Aalto University
School of Science

Dept. of Biomedical Engineering
and Computational Science

SITraN, University of Sheffield

Spatio-Temporal Gaussian Process Modeling

Arno Solin

Department of Biomedical Engineering and Computational Science
Aalto University, School of Science
arno.solin@aalto.fi

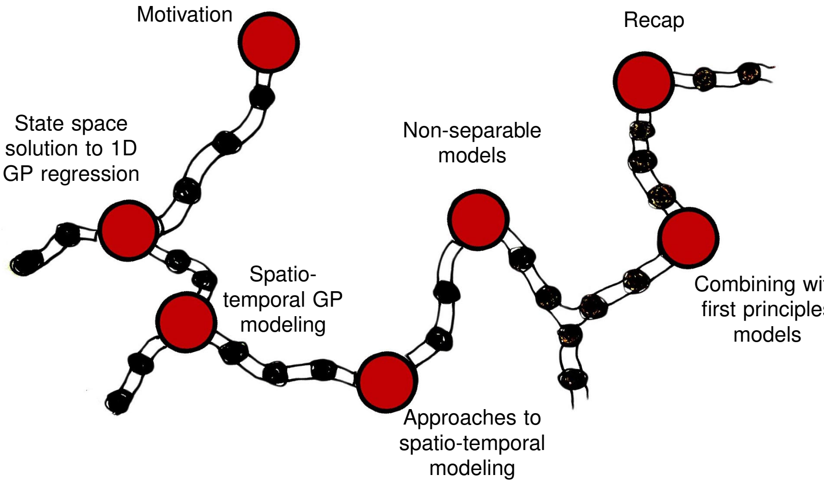
Nov 19, 2013

Introduction

Backround:

- Who?
  Arno Solin,
  second-year PhD student
- From?
  Bayesian Statistical Methods
  Group, Aalto University
- With whom?
  Instructor Dr. **Simo Särkkä**,
  visiting Prof. **Neil D. Lawrence**
- What?
  Signal processing,
  Gaussian processes

Presentation Outline



Motivation

GP regression

- Gaussian processes are elegant tools for modeling.
- Computational cost scales as $\mathcal{O}(n^3)$ in the number of observations.
- The direct GP methodology is not well suited for long or unbounded time series.
- The same applies to spatio-temporal models (typically a lot of observations in the temporal dimension).

Why the state space approach?

- Certain classes of GPs can be converted to equivalent(ish) state space models.
- Well suited for long or unbounded time-series.
- $\mathcal{O}(n)$ scaling in the number of temporal observations.
- In spatio-temporal models, the GP can be rewritten in terms of an infinite-dimensional state space model, where a spatial GP evolves in time.

Gaussian Process (GP) Regression in 1D

- Consider the GP regression model

$$f(t) \sim \mathcal{GP}(0, k(t,t'))$$
$$y_k = f(t_k) + \varepsilon_k,$$

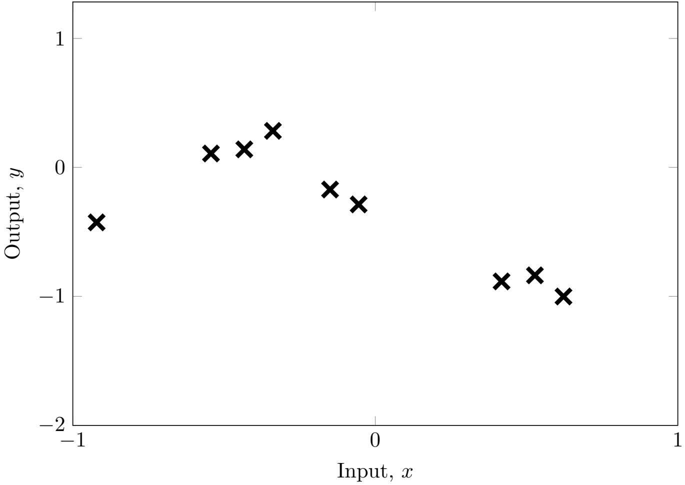
where $\varepsilon_k \sim \mathcal{N}(0, \sigma_\mathrm{n}^2)$.

