A hierarchical Bayesian non-asymptotic extreme value model for spatial data

Federica Stolf & Antonio Canale

Statistical methods and models for complex data — Poster Session: 21 September 2022

Introduction

We consider the problem of predicting **extreme values** for environmental data. Events such as extreme precipitation and storm wind speed are characterized by limited predictability and are spatial in extent.
The goal is to produce **maps of precipitation return levels** of the area of interest.
We propose a **spatial hierarchical Bayesian extreme value model (sHMEV)** where the asymptotic assumption, typical of the traditional extreme value theory, is relaxed.
Spatial dependence is characterized by geographical covariates and effects not fully described by the covariates are captured by spatial structure in the hierarchies.

Motivating application

- We analyze a collection of spatially distributed **daily time series of precipitation** in North Carolina (USA).
- The data refers to 27 different stations and a significant fraction of the available records is longer than 100 years.
- The data are extracted from the United States Historical Climatological Network (**USHCN**) data.
- Figure 1 shows the station locations.

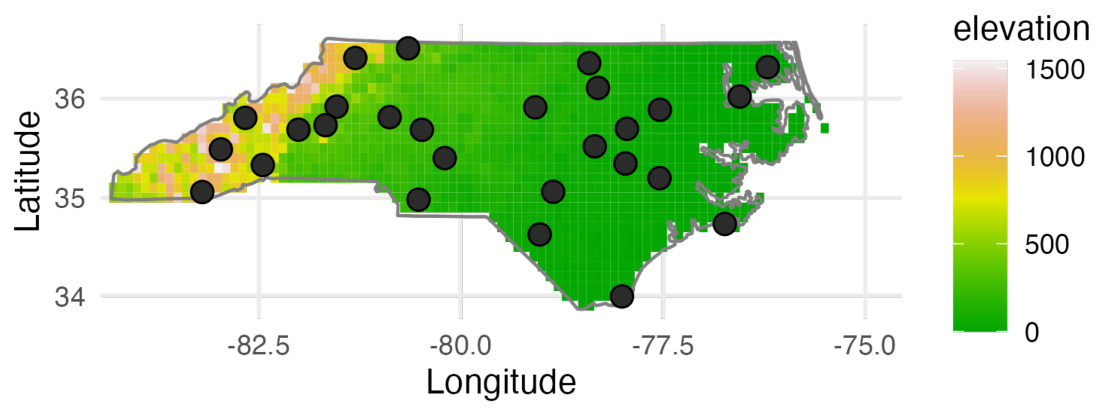

Figure 1. Map of North Carolina showing the sites and altitude in meters above sea level of the weather stations.

Background

Notation

- $x_{ij}(s)$ is the magnitude of the i-th event within the j-th block for the site s.
- J is the number of block, S is the number of sites and $n_j(s)$ is the number of events observed within the j-th block for the site s.
- The quantity of interest is the **cdf of block maxima** $Y_j(s) = \max_i\{X_{ij}(s)\}$:

$$\Pr(Y_j(s) \leq y) = F(y; \theta_j(s))^{n_j(s)}.$$

Traditional approaches

- block maxima with generalized extreme value distribution (GEV) (Gnedenko, 1943);
- peaks over threshold (POT) with generalized Pareto distribution (Pickands, 1975).

These methods are based on asymptotic arguments and focus only on a small portion of data.
Possible problems:

- × the number of events per block may be often not large enough for the asymptotic argument to hold;
- × the assumption of a constant parent distribution is unrealistic in many contexts.

A spatial Hierarchical Bayesian extreme value model

General formulation

- $x_{ij}(s)$ conditionally on unobserved latent processes are realizations of conditionally independent random variables with common cdf $F(\cdot;\theta_j(s))$, with $\theta_j(s)$ unknown parameter vector;
- $n_j(s)$ are realizations of random variables with probability distribution $p\{\cdot;\lambda(s)\}$, where $\lambda(s)$ is an unknown parameter vector.

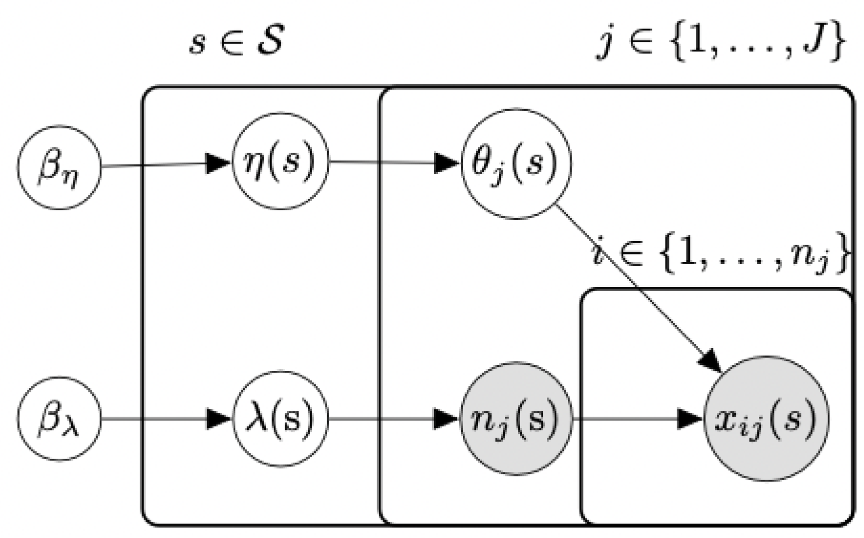

Figure 2. Graphical representation of the spatial hierarchical model.

