

## Research Article

# Bayesian Spatial and Trend Analysis on Ozone Extreme Data in South Korea: 1991–2015

Cheru Atsmegiorgis Kitabo 

Department of Statistics, College of Natural and Computational Sciences, Hawassa University, Hawassa, Ethiopia

Correspondence should be addressed to Cheru Atsmegiorgis Kitabo; cherueden@yahoo.com

Received 11 April 2020; Accepted 12 November 2020; Published 27 November 2020

Academic Editor: Herminia García Mozo

Copyright © 2020 Cheru Atsmegiorgis Kitabo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Background.** Extreme events like flooding, extreme temperature, and ozone depletion are happening in every corner of the world. Thus, the need to model such rare events having enormous damage has been getting priorities in most countries of the world. **Methods.** The dataset contains the ozone data from 29 representative air monitoring sites in South Korea collected from 1991 to 2015. Spatial generalized extreme value (GEV) using maximum likelihood estimation (MLE) and two max-stable and Bayesian kriging models are the statistical models used for analysis. Moreover, predictive performances of these statistical models are compared using measures like root-mean-squared error (RMSE), mean absolute error (MAE), relative bias (rBIAS), and relative mean separation (rMSEP) have been utilized. **Results.** From the time plot of ozone data, extreme ozone concentration is increasing linearly within the specified period. The return level of ozone concentration after 10, 25, 50, and 100 years have been forecasted and showed that there was an increasing trend in ozone extremes. High spatial variability of ozone extreme was observed, and those areas around the territories were having extreme ozone concentration than the centers. Moreover, Bayesian Kriging brought about relatively the minimum RMSE compared to the other models. **Conclusion.** The extreme ozone concentration has clearly showed a positive trend and spatial variation. Moreover, among the models considered in the paper, the Bayesian Kriging has been chosen as the better model.

## 1. Background

Extreme events like flooding, extreme temperature, and ozone depletion are happening in every corner of the world. Thus, the need to model such rare events is being getting prioritized in most countries of the world. These days, depletion of the stratospheric ozone layer by halogenated chemicals is a global problem. Since the issue was primarily a matter of scientific curiosity during the 1970s and early 1980s, recently, it became an urgent policy question for governments in both developed and developing countries.

There are various effects of ozone starting from human health to increasing the chance of existence of natural catastrophe. Some of them are the effects on human and animal health, terrestrial plants, aquatic ecosystems, biogeochemical cycles, air quality, materials, climate change, and ultraviolet radiation. According to Stedman et al., the

effect on human health like eye diseases, skin cancer, and infectious diseases is due to increased penetration of solar UV-B radiation [1]. The changes in plant form and secondary metabolism due to UV-B could have an important implication for plant competitive balance, plant pathogens, and biogeochemical cycles.

The climate and environmental applications used multivariate extreme distributions in order to take into account the extremal dependence [2–5]. For the study of extremes in a spatial context, max-stable processes are a relevant framework for modeling and inference. Different parametric models have been later proposed by various authors [6–8]. The extremal  $t$  process is a model of that family that extends the Schlather model [9, 10]. It has a supplementary parameter that lets the extremal coefficient function vary in a

wider range. Max-stable processes have been used in applications for modeling air pollution using ozone data, rainfall, temperature, snowfall, and snow depth [11–14].

Designing the methodology that brings accurate risk measures of these extreme events needed to minimize the casualties on human being and reduce the damage on infrastructure is indispensable. Thus, the importance of modeling such events and getting ready to take precaution measures will have great importance.

The main goal of this paper is to model and compare the predictive performance of the extreme values like spatial GEV with MLE estimation, two max-stable spatial models, and the other recent model Bayesian kriging. In most papers, Bayesian kriging has not been considered for comparison. The comparative criterion considered for comparing the predictive performance of Schlather's, extremal  $t$ , and Bayesian kriging is the 5-fold cross validation. The paper considers 29 air monitoring sites found South Korea, and the daily ozone data have been collected from the year 1991 to 2015. 8 hours' maximum ozone concentration has been selected from the daily maximum ozone concentration. Then, using this daily maximum ozone, the monthly maximum has been selected. From the monthly maximum, annual maximum have been selected. Actually, only ozone season months from May to September have been utilized for extracting data for the model fitting process.

## 2. Methodology

**2.1. Basics of Extreme Value Theory (EVT).** These days, the statistical modeling of univariate extremes has been well assessed. However, these tools to model spatial extreme data have been active research areas of the time.

**2.1.1. Univariate Extreme Value Models.** For each station in this study, the univariate extreme value model has been fitted using generalized extreme value (GEV) distribution using Bayesian perspective. The GEV defined as

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)^{-1/\xi}\right]\right\}, \quad (1)$$

on  $\{z: 1 + \xi((z - \mu)/\sigma) > 0\}$  and with parameters satisfying  $-\infty < \mu < \infty$ ,  $\sigma > 0$ , and  $-\infty < \xi < \infty$ .

The maximum value of ozone per site ( $Z_{t,i}$ )  $\sim$  GEV( $\mu_{t,i}$ ,  $\sigma$ ,  $\xi$ ) and the parameters of the model are location ( $\mu$ ), scale ( $\sigma$ ), and shape ( $\xi$ ) for each 29 air monitoring sites. Block maxima first approved in work by Fisher and Tippett approach has been utilized [15]. Time series plot of the ozone concentration versus time has been plotted. Trend has been observed. Thus, the location parameter has been expressed as a linear function of time:

$$\mu_{t,i} = \beta_{0,i} + \beta_{1,i}(t). \quad (2)$$

As the data contain ozone concentration in different sites, spatial dependence is inevitable.

**2.1.2. Generalized Extreme Value Distribution.** The generalized extreme value (GEV) distribution has a location parameter  $\mu$  and a scale parameter  $\sigma$  and combines the three

types of extreme value distributions by use of a shape parameter  $\xi$ . It is usually represented by GEV( $\mu$ ,  $\sigma$ ,  $\xi$ ). The distribution function is as follows: the case  $\xi < 0$  corresponds to the Weibull family; for the case  $\xi = 0$ , the limit of the GEV as  $\xi \rightarrow 0$  is used, which leads to the Gumbel family:

$$G(z) = \exp\left[-\exp\left\{-\left(\frac{z - \mu}{\sigma}\right)\right\}\right]. \quad (3)$$

Therefore, we now have a single formulation for modeling block maxima data which overcomes the issue of distribution choice from the three types by allowing the data to choose the appropriate tail behavior for the model. The uncertainty for this choice is now incorporated into the model within the uncertainty for the parameter estimate of  $\xi$ :

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)^{-1/\xi}\right]\right\}, \quad (4)$$

defined on  $\{z: 1 + \xi((z - \mu)/\sigma) > 0\}$  and with parameters satisfying  $-\infty < \mu < \infty$ ,  $\sigma > 0$  and  $-\infty < \xi < \infty$ . The case  $\xi > 0$  corresponds to the Fréchet family.

