Supporting Information for the manuscript: Bayesian non-asymptotic extreme value models for environmental data

Enrico Zorzetto, Antonio Canale, and Marco Marani

4/30/2020

Details on the Bayesian GEV and POT models implemented

Here we breafly review the main extreme value statistical models consistently with the notation used in the main manuscript. The two approaches detailed below are the Generalized Extreme Value distribution used as a model for block maxima series, and the Peak Over Threshold (POT) defined as a Poisson point process. For a complete discussion, see Coles (2001) or De Haan and Ferreira (2007). These models are commonly used for EV analysis and statistical software that implement these technique is available, such as for example the *extRemes* R package (Gilleland, Katz, and others 2016). The Bayesian (Hamiltonian Monte Carlo) estimation for the GEV and POT models used in our study is described below and implemented in the HMEV R package.

The Generalized Extreme Value distribution

The (GEV) distribution (Von Mises 1936) has cdf

$$Pr\left(Y \leq y\right) = F_{GEV}\left(y \mid \mu, \sigma, \xi\right) = \exp\left\{-\left(1 + \frac{\xi}{\sigma}\left(y - \mu\right)\right)_{+}^{-1/\xi}\right\}. \tag{S1}$$

where $\mu \in R$ and $\sigma \in R^{+}$ are location and scale parameters respectively, while $\xi \in R$ is a shape parameter, and $(\cdot)_{+} = \max\{0, \cdot\}$. Depending on the value of ξ, the GEV family encompasses a double exponential, an heavy-tailed, and an upper bounded distribution.

The GEV parameters are generally estimated by means of Maximum Likelihood (ML), Penalized ML (Martins and Stedinger 2000), L-Moments (Hosking 1990) or Bayesian methods (S. G. Coles and Tawn 1996, @coles2003fully). Confidence intervals can be obtained with the Delta method in the case of ML inference (Coles 2001) or using Bootstrap techniques when the L-moments are used to fit the distribution. Generally L-moments perform better than ML in the case of small samples, even though asyntotic theory for confidence intervals is not available. Bayesian Methods allow better characterization of the variability of extimated values (S. G. Coles and Tawn 1996; S. G. Coles and Powell 1996; Coles, Pericchi, and Sisson 2003; Stephenson and Tawn 2004). Here we use Bayesian methods for fitting the GEV model (we implemented a Stan model, sampling from the posterior using the Hamiltonian Monte Carlo sampler as done for HMEV). This gives us Bayesian probability intervals for the GEV quantiles for a given return time. We elicit the prior distributions for the three GEV parameters as follows: For the shape parameter, we select as prior distribution a normal distribution centered in 0.114 with a standard deviation $\sigma = 0.125$. This choice matches the expected value suggested globally for daily rainfall extremes (Koutsoyiannis 2004), while the overall shape of the Prior distribution closely match the Geophysical Prior proposed by Martins and Stedinger (2000) in order to guide inference towards reaslistc values of the shape parameter in our application. For the shape and scale parameters, we select informative priors centered around the mean and standard deviation of the annual maxima samples, in order to exploit prior knoiwledge on the expected value and characteristic variability of annual maxima values.

The Peak Over Threshold method

The GEV distribution also arises as limiting model for the block maxima of a point process with Poisson-distributed arrival of events with magnitudes distributed according to a Generalized Pareto distribution, which is often used to model exceedance over a high threshold (Davison and Smith 1990). The distribution of exceedances over a high threshold q is

$$Pr\left(Y > y\right) = Pr\left(X > x + q | X > q\right) \simeq \frac{1 - F(x+q)}{1 - F(u)} = 1 - F_{GPD}\left(x | q, \beta, \kappa\right) = \left(1 + \frac{\kappa}{\beta}\left(x - \mu\right)\right)^{-1/\kappa} \quad \text{(S2)}$$

In this case, the distribution of block maxima reads

$$F_{PP}\left(x\right) = \sum_{n=o}^{\infty} p_n(n|\lambda) F_{GPD}\left(x | \beta, q, \kappa\right) = 1 - \sum_{n=o}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!}\left(1 + \frac{\kappa}{\beta}\left(x - q\right)\right)^{-1/\kappa} = F_{GEV}\left(x | \mu, \sigma, \xi\right) \quad \text{(S3)}$$

where the parameters of the GEV are obtained as follows $\kappa = \xi$, $\beta = \sigma + \xi\left(q - \mu\right)$, and $\lambda = \left(1 + \xi\frac{\xi}{\beta}\left(q - \mu\right)\right)^{-1/\xi}$

Here we use Bayesian inference for this Poisson-GPD model for threshold exceedances. We elicit the prior distribution for the model parameters as follows: For the Pareto shape parameter, we choose the same informative prior elicited for the GEV - annual maxima model. For the Poisson rate, we select a prior distribution centered on 4 events/year, which appears reasonable value since we fix the threshold based on a fixed number of average exceedances in each block. For our analysis, we choose an automated threshold such that 5% of the non-zero values are above threshold.

