Non-separable Spatio-temporal Graph Kernels via SPDEs


Alexander Nikitin


ST John


Arno Solin


Samuel Kaski

FCAI Finnish Center for Artificial Intelligence

Aalto University

MANCHESTER 1824 The University of Manchester

TL;DR

- We extend methods from spatial statistics and, hence, link GPs to SPDEs for temporal signals on graphs
- Using the analogues of well-known SPDEs from spatial statistics, we derive non-separable spatio-temporal kernels on graphs
- We show effectiveness on synthetic data sets and applied machine learning problems: prediction of the distribution of chickenpox and the COVID-19 epidemic

Introduction

Gaussian processes are a non-parametric machine learning paradigm, in which we model the target function as a stochastic process, whose evaluation at any finite set of points has a joint Gaussian distribution. A Gaussian process $f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$ is defined by its mean function $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$.

The kernel encapsulates prior knowledge, and defining a good kernel is one of the key ingredients and challenges of setting up the GP model. GPs can be extended to vector-valued functions using multioutput GPs.

Problem. Standard GP toolchains include various kernels on continuous domains. However, the application of GPs to other domains is often restricted by the unavailability of principled kernels. We thus provide tools for spatio-temporal graph functions.

Methods

Framework: SPDE → graph kernel

i Define an SPDE, using prior knowledge about the underlying process
ii Convert the continuous SPDE to a graph counterpart
iii Solve the graph counterpart
iv Derive corresponding mean and covariance function of GP on graph

