Global temperature analysis with non-stationary random field models

Finn Lindgren, Håvard Rue
Norwegian University of Science and Technology, finn.lindgren@math.ntnu.no

Peter Guttorp
University of Washington and Norwegian Computing Center

Abstract: Analysis of regional and global mean temperatures based on instrumental observations has typically been based on aggregating temperature measurements to grid cells. Due to the uneven data coverage, this makes analysis of the associated uncertainties difficult. We here present an alternative model based approach, where the climate and weather are modelled as random fields generated by a stochastic partial differential equation. Using the efficient Markov representations developed by Lindgren et al. (2011), direct numerical optimisation and integration with the R-INLA software provides Bayesian temperature reconstructions and associated uncertainties.

Keywords: Global temperature analysis, Stochastic partial differential equation, Gaussian Markov random field

1 Introduction

When analysing past observed weather and climate, the Global Historical Climatology Network (GHCN) data set (Peterson and Vose, 1997) is commonly used. The data spans the period 1702 through 2010, though counting, for each year, only stations with no missing values, yearly averages can be calculated only as far back as 1835. The GHCN data is used to analyse regional and global temperatures in the GISS (Hansen et al., 1999) and HadCRUT3 (Brohan et al., 2006) global temperature series, together with additional data such as ocean based sea surface temperature measurements. Differing in detail, the analyses aggregate the data into grid boxes, which are combined into global averages. To reduce the influence of station specific effects, the methods are based on the temperature anomalies, defined as the difference in weather to the local climate, the latter defined as the average weather over a 30 year reference period. Due to the difficulty of assessing the statistical uncertainty of the resulting estimates, we instead choose to construct a stochastic model for the climate and anomalies, based on a non-stationary stochastic partial differential equation.

2 Model

I order to avoiding the computational difficulties associated with calculations based on covariance matrices, we use the link between the stochastic partial differential equation (SPDE) formulation of Matérn fields and Gaussian Markov random fields (GMRFs), as developed by Lindgren et al. (2011). Together with the INLA method (Rue et al., 2009) this allows us to perform a fully Bayesian analysis in a fraction of the time required by a traditional MCMC approach.

The climate (or expected weather) is $\boldsymbol{\mu}$, the yearly anomalies are $\boldsymbol{x}_t$, and the observations are $\boldsymbol{y}_t$. The anomalies are taken as solutions to the SPDE

$$\left(\kappa^2(\boldsymbol{u}) - \Delta\right)(\tau(\boldsymbol{u})x_t(\boldsymbol{u})) = \mathcal{W}(\boldsymbol{u}), \quad \boldsymbol{u} \in \mathbb{S}^2, \tag{1}$$

where $\mathcal{W}$ is a white noise process, Δ is the Laplacian, and κ and τ are spatially varying parameters. The prior distribution for the climate field is chosen as approximate solutions to the SPDE $\Delta\mu(\boldsymbol{u}) = \sigma_\mu \mathcal{W}(\boldsymbol{u})$, which are intrinsic random fields. The model is governed by a parameter vector $\boldsymbol{\theta} = \{\boldsymbol{\theta}_\kappa, \boldsymbol{\theta}_\tau, \boldsymbol{\theta}_s, \theta_\epsilon\}$, where $\boldsymbol{\theta}_\kappa$ and $\boldsymbol{\theta}_\tau$ controls the non-stationary dependence structure of the anomalies.

Introducing *observation matrices* $\boldsymbol{A}_t$, that extract the nodes from $\boldsymbol{x}_t$ for each observation, the full model is given by

$$(\boldsymbol{\mu}|\boldsymbol{\theta}) \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{Q}_{\boldsymbol{\mu}}^{-1}), \tag{2}$$

$$(\boldsymbol{x}_t|\boldsymbol{\theta}) \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{Q}_{\boldsymbol{x}}^{-1}), \tag{3}$$

$$(\boldsymbol{y}_t|\boldsymbol{\mu}, \boldsymbol{x}_t, \boldsymbol{\theta}) \sim \mathrm{N}(\boldsymbol{A}_t(\boldsymbol{\mu} + \boldsymbol{x}_t) + \boldsymbol{S}_t\boldsymbol{\theta}_s, \boldsymbol{Q}_{\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{x}}^{-1}), \tag{4}$$

where $\boldsymbol{S}_t\boldsymbol{\theta}_s$ are station specific effects (elevation), and the $\boldsymbol{Q}_{\cdot}$ matrices are the precision matrices corresponding to each conditional distribution, obtained with the finite element method (Lindgren et al., 2011).

