LUND UNIVERSITY

A Gaussian Markov random field model for total yearly precipitation over the African Sahel

Lindström, Johan; Lindgren, Finn

Published in:
Preprints in Mathematical Sciences

2008

Link to publication

Citation for published version (APA):
Lindström, J., & Lindgren, F. (2008). A Gaussian Markov random field model for total yearly precipitation over the African Sahel. Unpublished.

Total number of authors:
2

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

A Gaussian Markov Random Field Model for Total Yearly Precipitation over the African Sahel

Johan Lindström and Finn Lindgren

Preprints in Mathematical Sciences
2008:8

Lund Institute of Technology
Lund University

Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

A Gaussian Markov Random Field Model for Total Yearly Precipitation over the African Sahel

Johan Lindström Finn Lindgren

16th April 2008

Abstract

A spatio-temporal model is constructed to interpolate yearly precipitation data from 1982 to 1996 over the African Sahel. The precipitation data used in the analysis comes from the Global Historical Climatology Network.

The spatio-temporal model is based on a Gaussian Markov random field approach with AR(1)-dependence in time and a spatial component modeled using an approximation of a field with Matérn covariance. The model is defined on an irregular grid on a segment of the sphere, both avoiding the issue of matching observations to a regularly spaced grid, and handling the curvature of the Earth.

The model is estimated using a Markov chain Monte Carlo approach. The formulation as a Markov field allows for relatively efficient computations, even though the field has more than $3 \cdot 10^4$ nodes.

1 Introduction

One of the most widely used methods for interpolation of spatial data is Kriging, named after Krige (1951) and popularised when Matheron (1963) put linear interpolation in a geostatistical context. The Kriging predictor is a linear combination of observations; and thus suitable for Gaussian data, or data that is Gaussian after appropriate transformations (Box & Cox, 1964). The Kriging weights in the linear combination depend on the estimated mean and covariance structure of the data.

To allow for non-Gaussian observations of an unobserved, underlying Gaussian field Diggle *et al.* (1998) introduced a model based approach that uses Markov chain Monte Carlo (MCMC) methods to do Bayesian inference for the spatial predictors. The MCMC predictors also address the uncertainty due to parameter estimates. However, the approach is computationally intensive, requiring the inverse of the full covariance matrix at each iteration; for n observations, inverting the covariance matrix takes $\mathcal{O}\left(n^3\right)$ operations.

A computationally more effective method was introduced by Wikle *et al.* (1998). They let the spatial field be a Gaussian Markov random field (GMRF) defined on a rectangular grid in $\mathbb{R}^2$. The formulation of GMRFs using a sparse precision (inverse covariance) matrix lends itself to fast sampling from the predictive and posterior distributions (Rue, 2001). Due to the computational gains provided by the sparse precision matrix (Rue & Follestad, 2002) spatial modelling using GMRFs is popular and has been used for a wide variety of applications (e.g. Knorr-Held & Rue, 2002; Gamerman *et al.*, 2003; Bolin *et al.*, 2008).

Rue & Tjelmeland (2002) provide a link between GMRFs and Kriging by showing that a GMRF on a rectangular grid in $\mathbb{R}^2$ can be used to approximate fields with a wide class of covariance functions; the approximation is good even for fields with slowly decreasing correlation. A problem with defining the field on a rectangular grid is that our observations seldomly fall on the grid points. This can be remedied either by assigning each observation to the closest grid point (Hrafnkelsson & Cressie, 2003) or by letting values at the observations points be some (linear) interpolation of the values at nearby grid points (Werner Hartman, 2006).

Recently, Lindgren & Rue (2007) demonstrated that GMRFs also can be constructed directly to approximate fields with Matérn covariance functions. The approximating GMRF can be defined at irregular locations on a general, smooth manifold. Defining the field at irregular locations avoids the task of having to map observations to grid points. Additionally, since the GMRF is defined on a smooth manifold, it is possible to handle the curvature of the Earth; an effect that often is ignored by projecting the observation locations to $\mathbb{R}^2$ or by embedding the sphere in $\mathbb{R}^3$ and using covariance functions defined in $\mathbb{R}^3$.

Here we will use a spatio-temporal model, where the spatial component is based on the GMRF approximation proposed by Lindgren & Rue (2007), to interpolate precipitation data over the African Sahel, the region between Sahara and the rainforest. Parameter estimation is done using a MCMC approach.

Section 2 motivates the study of precipitation by giving a brief background to the environmental situation in the African Sahel, with 2.1 presenting the precipitation data; a brief overview of relevant theory is given in Section 3, the spatio-temporal model is described in Section 4, Section 5 discusses some details of the MCMC approach that is used for parameter estimation, and the analysis and results are detailed in Section 6. Finally, Section 7 summarises, and outlines some possible extensions.

2 The African Sahel

The African Sahel suffered a severe decline in rainfall following the 1960's (Hulme, 2001), and several droughts and famines have beset the area (Olsson, 1993). Recent studies indicate a vegetation recovery (Eklundh & Olsson, 2003; Olsson *et al.*, 2005; Bolin *et al.*, 2008) linked to an increase in precipitation (Hickler *et al.*, 2005). The region has attracted the interest of the remote sensing community for several years (Tucker *et al.*, 1985; Rasmussen, 1992; Prince *et al.*, 1998; Fensholt *et al.*, 2004), an interest fueled by the availability of remotely sensed vegetation data (Agbu & James, 1994; NASA, 2004) stretching back to 1982.

While the remotely sensed vegetation data covers the entire Sahel, precipitation measurements are restricted to point measurements at weather stations. Several analysis of the dependence between vegetation and precipitation have been done, however most of these either take a time series perspective, limiting the analysis to vegetation measurements close to each weather station and treating data from each station as independent time series (Los *et al.*, 2006) or give only qualitative (Nicholson *et al.*, 1990) analyses of relationships. As a first step in an effort towards building a spatio-temporal model for the dependence between vegetation and precipitation we interpolate the precipitation data between stations. Since studies of long term, year to year variations in vegetation (Eklundh & Olsson, 2003; Olsson *et al.*, 2005) primarily use measures of total biomass during each growing season, we will limit ourselves to interpolation of total yearly precipitation.

2.1 Precipitation data

The precipitation data used in our analysis comes from the Global Historical Climatology Network (GHCN) (NOAA, 2007; Vose *et al.*, 1992). The data in GHCN consists of monthly temperature, precipitation, and pressure measurements from a large number of weather stations around the globe, stretching back as far as the 18th century in some locations. Extracting the geographical area and time period of interests gives us precipitation measurements from a total of 550 measurement stations across the Sahel for 15 years of data (1982–1996). After 1996 the number of reporting stations decline sharply. The measurements are then summed to total yearly precipitation at each station. An example of data and measurement locations from 1982 is given in Figure 1. Finally we remove ten stations from the data set; these stations will be used to asses model accuracy after the analysis.

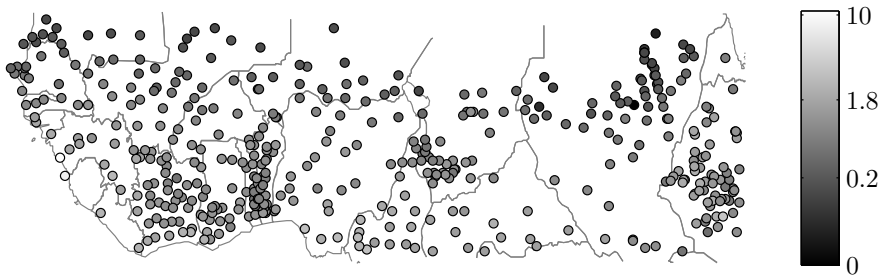

Figure 1: Yearly precipitation, in metres, for the 443 stations that have reported measurements for 1982. Note that the colour scale is non-linear.

3 Theory

Here the definition and properties of a Gaussian Markov random field are given, and the approximation of a spatial field with Matérn covariance using a GMRF is outlined.

