Application of regularization method to the nonstationary GEV distribution

Ijeong Han, Yeonseon Cho, Jong-June Jeon

University of Seoul

November 9, 2019

Introduction

- Extreme events are defined to be rare and unexpected.
- Extreme rainfall events are one of the most dangerous natural hazards.
- The prediction of severity of future extreme rainfall events is crucial to planning for disaster prevention.
- The generalized extreme value (GEV) distribution is one of the most popular models to address the heavy tail property.

Recent studies of the extreme rainfall analysis are divided into two main categories:

- **Non-stationary data analysis over time**
    - [Katz et al., 2002] use the regression model with linear trend over time.
    - [Méndez et al., 2007] employ a nonlinear model with seasonal trend and long-term trend.
- **Spatial data analysis over region**
    - [Beguería and Vicente-Serrano, 2006] model the spatial patterns of the hazard of extreme rainfall with regression-based interpolators.
    - [Szolgay et al., 2009] compare several interpolators (e.g., inverse distance weight, kriging, nearest neighbor) that map the annual maximum daily precipitation at the regional scale.

Shortcomings

- Site-specific
  They are inadequate to flexible modeling such as globally or locally regional trend modeling.

- Pattern-specific

  It is difficult to judge overfitting or underfitting of the estimated model.
  e.g. linear trends and linear interpolation.

$$\mu(t) = \mu_0 + \mu_1 t.$$

$$\mu(s) = \beta_0 + \beta_1 \mathsf{lat}_s + \beta_2 \mathsf{long}_s.$$

We consider a regularization method to construct a class of the nonstationary GEV distribution models at the spatial and temporal scale simultaneously.

- The regularization is popular method to efficiently control the complexity of model in fields of statistics and computer science [Tibshirani, 1996].
- It is helpful to construct a sufficiently large class of nonstationary models at the spatial and temporal scale and to develop an efficient estimation method for the models.
- The predictive performance can be improved.

Table of Contents

- The nonstationarity over time is captured by the trend filtering [Kim et al., 2009]. -Section 2
- The spatial dependence is modelled by the thin plate splines regression [Duchon, 1977]. -Section 3
- We propose an integrated method to take advantage of both spatial and temporal structure. -Section 4
- We present results of numerical studies for each model. -Section 5

The Generalized Extreme Value (GEV) distribution is a limiting distribution of sequences of appropriately normalized maxima of identically independent distributed (IID) random variables. Let $Y_n = \max(X_1, X_2, \cdots, X_n)$, Then as $n \to \infty$, the cumulative distribution function (cdf) of the GEV distribution is defined as

$$F(y; \mu, \sigma, \kappa) = \begin{cases} \exp\left(-\left[1 + \kappa(\frac{y-\mu}{\sigma})\right]_+^{-1/\kappa}\right) & \text{if } \kappa \neq 0, \\ \exp\left(-\exp(-\frac{(y-\mu)}{\sigma})\right) & \text{if } \kappa = 0, \end{cases}$$

where location $\mu \in \mathbb{R}$, scale $\sigma > 0$ and shape parameter $\kappa \in \mathbb{R}$.

Non-stationary Model

- The identical assumption can be relaxed by introducing time-dependent parameters [Towler et al., 2010].
- For example,

$$\mu_t = \beta_0 + \beta_1 t,$$

where t denotes a time index (e.g., an indicator of year).
- Let $y_t \sim \text{GEV}(\mu_t, \sigma, \kappa)$ for $t = 1, \cdots, T$.
- Let $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_T)^\top$ and $\theta = (\sigma, \kappa)^\top$.
- Then, the maximum likelihood estimators can be determined by the minimizing the negative log-likelihood,

$$L(\boldsymbol{\mu}, \theta) = -\sum_{t=1}^{T} \log f(y_t; \mu_t, \sigma, \kappa).$$

- In this study, the l_1-**trend filtering** penalty is applied to the location parameters, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_T)$.
- Trend filtering is a method for nonparametric regression that estimates underlying trends in time series data [Kim et al., 2009].
- The piecewise linear trend in the location parameter and the other parameters can be estimated by

$$(\hat{\boldsymbol{\mu}}, \hat{\theta}) = \underset{\mu,\theta}{\operatorname{argmin}}\, L(\boldsymbol{\mu}, \theta) + \lambda \|\mathbf{D}\boldsymbol{\mu}\|_1, \tag{1}$$

  where $\|\cdot\|_1$ denotes l_1 norm, $\lambda \geq 0$ is a regularization parameter and $\mathbf{D} \in \mathbb{R}^{(T-2)\times T}$ is the second-order difference matrix.
- $\|\mathbf{D}\boldsymbol{\mu}\|_1 = \sum_{t=2}^{T-1} |(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})|$.

