Gaussian-process factor analysis for modeling spatio-temporal data

Jaakko Luttinen

Helsinki University of Technology, TKK
Department of Information and Computer Science

26.11.2009

Supervisor: Professor Erkki Oja
Instructor: PhD Alexander Ilin

Overview

Introduction

-

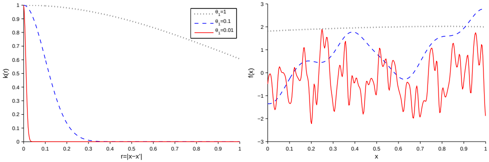
Gaussian processes

- Gaussian process is a distribution over functions.
- Any finite set of function values are multivariate normally distributed.
- The distribution

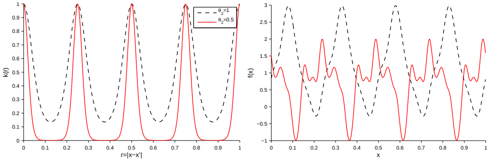
$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right)$$

- Covariance functions defines similarity betcontrol high-level properties, such as
- Computational cost scales cubically $O(N^3)$ with respect to the number N of observations.

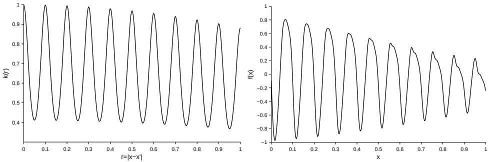
Gaussian processes – examples



(a) smooth functions



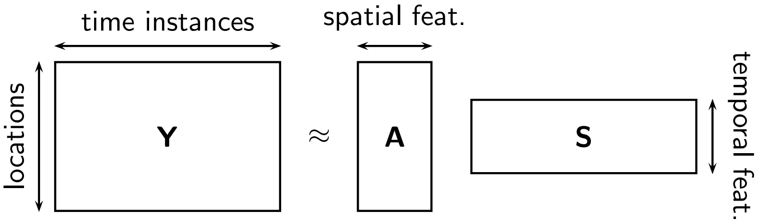
(b) periodic functions



(d) quasi-periodic functions

Gaussian-process factor analysis

- Model:



- The reconstruction error is modeled as Gaussian noise.

Gaussian-process factor analysis

- Data consists of observations at spatial locations $\{l_m\}_{m=1}^{M}$ at time instances $\{t_n\}_{n=1}^{N}$.
- Model for the observations

$$y(l_m, t_n) = \sum_{d=1}^{D} a_d(l_m) s_d(t_n) + \text{noise},$$

- Priors for the spatial and temporal feature functions:

$$a_d(l) \sim \mathcal{GP}\left(0, k_{a_d}(l, l')\right)$$
$$s_d(t) \sim \mathcal{GP}\left(0, k_{s_d}(t, t')\right)$$

- Covariance functions are chosen based on the prior knowledge.

Variational approximate inference

- True posterior $p(\mathbf{A}, \mathbf{S}|\mathbf{Y})$ is intractable.
- Approximate with a factorized distribution:

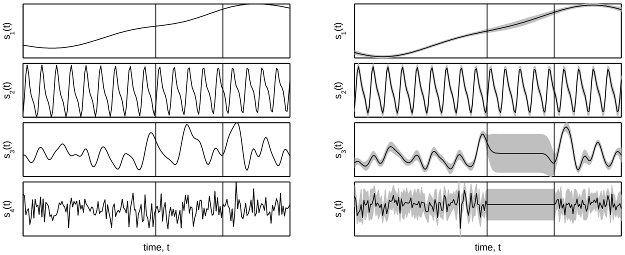
$$p(\mathbf{A}, \mathbf{S}|\mathbf{Y}) \approx q(\mathbf{A})q(\mathbf{S}).$$

- Optimize the approximation by minimizing the Kullback-Leibler divergence between the true and approximate distributions.
- In order to reduce computational cost, one can
  - factorize $q(\mathbf{A})$ and $q(\mathbf{S})$ with respect to the components.
  - use sparse approximations for the components.