3 Results

We implemented the model using `R-INLA`. The Bayesian analysis draws all its conclusions from the properties of the posterior distributions of $(\boldsymbol{\theta}|\boldsymbol{y})$, $(\boldsymbol{\mu}|\boldsymbol{y})$, and $(\boldsymbol{x}|\boldsymbol{y})$, so that all uncertainty about the weather anomaly $\boldsymbol{x}_t$ is included in the distribution for the model parameters $\boldsymbol{\theta}$, et cetera. Since $(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ is Gaussian, the Bayesian integration results are only approximate with regards to the numerical integration of the covariance parameters $(\boldsymbol{\theta}_\kappa, \boldsymbol{\theta}_\tau, \boldsymbol{\theta}_\epsilon)$. Due to the large size of the data set, the initial analysis is based on data only from the period 1970 through 1989, and the analysis took approximately one hour on a 12 core Linux system.

The spatial covariance parameters are harder to interpret individually, but we instead show the resulting spatially varying field standard deviations and correlation ranges in Figure 1, including pointwise 95% credible intervals. Both curves show a clear dependence on latitude, with both larger variance and correlation range near the poles, compared with the equator.

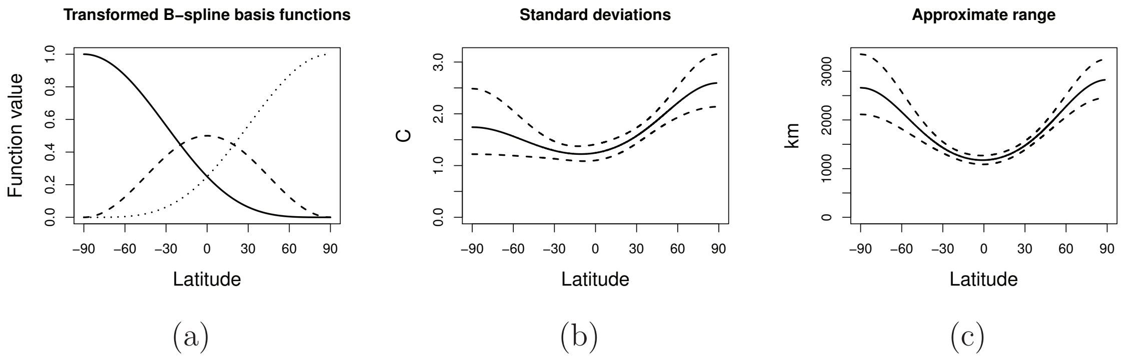

Figure 1: Three transformed B-spline basis functions of order 2 (a), and approximate 95% credible intervals for (b) standard deviation and (c) correlation range of the yearly weather, as functions of latitude.