- The direct solution to the GP regression problem gives predictions $p(f_* \mid t_*, \mathcal{D}) = \mathcal{N}(\mathbb{E}[f(t_*)], \mathbb{V}[f(t_*)])$:

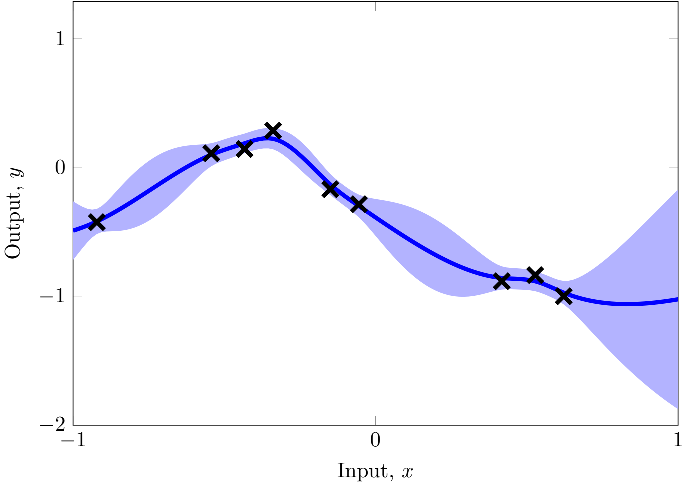
$$\mathbb{E}[f(t_*)] = \mathbf{k}_*^\mathsf{T}(\mathbf{K} + \sigma_\mathrm{n}^2\mathbf{I})^{-1}\mathbf{y},$$
$$\mathbb{V}[f(t_*)] = k(t_*, t_*) - \mathbf{k}_*^\mathsf{T}(\mathbf{K} + \sigma_\mathrm{n}^2\mathbf{I})^{-1}\mathbf{k}_*,$$

where $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{k}_*$ is an n-dimensional vector with the ith entry being $k(\mathbf{x}_*, \mathbf{x}_i)$, and $\mathbf{y}$ is a vector of the n observations.

Matérn prior covariance ($\nu = 3/2, \ell = 1, \sigma^2 = 1$)



Matérn prior covariance ($\nu = 3/2, \ell = 1, \sigma^2 = 1$)



State Space Solution to GP Regression

- For temporal GPs, we can work with the mathematical dual of the kernel formalism: the state space model.
- Certain classes of GPs can be written in state space form as stochastic differential equations:

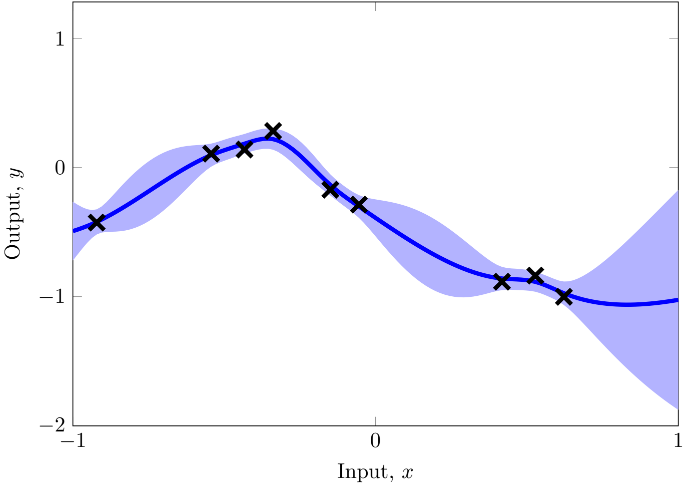
$$\frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} = \mathbf{F}\,\mathbf{f}(t) + \mathbf{L}\mathbf{w}(t)$$
$$y_k = \mathbf{H}\,\mathbf{f}(t_k) + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, \sigma_\mathrm{n}^2),$$

where $\mathbf{f}(t) = \big(f_1(t), f_2(t), \ldots, f_m(t)\big)^\mathsf{T}$ holds the m stochastic processes, and $\mathbf{w}(t)$ is a multi-dimensional white noise process with spectral density $\mathbf{Q}_\mathrm{c}$. The model is defined by the feedback matrix $\mathbf{F}$ and the noise effect matrix $\mathbf{L}$.