3.1 Gaussian Markov random fields

A random variable $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathsf{N}\left(\boldsymbol{\mu}, \boldsymbol{Q}^{-1}\right)$ is called a Gaussian Markov random field (see Rue & Held, 2005, for extensive details) if the joint distribution for $\boldsymbol{x}$ satisfies $\mathrm{p}(x_i|x_{-i}) = \mathrm{p}(x_i|x_{\mathcal{N}_i})\ \forall\ i$. Here x_{-i} denotes all elements in $\boldsymbol{x}$ except x_i and $\mathcal{N}_i$ is the neighbourhood of i; the neighbourhood, $\mathcal{N}_i$, is a set of pixels $j \in \mathcal{N}_i$ that, in some sense, are "close" to i. An important implication of the above definition is that if $i \neq j$, then

$$x_i \perp x_j | x_{-\{i,j\}} \Longleftrightarrow Q_{i,j} = 0 \Longleftrightarrow j \notin \mathcal{N}_i.$$

That is, the following properties are equivalent: 1) x_i and x_j are conditionally independent 2) the corresponding element in the precision matrix, $Q_{i,j}$ is zero and 3) i and j are not neighbours. Since $Q_{i,j} \neq 0$ only if i and j are neighbours, most GMRFs will have sparse precision matrices. If the sparse precision matrix is utilised when doing calculations on the GMRF, large gains in computational time are often possible (Rue, 2001; Rue & Follestad, 2002; Knorr-Held & Rue, 2002).

3.2 Approximation of Matérn fields

We are now ready to study the approximation of fields with Matérn covariances using GMRFs. As noted by Whittle (1963), the solutions, in $\mathbb{R}^d$, to the stochastic partial differential equation (SPDE)

$$(\Delta - \kappa^2)^{\alpha/2} \boldsymbol{x}(\boldsymbol{s}) = \mathcal{E}(\boldsymbol{s}), \tag{1}$$

where $\mathcal{E}$ is Gaussian white noise, $\alpha > 0$ and $\kappa > 0$, are Gaussian random fields with covariances that correspond to the Matérn family

$$\mathsf{C}(\boldsymbol{x}(\boldsymbol{s}), \boldsymbol{x}(\boldsymbol{s}+\boldsymbol{\tau})) \propto (\kappa \|\boldsymbol{\tau}\|)^{\nu} K_{\nu}(\kappa \|\boldsymbol{\tau}\|).$$

Here K_ν is a modified Bessel function and the relation between the regularity parameter, ν, and α in the SPDE is $\alpha = \nu + d/2$.

The relationship between the SPDE (1) and Gaussian random fields with covariances from the Matérn family is used by Lindgren & Rue (2007) to construct GMRF approximations to fields with Matérn covariances on $\mathbb{R}^d$. Lindgren & Rue argue that a natural generalisation of Matérn covariances to general manifolds can be found by solving (1) on said manifolds, and they proceed to construct a GMRF approximation of the solution on general manifolds. The approximation is found using the finite element method on a triangulation of irregularly spaced points. The resulting GMRF is defined on the points of the triangulation, making it suitable for modelling fields that are observed at irregular locations.

For $\alpha = 2$ in (1), the precision matrix of the approximating GMRF is

$$\boldsymbol{Q} = (\boldsymbol{H} - \kappa^2 \boldsymbol{C})^\top \boldsymbol{C}^{-1} (\boldsymbol{H} - \kappa^2 \boldsymbol{C}), \tag{2}$$

where the matrices $\boldsymbol{H}$ and $\boldsymbol{C}$ depend on the triangulated point configuration (see Lindgren & Rue, 2007, for details on how to calculate $\boldsymbol{H}$ and $\boldsymbol{C}$). The matrix $\boldsymbol{H}$ is sparse and $\boldsymbol{C}$ is diagonal, resulting in a sparse precision matrix $\boldsymbol{Q}$. Since we model precipitation on the sphere, i.e. $d = 2$, a GMRF with precision matrix according to (2) approximates a field with generalised Matérn covariance where the smoothness parameter is $\nu = 1$.

As a final detail it should be noted that Whittle studied the SPDE (1) in $\mathbb{R}^d$; When working with fields on a finite manifold, boundary effects will influence the field close to the edge of the manifold. In order to minimise these edge effects the manifold on which the field is defined is extended some distance outside of the area of interest.

4 Model

Given the GMRF approximation of fields with generalised Matérn covariance outlined above we are now ready to introduce the spatio-temporal model for the precipitation. First we will introduce the model in a general setting, thereafter Section 4.1 describes the spatio-temporal model using GMRFs, and finally Section 4.2 deduces the posterior distribution for the field given observations.

To simplify the following narrative we first introduce some notation. Below $\boldsymbol{I}$ denotes identity-matrices, $\mathbf{1}$ denotes unity (column) vectors and $\otimes$ is the Kronecker product. Further we will need the canonical representation

of Gaussian densities (Rue & Held, 2005):

$$\boldsymbol{x} \in \mathsf{N}\left(\boldsymbol{Q}^{-1}\boldsymbol{b}, \boldsymbol{Q}^{-1}\right) \iff \boldsymbol{x} \in \mathsf{N}_{\mathsf{C}}\left(\boldsymbol{b}, \boldsymbol{Q}\right).$$

In general terms our model is a latent field on a segment of the globe, Ω, and in time, i.e. $\boldsymbol{x}(\boldsymbol{s}, \tau)$ with $\boldsymbol{s} \in \Omega$ and $\tau \in \mathbb{R}$. The field is approximated by a GMRF evaluated at a number of discrete points, $\boldsymbol{X}(\boldsymbol{s}_i, t)$, $i = 1 \ldots N$ and $t = 1 \ldots T$; and the precipitation is modeled as observations of the latent field. The yearly precipitation measurements are highly skewed, but transforming the observations as

$$y = \tilde{y}^{1/3} \cdot \left(1 - 0.13\tilde{y}^{1/3}\right),$$

where $\tilde{y}$ is precipitation in metres, the transformed precipitation can be modelled as Gaussian observations of the latent field (Glasbey & Nevison, 1997). Each of the d observations is now taken as the field value at the corresponding measurement locations with additive, independent $\mathsf{N}\left(0, \sigma^2\right)$ errors

$$y_j = \boldsymbol{X}(\boldsymbol{s}_j, t_j) + \epsilon_j, \tag{3}$$

where $j = 1 \ldots d$ and ϵ_j are the errors.

4.1 Spatio-temporal model

Given the GMRF approximation of fields with Matérn covariances we introduce a precision matrix for the spatial dependence as $\boldsymbol{Q}_S = \chi \boldsymbol{Q}$ where $\boldsymbol{Q}$ is defined in (2) and χ is a variance scaling parameter. Further letting $\boldsymbol{X}_t$ denote the N-by-1 column vector representing the GMRF at each time point, the spatio-temporal field can be represented as $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{X}_1^\top & \cdots & \boldsymbol{X}_T^\top\end{bmatrix}^\top$.

Assuming an AR(1)-structure between consecutive years and a mean field, $\boldsymbol{\mu}(\boldsymbol{s})$, that is constant in time, the field $\boldsymbol{X}$, can be modeled as

$$(\boldsymbol{X}_t - \boldsymbol{\mu}) = a(\boldsymbol{X}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \tag{4}$$

where the innovations are independent in time but spatially correlated,

$$\boldsymbol{\eta}_t \in \mathsf{N}\left(\boldsymbol{0}, \boldsymbol{Q}_S^{-1}\right).$$

Further we take $\boldsymbol{X}_1 \in \mathsf{N}\left(\boldsymbol{\mu}, \boldsymbol{Q}_S^{-1}/(1-a^2)\right)$, analogously with the stationary distribution of an AR(1)-process. Combining the distribution of $\boldsymbol{X}_1$ with (4) the distribution for the spatio-temporal GMRF becomes

$$\boldsymbol{X} \in \mathsf{N}\left(\boldsymbol{1} \otimes \boldsymbol{\mu}, \left(\boldsymbol{Q}_T \otimes \boldsymbol{Q}_S\right)^{-1}\right). \tag{5}$$

where $\boldsymbol{Q}_T$ is the tri-diagonal T-by-T matrix

$$\boldsymbol{Q}_T = \begin{pmatrix} 1 & -a & 0 & & & & \\ -a & 1+a^2 & -a & 0 & & & \\ & & \ddots & & & & \\ & 0 & -a & 1+a^2 & -a & 0 & \\ & & & & \ddots & & \\ & & & 0 & -a & 1+a^2 & -a \\ & & & & 0 & -a & 1 \end{pmatrix}. \tag{6}$$

Note that $|\boldsymbol{Q}_T \otimes \boldsymbol{Q}_S| = |\boldsymbol{Q}_T|^N |\boldsymbol{Q}_S|^T = (1-a^2)^N |\boldsymbol{Q}_S|^T$, since $|\boldsymbol{Q}_T| = 1-a^2$, see Appendix A.