- When $\lambda \to 0$, the estimated parameter μ_t converges to maximum likelihood estimate.
- When $\lambda \to \infty$, the estimated μ_t converges to a point on a linear function.
- An appropriate selection of λ leads to a piecewise linear trend, which can be interpreted as abrupt change detection.

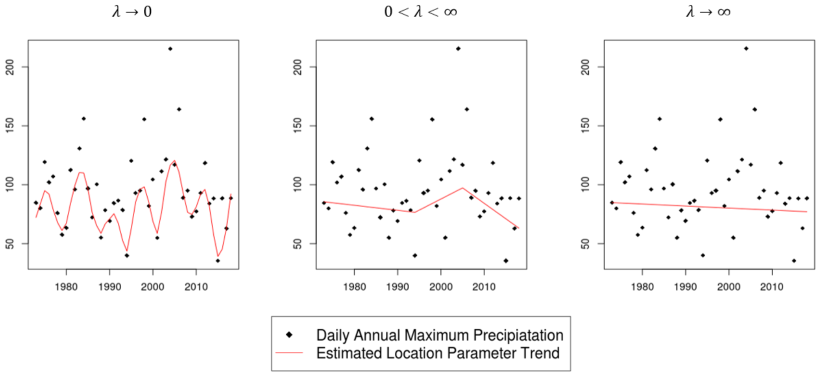


Figure: Results of Trend Filtering.

Computation

- We solve (1) by a Alternating Direction Method of Multipliers (ADMM) [Boyd et al., 2011].
- In the ADMM, the objective function (1) is reformulated as

$$\begin{aligned} \min_{\mu,\theta} \quad & L(\boldsymbol{\mu}, \theta) + \lambda \|\mathbf{v}\|_1 \\ \text{subject to} \quad & \mathbf{D}\boldsymbol{\mu} - \mathbf{v} = 0. \end{aligned} \tag{2}$$

  Here, $\mathbf{z}$ is an auxiliary variable on $\mathbb{R}^{T-2}$.
- The augmented Lagrangian function is defined by

$$L_\rho(\boldsymbol{\mu}, \theta, \mathbf{z}, \mathbf{u}) = L(\boldsymbol{\mu}, \theta) + \lambda \|\boldsymbol{v}\|_1 + \mathbf{u}^\top (\mathbf{D}\boldsymbol{\mu} - \mathbf{v}) + \frac{\rho}{2} \|\mathbf{D}\boldsymbol{\mu} - \boldsymbol{v}\|^2, \tag{3}$$

  where $\mathbf{u} \in \mathbb{R}^{T-2}$ is the dual parameter associated with the equality condition of (2) and $\rho > 0$ is augmented Lagrangian parameter.

Spatial Model

- Let $Y_{s,t}$ be random variable following the GEV distribution at site s in year t, $Y_{s,t} \sim \text{GEV}(\mu_s, \sigma_s, \kappa_s)$.
- In this study, the **thin-plate splines** (TPS) [Duchon, 1977] is employed and the location parameter is parametrized by

$$\mu_s = \mu_0 + h(\mathbf{x}_s), \quad s = 1, ..., S.$$

- μ_0 is a global location parameter across the considered sites.
- $\mathbf{x}_s = (x_{1s}, x_{2s})^\top$ is the coordinate vector whose elements denote longitude and latitude of the site s.
- $h : \mathbb{R}^2 \mapsto \mathbb{R}^1$ is a spatial map consisting of TPS basis functions,

$$h(\mathbf{x}_s) = \mathbf{B}(x_{1s}, x_{2s})^\top \boldsymbol{\beta} = \mathbf{z}_s^\top \boldsymbol{\beta},$$

with $\boldsymbol{\beta} \in \mathbb{R}^K$.

- Suppose that the data from each site are observed in the same period, $t = 1, \cdots, T$ and let $\theta_s = (\sigma_s, \kappa_s)$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_S) \in \mathbb{R}^{2S}$.
- Parameters of the spatial GEV distribution are estimated by minimizing the negative log-likelihood of $Y_{s,t}$,which is given by

$$L(\mu_0, \boldsymbol{\beta}, \boldsymbol{\theta}) = -\sum_{s=1}^{S} \sum_{t=1}^{T} \log f(y_{s,t}; \mu_0 + \mathbf{z}_s^\top \boldsymbol{\beta}, \theta_s),$$

where the total number of observations is $N = T \times S$.

