On Estimation and Prediction of Simple Model and Spatial Hierarchical Model for Temperature Extremes

Indriati Njoto Bisono
Department of Industrial Engineering
Petra Christian University
Surabaya, Indonesia
mlindri@petra.ac.id

***Abstract*—A simple independent generalized extreme value (GEV) model and a three-stage hierarchical model were applied to regional climate model outputs for temperature extremes over Tasmania, Australia. The parameters of each model were estimated using a maximum likelihood and a hybrid Markov chain Monte Carlo (MCMC) approach respectively. The two models were compared based on how well the models could predict extremes for 50 randomly selected locations that were withheld from fitting, using root mean squared prediction error (RMSPE), ten times. The RMSPEs of the two models show that the three-stage hierarchical model outperformed the simple model. We showed that the spatial hierarchical model has successfully smoothed the shape parameters. The high values tend to be pulled down, the low values to be pushed up.**

***Keywords*-generalized extreme value distribution; spatial hierarchical model; RMSPE**

I. Introduction

Extreme events occur rarely and we lack the observational data needed to learn about the characteristics of the event. In this case, simulation data generated from a global climate model (GCM), can be useful. A GCM helps us learn about weather phenomena in the long term, including climate change. A GCM uses physical and chemical equations and other parameters to predict climate behaviour. The model's predictions do not represent daily weather conditions but instead represent the overall behaviour of the climate. The coarse resolution of the GCM (up to hundreds of kilometres) of the model's predictions can be misleading. For example, within one grid cell with a range of 100 km, the topology can be very diverse and may contain mountains, valleys and plains that result in large variations in temperature. Because regional variations are not fully covered by GCM, an alternative approach is developing a higher resolution climate model, the RCM. While some RCMs are driven by GCM, RCM projections are available for a lattice grid of 0.1 degrees. References [1], [2] used these projections to evaluate extreme temperatures and produce a return level map for Tasmania.

Most atmospheric data including RCM outputs are highly spatial. The spatial dependence can be captured by directly modelling at the data level or at the process stage [3], [4], [5], [6], [7]. The latter arguably has the benefit of being more sensible and flexible because the spatial association is introduced as an adjustment to the explained covariates, usually through random effects [8]. A three-stage hierarchical model can be used as an alternative, and the Bayesian approach may be employed to estimate the parameters.

We applied and compared two models that have different assumptions about the spatial structure for predicted temperature extremes. The simple model (Model 1) assumed that the extreme values in each grid follow an independent generalized extreme value (GEV) distribution with particular parameters. The parameters were then independently estimated using maximum likelihood estimation (MLE). The more complex model (Model 2) is a three-stage hierarchical spatial model that is essentially the same as the model presented by [6]. The hierarchy allows for small-scale variation in regional effects. We compared the two models based on how well the models are able to predict unobserved locations via root mean squared prediction error (RMSPE). The hierarchical model (Model 2) will be briefly described in the next section.

II. The Hierarchical Model

We adopt a common three-level hierarchical model comprising a data level, process level and parameter level [4], [6], [7]. Let $S \subset \mathbf{R}^2$ be the area of interest, which consists of N locations or grid cells. Let $\{\boldsymbol{Y}(\mathbf{s}_i)\}$ be a subset of a spatial process $\boldsymbol{Y}(\mathbf{s}) : \mathbf{s} \in \mathbf{R}^2$, where $\boldsymbol{Y}(\mathbf{s}_i) = \{Y^{(1)}(\mathbf{s}_i), \ldots, Y^{(n)}(\mathbf{s}_i)\}$ is a vector of n years of independent data at location $\mathbf{s}_i$. Then assume $Y^{(k)}(\mathbf{s}_i), k = 1, \ldots, n$ follows a GEV distribution with particular parameters at each location, that is, $Y^{(k)}(\mathbf{s}_i) \sim \text{GEV}\big(\mu(\mathbf{s}_i), \sigma(\mathbf{s}_i), \xi(\mathbf{s}_i)\big)$. Furthermore, assume $\boldsymbol{Y}(\mathbf{s}_i)$ is conditionally independent of $\boldsymbol{Y}(\mathbf{s}_j)$ for $\mathbf{s}_i \neq \mathbf{s}_j$.