Once we draw from the posterior samples for these parameters $\lambda^{(s)}$, $\beta^{(s)}$, and $\kappa^{(s)}$, for $s = 1, \ldots, S$, we can compute the posterior probability for the quivalent parameters of the GEV model as follows:

$\mu^{(s)} = q + \frac{\beta^{(s)}}{\kappa^{(s)}}\left({\lambda^{(s)}}^{\kappa^{(s)}} - 1\right)$ $\sigma^{(s)} = \beta^{(s)} {\lambda^{(s)}}^{\kappa^{(s)}}$ $\xi^{(s)} = \kappa^{(s)}$

Definitions of lppd and lpml

Evaluating the predictive accuracy of extreme value models in estimating the right tail of the distribution is indeed an inherently challenging task, as high quantiles are, by definition, poorly represented in the available samples. For this reason, cross validation techniques are rarely used to assess the performance of fitted extreme value models. In our analysis, however, we harnessed the considerable length of the synthetic data sets available here in order to extensively test the performance of different methods using both in-sample and out-of-sample validation techniques. To this end, we employ both in-sample and out-of-sample measures. The log pointwise predictive density (lppd) (Gelman et al. 2013) computed both for the in-sample data and for the out-of-sample data is used as a measure of global performance of the models. This measure can be directly estimated from S draws from a MCMC sameple as

$$\widehat{lppd} = -\frac{1}{M}\sum_{i=1}^{M} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i | \theta^{(s)})\right) \quad \text{(S4)}$$

This quantity, if computed for in-sample annual maxima data y_i, $i = 1, \ldots, M$ is expected to overestimate the expected log predictive density (elpd) for the same data points. This overestimation is generally corrected by quantifying the overfit of the model using some estimate of the effective number of parameter of the model. Common corrections used in practice include the Deviance Information Crtiterion (DIC), Watanaibe-Aikake information criterion (WAIC) (Gelman et al. 2013), or leave one out techniques such as the log posterior marginal likelihood (LPML) (Gelfand and Dey 1994), or leave-one-out based on Pareto Smoothed Importance Sampling (PSIS) (Vehtari, Gelman, and Gabry 2017).

Here we used the logarithm of the pseudo-marginal likelihood (lpml), a convenient index that directly accounts, at no additional computational cost, for a leave-one-out cross validation measure (Gelfand and Dey 1994). Notably, since the lpml approximates the expected log pointwise predictive density, the difference between the in-sample lppd and the lpml represents the number of effective parameters of a model (Vehtari, Gelman, and Gabry 2017) and thus will be used to quantify the tendency of different models to overfitting.

$$lpml = -\frac{1}{M}\sum_{i=1}^{M}\log\left(CPO_i\right) \tag{S5}$$

where we add the factor $-1/M$ to reduce its variation with sample size. Here CPO_i is the *Conditional Predictive Ordinate* statistics introduced by (Gelfand, Dey, and Chang 1992) and (Gelfand and Dey 1994), which estimates the probability of observing a value y_i given that $\mathbf{y}_{-i}$ has been observed. CPO_i can be obtained as follows:

$$CPO_i = \left\{\int \frac{1}{p\left(y_i \mid \theta\right)} p\left(\theta \mid \mathbf{y_{-1}}\right) d\theta\right\}^{-1} \tag{S6}$$

The CPO can be computed as the geometric mean of the likelihood of the data (annual maxima) given the model. Sampling from the posterior, one can computeCPO_i as follows:

$$\widehat{CPO_i} = \left[\frac{1}{S}\sum_{s=1}^{S}\frac{1}{p\left(y_i \mid \theta^{(s)}\right)}\right]^{-1} \tag{S7}$$

Therefore CPO_i can be computed as the harmonic mean of the likelihood of the annual maxima y_i given the model. The best model amongst the ones tested here will be characterized by the largest value of *lpml*. Note that given the properties of this measure of predictive performance (in essence equivalent to a leave-one-out cross validation), we can use *lpml* when a single sample is available for calibration and testing of the EV models. Therefore, *lpml* can be used to decide which of these models is best for a given sample/ application. We therefore propose to use it when only short samples are available to decide which EV model we shouls use for the distribution of annual maxima. We note that *lpml* can be difficult to use for the GEV distribution when applied to a validation sample different from the one used to fit the distribution, since it is possible that the modelled posterior distribution assign zero probability to some of the observed values.