- We further assume that $\theta_j(s)$ are realizations of a stochastic process with conditional probability density function $g\{\cdot;\eta(s)\}$, where $\eta(s)$ is an unknown vector of parameters.
- The latent spatial processes, $\eta(s)$ and $\lambda(s)$ are deterministic functions of parameter β and spatial covariates.
- We can estimate the cdf of maxima marginalizing out the variables $n_j(s)$ and $\theta_j(s)$, i.e.

$$\sum_{n\geq 0}\int_\Theta F\{y;\theta(s)\}^{n(s)}\, p\{n(s);\lambda(s)\}\, g\{\theta(s);\eta(s)\}\mathrm{d}\theta(s).$$

Prior elicitation and posterior computation

- In defining the prior distributions we seek to harness information on the physical processes generating the data, avoiding strongly uninformative priors.
- Inference about the parameters and spatio-temporal predictions are obtained via MCMC simulation.

Main advantages

- ✓ The Bayesian approach allows the inclusion of valuable prior information that are often present in environmental modeling.
- ✓ Spatial modeling of extremes is expected to reduce the overall uncertainty in quantile estimates, by borrowing strength across different sites.
- ✓ Uncertainty measures on the quantile estimates result naturally from the sampling procedure in the Bayesian framework.
- ✓ By modeling the ordinary events the variability of estimates decreases, especially when the sample size is low.

Simulation study

- Scenarios: WEI correctly specified, GAM and WEI_{gp} misspecified.
- Competitors: sHMEV, HMEV (Zorzetto et al., 2020), Bayesian implementation of GEV.
- **Fractional square error** (FSE): evaluate the predictive accuracy in estimating the distribution of block maxima, considering both precision and variability of estimates.

Figure 3 shows the empirical distribution of the FSE over the sites, computed on the test set.

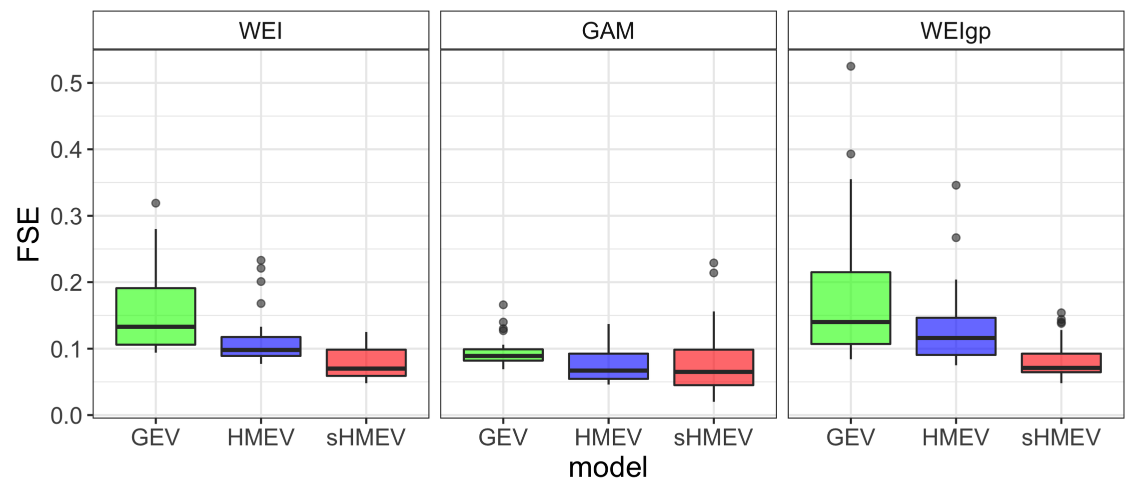

Figure 3. Fractional square error computed for the 3 different model specifications.

Figure 4 shows quantile versus return time plots for two sites.

