Available online at www.sciencedirect.com
ScienceDirect
Procedia Environmental Sciences 3 (2011) 8–13
Agriculture and Agricultural Science Procedia

Spatial Statistics 2011 – Mapping Global Change

Generalised extreme value geoadditive model analysis via variational Bayes

Sarah E. Neville[a*] & M.P. Wand[b]

[a]*School of Mathematics and Applied Statistics, University of Wollongong, Wollongong 2522, Australia*
[b]*School of Mathematical Sciences, University of Technology, Sydney 2007, Australia*

Abstract

We devise a variational Bayes algorithm for fast approximate inference in Bayesian Generalized Extreme Value additive model analysis. Such models are useful for flexibly assessing the impact of continuous predictor variables on sample extremes. The new methodology allows large Bayesian models to be fitted and assessed without the significant computing costs of Monte Carlo methods.

Selection and/or peer-review under responsibility of Alfred Stein.

Keywords: Auxiliary mixture sampling; Bayesian inference; generalized additive models; sample extremes; variational approximation.

1. Introduction

Regression analysis for sample extreme responses is a topic of considerable current interest, with climate research being one of the main driving forces. The last decade has seen additive models for sample extremes added to the regression armory. Relevant references are Davison & Ramesh [1], Chavez-Demoulin & Davison [2], Yee & Stephenson [3], Padoan & Wand [4] and Laurini & Pauli [5]. Each of them differ according to (a) whether a Bayesian or non-Bayesian approach is taken, (b) concentration on sample maxima/minima versus threshold exceedences, and (c) the method of fitting and inference. In this article we focus on Bayesian inference for sample maxima/minima, and introduce a new computational method for inference in models of these type: variational Bayes. The present article is

*Corresponding author. Tel.: +61 422 562 188; fax: +61 242 214 845.
E-mail address: sen045@uow.edu.au.

doi:10.1016/j.proenv.2011.02.003

concerned with Generalized Extreme Value (GEV) geoadditive models. We include an illustration of the variational Bayes methodology using maximum rainfall data from the Sydney hinterland region.

1.1. Notation

A random variable x has a Generalized Extreme Value distribution with parameters $\mu, \sigma > 0$ and ξ, denoted by $x \sim GEV(\mu, \sigma, \xi)$, if its density function is $p(x) = \frac{1}{\sigma} f_{GEV}\left(\frac{x-\mu}{\sigma}; \xi\right)$ where $f_{GEV}(x, \xi) \equiv (1+\xi x)^{-1/\xi-1} \exp\{-(1+\xi x)\}^{-1/\xi}$, $1+\xi x > 0$ is the $GEV(0,1,\xi)$ density function.

Similarly, x has an Inverse Gamma distribution with parameters $A, B > 0$, denoted by $x \sim IG(A,B)$, if its density function is $p(x) = B^A \Gamma(A)^{-1} x^{-A-1} e^{-B/x}$, $x > 0$. If y_i has distribution D_i for each $1 \le i \le n$, and the y_i are independent, then we write $y_i \overset{ind.}{\sim} D_i$.

2. Bayesian generalized extreme value geoadditive models

Let y_i, $1 \le i \le n$, be a set of response variables for which a $GEV(\mu_i, \sigma, \xi)$ distribution is appropriate. GEV geoadditive models assume that the means take the form

$$\mu_i = f_1(x_{1i}) + \ldots + f_d(x_{di}) + g(\mathbf{x}_i) \qquad (1)$$

where, for each $1 \le i \le n$, $(x_{1i}, \ldots, x_{di})$ is a vector of continuous predictor variables, $\mathbf{x}_i$ is a bivariate vector for geographic location, $f_1, \ldots, f_d$ are smooth univariate functions, and g is a smooth bivariate function. We adopt the mixed model-based penalized spline approach and model the right hand side of (1) as