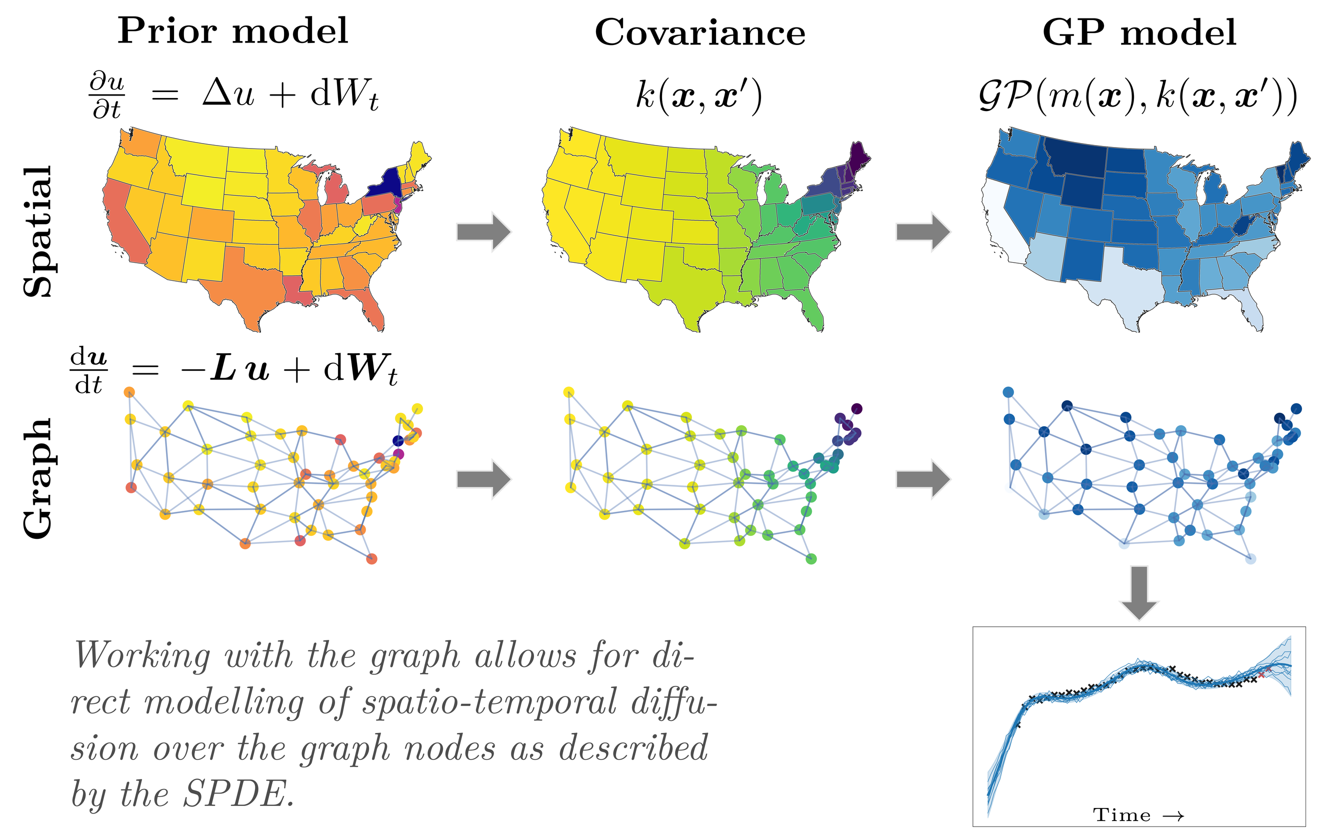


Fig. 1: Illustration of the proposed approach

Stochastic Heat Equation Kernel (SHEK)

The stochastic heat equation kernel (SHEK) on graphs can be defined by adding spatio-temporal white noise, or for convenient integration, as a *formal* differential of the Wiener process $\dot{\boldsymbol{W}}_t$:

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = -c\widetilde{\boldsymbol{L}}\boldsymbol{u} + \sigma\dot{\boldsymbol{W}}_t.$$

The solution is given by a Gaussian process:

$$\boldsymbol{u}(t) \sim \mathcal{GP}(\boldsymbol{\mu}(t), \mathrm{Cov}[\boldsymbol{u}(s), \boldsymbol{u}(t)]), \quad \text{with } \boldsymbol{\mu}(t) = e^{-c\widetilde{\boldsymbol{L}}t}\boldsymbol{u}(0),$$

$$\mathrm{Cov}[\boldsymbol{u}(t), \boldsymbol{u}(s)] = \frac{\sigma^2}{c} e^{-c\widetilde{\boldsymbol{L}}t - c\widetilde{\boldsymbol{L}}^\top s}(e^{c(\widetilde{\boldsymbol{L}}+\widetilde{\boldsymbol{L}}^\top)\min(t,s)} - \boldsymbol{I})(\widetilde{\boldsymbol{L}} + \widetilde{\boldsymbol{L}}^\top)^{-1}.$$

Or, when the matrix $\widetilde{\boldsymbol{L}}$ is self-adjoint (the graph is undirected), as

$$\boldsymbol{\mu}(t) = e^{-c\widetilde{\boldsymbol{L}}t}\boldsymbol{u}(0), \mathrm{Cov}[\boldsymbol{u}(t), \boldsymbol{u}(s)] = \frac{\sigma^2}{2c}\left(e^{-c\widetilde{\boldsymbol{L}}|t-s|} - e^{-c\widetilde{\boldsymbol{L}}(t+s)}\right)\widetilde{\boldsymbol{L}}^{-1}.$$

The kernel is parameterized by diffusivity c, noise scale σ, and parameters of the fractional Laplacian ν and κ. For matrix-valued white noise:

$$\mathrm{Cov}\left[\boldsymbol{u}(t), \boldsymbol{u}(s)\right] = \boldsymbol{P}^*\boldsymbol{C}(t,s)\boldsymbol{P},$$

where $\boldsymbol{P}$ is a unitary matrix: $\boldsymbol{P}\widetilde{\boldsymbol{L}}\boldsymbol{P}^* = \mathrm{diag}(\lambda_1, \ldots, \lambda_{|V|})$. The matrix $\boldsymbol{P}$ exists because $\widetilde{\boldsymbol{L}}$ is normal and positive definite. $\boldsymbol{C}(t,s)$ is defined for $t \geq s$ as:

$$\boldsymbol{C}(t,s)_{i,j} = \frac{1}{c}\frac{(\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top\boldsymbol{P}^*)_{i,j}}{\lambda_i + \lambda_j}(\exp(-c\lambda_i|t-s|) - \exp(-c(\lambda_i t + \lambda_j s))).$$