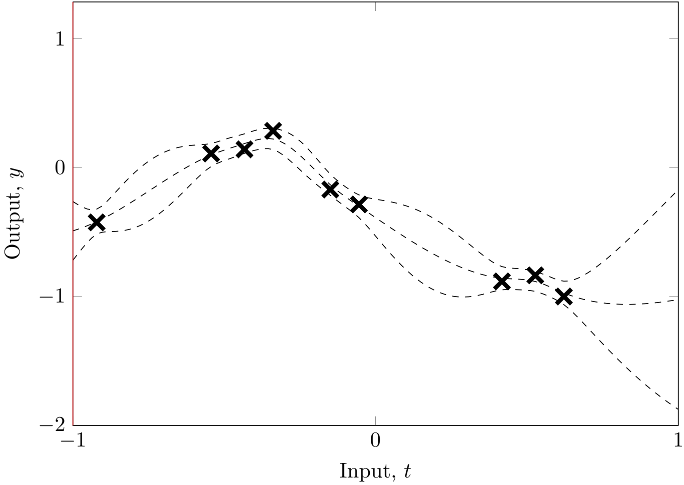
State Space Solution to GP Regression

- The model matrices $\mathbf{F}$, $\mathbf{L}$, $\mathbf{Q}_c$, $\mathbf{H}$ and the stationary covariance of the state $\mathbf{P}_\infty$ all depend on the prior covariance function.
- The marginal likelihood and the predictions given by the GP model can be calculated in closed-form using the Kalman filter and Rauch–Tung–Striebel smoother.
- The use of the filter/smoother makes the computational complexity linear with respect to the number of time steps.

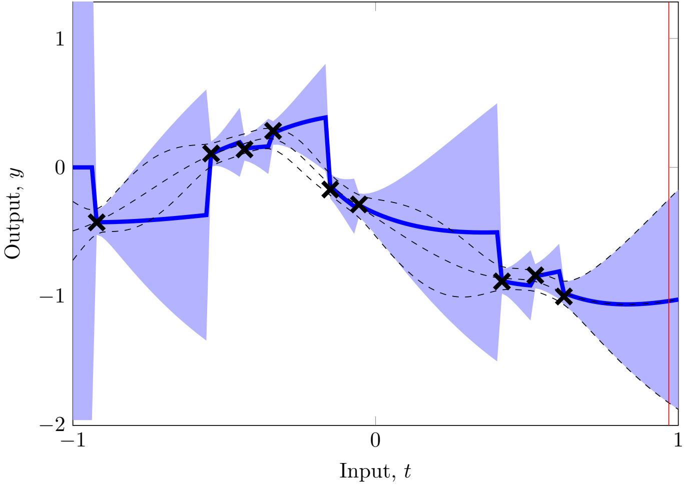
Matérn prior covariance ($\nu = 3/2, \ell = 1, \sigma^2 = 1$)