$$\sum_{l=1}^{d} f_l(x_l) + g(\mathbf{x}) = \beta_0 + \sum_{l=1}^{d} \left\{\beta_l x_l + \sum_{k=1}^{K_l} u_{l,k} z_{l,k}(x_l)\right\} + \boldsymbol{\beta}^{geo^T} \mathbf{x} + \sum_{k=1}^{K^{geo}} u_k^{geo} z_k^{geo}(\mathbf{x})$$

with $u_{l,1}, \ldots, u_{l,K_l} \mid \sigma_{ul}^2 \overset{ind.}{\sim} N(0, \sigma_{ul}^2)$ for each $1 \le l \le d$ and

$u_1^{geo}, \ldots, u_{K^{geo}}^{geo} \mid \sigma_{u,geo}^2 \overset{ind.}{\sim} N(0, \sigma_{u,geo}^2)$. Here, $\{z_{l,1}(\cdot), \ldots, z_{l,d}(\cdot)\}$ is a set of univariate spline basis functions for estimation of f_l and $\{z_k^{geo}(\cdot)\}$ is a set of bivariate spline basis functions for estimation of g. Define the matrices

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_d \\ \boldsymbol{\beta}^{geo} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1d} & \mathbf{x}_1^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nd} & \mathbf{x}_n^T \end{bmatrix}, \quad \mathbf{u}_l = \begin{bmatrix} u_{l,1} \\ \vdots \\ u_{l,K_l} \end{bmatrix}, \quad \mathbf{u}^{geo} = \begin{bmatrix} u_1^{geo} \\ \vdots \\ u_{K^{geo}}^{geo} \end{bmatrix},$$

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_d \\ \mathbf{u}^{geo} \end{bmatrix}, \quad \text{and } \mathbf{Z} = \begin{bmatrix} z_{l,1}(x_{1l}) & \cdots & z_{l,K_l}(x_{1l}) \\ \vdots & \ddots & \vdots \\ z_{l,1}(x_{nl}) & \cdots & z_{l,K_l}(x_{nl}) \end{bmatrix}.$$

The number and position of knots $\boldsymbol{\kappa}_k$ are generally chosen using a space filling algorithm as described in Ruppert et al. [6]. Form the matrices

$$\mathbf{Z}_K = \left[\left[\|\mathbf{x}_i - \boldsymbol{\kappa}_k\|^2 \log\|\mathbf{x}_i - \boldsymbol{\kappa}_k\| \right]_{1\le k,k'\le K^{geo}} \right]_{1\le i\le n} \quad \text{and} \quad \boldsymbol{\Omega} = \left[\|\boldsymbol{\kappa}_k - \boldsymbol{\kappa}_{k'}\|^2 \log\|\boldsymbol{\kappa}_k - \boldsymbol{\kappa}_{k'}\| \right]_{1\le k,k'\le K^{geo}}$$

and then find the singular value decomposition of $\boldsymbol{\Omega}$ using $\boldsymbol{\Omega} = \mathbf{U}diag(\mathbf{d})\mathbf{V}^T$ and use this to obtain the matrix square root of $\boldsymbol{\Omega}$, $\boldsymbol{\Omega}^{1/2} = \mathbf{U}diag(\sqrt{\mathbf{d}})\mathbf{V}^T$. We then compute $\mathbf{Z}^{geo} = \mathbf{Z}_K\boldsymbol{\Omega}^{-1/2}$ and define $\mathbf{Z} = \left[\mathbf{Z}_1 \ \cdots \ \mathbf{Z}_d \ \mathbf{Z}^{geo}\right]$. Then a Bayesian GEV geoadditive model is

$$y_i \mid \boldsymbol{\beta}, \mathbf{u}, \sigma_\varepsilon, \xi \overset{ind.}{\sim} GEV((\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})_i, \sigma_\varepsilon, \xi), \tag{2}$$