The yearly maximum values of the ozone concentration per geographic zone  $i$ ,  $Z_{1,i}, \dots, Z_{T,i}$  follows generalized extreme value (GEV) distribution of the following form:

$$Z_{t,i} = \text{GEV}(\mu_{t,i}, \sigma_{t,i}, \xi), \quad (5)$$

with the location parameter,

$$\mu_{t,i} = \beta_{0,i} + \beta_{1,i}(t), \quad \text{for } t = 1, \dots, 25 \text{ and } i = 1, 2, \dots, 29, \quad (6)$$

where  $\beta_{0,i}$  is the intercept and  $\beta_{1,i}$  is the trend in  $t$ .

**2.2. Return Level Estimation for Block Maxima Approach.** Finding the probability of occurrence for large events is the major motivation for modeling extreme data. This demands us to be able to predict the level of the process that we expect to exceed only once in the time period required, which is known as a return level.

For yearly maximum data, the value of the return level can be estimated from the GEV distribution function directly. If we let  $Z_p$  be the  $(1/P)$  year return level of the process, then

$$G(z_p) = \exp(-p), \quad (7)$$

where this expression comes from requiring the exceedance rate of  $z_p$  to average once every  $(1/P)$  years and considering a constant rate of exceedance. Accordingly, the approximate value of the  $(1/P)$  year return level for the GEV distribution is given as

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} \left[1 - \{-\log(1 - p)\}^{-\xi}\right], & \text{for } \xi \neq 0, \\ \mu - \sigma \log\{-\log(1 - p)\}, & \text{for } \xi = 0. \end{cases} \quad (8)$$

2.3. *Bayesian Kriging Model Fitting for Annual Maximum Ozone.* A spatial model with a simpler version that is considered in this paper utilizes a model with only one latent spatial process and no measurement errors:

$$\begin{aligned} \text{Level 1; } Y(u) &= X\beta + \sigma T(u), \\ \text{Level 2; } T(u) &\sim N(0, R_y(\mathcal{O})), \\ \text{Level 3; } \text{pr}(\beta, \sigma^2, \mathcal{O}). \end{aligned} \quad (9)$$

The likelihood function is given by

$$\begin{aligned} L(\beta, \sigma^2, \mathcal{O}|Y) &\propto (\sigma^2)^{(-n/2)} |R_y(\mathcal{O})|^{(-1/2)} \\ &\exp\left\{\frac{-1}{2}(y - X\beta)'(R_y(\mathcal{O}))^{-1}(y - X\beta)\right\}. \end{aligned} \quad (10)$$

2.3.1. *Posterior for the Mean Parameter  $\beta$ .* The conjugate prior for the mean parameter  $\beta$  has been considered. Assuming a normal conjugate prior for the mean parameter  $\beta$ , which means

$$(\beta|\sigma_*^2, \mathcal{O}_*) \sim N(m_\beta, \sigma_*^2 V_\beta). \quad (11)$$

The posterior distribution is given by

$$\begin{aligned} (\beta|Y, \sigma_*^2, \mathcal{O}_*) &\sim N\left((V_\beta^{-1} + X' \sigma_*^2 X)^{-1}\right. \\ &\left.(V_\beta^{-1} m_\beta + X' \sigma_*^2 y)\right) \sigma_*^2 \left((V_\beta^{-1} + X' \sigma_*^2 X)^{-1}\right) \sim N\left(\widehat{\beta}_N; \sigma_*^2 V_{\widehat{\beta}_N}\right). \end{aligned} \quad (12)$$

Thus, normal distribution is a conjugate prior.

2.3.2. *Posterior for the Model Parameter for  $\sigma^2$ .* The posterior distribution for  $\sigma^2$  is obtained by

$$\text{Pr}(\sigma^2|y, \beta_*, \mathcal{O}_*) \propto \text{Pr}(\sigma^2) \text{Pr}(y|\beta_*, t, n\sigma^2 q, h\mathcal{O}_*), \quad (13)$$

where the prior distribution is the first term and the second is the likelihood. The posterior distribution for all the three prior distributions is a scaled-inverse- $\chi^2$  of the form and is given by

$$(\sigma^2|y, \beta_*, \mathcal{O}_*) \sim \chi_{\text{Sci}}^2(v, Q). \quad (14)$$

The choice of prior distribution determines the parameters  $(v, Q)$ . In a particular case of the inverse-gamma distribution, the scaled-inverse  $\chi^2$  is the conjugate prior for  $\sigma^2$ . This conjugate distribution is specified by two hyperparameters:

$$(\sigma^2|\beta_*, \mathcal{O}_*) \sim \chi_{\text{Sci}}^2(n_\sigma, S_\sigma^2), \quad (15)$$

which corresponds to  $((n_\sigma S_\sigma^2)/\sigma^2) \sim \chi^2(n_\sigma)$

The posterior distribution is

$$(\sigma^2|y, \beta_*, \mathcal{O}_*) \sim \chi_{\text{Sci}}^2\left(n_\sigma + n, \frac{n_\sigma S_\sigma^2 + n\widehat{\sigma}^2}{n_\sigma + n}\right), \quad (16)$$

where  $\widehat{\sigma}^2 = (1/n)(y - X\beta)'R_y^{-1}(y - X\beta)$  is the maximum likelihood estimator for  $\sigma^2$

2.3.3. *Posterior for the Model Parameters  $\beta$  and  $\sigma^2$ .* The posterior distribution is obtained by

$$\text{Pr}(\beta, \sigma^2|y, \mathcal{O}_*) \propto \text{Pr}(\beta, \sigma^2|y, \mathcal{O}_*) \text{Pr}(y|\beta t, n\sigma^2 q, h\mathcal{O}_*). \quad (17)$$

The posterior distribution is a normal-scaled-inverse- $\chi^2$ , i.e., the product of normal and scaled-inverse- $\chi^2$  densities for both priors considered. It is given as follows:

$$(\beta, \sigma^2|y, \phi_*) \sim N(b, V) \chi_{\text{Sci}}^2(v, Q). \quad (18)$$

The prior choice determines the parameters  $(b, v, V, \text{ and } Q)$ .