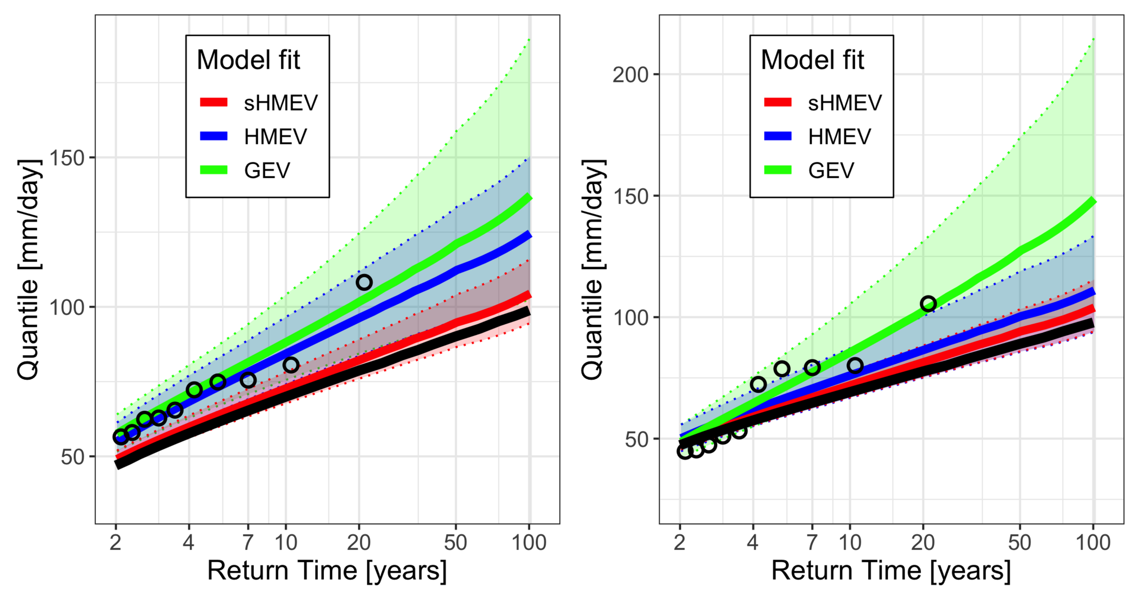

Figure 4. Expected value of the quantile for a given return time and 90% credibility intervals. Circles represent the observed block maxima on training set, while the black lines report the quantiles computed from the true sHMEV model.

North Carolina rainfall data analysis

Figure 5 shows the maps of the predictive pointwise 25 and 50 year return level estimates for rainfall (mm) obtained from the sHMEV model.

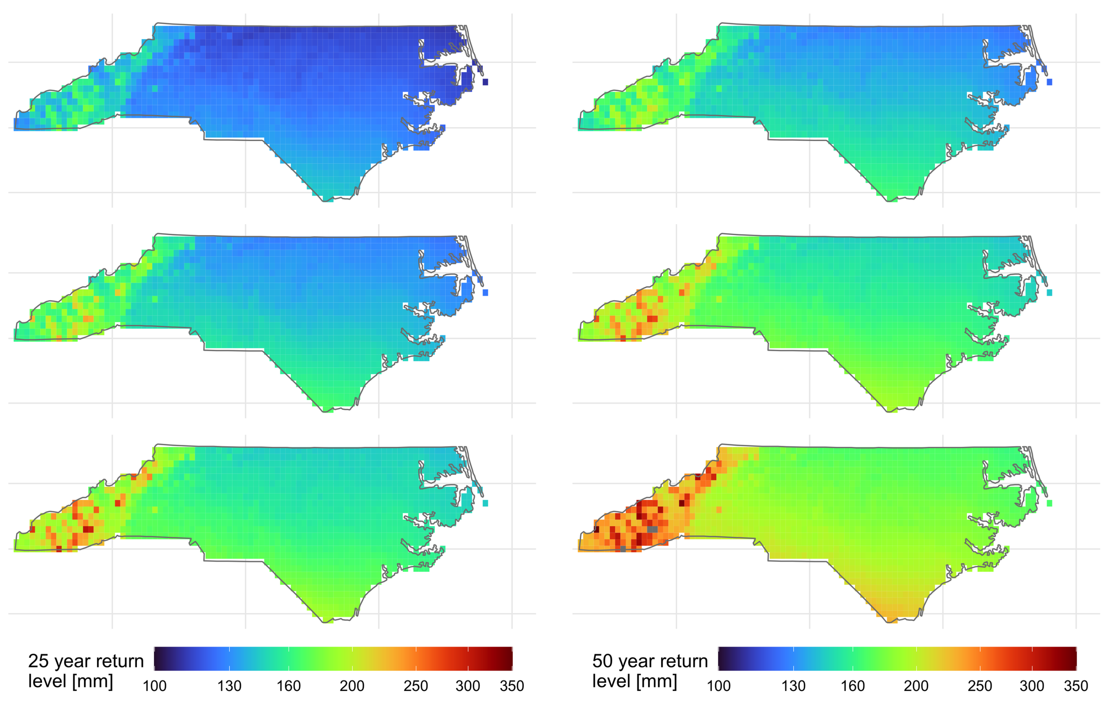

Figure 5. The upper and bottom rows show the lower and upper bounds of the 90% pointwise credible intervals, the middle row shows the predictive pointwise posterior mean.

References

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, **102**, 824 – 840.

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, **27 (2)**, 161 – 186.

Pickands III, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, **3 (1)**, 119 – 131.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, **44 (3)**, 423 – 453.

Stolf, F. and Canale, A. (2022). A hierarchical Bayesian non-asymptotic extreme value model for spatial data. *arXiv:2205.01499*.

Zorzetto, E., Canale, A., and Marani, M. (2020). Bayesian non-asymptotic extreme value models for environmental data. *arXiv:2005.12101*.

Contact information

Federica Stolf, Ph.D. student
Department of Statistical Sciences, University of Padova
federica.stolf@phd.unipd.it
federicastolf