The spatio-temporal covariance function of the field will be separable since the precision matrix for the joint spatio-temporal field can be written as a Kronecker product between the precision matrices for the temporal and spatial dependencies respectively.

4.2 Posterior density given observations

As previously mentioned we model the transformed precipitation as Gaussian observations of an underlying GMRF. Stacking all the observations in a d-by-1 vector $\boldsymbol{Y}$ the observation equation (3) can be formulated in matrix form as

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{\epsilon}, \tag{7}$$

where $\boldsymbol{\epsilon} \in \mathsf{N}\left(0, \boldsymbol{I}\sigma^2\right)$ and $\boldsymbol{A}$ is a d-by-NT observation matrix with one non-zero value on each row picking out the field value corresponding to that observation. The structure of $\boldsymbol{A}$ implies that $\boldsymbol{A}^\top\boldsymbol{A}$ is a diagonal matrix. It is also worth noticing that, comparing to a direct covariance formulation for the measurements, σ^2 represents a nugget effect.

The mean field $\boldsymbol{\mu}$ in (5) should handle large scale variations and is modeled as $\boldsymbol{\mu} = \boldsymbol{B}\boldsymbol{\theta}$ where $\boldsymbol{B}$ is a known matrix of regression basis vectors and $\boldsymbol{\theta}$ contains the unknown regression parameters. Selection of a suitable regression basis is discussed in Section 6.1. Using this mean field the prior distribution of the underlying GMRF is

$$\boldsymbol{X} \in \mathsf{N}\left(\boldsymbol{1} \otimes \boldsymbol{B}\boldsymbol{\theta}, \left(\boldsymbol{Q}_T \otimes \boldsymbol{Q}_S\right)^{-1}\right). \tag{8}$$

Finally we collect the model parameters as $\boldsymbol{\psi} = \{\chi, \kappa^2, a, \sigma^2\}$.

Given (7) and (8) the full posterior is

$$\mathrm{p}(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{Y}) \propto \mathrm{p}(\boldsymbol{Y}|\boldsymbol{X}, \sigma^2)\mathrm{p}(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{\psi})\mathrm{p}(\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\psi}), \tag{9}$$

where the priors for $\boldsymbol{\theta}$, and $\boldsymbol{\psi}$ are assumed to be independent and we select a conjugate Gaussian prior for $\boldsymbol{\theta}$, i.e. $\boldsymbol{\theta} \in \mathsf{N}\left(\boldsymbol{\mu}_\theta, \boldsymbol{Q}_\theta^{-1}\right)$.

An alternative formulation of (9) is

$$\mathrm{p}(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{Y}) = \mathrm{p}(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{Y})\mathrm{p}(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{Y})p(\boldsymbol{\psi}|\boldsymbol{Y}). \tag{10}$$

As shown in Appendix B, the densities in (10) are given by

$$(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{Y}) \in \mathsf{N}_{\mathsf{C}}\left((\boldsymbol{Q}_T\mathbf{1}\otimes\boldsymbol{Q}_S\boldsymbol{B})\boldsymbol{\theta}+\boldsymbol{b},\widehat{\boldsymbol{Q}}\right) \tag{11a}$$

$$(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{Y}) \in \mathsf{N}_{\mathsf{C}}\left(\boldsymbol{b}_\theta,\widehat{\boldsymbol{Q}}_\theta\right) \tag{11b}$$

$$\begin{aligned}\mathrm{p}(\boldsymbol{\psi}|\boldsymbol{Y}) \propto \exp&\left(-\frac{\boldsymbol{Y}^\top\boldsymbol{Y}}{2\sigma^2}+\frac{\boldsymbol{b}^\top\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{b}}{2}+\frac{\boldsymbol{b}_\theta^\top\widehat{\boldsymbol{Q}}_\theta^{-1}\boldsymbol{b}_\theta}{2}\right)\\ &\cdot\left(\frac{|\boldsymbol{Q}_S|^T(1-a^2)^N}{\left|\widehat{\boldsymbol{Q}}\right|\left|\widehat{\boldsymbol{Q}}_\theta\right||\sigma^2\boldsymbol{I}|}\right)^{1/2}\mathrm{p}(\boldsymbol{\psi}),\end{aligned} \tag{11c}$$

where

$$\begin{aligned}\boldsymbol{b} &= \frac{\boldsymbol{A}^\top\boldsymbol{Y}}{\sigma^2}, \qquad \widehat{\boldsymbol{Q}} = (\boldsymbol{Q}_T\otimes\boldsymbol{Q}_S)+\frac{\boldsymbol{A}^\top\boldsymbol{A}}{\sigma^2},\\ \boldsymbol{b}_\theta &= \boldsymbol{Q}_\theta\boldsymbol{\mu}_\theta+(\boldsymbol{Q}_T\mathbf{1}\otimes\boldsymbol{Q}_S\boldsymbol{B})^\top\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{b},\\ \widehat{\boldsymbol{Q}}_\theta &= \boldsymbol{Q}_\theta+(\boldsymbol{Q}_T\mathbf{1}\otimes\boldsymbol{Q}_S\boldsymbol{B})^\top\widehat{\boldsymbol{Q}}^{-1}\left(\frac{\boldsymbol{A}^\top\boldsymbol{A}}{\sigma^2}\right)(\mathbf{1}\otimes\boldsymbol{B}).\end{aligned}$$

The final element of the model is now to define the priors for $\boldsymbol{\psi}$. We assume independent priors and note that χ, κ^2, and σ^2 takes values in $\mathbb{R}^+$ and that $a\in(-1,1)$. Therefore we select Gamma priors for the first three variables and a uniform prior for a:

$$\chi\in\Gamma(\alpha_\chi,\beta_\chi),\quad \kappa^2\in\Gamma(\alpha_\kappa,\beta_\kappa),\quad \sigma^2\in\Gamma(\alpha_\sigma,\beta_\sigma),\quad a\in\mathsf{U}(-1,1). \tag{12}$$

Suitable values for the hyper-parameters are given in Appendix C. Figure 2 depicts a directed acyclic graph for our model with hyper-parameters.

5 Markov chain Monte Carlo

To estimate the posterior of the parameters, $\mathrm{p}(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{Y})$, and the posterior of the field, $\mathrm{p}(\boldsymbol{X}|\boldsymbol{Y})$, a Metropolis-Hastings based MCMC-algorithm is used (Metropolis *et al.*, 1953; Hastings, 1970). The proposal kernel, $\mathrm{q}(\cdot,\cdot)$, in the MCMC-algorithm, should generate a proposal for the variables $\{\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\psi}\}$ given variables at a previous iteration $\{\boldsymbol{X}_{\text{old}},\boldsymbol{\theta}_{\text{old}},\boldsymbol{\psi}_{\text{old}}\}$ and observed data $\boldsymbol{Y}$. The proposal is then accepted with a certain probability α_{MCMC}.