2.4. *Max-Stable Models.* A general representation of max-stable processes can be described by two components, a stochastic process  $\{X(s)\}$  and a Poisson process  $\Pi$  with the intensity  $d\zeta/\zeta^2$  on  $(0, \infty)$ . Let  $\{X_i(s)_{i \in N}\}$   $X(s)$  with  $E[X(s)] = 1$  and let  $\zeta_i \in \Pi$  be points of the Poisson process. Then, the spectral representation of Schlatter's model is given by

$$Z(s) = \max_{i \geq 1} \zeta_i X_i(s), \quad s \in S. \quad (19)$$

This is called Schlatter's spectral representation. A more flexible class of max-stable processes by taking  $X_i(s)$  to be any stationary Gaussian random field with finite expectation was suggested by Schlather [7]. A stationary max-stable process with unit Fréchet margins can be obtained by

$$Z(s) = \max_i \zeta_i \max\{0, X_i(s)\}, \quad (20)$$

where  $\mu = E\{\max(0, X_i(s))\} < \infty$  and  $\{\zeta_i\}$  denote the points of a Poisson process on  $(0, \infty)$  with intensity measure  $\mu^{-1}\zeta^{-1}d\zeta$ .

The correlation  $\rho(\cdot)$  and  $\mu^{-1} = \sqrt{2\Pi}$  is used if the random process is specified for a stationary isotropic Gaussian random field  $X_i(s)$  with unit variance. This process  $Z(s)$  is called an extremal Gaussian process, and the bivariate marginal distributions are given by

$$\text{Pr}(Z(s_1) \leq z_1, Z(s_2) \leq z_2) = \exp\left\{\frac{-1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left(1 + \sqrt{1 - 2(\rho(h) + 1)\frac{z_1 z_2}{(z_1 + z_2)^2}}\right)\right\}, \quad (21)$$

where  $h$  is the Euclidean distance between station  $s_1$  and  $s_2$ . For this paper, the power exponential correlation function has been utilized. Extremal dependence was discussed by various scholars in different disciplines [2–4, 16]. Various parametric models have been suggested by different experts [6–8, 17]. The extremal  $t$  process is a model of that family that extends the Schlather model [9, 10].

**2.4.1. Specification of Parameters of Schlater's Model.** The covariate specification for the location and scale parameters for fitting Schlater's model is

$$\begin{aligned}\mu(s) &= \beta_{01} + \beta_{11}\text{latitude} + \beta_{12}\text{longitude}, \\ \sigma(s) &= \beta_{02} + \beta_{21}\text{latitude}^2 + \beta_{22}\text{longitude}, \\ \text{Shape}(\xi) &\sim \text{constant}.\end{aligned}\quad (22)$$

**2.5. Model Choice Criteria Using 5-Fold Cross Validation.** To compare the precisions of predictions and forecasts obtained from the fitted models, we used some validation criteria [7, 18–20]. We used the root-mean-squared error (RMSE), mean absolute error (MAE), relative BIAS (rBIAS), and relative mean separation (rMSEP). These validation criteria are defined as

$$\begin{aligned}\text{RMSE} &= \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{z}_i - z_i)^2}, \\ \text{MAE} &= \frac{1}{M} \sum_{i=1}^m |\hat{z}_i - z_i|, \\ \text{rBIAS} &= \frac{1}{m\bar{z}} \sum_{i=1}^m (\hat{z}_i - z_i), \\ \text{rMSEP} &= \frac{\sum_{i=1}^m (\hat{z}_i - z_i)^2}{\sum_{i=1}^m (\bar{Z}_p - z_i)^2},\end{aligned}\quad (23)$$

where  $m$  is the total number of observations that we need to validate,  $z_i$  is the data indexed by  $i$ ,  $\hat{z}_i$  is the prediction value, and  $\bar{Z}_i$  and  $\bar{Z}_p$  are the arithmetic mean of the observations and predictions, respectively.

## 3. Results and Discussion

### 3.1. Results

**3.1.1. Air Monitoring Sites Selected for the Study.** Twenty-five years' (from 1991 to 2015) extreme ozone data have been collected from the 29 sites. The reason for considering 1991 as the start of data collection was the existence of large number of missing data in almost all 29 sites prior to 1991. The distribution of air monitoring sites in South Korea is displayed in Figure 1.

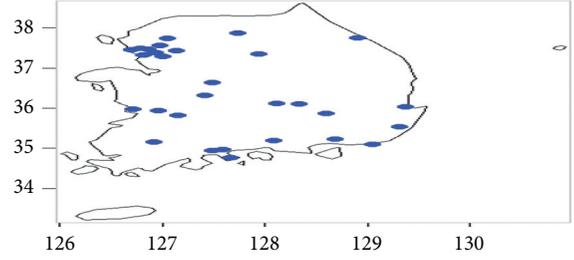


FIGURE 1: The distribution of the selected 29 air monitoring sites in South Korea.

**3.1.2. Descriptive Outputs of Maximum Ozone Concentration.** The minimum, first quartile, median, third quartile, and maximum are the descriptive measures used to portray the characteristics of extreme ozone concentrations in each 29 sites, and they are depicted in Figure 2.

Relative to the standard ozone concentration in South Korea which is 0.06 ppm, the distribution of maximum ozone concentration in all air monitoring sites was above the standards.

Time series plot of the maximum ozone concentration of all stations is displayed in Figure 3. As seen in Figure 3, we can clearly observe a positive trend in the ozone concentration with respect to time. This indicates that the location parameter of the generalized extreme value (GEV) distribution can be expressed as a linear combination of time.

**3.1.3. Station by Station Analysis by GEV.** We assumed that the maximum values of  $O_3$  per monitoring station,  $Z_1, \dots, Z_T$ , are independent and follow a GEV distribution of the following form:

$$\begin{aligned}Z_t &\sim \text{GEV}(\mu_t, \sigma, \xi), \\ \mu_t &= \beta_0 + \beta_1 t; \quad t = 1, 2, \dots, T = 29,\end{aligned}\quad (24)$$

where  $\beta_0$  is an intercept parameter while  $\beta_1$  represents a trend in  $t$  for the location parameter and  $t$  denotes the order in which the measurements were obtained. Moreover,  $\sigma$  and  $\xi$  are constants in time and denote the scale and shape parameters of the GEV distribution. The time series plot of all sites in Figure 3 can be taken as a justification to use a linear trend for the location parameter of generalized extreme value distribution of maximum ozone concentration.

**3.1.4. Prior Distribution of Parameter of GEV.** The prior is defined by the marginal distributions  $\mu \sim N(0, 10^2)$ ,  $\sigma \sim \log N(0, 10^2)$ , and  $\xi \sim N(0, 10^2)$  and the initial value for the prior of trend parameter is set to be 0.728. With the above prior specification, a Markov chain  $\theta_0, \dots, \theta_n$  with target distribution  $\pi(\theta|x)$  can be generated. Then, chains of length 10,000 have been generated using the initial values  $\theta_0 = (4, 1, 0.1, 0.1)$  and the proposal standard deviation  $s = (0.02, 0.1, 0.1, 0.1)$ . Convergence of the posteriors of all parameters has been checked using trace plots and Geweke test.

