

*Citation for published version:*

Roy, V, Evangelou, E & Zhu, Z 2015, Empirical Bayes methods for the transformed Gaussian random field model with additive measurement errors. in SK Upadhyay, U Singh, DK Dey & A Loganathan (eds), *Current Trends in Bayesian Methodology with Applications*. Chapman & Hall, pp. 521-536.  
<https://doi.org/10.1201/b18502-1>

*DOI:*

[10.1201/b18502-1](https://doi.org/10.1201/b18502-1)

*Publication date:*

2015

*Document Version*

Early version, also known as pre-print

[Link to publication](#)

Each copy containing our material must bear a credit line in the following format (insert information as appropriate: Copyright © 2015 From 'Current Trends in Bayesian Methodology with Applications', ed. by Satyanshu K. Upadhyay et al. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

This material is strictly for personal use only. For any other use, the user must contact Taylor & Francis directly at this address: & Francis directly at this address:

permissions.mailbox@taylorandfrancis.com. Printing, photocopying, sharing via any means is a violation of copyright.

Taylor & Francis stipulates that this material may not be printed or downloaded.

## University of Bath

### General rights

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.

### 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.

# Empirical Bayes methods for the transformed Gaussian random fields model with additive measurement errors

Vivekananda Roy<sup>1</sup>, Evangelos Evangelou<sup>2</sup> and Zhengyuan Zhu<sup>1</sup>

<sup>1</sup> Department of Statistics, Iowa State University, USA

<sup>2</sup> Department of Mathematical Sciences, University of Bath, UK

*Key words and phrases:* Box-Cox transformation, data augmentation, empirical Bayes, importance sampling, rainfall, spatial prediction, transformed Gaussian.

*Cite this document as:*

Roy, V., Evangelou, E. and Zhu, Z. (2015). Empirical Bayes methods for the transformed Gaussian random fields model with additive measurement errors. In Upadhyay, S. K., Singh, U., Dey, D. K., and Loganathan, A., editors, *Current Trends in Bayesian Methodology with Applications*, Boca Raton, FL, USA, CRC Press.

# 1 Introduction

If geostatistical observations are continuous but can not be modeled by the Gaussian distribution, a more appropriate model for these data may be the transformed Gaussian model. In transformed Gaussian models it is assumed that the random field of interest is a nonlinear transformation of a Gaussian random field (GRF). For example, De Oliveira et al. (1997) propose the Bayesian transformed Gaussian model where they use the Box-Cox family of power transformation (Box and Cox, 1964) on the observations. De Oliveira et al. (1997) show that prediction for unobserved random fields can be done through posterior predictive distribution where uncertainty about the transformation parameter is taken into account. More recently, Christensen et al. (2001) consider maximum likelihood estimation of the parameters and a “plug-in” method of prediction for transformed Gaussian model with Box-Cox family of transformations. Both De Oliveira et al. (1997) and Christensen et al. (2001) consider spatial prediction of rainfall to illustrate their model and method of analysis. A review of the Bayesian transformed Gaussian random fields model is given in De Oliveira et al. (2002). See also De Oliveira (2003) who discusses several issues regarding the formulation and interpretation of transformed Gaussian random field models, including the approximate nature of the model for positive data based on Box-Cox family of transformations, and the interpretation of the model parameters.

In their discussion, Christensen et al. (2001) mention that for analyzing rainfall data there “must be at least an additive measurement error” in the model. De Oliveira and Ecker (2002) consider a measurement error term for transformed data in their transformed Gaussian model. An alternative approach, as suggested by Christensen et al. (2001), is to assume measurement error in the original scale of the data, *not* in the transformed scale as done by De Oliveira and Ecker (2002). We propose a transformed Gaussian model where an additive measurement error term is used for the observed data, and the random fields, after a suitable transformation, is assumed to follow Gaussian distribution. In many practical situations this may be the more natural assumption. Unfortunately, the likelihood function is not available in closed form in this case and as mentioned by Christensen et al. (2001), this alternative model, although “is attractive”, it “raises technical complications”. In spite of the fact that the likelihood function is not available in closed form, we show that *data augmentation* techniques can be used for Markov chain Monte Carlo (MCMC) sampling from the target posterior density. These MCMC samples are then used for estimation of parameters of our proposed model as well as prediction of rainfall at new locations.

The Box-Cox family of transformations is defined for *strictly positive* observations. This implies that the transformed Gaussian random variables have restricted support and the corresponding likelihood function is not in closed form. Christensen et al. (2001) change the observed zeros in the data to a small positive number in order to have a closed form likelihood function. On the other hand, Stein

(1992) considers Monte Carlo methods for prediction and inference in a model where transformed observations are assumed to be a truncated Gaussian random field. In our proposed model we consider the transformation on random effects *instead of* observed data and consider a natural extension of the Box-Cox family of transformations for negative values of the random effect.

The so-called *full Bayesian analysis* of transformed Gaussian data requires specification of a joint prior distribution on the Gaussian random fields parameters as well as transformation parameters (see e.g. De Oliveira et al., 1997). Since a change in the transformation parameter value results in change of location and scale of transformed data, assuming GRF model parameters to be independent a priori of the transformation parameter would give nonsensical results (Box and Cox, 1964). Assigning an appropriate prior on covariance (of the transformed random field) parameters, like the range parameter, is also not easy as the choice of prior may influence the inference (see e.g. Christensen, 2004, p. 716). Use of improper prior on correlation parameters typically results in improper posterior distribution (Stein, 1999, p. 224). Thus it is difficult to specify a joint prior on all the model parameters of transformed GRF models. Here we consider an empirical Bayes (EB) approach for estimating the transformation parameter as well as the range parameter of our transformed Gaussian random field model. Our EB method avoids the difficulty of specifying a prior on the transformation parameter as well as the range parameter, which, as mentioned above, is problematic. In our EB method of analysis, we do not need to sample from the complicated nonstandard conditional distributions of these (transformation and range) parameters, which is required in the full Bayesian analysis. Further, an MCMC algorithm with updates on such parameters may not perform well in terms of mixing and convergence (see e.g. Christensen, 2004, p. 716). Recently, in some simulation studies in the context of spatial generalized linear mixed models for binomial data, Roy et al. (2014) observe that EB analysis results in estimates with less bias and variance than full Bayesian analysis. Roy (2014) uses an efficient importance sampling method based on MCMC sampling for estimating the link function parameter of a robust binary regression model (see also Doss, 2010). We use Roy's (2014) method for estimating the transformation and range parameters of our model.

The rest of the chapter is organized as follows. Section 2 introduces our transformed Gaussian model with measurement error. In Section 3 a method based on importance sampling is described for effectively selecting the transformation parameter as well as the range parameter. Section 4 discusses the computation of the Bayesian predictive density function. In Section 5 we analyze a data set using the proposed model and estimation procedure for constructing a continuous spatial map of rainfall amounts.