The second stage is the process level in which each of the GEV parameters is assumed to follow a normal distribution with means that comprise fixed effects ($\boldsymbol{X}^T\boldsymbol{\beta}$) and random effects (W). $\boldsymbol{X}_i$ is the covariate vector at grid $i, \boldsymbol{\beta}_\theta$ is a vector of unknown regression coefficients and θ is generically used

978-0-7695-6163-9/17 $31.00 © 2017 IEEE
DOI 10.1109/ICSIIT.2017.71

TABLE I
POSTERIOR MEANS OF β WITH CORRESPONDING STANDARD ERROR IN PARENTHESES

	ones	latitudes	longitudes	altitudes
$\boldsymbol{\beta}_\mu$	24.01	1.06	0.02	-0.01
	(0.06)	(0.06)	(0.02)	(0.00)
$\boldsymbol{\beta}_\sigma$	0.90	0.01	0.14	0.03
	(0.08)	(0.01)	(0.01)	(0.01)
$\boldsymbol{\beta}_\xi$	0.54	0.02	-0.20	0.00
	(0.22)	(0.01)	(0.00)	(0.00)

to stand for μ, σ, and ξ. Consequently, we can write

$$\begin{aligned} \mu(\mathbf{s}) &\sim \mathrm{N}\left(\boldsymbol{X}_\mu^T\boldsymbol{\beta}_\mu + W_\mu(\mathbf{s}),\ \tau_\mu^{-2}\right) \\ \log(\sigma(\mathbf{s})) &\sim \mathrm{N}\left(\boldsymbol{X}_\sigma^T\boldsymbol{\beta}_\sigma + W_\sigma(\mathbf{s}),\ \tau_\sigma^{-2}\right) \\ \xi(\mathbf{s}) &\sim \mathrm{N}\left(\boldsymbol{X}_\xi^T\boldsymbol{\beta}_\xi + W_\xi(\mathbf{s}),\ \tau_\xi^{-2}\right). \end{aligned} \tag{1}$$

A multivariate intrinsic conditional autoregressive model was applied to the random effects to capture the spatial dependence [6], [8]. Here, the effects were assumed to follow a function depending on the other effects in the neighbourhood. We chose conjugate priors for the hyperparameters to ease the computation (see [2] for details).

III. MODEL FITTING

The data are projections from the Commonwealth Scientific and Industrial Research Organisation's over Tasmania. The data are downscaled with high resolution of 0.1 degree and B1 emissions scenario, and were retrieved from the Tasmanian Partnership of Advanced Computing portal [9]. We opted to apply both models to the annual temperature maximum. In particular, we considered the annual temperature maximum in summer season (December-February) from 1991 to 2010. Hence, there are 20 data points at each site.

For Model 1, the MLE was used to estimate GEV parameters in each cell with help from ismev R-packages. The diagnostic check showed that the model fit well. For Model 2, a combination of Gibbs sampling and random walk Metropolis Hasting (MH) methods were used to sample the posterior parameters. The MH is used to sample the GEV parameters. The Gibbs sampler was used to sample the $\boldsymbol{\beta}, \boldsymbol{W}$ and $\mathbf{T}$ parameters because the posterior conditional distributions of these parameters can be fully developed.

We carried out $50,000$ iterations, left the first $20,000$ samples as a burn-in period and then thinned every 20th to eliminate autocorrelation. The running mean plot for each chain of simulated parameters behaves well and converges to each specific value. The trace plot of each simulated parameter converges with empirical density resembling the assumed normal distribution. It is not surprising that the autocorrelation plot decays quickly, indicating no autocorrelation, because every 20th sample was taken from the chain. Thus, the model passed the visual diagnostic checks.