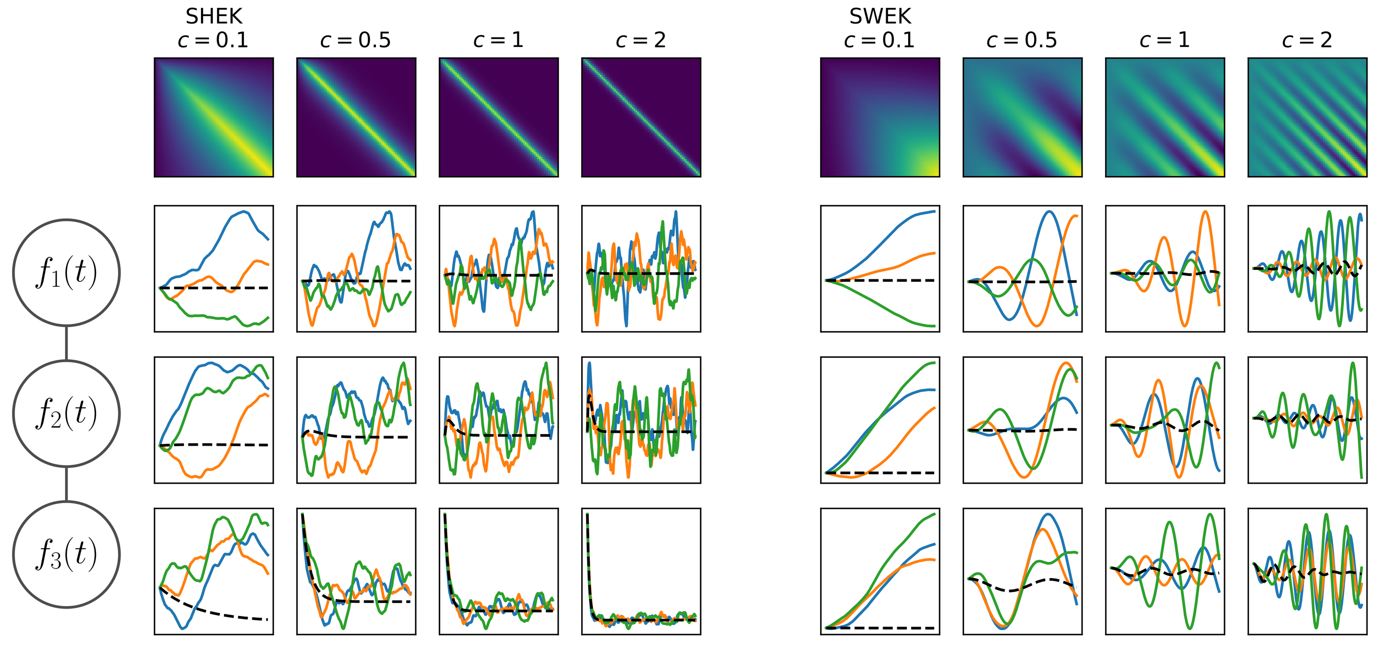


Fig. 2: SHEK and SWEK on a three-vertex graph

Stochastic Wave Equation Kernel (SWEK)