## 2 The transformed Gaussian model with measurement error

### 2.1 Model description

Let  $\{Z(s), s \in \mathcal{D}\}$ ,  $\mathcal{D} \in \mathbb{R}^l$  be the random field of interest. We observe a single realization from the random field with measurement errors at finite sampling locations  $s_1, \dots, s_n \in \mathcal{D}$ . Let  $\mathbf{y} = (y(s_1), \dots, y(s_n))$  be the observations. We assume that the observations are sampled according to the following model,

$$Y(s) = Z(s) + \epsilon(s),$$

where we assume that  $\{\epsilon(s), s \in \mathcal{D}\}$  is a process of mutually independent  $N(0, \tau^2)$  random variables, which is independent of  $Z(s)$ . The term  $\epsilon(s)$  can be interpreted as micro-scale variation, measurement error, or a combination of both. We assume that  $Z(s)$ , after a suitable transformation, follow a normal distribution. That is, for some family of transformations  $g_\lambda(\cdot)$ ,  $\{W(s) \equiv g_\lambda(Z(s)), s \in \mathcal{D}\}$  is assumed to be a Gaussian stochastic process with the mean function  $E(W(s)) = \sum_{j=1}^p f_j(s)\beta_j$ ;  $\beta = (\beta_1, \dots, \beta_p)' \in \mathbb{R}^p$  are the unknown regression parameters,  $f(s) = (f_1(s), \dots, f_p(s))$  are known location dependent covariates, and  $\text{cov}(W(s), W(u)) = \sigma^2 \rho_\theta(\|s - u\|)$ , where  $\|s - u\|$  denotes the Euclidean distance between  $s$  and  $u$ . Here,  $\theta$  is a vector of parameters which controls the range of correlation and the smoothness/roughness of the random field.

We consider  $\rho_\theta(\cdot)$  as a member of the Matérn family (Matérn, 1986)

$$\rho(u; \phi, \kappa) = \{2^{\kappa-1} \Gamma(\kappa)\}^{-1} (u/\phi)^\kappa K_\kappa(u/\phi),$$

where  $K_\kappa(\cdot)$  denotes the modified Bessel function of order  $\kappa$ . In this case  $\theta \equiv (\phi, \kappa)$ . This two-parameter family is very flexible in that the integer part of the parameter  $\kappa$  determines the number of times the process  $W(s)$  is mean square differentiable, that is,  $\kappa$  controls the smoothness of the underlying process, while the parameter  $\phi$  measures the scale (in units of distance) on which the correlation decays. We assume that  $\kappa$  is known and fixed and estimate  $\phi$  using our empirical Bayes approach.

A popular choice for  $g_\lambda(\cdot)$  is the Box-Cox family of power transformations (Box and Cox, 1964) indexed by  $\lambda$ , that is,

$$g_\lambda^0(z) = \begin{cases} \frac{z^\lambda - 1}{\lambda} & \text{if } \lambda > 0 \\ \log(z) & \text{if } \lambda = 0 \end{cases}. \quad (1)$$

Although the above transformation holds for  $\lambda < 0$ , in this chapter, we assume that  $\lambda \geq 0$ . As mentioned in the Introduction, the Box-Cox transformation (1) holds for  $z > 0$ . This implies that the image of the transformation is  $(-1/\lambda, \infty)$ , which contradicts the Gaussian assumption. Also the normalizing

constant for the pdf of the transformed variable is not available in closed form. Note that the inverse transformation of (1) is given by

$$h_\lambda^0(w) = \begin{cases} (1 + \lambda w)^{\frac{1}{\lambda}} & \text{if } \lambda > 0 \\ \exp(w) & \text{if } \lambda = 0 \end{cases}. \quad (2)$$

The transformation (2) can be extended naturally to the whole real line to

$$h_\lambda(w) = \begin{cases} \text{sgn}(1 + \lambda w)|1 + \lambda w|^{\frac{1}{\lambda}} & \text{if } \lambda > 0 \\ \exp(w) & \text{if } \lambda = 0, \end{cases} \quad (3)$$

where  $\text{sgn}(x)$  denotes the sign of  $x$ , taking values  $-1$ ,  $0$ , or  $1$  depending on whether  $x$  is negative, zero, or positive respectively. The proposed transformation (3) is monotone, invertible and is continuous at  $\lambda = 0$  with the inverse, the *extended Box-Cox transformation* given in (Bickel and Doksum, 1981)

$$g_\lambda(z) = \begin{cases} \frac{(\text{sgn}(z)|z|^\lambda - 1)}{\lambda} & \text{if } \lambda > 0 \\ \log(z) & \text{if } \lambda = 0 \end{cases}.$$

Let  $\mathbf{w} = (w(s_1), \dots, w(s_n))^T$ ,  $\mathbf{z} = (z(s_1), \dots, z(s_n))^T$ ,  $\gamma \equiv (\lambda, \phi)$ , and  $\psi \equiv (\beta, \sigma^2, \tau^2)$ . The reason for using different notation for  $(\lambda, \phi)$  than other parameters will be clear later. Since  $\{W(s) \equiv g_\lambda(Z(s)), s \in \mathcal{D}\}$  is assumed to be a Gaussian process, the joint posterior density of  $(\mathbf{y}, \mathbf{w})$  is given by

$$\begin{aligned} f(\mathbf{y}, \mathbf{w} | \psi, \gamma) &= (2\pi)^{-n} (\sigma\tau)^{-n} \exp\left\{-\frac{1}{2\tau^2} \sum_{i=1}^n (y_i - h_\lambda(w_i))^2\right\} |R_\theta|^{-1/2} \\ &\quad \exp\left\{-\frac{1}{2\sigma^2} (\mathbf{w} - F\beta)^T R_\theta^{-1} (\mathbf{w} - F\beta)\right\}, \end{aligned} \quad (4)$$

where  $\mathbf{y}, \mathbf{w} \in \mathbb{R}^n$ ,  $F$  is the known  $n \times p$  matrix defined by  $F_{ij} = f_j(s_i)$ ,  $R_\theta \equiv H_\theta(s, s)$  is the correlation matrix with  $R_{\theta,ij} = \rho_\theta(\|s_i - s_j\|)$ , and  $h_\lambda(\cdot)$  is defined in (3). The likelihood function for  $(\psi, \gamma)$  based on the observed data  $\mathbf{y}$  is given by

$$L(\psi, \gamma | \mathbf{y}) = \int_{\mathbb{R}^n} f(\mathbf{y}, \mathbf{w} | \psi, \gamma) d\mathbf{w}. \quad (5)$$