Artificial experiment

- Generated data by using the presented model with $D = 4$ latent components.
- The four components had different characteristics.
- $M = 30$ spatial locations.
- $N = 200$ time instances.
- 90% of the data was discarded, resulting in approximately 450 noisy observations for training.

Artificial experiment – temporal components

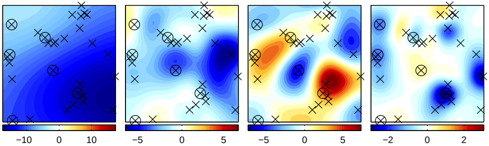


(a) (b)

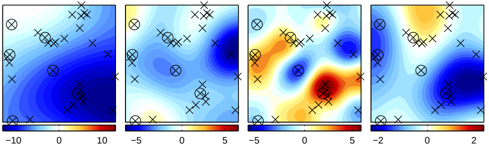
- (a) The true latent signals $s_d(t)$ used to generate the data.
- (b) The posteriors of the four latent signals $s_d(t)$.
- Vertical lines show a gap with no training observations.

Artificial experiment – spatial components

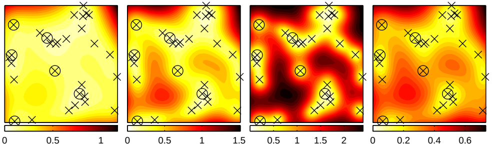
- The true spatial loadings $a_d(l)$ used to generate the data:



- The posterior means of the loadings:

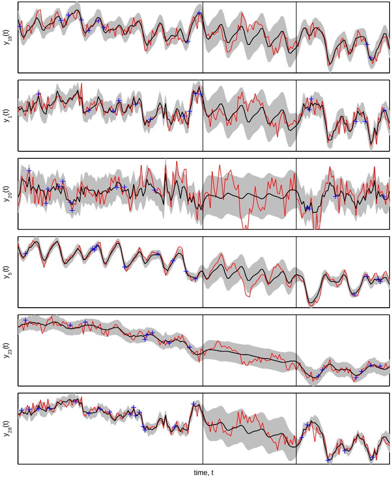


- The standard deviations computed from the posterior:



Artificial experiment – predictive distribution

- Posterior predictive distribution for six randomly selected locations.



Historical sea surface data

- Historical sea surface temperature dataset:
  - monthly temperature averages over 1856–1991
  - $5^\circ \times 5^\circ$ longitude-latitude bins
  - 55% of the values missing
- Estimated $D = 80$ components:
  - 5 very slow components
  - 5 smooth interannual components
  - 5 quasi-periodic components
  - 65 fast varying components

Historical sea surface data

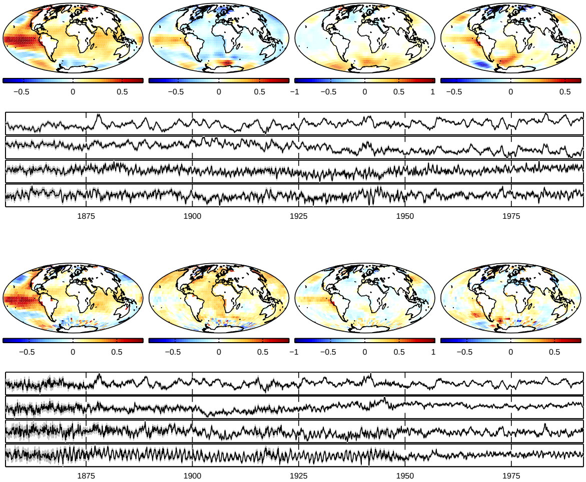


Figure: Experimental results for the MOHSST5 dataset. The spatial and temporal patterns of the four most dominating principal components for GPFA (above) and VB-PCA (below). The solid lines and gray color in the time series show the mean and two standard deviations of the

Conclusions

- A novel method for spatio-temporal modeling and exploratory analysis.
- Spatial and temporal structure modeled with Gaussian processes.
- Computational savings compared to standard GPs.