Additional tables

%

Stations: Effective number of parameters

As we can see in Figure S1 we report the effective number of paramaters computed for observed data. Additionally, we include the same quantity computed in a different way (Watanabe-Aikake information criterion) to show that results do not change too much.

Table S1: Values of the constants used to elicit the prior hyperparameters of the model for event magnitudes.

Parameter	Value		
$\alpha_{\mu\gamma0} =$	$i_{\mu\gamma0}$		
$\beta_{\mu\gamma0} =$	$i_{\mu\gamma0} \cdot e_{\mu\gamma0}$		
$\alpha_{\mu\delta0} =$	$i_{\mu\delta0}$		
$\beta_{\mu\delta0} =$	$i_{\mu\delta0} \cdot e_{\mu\delta0}$		
$\alpha_{\sigma\gamma0} =$	$i_{\sigma\gamma0}$		
$\beta_{\sigma\gamma0} =$	$i_{\sigma\gamma0} \cdot e_{\sigma\gamma0} \cdot v_{\sigma\gamma0}$		
$\alpha_{\sigma\delta0} =$	$i_{\sigma\delta0}$		
$\beta_{\sigma\delta0} =$	$i_{\sigma\delta0} \cdot e_{\sigma\delta0} \cdot v_{\sigma\delta0}$		
Constant	Value	Meaning	Empirical Prior
$i_{\mu\gamma0}$	10	shape informativeness	no
$i_{\mu\delta0}$	10	scale informativeness	no
$i_{\sigma\gamma0}$	10	shape informativeness	no
$i_{\sigma\delta0}$	10	scale informativeness	no
$e_{\mu\gamma0}$	0.7	expected value shape	no
$e_{\mu\delta0}$	$10/\Gamma\,(1+1/0.7)$	expected value scale	no (possible)
$v_{\sigma\gamma0}$	0.05	expected variability shape	no
$v_{\sigma\delta0}$	0.25	expected variability scale	no (possible)