## 2.2 Posterior density and MCMC

The likelihood function  $L(\psi, \gamma | \mathbf{y})$  defined in (5) is not available in closed form. A full Bayesian analysis requires specification of a joint prior distribution on the model parameters  $(\psi, \gamma)$ . As mentioned before, assigning a joint prior distribution is difficult in this problem. Here, we estimate  $\gamma$  using the method described in Section 3. For other model parameters  $\psi$ , we assume the following conjugate priors,

$$\beta | \sigma^2 \sim \mathbf{N}_p(m_b, \sigma^2 V_b), \quad \sigma^2 \sim \chi_{ScI}^2(n_\sigma, a_\sigma), \quad \text{and} \quad \tau^2 \sim \chi_{ScI}^2(n_\tau, a_\tau), \quad (6)$$

where  $m_b, V_b, n_\sigma, a_\sigma, n_\tau, a_\tau$  are assumed to be known hyperparameters. (We say  $W \sim \chi_{ScI}^2(n_\sigma, a_\sigma)$  if the pdf of  $W$  is  $f(w) \propto w^{-(n_\sigma/2+1)} \exp(-n_\sigma a_\sigma/(2w))$ .) The posterior density of  $\psi$  is given by

$$\pi_\gamma(\psi|\mathbf{y}) = \frac{L_\gamma(\psi|\mathbf{y})\pi(\psi)}{m_\gamma(\mathbf{y})}, \quad (7)$$

where  $L_\gamma(\psi|\mathbf{y}) \equiv L(\psi, \gamma|\mathbf{y})$  is the likelihood function, and  $m_\gamma(\mathbf{y}) = \int_\Omega L_\gamma(\psi|\mathbf{y})\pi(\psi)d\psi$  is the normalizing constant, with  $\pi(\psi)$  being the prior on  $\psi$  and the support  $\Omega = \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}_+$ . Since the likelihood function (5) is not available in closed form, it is difficult to obtain MCMC sample from the posterior distribution  $\pi_\gamma(\psi|\mathbf{y})$  directly using the expression in (7).

Here we consider the following so-called *complete* posterior density

$$\pi_\gamma(\psi, \mathbf{w}|\mathbf{y}) = \frac{f(\mathbf{y}, \mathbf{w}|\psi, \gamma)\pi(\psi)}{m_\gamma(\mathbf{y})},$$

based on the joint density  $f(\mathbf{y}, \mathbf{w}|\psi, \gamma)$  defined in (4). Note that, integrating the complete posterior density  $\pi_\gamma(\psi, \mathbf{w}|\mathbf{y})$  we get the target posterior density  $\pi_\gamma(\psi|\mathbf{y})$ , that is,

$$\int_{\mathbb{R}^n} \pi_\gamma(\psi, \mathbf{w}|\mathbf{y})d\mathbf{w} = \pi_\gamma(\psi|\mathbf{y}).$$

So if we can generate a Markov chain  $\{\psi^{(i)}, \mathbf{w}^{(i)}\}_{i=1}^N$  with stationary density  $\pi_\gamma(\psi, \mathbf{w}|\mathbf{y})$ , then the marginal chain  $\{\psi^{(i)}\}_{i=1}^N$  has the stationary density  $\pi_\gamma(\psi|\mathbf{y})$  defined in (7). This is the standard technique of data augmentation and here  $\mathbf{w}$  is playing the role of ‘‘latent’’ variables (or ‘‘missing data’’) (Tanner and Wong, 1987).

Since we are using conjugate priors for  $(\beta, \sigma^2)$  in (6), integrating  $\pi_\gamma(\psi, \mathbf{w}|\mathbf{y})$  with respect  $\beta$  we have

$$\begin{aligned} \pi_\gamma(\sigma^2, \tau^2, \mathbf{w}|\mathbf{y}) &\propto (\sigma\tau)^{-n} \exp\left\{-\frac{1}{2\tau^2} \sum_{i=1}^n (y_i - h_\lambda(w_i))^2\right\} \tau^{-(n_\tau+2)} \exp(-n_\tau a_\tau/2\tau^2) |\Lambda_\theta|^{-1/2} \\ &\exp\left\{-\frac{1}{2\sigma^2} (\mathbf{w} - Fm_b)^T \Lambda_\theta^{-1} (\mathbf{w} - Fm_b)\right\} \sigma^{-(n_\sigma+2)} \exp(-n_\sigma a_\sigma/2\sigma^2), \quad (8) \end{aligned}$$

where  $\Lambda_\theta = FV_bF^T + R_\theta$ .

We use a Metropolis within Gibbs algorithm for sampling from the posterior density  $\pi_\gamma(\sigma^2, \tau^2, \mathbf{w}|\mathbf{y})$  given in (8). Note that

$$\sigma^2|\tau^2, \mathbf{w}, \mathbf{y} \sim \chi_{ScI}^2(n'_\sigma, a'_\sigma), \text{ and } \tau^2|\sigma^2, \mathbf{w}, \mathbf{y} \sim \chi_{ScI}^2(n'_\tau, a'_\tau),$$

where  $n'_\sigma = n + n_\sigma$ ,  $a'_\sigma = (n_\sigma a_\sigma + (\mathbf{w} - Fm_b)^T \Lambda_\theta^{-1} (\mathbf{w} - Fm_b))/(n + n_\sigma)$ ,  $n'_\tau = n + n_\tau$ , and  $a'_\tau = (n_\tau a_\tau + \sum_{i=1}^n (y_i - h_\lambda(w_i))^2)/(n + n_\tau)$ . On the other hand, the conditional distribution of  $\mathbf{w}$  given  $\sigma^2, \tau^2, \mathbf{y}$  is not a standard distribution. We use a Metropolis-Hastings algorithm given in Zhang (2002) for sampling from this conditional distribution.

### 3 Estimation of transformation and correlation parameters

Here we consider an empirical Bayes approach for estimating the transformation parameter  $\lambda$  and the range parameter  $\phi$ . That is, we select that value of  $\gamma \equiv (\lambda, \phi)$  which maximizes the marginal likelihood of the data  $m_\gamma(\mathbf{y})$ . For selecting models that are better than other models when  $\gamma$  varies across some set  $\Gamma$ , we calculate and subsequently compare the values of  $B_{\gamma, \gamma_1} := m_\gamma(\mathbf{y})/m_{\gamma_1}(\mathbf{y})$ , where  $\gamma_1$  is a suitably chosen fixed value of  $(\lambda, \phi)$ . Ideally, we would like to calculate and compare  $B_{\gamma, \gamma_1}$  for a large number of values of  $\gamma$ . Roy (2014) used a method based on importance sampling for selecting link function parameter in a robust regression model for binary data by estimating a large family of Bayes factors. Here we apply Roy's (2014) method to efficiently estimate  $B_{\gamma, \gamma_1}$  for a large set of possible values of  $\gamma$ . Recently Roy et al. (2014) successfully used this method for estimating parameters in spatial generalized linear mixed models.