As it is depicted in Table 1, all the values of  $\beta_1$  for all the stations in the country are positive indicating that there is a positive trend in the maximum ozone concentration with time.

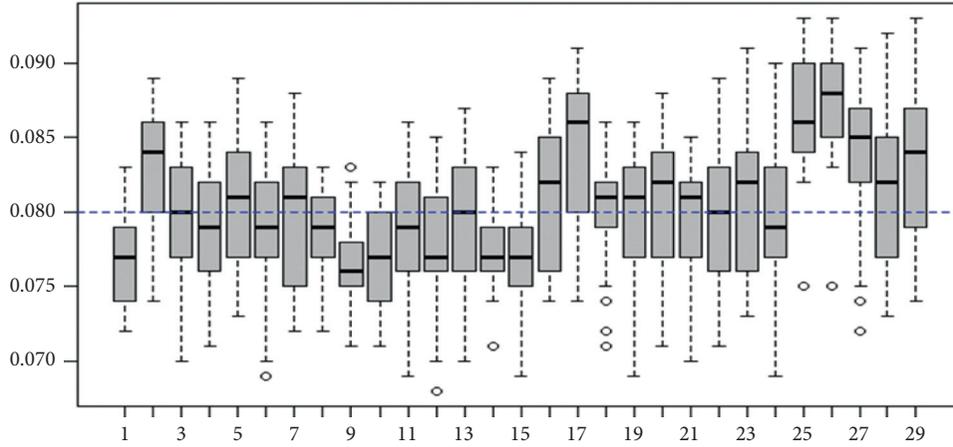


FIGURE 2: Box plot of the distribution of ozone concentration in the 29 air monitoring sites.

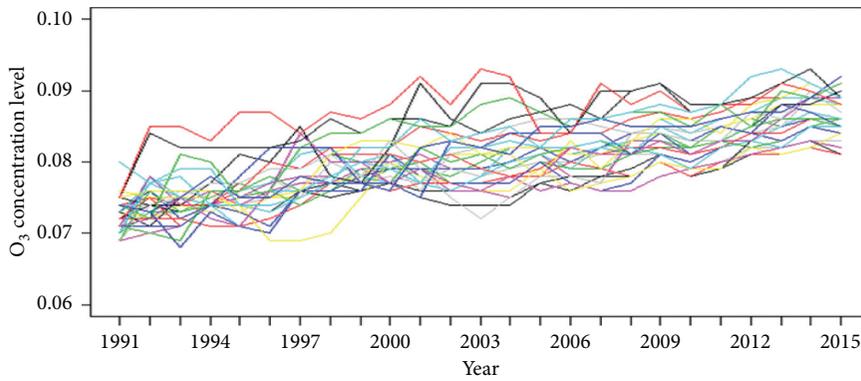


FIGURE 3: Time series plots of ozone concentration levels in the 29 air monitoring sites in South Korea.

3.1.5. *Spatial Generalized Extreme Value Analysis Assuming No Spatial Dependence.* In this perspective, the spatial extreme value analysis does not assume spatial dependence among stations in the study areas. Using this specification, the following likelihood is used for this analysis:

$$l(Z; \psi) = \sum_{i=1}^n \sum_{j=1}^k \log f_{\text{GEV}}(Z_i(x_j), \mu(Z_j), \sigma(Z_j), \xi(Z_j)). \quad (25)$$

3.1.6. *Exploratory Plot of Location Parameter versus Geographic Coordinates.* We used the exploratory plot in Figure 4 to show the relation between the dependent variable ozone extreme value with two covariates latitude and longitude. The plot shows approximately linear relation.

As seen in the above plot, there seems to be a linear trend between the covariates latitude and longitude and the ozone concentration.

3.1.7. *Exploratory Plot of Scale Parameter versus Geographic Coordinates.* Similarly, the relationship between the scale parameter of the spatial GEV model and the geographic covariates latitude and longitude has been displayed in Figure 5.

3.1.8. *Model Selection Criterion Using Takeuche Information Criteria.* Based on the linear trend relation between the ozone concentration with that of latitude and longitude, the spatial generalized extreme value analysis has been fitted. Therefore, a various combinations of covariates have been used to model a spatial GEV model. Using this specification, 9 models have been fitted. Takeuche information criteria (TIC) were used for selection. Based on this criterion, the model having small TIC would be preferable. TIC of the combinations of covariates is displayed in Table 2.

The second combination has been selected as good model for both the location and scale parameters with the consideration of the type of relationships displayed in Figures 4 and 5 and actually the smaller TIC value as well. TIC has been used as model selection criteria since Takeuchi [21] proposed Takeuchi information criterion (TIC) as a robust AIC in the spatial GEV model.

3.1.9. *Specification of Location and Scale Parameter.* The response surface with the covariates latitude and longitude has been fitted as follows:

$$\begin{aligned} \mu(s) &= \beta_{01} + \beta_{11} \text{lat} + \beta_{12} \text{lon}, \\ \sigma(s) &= \beta_{02} + \beta_{21} \text{lat}^2 + \beta_{22} \text{lon}, \\ \text{Shape}(\xi) &\sim \text{constant}. \end{aligned} \quad (26)$$

TABLE 1: The slope ( $\beta_1$ ) and its standard deviation (SD) of the trend line fitted for Bayesian extreme value analysis at 5% level of significance in all 29 air monitoring sites.

Stations	$\beta_1$ (SD)
Seoul	0.014 (0.719)
Busan	0.011 (0.312)
Daegu	0.058 (0.787)
Incheon	0.047 (0.740)
Anyang	0.017 (0.356)
Songnam	0.078 (0.810)
Uijeongbu	0.062 (0.570)
Kuangmyeon	0.029 (0.450)
Ansan	0.018 (0.630)
Gimcheon	0.069 (0.490)
Bucheon	0.009 (0.060)
Chuncheon	0.059 (0.740)
Kang-neung	0.018 (0.390)
Wonju	0.017 (0.027)
Changwon	0.038 (0.640)
Gwangju	0.027 (0.623)
Daejon	0.083 (0.590)
Ulsan	0.012 (0.777)
Suwon	0.010 (0.705)
Cheongju	0.005 (0.479)
Chinju	0.027 (0.014)
Suncheon	0.057 (0.710)
Jeonju	0.077 (0.640)
Kungsan	0.007 (0.030)
Iksan	0.097 (0.840)
Yeosu	0.128 (0.940)
Gwangyang	0.076 (0.690)
Pohang	0.004 (0.150)
Gumi	0.061 (0.736)

Referring the results in Table 3, the model coefficient for latitude covariate is negative indicating that there is an inverse relation between ozone concentration and latitude. The negative relation between ozone concentration dispersion and latitude covariate has been observed.