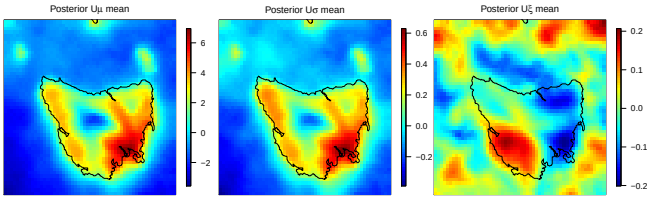

Figure 1. The posterior random effects of location, scale and shape parameters

The posterior parameters were summarised using means from the Markov chain Monte Carlo (MCMC) chains. Table I provides the posterior means of $\boldsymbol{\beta}$, and the values in parentheses are the corresponding standard errors. Some of the regression coefficients are not significant: longitude is not significant for location parameter μ, latitude is not significant for σ and surprisingly altitude is only near significant for σ. The posterior means of the random effects $\boldsymbol{W}(\mathbf{s})$ are plotted as maps in Fig. 1. Comparing the three plots, we can see that the random effect values for location parameters have similar patterns but much larger values than the scale parameters, while the values of the shape parameters exhibit different patterns with relatively small values.

The spatial hierarchical model has successfully smoothed the shape parameters, which are crucial for extreme values. With only 20 data points at each site, the tails of the distribution are likely to be thinner than that of a site with a large sample size. Therefore, the GEV distribution fitted to such data provides a large value for the shape parameter. For example, the MLE estimates of the shape parameters for the annual extremes in summer periods (with 20 data points at each site) range from -1.11 to 0.97. For monthly extremes from the same projections (with 69 data at each site), the values range from -0.58 to 0.04. Model 2 has greatly smoothed the shape parameters. For the summer annual extremes, Model 2 smoothed the shape parameters to a range of $(-0.57, 0.27)$; see Fig. 2). For the monthly extremes, with sample sizes of 69, the model does a little smoothing and only reduced the values to a range of $(-0.55, 0.02)$. The advantages of this reduction or smoothing are apparent in computing return levels. The random effects shown in Fig. 1 may explain this occurrence. The random effect values for the location and the scale parameters, $W_\mu(\mathrm{s})$ and $W_\sigma(\mathrm{s})$, exhibit similar patterns to the corresponding process values. Meanwhile, the random effect values for the shape parameter smooth the shape posteriors. The scatter plot of the shape parameters of Model 1 and Model 2 (Fig. 2) shows that for Model 2, the high values tend to be pulled down and the low values to be pushed up toward usual values for temperature extremes (i.e., about -0.2). Hence, the shape parameters of Model 2 are less spread than those of Model 1.

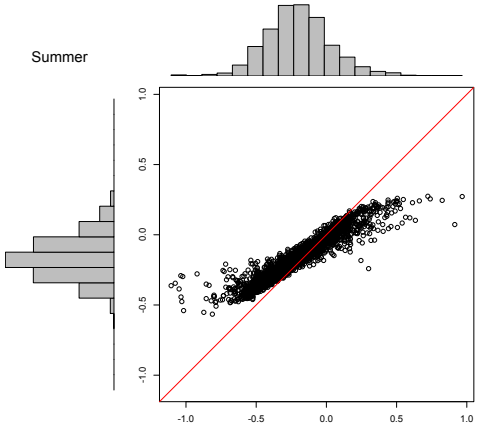

Figure 2. Shape parameter estimates Model 1 (x-axis) vs. shape posteriors Model 2 (y-axis)

IV. Prediction and Comparison

Fifty locations were selected at random and the data in these locations were considered missing. The two models were fitted to the rest of the data, and then the data from the randomly selected locations were used to validate the models. To study the performance of the two models, we computed RMSPE over $L=50$ locations. The smaller the RMSPE, the better the model. We repeated the process ten times.