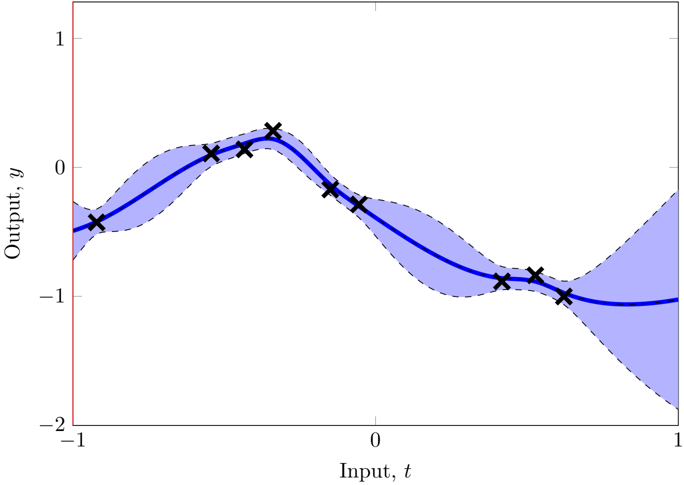
Filter Estimate



Smoother Estimate



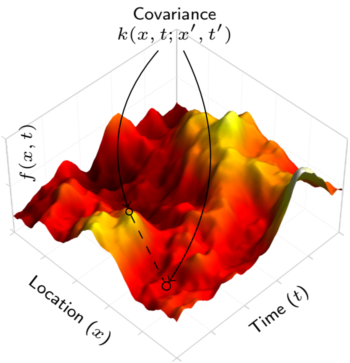
Smoother Estimate

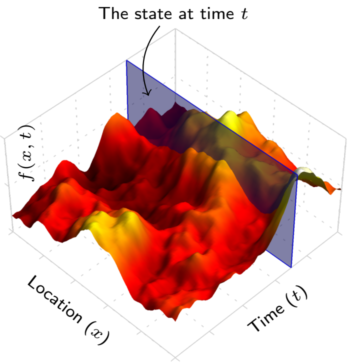


Spatio-Temporal GP Modeling

- A spatio-temporal phenomenon is subject to variability in both space and time.
- We characterize such phenomena as $\mathbf{f}(\mathbf{x}, t) : \Omega \times \mathbb{R}_+ \to \mathbb{R}^n$.
- Often the physical phenomenon provides some prior knowledge of spatial and temporal structures that we wish to include in the model
- ... and often the observations are subject to noise.

Spatio-Temporal GP Modeling





$$f(\mathbf{x}, t) \sim \mathcal{GP}(0, k(\mathbf{x}, t; \mathbf{x}', t'))$$
$$y_i = f(\mathbf{x}_i, t_i) + r_i$$

$$\frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial t} = \mathcal{F}\, \mathbf{f}(\mathbf{x}, t) + \mathcal{L}\, w(x, t)$$
$$y_i = \mathcal{H}_i\, f(\mathbf{x}, t_i) + r_i$$

Approaches to Spatio-Temporal GP Modeling

Model each spatial time series independently

Consider the process $f(\mathbf{x}, t)$ to be a set of independent GPs for each spatial location $\mathbf{x}_i$.

Separable covariance functions in space and time

Consider the covariance function to be separable such that $k(\mathbf{x}, t; \mathbf{x}', t') = k_1(\mathbf{x}, \mathbf{x}')\, k_2(t, t')$.

Non-separable covariance functions

The covariance function $k(\mathbf{x}, t; \mathbf{x}', t')$ cannot be written as above.

Extensions:

Non-stationary covariance functions, combining with first principles models, *et cetera*.

Non-Separable Spatio-Temporal GP Regression

- Consider the Matérn covariance function ($r = \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|$, $\boldsymbol{\xi} = (x_1, x_2, \ldots, x_{d-1}, t)$)

$$k(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{r}{\ell}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\frac{r}{\ell}\right),$$

  where σ^2 is the magnitude scale, ℓ the length-scale, and ν the smoothness parameter.
- This covariance function is not separable in $\mathbf{x}$ and t.

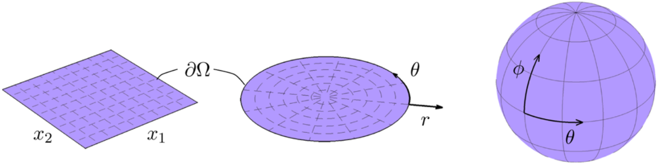
Non-Separable Spatio-Temporal GP Regression

- Similarly as in the 1D case, we write the corresponding state space form.
- It is a stochastic partial differential equation, and if $d = 2$ and $\nu = 1$ we get

$$\frac{\partial \mathbf{f}(x,t)}{\partial t} = \begin{pmatrix} 0 & 1 \\ \nabla^2 - \lambda^2 & -2\sqrt{\lambda^2 - \nabla^2} \end{pmatrix} \mathbf{f}(x,t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(x,t)$$

  where $\lambda = \sqrt{2\nu}/\ell$.
- For details, see Särkkä, Solin, Hartikainen (2013).

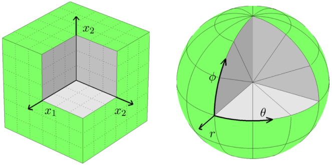
How to deal with the remaining spatial operators?