Due to the structure of the posterior density in (10) a suitable form of the proposal kernel is

$$\mathrm{q}(\{\boldsymbol{X}_{\text{old}},\boldsymbol{\theta}_{\text{old}},\boldsymbol{\psi}_{\text{old}}\},\{\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\psi}\};\boldsymbol{Y}) = \mathrm{p}(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{Y})\mathrm{p}(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{Y})\mathrm{q}_\psi(\boldsymbol{\psi}_{\text{old}},\boldsymbol{\psi}),$$

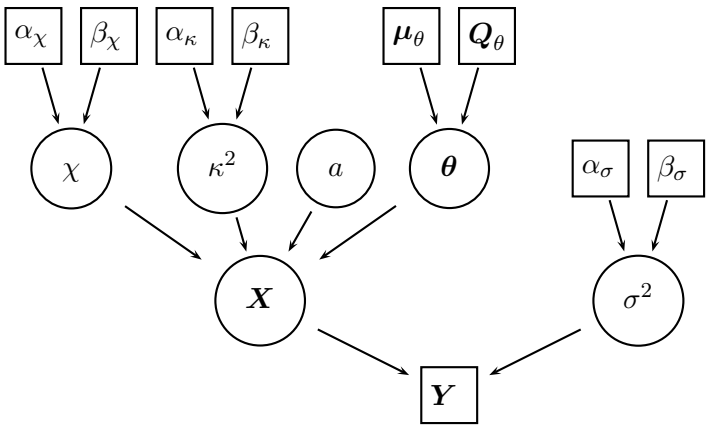

Figure 2: Directed acyclic graph for the hierarchical model with hyper-parameters.

where $\mathrm{p}(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{Y})$ and $\mathrm{p}(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{Y})$ are taken from (11). Using this proposal the acceptance probability for the new proposal, $\{\boldsymbol{X}\boldsymbol{\theta},\boldsymbol{\psi}\}$, reduces to

$$\begin{aligned}\alpha_{\mathrm{MCMC}} &= \min\left(1, \frac{\mathrm{p}(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{Y})\mathrm{q}(\{\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\psi}\},\{\boldsymbol{X}_{\mathrm{old}},\boldsymbol{\theta}_{\mathrm{old}},\boldsymbol{\psi}_{\mathrm{old}}\};\boldsymbol{Y})}{\mathrm{p}(\boldsymbol{X}_{\mathrm{old}},\boldsymbol{\theta}_{\mathrm{old}},\boldsymbol{\psi}_{\mathrm{old}}|\boldsymbol{Y})\mathrm{q}(\{\boldsymbol{X}_{\mathrm{old}},\boldsymbol{\theta}_{\mathrm{old}},\boldsymbol{\psi}_{\mathrm{old}}\},\{\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\psi}\};\boldsymbol{Y})}\right)\\ &= \min\left(1, \frac{\mathrm{p}(\boldsymbol{\psi}|\boldsymbol{Y})\mathrm{q}_\psi(\boldsymbol{\psi},\boldsymbol{\psi}_{\mathrm{old}})}{\mathrm{p}(\boldsymbol{\psi}_{\mathrm{old}}|\boldsymbol{Y})\mathrm{q}_\psi(\boldsymbol{\psi}_{\mathrm{old}},\boldsymbol{\psi})}\right).\end{aligned}$$

Note that the proposal for $\{\boldsymbol{X},\boldsymbol{\theta}\}$ does not effect the acceptance probability and thus we only need to generate a new proposal for $\boldsymbol{\psi}$ before doing the accept/reject step. The proposal for $\{\boldsymbol{X},\boldsymbol{\theta}\}$ is generated only if the new $\boldsymbol{\psi}$ is accepted, resulting in corresponding savings in computation time.

Initial MCMC-runs indicated a strong correlation between the components in $\boldsymbol{\psi}$ (see Figure 3). Transforming the variables as

$$\begin{aligned}\widetilde{\sigma}^2 &= \log(\sigma^2), & \widetilde{a} &= \log(1+a) - \log(1-a),\\ \widetilde{\kappa}^2 &= \log(\kappa^2), & \widetilde{\chi} &= \log(\chi),\end{aligned} \tag{13}$$

we obtain variables which can take values in $\mathbb{R}$, so that a random walk proposal on the transformed variables can be used

$$\begin{bmatrix}\widetilde{\sigma}^2\\ \widetilde{a}\\ \widetilde{\kappa}^2\\ \widetilde{\chi}\end{bmatrix} \in \mathsf{N}\left(\begin{bmatrix}\widetilde{\sigma}^2_{\mathrm{old}}\\ \widetilde{a}_{\mathrm{old}}\\ \widetilde{\kappa}^2_{\mathrm{old}}\\ \widetilde{\chi}_{\mathrm{old}}\end{bmatrix}, \boldsymbol{\Sigma}_{\mathrm{prop}}\right). \tag{14}$$

Here the covariance matrix $\boldsymbol{\Sigma}_{\mathrm{prop}}$ takes the correlation between variables into account (suitable values are given in Appendix C). Using the above

proposal and transforming back to the original coordinates the ratio of the proposal densities is

$$\frac{\mathrm{q}_\psi(\boldsymbol{\psi}, \boldsymbol{\psi}_{\mathrm{old}})}{\mathrm{q}_\psi(\boldsymbol{\psi}_{\mathrm{old}}, \boldsymbol{\psi})} = \frac{\sigma^2\kappa^2\chi(1-a^2)}{\sigma^2_{\mathrm{old}}\kappa^2_{\mathrm{old}}\chi_{\mathrm{old}}(1-a^2_{\mathrm{old}})}.$$

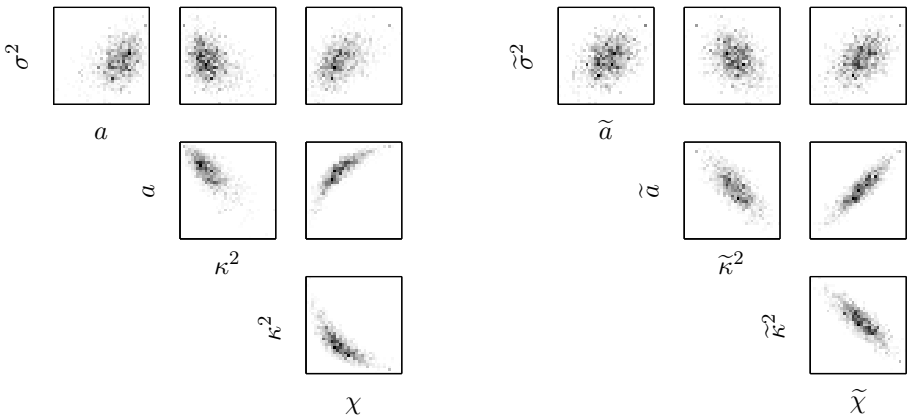

Figure 3: Two-dimensional histograms illustrating the dependence between the components of $(\boldsymbol{\psi}|\boldsymbol{Y})$ before (left pane) and after (right pane) the transformation given above in (13).

6 Analysis

Having presented the spatio-temporal model and some details of the MCMC estimation, we are now ready to interpolate the precipitation data. In Section 6.1 we introduce the spatial regression basis that characterises our mean field; hereafter the triangulation of the measurement locations is presented in Section 6.2. Results from the MCMC are compared to measurements of the field in Section 6.3, and in Section 6.4 covariance functions estimated from data are compared to those obtained from the MCMC. Finally, in Section 6.5, we make some notes and comments regarding the computational cost of the algorithm.

6.1 Linear Regression basis

In Section 4.2 the prior mean of $\boldsymbol{X}$ was specified using a matrix of regression basis vectors, $\boldsymbol{\mu} = \boldsymbol{B\theta}$. It is now time to specify the exact form of the regression basis vectors. Studying the data the latitude (north-south) dependence is described by a broken trend, which is modeled using two piecewise linear functions, with the break point found using a least squares criterion. Additionally, large scale variations in precipitation are modeled

using a cubic B-spline surface with five knots in latitude and seven knots in longitude under the additional constraint of zero derivatives at the edges, resulting in 15 basis surfaces, see Figure 4. The latitude knots are selected as the 0%, 25%, 50%, 75%, and 100% quantiles of the latitude coordinates; selection of the longitudinal knots is done in a similar fashion.

6.2 Triangulation

As mentioned in Section 3.2, the GMRF used to model the data is defined on a set of irregular, triangulated, points. This set of points includes the measurement stations, but can also be extended by additional points to obtain a suitable spatial resolution where we have few measurement stations. The set of points should also include a few points surrounding the region of interest to counteract the possible edge effects discussed in Section 3.2. Here we have applied the triangulation algorithm written by Shewchuk (1996) to latitude and longitude coordinates of the measurement stations, see Figure 5. A dense grid without any extremely obtuse triangles is obtained by setting limitations on area and minimal angle of each triangle within the area of interest, allowing the algorithm to add points as needed.