$$\mathbf{u} \mid \sigma_{u1}^2, \ldots, \sigma_{ud}^2, \sigma_{u,geo}^2 \sim N(\mathbf{0}, blockdiag(\sigma_{u1}^2\mathbf{I}, \ldots, \sigma_{ud}^2\mathbf{I}, \sigma_{u,geo}^2\mathbf{I})),$$

$$\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}),\ \sigma_\varepsilon^2 \sim IG(A_\varepsilon, B_\varepsilon),$$

$$\sigma_{ul}^2 \overset{ind.}{\sim} IG(A_{ul}, B_{ul}),\ \sigma_{u,geo}^2 \sim IG(A_{u,geo}, B_{u,geo}) \text{ and } \xi \sim p(\xi),\ \xi \in \Xi.$$

3. Variational Bayes inference

Our approach to variational Bayes inference for GEV geoadditive models consists of three stages, The first involves finite normal mixture approximation of the $GEV(0,1,\xi)$ density function over each $\xi \in \Xi$. The second stage involves variational Bayes inference for Bayesian finite normal mixture geoadditive models. Such models take the same form as (2), but with a finite normal mixture distribution used to model the responses. Section 3.2 describes such models and a variational Bayes fitting algorithm. For fixed ξ this algorithm results in approximate posteriors for the geoadditive model parameters $\boldsymbol{\beta}, \mathbf{u}, \sigma_\varepsilon^2$ σ_{ul}^2, $1 \le l \le d$ and $\sigma_{u,geo}^2$. The final stage is to combine the results across all fits to make approximate Bayesian inference for all model parameters, including the shape parameter ξ.

3.1. Basic principles of variational Bayes

Consider a generic Bayesian model, with observed data vector $\mathbf{y}$ and parameter vector $\boldsymbol{\theta}$. Suppose that $\boldsymbol{\theta}$ is continuous over the parameter space $\boldsymbol{\Theta}$. The posterior density function $p(\boldsymbol{\theta} \mid \mathbf{y})$ is often intractable. Variational Bayes overcomes this intractability by postulating that $p(\boldsymbol{\theta} \mid \mathbf{y})$ can be well approximated by the product density forms. An example is

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \approx q_1(\boldsymbol{\theta}_1)q_2(\boldsymbol{\theta}_2)q_3(\boldsymbol{\theta}_3) \tag{3}$$

where $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3\}$ is a partition of the entries of $\boldsymbol{\theta}$. The choice of partition is usually made on tractability grounds. Each q_i is a desity function in $\boldsymbol{\theta}_i$ ($i = 1,2,3$) and they are chosen to minimize the Kullback-Leibler distance between the left and right hand sides of (3). Minimisation of the Kullback-Leibler distance is equivalent to maximization of

$$\underline{p}(\mathbf{y};q) = \int q_1(\boldsymbol{\theta}_1)q_2(\boldsymbol{\theta}_2)q_3(\boldsymbol{\theta}_3)\log\left\{\frac{p(\boldsymbol{\theta},\mathbf{y})}{q_1(\boldsymbol{\theta}_1)q_2(\boldsymbol{\theta}_2)q_3(\boldsymbol{\theta}_3)}\right\}d\boldsymbol{\theta}$$

and an iterative convex optimization algorithm (e.g. Luenberger & Ye [7]) is available for obtaining the solution. Upon convergence, the optimal q_i densities can be used for approximate Bayesian inference. The quality of the approximation depends on the reasonableness of (3).

3.2. Finite normal mixture responses

In the case of the GEV distribution, the direct variational Bayes approach fails. This is due to the GEV likelihood's complicated dependence on parameters. For this reason, we use the auxiliary mixture approach to handling GEV responses as developed by Wand et al. [8]. For each discrete value of ξ, the skewed GEV density is approximated by an extremely accurate finite mixture of normal densities. This is achieved by minimizing the Kullback-Leiber distance between the true function and the mixture approximation.

