma_t \frac{\partial}{\partial t} - \Delta\right] X(\mathbf{s}, t) = \tau W(\mathbf{s}, t). \tag{2.14}$$

The symbol function is

$$(\boldsymbol{\xi}, \omega) \to [\gamma_t i\omega + \|\boldsymbol{\xi}\|^2].$$

One can only modulate how much comes from time and space by the parameter γ_t. Hence, the stochastic heat equation defines an *intrinsic random field*. Equation (2.14) does not respect the uniqueness condition, since $g(\boldsymbol{\xi}) = \|\boldsymbol{\xi}\|^2$ has a zero at the origin (see Remark 2.2.1). However, there exist solutions to the Stochastic Heat Equation for spatial dimensions $d \geq 3$, which can only be conceived as GeRFs and never as random functions continuous in mean-square (Carrizo-Vergara et al., 2022).

2.2.5 Damped Heat Equation (Krainski, 2018)

Krainski (2018) proposed the *Damped Heat Equation*, which can be obtained as a special case of the Jones and Zhang model (2.8) with $\alpha = 2$ or as a generalization of the Stochastic Heat Equation (2.14) with the adding of the damping parameter κ^2. It hence writes

$$\left[\gamma_t \frac{\partial}{\partial t} + (\kappa^2 - \Delta)\right] X(\mathbf{s}, t) = \tau W(\mathbf{s}, t). \tag{2.15}$$

Its spatio-temporal symbol function is

$$(\boldsymbol{\xi}, \omega) \to [(\gamma_t i\omega) + (\kappa^2 + \|\boldsymbol{\xi}\|^2)] \tag{2.16}$$

and the spectral density of X is

$$S(\boldsymbol{\xi}, \omega) = \frac{\tau^2}{(2\pi)^{d+1}[\gamma_t^2\omega^2 + (\kappa^2 + \|\boldsymbol{\xi}\|^2)^2]}. \tag{2.17}$$

Equation (2.15) verifies the SCEU.

2.2.6 Evolution Equation (Carrizo-Vergara et al., 2022)

In Carrizo-Vergara et al. (2022) new spatio-temporal models were obtained from known PDEs describing physical processes, such as diffusion, advection, and oscillations with stochastic forcing terms. The SPDE Equation (2.6) was called *Evolution Equation*. When it respects the SCEU, the spectral density of its stationary solution is given by

$$S(\boldsymbol{\xi}, \omega) = \frac{1}{(2\pi)^{d+1}\left[|\omega|^{2\beta} + 2|\omega|^{\beta}\left(g_R(\boldsymbol{\xi})\cos\left(\frac{\beta\pi}{2}\right) + \text{sgn}(\omega)g_I(\boldsymbol{\xi})\sin\left(\frac{\beta\pi}{2}\right)\right) + |g(\boldsymbol{\xi})|^2\right]}, \tag{2.18}$$

It gives a separable model if $g_I = 0$ and g_R is a constant function. The presence of the function sgn in Equation (2.18) implies that in some cases the spectral density depends on $|\omega|$ with a negative sign. This means that the symmetry of the model can be controlled. A symmetric model is obtained when β is an even integer or when the function g_I is null. An asymmetric model is obtained otherwise. The marginal

temporal smoothness depends on the parameter β. Thus, this model allows a practical control of the separability, symmetry and smoothness conditions.

Two specific cases of Evolution Equations are reported here, for $\beta = 1$ and $\beta = 2$. When $\beta = 1$, the *first order evolution model* has a spectral density equal to

$$S(\boldsymbol{\xi}, \omega) = \frac{1}{(2\pi)^{d+1}[(\omega + g_I(\boldsymbol{\xi}))^2 + g_R^2(\boldsymbol{\xi})]}, \tag{2.19}$$

while, when $\beta = 2$, the *second order evolution model* has a spectral density equal to

$$S(\boldsymbol{\xi}, \omega) = \frac{1}{(2\pi)^{d+1}[(\omega^2 - g_R(\boldsymbol{\xi}))^2 + g_I^2(\boldsymbol{\xi})]}. \tag{2.20}$$

2.2.7 Evolving Matérn (Carrizo-Vergara et al., 2022)

An *Evolving Matérn* model is a stationary solution to the evolution equation (2.6) such that its spatial trace follows a Matérn model. Evolving Matérn models can be obtained by controlling the spatial symbol function g, but also the structure of the stochastic forcing term Z. We assume that the SCEU is verified, i.e., the parameters are chosen such that g_R satisfies the SCEU and $g_R \cos\left(\frac{\beta\pi}{2}\right) \geq 0$.

When the stochastic forcing term is a spatio-temporal white noise, i.e., $Z(\mathbf{s}, t) = W(\mathbf{s}, t)$, the Evolving Matérn SPDE writes

$$\left[\frac{\partial^\beta}{\partial t^\beta} + s_\beta a(\kappa^2 - \Delta)^{\alpha/2}\right] X(\mathbf{s}, t) = W(\mathbf{s}, t), \tag{2.21}$$

where s_β is a parameter that takes the value 1 or -1 depending conveniently on β in order to obtain the conditions of the SCEU. Its spatio-temporal symbol function is

$$(\boldsymbol{\xi}, \omega) \to (i\omega)^\beta + g(\boldsymbol{\xi}) = (i\omega)^\beta + s_\beta a(\kappa^2 + \|\boldsymbol{\xi}\|^2)^{\alpha/2} \tag{2.22}$$

and its spectral density is

$$S(\boldsymbol{\xi}, \omega) = \frac{1}{(2\pi)^{d+1}\left[|\omega|^{2\beta} + 2|\omega|^\beta a(\kappa^2 + \|\boldsymbol{\xi}\|^2)^{\alpha/2}|\cos\left(\frac{\beta\pi}{2}\right)| + a^2(\kappa^2 + \|\boldsymbol{\xi}\|^2)^\alpha\right]}. \tag{2.23}$$

The solution to (2.21) with $\beta = 1$ is a symmetric nonseparable model which is a mixture of a J-Bessel model in space with an exponential model in time; it corresponds to the model of Jones and Zhang (1997) detailed in Section 2.2.2. If $\beta \in \mathbb{N}$ these models are a particular case of Stein models (Stein, 2005) (see Section 2.2.3).

2.2.8 DEMF (Lindgren et al., 2020)

The *Diffusion-based Extension of the Matérn Field* (DEMF) (Lindgren et al., 2020), which was introduced in a particular case for the first time in Krainski (2018) as

Iterated Heat Equation, is the SPDE

$$\left[\gamma_t \frac{\partial}{\partial t} + (\kappa^2 - \Delta)^{\alpha/2}\right]^{\delta} X(\mathbf{s}, t) = \tau Z(\mathbf{s}, t). \tag{2.24}$$

The symbol function of the left-hand-side operator is

$$(\boldsymbol{\xi}, \omega) \to [(\gamma_t i\omega) + (\kappa^2 + \|\boldsymbol{\xi}\|^2)^{\alpha/2}]^{\delta} \tag{2.25}$$

and the right-hand-side stochastic forcing term is a Gaussian noise, white in time and correlated in space, i.e., $Z(\mathbf{s}, t) = W_T(t) \otimes Z_S(\mathbf{s})$, where $Z_S(\mathbf{s})$ is solution to the spatial SPDE

$$(\kappa^2 - \Delta)^{\alpha_S/2} Z_S(\mathbf{s}) = W_S(\mathbf{s}).$$

The spectral density of X is then

$$S(\boldsymbol{\xi}, \omega) = \frac{\tau^2}{(2\pi)^{d+1}[\gamma_t^2\omega^2 + (\kappa^2 + \|\boldsymbol{\xi}\|^2)^{\alpha}]^{\delta}(\kappa^2 + \|\boldsymbol{\xi}\|^2)^{\alpha_S}}. \tag{2.26}$$

Equation (2.24) verifies the SCEU and defines a family of spatio-temporal processes which, when restricted to space given that the smoothness parameters are chosen appropriately, gives the Matérn field. It allows us to choose how differentiable the stochastic process should be through α, α_S and δ. The type of nonseparability also changes based on the chosen parameters. The DEMF family contains both a subfamily of separable models (when $\alpha = 0$) and a subfamily of the Stein family (Stein, 2005) (when $\alpha_S = 0$), including Jones and Zhang model (Jones and Zhang, 1997) (when $\delta = 1$ and $\alpha_S = 0$).

2.3 Unsteady advection-diffusion SPDE

As detailed in the previous section, Carrizo-Vergara et al. (2022), among others, built models of spatio-temporal random fields that are stationary solutions, when they exist, of a very large class of SPDEs. They constructed very general models, that include and encompass existing ones, for example by proposing asymmetric models. They offered the possibility to build and characterize models far beyond the Matérn family which is currently the covariance model considered within most SPDE implementations (Krainski, 2018; Lindgren et al., 2020).

In this work, we will focus on the unsteady *advection-diffusion* SPDE, which is explained from a physical point of view in the next section and then detailed in a statistical framework in the following sections. *Unsteady* (S)PDEs relate to the rate of changes of spatially distributed physical quantities interacting with one another and predict their evolution over time. Hence, at least one derivative with respect to time is present in the (S)PDE and the solution is a function of time. The *steady* version of the same (S)PDE is a time-independent function and it is obtained by setting the partial derivative with respect to time to constant zero.

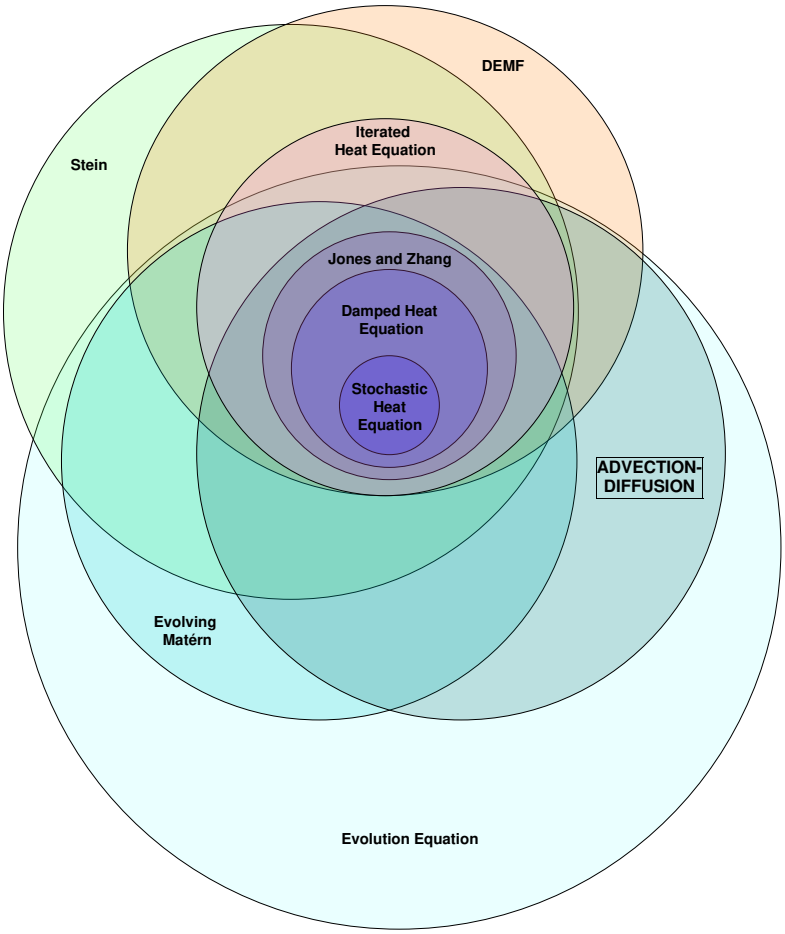


FIGURE 2.1: Venn Diagram of nonseparable SPDE models proposed in the literature.

In Figure 2.1, we illustrate the nonseparable models we've presented in the last sections, along with the advection-diffusion model proposed in this work. We remark how statisticians have been trying to enlarge the family of nonseparable SPDE models to capture higher degrees of complexity.

2.3.1 Diffusion and advection

We here detail the physical characteristics of the advection-diffusion (S)PDE.

Diffusion is a mass transfer phenomenon that causes the distribution of a species to become more uniform in space as time passes. The species can be, for example, a chemical dissolved in a solvent. The driving force for diffusion is the thermal motion of molecules, i.e., the fact that molecules are never at rest at temperatures above absolute zero. When molecules collide with each other frequently, the direction of the motion becomes randomized. Although diffusion occurs because of statistical effects, usually diffusion is modeled through continuous partial differential equations. These

PDEs include Fick's laws (Fick, 1855), the advection-diffusion equation, or more complex methods for concentrated mixtures, like Maxwell-Stefan diffusion (Maxwell, 1867; Stefan, 1871). Fick's laws contain only one parameter, the diffusion coefficient, or *diffusivity*, which is the proportionality constant between the molar flux due to molecular diffusion and the negative value of the gradient in the concentration of the species. Its unit of measurement is m^2/s.

Advection is a mass transfer due to the bulk motion of a fluid. For example, the flow of water transports molecules or ions dissolved in the water, or the flow of air transports molecules present in air. Sometimes the words *convection* or *transport* are used as synonyms of advection. Even though their signification is different in the context of heat transfer (where the term convection usually refers to the combination of heat transfer by conduction and advection, while the term advection refers to heat transfer due to bulk fluid motion), from now on we will use those terms as equivalent.

The velocity of a molecule undergoing mass transfer incorporates both an advective and diffusive component, since all chemical species have a nonzero diffusivity in reality. To summarize, advection is the mass transport due to the average velocity of all molecules, and diffusion is the mass transport due to the instantaneously varying, randomized velocity of individual molecules, compared to the average velocity of the fluid as a whole.

2.3.2 Unsteady advection-diffusion SPDE model

In this work we study the unsteady advection-diffusion SPDE, where the randomness is given by a stochastic forcing term. This SPDE writes

$$\left[\frac{\partial}{\partial t} + \frac{1}{c}(\kappa^2 - \nabla \cdot \mathbf{H}\,\nabla)^{\alpha} + \frac{1}{c}\boldsymbol{\gamma}\cdot\nabla\right] X(\mathbf{s},t) = \frac{\tau}{\sqrt{c}} Z(\mathbf{s},t), \tag{2.27}$$

where

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

This equation was mentioned in Lindgren et al. (2011), Carrizo-Vergara et al. (2022) and Lindgren et al. (2020), and was analyzed using spectral approaches in Sigrist et al. (2015) and Liu et al. (2020). It is a particular case of the evolution equation defined in Carrizo-Vergara et al. (2022) (see Section 2.2.6), as illustrated in

Figure 2.1. The stochastic forcing term $Z(\mathbf{s}, t)$ is assumed separable with

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

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

$$(\kappa^2 - \nabla \cdot \mathbf{H}\, \nabla)^{\alpha_S/2} Z_S(\mathbf{s}) = W_S(\mathbf{s}), \tag{2.28}$$

where W_S is a Gaussian white noise. Notice that the parameter κ^2 in the forcing term has been set identical to that in the diffusion term in the left-hand-side of (2.27) to ensure that the spatial marginalization of the process is a Matérn field, as detailed below.

When $\alpha > 0$, $X(\mathbf{s}, t)$ is a stationary nonseparable spatio-temporal field with covariance function $C_{ST}(\mathbf{h}, u)$, with $(\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$. The advection-diffusion equation (2.27) is a particular first order evolution model as in Equation (2.6) with $\beta = 1$. Its spatial symbol function

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

verifies the sufficient condition for existence and uniqueness of a stationary solution (SCEU) recalled in Section 2.2. Carrizo-Vergara et al. (2022) showed that the advection term does not affect the spatial trace $X(\cdot, t)$ of the solution (see also Section 2.3.3). For some specific values of the parameters, the spatial trace of the solution to (2.27) is a Matérn field, as detailed in Proposition 2.1. In the following $|\mathbf{H}|$ denotes the determinant of the square matrix $\mathbf{H}$.

Proposition 2.1. *Let $Z(\mathbf{s}, t)$ be a spatio-temporal noise colored in space with $Z_S(\mathbf{s})$ satisfying (2.28), and let $\alpha_{tot} = \alpha + \alpha_S$. If $\alpha_{tot} > d/2$, the spatial trace of the stationary solution $X(\mathbf{s}, t)$ of the SPDE (2.27) is the Gaussian Matérn field with covariance*

$$C_S(\mathbf{h}) = C_{ST}(\mathbf{h}, 0) = \frac{\tau^2 \Gamma(\alpha_{tot} - d/2)}{2\Gamma(\alpha_{tot})(4\pi)^{d/2} \kappa^{2(\alpha_{tot}-d/2)} |\mathbf{H}|^{1/2}} C^{\mathcal{M}}_{\alpha_{tot}-d/2}\left(\kappa \left\| \mathbf{H}^{-1/2}\, \mathbf{h} \right\|\right). \tag{2.29}$$

where $\mathbf{h} = \mathbf{s} - \mathbf{s}'$ is the spatial lag and $C^{\mathcal{M}}_{\alpha_{tot}-d/2}(\cdot)$ is the unit variance and scale Matérn covariance function defined in (2.5) with smoothness parameter equal to $\nu = \alpha_{tot} - d/2$.

Proposition 2.1 is adapted from Proposition 1 in Lindgren et al. (2020). A proof is reported below.

Proof. The covariance function of the spatial trace between $X(\mathbf{s}, t)$ and $X(\mathbf{s}', t)$ for a spatial lag $\mathbf{h} = \mathbf{s} - \mathbf{s}'$ does not depend on the imaginary part of the spatial symbol

function (Carrizo-Vergara et al., 2022), hence it can be written as

$$\begin{aligned} C_S(\mathbf{h}) = C_{ST}(\mathbf{h}, 0) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp\Big(i\, \mathbf{h}^\top \boldsymbol{\xi}\Big) S(\boldsymbol{\xi}, \omega)\, \mathrm{d}\omega\, \mathrm{d}\,\boldsymbol{\xi} \\ &= \int_{\mathbb{R}^d} \exp\Big(i\, \mathbf{h}^\top \boldsymbol{\xi}\Big) \left[\int_{\mathbb{R}} S(\boldsymbol{\xi}, \omega)\, \mathrm{d}\omega\right] \mathrm{d}\,\boldsymbol{\xi} \\ &= \int_{\mathbb{R}^d} \exp\Big(i\, \mathbf{h}^\top \boldsymbol{\xi}\Big) S_S(\boldsymbol{\xi})\, \mathrm{d}\,\boldsymbol{\xi}\,, \end{aligned} \tag{2.30}$$

where $S(\boldsymbol{\xi}, \omega)$ is the spectral density defined as

$$S(\boldsymbol{\xi}, \omega) = \frac{\tau^2}{(2\pi)^{d+1} \left[\omega^2 + c^{-2}(\kappa^2 + \boldsymbol{\xi}^\top \mathbf{H}\, \boldsymbol{\xi})^{2\alpha}\right] c(\kappa^2 + \boldsymbol{\xi}^\top \mathbf{H}\, \boldsymbol{\xi})^{\alpha_S}}.$$

Integrating over ω, we obtain the spatial spectral density

$$\begin{aligned} S_S(\boldsymbol{\xi}) &= \frac{\tau^2}{(2\pi)^d c(\kappa^2 + \boldsymbol{\xi}^\top \mathbf{H}\, \boldsymbol{\xi})^{\alpha_S}} \int_{\mathbb{R}} \frac{1}{2\pi \left[\omega^2 + c^{-2}(\kappa^2 + \boldsymbol{\xi}^\top \mathbf{H}\, \boldsymbol{\xi})^{2\alpha}\right]}\, \mathrm{d}\omega \\ &= \frac{\tau^2}{(2\pi)^d c(\kappa^2 + \boldsymbol{\xi}^\top \mathbf{H}\, \boldsymbol{\xi})^{\alpha_S}} \frac{1}{2\left[c^{-2}(\kappa^2 + \boldsymbol{\xi}\, \mathbf{H}\, \boldsymbol{\xi})^{2\alpha}\right]^{1/2}} \\ &= \frac{\tau^2}{2(2\pi)^d (\kappa^2 + \boldsymbol{\xi}^\top \mathbf{H}\, \boldsymbol{\xi})^{\alpha_{tot}}}. \end{aligned} \tag{2.31}$$

Using the change of variable $\boldsymbol{\xi} = \kappa\, \mathbf{H}^{-1/2}\, \mathbf{w}$ and plugging Equation (2.31) into (2.30), we obtain

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

The last result comes from the computation of $\int_{\mathbb{R}^d} (1 + \mathbf{w}^\top\, \mathbf{w})^{-\alpha_{tot}}\, \mathrm{d}\,\mathbf{w}$ with polar coordinates.

□

The model reduces to a separable one in a particular case stated in the corollary below.

Corollary 2.1. *Let the coefficients of the SPDE (2.27) be such that $\alpha = 0$ and $\boldsymbol{\gamma} = \mathbf{0}$; the spatial operator applied to the spatio-temporal field $X(\mathbf{s}, t)$ is then the constant value c^{-1}. Let $Z(\mathbf{s}, t)$ be a spatio-temporal noise colored in space, with $Z_S(\mathbf{s})$ satisfying (2.28). If $\alpha_S > d/2$, the stationary solution to the SPDE is a separable*

spatio-temporal field with covariance

$$C_{ST}(\mathbf{h}, u) = \frac{\tau^2 \Gamma(\alpha_S - d/2)}{2\Gamma(\alpha_S)(4\pi)^{d/2}\kappa^{2(\alpha_S - d/2)}|\mathbf{H}|^{1/2}} C^{\mathcal{M}}_{\alpha_S - d/2}\left(\kappa \left\| \mathbf{H}^{-1/2}\,\mathbf{h} \right\|\right) \exp\left(\frac{u}{c}\right),$$

with smoothness parameter equal to $\nu = \alpha_S - d/2$.

2.3.3 Relationship between diffusion and advection-diffusion SPDE

Let us consider the stationary solution to the spatio-temporal SPDE

$$\left[\frac{\partial}{\partial t} + \mathcal{L}_g + \boldsymbol{\gamma} \cdot \nabla\right] X(\mathbf{s}, t) = W(\mathbf{s}, t),$$

where $\mathcal{L}_g$ is a purely spatial symbol operator that satisfies the SCEU. Given the stationary solution $X(\mathbf{s}, t)$, if we shift it by defining $X_{\boldsymbol{\gamma}}(s, t) = X(\mathbf{s} - \boldsymbol{\gamma}\, t, t)$, then $X_{\boldsymbol{\gamma}}(\mathbf{s}, t)$ satisfies the SPDE

$$\left[\frac{\partial}{\partial t} + \mathcal{L}_g\right] X_{\boldsymbol{\gamma}}(\mathbf{s}, t) = W_2(\mathbf{s}, t),$$

where $W_2(\mathbf{s}, t)$ is a shifted form of $W(\mathbf{s}, t)$, but still a white noise with the same covariance structure. For this reason, the solutions $X(\mathbf{s}, t)$ and $X_{\boldsymbol{\gamma}}(\mathbf{s}, t)$ have the same spatial trace and the spatio-temporal structure is influenced only by a shift term depending on $\boldsymbol{\gamma}$. If we wanted to simulate the solution $X(\mathbf{s}, t)$, we could simulate the solution $X_{\boldsymbol{\gamma}}(\mathbf{s}, t)$ of the SPDE without advection and shift it, by ensuring that the domain is shifted too. This property will be illustrated in Chapter 4 through simulations.

Chapter 3

Discretization of spatio-temporal SPDEs

Résumé

Dans ce chapitre nous décrivons les approches de discrétisation usuelles pour les EDP et les EDPS, puis nous présentons l'approche choisie pour la classe d'EDPS spatio-temporelles d'advection-diffusion. Nous discutons la notion de domination d'advection et proposons des méthodes de stabilisation pour les EDPS qui présentent cette caractéristique. Enfin, nous présentons la forme approchée du champ aléatoire Gaussien de Markov et la structure creuse de sa matrice de précision.

3.1 Discretization approaches for (S)PDEs

In numerical analysis, different discretization approaches can be chosen to approximate PDEs. The main ones are the *Finite Difference Method* (FDM), the *Finite Element Method* (FEM), the *Finite Volume Method* (FVM) and spectral methods.

The principle of the FDM is to assign one discrete unknown per each discretization point over a mesh and to write one equation per point. At each discretization point, the derivatives of the unknown are approximated by finite differences through the use of Taylor expansions. The FDM becomes difficult to use when the coefficients involved in the equation are discontinuous, or when the geometry of the domain is complex. For simplified geometries, FDM is an easy method to implement and likely leads to faster codes than FEM and FVM.

The FEM (see e.g. Quarteroni (2008)) is based on a variational formulation, which is written for both the continuous and the discrete problems. The variational formulation is obtained by multiplying the original equation by a *test function*. The continuous unknown is then approximated by a linear combination of the test functions for the discrete variational formulation (this is the so called *Galerkin method*); the resulting equation is integrated over the domain. The choice of the FEM to discretize spatial derivatives is, in general, directed by the consideration that the method fits naturally to a general spatial domain, even for complex geometries, since the triangulation of the domain can be much more precise than with FDM (Boulakia et al., 2015).

The FVM evaluates the unknowns at discrete locations in the spatial domain by using an integral formulation of the balance equations and a successive approximation of integrals by numerical integration at each cell of the decomposition of the domain. The method is based on conservation laws.

Spectral methods consist in writing the solution of the differential equation as a sum of basis functions (for example, as a Fourier series) and choosing the coefficients of the terms of the sum in order to satisfy the PDE as well as possible.

In the following sections, we review the methods used in the statistical literature to solve the SPDEs whose solutions are spatio-temporal random fields; then, we present the discretization method chosen in our work.

3.1.1 Finite Element Method

Lindgren et al. (2020) represented the continuously indexed random process $X(\mathbf{s}, t)$, solution to the spatio-temporal DEMF SPDE (2.24), by a discretely indexed GMRF, to enable fast inference. They considered a discretization approach based entirely on the FEM, both for temporal and spatial approximation. This approach directly generalizes the spatial discretization in Lindgren et al. (2011).

Specifically, they gave a GMRF representation of the process $X(\mathbf{s}, t)$ in Equation (2.24) on the domain $\Omega \times [a, b]$, where $\Omega \subset \mathbb{R}^d$ is a polygonal domain and $[a, b] \subset \mathbb{R}$. For the spatial domain, they used the Neumann boundary condition

$$\nabla X \cdot \hat{\mathbf{n}} = 0 \quad \text{on } \partial\Omega \tag{3.1}$$

where $\hat{\mathbf{n}}$ is the normal vector to the boundary. They assumed that the domain is extended away from the region of interest to account for possible errors at the boundaries. For the temporal boundary, they ensured that the stochastic process is stationary in time, i.e.,

$$\text{Cov}(X(\mathbf{s}_1, t_1), X(\mathbf{s}_2, t_2)) = \text{Cov}(X(\mathbf{s}_1, t_1 + u), X(\mathbf{s}_2, t_2 + u)), \tag{3.2}$$

whenever $t_1, t_2, t_1 + u, t_2 + u$ are in the domain.

Let $\Phi_T = \{\phi_j\}_{j=1}^{N_T}$ denote a set of piecewise linear temporal basis functions over a regular discretization of $[a, b]$. Let $\Psi_S = \{\psi_i\}_{i=1}^{N_S}$ denote a set of piecewise linear spatial basis functions obtained by a triangulation of the spatial domain. They approximated the process $X(\mathbf{s}, t)$ in (2.24), with boundary conditions as in (3.1) and (3.2), in the Kronecker basis $\Psi_S \otimes \Phi_T$, as

$$X_h(\mathbf{s}, t) = \sum_{i=1}^{N_S} \sum_{j=1}^{N_T} x_{ij} \psi_i(\mathbf{s}) \phi_j(t). \tag{3.3}$$

This is the projection of the solution $X(\mathbf{s}, t)$ into the finite Hilbert space spanned by Ψ_S and Φ_T. They proved that the coefficients x_{ij} in (3.3) have a precision matrix

that is expressed as a sum of Kronecker products. More details on this approach are available in Lindgren et al. (2020).

3.1.2 Finite Volume Method

Other statisticians dealing with SPDEs prefer to use the FVM, that is well suited for the numerical simulation of various types (elliptic, parabolic, or hyperbolic) of conservation laws (Eymard et al., 2000). Some of the important features of the FVM are similar to those of the FEM: it may be used on arbitrary geometries, using structured or unstructured meshes, and it leads to robust schemes. Fuglstad (2010) considered the model given in Equation (2.14) in the one-dimensional space domain to derive the precision matrix using the FVM. Then, in following works (Fuglstad et al., 2015; Fuglstad and Castruccio, 2020; Hu et al., 2022; Berild and Fuglstad, 2023), the FVM was adopted for spatial SPDE models in 2D and 3D. The extension of the FVM to spatio-temporal models in 2D and 3D was never considered, and usually the temporal spatial repetitions were supposed to be independent.

3.1.3 Spectral methods

Spectral methods have been used in spatio-temporal statistics, mostly for approximating or solving deterministic integro-difference equations or PDEs. Wikle and Cressie (1999) introduced a dynamic spatio-temporal model obtained from an integro-difference equation that is approximated by using a reduced dimensional spectral basis. Cressie and Wikle (2011, Chapter 7) gave an overview of basis function expansions in spatio-temporal statistics.

In a stochastic framework, spectral methods are widely used to analyze approximations of solutions to specific SPDEs. When the FEM and the FVM lack in generality for cases with complicated operators involved, spectral methods can be used, even if they require a particular basis of functions to be fixed. This requisite implies that the chosen basis can be useful for a specific setting and less adapted to different equations. One of the possible basis is the *Fourier basis* (see Chilès and Delfiner (1999) for a comprehensive description of the method). Efficient simulations of the SPDE models can be easily conceived using a Fourier spectral method as proposed in Lang and Potthoff (2011) and extended in Carrizo Vergara (2018). Sigrist et al. (2015) showed how to obtain a space–time Gaussian process by solving an advection–diffusion SPDE using the real Fourier transform. They also showed that, by solving the SPDE using Fourier functions, they obtained computationally efficient statistical inference methods. This computational efficiency is due to the temporal Markov property and the fact that Fourier functions are eigenfunctions of the spatial differential operators. The overall computational cost is determined by the fast Fourier transform (FFT) (Cooley and Tukey, 1965), which is $\mathcal{O}(N_T \log(N_S))$. However, it is difficult to fit this model to sparse real data, since usually we cannot obtain a precise representation of the data in the Fourier basis. Another downside of spectral methods is that, since Fourier

terms are global functions, stationarity is a necessary assumption, while this is not a requirement for the FEM and the FVM (see Chapter 7).

3.2 FEM for advection-dominated PDEs

In this section, we consider PDEs where both diffusion (second order) and advection (first order) terms are present, as in the case of Equation (2.27). We present how the FEM, used to discretize the spatial derivatives, is modified to take into account a particular characteristic of the PDE: the advection-dominated state.

The steady advection-diffusion PDE reads

$$-\lambda \Delta X + \boldsymbol{\gamma} \cdot \nabla X = f,$$

with $\lambda > 0$ and $\boldsymbol{\gamma} \in \mathbb{R}^d$.

If the problem is diffusion-dominated ($\lambda \gg \|\boldsymbol{\gamma}\|$), we can safely apply the Galerkin technique, i.e., we can find a unique solution into a finite Hilbert space spanned by the basis functions $\{\psi_i\}_{i=1}^{N_S}$ defined on a triangulation with N_S spatial points. However, if we consider advection-dominated problems, applying the Galerkin technique introduces significant instabilities (Quarteroni, 2008, Chapter 5). To investigate this kind of phenomena, we analyze the 1D problem ($\gamma \in \mathbb{R}$)

$$-\lambda \frac{\partial^2 X}{\partial x^2} + \gamma \frac{\partial X}{\partial x} = f. \tag{3.4}$$

We define the *Péclet number* as $\mathrm{Pe}^h = \frac{\gamma h}{2\lambda}$, where h is a measure of the mesh of the spatial domain (e.g. the average length of the intervals in 1D). Pe^h is defined to be the ratio of the rate of advection of a physical quantity by the flow to the rate of diffusion of the same quantity driven by an appropriate gradient. If $\mathrm{Pe}^h > 1$, that is, if $h > 2\lambda/\gamma$, the PDE is said to be *advection-dominated* and it can be shown that the numerical solution X to (3.4) is oscillating (Quarteroni, 2008, Chapter 5), thus it is not physically admissible. Then, h must be set to a small value such that $\mathrm{Pe}^h < 1$. However, the condition $h < 2\lambda/\gamma$ is too strict in practice, especially for advection-dominated problems.

A way of stabilizing the advection operator $\gamma \cdot \nabla$ is to replace the diffusion coefficient λ with $\tilde{\lambda} = \lambda(1 + \phi(\mathrm{Pe}_h))$, where $\lim_{h\to 0} \phi(\mathrm{Pe}_h) = 0$ and $\phi(\mathrm{Pe}_h) \geq \mathrm{Pe}_h$. This method is called *upwind* (U) in the simplified case where $\phi(\mathrm{Pe}_h) = \mathrm{Pe}_h$. The idea behind the stabilization method is to add an artificial diffusion term equal to $\lambda\phi(\mathrm{Pe}_h)\frac{\partial^2 X}{\partial x^2}$ that depends on the size of the discretization mesh h and on the Péclet number. In this way, the equation with the additional stabilization term reads

$$-\lambda(1 + \phi(\mathrm{Pe}_h))\frac{\partial^2 X}{\partial x^2} + \gamma \frac{\partial X}{\partial x} = f$$

and it is not advection-dominated anymore, since its Péclet number $\widetilde{\mathrm{Pe}}_h$ is now equal to

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

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

The problem extended to dimension d, where we add also a reaction term given by bX, is the following:

$$-\nabla\cdot(\lambda\nabla X) + \boldsymbol{\gamma}\cdot\nabla X + bX = f. \tag{3.5}$$

The weak solution $X \in \mathcal{V}$ to (3.5), where $\mathcal{V}$ is a Hilbert space, satisfies

$$\underbrace{\int_\Omega \lambda\nabla X\cdot\nabla v\,\mathrm{d}\,\mathbf{s} + \int_\Omega \boldsymbol{\gamma}\cdot\nabla X v\,\mathrm{d}\,\mathbf{s} + \int_\Omega bXv\,\mathrm{d}\,\mathbf{s}}_{\mathcal{A}(X,v)} = \underbrace{\int_\Omega fv\,\mathrm{d}\,\mathbf{s}}_{\mathcal{G}(v)} \quad \forall v\in\mathcal{V}. \tag{3.6}$$

We can have a possible pathological situation, when the advection coefficient $\boldsymbol{\gamma}$ is much larger than the diffusion coefficient λ. In this case, we can resort to a modified version of the problem, with augmented diffusion, similarly to the 1D case. Inspired by the 1D case, we consider the Galerkin method that searches $X_h \in \mathcal{V}_h \subset \mathcal{V}$ such that

$$\mathcal{A}_h(X_h, v_h) = \mathcal{G}(v_h) \quad \forall v_h \in \mathcal{V}_h.$$

To choose $\mathcal{A}_h(\cdot,\cdot)$, the general idea is to pick up

$$\mathcal{A}_h(\cdot,\cdot) = \mathcal{A}(\cdot,\cdot) + \mathcal{S}_h(\cdot,\cdot)$$

The additional term $\mathcal{S}_h(\cdot,\cdot)$ has the purpose of eliminating, or at least reducing, the numerical oscillations produced by the Galerkin method when the grid is not fine enough, and are therefore named stabilization terms. By defining h as a measure of the size of the triangulation of the spatial domain (e.g. the average diameter of the triangles), the term $\mathcal{S}_h(\cdot,\cdot)$ has to vanish at the limit $h \to 0$ to ensure *weak consistency*, i.e., the fact that $\lim_{h\to 0}\mathcal{A}_h(X - X_h, v_h) = 0, \quad \forall v_h \in \mathcal{V}_h$. We have several possible numerical approaches:

1. *Upwind* (U)

   $\mathcal{S}_{U,h}(w_h, v_h) = Qh\int_\Omega \nabla w_h\cdot\nabla v_h\,\mathrm{d}\,\mathbf{s}\,, \qquad Q>0.$

2. *Streamline Diffusion* (SD) (Hughes and Brooks, 1981)

   $\mathcal{S}_{SD,h}(w_h, v_h) = Qh\int_\Omega(\boldsymbol{\gamma}\cdot\nabla w_h)(\boldsymbol{\gamma}\cdot\nabla v_h)\,\mathrm{d}\,\mathbf{s}\,, \qquad Q>0.$

   $\boldsymbol{\gamma}\cdot\nabla w$ is called *stream-line*; it refers to all diffusion going on along the advection direction. Generally Q is set to $Q = \|\boldsymbol{\gamma}\|^{-1}$.

The scaling coefficient h is necessary to recover consistency. Indeed, element-size, since $\nabla w_h \sim h^{-1}$, if h was not present, we would have both $\mathcal{S}_{U,h} \sim h^{-1}$ and

$\mathcal{S}_{SD,h} \sim h^{-1}$, which implies no consistency. Both methods are only weakly consistent and provide an error that is $\mathcal{O}(h)$, i.e., they are first order convergent. The Upwind stabilization term can be considered as an additional artificial diffusion equal to $Qh\Delta X$ in the SPDE (3.5). This extra diffusion is not only in the direction of the transport, where we aim to reduce the oscillations, but also in the orthogonal direction, where there is no problem of convergence. The SD method, on the other hand, considers only an artificial diffusion along the advection direction and it acts as an extra anisotropic diffusion $Qh\nabla \cdot \boldsymbol{\gamma}\,\boldsymbol{\gamma}^\top \nabla X$ in (3.5).

It is also possible to add stabilization terms that ensure *strong consistency*, i.e., such that $\mathcal{A}_h(X - X_h, v_h) = 0, \quad \forall v_h \in \mathcal{V}_h, \quad \forall h$. These methods obtain a perturbed problem in the strong form. In this case, the Galerkin method searches $X_h \in \mathcal{V}_h$ such that

$$\mathcal{A}(X_h, v_h) + \widetilde{\mathcal{S}}_h(X_h, f, v_h) = \mathcal{G}(v_h) \quad \forall v_h \in \mathcal{V}_h$$

where $\widetilde{\mathcal{S}}_h$ is such that $\widetilde{\mathcal{S}}_h(X_h, f, v_h) = 0$, to enforce strong consistency. Two typical choices for $\widetilde{\mathcal{S}}_h$ are Galerkin Least Squares (GLS) and Streamline Upwind Petrov-Galerkin (SUPG) methods, detailed below.

1. *Galerkin Least Squares* (GLS)

   $\widetilde{\mathcal{S}}_{GLS,h}(w_h, f, v_h) = \sum_{\mathrm{Tr}\in\mathcal{T}} \tau_{\mathrm{Tr}} \int_\Omega (\mathcal{L}w_h - f)\mathcal{L}v_h \,\mathrm{d}\,\mathbf{s}\,,$
   with $\tau_{\mathrm{Tr}} = \frac{\delta h_{\mathrm{Tr}}}{|\boldsymbol{\gamma}|}, \quad \delta > 0, \quad h_{\mathrm{Tr}} = \mathrm{diam}(\mathrm{Tr})$, with Tr being a triangle in the triangulation $\mathcal{T}$
   and where $\mathcal{L}u = -\lambda\Delta u + \boldsymbol{\gamma}\cdot\nabla u + bu$.

2. *Streamline Upwind Petrov-Galerkin* (SUPG)

   $\widetilde{\mathcal{S}}_{SUPG,h}(w_h, f, v_h) = \sum_{\mathrm{Tr}\in\mathcal{T}} \tau_{\mathrm{Tr}} \int_\Omega (\mathcal{L}w_h - f)(\mathcal{L}^{SS}v_h) \,\mathrm{d}\,\mathbf{s}\,,$
   where $\mathcal{L}^{SS}$ is the skew-symmetric part of $\mathcal{L}$, i.e., $\mathcal{L}^{SS}u = \frac{1}{2}(\mathrm{div}(\boldsymbol{\gamma}\,u) + \boldsymbol{\gamma}\cdot\nabla u)$, as explained in the following remark.

Remark 3.2.1. We can always write an operator as the sum of a symmetric $\mathcal{L}^S$ and a skew-symmetric $\mathcal{L}^{SS}$ part: $\mathcal{L}u = \mathcal{L}^S u + \mathcal{L}^{SS}u$. Let the advection-reaction-diffusion operator $\mathcal{L}u = -\lambda\Delta u + \mathrm{div}(\boldsymbol{\gamma}\,u) + bu$. Since

$$\mathrm{div}(\boldsymbol{\gamma}\,u) = \frac{1}{2}\,\mathrm{div}(\boldsymbol{\gamma}\,u) + \frac{1}{2}\,\mathrm{div}(\boldsymbol{\gamma}\,u) = \frac{1}{2}\,\mathrm{div}(\boldsymbol{\gamma}\,u) + \frac{1}{2}u\,\mathrm{div}(\boldsymbol{\gamma}) + \frac{1}{2}\,\boldsymbol{\gamma}\cdot\nabla u,$$

we can decompose the operator in the following way:

$$\mathcal{L}u = \left[-\lambda\Delta + b + \frac{1}{2}\,\mathrm{div}(\boldsymbol{\gamma})\right] u + \frac{1}{2}\left[\mathrm{div}(\boldsymbol{\gamma}\,u) + \boldsymbol{\gamma}\cdot\nabla u\right].$$

Note that the reaction coefficient b has become $b^\star = b + \frac{1}{2}\operatorname{div}(\boldsymbol{\gamma})$. The first bracket is a symmetric term, since

$$\begin{aligned}\langle \mathcal{L}^S u, v\rangle_{\mathcal{V}',\mathcal{V}} &= \lambda(\nabla u, \nabla v) + (b^\star u, v)\\ &= -\lambda\langle u, \Delta v\rangle_{\mathcal{V},\mathcal{V}'} + (u, b^\star v)\\ &= \langle u, \mathcal{L}^S v\rangle_{\mathcal{V},\mathcal{V}'}, \qquad \forall u, v \in \mathcal{V},\end{aligned}$$

and the second bracket is a skew-symmetric term, since

$$\begin{aligned}\langle \mathcal{L}^{SS} u, v\rangle_{\mathcal{V}',\mathcal{V}} &= \frac{1}{2}(\operatorname{div}(\boldsymbol{\gamma}\, u), v) + \frac{1}{2}(\boldsymbol{\gamma}\cdot\nabla u, v)\\ &= -\frac{1}{2}(\boldsymbol{\gamma}\, u, \nabla v) + \frac{1}{2}(\nabla u, \boldsymbol{\gamma}\, v)\\ &= -\frac{1}{2}(u, \boldsymbol{\gamma}\cdot\nabla v) - \frac{1}{2}(u, \operatorname{div}(\boldsymbol{\gamma}\, v))\\ &= -\langle u, \mathcal{L}^{SS} v\rangle_{\mathcal{V},\mathcal{V}'}, \qquad \forall u, v \in \mathcal{V}.\end{aligned}$$

GLS and SUPG approximate the solution with error rate $\mathcal{O}(h^{m+\frac{1}{2}})$, where m is the degree of the polynomial approximation used in the approximation. In the simple case of piecewise linear functions, $m = 1$. One can ask how to choose between the several methods proposed before. On the one hand, the SD method is easy to implement, but it is only first order accurate with respect to the mesh size and much more diffusive with respect to strongly consistent methods. On the other hand, it is well known that the SUPG method has the capability of improving numerical stability for convection-dominated flows, while satisfying a strong consistency property (Tezduyar, 1992). Like many other conforming stabilized methods, the SUPG scheme contains an element-wise stabilization parameter τ_{Tr} that has to be tuned in practice, and, except for simplified situations, the optimal value is not known. Moreover, discontinuous methods are naturally well suited for problems whose solution have low regularity, and hence they work very well for advection-diffusion problems, both in advection-dominated and diffusion-dominated regimen.

In our work, we accept to have a first order convergence and we opt for the Streamline Diffusion method, which is easier to implement. We define $\mathcal{S} = \mathcal{S}_{SD}$ for ease of notation.

In Figure 3.1, we represent how a stabilization method works. The 2D problem considered here is Equation (3.5) with $b = 0$, $\lambda = 0.1$, $f = 1$. The domain is $\Omega = [-1, 1] \times [-1, 1]$ with Dirichlet boundary conditions ($X = 0$ on $\delta\Omega$) and the mesh is a grid with $h = 0.1$. $\boldsymbol{\gamma}$ is set to $\boldsymbol{\gamma}_1 = [0.5, 0.5]^\top$ for a diffusion-dominated problem and to $\boldsymbol{\gamma}_2 = [5, 5]^\top$ for an advection-dominated problem. In fact, $\mathrm{Pe}_{h,1} = 0.35$, while $\mathrm{Pe}_{h,2} = 3.5$. Figure 3.1 shows the FEM solution to the PDE with $\boldsymbol{\gamma}_1$ on the left, the FEM solution to the PDE with $\boldsymbol{\gamma}_2$ without any stabilization in the middle and the FEM solution to the PDE with $\boldsymbol{\gamma}_2$ with the Streamline Diffusion stabilization term on the right. In the bottom line, we show the solution along the $[1, 1]$-diagonal. We remark how the stabilization method in the right plot remove the oscillations present

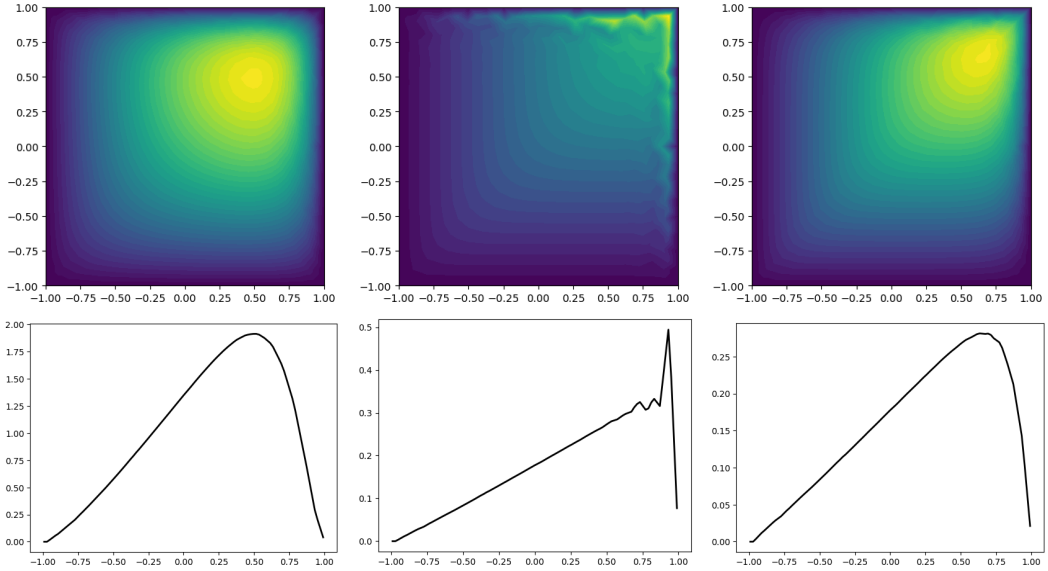


FIGURE 3.1: FEM solution to the PDE (3.5) with $b = 0$, $\lambda = 0.1$, $f = 1$, $h = 1$ and Dirichlet BCs. Left: $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 = [0.5, 0.5]^\top$ without stabilization; Middle: $\boldsymbol{\gamma} = \boldsymbol{\gamma}_2 = [5, 5]^\top$ without stabilization. Right: $\boldsymbol{\gamma} = \boldsymbol{\gamma}_2 = [5, 5]^\top$ with SD stabilization.

in the middle plot.

3.3 Discretization of the unsteady advection-diffusion SPDE

In the numerical analysis literature, when the (S)PDE presents a first-order derivative in time, generally it is discretized in space using either the FDM, the FEM or the FVM, and discretized in time using a standard implicit solver like Euler. The discretization of SPDE by finite differences in time and finite elements in space has been considered by several authors in theoretical studies, see for example Boulakia et al. (2015); Cao et al. (2007); Hausenblas (2003).

In the *method of lines*, the (S)PDE is discretized first in space, obtaining a semi-discrete equation (system of ODEs), and then in time. This approach allows switching from one temporal discretization to another without much effort.

In the *method of Rothe*, the (S)PDE is discretized first in time, and then in space. This is useful when one needs to adapt the mesh between time steps. In fact, with the Rothe method, one can choose the spatial discretization independently at each time step. For unsteady (S)PDEs, generally simple time stepping methods, such as, for example, Crank-Nicolson or the implicit and explicit Euler schemes, are efficient enough. For these simple time integrators, it is much easier to hand-code the time integration than the spatial discretization. That might explain why most people in the finite element world prefer to think along the Rothe method.

As we have already seen in Section 3.1.1, the third possibility is to discretize the (S)PDE both in space and time simultaneously (also known as space-time discretization), as it was proposed by Lindgren et al. (2020).

Following the considerations above, the unsteady advection-diffusion SPDE (2.27) is discretized with the Rothe method, first in time and then in space, using the Finite Difference Method (FDM) and the Finite Element Method (FEM), respectively. The temporal domain $[1, T]$ is discretized in N_T regular time steps of length $dt = T/N_T$. Since implicit solvers are usually less sensitive to numerical instability than explicit solvers, the implicit Euler scheme is chosen for the temporal discretization. This choice implies unconditional stability, hence convergence towards the stationary solution. Moreover, we will show in Section 3.3.2 that the implicit Euler scheme is necessary to obtain a sparse structure of the global spatio-temporal precision matrix. The FEM method for the spatial discretization is the Galerkin method with Neumann boundary conditions, as detailed in Lindgren et al. (2011).

The solution in two dimensions is now detailed. The solution in three dimensions involve geometrical technicalities, but is otherwise very similar. Let $\Omega \subset \mathbb{R}^2$ be a compact and connected domain of $\mathbb{R}^2$. Ω is meshed using a triangulation $\mathcal{T}$ with N_S vertices $\{\mathbf{s}_1, \dots, \mathbf{s}_{N_S}\} \subset \Omega$. Let $h := \max_{\mathrm{Tr} \in \mathcal{T}} h_{\mathrm{Tr}}$, where h_{Tr} is the diameter of the triangle $\mathrm{Tr} \in \mathcal{T}$. As before, we use a first order finite element representation X_h of the solution to the spatial SPDE, i.e., $X_h = \sum_{i=1}^{N_S} x_i \psi_i$. The weights $\{x_i\}_{i=1}^{N_S}$ define uniquely the values of the field at the vertices, while the values in the interior of the triangles are determined by linear interpolation. The Galerkin solution is then obtained by finding the weights that fulfill the weak formulation of Equation (2.27) for test functions belonging to the space $\mathcal{V}_h$ spanned by $\{\psi_i\}_{i=1}^{N_S}$.

Proposition 3.1. *Let $X(\mathbf{s}, t)$ be the spatio-temporal process solution to Equation (2.27) with $\alpha \in \{0, 1\}$ and spatio-temporal white noise, i.e. $Z(\mathbf{s}, t) = W(\mathbf{s}, t) = W_T(t) \otimes W_S(\mathbf{s})$. Let $\mathcal{T}$ be a triangulation of Ω and $\{\psi_i\}_{i=1}^{N_S}$ be the piecewise linear basis functions defined over $\mathcal{T}$. Let us define the mass matrix $\mathbf{M} = [M_{ij}]_{i,j=1}^{N_S}$, the stiffness matrix $\mathbf{G} = [G_{ij}]_{i,j=1}^{N_S}$, the advection matrix $\mathbf{B} = [B_{ij}]_{i,j=1}^{N_S}$ and the matrix $\mathbf{K} = [K_{ij}]_{i,j=1}^{N_S}$ as follows:*

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

Then, at each time step, the continuous Galerkin finite element solution vector $\mathbf{x}^{(k+1)} = \{x_i^{(k+1)}\}_{i=1}^{N_S}$ satisfies

$$
\left(\mathbf{M} + \frac{dt}{c} (\mathbf{K} + \mathbf{B}) \right) \mathbf{x}^{(k+1)} = \mathbf{M}\, \mathbf{x}^{(k)} + \frac{\tau \sqrt{dt}}{\sqrt{c}} \, \mathbf{M}^{1/2}\, \mathbf{z}^{(k+1)}, \tag{3.7}
$$

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

$W_T(t) \otimes Z_S(\mathbf{s})$, the discretization reads

$$\left(\mathbf{M} + \frac{dt}{c}(\mathbf{K} + \mathbf{B})\right)\mathbf{x}^{(k+1)} = \mathbf{M}\,\mathbf{x}^{(k)} + \frac{\tau\sqrt{dt}}{\sqrt{c}}\,\mathbf{M}\,\mathbf{L}_S^{\top}\,\mathbf{z}^{(k+1)},$$

where $\mathbf{L}_S$ is the Cholesky decomposition of $\mathbf{Q}_S^{-1}$, the covariance matrix of the discretized solution $\mathbf{Z}_S$ of the spatial SPDE (2.28), obtained with the continuous Galerkin FEM (Lindgren et al., 2011).

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

$$\left[\frac{\partial}{\partial t} + \frac{1}{c}(\kappa^2 - \Delta) + \frac{1}{c}\boldsymbol{\gamma}\cdot\nabla\right] X(\mathbf{s}, t) = \frac{\tau}{\sqrt{c}} W(\mathbf{s}, t). \tag{3.8}$$

For the discretization of the temporal derivative in Equation (3.8), we opt for the implicit Euler scheme, which considers the differential equation

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

with initial value $X^{(1)} = X(t_0)$. The method produces a sequence $\{X^{(k)}\}_{k=1}^{N_T}$, such that $X^{(k)}$ approximates $X(t_0 + kdt)$, where dt is the time step size. The approximation reads

$$X^{(k+1)} = X^{(k)} + dt f(t^{(k+1)}, X^{(k+1)}). \tag{3.9}$$

In the specific case of Equation (3.8), the implicit Euler discretization step reads

$$X^{(k+1)}(\mathbf{s}) - X^{(k)}(\mathbf{s}) + dt\left[\frac{1}{c}(\kappa^2 - \Delta) + \frac{1}{c}\boldsymbol{\gamma}\cdot\nabla\right] X^{(k+1)}(\mathbf{s}) = \frac{\sqrt{dt}\tau}{\sqrt{c}} W_S^{(k+1)}(\mathbf{s}), \tag{3.10}$$

where $W_S^{(k+1)}(\mathbf{s})$ is a spatial white noise obtained by integrating out the temporal white noise. In fact, the cumulative time-integral process between 0 and $a > 0$, $\int_0^a W(\mathbf{s}, t)dt$, is a purely spatial Gaussian white noise multiplied by the standard deviation of the temporal increment, which is equal to $1/\sqrt{a}$, hence the $\sqrt{dt}$ in the right-hand side term.

For ease of notation, we denote $X^{(k+1)} = X^{(k+1)}(\mathbf{s})$, $X^{(k)} = X^{(k)}(\mathbf{s})$ and $W_S = W_S^{(k+1)}(\mathbf{s})$, since the spatial noise is independent of the temporal step k.

At each time step of the temporal discretization, a spatial Finite Element Method method is applied. In our case, we use the continuous Galerkin with Neumann boundary condition. The weak form of Equation (3.10) is

$$\int_\Omega X^{(k+1)} v \,\mathrm{d}\,\mathbf{s} + \frac{dt}{c}\left(\int_\Omega \kappa^2 X^{(k+1)} v \,\mathrm{d}\,\mathbf{s} - \int_\Omega \Delta X^{(k+1)} v \,\mathrm{d}\,\mathbf{s} + \int_\Omega \boldsymbol{\gamma}\cdot\nabla X^{(k+1)} v \,\mathrm{d}\,\mathbf{s}\right) = \\ = \int_\Omega X^{(k)} v \,\mathrm{d}\,\mathbf{s} + \frac{\sqrt{dt}\tau}{\sqrt{c}} W_S(v), \quad \forall v \in \mathcal{V}, \tag{3.11}$$

where $\mathcal{V}$ is the Hilbert space in which we search the solution and $W_S(v)$ is the white noise applied to the test function v.