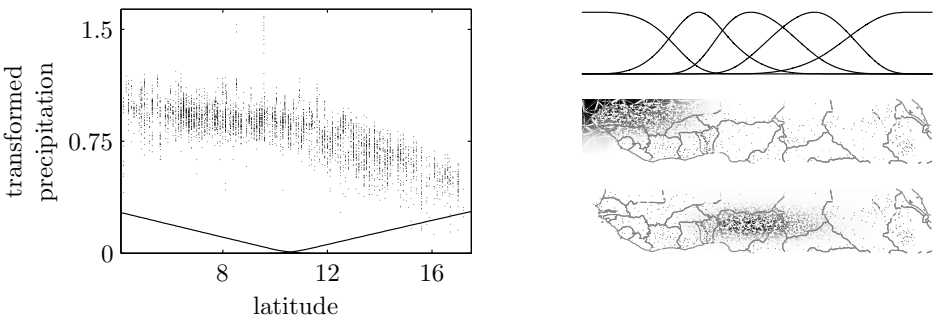

Figure 4: The left pane illustrates how data depends on the latitude, the two basis functions used to model this dependence are given in the lower part of the figure. The three right hand panes illustrate some of the 15 B-spline surface basis functions. The top pane depicts the five longitudinal B-splines while the mid and bottom panes show two of the resulting surfaces; measurement locations and national boundaries included for reference.

6.3 Posterior distributions

Given the spatio-temporal model described in Section 4.1 the MCMC-algorithm outlined in Section 5 was run on the transformed yearly precipitation data for 5000 iterations. The run was limited to 5000 iterations due to the computational cost, see Section 6.5. The Markov chain rapidly reached stationarity and exhibited good mixing, see Figure 6, with an average acceptance probability of 32.5%. Additionally, the estimated autocorrelation functions for

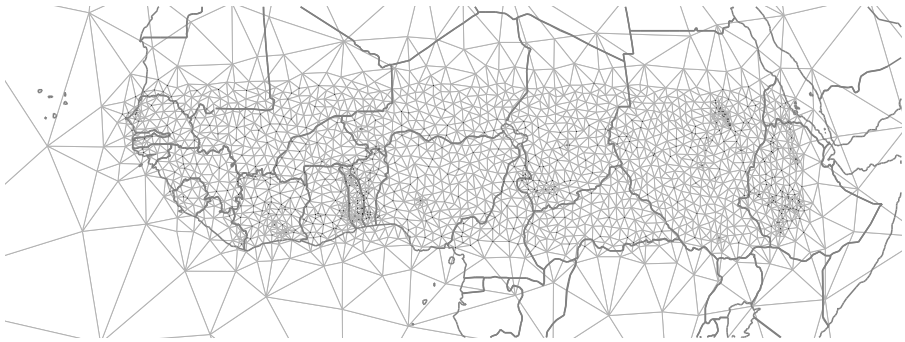

Figure 5: Triangulation (light grey) of the measurement locations (black dots), adding extra points in the interior. Note that the triangulation extends beyond the measurement locations (and this figure) to handle edge effects.

all four parameter trajectories decline below 0.1 somewhere between lag 25 and 55, and to zero between lag 40 and 70.

Given MCMC samples of $(\boldsymbol{X} \mid \boldsymbol{Y})$, the posterior mean, $\mathsf{E}(\boldsymbol{X} \mid \boldsymbol{Y})$, and 95% pointwise confidence intervals were constructed. Studying the interpolated field from 1982 shows a good correlation between measurements and the interpolation, see Figure 7. Results for the other years are similar. Worth noticing is that, due to the transformation of the precipitation data, back transformed confidence intervals are wider for large data values, as well as being skewed towards larger values.

The models ability to reproduce data at the ten stations omitted from the analysis is illustrated in Figure 8. In most cases the observations fall within the 95% confidence bands.

6.4 Covariance functions

As a last step in assessing the model we study the spatial covariance functions of the temporal residuals

$$\boldsymbol{e}_t = \big(\boldsymbol{Y}_t - \widehat{\boldsymbol{A}}_t \boldsymbol{B}\boldsymbol{\theta}\big) - a\big(\boldsymbol{Y}_{t-1} - \widehat{\boldsymbol{A}}_t \boldsymbol{B}\boldsymbol{\theta}\big). \tag{15}$$

Here $\widehat{\boldsymbol{A}}_t$ is a matrix that extracts fields points that have been observed both at time t and $t-1$. When actually calculating the residuals the true parameter values are unknown and we replace them with the corresponding posterior mean estiamtes. The residuals $\boldsymbol{e}_t$ should now essentially represent the spatially dependent innovations $\boldsymbol{\eta}_t$ in the AR(1) formulation (4), with additional independent measurement noise due to the errors $\boldsymbol{\epsilon}$ in (7). It is easy to show that the distribution of $\boldsymbol{e}_t$ is

$$\boldsymbol{e}_t \in \mathsf{N}\left(0, \widehat{\boldsymbol{A}}_t \boldsymbol{Q}_S^{-1} \widehat{\boldsymbol{A}}_t^\top + \sigma^2(1+a^2)\boldsymbol{I}\right).$$

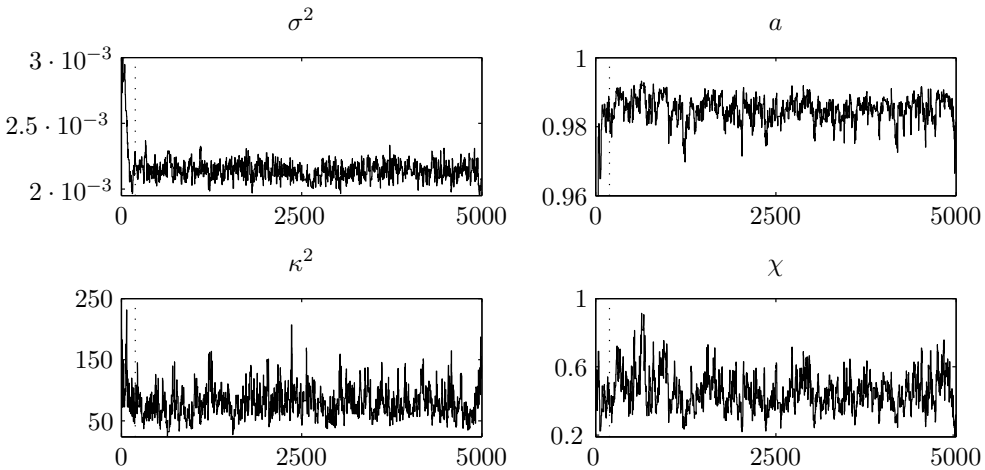

Figure 6: The four panes illustrate the parameter trajectories for the 5000 MCMC simulations, with the dotted vertical lines indicating the end of the burn in period.

The covariance function for $\boldsymbol{e}_t$ was estimated and compared to 1000 covariance function estimates based on simulations from the model given by (7) and (5) and using the posterior mean of the parameters. The estimates are given in Figure 9 and it is clear that the covariance function estimated from data does not deviate further from the theoretical covariance function than what can be expected based on the estimates obtained from simulated data.

6.5 Computational costs

The dominating computational cost for each MCMC iteration is evaluation of the likelihood (11c) and the most costly operation when evaluating the likelihood is calculation of the Cholesky factorisation of $\widehat{\boldsymbol{Q}}$. Given the Cholesky factor, calculation of the determinant is trivial and vector calculations involving $\widehat{\boldsymbol{Q}}^{-1}$ can be done efficiently using back-substitution (Rue, 2001).

Given a spatial GMRF (no time dependence) defined on a regular square lattice consisting of n points (i.e. a side of $\sqrt{n}$) calculation of the Cholesky factor takes $\mathcal{O}\left(n^{3/2}\right)$ operations (Rue & Held, 2005, Chapter 2.4), compared to $\mathcal{O}\left(n^3\right)$ operations when inverting a full covariance matrix. The $\mathcal{O}\left(n^{3/2}\right)$ complexity seems to hold also for the GMRFs defined on irregular grids used here. However we pay a price for extending the GMRF to a spatio-temporal model; for a spatio-temporal GMRF defined on a regular cube with of n points (i.e. a side of $\sqrt[3]{n}$) calculation of the Cholesky factor takes $\mathcal{O}\left(n^2\right)$ operations (Rue & Held, 2005, Chapter 2.4), and this complexity seems to be approximately true for our spatio-temporal GMRF.

