ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Environmental Sciences 7 (2011) 264–268

Procedia Environmental Sciences

1st Conference on Spatial Statistics 2011- Mapping Global Change

A Comparison of Empirical Bayes and Reference Prior Methods for Spatio-Temporal Data Analysis

Firoozeh Rivaz a*

Faculty of Mathematical Sciences, Shahid Beheshti University, G.C., Iran

Abstract

In Bayesian analysis of spatio-temporal data, the problem of selecting prior distribution for model parameters is of great demand. This paper considers two most popular approaches, empirical Bayes and reference prior, for Bayesian inference. We then use simulation to compare the frequentist properties of these two methods. Since, posterior propriety of the reference prior is only established under separable correlation models, this comparison is concentrated on this case.

Selection and peer-review under responsibility of Spatial Statistics 2011

Keywords: Spatio-temporal data; Bayesian Inference; Reference prior; Empirical Bayes.

1. Introduction

The Bayesian approach for analyzing spatio-temporal data has been seen gaining popularity, especially when the goal is prediction (e.g. Hooten and Wikle, 2008; Sahu et al, 2006; Banerjee et al, 2004; Wikle et al, 2001, 1998; Berliner et al, 1999). The main advantage of this approach is that the uncertainty of parameters is fully accounted for when performing prediction. In this setting, the problem of selecting prior distribution for model parameters is of great demand. Subjective specification of the prior distribution for the parameters is somewhat delicate. First, spatial and temporal parameters are in general difficult to interpret and consequently it is difficult to elicit their prior distributions. Second, a naive specification of the prior for the model parameters may give rise to improper posterior distributions (Berger et al, 2001). To overcome the above difficulties, we consider two Bayesian analysis approaches.

* Corresponding author. Tel.: +98-21-29903008.
E-mail address: f_rivaz@sbu.ac.ir.

Selection and peer-review under responsibility of Spatial Statistics 2011
doi:10.1016/j.proenv.2011.07.046

The first approach is "empirical Bayes analysis". In this setting, we consider proper prior distributions which depend on some unknown hyperparameters for spatial and temporal parameters. Then, hyperparameters are estimated. For this purpose, one may obtain the marginalized maximum likelihood (ML-II) estimates, which has to be accomplished numerically, as no explicit solution is available. However, there are some shortcomings with regard to maximize the marginal likelihood, which make the numerical identification of the maximum more difficult. To overcome the challenges related to ML-II method in spatio-temporal analysis, Rivaz et al. (2011) proposed a sampling-based method for estimation of hyperparameters. Specifically, they estimated hyperparameters based on producing some realizations from the prior distributions. The other approach that has seen significant development in recent years is to use an "objective Bayesian analysis", where by this we mean the use of default priors derived by formal rules that use the structure of the problem at hand but do not require subjective prior elicitation. One of the most popular of these methods that we will study here are the reference prior. In this paper, we use simulation to compare the frequentist properties of these two methods. Since, posterior propriety of the reference prior is only established under separable correlation models (Paulo, 2005), this comparison is concentrated on this case.

The article is organized as follows. In Section 2, we introduce the Gaussian space-time model. Section 3 illustrates empirical and reference Bayes methods. Section 4 explores a simulation experiment to assess the effectiveness of these two methods on parameters estimation. Concluding remarks are presented in Section 5.

2. Statistical Model

Let $Y(\cdot,\cdot)=\{Y(s,t); s\in D\subseteq \mathcal{R}^d, t\in T\subseteq \mathcal{R}\}$, $d\geq 1$ be a Gaussian random field with mean $E[Y(s,t)]=f'(s,t)\beta$ and covariance function

$$Cov[Y(s,t),Y(s',t')]=\sigma^2\rho(s-s',t-t';\theta);\;\; s,s'\in D,t,t'\in T$$

where $f(s,t)=(f_1(s,t),\cdots,f_p(s,t))'$ denotes the space-time dependent covariates, $\beta=\left(\beta_1,\cdots,\beta_p\right)'$ is unknown regression parameters vector, $\sigma^2=Var[Y(s,t)]$ is the fixed variance of the random field, $\rho(s-s',t-t';\theta)$ is the stationary space-time correlation function with parameter vector θ. Assuming that $\rho(s-s',t-t';\theta)=\rho_S(s-s';\theta_S)\rho_T(t-t';\theta_T)$ where $\theta_S\in\mathcal{R}^{q_S}$ and $\theta_T\in\mathcal{R}^{q_T}$. For spatial and temporal correlation functions, we concentrate on the rich and flexible Matérn family, which involves smoothness parameter ν in addition to the range parameter ϑ, and is given as:

$$\rho(d;\vartheta,\nu)=\{2^{\nu-1}\Gamma(\nu)\}^{-1}(d/\vartheta)^{\nu}\,K_\nu(\frac{d}{\vartheta}),\qquad \vartheta>0$$

where d is the distance between two locations or time instants. This family is strongly recommended by Stein (1999). In the sequel, we consider the smoothness parameter ν as a fixed parameter.