By applying Green's first identity, i.e., by writing

$$\int_\Omega \Delta X^{(k+1)} v \,\mathrm{d}\,\mathbf{s} = -\int_\Omega \nabla X^{(k+1)} \cdot \nabla v \,\mathrm{d}\,\mathbf{s} + \int_{\partial\Omega} v \cdot (\nabla X^{(k+1)} \cdot \hat{\mathbf{n}}) \,\mathrm{d}\sigma\,,$$

with $\hat{\mathbf{n}}$ being the normal vector on the boundary, and by simplifying the second term thanks to the Neumann boundary condition, we obtain

$$\underbrace{\int_\Omega X^{(k+1)} v \,\mathrm{d}\,\mathbf{s} + \frac{dt}{c}\left(\int_\Omega \kappa^2 X^{(k+1)} v \,\mathrm{d}\,\mathbf{s} + \int_\Omega \nabla X^{(k+1)} \cdot \nabla v \,\mathrm{d}\,\mathbf{s} + \int_\Omega \boldsymbol{\gamma} \cdot \nabla X^{(k+1)} v \,\mathrm{d}\,\mathbf{s}\right)}_{\mathcal{A}(X^{(k+1)}, v)}$$
$$= \underbrace{\int_\Omega X^{(k)} v \,\mathrm{d}\,\mathbf{s}}_{\mathcal{C}(X^{(k)}, v)} + \underbrace{\frac{\sqrt{dt}\tau}{\sqrt{c}} W_S(v)}_{\mathcal{E}(v)}, \quad \forall v \in \mathcal{V}.$$

Let $\mathcal{V}_h$ be the space of finite element solutions spanned by the basis functions $\{\psi_i\}_{i=1}^{N_S}$. The Galerkin method allows us to find an approximated solution $X_h^{(k+1)} \in \mathcal{V}_h \subset \mathcal{V}$ to the SPDE, such that

$$\mathcal{A}(X_h^{(k+1)}, v_h) = \mathcal{C}(X_h^{(k)}, v_h) + \mathcal{E}(v_h) \quad \forall v_h \in \mathcal{V}_h. \tag{3.12}$$

The functions $X_h^{(k+1)}$, $X_h^{(k)}$ and v_h are linear combinations of the basis functions, with

$$X_h^{(k+1)} = \sum_{i=1}^{N_S} x_i^{(k+1)} \psi_i; \quad X_h^{(k)} = \sum_{i=1}^{N_S} x_i^{(k)} \psi_i; \quad v_h = \sum_{i=1}^{N_S} v_i \psi_i.$$

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

$$\sum_{i=1}^{N_S} \mathcal{A}(\psi_i, v_h) x_i^{(k+1)} = \sum_{i=1}^{N_S} \mathcal{C}(\psi_i, v_h) x_i^{(k)} + \mathcal{E}(v_h), \quad \forall v_h \in \mathcal{V}_h, \tag{3.13}$$

where

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

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

Since any v_h can be written as a linear combination of basis functions, the formulation (3.13) is equivalent to

$$\sum_{i=1}^{N_S} \mathcal{A}(\psi_i, \psi_j) x_i^{(k+1)} = \sum_{i=1}^{N_S} \mathcal{C}(\psi_i, \psi_j) x_i^{(k)} + \mathcal{E}(\psi_j), \quad \forall j. \tag{3.14}$$

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

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

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

by following the definition of white noise in Equation (2.1).

If $\mathbf{z}^{(k+1)}$ is a (N_S)-Gaussian vector such that $\mathbf{z}^{(k+1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_S})$, $\mathbf{x}^{(k+1)}$ is the vector containing the values $\{x_i^{(k+1)}\}_{i=1}^{N_S}$ and $\mathbf{x}^{(k)}$ is the vector containing the values $\{x_i^{(k)}\}_{i=1}^{N_S}$, then the sparse linear system corresponding to Equation (3.14) reads

$$\mathbf{M}\,\mathbf{x}^{(k+1)} + \frac{dt}{c}(\mathbf{K}+\mathbf{B})\mathbf{x}^{(k+1)} = \mathbf{M}\,\mathbf{x}^{(k)} + \frac{\sqrt{dt}\tau}{\sqrt{c}}\,\mathbf{M}^{1/2}\,\mathbf{z}^{(k+1)}, \tag{3.15}$$

where $\mathbf{K} = \kappa^2\,\mathbf{M} + \mathbf{G}$ and $\mathbf{M}^{1/2}$ is any matrix such that $\mathbf{M}^{1/2}\,\mathbf{M}^{1/2} = \mathbf{M}$.

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

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

and it satisfies

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

Hence,

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

If $\mathbf{z}_S = \{z_{S,i}\}_{i=1}^{N_S}$ has precision matrix equal to $\mathbf{Q}_S$, then the sparse linear system is

$$\mathbf{M}\,\mathbf{x}^{(k+1)} + \frac{dt}{c}(\mathbf{K}+\mathbf{B})\mathbf{x}^{(k+1)} = \mathbf{M}\,\mathbf{x}^{(k)} + \frac{\sqrt{dt}\tau}{\sqrt{c}}\,\mathbf{M}\,\mathbf{L}_S^\top\,\mathbf{z}^{(k+1)}, \tag{3.16}$$

where $\mathbf{z}^{(k+1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_S})$ and $\mathbf{L}_S$ is the Cholesky decomposition of $\mathbf{Q}_S^{-1}$.

□

Remark 3.3.1. When the diffusion term includes an anisotropy matrix $\mathbf{H}$, i.e., when Δ is replaced by $\nabla \cdot \mathbf{H}\,\nabla$, the stiffness operator becomes $\mathcal{G}(v, w) = \int_\Omega \mathbf{H}\,\nabla v \cdot \nabla w \,\mathrm{d}\,\mathbf{s}$, and the stiffness matrix changes consequently.

Remark 3.3.2. The elements of the matrices $\mathbf{M}$, $\mathbf{G}$, $\mathbf{B}$ and $\mathbf{K}$ are non-zero only for pairs of basis functions which share common triangles. This implies that the matrix $(\mathbf{M} + \frac{dt}{c}(\mathbf{K}+\mathbf{B}))$ is sparse and that Equation (3.7) can be solved by LU decomposition (Turing, 1948) in an efficient way.

Remark 3.3.3. The discretization error introduced by the scheme (3.12) is defined by

$$E_h^{(k)} = e_h^{(k)} + p_h^{(k)},$$

where

$$e_h^{(k)} = X_h^{(k)} - \Pi_h X_{t_k}, \quad p_h^{(k)} = \Pi_h X_{t_k} - X_{t_k}$$

for $1 \leq k \leq N_T$.

For $k \in \{1, \ldots, N_T\}$, the error $e_h^{(k)}$ is the difference between the approximated solution given by the scheme and the elliptic projection $\Pi_h(\cdot)$ on $\mathcal{V}_h$ of the exact solution at time kdt. The error $p_h^{(k)}$ is the difference between the projection of the exact solution on $\mathcal{V}_h$ at time kdt and the exact solution at kdt. We have the following covergence result (Boulakia et al., 2015):

$$\sqrt{\mathbb{E}(||E_h^{(k)}||^2)} \leq K(h + \sqrt{dt}), \tag{3.17}$$

where K is a constant.

3.3.1 Stabilization of the unsteady advection-dominated SPDE

When the advection term is too strong with respect to the diffusion term in Equation (2.27), advection-domination occurs. In the framework outlined above, when $\alpha = 1$, the asymmetric matrix $\left[\mathbf{M} + \frac{dt}{c}(\mathbf{K} + \mathbf{B})\right]$ becomes ill-conditioned, which induces oscillations and unstable solutions for the continuous Galerkin approximation. Specifically, the advection-domination occurs when the Péclet number $\mathrm{Pe}^h = \frac{\|\gamma\|h}{2\lambda} > 1$, where λ is the coefficient of the isotropic Laplacian operator (see, for example, Mekuria and Rao (2016) or Quarteroni (2008, Chapter 5)).

We saw in Section 3.2 that one possible solution is to decrease the diameter h, i.e., to refine the triangulation, until the advection no longer dominates on the element-level, with $\mathrm{Pe}^h < 1$. However, in many cases this is not a feasible solution because it would increase the number of vertices beyond computation limits. Another solution, adopted here, is to introduce a stabilization term. Many stabilization approaches are possible, some being more accurate than others (see Section 3.3.1 or Quarteroni (2008, Chapter 5)). In our case, we opt for the Streamline Diffusion (SD) stabilization approach (Hughes and Brooks, 1981), considered as a good trade-off between accuracy and computational complexity. Essentially, the SD approach consists in stabilizing the advection by introducing an artificial diffusion term along the advection direction. The following proposition presents the stabilized solution to (2.27).

Proposition 3.2. *Assume the same hypotheses as in Proposition 3.1 with $\alpha = 1$. The solution to Equation (2.27) in presence of Streamline Diffusion stabilization is*

$$\left(\mathbf{M} + \frac{dt}{c}(\mathbf{K} + \mathbf{B} + \mathbf{S})\right)\mathbf{x}^{(k+1)} = \mathbf{M}\,\mathbf{x}^{(k)} + \frac{\tilde{\tau}\sqrt{dt}}{\sqrt{c}}\,\mathbf{M}^{1/2}\,\mathbf{z}^{(k+1)}, \tag{3.18}$$

where $\mathbf{S} = [S_{ij}]_{i,j=1}^{N_S}$ *is the matrix of the Streamline Diffusion stabilization operator* $\mathcal{S}$*, such that*

$$S_{ij} = \mathcal{S}(\psi_i, \psi_j) = h\|\boldsymbol{\gamma}\|^{-1} \int_\Omega (\boldsymbol{\gamma}\cdot\nabla\psi_i)(\boldsymbol{\gamma}\cdot\nabla\psi_j)\,\mathrm{d}\,\mathbf{s}\,,$$

and $\tilde{\tau} = \tau \left(|\mathbf{H} + h\|\boldsymbol{\gamma}\|^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top|\right)^{-1/4} (|\mathbf{H}|)^{1/4}$*. When the noise on the right-hand side of Equation (2.27) is colored in space, the discretization becomes*

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

where $\mathbf{z}^{(k+1)}$ *and* $\mathbf{L}_S$ *are as in Proposition 3.1.*

The proof of the discretized equation follows the same reasoning as that of Proposition 3.1 with the addition of the matrix $\mathbf{S}$. The Streamline Diffusion approach can be seen as a perturbation of the original SPDE (Bank et al., 1990). Indeed, by making the classical hypothesis of Neumann boundary condition on Ω and by using the Green's first identity, we get

$$\int_\Omega (\boldsymbol{\gamma}\cdot\nabla X^{(k+1)})(\boldsymbol{\gamma}\cdot\nabla v)\,\mathrm{d}\,\mathbf{s} = -\int_\Omega \nabla\cdot(\boldsymbol{\gamma}\boldsymbol{\gamma}^\top)\nabla X^{(k+1)} v\,\mathrm{d}\,\mathbf{s}\,.$$

As a consequence, the original SPDE (2.27) can be rewritten with an additional diffusion term as

$$\left[\frac{\partial}{\partial t} + \frac{1}{c}\left[\kappa^2 - \nabla\cdot\left(\mathbf{H} + h\|\boldsymbol{\gamma}\|^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top\right)\nabla + \boldsymbol{\gamma}\cdot\nabla\right]\right] X(\mathbf{s},t) = \frac{\tau}{\sqrt{c}} Z(\mathbf{s},t). \tag{3.19}$$

The term $(h\|\boldsymbol{\gamma}\|^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top)$ acts as an anisotropic "diffusion" matrix that is added to the anisotropy matrix $\mathbf{H}$ of the original diffusion. This extra diffusion stabilizes the advection directed along the direction $\boldsymbol{\gamma}$. By following the proof of Proposition 2.1, we find that the marginal variance of the spatial field $X(\cdot,t)$ of Equation (3.19) is equal to

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

For the variance to be equal to the variance in Proposition 2.1, τ must be replaced by $\tilde{\tau} = \tau\left(|\mathbf{H} + h\|\boldsymbol{\gamma}\|^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top|\right)^{1/4}(|\mathbf{H}|)^{-1/4}$.

3.3.2 Spatio-temporal GMRF approximation

Proposition 3.3. *In presence of an advection-dominated flow and a spatio-temporal white noise on the right-hand side of Equation (2.27), the discretized vector* $\mathbf{x}^{(k+1)}$ *on the mesh* $\mathcal{T}$ *at each time step is the solution to the following equation:*

$$\begin{aligned} \mathbf{x}^{(1)} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \\ \mathbf{x}^{(k+1)} &= \mathbf{D}\,\mathbf{x}^{(k)} + \mathbf{E}\,\mathbf{z}^{(k+1)}, \end{aligned} \tag{3.20}$$

where

$$\begin{aligned}\mathbf{D} &= \left(\mathbf{M}+\frac{dt}{c}(\mathbf{K}+\mathbf{B}+\mathbf{S})\right)^{-1}\mathbf{M},\\ \mathbf{E} &= \frac{\tilde{\tau}\sqrt{dt}}{\sqrt{c}}\left(\mathbf{M}+\frac{dt}{c}(\mathbf{K}+\mathbf{B}+\mathbf{S})\right)^{-1}\mathbf{M}^{1/2},\end{aligned} \tag{3.21}$$

and $\mathbf{z}^{(k+1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_S})$ *is independent of* $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k+1)}$*. In presence of a spatio-temporal noise colored in space on the right-hand side of Equation (2.27), the matrix* $\mathbf{E}$ *reads*

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

where $\mathbf{L}_S$ *is defined in Proposition 3.1.*

Proof. Starting from Equation (3.18), which represents the numerical scheme for the advection-diffusion spatio-temporal SPDE with stabilization, it is straightforward to obtain (3.20). □

When the SPDE is not advection-dominated, which implies that no stabilization term is needed, Equation (3.21) is replaced by the similar equation where the matrix $\mathbf{S}$ is deleted and $\tilde{\tau}$ is replaced by τ.

$\boldsymbol{\Sigma}$, which is the covariance matrix of the spatial trace $\mathbf{x}^{(1)}$ at the first time step, can be taken to be equal to any admissible positive definite matrix. The closer $\boldsymbol{\Sigma}$ is to the covariance C_S of $X(\cdot, t)$, the faster the stationary solution is obtained. When the hypotheses of Proposition 2.1 are satisfied, an efficient option is to choose $\boldsymbol{\Sigma}$ as the Matérn covariance of Equation (2.29).

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

Proposition 3.4. *Let* $\mathbf{x}_{1:N_T} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N_T)}]^{\top}$ *be the vector containing all spatial solutions until time step* N_T *of Equation (3.20). The global precision matrix* $\mathbf{Q}$ *of the vector* $\mathbf{x}_{1:N_T}$ *of size* $(N_S N_T, N_S N_T)$ *reads*

$$\mathbf{Q} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1}+\mathbf{D}^{\top}\mathbf{F}^{-1}\mathbf{D} & -\mathbf{D}^{\top}\mathbf{F}^{-1} & 0 & \ldots & 0 \\ -\mathbf{F}^{-1}\mathbf{D} & \mathbf{F}^{-1}+\mathbf{D}^{\top}\mathbf{F}^{-1}\mathbf{D} & -\mathbf{D}^{\top}\mathbf{F}^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -\mathbf{F}^{-1}\mathbf{D} & \mathbf{F}^{-1}+\mathbf{D}^{\top}\mathbf{F}^{-1}\mathbf{D} & -\mathbf{D}^{\top}\mathbf{F}^{-1} \\ 0 & \ldots & 0 & -\mathbf{F}^{-1}\mathbf{D} & \mathbf{F}^{-1} \end{pmatrix}, \tag{3.22}$$

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

Proof. Let us denote $\mathbf{x}_{1:N_T} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N_T)}]^\top$ the vector containing all spatial solutions until time step N_T. Then,

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

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

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

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

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

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

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

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

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

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

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

□

3.3.3 Sparsity pattern of the precision matrix

Here, we report two examples of sparsity patterns of the spatio-temporal precision matrix, computed with $N_S = 5^2$ and $N_T = 5$ and $N_S = 20^2$ and $N_T = 5$ respectively. The sparsity is due both to the discretization in space with the FEM and to the use of the implicit Euler scheme.

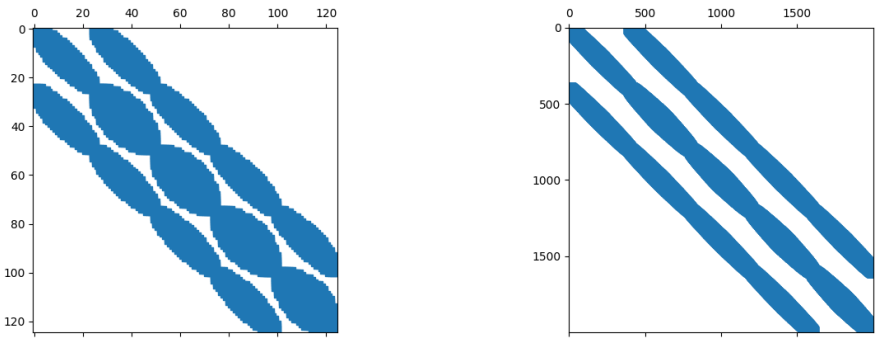


FIGURE 3.2: Sparsity pattern of precision matrix. Left: $N_S = 5^2$ and $N_T = 5$; Right: $N_S = 20^2$ and $N_T = 5$.

Chapter 4

Simulation

Résumé

Dans ce chapitre, nous présentons des méthodes destinées à générer des simulations non conditionnelles de champs Gaussiens spatio-temporels stationnaires solutions de la classe d'EDPS d'advection-diffusion présentée au chapitre précédent. La principale méthode utilisée est la discrétisation aux Éléments Finis en espace et aux Différences Finies en temps. Nous comparons cette méthode à une méthode spectrale basée sur l'approximation de la transformée de Fourier du champ par la transformée de Fourier rapide. Enfin, nous montrons comment la domination d'advection influence la solution de l'EDPS et comment une méthode de stabilisation peut résoudre ce problème.

In this chapter we present some examples of non-conditional simulation of stationary spatio-temporal Gaussian fields, solutions to the unsteady advection-diffusion SPDE (2.27). The main method that will be used is the FEM/FDM discretization described in Section 3.3. We will compare this method to a spectral method based on the approximation of the Fourier Transform of the field by Fast Fourier Transform. Finally, we will show how the advection-domination influences the solution of the SPDE (2.27) and how the Streamline Diffusion stabilization can tackle this issue. Some implementation details are available in Appendix A.1.

4.1 Simulation of spatio-temporal SPDEs

We consider the advection-diffusion SPDE (2.27) with $d = 2$. The spatio-temporal domain $\Omega \times [1, T]$ of $X(\mathbf{s}, t)$, with $\Omega \subset \mathbb{R}^2$, is discretized in space with a triangulation $\mathcal{T}$ with N_S nodes and discretized in time by means of N_T regular time steps. This space-time discretization is denoted $\mathcal{T}' = \mathcal{T} \times \{T/N_T, \ldots, T\}$.

When considering the SPDE (2.27) on a spatial bounded domain Ω, one has to choose some boundary conditions. Lindgren et al. (2011) and Khristenko et al. (2019) showed that, when Neumann boundary conditions are used, one can obtain a satisfactory solution to SPDE (2.3) on Ω by using a domain $\tilde{\Omega}$ that is extended by a distance which is at least two times the practical correlation range $r = \sqrt{8(\alpha_{tot} - 1)}/\kappa$ outside the domain of interest Ω. They further showed that the error between the covariance of the solution and the covariance of the solution inside the domain Ω

decreases exponentially as r increases. We chose to use the same practical framework. When $\alpha_{tot} \leq 1$, r cannot be computed; in this case, we consider that $r = \sqrt{0.8}/\kappa$.

A non-conditional simulation of the spatio-temporal Gaussian vector $\mathbf{x}_{1:N_T}$ on the discretized domain $\mathcal{T}'$ with known precision matrix $\mathbf{Q}$ as in Equation (3.22) can be obtained through

$$\mathbf{x}_{1:N_T} = \mathbf{L}\,\mathbf{z}, \tag{4.1}$$

where $\mathbf{z}$ is a $N_S N_T$ vector with independent zero-mean, unit variance and normally distributed random components, and $\mathbf{L}$ is a matrix such that (Gentle, 2009)

$$\mathbf{L}\,\mathbf{L}^\top = \mathbf{Q}^{-1}\,. \tag{4.2}$$

Since $\mathbf{Q}^{-1}$ is not known, we can compute $\mathbf{Q}_{chol}$ as the Cholesky decomposition of the precision matrix $\mathbf{Q}$. Then, the simulated vector $\mathbf{x}_{1:N_T}$ is computed as the solution of the linear system $\mathbf{Q}_{chol}^\top\,\mathbf{x}_{1:N_T} = \mathbf{z}$. Two performance issues arise from this approach. First, the computation of the Cholesky decomposition of $\mathbf{Q}$ is intractable for large problems or when the matrix is not sparse enough; then, once computed, this decomposition must be stored, and is used to solve a linear system. Both these tasks get increasingly expensive as the size or the filling of $\mathbf{Q}_{chol}$ increases.

However, we know that, by construction of $\mathbf{Q}$, we only need to store the sparse spatial matrices of the tridiagonal block structure, which results in a cheaper storage. Likewise, a non-conditional simulation of the spatio-temporal field $X(\mathbf{s}, t)$ on the discretized domain $\mathcal{T}'$, i.e., the vector $\mathbf{x}_{1:N_T}$, can be obtained by solving the sparse linear system (3.20) governing the field in space and updating the Euler scheme at each time step. In this case, only sparse (N_S, N_S) spatial matrices are used, which reduces the computational cost.

The convergence of the spatio-temporal field to the stationary solution obtained from the theory is observed by comparing the empirical spatio-temporal covariance with the theoretical one.

First, we compare the empirical covariance of the spatial trace with the theoretical covariance of the corresponding spatial model, when it exists (see Proposition 2.1). In Figure 4.1 we show the spatial trace of the stationary solution of an advection-diffusion SPDE with $\alpha = 1$, $\alpha_S = 0$, $\kappa = 0.1$, $c = 1$, $\boldsymbol{\gamma} = [1, 0]^\top$, $\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$, at 4 time steps separated by $4dt$. The solution is the approximated field obtained with the FEM/FDM approach detailed in Section 3.3. We remark how the anisotropic matrix influences the form of the spatial field, giving an elongated vertical dependence structure, and how the spatial field moves to the right due to the advection term.

Now, we restrict ourselves to the purely isotropic diffusive SPDE ($\boldsymbol{\gamma} = [0, 0]$ and $\mathbf{H} = \mathbf{I}$) with $\kappa = 0.3$, $c = 1$, $\tau = 1$, and we compare the spatio-temporal models obtained with different combinations of α and α_S. The combinations are summarized in Table 4.1, along with the marginal spatial variance and the value of the practical range (when available) computed from the results of Proposition 2.1. The models (A)

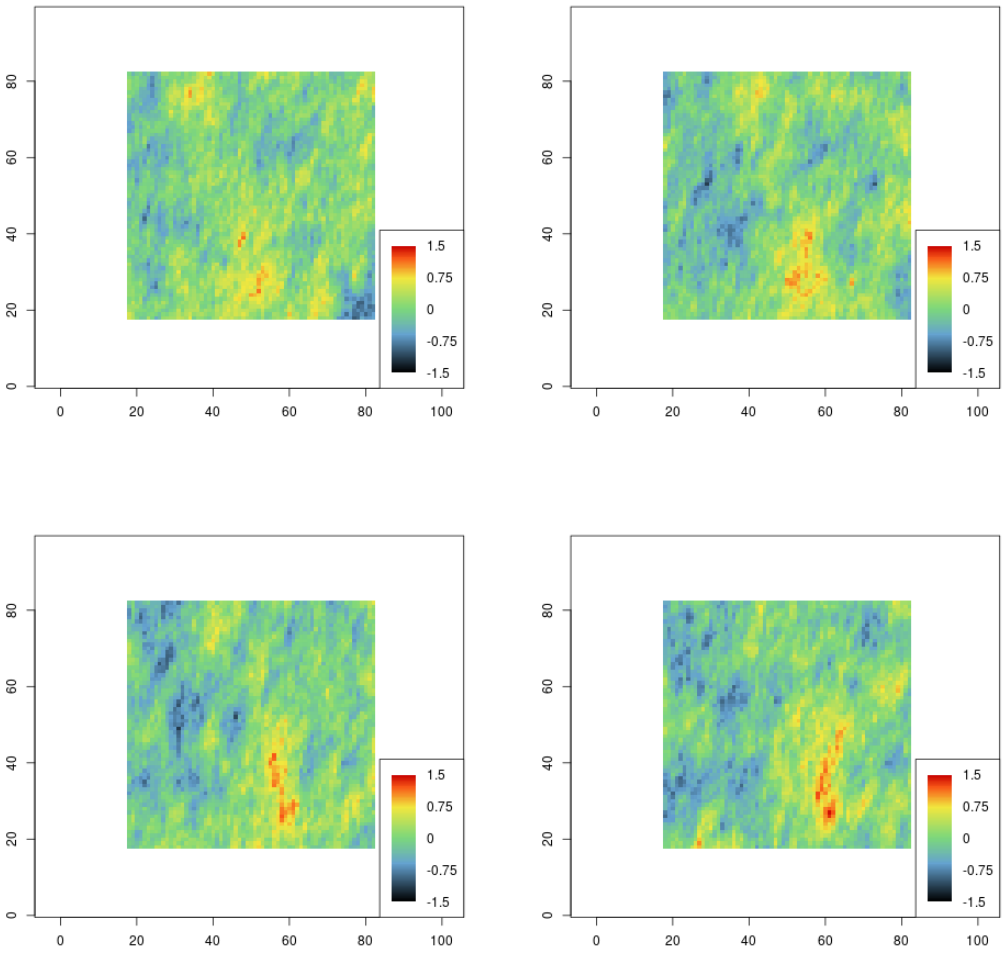


FIGURE 4.1: Spatial trace at 4 time steps (separated by $4dt$) of the simulation of advection-diffusion SPDE for $\alpha = 1$, $\alpha_S = 0$, $\kappa = 0.1$, $c = 1$, $\boldsymbol{\gamma} = [1, 0]^\top$, $\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$.

and (B) are separable models.

Model name	α	α_S	σ^2	$r = \sqrt{8\nu}/\kappa$
(A)	0	2	0.442	9.428
(B)	0	4	18.193	16.330
(C)	1	0	/	/
(D)	2	0	0.442	9.428
(E)	1	2	2.456	13.333
(F)	2	2	18.193	16.330

TABLE 4.1: Different diffusive models obtained with combinations of α and α_S, along with corresponding marginal spatial variance and practical range.

In Figures 4.2 and 4.3 we plot the experimental spatio-temporal variograms of the simulations of the diffusive SPDE models with the parameters of Table 4.1. The plots show that the marginal variance is slightly underestimated.

4.2 Comparison FEM/FDM with spectral method

Understanding the marginal properties of the model which is solution to the SPDE is important, as they provide additional information about the process. We already know that the covariance can be directly computed from the spectral density, even though this is not always possible analytically. We consider the spectral density of the nonseparable model in order to compute the space-time covariance and marginal spatial and temporal ones as well.

An approximation of the covariances of X, solution to the advection-diffusion SPDE (2.27), can be obtained using a procedure related to the classical development on Fourier basis, for which the computations can be efficiently obtained through the use of the Fast Fourier Transform (FFT), as proved by Sigrist et al. (2015) and Carrizo Vergara (2018). This method can be applied to any evolution equation presented in Section 2.2.6. We refer to Carrizo Vergara (2018, Chapter 6) for a detailed exposition of the method.

As an illustration of the method, in Figure 4.4 we plot the marginal spatial and temporal covariance functions obtained with the FFT approach from the SPDE (2.27) with $\alpha = 0$, $\alpha_S = 2$, $\kappa = 0.1$, $\tau = 1$, $c = 0.2$, along with the theoretical covariances of the separable model with spatial covariance equal to a Matérn with $\nu = \alpha_S - 1 = 1$ and temporal covariance equal to an exponential with scale c (see Proposition 2.1). We note that the spatial and temporal covariances computed with the FFT approximation fit well the theoretical covariances.

In Figure 4.5 we plot the marginal spatial covariance functions obtained with both the FEM/FDM and FFT approaches from the SPDE (2.27) with spatio-temporal white noise ($\alpha_S = 0$), $\kappa = 0.5$, $\tau = 1$, $c = 1$. α is set to $\alpha_1 = 2$ and $\alpha_2 = 3$ in order to obtain a Matérn spatial trace with $\nu_1 = 1$ and $\nu_2 = 2$, respectively. The

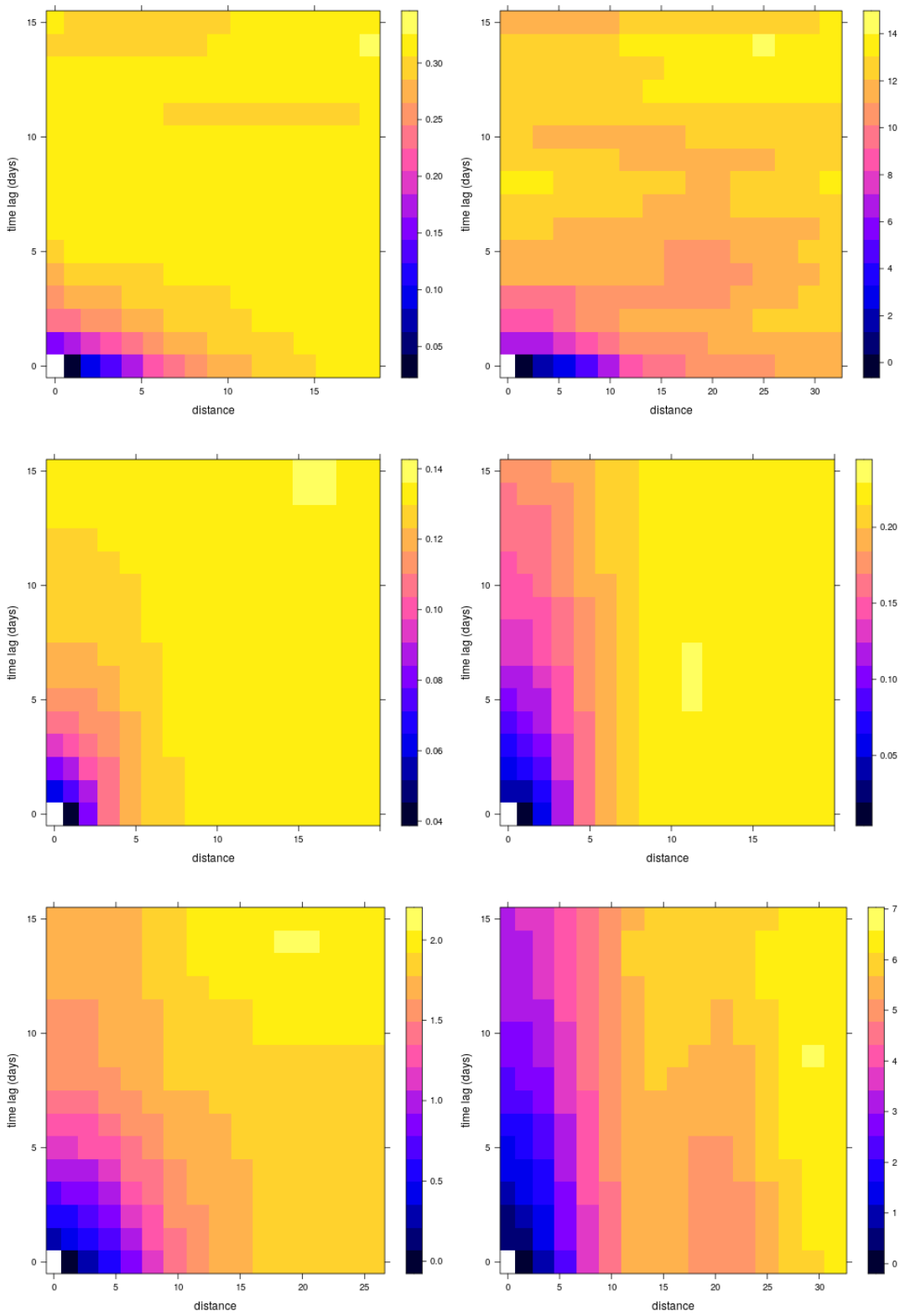


FIGURE 4.2: Spatio-temporal variogram maps of simulation of models of Table 4.1. Top row: (A), (B). Middle row: (C), (D). Bottom row: (E), (F).

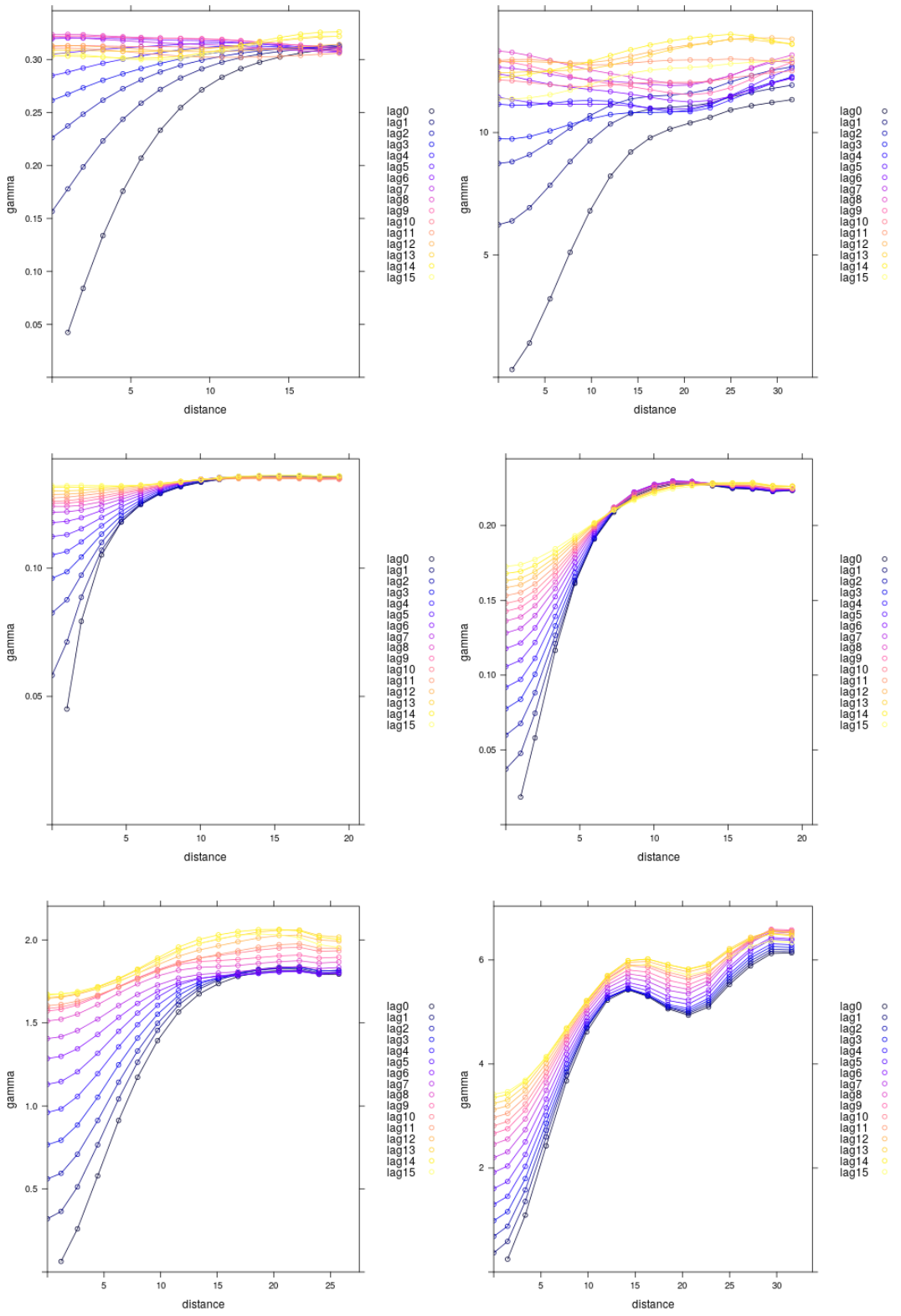


FIGURE 4.3: Spatio-temporal variogram of simulation of models of Table 4.1. Top row: (A), (B). Middle row: (C), (D). Bottom row: (E), (F).

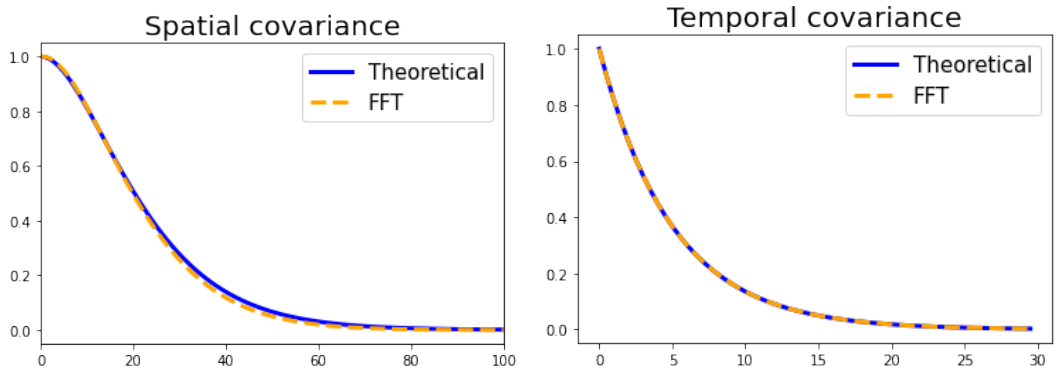


FIGURE 4.4: Separable spatio-temporal model ($\alpha_S = 2$, $\alpha = 0$, $\kappa = 0.1$, $\tau = 1$, $c = 0.2$, $\boldsymbol{\gamma} = \mathbf{0}$). Left: spatial trace obtained with FFT and theoretical Matérn covariance ($\nu = \alpha_S - 1 = 1$). Right: temporal trace obtained with FFT and theoretical exponential covariance.

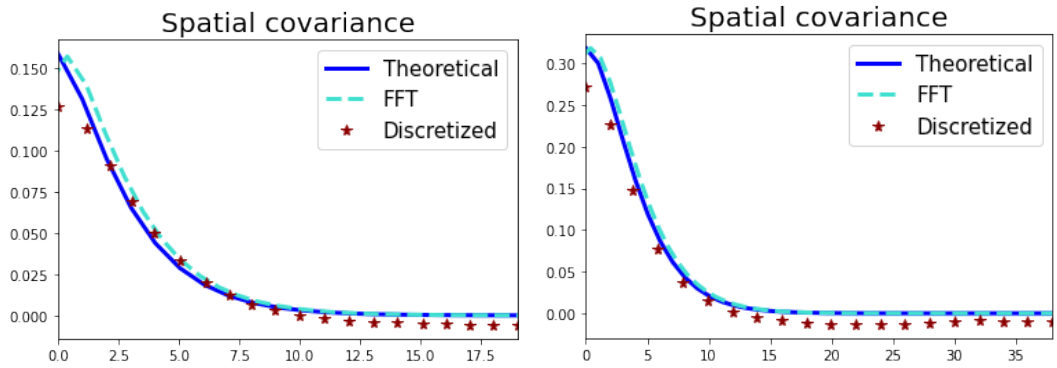


FIGURE 4.5: Covariances of stationary spatial trace of nonseparable spatio-temporal model ($\alpha_S = 0$, $\kappa = 0.5$, $\tau = 1$, $c = 1$, $\boldsymbol{\gamma} = \mathbf{0}$) computed with FEM/FDM ("discretized"), FFT and theoretical Matérn. Left: $\alpha = 2$, hence $\nu = 1$. Right: $\alpha = 3$, hence $\nu = 2$.

theoretical marginal variance and practical range, computed from Equation (2.29), are $\sigma_1^2 = 0.159$ and $r_1 = 5.657$ (and $\sigma_2^2 = 0.318$ and $r_2 = 8$). These values are well retrieved by both the approximation approaches, with a little error for the variances estimated by the FEM/FDM approach. This is visible in the first star on the left, which is lower than the dashed and solid lines at the same abscissa. We already remarked that the marginal variances were slightly underestimated in Figure 4.3. This is probably due to the discretization in space, which, being a bit coarse here, smooths the behavior at the origin.

4.3 Simulation of advection-diffusion SPDEs as shifted diffusion SPDEs

In the following, we illustrate with simulations the result of Section 2.3.3, i.e., the fact that the result of the advection-diffusion SPDE is equal to a shifted solution of the classical diffusion SPDE. We consider the solution of the advection-diffusion SPDE

with $\alpha = 1$, $\alpha_S = 0$, $dt = 1$, $\tau = 1$, $c = 1$, $\kappa = 0.3$ and $\boldsymbol{\gamma} = [1, 0]^\top$, along with the solution of the corresponding diffusion SPDE (with same parameters α, α_S, κ, τ and c), which has been computed on a larger grid, so that we can shift the solution along the x-axis of an interval $dx = dt\gamma_x = 1$ every time step.

In Figure 4.6, we plot the covariance function of the solutions of the advection-diffusion SPDE between 4 following time steps, when the simulation has already converged. We remark how the higher values of the covariance map follow the direction $[1, 0]^\top$, as expected from the advection term with velocity parameter $\boldsymbol{\gamma} = [1, 0]^\top$. In Figure 4.7, we plot the covariance function of the solutions of the purely diffusive SPDE with the same parameters, after being shifted along the x-axis of an interval $dx = dt\gamma_x = 1$ every time step. Moreover, in Table 4.2, we report the mean and the standard devation of the differences between the two experimental spatial covariances (shifted diffusion SPDE and corresponding advection-diffusion SPDE) over 4 following time steps. We note that the plots are very similar and the differences very small, which empirically illustrates the result of Section (2.3.3).

time steps	t1,t1	t1,t2	t1,t3	t1,t4
mean (std) $\times 10^{-3}$	1.19 (4.54)	-0.41 (4.91)	0.86 (3.61)	-1.60 (4.12)

TABLE 4.2: Mean (std) $\times 10^{-3}$ of differences between experimental spatial covariances over 4 following time steps with shifted diffusion SPDE and corresponding advection-diffusion SPDE.

4.4 Simulation of advection-dominated SPDEs

In the following, we simulate a spatio-temporal field from the SPDE (2.27) when the advection term is dominant with respect to the diffusive term (see Section 3.3.1. We consider the solution of the SPDE (2.27) with $\alpha = 1$, $\alpha_S = 0$, $dt = 1$, $\tau = 1$, $c = 1$, $\kappa = 0.3$ and $\boldsymbol{\gamma} = [0, 5]^\top$. We plot in Figure 4.8 the spatial trace of the solution to the SPDE at 4 following steps. We remark how the strong advection coming from the left side influences the spatial trace, causing problems in the solution. To better capture the amplitude of the fluctuations, we report here the average variance over 50 time steps. In the purely diffusive case, the average variance is 0.147, in the not stabilized advection-dominated case it is 0.958, while in the stabilized advection-dominated case it is 0.167. In Figure 4.9 we plot the same spatial traces but with the introduction of the Streamline Diffusion stabilization term into the FEM. We note that now the solution is stable and we remark the influence of the stabilization term on the diffusive behavior. Indeed, the additional diffusive term $h/\|\boldsymbol{\gamma}\|\nabla \cdot \boldsymbol{\gamma}\,\boldsymbol{\gamma}^\top \nabla X$ in the SPDE is visible as an anisotropic covariance structure that the spatial trace clearly shows.

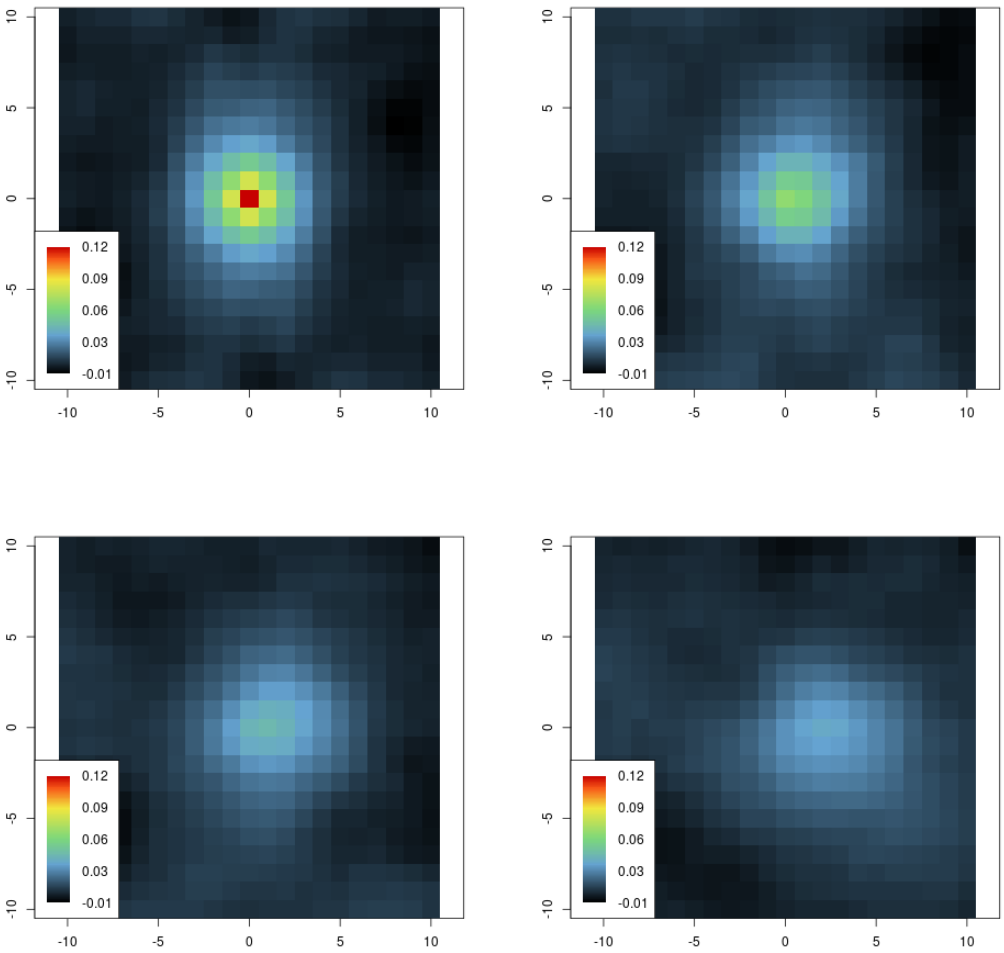


FIGURE 4.6: Experimental spatial covariance over 4 following time steps of the simulation of advection-diffusion SPDE for $\kappa = 0.3$, $\tau = 1$, $c = 1$, $\alpha = 1$ and $\boldsymbol{\gamma} = [0, 1]^\top$.

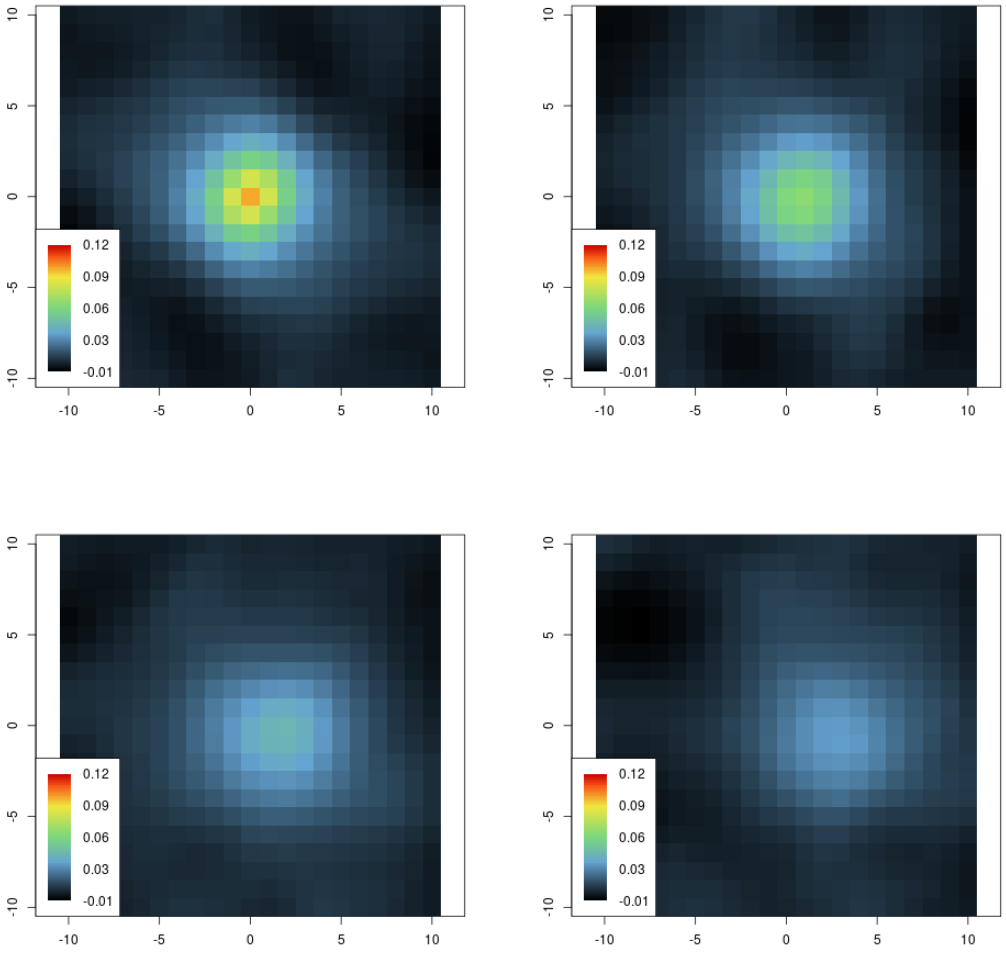


FIGURE 4.7: Experimental spatial covariance over 4 following time steps of the simulation of diffusion SPDE for $\kappa = 0.33$, $\alpha = 1$, $\tau = 1$ and $c = 1$, shifted at each time step of $dx = 1$ on the x-axis.

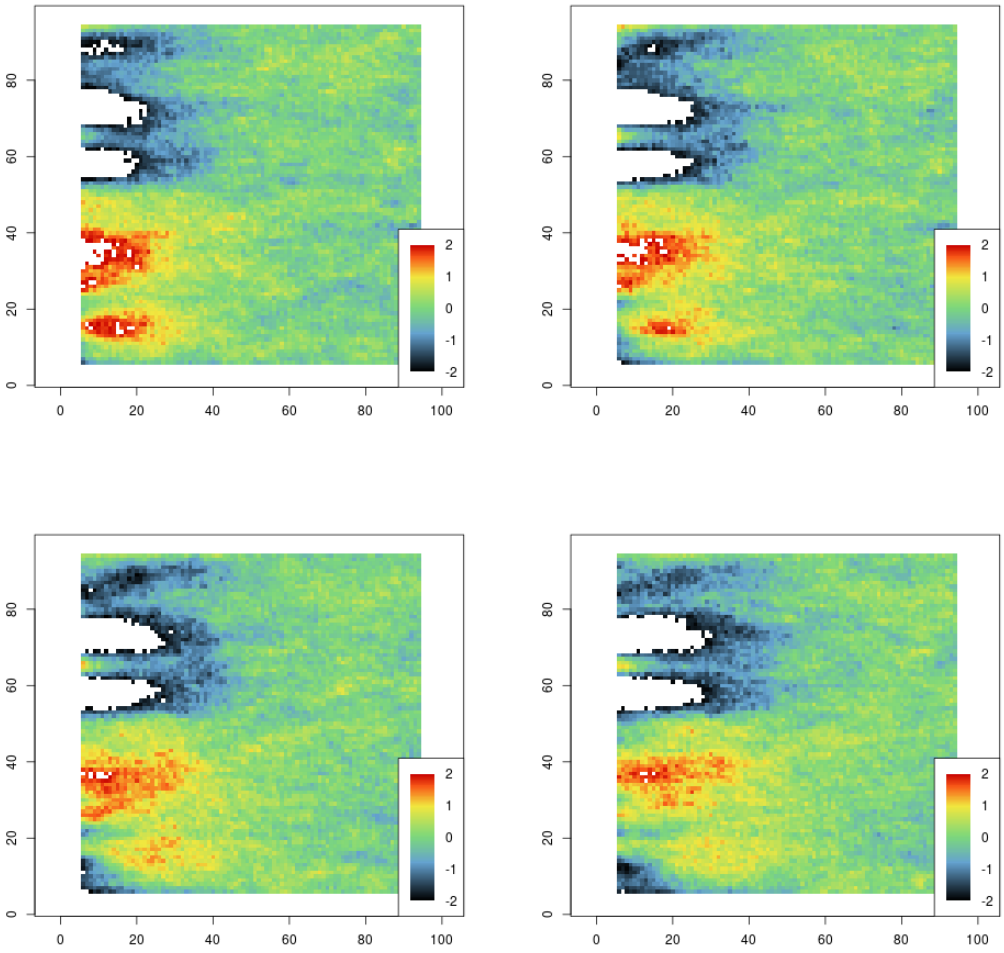


FIGURE 4.8: Advection-dominated SPDE. Spatial trace at 4 time steps (separated by $dt = 1$) of the simulation of advection-diffusion SPDE for $\alpha = 1$, $\alpha_S = 0$, $\kappa = 0.3$, $\tau = 1$, $c = 1$, $\boldsymbol{\gamma} = [5, 0]^\top$, without stabilization.

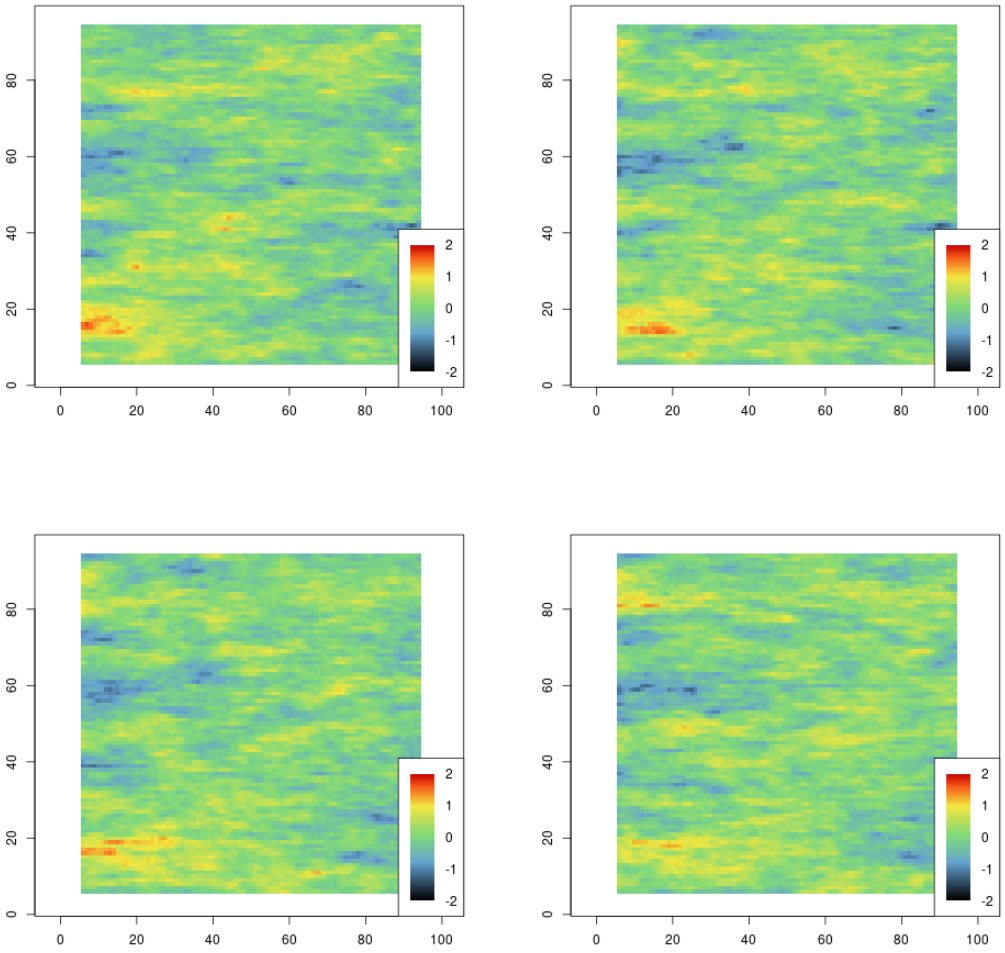


FIGURE 4.9: Stabilized advection-dominated SPDE. Spatial trace at 4 time steps (separated by $dt = 1$) of the simulation of advection-diffusion SPDE for $\alpha = 1$, $\alpha_S = 0$, $\kappa = 0.3$, $\tau = 1$, $c = 1$, $\boldsymbol{\gamma} = [5, 0]^\top$, with Streamline Diffusion stabilization.

Chapter 5

Estimation, prediction and conditional simulation

Résumé

Dans ce chapitre nous nous intéressons au problème de l'estimation des paramètres d'un modèle statistique où la partie aléatoire est la solution d'une EDPS d'advection-diffusion. Il s'agit d'inférer les propriétés statistiques du processus à partir de données spatio-temporelles partiellement observées et bruitées. Nous présentons une approche basée sur la maximisation de la vraisemblance. Les calculs s'appuient sur la décomposition de Cholesky de matrices creuses ou, dans le cas de grands jeux de données, sur des approches dites "matrix-free", qui ne font intervenir que des produits matrices-vecteurs. Puis, nous présentons la prédiction spatio-temporelle et la simulation conditionnelle par krigeage. Enfin, nous étudions les techniques proposées sur un jeu de données simulées.