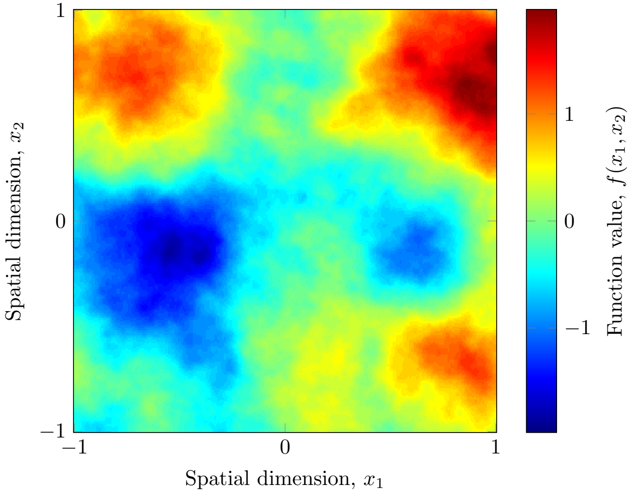
The eigendecomposition of the negative Laplace operator

$$-\nabla^2 \phi_n(\mathbf{x}) = \lambda_n \phi_n(\mathbf{x})$$

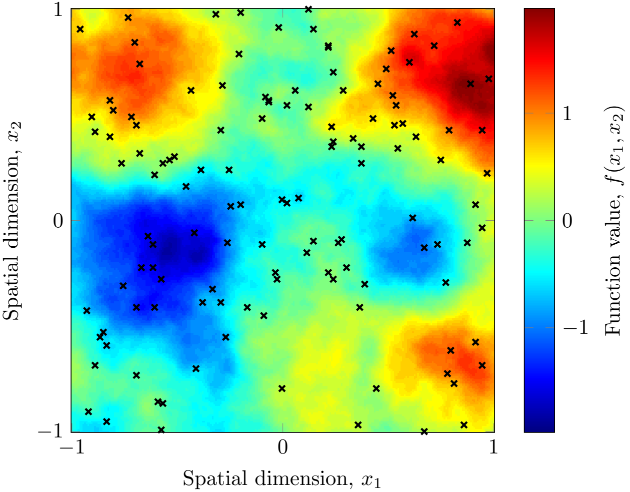
can be solved in closed form in various domains.



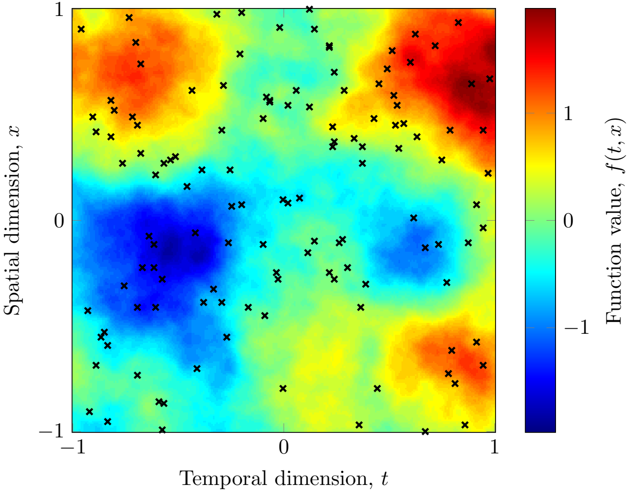
Spatial Matérn Field ($\nu = 1, \ell = 1, \sigma^2 = 1$)



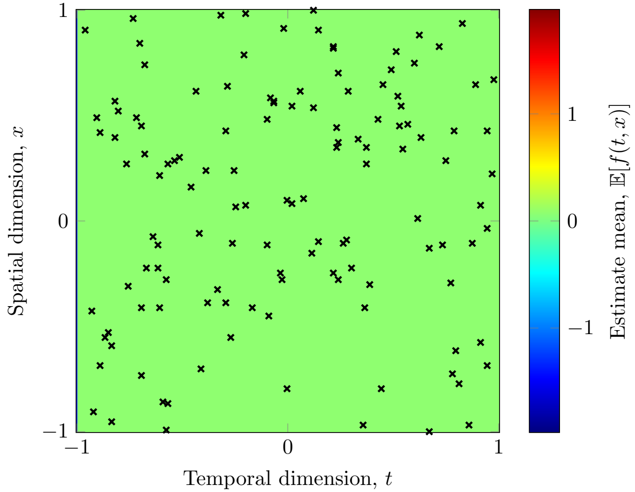
Spatial Matérn Field ($\nu = 1, \ell = 1, \sigma^2 = 1$)