Suppose that $\boldsymbol{Y}=(Y(s_1,t_1),\cdots,Y(s_{n_S},t_{n_T})'$ be a n_Sn_T-vector represents the data measured at the sampling locations $s_1,\cdots,s_{n_S}\in D$ and time instants $t_1,\cdots,t_{n_T}\in T$. By the stated assumptions, Y has multivariate normal distribution

$$Y\sim N_{n_Sn_T}(X\beta,\sigma^2\Sigma_\theta), \tag{1}$$

where $X=(f'(s_1,t_1),\cdots,f'(s_{n_S},t_{n_T}))'$ is the known full rank $n_Sn_T\times p$ matrix, $n_Sn_T>p$ and $\Sigma_\theta=\rho(s-s',t-t';\theta)$ with $\theta=(\theta_S,\theta_T)$. The likelihood function of the model parameters $\varphi=(\beta,\sigma^2,\theta)$ based on the observed data $\boldsymbol{y}=(y(s_1,t_1),\cdots,y(s_{n_S},t_{n_T}))$ is given as:

$$L(\varphi;\boldsymbol{y})=\left(\frac{1}{2\pi\sigma^2}\right)^{n_Sn_T/2}|\Sigma_\theta|^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-X\beta)'\Sigma_\theta{}^{-1}(\boldsymbol{y}-X\beta)\right\}.$$

Applying the observed data, we are intended in predicting random field $Y(\cdot,\cdot)$ at arbitrary location s_0 and time t_0. It can be shown that the predictive distribution of $Y(s_0,t_0)$ given $\boldsymbol{y}$ and φ is normally distributed as

$$(Y(s_0,t_0)|\boldsymbol{y},\varphi)\sim N(\mu_1,\sigma^2\rho_1),$$

where

$$\mu_1 = f'(s_0,t_0)\beta + r_\theta^{\prime\Sigma_\theta^{-1}(y-X\beta)},$$

$\rho_1 = 1 - r_\theta'\Sigma_\theta^{-1}r_\theta$ and $r_\theta = \big(\rho(s_i - s_0, t_i - t_0;\theta)\big)$. Since, in practice, all of the model parameters are unknown, present study uses the Bayesian approach for space-time prediction.

3. The Empirical and Reference Priors

The Bayesian analysis requires prior distributions for all the unknown parameters. We assume that the correlation parameters are independent of each other, the vector θ is independent of the parameters β and σ^2, and $\pi(\beta,\sigma^2) \propto \frac{1}{\sigma^2}$. Since, the correlation parameters can be difficult to interpret, hence, eliciting reasonable prior distributions for them is an important issue in this context. In this setting, two approaches are often used. The first one is empirical Bayes, i.e. proper priors, which depend on some unknown hayperparameters, are used for the correlation parameters. By this way, one does not need to verify posterior existence. The other approach is objective Bayesian analysis that uses default priors. In this paper, we present the Bayesian analysis of spatio-temporal data based on these two approaches.

3.1 Empirical Bayes Prior

In order to assure a proper posteriors, we consider independent proper priors as $G(a_S,b_S)$, $G(a_T,b_T)$ for θ_S and θ_T, respectively. So, the prior densities satisfy

$$\pi(\varphi|\lambda) = \pi(\beta,\sigma^2,\theta_S,\theta_T|\lambda) \propto \frac{\pi(\theta_S|\lambda_S)\pi(\theta_T|\lambda_T)}{\sigma^2}$$

where $\lambda = (\lambda_S,\lambda_T)$ is the unknown hyperparameters vector with $\lambda_S = (a_S,b_S)$ and $\lambda_T = (a_T,b_T)$. To estimate the hyperparameters vector, λ, one may use the ML-II method. However, there are some shortcomings with regard to maximize the marginal likelihood, which make the numerical identification of the maximum more difficult. To overcome the challenges related to ML-II method in spatio-temporal analysis, Rivaz et al. (2011) proposed a sampling-based method for estimation of hyperparameters. Specifically, they considered the estimates of θ_S in n_T time and θ_T in n_S sites as samples of the prior distributions $\pi(\theta_S|\lambda_S)$ and $\pi(\theta_T|\lambda_T)$, respectively. Then estimation of hyperparameters, i.e. $\widehat{\lambda_S}$ and $\widehat{\lambda_T}$ were determined using moment method. According to the estimated priors, the posterior distribution is given by