This section presents an efficient implementation for parameter estimation, spatio-temporal prediction and conditional simulation within the spatio-temporal SPDE framework. We consider the advection-diffusion SPDE (2.27) with $d = 2$, $\alpha = 1$, $\mathbf{H} = \mathbf{I}$ (isotropic diffusion) and colored noise in space with $\alpha_S = 2$. Similar computations can be generalized to other values of α_S such that $\alpha_S/2$ is integer or to anisotropic diffusion.

The spatio-temporal domain $\Omega \times [1, T]$ is discretized in space with a triangulation $\mathcal{T}$ with N_S nodes and discretized in time by means of N_T regular time steps. This space-time discretization is denoted $\mathcal{T}' = \mathcal{T} \times \{T/N_T, \ldots, T\}$. At each time step $k = 1, \ldots, N_T$ there are $n^{(k)}$ observations scattered in the spatial domain Ω. There is thus a total of $n = \sum_{k=1}^{N_T} n^{(k)}$ spatio-temporal data collected in the vector $\mathbf{y}_{1:N_T} = [(\mathbf{y}^{(1)})^\top, \ldots, (\mathbf{y}^{(N_T)})^\top]^\top$.

We consider a statistical model with fixed and random effects. The fixed effect is a regression on a set of covariates and the random effect is modeled as the FEM discretization of a random field described by the SPDE (2.27) with the addition of a random noise:

$$\mathbf{y}_{1:N_T} = \boldsymbol{\eta}\, \mathbf{b} + \mathbf{A}^\top \mathbf{x}_{1:N_T} + \sigma_0\, \boldsymbol{\varepsilon}, \tag{5.1}$$

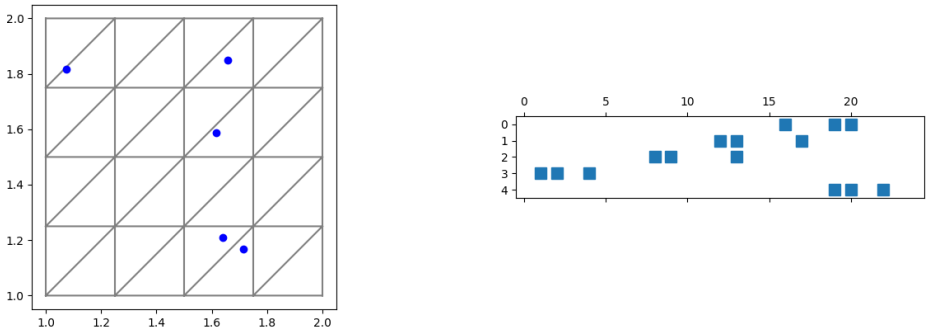


FIGURE 5.1: Left: Triangulated mesh $\mathcal{T}$ with $N_S = 5^2$ and $n_{obs} = 5$ scattered data (blue). Right: spatial projection matrix $\mathbf{A}_S^\top$ between scattered points and points $\mathcal{T}$ (size (n, N_S)).

where $\mathbf{b}$ is the vector of q fixed effects and $\boldsymbol{\eta}$ is a (n, q) matrix of covariates with $[\boldsymbol{\eta}]_{jk} = \eta_k(\mathbf{s}_j, t_j)$, $j = 1 \ldots, n$ and $k = 1, \ldots, q$. The matrix $\mathbf{A}$ is the $(N_S N_T, n)$ projection matrix between the points in $\mathcal{T}'$ and the data, and $\boldsymbol{\varepsilon}$ is a standard Gaussian random vector with independent components, independent of $\mathbf{x}_{1:N_T}$. When the observation locations do not change during the time window, $\mathbf{A}\,\mathbf{A}^\top$ is a $(N_S N_T, N_S N_T)$ block-diagonal matrix with all (N_S, N_S) equal blocks.

In the left panel of Figure 5.1, we represent a square spatial domain $[1, 2] \times [1, 2]$, discretized in space with a triangulation $\mathcal{T}$ with $N_S = 5^2$, and $n_S = 5$ scattered observations (in blue). In the right panel, we show the sparsity pattern of the projection matrix $\mathbf{A}_S^\top$ between $\mathcal{T}$ and the scattered data. For each row i of the matrix $\mathbf{A}_S^\top$, three values are different from 0: they correspond to the three vertices of the triangle in which the i-observation is located. For example, the first and the last rows correspond to the two blue points in the bottom right part of the spatial domain. Indeed, they share two common vertices. The values of the matrix are the coefficients of a convex combination (the barycentric interpolation) of the values of the discretized field at each one of the three vertices. In Figure 5.2 we present the sparsity pattern of the matrix $\mathbf{A}_S\,\mathbf{A}_S^\top$ which is the diagonal block of the spatio-temporal matrix $\mathbf{A}\,\mathbf{A}^\top$ with $N_T = 5$, when the observation locations do not change with time.

5.1 Estimation of the parameters

The parameters of the SPDE are estimated using Maximum Likelihood. We collect the parameters of the SPDE in the vector $\boldsymbol{\theta}^\top = [\kappa, \gamma_1, \gamma_2, c, \tau]$, while all the parameters of the statistical model are collected in $\boldsymbol{\psi}^\top = [\boldsymbol{\theta}^\top, \mathbf{b}^\top, \sigma_0]$. Following (5.1), $\mathbf{y}_{1:N_T}$ is a Gaussian vector with expectation $\boldsymbol{\eta}\,\mathbf{b}$ and covariance matrix

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

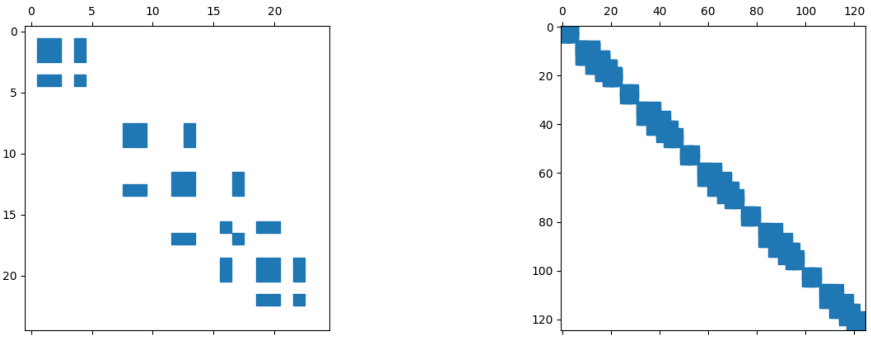


FIGURE 5.2: Left: spatial matrix $\mathbf{A}_S\,\mathbf{A}_S^\top$ (size (N_S, N_S) with $N_S = 5^2$). Right: spatio-temporal block-diagonal matrix $\mathbf{A}\,\mathbf{A}^\top$ (size $(N_S N_T, N_S N_T)$ with $N_T = 5$).

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

$$\mathcal{L}(\boldsymbol{\psi}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{\mathbf{y}_{1:N_T}}| - \frac{1}{2}(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\,\mathbf{b})^\top\,\boldsymbol{\Sigma}^{-1}_{\mathbf{y}_{1:N_T}}(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\,\mathbf{b}). \quad (5.2)$$

We use the Broyden, Fletcher, Goldfarb, and Shanno optimization algorithm (Nocedal and Wright, 2006), that makes use of the first and second-order derivatives of the objective function. The gradients of the log-likelihood function (5.2) with respect to the different parameters included in $\boldsymbol{\psi}$ are approximated with finite differences. We now propose a computationally efficient formulation of each term of the log-likelihood (5.2).

Proposition 5.1. *In the framework outlined above, we have*

$$\log|\boldsymbol{\Sigma}_{\mathbf{y}_{1:N_T}}| = n\log\sigma_0^2 - \log|\mathbf{Q}| + \log|\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top|. \quad (5.3)$$

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

$$\boldsymbol{\Sigma}_c = \begin{pmatrix} \mathbf{Q}^{-1} & \mathbf{Q}^{-1}\,\mathbf{A} \\ \mathbf{A}^\top\,\mathbf{Q}^{-1} & \boldsymbol{\Sigma}_{\mathbf{y}_{1:N_T}} \end{pmatrix}. \quad (5.4)$$

Hence,

$$\mathbf{Q}_c = \boldsymbol{\Sigma}_c^{-1} = \begin{pmatrix} \mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top & -\sigma_0^{-2}\,\mathbf{A} \\ -\sigma_0^{-2}\,\mathbf{A}^\top & \sigma_0^{-2}\,\mathbf{I} \end{pmatrix}. \quad (5.5)$$

Using block formulas, we have

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

and

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

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

Proposition 5.2. *The term* $\log|\mathbf{Q}|$ *in Equation* (5.3) *can be computed with the computationally efficient formula*

$$\log|\mathbf{Q}| = \log|\boldsymbol{\Sigma}^{-1}| + (N_T - 1)\log|\mathbf{F}^{-1}|, \tag{5.6}$$

where $\boldsymbol{\Sigma}^{-1}$ *is the precision matrix of the spatial trace* $\mathbf{x}^{(1)}$ *at the first time step, and*

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

where $\mathbf{Q}_S$ *is the precision matrix of the discretized spatial noise* $\mathbf{Z}_S$ *defined in Proposition 3.1.*

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

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

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

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

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

$$|\mathbf{Q}| = |\boldsymbol{\Sigma}^{-1}||\mathbf{F}^{-1}|^{N-1}. \tag{5.7}$$

Applying the logarithm, we obtain Equation (5.6). □

Remark 5.1.1. Note that $\mathbf{F}^{-1}$ is a (N_S, N_S) sparse, symmetric and positive definite matrix. The computation of its determinant can be obtained by Cholesky decomposition of $\mathbf{F}^{-1}$.

We here report how the log-determinant of a matrix $\mathbf{N}$ is computed by Cholesky decomposition. We have

$$\log|\mathbf{N}| = \log|\mathbf{N}_{chol}\,\mathbf{N}_{chol}^\top| = 2\log\left(\prod_{i=1}^{n} N_{chol,ii}\right) = 2\sum_{i=1}^{n}\log(N_{chol,ii}),$$

since $\mathbf{N}_{chol}$ is a triangular matrix, whose determinant is the product of the diagonal elements.

The term $\log|\mathbf{Q}+\sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top|$ requires a detailed analysis. The term $\sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top$ is an $(N_S N_T, N_S N_T)$ diagonal block matrix, whose (N_S, N_S) blocks are sparse (see Figure 5.2). The computation of $\log|\mathbf{Q}+\sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top|$ is not as straightforward as in the case of $\log|\mathbf{Q}|$, because there is no way of reducing the computation to purely spatial matrices. Depending on the size $N_S N_T$, we can either apply a Cholesky decomposition of the $(N_S N_T, N_S N_T)$ matrix $(\mathbf{Q}+\sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)$ or the matrix-free approach proposed in Pereira et al. (2022), here sketched. The logarithm function is first approximated by a Chebyschev polynomial $P(\cdot)$ (Chebyshev, 1853), then the Hutchinson's estimator (Hutchinson, 1990) is used to obtain a stochastic estimate of $\mathrm{tr}[P(\mathbf{Q}+\sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)]$. The method is detailed in Section 5.2.

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

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

Hence

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

The second term can be computed either by Cholesky decomposition or using the Conjugate Gradient (CG) method, outlined in Section 5.2.1. This latter method solves $\mathbf{N}\,\mathbf{v}=\mathbf{w}$ with respect to $\mathbf{v}$ and computes $\mathbf{v}_{sol}=\mathbf{w}^\top\,\mathbf{v}$, with $\mathbf{N}=(\mathbf{Q}+\sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)$ and $\mathbf{w}=\mathbf{A}(\mathbf{y}_{1:N_T}-\boldsymbol{\eta}\,\mathbf{b})$. In this case, it is useful to find a good preconditioner for the matrix $(\mathbf{Q}+\sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)$ to ensure fast convergence of the CG method. We found that a temporal block Gauss-Seidel preconditioner (Young, 1971, Chapter 3) was a good choice in this case (see Section 5.2.2).

5.2 Matrix-free methods

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

We will benefit from these methods for the computation of the terms of the log-likelihood to obtain fast computations even for very large matrices $\mathbf{Q}$, i.e., large $N_S N_T$. Some implementation details are available in Appendix A.1.

5.2.1 Conjugate Gradient method

Suppose we want to solve the system of linear equations

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

for the vector $\mathbf{x}$, where the known (n, n) matrix $\mathbf{N}$ is symmetric, positive-definite, and real, and $\mathbf{b}$ is known as well. We denote the unique solution of this system by $\mathbf{x}^\star$. The Conjugate Gradient (CG) method (Hestenes and Stiefel, 1952) is an iterative method that allows us to approximately solve systems where n is so large that the direct method would take too much time. It is in this sense a matrix-free method.

The solution $\mathbf{x}^\star$ is the unique minimizer of the following quadratic function

$$f(\mathbf{x}) = \frac{1}{2}\,\mathbf{x}^\top\,\mathbf{N}\,\mathbf{x} - \mathbf{x}^\top\,\mathbf{b}, \quad \mathbf{x} \in \mathbb{R}^n\,. \tag{5.8}$$

The idea behind the CG method is to take the approximated solution $\mathbf{x}_k$ at each iteration k along always different search directions $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{n-1}$. We take the first basis vector $\mathbf{p}_0$ to be the negative of the gradient of f at $\mathbf{x} = \mathbf{x}_0$, leading to $\mathbf{p}_0 = \mathbf{b} - \mathbf{N}\,\mathbf{x}_0$, which is also equal to the residual $\mathbf{r}_0 = \mathbf{b} - \mathbf{N}\,\mathbf{x}_0$. The direction at iteration $k+1$ is constructed by conjugation of all the residuals $\mathbf{r}_i = \mathbf{b} - \mathbf{N}\,\mathbf{x}_i$ for $i \leq k$. At each iteration, the residual $\mathbf{r}_k$ is orthogonal to the previous search directions, so it is guaranteed to produce a new, linearly independent search direction unless the residual is zero, in which case the problem is solved. The conjugation constraint is an orthonormal-type constraint, which makes the algorithm an example of Gram-Schmidt orthonormalization. The number of iterations K can be either set a priori or chosen by defining an error tolerance. The algorithm is detailed in 1.

Algorithm 1 Conjugate Gradient

Require: Matrix $\mathbf{N} \in \mathbb{R}^n \times \mathbb{R}^n$. Vector $\mathbf{b} \in \mathbb{R}^n$.

Ensure: The approximated solution $\mathbf{x}_K \approx \mathbf{x}$ of $\mathbf{N}\,\mathbf{x} = \mathbf{b}$.

Set $\mathbf{x}_0 = 0, \quad \mathbf{r}_0 = \mathbf{b}$

Set $\mathbf{p}_0 = \mathbf{r}_0$

for $k = 1, 2, \ldots, K$ **do**

Compute a step length $\alpha_k = \frac{\mathbf{r}_{k-1}^\top\,\mathbf{r}_{k-1}}{\mathbf{p}_{k-1}^\top\,\mathbf{N}\,\mathbf{p}_{k-1}}$

Update the approximate solution $\mathbf{x}_k = \mathbf{x}_{k-1} + \alpha_k\,\mathbf{p}_{k-1}$

Update the residual $\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_k\,\mathbf{N}\,\mathbf{p}_{k-1}$

Compute a gradient correction factor $\beta_k = \frac{\mathbf{r}_k^\top\,\mathbf{r}_k}{\mathbf{r}_{k-1}^\top\,\mathbf{r}_{k-1}}$

Set the new search direction $\mathbf{p}_k = \mathbf{r}_k + \beta_k\,\mathbf{p}_{k-1}$

end for

The algorithm seems to require storage of all previous searching directions and

residual vectors, as well as many matrix-vector multiplications, leading to expensive computations. However, we can show that this is not true. We define $\mathcal{P}_k = \text{span}\{\mathbf{p}_0, \dots \mathbf{p}_{k-1}\}$ the k-dimensional subspace spanned by the searching directions until iteration $k-1$. Since the searching vectors are built from the residuals, the subspace $\mathcal{R}_k = \text{span}\{\mathbf{r}_0, \dots \mathbf{r}_{k-1}\}$ is equal to $\mathcal{P}_k$. In fact, as each residual is orthogonal to the previous search directions, it is also orthogonal to the previous residuals, i.e., $\mathbf{r}_i^\top \mathbf{r}_j = 0$ for $i \neq j$. Recalling that $\mathbf{p}_k$ belongs to $\mathcal{P}_{k+1}$, this fact implies that each new subspace $\mathcal{P}_{k+1}$ is formed from the union of the previous subspace $\mathcal{P}_k$ and the subspace $\mathbf{N}\,\mathcal{P}_k$. Hence,

$$\begin{aligned}\mathcal{P}_k &= \text{span}\{\mathbf{p}_0, \mathbf{N}\,\mathbf{p}_0, \mathbf{N}^2\,\mathbf{p}_0, \dots, \mathbf{N}^{k-1}\,\mathbf{p}_0\} \\ &= \text{span}\{\mathbf{r}_0, \mathbf{N}\,\mathbf{r}_0, \mathbf{N}^2\,\mathbf{r}_0, \dots, \mathbf{N}^{k-1}\,\mathbf{r}_0\}\end{aligned}$$

This subspace is a *Krylov subspace*, as it is created by repeatedly applying a matrix to a vector, and it has a good property: since $\mathbf{N}\,\mathcal{P}_k$ is included in $\mathcal{P}_{k+1}$, the fact that the next residual $\mathbf{r}_{k+1}$ has to be orthogonal to $\mathcal{P}_{k+1}$ implies that $\mathbf{r}_{k+1}$ has to be $\mathbf{N}$-orthogonal to $\mathcal{P}_k$. Gram-Schmidt conjugation becomes easy, because $\mathbf{r}_{k+1}$ is already $\mathbf{N}$-orthogonal to all the previous search directions except $\mathbf{p}_k$. Hence, it is no longer necessary to store old search vectors to ensure the $\mathbf{N}$-orthogonality of new search vectors. This feature is what makes CG a matrix-free approach, because both the space complexity and time complexity per iteration are reduced from $\mathcal{O}(n^2)$ to $\mathcal{O}(m)$, where m is the number of nonzero entries of $\mathbf{N}$.

The *condition number* of a problem measures how sensitive the answer is to perturbations in the input data and to roundoff errors made during the solution process (Demmel, 1987). A problem with a low condition number is said to be *well-conditioned*, while a problem with a high condition number is said to be *ill-conditioned*. The condition number with respect to inversion of a nonsingular square matrix $\mathbf{N}$ is

$$\kappa(\mathbf{N}) = \|\mathbf{N}\|\left\|\mathbf{N}^{-1}\right\|.$$

For a condition number $\kappa(\mathbf{N})$, it turns out that the CG algorithm converges in $\mathcal{O}(\sqrt{\kappa(\mathbf{N})})$ iterations. However, in practice, the CG algorithm often converges faster than predicted by this upper bound.

5.2.2 Gauss-Seidel preconditioner

The general concept behind a preconditioner is to reduce the condition number of an ill-conditioned matrix to obtain less sensitive solutions to a linear system. Given a linear system $\mathbf{N}\,\mathbf{x} = \mathbf{b}$, we want to find the matrix $\mathbf{P}_R$ and/or $\mathbf{P}_L$ such that the condition number of $\mathbf{N}\,\mathbf{P}_R^{-1}$ (right preconditioner) or $\mathbf{P}_L^{-1}\,\mathbf{N}$ (left preconditioner) or $\mathbf{P}_L^{-1}\,\mathbf{N}\,\mathbf{P}_R^{-1}$ is lower than that of $\mathbf{N}$, and that we can easily solve $\mathbf{P}_L\,\mathbf{w} = \mathbf{g}$ or

$\mathbf{P}_R\,\mathbf{w} = \mathbf{g}$ for any $\mathbf{g}$. If this matrix exists, then we can solve the systems

$$\begin{aligned}
&\mathbf{N}\,\mathbf{P}_R^{-1}\,\mathbf{y} = \mathbf{b}, && \mathbf{P}_R\,\mathbf{x} = \mathbf{y} \quad \text{(right preconditioner)}\\
&\mathbf{P}_L\,\mathbf{y} = \mathbf{b}, && \mathbf{P}_L^{-1}\,\mathbf{N}\,\mathbf{x} = \mathbf{y} \quad \text{(left preconditioner)}\\
&\mathbf{P}_L^{-1}\,\mathbf{N}\,\mathbf{P}_R^{-1}\,\mathbf{y} = \mathbf{P}_L^{-1}\,\mathbf{b}, && \mathbf{P}_R\,\mathbf{x} = \mathbf{y} \quad \text{(two-side preconditioner)}
\end{aligned}$$

instead of $\mathbf{N}\,\mathbf{x} = \mathbf{b}$.

Let us consider only the two-side preconditioner. The following reasoning can be adapted easily to any other type of preconditioner, by setting $\mathbf{P}_L = \mathbf{I}$ for a right preconditioner and $\mathbf{P}_R = \mathbf{I}$ for a left preconditioner. The preconditioned CG method converges in $\mathcal{O}(\sqrt{\kappa(\mathbf{P}_L^{-1}\,\mathbf{N}\,\mathbf{P}_R^{-1})})$. Since this condition number is lower than the condition number of $\mathbf{N}$, this implies fewer iterations to converge. However, each iteration is more expensive, due to the extra cost of computing $\mathbf{P}_R\,\mathbf{x} = \mathbf{y}$ or $\mathbf{P}_L\,\mathbf{y} = \mathbf{b}$. This computation does not need to be explicit; it is only necessary to be able to compute the effect of applying $\mathbf{P}^{-1}$ to a vector, i.e., it is matrix-free.

Intuitively, preconditioning is an attempt to stretch the quadratic form (5.8) to make it appear more spherical, so that the eigenvalues are close to each other. The best choice for a preconditioner would be of course $\mathbf{P}_R = \mathbf{I}$ and $\mathbf{P}_L = \mathbf{N}$ (or vice-versa), since in this case the condition number of $\mathbf{P}_L^{-1}\,\mathbf{N}\,\mathbf{P}_R^{-1}$ would be equal to 1 and the quadratic form would be perfectly spherical; however, this does not make life easier, because $\mathbf{N}$ is ill-conditioned and the solution of $\mathbf{P}_L\,\mathbf{y} = \mathbf{b}$ would still be a problem. The simplest preconditioner is a diagonal matrix $\mathbf{D}$ whose diagonal entries are identical to those of $\mathbf{N}$ (i.e., $\mathbf{P}_R = \mathbf{I}$ and $\mathbf{P}_L = \mathbf{D}$). The process of applying this preconditioner, known as *Jacobi preconditioning*, is equivalent to scaling the quadratic form along the coordinate axes (by comparison, the perfect preconditioner $\mathbf{P}_L = \mathbf{N}$ scales the quadratic form along its eigenvector axes). A diagonal matrix is trivial to invert, but is often only a mediocre preconditioner.

Another possibility is to use other iterative methods as preconditioners, such as the Gauss-Seidel method or the SOR(ω) method (Successive over-relaxation) (Young, 1954).

A more elaborate preconditioner is the *incomplete Cholesky preconditioning*. Incomplete Cholesky factorization is a variant of the Cholesky decomposition in which little or no fill is allowed; $\mathbf{N}$ is approximated by the product $\tilde{\mathbf{L}}\tilde{\mathbf{L}}^\top$, where $\tilde{\mathbf{L}}$ might be restricted to have the same pattern of nonzero elements as $\mathbf{N}$; other elements of $\mathbf{L}$ are thrown away. To use $\mathbf{P}_L = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^\top$ as a preconditioner, the solution to $\tilde{\mathbf{L}}\tilde{\mathbf{L}}^\top\,\mathbf{y} = \mathbf{b}$ is computed by backsubstitution, i.e., the inverse of $\tilde{\mathbf{L}}\tilde{\mathbf{L}}^\top$ is never explicitly computed. Unfortunately, incomplete Cholesky preconditioning is not always stable.

The preconditioner that we will use in this work is the *Gauss-Seidel* (GS) preconditioner, whose matrix form is here detailed. Given the matrix $\mathbf{N}$, we have

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

where $\mathbf{D}$ is the diagonal of $\mathbf{N}$ and $\mathbf{L}^\star$ and $\mathbf{U}^\star$ are the lower-triangular part and upper-triangular part with zero on the diagonal, respectively. One iteration of the GS method reads

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

and we refer to the preconditioner $\mathbf{P}_L = \mathbf{L}^\star + \mathbf{D}$ (and $\mathbf{P}_R = \mathbf{I}$) as Gauss-Seidel preconditioner. A good property of this preconditioner is that $\rho(\mathbf{I} - (\mathbf{L}^\star + \mathbf{D})^{-1}\,\mathbf{N}) < 1$, where ρ is the spectral radius, i.e., for a positive definite matrix the GS method always converges. The preconditioned Conjugate Gradient method is described in Algorithm 2.

In the specific case of the precision matrix $\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top$, all the blocks on the diagonal are equal (except the first and the last one), and the same occurs also on the upper and lower diagonals. For this reason, the Gauss-Seidel preconditioner $\mathbf{P}_L$ can be divided in N_T temporal blocks $\mathbf{P}_L = [\mathbf{P}_{L,1}, \dots, \mathbf{P}_{L,N_T}]$ with the following properties. For the first time step, $\mathbf{P}_{L,1} = \mathbf{D}_1 = \mathbf{Q}^{(1,1)} + \sigma_0^{-2}\,\mathbf{A}_S\,\mathbf{A}_S^\top$, where $(1,1)$ is the index of the temporal block in the matrix $\mathbf{Q}$. For the time steps $i = 2, \dots, N_T - 1$, $\mathbf{P}_{L,i} = \mathbf{L}_i^\star + \mathbf{D}_i = \mathbf{Q}^{(i,i-1)} + \mathbf{Q}^{(i,i)} + \sigma_0^{-2}\,\mathbf{A}_S\,\mathbf{A}_S^\top$, and the matrices $\mathbf{N}_i$ are equal for all i. Finally, for the last temporal step, $\mathbf{P}_{L,N_T} = \mathbf{L}_{N_T}^\star + \mathbf{D}_{N_T} = \mathbf{Q}^{(N_T,N_T-1)} + \mathbf{Q}^{(N_T,N_T)} + \sigma_0^{-2}\,\mathbf{A}_S\,\mathbf{A}_S^\top$, which is different from the previous blocks. This means that only a few number of spatial matrices must be stored and that the computation of $\mathbf{P}_L\,\mathbf{y} = \mathbf{b}$ can be obtained by computing $\mathbf{P}_{L,j}\,\mathbf{y}_j = \mathbf{b}_j$ for each temporal block $j = 1, \dots, N_T$ of the matrix and of the vectors, before reconstructing the vector $\mathbf{y}$ by stacking all $\mathbf{y}_j$. This leads to N_T sparse systems of size (N_S, N_S), that can be solved with a LU decomposition (Turing, 1948) of the matrix $\mathbf{P}_{L,j}$, by remarking that the matrix $\mathbf{P}_{L,j}$ is not symmetric.

Algorithm 2 Preconditioned Conjugate Gradient

Require: Matrix $\mathbf{N} \in \mathbb{R}^n \times \mathbb{R}^n$. Left preconditioner $\mathbf{P} \in \mathbb{R}^n \times \mathbb{R}^n$. Vector $\mathbf{b} \in \mathbb{R}^n$.

Ensure: The approximated solution $\mathbf{x}_K \approx \mathbf{x}$ of $\mathbf{N}\,\mathbf{x} = \mathbf{b}$.

Set $\mathbf{x}_0 = 0, \quad \mathbf{r}_0 = \mathbf{b}$

Compute $\tilde{\mathbf{r}}_0$ s.t. $\mathbf{P}\,\tilde{\mathbf{r}}_0 = \mathbf{r}_0$

Set $\mathbf{p}_0 = \tilde{\mathbf{r}}_0$

for $k = 1, 2, \dots, K$ **do**

Compute a step length $\alpha_k = \frac{\mathbf{r}_{k-1}^\top \tilde{\mathbf{r}}_{k-1}}{\mathbf{p}_{k-1}^\top \mathbf{N}\,\mathbf{p}_{k-1}}$ (where $\mathbf{P}\,\tilde{\mathbf{r}}_{k-1} = \mathbf{r}_{k-1}$)

Update the approximate solution $\mathbf{x}_k = \mathbf{x}_{k-1} + \alpha_k\,\mathbf{p}_{k-1}$

Update the residual $\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_k\,\mathbf{N}\,\mathbf{p}_{k-1}$

Solve $\mathbf{P}\,\tilde{\mathbf{r}}_k = \mathbf{r}_k$

Compute a gradient correction factor $\beta_k = \frac{\mathbf{r}_k^\top \tilde{\mathbf{r}}_k}{\mathbf{r}_{k-1}^\top \tilde{\mathbf{r}}_{k-1}}$ (where $\mathbf{P}\,\tilde{\mathbf{r}}_{k-1} = \mathbf{r}_{k-1}$)

Set the new search direction $\mathbf{p}_k = \tilde{\mathbf{r}}_k + \beta_k\,\mathbf{p}_{k-1}$ (where $\mathbf{P}\,\tilde{\mathbf{r}}_k = \mathbf{r}_k$)

end for

5.2.3 Chebyshev algorithm

Approximating a function applied to a matrix can be performed in three steps using a Chebyshev algorithm (Fox and Parker, 1972; Higham, 2008). First, an interval $[a, b]$ that contains all the eigenvalues of the matrix is defined. Then a polynomial approximation of the function over $[a, b]$ is derived using Chebyshev sums (Chebyshev, 1853; Clenshaw, 1955). Finally, the matrix operation is applied to this polynomial instead of the original function, using an iterative method that only involves matrix-vector products. The first step is done by considering the following property for symmetric matrices.

Theorem 5.1. *(Gerschgorin circle theorem (Gerschgorin, 1931)) Any eigenvalue λ of a symmetric matrix $\mathbf{N} \in \mathbb{R}^n \times \mathbb{R}^n$ satisfies*

$$\lambda \in \bigcup_{i \in [1:n]} [N_{ii} - r_i, N_{ii} + r_i],$$

where $r_i = \sum_{i \neq j} |N_{ij}|$. Hence, all the eigenvalues of $\mathbf{N}$ are contained in the interval

$$\left[\min_{i \in [1,n]} (N_{ii} - r_i), \max_{i \in [1,n]} (N_{ii} + r_i) \right], \quad i \in [1, n], j \in [1, n].$$

If $\mathbf{N}$ is positive (semi)definite, then $a = 0$ is a lower bound of its eigenvalues. Consequently, the interval for the eigenvalues of $\mathbf{Q} + \sigma_0^{-2} \mathbf{A}\mathbf{A}^\top$ can be taken as $\left[0, \max_{i \in [1,n]} \sum_{j=1}^{n} |(\mathbf{Q} + \sigma_0^{-2} \mathbf{A}\mathbf{A}^\top)_{ij}|\right]$ (if the diagonal elements of $\mathbf{Q} + \sigma_0^{-2} \mathbf{A}\mathbf{A}^\top$ are non-negative). A different lower bound must be found for the functions that are not defined at 0, such as the logarithm. In our case, we will set a lower bound close to 0, i.e., $a = 0.001$. For all the tests where we could compute the eigenvalues directly, this bound was found to be adapted.

The second step, consisting in finding a polynomial approximation of the function, is obtained by computing the Chebyshev sum of the function. The Chebyshev polynomials of the first kind are a set of orthogonal polynomials defined as the solutions to the Chebyshev differential equation and denoted $T_n(x)$. They are obtained from the recurrence relation

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_{n+1}(x) &= 2x\, T_n(x) - T_{n-1}(x). \end{aligned} \tag{5.9}$$

The Chebyshev polynomials are important in approximation theory because the roots of $T_n(x)$, which are also called Chebyshev nodes, are used as matching points for optimizing polynomial interpolation. The resulting interpolation polynomial provides an approximation that is close to the best polynomial approximation to a continuous function under the maximum norm.

The last step of the algorithm is outlined in Algorithm 3. This approach can be considered as a matrix-free algorithm. Indeed, it does not actually require the operator $\mathbf{N}$ to be stored in memory. Rather, it relies only on the product between the operator and vectors. This property is clearly desirable in a context where the size of the vectors and matrices may be so large that any gain in memory is appreciated.

Algorithm 3 Chebyshev algorithm

Require: Matrix $\mathbf{N} \in \mathbb{R}^n \times \mathbb{R}^n$. Vector $\mathbf{x} \in \mathbb{R}^n$. A set of Chebyshev coefficients $c_0, \dots, c_m \in \mathbb{R}$ for the function $h(\mathbf{N})$.

Ensure: The product $\mathbf{y} = P_h(\mathbf{N})\,\mathbf{x} = \left(\frac{1}{2}c_0 T_0(\mathbf{N}) + \sum_{k=1}^{m} c_k T_k(\mathbf{N})\right)\mathbf{x}$, approximation of $h(\mathbf{N})\,\mathbf{x}$.

Set $\mathbf{u}^{(-2)} = \mathbf{u}^{(-1)} = \mathbf{u} = \mathbf{y} = \mathbf{0}$
$P = \frac{2}{a-b}\,\mathbf{N} - \frac{b+a}{b-a}\,\mathbf{I}$
Set $\mathbf{u}^{(-2)} = \mathbf{x}$
Set $\mathbf{u}^{(-1)} = P\,\mathbf{x}$
Set $\mathbf{y} = \frac{1}{2}\,\mathbf{u}^{(-2)} + c_1\,\mathbf{u}^{(-1)}$
for $k = 2, \dots, m$ **do**
  $\mathbf{u} = P\,\mathbf{u}^{(-1)} - \mathbf{u}^{(-2)}$
  $\mathbf{y} = \mathbf{y} + c_k\,\mathbf{u}$
  $\mathbf{u}^{(-2)} = \mathbf{u}^{(-1)}$
  $\mathbf{u} = \mathbf{u}^{(-1)}$
end for

This algorithm is used to approximately compute $\log\left(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top\right)$. In fact, $\log\left(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top\right)$ needs to be computed by spectral decomposition of $\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top = \mathbf{U}^\top\,\mathbf{\Lambda}\,\mathbf{U}$, which is computationally expensive for large matrices. We know that $f(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top) = \mathbf{U}^\top\,f(\mathbf{\Lambda})\,\mathbf{U}$, but, when f is a polynomial such that $f(x) = P(x) = \sum_{i=1}^{p} a_i x^i$, the function applied to the matrix $\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top$ becomes

$$P(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top) = \sum_{i=1}^{p} a_i (\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)^i.$$

For this reason, it is useful to approximate the function $f(\cdot) = \log(\cdot)$ by a polynomial P on an interval that contains all the eigenvalues $\{\lambda_i\}_{i=1}^{N_S N_T}$ of $\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top$. The chosen interval $[a, b]$ must be positive. More details are available in Pereira and Desassis (2019).

5.2.4 Hutchinson's estimator

The Hutchinson's estimator is an unbiased estimator for the trace of a matrix (Hutchinson, 1990). Consider the term $\mathrm{tr}(\mathbf{N})$. This can be approximated with the Hutchinson's

estimator as

$$\mathrm{tr}\,(\mathbf{N}) \approx \frac{1}{m}\sum_{j=1}^{m} \mathbf{z}_j^\top\, \mathbf{N}\, \mathbf{z}_j, \tag{5.10}$$

where $\mathbf{z}_j$ is a vector of independent Gaussian random elements or $\mathbf{z}_j = 2\mathcal{I}(\mathbf{u} > 0.5) - 1$ for $\mathbf{u} \sim \mathcal{U}(0,1)$ such that $\mathbf{z}_j$ is a vector of independent random elements -1 or 1 of equal probability (the Rademacher distribution). The approximation gets better as m increases, but even a few number of terms in the sum allows finding a good approximation of the trace.

This result implies that we do not have to compute the eigenvalues of the matrix $\mathbf{N}$ when it is computationally demanding, but we can only use matrix-vector multiplication between $\mathbf{N}$ and the vectors $\mathbf{z}_j$. In our case, this approximation is used to approximately compute $\mathrm{tr}[P(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)]$.

5.3 Prediction by Kriging

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

In the space-time interpolation setting, the spatio-temporal vector $\mathbf{x}_{1:N_T}$ is predicted on the entire spatial mesh during the time window $[1, T]$, i.e. on $\mathcal{T}'$, using the data $\mathbf{y}_{1:N_T}$ defined in Equation (5.1). The kriging predictor is directly read from Equation (5.5):

$$\mathbf{x}^\star_{1:N_T} = \mathbb{E}(\mathbf{x}_{1:N_T} \mid \mathbf{y}_{1:N_T}) = \sigma_0^{-2}(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)^{-1}\,\mathbf{A}(\mathbf{y}_{1:N_T} - \boldsymbol{\eta}\widehat{\mathbf{b}}). \tag{5.11}$$

The computation of (5.11) requires the resolution of a system with matrix $(\mathbf{Q} + \sigma_0^{-2}\,\mathbf{A}\,\mathbf{A}^\top)$, as detailed in Section 5.1. The conditional variance, also called kriging variance, is

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

The computation of the diagonal of an inverse matrix is not straightforward when only the Cholesky decomposition of the matrix is available. Among the existing methods there is the Takahashi recursive algorithm described in Takahashi et al. (1973) and Erisman and Tinney (1975). Another way of computing the kriging variance is through conditional simulations, as detailed in Section 5.4.

In the temporal extrapolation setting, the vector $\mathbf{x}^{(N_T+1)}$ is predicted at time step $(N_T + 1)$ on $\mathcal{T}$ using all the data available until time T, i.e. from $\mathbf{y}_{1:N_T}$. Following Equation (3.20), we have

$$\mathbf{x}^{(N_T+1)} = \mathbf{D}\,\mathbf{x}^{(N_T)} + \mathbf{E}\,\mathbf{z}^{(N_T+1)}, \tag{5.12}$$

where $\mathbf{z}^{(N_T+1)}$ is a standardized Gaussian vector. The kriging predictor $\mathbf{x}^{\star(N_T+1)}$ is

$$\mathbf{x}^{\star(N_T+1)} = \mathbb{E}(\mathbf{x}^{(N_T+1)} \mid \mathbf{y}_{1:N_T}) = \mathbf{D}\,\mathbb{E}(\mathbf{x}^{(N_T)} \mid \mathbf{y}_{1:N_T}) = \mathbf{D}\,\mathbf{x}^{\star(N_T)}, \quad (5.13)$$

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

5.4 Conditional simulations

To perform a conditional simulation, we use the conditional kriging paradigm presented below. This approach relies on the fact that kriging predictors and kriging residuals are uncorrelated (independent under Gaussian assumption, see Chilès and Delfiner (1999, Chapter 7)). First, a non-conditional simulation $\mathbf{x}^{(NC)}_{1:N_T}$ is performed on the spatio-temporal grid $\mathcal{T}'$. From this simulation, kriging residuals

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

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

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

Conditional simulations at further time steps are obtained by iteratively computing $\mathbf{x}^{(C)}_{N_T+k}$ using the propagation equation (5.12) with $k \geq 1$. Multiple independent realizations of conditional simulations can then be used to compute approximate conditional variances or other quantities, such as probability maps of threshold exceedance.

5.5 Simulation study

Several tests will be carried out to validate the estimation technique with both methods (Cholesky and matrix-free).

We report here some results regarding the estimation of the parameters $\boldsymbol{\theta}^\top = [\kappa, \gamma_1, \gamma_2, c, \tau]$ for a spatio-temporal model simulated with the SPDE (2.27). We set $\mathbf{H} = \mathbf{I}$, $\alpha = 1$ and $\alpha_S = 2$. The spatial domain is the $[0, 30]^2$ square with a grid triangulation of $N_S = 900$ spatial points. The time window is $[1, 10]$ with unit time step and $N_T = 10$. The $n_S = 100$ observations are randomly located into the spatial domain and their position do not change during the N_T time steps (hence $n = 1000$). Since the sizes of both the dataset and the spatio-temporal mesh are reasonable, we report the estimations computed with both the Cholesky decomposition approach and the matrix-free approach.

As initial values, we use estimated values obtained from the variograms of the spatial and temporal traces of the process. Specifically, the initial value for κ is the estimated scale parameter of a Matérn covariance function with smoothness parameter $\nu = \alpha + \alpha_S - 1 = 2$ considering independent temporal repetitions, the initial value for c is deduced from the estimated parameter of n_S independent repetitions of AR(1) processes of length N_T and τ^2 is computed from Equation (2.29) with σ^2 being the empirical variance computed on the data. Finally, the initial value for $\boldsymbol{\gamma}$ is the null vector. The parameters for the matrix-free approach are set to the following: the order of the Chebychev polynomial to approximate the logarithm is set to 30 and the number of terms in the sum of the Hutchinson's estimator is set to 10. The results are reported in Table 5.1. They show that all parameters are accurately estimated with both the approaches. In almost all cases, the true value of the parameter is within the mean $\pm$ 2 standard deviations interval. We remark how the matrix-free approach takes more time to estimate the parameters. This is due to the iterative computations, that increase the computational time. However, we know that the benefit of the matrix-free approach is the possibility of applying it to much larger spatio-temporal meshes, where the Cholesky decomposition cannot be applied at all.

method	κ	γ_1	γ_2	c	τ	average time (s)
correct	0.5	2	2	1	1	
Cholesky	0.610 (0.047)	2.354 (0.515)	2.325 (0.421)	1.037 (0.218)	1.072 (0.040)	194
Matrix-free	0.483 (0.029)	1.904 (0.178)	1.906 (0.147)	1.027 (0.046)	0.998 (0.024)	960
correct	0.7	1	-1	2	0.5	
Cholesky	0.695 (0.067)	1.056 (0.659)	-1.134 (0.631)	2.090 (0.352)	0.503 (0.022)	172
Matrix-free	0.669 (0.052)	0.967 (0.123)	-1.164 (0.142)	1.954 (0.118)	0.485 (0.013)	863

TABLE 5.1: Mean (and standard deviation) of ML estimates $\hat{\boldsymbol{\theta}}^\top = [\hat{\kappa}, \hat{\gamma}_1, \hat{\gamma}_2, \hat{c}, \hat{\tau}]$ over 50 simulations for two different subsets of advection-diffusion model parameters and two different estimation approaches (Cholesky decomposition and matrix-free approach).

Chapter 6

Applications

Résumé

Dans ce chapitre nous présentons une application de la méthode aux EDPS d'advection-diffusion détaillée dans les chapitres précedents pour la prédiction spatio-temporelle à court terme de données de radiation solaire et de vitesse de vent. Nous montrons la flexibilité de la méthode et son interprétation statistique et physique.

6.1 Spatio-temporal statistics for environmental data

Environmental data are usually characterized by spatial, temporal, and spatio-temporal correlations. Capturing these dependencies is extremely important to model the underlying phenomena precisely. We propose to apply the SPDE methodological framework developed in the previous chapters to predict a spatio-temporal field on a fine mesh using spatially irregularly spatial data measured at regular time steps. Our approach should be able to reduce the computational demand of geostatistical methods without giving up a precise description of the spatio-temporal dependencies.

We will consider two different datasets: first, a solar radiation dataset of few spatial data (99 stations) sparse over a pretty small region (160 km^2) measured every minute; then, a wind speed dataset of 53333 spatial data measured every hour coming from a numerical simulation over the entire territory of Saudi Arabia (2.15 million km^2). In both cases, the objective is the short-term forecast at 1,2 and 3 time steps ahead on a fine spatial grid.

6.2 Application to a solar radiation dataset

The approach detailed in the previous chapters is now applied to a solar radiation dataset for which experts agree on the presence of advection due to Western prevailing winds transporting clouds from one side of the domain to the other. The HOPE campaign (Macke et al., 2017) recorded Global Horizontal Irradiance (GHI) (also called SSI, Surface Solar Irradiance) over a 10×16 km^2 region in West Germany near the city of Jülich from April 2 to July 2, 2013. The sensors were located at 99 stations located as pictured in Figure 6.1 and GHI was recorded every 15 seconds. A detailed description of the campaign can be found in Macke et al. (2017).

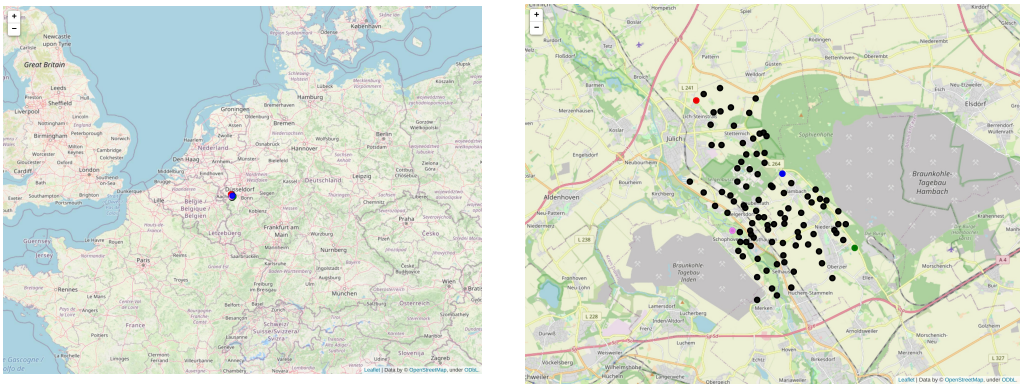


FIGURE 6.1: Stations capturing GHI over the spatial domain.

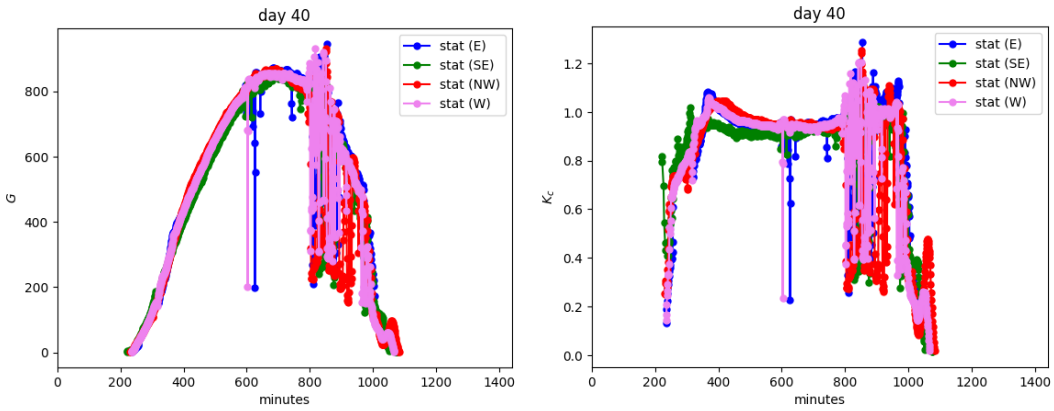


FIGURE 6.2: GHI G and Clear Sky Index K_c for 4 different stations on May 28, 2013.

The dataset was cleaned for outlying values and non-operating sensors, and the temporal resolution was reduced from 15 seconds to 1 minute. Figure 6.2 (left panel) shows GHI as a function of time (in minute, during a full day – May 28, 2013) at 4 different stations. These stations, represented in color in Figure 6.1, are located at the border of the domain, far from each other. The GHI starts close to 0, increases after sunrise, peaks at midday and tends to 0 at sunset. The maximal theoretical amount of irradiance reaching the sensor follows an ideal concave curve. The divergence between the measured irradiance and the optimal curve can be slight or important, depending on the presence of clouds. One can see on this example that the evolution among the 4 stations is similar, with variations accounting for spatio-temporal variations of the clouds.

A first preprocessing was made in order to stationarize the phenomenon, in collaboration with researchers from the OIE Center of Mines Paris, who made the datasets and their preprocessing codes available for our work.

Oumbe et al. (2014) showed that the solar irradiance at ground level, GHI (denoted G for short from now on), computed by a radiative transfer model can be approximated by the product of the irradiance under clear atmosphere (called Clear Sky GHI, or G_c) and a modification factor due to cloud properties and ground albedo

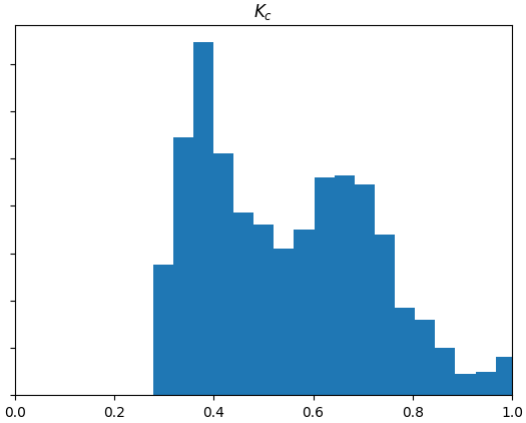


FIGURE 6.3: Histogram of K_c over 20 time steps.

only (Clear Sky Index, or K_c, Beyer et al. (1996)):

$$G \simeq G_c K_c. \tag{6.1}$$

The error made using this approximation depends mostly on the solar zenith angle, the ground albedo and the cloud optical depth. In most cases, the maximum errors (95th percentile) on global and direct surface irradiances are less than 15 Wm^{-2} and less than 2 to 5 % in relative value, as recommended by the World Meteorological Organization for high-quality measurements of the solar irradiance (Oumbe et al., 2014). Practically, it means that a model for fast calculation of surface solar irradiance may be separated into two distinct and independent models: a deterministic model for G, under clear-sky conditions, as computed according to Gschwind et al. (2019), considered as known in this study, and a model for K_c, which accounts for cloud influence on the downwelling radiation and is expected to change in time and space. K_c is modeled as a random spatio-temporal process and will be the subject of our analysis. Figure 6.2 (right panel) shows the variable K_c corresponding to the variable G shown on the left panel. In general, K_c lies between 0 and 1, but in rare occasions, values above 1 can be observed. This phenomenon is called *overshooting* (Schade et al., 2007) and is due to light reflection by surrounding clouds.

A time window of 20 minutes around 4 p.m. on May 28, 2013 is extracted with observations every minute at the 73 stations with well recorded values. Parameters are estimated on this 20-minute window using the method described in Section 5.1. The spatio-temporal grid contains $N_T = 20$ one-minute time steps, from $t = 1$ to $t = 20$ and $N_S = 900$ regular spatial mesh points.

6.2.1 Estimation and prediction

Six different models are fitted to the data and used for prediction: 3 models with advection (called "adv-diff") and 3 models without advection (called "diff") obtained by setting $\boldsymbol{\gamma} = \mathbf{0}$. Both groups contain the three following sub-models: (i) a model with diffusion included only in the stochastic forcing term, with a Matérn spatial trace

with $\nu = 1$; (ii) a nonseparable model whose spatial trace has no known closed form expression for the covariance function (the closed form exists only if $\alpha + \alpha_S > d/2$, where $d = 2$ in this case, see Proposition 2.1); (iii) a nonseparable model with a Matérn spatial trace with $\nu = 2$. In the general model of Equation (2.27) they correspond respectively to $(\alpha, \alpha_S) = (0, 2), (1, 0), (1, 2)$. The parameters of the SPDE are estimated for each model separately. The results are reported in Table 6.1.

The log-likelihoods of the models that include advection are within a range of variations of 10 log-likelihood units and are between 34 and 80 units larger than those with diffusion only. As a point of comparison, if all spatio-temporal dependencies were ignored, the BIC penalization for the advection parameters would be $2\ln(1460) \simeq 14.5$. These results indicate strong evidence in favor of models with advection, but no significant differences among them. The parameters vary substantially from one model to the other, but it must be remembered that, when considered independently, their physical interpretation is model dependent. Some combinations are interpretable however. For example, following Proposition 2.1, the overall variance is equal to $(8\pi)^{-1}\tau^2\kappa^{-2}$ (or $(8\pi)^{-1}\tau^2\kappa^{-2}|\mathbf{I}+h\|\boldsymbol{\gamma}\|^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top|^{-1/2}$ in the stabilized case) when $\alpha_{tot} = \alpha + \alpha_S = 2$ and it is equal to $(16\pi)^{-1}\tau^2\kappa^{-4}$ (or $(16\pi)^{-1}\tau^2\kappa^{-4}|\mathbf{I}+h\|\boldsymbol{\gamma}\|^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^\top|^{-1/2}$) when $\alpha_{tot} = 3$. Accordingly, the estimated standard deviations for models (1), (3), (4) and (6) are equal to 0.160, 0.119, 0.174 and 0.199 respectively, with the experimental standard deviation being equal to 0.184. The advection-diffusion models (1) and (3) are advection-dominated, hence the formula for the variance is the stabilized one. For the same models, the practical ranges computed as $\sqrt{8\nu}/\kappa$ (Lindgren et al., 2011) are equal to 1.915, 3.079, 2.281 and 4.255 respectively. Notice that among pairs of models that differ by the presence or absence of advection, the estimated range is larger for those without advection in an attempt to account for the larger correlation distance due to transport.

	Model	α	α_S	log-likelihood	$\hat{\kappa}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{c}$	$\hat{\tau}$	$\hat{\sigma}_0$	$\hat{b}$
(1)	adv-diff	0	2	2587	1.477	9.642	-5.382	11.659	2.254	0.052	0.570
(2)	adv-diff	1	0	2577	0.237	4.718	-0.928	9.315	0.458	0.045	0.598
(3)	adv-diff	1	2	2579	1.299	17.325	-8.442	41.017	3.072	0.058	0.574
(4)	diff	0	2	2507	1.240	0	0	12.558	1.081	0.059	0.569
(5)	diff	1	0	2545	0.246	0	0	6.594	0.436	0.047	0.577
(6)	diff	1	2	2512	0.940	0	0	34.607	1.248	0.064	0.580

TABLE 6.1: Estimated parameters and log-likelihood for 6 different models from all data on a 20-minute window.

We then perform prediction with two different validation settings containing 80% of conditioning data and 20% of validation data. In the first case (called "Uniform") the validation locations are uniformly randomly selected. In the second case (called "South-East") the validation locations are located downwind (i.e. South-East) with respect to the estimated advection direction. See Figure 6.4 for a representation of the validation settings.

Recall that a time window containing 20 time-steps, from $T = 1$ to $T = 20$,

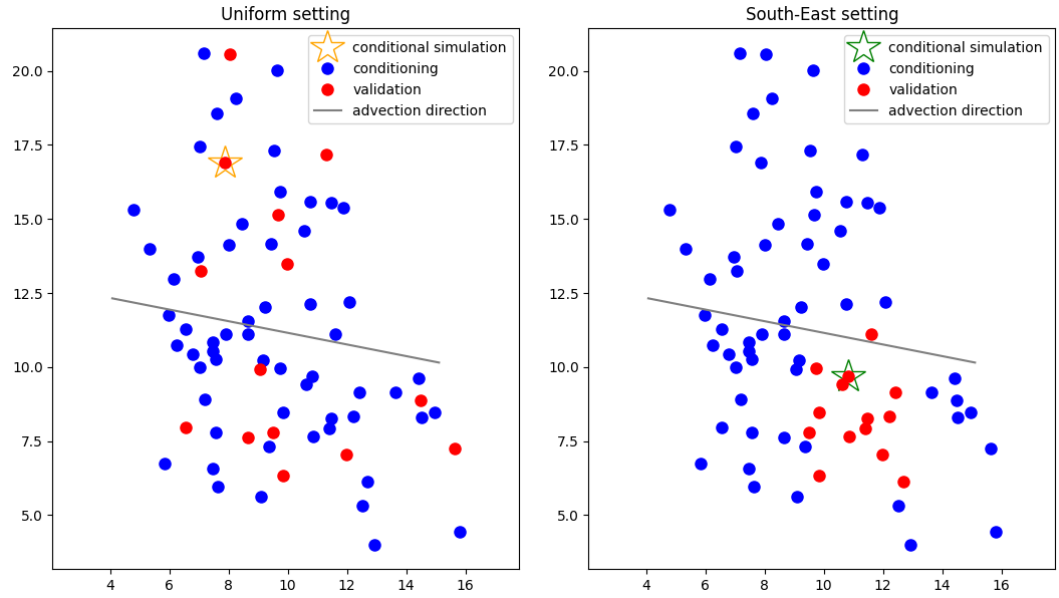


FIGURE 6.4: Validation settings: Uniform (left) and South-East (right).

has been selected. For each validation setting and from $T = 11$ to $T = 20$, three prediction configurations using conditioning data from time $(T-9)$ to T are computed and compared to the real values, allowing us to compute a Root Mean Square Error (RMSE) validation score. First, the kriging is performed spatially only (hereafter referred to as "S" kriging). Second, a temporal extrapolation is computed at the conditioning locations at time horizons $(T+1)$, $(T+2)$ and $(T+3)$ ("T1", "T2", "T3" kriging). Third, the spatio-temporal prediction is computed at the validation locations at time horizons $(T+1)$, $(T+2)$ and $(T+3)$ ("ST1", "ST2", "ST3" kriging). We thus have a total of 6 models $\times$ 2 validation settings $\times$ 3 prediction configurations. RMSEs are averaged over the 10 repetitions. Results are shown in Table 6.2.

For all tested validation settings and prediction configurations, the models with advection show better RMSE scores than models without advection. This result is a confirmation of the results already observed on log-likelihoods. Models with advection have similar prediction scores in the prediction configurations S and T, model (1) having slightly better performances in the configuration S. In the T and ST configurations, models (2) and (3) have in general quite similar RMSEs, except in the South-East setting with ST configuration where model (2) is clearly the best model. In this case, prediction is made in a space-time domain lying downstream with respect to the advection. It is thus expected that the model best representing the underlying physics should lead to the best prediction performances.