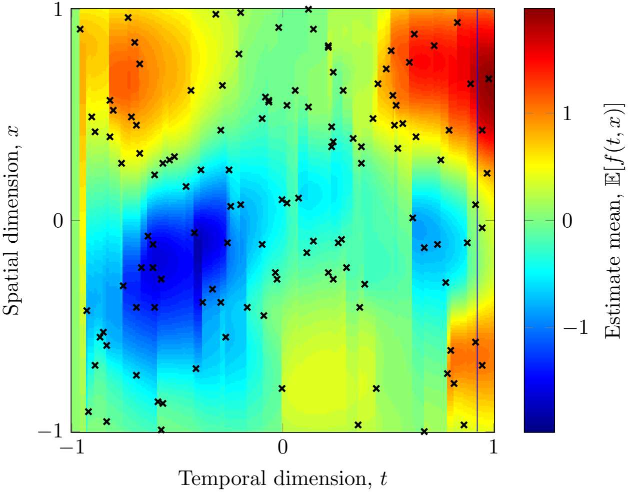
Spatio-Temporal Matérn Field



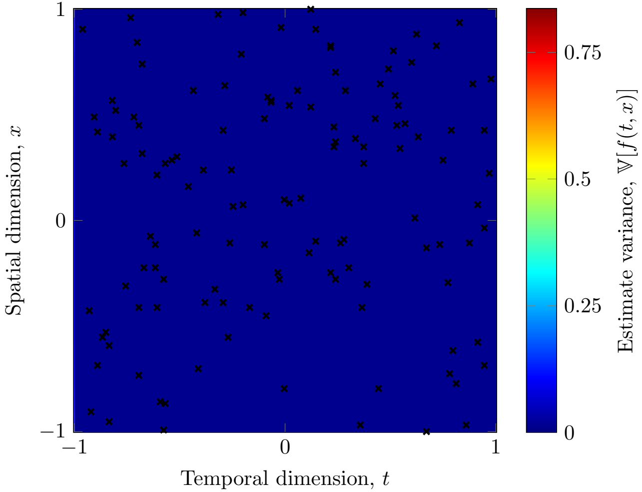
Filter Mean Estimate



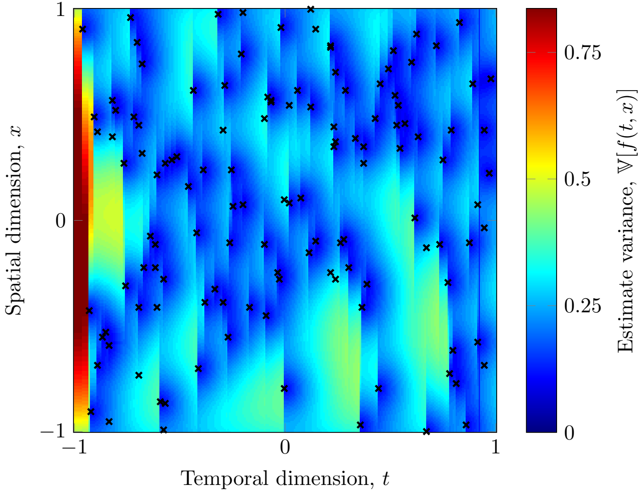
Smoother Mean Estimate



Filter Uncertainty Estimate



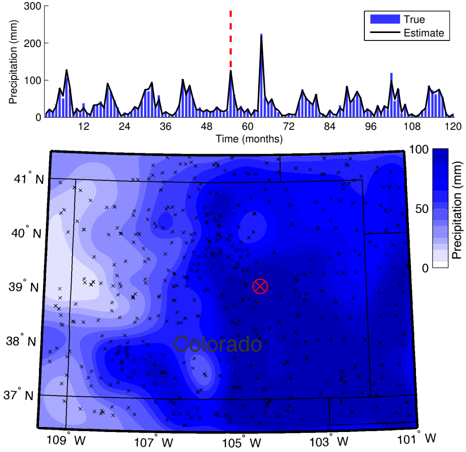
Smoother Uncertainty Estimate



Modeling of Precipitation in Colorado

- We consider spatio-temporal ($\mathbf{x} \in \mathbb{R}^2, t \in \mathbb{R}_+$) interpolation of monthly precipitation levels using a GP model.
- We use a non-separable Matérn covariance ($\nu = 3/2$, separate length-scale for the temporal dimension).
- Data collected over ten years (120 months) at about 400 stations around Colorado, US (55,410 observations altogether).
- The filters and smoothers were approximated by truncated eigenfunction expansion of the Laplace operator with 384 eigenfunctions.