Let  $f(\mathbf{y}, \mathbf{w}|\gamma) \equiv \int_{\Omega} f(\mathbf{y}, \mathbf{w}|\psi, \gamma)\pi(\psi)d\psi$ . Since we are conjugate priors for  $(\beta, \sigma^2)$  and  $\tau^2$  in (6), the marginal density  $f(\mathbf{y}, \mathbf{w}|\gamma)$  is available in closed form. In fact, from standard Bayesian analysis of normal linear model we have

$$f(\mathbf{y}, \mathbf{w}|\gamma) \propto \{a_\tau n_\tau + (\mathbf{y} - h_\lambda(\mathbf{w}))^T (\mathbf{y} - h_\lambda(\mathbf{w}))\}^{-\frac{n_\tau+n}{2}} |\Lambda_\theta|^{-1/2} \{a_\sigma n_\sigma + (\mathbf{w} - Fm_b)^T \Lambda_\theta^{-1} (\mathbf{w} - Fm_b)\}^{-\frac{n_\sigma+n}{2}}. \quad (9)$$

Note that

$$m_\gamma(\mathbf{y}) = \int_{\mathbb{R}^n} f(\mathbf{y}, \mathbf{w}|\gamma)d\mathbf{w}.$$

Let  $\{(\sigma^2, \tau^2)^{(l)}, \mathbf{w}^{(l)}\}_{l=1}^N$  be the Markov chain (with stationary density  $\pi_{\gamma_1}(\sigma^2, \tau^2, \mathbf{w}|\mathbf{y})$ ) underlying the MCMC algorithm presented in Section 2.2. Then by ergodic theorem we have a simple consistent estimator of  $B_{\gamma, \gamma_1}$ ,

$$\frac{1}{N} \sum_{i=1}^N \frac{f(\mathbf{y}, \mathbf{w}^{(i)}|\gamma)}{f(\mathbf{y}, \mathbf{w}^{(i)}|\gamma_1)} \xrightarrow{a.s.} \int_{\mathbb{R}^n} \frac{f(\mathbf{y}, \mathbf{w}|\gamma)}{f(\mathbf{y}, \mathbf{w}|\gamma_1)} \pi_{\gamma_1}(\mathbf{w}|\mathbf{y})d\mathbf{w} = \frac{m_\gamma(\mathbf{y})}{m_{\gamma_1}(\mathbf{y})}, \quad (10)$$

as  $N \rightarrow \infty$ , where  $\pi_{\gamma_1}(\mathbf{w}|\mathbf{y}) = \int_{\Omega} \pi_{\gamma_1}(\psi, \mathbf{w}|\mathbf{y})d\psi$ . Note that in (10) a single Markov chain  $\{\mathbf{w}^{(l)}\}_{l=1}^N$  with stationary density  $\pi_{\gamma_1}(\mathbf{w}|\mathbf{y})$  is used to estimate  $B_{\gamma, \gamma_1}$  for different values of  $\gamma$ . As mentioned in Roy (2014) the estimator (10) can be unstable and following Roy (2014) we consider the following method for estimating  $B_{\gamma, \gamma_1}$ .

Let  $\gamma_1, \gamma_2, \dots, \gamma_k \in \Gamma$  be  $k$  appropriately chosen skeleton points. Let  $\{\psi_j^{(l)}, \mathbf{w}^{(j;l)}\}_{l=1}^{N_j}$  be a Markov chain with stationary density  $\pi_{\gamma_j}(\psi, \mathbf{w}|\mathbf{y})$  for  $j = 1, \dots, k$ . Define  $r_i = m_{\gamma_i}(\mathbf{y})/m_{\gamma_1}(\mathbf{y})$  for  $i =$



2, 3, ..., k, with  $r_1 = 1$ . Then  $B_{\gamma, \gamma_1}$  is consistently estimated by

$$\hat{B}_{\gamma, \gamma_1} = \sum_{j=1}^k \sum_{l=1}^{N_j} \frac{f(\mathbf{y}, \mathbf{w}^{(j;l)} | \gamma)}{\sum_{i=1}^k N_i f(\mathbf{y}, \mathbf{w}^{(j;l)} | \gamma_i) / \hat{r}_i}, \quad (11)$$

where  $\hat{r}_1 = 1$ ,  $\hat{r}_i$ ,  $i = 2, 3, \dots, k$  are consistent estimator of  $r_i$ 's obtained by the ‘‘reverse logistic regression’’ method proposed by Geyer (1994). (See Roy (2014) for details about the above method of estimation and how to choose the skeleton point  $\gamma_i$ 's and sample size  $N_i$ 's.) The estimate of  $\gamma$  is obtained by maximizing (11), that is,  $\hat{\gamma} = \arg \max_{\gamma \in \Gamma} \hat{B}_{\gamma, \gamma_1}$ .

## 4 Spatial prediction

We now discuss how we make prediction about  $Z_0$ , the values of  $Z(s)$  at some locations of interest, say  $(s_{01}, s_{02}, \dots, s_{0k})$ , typically a fine grid of locations covering the observed region. We use the posterior predictive distribution

$$f(\mathbf{z}_0 | \mathbf{y}) = \int_{\Omega} \int_{\mathbb{R}^n} f_{\gamma}(\mathbf{z}_0 | \mathbf{w}, \psi) \pi_{\gamma}(\psi, \mathbf{w} | \mathbf{y}) d\mathbf{w} d\psi, \quad (12)$$

where  $\mathbf{z}_0 = (z(s_{01}), z(s_{02}), \dots, z(s_{0k}))$ . Let  $\mathbf{w}_0 = g_{\lambda}(\mathbf{z}_0) = (g_{\lambda}(z(s_{01})), g_{\lambda}(z(s_{02})), \dots, g_{\lambda}(z(s_{0k})))$ .