An example of prediction maps on $\mathcal{T}$ at time horizons $(T+1)$, $(T+2)$ and $(T+3)$ is reported in Figure 6.5, along with the observed values (black contoured dots). The white contoured dots are the locations used for validation in the Uniform setting.

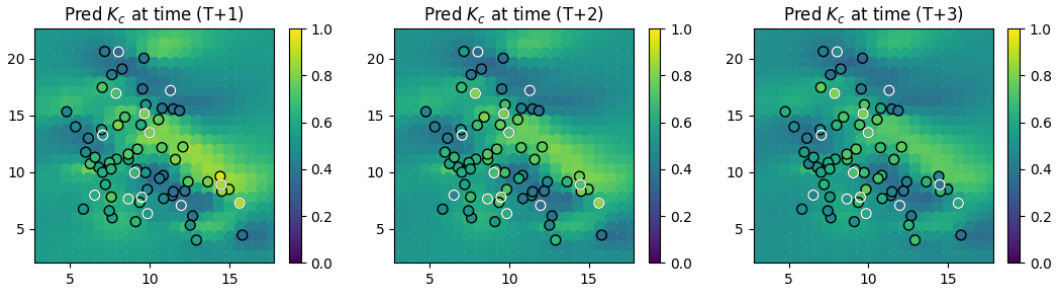


FIGURE 6.5: Predictions of K_c at $(T+1)$, $(T+2)$ and $(T+3)$ with model (3) ("adv-diff" with $\alpha = 1$ and $\alpha_S = 2$). The black contoured dots are the conditioning locations and the white contoured dots are the validation locations for the Uniform setting.

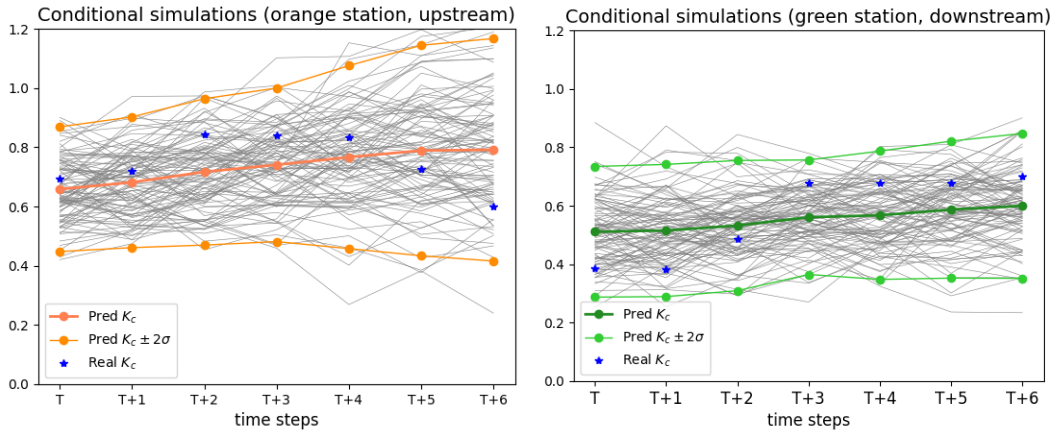


FIGURE 6.6: Real K_c, mean of conditional simulations of K_c and $\pm 2\sigma$ envelope at time horizons $T, (T+1), (T+2), \ldots, (T+6)$. Left: orange station. Right: green station.

6.2.2 Conditional simulations

Figure 6.6 shows 100 conditional simulations of K_c computed at time $T = 11$ and horizons $(T+1), (T+2), \ldots, (T+6)$ with the advection-diffusion model (3). Two validation stations have been selected: one in the North-West part of the domain (the orange star in the left panel of Figure 6.4) and one in the South-East part of the domain (the green star in the right panel of Figure 6.4). Given that there is an advection from North-West to South-East, it is therefore expected that the advection-diffusion model should be able to transport the information. The mean of the 100 simulations and the envelopes corresponding to twice the pointwise standard deviation have also been represented, along with the true values. As expected, most of the conditional simulations lie within the envelopes in both cases and at all time horizons. The remarkable result is that the variance of the conditional simulations at the green station is smaller than that at the orange one at every time step, especially when the time horizon increases. This is due to the advection term in model (3), able to propagate information from North-West to South-East.

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

TABLE 6.2: Averaged RMSE computed at 10 successive time steps for 6 different models, 2 validation settings (Uniform and South-East) and 3 prediction configurations (S, T and ST); see text for details. In each case, the best score among the models is in bold font.

6.2.3 Discussion

A reliable and fast short-term prediction of solar radiation is not only interesting for environmental research, but also for the management of the electrical network in the

context of the increase of renewable energies in the energy mix. The evolution of solar radiation during a short period (from one minute to one hour) is essentially generated by the movement of clouds on the considered spatial domain. This phenomenon is governed by advection and diffusion processes. We therefore took this phenomenon into account by our statistical approach based on physical SPDEs.

On the HOPE dataset, we have shown that the advection-diffusion model developed in the PhD presented better prediction performances compared to simpler models, while remaining computationally fast. An in-depth analysis of the HOPE solar radiation dataset, along with the OIE research team of Mines Paris, must be undertaken. We conceive possible transformations of the dataset, which is not perfectly Gaussian, or modifications of our current approach, in order to adapt the spatio-temporal SPDE approach to this specific dataset. This work will eventually be left to a new PhD project.

6.3 Application to a wind speed dataset

The SPDE approach is now applied to a wind speed dataset, already analyzed by means of several prediction techniques in Huang et al. (2022). The choice of this dataset is due to a collaboration with Marc Genton and the Spatio-Temporal Statistics and Data Science (STSDS) research group of the King Abdullah University of Science and Technology (KAUST) during a 5-week-long scientific visit there in October-November 2022.

The data are hourly high-resolution simulations from the Weather Research and Forecasting model (WRF; Skamarock et al. (2008)) over the territory of Saudi Arabia. WRF was simulated from 2013 to 2016 by Giani et al. (2020). The hourly wind field on a 2.15 million km^2 wide country is extremely difficult to model, due especially to the nonlinear dynamics implied by the topographic characteristics of the spatial domain (two coastal areas, mountain ranges and sandy and rocky deserts) and by the massive amount of data in space and time. Standard statistical approaches such as the autoregressive integrated moving average (ARIMA) model are effective in capturing a short-range linear dependence; machine learning approaches, such as artificial neural networks, are more adapted to capture nonlinear dynamics because of their flexibility and the possibility of defining a recursive structure (recurrent neural networks, RNNs) for temporal data. The downside of RNNs is that their inference is very challenging, due to the high dimensionality of the parameter space and the impossibility to directly apply the iterative gradient computations of traditional neural networks. To reduce the computational cost of the inference, Jaeger (2001) proposed the echo state networks (ESNs), which is grounded on the use of sparse, random matrices instead of the dense unknown ones of RNNs. In ESNs the inputs of the neural networks are mapped to high-dimensional hidden states through fixed nonlinear dynamics. These hidden states consider sequential linkages and thus allow for a nonlinear transformation of the input history. In order to reduce the spatial dimensionality of the ESNs

by preserving the spatial variability, one could use a dimensionality reduction technique, such as the empirical orthogonal function (EOF) method (grounded on the eigenvectors of the spatial covariance matrix of the space-time data) or the principal component analysis (PCA). Another possibility to reduce the spatial dimension is to consider only a small collection of knots and to recover the entire field by spatial interpolation.

Huang et al. (2022) proposed a dimensionally reduced ESN model, describing the dynamics of a spatial field through a set of knots, some of which are sampled on a fix grid and others sampled in areas with complex patterns. The spatio-temporal forecast was then obtained by using the ESN for temporal forecast and a spatial interpolation approach for spatial prediction. To reflect the spatially varying dependence structure dictated by the different topographical features in Saudi Arabia, they used a nonstationary Matérn covariance function (Paciorek and Schervish, 2006) fitted to data with the R package `convoSPAT` (Risser and Calder, 2017).

The model was compared to other approaches, including ARIMA, EOF-based ESN, Fixed Rank Kriging (FRK) (Cressie and Johannesson, 2008; Zammit-Mangion and Cressie, 2021) and the persistence forecast, i.e., the field where the values at each spatial location at 1,2 and 3 steps ahead are set to the last known values at the same location. The ESN method proposed by Huang et al. (2022) was found to have the best prediction score (MSE) among the analyzed methods at 2 and 3 steps ahead, even though it takes into account only a subset of observations.

Here, we will apply the SPDE approach on a regular triangulation of $N_S = 40^2 = 1600$ spatial knots by using the entire set of observations as known data. These are $n_S = 53333$ values. The predictions will be "short-term", at 3 steps ahead, i.e., 3 hours ahead. We do not pretend to obtain better performances than the ESN proposed by Huang et al. (2022), since Huang et al. (2022) already proved that their deep learning approach is capable of capturing the nonlinear temporal dynamics of the dataset. We only want to illustrate how the spatio-temporal advection-diffusion SPDE model works on an environmental dataset to point out the advantages and disadvantages of the methodology. For this reason, and also for computational reasons, we will restrict the temporal domain to a day, i.e., 24 temporal measurements (in particular, January 2, 2016).

We refer to Huang et al. (2022, Section 3) for the preprocessing of the dataset. The dynamics of the square root of the wind speed is described in his deterministic part by a harmonics depending on the spatial locations and in his random part by the multiplication of a spatially varying scaling coefficient by a residual spatio-temporal random field $Y(\mathbf{s}, t)$, which has unit variance at each location. The square root applied to the wind speed transforms the right-skewed distribution of the wind speed to normality (Gneiting, 2002). In this work we treat the spatio-temporal random field $Y(\mathbf{s}, t)$ as defined in Equation (1.1) and apply the estimation and prediction approach developed in Chapter 5.

In Figure 6.7 (left panel) we plot the entire set of observations over the spatial

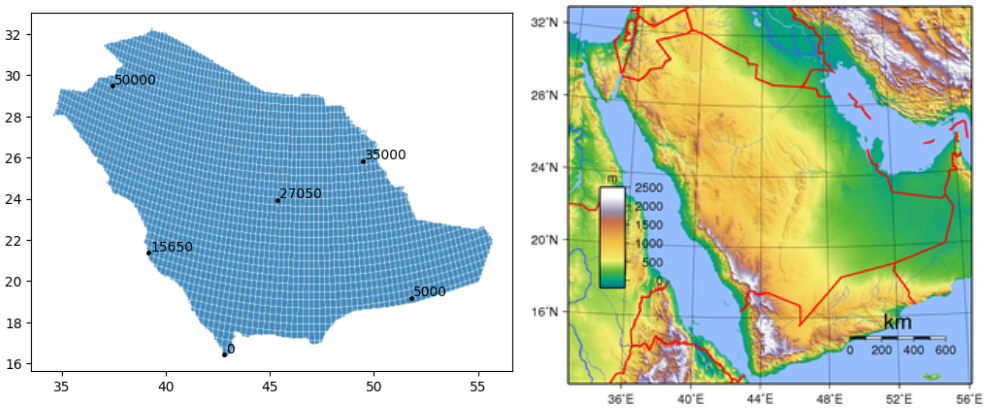


FIGURE 6.7: Left: spatial locations of WRF wind simulations over the entire spatial domain. Right: topography of Saudi Arabia.

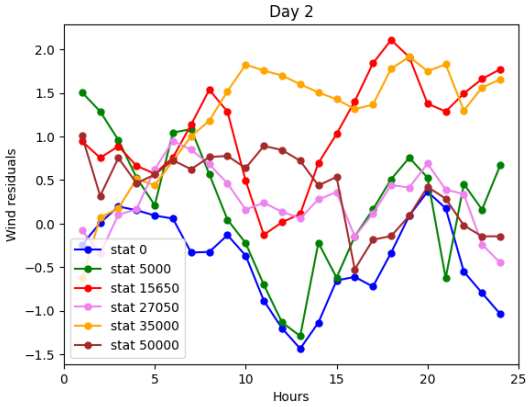


FIGURE 6.8: Wind residuals during January 2, 2016 at 6 different spatial locations.

domain, where we pick 6 locations in different areas of the domain to assess the behavior of the wind. The right panel shows the topography of Saudi Arabia. Figure 6.8 shows the wind residual at the 6 picked locations; we remark that the evolution over 24 hours changes depending on the spatial location, with some similar behaviors for stations not far from the west border of the country, i.e., the Red Sea coast (stations 0 (blue), 5000 (green) and 15650 (red)). Figure 6.9 shows the evolution of the wind residual over 24 hours during 4 following days for each of the 6 stations. It is quite hard to define if some behaviors occur repetitively at each station. It seems that station 15650 and station 35000 show a daily quasiperiodic behavior, with low and high values of wind residuals approximately at the same hours of the day every day. The other stations are much harder to analyze. It can be noted that when the wind residual is at its highest in the south of Saudi Arabia (station 5000 on day 3), it is not the case in the north of the country (station 50000 on day 3) and vice-versa.

6.3.1 Estimation and prediction

The parameters of the statistical model, including the parameters of the SPDE (2.27), are estimated at each time steps from the known values at the spatial observations

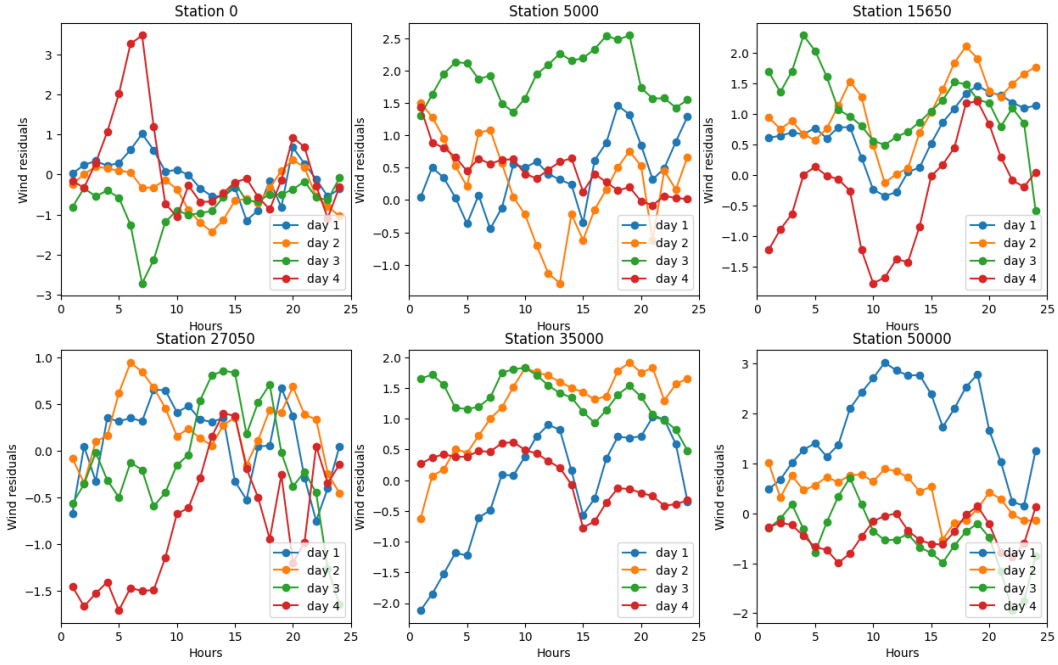


FIGURE 6.9: Wind residuals at 6 different spatial locations during January 1-4, 2016.

during the 5 hours before. The spatio-temporal predictions at 1,2 and 3 hours ahead are then obtained by kriging and by application of three Euler steps, as described in Section 5.3.

Different advection-diffusion SPDE models are tested and compared to the persistence model (where $\mathbf{y}^{(T+3)} = \mathbf{y}^{(T+2)} = \mathbf{y}^{(T+1)} = \mathbf{y}^{(T)}$) and the FRK model predictions, obtained with the R package `FRK`. Fixed Rank Kriging (Cressie and Johannesson, 2008; Zammit-Mangion and Cressie, 2021) is a scalable spatio-temporal modelling and prediction framework based on the use of a spatial random effects (SRE) model, in which a zero-mean random process is decomposed using a linear combination of spatio-temporal basis functions with random weights plus a term that captures the random process fine-scale variations. The number of basis functions defines the dimensionality reduction and the computational efficiency of the prediction. This method results in a nonstationary spatio-temporal field, which is a good asset with respect to the stationary SPDE model.

Table 6.3 reports the MSE of the different models. The results show a better performance of all the SPDE models with respect to the persistence prediction for 2 and 3 hours ahead, and similar performances to FRK. The best SPDE model are models (2) and (3), i.e., the advection-diffusion models that result in nonseparable spatio-temporal fields.

Figure 6.10 shows the forecasted wind residuals at $(T+1)$, $(T+2)$ and $(T+3)$ with model (3) and Figure 6.11 shows the difference between the MSE of model (3) and the MSE of the persistence model. The main remark is that model (3) is not good at predicting the value of the wind residual in a region in the South-West part

of the country. This area is characterized by high mountains (see right panel of Figure 6.7); this topographic feature results in a stronger wind with respect to other regions, but the SPDE model, which is stationary, cannot really capture this behavior. A nonstationary model would certainly have better predictive performances in this specific area. We refer to Chapter 7 for some first results about the nonstationary extension of the spatio-temporal SPDE approach.

We know that the SPDE model performance depends on the number of mesh points used in the triangulation, and on their position. We thus perform a test with a higher number of mesh points ($N_S = 50^2$ and $N_S = 60^2$), always on a regular grid. We remark in Table 6.4 that the MSE decreases with the increase of N_S and we presume that the function is decreasing with the number of mesh points. Since we know that the topographic features influence the spatial dependence structure, we decide to test a custom-made mesh, made of $N_S = 3273$, where the majority of the points are located on the mountain areas of the map (see Figure 6.13). The results of the estimation with this irregular mesh gives the best predictive results for $(T+1)$ and $(T+2)$. This improvement can be observed especially in the mountain areas by looking at the map of the difference between the MSEs of the SPDE model and the persistence model in Figure 6.12. The yellow areas clearly visible in Figure 6.11 are now much less evident.

	Model	α	α_S	T+1	T+2	T+3
(1)	adv-diff	0	2	0.312	0.526	0.738
(2)	adv-diff	1	0	0.309	0.521	0.719
(3)	adv-diff	1	2	0.311	**0.519**	0.724
(4)	diff	0	2	0.312	0.533	0.749
(5)	diff	1	0	0.310	0.528	0.738
(6)	diff	1	2	0.314	0.528	0.734
(7)	persistence			**0.244**	0.544	0.806
(8)	FRK			0.420	0.546	**0.696**

TABLE 6.3: Average MSE at (T+1), (T+2), (T+3) from $n_S = 53333$ observations during 24 hours. Persistence, FRK, and various SPDE models (2.27) with different combinations of α and α_S, with or without advection, and $N_S = 40^2$. The best score among the models is in bold font.

Model	α	α_S	N_S	T+1	T+2	T+3
adv-diff	1	2	40^2	0.311	0.519	0.724
adv-diff	1	2	50^2	0.293	0.512	0.723
adv-diff	1	2	60^2	0.281	0.507	**0.721**
adv-diff	1	2	3273 (custom-made)	**0.256**	**0.506**	0.732

TABLE 6.4: Average MSE at (T+1), (T+2), (T+3) from $n_S = 53333$ observations during 24 hours. Advection-diffusion SPDE model (2.27) with $\alpha = 1$ and $\alpha_S = 2$, and $N_S = 40^2$, $N_S = 50^2$, $N_S = 60^2$ and custom-made mesh. The best score among the different meshes is in bold font.

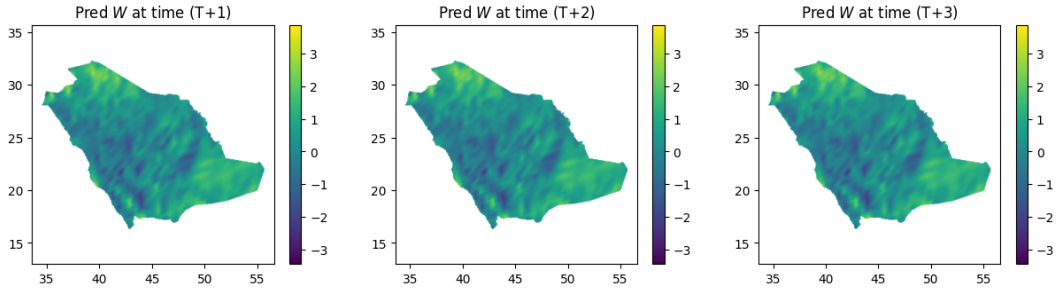


FIGURE 6.10: Predictions of W at $(T+1)$, $(T+2)$ and $(T+3)$ with model (3) ("adv-diff" with $\alpha = 1$ and $\alpha_S = 2$).

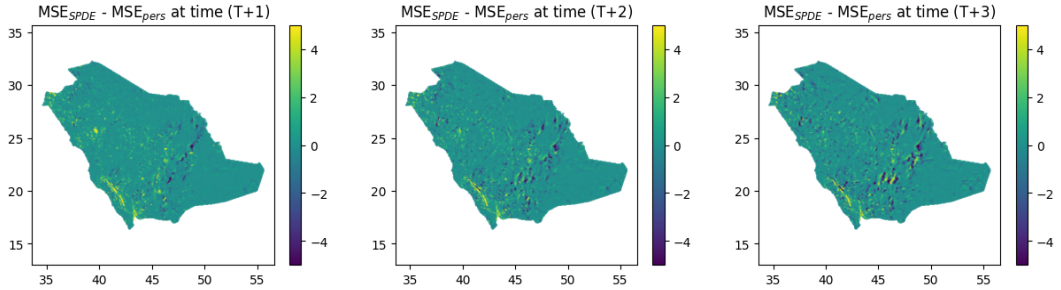


FIGURE 6.11: Difference between MSE of model (3) (regular mesh with $N_S = 900$) and MSE of persistence model at $(T+1)$, $(T+2)$ and $(T+3)$.

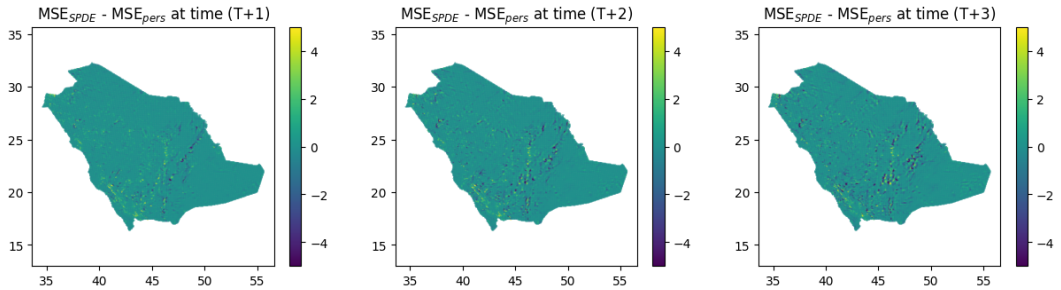


FIGURE 6.12: Difference between MSE of model (3) (handmade mesh with $N_S =$) and MSE of persistence model at $(T+1)$, $(T+2)$ and $(T+3)$.

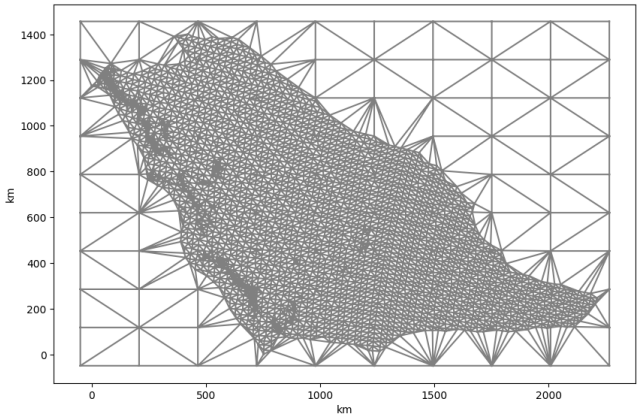


FIGURE 6.13: Handmade mesh to capture complex topographic areas (the spatial domain is given in $km \times km$).

6.4 Concluding remarks

The amount of environmental data has recently dramatically increased, making classical geostatistical approaches to fail in most cases (see the "big n problem", Section 1.4).

On the one hand, the generality of the SPDE approach we propose, and its adaptability to large datasets, favors its application to a wide class of environmental and geoscience data (wind, fine particle concentration, water resources, etc.). We pointed out the advantages of this method in terms of estimation and predictive performances on two different environmental datasets.

On the other hand, there have been several attempts to apply deep learning techniques to environmental problems. In the specific case of wind forecasting (Section 6.3), the predictive approach based on echo state networks proposed by Huang et al. (2022) seems to obtain better forecast performances. This is due especially to the fact that the complexity of the dataset resides mainly in the temporal dynamics, that deep neural networks usually capture well, and in the nonstationary spatial dependence structure, that the current advection-diffusion SPDE model cannot reproduce.

When using deep learning techniques, researchers have generally transposed methods from computer vision to environmental raster data, coming from satellites for earth observation or from climate models outputs. Nevertheless, especially at the local scales, it is preferable to work with direct ground measurements, i.e., on a set of irregular points in space. Indeed, ground data are more precise than satellite data and climate models output, generally with a higher temporal sampling frequency of the phenomena under study and without missing data. The direct adaptation of models coming from traditional deep learning is more complex in the case of spatio-temporal data collected over sparsely distributed measurement stations, as it is the case of the solar radiation dataset analyzed in Section 6.2.

The developments of the nonstationary SPDE approach (see Chapter 7) and of computationally faster techniques for estimation and prediction will probably lead to the possibility of applying the SPDE approach to more complex scenarios, both on the already studied datasets and on new datasets. For example, the comparison between the methods applied to the wind speed dataset would be more interesting when the nonstationary behavior will be captured also by the SPDE approach. Indeed, the SPDE approach should improve its predictive performances, especially in the areas with complex topographic characteristics. This extension will be left for future work.

Chapter 7

Generalization to nonstationary fields

Résumé

Dans ce chapitre, nous proposons une généralisation du modèle d'EDPS spatio-temporelle d'advection-diffusion pour les champs spatio-temporels non stationnaires. Nous passons d'abord en revue les techniques les plus utilisées pour traiter les non stationnarités en statistiques spatiales et proposons ensuite une manière d'étendre notre cadre méthodologique pour la simulation, l'estimation et la prédiction d'EDPS avec des paramètres variant dans l'espace et/ou dans le temps. Ce travail est encore en cours.

We here sketch a possible generalization of the unsteady advection-diffusion SPDE model to nonstationary spatio-temporal fields. We first review the most used techniques to tackle nonstationarities in the spatial statistics community and then propose a way of extending our methodological framework for simulation, estimation and prediction of SPDEs with spatially and/or temporally varying parameters. This work is still ongoing.

7.1 Nonstationarity: state of the art

In the *nonstationary* case, the covariance function can no longer be expressed as a simple function of the distance between the points as in Equation (1.2). Instead, it has an expression that depends on both the location and the relative position of the considered pair of points. The first challenge that appears is to determine the expression of this covariance function from the observed data. One of the ways to tackle this problem is by imposing that the random field can be modeled in a way that allows incorporating the prior structural information about the field. In general, the different methods try to derive an expression of the covariance function for any pairs of points in the domain. A detailed review of the methods used to model nonstationary random fields is available in Fouedjio (2017). We here sketch the more popular approaches in the case of a spatial field.

7.1.1 Space deformation

An early approach is the deformation method (Sampson and Guttorp, 1992). A nonstationary field observed on a spatial domain can be turned into a stationary one after applying a (nonlinear) deformation to the domain. The problem is thus reformulated in a stationary framework after deformation. The covariance function of the nonstationary process is written as the composition of a stationary covariance model with a deformation function. This approach has been further developed in Schmidt and O'Hagan (2003), Schmidt et al. (2011), Bornn et al. (2012) and Fouedjio et al. (2015). The characterization of the deformation function can be made through a set of parametrized functions which are fitted to the data by minimizing an objective function (Perrin and Senoussi, 2000). These approaches are computationally expensive, which limit their applicability for large-scale datasets. Moreover, they do not allow to easily take into account prior structural information about the nonstationarity. Indeed, they all seek to directly characterize the overall deformation while only considering the location and the value of the data points.

Zammit-Mangion et al. (2022) tackled the computational problem, by expressing spatial processes with nonstationary and anisotropic covariance structure as ones that have stationary and isotropic covariance structure on a warped spatial domain as in Sampson and Guttorp (1992), but modeling the injective warping function through a composition of multiple elemental injective functions in a deep-learning framework. They employed approximate Bayesian methods to make inference with these models using recent methodological and technological advances in deep learning and deep Gaussian processes.

7.1.2 Convolution model

Another approach is the kernel convolution method (Higdon et al., 1999; Paciorek and Schervish, 2006; Kleiber and Nychka, 2012; Fouedjio et al., 2016). The idea behind the method is that a nonstationary field at a given point of the spatial domain can be seen as the result of the spatial convolution over the domain of a deterministic function, called kernel function, with a white noise. The kernel functions defining the process covariance can easily be made spatially varying; they control the anisotropic behavior of the process in a small neighborhood of each point. With this method, even the smoothness of the spatial field can vary across space, which is not the case for the space deformation approach. The downside of this framework is that only the information of the anisotropy at each point is taken into account in their covariance, and not the overall structure of the anisotropy, which in practice might influence the covariance.

7.1.3 Dimension expansion

Bornn et al. (2012) proposed a different idea to model nonstationarity which assumed that the nonstationary process is a projection of a stationary process in higher dimensions. They justified their dimension augmentation procedure by stating that a realization of a nonstationary Gaussian process in $\mathbb{R}^d$ may be interpreted as a realization of a stationary field in $\mathbb{R}^{d+p}$ for $p > 0$ under appropriate moment constraints.

7.1.4 Computational complexity

All the methods mentioned above involve the covariances between any pair of locations. This means standard implementations are infeasible for large datasets. In general, increasing the degree of flexibility in the covariance structure requires increasing the number of parameters. The common isotropic Matérn covariance functions are parametrized through 3 parameters: marginal variance, range, and smoothness. Flexible models can have dozens or even hundreds of parameters. This implies that the estimation of these parameters can be hard and computationally cumbersome. There exist some ways to overcome these computational issues (Heaton et al., 2019). An appealing way to reduce dimensionality is to describe the covariance structure through covariates (Ingebrigtsen et al., 2014; Schmidt et al., 2011; Risser and Calder, 2017).

7.1.5 Spatio-temporal nonstationarity

Compared to the vast literature for modeling nonstationarity in space, nonstationarity in both space and time is studied much less extensively. Ma (2002) derived nonstationary space-time covariance functions by applying a spatially and temporally varying kernel to stationary covariance functions. Garg et al. (2021) extended the idea of convolution processes by Higdon et al. (1999) and Paciorek and Schervish (2006) to model a nonstationary spatio-temporal Gaussian process. Set in a Bayesian framework, Sigrist et al. (2012) proposed a dynamic nonstationary spatio-temporal model for short term precipitation by linking the advection parameter, describing the horizontal transport of rainfall, of the convolution kernel to an external wind vector, measured independently of precipitation. Huang and Hsu (2004) extended Wikle and Cressie (1999) to develop a space-time Kalman filter where the spatio-temporal covariance function depends on covariates. Shand and Li (2017) proposed modeling nonstationarity in both space and time for a spatio-temporal process by applying the dimension expansion technique described in Bornn et al. (2012). The computational cost of all these methods remains the main problem.

7.2 Nonstationarity in the unsteady advection-diffusion SPDE

One of the main advantages of the SPDE approach is that it directly gives rise to computationally efficient models and easily enables Gaussian random field models with spatially and/or temporal varying parameters, by offering a framework where dependence structures can be motivated from physical properties of the spatial process. Lindgren et al. (2011) noted that the deformation method for nonstationary models (Sampson and Guttorp, 1992) can be reformulated as a nonstationary SPDE in the special case of deformations within a plane. However, for deformations into a higher dimensional space the two approaches differ considerably. Since the nonstationary parameters control the local anisotropy, they must be defined on the discretization mesh. From a technical point of view, having spatially and temporal dependent parameters only changes the elements of the precision matrix of the Markov representation of the Gaussian random field.

Ingebrigtsen et al. (2014) introduced a parametric model with explanatory variables in the parameters that define the SPDE model, which results in a nonstationary spatial model. They used knowledge about the spatial process to parameterize the nonstationary dependence structure, leading to a model that can be interpreted, and they could speed up computations by sampling from the model with a fully Bayesian inference thanks to the SPDE–GMRF–INLA link.

Pereira (2019) defined generalized Gaussian fields on Riemannian manifolds that ally the ease of taking into account fields of local anisotropies (as the convolution model by Paciorek and Schervish (2006)) to the definition of covariance functions that assimilate them as a whole (as the space deformation model by Sampson and Guttorp (1992)). Their approach allowed to easily take into account local anisotropies in a global model of covariance. However, they lost the closed-form expression of the covariance model, which can only be computed numerically.

Hu et al. (2022) introduced spatial nonstationarity as a SPDE governed by a spatially varying metric tensor that is a combination of general functions and spatial covariates though a function which is either linear or nonlinear (neural network). Compared to a model with no covariate (Fuglstad and Castruccio, 2020), the function allows to incorporate external information about the local variation not captured by the large scale general basis functions.

Nonstationary fields can thus be defined by letting the parameters ($\kappa^2(\mathbf{s}, t)$, $\tau(\mathbf{s}, t)$, $c(\mathbf{s}, t)$ and $\boldsymbol{\gamma}(\mathbf{s}, t)$) in the unsteady advection-diffusion SPDE (2.27) be space-time dependent. This generalization implies only minimal changes to the method used in the stationary case concerning the simulation, but needs more work when it must be applied to inference and prediction. We can also incorporate models of spatially varying anisotropy by replacing the Laplacian operator $\nabla \cdot \mathbf{H} \nabla$ with the more general operator $\nabla \cdot \mathbf{H}(\mathbf{s}, t) \nabla$, where $\mathbf{H}$ depends on space and/or time.

The introduction of those nonstationarities in the simulation, inference and prediction approaches is the next big target of this work. This generalization could be useful for those phenomena where anisotropies are clearly present, for example for wind velocity fields. We here consider only the spatial nonstationarity. Indeed, we consider that the temporal nonstationarity can be removed by dividing the temporal domain in short time windows, where all the parameters of the SPDE can be considered temporally invariant, leading to stationarity in time on the temporal subdomain.

7.2.1 Simulation

Including spatially varying parameters into the non-conditional simulation approach detailed in Chapter 4 is quite straightforward. Since the simulation relies on the FEM/FDM discretization of the spatio-temporal SPDE, if we consider that the parameters do not vary in space inside a single triangle of the triangulation, the numerical solution of the SPDE is obtained with the same approach where the spatially varying parameters possibly change from one triangle to the other, but are constant on a single triangle. With any possible parameterization on the variables $\kappa(\mathbf{s})$, $\boldsymbol{\gamma}(\mathbf{s})$, $\tau(\mathbf{s})$ and $\mathbf{H}(\mathbf{s})$, it is sufficient to know their values on each triangle and use those in the solution of the spatial sparse system (3.7).

Here we report two different examples of spatio-temporal fields simulated from the SPDE with spatially varying parameters. We present a diffusion SPDE with varying $\kappa(\mathbf{s})$ and an advection-diffusion SPDE with varying $\boldsymbol{\gamma}(\mathbf{s})$. Both simulations show the nonstationary pattern in the spatial field. In Figure 7.1 we plot the spatial trace of the two spatio-temporal fields after 20 time steps of temporal simulation.

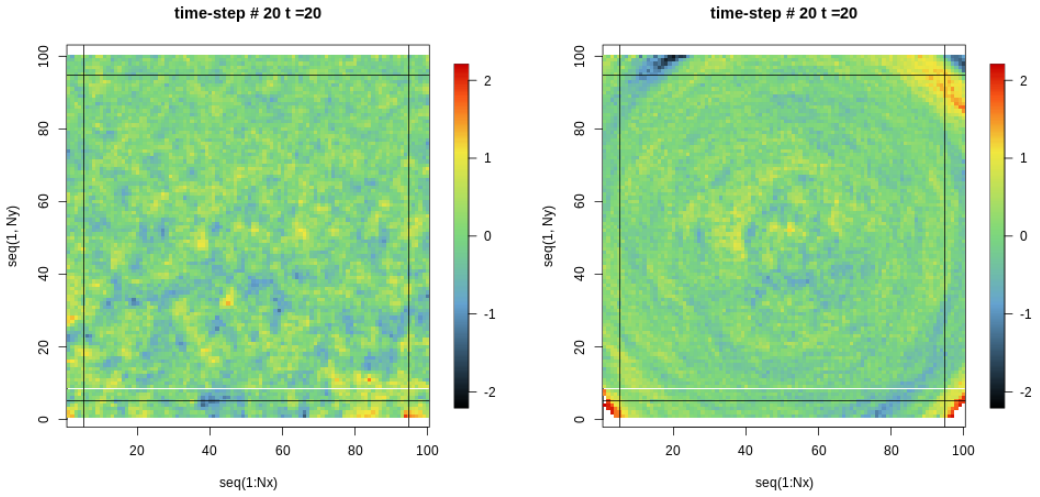


FIGURE 7.1: Nonstationary spatial trace given by spatially varying $\kappa(\mathbf{s}) = s_2/200$ (left) and spatially varying $\boldsymbol{\gamma}(\mathbf{s}) = [-0.5(s_2 - 50), 0.5(s_1 - 50)]^\top$ (right). The other parameters are set to $\kappa = 0.33$, $c = 1$, $\tau = 1$, $dt = 1$, $h = 1$.

7.2.2 Estimation

Introducing nonstationarity in the estimation technique needs to take into account the more complex structures derived by the space-time varying parameters. Different possible ways of including nonstationarities into the estimation technique can be envisaged, such as a parameterization of the parameters based on global basis functions and on external covariates or a spatio-temporal moving neighborhood integrated into an interpolation approach.

We here consider that the only spatially varying parameter in the advection-diffusion SPDE (2.27) is the advection vector $\boldsymbol{\gamma}$. This choice is made based on two considerations: first, because the spatially varying diffusion has already been considered in several papers (Lindgren et al., 2011; Fuglstad et al., 2015; Berild and Fuglstad, 2023); second, because we think that the advection field in environmental datasets could be the main source of a nonstationary behavior. The complex behaviors of the advection field should be taken into account in the statistical model to obtain better predictive performances.

As a first illustrative example, we assume that the advection field is a rotation around a point on the spatial domain and parameterize the 2-dimension advection vector as $\boldsymbol{\gamma}(\mathbf{s}) = [-v_1(s_2 - x_1), v_2(s_1 - x_2)]^\top$, where v_1, v_2, x_1, x_2 are 4 parameters to estimate. v_1 and v_2 define the velocity of the rotation, while x_1 and x_2 define the point (x_1, x_2) on the spatial domain around which the rotation occurs. After setting the 4 parameters of the spatially varying advection vector and all the other constant parameters of the SPDE, we simulate the spatio-temporal field solution to the SPDE over the spatio-temporal domain $[0, 30] \times [0, 30] \times [1, 10]$, with a triangulation of $N_S = 30^2$ spatial points and $N_T = 10$ time steps. We pick $n_S = 300$ spatial locations, uniformly selected over the spatial domain and consider their values as known. We then estimate the parameters of the model from these observations and report the results of 10 different simulations in Table 7.1. We remark that the parameters are pretty well estimated; only the parameter κ is slightly overestimated. This means that the model is able to capture the rotational nonstationary dependence structure.

	v_1	v_2	x_1	x_2	κ	c	τ	avg time (s)
true	0.5	0.5	12	18	0.7	1.5	1	
estimated	0.570 (0.057)	0.517 (0.072)	12.198 (0.534)	18.403 (0.688)	0.824 (0.060)	1.601 (0.160)	1.043 (0.036)	254
true	-0.3	-0.3	15	15	0.5	2	1	
estimated	-0.323 (0.038)	-0.311 (0.044)	14.942 (0.946)	15.386 (0.915)	0.695 (0.048)	2.139 (0.164)	1.031 (0.029)	257

TABLE 7.1: Mean (and standard deviation) of ML estimates $\hat{\boldsymbol{\theta}}^\top = (\hat{\kappa}, \hat{v}_1, \hat{v}_2, \hat{x}_1, \hat{x}_2, \hat{c}, \hat{\tau})$ over 10 simulations for two different subsets of nonstationary advection-diffusion model parameters (rotational advection).

However, it is not always possible to know a priori the structure of the nonstationarity, nor to parameterize it by a few parameters as in this case. If we do not have any information about the form of the nonstationarity, we can parameterize the space dependence of the velocity vector $\boldsymbol{\gamma}$ through a finite number of parameters by using basis functions. Let $g : \mathbb{R}^2 \to \mathbb{R}$ denote a generic function that we want to

expand in a basis, and let $p > 0$ be the number of basis functions. We use basis splines similarly to Fuglstad et al. (2015) and set

$$g(\mathbf{s}) = \mathbf{f}(\mathbf{s})^\top \boldsymbol{\alpha}_g,$$

where $\boldsymbol{\alpha}_g \in \mathbb{R}^p$, and $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}), \dots, f_p(\mathbf{s}))^\top$ is a p-dimensional vector with the basis functions evaluated at location $\mathbf{s}$. If the domain D is rectangular, we can construct a basis as a tensor product of two one-dimensional B-splines. This means that $p = m^2$, where $m > 0$ is the number of basis functions used in each dimension. We use clamped splines where the derivative is 0 at each boundary.

Let $B_{x,i}$ denote the i-th basis function of the second-order basis in the x-dimension, and similarly $B_{y,j}$ for the y-dimension. The resulting two-dimensional basis is then

$$f_{ij}(\mathbf{s}) = B_{x,i}(s_1)B_{y,j}(s_2), \quad \mathbf{s} = (s_1, s_2)^\top \in D,$$

for all combinations $i, j \in \{1, \dots, m\}$. This means that $\boldsymbol{\alpha}_g \in \mathbb{R}^{m^2}$, and m^2 parameters must be estimated for each of the arguments of the velocity vector $\boldsymbol{\gamma}$.

We use $m = 3$, hence $p = m^2 = 9$, for a total of 18 parameters for the two functions $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$. We use the same parameters of the second rotation example, obtain a simulation with rotational advection field and estimate the parameters of the model (the constant ones and the $p = 18$ parameters of the B-splines basis functions) from $n_S = 300$ observations. The constant parameters are well estimated. In Figure 7.2 we plot the two-dimensional basis made of 3 clamped B-splines on each axis on the spatial domain $[0, 30] \times [0, 30]$. The elements $\gamma_1(\mathbf{s})$ and $\gamma_2(\mathbf{s})$ of the advection vector are represented on this basis. To determine if the advection field is well estimated, we plot the advection field recovered from the estimation with B-splines and compare it to the original one in Figure 7.3. We remark that the rotation is well retrieved, even if the center point and the intensity are slightly different.

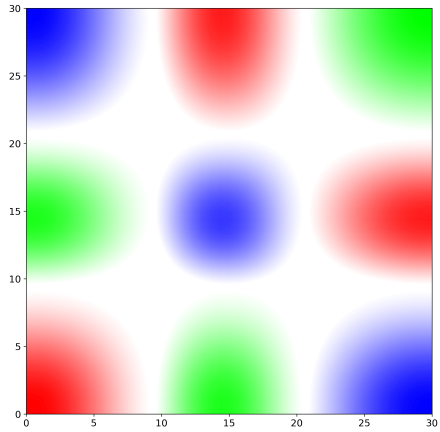


FIGURE 7.2: 2D clamped B-splines basis in which are represented the spatially varying parameters $\gamma_1(\mathbf{s})$ and $\gamma_2(\mathbf{s})$.

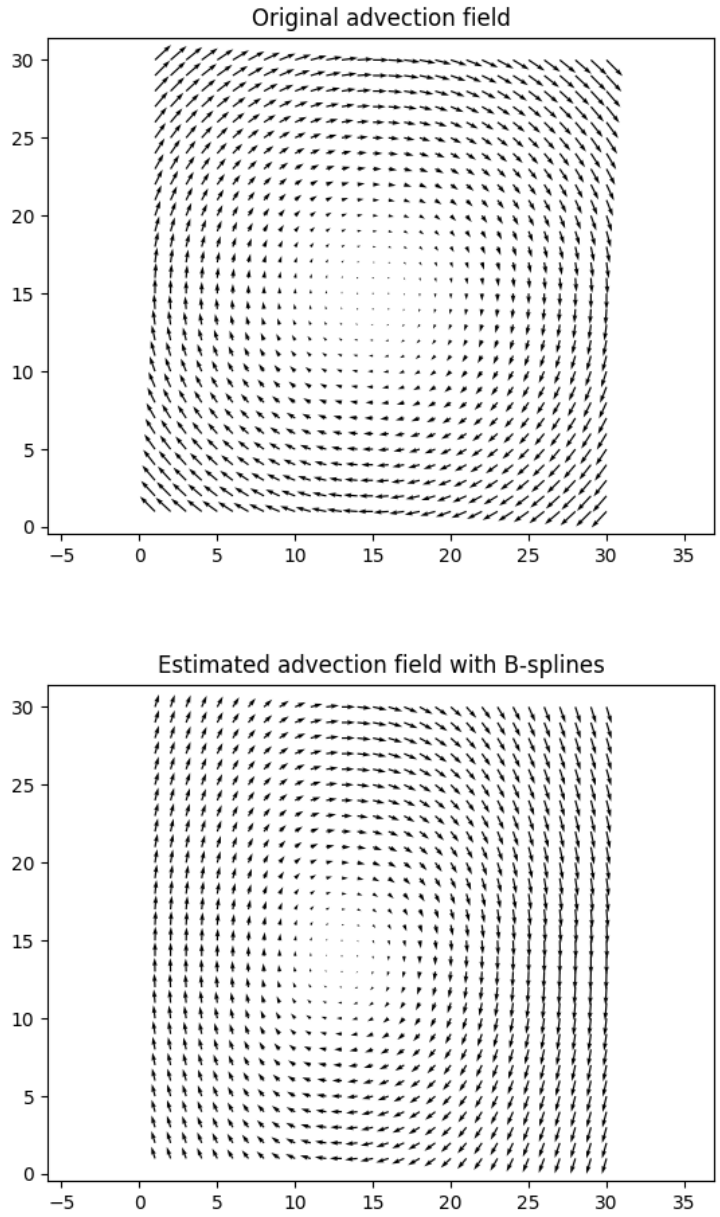


FIGURE 7.3: Rotational advection field. Top row: simulated field ($\boldsymbol{\gamma}(\mathbf{s}) = [0.3(s_2 - 15), -0.3(s_1 - 15)]^\top$). Bottom row: estimated field with parameterization based on B-splines.

7.3 Concluding remarks

Letting the spatial and spatio-temporal models be more flexible through nonstationary covariance functions is vital to be able to describe complex dependence structures. This flexibility can be obtained in the SPDE approach by making the parameters of the equation to be spatially and/or temporally varying. This extension is easily introduced in the simulation of SPDE models, but needs more work when it comes to estimation of the parameters. We have proposed two different ways to estimate a

spatially varying advection vector: first, a parameterization of the vector based on prior information about the structure of the nonstationarity, then a global parameterization based on a basis made of clamped B-splines over the two-dimensional domain. In both case, the spatial field of the advection vector is well captured by the estimation technique. However, both parameterization are based on a small number of parameters (4 and 18 respectively). The implementation details of the nonstationary extension are available in Appendix A.1.

As the number of parameters to be estimated increases with the complexity of the spatial nonstationarity, the likelihood maximization becomes harder. Moreover, in our specific case, the gradients of the log-likelihood are obtained by finite differences. This computation is cumbersome for numerous parameters. A solution would be to compute the analytical gradients of the log-likelihood, as sketched briefly in Appendix A.2.

Another possible approach to deal with spatially and/or temporally varying parameters is to use neural networks for parameter estimation. This kind of estimation approach has already been proposed by Zammit-Mangion et al. (2022).

More work is needed to obtain proper results about the estimation of nonstationary fields in the SPDE approach, and to determine if the extension to nonstationary fields is interesting for prediction performances on environmental datasets as the ones we have considered in this thesis. However, the promising results we have shown tell us that we are going in the right direction.

Chapter 8

Other contributions

Résumé

Dans ce chapitre nous présentons deux autres contributions de la thèse. Tout d'abord, nous présentons la généralisation entièrement non séparable de la classe de Gneiting des fonctions de covariance spatio-temporelles multivariées. Nous discutons de la précision de l'inférence et de la prédiction du modèle proposé sur un ensemble de données météorologiques trivariées, en la comparant à un modèle plus parcimonieux. Le principal atout de l'approche proposée est la possibilité d'obtenir des modèles entièrement non séparables dans un cadre multivarié. Ensuite, nous présentons des méthodes d'estimation et de prédiction approchées pour les données spatiales et spatio-temporelles, motivées par l'objectif de trouver le bon compromis entre la capacité prédictive et la complexité de calcul pour les grands jeux de données spatiales et spatio-temporelles. Certaines de ces méthodes se sont avérées efficaces pour l'estimation des paramètres et la prédiction dans le contexte de la "Compétition de statistique spatiale pour les grands jeux de données" organisée par la King Abdullah University of Science and Technology (KAUST) en 2021 et 2022.

In this chapter, we present two other contributions made as a PhD student. First, we present the fully nonseparable generalization of the Gneiting class of multivariate space-time covariance functions (Allard et al., 2022). We discuss the fitting and predictive scores of the proposed model on a weather trivariate dataset, by comparing it to a more parsimonious model. The main potential of the proposed approach is the possibility to obtain entirely nonseparable models in a multivariate setting. Then, we present some methods for approximate estimation and prediction for spatial and spatio-temporal data, motivated by the objective of reaching a trade-off between statistical efficiency and computational complexity for large spatial and spatio-temporal datasets. Some of these methods proved to be effective for parameter estimation and prediction in the context of the "Spatial Statistics Competition for Large Datasets" organized by the King Abdullah University of Science and Technology (KAUST) in 2021 and 2022.

8.1 Fully nonseparable Gneiting covariance functions for multivariate space-time data

In this section, we report the main points of the paper Allard et al. (2022) co-authored with Denis Allard and Xavier Emery during my PhD, and available as Supplementary Material in S.1. We broaden the well-known Gneiting class of space-time covariance functions by introducing a very general parametric class of fully nonseparable direct and cross-covariance functions for multivariate random fields, where each component has a spatial covariance function from the Matérn family with its own smoothness and scale parameters and, unlike all currently available models, its own correlation function in time. The application of the proposed model is illustrated on a weather trivariate dataset over France. Our new model yields better fitting and better predictive scores compared to a more parsimonious model with common temporal correlation function. In particular, I participated in the construction of the estimation technique, in the coding of the application of the proposed approach to real data, in the general discussions with the co-authors and in the final editing of the paper.

8.1.1 Introduction and background material

An increasing amount of multivariate data indexed by Euclidean space-time coordinates are available in numerous scientific and engineering applications, including atmospheric and environmental sciences, geosciences, as well as geological and mining engineering. For statisticians analyzing these data, one of the key issues is to model the space-time dependence structure not only for each variable separately, but also across the variables. This requires versatile models that can account for different scale and smoothness parameters for each variable, and yet whose parameters can be accurately estimated. Bourotte et al. (2016) proposed valid multivariate space-time classes based on Matérn and Cauchy spatial covariance functions, inspired by the multivariate Matérn model in Gneiting et al. (2010) and Apanasovich et al. (2012). For both classes, each variable has its own scale and own degree of smoothness in space, while allowing for some degree of cross-correlation. However, in this construction, the marginal temporal correlation function is identical for all variables, which, as discussed in Bourotte et al. (2016), was already seen as a restrictive assumption because it was found that the time correlations of the three variables in the analyzed dataset were different from one variable to the other. To account for different time correlations, space-time linear models of coregionalizations can be proposed, as in Rouhani and Wackernagel (1990), De Iaco et al. (2003), Choi et al. (2009) or Finazzi et al. (2013). However, there are two major drawbacks to this construction. First, the smoothness of any component of the multivariate field is restricted to that of the roughest underlying univariate field (Gneiting et al., 2010). Second, the number of parameters to be estimated increases quickly as the number of components of the model increases, thus raising issues of robustness and over-fitting. Gelfand et al. (2005) Apanasovich and Genton (2010), Rodrigues and Diggle (2010), Ip and

Li (2016) and Gelfand (2021) introduced Bayesian dynamic, latent dimension, convolution and spectral approaches, respectively, to build valid matrix-valued space-time covariance functions with entries that can have different smoothness parameters. Recent comprehensive accounts of space-time univariate and multivariate covariance modeling can be found in Chen et al. (2021) and in Porcu et al. (2021).

We consider a multivariate space-time random field $\mathbf{X}(\mathbf{s},t) = [X_i(\mathbf{s},t)]_{i=1}^p$ defined in $\mathbb{R}^d \times \mathbb{R}$, where d is the space dimension and p the number of random field components, each being real-valued. Without loss of generality, we assume that all random field components are centered, i.e. $\mathbb{E}[X_i(\mathbf{s},t)] = 0\ \forall(\mathbf{s},t) \in \mathbb{R}^d \times \mathbb{R}$, $\forall i = 1,\ldots,p$. It will also be assumed that the multivariate random field $\mathbf{X}(\mathbf{s},t)$ is second-order stationary, so that its covariance functions exist and depend only on the space-time lag $(\mathbf{h},u) \in \mathbb{R}^d \times \mathbb{R}$ (Chilès and Delfiner, 1999):

$$C_{ij}(\mathbf{h},u) = \mathrm{Cov}(X_i(\mathbf{s},t), X_j(\mathbf{s}+\mathbf{h},t+u)), \tag{8.1}$$

for any pair $i,j = 1,\ldots,p$, $\forall(\mathbf{s},\mathbf{s}+\mathbf{h}) \in \mathbb{R}^d \times \mathbb{R}^d$ and $\forall(t,t+u) \in \mathbb{R} \times \mathbb{R}$. The functions C_{ij} are called direct covariance functions when $i = j$ and cross-covariance functions otherwise. The matrix-valued covariance function $\mathbf{C}$, which associates each space-time lag $(\mathbf{h},u)$ with the (p,p) matrix $\mathbf{C}(\mathbf{h},u) = [C_{ij}(\mathbf{h},u)]_{i,j=1}^p$, is positive semidefinite in $\mathbb{R}^d \times \mathbb{R}$.

The separability and nonseparability concepts were defined in Section 1.2 for each $C_{ij}(\mathbf{h},u)$ such that $i = j$. The separability functions (see Equation (1.3)) are $s_{ii}(\mathbf{h},u)$, for all $i = 1,\ldots,p$. A matrix-valued space–time covariance function is said to be a proportional model when it is obtained as the product of a (p,p) covariance matrix $\mathbf{R}$ and a univariate space–time correlation function ρ, i.e. $\mathbf{C}(\mathbf{h},u) = \mathbf{R}\ \rho(\mathbf{h},u)$, $\forall(\mathbf{h},u) \in \mathbb{R}^d \times \mathbb{R}$. It is proportional-in-time if $\mathbf{C}_T(u) = \mathbf{R}\ \rho_T(u)$ and proportional-in-space if $\mathbf{C}_S(\mathbf{h}) = \mathbf{R}\ \rho_S(\mathbf{h})$, $\forall(\mathbf{h},u) \in \mathbb{R}^d \times \mathbb{R}$. Bourotte et al. (2016) proposed nonproportional matrix-valued space-time covariance functions in which the purely spatial part is a nonproportional model but the purely temporal part is a proportional model, i.e., the model is proportional-in-time but not in space.

Instead, we build valid and flexible parametric classes of matrix-valued space–time covariance functions such that both the spatial and temporal components are nonproportional. Such matrix-valued space-time covariances are referred to as being *fully nonseparable*. The proposed class of multivariate space-time covariance functions with Matérn spatial traces is of the form

$$C_{ij}(\mathbf{h},u) = \sigma_{ij}(u)\mathcal{M}(\mathbf{h};r_{ij}(u),\nu_{ij}(u)), \qquad (\mathbf{h},u) \in \mathbb{R}^d \times \mathbb{R},\ i,j = 1,\ldots,p, \tag{8.2}$$

where $\mathcal{M}(\cdot;r,\nu)$ denotes the univariate Matérn covariance function with scale and smoothness parameters r and ν, i.e., $\mathcal{M}(\mathbf{h};r,\nu) = C_\nu^{\mathcal{M}}\left(r\|\mathbf{h}\|\right)$, where $C^{\mathcal{M}}$ is the unit variance and scale Matérn covariance function defined in (2.5). In order to build multivariate models that are neither proportional in space nor in time, an essential building block is to define admissible and relevant temporal matrix-valued covariances

of the form $[(\gamma_{ij}(u)+1)^{-d/2}]_{i,j=1}^{p}$.