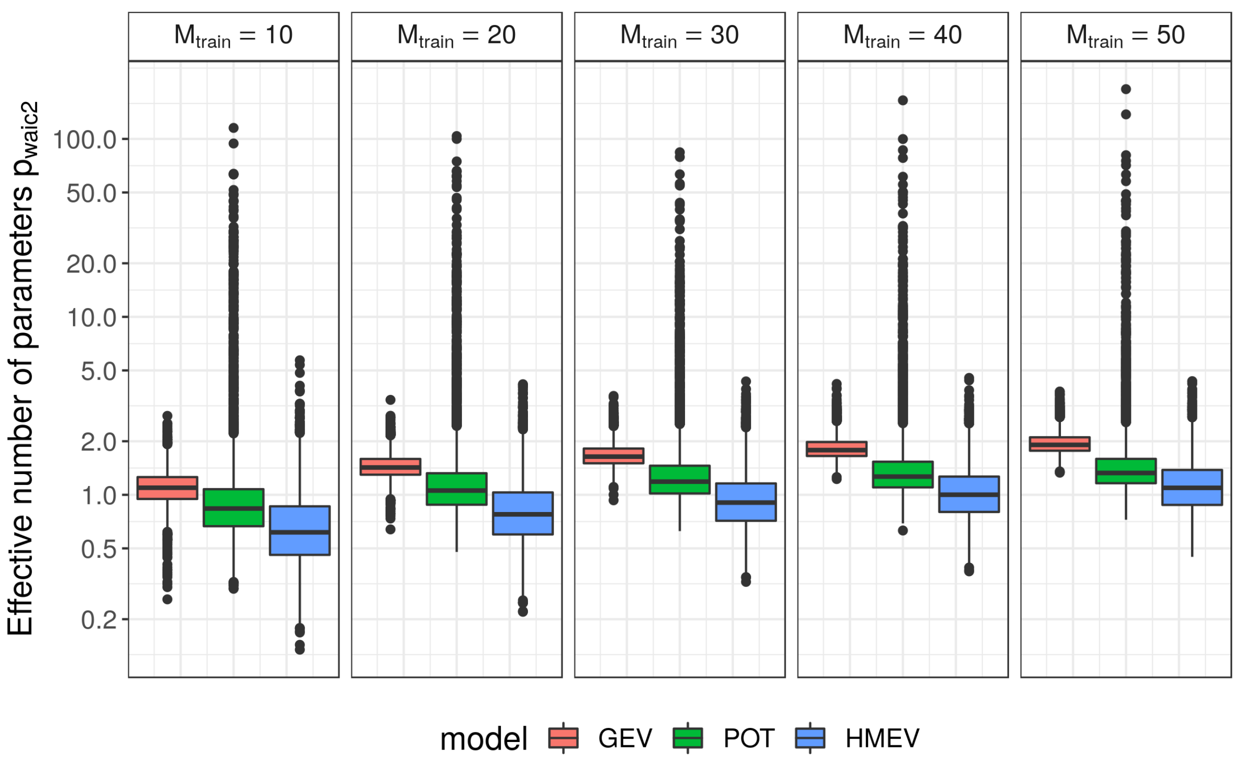

Figure S1: Effective number of parameters for the stations in the USHCN dataset.

Table S2: Summary of 4 model specifications used to generate synthetic datasets in the simulation study.

Model for x_{ij}	parameters
GP	$\xi = 0.1, \quad \sigma = 8$
GAM	$\alpha = 1.2, \quad \beta = 0.12$
WEI	$\gamma = 0.6, \quad \delta = 8$
WEI_G	$\mu_\delta = 7, \quad \sigma_\delta = 1, \quad \mu_\gamma = 1, \quad \sigma_\gamma = 0.1,$
Model for n_j	parameters
BBN	$\mu_n = 100, \quad \sigma_n^2 = 150$

Table S3: Summary of the model for the number of arrivals.

Model	Outer level	Inner Level	Prior
Binomial	$n_j \sim Bin(\pi_0)$		$\pi_0 \sim Beta\,(2, 2)$
Beta Binomial	$n_j \sim Bin(\pi_0)$	$\pi_0 \sim Beta\,(\alpha_n, \beta_n)$	$\mu_n\ \Gamma\,(10, 0.1)$
			$\omega_n\ \Gamma\,(0.2, 0.2)$
			$\alpha_n = \left(1 - \frac{(\omega_n+1)}{(N_t-\mu_n)}\right) / \left(\frac{(\omega_n+1)}{(N_t-\mu_n)} N_t - 1\right)$
			$\beta_n = \frac{\alpha_n}{\mu_n}\,(N_t - \mu_n)$

Spatial distribution of the results for LPPD and FSE

As we can see in Figure S2, here are the results for the three models evaluated using the LPPD

Here we provide a spatially-explicit representation of model performances, by mapping, in Figures S2 and S3, the best model for each station as evaluated through the lppd pr FSE measures respectively. This representation of the results of our analysis shows again the interesting difference observed for the in-sample analysis, which tend to favor the POT method, and the out-of-sample results where HMEV appears to be selected most often as preferred model. The frequency of HMEV being the model of choice is higher for smaller sample sizes.

And in Figure S3 we report the same result for the FSE

in Figures S4, S6, and S5 we report examples using the Weibull, Gamma, and Generalized Pareto specifications respectively.

References

Coles, Stuart. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Vol. 208. Springer.

Coles, Stuart G, and Elwyn A Powell. 1996. "Bayesian Methods in Extreme Value Modelling: A Review and New Developments." *International Statistical Review/Revue Internationale de Statistique*. JSTOR, 119–36.

Coles, Stuart G, and Jonathan A Tawn. 1996. "A Bayesian Analysis of Extreme Rainfall Data." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 45 (4). Wiley Online Library: 463–78.

Coles, Stuart, Luis Raúl Pericchi, and Scott Sisson. 2003. "A Fully Probabilistic Approach to Extreme Rainfall Modeling." *Journal of Hydrology* 273 (1-4). Elsevier: 35–50.

Davison, Anthony C, and Richard L Smith. 1990. "Models for Exceedances over High Thresholds." *Journal of the Royal Statistical Society. Series B (Methodological)*. JSTOR, 393–442.

De Haan, Laurens, and Ana Ferreira. 2007. *Extreme Value Theory: An Introduction*. Springer Science & Business Media.