- In the TPS regression, the curvature is also considered as the complexity of the estimated model.
- The TPS penalty function to $h(\cdot)$ is given by

$$J[h(\cdot)] = \iint_{\mathbb{R}^2} \left(\frac{\partial^2 h(\mathbf{x})}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial^2 h(\mathbf{x})}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 h(\mathbf{x})}{\partial x_2^2} \right)^2 dx_1 dx_2.$$

- We estimate the spatial function by minimizing the penalized negative log-likelihood,

$$L_\lambda(\mu_0, \boldsymbol{\beta}, \boldsymbol{\theta}) = L(\mu_0, \boldsymbol{\beta}, \boldsymbol{\theta}) + \lambda J[h(\cdot; \beta)], \qquad (4)$$

where $\lambda \geq 0$ is a regularization (smoothing) parameter.
- The penalized MLE of (4) is simply obtained by Newton-Raphson method.

- When $\lambda \to 0$, μ_s converges to an interpolating estimate.
- When $\lambda \to \infty$, it leads to just fitting least squares plane estimate.

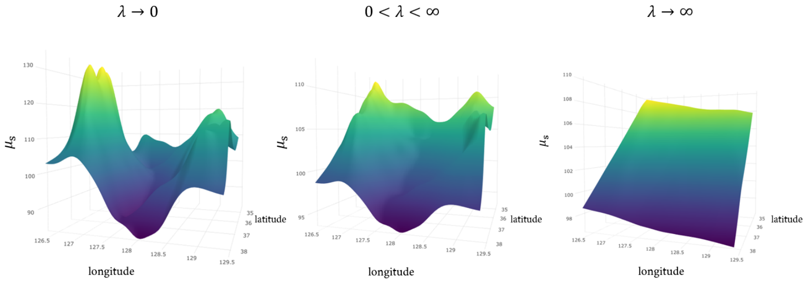


Figure: Results of Spatial Model.

Spatial-Temporal Model

- We propose an integrated method to take advantage of both spatial and temporal structure by regularization.
- Let $\mu_{s,t} = \mu_{0,t} + h(x_{1s}, x_{2s})$.
- We apply the trend filtering to $\mu_{0,t}$ for $t = 1, \cdots, T$.
- And we define the objective function

$$\min L(\boldsymbol{\mu}_0, \beta, \boldsymbol{\theta}) + \lambda_1 \|\mathbf{D}\boldsymbol{\mu}_0\|_1 + \lambda_2 J[h(\cdot;\beta)], \qquad (5)$$

  where $\boldsymbol{\mu}_0 = (\mu_{0,1}, ..., \mu_{0,T})$.
- In the ADMM, the objective function (5) is reformulated as

$$\begin{aligned} \min_{\mu_0,\beta,\theta} \quad & L(\boldsymbol{\mu}, \theta) + \lambda_1 \|\mathbf{v}\|_1 + \lambda_2 J[h(\cdot;\beta)] \\ \text{subject to} \quad & \mathbf{D}\boldsymbol{\mu_0} - \mathbf{v} = 0. \end{aligned}$$

- We use model selection criteria like AIC (Akaike information criterion).
- For the non-stationary model, df$(\hat{\boldsymbol{\mu}}) = \|\mathbf{D}\hat{\boldsymbol{\mu}}\|_0 + 1$.
- For the spatial model, df$(\hat{\beta}) = \text{Tr}\left(\mathbf{H}_\lambda\right)$, where $\mathbf{H}_\lambda = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z} + \lambda\boldsymbol{\Omega}_{\mathbf{Z}})^{-1}\mathbf{Z}^\top$.
- Then the AIC of the estimated spatial-temporal model are as follows.

$$\text{AIC}(\hat{\boldsymbol{\mu}}_0, \hat{\beta}, \hat{\boldsymbol{\theta}}) = 2L(\hat{\boldsymbol{\mu}}_0\hat{\beta}, \hat{\boldsymbol{\theta}}) + 2\text{df}(\hat{\boldsymbol{\mu}}_0, \hat{\beta}, \hat{\boldsymbol{\theta}}).$$

Here, df$(\hat{\boldsymbol{\mu}}_0, \hat{\beta}, \hat{\boldsymbol{\theta}}) = (\|\mathbf{D}\hat{\boldsymbol{\mu}}_0\|_0 + 1) + \text{Tr}\left(\mathbf{H}_{\lambda_2}\right) + 2S$.