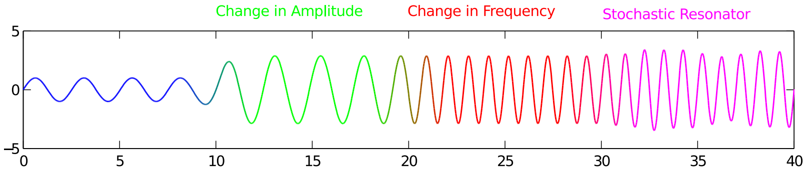
Modeling of Precipitation in Colorado



Combining First Principles Models with GPs

- In many cases a mathematical model in terms of differential equations exists or can be written out for modeling a phenomenon.
- Combining the infinite-dimensional state space formalism with such models is straight forward(ish).
- This provides a framework for combining GP models with stochastic partial differential equation models.

Modeling Spatio-Temporal Oscillations



- The dynamics are presented as a superposition, $f(\mathbf{x}, t) = \sum_j f_j(\mathbf{x}, t)$, of several resonators
- with **known frequencies**
- but **unknown phases and amplitudes**.
- We construct a stochastic model of form

$$\frac{\partial^2 f_j(\mathbf{x}, t)}{\partial t^2} + \mathcal{A}_j \frac{\partial f_j(\mathbf{x}, t)}{\partial t} + \mathcal{B}_j f_j(\mathbf{x}, t) = \xi_j(\mathbf{x}, t)$$

with arbitrary linear operators $\mathcal{A}_j$ and $\mathcal{B}_j$ affecting both the oscillation and damping.

- We consider the following operator structure

$$\begin{aligned}
\mathcal{A}_j &= \gamma_j \mathcal{I} - \chi_j \nabla^2 \\
\mathcal{B}_j &= \frac{\gamma_j^2}{2} - \gamma_j \chi_j \nabla^2 + \frac{\chi_j^2}{2} \nabla^4 + \omega_j^2 \\
&= \frac{1}{2}(\gamma_j - \chi_j \nabla^2)^2 + \omega_j^2,
\end{aligned}$$

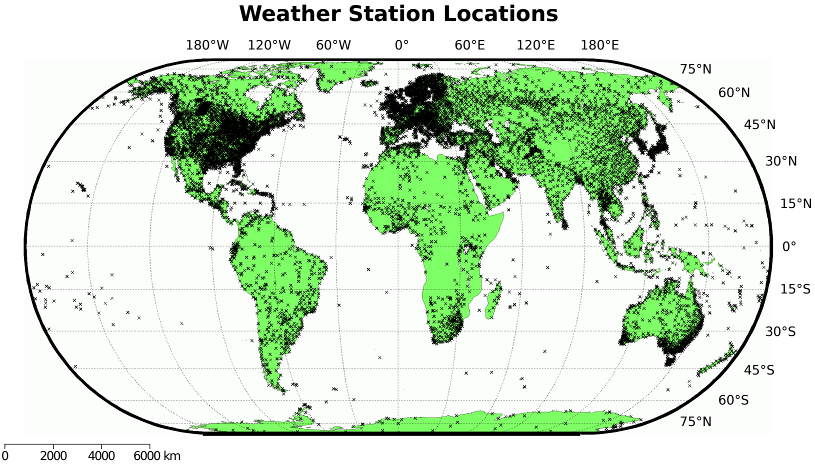
where $\gamma_j, \chi_j \geq 0$ are some non-negative constants.

- The perturbation term $\xi_j(\mathbf{x}, t)$ is a GP.