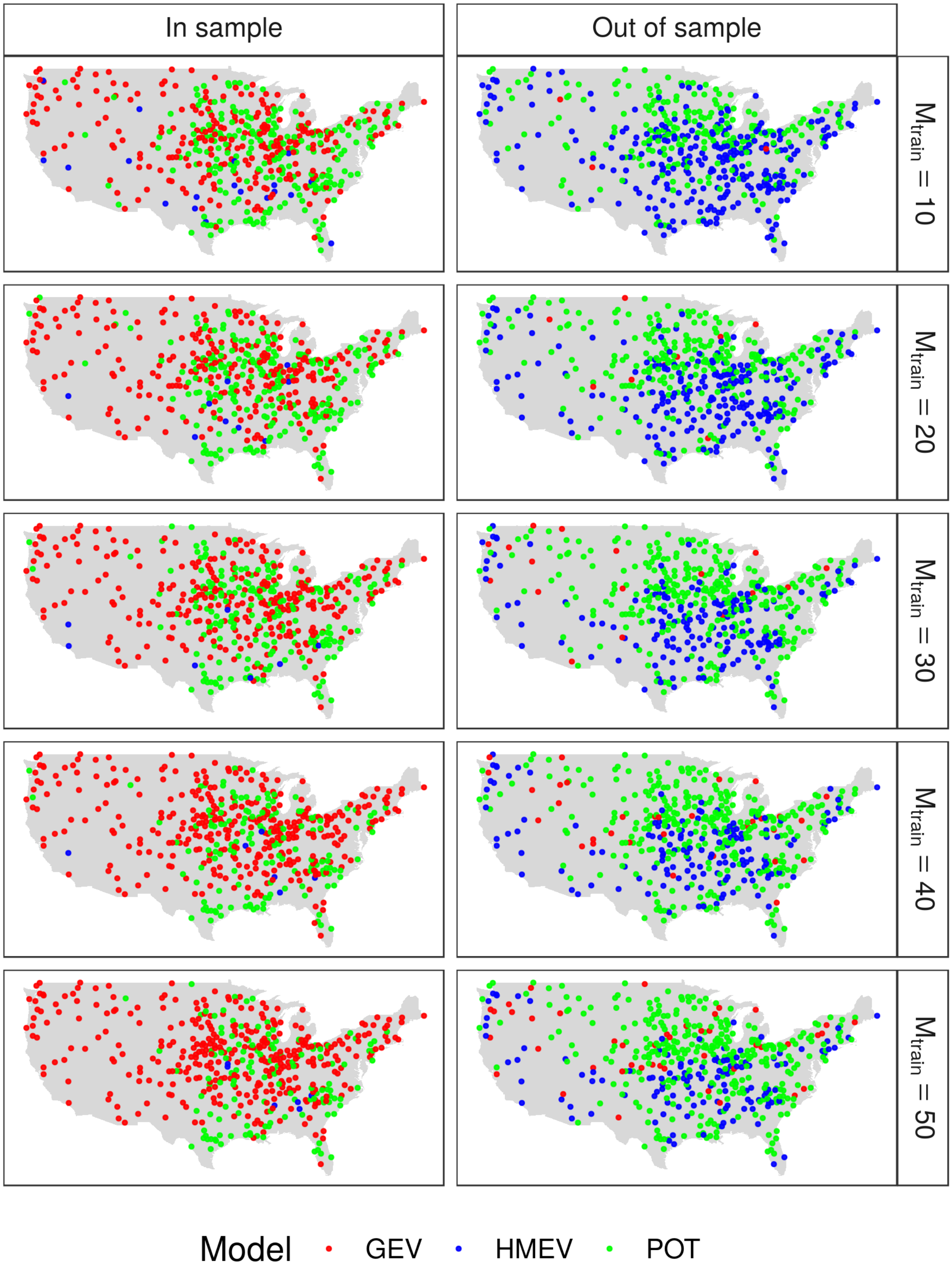

Figure S2: Best model for each station, as evaluated through the LPPD measure (in-sample and out-of-sample).