As an example, calculation of the Cholesky factor of the pure spatial

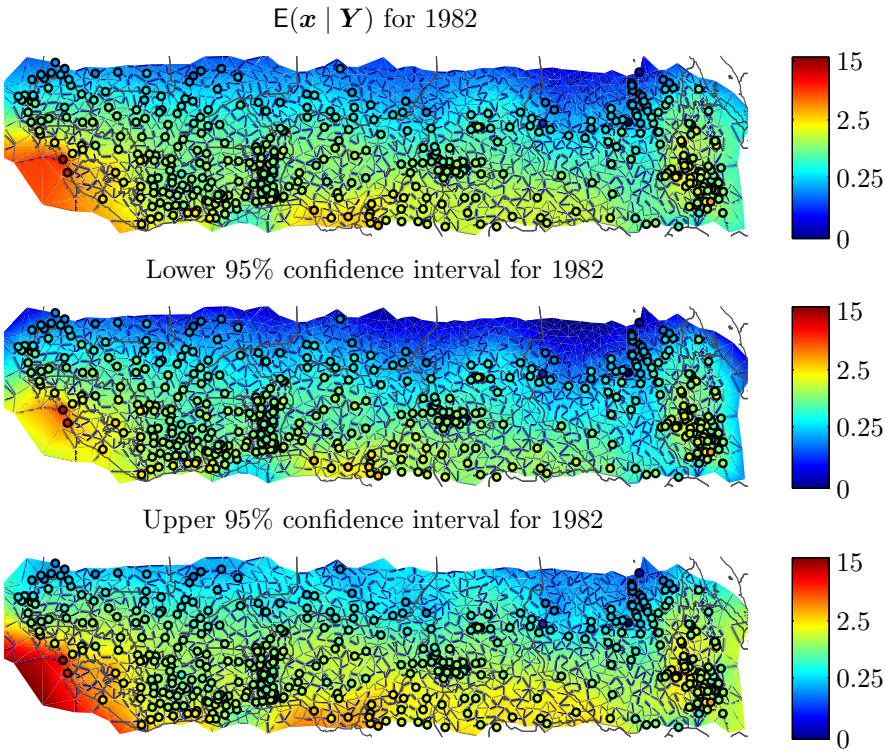

Figure 7: The top pane shows $\mathsf{E}(\boldsymbol{X} \mid \boldsymbol{Y})$ for the precipitation data from 1982 obtained from MCMC-simulations. The following two panes gives the boundaries of point-wise 95% confidence intervals of $(\boldsymbol{X} \mid \boldsymbol{Y})$ obtained as quantiles from the MCMC-simulations. The circles represent the measurement locations and corresponding measurement values.

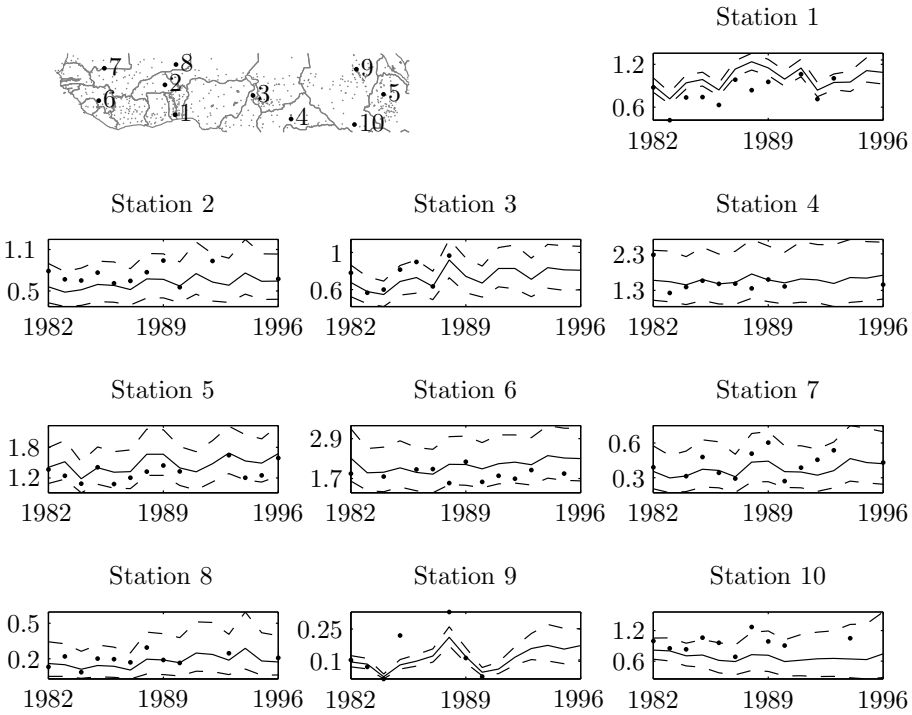

Figure 8: The top left pane shows the locations of the 10 stations omitted from the analysis. The remaining 10 panes show $\mathsf{E}(\boldsymbol{X} \mid \boldsymbol{Y})$ (solid lines), 95% confidence intervals (dashed lines) and actual measurements (dots) for each of the stations from 1982 to 1996. All the data has been back transformed to yearly precipitation in metres.

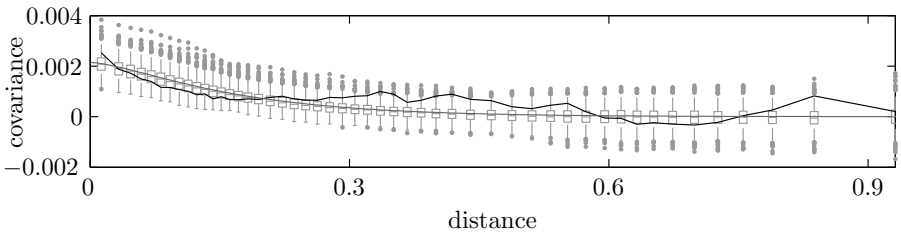

Figure 9: Covariance function for the residuals. The black line is the spatial, isotropic covariance function for the residuals (15), the box-plots describe the covariance functions estimated using a 1000 simulations from the model, the grey line is the theoretical covariance function given the posterior means of the parameters.

2039-by-2039 precision matrix in (2) takes 0.05s on a 2.8GHz Celeron. Extending the field to include temporal AR(1)-dependence for the 15 years of data calculation of the Cholesky factor of $\widehat{\boldsymbol{Q}}$ in (11c) takes 46.8s, and consequently each MCMC iteration takes roughly $60s$. As a comparison, the full spatio-temporal field has $2039 \cdot 15 = 30585$ nodes, if we instead study a pure spatial field with 30585 nodes calculation of the Cholesky factor takes only 2.48s, clearly illustrating the steep computational cost incurred by the temporal dependencies.

7 Conclusion and discussion

The use of GMRFs to model spatial and spatio-temporal dependencies allows for fast computations using algorithms for sparse matrices Rue (2001). This work uses the computational efficiency of GMRFs together with the possibility to approximate fields with Matérn covariances using GMRFs (Lindgren & Rue, 2007) to interpolate yearly precipitation over the African Sahel. The resulting model and method presented here allows for spatio-temporal estimation and interpolation of very large fields defined on irregular locations using reasonable computational resources.

Model assessment using measurements at stations left out of the analysis, as well as a comparison between the empirical and theoretical spatial covariance functions indicate that the model describes the data well and provides a good interpolation of the precipitation data. This interpolation can be used in future work examining the interactions between vegetation and precipitation in the region.

Another interesting extension would be to model monthly precipitation. Doing so would introduce a large number of zero measurements, especially during the dry winter months. To handle these zero measurements non-Gaussian observation model would have to be employed. Extending the model to use non-Gaussian observations will add a small computational burden. However, given the large literature of latent GMRFs with non-Gaussian observations (Knorr-Held & Rue, 2002; Hrafnkelsson & Cressie, 2003; Rue & Martino, 2007, etc.) an extension should not present any theoretical challenges. The computational burden incurred by using an AR(1)-dependence in time could possibly be lessened by the use of temporal basis functions.