3.3. Generalized extreme value responses

For each fixed $\xi \in \Xi$, we obtain variational approximations to the conditional posteriors of the model parameters. We then use Bayesian averaging to combine the results specific to each value of ξ to an overall expression for the optimal density of each model parameter. We use the resultant optimal densities for approximate Bayesian inference.

4. Application

We now provide illustration of the methodology via the Sydney hinterland maximum rainfall data. The response variable *winter maximum rainfall* was modelled as a function of the explanatory variables *year*, *day in season*, *Southern Oscillatory Index (SOI)*, *Ocean Heat content Anomaly (OHA)*, *Pacific Decadal Oscillation (PDO)*, *longitude* and *latitude*. SOI, OHA and PDO are included since they are considered possible climate drivers for rainfall. Year and geographical location are included to address potential temporal and spatial correlation. The following GEV geoadditive model was fitted:

$$\begin{aligned}&\text{winter max. rainfall}_i\\ &\quad\overset{ind.}{\sim} GEV\{f_1(\text{year}_i) + f_2(\text{day in season}_i) + f_3(\text{OHA}_i) + f_4(\text{SOI}_i) + f_5(\text{PDO}_i)\\ &\qquad\qquad + g(\text{longitude}_i, \text{latitude}_i), \sigma_\varepsilon^2, \xi\}\end{aligned} \qquad (4)$$

for $1 \le i \le n$, where $n = 1874$ is the total number of winter maximum rainfall measurements from 50 weather stations between the years 1955 and 2003 (not all stations had this full set of years). Figures 1 and 2 show, respectively, the estimated univariate functions and bivariate function resulting from the variational Bayes fitting.

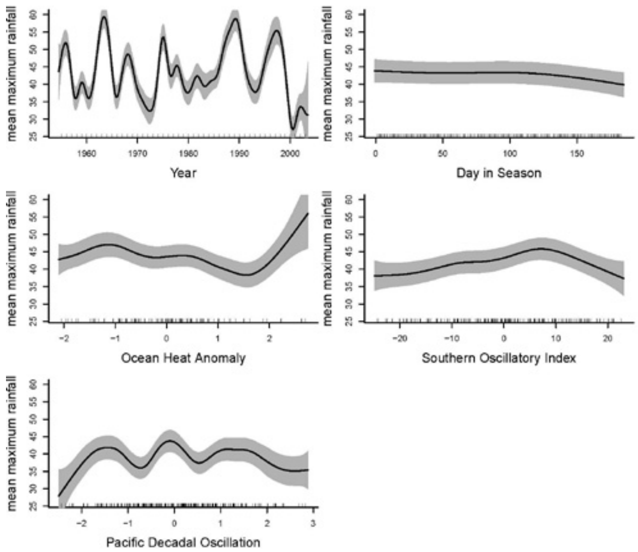

Fig. 1. Variational Bayes univariate functional fits in the GEV geoadditive model (4) for the Sydney Hinterland rainfall data. The grey region corresponds to approximate pointwise 95% credible sets.

The smooth function of year shows pronounced oscillation, corresponding to drought and wet periods in the Sydney hinterland.

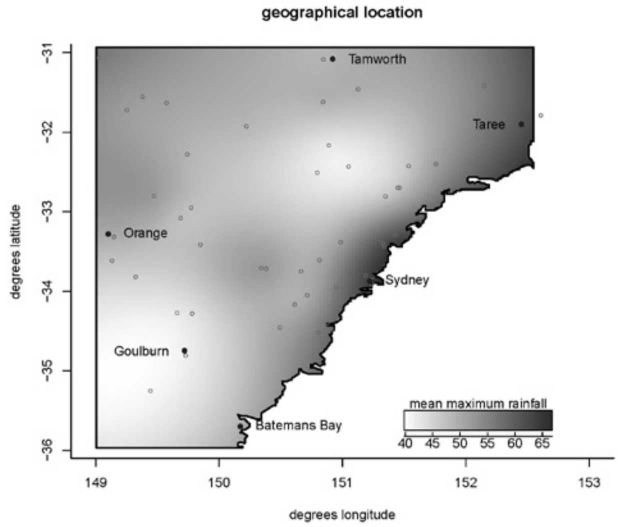

Fig. 2. Variational Bayes bivariate functional fit for geographical location in the GEV geoadditive model (4) for the Sydney hinterland rainfall data.