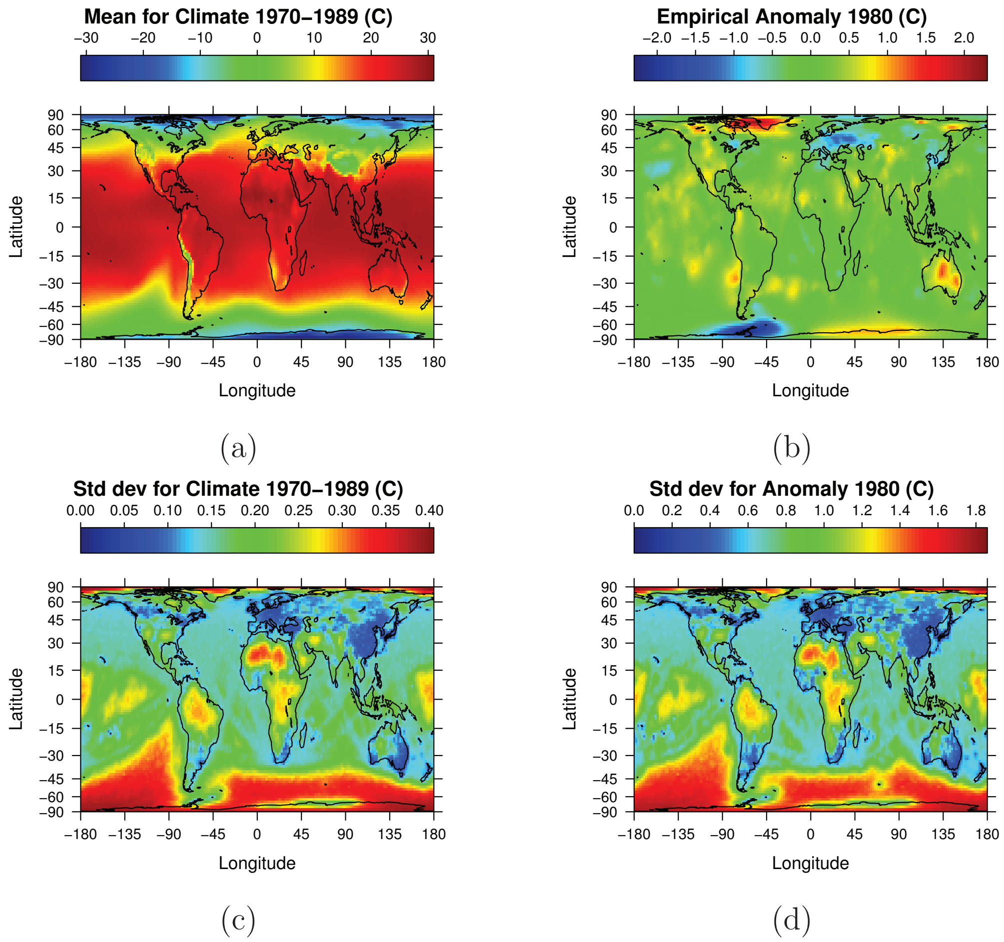

Figure 2: Posterior means for the empirical 1970–1989 climate (a) and for the empirical mean anomaly 1980 (b), together with the corresponding posterior standard deviations in (c) and (d). The climate includes the estimated effect of elevation. An area-preserving cylindrical projection is used.

In Figure 2(a) and (b), the posterior expectation of the empirical climate, $\mathrm{E}(\boldsymbol{\mu}|\boldsymbol{y})$, is shown (with the estimated effect of elevation added), together with the posterior expectation of the temperature anomaly for 1980, $\mathrm{E}(\boldsymbol{x}_{1980}|\boldsymbol{y})$. The spatial dependence model was based on the GHCN data, but these Kriging estimates also include ocean-based data. A preliminary analysis indicates that the dependence structure is different for land and ocean, which can be handled by adding appropriate basis functions to the κ and τ models. The pre-gridded ocean data is also a good example of how the observation matrix $\boldsymbol{A}_t$ can solve the problem of "misaligned" data, since it decouples the spatial model from the data locations, allowing arbitrary linear measurement equations from one spatial model.

References

Brohan, P., Kennedy, J., Harris, I., Tett, S., and Jones, P. (2006). Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *Journal of Geophysical Research*, 111.

Hansen, J., Ruedy, R., Glascoe, J., and Sato, M. (1999). GISS analysis of surface temperature change. *Journal of Geophysical Research*, 104:30997–31022.

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 approach (with discussion). *Journal of the Royal Statistical Society, Series B*. In press.

Peterson, T. and Vose, R. (1997). An overview of the Global Historical Climatology Network temperature database. *Bulletin of the American Meteorological Society*, 78(12):2837–2849.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71(2):319–392.