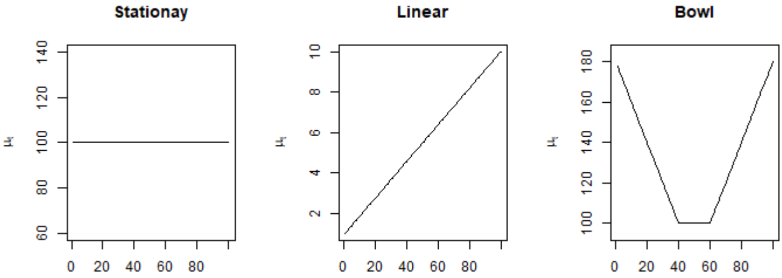
- We exploit predictive performance of estimated models according to regularization parameters.
- We evaluate results of AIC based on the predictive performance achieved by an optimal tuning parameter.
- To assess the performance of the model, root mean-squared error (RMSE) are computed.

$$\text{RMSE} = \sqrt{\frac{1}{S(T+2)} \sum_{s=1}^{S} (\eta_s - \hat{\eta}_s)^2},$$

where $\eta_s = (\boldsymbol{\mu}_s, \sigma_s, \kappa_s)$.

Simulation 1: Non-stationary Model

- Simulation 1 explores in three types of scenarios for μ_t.
  - Scenario 1: $\mu_t = 100$.
  - Scenario 2: $\mu_t = 9 + 0.1t$.
  - Scenario 3: $\mu_t = (180 - 2t)I_{(t<40)} + 100I_{(40 \le t<60)} + (-20 + 2t)I_{(t \ge 60)}$.
- For 100 replicates, we simulated datasets consisting of $T = 100$ observations, $y_t \sim GEV(\mu_t, 40, 0.1)$ for $t = 1, ..., T$.



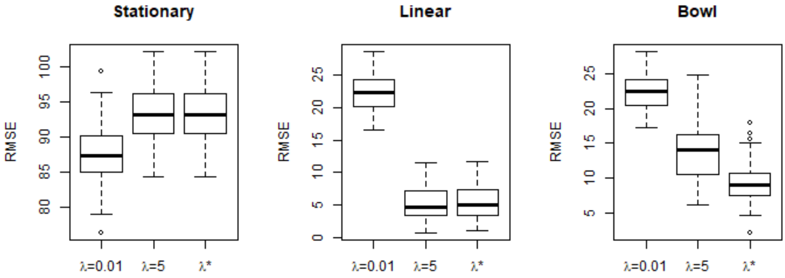


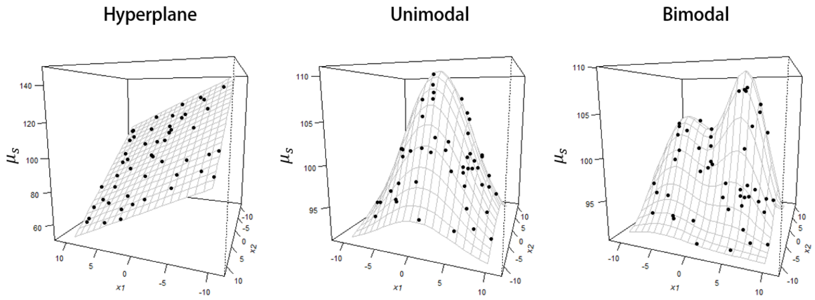
Figure: Boxplot of RMSE values based on 100 runs for the estimated GEV parameter in the simulation 1. λ^* is the most frequent selected lambda.

Simulation 2: Spatial Model

- For 100 replicates, we obtain $y_{s,1}, ..., y_{s,T}$ from

$$y_{s,t} \sim GEV(\mu_s, \sigma_s, \kappa_s), \qquad t = 1, ..., T.$$

  Here, $\sigma_s \sim \mathrm{N}(40, 2^2)$ and $\kappa_s \sim U(0.1, 0.25)$.
- We used $S = 50$ sites randomly located on (x_1, x_2) within a spatial domain of $\mathbb{R}^2 = [-10, 10]^2$ for all scenario.
- Simulation 2 explores in three types of scenarios for μ_s.



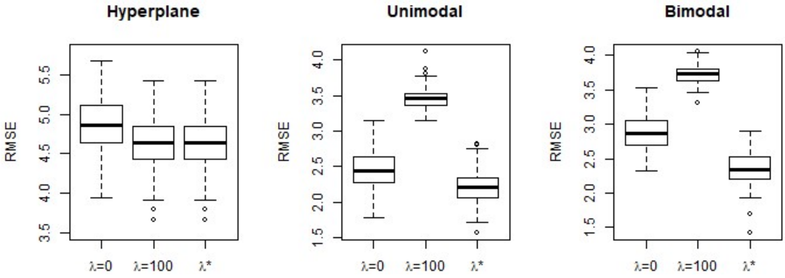


Figure: Boxplot of root mean square error (RMSE) based on 100 runs for the GEV parameters in the simulation 2. λ^* is the most frequent selected lambda.