$$\pi\big(\varphi\big|y,\hat{\lambda}\big) = \pi(\beta|y,\sigma^2,\theta)\pi(\sigma^2|y,\theta)\pi(\theta|y,\hat{\lambda})$$

such that $(\beta|y,\sigma^2,\theta)\sim N_p(\hat{\beta},\sigma^2\hat{V})$ and $(\sigma^2|y,\theta)\sim\chi^2{}_{INV}(n_S n_T - p, S^2)$ where $\hat{\beta} = (X'\Sigma_\theta^{-1}X)^{-1}X'\Sigma_\theta^{-1}y$, $\hat{V} = (X'\Sigma_\theta^{-1}X)^{-1}$ and $S^2 = \frac{y'\Sigma_\theta^{-1}y - \hat{\beta}'\hat{V}^{-1}\hat{\beta}}{n_S n_T - p}$. To compute $\pi(\theta|y,\hat{\lambda})$, we note

$$\begin{aligned}\pi\big(\theta\big|y,\hat{\lambda}\big) &= \frac{\pi(\beta,\sigma^2,\theta|y)}{\pi(\beta|y,\sigma^2,\theta)\pi(\sigma^2|y,\theta)}\\ &= \frac{f(y|\beta,\sigma^2,\theta)\pi(\beta|\sigma^2)\pi(\sigma^2)}{\pi(\beta|y,\sigma^2,\theta)\pi(\sigma^2|y,\theta)}\pi\big(\theta\big|\hat{\lambda}\big)\\ &\propto h(y|\theta)\pi\big(\theta\big|\hat{\lambda}\big)\end{aligned}$$

where

$$h(y|\theta) = |\hat{V}|^{\frac{1}{2}}|\Sigma_\theta|^{-\frac{1}{2}}(S^2)^{-\frac{n_S n_T - p}{2}},$$

and the proportionality constant is independent of θ. However, this expression does not define a standard probability distribution. In order to generate samples from the posterior distribution $\pi(\theta|y,\hat{\lambda})$, we run an

MCMC algorithm. For this, we exploit the full conditional distributions as usually done in Gibbs sampling (Gelfand and Smith, 1990) framework. Then, to sample from the conditional distributions, we use random-walk Metropolis-Hastings algorithm, tuned to give a reasonable acceptance rate.

3.2 Reference Prior

The other prior that we can choose for φ is the default priors derived by formal rules that use the structure of the problem at hand but do not require subjective prior elicitation. One of the most popular of these methods is the reference prior. Paulo (2005) derived explicit expression for reference prior and established results on propriety of the resulting posterior distribution under some restricted assumptions. More precisely, he obtained the reference prior as:

$$\pi^R(\varphi) \propto \frac{\pi^R(\theta)}{(\sigma^2)^a} \qquad (2)$$

with $a = 1$ and $\pi^R(\theta) \propto |I(\theta)|^{1/2}$, where $|I(\theta)|$ is the Fisher information matrix. Also, the resulting posterior was shown to be proper when $p = 1$ and the correlation structure is separable. According to reference prior (2), the posterior distribution is given by

$$\pi(\varphi|y) = \pi(\beta|y, \sigma^2, \theta)\pi(\sigma^2|y, \theta)\pi(\theta|y)$$

where $\pi(\beta|y, \sigma^2, \theta)$ and $\pi(\sigma^2|y, \theta)$ are similar to empirical Bayes prior, but

$$\pi(\theta|y) \propto L(\theta; y)\pi(\theta)$$

where $L(\theta; y)$ is the integrated likelihood of θ. To sample from the posterior distribution $\pi(\theta|y)$, we use the Metropolis-Hastings algorithm.

4. Comparison of the Empirical Bayes and Reference Priors

In this section, we use simulation to assess the frequentist properties of Bayesian procedures based on reference and empirical Bayes priors. In our simulation studies, the spatial region of interest is the unit square $D = [0,1] \times [0,1]$ with allowable spatial sampling points of 25 where they are selected regularly in D with coordinates $s(i, j) = (i - 0.5, j - 0.5)/5$, $i, j = 1, \cdots, 5$. Data are simulated at 25 locations on 15 temporal instants $t = 1, \cdots, 15$ from model (1). To check how well the Bayesian methods estimate the parameters, we consider different parameters values. In fact, separable Gaussian random fields based on model (1) are simulated with zero mean, $\sigma^2 = 1$ and spatial and temporal correlation functions of the Matérn family with $\theta_S, \theta_T \in \{0.5, 1.5\}$ as corresponding to small and large spatial and temporal range parameters and ν_S and ν_T, the smoothness parameters belonging to $\{0.5, 1\}$. For each set of parameters we generated 100 Gaussian random fields and from each simulated data set we computed the posterior mean of μ, σ^2, θ_S and θ_T using both empirical Bayes and reference priors. Then, for each parameter we computed the MSE of both Bayesian estimators.