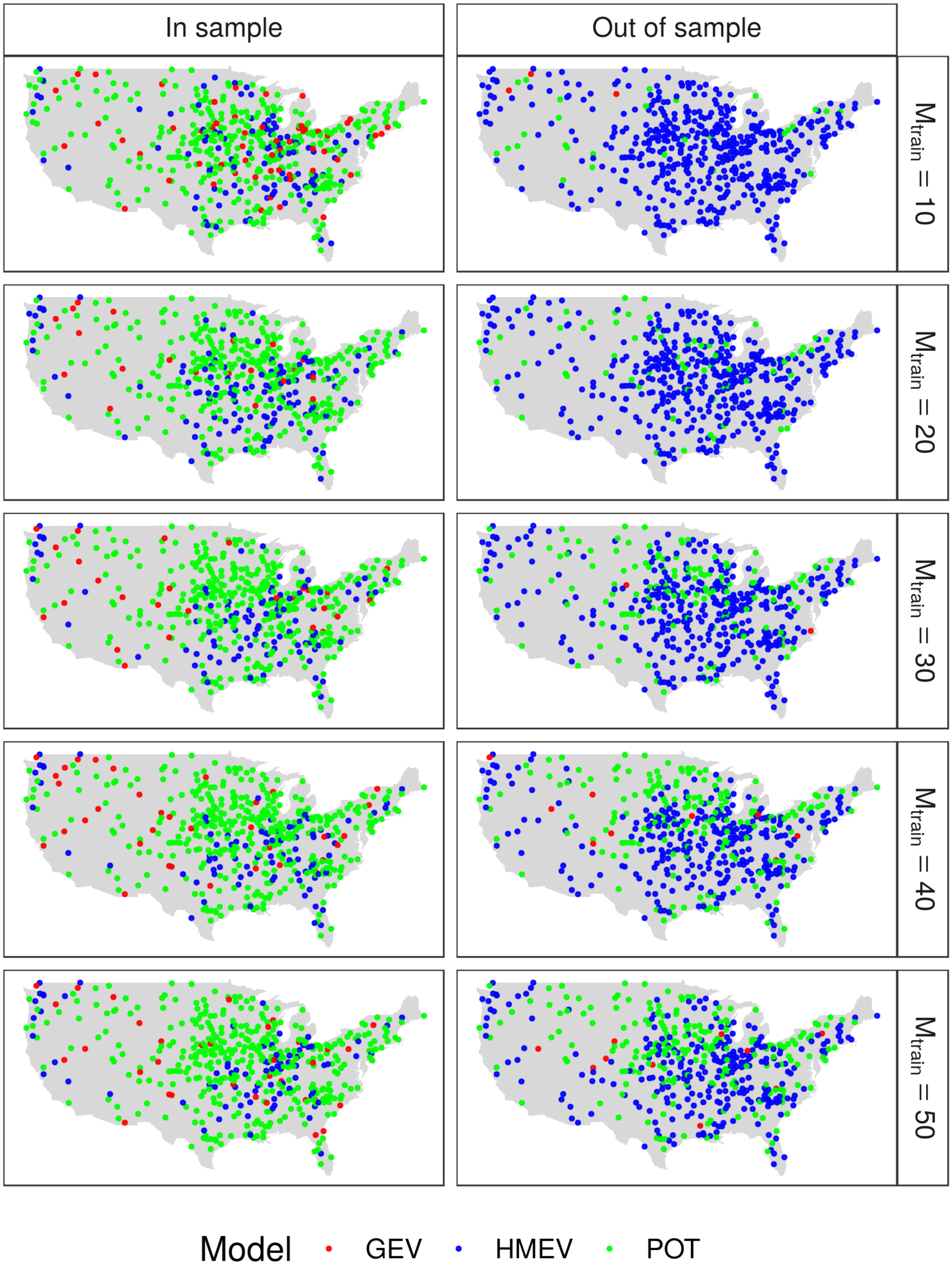

Figure S3: Best model for each station, as evaluated through the FSE measure (in-sample and out-of-sample).