### 3.1.10. Interpolation of Return Level for Spatial GEV (No Spatial Dependence)

(1) *Interpolation of 5-Year Return Level.* For annual maxima data, the value of the return level can be estimated from the GEV distribution function directly.

As depicted in Figure 6, different colors show the intensity of the extreme ozone concentrations in various air monitoring sites of South Korea. The air monitoring sites found in territories having higher longitude have higher 5-year return level compared to other areas. The 5-year return level is the ozone concentration that is expected to be exceeded once in 5 years.

(2) *Interpolation of 10 Years' Return Level.* The 10 years' return levels (Figure 7) have been interpolated for the study areas using the spatial generalized extreme value analysis. A similar pattern has been observed as the 5 years' return level. Those areas having smaller longitude have a lower return level.

The higher return periods have higher return levels. To confirm this relation between return periods and return level, we interpolated the 25, 50, and 100 years' return levels. These return levels are the magnitude of the extreme ozone concentrations that are expected to be exceeded once in 25, 50, and 100 years, respectively.

(3) *Interpolation of 25 Years' Return Level.* The areas in the upper left and lower right corner of the 25 years return level plot have higher interpolated return level relative to the other level. Similarly, the areas at the center of the study area South Korea depicted smaller 25 years' return level. Moreover, areas having longitude between 127.7 and 128.4, irrespective of the latitude, have smaller 25 years' return level compared to other areas as depicted in Figure 8.

(4) *Interpolation of 50 Years' Return Level.* Similarly, the interpolations of higher return periods like 50 years and 100 years have been analyzed, and accordingly, the results are displayed in Figures 9 and 10, respectively. As seen in the 5, 10, and 25 years' return levels, the same pattern of interpolated return levels for ozone concentration has been observed for both 50 and 100 years' return periods.

### (5) Interpolation of 100 Years' Return Level

#### 3.1.11. Model Parameter Estimation of Schlater's Model.

The estimated parameter values based on Schlater's model is displayed in Table 4. These values of model parameters are used for interpolation of return levels of extreme ozone in the specified period of time. Moreover, the standard error of each parameter of Schlater's model depicts the level of dispersion of the possible model parameters in sampling distributions.

#### 3.1.12. Model Parameter Estimation of Extremal $t$ Model.

The estimated model parameter and the corresponding standard error for each parameters value based on the extremal  $t$  model specifications have been displayed in Table 5.

#### 3.1.13. Model Comparison Using 5-Fold Cross Validation.

Random classification of these 29 stations in to 5 folds has been done. Four of the folds have 6 air monitoring sites and the last fold has 5 sites. Using 5 years' return levels as actual ozone concentration and the interpolated ozone concentration as predicted values, the 7 performance measures have been computed for each fold, and we took the average of it (Table 6).

We used the root-mean-squared error (RMSE) as a measure of predictive performance. In Table 6, using the 5 years' return level, the RMSE for the Bayesian kriging model brought the minimum RMSE. This indicates that it has good prediction performance relative to the other models. Similar results have been found for the 10, 25, 50, and 100 years' return levels.

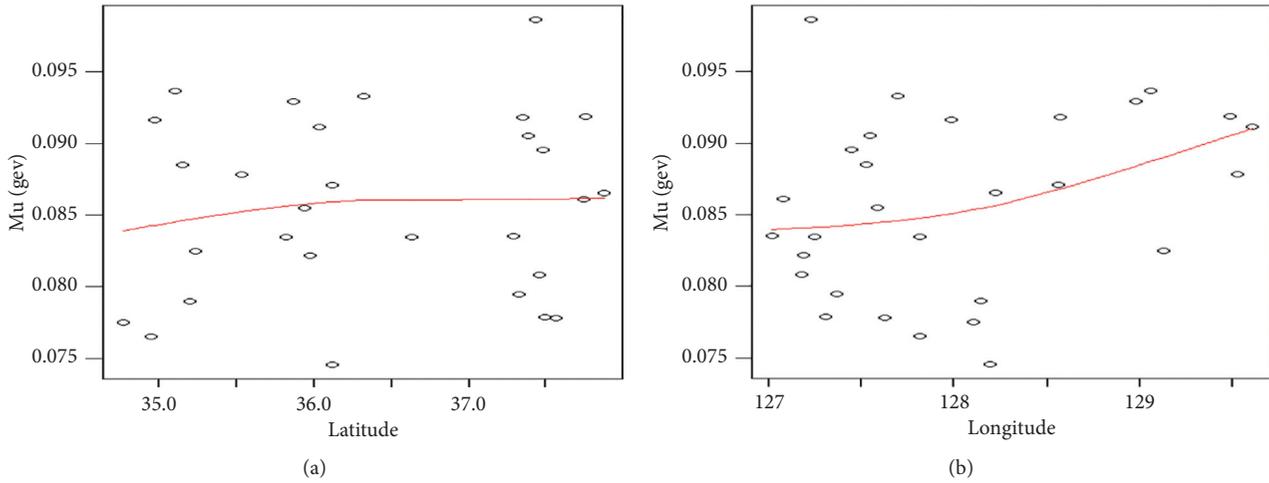


FIGURE 4: The relation between the location parameter of the spatial GEV model and the covariates latitude and longitude.

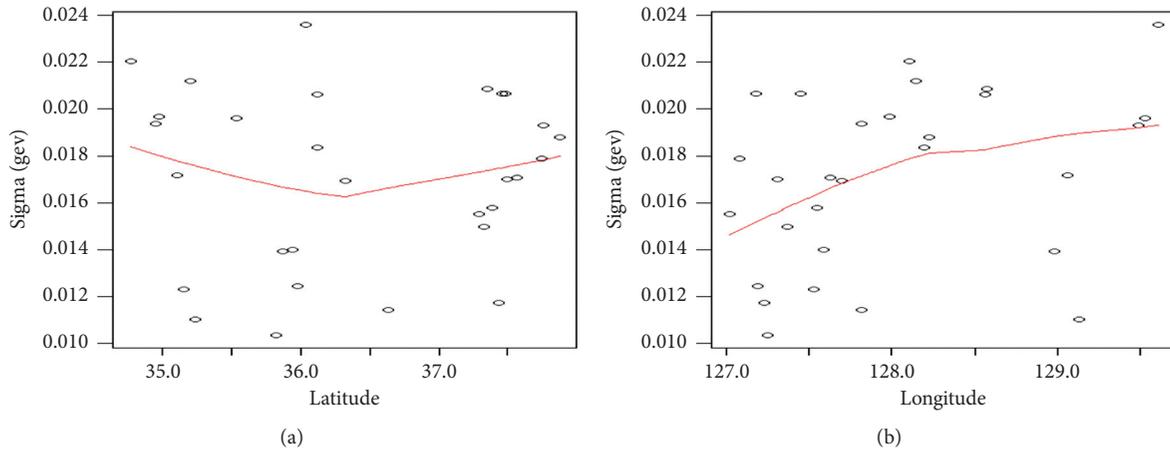


FIGURE 5: The relationship between the scale parameter of the spatial GEV model and the covariates latitude and longitude.