Let $D_M=\{\mathrm{m}_1,\ldots,\mathrm{m}_L\}$ be the set of locations where data are omitted and kept as validation data. Denote the validation data mean at each location as $\bar{y}(\mathrm{m}_i)$ and denote the associate predicted mean as $\tilde{y}_l(\mathrm{m}_i)$, where $l=1$ for Model 1 and $l=2$ for Model 2. The RMSPE is given as

$$\mathrm{RMSPE}_l=\sqrt{\frac{1}{L}\sum_{i=1}^{L}\big(\tilde{y}_l(\mathrm{m}_i)-\bar{y}(\mathrm{m}_i)\big)^2}. \tag{2}$$

where

$$\tilde{y}_l(\mathrm{m}_i)=\begin{cases}\tilde{\mu}_l(\mathrm{m}_i)+\dfrac{\tilde{\sigma}_l(\mathrm{m}_i)}{\tilde{\xi}_l(\mathrm{m}_i)}\big(\Gamma(1-\tilde{\xi}_l(\mathrm{m}_i))-1\big) \\ \qquad\qquad \text{if } \tilde{\xi}_l(\mathrm{m}_i)\neq 0,\ \tilde{\xi}_l(\mathrm{m}_i)<1 \\ \tilde{\mu}_l(\mathrm{m}_i)+\tilde{\sigma}_l(\mathrm{m}_i)\gamma \\ \qquad\qquad \text{if } \tilde{\xi}_l(\mathrm{m}_i)=0,\ \gamma \text{ is Euler's constant} \\ \infty \qquad\qquad \text{if } \tilde{\xi}_l(\mathrm{m}_i)\geq 1,\end{cases} \tag{3}$$

that is, the mean of a GEV distribution.

Prediction for a hierarchical model is straightforward. Let $Y(\mathrm{m}_i)$ be a random variable at an unobserved location m_i that we want to predict given $y=\{y(\mathrm{s}_1),\ldots,y(\mathrm{s}_n)\}$, and let $f(y)$ be a function that minimises the mean squared prediction error, $E\big[\big(Y(\mathrm{m}_i)-f(y)\big)^2|y\big]$. By adding and subtracting the conditional mean $E[Y(\mathrm{m}_i)|y]$ and decomposing the terms inside the parentheses, we obtain

$$\begin{aligned}E\big[\big(Y(\mathrm{m}_i)-f(y)\big)^2|y\big] =& E\big[\big(Y(\mathrm{m}_i)-E[Y(\mathrm{m}_i)|y]\big)^2|y\big] \\ &+\big(E[Y(\mathrm{m}_i)|y]-f(y)\big)^2.\end{aligned} \tag{4}$$

The last term in Equation (4) is always positive, hence

$$E\big[\big(Y(\mathrm{m}_i)-f(y)\big)^2|y\big]\geq E\big[\big(Y(\mathrm{m}_i)-E[Y(\mathrm{m}_i)|y]\big)^2|y\big]. \tag{5}$$

Therefore, $f(y)=E[Y(\mathrm{m}_i)|y]$ must be the predictor that minimises the prediction error, and this is just the posterior mean of $Y(\mathrm{m}_i)$; see [8] for more details.

Here, the spatial prediction was conducted by first summarising the outputs of the MCMC runs, the posterior hyperparameters $\{\hat{\boldsymbol{\beta}},\hat{\boldsymbol{W}}(\mathrm{s}),\hat{\boldsymbol{T}}\}$, by their means. The random effect structure captures the spatial dependence through the intrinsically autoregressive model, that is, by averaging the first-order neighbour values. We used the same structure to predict the random effects at the unobserved locations by computing the predicted GEV parameters $\{\tilde{\mu}_2(\mathrm{m}_i),\tilde{\sigma}_2(\mathrm{m}_i),\tilde{\xi}_2(\mathrm{m}_i)\}$ given the covariates at the location to be predicted. Likewise, we used the mean of the neighbours to predict the GEV parameters at unobserved locations for Model 1, $\{\tilde{\mu}_1(\mathrm{m}_i),\tilde{\sigma}_1(\mathrm{m}_i),\tilde{\xi}_1(\mathrm{m}_i)\}$. If ∂_i is the set of first-order neighbours of m_i and d_i is the number of ∂_i members, $d_i=\mathrm{n}(\partial_i)$, then

