Applications in climate science

Alexander Ilin, Jaakko Luttinen, and Jarmo Saarimäki

We applied the Bayesian methodology for several problems in climate science.

In our papers [1, 3], we consider the problem of historical reconstruction of climate fields, which is a problem of infilling missing values in the observational data. We take the statistical approach and propose a probabilistic model called Gaussian-process factor analysis (GPFA). The model is based on standard matrix factorization

$$\mathbf{Y} = \mathbf{W}\mathbf{X} + \text{noise} = \sum_{d=1}^{D} \mathbf{w}_{:d}\mathbf{x}_{d:}^{\mathrm{T}} + \text{noise},$$

where $\mathbf{Y}$ is a data matrix in which each row contains measurements in one spatial location and each column corresponds to one time instance. The goal is to learn the model parameters $\mathbf{W}$, $\mathbf{X}$ from available observations in order to reconstruct the missing values in $\mathbf{Y}$. Each $\mathbf{x}_{d:}$ is a vector representing the time series of one of the D factors whereas $\mathbf{w}_{:d}$ is a vector of loadings which are spatially distributed. We assume that both factors $\mathbf{x}_{d:}$ and corresponding loadings $\mathbf{w}_{:d}$ have prominent structures that we model using the Gaussian process methodology [2]. The model is identified in the framework of variational Bayesian learning and high computational cost of GP modeling is reduced by using sparse approximations derived in the variational methodology.

Another problem studied in our group is parametric tuning of climate models. Climate models contain closure parameters which can act as effective "tuning handles" of the simulated climate. These appear in physical parameterization schemes where unresolved variables are expressed by predefined parameters rather than being explicitly modeled. In the current climate model tuning process, best expert knowledge is used to define the optimal closure parameter values, based on observations, process studies, large eddy simulations, etc.

Our research group participates in the Academy of Finland project called "Novel advanced mathematical and statistical methods for understanding climate" (NOVAC, 2010-2013), whose goal is to develop algorithmic ways for closure parameter estimation. We focus on the atmospheric model ECHAM5 but the methodology is generic and applicable in any multi-scale problem with similar closure parameters [4].

The uncertainties of the closure parameters are estimated using Markov chain Monte Carlo (MCMC) simulations [5]. The MCMC approach is, however, computationally very expensive and only maximally optimized MCMC techniques make the approach realistic in practice. We develop new tools based on adaptive algorithms, multiple computational grids, parallel chains as well as methods based on early rejection.

The central problem in closure parameter estimation is how to formulate the likelihood function. This task is not trivial because of the chaotic nature of climate models. Climate model simulations quickly diverge from observations, which makes classical parameter estimation based on direct comparison of model simulations and observations inefficient. Our initial approach to circumvent the chaoticity problem was to formulate the likelihood function in terms of summary statistics. In [5], the likelihood is evaluated by comparing some temporal and spatial averages of observed and simulated data. Several summary statistics potentially useful for climate model tuning have been studied in [6].

References

[1] J. Luttinen and A. Ilin. Variational Gaussian-process factor analysis for modeling spatio-temporal data. In *Advances in Neural Information Processing Systems 22*, 2009.

[2] C. E. Rasmussen, C. K. I. Williams, *Gaussian processes for machine learning*. MIT Press, 2006.

[3] A. Ilin and J. Luttinen. Variational Gaussian-process factor analysis for modeling spatio-temporal data. In *Proc. of the 11th Meeting on Statistical Climatology*, Edinburgh, Scotland, 2010.

[4] H. Haario, E. Oja, A. Ilin, H. Järvinen and J. Tamminen Novel advanced mathematical and statistical methods for understanding climate (NOVAC). In *Proc. of the 11th Meeting on Statistical Climatology*, Edinburgh, Scotland, 2010.

[5] H. Järvinen, P. Räisänen, M. Laine, J. Tamminen, A. Ilin, E. Oja, A. Solonen, and H. Haario. Estimation of ECHAM5 climate model closure parameters with adaptive MCMC. *Atmospheric Chemistry and Physics Discussion*, Vol. 10, No 5, pp. 11951–11973, 2010.

[6] J. Saarimäki. Performance Metrics for the Atmospheric Model ECHAM5. Master's thesis, Aalto University, 2010