The stochastic wave equation kernel (SWEK) on undirected graphs is defined by the second-order matrix differential equation

$$\frac{\mathrm{d}^2\boldsymbol{u}}{\mathrm{d}t^2} = -c^2\widetilde{\boldsymbol{L}}\boldsymbol{u} + \sigma\dot{\boldsymbol{W}}_t,$$

and a solution to this equation for undirected graphs can be expressed by the Gaussian process:

$$\boldsymbol{u}(t) \sim \mathcal{GP}(\boldsymbol{\mu}, \mathrm{Cov}[\boldsymbol{u}(s), \boldsymbol{u}(t)], \quad \text{with}$$

$$\boldsymbol{\mu}(t) = \frac{1}{c}\widetilde{\boldsymbol{L}}^{-\frac{1}{2}}\sin(c\sqrt{\widetilde{\boldsymbol{L}}}t)\boldsymbol{P}\dot{\boldsymbol{u}}(0) + \cos(c\sqrt{\widetilde{\boldsymbol{L}}}t)\boldsymbol{P}\boldsymbol{u}(0),$$

$$\mathrm{Cov}[\boldsymbol{u}(s), \boldsymbol{u}(t)] = \sigma^2\boldsymbol{\Theta}^{-2}\Big(\cos(\boldsymbol{\Theta}(t-s))\min(t,s) - \frac{1}{2}\cos(\boldsymbol{\Theta}\max(t,s))\sin(\boldsymbol{\Theta}\min(t,s))\boldsymbol{\Theta}^{-1}\Big),$$

where $\boldsymbol{\Theta} = c\sqrt{\widetilde{\boldsymbol{L}}}$ and $\boldsymbol{P}$ is defined by the diagonalization of the fractional Laplacian matrix: $\widetilde{\boldsymbol{L}} = \boldsymbol{P}^{-1}\widetilde{\boldsymbol{L}}_d\boldsymbol{P}$.

Experiments

Tasks:

i interpolation of a spatiotemporal graph signal
ii extrapolation of a spatiotemporal graph signal

Domains:

i heat and wave distribution over a one-dimensional line
ii spreading of COVID-19 cases across the United States
iii spreading of chickenpox cases over Hungarian counties