$$\begin{aligned}\tilde{\mu}_1(\mathrm{m}_i) &= \frac{1}{d_i}\sum_j \hat{\mu}_{\mathrm{MLE}}(\mathrm{s}_j),\ \ \mathrm{s}_j\in\partial_i \\ \tilde{\mu}_2(\mathrm{m}_i) &= X(\mathrm{m}_i)\hat{\beta}_\mu+\rho\sum_j\frac{w_{ij}}{w_{i+}}\hat{W}_\mu(\mathrm{s}_j),\ \ \mathrm{s}_j\in\partial_i.\end{aligned} \tag{6}$$

$\tilde{\sigma}_l(\mathrm{m}_i)$ and $\tilde{\xi}_l(\mathrm{m}_i)$ were similarly defined for $l=1,2$. Finally, we computed the expected values (Equation (3)) and the RMSPEs (Equation (2)).

TABLE II
The RMSPE for Model 1 and Model 2

	Model 1	Model 2
1	6.55	6.29
2	7.59	6.45
3	6.95	6.85
4	7.89	7.27
5	7.56	6.89
6	8.41	7.90
7	6.80	6.38
8	6.85	6.11
9	7.68	6.84
10	7.50	6.47
mean	7.38	6.66
sd	0.58	0.37

Table II provides the RMSPE for each replication for both models. The spatial predictions based on averaging neighbouring values for Model 1 and Model 2 give smaller RMSPEs for Model 2 than Model 1. The RMSPE means of both models are 7.38 and 6.66, while the standard deviations are 0.58 and 0.37, respectively.

V. Summary

The simple model (Model 1) and the three-stage hierarchical model (Model 2) were applied to the summer temperature maximum over Tasmania. For Model 1, we assumed that the data at each site follow the GEV distribution with particular parameters, and we independently estimated the GEV parameters using MLE. For Model 2, we assumed that the GEV parameters in Model 1 follow normal and log-normal distributions for which the mean structures consist of both fixed and random effects. Further, the random effects were modelled by the multivariate intrinsically autoregressive model to capture the spatial dependence. We sampled the posteriors using the MCMC and found the posteriors are very close to the MLE estimate counterparts, which, by and large, is the case when the sample sizes are relatively large.

Model 2 is prominent as the sample size decreases. With a small number of data points at each site, there is a high of getting a distribution with a thin tail at some sites, and accordingly, large shape parameter values. The spatial hierarchical model (Model 2) successfully smoothed the shape parameters. The high values tend to be pulled down, and the low values tend to be pushed up. This notion of smoothing (Fig. 2) may due to the random effects shown in Fig. 1. The random effect values for the shape parameters $W_\xi(\mathrm{s})$ exhibit a smoothing spatial pattern that picks up a similar pattern from the corresponding parameter values. The notion of smoothing is also seen in the random effects for the location and scale parameters. The patterns of these random effects resemble the topography of Tasmania. The highland areas get negative values and the coastal areas get positive values.

The spatial predictions for Model 2 give, on average, smaller RMSPE than the predictions for Model 1. The means of the RMSPE of each model are 7.38 and 6.66, respectively, which are relatively high. The high RMSPE values indicate that the spatial hierarchical model is less able to predict values at unobserved locations. These results agree with the results of [10], who found that the hierarchical model gives unrealistic spatial predictions. References [10] and [11] pointed out that the reason for this drawback is the conditional independence assumption underlying the model. For weather data especially, it is unreasonable to assume that the data are conditionally independent because the occurence of an individual event may affect its neighbours. Improvement can be sought through modelling dependence in residuals. This has been attempted by [12] using copula, [13] using a Dirichlet process and [14] using a max-stable process.