A Determinant of the precision for an AR(1)-process

Given an AR(1)-process, $x_t = ax_{t-1} + \eta_t$, where η_t are i.i.d. $\mathsf{N}(0,1)$, $|a| < 1$ and x_1 belongs to the stationary distribution $\mathsf{N}\left(0, 1/(1-a^2)\right)$ the joint distribution for $\{x_t\}_{t=1\ldots T}$ is $\mathsf{N}\left(0, \boldsymbol{Q}_T^{-1}\right)$ where $\boldsymbol{Q}_T$ is the tri-diagonal T-by-T matrix from (6). To calculate the determinant, $|\boldsymbol{Q}_T|$, we first need the following lemma:

Lemma 1. *Let $\boldsymbol{A}_T$ be the tri-diagonal T-by-T matrix*

$$\boldsymbol{A}_T = \begin{pmatrix} 1+a^2 & -a & 0 & & & & \\ -a & 1+a^2 & -a & 0 & & & \\ & & \ddots & & & & \\ & 0 & -a & 1+a^2 & -a & 0 & \\ & & & & \ddots & & \\ & & & 0 & -a & 1+a^2 & -a \\ & & & & 0 & -a & 1 \end{pmatrix},$$

with $|a| < 1$. Further let $\boldsymbol{A}_1 = 1$ and

$$\boldsymbol{A}_2 = \begin{pmatrix} 1+a^2 & -a \\ -a & 1 \end{pmatrix}.$$

The determinant of $\boldsymbol{A}_T$ is $|\boldsymbol{A}_T| = 1$, for any $T \geq 1$

Proof. It is obvious that $|\boldsymbol{A}_1| = 1$ and $|\boldsymbol{A}_2| = 1$. For $T > 2$ expansion of the determinant along the first column followed by expansion of the second term along the first row gives

$$|\boldsymbol{A}_T| = (1+a^2)\,|\boldsymbol{A}_{T-1}| - a^2\,|\boldsymbol{A}_{T-2}|\,, \tag{16}$$

and Lemma 1 follows by induction. □

Proposition 2. *The determinant of $\boldsymbol{Q}_T$, defined by (6), is $1-a^2$ for any $T \geq 1$.*

Proof. For $T = 1$ and $T = 2$ the result is trivial. For $T > 2$ we expand the determinant along the first column followed by expansion of the second term along the first row which gives

$$|\boldsymbol{Q}_T| = |\boldsymbol{A}_{T-1}| - a^2\,|\boldsymbol{A}_{T-2}| = 1 - a^2,$$

where the last equality is obtained using Lemma 1. □

B Calculation of the posterior distribution

Expanding (9) the full posterior becomes

$$\begin{aligned}\mathrm{p}(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{Y}) \propto \left(\frac{|\boldsymbol{Q}_S|^T(1-a^2)^N}{|\sigma^2\boldsymbol{I}|}\right)^{1/2} \mathrm{p}(\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\psi}) \\ \exp\Big(-\frac{1}{2}\big(\boldsymbol{X}-\boldsymbol{1}\otimes\boldsymbol{B}\boldsymbol{\theta}\big)^\top\big(\boldsymbol{Q}_T\otimes\boldsymbol{Q}_S\big)\big(\boldsymbol{X}-\boldsymbol{1}\otimes\boldsymbol{B}\boldsymbol{\theta}\big)\Big) \\ \exp\Big(-\frac{1}{2\sigma^2}\big(\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{X}\big)^\top\big(\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{X}\big)\Big).\end{aligned} \tag{17}$$

Expanding the terms in the exponents and collecting terms containing $\boldsymbol{X}$ we obtain

$$\begin{aligned}-\frac{1}{2}\Bigg(\boldsymbol{X}^\top\Big(\underbrace{\boldsymbol{Q}_T\otimes\boldsymbol{Q}_S+\frac{\boldsymbol{A}^\top\boldsymbol{A}}{\sigma^2}}_{\widehat{\boldsymbol{Q}}}\Big)\boldsymbol{X}-2\boldsymbol{X}^\top\Big(\big(\boldsymbol{Q}_T\boldsymbol{1}\otimes\boldsymbol{Q}_S\boldsymbol{B}\big)\boldsymbol{\theta}+\underbrace{\frac{\boldsymbol{A}^\top\boldsymbol{Y}}{\sigma^2}}_{\boldsymbol{b}}\Big) \\ +\frac{\boldsymbol{Y}^\top\boldsymbol{Y}}{\sigma^2}+\boldsymbol{\theta}^\top\big(\boldsymbol{Q}_T\boldsymbol{1}\otimes\boldsymbol{Q}_S\boldsymbol{B}\big)^\top\big(\boldsymbol{1}\otimes\boldsymbol{B}\big)\boldsymbol{\theta}\Bigg).\end{aligned}$$

Identifying the first two terms as part of a Gaussian distribution for $\boldsymbol{X}$ we can rewrite (17) as

$$\begin{aligned}\mathrm{p}(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{Y}) \propto \mathrm{p}(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{Y})\mathrm{p}(\boldsymbol{\psi})\left(\frac{|\boldsymbol{Q}_S|^T(1-a^2)^N}{\left|\widehat{\boldsymbol{Q}}\right||\sigma^2\boldsymbol{I}|}\right)^{1/2} \\ \exp\Bigg(-\frac{1}{2}\Bigg(\frac{\boldsymbol{Y}^\top\boldsymbol{Y}}{\sigma^2}+\boldsymbol{\theta}^\top\big(\boldsymbol{Q}_T\boldsymbol{1}\otimes\boldsymbol{Q}_S\boldsymbol{B}\big)^\top\big(\boldsymbol{1}\otimes\boldsymbol{B}\big)\boldsymbol{\theta} \\ -\Big(\big(\boldsymbol{Q}_T\boldsymbol{1}\otimes\boldsymbol{Q}_S\boldsymbol{B}\big)\boldsymbol{\theta}+\boldsymbol{b}\Big)^\top\widehat{\boldsymbol{Q}}^{-1}\Big(\big(\boldsymbol{Q}_T\boldsymbol{1}\otimes\boldsymbol{Q}_S\boldsymbol{B}\big)\boldsymbol{\theta}+\boldsymbol{b}\Big) \\ +\big(\boldsymbol{\theta}-\boldsymbol{\mu}_\theta\big)^\top\boldsymbol{Q}_\theta\big(\boldsymbol{\theta}-\boldsymbol{\mu}_\theta\big)\Bigg)\Bigg),\end{aligned} \tag{18}$$

where $\mathrm{p}(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{Y})$ is the density function for the posterior $(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{Y})$ defined in (11a). Note that we have divided by the determinant of $\widehat{\boldsymbol{Q}}$ to cancel the corresponding determinant which arises in $\mathrm{p}(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{Y})$; the quadratic form involving $\widehat{\boldsymbol{Q}}^{-1}$ in the third row arise for the same reason.

We again study the exponent but this time we collect terms containing

$\boldsymbol{\theta}$, obtaining

$$
\begin{aligned}
-\frac{1}{2}\Bigg(&\boldsymbol{\theta}^\top\Bigg(\underbrace{\boldsymbol{Q}_\theta + (\boldsymbol{Q}_T\mathbf{1}\otimes\boldsymbol{Q}_S\boldsymbol{B})^\top\Big((\mathbf{1}\otimes\boldsymbol{B}) - \widehat{\boldsymbol{Q}}^{-1}(\boldsymbol{Q}_T\mathbf{1}\otimes\boldsymbol{Q}_S\boldsymbol{B})\Big)}_{\widehat{\boldsymbol{Q}}_\theta}\Bigg)\boldsymbol{\theta} \\
&- 2\boldsymbol{\theta}^\top\Big(\underbrace{\boldsymbol{\mu}_\theta\boldsymbol{Q}_\theta + (\boldsymbol{Q}_T\mathbf{1}\otimes\boldsymbol{Q}_S\boldsymbol{B})^\top\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{b}}_{\boldsymbol{b}_\theta}\Big) \\
&- \boldsymbol{b}^\top\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{b} + \frac{\boldsymbol{Y}^\top\boldsymbol{Y}}{\sigma^2} + \boldsymbol{\mu}_\theta^\top\boldsymbol{Q}_\theta\boldsymbol{\mu}_\theta\Bigg).
\end{aligned}
$$