From Section 2.1, it follows that

$$(\mathbf{w}_0, \mathbf{w} | \psi) \sim N_{k+n} \left( \begin{pmatrix} F_0 \beta \\ F \beta \end{pmatrix}, \sigma^2 \begin{pmatrix} H_{\theta}(s_0, s_0) & H_{\theta}(s_0, s) \\ H_{\theta}^T(s_0, s) & H_{\theta}(s, s) \end{pmatrix} \right),$$

where  $F_0$  is the  $k \times p$  matrix with  $F_{0ij} = f_j(s_{0i})$ , and  $H_{\theta}(s_0, s)$  is the  $k \times n$  matrix with  $H_{\theta, ij}(s_0, s) = \rho_{\theta}(\|s_{0i} - s_j\|)$ . So  $\mathbf{w}_0 | \mathbf{w}, \psi \sim N_k(c_{\gamma}(\mathbf{w}, \psi), \sigma^2 D_{\gamma}(\mathbf{w}))$  where

$$c_{\gamma}(\mathbf{w}, \psi) = F_0 \beta + H_{\theta}(s_0, s) H_{\theta}^{-1}(s, s) (\mathbf{w} - F \beta), \text{ and } D_{\gamma}(\psi) = H_{\theta}(s_0, s_0) - H_{\theta}(s_0, s) H_{\theta}^{-1}(s, s) H_{\theta}^T(s_0, s).$$

Suppose, we want to estimate  $E(t(\mathbf{z}_0) | \mathbf{y})$  for some function  $t$ . Let  $\{\psi^{(i)}, \mathbf{w}^{(i)}\}_{i=1}^N$  be a Markov chain with stationary density  $\pi_{\gamma}(\psi, \mathbf{w} | \mathbf{y})$ , where  $\hat{\gamma}$  is the estimate of  $\gamma$  obtained using the method described in Section 3. We then simulate  $\mathbf{w}_0^{(i)}$  from  $f(\mathbf{w}_0 | \mathbf{w}^{(i)}, \psi^{(i)})$  for  $i = 1, \dots, N$ . Finally, we calculate the following approximate minimum mean squared error predictor

$$E(t(\mathbf{z}_0) | \mathbf{y}) \approx \frac{1}{N} \sum_{i=1}^N t(h_{\hat{\lambda}}(\mathbf{w}_0^{(i)})).$$

We can also estimate the predictive density (12) using these samples  $\{\mathbf{w}_0^{(i)}\}_{i=1}^N$ . In particular, we can estimate the quantiles of the predictive distribution of  $t(\mathbf{z}_0)$ .

## 5 Example: Swiss rainfall data

To illustrate our model and method of analysis we apply it to a well-known example. This dataset consists of the rainfall measurements that occurred on May 8, 1986 at the 467 locations in Switzerland shown in Figure 1. This dataset is available in the `geoR` package (Ribeiro and Diggle, 2012) in R and it has been analyzed before using a transformed Gaussian model by Christensen et al. (2001) and Diggle et al. (2003). The scientific objective is to construct a continuous spatial map of the average rainfall using the observed data as well as predict the proportion over the total area that the amount of rainfall exceeds a given level. The original data range from 0.5 to 585 but for our analysis these values are scaled by their geometric mean (139.663). Scaling the data helps avoid numerical overflow when computing the likelihood when the  $w$ 's are simulated from different  $\lambda$ . Following Christensen et al. (2001) we use

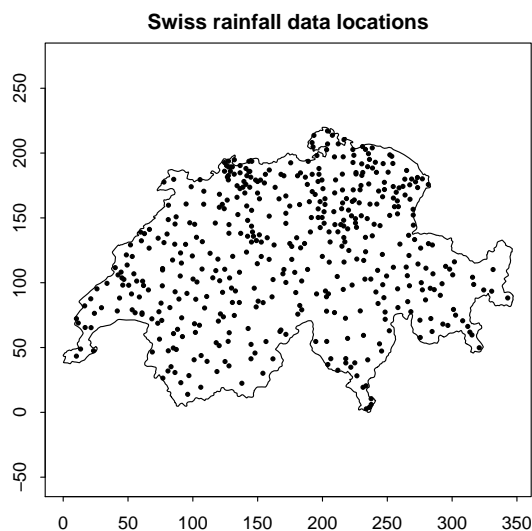


Figure 1: Sampled locations for the rainfall example.

the Matérn covariance and a constant mean  $\beta$  in our model for analyzing the rainfall data. Christensen et al. (2001) mention that  $\kappa = 1$  “gives a better fit than  $\kappa = 0.5$  or  $\kappa = 2$ ” (see also Diggle et al., 2003). We also use  $\kappa = 1$  in our analysis. We estimate  $(\lambda, \phi)$  using the method proposed in Section 3. In particular, we use the estimator (11) for estimating  $B_{\gamma, \gamma_1}$ . For the skeleton points we take all pairs of values of  $(\lambda, \phi)$ , where

$$\lambda \in \{0, 0.10, 0.20, 0.25, 0.33, 0.50, 0.67, 0.75, 1\} \text{ and } \phi \in \{10, 15, 20, 30, 50\}. \quad (13)$$

The first combination  $(0, 10)$  is taken as the baseline point  $\gamma_1$ . MCMC samples of size 3,000 at each of the 45 skeleton points are used to estimate the Bayes factors  $r_i$ 's at the skeleton points using the reverse

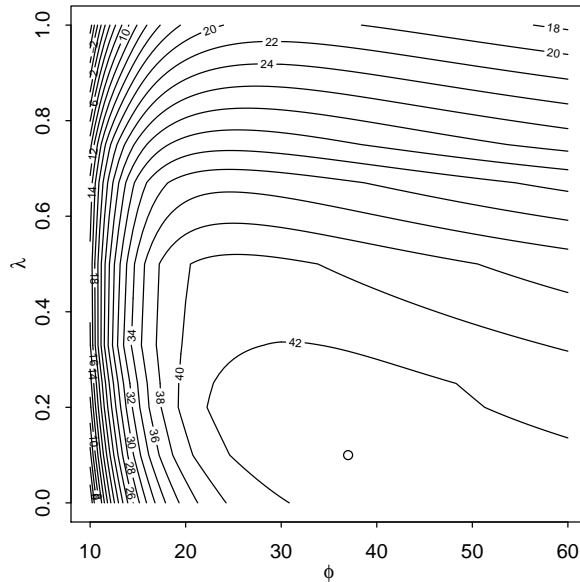


Figure 2: Contour plot of estimates of  $B_{\gamma, \gamma_1}$  for the rainfall dataset. The plot suggests  $\hat{\lambda} = 0.1$  and  $\hat{\phi} = 37$ . Here the baseline value corresponds to  $\lambda_1 = 0$  and  $\phi_1 = 10$ .

logistic regression method. Here we use the Metropolis within Gibbs algorithm mentioned in Section 2.2 for obtaining MCMC samples. These samples are taken after discarding an initial burn-in of 500 samples and keeping every 10th draw of subsequent random samples. Since the whole  $n$ -dimensional vectors  $\mathbf{w}^{(i)}$ 's must be stored, it is advantageous to make thinning of the MCMC sample such that saved values of  $\mathbf{w}^{(i)}$  are approximately uncorrelated (see e.g. Christensen, 2004, p. 706). Next we use *new* MCMC samples of size 500 corresponding to the 45 skeleton points mentioned in (13) to compute the Bayes factors  $B_{\gamma, \gamma_1}$  at other points. The estimate  $(\hat{\lambda}, \hat{\phi})$  is taken to be the value of  $(\lambda, \phi)$  where  $\hat{B}_{\gamma, \gamma_1}$  attains its maximum. Here again we collect every 10th sample after initial burn-in of 500 samples. For the entire computation it took about 70 minutes on a computer with 2.8 GHz 64-bit Intel Xeon processor and 2 Gb RAM. The computation was done using Fortran 95. Figure 2 shows the contour plot of the Bayes factor estimates. From the plot we see that  $\hat{B}_{\gamma, \gamma_1}$  attains maximum at  $\hat{\gamma} = (\hat{\lambda}, \hat{\phi}) = (0.1, 37)$ . The Bayes factors for a selection of fixed  $\lambda$  and  $\phi$  values is also shown in Figure 3.