References

[1] I.N. Bisono and A.P. Robinson, "Spatial Bayesian Model for Maximum Temperature," *Int. J. Applied Mathematics and Statistics*, vol. 53, no. 6, pp. 137–144, 2015.

[2] I.N. Bisono, *On Modelling Extreme Values*, Ph.D. Thesis, School of Mathematics and Statistics, University of Melbourne, Melbourne (Australia), 2016.

[3] D. Cooley, D. Nychka, and P. Naveau, "Bayesian Spatial Modeling of Extreme Precipitation Return Levels," *J. American Statistical Association*, vol. 102, no. 479, pp. 824–840, 2007, doi: 10.1198/016214506000000780.

[4] H. Sang and A.E. Gelfand, "Hierarchical Modeling for Extreme Values Observed over Space and Time," *Environmental and Ecological Statistics*, vol. 16, no. 3, pp. 407–426, Sep. 2009, doi: 10.1007/s10651-007-0078-0.

[5] B. Hrafnkelsson, J.S. Morris, and V. Baladandayuthapani, "Spatial Modeling of Annual Minimum and Maximum Temperatures in Iceland," *Meteorology and Atmospheric Physics*, vol. 116, no. 1–2, pp. 43–61, Apr. 2012, doi: 10.1007/s00703-010-0101-0.

[6] E.M. Schliep, D. Cooley, S.R. Sain, and J.A. Hoeting, "A Comparison Study of Extreme Precipitation from Six Different Regional Climate Models via Spatial Hierarchical Modeling," *Extremes*, vol. 13, no. 2, pp. 219–239, Jun. 2010, doi: 10.1007/s10687-009-0098-2.

[7] E.L. Kang and N. Cressie, "Bayesian Inference for the Spatial Random Effects Model," *J. American Statistical Association*, vol. 106, no. 495, pp. 972–983, 2011, doi: 10.1198/jasa.2011.tm09680.

[8] S. Banerjee, B.P. Carlin, and A.E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, Monographs on Statistics and Applied Probability, CRC Press, 2004.

[9] TPAC, Oceans & Climate Digital Library Portal, Tasmanian Partnership for Advanced Computing. [Online] Available: http://dl.tpac.org.au/tpacportal/.

[10] A.C. Davison, S.A. Padoan, and M. Ribatet, "Statistical Modeling of Spatial Extremes," *Statistical Science*, vol. 27, no. 2, pp. 161–186, May 2012.

[11] D. Cooley, J. Cisewski, R.J. Erhardt, S. Jeon, E. Mannshardt, B.O. Omolo, and Y. Sun, "A Survey of Spatial Extremes: Measuring Spatial Dependence and Modeling Spatial Effects," *REVSTAT Statistical Journal*, vol. 10 no. 1, pp. 135–165, Mar. 2012.

[12] H. Sang and A.E. Gelfand, "Continuous Spatial Process Models for Spatial Extreme Values," *J. Agricultural, Biological, and Environmental Statistics*, vol. 15, no. 1, pp. 49–65, Mar. 2010, doi: 10.1007/s13253-009-0010-1.

[13] M. Fuentes, J. Henry, and B. Reich, "Nonparametric Spatial Models for Extremes: Application to Extreme Temperature Data," *Extremes*, vol. 16, no. 1, pp. 75–101, Mar. 2013, doi: 10.1007/s10687-012-0154-1.

[14] M. Ribatet, D. Cooley, and A.C. Davison, "Bayesian Inference for Composite Likelihood with an Application to Spatial Extremes," *Statistica Sinica*, vol. 22, no. 2, pp. 813–845, Apr. 2012, doi: 10.5705/ss.2009.248.