Simulation 3: Spatial-Temporal Model

- In Simulation 3, we consider the spatial-temporal model that assumes

$$\mu_{s,t} = \mu_{0,t} + h(x_{1s}, x_{2s}).$$

- For simulation setting, the pattern of bowl having 2 change points of means are considered for $\mu_{0,t}$ as described in Simulation 1.
- And the unimodal surface are considered for $h(x_{1s}, x_{2s})$ as described in Simulation 2.
- For 100 replicates, we sampled from

$$y_{s,t} \sim GEV(\mu_{s,t}, \sigma_s, \kappa_s), \qquad t = 1, ..., T.$$

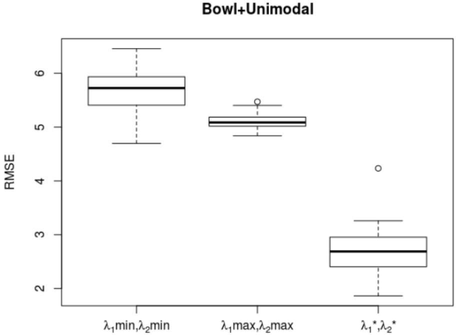


Figure: Boxplot of root mean square error (RMSE) based on 100 runs for the GEV parameters in the simulation 3. A pair of λ_1^*, λ_2^* is the most frequent selected lambda.

Real Data Analysis: Spatial-Temporal Model

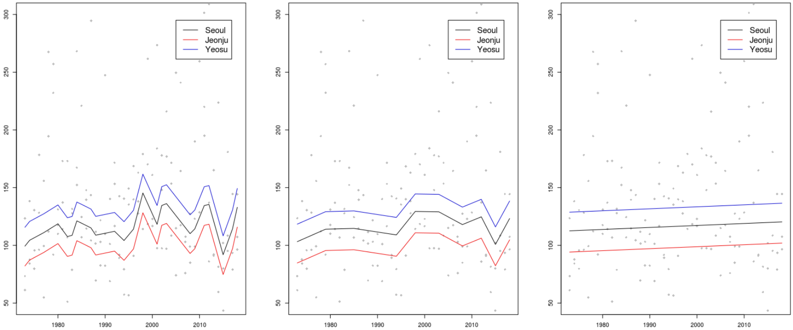


Figure: Results of Spatial-Temporal Model: points denote the observed daily AMPs of the South Korea; lines are estimated trend of the each site; λ_2 is fixed.

Conclusion

- There are relatively few example in hydrology where characteristics of extreme value have been studied with regularization approaches.
- Numerical analysis was conducted and the predictive performances from the model selected by AIC were better for RMSE.
- The proposed GEV spatial-temporal model in this study can be applied to flexibly find the global and local regional trend.

Beguería, S. and Vicente-Serrano, S. M. (2006).
Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques.
Journal of applied meteorology and climatology, 45(1):108–124.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011).
Distributed optimization and statistical learning via the alternating direction method of multipliers.
Foundations and Trends® in Machine learning, 3(1):1–122.

Duchon, J. (1977).
Splines minimizing rotation-invariant semi-norms in sobolev spaces.
In *Constructive theory of functions of several variables*, pages 85–100. Springer.

Katz, R. W., Parlange, M. B., and Naveau, P. (2002).
Statistics of extremes in hydrology.
Advances in water resources, 25(8-12):1287–1304.

Kim, S.-J., Koh, K., Boyd, S., and Gorinevsky, D. (2009).
ℓ_1 trend filtering.
SIAM review, 51(2):339–360.

Méndez, F. J., Menéndez, M., Luceño, A., and Losada, I. J. (2007).
Analyzing monthly extreme sea levels with a time-dependent gev model.
Journal of Atmospheric and Oceanic Technology, 24(5):894–911.

Szolgay, J., Parajka, J., Kohnová, S., and Hlavčová, K. (2009).
Comparison of mapping approaches of design annual maximum daily precipitation.
Atmospheric Research, 92(3):289–307.

Tibshirani, R. (1996).
Regression shrinkage and selection via the lasso.
Journal of the Royal Statistical Society: Series B (Methodological), 58(1):267–288.

Towler, E., Rajagopalan, B., Gilleland, E., Summers, R. S., Yates, D., and Katz, R. W. (2010).
Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory.
Water Resources Research, 46(11).