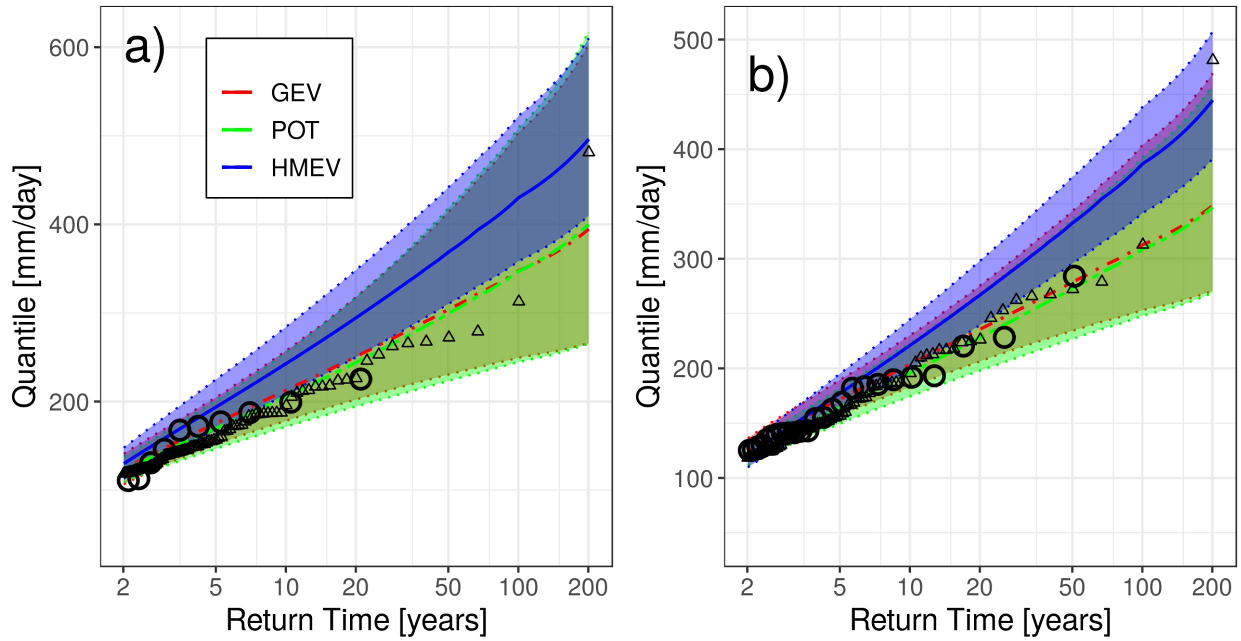

Figure S4: Example of fit to samples of 20 and 50 yearly blocks of data generated according to the Weibull specification.

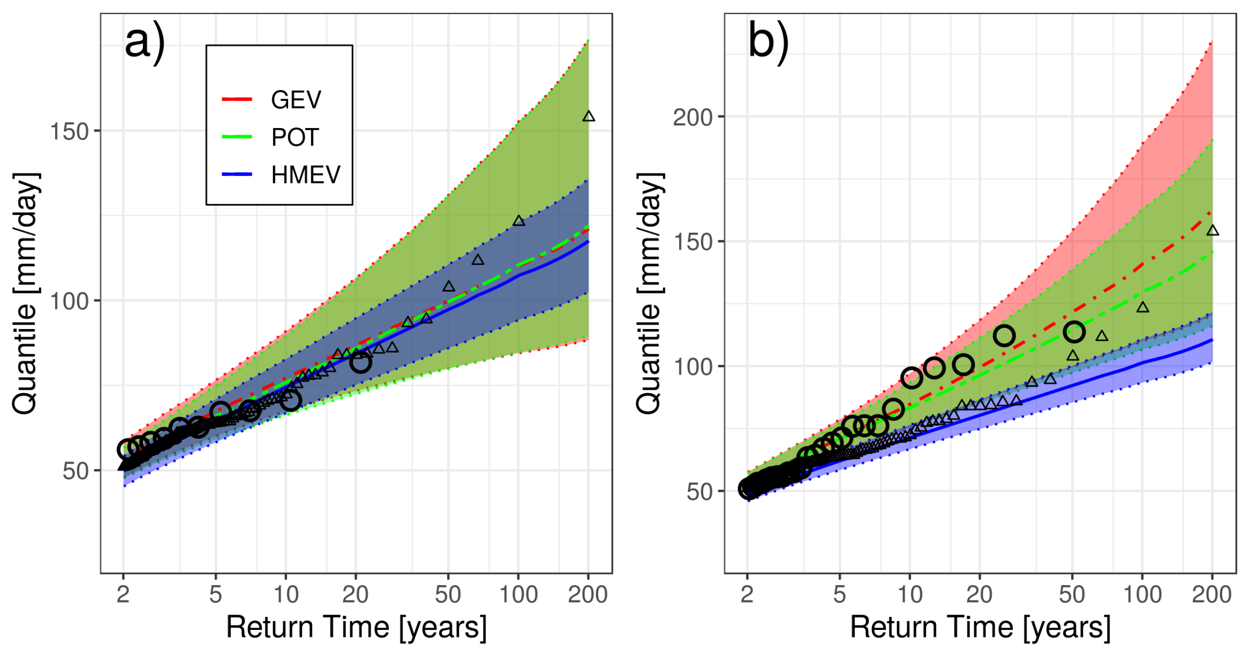

Figure S5: Example of fit to samples of 20 and 50 yearly blocks of data generated according to the Generalized Pareto specification.