Next, we fix  $\lambda$  and  $\phi$  at their estimates and estimate  $\beta$ ,  $\sigma^2$  and  $\tau^2$ , as well as the random field  $\mathbf{z}$  at the observed and prediction locations. The prediction grid consists of a square grid of length and width equal to 5 kilometers. The prior hyperparameters were as follows: prior mean for  $\beta$ ,  $m_b = 0$ , prior variance for  $\beta$ ,  $V_b = 100$ , degrees of freedom parameter for  $\sigma^2$ ,  $n_\sigma = 1$ , scale parameter for  $\sigma^2$ ,  $a_\sigma = 1$ , degrees of freedom parameter for  $\tau^2$ ,  $n_\tau = 1$ , and scale parameter for  $\tau^2$ ,  $a_\tau = 1$ . A MCMC sample of

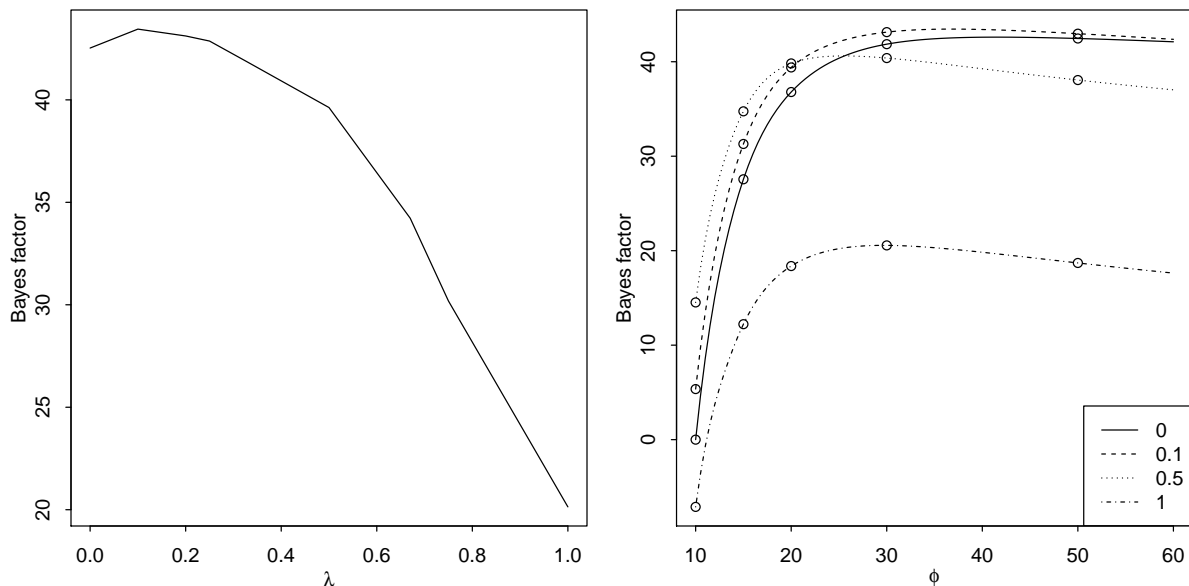


Figure 3: Estimates of  $B_{\gamma, \gamma_1}$  against  $\lambda$  for fixed value of  $\phi = 37$  (left panel) and against  $\phi$  for fixed values of  $\lambda$  (right panel). The circles show some of the skeleton points.

size 3,000 is used for parameter estimation and prediction. Like before we discard initial 500 samples as burn-in and collected every 10th sample. Let  $\{\sigma^{2(i)}, \tau^{2(i)}, \mathbf{w}^{(i)}\}_{i=1}^N$  be the MCMC samples (with invariant density  $\pi_{\hat{\gamma}}(\sigma^2, \tau^2, \mathbf{w}|\mathbf{y})$ ) obtained using the Metropolis within Gibbs algorithm mentioned in Section 2.2. Then we simulate  $\beta^{(i)}$  from its full conditional density  $\pi_{\hat{\gamma}}(\beta|\sigma^{2(i)}, \tau^{2(i)}, \mathbf{w}^{(i)}, \mathbf{y})$ , which is a normal density, to obtain MCMC samples for  $\beta$ . This part of the algorithm took no more than 2 minutes to run on the same computer. The estimates of posterior means of the parameter are given in Table 1. The standard errors of the MCMC estimators are computed using the method of overlapping batch means (Geyer, 2011) and also given in Table 1. Predictions of  $Z(s)$  and the corresponding prediction standard deviations are presented in Figure 4. Note that for the prediction, the MCMC sample is scaled back to the original scale of the data.

	$\beta$	$\sigma^2$	$\tau^2$
Estimate	-0.23	0.74	0.05
St Error	0.00483	0.00701	0.00032

Table 1: Posterior estimates of model parameters.

Using the model discussed in Christensen et al. (2001) fitted to the scaled data, the maximum likelihood estimates of the parameters for fixed  $\kappa = 1$  and  $\lambda = 0.5$  are  $\hat{\beta} = -0.13$ ,  $\hat{\sigma}^2 = 0.75$ ,  $\hat{\tau}^2 = 0.05$ ,

and  $\hat{\phi} = 35.8$ . These are not very different from our estimates although we emphasize that the interpretation of  $\tau^2$  in our model is different. We use the `krige.conv` function (used also by Christensen et al. (2001), personal communication) in the `geoR` package (Ribeiro and Diggle, 2012) in R to reproduce the prediction map of Christensen et al.’s (2001) and is given in Figure 5. From Figure 5 we see that the prediction map is similar to the map in Figure 4 obtained using our model. Note that, Figure 4 is prediction map for  $Z(\cdot)$  whereas Figure 5 is prediction map for  $Y$  as done in Christensen et al. (2001). On the other hand, if the parameter nugget  $\tau^2$  in Christensen et al.’s (2001) model is interpreted as measurement error (in the transformed scale), it is not obvious how to define the target of prediction, that is, the signal part of the transformed Gaussian variables without noise when  $g_\lambda(\cdot)$  is *not* the identity link function (Christensen, 2004, p. 710). (See also De Oliveira and Ecker (2002) who consider prediction for transformed Gaussian random fields where the parameter  $\tau^2$  is interpreted as measurement error.) When adding the error variance term to the map of the prediction variance for our model we get a similar pattern as the prediction variance plot corresponding to Christensen et al.’s (2001) model with slightly larger values. The difference in the variance is expected since our Bayesian model accounts for the uncertainty in the model parameters while the plug-in prediction in Christensen et al. (2001) does not.