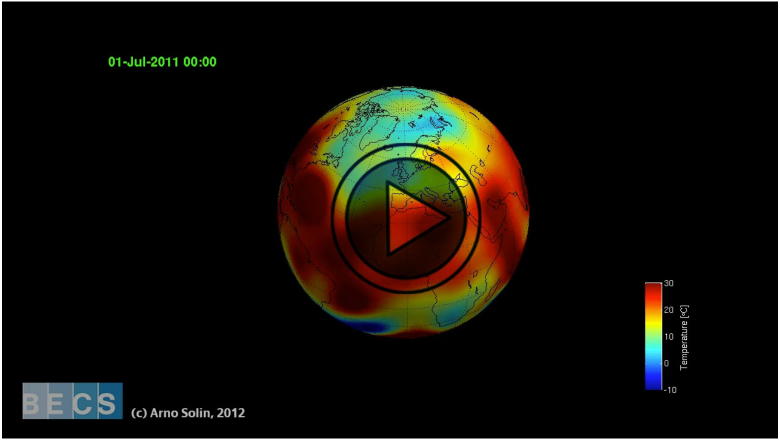
Modeling the Daily Temperature Variation

- We apply the spatio-temporal resonator model to temperature data for July 2011:
  - 11 344 locations
  - readings once an hour
  - 5 637 501 measurements
- Resonator model with frequency 1/day (+ one harmonic).
- We include a slowly moving bias to account for the daily mean temperature.
- The perturbation term has a Matérn covariance function with $\nu = 3/2$.

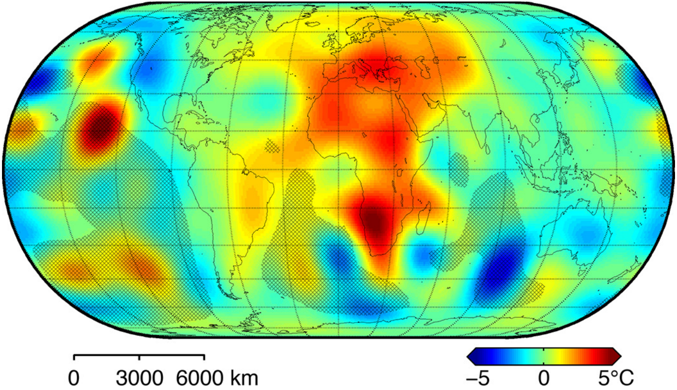
Modeling Daily Temperature Variation



Modeling Daily Temperature Variation



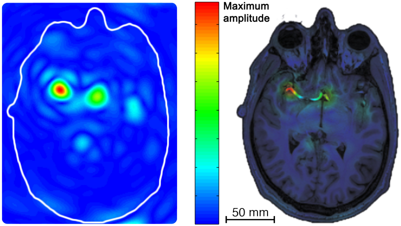
Modeling Daily Temperature Variation



Temperature deviations from the slow moving mean temperature on July 8, 2011 at noon (GMT). The hatched areas are regions of high uncertainty.

Non-Stationary Modeling of Physiological Noise

- We model the periodic cardiac- and respiration-induced noise in functional magnetic resonance imaging brain data.
- We apply the model to fMRI data, with matrix size 64×64, repetition time 0.1 s, and total length 26 s (1,064,960 observations).
- Resonator model with time-dependent frequencies from reference signals.



The average amplitude of the heart beat induced noise.

Recap

- Certain classes of one-dimensional GP regression problems can be written as state space models.
- Spatio-temporal Gaussian processes can in many cases be converted into stochastic evolution equations (inf.-dim. state space models).
- The infinite-dimensional Kalman filter and RTS smoother can be used for efficient inference in spatio-temporal Gaussian process models.
- The infinite-dimensional evolution equations (SPDEs) can be combined with other models, which provides a framework for combining them with GPs.

References

- Särkkä, S., Solin, A. and Hartikainen, J. (2013). *Spatio-Temporal Learning via Infinite-Dimensional Bayesian Filtering and Smoothing*. IEEE Signal Processing Magazine. 30(4):51–61.
- Solin, A. and Särkkä, S. (2013). *Infinite-Dimensional Bayesian Filtering for Detection of Quasi-Periodic Phenomena in Spatio-Temporal Data*. Physical Review E, 88(5):052909.
- Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press.