TABLE 2: Model selection criteria for spatial GEV model with various combinations of covariates using Takeuche information criteria (TIC).

Combinations	Covariate combinations	TIC values
1	$\mu = \sigma = \beta_{01} + \beta_{11}\text{lat} + \beta_{21}\text{lon}$	2738.585
2	$\mu = \beta_{01} + \beta_{11}\text{lat} + \beta_{12}\text{lon}$ $\sigma = \beta_{02} + \beta_{21}\text{lat}^2 + \beta_{22}\text{lon}$	2741.467
3	$\mu = \beta_{01} + \beta_{11}\text{lat} + \beta_{21}\text{lon}$ $\sigma = \beta_{02} + \beta_1\text{lat}$	2741.862
4	$\mu = \beta_{01} + \beta_{11}\text{lat}, \sigma = \beta_{02} + \beta_{22}\text{lon}$	2746.124
5	$\mu = \beta_{01} + \beta_{11}\text{lat}$ $\sigma = \beta_{02} + \beta_{21}\text{lat} + \beta_{22}\text{lon}$	2742.389
6	$\mu = \beta_{01} + \beta_{11}\text{lon}$ $\sigma = \beta_{02} + \beta_{21}\text{lat} + \beta_{22}\text{lon}$	2743.248
7	$\mu = \beta_{01} + \beta_{11}\text{lat}, \sigma = \beta_{02} + \beta_{12}\text{lat}$	2745.995
8	$\mu = \beta_{01} + \beta_{11}\text{lat}, \sigma = \beta_{02} + \beta_{22}\text{lon}$	2745.995
9	$\mu = \beta_{01} + \beta_{11}\text{lon}, \sigma = \beta_{02} + \beta_2\text{lon}$	2747.262

lon = longitude, lat = latitude, loc = location,  $\mu$  = location,  $s$  = scale,  $\xi$  = considered constant.

TABLE 3: Spatial GEV response surface coefficients estimation and its standard deviation (SD) using maximum likelihood estimation with covariates latitude and longitude and constant shape parameter.

Location parameter		
Constant (SD)	Latitude (SD)	Longitude (SD)
-0.051 (0.001)	-0.001 (0.00628)	0.0013 (0.00185)
Scale parameter		
Constant (SD)	Latitude (SD)	Shape (SD)
0.063 (0.00363)	-0.003 (0.00175)	-0.4898 (15.6)

SD = standard deviation of the parameter.

3.2. Discussion. Quentin conducted a study with the title ‘‘Comparison of spatial extreme value models application to precipitation data’’ [22]. This paper compares six models from max-stable process and other hierarchical models using the return level. But, the uniqueness of paper mainly lies on the consideration of other means of model comparison called k-fold cross validation in addition to return level measure.

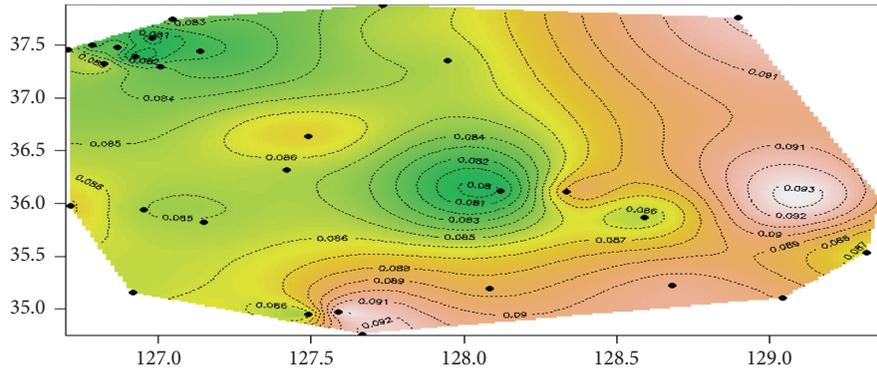


FIGURE 6: Interpolations of the 5 years' return level for extreme ozone concentration in 29 air monitoring sites versus longitude ( $x$ -axis) and latitude ( $y$ -axis) in South Korea.

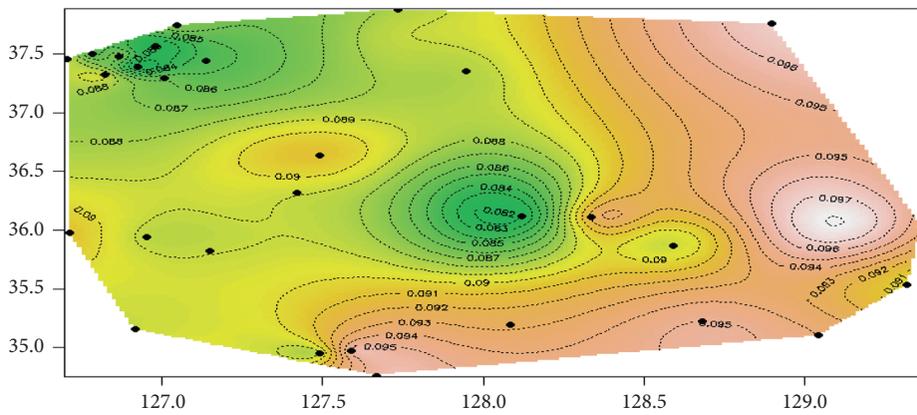


FIGURE 7: Interpolation of the 10 years' return level for ozone concentration in 29 air monitoring sites versus longitude ( $x$ -axis) and latitude ( $y$ -axis) in South Korea.

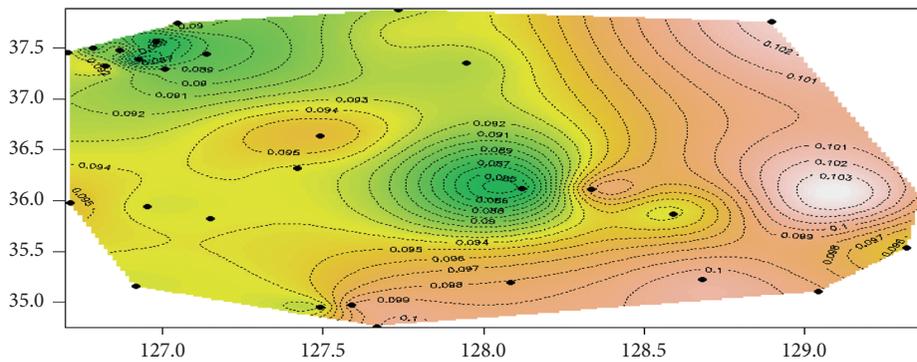


FIGURE 8: Interpolations of the 25 years' return levels for extreme ozone concentrations in 29 air monitoring sites versus longitude ( $x$ -axis) and latitude ( $y$ -axis) in South Korea.