Next, we consider a cross validation study to compare the performance of our model and Christensen et al.’s (2001) model. We remove 15 randomly chosen observations and predict these values using the remaining 452 observations. We repeat this procedure 31 times, each time removing 15 randomly chosen observations and predicting them with the remaining 452 data. For both our model as well as Christensen et al.’s (2001) model, we keep  $\lambda$  and  $\phi$  parameters fixed at their estimates when all data are observed. The average (over all  $15 \times 31$  deleted observations) root mean squared error (RMSE) for our model is 7.55 and for Christensen et al.’s (2001) model is 7.48. We use 2000 samples from the predictive distribution (posterior predictive distribution in the case of our model) in order to estimate RMSE at each location. We also compute the proportion of these samples that fall below the observed (deleted) value at each of the  $15 \times 31$  locations. These proportions are subtracted from 0.5 and the average of their absolute values across all locations for our model is 0.239 and for Christensen et al.’s (2001) model is 0.238. Lastly, we compute the proportions of one-sided prediction intervals that capture the observed (deleted) value. That is, we estimate the prediction intervals of the form  $(-\infty, z_{0\alpha})$ , where  $z_{0\alpha}$  corresponds to the  $\alpha$ th quantile of the predictive distribution. Table 2 shows the coverage probability of prediction intervals for different  $\alpha$  values corresponding to our model and Christensen et al.’s (2001) model. From Table 2 we see that the coverage probabilities of prediction intervals for the two models are similar.

0.010	0.017	0.019
0.025	0.034	0.030
0.050	0.045	0.054
0.100	0.067	0.080
0.500	0.510	0.488
0.900	0.899	0.897
0.950	0.942	0.946
0.975	0.978	0.976
0.990	0.987	0.985

Table 2: Coverage probability of one-sided prediction intervals  $(-\infty, z_{0\alpha})$  for different values of  $\alpha$  (first column) corresponding to our model (second column) and Christensen et al.’s (2001) model (third column).

Finally, as mentioned in Christensen et al. (2001), the relative area where  $Z(s) \geq c$  for some constant  $c$  is of practical significance. The proportion of locations that exceed the level 200 is computed using

$$\hat{E}[I(s \in \tilde{A}, Z(s) \geq 200)]/\#\tilde{A} = \frac{1}{N} \sum_{i=1}^N \#\{s \in \tilde{A}, h_{\hat{\lambda}}(W^{(i)}(s)) \geq 200\}/\#\tilde{A},$$

where  $I(\cdot)$  is the indicator function,  $\tilde{A}$  is the square grid of length and width equal to 5 kilometers and  $\{W^{(i)}(s)\}_{i=1}^N$  is the posterior predictive sample as described in Section 4. The histogram of samples of these proportions is shown in Figure 6.

## 6 Discussion

For Gaussian geostatistical models, estimation of unknown parameters as well as minimum mean squared error prediction at unobserved location can be done in closed form. On the other hand, many datasets in practice show non-Gaussian behavior. Certain types of non-Gaussian random fields data may be adequately modeled by transformed Gaussian models. In this chapter, we present a flexible transformed Gaussian model where an additive measurement error as well as a component representing smooth spatial variation is considered. Since specifying a joint prior distribution for all model parameters is difficult, we consider an empirical Bayes method here. We propose an efficient importance sampling method based on MCMC sampling for estimating the transformation and range parameters of our model. Although, we consider an extended Box-Cox transformation in our model, other types of transforma-

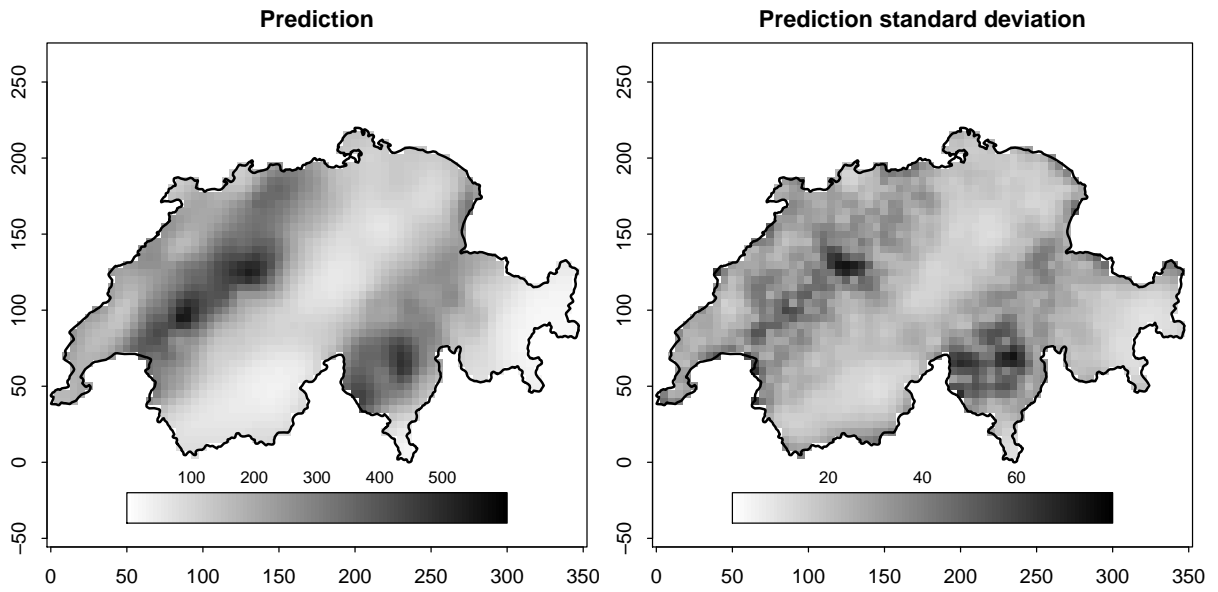


Figure 4: Maps of predictions (left panel) and prediction standard deviation (right panel) for the rainfall dataset.