8.1.2 A general fully nonseparable Gneiting-Matérn class

Consider a p-variate temporal random field $\mathbf{X}(t) = [X_i(t)]_{i=1}^{p}$, $t \in \mathbb{R}$. Under the assumption that the direct and cross-increments are jointly stationary, the matrix-valued pseudo-variogram $\boldsymbol{\eta} = [\eta_{ij}]_{i,j=1}^{p}$ is defined as (Papritz et al., 1993)

$$\eta_{ij}(u) \quad = \quad \frac{1}{2}\mathrm{Var}[X_i(t+u) - X_j(t)], \tag{8.3}$$

for all $t, u \in \mathbb{R}$ and any pair $i, j = 1, \ldots, p$. Notice that the pseudo-variogram has nonnegative entries and is not necessarily an even function. For any $i = 1, \ldots, p$, η_{ii} (direct variogram of X_i) is a conditionally negative semidefinite function ($\eta_{ii} \preceq_c 0$), i.e., such that $\sum_{i=1}^{p}\sum_{j=1}^{p} \omega_i\, \omega_j\, \eta_{ij} \leq 0, \forall\, \omega_1, \ldots, \omega_p \in \mathbb{R}$ such that $\sum_{i=1}^{p} \omega_i = 0$ (Chilès and Delfiner, 1999). Necessary and sufficient conditions for a matrix-valued function to be a pseudo-variogram are provided in Dörr and Schlather (2021).

Theorem 8.1. *Let $\boldsymbol{\nu}$, $\mathbf{b}^2$ and $\mathbf{a}^2$ be (p,p) symmetric conditionally negative semidefinite matrices, all with positive entries, such that $\mathbf{a}^2 - \boldsymbol{\nu} \preceq_c 0$. Let $u \mapsto \boldsymbol{\rho}(u)$ be a (p,p) matrix-valued covariance function on $\mathbb{R}$ and let $u \mapsto \boldsymbol{\eta}(u)$ be a (p,p) matrix-valued pseudo-variogram on $\mathbb{R}$. Denoting $\mathbf{R}(u) = [r_{ij}(u)]_{i,j=1}^{p}$ and $\boldsymbol{\sigma}(u) = [\sigma_{ij}(u)]_{i,j=1}^{p}$, the matrix-valued function $\mathbf{C} : (\mathbf{h}, u) \mapsto [C_{ij}(\mathbf{h}, u)]_{i,j=1}^{p}$ with*

$$C_{ij}(\mathbf{h}, u) = \sigma_{ij}(u)\, \mathcal{M}(\mathbf{h}; r_{ij}(u), \nu_{ij}), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R},$$

is positive semidefinite in $\mathbb{R}^d \times \mathbb{R}$ for

$$\mathbf{R}(u) = \sqrt{\frac{\alpha\boldsymbol{\eta}(u) + \mathbf{a}^2}{\beta\boldsymbol{\eta}(u) + \mathbf{b}^2}} \;\; \textit{and} \;\; \boldsymbol{\sigma}(u) = \frac{\boldsymbol{\rho}(u)\Gamma(\boldsymbol{\nu})\exp(\boldsymbol{\nu})}{(\alpha\boldsymbol{\eta}(u) + \mathbf{a}^2)^{\boldsymbol{\nu}} \left(\beta\boldsymbol{\eta}(u) + \mathbf{b}^2\right)^{d/2}}, \tag{8.4}$$

where α and β are nonnegative.

Corollary 8.1. *Let $\boldsymbol{\nu}$, $\mathbf{b}^2$, $\mathbf{a}^2$, $\boldsymbol{\eta}$, α and β be as in Theorem 8.1. Let $\boldsymbol{\tau}$ be a (p,p) symmetric real matrix. The matrix-valued function $\mathbf{C} : (\mathbf{h}, u) \mapsto [C_{ij}(\mathbf{h}, u)]_{i,j=1}^{p}$ with*

$$C_{ij}(\mathbf{h}, u) = \frac{\tau_{ij}}{(\beta\eta_{ij}(u) + b_{ij}^2)^{d/2}(\alpha\eta_{ij}(u) + a_{ij}^2)^{\nu_{ij}}} \mathcal{M}\left(\mathbf{h}; \sqrt{\frac{\alpha\eta_{ij}(u) + a_{ij}^2}{\beta\eta_{ij}(u) + b_{ij}^2}}, \nu_{ij}\right), \tag{8.5}$$

is positive semidefinite in $\mathbb{R}^d \times \mathbb{R}$ if the matrix $\boldsymbol{\tau} e^{-\boldsymbol{\nu}}/\Gamma(\boldsymbol{\nu})$ is positive semidefinite.

We refer to Allard et al. (2022) for a proof and more mathematical details. The matrix-valued space-time covariance function (8.5) is of Gneiting-Matérn type. When $\alpha = 0$, accounting for the fact that $\exp(-\boldsymbol{\nu})$ is positive semidefinite, the models in (8.4) and (8.5) generalize the Gneiting-Matérn model (Bourotte et al., 2016)

$$C_{ij}^{\mathcal{M}}(\mathbf{h}, u) = \frac{\sigma_i \sigma_j \rho_{ij}}{(\gamma(u)+1)^{d/2}} \mathcal{M}\left(\mathbf{h}; \frac{r_{ij}}{(\gamma(u)+1)^{1/2}}, \nu_{ij}\right), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}. \tag{8.6}$$

If, furthermore, the direct variograms η_{ii} are increasing functions for $i = 1, \ldots, p$, the scale parameters $r_{ii}(u) = a_{ii}(\beta\eta_{ii}(u) + b_{ii}^2)^{-1/2}$ decrease as u increases. The space-time direct correlations are thus higher than what would happen for a separable space-time covariance function, i.e., they are positively nonseparable. When $\beta = 0$ and under the same monotonicity assumption for the direct variograms, the scale parameters $r_{ii}(u) = \sqrt{\alpha\eta_{ii}(u) + a_{ii}^2}/b_{ii}$ increase with u. This a very original feature, since for all Gneiting-type spatio-temporal models of our knowledge, the scale parameter decreases with u. Applying the same reasoning as above, the resulting space-time direct covariances are thus negatively nonseparable. When α and β are both positive, the model is positively or negatively space-time nonseparable, depending on the relative values of the other parameters involved in $\mathbf{R}(u)$. However, in the common case where $\eta_{ij}(u) \to \infty$ as $u \to \infty$, $\mathbf{R}(u)$ asymptotically tends to a matrix with all elements equal to $\sqrt{\alpha/\beta}$. The space-time covariance is thus asymptotically separable as $|u| \to \infty$.

8.1.3 Application to a weather dataset

Data and parameterization

We now illustrate the model presented in Section 8.1.2 on a weather dataset consisting of three daily variables (solar radiation R, temperature T, and humidity H) recorded at 13 stations in Western France from 2003 to 2012. We shall compare a specific fully nonseparable multivariate model (FULL), as proposed in (8.5), to a Proportional-In-time (PIT) model as in (8.6). Both models will be applied to residuals, after centering in time and space, in order to filter out any seasonal and regional effect. For each of the three variables, the separability index $S(\mathbf{h}, u) = C(\mathbf{h}, u)C(\mathbf{0}, 0) - C_S(\mathbf{h})C_T(u)$ was computed. Empirical covariances were derived from the space-time empirical variograms computed with the function `EVariogram` from the package `GeoModels` (Padoan and Bevilacqua, 2015), using spatial bins of 50 m. Since the variograms have been computed on residuals, $C(\mathbf{0}, 0) = 1$ and the separability index must be in the interval $[-1, 1]$, with $S(\mathbf{h}, u) = 0$ indicating separability of $C(\mathbf{h}, u)$. These analyses showed clearly that the residuals are uniformly positively space-time nonseparable for $|\,\mathbf{h}\,| \leq 500$ m and $u \leq 3$, which is actually often the case for climate variables. The parameter α in (8.5) is thus set to $\alpha = 0$. For identifiability reasons, we further set $\beta = 1$ and $b_{ij}^2 = 1 - R_{ij}^0$ with $\max_i\{R_{ii}(0)\} < 1$.

The pseudo-variogram $\boldsymbol{\eta}$ needs to be specified. For the unbounded part common to all components, we choose $\gamma_0(u) = (1 + |cu|^{2a})^b - 1, \quad u \in \mathbb{R}$, with $c > 0, 0 < a \leq 1$ and $0 \leq b \leq 1$, for ease of comparison with Bourotte et al. (2016). This variogram is reminiscent of the generalized Cauchy covariance model that will also be used for $\mathbf{R}$: $R_{ij}(u) = A_i A_j(1 + |ru|^{2\lambda})^{-b}$, $u \in \mathbb{R}$, with $r > 0$, $0 < \lambda < 1$ and $0 \leq A_i < 1$ for $i = 1, \ldots, p$, and where for the sake of parsimony the parameter b is set identical to the parameter b in γ_0. In this parameterization, the space–time nonseparability parameter b acts both inside and outside the Matérn function $\mathcal{M}$, implying a constant

perfect correlation in time if $b = 0$. Following Gneiting (2002) and Bourotte et al. (2016), a reparameterization is thus useful. Multiplying the above equation by the temporal covariance function $\tilde{\rho}_{ij}(u) = \left[(1 + c|u|^{2a})^{\delta} - A_i A_j (1 + r|u|^{2\lambda})^{-\delta}\right]^{-1}$ with $0 \leq \delta \leq 1$ leads to the parametric family that will be used in the rest of this section. The parameterization is further simplified by setting $\tau_{ii} = \sigma_i^2 a_{ii}^{2\nu_{ii}}(1 - A_i^2)^2$ and, for $i \neq j$, $\tau_{ij} = \rho_{ij}\sqrt{\tau_{ii}\tau_{jj}}$, with $[\rho_{ij}]_{i,j=1,\ldots,p}$ being a correlation matrix. With this parameterization, the matrix $\boldsymbol{\tau}$ is positive semidefinite by construction. Finally, the model reads

$$C_{ij}(\mathbf{h}, u) = \frac{1}{(1 + |cu|^{2a})^{\delta} - A_i A_j (1 + |ru|^{2\lambda})^{-\delta}} \frac{\rho_{ij} a_{ij}^{-2\nu_{ij}} \sqrt{\tau_{ii}\tau_{jj}}}{(1 + |cu|^{2a})^{b} - A_i A_j (1 + |ru|^{2\lambda})^{-b}} \times \mathcal{M}\left(\mathbf{h}; \frac{a_{ij}}{\left[(1 + |cu|^{2a})^{b} - A_i A_j (1 + |ru|^{2\lambda})^{-b}\right]^{1/2}}, \nu_{ij}\right), \quad (8.7)$$

with $(\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$ and $i, j = 1, \ldots, p$. The model in (8.7) provides interpretable parameters for a fully nonseparable Gneiting-Matérn model.

Estimation

Composite likelihood methods have proven efficient in the context of spatio-temporal data, with less time-consuming steps and good asymptotic properties (Bevilacqua et al., 2012). Pairwise Likelihood (PL) is the product of marginal Gaussian likelihoods computed on all pairs $\{Z_i(\mathbf{s}_\alpha, t_\alpha), Z_j(\mathbf{s}_\beta, t_\beta)\}$ such that $|\mathbf{s}_\alpha - \mathbf{s}_\beta| \leq d_m$ and $|t_\alpha - t_\beta| \leq t_m$, where $i, j = 1, \ldots, p$ and $\alpha, \beta = 1, \ldots, N$. The data analyzed here are standardized residuals after centering for spatial and temporal trends. We thus set $\sigma_i = 1$, for $i = 1, \ldots, p$. There is a total of $p(p + 2) + 6$ parameters to estimate, where p is the number of variables. Here $p = 3$, leading to a total of 21 parameters. PL was maximized sequentially in subspaces of the parameter space corresponding to blocks of related parameters, while keeping all other parameters fixed to the values previously attained. Finally $\hat{b}$ is the value of b among $\{0, 0.1, \ldots, 0.9, 1\}$ corresponding to the highest maximized PL. To perform the maximization in the subspaces of the parameter space, we used the package `nlminb` implemented in R with lower and upper bounds for the parameters when mathematically necessary.

All 13 stations were used to estimate the parameters of both the FULL and PIT models. Figure 8.1(left) shows the maximum log-PL as a function of the separability parameter b for the fully nonseparable Gneiting-Matérn model, referred to as FULL. The maximum is reached for $\hat{b} = 0.1$ with $\log \mathrm{PL}_{\mathrm{FULL}}(\hat{\boldsymbol{\theta}}_{b=0.1}) = -933\,337.6$. For comparison purpose, PL was also maximized for the PIT model, i.e. when $\mathbf{a} = \mathbf{0}$. For this simpler model, $\log \mathrm{PL}_{\mathrm{PIT}}(\hat{\boldsymbol{\theta}}_b)$ is monotonically decreasing from $-933\,961.3$ to $-934\,072.6$ as b varies from 0 to 1. FL was also computed for these two models (remember that FL is easy to compute but very long to optimize): $\log \mathrm{FL}_{\mathrm{FULL}}(\hat{\boldsymbol{\theta}}_{b=0.1}) = -2\,652.7$ and $\log \mathrm{FL}_{\mathrm{PIT}}(\hat{\boldsymbol{\theta}}_{b=0.1}) = -2\,819.1$, indicating strong evidence in favor of the more complex model, considering that the number of data

is moderate with a total of 10 independent repetitions of 1209 correlated data. As a point of comparison, the (BIC) penalization to the full likelihood of the more complex model is equal to $2.5 \ln(12090) \simeq 23.5$.

Looking at the temporal marginals, the models show very different behaviors – as we expected. Figure 8.1(right) displays the temporal empirical covariances for variables R and H with their fitted models. Thanks to the flexibility of the FULL model, the temporal covariance functions fit the empirical values for both variables. In contrast, the unique covariance function of the PIT model lies somewhere between the empirical covariances of the two variables, thus being unable to provide a satisfactory fit to any of them. Table 8.1 reports the estimated parameters maximizing PL for both models. It is interesting to note that the parameter λ is equal 0 in the FULL model, thus implying a sort of temporal nugget effect, visible on the temporal covariances in Figure 8.1(right). Figure 8.2 shows the spatial (at time lag $u = 0$) and spatio-temporal covariances (at time lags $u = 1, 2$) for all variables and pairs of variables. The empirical covariances between all pairs of stations and the FULL model with estimated parameters from Table 8.1 are shown. Overall, the fit is good. On the top left panel representing the covariance function for the variable R, the very high smoothness of the covariance function near the origin is clearly visible.

	Temporal parameters							
	c	a	r	λ	δ	A_1	A_2	A_3
FULL	0.528	0.993	0.341	0.0	1.000	0.871	0.608	0.646
PIT	4.31	0.997	N/A	N/A	0.638	0	0	0
	Spatial parameters							
	a_1	a_2	a_3	ν_1	ν_2	ν_3		
FULL	$15.8\ 10^{-3}$	$6.4\ 10^{-3}$	$12.4\ 10^{-3}$	3.76	0.69	0.51		
PIT	$18.9\ 10^{-3}$	$8.2\ 10^{-3}$	$17.0\ 10^{-3}$	1.61	0.71	0.53		
	Correlation parameters							PL
	ρ_{12}	ρ_{13}	ρ_{23}					
FULL	−0.124	−0.330	−0.099					−933 337.6
PIT	−0.151	−0.324	−0.103					−933 961.3

TABLE 8.1: PL parameter estimates for the weather dataset with $d_m = 250$ km, $t_m = 2$ days. FULL: fully nonseparable model as in (8.7) with $b = 0.1$. PIT: simplified Proportional-In-Time model with $b = 0$. The indices 1, 2 and 3 refer to R, T and H, respectively.

Prediction

We consider two different prediction settings. In the first setting, called *Spatial Interpolation* , prediction of the three variables at a validation station at day t is made knowing the trivariate data at the 11 estimation stations at day t, $t-1$ and $t-2$. In the second setting, called *Temporal Prediction*, prediction is made with data from days $t-1$ and $t-2$ only at all 13 stations (estimation and validation stations). Other settings, such as Spatio-temporal prediction (using data at 11 stations from days $t-1$ and $t-2$ for predicting at the 2 validations stations at day t) is possible, but not shown here because they are less common in practical situations. Predictions are computed from January 3 to January 31. In the Gaussian framework, the conditional

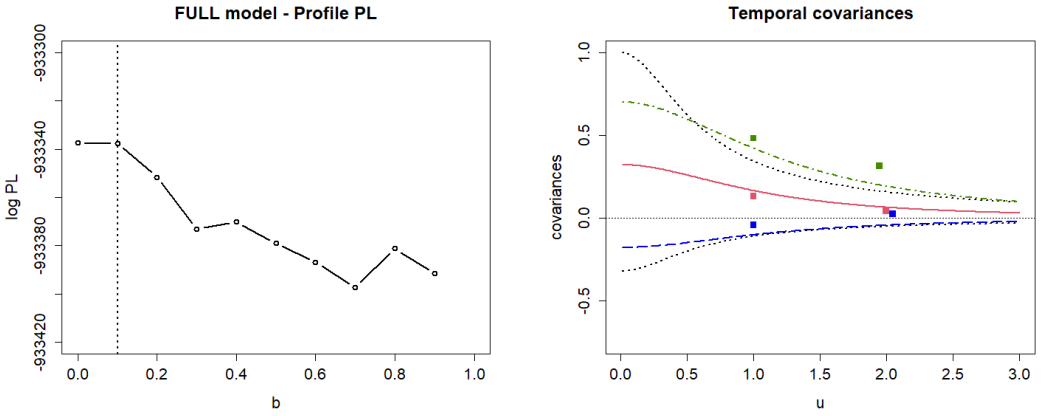


FIGURE 8.1: Left: PL as a function of the space-time separability parameter b. Right: (R,H) temporal correlation functions. PIT model with unique direct correlation (positive) and cross-correlation (negative) in dotted lines. FULL model with direct correlation (positive) for R and H in solid and dotdashed lines, respectively, and cross-correlation (negative) in longdashed line. Squares: empirical correlations.

expectation is the best linear unbiased prediction, also called (co)kriging in the geostatistics literature, while the conditional covariance matrix is the covariance matrix of the cokriging errors. The conditional variances are identical for all days, since the configuration of the conditioning data remains identical. The two models (FULL and PIT) are compared by means of the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the Continuous Ranked Probability Score (CRPS) and the Logarithmic Score (LogS). Table 8.2 reports our results. As a benchmark, we also report the scores obtained with a trivial prediction where all predicted values are set to the expectation (equal to 0) and all variances to the theoretical variance (equal to 1).

	RMSE			MAE			CRPS			LogS		
	R	T	H	R	T	H	R	T	H	R	T	H
	Spatial interpolation											
FULL	**0.746**	**0.582**	**0.864**	**0.587**	0.464	0.732	**0.427**	**0.335**	**0.509**	**0.268**	0.015	**0.393**
PIT	0.751	0.584	**0.864**	0.593	**0.462**	**0.731**	0.430	**0.335**	**0.509**	0.271	**0.010**	**0.393**
	Temporal prediction											
FULL	**0.919**	0.837	**0.813**	0.784	**0.706**	**0.667**	**0.531**	**0.482**	**0.470**	**0.421**	**0.329**	**0.311**
PIT	0.933	**0.835**	0.825	**0.780**	0.714	0.684	0.537	0.483	0.479	0.437	0.331	0.330
	Trivial interpolation											
	0.933	0.893	0.919	0.801	0.777	0.782	0.540	0.521	0.538	0.439	0.403	0.443

TABLE 8.2: Prediction scores at the validation stations at day t, using ML estimates. FULL: fully nonseparable model as in (8.7) with $b = 0.1$. PIT: simplified Proportional-In-Time model with $b = 0$. The best scores among FULL or PIT are shown in bold font.

For illustration purpose, Figure 8.3 shows the predictions of the three variables from January 3 to January 31, 2003 at one station (Le Rheu) following the Spatial Interpolation setting. The predicted residuals were back-transformed to the original scales to be compared to the real values. An envelope of ± 2 conditional standard

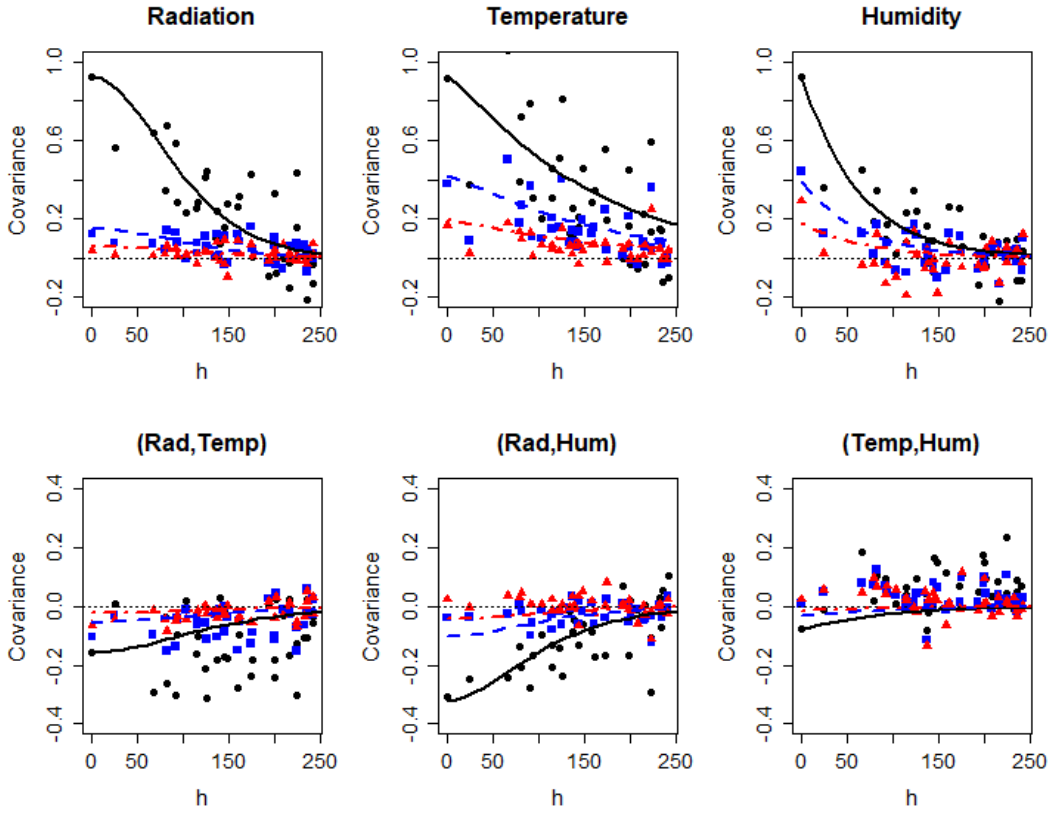


FIGURE 8.2: Empirical and fitted spatial-temporal correlation at $u = 0$ (black solid lines and circles), $u = 1$ (blue dashed lines and squares) and $u = 2$ (red dotdashed lines and triangles).

deviations was added to the plot.

8.1.4 Concluding remarks

We have proposed new parametric classes of matrix-valued covariance functions for multivariate spatio-temporal random fields, where each component has its own smoothness and scale parameter in space and its own correlation function in time. Our constructions generalize the Gneiting class of space-time covariance functions and are fully nonseparable, in the sense that they are space-time nonseparable and that they are neither proportional-in-space nor proportional-in-time. This is a major improvement to most of the models proposed up to date (with the exceptions of the models in Dörr and Schlather (2021) and Porcu et al. (2022)), where the marginal temporal correlation function is identical for all the variables, which is overly restrictive. The main ingredient for achieving this has been the use of pseudo-variograms, as multivariate analogous of Bernstein functions as pointed out by Dörr and Schlather (2021) and Qadir and Sun (2022) in their discussion.

The main class proposed in Section 8.1.2 allows for a general behavior for the spatial scale parameter as a function of the time lag u, depending on the values taken by the parameters. It is well known that the Gneiting class of models is positively

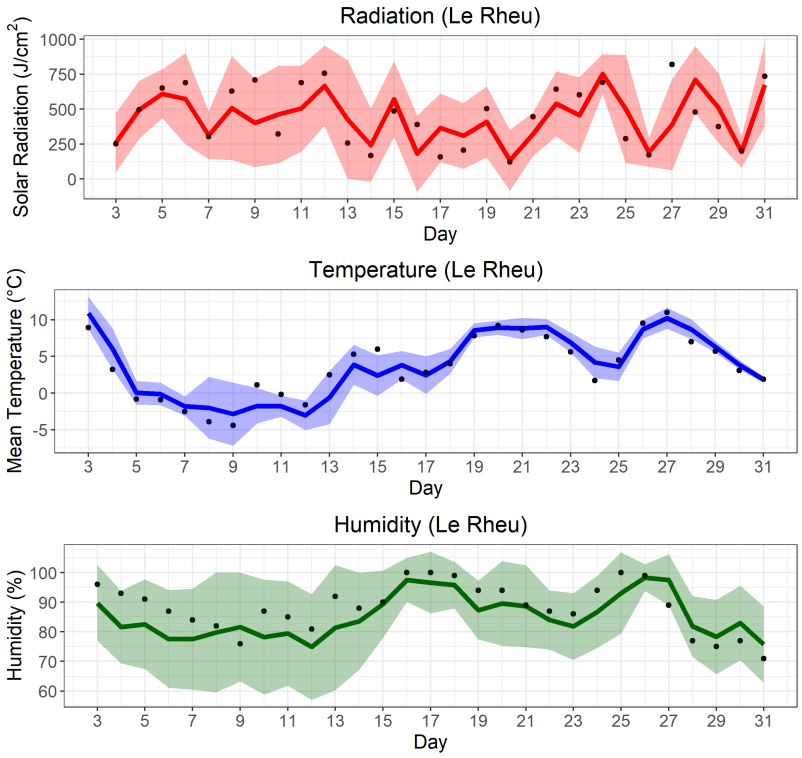


FIGURE 8.3: Predictions of R, T and H at Le Rheu from January 3 to January 31, 2003 following the *Spatial Interpolation* setting. Predictions (continuous line), real values (dots) and envelope of ± 2 conditional standard deviations (colored ribbon).

nonseparable. A very original feature of our construction is that the models can be negatively nonseparable or asymptotically separable as $|u| \to \infty$. Since these models are continuous over $\mathbb{R}^d \times \mathbb{R}$, their application is not limited to data organized as regular records at measurement stations. More complex designs are in principle possible, such as irregular or incomplete records, heterotopic designs or even mobile measurement devices. We hope to see applications of our models to such datasets in a near future.

From a statistical point of view, the main challenge is to find the right balance between the flexibility of the model, which must be able to account for the complex interactions between space and time across several variables, and the number of parameters that should remain interpretable and not too numerous. More complex models should only be preferred if supported by the data. The dataset analyzed in this work is of moderate size with 10 independent repetitions of $13 \times 31 = 403$ space-time trivariate data. Its analysis has shown that our model improves the fitting and the predictive performances over the more parsimonious model in Bourotte et al. (2016), in particular when the time dimension is important for prediction. In Section 8.1.3, it was shown that the maximum PL difference is around 600. Combined with the improved prediction scores in time, this provides evidence in favor of our new model for this dataset. The analysis of other datasets, in particular with more spatial locations, is necessary for confirming this first result.

Estimating the parameters of a parametric model such as the one in Section 8.1.3 is a challenge. Here, we have used Pairwise Likelihood because it was proven efficient on the proportional-in-time model in Bourotte et al. (2016). It is however relatively slow: reaching convergence with our optimization algorithm and the function `nlmbinb` takes typically several hours on a recent laptop computer. There is thus certainly room for improvement for an easier use of such complex models. At the cost of even larger computing times, composite likelihood with larger sets of elements than pairs and full likelihood might be interesting options for a more accurate estimation of the parameters, considering the complex interactions between space, time and the variables. Conducting a thorough comparison of several likelihood approaches was out of the scope of this work and left for future research. In our opinion, more interesting improvements can be expected by using Vecchia Gaussian process approximation (Katzfuss and Guinness, 2021). We are looking forward to (multivariate) spatio-temporal versions of the `GpGp` package.

All the models presented in this work can easily be generalized to $\mathbb{R}^d \times \mathbb{R}^\ell$, where ℓ is an integer greater than one. Other constructions based on characterizations established in Porcu et al. (2022) are also possible. Extensions of our models to other spaces that are of interest in environmental sciences, atmospheric sciences and geosciences, such as a sphere crossed with an Euclidean space ($\mathbb{S}^d \times \mathbb{R}^\ell$), are left to further research.

8.2 Spatial and spatio-temporal statistics for large datasets

In this section we discuss a project of my PhD related to the approximate approaches for parameter estimation and prediction for spatial and spatio-temporal data. In particular, we examine the participation of the RESSTE team in 2021 and 2022 to the "Spatial Statistics Competition for Large Datasets" organized by the King Abdullah University of Science and Technology (KAUST) located in Thuwal, Saudi Arabia. The RESSTE team was composed of 4 researchers from Mines Paris and INRAE Avignon (Denis Allard, Thomas Romary, Thomas Opitz and myself). We present the techniques and reasoning that led the team to rank in the first positions in both the editions of the competition.

The interest in approximate approaches is also motivated by the other research projects of the PhD, where those techniques were either part of the approach itself (the GMRF approximation of the random fields solution to the SPDE, see Chapter 3) or were found to be necessary to reduce the computational cost of the maximum likelihood (the pairwise likelihood used in Section 8.1).

8.2.1 State of the art

Most of the prediction approaches and estimation techniques for spatial and spatio-temporal data that have been proposed in the last decade are motivated by reaching a compromise between statistical efficiency and computational complexity (Porcu et al., 2021). This is due to the fact that massive (and/or multivariate) data sets are now available. The main estimation techniques for massive spatial datasets were gathered and reviewed by Sun et al. (2012). They include, among others, covariance matrix tapering, composite likelihoods, spectral techniques, and approximation of the random field with a GMRF. Estimation methods for space(–time) covariances for large data are conceptually similar to classical approaches, but they have to deal with intractable computation of likelihoods. It is statistically difficult to decide which method is the best, because approximating covariance functions with computationally tractable ones is a form of deliberate misspecification of the true underlying covariance structure. For example, in the covariance tapering framework (Furrer et al., 2006; Kaufman et al., 2008), a direct product of the true covariance function and a compactly supported correlation function is used. Composite likelihood methods (Varin et al., 2011) are built on the consideration of the likelihood function as a product of marginal or conditional densities. Low-rank approximation methods (Banerjee et al., 2008; Cressie and Johannesson, 2008) project the entire random process to a low-dimensional space and use the low-rank representation to approximate the original process. Spectral methods rely on truncation of the spectral expansion, hence on the loss of information on the geometric properties of the associated random field. Finally, Markov random fields coupled with SPDEs (Lindgren et al., 2011; Rue et al., 2009) approximate a continuous process with a process defined over a mesh for which

the conditional distributions only depend on neighbors, leading to sparseness of the precision matrix.

The Spatio-Temporal Statistics and Data Science (STSDS) department of KAUST launched the first KAUST "Spatial Statistics Competition for Large Datasets" to the international spatial statistics community in 2021, followed by the second and third editions in 2022 and 2023. Several teams from all over the world participated in the competition, 21 in 2021 and 14 in 2022. The STSDS team has been exploiting the modern High-Performance Computing (HPC) systems to make large-scale exact computation of likelihoods possible and to allow processing larger data sizes more easily than before. They considered that the lack of large-scale exact computation has led to an inefficient assessment of spatial modeling approximation methods in the last decades. Indeed, on the one hand, every research group has used different datasets to assess the proposed methods; on the other hand, the few comparative studies have been limited to small and medium-sized datasets. Thus, they decided to assess the methods developed worldwide based on generated synthetic datasets and on the true and the exact estimated parameters, obtained with the `ExaGeoStat` software they have been developing (Abdulah et al., 2018). This software can deal with millions of locations, by relying on parallel linear algebra libraries to allow fast and efficient computation of the log-likelihood function and by using modern runtime systems to make the code portable on the different parallel architectures, including GPUs. `ExaGeoStat` has been coupled with a rich class of covariance models to obtain exact parameter estimation and predictions for large-scale spatial and spatio-temporal data, see e.g. Salvaña and Genton (2021).

In the end, the goal of the competition was to achieve a comprehensive comparison between as many different methods as possible for parameter estimation and prediction on massive datasets.

8.2.2 2021 Edition

The 2021 edition of the "Spatial Statistics Competition for Large Datasets" featured 4 sub-competitions. The objective was to evaluate existing approximation methods on large spatial datasets. A collection of large synthetic datasets (100K/1M data) was generated from a set of real models using `ExaGeoStat` and each dataset was split into training and validation sets. The selected models covered disparate spatial properties to ensure a fair comparison between all competing methods. All the details on the datasets and on the scoring rules, and the main methods used by the participants are available in Huang et al. (2021). The RESSTE team was ranked first at 3 sub-competitions out of the 4 proposed.

In sub-competitions 1a and 1b, 16 datasets of 100K data each were generated from a Gaussian spatial field with a Matérn covariance function as defined in (8.2). The goal was to estimate the parameters of the covariance function from 90K training data and to make spatial predictions at 10K test locations. The four parameters determining the covariance structure are the partial sill σ^2, the range $1/r > 0$, the smoothness

$\nu > 0$, and the nugget τ^2. Different smoothness scenarios were considered, from smooth to rough, and, by varying the range parameter, different resulting effective range were proposed, for weak, medium, and strong dependence. The 16 datasets thus covered a wide variety of covariance properties. The Mean Loss of Efficiency (MLOE) and Mean Misspecification of the Mean Square Error (MMOM) were used to evaluate the performances of the parameter estimation in the sub-competition 1a. MLOE characterizes the average loss of prediction efficiency when the approximated model is used for predictions instead of the true model. MMOM characterizes the average misspecification of the mean square error when calculated under the approximated model. RMSE was used to evaluate the prediction accuracy in the sub-competition 1b.

As RESSTE team, we opted here for block approaches, which satisfy the following principles: (i) in each block, the estimation (by maximum likelihood) or the prediction (by kriging) is optimal; (ii) the blocks should be as large as possible while taking into account the computational problems; (iii) the blocks are assumed to be independent to each other. The approximation lies entirely in (iii), and it is easy to understand that (ii) is the key to success. We ranked first in the sub-competition 1b, related to spatial prediction, and third in the parameter estimation of sub-competition 1a. An approximation approach proposed by another team that gave good results in terms of parameter estimation and spatial prediction was the Vecchia approximations method detailed below. Generally, the parameter estimation was more difficult when the process was smoother (higher smoothness parameter) and had stronger dependence (higher effective range). Even if we did not find the best estimates of the parameters, however our prediction approach gave better results than the other teams (even those that ranked first and second in the sub-competition 1a) and than the exact kriging obtained with `ExaGeoStat` without any approximation. One possible reason is that the approximation of the other teams was inadequate in kriging, even though the underlying model they used was more accurate. This implies that both the model inference and the number of neighbors considered are important for local kriging predictions.

The datasets for sub-competitions 2a and 2b were univariate non-Gaussian spatial data on a unit square containing 100K and 1M observations respectively. Without giving the information about the generating process to the participants in the competition, the data were generated by the Tukey g-and-h random fields (Xu and Genton, 2017), which generalize Gaussian random fields to account for skewness and heavy tails. For a generated Gaussian field $X(\mathbf{s})$, the Tukey g-and-h random process $T(\mathbf{s})$ is defined by marginal transformation at each location $\mathbf{s}$ as

$$T(\mathbf{s}) = \xi + \omega \frac{e^{gX(\mathbf{s})} - 1}{g} \frac{e^{hX^2(\mathbf{s})}}{2},$$

where ξ and ω are the location and scale parameters, respectively, g controls the skewness and $h \geq 0$ determines the tail-heaviness. They chose two sets of values g and

h to consider random processes with medium and strong departure from Gaussianity. RMSE was used to evaluate the prediction accuracy in the sub-competitions 2a and 2b.

To process these data, we first considered several methods, either from more classical geostatistical analyses or from machine learning techniques. We recognized the data distribution, as being generated by the Tukey g-and-h random fields. We thus transformed the data into Gaussian data by estimating the parameters of the Tukey g-and-h transformation and by using the inverse transform, we then estimated the parameters of the Gaussian model and finally obtained the predictions by kriging before transforming the data back with the direct transform. The estimation of the parameters and the prediction were done with the R package `GpGp` which uses the Vecchia approximations (Vecchia, 1988). The principle of these approximations is similar to that implemented in the composite likelihood methods. The computations are based on the Cholesky factor of the inverse covariance matrix which can be computed explicitly and is sparse by construction. They therefore allow numerically efficient inference and prediction. The `GpGp` library provides an implementation of the Vecchia approximation that uses an elaborate method for ordering and clustering data into conditionally independent blocks (Guinness, 2018; Katzfuss et al., 2020). We obtained the best predictive scores in both the sub-competitions 2a and 2b.

Full details on the methods that we used are provided in the Discussion on "Competition on Spatial Statistics for Large Datasets" (Allard et al., 2021), available as Supplementary Material in S.2.

8.2.3 2022 Edition

The 2022 edition of the "Spatial Statistics Competition for Large Datasets" featured 6 sub-competitions. The objective was to evaluate existing approximate prediction methods on large univariate and multivariate spatial and spatio-temporal datasets. A collection of large synthetic datasets (100K/1M data) was generated from a set of real models (2 univariate models, 2 spatio-temporal models, 2 bivariate models) using `ExaGeoStat` and each dataset was split into training and validation sets. The models exhibited spatial nonstationarity. All the details on the datasets and on the scoring rules, and the main methods used by the participants, are available in Abdulah et al. (2022).

We already know that the computation of the log-likelihood is prohibitive for large sample size N_S of spatial data, as the complexity of computing the inverse of the covariance matrix is $\mathcal{O}(N_S^3)$ and requires $\mathcal{O}(N_S^2)$ memory. This computation becomes even more expensive when considering the multivariate and spatio-temporal case because the sample size becomes pN_SN_T, where p is the number of spatial variables and N_T is the number of time points.

We participated in the sub-competitions 1a and 1b (univariate spatial data): we used the R package `GpGp` described in the previous section for parameter estimation and spatial prediction. We noted that the nonstationary structure changed from one

area to the other in the spatial domain. We therefore decided to split the spatial domain with different splitting configurations and fit an isotropic or anisotropic Matérn model for each considered region. The selection of the best splitting configuration and the best Matérn model was made by cross-validation. The option with the lowest mean square prediction error values was then finally chosen. The RESSTE team was ranked first in the sub-competition 1a and third in the sub-competition 1b. The team that obtained the best prediction accuracy in 1b adopted a deformation technique to map the locations from the original space to a latent space depending on two variables $r = \sqrt{s_x + s_y}$ and $\theta = \tan^{-1}\left(\sqrt{s_y/s_x}\right)$, where the data were assumed to be stationary. The prediction was performed using a deep feed-forward neural network model with two inputs (r and θ), four hidden layers, and one output (the objective function, i.e., the mean square predictive error). They used the Adam optimizer (Kingma and Ba, 2015) for the optimization of the objective function.

The spatio-temporal datasets of the sub-competition 2a and 2b were generated with 1K and 10K locations, and 100 time-slots. The spatio-temporal models were nonseparable, with different spatial and temporal dependencies. The teams were asked to make predictions in three different scenarios of leaving out space-time points: random spatial locations with all times left out (RS), random locations in space/time left out (RST) and all spatial locations left out on the last 10 time coordinates (T10).

We first performed an exploratory data analysis step and found that the data were generated from a positively non-separable space-time kernel. We then chose a covariance function from the Gneiting class, with a Matérn spatial covariance and a Cauchy temporal covariance. Due to the size of the given datasets, we used a block-composite likelihood to estimate the spatial and temporal parameters separately. Then, we used the estimated parameters to fit the space-time model and estimate the full set of parameters. To predict at the test locations we applied ordinary kriging conditioned on a subset of nearest data points in space and 3 preceding, current, and 3 following time-slots in time. We were ranked third in both the sub-competitions, but with a RMSE quite lower than the first one. Our results were good for the first two prediction scenarios (RS and RST), but not so good for the third scenario (T10). Indeed, the third scenario could be considered more like a multivariate time series forecast, where kriging is not necessarily the best prediction technique. The winning team, in fact, proposed a deep neural network for spatio-temporal predictions, by using basis functions to capture the spatio-temporal dependence. They used regression for the spatial interpolation step and they trained a long-short term memory network for the temporal forecast step. Their results for the first two predictive scenarios were worse than ours, but the LSTM made their temporal forecasts much more precise.

8.2.4 Concluding remarks

For decades, the "big n problem" in spatial data has been an active research area due to the challenges engendered by the real datasets, which often contain millions of observations, such as remote sensing data or numerical model outputs. Those datasets

are sometimes multivariate and/or spatio-temporal and need efficient methods for their prediction, both in the spatial interpolation and in the temporal forecasting frameworks. The competitions proposed by KAUST in the last years are a great tool to provide a unified framework for evaluating the performances of existing approximation methods and for suggesting future research paths. We appreciated the fact that dozens of research teams participated worldwide, each one proposing different techniques. Some of them proved to be great in some contexts and worse in others, and all the pros and cons of each method are now comparable in terms of prediction accuracy on the same synthetic datasets.

The computational time was never compared in the proposed competitions, because the execution time of each participating team on different machines was not directly comparable. However, this matter should be taken into account to better evaluate a given approach. In fact, often researchers and companies do not own machines that are able to treat massive datasets, or they do not want to use excessively consuming techniques for sustainable reasons. Sometimes, as we have shown in the previous sections, approximate methods have similar performances to exact methods with a much lower computation cost, which is a great asset for a thoughtful research tool.

The spatial and spatio-temporal models proposed in the first two editions of the "Spatial Statistics Competition for Large Datasets" are only a small part of possible and more challenging models. Assessing the results of the approximate methods proposed worldwide on real environmental datasets, that show different behaviors and dependence structures, could be a great tool to choose the best approach for each peculiar type of data. Another option would be to choose a benchmark dataset that at least presents a variety of issues that arise in practice. Another important point to take into account is the experimental design: in both the competitions the data were point data. It could be interesting to evaluate if the predictive performances of the methods change in presence of gridded data or clustered data. The case of gap filling, i.e., large areas with missing data, should also be considered.

This year, KAUST hosts a third competition that focuses on the construction of confidence and prediction intervals, instead of point estimation and prediction. The datasets were generated from stationary Gaussian random fields with an isotropic Matérn covariance function, with designs of irregularly spaced locations. The team RESSTE does not participate in the 2023 competition, but the results will certainly be interesting to get useful insights on approximate methods on those datasets.

There is still room for improvement in spatial statistics to obtain more efficient approximated methods for large spatial and spatio-temporal datasets, and this is a significant goal for the research community.

Chapter 9

Discussion and perspectives

Résumé

Dans ce dernier chapitre, nous resumons les résultats de la thèse en détaillant les contributions principales. Ensuite, nous proposons des pistes de recherche futures liées aux méthodes presentées.

9.1 Discussion

This dissertation explored some approaches to deal with spatio-temporal data in environmental science. Here we make some concluding remarks on the proposed approaches and propose interesting perspectives for future work.

The spatio-temporal SPDE approach based on unsteady advection-diffusion equations that we have proposed in Chapter 2 combines elements of physics, numerical analysis and statistics. It can be seen as a first step toward *physics-informed geostatistics*, which introduces physical dynamics into a statistical model, accounting for possible hidden structures governing the evolution of the spatio-temporal phenomenon. The different terms of the SPDE (advection, diffusion) directly influence the spatio-temporal dependencies of the process, by controlling its variability in space and time. Compared to spatio-temporal models built on covariance functions such as the Gneiting class (Gneiting, 2002), we gain in interpretability since the parameters of the model can be linked to the physical coefficients of SPDEs.

In Chapter 3 we showed that it is possible to build an accurate space-time approximation of the process driven by the advection-diffusion SPDE using a combination of FEM in space and implicit Euler scheme in time. We dealt with the advection-domination flow, by introducing a stabilization technique in the discretization approach. It leads to sparse structured linear systems. We obtained promising results for the simulation (Chapter 4), the estimation and the prediction (Chapter 5) of processes both in terms of precision and speed. When the size of the spatio-temporal mesh is moderate, direct matrix implementation is possible. We showed how matrix-free methods can be implemented in order to obtain scalable computations even for very large meshes. The application to the solar radiation and to the wind speed datasets (Chapter 6) demonstrated that the nonseparable advection-diffusion model

exhibited the best prediction performances on phenomenona that are governed by advection and diffusion processes.

The main results of these chapters are available in the arXiv paper Clarotto et al. (2022).

One of the main advantages of the SPDE formulation is that it is easy to generalize it to nonstationary settings. Nonstationary fields can be defined by letting the parameters of the advection-diffusion SPDE ($\kappa(\mathbf{s},t)$, $\boldsymbol{\gamma}(\mathbf{s},t)$, etc.) be space-time-dependent. This generalization implies only minimal changes to the method used in the stationary case concerning non-conditional simulations, but needs more work for estimation and prediction, since the maximum likelihood approach becomes much more expensive. We can also incorporate models of spatially varying anisotropy by modifying the general operator $\nabla \cdot \mathbf{H}(\mathbf{s},t)\nabla$ with a nonstationary anisotropic matrix $\mathbf{H}(\mathbf{s},t)$. The introduction of nonstationarities could allow us to better describe phenomena where local variations are clearly present, such as in the case of wind speed on complex topographic regions (see Section 6.3.1). The generalization of the approaches by Fuglstad et al. (2015) and Pereira et al. (2022) are being investigated and generalized to the spatio-temporal framework. A first sketch of ongoing work was presented in Chapter 7.

Further work is necessary to better assess the prediction accuracy and the computational complexity. Applications to larger and more complex datasets, in particular using the matrix-free approach, will be considered. Comparison to models expressing the advection in a Lagrangian framework (Ailliot et al., 2011; Benoit et al., 2018b; Salvaña and Genton, 2021) should also be performed. A frequentist maximum likelihood was implemented. As a follow-up work, it would be interesting to implement this space-time model as part of a Bayesian hierarchical construction, possibly within the INLA/SPDE framework (Rue et al., 2009; Krainski et al., 2018).

The models and the methodological framework developed in this thesis were implemented in Python by making use of the `FEniCS` library (Logg et al., 2012; Alnaes et al., 2015) for the task of computing the FEM solution of (S)PDEs. The implementation details are available in Appendix A.1.

A short-term goal is to implement the methods in open-source softwares, in particular in the library `gstlearn` (Python, R, C++) developed by the Geostatistics team of Mines Paris, where most of the building blocks of the proposed method are already present, and in the library `R-INLA` (Lindgren and Rue, 2015). Some spatio-temporal models, such as the DEMF model described in Section 2.2.8, are already implemented in `R-INLA`; extending this class of models with the unsteady advection-diffusion SPDE model would be interesting and would only require the coding of the ingredients that make our model different from DEMF (such as the FEM advection matrix, the stabilization, etc.). Linking our model to the INLA framework would benefit from all the tools already available in the library, such as the possibility of managing non-Gaussian models and Bayesian hierarchical models. Moreover, the model would be accessible to other research groups willing to model spatio-temporal

datasets with advection.

Another perspective is the development of computationally efficient algorithms for inference and prediction. We envisage the possibility of introducing parallelization using MPI sub-communicators as a way of obtaining faster computations in maximum likelihood estimation and prediction. We showed in Section 8.2 that some research teams in the spatial statistics community are already dealing with parallel codes and using modern runtime systems to allow efficient computations for dense matrices (Abdulah et al., 2018). Thus, we presume that extending these techniques to sparse matrices could be very useful for future applications, not only in the SPDE framework.

Another interesting consequence of defining the models through local stochastic partial differential equations is that the SPDEs still make sense when $\mathbb{R}^d$ is replaced by a space that is only locally flat. We can define nonstationary Gaussian fields on manifolds, and still obtain a GMRF representation. Important improvements were obtained in the spatial case (Pereira et al., 2022). The generalization to space-time processes could be explored further. Possible generalization to spatio-temporal SPDEs with a fractional exponent in the diffusion term could also be considered. A development of the methods proposed by Bolin and Kirchner (2020) and Vabishchevich (2015) should be explored.

A further interesting direction is the extension of the approach to new SPDEs able to better describe some peculiar physical phenomena. The current approach allows finding a sparse structure of the precision matrix of a space-time field solution of any SPDE that can be written as a sum of differential operators. If we had any information about the physics of a given spatio-temporal phenomenon we want to model, we could replace the advection-diffusion SPDE by another SPDE closer to the true physical equation of the process. An example could be the modeling of the motion of sea waves by the wave equation, as recently proposed in Henderson et al. (2023).

The contributions detailed in Chapter 8 enlarge the view on spatio-temporal modeling and prediction. The class of multivariate space-time covariance functions that we have proposed in Allard et al. (2022) (see Section 8.1 and Supplementary Material S.1) allows a generalization of the Gneiting class of space-time covariance functions to fully nonseparable functions, i.e., functions that are space-time nonseparable and that are neither proportional-in-space nor proportional-in-time. This is a major improvement to most of the models proposed up to date, where the marginal temporal correlation function is identical for all the variables, which is an overly restrictive hypothesis. The model we have proposed is continuous over $\mathbb{R}^d \times \mathbb{R}$, hence it is not restricted to a regular design of observations. More complex designs are in principle possible, such as irregular or incomplete records, heterotopic designs or even mobile measurement devices. We hope to see applications of our model to such datasets in a near future.

This work allowed me to see in practice how important it is to find a good trade-off between the flexibility of a multivariate spatio-temporal model, which must be able

to account for the complex interactions between space and time across several variables, and the number of parameters of the model, which should remain interpretable and not too numerous. A large number of parameters results in a time-consuming estimation, which often needs to be tackled with approximate techniques, such as the pairwise likelihood method that we used in our application. More efficient methods for parameter estimation probably exist (see for example the recent approaches by Lenzi et al. (2023) and Sainsbury-Dale et al. (2023)): finding the best one for this type of model would inevitably lead to faster computations and possible applications to larger datasets. In general, we should bear in mind that a more complex model should only be preferred if it is supported by the data.

Participating in the "Spatial Statistics Competition for Large Datasets" proposed by the King Abdullah University of Science and Technology (KAUST) in 2021 and 2022 (see Section 8.2) was a great way to get to know the current trends in spatial and spatio-temporal statistics for massive datasets, to propose approximate methods for parameter estimation and prediction and compare them with those of several statistics teams worldwide, to work in a research team and to enjoy a healthy competition. It was really important for my education, because it was the first scientific project where I was a researcher at the same level as the other members of the team and where my scientific ideas counted as much as theirs. Moreover, due to the good ranking of the RESSTE team in both the editions of the competition, I had the chance to spend a visiting period of 5 weeks at KAUST, working in the STDSD research group, under the supervision of Marc Genton, and to participate in the KAUST Workshop on Statistics in November 2022. During this research visit, I worked in collaboration with Marc Genton on applications of the SPDE method to a dataset analyzed in his team (wind speed in Saudi Arabia, see Section 6.3) and on comparisons with alternative methods. I participated in the daily activities of the team (team and department seminars, workshops...) and I exchanged with researchers from the other statistics teams (Håvard Rue, David Bolin, Ying Sun, Raphaël Huser) on topics of common interest, that could possibly lead to future collaborations.

We are increasingly interested in deep statistical learning for spatio-temporal environmental data. Although none of the works we have presented in this thesis make use of such methods, the interest of the scientific community in spatio-temporal statistics on this subject is growing and some approaches have already proven to be effective and efficient on large spatio-temporal datasets (Zammit-Mangion et al., 2022; Huang et al., 2022; Hu et al., 2022). We remark that, especially when we deal with a nonlinear temporal dynamics as it was the case for the wind data (see Section 6.3), deep neural networks are able to better capture a complex dynamics. As another example, we envisage to use deep learning methods to infer nonstationarities in the SPDE framework, hoping to obtain faster and more accurate estimations than with classical optimization methods. More generally, we intend to take advantage of the

predictive capabilities and computational speed of deep statistical learning to deal with high-dimensional spatio-temporal datasets.

Bibliography

Abdulah, S., Alamri, F., Nag, P., Sun, Y., Ltaief, H., Keyes, D., and Genton, M. G. (2022). The second competition on spatial statistics for large datasets. *Journal of Data Science*, 20(4):439–460.

Abdulah, S., Ltaief, H., Sun, Y., Genton, M. G., and Keyes, D. E. (2018). Exageostat: A high performance unified software for geostatistics on manycore systems. *IEEE Transactions on Parallel and Distributed Systems*, 29(12):2771–2784.

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

Allard, D., Clarotto, L., and Emery, X. (2022). Fully nonseparable gneiting covariance functions for multivariate space-time data. *Spatial Statistics*, 52:100706.

Allard, D., Clarotto, L., Opitz, T., and Romary, T. (2021). Discussion on "competition on spatial statistics for large datasets". *Journal of Agricultural, Biological and Environmental Statistics*, 26.

Alnaes, M. S., Blechta, J., J. Hake, A. J., Kehlet, B., Logg, A., Richardson, C., J. Ring, M. E. R., and Wells, G. N. (2015). The FEniCS project version 1.5. *Archive of Numerical Software*, 3.

Apanasovich, T. and Genton, M. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika*, 97(1):15–30.

Apanasovich, T. V., Genton, M. G., and Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107(497):180–193.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *WIREs Computational Statistics*, 10(6):e1443.

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

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(4):825–848.

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

Benoit, L., Allard, D., and Mariethoz, G. (2018a). Stochastic rainfall modeling at sub-kilometer scale. *Water Resources Research*, 54(6):4108–4130.

Benoit, L., Allard, D., and Mariethoz, G. (2018b). Stochastic rainfall modeling at sub-kilometer scale. *Water Resources Research*, 54(6):4108–4130.

Berild, M. O. and Fuglstad, G.-A. (2023). Spatially varying anisotropy for gaussian random fields in three-dimensional space. *Spatial Statistics*, 55:100750.

Bertino, L., Evensen, G., and Wackernagel, H. (2003). Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71(2):223–241.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Bevilacqua, M., Faouzi, T., Furrer, R., and Porcu, E. (2019). Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. *The Annals of Statistics*, 47(2):828 – 856.

Bevilacqua, M., Gaetan, C., Mateu, J., and Porcu, E. (2012). Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. *Journal of the American Statistical Association*, 107(497):268–280.

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

Bolin, D. and Kirchner, K. (2017). The rational SPDE approach for gaussian random fields with general smoothness. *Journal of Computational and Graphical Statistics*.

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

Bornn, L., Shaddick, G., and Zidek, J. (2012). Modeling nonstationary processes through dimension expansion. *Journal of The American Statistical Association*, 107:281–289.

Boulakia, M., Genadot, A., and Thieullen, M. (2015). Simulation of SPDEs for Excitable Media Using Finite Elements. *Journal of Scientific Computing*, 65:171–195.

Bourotte, M., Allard, D., and Porcu, E. (2016). A flexible class of non-separable cross-covariance functions for multivariate space–time data. *Spatial Statistics*, 18:125–146.

Brown, P. E., Roberts, G. O., Kåresen, K. F., and Tonellato, S. (2000). Blur-generated non-separable space–time models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):847–860.

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

Cao, Y., Yang, H., and Yin, L. (2007). Finite element methods for semilinear elliptic stochastic partial differential equations. *Numerische Mathematik*, 106:181–198.

Carrizo Vergara, R. (2018). *Development of geostatistical models using stochastic partial differential equations*. Theses, Université Paris sciences et lettres.

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

Chebyshev, P. L. (1853). *Théorie des mécanismes connus sous le nom de parallélogrammes*. Imprimerie de l'Académie impériale des sciencesCY - St. Pétersbourg.

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

Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New York.

Choi, J., Fuentes, M., Reich, B., and Davis, J. (2009). Multivariate spatial–temporal modeling and prediction of speciated fine particles. *Journal of Statistical Theory and Practice*, 3(2):407—418.

Clarotto, L., Allard, D., Romary, T., and Desassis, N. (2022). The spde approach for spatio-temporal datasets with advection and diffusion. *arXiv:2208.14015v3*.

Clenshaw, C. W. (1955). A note on the summation of chebyshev series. *Mathematics of Computation*, 9:118–120.

Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19:297–301.

Cox, D. R. and Isham, V. (1988). A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 415(1849):317–328.

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

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.

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

De Iaco, S., Myers, D., and Posa, D. (2003). The linear coregionalization model and the product–sum space–time variogram. *Mathematical Geology*, 35(1):25–38.

De Iaco, S. and Posa, D. (2013). Positive and negative non-separability for space–time covariance models. *Journal of Statistical Planning and Inference*, 143(2):378–391.

Demmel, J. W. (1987). On condition numbers and the distance to the nearest ill-posed problem. *Numerische Mathematik*, 51:251–290.

Donoghue, W. F. (1969). *Distributions and Fourier transforms*. Academic Press.

Dörr, K. and Schlather, M. (2021). Characterization theorems for pseudo-variograms. *arXiv:2112.04026*.

Edwards, M., Castruccio, S., and Hammerling, D. (2019). A Multivariate Global Spatiotemporal Stochastic Generator for Climate Ensembles. *Journal of Agricultural, Biological and Environmental Statistics*, 24(3):464–483.