We recognise the first two terms as belonging to the posterior $(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{Y})$ given in (11b). We can now obtain (11c) by rewriting (18) in a similar way as previously used when obtaining (18) from (17). Finally we rewrite $\widehat{\boldsymbol{Q}}_\theta$ as

$$
\begin{aligned}
\widehat{\boldsymbol{Q}}_\theta &= \boldsymbol{Q}_\theta + (\boldsymbol{Q}_T\mathbf{1}\otimes\boldsymbol{Q}_S\boldsymbol{B})^\top\Big(\widehat{\boldsymbol{Q}}^{-1}\widehat{\boldsymbol{Q}} - \widehat{\boldsymbol{Q}}^{-1}(\boldsymbol{Q}_T\otimes\boldsymbol{Q}_S)\Big)(\mathbf{1}\otimes\boldsymbol{B}) \\
&= \boldsymbol{Q}_\theta + (\boldsymbol{Q}_T\mathbf{1}\otimes\boldsymbol{Q}_S\boldsymbol{B})^\top\widehat{\boldsymbol{Q}}^{-1}\Big(\frac{\boldsymbol{A}^\top\boldsymbol{A}}{\sigma^2}\Big)(\mathbf{1}\otimes\boldsymbol{B}).
\end{aligned}
$$

C Parameter values

C.1 Hyper-parameters

The hyper-parameters of the priors in (12) are selected to make the priors wide, while at the same time making values close to zero unlikely. Avoiding zero values is reasonable since $\chi = 0$ implies a filed with infinite variance; and $\kappa^2 = 0$ gives infinite range in the spatial covariance function. Thus, we take the shape parameters to be

$$
\alpha_\sigma = 2, \qquad \alpha_\kappa = 2, \qquad \alpha_\chi = 1.5,
$$

and the scale parameters are set to

$$
\beta_\sigma = 500, \qquad \beta_\kappa = 0.0025, \qquad \beta_\chi = 5.
$$

C.2 Proposal distribution

The covariance of the proposal distribution (14) is set through the standard deviation of each parameter and the correlation between parameters, i.e.

$$
\boldsymbol{\Sigma}_{\text{prop}} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \rho_{14}\sigma_1\sigma_4 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \rho_{24}\sigma_2\sigma_4 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 & \rho_{34}\sigma_3\sigma_4 \\ \rho_{14}\sigma_1\sigma_4 & \rho_{24}\sigma_2\sigma_4 & \rho_{34}\sigma_3\sigma_4 & \sigma_4^2 \end{pmatrix}.
$$

A short initial MCMC run on a triangulation of coarser resolution and using data from 1982 to 1987 indicated reasonable values of the parameters to be $\sigma_1 = 0.15$, $\sigma_2 = 0.15$, $\sigma_3 = 0.03$, $\sigma_4 = 0.30$ and

$$\begin{aligned} \rho_{12} &= 0.6 & \rho_{13} &= 0.2 & \rho_{14} &= -0.4 \\ & & \rho_{23} &= 0.4 & \rho_{24} &= -0.6 \\ & & & & \rho_{34} &= -0.1. \end{aligned}$$

References

Agbu, P. & James, M. (1994). *The NOAA/NASA Pathfinder AVHRR land data set user's manual.* Goddard Distributed Active Archive Center, NASA, Goddard Space Flight Center, Greenbelt.

Bolin, D., Lindström, J., Eklundh, L. & Lindgren, F. (2008). Fast estimation of spatially dependent temporal vegetation trends using Gaussian Markov random fields. *Submitted to Comput. Statist. and Data Anal.* .

Box, G. & Cox, D. (1964). An analysis of transformations. *J. Roy. Statist. Soc. Ser. B* **26**, 211–252.

Diggle, P., Tawn, J. & Moyeed, R. (1998). Model-based geostatistics. *J. Roy. Statist. Soc. Ser. C* **47**, 299–350.

Eklundh, L. & Olsson, L. (2003). Vegetation index trends for the African Sahel 1982-1999. *Geophys. Res. Lett.* **30**, 1430–1433.

Fensholt, R., Sandholt, I. & Rasmussen, M. (2004). Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. *Remote Sensing of Environ.* **91**, 490–507.

Gamerman, D., Moreira, A. & Rue, H. (2003). Space-varying regression models: specifications and simulation. *Comput. Statist. and Data Anal.* **42**, 513–533.

Glasbey, C. & Nevison, I. (1997). Rainfall modelling using a latent Gaussian variable. *Lect. Notes Statist.* **122**, 233–242.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Hickler, T., Eklundh, L., Seaquist, J., Smith, B., Ardö, J., Olsson, L., Sykes, M. & Sjöström, M. (2005). Precipitation controls Sahel greening trend. *Geophys. Res. Lett.* **32**.

Hrafnkelsson, B. & Cressie, N. (2003). Hierarchical modeling of count data with application to nuclear fall-out. *Environ. and Ecological Statist.* **10**, 179–200.

Hulme, M. (2001). Climatic perspectives on Sahelian desiccation: 1973-1998. *Global Environ. Change* **11**, 19–29.

Knorr-Held, L. & Rue, H. (2002). On block updating in Markov random field models for disease mapping. *Scand. J. Statist.* **29**, 597–614.

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. Chem., Metal. and Mining Soc. South Africa* **52**, 119–139.

Lindgren, F. & Rue, H. (2007). Explicit construction of GMRF approximations to generalised Matérn fields on irregular grids. Tech. Rep. 12, Centre for Mathematical Sciences, Lund University, Lund, Sweden.

Los, S., Weedon, G., North, P., Kaduk, J., Taylor, C. & Cox, P. (2006). An observation-based estimate of the strength of rainfall-vegetation interactions in the Sahel. *Geophys. Res. Lett.* **33**.

Matheron, G. (1963). Principles of geostatistics. *Econom. Geol.* **58**, 1246–1266.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092.

NASA (2004). Global land cover characteristics data base.

Nicholson, S., Davenport, M. & Malo, A. (1990). A comparison of the vegetation response to rainfall in the Sahel and East Africa, using normalized difference vegetation index NOAA AVHRR. *Climate Change* **17**, 209–241.

NOAA (2007). The Global Historical Climatology Network (GHCN-monthly), ver. 2. Accessed 2007-12-03.

Olsson, L. (1993). On the causes of famine – drought, desertification and market failure in the Sudan. *Ambio* **22**, 395–403.

Olsson, L., Eklundh, L. & Ardö, J. (2005). A recent greening of the Sahel-trends, patterns and potential causes. *J. Arid Environ.* **63**, 556–566.

Prince, S., de Colstoun, E. & Kravitz, L. (1998). Evidence from rain-use efficiencies does not indicate extensive Sahelian desertification. *Global Change Biology* **4**, 359–374.

Rasmussen, M. (1992). Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from the AVHRR. *Internat. J. Remote Sensing* **13**, 3431–3442.

Rue, H. (2001). Fast sampling from Gaussian Markov random fields. *J. Roy. Statist. Soc. Ser. B* **63**, 325–338.

Rue, H. & Follestad, T. (2002). GMRFLib: a C-library for fast and exact simulation of Gaussian Markov random fields. Tech. Rep. 1, Department of Statistics, Norwegian University of Science and Technology, Trondheim, Norway.

Rue, H. & Held, L. (2005). *Gaussian Markov random fields; theory and applications*, vol. 104 of *Monographs on Statistics and Applied Probability.* Chapman & Hall/CRC.

Rue, H. & Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *J. Statist. Plann. and Inference* **137**, 3177–3192.

Rue, H. & Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scand. J. Statist.* **29**, 31–49.

Shewchuk, J. R. (1996). Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator. In M. C. Lin & D. Manocha, eds., *Applied computational geometry: Towards geometric engineering*, vol. 1148 of *Lecture Notes in Computer Science*. Springer-Verlag, pp. 203–222.

Tucker, C., Vanpraet, C., Sharman, M. & van Ittersum, G. (1985). Satellite remote sensing of total herbaceous biomass production in the Sengalese Sahel: 1980-1984. *Remote Sensing of Environ.* **17**, 233–249.

Vose, R., Schmoyer, R. L., Steurer, P. M., Peterson, T. C., Heim, R., Karl, T. R. & Eischeid, J. (1992). The Global Historical Climatology Network: long-term monthly temperature, precipitation, sea level pressure, and station pressure data. Tech. Rep. NDP-041, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, US.

Werner Hartman, L. (2006). Bayesian modelling of spatial data using Markov random fields, with application to elemental composition of forest soil. *Math. Geol.* **38**, 113–133.

Whittle, P. (1963). Stochastic processes in several dimensions. *Bull. Internat. Statist. Inst.* **40**, 974–994.

Wikle, C. K., Berliner, L. M. & Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environ. and Ecological Statist.* **5**, 117–154.

Preprints in Mathematical Sciences 2008:8
ISSN 1403-9338

LUTFMS-5074-2008

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/