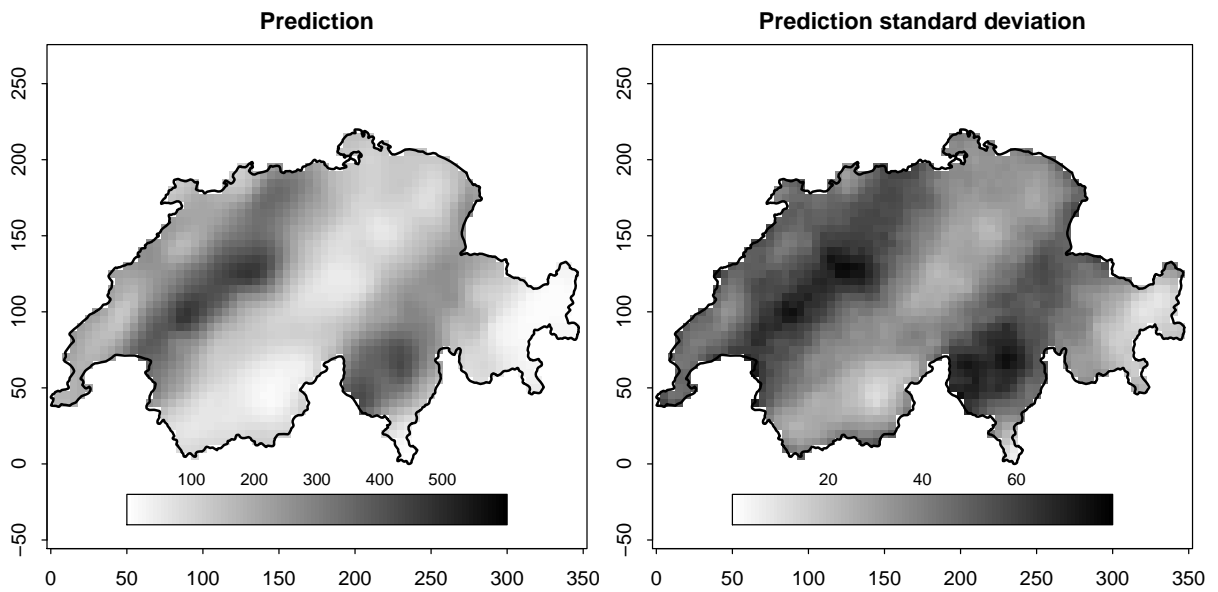


Figure 5: Maps of predictions (left panel) and prediction standard deviation (right panel) for the rainfall dataset using Christensen et al.'s (2001) model.

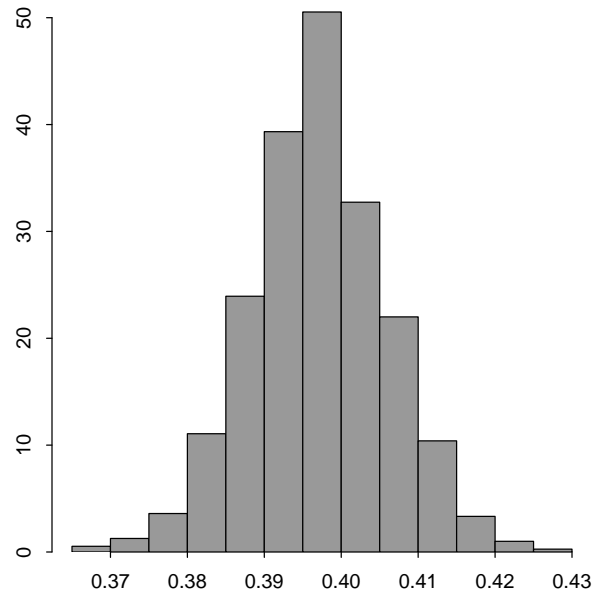


Figure 6: Histogram of random samples corresponding to the proportion of the area with rainfall larger than 200.

tions can be used. For example, the exponential family of transformations proposed in Manly (1976), or the flexible families of transformations for binary data presented in Aranda-Ordaz (1981) can be assumed. The method of estimating transformation parameter presented in this chapter can be used in these models also.

## Acknowledgments

The authors thank two anonymous reviewers for helpful comments and valuable suggestions which led to several improvements in the manuscript.

## References

- Aranda-Ordaz, F. L. (1981). On two families of transformations to additivity for binary response data. *Biometrika*, 68:357–363.
- Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association*, 76:296–311.



- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26:211–252.
- Christensen, O. F. (2004). Monte Carlo maximum likelihood in model based geostatistics. *Journal of Computational and Graphical Statistics*, 13:702–718.
- Christensen, O. F., Diggle, P. J., and Ribeiro, P. J. (2001). Analyzing positive-valued spatial data: the transformed gaussian model. In Monestiez, P., Allard, D., and Froidevaux, R., editors, *geoENV III-Geostatistics for Environmental Applications*, pages 287–298. Springer.
- De Oliveira, V. (2003). A note on the correlation structure of transformed gaussian random fields. *Australian and New Zealand Journal of Statistics*, 45:353–366.
- De Oliveira, V. and Ecker, M. D. (2002). Bayesian hot spot detection in the presence of a spatial trend: application to total nitrogen concentration in chesapeake bay. *Environmetrics*, 13:85–101.
- De Oliveira, V., Fokianos, K., and Kedem, B. (2002). Bayesian transformed gaussian random field: A review. *Japanese Journal of Applied Statistics*, 31:175–187.
- De Oliveira, V., Kedem, B., and Short, D. A. (1997). Bayesian prediction of transformed gaussian random fields. *Journal of the American Statistical Association*, 92:1422–1433.
- Diggle, P. J., Ribeiro, P. J., and Christensen, O. F. (2003). An introduction to model-based geostatistics. In *Spatial statistics and computational methods. Lecture notes in statistics*, pages 43–86. Springer.
- Doss, H. (2010). Estimation of large families of Bayes factors from Markov chain output. *Statistica Sinica*, 20:537–560.
- Geyer, C. J. (1994). Estimating normalizing constants and reweighting mixtures in Markov chain Monte Carlo. Technical Report 568, School of Statistics, University of Minnesota.
- Geyer, C. J. (2011). *Handbook of Markov chain Monte Carlo*, chapter Introduction to Markov chain Monte Carlo, pages 3–48. CRC Press, Boca Raton, FL.
- Manly, B. F. J. (1976). Exponential data transformation. *The Statistician*, 25:37–42.
- Matérn, B. (1986). *Spatial Variation*. Springer, Berlin, 2nd. edition.
- Ribeiro, P. J. and Diggle, P. J. (2012). *geoR*. R package version 1.7-4.

- Roy, V. (2014). Efficient estimation of the link function parameter in a robust Bayesian binary regression model. *Computational Statistics and Data Analysis*, 73:87–102.
- Roy, V., Evangelou, E., and Zhu, Z. (2014). Efficient estimation and prediction for the Bayesian spatial generalized linear mixed models with flexible link functions. Technical report, Iowa State University.
- Stein, M. L. (1992). Prediction and inference for truncated spatial data. *Journal of Computational and Graphical Statistics*, 1:91–110.
- Stein, M. L. (1999). *Interpolation of Spatial Data*. Springer Verlag, New York.
- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, 82:528–550.
- Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models. *Biometrics*, 58:129–136.