When using the 5 years' return level as an observed value, the RMSE for the Bayesian kriging models brought about the RMSE of 0.063, and it is relatively the minimum compared to 0.077 for Schlater's model, 0.118 for the spatial GEV with MLE, and 0.083 for the extremal  $t$  model. Similarly, using the 25 years' return level as observed value, the RMSE for Schlater's, spatial GEV (MLE), Bayesian kriging, and extremal  $t$  are 0.083, 0.130, 0.071, and 0.089, respectively. The last return level considered as an observed value was the

50 years' return level. Using this value, it has brought about an RMSE of 0.089, 0.137, 0.077, and 0.091 for Schlater's, spatial GEV (MLE), Bayesian kriging, and extremal  $t$  models, respectively.

Zahidetal conducted a study on spatial prediction and concluded that Bayesian universal kriging fits better than universal kriging [23]. Based on the result from the study by Cui, Bayesian kriging was more precise as compared to those obtained with ordinary kriging [24]. A simulation study

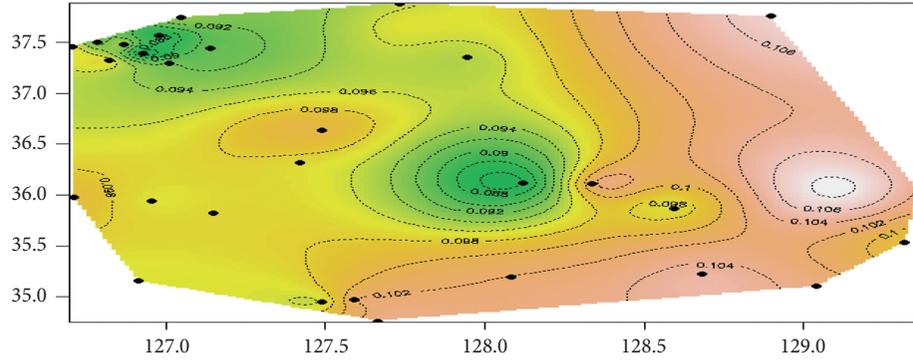


FIGURE 9: Interpolations of the 50 years' return levels for extreme ozone concentrations in 29 air monitoring sites versus longitude ( $x$ -axis) and latitude ( $y$ -axis) in South Korea.

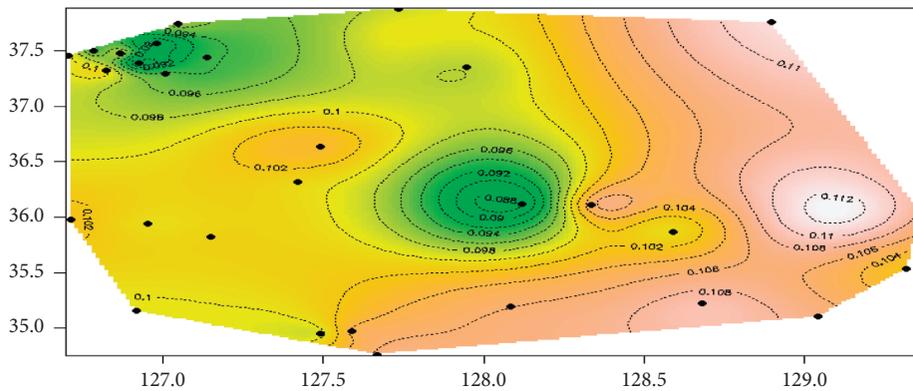


FIGURE 10: Interpolations of the 100 years' return level for extreme ozone concentrations in 29 air monitoring sites versus longitude ( $x$ -axis) and latitude ( $y$ -axis) in South Korea.

TABLE 4: Schlater's model parameter estimation using the linear correlation between location and scale parameter with that of latitude and longitude (shape parameter considered constant).

Nugget (SD)	Range (SD)	Smooth (SD)	$\mu_{\text{Constant}}$ (SD)	$\mu_{\text{Lon}}$ (SD)
0.003 (0.0012)	0.015 (0.0035)	61.825 (29.3)	-0.049 (0.00053)	-0.001 (0.00083)
$\mu_{\text{Lat}}$ (SD)	$\sigma_{\text{Cons}}$ (SD)	$\sigma_{\text{Lon}}$ (SD)	$\sigma_{\text{Lat}^2}$ (SD)	$\xi_{\text{Cons}}$ (SD)
0.001 (0.0071)	-0.064 (0.0061)	-0.001 (0.0036)	0.001 (0.00083)	-0.270 (0.034)

Lon = longitude, Lat = latitude, Cons = constant, SD = standard deviation,  $\mu$  = mean,  $\sigma$  = population standard deviation.

TABLE 5: Extremal  $t$  max-stable model parameter estimation using the location and scale with geographic coordinates (shape parameter considered constant).

Nugget (SD)	Range (SD)	Smooth (SD)	$\mu_{\text{Constant}}$ (SD)	$\mu_{\text{Lon}}$ (SD)
0.197 (0.00053)	0.546 (0.0018)	92.897 (18.9)	-0.045 (0.00016)	-0.001 (0.00073)
$\mu_{\text{Lat}}$ (SD)	$\sigma_{\text{Cons}}$ (SD)	$\sigma_{\text{Lon}}$ (SD)	$\sigma_{\text{Lat}^2}$ (SD)	$\xi_{\text{Cons}}$ (SD)
0.002 (0.0038)	-0.044 (0.013)	-0.001 (0.091)	0.005 (0.000045)	-0.222 (0.026)

Lon = longitude, Lat = latitude, Cons = constant, SD = standard deviation,  $\mu$  = Mean,  $\sigma$  = population standard deviation.

TABLE 6: Predictive performance measures of various models using 5 years' return level as observed maximum ozone value.

Performance measures	Schlather's model	Spatial GEV (MLE)	Bayesian kriging	Extremal- $t$ model
MSE	0.006	0.014	0.004	0.007
RMSE	0.077	0.118	0.063	0.083
MAE	0.008	0.089	0.002	0.009
MAPE	9.184	14.338	9.034	10.336
BIAS	-0.008	-0.030	-0.078	-0.009
rBIAS	-0.092	-0.118	-0.088	-0.104
rMSEP	1.004	1.182	1.003	1.009

highlighted that the methodology of conditional simulations for the extremal  $t$  process was effective for a large range of  $\alpha \in [1, 6]$  values which characterizes the spatial asymptotic dependence [25].