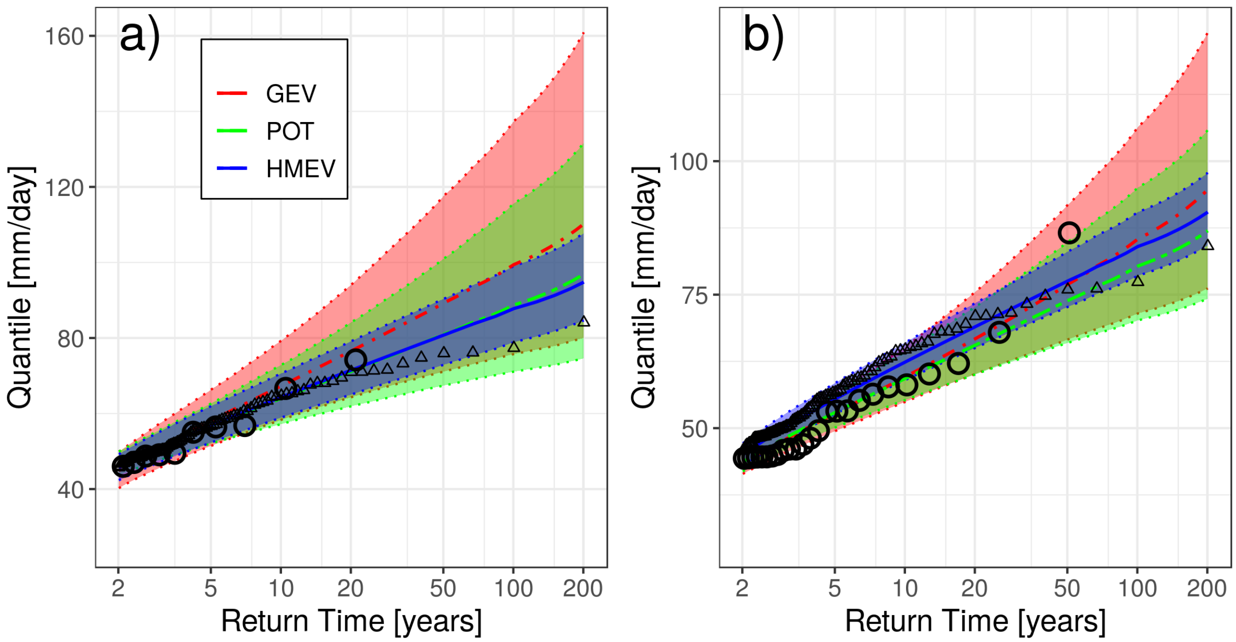

Figure S6: Example of fit to samples of 20 and 50 yearly blocks of data generated according to the Gamma specification.

Gelfand, Alan E, and Dipak K Dey. 1994. "Bayesian Model Choice: Asymptotics and Exact Calculations." *Journal of the Royal Statistical Society. Series B (Methodological)*. JSTOR, 501–14.

Gelfand, Alan E, Dipak K Dey, and Hong Chang. 1992. "Model Determination Using Predictive Distributions with Implementation via Sampling-Based Methods." STANFORD UNIV CA DEPT OF STATISTICS.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. 2013. *Bayesian Data Analysis*. Chapman; Hall/CRC.

Gilleland, Eric, Richard W Katz, and others. 2016. "ExtRemes 2.0: An Extreme Value Analysis Package in R." *Journal of Statistical Software* 72 (8). Foundation for Open Access Statistics: 1–39.

Hosking, Jonathan RM. 1990. "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics." *Journal of the Royal Statistical Society: Series B (Methodological)* 52 (1). Wiley Online Library: 105–24.

Koutsoyiannis, Demetris. 2004. "Statistics of Extremes and Estimation of Extreme Rainfall: I. Theoretical Investigation/Statistiques de Valeurs Extrêmes et Estimation de Précipitations Extrêmes: I. Recherche Théorique." *Hydrological Sciences Journal* 49 (4). Taylor & Francis.

Martins, Eduardo S, and Jery R Stedinger. 2000. "Generalized Maximum-Likelihood Generalized Extreme-Value Quantile Estimators for Hydrologic Data." *Water Resources Research* 36 (3). Wiley Online Library: 737–44.

Stephenson, Alec, and Jonathan Tawn. 2004. "Bayesian Inference for Extremes: Accounting for the Three Extremal Types." *Extremes* 7 (4). Springer: 291–307.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. "Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and Waic." *Statistics and Computing* 27 (5). Springer: 1413–32.

Von Mises, Richard. 1936. "La Distribution de La Plus Grande de N Valeurs." *Rev. Math. Union Interbalcanique* 1 (1).