The fitted surface for geographical location reflects well known geographical patterns such as higher rainfall along the New South Wales coastal plain and orographic effects due to the Great Dividing Range.

5. Comparisons with MCMC

Speed is the main attraction of variational Bayes when compared with MCMC. Our R programme for performing the analysis of the Sydney hinterland rainfall data takes about 4.5 minutes to run on the third author's laptop (Mac OS X; 2.66 GHz processor, 43 GBytes of random access memory). On the the hand, MCMC implementation of the same model via the R package BRugs (Ligges et al. [9]), and with 10000 MCMC iterations, took just over 21 hours to run on the same computer. Hence, GEV additive model analyses based on MCMC can be quite difficult due to the long waiting period between model fits.

We have done some cursory accuracy comparisons between variational Bayes and MCMC. Overall, the function estimates seem to be very close. However, the pointwise credible intervals differ substantially. In particular, those based on variational Bayes are overly narrow. This observation is in keeping with those made by Wand et al. [8] for simpler GEV models. This behaviour is typical of all the varational Bayes versus MCMC comparisons we have performed for GEV additive models. It suggests that variational Bayes leads to accurate recovery of the mean structure, but that the credible interval bands are not amenable to valid pointwise inference for the mean function. Additionally, the accuracy of variational Bayes for inference concerning ξ is quite good (83% accuracy).

6. Conclusion

We have developed a new method for GEV geoadditive model analysis. Comparison studies suggest that the variational Bayes estimation of the additive model components and shape parameter is highly accurate, but the credible sets are overly narrow. Nevertheless, it facilitates approximate Bayesian inference for the such analysis in a fraction of the time taken by MCMC, and has viability advantages for larger models and sample sizes.

Ackowledgements

The authors are grateful for comments received from a Commonwealth Scientific and Industrial Research Organisation internal reviewer. This research was partially supported by Australian Research Council Discovery Project DP0877055.

References

[1] Davison, A. C. and Ramesh, N. I. (2000). Local likelihood smoothing of sample extremes. Journal of the Royal Statistical Society, Series B, 62, 191–208.

[2] Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. Applied Statistics, 54, 207–222.

[3] Yee, T.W. and Stephenson, A.G. (2007). Vector generalized linear and additive extreme value models. Extremes, 10, 1–19.

[4] Padoan, S.A. and Wand, M.P. (2008). Mixed-model based additive models for sample extremes. Statistics and Probability Letters, 78, 2850–2858.

[5] Laurini, F. & Pauli, F. (2009). Smoothing sample extremes: the mixed model approach. Computational Statistics and Data Analysis, 53, 3842–3854.

[6] Ruppert, D.,Wand, M.P.&Carroll, R.J. (2003). Semiparametric Regression. Cambridge, UK: Cambridge University Press.

[7] Luenberger, D.G & Ye, Y. (2008). Linear and Nonlinear Programming, Third Edition. New York: Springer.

[8] Wand, M.P., Ormerod, J.T., Padoan, S.A. & Frühwirth, R.(2010). Variational Bayes for elaborate distributions. Unpublished.

[9] Ligges, U., Thomas, A., Spiegelhalter, D., Best, N. Lunn, D., Rice, K. & Sturtz, S. (2009). BRugs 0.5: OpenBUGS and its R/S-PLUS interface BRugs. http://www.stats.ox.ac.uk/pub/RWin/src/contrib/