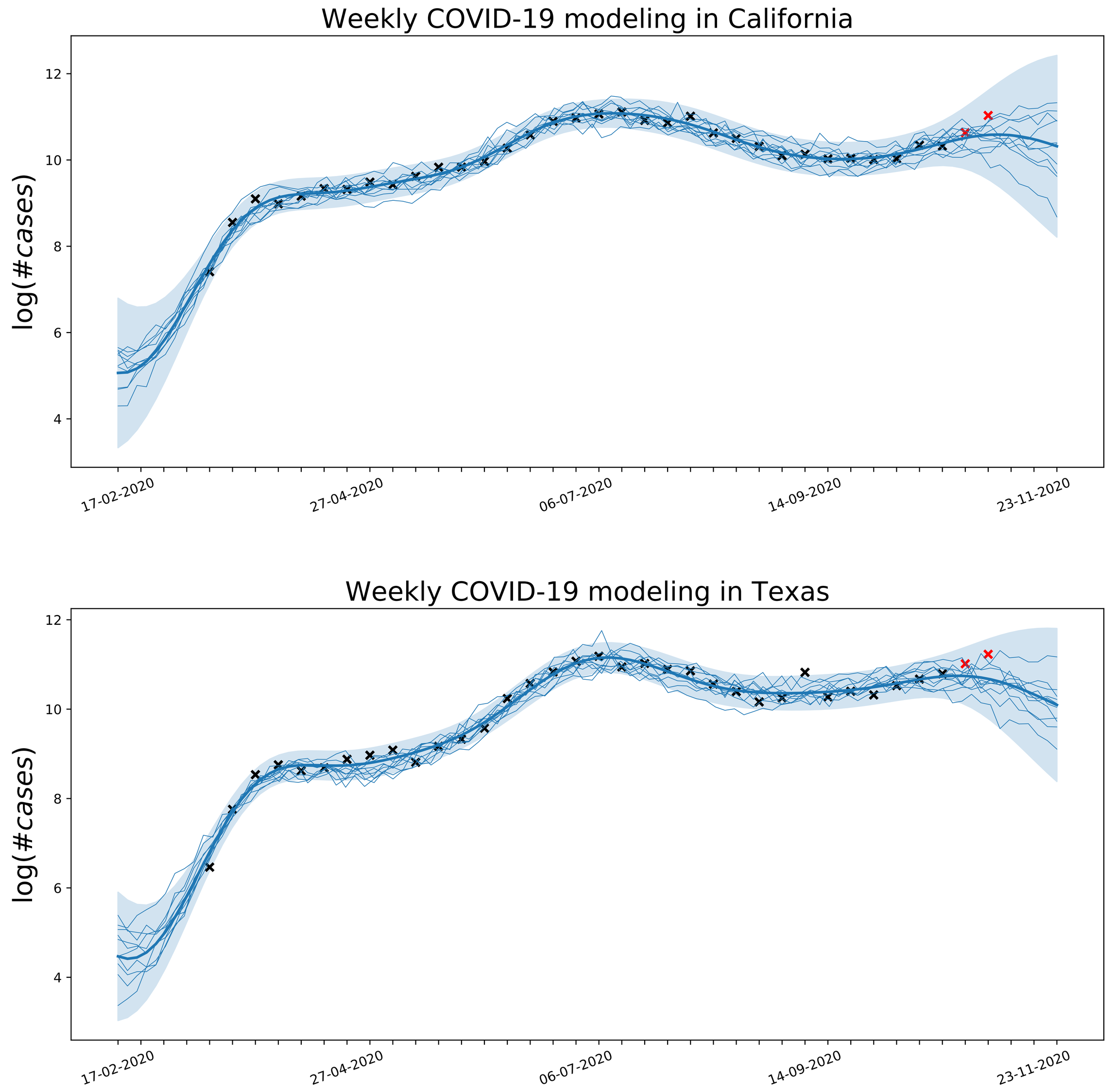
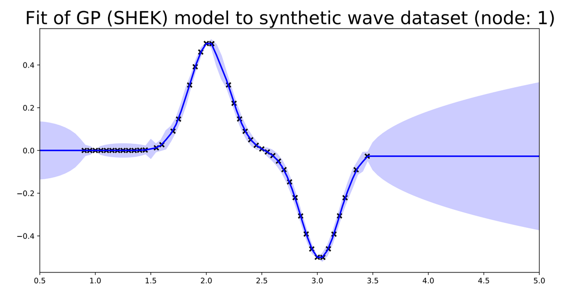
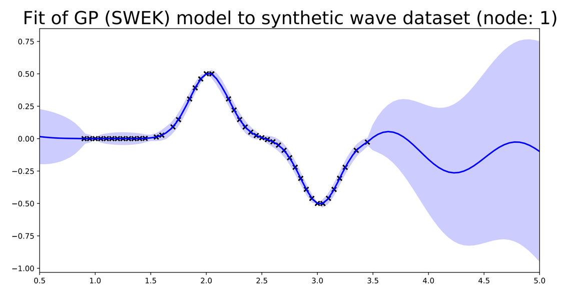


Fig. 4: SHEK fit to COVID-19 data for two states.



SHEK on wave dataset (node #1).



SWEK on wave dataset (node #1).

Fig. 5: SHEK and SWEK fit to synthetic wave data set.

Bibliography

Viacheslav Borovitskiy, Iskander Azangulov, Alexander Terenin, Peter Mostowsky, Marc Deisenroth, and Nicolas Durrande. Matérn Gaussian processes on graphs. In *International Conference on Artificial Intelligence and Statistics*, pages 2593–2601. PMLR, 2021.

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

CE. Rasmussen and CKI. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA, January 2006.

Simo Särkkä and Arno Solin. *Applied Stochastic Differential Equations*. Cambridge University Press, 2019.

Alexander J Smola and Risi Kondor. Kernels and regularization on graphs. In *Learning Theory and Kernel Machines*, pages 144–158. Springer, 2003.

Arno Solin. *Stochastic Differential Equation Methods for Spatio-Temporal Gaussian Process Regression*. PhD thesis, Aalto University, 2016.

alexander.nikitin@aalto.fi