Tables 1 summarizes the results related to the mean squared error (MSE) of mean, variance and range parameters based on 1000 replications. As seen, based on the reference method, estimation of variance is better than empirical Bayes. But, generally both methods have similar manner in parameters estimation. Although, this study is too limited to establish that the reference prior and empirical Bayes method generally yields inferences with satisfactory frequentist performance, but on the basis that the reference prior is computationally more demanding, we recommend the empirical Bayes prior for analysis of spatio-temporal Gaussian models.

Table 1. Mean squared error of mean, variance and range parameters based on empirical Bayes prior $(\hat{\mu}_{EB}, \hat{\sigma}^2{}_{EB}, (\hat{\theta}_S, \hat{\theta}_T)_{EB})$ and reference prior $(\hat{\mu}_R, \hat{\sigma}^2{}_R, (\hat{\theta}_S, \hat{\theta}_T)_R)$

		(θ_S, θ_T)			
(ν_S, ν_T)		(0.5,0.5)	(0.5,1.5)	(1.5,0.5)	(1.5,1.5)
(0.5,0.5)	$\hat{\mu}_{EB}$	0.0014	0.0010	0.0008	0.0013
	$\hat{\mu}_R$	0.0012	0.0015	0.0012	0.0011
(1,1)	$\hat{\mu}_{EB}$	0.0018	0.0015	0.0010	0.0017
	$\hat{\mu}_R$	0.0011	0.0013	0.0016	0.0014
(0.5,0.5)	$\hat{\sigma}^2{}_{EB}$	0.060	0.074	0.071	0.082
	$\hat{\sigma}^2{}_R$	0.049	0.048	0.063	0.66
(1,1)	$\hat{\sigma}^2{}_{EB}$	0.066	0.069	0.081	0.077
	$\hat{\sigma}^2{}_R$	0.059	0.064	0.070	0.081
(0.5,0.5)	$(\hat{\theta}_S, \hat{\theta}_T)_{EB}$	0.053	0.059	0.061	0.057
	$(\hat{\theta}_S, \hat{\theta}_T)_R$	0.049	0.054	0.063	0.060
(1,1)	$(\hat{\theta}_S, \hat{\theta}_T)_{EB}$	0.058	0.050	0.56	0.063
	$(\hat{\theta}_S, \hat{\theta}_T)_R$	0.066	0.061	0.055	0.060

5. Concluding Remarks

In this study, we have proposed two Bayesian procedures for analysis a Gaussian spatio-temporal model. In the simulation study, the performance of two approaches, empirical Bayes and reference prior methods, has been investigated on parameters estimation. It has been found that the empirical Bayes prior has approximately similar frequentist properties to reference prior. Although, this empirical Bayes method could not serve as a general methodology for hyperparameters estimation (Rivaz et al. 2011), However, in compare to reference priord in spatio-temporal setting, its determination is more simple. Moreover, it is a source of debate as whether to extend the reference prior method to more general settings such as nonseparable models. So, on the basis that the reference prior is computationally more demanding, we recommend the empirical Bayes prior for analysis of spatio-temporal Gaussian models.

References

[1] Banerjee S, Carlin BP, Gelfand AE. *HierarchicalModeling and Analysis for Spatial Data* . Boca Raton, FL: Chapman and Hall/CRC; 2004.

[2] Berger JO, De OliveiraV, and Sanso B. Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association* 2001. **96** :1361-74.

[3] Berliner LM, Royle JA, Wikle CK, Millif RF, Bayesian Methods in the Atmospheric Sciences, *Bayes Stat* 1999. **6**: 83-100.

[4] Paulo R. Default priors for Gaussian processes, *Annals of Statistics* 2005. **33**: 556-82.

[5] Gelfand AE., Smith AFM., Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* 1990. **85**:398-409.

[6] Hooten MB., Wikle CK. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the eurasian collared-dove, *Environ. Ecol. Stat.* 2008; **15**: 59–70.

[7] Rivaz F, Mohammadzadeh M, Khaledi M.J. Spatio-temporal modeling and prediction of CO concentrations in Tehran city, *Journal of Applied Statistics* 2011, In press.

[8] Sahu SK., Gelfand AE., Holland DM. Spatio-temporal modeling of fine particulate matter, *J. Agric. Biol. Environ. Statist.* 2006; **11**: 61–86.

[9] Stein M. *Spatial Interpolation: Some Theory for Kriging.* Springer-Verlag, New York; 1999.

[10] Wikle CK., Milliff RF, Nychka D, Berliner LM. Spatiotemporal Hierarchical Bayesian Modeling. Tropical Ocean Surface Winds, *Journal of the American Statistical Association* 2001; **96**: 382-97.