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

Eymard, R., Gallouët, T., and Herbin, R. (2000). Finite volume methods. In *Solution of Equation in* $\mathbb{R}^n$ *(Part 3), Techniques of Scientific Computing (Part 3)*, volume 7 of *Handbook of Numerical Analysis*, pages 713–1018. Elsevier.

Fick, A. (1855). On liquid diffusion. *Annalen der Physik und Chemie*, 94(59).

Finazzi, F., Scott, E. M., and Fassò, A. (2013). A model-based framework for air quality indices and population risk evaluation, with an application to the analysis of Scottish air quality data. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 62(2):287.

Fouedjio, F. (2017). Second-order non-stationary modeling approaches for univariate geostatistical data. *Stochastic Environmental Research and Risk Assessment*, 31.

Fouedjio, F., Desassis, N., and Rivoirard, J. (2016). A generalized convolution model and estimation for non-stationary random functions. *Spatial Statistics*, 16:35–52.

Fouedjio, F., Desassis, N., and Romary, T. (2015). Estimation of space deformation model for non-stationary random functions. *Spatial Statistics*, 13:45–61.

Fox, L. and Parker, I. (1972). *Chebyshev Polynomials in Numerical Analysis*. Oxford mathematical handbooks. Oxford University Press.

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*, 102(477):321–331.

Fuglstad, G.-A. (2010). Approximating solutions of stochastic differential equations with gaussian markov random fields. Technical report, NTNU.

Fuglstad, G.-A. and Castruccio, S. (2020). Compression of climate simulations with a nonstationary global spatiotemporal SPDE model. *The Annals of Applied Statistics*, 14(2):542 – 559.

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

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

Garg, S., Singh, A., and Ramos, F. (2021). Learning non-stationary space-time models for environmental monitoring. *Proceedings of the AAAI Conference on Artificial Intelligence*, 26(1):288–294.

Gelfand, A. (2021). Multivariate spatial process models. In Fischer, M. M. and Nijkam, P., editors, *Handbook of Regional Science*, pages 1985–2016, Dordrecht. Springer.

Gelfand, A., Banerjee, S., and Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, 16(5):465—479.

Gentle, J. (2009). *Computational Statistics*. Statistics and Computing Series. Springer-Verlag.

Genton, M. G. and Kleiber, W. (2015). Cross-Covariance Functions for Multivariate Geostatistics. *Statistical Science*, 30(2):147 – 163.

Gerschgorin, S. (1931). Uber die abgrenzung der eigenwerte einer matrix. *Izvestija Akademii Nauk SSSR, Serija Matematika*, 7(3):749–754.

Giani, P., Tagle, F., Genton, M. G., Castruccio, S., and Crippa, P. (2020). Closing the gap between wind energy targets and implementation for emerging countries. *Applied Energy*, 269:115085.

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

Gneiting, T., Genton, M., and Guttorp, P. (2006). Geostatistical space-time models, stationarity, separability and full symmetry. *Statistical Methods for Spatio-temporal Systems (Monographs on Statistics and Applied Probability)*, 107.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491):1167–1177.

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

Guinness, J. (2018). Permutation and grouping methods for sharpening gaussian process approximations. *Technometrics*, 60(4):415–429.

Hausenblas, E. (2003). Approximation for semilinear stochastic evolution equations. *Potential Analysis*, 18:141–186.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., and Lindgr, F. (2019). A Case Study Competition Among Methods for Analyzing Large Spatial Data. *Journal of Agricultural, Biological and Environmental Statistics*, 24(3):398–425.

Heine, V. (1955). Models for two-dimensional stationary stochastic processes. *Biometrika*, 42:170–78.

Henderson, I., Noble, P., and Roustant, O. (2023). Characterization of the second order random fields subject to linear distributional pde constraints. *arXiv:2301.06895*.

Hestenes, M. R. and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards*, 49:409–435.

Higdon, D., Swall, J., and Kern, J. (1999). Non-stationary spatial modeling. *Bayesian Statistics*, 6(761-768).

Higham, N. (2008). *Functions of Matrices: Theory and Computation*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104).

Hu, W., Fuglstad, G.-A., and Castruccio, S. (2022). A stochastic locally diffusive model with neural network-based deformations for global sea surface temperature. *Stat*, 11(1):e431.

Huang, H., Abdulah, S., Sun, Y., Ltaief, H., Keyes, D. E., and Genton, M. G. (2021). Competition on Spatial Statistics for Large Datasets. *Journal of Agricultural, Biological and Environmental Statistics*, 26(4):580–595.

Huang, H., Castruccio, S., and Genton, M. G. (2022). Forecasting high-frequency spatio-temporal wind power with dimensionally reduced echo state networks. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 71(2):449–466.

Huang, H.-C. and Hsu, N.-J. (2004). Modeling transport effects on ground-level ozone using a non-stationary space–time model. *Environmetrics*, 15(3):251–268.

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

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

Ingebrigtsen, R., Lindgren, F., and Steinsland, I. (2014). Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, 8:20.

Ip, R. and Li, W. (2016). Matérn cross-covariance functions for bivariate spatio-temporal random fields. *Spatial Statistics*, 17(1):22–37.

Itô, K. (1954). Stationary random distributions. *Memoirs of the College of Science, University of Kyoto. Series A: Mathematics*, 28(3):209–223.

Jaeger, H. (2001). The "echo state" approach to analysing and training recurrent neural networks-with an erratum note'. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148.

Jones, R. H. and Zhang, Y. (1997). Models for continuous stationary space-time processes. In Gregoire, T. G., Brillinger, D. R., Diggle, P. J., Russek-Cohen, E., Warren, W. G., and Wolfinger, R. D., editors, *Modelling Longitudinal and Spatially Correlated Data*, pages 289–298. Springer New York.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440.

Katzfuss, M. and Guinness, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statistical Science*, 36(1):124–141.

Katzfuss, M., Jurek, M., Zilber, D., Gong, W., Guinness, J., Zhang, J., and Schäfer, F. (2020). GPvecchia: Fast Gaussian-process inference using Vecchia approximations. *R package version 0.1*, 3.

Kaufman, C. G., Schervish, M. J., and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555.

Khristenko, U., Scarabosio, L., Swierczynski, P., Ullmann, E., and Wohlmuth, B. (2019). Analysis of boundary effects on pde-based sampling of whittle–matérn random fields. *SIAM/ASA Journal on Uncertainty Quantification*, 7:948–974.

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Kleiber, W. and Nychka, D. (2012). Nonstationary modeling for multivariate spatial processes. *Journal of Multivariate Analysis*, 112:76–91.

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall/CRC.

Krainski, E. T. (2018). *Statistical Analysis of Space-time Data: New Models and Applications*. PhD thesis, NTNU.

Lang, A. and Potthoff, J. (2011). Fast simulation of gaussian random fields. In *Monte Carlo Methods Appl.*

Lenzi, A., Bessac, J., Rudi, J., and Stein, M. L. (2023). Neural networks for parameter estimation in intractable models. *Computational Statistics & Data Analysis*, page 107762.

Lenzi, A. and Genton, M. G. (2020). Spatiotemporal probabilistic wind vector forecasting over Saudi Arabia. *The Annals of Applied Statistics*, 14(3):1359 – 1378.

Lindgren, F., Bakka, H., Bolin, D., Krainski, E., and Rue, H. (2020). A diffusion-based spatio-temporal extension of gaussian matérn fields. *arXiv:2006.04917v3*.

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

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

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

Linka, K., Schäfer, A., Meng, X., Zou, Z., Karniadakis, G. E., and Kuhl, E. (2022). Bayesian physics informed neural networks for real-world nonlinear dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, 402:115346. A Special Issue in Honor of the Lifetime Achievements of J. Tinsley Oden.

Liu, X., Yeo, K., and Lu, S. (2020). Statistical modeling for spatio-temporal data from stochastic convection-diffusion processes. *Journal of the American Statistical Association*, pages 1–37.

Logg, A., Mardal, K., and Wells, G. N. (2012). *Automated Solution of Differential Equations by the Finite Element Method*. Springer.

Ma, C. (2002). Spatio-temporal covariance functions generated by mixtures. *Mathematical Geology*, 34(8):965–975.

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

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

Matern, B. (1986). *Spatial Variation*. Lecture Notes in Statistics. Springer New York.

Maxwell, J. C. (1867). Iv. on the dynamical theory of gases. *Philosophical Transactions of the Royal Society of London*, 157:49–88.

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

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

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

Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17.

Paciorek, C. J., Yanosky, J. D., Puett, R. C., Laden, F., and Suh, H. H. (2009). Practical large-scale spatio-temporal modeling of particulate matter concentrations. *The Annals of Applied Statistics*, 3(1):370–397.

Padoan, S. A. and Bevilacqua, M. (2015). Analysis of random fields using CompRandFld. *Journal of Statistical Software*, 63:1–27.

Papritz, A., Künsch, H., and Webster, R. (1993). On the pseudo cross-variogram. *Mathematical Geology*, 25(8):1015–1026.

Pereira, M. (2019). *Generalized random fields on Riemannian manifolds : theory and practice*. PhD thesis, Mines Paris, PSL University.

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

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

Perrin, O. and Senoussi, R. (2000). Reducing non-stationary random fields to stationarity and isotropy using a space deformation. *Statistics & Probability Letters*, 48(1):23–32.

Porcu, E., Emery, X., and Mery, N. (2022). Criteria and characterizations for spatially isotropic and temporally symmetric matrix-valued covariance functions. *Computational and Applied Mathematics*, 41:223.

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

Porcu, E., Gregori, P., and Mateu, J. (2006). Nonseparable stationary anisotropic space–time covariance functions. *Stochastic Environmental Research and Risk Assessment*, 21:113–122.

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

Qadir, G. A. and Sun, Y. (2022). Modeling and predicting spatio-temporal dynamics of pm _{2.5} concentrations through time-evolving covariance models. *arXiv:2202.12121*.

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

Raissi, M., Perdikaris, P., and Karniadakis, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707.

Risser, M. D. and Calder, C. A. (2017). Local likelihood estimation for covariance functions with spatially-varying parameters: The convospat package for r. *Journal of Statistical Software*, 81(14):1–32.

Rodrigues, A. and Diggle, P. (2010). A class of convolution based models for spatio-temporal processes with non-separable covariance structure. *Scandinavian Journal of Statistics*, 37(4):553–567.

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

Rouhani, S. and Wackernagel, H. (1990). Multivariate geostatistical approach to space–time data analysis. *Water Resources Research*, 36(4):585—591.

Rozanov, I., Rozanov, J., Elson, C., and Rozanov, Y. (1982). *Markov Random Fields*. Applications of mathematics. Springer New York.

Rozanov, Y. A. (1977). *Random Fields and Stochastic Partial Differential Equations*. Mathematics and Its Applications. Springer, Dordrecht, 1 edition.

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

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

Sainsbury-Dale, M., Zammit-Mangion, A., and Huser, R. (2023). Fast optimal estimation with intractable models using permutation-invariant neural networks. *arXiv:2208.12942v3*.

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

Sampson, P. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of The American Statistical Association*, 87:108–119.

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

Schmidt, A. M., Guttorp, P., and O'Hagan, A. (2011). Considering covariates in the covariance structure of spatial processes. *Environmetrics*, 22(4):487–500.

Schmidt, A. M. and O'Hagan, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(3):743–758.

Schwartz, L. (1966). *Théorie des distributions*. Hermann.

Shand, L. and Li, B. (2017). Modeling nonstationarity in space and time. *Biometrics*, 73(3):759–768.

Sigrist, F., Künsch, H. R., and Stahel, W. A. (2011). An autoregressive spatio-temporal precipitation model. *Procedia Environmental Sciences*, 3:2–7. 1st Conference on Spatial Statistics 2011 – Mapping Global Change.

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

Sigrist, F., Künsch, H. R., and Stahel, W. A. (2015). Stochastic partial differential equation based modelling of large space-time data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):3–33.

Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Barker, D. M., Duda, G., Huang, X., Wang, W., and Powers, G. (2008). A description of the advanced research wrf version 3. Technical report, University Corporation for Atmospheric Research.

Stefan, J. (1871). Über das gleichgewicht und die bewegung, insbesondere die diffusion von gasgemengen. *Sitzungsberichte der Mathematisch-Naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenschaften Wien, 2te Abteilung*, 63:63 – 124.

Stein, M. L. (1999). *Interpolation of spatial data*. Springer Series in Statistics. Springer-Verlag.

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

Stein, M. L., Chi, Z., and Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2):275–296.

Sun, Y., Li, B., and Genton, M. G. (2012). Geostatistics for large datasets. In Porcu, E., Montero, J.-M., and Schlather, M., editors, *Advances and Challenges in Space-time Modelling of Natural Events*, pages 55–77, Berlin, Heidelberg. Springer Berlin Heidelberg.

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

Tezduyar, T. (1992). Stabilized finite element formulations for incompressible flow computations. *Advances in Applied Mechanics*, 28:1–44.

Turing, A. M. (1948). Rounding-off errors in matrix processes. *The Quarterly Journal of Mechanics and Applied Mathematics*, 1(1):287–308.

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

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):297–312.

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

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

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

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

Xiong, Z., Simas, A. B., and Bolin, D. (2022). Covariance-based rational approximations of fractional SPDEs for computationally efficient bayesian inference. *arXiv:2209.04670v2*.

Xu, G. and Genton, M. G. (2017). Tukey g-and-h random fields. *Journal of the American Statistical Association*, 112(519):1236–1249.

Young, D. M. (1954). Iterative methods for solving partial difference equations of elliptic type. *Transactions of the American Mathematical Society*, 76:92–111.

Young, D. M. (1971). *Iterative Solution of Large Linear Systems*. Academic Press.

Zammit-Mangion, A. and Cressie, N. (2021). Frk: An r package for spatial and spatio-temporal prediction with large datasets. *Journal of Statistical Software*, 98(4):1–48.

Zammit-Mangion, A., Ng, T. L. J., Vu, Q., and Filippone, M. (2022). Deep compositional spatial models. *Journal of the American Statistical Association*, 117(540):1787–1808.

Appendices

A.1 Implementation

We here sketch the main implementation features of the methodology proposed in the dissertation. The main points of the implementation concern the following procedures:

- Simulation of (S)PDE models (including the unsteady advection-diffusion model with and without stabilization);
- Estimation of SPDE model parameters;
- Prediction and conditional simulations of spatio-temporal SPDE models applied to real data;
- Nonstationary extension of the approach.

Reproducible scripts will be made available at `http://github.com/luciclar/ST_advdiff_SPDE` by October 2023.

A.1.1 Simulation

As detailed in Chapter 4, the simulation of SPDE models is based on the discretization approach of Section 3.3. The discretization based on the FEM in space and the FDM in time is obtained by coding the implicit Euler scheme directly in Python and by approximately solving the purely spatial SPDE at each time step by finite elements.

This last task is made using the `FEniCS` software (Logg et al., 2012; Alnaes et al., 2015). `FEnICS` is an open-source software that allows for automated ways of implementing the FEM code for partial differential equations from the variational formulation of a differential equation. Using the `FEnICS` platform in Python we implement the spatial (S)PDE by writing the weak form detailed in Equation (3.11). We set the parameters of the SPDE and the parameters of the numerical approximation (size of the spatial domain and of the time window, mesh structure, mesh size h, boundary conditions). The FEM matrices (mass matrix $\mathbf{M}$, lumped mass matrix $\tilde{\mathbf{M}}$, stiffness matrix $\mathbf{G}$ (isotropic or anisotropic), advection matrix $\mathbf{B}$ and Streamline Diffusion stabilization matrix $\mathbf{S}$, defined in Section 3.3) are computed by `FEnICS` and stored as sparse matrices with the package `scipy.sparse.coo_matrix`. The rest of the implementation is made in Python, by using the sparse matrices retrieved from `FEnICS`. For the noise term it is sufficient to sample a Gaussian standard vector and to apply the corresponding multiplicative matrix. The LU decomposition of

the sparse matrix in the left-hand side of system (3.7) is made using the function `scipy.sparse.linalg.splu`. Concerning the sparse global precision matrix $\mathbf{Q}$ defined in Equation (3.22), either we store it in the default Python sparse format (not necessarily the most convenient format in our specific case) or we store each of its building blocks (3 different blocks for the diagonal and 1 block for the upper and lower diagonal) in a sparse format itself.

Simulations of a spatio-temporal field, solution to the SPDE, are obtained by setting the parameters of the SPDE to estimated or chosen values. The features of the covariance structures of those fields are studied by means of comparisons with known covariance functions (the Matérn, the exponential) and with the Fourier basis functions approximation, detailed in Section 4.2 and coded entirely in Python.

A.1.2 Estimation

The estimation approach proposed in Chapter 5 is implemented in Python by coding the different functions necessary to compute the terms of the log-likelihood function.

First, the projection matrix $\mathbf{A}$ is created from the mesh defined in `FEnICS` and the locations of the data points.

The matrix-free methods (matrix-vector multiplication for block-structured matrices, temporal-block Gauss-Seidel preconditioner, Chebyschev approximation of the logarithm function, Hutchinson's estimator of the trace) are coded from scratch. The matrix-vector multiplication for block-structured matrices and the temporal-block Gauss-Seidel preconditioner are coded using the `functools.partial` function and by coding the single operations between each block in a custom-made function. Then, both matrix-free methods are transformed to the LinearOperator format by using the function `scipy.sparse.linalg.LinearOperator`. The Conjugate Gradient method is available in the function `scipy.sparse.linalg.cg`, where the matrix-free product and the matrix-free preconditioner are input in the LinearOperator format.

The code allows switching from the direct matrix implementation based on the Cholesky decomposition of the spatio-temporal matrices (obtained with the function `scipy.sparse.linalg.cholesky`) to the matrix-free approach, when the spatio-temporal mesh is too large for the Cholesky decomposition. The maximization of the log-likelihood is made with the `scipy.optimize` package with the BFGS method.

The variogram fit for defining the initial values of the simulation study is obtained with the Python library `gstlearn`.

A.1.3 Prediction and application to real dataset

The spatial meshes are constructed using the `FEniCS` implemented `mesh` function, either in the default mode or by editing it to obtain more complex meshes.

The ML optimization is obtained by using the functions detailed in the previous section.

The kriging and the conditional simulations are coded in Python by following the schemes detailed in Section 5.3 and 5.4.

The plots are made with `matplotlib`.

A.1.4 Nonstationarity

The simulation approach in the nonstationary extension of the SPDE model is obtained by including spatially varying parameters in the SPDE with the `Expression` function available in `FEniCS`. The corresponding spatial FEM matrices are then obtained.

The estimation approach is similar to the one detailed above, but includes more parameters in the case of the parameterization of the advection vector with a defined function of with the B-splines basis. The basic B-splines are computed by `scipy.interpolate.BSpline`. The 2D B-splines basis is custom-made.

A.2 Analytical gradient of the log-likelihood

We here sketch the analytical formula of the different terms of the gradient of the log-likelihood 5.2, reported here:

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

By setting $\boldsymbol{\eta}\,\mathbf{b} = \boldsymbol{\mu}$, $\mathbf{Q}_C = \mathbf{Q} + \sigma_0^{-2}\mathbf{A}\mathbf{A}^\top$ and $\boldsymbol{\mu}_C = \sigma_0^{-2}\mathbf{Q}_C^{-1}\mathbf{A}(\mathbf{y}_{1:N_T} - \boldsymbol{\mu})$, we can rewrite the equation above as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\psi}) = -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log\left(\sigma_0^{-2}\right) + \frac{1}{2}\left(\log|\mathbf{Q}| - \log|\mathbf{Q}_C|\right) - \frac{\sigma_0^{-2}}{2}(\mathbf{y}_{1:N_T} - \boldsymbol{\mu})^\top(\mathbf{y}_{1:N_T} - \boldsymbol{\mu}) \\ + \frac{\sigma_0^{-4}}{2}(\mathbf{y}_{1:N_T} - \boldsymbol{\mu})^\top\,\mathbf{A}^\top\,\mathbf{Q}_C^{-1}\,\mathbf{A}(\mathbf{y}_{1:N_T} - \boldsymbol{\mu}).\end{aligned}$$

By observing that

$$\frac{\partial \mathbf{Q}_C}{\partial \theta_i} = \frac{\partial \mathbf{Q}}{\partial \theta_i} \quad \text{and} \quad \frac{\partial \mathbf{Q}_C}{\partial \log\left(\sigma_0^{-2}\right)} = \sigma_0^{-2}\mathbf{A}\mathbf{A}^\top,$$

we have the following formulas for the gradients of each term of the log-likelihood with respect to each θ_i and to $\log\left(\sigma_0^{-2}\right)$.

$$\frac{\partial}{\partial \theta_i}\left(\log|\mathbf{Q}| - \log|\mathbf{Q}_C|\right) = \operatorname{tr}\left(\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \theta_i}\right) - \operatorname{tr}\left(\mathbf{Q}_C^{-1}\frac{\partial \mathbf{Q}}{\partial \theta_i}\right) = \operatorname{tr}\left((\mathbf{Q}^{-1} - \mathbf{Q}_C^{-1})\frac{\partial \mathbf{Q}}{\partial \theta_i}\right)$$

$$\frac{\partial}{\partial \log\left(\sigma_0^{-2}\right)}\left(\log|\mathbf{Q}| - \log|\mathbf{Q}_C|\right) = -\operatorname{tr}\left(\mathbf{Q}_C^{-1}\frac{\partial}{\partial \log\left(\sigma_0^{-2}\right)}\mathbf{Q}_C\right) = -\operatorname{tr}\left(\mathbf{Q}_C^{-1}\sigma_0^{-2}\mathbf{A}\mathbf{A}^\top\right)$$

$$\frac{\partial}{\partial \theta_i}\sigma_0^{-2}(\mathbf{y}_{1:N_T} - \boldsymbol{\mu})^\top(\mathbf{y}_{1:N_T} - \boldsymbol{\mu}) = 0$$

$$\frac{\partial}{\partial \log\left(\sigma_0^{-2}\right)}\sigma_0^{-2}(\mathbf{y}_{1:N_T} - \boldsymbol{\mu})^\top(\mathbf{y}_{1:N_T} - \boldsymbol{\mu}) = \sigma_0^{-2}(\mathbf{y}_{1:N_T} - \boldsymbol{\mu})^\top(\mathbf{y}_{1:N_T} - \boldsymbol{\mu})$$

$$
\begin{aligned}
&\frac{\partial}{\partial \theta_i}\sigma_0^{-4}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu})^\top \mathbf{A}^\top \mathbf{Q}_C^{-1}\mathbf{A}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu}) = \\
&\quad = -\sigma_0^{-4}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu})^\top \mathbf{A}^\top \mathbf{Q}_C^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \theta_i}\right)\mathbf{Q}_C^{-1}\mathbf{A}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu}) \\
&\quad = -\boldsymbol{\mu}_C^\top \left(\frac{\partial \mathbf{Q}}{\partial \theta_i}\right)\boldsymbol{\mu}_C \\
&\frac{\partial}{\partial \log\left(\sigma_0^{-2}\right)}\sigma_0^{-4}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu})^\top \mathbf{A}^\top \mathbf{Q}_C^{-1}\mathbf{A}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu}) = \\
&\quad = -\sigma_0^{-4}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu})^\top \mathbf{A}^\top \mathbf{Q}_C^{-1}\left(\frac{\partial \mathbf{Q}_C}{\partial \log\left(\sigma_0^{-2}\right)}\right)\mathbf{Q}_C^{-1}\mathbf{A}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu}) \\
&\quad = -\sigma_0^{-6}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu})^\top \mathbf{A}^\top \mathbf{Q}_C^{-1}\mathbf{A}\mathbf{A}^\top \mathbf{Q}_C^{-1}\mathbf{A}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu}) \\
&\quad = -\sigma_0^{-2}\boldsymbol{\mu}_C^\top \mathbf{A}\mathbf{A}^\top \boldsymbol{\mu}_C
\end{aligned}
$$

Hence, finally, the gradient of the log-likelihood with respect to θ_i is

$$
\frac{\partial}{\partial \theta_i}\mathcal{L}(\boldsymbol{\psi}) = \frac{1}{2}\mathrm{tr}\left((\mathbf{Q}^{-1}-\mathbf{Q}_C^{-1})\frac{\partial \mathbf{Q}}{\partial \theta_i}\right) - \frac{1}{2}\boldsymbol{\mu}_C^{\mathrm{T}}\left(\frac{\partial \mathbf{Q}}{\partial \theta_i}\right)\boldsymbol{\mu}_C, \tag{A.1}
$$

and the gradient with respect to σ_0^{-2} is

$$
\begin{aligned}
\frac{\partial \mathcal{L}(\boldsymbol{\psi})}{\partial \log\left(\sigma_0^{-2}\right)} =& \frac{n}{2} - \frac{1}{2}\mathrm{tr}\left(\mathbf{Q}_C^{-1}\sigma_0^{-2}\mathbf{A}\mathbf{A}^\top\right) - \frac{\sigma_0^{-2}}{2}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu})^{\mathrm{T}}(\mathbf{y}_{1:N_T}-\boldsymbol{\mu}) + \\
&- \frac{\sigma_0^{-2}}{2}\boldsymbol{\mu}_C^{\mathrm{T}}\mathbf{A}\mathbf{A}^\top\boldsymbol{\mu}_C
\end{aligned} \tag{A.2}
$$

We remark that in Equation (A.1), we need to compute the derivative of $\mathbf{Q}$ with respect to θ_i. By calling $\mathbf{J} = \left(\mathbf{M} + \frac{dt}{c}(\mathbf{K}+\mathbf{B}+\mathbf{S})\right)$, we recall that the matrices $\mathbf{D}$ and $\mathbf{F}$, which are the building blocks of the precision matrix $\mathbf{Q}$ and were defined in Section 3.3.2, satisfy

$$
\begin{aligned}
\mathbf{D} &= \mathbf{J}^{-1}\,\mathbf{M}, \\
\mathbf{F}^{-1} &= \frac{c}{\tau^2 dt}\,\mathbf{J}^\top\,\mathbf{M}^{-1}\,\mathbf{Q}_S\,\mathbf{M}^{-1}\,\mathbf{J}, \\
\mathbf{D}^\top\,\mathbf{F}^{-1} &= \frac{c}{\tau^2 dt}\,\mathbf{Q}_S\,\mathbf{M}^{-1}\,\mathbf{J}\,.
\end{aligned}
$$

When computing the derivative of $\mathbf{Q}$ with respect to θ_i, we can work on each block of the matrix. We observe that the derivative of the first temporal block $\mathbf{Q}^{(1,1)}$ is

$$
\frac{\partial}{\partial \theta_i}\,\mathbf{Q}^{(1,1)} = \frac{c}{\tau^2 dt}\frac{\partial \mathbf{Q}_S}{\partial \theta_i}
$$

The derivative of the upper diagonal blocks $\mathbf{Q}^{(i,j)}$ for $j = i + 1$ is

$$\frac{\partial}{\partial \theta_i}\mathbf{Q}^{(i,j)} = \frac{\partial}{\partial \theta_i}\frac{c}{\tau^2 dt}(-\mathbf{D}^\top \mathbf{F}^{-1}) = \frac{\partial}{\partial \theta_i}\frac{c}{\tau^2 dt}\left(-\mathbf{Q}_S \mathbf{M}^{-1}\mathbf{J}\right) = \\ = -\frac{c}{\tau^2 dt}\mathbf{M}^{-1}\left(\frac{\partial \mathbf{Q}_S}{\partial \theta_i}\mathbf{J} + \mathbf{Q}_S\frac{\partial \mathbf{J}}{\partial \theta_i}\right),$$

and similarly for the transposed blocks $\mathbf{Q}^{(i,j)}$ for $i = j + 1$. For the blocks on the diagonal, $\mathbf{Q}^{(i,i)}$ for $2 \leq i \leq N_T - 1$, we have

$$\frac{\partial}{\partial \theta_i}\mathbf{Q}^{(i,i)} = \frac{\partial}{\partial \theta_i}\frac{c}{\tau^2 dt}\left(\mathbf{J}^\top \mathbf{M}^{-1}\mathbf{Q}_\mathrm{S}\mathbf{M}^{-1}\mathbf{J} + \mathbf{Q}_\mathrm{S}\right) = \\ = 2\frac{c}{\tau^2 dt}\left(\frac{\partial \mathbf{J}}{\partial \theta_i}\right)^\top \mathbf{M}^{-1}\mathbf{Q}_\mathrm{S}\mathbf{M}^{-1}\mathbf{J} + \mathbf{J}^\top \mathbf{M}^{-1}\left(\frac{\partial \mathbf{Q}_\mathrm{S}}{\partial \theta_i}\right)\mathbf{M}^{-1}\mathbf{J} + \frac{\partial \mathbf{Q}_\mathrm{S}}{\partial \theta_i},$$

and for the last diagonal block $\mathbf{Q}^{(N_T,N_T)}$, the last term in the above expressions is removed.

The derivatives that arise from the previous equations are those of $\mathbf{Q}_S$ and $\mathbf{J}$. If the parameter controls part of the temporal range, e.g., τ or c, the derivative of the former is zero, and the derivative of the latter is a simple derivative of one of its matrix terms. If the parameter is part of the diffusion, then we have

$$\frac{\partial \mathbf{Q}_S}{\partial \theta_i} = 2\left(\frac{\partial \mathbf{K}}{\partial \theta_i}\right)^\top \mathbf{M}^{-1}\mathbf{K},$$

so we would have further nested derivatives here.

We don't go into further details, but there is an analytical way of computing those gradients for each parameter θ_i. If we were also able to compute those gradients with respect to each of the parameters parameterizing any spatially and/or temporally varying parameters field, we could obtain analytical formulas even in the case nonstationary SPDE models. This extension is left for future work.

Supplementary material

S.1 Fully nonseparable Gneiting covariance functions for multivariate space-time data

Spatial Statistics 52 (2022) 100706

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Fully nonseparable Gneiting covariance functions for multivariate space–time data

Denis Allard [a,*], Lucia Clarotto [b], Xavier Emery [c,d,1]

[a] *Biostatistics and Spatial Processes (BioSP), INRAE, 84914 Avignon, France*
[b] *Centre for geosciences and geoengineering, Mines Paris, PSL University, 77300 Fontainebleau, France*
[c] *Department of Mining Engineering, University of Chile, Avenida Beauchef 850, Santiago 8370448, Chile*
[d] *Advanced Mining Technology Center, University of Chile, Avenida Beauchef 850, Santiago 8370448, Chile*

ARTICLE INFO

Article history:
Received 28 May 2022
Received in revised form 14 September 2022
Accepted 14 September 2022
Available online 21 September 2022

Keywords:
Spatio-temporal modeling
Matrix-valued covariance function
Pseudo-variogram
Matérn covariance
Spectral simulation

ABSTRACT

We broaden the well-known Gneiting class of space–time covariance functions by introducing a very general parametric class of fully nonseparable direct and cross-covariance functions for multivariate random fields, where each component has a spatial covariance function from the Matérn family with its own smoothness and scale parameters and, unlike most of currently available models, its own correlation function in time. We present sufficient conditions that result in valid models with varying degrees of complexity and we discuss the parameterization of those. Continuous-in-space and discrete-in-time simulation algorithms are also given, which are not limited by the number of target spatial coordinates and allow tens of thousands of time coordinates. The application of the proposed model is illustrated on a weather trivariate dataset over France. Our new model yields better fitting and better predictive scores in time compared to a more parsimonious model with a common temporal correlation function.

1. Introduction

An increasing amount of multivariate data indexed by Euclidean space–time coordinates is available in numerous scientific and engineering applications, including atmospheric and

* Corresponding author.
E-mail address: denis.allard@inrae.fr (D. Allard).

[1] The author acknowledges the funding of the National Agency for Research and Development of Chile, through grants ANID/FONDECYT/REGULAR/No. 1210050 and ANID PIA AFB180004.

https://doi.org/10.1016/j.spasta.2022.100706
2211-6753/

environmental sciences, geosciences, as well as geological and mining engineering. For statisticians analyzing these data, one of the key issues is to model the space–time dependence structure not only for each variable separately, but also across the variables. This requires versatile models that can account for different scale and smoothness parameters for each variable, and yet whose parameters can be accurately estimated. Bourotte et al. (2016) proposed valid multivariate space–time classes based on Matérn and Cauchy spatial covariance functions, inspired by the multivariate Matérn model in Gneiting et al. (2010) and Apanasovich et al. (2012). For both classes, each variable has its own scale and own degree of smoothness in space, while allowing for some degree of cross-correlation. However, in this construction, the marginal temporal correlation function is identical for all variables, which, as discussed in Bourotte et al. (2016), was already seen as a restrictive assumption because it was found that the time correlations of the three variables in the analyzed dataset were different from one variable to the other. To account for different time correlations, space–time linear models of coregionalizations can be proposed, as in Rouhani and Wackernagel (1990), De Iaco et al. (2003), Choi et al. (2009) or Finazzi et al. (2013). However, there are two major drawbacks to this construction. First, the smoothness of any component of the multivariate field is restricted to that of the roughest underlying univariate field (Gneiting et al., 2010). Second, the number of parameters to be estimated increases quickly as the number of components of the model increases, thus raising issues of robustness and over-fitting. Gelfand et al. (2005), Apanasovich and Genton (2010), Rodrigues and Diggle (2010), Ip and Li (2016) and Gelfand (2021) introduced Bayesian dynamic, latent dimension, convolution and spectral approaches, respectively, to build valid matrix-valued space–time covariance functions with entries that can have different smoothness parameters. Recent comprehensive accounts of space–time univariate and multivariate covariance modeling can be found in Chen et al. (2021) and in Porcu et al. (2021).

In this work, we propose new classes of matrix-valued space–time covariance functions with Matérn spatial traces that allow, for each variable, different correlation functions in time and different smoothness and scale parameters in space. Our main contribution is to establish sufficient validity conditions for a very general class of covariance functions of the type

$$C_{ij}(\boldsymbol{h}, u) = \sigma_{ij}(u)\mathcal{M}(\boldsymbol{h}; r_{ij}(u), \nu_{ij}(u)), \qquad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R},\ i, j = 1, \ldots, p,$$

where $\mathcal{M}(\cdot; r, \nu)$ denotes a univariate Matérn covariance function with scale and smoothness parameters r and ν. We give several constructions that offer new modeling possibilities, in particular negatively nonseparable covariance functions.

The outline of the paper is the following. Section 2 provides the necessary background material on matrix-valued space–time covariance functions and on matrix-valued pseudo-variograms. Section 3 proposes a first, parsimonious construction based on the substitution approach presented in Allard et al. (2020), while Section 4 proposes a more general, fully nonseparable, multivariate Gneiting–Matérn class. Section 5 shows how this model can be implemented on a trivariate weather dataset and improves the fitting and prediction performances over an existing more parsimonious model. Concluding remarks follow in Section 6. Other general forms of the Gneiting–Matérn model together with simulation algorithms, technical lemmas and proofs are given in appendices.

2. Background material

Throughout this work, we use roman letters for scalars and bold letters for vectors, matrices and matrix-valued functions. In particular, $\mathbf{0}$ and $\mathbf{1}$ denote vectors or matrices of all-zeros and all-ones of appropriate dimension, respectively. Also, all matrix operations (product, ratio, power, square root, exponential, etc.) are understood as elementwise.

2.1. Some matrix definitions and notation

For $\boldsymbol{a} = [a_{ij}]_{i,j=1}^{p}$ a symmetric real matrix, we write $\boldsymbol{a} \succeq 0$ (resp. $\boldsymbol{a} \preceq 0$) to indicate that $\boldsymbol{a}$ is positive (resp. negative) semidefinite, i.e., $\sum_{i=1}^{p}\sum_{j=1}^{p} \omega_i \omega_j a_{ij} \geq 0$ (resp. ≤ 0), $\forall\, \omega_1, \ldots, \omega_p \in \mathbb{R}$. Likewise, $\boldsymbol{a} \succeq_c 0$ (resp. $\boldsymbol{a} \preceq_c 0$) indicates that $\boldsymbol{a}$ is conditionally positive (resp. negative) semidefinite,

i.e., $\sum_{i=1}^{p}\sum_{j=1}^{p}\omega_i\,\omega_j\,a_{ij}\geq 0$ (resp. ≤ 0), $\forall\,\omega_1,\ldots,\omega_p\in\mathbb{R}$ such that $\sum_{i=1}^{p}\omega_i=0$. A conditionally null definite matrix is both conditionally positive and conditionally negative semidefinite. An addition separable (a-separable) matrix is such that $2a_{ij}=a_{ii}+a_{jj}$ for all $i,j=1,\ldots,p$, and a product separable (p-separable) matrix is such that $a_{ij}^2=a_{ii}a_{jj}$ for all $i,j=1,\ldots,p$. It is straightforward to prove that any a-separable matrix is conditionally null definite and that any p-separable matrix with nonnegative entries is positive semidefinite, see Appendix D.1.

2.2. Matrix-valued space–time covariance functions

We consider a multivariate space–time random field $\boldsymbol{Z}(\boldsymbol{s},t)=[Z_i(\boldsymbol{s},t)]_{i=1}^{p}$ defined in $\mathbb{R}^d\times\mathbb{R}$, where d is the space dimension and p the number of random field components, each being real-valued. A space–time coordinate will be denoted $(\boldsymbol{s},t)$, with $\boldsymbol{s}\in\mathbb{R}^d$ and $t\in\mathbb{R}$. For the sake of light notation, the Euclidean norm of $\boldsymbol{h}$ will be denoted $|\boldsymbol{h}|$. Without loss of generality, we assume that all random field components are centered, i.e., $\mathbb{E}[Z_i(\boldsymbol{s},t)]=0\ \forall(\boldsymbol{s},t)\in\mathbb{R}^d\times\mathbb{R}$, $\forall i=1,\ldots,p$. It will also be assumed that the multivariate random field $\boldsymbol{Z}(\boldsymbol{s},t)$ is second-order stationary, so that its covariance functions exist and depend only on the space–time lag $(\boldsymbol{h},u)\in\mathbb{R}^d\times\mathbb{R}$ (Chilès and Delfiner, 2012):

$$\mathrm{Cov}(Z_i(\boldsymbol{s},t),Z_j(\boldsymbol{s}+\boldsymbol{h},t+u))=C_{ij}(\boldsymbol{h},u), \tag{1}$$

for any pair $i,j=1,\ldots,p$, $\forall(\boldsymbol{s},\boldsymbol{s}+\boldsymbol{h})\in\mathbb{R}^d\times\mathbb{R}^d$ and $\forall(t,t+u)\in\mathbb{R}\times\mathbb{R}$. The functions C_{ij} are called direct covariance functions when $i=j$ and cross-covariance functions otherwise. The matrix-valued covariance function $\boldsymbol{C}$, which associates each space–time lag $(\boldsymbol{h},u)$ with the $p\times p$ matrix $\boldsymbol{C}(\boldsymbol{h},u)=\left[C_{ij}(\boldsymbol{h},u)\right]_{i,j=1}^{p}$, is positive semidefinite in $\mathbb{R}^d\times\mathbb{R}$, that is, for any finite collection of space–time coordinates $(\boldsymbol{s}_k,t_k)_{k=1}^{N}$, the matrix $\left[[C_{ij}(\boldsymbol{s}_l-\boldsymbol{s}_k,t_l-t_k)]_{i,j=1}^{p}\right]_{k,l=1}^{N}$ is positive semidefinite (Wackernagel, 2003). Setting $u=0$ in (1) defines the purely spatial matrix-valued covariance function $\boldsymbol{C}_S(\boldsymbol{h}):=\boldsymbol{C}(\boldsymbol{h},0)$. Likewise, $\boldsymbol{C}_T(u):=\boldsymbol{C}(\boldsymbol{0},u)$ is a purely temporal matrix-valued covariance function.

A matrix-valued space–time covariance function is space–time separable if it is the elementwise product of a matrix-valued spatial covariance function with a matrix-valued temporal covariance function, i.e., when $C_{ij}(\boldsymbol{h},u)=C_{ij}(\boldsymbol{0},0)^{-1}\,C_{S,ij}(\boldsymbol{h})\,C_{T,ij}(u)$ for all $i,j=1,\ldots,p$. The direct covariance C_{ii} is said to be positively (resp. negatively) space–time nonseparable if $S_{ii}(\boldsymbol{h},u)=C_{ii}(\boldsymbol{h},u)C_{ii}(\boldsymbol{0},0)-C_{S,ii}(\boldsymbol{h})\,C_{T,ii}(u)\geq 0$ (resp. ≤ 0) for all $(\boldsymbol{h},u)\in\mathbb{R}^d\times\mathbb{R}$ (Rodrigues and Diggle, 2010; De Iaco and Posa, 2013). The functions $S_{ii}(\boldsymbol{h},u)$ will be referred to as separability functions. Their empirical versions can be used as exploratory tools to characterize the type of nonseparability in a dataset.

A matrix-valued space–time covariance function is said to be a proportional model when it is obtained as the product of a $p\times p$ covariance matrix $\boldsymbol{R}$ and a univariate space–time correlation function ρ, i.e., $\boldsymbol{C}(\boldsymbol{h},u)=\boldsymbol{R}\,\rho(\boldsymbol{h},u)$, $\forall(\boldsymbol{h},u)\in\mathbb{R}^d\times\mathbb{R}$. It is proportional-in-time if $\boldsymbol{C}_T(u)=\boldsymbol{R}\,\rho_T(u)$ and proportional-in-space if $\boldsymbol{C}_S(\boldsymbol{h})=\boldsymbol{R}\,\rho_S(\boldsymbol{h})$, $\forall(\boldsymbol{h},u)\in\mathbb{R}^d\times\mathbb{R}$. Bourotte et al. (2016) proposed nonproportional matrix-valued space–time covariance functions in which the purely spatial part is a nonproportional model but the purely temporal part is a proportional model, i.e., the model is proportional-in-time but not in space. In this work, we will build valid and flexible parametric classes of matrix-valued space–time covariance functions such that both the spatial and temporal components are nonproportional. Such matrix-valued space–time covariances are referred to as being *fully nonseparable*.

2.3. The Gneiting class of space–time covariance functions

We first restrict ourselves to the univariate setting, i.e., $p=1$. We thus choose to drop temporarily the use of subscripts for the ease of notation. The Gneiting class of spatio-temporal covariance functions was originally presented in Gneiting (2002). For an easier exposition of our results, we follow (Allard et al., 2020) and consider the extended class of functions of the form

$$C(\boldsymbol{h},u)=\frac{1}{(\gamma(u)+1)^{d/2}}\varphi\left(\frac{|\boldsymbol{h}|^2}{\gamma(u)+1}\right),\qquad(\boldsymbol{h},u)\in\mathbb{R}^d\times\mathbb{R}, \tag{2}$$

which is a second-order stationary covariance function for any continuous and completely monotone function φ on $[0,\infty)$ if and only if the function $\gamma : \mathbb{R} \to [0,\infty)$ is a variogram on $\mathbb{R}$, i.e., a conditionally negative semidefinite function (Chilès and Delfiner, 2012). The Gneiting class of spatio-temporal covariances thus involves two functions φ and γ: the former is associated with the spatial structure, since $C_S(\boldsymbol{h}) = \varphi(|\boldsymbol{h}|^2)$, whereas the latter is associated with the temporal structure, with $C_T(u) = (\gamma(u)+1)^{-d/2}$. Notice that $C_T(u) \to 0$ as $|u| \to \infty$ if and only if $\gamma(u)$ is unbounded. From a modeling point of view, the formulation (2) offers more flexibility than the original formulation in Gneiting (2002) and provides a direct geostatistical interpretation in the temporal dimension. Examples of such classes include the Gneiting–Matérn and Gneiting–Cauchy covariance functions, in the cases where $\varphi(|\boldsymbol{h}|^2)$ is the Matérn or the Cauchy spatial covariance, respectively. The expression of the Matérn covariance is

$$\mathcal{M}(\boldsymbol{h}; r, \nu) = \frac{2^{1-\nu}}{\Gamma(\nu)} (r|\boldsymbol{h}|)^{\nu} \mathcal{K}_{\nu}(r|\boldsymbol{h}|), \quad \boldsymbol{h} \in \mathbb{R}^d, \tag{3}$$

where $r > 0$ is a scale parameter ($1/r$ is called the range), $\nu > 0$ is a smoothness parameter and $\mathcal{K}_\nu$ denotes the modified Bessel function of the second kind of order ν. With $\alpha, r > 0$, the Cauchy covariance is $\mathcal{C}(\boldsymbol{h}; r, \alpha) = \left(1 + r|\boldsymbol{h}|^2\right)^{-\alpha}$, $\mathbf{h} \in \mathbb{R}^d$. Rodrigues and Diggle (2010) and De Iaco and Posa (2013) showed that the Gneiting class of covariance functions (2) cannot accommodate negative nonseparability.

2.4. Proportional-in-time multivariate Gneiting classes

Going back to the multivariate setting, Bourotte et al. (2016) proposed classes of matrix-valued space–time covariance functions that are extensions of the Gneiting class in (2), where the completely monotone real-valued function φ is replaced by a matrix-valued function $\boldsymbol{\varphi} = [\varphi_{ij}]_{i,j=1}^p$, with each component φ_{ij} having specific parameters. In particular, they showed that the Gneiting–Matérn space–time model $\mathbf{C}^{\mathcal{M}} = \left[C_{ij}^{\mathcal{M}}\right]_{i,j=1}^p$ with

$$C_{ij}^{\mathcal{M}}(\boldsymbol{h}, u) = \frac{\sigma_i \sigma_j \rho_{ij}}{(\gamma(u)+1)^{d/2}} \mathcal{M}\left(\boldsymbol{h}; \frac{r_{ij}}{(\gamma(u)+1)^{1/2}}, \nu_{ij}\right), \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \tag{4}$$

is a valid second-order stationary matrix-valued covariance function if, for all $i, j = 1, \ldots, p$,

$$2r_{ij}^2 = r_{ii}^2 + r_{jj}^2; \quad 2\nu_{ij} = \nu_{ii} + \nu_{jj}; \quad \rho_{ij} = \beta_{ij} \frac{\Gamma(\nu_{ij})}{\Gamma(\nu_{ii})^{1/2}\Gamma(\nu_{jj})^{1/2}} \frac{r_{ii}^{\nu_{ii}} r_{jj}^{\nu_{jj}}}{r_{ij}^{2\nu_{ij}}}, \tag{5}$$

where Γ is the gamma function, $\boldsymbol{\beta} = \left[\beta_{ij}\right]_{i,j=1}^p$ is a correlation matrix and $r_{ii}, \nu_{ii} > 0$ for $i = 1, \ldots, p$. Using matrix notations, the model in (4) is valid if the matrices $\boldsymbol{\nu} = [\nu_{ij}]_{i,j=1}^p$ and $\boldsymbol{r}^2 = [r_{ij}^2]_{i,j=1}^p$ have positive entries and are a-separable, and if $\boldsymbol{\rho}\boldsymbol{r}^{2\boldsymbol{\nu}}/\Gamma(\boldsymbol{\nu}) \succeq 0$, with $\boldsymbol{\rho} = [\rho_{ij}]_{i,j=1}^p$. Similarly, the Gneiting–Cauchy space–time model $\mathbf{C}^{\mathcal{C}} = \left[C_{ij}^{\mathcal{C}}\right]_{i,j=1}^p$ with

$$C_{ij}^{\mathcal{C}}(\boldsymbol{h}, u) = \frac{\sigma_i \sigma_j \rho_{ij}}{(\gamma(u)+1)^{d/2}} \mathcal{C}\left(\boldsymbol{h}; \frac{r_{ij}}{(\gamma(u)+1)^{1/2}}, \alpha_{ij}\right), \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \tag{6}$$

is a valid matrix-valued covariance function if, for all $i, j = 1, \ldots, p$,

$$2r_{ij}^{-1} = r_{ii}^{-1} + r_{jj}^{-1}, \quad 2\alpha_{ij} = \alpha_{ii} + \alpha_{jj}, \quad \rho_{ij} = \beta_{ij} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ii})^{1/2}\Gamma(\alpha_{jj})^{1/2}} \frac{r_{ij}^{\alpha_{ij}}}{(r_{ii}^{\alpha_{ii}} r_{jj}^{\alpha_{jj}})^{1/2}}, \tag{7}$$

with $r_{ii}, \alpha_{ii} > 0$ for $i = 1, \ldots, p$ and $\boldsymbol{\beta} = [\beta_{ij}]_{i,j=1}^p$ a positive semidefinite matrix. The sufficient conditions in (7) are thus that $\boldsymbol{\alpha}$ and $\boldsymbol{r}^{-1}$ are a-separable and $\boldsymbol{\rho}\boldsymbol{r}^{-\boldsymbol{\alpha}}/\Gamma(\boldsymbol{\alpha}) \succeq 0$. Proofs are given in Bourotte et al. (2016). Setting $\boldsymbol{h} = \mathbf{0}$ in both cases, the entries of the associated temporal matrix-valued covariance functions are

$$C_{ij}^{\mathcal{M}}(\mathbf{0}, u) = C_{ij}^{\mathcal{C}}(\mathbf{0}, u) = \frac{\sigma_i \sigma_j \rho_{ij}}{(\gamma(u)+1)^{d/2}}, \quad u \in \mathbb{R}.$$

As discussed in Section 2.2, these are proportional-in-time models, since they are the product of a covariance matrix with a single time correlation function $\rho_T(u) = (\gamma(u)+1)^{-d/2}$. In order to build multivariate models that are neither proportional in space nor in time, an essential building block is to define admissible and relevant temporal matrix-valued covariances of the form $[(\gamma_{ij}(u)+1)^{-d/2}]_{i,j=1}^p$. To address this point, matrix-valued pseudo-variograms are now introduced.

2.5. Matrix-valued pseudo-variograms

Consider a p-variate temporal random field $\boldsymbol{Z}(t) = [Z_i(t)]_{i=1}^p$, $t \in \mathbb{R}$. Under the assumption that the direct and cross-increments are jointly stationary, the matrix-valued pseudo-variogram $\boldsymbol{\eta} = [\eta_{ij}]_{i,j=1}^p$ is defined as in Papritz et al. (1993)

$$\eta_{ij}(u) = \frac{1}{2}\mathrm{Var}\big[Z_i(t+u) - Z_j(t)\big], \tag{8}$$

for all $t, u \in \mathbb{R}$ and any pair $i, j = 1, \ldots, p$. Notice that the pseudo-variogram has nonnegative entries and is not necessarily an even function. For any $i = 1, \ldots, p$, η_{ii} (direct variogram of Z_i) is a conditionally negative semidefinite function (Chilès and Delfiner, 2012). Necessary and sufficient conditions for a matrix-valued function to be a pseudo-variogram have been provided in Dörr and Schlather (2021).

3. Parsimonious Gneiting–Matérn and Gneiting–Cauchy multivariate models

In this section, we construct matrix-valued space–time covariance functions that are neither proportional in space nor in time. The construction is based on the substitution approach proposed in Allard et al. (2020) for simulating univariate space–time Gaussian random fields with Gneiting-type covariance functions. It also uses the integral representation of completely monotone functions. Compared to the model in (4), the univariate variogram is generalized into a matrix-valued pseudo-variogram, as defined in (8). This construction is less general than the class that will be presented in Section 4, but it is more direct and parsimonious.

Theorem 1. *Let $\boldsymbol{\sigma} = [\sigma_{ij}]_{i,j=1}^p$ be a symmetric positive semidefinite matrix, $u \mapsto \boldsymbol{\eta}(u) = [\eta_{ij}(u)]_{i,j=1}^p$ be a matrix-valued pseudo-variogram on $\mathbb{R}$, and $t \mapsto \boldsymbol{\varphi}(t) = [\varphi_{ij}(t)]_{i,j=1}^p$ be a matrix of completely monotone functions on $[0,\infty)$ such that $\varphi_{ij}(t) = \int_0^\infty e^{-rt}(f_{ii}(r)f_{jj}(r))^{1/2}dr$, where f_{ii} and f_{jj} are probability density functions on $(0,\infty)$. Then, the matrix-valued function $\mathbf{C} : (\boldsymbol{h}, u) \mapsto [C_{ij}(\boldsymbol{h}, u)]_{i,j=1}^p$ with*

$$C_{ij}(\boldsymbol{h}, u) = \frac{\sigma_{ij}}{(\eta_{ij}(u)+1)^{d/2}}\varphi_{ij}\left(\frac{|\boldsymbol{h}|^2}{\eta_{ij}(u)+1}\right), \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \tag{9}$$

is positive semidefinite in $\mathbb{R}^d \times \mathbb{R}$.

The following two examples generalize (4) and (6) with a matrix-valued pseudo-variogram $\boldsymbol{\eta}$ instead of a univariate variogram γ.

Example 1 (*Gneiting–Matérn Model*)**.** Let $f_{ii} : r \mapsto \Gamma(\nu_{ii})^{-1}(r_{ii}/2)^{2\nu_{ii}} r^{-\nu_{ii}-1}\exp(-r_{ii}^2/(4r))$ be the probability density of an inverse gamma distribution on $(0,\infty)$ with shape parameter $\nu_{ii} > 0$ and scale parameter $r_{ii}^2/4 > 0$. Based on the fact that an inverse gamma mixture of Gaussian covariances is a Matérn covariance (Emery and Lantuéjoul, 2006), the direct and cross-covariances are found to belong to the Gneiting–Matérn family:

$$C_{ij}(\boldsymbol{h}, u) = \frac{\Gamma(\nu_{ij})}{\sqrt{\Gamma(\nu_{ii})\Gamma(\nu_{jj})}}\frac{r_{ii}^{\nu_{ii}} r_{jj}^{\nu_{jj}}}{r_{ij}^{2\nu_{ij}}}\frac{\sigma_{ij}}{(\eta_{ij}(u)+1)^{d/2}}\mathcal{M}\left(\boldsymbol{h}; \frac{r_{ij}}{(\eta_{ij}(u)+1)^{1/2}}, \nu_{ij}\right), \tag{10}$$

with $2\nu_{ij} = \nu_{ii} + \nu_{jj}$, $2r_{ij}^2 = r_{ii}^2 + r_{jj}^2$ and $\mathcal{M}$ the Matérn covariance defined in (3).

Example 2 (*Gneiting–Cauchy Model*). Let $f_{ii} : r \mapsto \Gamma(\alpha_{ii})^{-1} \exp(-r/r_{ii})\, r_{ii}^{-\alpha_{ii}}\, r^{\alpha_{ii}-1}$ be a gamma probability density on $(0, \infty)$ with shape parameter $\alpha_{ii} > 0$ and scale parameter $r_{ii} > 0$. Based on the fact that a gamma mixture of Gaussian covariances is a Cauchy covariance (Emery and Lantuéjoul, 2006), the direct and cross-covariances are found to belong to the Gneiting–Cauchy family:

$$C_{ij}(\boldsymbol{h}, u) = \frac{\Gamma(\alpha_{ij})}{\sqrt{\Gamma(\alpha_{ii})\,\Gamma(\alpha_{jj})}} \frac{r_{ij}^{\alpha_{ij}}}{r_{ii}^{\alpha_{ii}/2} r_{jj}^{\alpha_{jj}/2}} \frac{\sigma_{ij}}{(\eta_{ij}(u)+1)^{d/2}} \mathcal{C}\left(\frac{\boldsymbol{h}}{(\eta_{ij}(u)+1)^{1/2}}; r_{ij}, \alpha_{ij}\right), \tag{11}$$

with $2\alpha_{ij} = \alpha_{ii} + \alpha_{jj}$, $2r_{ij}^{-1} = r_{ii}^{-1} + r_{jj}^{-1}$ and $\mathcal{C}$ the Cauchy covariance defined above.

The advantage of this construction is that it is easy to understand and that it is straightforward to simulate realizations of the multivariate random field $\boldsymbol{Z}$, as shown in Appendix B. However, an implicit assumption is that $\varphi_{ij}(t) = \int_0^\infty e^{-tr} f_{ij}(r) dr$, with f_{ij} being p-separable: $f_{ij} = f_{ii}^{1/2} f_{jj}^{1/2}$. This is a parsimonious parameterization, but also a restrictive condition. In Section 4, we propose a more general construction that encompasses the model presented in Example 1.

4. A general fully nonseparable Gneiting–Matérn class

We now present a very general class of matrix-valued space–time covariance functions that are neither proportional in space nor in time. Their purely spatial part belongs to the Matérn family. The key of our construction (see Appendix D.2) is to convert a p-variate spatio-temporal random field at finitely many (say, n_T) time coordinates into a $(p \times n_T)$-variate spatial random field, and to find conditions ensuring that the latter has a multivariate Matérn covariance function for any choice of n_T, based on the results established in Emery et al. (2022). More general classes of space–time covariance functions, with the spatial smoothness parameter depending on the temporal lag, can be elaborated in a similar way, based on findings by Gneiting et al. (2010), Apanasovich et al. (2012) and Emery et al. (2022) (see Theorem 3 in Appendix A).