The finding from the analysis indicated that the Bayesian kriging model brought about the minimum predictive performance measure compared to other extreme value models considered in this research. For researchers interested to extend this research, I would suggest to consider the criterion of amount of time consumed for estimating model parameters and the CPU used for model comparison.

#### 4. Conclusion

In this paper, spatial GEV using MLE estimation, two models from max-stable process, Schlather's and extremal  $t$ , and the Bayesian kriging have been put in comparison with respect to the predictive performance measures.

The station by station analysis using the Bayesian approach depicted that all the 29 stations considered in this paper brought about a positive trend in the annual extreme ozone concentration. This would indicate that the ozone concentration of all these air monitoring sites has an increasing trend with time. From the interpolation of the 5, 10, 25, 50, and 100 years' return level, most stations found in the territories will have a higher ozone concentration.

#### Abbreviations

EVT:	Extreme value theory
GEV:	Generalized extreme value
MAE:	Mean absolute error
MLE:	Maximum likelihood estimation
rBIAS:	Relative bias
rMSEP:	Relative mean separation
RMSE:	Root-mean-squared error
SD:	Standard deviation
TIC:	Takeuchi information criteria
UV-B:	Ultraviolet B.

#### Data Availability

All the data used in this study are available from the corresponding author upon request.

#### Additional Points

The use of national data that can represent South Korea and the utilization of the long period (25 years) extreme ozone concentration data are among the strengths of this study. The use of national data can help us to make inference to the national level. Although the data source where I extracted the data is a reliable source, the use of secondary data by itself has some limitation.

#### Conflicts of Interest

The author declares no conflicts of interest.

#### Acknowledgments

The author is grateful to Prof. Jong Tae Kim for locating the source of the data for ozone extreme observations in all air monitoring sites selected for this research from online source.

#### References

- [1] D. H. Stedman, J. J. Margitan, and D. H. Stedman, "Atomic chlorine and the chlorine monoxide radical in the stratosphere: three in situ observations," *Science*, vol. 198, no. 4316, pp. 501–503, 1981.
- [2] S. G. Coles and J. A. Tawn, "Statistical methods for multivariate extremes: an application to structural design," *Applied Statistics*, vol. 43, no. 1, pp. 1–48, 1994.
- [3] L. De Haan and J. de Ronde, "Sea and wind: multivariate extremes at work," *Extremes*, vol. 1, no. 1, p. 7, 1998.
- [4] M. Schlather and J. A. Tawn, "A dependence measure for multivariate and spatial extreme values: properties and inference," *Biometrika*, vol. 90, no. 1, pp. 139–156, 2003.
- [5] L. Fawcett and D. Walshaw, "Sea-surge and wind speed extremes: optimal estimation strategies for planners and engineers," *Stochastic Environmental Research and Risk Assessment*, vol. 30, no. 2, pp. 463–480, 2015.
- [6] R. L. Smith, J. A. Tawn, and H. K. Yuen, "Statistics of multivariate extremes," *International Statistical Review/Revue Internationale de Statistique*, vol. 58, no. 1, pp. 47–58, 1990.
- [7] M. Schlather, "Models for stationary max-stable random fields," *Extremes*, vol. 5, no. 1, pp. 33–44, 2002.
- [8] Z. Kabluchko, M. Schlather, and L. de Haan, "Stationary max-stable fields associated to negative definite functions," *Annals of Probability*, vol. 37, no. 5, pp. 2042–2065, 2009.
- [9] P. J. Ribeiro and P. J. Diggle, "Bayesian inference in Gaussian model-based geostatistics," Technical report ST-99-08, Lancaster University, Lancashire, England, 2011.
- [10] T. Opitz, "Extremal  $t$  processes: elliptical domain of attraction and a spectral representation," *Journal of Multivariate Analysis*, vol. 122, pp. 409–413, 2013.
- [11] A. C. Davison, R. Huser, and E. Thibaud, "Geostatistics of dependent and asymptotically independent extremes," *Mathematical Geosciences*, vol. 45, no. 5, pp. 511–529, 2013.
- [12] A. C. Davison and M. M. Gholamrezaee, "Geostatistics of extremes," *Proceedings of the Royal Society A Mathematical Physical and Engineering Sciences*, vol. 468, no. 2138, 2012.
- [13] Y. Lee, S. Yoon, M. S. Murshed et al., "Spatial modeling of the highest daily maximum temperature in Korea via max-stable processes," *Advances in Atmospheric Sciences*, vol. 30, no. 6, pp. 1608–1620, 2013.
- [14] J. Blanchet and A. C. Davison, "Spatial modeling of extreme snow depth," *The Annals of Applied Statistics*, vol. 5, no. 3, pp. 1699–1725, 2011.
- [15] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, no. 2, pp. 180–190, 1928.
- [16] L. Fawcett and D. Walshaw, "Estimating return levels from serially dependent extremes," *Environmetrics*, vol. 23, no. 3, pp. 272–283, 2012.
- [17] L. De Haan, "A spectral representation for max-stable processes," *Annals of Probability*, vol. 12, no. 4, pp. 1194–1204, 1984.
- [18] Atkinson and Lloyd, *What should be done in a science assessment in protecting the ozone layer: lessons, models, and*

- prospects*, Atmospheric and Climate Sciences, vol. 6, no. 1, pp. 327–345, 1998.
- [19] Z. Kabluchko, M. Schlather, and L. de Haan, “Stationary max-stable fields associated to negative definite functions,” *The Annals of Probability*, vol. 37, no. 5, pp. 2042–2065, 2009.
- [20] H. Zhang, “Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics,” *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 250–261, 2009.
- [21] K. Takeuchi, “Distribution of information statistics and criteria for adequacy of models,” *Mathematical Sciences*, vol. 153, pp. 12–18, 1976.
- [22] Q. Sebillé, A. L. Fougères and C. Mercadier, *A comparison of spatial extreme value models. Application to precipitation data*, Ph.D. thesis, Institut Camille Jordan, Villeurbanne, France, 2016.
- [23] E. Zahidetal, “Statistical modeling of spatial extremes,” *Statistical Science*, vol. 27, no. 2, pp. 161–186, 2009.
- [24] H. Cui, “Bayesian inference in Gaussian model-based geostatistics,” Technical report ST-99-08, Lancaster University, Lancashire, England, 1995.
- [25] A. Verdin, B. Rajagopalan, W. Kleiber, and C. Funk, “A Bayesian kriging approach for blending satellite and ground precipitation observations,” *Water Resources Research*, vol. 51, no. 2, pp. 908–921, 2014.