4.1. Main result

Theorem 2. *Let $\boldsymbol{\nu}$, $\boldsymbol{b}^2$ and $\boldsymbol{a}^2$ be $p \times p$ symmetric conditionally negative semidefinite matrices, all with positive entries, such that $\boldsymbol{a}^2 - \boldsymbol{\nu} \preceq_c 0$. Let $u \mapsto \boldsymbol{\rho}(u)$ be a $p \times p$ matrix-valued covariance function on $\mathbb{R}$ and let $u \mapsto \boldsymbol{\eta}(u)$ be a $p \times p$ matrix-valued pseudo-variogram on $\mathbb{R}$. Then, the matrix-valued function $\boldsymbol{C} : (\boldsymbol{h}, u) \mapsto [C_{ij}(\boldsymbol{h}, u)]_{i,j=1}^p$ with*

$$C_{ij}(\boldsymbol{h}, u) = \sigma_{ij}(u)\, \mathcal{M}(\boldsymbol{h}; r_{ij}(u), \nu_{ij}), \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R},$$

is positive semidefinite in $\mathbb{R}^d \times \mathbb{R}$ for

$$\begin{aligned} \boldsymbol{r}(u) &= [r_{ij}(u)]_{i,j=1}^p = \sqrt{\frac{\alpha\boldsymbol{\eta}(u) + \boldsymbol{a}^2}{\beta\boldsymbol{\eta}(u) + \boldsymbol{b}^2}} \quad \text{and} \\ \boldsymbol{\sigma}(u) &= [\sigma_{ij}(u)]_{i,j=1}^p = \frac{\boldsymbol{\rho}(u)\Gamma(\boldsymbol{\nu})\exp(\boldsymbol{\nu})}{\left(\alpha\boldsymbol{\eta}(u) + \boldsymbol{a}^2\right)^{\boldsymbol{\nu}} \left(\beta\boldsymbol{\eta}(u) + \boldsymbol{b}^2\right)^{d/2}}, \end{aligned} \tag{12}$$

where α and β are nonnegative.

Corollary 1. *Let $\boldsymbol{\nu}$, $\boldsymbol{b}^2$, $\boldsymbol{a}^2$, $\boldsymbol{\eta}$, α and β be as in Theorem 2. Let $\boldsymbol{\tau}$ be a $p \times p$ symmetric real matrix. The matrix-valued function $\mathbf{C} : (\boldsymbol{h}, u) \mapsto [C_{ij}(\boldsymbol{h}, u)]_{i,j=1}^p$ with*

$$C_{ij}(\boldsymbol{h}, u) = \frac{\tau_{ij}}{(\beta\eta_{ij}(u) + b_{ij}^2)^{d/2} (\alpha\eta_{ij}(u) + a_{ij}^2)^{\nu_{ij}}} \mathcal{M}\left(\boldsymbol{h}; \sqrt{\frac{\alpha\eta_{ij}(u) + a_{ij}^2}{\beta\eta_{ij}(u) + b_{ij}^2}}, \nu_{ij}\right), \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \tag{13}$$

is positive semidefinite in $\mathbb{R}^d \times \mathbb{R}$ if the matrix $\boldsymbol{\tau}e^{-\boldsymbol{\nu}}/\Gamma(\boldsymbol{\nu})$ is positive semidefinite.

The matrix-valued space–time covariance function (13) is of Gneiting–Matérn type. The sufficient conditions on its parameters call for several comments.

1. A sufficient condition for $\boldsymbol{a}^2 - \boldsymbol{\nu} \preceq_c 0$ is that $\boldsymbol{a}^2 \preceq_c 0$ and $\boldsymbol{\nu}$ is a-separable (see the technical lemmas in Appendix D.1). This more restrictive condition is also found in Apanasovich et al. (2012) and Bourotte et al. (2016). Even though a-separability is not necessary, it is convenient because it is easy to satisfy. From a statistical point of view, it reduces the number of parameters from $p(p+1)/2$ to p and therefore leads to a more parsimonious model.
2. The conditional negative semidefiniteness condition on $\boldsymbol{a}^2$ is similar to that in Apanasovich et al. (2012) in a spatial context. It is weaker than the a-separability condition in (5) that applies to the Gneiting–Matérn model in Bourotte et al. (2016).
3. Since $\boldsymbol{\nu} \preceq_c 0$, the matrix $\exp(-\boldsymbol{\nu})$ is positive semidefinite (see Appendix D.1). Hence, based on the Schur product theorem, a sufficient condition for $\boldsymbol{\tau} e^{-\boldsymbol{\nu}}/\Gamma(\boldsymbol{\nu})$ to be positive semidefinite is that $\boldsymbol{\tau}/\Gamma(\boldsymbol{\nu})$ is positive semidefinite.
4. The positive semidefiniteness condition on $\boldsymbol{\tau}$ only depends on the smoothness parameters $\boldsymbol{\nu}$. In contrast, the positive semidefiniteness condition (5) in Bourotte et al. (2016) – also necessary in the parsimonious approach presented in Section 3 – depends on both $\boldsymbol{\nu}$ and $\boldsymbol{r}(u)$, which is a stronger condition. Theorem 2 therefore offers more flexibility for modeling multivariate space–time data.

The Gneiting–Matérn class (12) is quite versatile. The space–time properties depend very much upon the values of α and β. Obviously, if $\alpha = \beta = 0$, the model is purely spatial. In this context, by setting $\boldsymbol{b}$ to be the all-ones matrix, the condition on $\boldsymbol{\sigma}$ reduces to the condition of Theorem 3B in Emery et al. (2022).

When $\alpha = 0$, accounting for the fact that $\exp(-\boldsymbol{\nu})$ and all p-separable matrices with nonnegative entries are positive semidefinite (see technical lemmas in Appendix D.1), the models in (12) and (13) generalize the Gneiting–Matérn model (4). If, furthermore, the direct variograms η_{ii} are increasing functions for $i = 1, \ldots, p$, the scale parameters $r_{ii}(u) = a_{ii}(\beta\eta_{ii}(u) + b_{ii}^2)^{-1/2}$ decrease as u increases. The space–time direct correlations are thus higher than what would happen for a separable space–time covariance function, i.e., they are positively nonseparable. When $\beta = 0$ and under the same monotonicity assumption for the direct variograms, the scale parameters $r_{ii}(u) = \sqrt{\alpha\eta_{ii}(u) + a_{ii}^2/b_{ii}}$ increase with u. This a very original feature, since for all Gneiting-type spatio-temporal models of our knowledge, the scale parameter decreases with u. Applying the same reasoning as above, the resulting space–time direct covariances are thus negatively nonseparable. When α and β are both positive, the model is positively or negatively space–time nonseparable, depending on the relative values of the other parameters involved in $\boldsymbol{r}(u)$. However, in the common case where $\eta_{ij}(u) \to \infty$ as $u \to \infty$, $\boldsymbol{r}(u)$ asymptotically tends to a matrix with all elements equal to $\sqrt{\alpha/\beta}$. The space–time covariance is therefore asymptotically separable as $|u| \to \infty$.

This model covers thus a large variety of nonseparability situations. When analyzing multivariate space–time data, an exploratory analysis can reveal the type of nonseparability that best describes the dataset, offering the possibility to set $\alpha = 0$ (resp. $\beta = 0$) if uniformly positive (resp. negative) nonseparability is established. This point will be illustrated in Section 5.

4.2. Models for unbounded pseudo-variograms

The temporal structure is parameterized by a matrix-valued pseudo-variogram $\boldsymbol{\eta} : u \mapsto [\eta_{ij}(u)]_{i,j=1}^p$. As discussed in Section 2.3, it is critical that this pseudo-variogram is unbounded for the direct and cross-covariances $C_{ij}(\boldsymbol{h}, u)$ tends to 0 as $|u| \to \infty$. On the one hand, Chen and Genton (2019) showed that the well-known linear model of coregionalization cannot be used to construct matrix-valued pseudo-variograms with different diagonal entries, which is required for a fully nonseparable multivariate model. On the other hand, in Papritz et al. (1993) it is shown that, for a matrix-valued pseudo-variogram $\boldsymbol{\eta}$, the following large distance behavior holds: $\lim_{|u|\to\infty} \eta_{ij}(u)/\eta_{ii}(u) = 1$, for any pair $i, j = 1, \ldots, p$. In other words, all entries of a matrix-valued pseudo-variogram must have the same behavior for a very large lag separation u.

Building valid unbounded matrix-valued pseudo-variograms with different diagonal entries (direct variograms) is still an open question. As a simple model for this, we propose the following construction, which is a straightforward generalization of the bivariate construction of Oesting et al. (2017). Let us define $W_i(u) = Y_0(u) + Y_i(u),\ u \in \mathbb{R}$, where $Y_1, \ldots, Y_p$ are second-order stationary random fields with direct and cross-covariance functions $\boldsymbol{R} : u \mapsto [R_{ij}(u)]_{i,j=1}^p$ and Y_0 is an independent intrinsic random field with unbounded variogram γ_0. Then, the pseudo-variogram of $\boldsymbol{W} = [W_i]_{i=1}^p$ is such that $\eta_{ij}(u) = \gamma_0(u) + (R_{ii}(0) + R_{jj}(0))/2 - R_{ij}(u),\ u \in \mathbb{R}$. In matrix notation, a valid class of unbounded matrix-valued pseudo-variogram is thus

$$\boldsymbol{\eta}(u) = \gamma_0(u)\mathbf{1} + \boldsymbol{R}^0 - \boldsymbol{R}(u), \quad u \in \mathbb{R}, \tag{14}$$

where $\boldsymbol{R}$ is any matrix-valued covariance function and $\boldsymbol{R}^0$ is the a-separable matrix with entries $R^0_{ij} = (R_{ii}(0) + R_{jj}(0))/2$.

A more general model consists in defining $W_i(u) = Y_0(u + \delta_i) + Y_i(u)$, $u \in \mathbb{R}$ and $\delta_i \in \mathbb{R}$, which leads to the pseudo-variogram $\boldsymbol{\eta}(u) = [\gamma_0(u + \delta_i - \delta_j)]_{i,j=1}^p + \boldsymbol{R}^0 - \boldsymbol{R}(u),\ u \in \mathbb{R}$.

4.3. Illustration

We illustrate the flexibility of our proposed model in the univariate and bivariate case. We focus specifically on the nonseparability feature. In the univariate setting ($p = 1$) the space–time covariance function (13) is a generalization of the usual Gneiting class with

$$C(\boldsymbol{h}, u) = \frac{1}{(\beta\eta(u) + b^2)(\alpha\eta(u) + a^2)^\nu} \mathcal{M}\left(\boldsymbol{h}; \sqrt{\frac{\alpha\eta(u) + a^2}{\beta\eta(u) + b^2}}, \nu\right), \tag{15}$$

where η is a variogram. While the Gneiting class is uniformly positively nonseparable (Rodrigues and Diggle, 2010), the model in (15) allows for different nonseparability situations, depending on the values of parameters α, β, a and b, as discussed above. Fig. 1 shows the contour plots for the space–time correlation function (15) for different values of these parameters.

In the bivariate case, we will consider the general model in Corollary 1 with $\nu_{11} = 1.5$, $\nu_{12} = \nu_{21} = 1$ and $\nu_{22} = 0.5$; $\tau_{11} = \tau_{22} = 1$ and $\tau_{12} = \tau_{21} = 0.75$. For these values, the condition $\boldsymbol{\tau} e^{-\boldsymbol{\nu}}/\Gamma(\boldsymbol{\nu}) \succeq 0$ is verified. We will consider the same nonseparability scenarios as in Fig. 1, i.e., (α, β) being successively equal to $(1, 1)$ (separable), $(0, 1)$ and $(1, 0)$ (resp. positively and negatively nonseparable) and $(0.3, 0.7)$. We further set $\gamma_0(u) = 0.5|u|$, $R_{ij}(u) = A_i A_j (1 + 0.5|u|)^{-1}$ with $A_1 = 0.8$ and $A_2 = 0.5$, and $a_{ij} = b_{ij} = 1 - R^0_{ij}$. Simulations (see Supplementary Material) are performed on a 200×200 grid covering the square $[0, 2] \times [0, 2]$ ($d = 2$) with 20 regular times steps in the interval $[0, 0.2]$. We used the spectral approach given in Section Appendix C (Algorithm 3) with $M = 20{,}000$ cosine waves whose frequencies are drawn from the spectral density of a Matérn covariance function with $\nu = 0.5$ and $r = 2$.

5. Application to a weather dataset

5.1. Introduction

We now illustrate the use of the fully nonseparable multivariate Gneiting–Matérn model proposed in (13) on a weather dataset consisting of three daily variables (solar radiation R, temperature T, and humidity H) recorded at $n_S = 13$ stations in Western France from 2003 to 2012. In order to restrict ourselves to data being stationary in time, we only analyze data recorded in January. In order to filter out any seasonal and regional effect, the daily values are first centered in time and space according to

$$\tilde{y}_i(\boldsymbol{s}, t) = y_i(\boldsymbol{s}, t) - \overline{y_i}(\cdot, t) - \overline{y_i}(\boldsymbol{s}, \cdot) + \overline{y_i}(\cdot, \cdot), \quad i = 1, 2, 3,$$

where $\overline{y_i}(\cdot, t)$, $\overline{y_i}(\boldsymbol{s}, \cdot)$ and $\overline{y_i}(\cdot, \cdot)$ are the spatial average at day t, temporal average at location $\boldsymbol{s}$ and overall average of the ith variable, respectively. These centered values are then standardized every

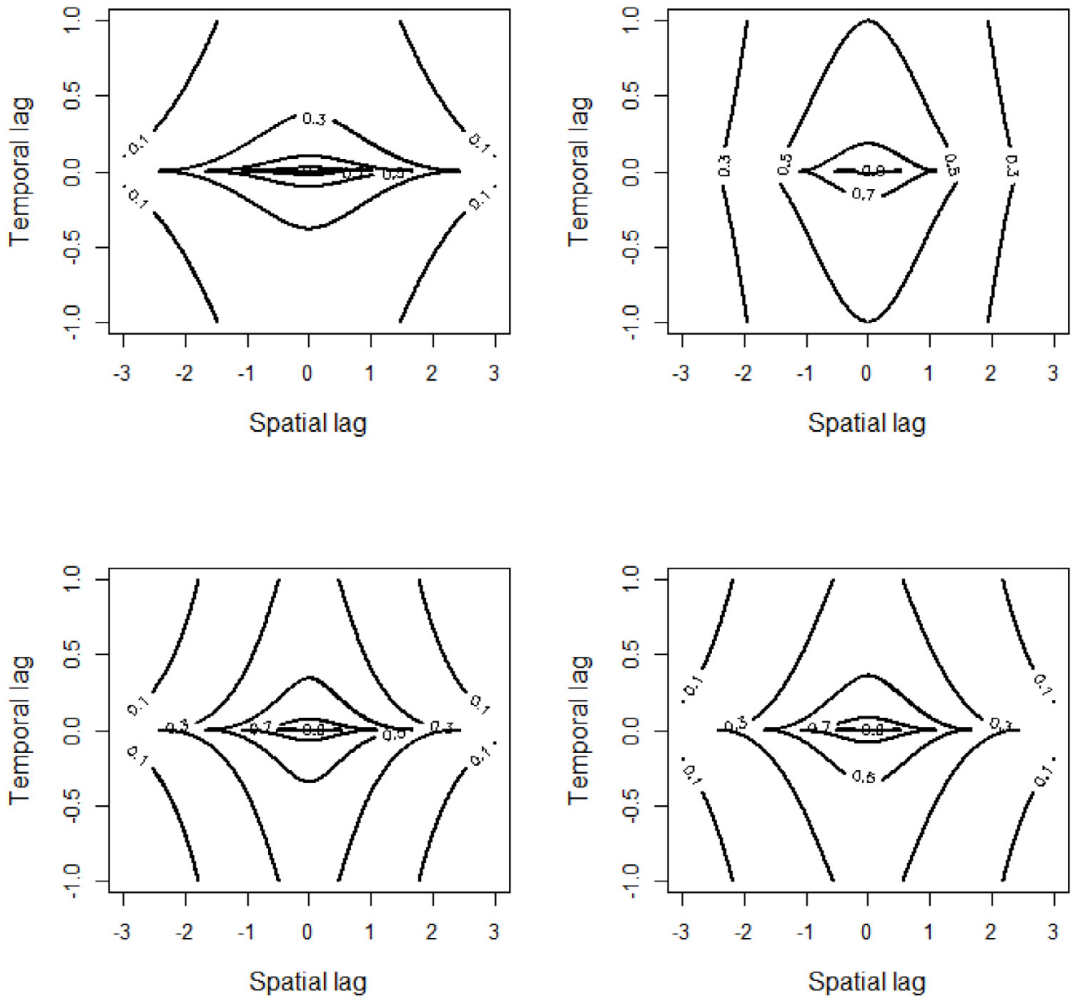


Fig. 1. Contour plots for the space–time covariance (15) with $\eta(u) = \sqrt{|u|}$, $\nu = 1.5$, $a^2 = b^2 = 1$: separability ($(\alpha, \beta) = (1, 1)$, upper left), positive nonseparability ($(\alpha, \beta) = (0, 1)$, upper right), negative nonseparability ($(\alpha, \beta) = (1, 0)$, lower left) and general nonseparability ($(\alpha, \beta) = (0.7, 0.3)$, lower right).

day t according to $z_i(\boldsymbol{s}, t) = \tilde{y}_i(\boldsymbol{s}, t)/S_i(t)$ with $S_i^2(t)$ the experimental variance of the ith variable at day t.

We thus have 10 repetitions of 13×31 of trivariate standardized vectors, which will be supposed independent. For a detailed exposition of this dataset, we refer to a previous analysis (Bourotte et al., 2016) which has shown that a Proportional-In-Time (PIT) multivariate Gneiting–Matérn model improved the fitting and the prediction scores as compared to a model with a separable spatio-temporal covariance function. In this analysis, we shall compare a specific fully nonseparable multivariate model (FULL), as proposed in (13), to the PIT model.

For each of the three variables, the separability function $S_{ii}(\boldsymbol{h}, u) = C_{ii}(\boldsymbol{h}, u)C_{ii}(\boldsymbol{0}, 0) - C_{S,ii}(\boldsymbol{h})C_{T,ii}(u)$ was computed. Empirical covariances were derived from the space–time empirical variograms computed with the function `EVariogram` from the package `CompRandFld` (Padoan and Bevilacqua, 2015), using spatial bins of 50 m. Since the variograms have been computed on residuals, the separability function must be in the interval $[-1, 1]$, with $S(\boldsymbol{h}, u) = 0$ indicating separability of $C(\boldsymbol{h}, u)$. Fig. 2 shows the separability function of all variables at the first three time steps. Without having to compute a formal test for separability as for example in Mitchell et al. (2006), it is clear from these plots that the residuals are uniformly positively space–time nonseparable for

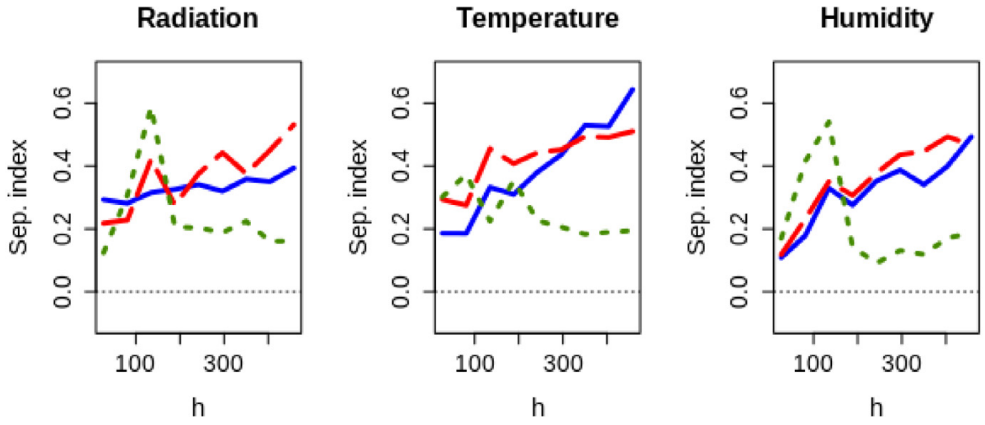


Fig. 2. Separability function $S(\boldsymbol{h}, u)$ computed at $u = 1$ (blue solid lines), $u = 2$ (red dashed lines) and $u = 3$ (green dotted lines).

$|\boldsymbol{h}| \leq 500$ m and $u \leq 3$, which is actually often the case for climate variables. The same analysis was performed using other values for the spatial bins and similar results (not shown) were consistently obtained.

5.2. Model specification

Given the positive nonseparability of the residuals, and in light of the discussion in Section 4, the parameter α in (13) is set to $\alpha = 0$. The temporal structure is described by a matrix-valued unbounded pseudo-variogram as given in (14). Since $d = 2$ here, the entries of the matrix-valued covariance function read:

$$C_{ij}(\boldsymbol{h}, u) = \frac{\tau_{ij} a_{ij}^{-2\nu_{ij}}}{\beta \eta_{ij}(u) + b_{ij}^2} \mathcal{M}\left(\boldsymbol{h}; \frac{a_{ij}}{\sqrt{\beta \eta_{ij}(u) + b_{ij}^2}}, \nu_{ij}\right), \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \tag{16}$$

with the condition that $\boldsymbol{\tau} e^{-\boldsymbol{\nu}} / \Gamma(\boldsymbol{\nu}) \succeq 0$. Setting $\boldsymbol{h} = \boldsymbol{0}$ and $u = 0$ in (16), one gets the generic entry $C_{ij}(\boldsymbol{0}, 0) = \tau_{ij} a_{ij}^{-2\nu_{ij}} / (\beta [R_{ij}^0 - R_{ij}(0)] + b_{ij}^2)$ of the collocated covariance matrix. For identifiability reasons, we further set $\beta = 1$ and $b_{ij}^2 = 1 - R_{ij}^0$ with $\max_i \{R_{ii}(0)\} < 1$. The matrix $\boldsymbol{b^2}$ is thus conditionally negative definite, as required. As a consequence, we get the simpler expression $C_{ij}(\boldsymbol{0}, 0) = \tau_{ij} a_{ij}^{-2\nu_{ij}} / [1 - R_{ij}(0)]$.

The pseudo-variogram $\boldsymbol{\eta}$ needs to be specified. For the unbounded part common to all components we choose $\gamma_0(u) = (1 + |cu|^{2a})^b - 1, \quad u \in \mathbb{R}$ with $c > 0, 0 < a \leq 1$ and $0 \leq b \leq 1$, for ease of comparison with Bourotte et al. (2016), also advocated in Gneiting (2002). This variogram is reminiscent of the generalized Cauchy covariance model that will also be used for $\boldsymbol{R}$: $R_{ij}(u) = A_i A_j (1 + |ru|^{2\lambda})^{-b}, \ u \in \mathbb{R}$, with $r > 0, 0 < \lambda < 1$ and $0 \leq A_i < 1$ for $i = 1, \ldots, p$, and where for the sake of parsimony the parameter b is set identical to the parameter b in γ_0. Notice that the matrix $[A_i A_j]_{i,j=1}^p$ is p-separable and has nonnegative entries, thus it is positive semidefinite. $\boldsymbol{R}$ is therefore a valid matrix-valued covariance function. The covariance model that we obtain is

$$C_{ij}(\boldsymbol{h}, u) = \frac{\tau_{ij} a_{ij}^{-2\nu_{ij}}}{(1 + |cu|^{2a})^b - A_i A_j (1 + |ru|^{2\lambda})^{-b}} \times \mathcal{M}\left(\boldsymbol{h}; \frac{a_{ij}}{\left[(1 + |cu|^{2a})^b - A_i A_j (1 + |ru|^{2\lambda})^{-b}\right]^{1/2}}, \nu_{ij}\right), \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}.$$

In the above parameterization, the space–time nonseparability parameter b acts both inside and outside the Matérn function $\mathcal{M}$, implying a constant perfect correlation in time if $b = 0$. Following Gneiting (2002) and Bourotte et al. (2016), a reparameterization is thus useful. Multiplying the above equation by the temporal covariance function $\tilde{\rho}_{ij}(u) = \left[(1 + c|u|^{2a})^{\delta} - A_i A_j (1 + r|u|^{2\lambda})^{-\delta}\right]^{-1}$ with $0 \leq \delta \leq 1$ leads to the parametric family that will be used in the rest of this section. In order to get $C_{ii}(\mathbf{0}, 0) = \sigma_i^2$, we set $\tau_{ii} = \sigma_i^2 a_{ii}^{2\nu_{ii}}(1 - A_i^2)^2$ and, for $i \neq j$, $\tau_{ij} = \rho_{ij}\sqrt{\tau_{ii}\tau_{jj}}$, with $[\rho_{ij}]_{i,j=1}^{p}$ being a correlation matrix. With this parameterization, the matrix $\boldsymbol{\tau}$ is positive semidefinite by construction. Finally, the fully nonseparable multivariate Gneiting–Matérn covariance model reads

$$C_{ij}(\boldsymbol{h}, u) = \frac{1}{(1 + |cu|^{2a})^{\delta} - A_i A_j (1 + |ru|^{2\lambda})^{-\delta}} \frac{\rho_{ij} a_{ij}^{-2\nu_{ij}} \sqrt{\tau_{ii}\tau_{jj}}}{(1 + |cu|^{2a})^{b} - A_i A_j (1 + |ru|^{2\lambda})^{-b}} \times \mathcal{M}\left(\boldsymbol{h}; \frac{a_{ij}}{\left[(1 + |cu|^{2a})^{b} - A_i A_j (1 + |ru|^{2\lambda})^{-b}\right]^{1/2}}, \nu_{ij}\right), \tag{17}$$

with $(\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$ and $i, j = 1, \ldots, p$. The model in (17) provides interpretable parameters for a fully nonseparable Gneiting–Matérn model:

- The collocated covariance matrix $[C_{ij}(\mathbf{0}, 0)]_{i,j=1}^{p}$ has elements

$$C_{ij}(\mathbf{0}, 0) = \rho_{ij}\sigma_i\sigma_j \frac{a_{ii}^{\nu_{ii}} a_{jj}^{\nu_{jj}}}{a_{ij}^{2\nu_{ij}}} \frac{(1 - A_i^2)(1 - A_j^2)}{(1 - A_i A_j)^2}, \tag{18}$$

  which simplify to $C_{ii}(\mathbf{0}, 0) = \sigma_i^2$ if $i = j$. This is the elementwise product of a covariance matrix $[\rho_{ij}\sigma_i\sigma_j]_{i,j=1}^{p}$ with a weight matrix whose elements are ratios combining the other parameters. Notice that the quantities $\sigma_i a_{ii}^{\nu_{ii}}$ appearing in (18) are the square roots of the microergodic quantities that can be consistently estimated by Maximum Likelihood under a Gaussian assumption in an infill asymptotics setting (Zhang, 2004).
- The vector $\boldsymbol{A} = [A_i]_{i=1}^{p}$ describes the multivariate aspect of the temporal covariance function. When $A_i = 0$ for $i = 1, \ldots, p$, the above model reduces to the model in Bourotte et al. (2016, Equation 9), up to a slight reparameterization.
- The temporal parameters are, on the one hand, (c, a) that characterize the common part of the temporal covariance function and, on the other hand, $(r, \lambda, \boldsymbol{A})$ that characterize the multivariate part of it. The parameters (c, r) are scale parameters, while (a, λ) are shape parameters relating to the regularity at the origin of $C_{T,ij}(u)$. Note that the parameters (b, δ) act on both parts of the temporal covariance function.
- The parameter b, which appears in both the spatial and temporal terms, is the separability parameter. It can vary from $b = 0$ (separability) to $b = 1$ (full spatio-temporal nonseparability).
- The spatial parameters are the scale matrix $\boldsymbol{a}$, with $\boldsymbol{a}^2 \preceq_c 0$, and the smoothness matrix $\boldsymbol{\nu}$, which must verify $\boldsymbol{a}^2 - \boldsymbol{\nu} \preceq_c 0$. For simplicity, and for a fair comparison to Bourotte et al. (2016), $\boldsymbol{a}^2$ and $\boldsymbol{\nu}$ will be chosen as a-separable. The two above conditions on $\boldsymbol{a}$ and $\boldsymbol{a}^2 - \boldsymbol{\nu}$ are thus always verified.

5.3. Estimation of the parameters

Recall that the data analyzed here are standardized residuals after centering for spatial and temporal trends. We thus set $\sigma_i = 1$, for $i = 1, \ldots, p$. There is a total of $p(p + 2) + 6$ parameters to estimate, where p is the number of variables. Here $p = 3$, leading to a total of 21 parameters. Compared to the model in Bourotte et al. (2016), our more flexible model contains only 6 additional parameters describing the multivariate nature of the temporal covariance. Let us denote $\boldsymbol{\theta}$ the vector of all parameters, and let Θ be the space of parameters for which the model (17) is valid. Optimizing a likelihood function in Θ is not straightforward. First, the space Θ has a complex shape owing to the positive semidefiniteness condition on $\boldsymbol{\tau} e^{-\boldsymbol{\nu}}/\Gamma(\boldsymbol{\nu})$. Second, employing blindly an optimization function to maximize the likelihood in the high-dimensional space Θ is likely to fail. As a way to

alleviate this problem, the likelihood was maximized sequentially in subspaces of Θ corresponding to blocks of related parameters, while keeping all other parameters fixed to the values previously attained.

Maximum likelihood estimation requires inverting a 1209×1209 matrix each time a parameter is updated, which is prohibitively expensive considering the very large number of calls entailed by the optimization in a 21-dimensional space. Composite likelihood methods have been found to be efficient in the context of spatio-temporal data, with less time-consuming steps and good asymptotic properties (Bevilacqua et al., 2012). Pairwise Likelihood (PL) is the product of marginal Gaussian likelihoods computed on all pairs with a spatial separation less than d_m and a temporal separation less than t_m, where d_m and t_m are fixed thresholds. On a simulation study mimicking the sampling design of the dataset, it was shown in Bourotte et al. (2016) that, for the smoothness and scale parameters, Pairwise Likelihood (PL) provides estimates with only a slight loss in efficiency as compared to a Full Likelihood (FL) approach, with a significant gain in terms of computation. Moreover, for prediction, the difference between PL and FL was found very small on all tested prediction scores. We have therefore decided to estimate the parameters of our model using PL in this same way. We have set $d_m = 250$ km and $t_m = 2$ days because these values have been shown to minimize the trace of the estimated covariance matrix, thus providing optimal estimation for all parameters (Bourotte et al., 2016). The following procedure for maximizing the likelihood builds on the one presented in Bourotte et al. (2016), which was proven to be efficient.

1. The $p \times p$ marginal empirical covariance matrix with elements $\hat{c}_{ij},\ i, j = 1, \ldots, p$ is computed.
2. The separability parameter b, known to be difficult to estimate, is successively fixed to $0, 0.1, \ldots, 1$.
3. For every fixed value of b:

   (a) Initial temporal parameters $(c, a, r, \lambda, \delta, \boldsymbol{A})$ are estimated by maximizing the PL for the temporal covariance model $\boldsymbol{C}_T$. Using (17) and (18), one gets

   $$C_{T,ij}(u) = \frac{1}{(1+|cu|^{2a})^{\delta} - A_i A_j (1+|ru|^{2\lambda})^{-\delta}} \frac{\hat{c}_{ij}(1-A_iA_j)^2}{(1+|cu|^{2a})^{b} - A_i A_j (1+|ru|^{2\lambda})^{-b}}.$$

   (b) Using the estimates $\hat{\boldsymbol{A}}$ from the previous step, initial spatial parameters $(\boldsymbol{a}, \boldsymbol{\nu})$ are estimated by maximizing the PL for the spatial matrix-valued covariance $\boldsymbol{C}_S$ with entries

   $$C_{S,ij}(\boldsymbol{h}) = \hat{c}_{ij}\, \mathcal{M}\left(\boldsymbol{h};\ \frac{a_{ij}}{\left(1-\hat{A}_i\hat{A}_j\right)^{1/2}},\ \nu_{ij}\right).$$

   (c) The temporal parameters are updated, considering all other parameters fixed, by maximizing the PL of the multivariate spatio-temporal model (17).

   (d) The spatial parameters are updated, considering all other parameters fixed by maximizing the PL of the multivariate spatio-temporal model (17).

   (e) Steps (c)–(e) are iterated until a stopping criterion is reached.

In steps 3.(a) to 3.(d), the optimization is performed by calling the R function `nlminb`. At each iteration of the optimization, the estimate $\hat{\boldsymbol{\rho}}$ is computed using (18) given $\hat{\boldsymbol{a}}, \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{A}}$, and the definite positiveness of $\hat{\boldsymbol{\nu}}$ and $\hat{\boldsymbol{\tau}} e^{-\hat{\boldsymbol{\nu}}}/\Gamma(\hat{\boldsymbol{\nu}})$ are enforced. Iterations are stopped when PL is increased by less than 1 unit after one iteration from (c) to (e). Finally, $\hat{b}$ is the value of b among $\{0, 0.1, \ldots, 0.9, 1\}$ corresponding to the highest maximized PL. To perform the maximization in the subspaces of Θ, we used the package `nlminb` implemented in R with lower and upper bounds for the parameters when mathematically necessary. In addition, for the sake of numerical stability, the smoothness parameter ν was upper bounded at 6.0 – which was never reached.

All 13 stations were used to estimate the parameters of both the FULL and PIT models. Fig. 3(left) shows the maximum log-PL as a function of the separability parameter b for the fully nonseparable Gneiting–Matérn model, referred to as FULL. The maximum is reached for $\hat{b} = 0.1$ with

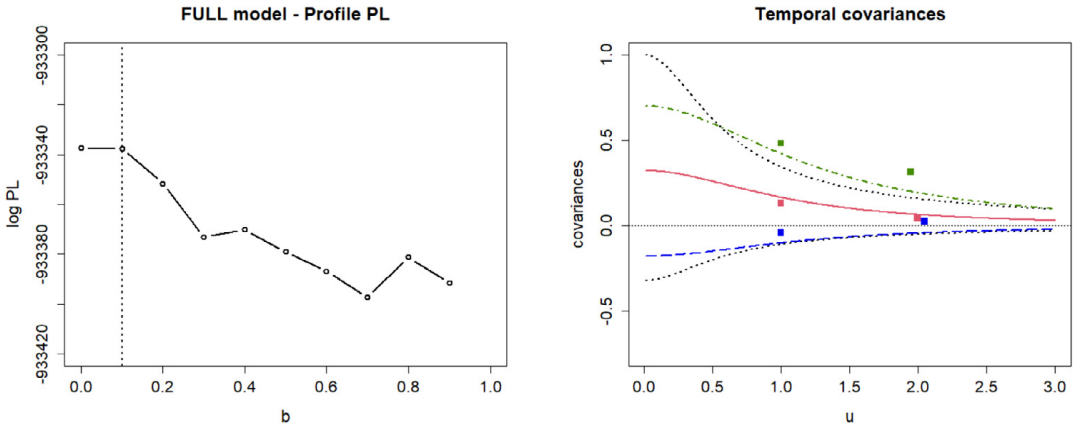


Fig. 3. Left: PL as a function of the space–time separability parameter b. Right: (R, H) temporal correlation functions. PIT model with unique direct correlation (positive) and cross-correlation (negative) in dotted lines. FULL model with direct correlation (positive) for R and H in solid and dotdashed lines, respectively, and cross-correlation (negative) in longdashed line. Squares: empirical correlations.

Table 1
PL parameter estimates for the weather dataset with $d_m = 250$ km, $t_m = 2$ days. FULL: fully nonseparable model as in (17) with $b = 0.1$. PIT: simplified Proportional-In-Time model with $b = 0$. The indices 1, 2 and 3 refer to R, T and H, respectively.

	Temporal parameters							
	c	a	r	λ	δ	A_1	A_2	A_3
FULL	0.528	0.993	0.341	0.0	1.000	0.871	0.608	0.646
PIT	4.31	0.997	N/A	N/A	0.638	0	0	0
	Spatial parameters							
	a_1	a_2	a_3	ν_1	ν_2	ν_3		
FULL	$15.8\ 10^{-3}$	$6.4\ 10^{-3}$	$12.4\ 10^{-3}$	3.76	0.69	0.51		
PIT	$18.9\ 10^{-3}$	$8.2\ 10^{-3}$	$17.0\ 10^{-3}$	1.61	0.71	0.53		
	Correlation parameters							PL
	ρ_{12}	ρ_{13}	ρ_{23}					
FULL	−0.124	−0.330	−0.099					−933 337.6
PIT	−0.151	−0.324	−0.103					−933 961.3

$\log \mathrm{PL}_{\mathrm{FULL}}(\hat{\boldsymbol{\theta}}_{b=0.1}) = -933\,337.6$. For comparison purpose, PL was also maximized for the PIT model, i.e., when $\boldsymbol{A} = \mathbf{0}$. For this simpler model, $\log \mathrm{PL}_{\mathrm{PIT}}(\hat{\boldsymbol{\theta}}_b)$ is monotonically decreasing from $-933\,961.3$ to $-934\,072.6$ as b varies from 0 to 1. FL was also computed for these two models (remember that FL is easy to compute but very long to optimize): $\log \mathrm{FL}_{\mathrm{FULL}}(\hat{\boldsymbol{\theta}}_{b=0.1}) = -2\,652.7$ and $\log \mathrm{FL}_{\mathrm{PIT}}(\hat{\boldsymbol{\theta}}_{b=0.1}) = -2\,819.1$, indicating strong evidence in favor of the more complex model, considering that the number of data is moderate with a total of 10 independent repetitions of 1209 correlated data. As a point of comparison, the (BIC) penalization to the full likelihood of the more complex model is equal to $2.5 \ln(12090) \simeq 23.5$.

Looking at the temporal marginals, the models show very different behaviors — as we expected. Fig. 3(right) displays the temporal empirical covariances for variables R and H with their fitted models. Thanks to the flexibility of the FULL model, the temporal covariance functions fit the empirical values for both variables. In contrast, the unique covariance function of the PIT model lies somewhere between the empirical covariances of the two variables, thus being unable to provide a satisfactory fit to any of them. Table 1 reports the estimated parameters maximizing PL for both models. It is interesting to note that the parameter λ is equal to 0 in the FULL model, thus implying a sort of temporal nugget effect, visible on the temporal covariances in Fig. 3(right). Fig. 4 shows the

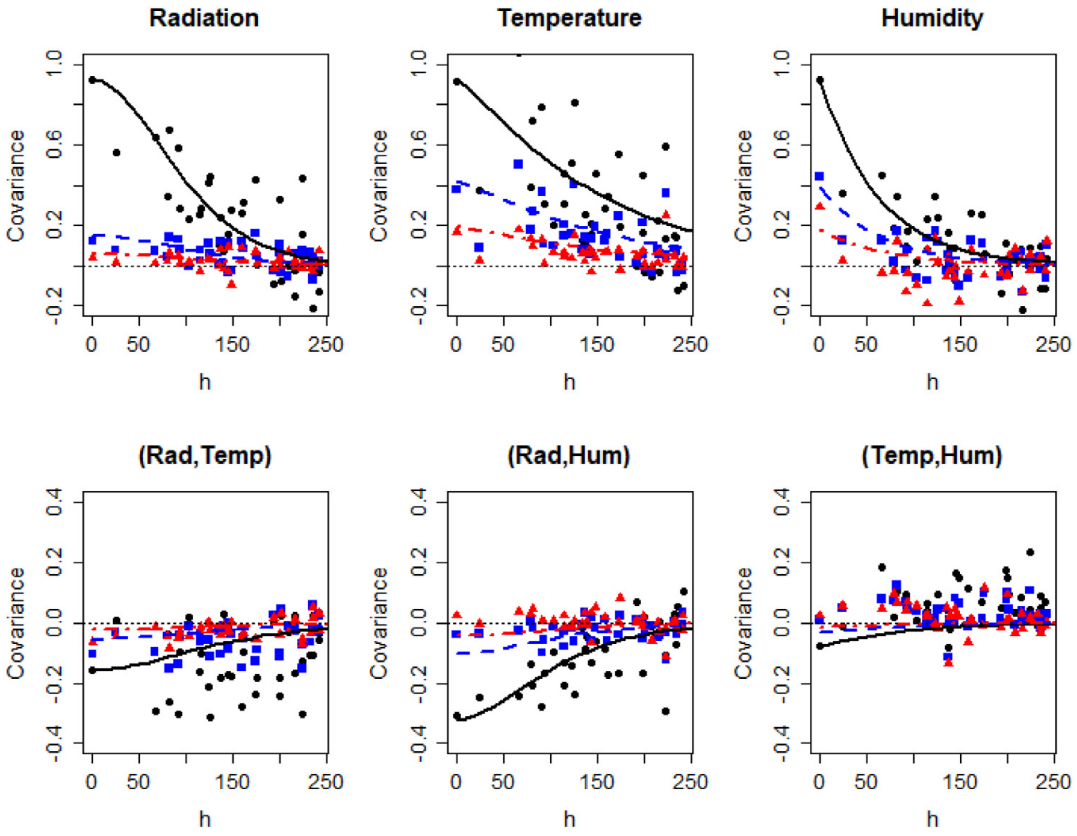


Fig. 4. Empirical and fitted spatial–temporal correlation at $u = 0$ (black solid lines and circles), $u = 1$ (blue dashed lines and squares) and $u = 2$ (red dotdashed lines and triangles).

spatial (at time lag $u = 0$) and spatio-temporal covariances (at time lags $u = 1, 2$) for all variables and pairs of variables. The empirical covariances between all pairs of stations and the FULL model with estimated parameters from Table 1 are shown. Overall, the fit is good. On the top left panel representing the covariance function for the variable R, the very high smoothness of the covariance function near the origin is clearly visible.

5.4. Prediction

We consider two different prediction settings. In the first setting, prediction of the three variables at a validation station at day t is made knowing the trivariate data at the other 12 estimation stations at day t, $t-1$ and $t-2$, each station being used as validation station in turn. This setting is called *Spatial Interpolation* because data from the same day at different locations are used (as well as data from previous days). In the second setting, prediction is made with data from days $t-1$ and $t-2$ only at all 13 stations. In this second setting, called *Temporal Prediction*, purely temporal covariances are thus introduced. The first setting corresponds to interpolation at an ungauged location while the second setting corresponds to the classical one-day ahead forecast. Predictions are computed from January 3rd to January 31st since the two previous days must be used as conditioning data. In the Gaussian framework, the conditional expectation is the best linear unbiased prediction, also called (co)kriging in the geostatistics literature, while the conditional covariance matrix is the covariance matrix of the cokriging errors (Chilès and Delfiner, 2012). The conditional variances are identical for all days since the configuration of the conditioning data remains identical.

The two models (FULL and PIT) are compared by means of four scores: the Mean Square Error (MSE), the Mean Absolute Error (MAE), the Continuous Ranked Probability Score (CRPS) and the

Table 2
MSE, MAE, CRPS and LogS of predicted values. "FULL": fully nonseparable model as in (17) with $b = 0.1$ and ML estimators. "PIT": simplified Proportional-In-Time model with $b = 0$ and ML estimators. "All 0s": trivial prediction with 0. The best scores among FULL or PIT are shown in bold font.

	MSE			MAE			CRPS			LogS		
	R	T	H	R	T	H	R	T	H	R	T	H
	Spatial interpolation											
FULL	0.594	**0.482**	**0.814**	0.574	**0.508**	**0.714**	0.412	**0.371**	**0.505**	0.188	**0.091**	0.382
PIT	**0.585**	0.490	**0.814**	**0.570**	**0.508**	0.716	**0.409**	**0.371**	**0.505**	**0.157**	0.092	**0.381**
	Temporal prediction											
FULL	**0.903**	**0.757**	**0.708**	**0.790**	**0.705**	**0.673**	**0.542**	**0.493**	**0.475**	**0.449**	**0.361**	**0.329**
PIT	0.943	0.763	0.722	0.795	0.717	0.684	0.553	0.497	0.481	0.474	0.367	0.342
	Trivial prediction											
All 0s	0.923	0.923	0.923	0.805	0.801	0.783	–	–	–	–	–	–
	Estimation standard deviations for FULL											
	0.0194	0.0176	0.0173	0.0091	0.0088	0.0087	–	–	–	–	–	–

Logarithmic Score (LogS) (Gneiting and Raftery, 2007). The first two scores, MSE and MAE, compare the conditional expectation to the true value. The other two scores, CRPS and LogS, assess not only the prediction but the prediction variance as well. They both measure the discrepancy between the predictive cumulative distribution function and the true value. More details can be found in Gneiting and Raftery (2007). Table 2 reports our results. As a benchmark, we also report the scores obtained with a trivial prediction where all predicted values are set to the expectation equal to 0. Since no prediction variance is attached in this trivial setting, CRPS and LogS are not available. Recall that lower scores indicate a better adequacy between the model and the data and that the first three scores are bounded from below by 0 whilst LogS is unbounded.

For variables R and T, Spatial Interpolation is more accurate than Temporal Prediction. For variable H, the situation is reverse, in relation with the fact that the associated MSE is close to that of the trivial prediction. Usually, spatial correlations are higher than one-day-ahead temporal correlations, but for H the spatial correlation are much weaker than for R and T (see Fig. 4). In the Spatial Interpolation setting, the two models can hardly be distinguished: the validation scores are often equal, and when they are not, they are extremely close.

In the Temporal Prediction setting however, the FULL model clearly outperforms PIT for all variables and for all scores, in accordance with the fact that the FULL model is able to better fit the data in time thanks to the extra parameters in the temporal matrix-valued covariance function. In this setting, we also computed the standard deviations of the squared and absolute prediction errors for the FULL model (those of the PIT model are very close) and divided them by $\sqrt{11310}$, where $11310 = 13 \times 10 \times 29 \times 3$ is the total number of predictions. From this, we see that the MSE score difference is larger than twice this value for variable R. In addition, the absolute error is smaller with FULL than with PIT for 6151 predictions out of 11310, which represents 54.4% of the predictions. As a matter of comparison, the centered 99% interval of a binomial random variable with probability of success 0.5 and sample size 11310 is $[5518, 5792]$. The improvement brought by the FULL model as compared to the PIT model can thus be considered as significant.

For illustration purpose, Fig. 5 shows the predictions of the three variables from January 3rd to January 31st 2003 at one station (Le Rheu) following the Spatial Interpolation setting. The predicted residuals were back-transformed to the original scales to be compared to the real values. An envelope of ± 2 conditional standard deviations was added to the plot.

6. Concluding remarks

We have proposed new parametric classes of matrix-valued covariance functions for multivariate spatio-temporal random fields, where each component has its own smoothness and scale parameter

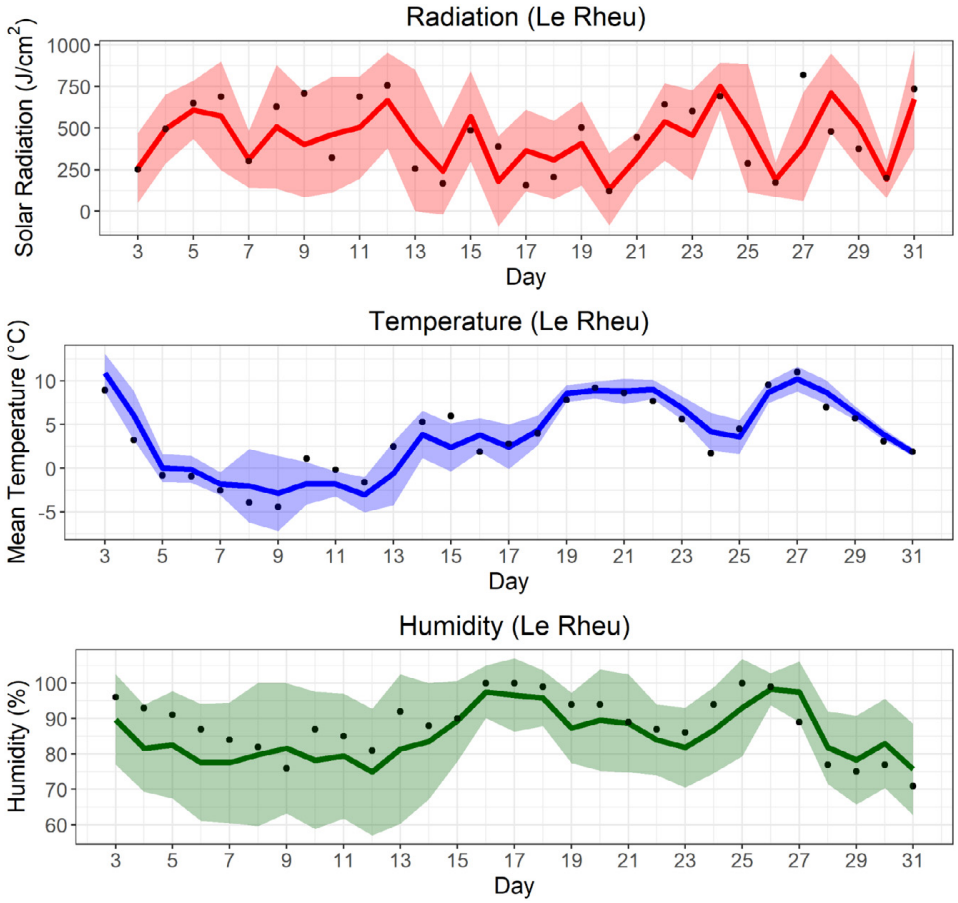


Fig. 5. Predictions of R, T and H at Le Rheu from January 3rd to January 31st 2003 following the *Spatial Interpolation* setting. Predictions (continuous line), real values (dots) and envelope of ± 2 conditional standard deviations (colored ribbon).

in space and its own correlation function in time. Our constructions generalize the Gneiting class of space–time covariance functions and are fully nonseparable, in the sense that they are space–time nonseparable and that they are neither proportional-in-space nor proportional-in-time. This is a major improvement to most of the models proposed up to date (with the exceptions of the models in Dörr and Schlather, 2021 and Porcu et al., 2022), where the marginal temporal correlation function is identical for all the variables, which is overly restrictive. The main ingredient for achieving this has been the use of pseudo-variograms, as multivariate analogous of Bernstein functions as pointed out in Dörr and Schlather (2021) and Qadir and Sun (2022) in their discussion sections.

The main class proposed in Section 4 allows for a general behavior for the spatial scale parameter as a function of the time lag u, depending on the values taken by the parameters. It is well known that the Gneiting class of models is positively nonseparable. A very original feature of our construction is that the models can be negatively nonseparable or asymptotically separable as $|u| \to \infty$. Since these models are continuous over $\mathbb{R}^d \times \mathbb{R}$, their application is not limited to data organized as regular records at measurement stations. More complex designs are in principle possible, such as irregular or incomplete records, heterotopic designs or even mobile measurement devices. We hope to see applications of our models to such datasets in a near future.

From a statistical point of view, the main challenge is to find the right balance between the flexibility of the model, which must be able to account for the complex interactions between space

and time across several variables, and the number of parameters, that should remain interpretable and not too numerous. More complex models should only be preferred if supported by the data. The dataset analyzed in this work is of moderate size with 10 independent repetitions of $13 \times 31 = 403$ space–time trivariate data. Its analysis has shown that our model improves the fitting and the predictive performances over the more parsimonious model in Bourotte et al. (2016), in particular when the time dimension is important for prediction. In Section 5.3, it was shown that the maximum PL difference is around 600. Combined with the improved prediction scores in time, this provides evidence in favor of our new model for this dataset. The analysis of other datasets, in particular with more spatial locations, is necessary for confirming this first result.

Estimating the parameters of a parametric model such as the one in Section 5 is a challenge. Here, we have used Pairwise Likelihood because it was proven efficient on the proportional-in-time model in Bourotte et al. (2016). It is however relatively slow: reaching convergence with our optimization algorithm and the function `nlmbinb` takes typically several hours on a recent laptop computer. There is thus certainly room for improvement for an easier use of such complex models. At the cost of even larger computing times, composite likelihood with larger sets of elements than pairs and full likelihood might be interesting options for a more accurate estimation of the parameters, considering the complex interactions between space, time and the variables. Conducting a thorough comparison of several likelihood approaches was out of the scope of this work and left for future research. In our opinion, more interesting improvements can be expected by using Vecchia's Gaussian process approximation (Katzfuss and Guinness, 2021). We are looking forward to (multivariate) spatio-temporal versions of the `GpGp` package.

In Appendix A we have proposed even more general constructions where both the spatial scale and spatial smoothness parameters can vary with the time lag u. We have also designed (see Appendix C) computationally efficient algorithms for simulating Gaussian random fields with these covariance functions, which are not restricted in the number of target spatial coordinates and allow tens of thousands of time coordinates, hence are applicable to large-scale space–time prediction problems.

All the models presented in this work can easily be generalized to $\mathbb{R}^d \times \mathbb{R}^\ell$, where ℓ is an integer greater than one. Other constructions based on characterizations established in Porcu et al. (2022) are also possible. Extensions of our models to other spaces that are of interest in environmental sciences, atmospheric sciences and geosciences, such as a sphere crossed with an Euclidean space ($\mathbb{S}^d \times \mathbb{R}^\ell$), are left to further research.

Appendix A. General forms of Gneiting–Matérn covariances with temporal-lag dependent parameters

The following theorem provides sufficient validity conditions for a nonseparable matrix-valued space–time covariance function in which the spatial direct and cross-covariances belong to the Matérn family, with collocated correlation coefficients, scale and/or smoothness factors that depend on the temporal lag. The proof follows the line of that of Theorem 2 and is based on the sufficient validity conditions for matrix-valued spatial Matérn models given in Gneiting et al. (2010, Theorem 1), Apanasovich et al. (2012, Theorem 1) and Emery et al. (2022, Theorems 1, 2, 3). Also, the spectral algorithm given in Appendix C can be adapted to simulate space–time random fields having such a covariance.

Theorem 3. *Let d, p be positive integers, r and ν be positive real numbers. Let $\boldsymbol{\mu}_0$, $\boldsymbol{\nu}_0$ and $\boldsymbol{\psi}_0$ be $p \times p$ symmetric conditionally negative semidefinite matrices, the first two ones with positive entries and the last one with nonnegative entries. Also, let $u \mapsto \boldsymbol{\mu}(u) - \boldsymbol{\mu}_0$, $u \mapsto \boldsymbol{\nu}(u) - \boldsymbol{\nu}_0$ and $u \mapsto \boldsymbol{\psi}(u) - \boldsymbol{\psi}_0$ be $p \times p$ matrix-valued pseudo-variograms in $\mathbb{R}$ and $u \mapsto \boldsymbol{\rho}(u)$ be a $p \times p$ matrix-valued covariance in $\mathbb{R}$. Then, the matrix-valued function $\boldsymbol{C} : (\boldsymbol{h}, u) \mapsto \boldsymbol{\sigma}(u)\,\mathcal{M}(\boldsymbol{h}; \boldsymbol{r}(u), \boldsymbol{\nu}(u)) = [\sigma_{ij}(u)\,\mathcal{M}(\boldsymbol{h}; r_{ij}(u), \nu_{ij}(u))]_{i,j=1}^{p}$ with $(\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$ and $\mathcal{M}$ the Matérn covariance defined in (3), is positive semidefinite in $\mathbb{R}^d \times \mathbb{R}$ if one of the following sets of sufficient conditions holds:*

(A) $\boldsymbol{r}(u) = \sqrt{\frac{\boldsymbol{\psi}(u)+\boldsymbol{\nu}(u)}{\boldsymbol{\mu}(u)}}$ and $\boldsymbol{\sigma}(u) = \boldsymbol{\rho}(u)\boldsymbol{r}(u)^{-2\boldsymbol{\nu}(u)}\Gamma(\boldsymbol{\nu}(u))\exp(\boldsymbol{\nu}(u))\boldsymbol{\mu}(u)^{-\boldsymbol{\nu}(u)-d/2}$

(B) $\boldsymbol{r}(u) = r\sqrt{\boldsymbol{\psi}(u)+\boldsymbol{\nu}(u)}$ and $\boldsymbol{\sigma}(u) = \boldsymbol{\rho}(u)\Gamma(\boldsymbol{\nu}(u))\exp(\boldsymbol{\nu}(u))\left(\frac{\boldsymbol{r}(u)}{r}\right)^{-2\boldsymbol{\nu}(u)}$

(C) $\boldsymbol{r}(u) = \sqrt{\frac{\boldsymbol{\nu}(u)}{\boldsymbol{\mu}(u)}}$ and $\boldsymbol{\sigma}(u) = \boldsymbol{\rho}(u)\boldsymbol{r}(u)^{d}\Gamma(\boldsymbol{\nu}(u))\exp(\boldsymbol{\nu}(u))\boldsymbol{\nu}(u)^{-\boldsymbol{\nu}(u)-d/2}$

(D) $\boldsymbol{\nu}(u) \leq \nu$, $\boldsymbol{r}(u) = \sqrt{\boldsymbol{\mu}(u)}$ and $\boldsymbol{\sigma}(u) = \boldsymbol{\rho}(u)\frac{\Gamma(\boldsymbol{\nu}(u))}{\Gamma(\boldsymbol{\nu}(u)+d/2)}\left[\frac{\Gamma(\nu_{ii}(u)+\nu_{jj}(u)+d/2)}{r_{ij}(u)^{2\nu+\nu_{ii}(u)+\nu_{ii}(u)}}\right]_{i,j=1}^{p}$

(E) $\boldsymbol{\nu}(u) = \nu\mathbf{1}$, $\boldsymbol{r}(u) = \boldsymbol{\mu}(u)$ and $\boldsymbol{\sigma}(u) = \boldsymbol{\rho}(u)\boldsymbol{r}(u)^{-\lfloor\frac{d+1+3\lceil 2\nu\rceil}{2}\rfloor}$

(F) $\boldsymbol{r}(u) = r\mathbf{1}$ and $\boldsymbol{\sigma}(u) = \boldsymbol{\rho}(u)\frac{\Gamma(\boldsymbol{\nu}(u))}{\Gamma(\boldsymbol{\nu}(u)+d/2)}$

(G) $\nu_{ij}(u) = \frac{\nu_{ii}(u)+\nu_{jj}(u)}{2}$, $\boldsymbol{r}(u) = r\mathbf{1}$ and $\boldsymbol{\sigma}(u) = \frac{\boldsymbol{\rho}(u)\Gamma(\boldsymbol{\nu}(u))}{\Gamma(\boldsymbol{\nu}(u)+d/2)}\left[\sqrt{\frac{\Gamma(\nu_{ii}(u)+d/2)\Gamma(\nu_{jj}(u)+d/2)}{\Gamma(\nu_{ii}(u))\Gamma(\nu_{jj}(u))}}\right]_{i,j=1}^{p}$,

with $\lfloor\cdot\rfloor$ *and* $\lceil\cdot\rceil$ *standing for the floor and ceil functions.*

Appendix B. A simulation algorithm for parsimonious models

We provide a simulation algorithm for the parsimonious Gneiting–Matérn and Gneiting–Cauchy multivariate models of Section 3.

A realization of a p-variate zero-mean random field $\boldsymbol{Z} = [Z_i]_{i=1}^{p}$ with approximate Gaussian distribution and with covariance as in (10) can be constructed from the standardized sum of a large number M of independent copies of basic random fields, each of which depending only on two random variables R and Φ, two random vectors $\boldsymbol{V}$ and $\boldsymbol{\Omega}$ and a temporal p-variate random field $\boldsymbol{W}$. This is summarized in Algorithm 1 hereinafter, where $\langle\cdot,\cdot\rangle$ stands for the usual scalar product in $\mathbb{R}^d$, $R \sim f$ is the short notation for "R is distributed according to the distribution f", $\mathcal{U}(0, 2\pi)$ is the uniform distribution on $(0, 2\pi)$, $\mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$ is the d-dimensional normal distribution with expectation $\mathbf{0}$ and variance–covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{I}_d$ is the $d \times d$ identity matrix. We refer to Arroyo and Emery (2021) for criteria on the choice of the distribution f.

The multivariate random field $\boldsymbol{W}$ can be simulated with Algorithm 2 hereinafter, which is applicable as long as the number of target time coordinates is not too large (in practice, up to a few tens of thousands), so that the positive semidefinite square root matrix can be computed (Horn and Johnson, 2013, Theorem 7.2.6). For larger numbers, iterative methods based on Gibbs sampling (Lantuéjoul and Desassis, 2012; Arroyo and Emery, 2020) can be used instead. Alternatively, for pseudo-variograms in the class presented in Section 4.2, $\boldsymbol{W}$ can be constructed by combinations of stationary Gaussian random fields and intrinsic random fields with Gaussian increments, for which discrete (Dietrich and Newsam, 1997; Stein, 2002) and continuous (Emery et al., 2016; Arroyo and Emery, 2017) spectral simulation algorithms are available.

Appendix C. A spectral simulation algorithm

A realization of a p-variate Gaussian random field $\boldsymbol{Z} = [Z_i]_{i=1}^{p}$ with zero mean and matrix-valued covariance function as in Theorem 2 can be constructed with a central limit approximation, by rescaling the sum of a large number M of independent copies of basic random fields. Specifically, let us consider that $\boldsymbol{Z}$ is to be simulated on a grid of n_S nodes in $\mathbb{R}^d$ crossed with a grid of n_T nodes $t_1, \ldots, t_{n_T}$ in $\mathbb{R}$. Let $q = p\,n_T$ and, for $\boldsymbol{\omega} \in \mathbb{R}^d$, define the $q \times q$ matrix $\boldsymbol{F}(\boldsymbol{\omega}) = [[F_{i,m;j,n}(\boldsymbol{\omega})]_{i,j=1}^{p}]_{m,n=1}^{n_T}$ whose generic entry is the spectral density of the Matérn covariance $\boldsymbol{h} \mapsto \sigma_{ij}(u_{mn})\,\mathcal{M}(\boldsymbol{h}; r_{ij}(u_{mn}), \nu_{ij})$, with $u_{mn} = t_m - t_n$, evaluated at $\boldsymbol{\omega}$ (Chilès and Delfiner, 2012):

$$F_{i,m;j,n}(\boldsymbol{\omega}) = \sigma_{ij}(u_{mn})\frac{\pi^{\frac{d}{2}}\,2^d\,\Gamma\left(\nu_{ij}+\frac{d}{2}\right)}{r_{ij}(u_{mn})^d\,\Gamma(\nu_{ij})}\left(1+\frac{|2\pi\boldsymbol{\omega}|^2}{r_{ij}^2(u_{mn})}\right)^{-\frac{d}{2}-\nu_{ij}}.$$

In the same way as in the proof of Theorem 2 given in Appendix D.2, the p-variate random field $\boldsymbol{Z}$ to simulate at the $n_S \times n_T$ target nodes of $\mathbb{R}^d \times \mathbb{R}$ can be viewed as a q-variate random field

Algorithm 1 Substitution simulation algorithm

Require: Admissible Matérn or Cauchy parameters as given in Theorem 1
Require: Admissible pseudo-variogram $\boldsymbol{\eta}$
Require: A probability density function f on $(0, \infty)$ whose support contains the supports of $f_1, \ldots, f_p$
Require: A large integer M
1: Initialize $Z_i(\boldsymbol{s}, t) = 0$, $\forall i$ and at each target space–time coordinate $(\boldsymbol{s}, t) \in \mathbb{R}^d \times \mathbb{R}$
2: **for** $m = 1$ to M **do**
3: Simulate $R \sim f$
4: Simulate $\boldsymbol{\Omega} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{I}_d)$
5: Simulate $\Phi \sim \mathcal{U}(0, 2\pi)$
6: Simulate $\boldsymbol{V} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\sigma})$
7: At each target time coordinate t, simulate a p-variate random field $\boldsymbol{W} = [W_i]_{i=1,\ldots,p}$ with stationary Gaussian direct and cross-increments, with zero mean and pseudo-variogram $\boldsymbol{\eta}$
8: **for** each target space–time coordinate $(\boldsymbol{s}, t)$ **do**
9: Compute

$$Z_i(\boldsymbol{s}, t) = Z_i(\boldsymbol{s}, t) + \sqrt{\frac{2}{M}} \sqrt{\frac{f_i(R)}{f(R)}}\, V_i \cos\left(\sqrt{2R}\, \langle \boldsymbol{\Omega}, \boldsymbol{s} \rangle + \frac{|\boldsymbol{\Omega}|}{\sqrt{2}} W_i(t) + \Phi\right), \quad i = 1, \ldots, p$$

10: **end for**
11: **end for**

Algorithm 2 Matrix decomposition algorithm

Require: Admissible pseudo-variogram $\boldsymbol{\eta}$
Require: Set of time coordinates $t_1, \ldots, t_{n_T}$ targeted for simulation
1: Calculate the $(p\, n_T) \times (p\, n_T)$ variance–covariance matrix (Papritz et al., 1993, Eq. 6)

$$\boldsymbol{C}_W = [[\eta_{i1}(t_m) + \eta_{1j}(-t_n) - \eta_{ij}(t_m - t_n)]_{i,j=1}^{p}]_{m,n=1}^{n_T}$$

2: Simulate $\boldsymbol{X} \sim \mathcal{N}_{p\, n_T}(\mathbf{0}, \boldsymbol{I}_{p\, n_T})$
3: Compute $\boldsymbol{W} = [[W_i(t_n)]_{i=1}^{p}]_{n=1}^{n_T} = \boldsymbol{C}_W^{1/2} \boldsymbol{X}$, with $\boldsymbol{C}_W^{1/2}$ the positive semidefinite square root of $\boldsymbol{C}_W$

$\boldsymbol{Y}$ at n_S nodes of $\mathbb{R}^d$ with a multivariate Matérn covariance with parameters $\left[[r_{ij}(u_{mn})]_{i,j=1}^{p}\right]_{m,n=1}^{n_T}$, $\left[[\nu_{ij}]_{i,j=1}^{p}\right]_{m,n=1}^{n_T}$ and $\left[[\sigma_{ij}(u_{mn})]_{i,j=1}^{p}\right]_{m,n=1}^{n_T}$. The simulation can be done by means of the continuous spectral algorithm presented in Emery et al. (2016), which relies on importance sampling and on the square root decomposition of the $q \times q$ positive semidefinite matrix $\boldsymbol{F}(\boldsymbol{\Omega})$ evaluated at a random vector $\boldsymbol{\Omega}$. In addition to parameters verifying the conditions of Theorem 2, a probability density function g that does not vanish on $\mathbb{R}^d$ is thus required to define $\boldsymbol{\Omega}$. This is summarized in Algorithm 3, where $\langle \cdot, \cdot \rangle$ stands for the usual scalar product in $\mathbb{R}^d$, $R \sim f$ is the short notation for "R is distributed according to the distribution f" and $\mathcal{U}(0, 2\pi)$ is the uniform distribution on $(0, 2\pi)$.

The spectral simulation algorithm is applicable when q is not too large, say, less than a few tens of thousands, so that the square root decomposition of the spectral density matrix $\boldsymbol{F}(\boldsymbol{\Omega})$ can be achieved. There is no such restriction on the number of spatial nodes n_S, which can be much higher, insofar as the simulation in space is achieved in a continuous manner by using cosine waves that can be computed at as many coordinates as desired. We refer to Arroyo and Emery (2021) for a proper choice of the distribution g.

Algorithm 3 Spectral simulation algorithm

Require: Admissible parameters for $\boldsymbol{\nu}$, $\boldsymbol{b}^2$ and $\boldsymbol{a}^2$
Require: Admissible pseudo-variogram $\boldsymbol{\eta}$
Require: A probability density function g not vanishing on $\mathbb{R}^d$
Require: A large integer M
Initialize $\boldsymbol{Y}(\boldsymbol{s}_k) = \boldsymbol{0}$ for each target spatial coordinate $\boldsymbol{s}_k$, $k = 1, \ldots, n_S$
for $m = 1$ to M **do**
  Simulate $\boldsymbol{\Omega} \sim g$
  Compute the positive semidefinite square root $\boldsymbol{H}(\boldsymbol{\Omega})$ of matrix $2\boldsymbol{F}(\boldsymbol{\Omega})/g(\boldsymbol{\Omega})$
  Simulate $\Phi \sim \mathcal{U}(0, 2\pi)$
  Simulate a random integer Q uniformly distributed in $\{1, \ldots, q\}$
  for each target spatial coordinate $\boldsymbol{s}_k$ **do**
    Compute

$$\boldsymbol{Y}(\boldsymbol{s}_k) = \boldsymbol{Y}(\boldsymbol{s}_k) + \sqrt{\frac{q}{M}}\boldsymbol{H}(\boldsymbol{\Omega}; Q)\cos(2\pi\langle\boldsymbol{\Omega}, \boldsymbol{s}_k\rangle + \Phi)$$

    where $\boldsymbol{H}(\boldsymbol{\Omega}; Q)$ denotes the Q-th column of $\boldsymbol{H}(\boldsymbol{\Omega})$
  end for
end for
$\boldsymbol{Z}$:= re-ordering of $\boldsymbol{Y}$ into a p-variate random field at $n_S \times n_T$ space–time coordinates

Appendix D. Technical lemmas and proofs

D.1. Technical lemmas

Lemma 1. *Let $\boldsymbol{a} = [a_{ij}]_{i,j=1}^p$ be a real, symmetric conditionally negative semidefinite matrix. Then, there exist random variables $Y_1, \ldots, Y_p$ such that:*

$$a_{ij} = \frac{a_i + a_j}{2} + \frac{1}{2}\mathrm{Var}(Y_i - Y_j), \quad i, j = 1, \ldots, p.$$

Proof. The matrix $\boldsymbol{b} = [b_{ij}]_{i,j=1}^p$ with $b_{ij} = a_{ip} + a_{pj} - a_{ij} - a_{pp}$ is positive semidefinite (Berg et al., 1984, Chapter 3, Lemma 2.1). Let $Y_1, \ldots, Y_p$ be Gaussian random variables with variance–covariance matrix $\boldsymbol{b}$. Then

$$\frac{1}{2}\mathrm{Var}(Y_i - Y_j) = \frac{b_{ii} + b_{jj}}{2} - b_{ij} = a_{ij} - \frac{a_{ii} + a_{jj}}{2}, \quad i, j = 1, \ldots, p. \quad \square$$

Property 1. *A matrix is conditionally null definite if and only if it is a-separable.*

Property 2. *All p-separable matrices with nonnegative entries are positive semidefinite.*

Property 3. *If $\boldsymbol{a} \preceq_c 0$, then $\exp(-t\boldsymbol{a}) \succeq 0$ for all $t \geq 0$.*

Property 4. *If $\boldsymbol{a} \preceq_c 0$ and $\boldsymbol{b}$ is a-separable, with $\boldsymbol{a}$ and $\boldsymbol{b}$ of size $p \times p$, then $\boldsymbol{a} + \boldsymbol{b} \preceq_c 0$ and $\boldsymbol{a} - \boldsymbol{b} \preceq_c 0$. In particular, $\boldsymbol{1} - \boldsymbol{b} \preceq_c 0$, where $\boldsymbol{1}$ is the all-ones matrix of size $p \times p$.*

Proof. The fourth property derives from the first one, which can be proven by using Lemma 1, and from the fact that the sum of two conditionally negative (positive) semidefinite matrices is still conditionally negative (positive) semidefinite. For a proof of the second and third properties, see Berg et al. (1984, Chapter 3, Property 1.9 and Theorem 2.2). $\square$

Lemma 2. *Let $\boldsymbol{C} : u \mapsto [C_{ij}(u)]_{i,j=1}^p$, $u \in \mathbb{R}$, be the matrix-valued covariance function of a p-variate second-order stationary random field $\boldsymbol{Y}$ in $\mathbb{R}$. Then, for any positive integer n_T and real values $t_1, \ldots, t_{n_T}$, the matrix $\boldsymbol{C}_0 = [[C_{ij}(t_m - t_n)]_{i,j=1}^p]_{m,n=1}^{n_T}$ is positive semidefinite.*

Proof. It is immediate because $\boldsymbol{C}_0$ is the variance–covariance matrix of the random vector $[\boldsymbol{Y}(t_n)]_{n=1}^{n_T}$. □

Lemma 3. *Let $\boldsymbol{a} = [a_{ij}]_{i,j=1}^{p}$ be a conditionally negative semidefinite matrix and $\boldsymbol{\eta} : u \mapsto [\eta_{ij}(u)]_{i,j=1}^{p}$, $u \in \mathbb{R}$, be the matrix-valued pseudo-variogram (dependent only on the lag vector u) of a multivariate random field $\boldsymbol{Y} = [Y_i]_{i=1}^{p}$ in $\mathbb{R}$ with stationary direct and cross-increments. Then, for any positive integer n_T and real values $t_1, \ldots, t_{n_T}$, the matrix $\boldsymbol{\eta}_0 = [[a_{ij} + \eta_{ij}(t_m - t_n)]_{i,j=1}^{p}]_{m,n=1}^{n_T}$ is conditionally negative semidefinite.*

Proof. Let $\boldsymbol{\Lambda} = [[\lambda_{m,i}]_{i=1}^{p}]_{m=1}^{n_T}$ be a real vector with entries adding to zero. Define $\widetilde{\boldsymbol{\Lambda}} = [\widetilde{\lambda}_i]_{i=1}^{p}$ and $\widetilde{\boldsymbol{C}}_0 = [[C_{i,j;m,n}]_{i,j=1}^{p}]_{m,n=1}^{n_T}$ with

$$\widetilde{\lambda}_i = \sum_{m=1}^{n_T} \lambda_{m,i}, \quad i = 1, \ldots, p$$

and

$$C_{i,j;m,n} = \mathrm{Cov}(Y_i(t_m) - Y_1(t_1), Y_j(t_n) - Y_1(t_1)), \quad i, j = 1, \ldots, p, \quad m, n = 1, \ldots, n_T.$$

Then:

$$\begin{aligned}
&\sum_{m=1}^{n_T}\sum_{n=1}^{n_T}\sum_{i=1}^{p}\sum_{j=1}^{p} \lambda_{m,i}\lambda_{n,j}[a_{ij} + \eta_{ij}(t_m - t_n)] \\
&= \sum_{m=1}^{n_T}\sum_{n=1}^{n_T}\sum_{i=1}^{p}\sum_{j=1}^{p} \lambda_{m,i}\lambda_{n,j}[a_{ij} + \frac{1}{2}\mathrm{Var}(Y_i(t_m) - Y_1(t_1) + Y_1(t_1) - Y_j(t_n))] \\
&= \sum_{m=1}^{n_T}\sum_{n=1}^{n_T}\sum_{i=1}^{p}\sum_{j=1}^{p} \lambda_{m,i}\lambda_{n,j}[a_{ij} + \eta_{i1}(t_m - t_1) + \eta_{j1}(t_n - t_1) \\
&\quad - \mathrm{Cov}(Y_i(t_m) - Y_1(t_1), Y_j(t_n) - Y_1(t_1))] \\
&= \left(\sum_{n=1}^{n_T}\sum_{j=1}^{p} \lambda_{n,j}\right)\left(\sum_{m=1}^{n_T}\sum_{i=1}^{p} \lambda_{m,i}\,\eta_{i1}(t_m - t_1)\right) + \left(\sum_{m=1}^{n_T}\sum_{i=1}^{p} \lambda_{m,i}\right)\left(\sum_{n=1}^{n_T}\sum_{j=1}^{p} \lambda_{n,j}\,\eta_{j1}(t_n - t_1)\right) \\
&\quad + \sum_{i=1}^{p}\sum_{j=1}^{p} \widetilde{\lambda}_i\widetilde{\lambda}_j a_{ij} - \sum_{m=1}^{n_T}\sum_{n=1}^{n_T}\sum_{i=1}^{p}\sum_{j=1}^{p} \lambda_{m,i}\lambda_{n,j}\,C_{i,j;m,n}.
\end{aligned}$$

Since the entries of $\boldsymbol{\Lambda}$ add to zero, the first two terms in the last equality are zero, while the last two terms are negative or zero since, on the one hand, the entries of $\widetilde{\boldsymbol{\Lambda}}$ add to zero and $\boldsymbol{a}$ is conditionally negative semidefinite and, on the other hand, $\widetilde{\boldsymbol{C}}_0$ is a variance–covariance matrix, hence positive semidefinite. □

D.2. Proofs

Proof of Theorem 1. The proof is constructive and based on an extension of the substitution approach proposed in Allard et al. (2020). Consider the p-variate space–time random field $\boldsymbol{Z} = [Z_i]_{i=1}^{p}$ defined as

$$Z_i(\boldsymbol{s}, t) = \sqrt{\frac{2f_i(R)}{f(R)}}\, V_i \cos\left(\sqrt{2R}\,\langle \boldsymbol{\Omega}, \boldsymbol{s}\rangle + \frac{|\boldsymbol{\Omega}|}{\sqrt{2}} W_i(t) + \Phi\right), \qquad (\boldsymbol{s}, t) \in \mathbb{R}^d \times \mathbb{R},$$

where R is a nonnegative random variable with a probability density f whose support contains that of $f_1, \ldots, f_p$, $\boldsymbol{V} = [V_i]_{i=1}^{p}$ is a centered Gaussian random vector with covariance matrix $\boldsymbol{\sigma} = [\sigma_{ij}]_{i,j=1}^{p}$, $\boldsymbol{\Omega}$ is a standard Gaussian random vector with d independent components, Φ is a uniform random

variable in $(0, 2\pi)$, and W_i, $i = 1, \ldots, p$, are random fields on $\mathbb{R}$ whose direct and cross-increments are second-order stationary, characterized by the matrix-valued pseudo-variogram $\boldsymbol{\eta}$, and Gaussian. The random variable R, the random vectors $\boldsymbol{\Omega}$, Φ and $\boldsymbol{V}$ and the random field $\boldsymbol{W} = [W_i]_{i=1}^p$ are independent.

Following Allard et al. (2020, Theorem 3), it is straightforward to show that, for any $i, j = 1, \ldots, p$, Z_i and Z_j have a zero mean and that the covariance function between Z_i and Z_j is

$$\begin{aligned}
&\mathbb{E}[Z_i(\boldsymbol{s}, t)Z_j(\boldsymbol{s}', t')] \\
&= \frac{\sigma_{ij}}{(2\pi)^{d/2}} \int_0^\infty \int_{\mathbb{R}^d} \sqrt{f_i(r)f_j(r)} \cos(\sqrt{2r}\, \langle \boldsymbol{\omega}, \boldsymbol{s} - \boldsymbol{s}' \rangle) \exp\left(-|\boldsymbol{\omega}|^2 \frac{\eta_{ij}(t - t') + 1}{2}\right) \mathrm{d}\boldsymbol{\omega}\mathrm{d}r \\
&= \frac{\sigma_{ij}}{(\eta_{ij}(t - t') + 1)^{d/2}} \varphi_{ij}\left(\frac{|\boldsymbol{s} - \boldsymbol{s}'|^2}{\eta_{ij}(t - t') + 1}\right), \qquad (\boldsymbol{s}, t) \in \mathbb{R}^d \times \mathbb{R}, (\boldsymbol{s}', t') \in \mathbb{R}^d \times \mathbb{R}.
\end{aligned}$$

The cross-covariance between Z_i and Z_j depends only on the spatial and temporal lags $\boldsymbol{h} = \boldsymbol{s} - \boldsymbol{s}'$ and $u = t - t'$. Accordingly, the random field $\boldsymbol{Z}$ is second-order stationary with zero mean and covariance function $\boldsymbol{C} : (\boldsymbol{h}, u) \mapsto \left[C_{ij}(\boldsymbol{h}, u)\right]_{i,j=1}^p$, with $C_{ij}(\boldsymbol{h}, u)$ given as in Theorem 1. □

Proof of Theorem 2. Without loss of generality, let us consider a regular design of $n_S \times n_T$ space–time coordinates $(\boldsymbol{s}_1, t_1), \ldots, (\boldsymbol{s}_{n_S}, t_1), \ldots, (\boldsymbol{s}_{n_S}, t_{n_T}) \in \mathbb{R}^d \times \mathbb{R}$. Define $u_{mn} = t_m - t_n$, $\boldsymbol{h}_{kl} = \boldsymbol{s}_k - \boldsymbol{s}_l$ and

$$\boldsymbol{M} = [[[\sigma_{ij}(u_{mn})\, \mathcal{M}(\boldsymbol{h}_{kl}; r_{ij}(u_{mn}), \nu_{ij})]_{k,l=1}^{n_S}]_{i,j=1}^{p}]_{m,n=1}^{n_T}.$$

For $\boldsymbol{C}$ to be a positive semidefinite function, we need to show that $\boldsymbol{M}$ is positive semidefinite for any such design in $\mathbb{R}^d \times \mathbb{R}$. This can be done by viewing the p-variate spatio-temporal random field at n_T time coordinates as a q-variate spatial random field, with $q = p\, n_T$, and using the sufficient validity conditions for multivariate spatial Matérn models given in Emery et al. (2022). Specifically, if one sets $\boldsymbol{\psi}(u) = \beta\boldsymbol{\eta}(u) + \boldsymbol{b}^2$, under the conditions of Theorem 2, the $q \times q$ matrices $\left[[\psi_{ij}(u_{mn})]_{i,j=1}^p\right]_{m,n=1}^{n_T}$ and $\left[[r_{ij}^2(u_{mn})\psi_{ij}(u_{mn}) - \nu_{ij}]_{i,j=1}^p\right]_{m,n=1}^{n_T}$ are conditionally negative semidefinite (see Lemma 3). Furthermore, the matrix

$$\left[[\sigma_{ij}(u_{mn})e^{-\nu_{ij}} r_{ij}(u_{mn})^{2\nu_{ij}} \psi_{ij}(u_{mn})^{\nu_{ij}+d/2} / \Gamma(\nu_{ij})]_{i,j=1}^p\right]_{m,n=1}^{n_T} = \left[[\rho_{ij}(u_{mn})]_{i,j=1}^p\right]_{m,n=1}^{n_T}$$

is positive semidefinite (see Lemma 2). Hence, according to Emery et al. (2022, Theorem 2B), the matrix-valued function $\boldsymbol{h} \mapsto \sigma_{ij}(u_{mn})\, \mathcal{M}(\boldsymbol{h}; r_{ij}(u_{mn}), \nu_{ij})$ is the covariance function of a q-variate random field $\boldsymbol{Y} = [Y_v]_{v=1}^q$ in $\mathbb{R}^d$. $\boldsymbol{M}$ is therefore positive semidefinite since it is the variance–covariance matrix of the random vector $[[Y_v(\boldsymbol{s}_k)]_{k=1}^{n_S}]_{v=1}^q$. □

Proof of Corollary 1. In Theorem 2, $\boldsymbol{\rho}(u)$ must be a matrix-valued covariance function. Here, we set $\boldsymbol{\rho}(u) = \boldsymbol{\rho}$, for all $u \in \mathbb{R}$. By comparing (13) with (12), it is thus required that $\boldsymbol{\tau} e^{-\boldsymbol{\nu}} / \Gamma(\boldsymbol{\nu}) = \boldsymbol{\rho}$ is a positive semidefinite matrix. □

Appendix E. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spasta.2022.100706.

Space–time simulations of the proposed model with the parameters described in Section 4.3. We used Algorithm 3 with $M = 20{,}000$ cosine waves. The density g is the spectral density of a Matérn covariance function with $\nu = 0.5$ and $r = 2$. All simulations have been obtained with the same seed for random number generation. They are thus comparable, except in terms of space–time non-separability.

References

Allard, D., Emery, X., Lacaux, C., Lantuéjoul, C., 2020. Simulating space–time random fields with nonseparable Gneiting-type covariance functions. Stat. Comput. 30 (5), 1479–1495.

Apanasovich, T., Genton, M., 2010. Cross-covariance functions for multivariate random fields based on latent dimensions. Biometrika 97 (1), 15–30.

Apanasovich, T.V., Genton, M.G., Sun, Y., 2012. A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. J. Amer. Statist. Assoc. 107 (497), 180–193.

Arroyo, D., Emery, X., 2017. Spectral simulation of vector random fields with stationary Gaussian increments in d-dimensional Euclidean spaces. Stoch. Environ. Res. Risk Assess. 31 (7), 1583–1592.

Arroyo, D., Emery, X., 2020. Iterative algorithms for non-conditional and conditional simulation of Gaussian random vectors. Stoch. Environ. Res. Risk Assess. 34 (10), 1523–1541.

Arroyo, D., Emery, X., 2021. Algorithm 1013: An R implementation of a continuous spectral algorithm for simulating vector Gaussian random fields in Euclidean spaces. ACM Trans. Math. Software 47 (1), 8.

Berg, C., Christensen, J.P.R., Ressel, P., 1984. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions. Springer-Verlag.

Bevilacqua, M., Gaetan, C., Mateu, J., Porcu, E., 2012. Estimating space and space–time covariance functions for large data sets: A weighted composite likelihood approach. J. Amer. Statist. Assoc. 107 (497), 268–280.

Bourotte, M., Allard, D., Porcu, E., 2016. A flexible class of non-separable cross-covariance functions for multivariate space–time data. Spatial Stat. 18, 125–146.

Chen, W., Genton, M.G., 2019. Parametric variogram matrices incorporating both bounded and unbounded functions. Stoch. Environ. Res. Risk Assess. 33 (10), 1669–1679.

Chen, W., Genton, M.G., Sun, Y., 2021. Space–time covariance structures and models. Annu. Rev. Stat. Appl. 8, 191–215.

Chilès, J.-P., Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty, second ed. John Wiley & Sons.

Choi, J., Fuentes, M., Reich, B., Davis, J., 2009. Multivariate spatial–temporal modeling and prediction of speciated fine particles. J. Stat. Theory Pract. 3 (2), 407–418.

De Iaco, S., Myers, D., Posa, D., 2003. The linear coregionalization model and the product–sum space–time variogram. Math. Geol. 35 (1), 25–38.

De Iaco, S., Posa, D., 2013. Positive and negative non-separability for space–time covariance models. J. Statist. Plann. Inference 143 (2), 378–391.

Dietrich, C., Newsam, G., 1997. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. SIAM J. Sci. Comput. 18 (4), 1088–1107.

Dörr, K., Schlather, M., 2021. Characterization theorems for pseudo-variograms. arXiv, arXiv:2112.04026.

Emery, X., Arroyo, D., Porcu, E., 2016. An improved spectral turning-bands algorithm for simulating stationary vector Gaussian random fields. Stoch. Environ. Res. Risk Assess. 30 (7), 1863–1873.

Emery, X., Lantuéjoul, C., 2006. TBSIM: A computer program for conditional simulation of three-dimensional Gaussian random fields via the turning bands method. Comput. Geosci. 32 (10), 1615–1628.

Emery, X., Porcu, E., White, P., 2022. New validity conditions for the multivariate Matérn coregionalization model, with an application to exploration geochemistry. Math. Geosci. 54, 1043–1068.

Finazzi, F., Scott, E.M., Fassò, A., 2013. A model-based framework for air quality indices and population risk evaluation, with an application to the analysis of Scottish air quality data. J. R. Stat. Soc. Ser. C. Appl. Stat. 62 (2), 287.

Gelfand, A., 2021. Multivariate spatial process models. In: Fischer, M.M., Nijkam, P. (Eds.), HandBook of Regional Science. Springer, Dordrecht, pp. 1985–2016.

Gelfand, A., Banerjee, S., Gamerman, D., 2005. Spatial process modelling for univariate and multivariate dynamic spatial data. Environmetrics 16 (5), 465–479.

Gneiting, T., 2002. Nonseparable, stationary covariance functions for space–time data. J. Amer. Statist. Assoc. 97 (458), 590–600.

Gneiting, T., Kleiber, W., Schlather, M., 2010. Matérn cross-covariance functions for multivariate random fields. J. Amer. Statist. Assoc. 105 (491), 1167–1177.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Amer. Statist. Assoc. 102 (477), 359–378.

Horn, R.A., Johnson, C.R., 2013. Matrix Analysis. Cambridge University Press.

Ip, R., Li, W., 2016. Matérn cross-covariance functions for bivariate spatio-temporal random fields. Spatial Stat. 17 (1), 22–37.

Katzfuss, M., Guinness, J., 2021. A general framework for Vecchia approximations of Gaussian processes. Statist. Sci. 36 (1), 124–141.

Lantuéjoul, C., Desassis, N., 2012. Simulation of a Gaussian random vector: A propagative version of the Gibbs sampler. In: Presented at: Ninth International Geostatistics Congress, Held in Oslo. Norway, pp. 1–8.

Mitchell, M.W., Genton, M.G., Gumpertz, M.L., 2006. A likelihood ratio test for separability of covariances. J. Multivariate Anal. 97 (5), 1025–1043.

Oesting, M., Schlather, M., Friederichs, P., 2017. Statistical post-processing of forecasts for extremes using bivariate Brown-Resnick processes with an application to wind gusts. Extremes 20 (2), 309–332.

Padoan, S.A., Bevilacqua, M., 2015. Analysis of random fields using CompRandFld. J. Stat. Softw. 63, 1–27.

Papritz, A., Künsch, H., Webster, R., 1993. On the pseudo cross-variogram. Math. Geol. 25 (8), 1015–1026.

Porcu, E., Emery, X., Mery, N., 2022. Criteria and characterizations for spatially isotropic and temporally symmetric matrix-valued covariance functions. Comput. Appl. Math. 41 (223).

Porcu, E., Furrer, R., Nychka, D., 2021. 30 Years of space–time covariance functions. WIREs Comput. Stat. 13 (2), e1512.

Qadir, G.A., Sun, Y., 2022. Modeling and predicting spatio-temporal dynamics of PM_{2.5} concentrations through time-evolving covariance models. arXiv preprint arXiv:2202.12121.

Rodrigues, A., Diggle, P., 2010. A class of convolution based models for spatio-temporal processes with non-separable covariance structure. Scand. J. Stat. 37 (4), 553–567.
Rouhani, S., Wackernagel, H., 1990. Multivariate geostatistical approach to space–time data analysis. Water Resour. Res. 36 (4), 585–591.
Stein, M., 2002. Fast and exact simulation of fractional Brownian surfaces. J. Comput. Graph. Statist. 11 (3), 587–599.
Wackernagel, H., 2003. Multivariate Geostatistics: An Introduction with Applications. Springer Science & Business Media.
Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J. Amer. Statist. Assoc. 99 (465), 250–261.

S.2 Discussion on "Competition on Spatial Statistics for Large Datasets"

Supplementary materials for this article are available at https://doi.org/10.1007/s13253-021-00462-2.

Discussion on "Competition on Spatial Statistics for Large Datasets"

Denis ALLARD, Lucia CLAROTTO, Thomas OPITZ, and Thomas ROMARY

We discuss the methods and results of the RESSTE team in the competition on spatial statistics for large datasets. In the first sub-competition, we implemented block approaches both for the estimation of the covariance parameters and for prediction using ordinary kriging. In the second sub-competition, a two-stage procedure was adopted. In the first stage, the marginal distribution is estimated neglecting spatial dependence, either according to the flexible Tuckey g and h distribution or nonparametrically. In the second stage, estimation of the covariance parameters and prediction are performed using Kriging. Vecchias's approximation implemented in the `GpGp` package proved to be very efficient. We then make some propositions for future competitions.

Key Words: Composite likelihood; Block likelihood; Block approach; Tuckey g and h; Vecchia's approximation.

1. INTRODUCTION

We congratulate the authors for organizing such a great and challenging competition. Our team brings together researchers from two French groups that have long-standing collaborations: the BioSP research unit at INRAE and the Geostatistics team at Mines ParisTech (formerly known as Ecole des Mines de Paris). They both belong to the RESSTE network[1] funded by INRAE. RESSTE organizes scientific animation around models, methods and algorithms for spatiotemporal data, and it fosters collaborations between statisticians and other scientists sharing interest in spatial and spatiotemporal data. Forced to be physically distant due to the Covid-19 sanitary crisis, we set up an efficient working environment with the help of collaborative online platforms for code, text and vivid discussions. We were thus able to contribute to all sub-competitions, including with multiple submissions for sub-competition 2. We enjoyed very much participating to this exercise. In addition to

Authors have contributed equally and are listed in alphabetical order.

D. Allard (✉) and T. Opitz, Biostatistics and Spatial Processes (BioSP), INRAE, 84914 Avignon, France (E-mail: *denis.allard@inrae.fr*). L. Clarotto and T. Romary, MINES ParisTech, PSL University, Centre de Geosciences, 77300 Fontainebleau, France.

Journal of Agricultural, Biological, and Environmental Statistics
https://doi.org/10.1007/s13253-021-00462-2

[1] https://informatique-mia.inrae.fr/reseau-resste/

Published online: 23 July 2021

congratulate Huang, Sameh, Ying, Hatem, David and Marc for the excellent organization, we sincerely wish to thank them for the nice moments it brought to the four of us.

2. COMPETITION 1: BLOCK APPROACHES FOR GAUSSIAN PROCESSES

Data are known to have been generated from a Gaussian process with Matérn covariance function. Maximum likelihood (ML) would thus be the most efficient estimation method, and conditional expectation, also referred to as Kriging, is optimal for prediction. However, due to the size of the dataset, neither full ML nor unique neighborhood Kriging can be achieved, at the exception of *ExaGeoStat*. Efficient approximations must be sought. Here, we have opted for block approaches that satisfy the following principles: (i) in each block, estimation (ML) or prediction (Kriging) is optimal; (ii) blocks should be as large as possible while taking into account computational issues; (iii) blocks are assumed to be independent among each other. The approximation lies entirely in point (iii), and it is easy to understand that (ii) is the key for achieving good performances. What "large" means has slightly different meaning for estimation and for prediction. As regards estimation, some parameters control local properties, while others are global. Therefore, blocks for estimation need to have a large spatial extent. For plug-in Kriging, only local information is necessary. Prediction blocks are thus local, containing as many data points as possible. Details are provided below.

2.1. ESTIMATION

First, a rough estimation of the parameters was performed on each dataset with weighted least squares fits for experimental variograms using the package `RGeostats`. These estimates, from which an approximate effective range $ER = \hat{\beta}\sqrt{12\hat{\nu}}$ was computed, allowed us to gain a general picture of the experimental design similar to that shown in Table 1 in Huang et al. (2021). In particular, ER was clearly close to the size of the domain for some datasets.

In a second stage, estimation of the parameters was performed using a maximum composite block-likelihood (BL) method. Blocks are characterized by their size (number of data points, N_D), shape and location of the data within the blocks. Nugget ($\tau^2 \geq 0$) and smoothness ($\nu > 0$) are local parameters, whilst range ($\beta > 0$) and sill ($\sigma^2 > 0$) are non-local. When $\tau^2 = 0$, Zhang (2004) showed that the only quantity that can be efficiently estimated in an in-fill asymptotic framework is $\sigma^2\beta^{-2\nu}$. Efficient estimation of all parameters thus requires a "large domain" framework that allows sampling small distances for estimating τ^2 and ν, and intermediate to large distances with respect to ER for estimating β and σ^2. In each block, the sub-sample must be built so as to sample all distances from 0 to a multiple of the practical range. Data separated by a distance larger than 2 to 3 times the practical range are useless for estimating nugget, regularity and range parameters and can be excluded. Here, blocks were disks with a radius set to $1.5ER$, centered on a regular $B \times B$ grid covering the domain. N_D points were then sampled at random within each disk with a weight decreasing linearly from the center (where it is equal to 1) to the edge of the disk (where it is equal to 0).

After some experiments, the final setting was $B = 15$ and $N_D = 500$. For datasets G4/G12, when the smoothness parameter is large, the covariance matrix was seen as singular in `R`; in these cases, N_D was set to 400. The function `nlbmin` was used for minimizing the negative log-BL. Initial values for the optimization were given by the WLS estimates of the first stage. Since blocks are random, this operation is repeated 5 times for each dataset. For each parameter, the average of the estimates was computed. This average is the final estimate.

2.2. PREDICTION

A "local unique neighborhood" technique is adopted, in which Kriging is performed at all locations belonging to the same block using a common neighborhood. To this end, a $K \times K$ regular grid is defined on the domain, and one Kriging system is built for each mesh of that grid, hereafter referred to as target block. The neighborhood of a given target block is a disk of radius $1.5\sqrt{2}/K$, so that all data belonging to the 3×3 meshes surrounding the center of the target block are part of the Kriging system. Smaller K yields higher precision but is computationally more demanding. Larger K leads to smaller neighborhood and lower precision. Computing time decreases rapidly as K increases, roughly at a K^{-4} rate. A good trade-off between performance and speed was obtained with $K = 31$. The average number of training data per Kriging system was therefore around 1200.

2.3. DISCUSSION

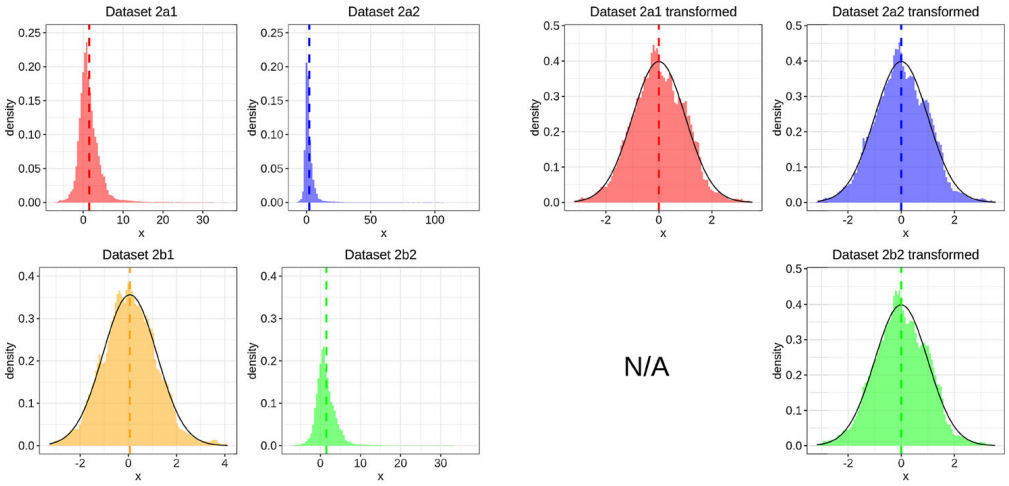
Overall these block approaches performed relatively well, ranking fourth and second in sub-competitions 1a and 1b, respectively. It is noticeable that in sub-competition 1b the plug-in block Kriging described above was only outperformed by *ExaGeoStat* using the true model, whereas *ExaGeoStat* with the estimated parameters performed slightly less well. As expected, we experienced some difficulties for the simultaneous estimation of β, ν and σ^2 for smooth GPs with large effective range, i.e., when $\nu > 0.6$ and $ER > 0.1$. G4 is particularly poorly estimated, with simultaneous underestimation of the sill ($\hat{\sigma}^2 = 1.2092$) and underestimation of the range ($\hat{\beta} = 0.0486$), resulting in a high MMOM and RMSE—even though MLOE is relatively small. Estimating ν when the regularity is high in the presence of a nugget is very challenging (G12, G13 and G15), as can be seen in Figures 2 and S1. In these cases, the poor estimation translated into high RMSEs for plug-in Kriging.

3. COMPETITION 2: TRANS-GAUSSIAN MODELS

In this competition, the generating mechanism of the data was supposed to be unknown. Therefore, we first explored marginal distributions of the data using tools such as boxplots and histograms to check if a Gaussian model for the marginal distributions makes sense. If not, we followed the well-established statistical practice of using a marginally transformed Gaussian model. This approach, among which the Box–Cox transformation is the most popular representative, allows accommodating data features such as heavier-than-Gaussian tails or asymmetry of upper and lower tails.

Table 1. Mean, minimal and maximal value of the datasets of competition 2

Dataset	Min	Mean	Max
2a1	−7.1426	1.5141	36.0287
2a2	−7.3185	2.2249	126.789
2b1	−3.2676	0.0652	4.1168
2b2	−7.2361	1.504	37.1778



(a) 4 original datasets

(b) 4 datasets after Tukey g and h transform

Figure 1. Histograms of the datasets of competition 2 with superimposed Gaussian density when adequate .

3.1. Exploration of the Data

The first step of the solution to sub-competitions 2a and 2b is related to the exploratory analysis of the dataset. Some statistical quantities such as the mean, the minimum and maximum values (Table 1) were computed along with the histograms for each of the four datasets (Fig. 1a). It is clearly visible that only one of the four datasets (2b1) could be considered as marginally Gaussian distributed. The histograms of the other three datasets (2a1, 2a2, 2b2) present heavy tails toward high values. This fact is highlighted also by the maximal values of the datasets, which are extremely far from the mean if compared to the corresponding minimum values, showing asymmetry of tails. These insights suggest that the non-Gaussian datasets could be transformed to Gaussian models before applying geostatistical inference and prediction. Various types of transformations were inspected as reported in the next section.

3.2. Transformed Margins

3.2.1. Tukey g and h Transform

A flexible parametric marginal transform of Gaussian variables was proposed by J.W. Tukey and is known as the g and h distribution (Jorge and Boris 1984). It has been

recently studied for spatial Gaussian fields by Xu and Genton (2017). Tukey g and h transformation function is strictly monotonic and defined as follows:

$$\tau_{g,h}(x) = \begin{cases} \frac{1}{g}\left(\exp(gx) - 1\right)\exp\left(hx^2/2\right), & g \neq 0, \\ x\exp\left(hx^2/2\right), & g = 0, \end{cases} \quad x \in \mathbb{R}, \quad h \geq, \ g \in \mathbb{R}. \tag{1}$$

Given a standard Gaussian variable W, the Tukey g and h distribution is constructed as

$$Z = \beta_0 + \beta_1 \tau_{g,h}(W),$$

with parameters to control location (β_0), scale (β_1), asymmetry (g) and tail heaviness (h).

No external predictor variables were provided for the dataset, and visually we could not detect any other trends or anisotropies in the data. Therefore, we used a stationary and isotropic Tukey g and h random field model, which is obtained by applying the transformation (1) to a standard Gaussian random field with the Matérn covariance. We estimated the four parameters for each of the fields through the independence likelihood (neglecting spatial dependence) using the `R` library `OpVaR`; histograms of data after the inverse transformation to the standard Gaussian margins are shown in Figure 1b and correspond well to the superposed standard Gaussian density.

3.2.2. Nonparametric Transform

The estimation of the parameters of the Tukey g and h transform relies on an approximate parameter estimation procedure described below, which may conduct to underestimation or overestimation of the transformation parameters. Therefore, we also investigated the use of simple nonparametric transforms, namely a log transform for datasets 2a1 and 2b2 and a log-log transform for dataset 2a2. Their adequacy was checked by a visual inspection of the histograms of the transformed data (not represented here), in particular their symmetry.

3.3. Estimation and Prediction

Joint estimation of marginal and dependence parameters can be useful to allow for transfer of information between the models for margins and dependence, and for very accurate assessment of uncertainty in estimates. However, two-step approaches with separate estimation of marginal parameters, followed by marginal transformation to the standard Gaussian scale and estimation of the Gaussian correlation function, have the advantage of being more robust. In particular, they allow for the combination of different estimation techniques for margins and dependence. We here adopt two-step approaches. In the first step, which is common to two of our three approaches, we implement two substeps: (1) estimation of the marginal parameters β_0, β_1, g, h using the independence likelihood (i.e., by neglecting spatial dependence); (2) marginal transformation to the standard Gaussian scale using the parameters estimated in substep 1. In the last approach, we used a nonparametric transform. The following two subsections detail estimation after transformation of data to the standard Gaussian margins.

3.3.1. Bootstrap Approach

This approach is designed to run fast with moderate computing resources, such as those available on a personal computer. We proceed as follows using the pretransformed data:

1. Bootstrap estimations (25 samples) of Matérn correlation parameters using marginally pretransformed data: subsampling (without replacement) of 500 observations, and estimation of the scale and shape parameters of the Matérn correlation function.
2. The final estimated Matérn correlation parameters are set to the median of the bootstrap estimates.
3. Simple Kriging prediction is performed on the standard Gaussian scale by using k nearest neighbors of observed locations around the location to predict.
4. Standard Gaussian predictions are transformed back to the original scale by using the direct Tukey g and h transformation in (1) with parameters estimated in Step 1.

We have used validation data to choose among several values $k = 25, 50, 100$ of nearest neighbors in Step 5. The implementation was realized using the `R` library `CompRandFld` for estimation of covariance parameters and Kriging.

3.3.2. GpGp

Vecchia approximations are a particular case of composite likelihood methods. They can also provide an approximation of the parent Gaussian process (Katzfuss and Guinness 2021). The computations are based on the Cholesky factor of the inverse covariance matrix that can be computed explicitly and that is sparse by construction. Therefore, they allow for numerically efficient inference and prediction. The package `GpGp` proposes an implementation of a Vecchia approximation that uses an elaborate way to order and group the data into conditionally independent blocks (Guinness 2018). It also provides an implementation of the Fisher scoring algorithm for the ML estimation of the parameters (Guinness 2021). We have used this package for parameter estimation and prediction for marginally pretransformed data in datasets 2a1, 2a2 and 2b2, or directly for dataset 2b1. Then, the predictions were transformed back into their original scale. The only parameter to be set is the number of neighbors to be considered in the groups for estimation and prediction. It has been set by trial and error regarding the prediction performances on out-of-sample validation data.

3.4. VALIDATION

After transformation of the marginals (either through a nonparametric approach or through Tukey g and h), two parameters had to be set for the Vecchia approximation approach: the number of neighbors in the estimation step, n_e, and the number of neighbors in the prediction step, n_p. A holdout validation method was used to define the best n_e and n_p. Each time 70000 (resp. 700000) data were selected as training points in the datasets of sub-competition 2a (resp. 2b), and RMSE was computed over the 20000 (resp. 200000) remaining points. The values of n_e and n_p leading to the best RMSE were $n_e = 50$ for datasets in 2a, $n_e = 30$ for datasets in 2b and $n_p = 100$ for all datasets. In the bootstrap

Table 2. Submissions of the RESSTE team for competition 2

Submission	Margins (except 2b1)	Estimation+Prediction	Rank 2a	Rank 2b
Tukey-g-h-trans-bootstrap	Tukey g-h	Bootstrap (3.3.1)	5	6
Tukey-g-h-trans-GPGP	Tukey g-h	GpGp (3.3.2)	1	1 (ties)
nonpara-trans-GPGP	Non-parametric	GpGp (3.3.2)	2 (ties)	1 (ties)

approach (Sect. 3.3.1), the size of the nearest neighbor Kriging neighborhood was set to 25 (Table 2).

3.5. Discussion

Several methods were initially considered to treat the datasets of sub-competitions 2a and 2b, coming either from more classical geostatistical analyses or from machine learning techniques. Data-based methods such as random forests and neural networks, even when the spatial coordinates were combined with the addition of local features (mean/min/max values computed on K nearest neighbors), led to meager results in prediction.

Regarding competition 2, the type of point prediction to use depends on the score to be optimized. The conditional mean is known to minimize mean-squared error (MSE), and it corresponds to Kriging predictions. However, when marginal transformations are involved, the transformed conditional mean prediction is not equal to the conditional mean on the transformed scale. When the target is to minimize mean absolute errors, then conditional medians provide optimal predictions. With Gaussian data, conditional means and medians coincide. To compute conditional medians, we can simply transform data to the Gaussian scale, predict on the Gaussian scale, and then transform back to the original data scale. In competition 2, the target score was MSE. Due to very small prediction variances on the Gaussian scale, we found only very small differences between conditional median and conditional mean predictions on the non-Gaussian marginal scale of the original data. In some approaches (e.g., the bootstrap approach), we have therefore submitted marginally transformed Gaussian Kriging predictions.

4. FUTURE DIRECTIONS

This competition has explored different methods for the estimation of the parameters, and for the prediction (Kriging), of Gaussian and Tukey g and h trans-Gaussian random fields. Among these methods, several have achieved very good performance, as shown in Huang et al. (2021). More challenging setups than classical point data could also be considered, such as preferential sampling or the addition of location errors. An interesting question is whether the methods described in this paper would be efficient on gridded data, or whether grid-specific approaches would perform better. In particular, the case of gap filling (large areas without observations) could also be investigated.

Future competitions could also consider more challenging types of trans-Gaussian models arising in the analysis of agricultural, biological and environmental data, such as zero-inflated data, count data and other discrete data, or compositional data. Non-Gaussian margins arising from non (trans-)Gaussian random fields such as skew-Gaussian, skew-elliptical or max-stable random fields could also be considered.

Another very interesting extension would be to increase the dimensionality of the data by considering multivariate random fields, spatiotemporal random fields, or both. In different communities, such as machine learning, sensitivity analysis and uncertainty quantification, Gaussian processes are used in high-dimensional spaces. Testing the methods that have been developed successfully in spatial statistics to the problems faced by these communities offers interesting perspectives.

To provide a more realistic setting of real data analyses, future comparisons of prediction methods could also explore nonstationary settings with trend functions that may depend on external predictors, and with nonstationary covariance functions.

ACKNOWLEDGEMENTS

We would like to warmly thank several colleagues who have helped to improve and implement our prediction approach through discussions about methods and through technical support: N. Desassis, X. Freulon, L. Houde, M. Pereira and D. Renard.

[Received June 2021. Revised June 2021. Accepted July 2021.]

REFERENCES

Guinness J (2018) Permutation and grouping methods for sharpening Gaussian process approximations. Technometrics 60(4):415–429

Guinness J (2021) Gaussian process learning via Fisher scoring of Vecchia's approximation. Statist Comput 31(3):1–8

Huang H, Abdulah S, Sun Y, Ltaief H, Keyes D, Genton MG (2021) Competition on spatial statistics for large datasets. J Agric Biol Environ Stat

Jorge M, Boris I (1984) Some properties of the Tukey g and h family of distributions. Commun Stat Theory Methods 13(3):353–369

Katzfuss M, Guinness J (2021) A general framework for Vecchia approximations of Gaussian processes. Stat Sci 36(1):124–141

Xu G, Genton MG (2017) Tukey g-and-h random fields. J Am Stat Assoc 112(519):1236–1249

Zhang H (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J Am Stat Assoc 99(465):250–261

RÉSUMÉ

L'introduction de modèles statistiques inspirés de la physique des phénomènes sous-jacents et numériquement efficaces est d'un intérêt croissant pour la prédiction de processus spatio-temporels en sciences environnementales. Les grands jeux de données spatio-temporelles nécessitent de nouvelles méthodes numériques efficaces. L'approche par Equations aux Dérivées Partielles Stochastiques (EDPS) s'est avérée efficace pour l'estimation et la prédiction dans un contexte spatial. Nous présentons ici une EDPS d'advection-diffusion avec une dérivée de premier ordre en temps qui définit une grande classe de modèles spatio-temporels non séparables. On construit une approximation de la solution de l'EDPS par un champ aléatoire Markovien Gaussien en discrétisant la dérivée temporelle par la méthode des différences finies (Euler implicite) et en résolvant l'EDPS spatiale par la méthode des éléments finis (Galerkin continu) à chaque pas de temps. La technique de stabilisation "Streamline Diffusion" est introduite lorsque le terme d'advection domine la diffusion. Des méthodes de calcul efficaces sont proposées pour estimer les paramètres de l'EDPS et pour prédire le champ spatio-temporel par krigeage, ainsi que pour effectuer des simulations conditionnelles. L'approche est appliquée à des jeux de données de rayonnement solaire et de vitesse du vent. Ses avantages et ses limites sont examinées, et de nouvelles perspectives de travail sont envisagées, notamment afin de proposer une extension dans un cadre non stationnaire. On présente également un travail portant sur la généralisation non séparable de la classe de Gneiting des fonctions de covariance spatio-temporelles multivariées. Le principal potentiel de l'approche est la possibilité d'obtenir des modèles entièrement non séparables dans un cadre multivarié, et les avantages sont illustrés sur un ensemble de données météorologiques trivariées. De plus, on propose une analyse de méthodes d'estimation et de prédiction approximées pour les données spatiales et spatio-temporelles, motivée par l'objectif de parvenir à un compromis entre l'efficacité statistique et la complexité computationnelle. Ces méthodes se sont avérées efficaces pour l'estimation des paramètres et la prédiction dans le contexte de la "Compétition de statistiques spatiales pour les grands jeux de données" organisée par la King Abdullah University of Science and Technology (KAUST) en 2021 et 2022. Enfin, d'autres pistes de recherche sont envisagées et examinées.

MOTS CLÉS

Géostatistique; Modélisation spatio-temporelle; Analyse numérique; SPDE; Statistique environmentale.

ABSTRACT

In the task of predicting spatio-temporal fields in environmental science using statistical methods, introducing statistical models inspired by the physics of the underlying phenomena that are numerically efficient is of growing interest. Large space-time datasets call for new numerical methods to efficiently process them. The Stochastic Partial Differential Equation (SPDE) approach has proven to be effective for the estimation and the prediction in a spatial context. We present here the unsteady advection-diffusion SPDE which defines a large class of nonseparable spatio-temporal models. A Gaussian Markov random field approximation of the solution to the SPDE is built by discretizing the temporal derivative with a finite difference method (implicit Euler) and by solving the spatial SPDE with a finite element method (continuous Galerkin) at each time step. The Streamline Diffusion stabilization technique is introduced when the advection term dominates the diffusion. Computationally efficient methods are proposed to estimate the parameters of the SPDE and to predict the spatio-temporal field by kriging, as well as to perform conditional simulations. The approach is applied to solar radiation and wind speed datasets. Its advantages and limitations are discussed, and new perspectives for future work are envisaged, especially involving a nonstationary extension of the approach. As a further contribution of the PhD, the nonseparable generalization of the Gneiting class of multivariate space-time covariance functions is presented. The main potential of the approach is the possibility to obtain entirely nonseparable models in a multivariate setting, and this advantage is shown on a weather trivariate dataset. Finally, a review of some methods for approximate estimation and prediction for spatial and spatio-temporal data is proposed, motivated by the objective of reaching a trade-off between statistical efficiency and computational complexity. These methods proved to be effective for parameter estimation and prediction in the context of the "Spatial Statistics Competition for Large Datasets" organized by the King Abdullah University of Science and Technology (KAUST) in 2021 and 2022. Lastly, possible further research directions are discussed.

KEYWORDS

Geostatistics; Spatio-temporal modeling; Numerical analysis; SPDE; Environmental statistics.