

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 briefly 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) model with the frequency of excesses over threshold is described through 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 techniques 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

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

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; Coles, Pericchi, and Sisson 2003). Generally L-moments perform better than ML in the case of small samples, even though asymptotic theory for confidence intervals is not available. Bayesian Methods allow better characterization of the variability of estimated 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 matches the *Geophysical prior* proposed by Martins and Stedinger (2000) in order to guide inference towards realistic 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 available knowledge on the expected value and characteristic variability of the 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). If the random variable $X \in R^+$ represents daily rainfall magnitude, and $Y = X - q$ the excess over a threshold q , the cumulative distribution of excesses over threshold reads

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

In this case, the distribution of block maxima reads

$$F_{PP}(x) = \sum_{n=0}^{\infty} p_n(n|\lambda) F_{GPD}(x|\beta, q, \kappa) = 1 - \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \left(1 + \frac{\kappa}{\beta}(x - q)\right)^{-1/\kappa} = F_{GEV}(x|\mu, \sigma, \xi) \quad (\text{S3})$$

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

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 $s = 1, \dots, S$, we can compute the posterior probability for the GEV parameters through the relations

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

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. 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. For a sample of block maxima y_i , $i = 1, \dots, M$, this measure can be directly estimated from S MCMC draws as

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

where we multiply by the factor $-1/M$ for plotting purposes, so as to reduce its variation with sample size M . This quantity, if computed for in-sample annual maxima data y_i , $i = 1, \dots, M$ 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 Criterion (DIC), Watanabe-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.

$$lpml = -\frac{1}{M} \sum_{i=1}^M \log(CPO_i) \quad (\text{S5})$$

where again we multiply by 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:

$$CPO_i = \left\{ \int \frac{1}{p(y_i | \theta)} p(\theta | \mathbf{y}_{-i}) d\theta \right\}^{-1} \quad (\text{S6})$$

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 compute CPO_i as follows:

$$\widehat{CPO}_i = \left[\frac{1}{S} \sum_{s=1}^S \frac{1}{p(y_i | \theta^{(s)})} \right]^{-1} \quad (\text{S7})$$

and thus $\widehat{lpml} = -\frac{1}{M} \sum_{i=1}^M \log(\widehat{CPO}_i)$. Therefore the CPO_i can be computed as the harmonic mean of the likelihood of the annual maxima y_i given the model.

Stations: Effective number of parameters

As we can see in Figure S1 we report the effective number of parameters computed for observed data, computed as the difference between $lppd$ and $lpml$. Additionally, we include the same quantity computed in a different way (Watanabe-Aikake information criterion, WAIC, in Figure S2) to show that results are not overly sensitive to the specific choice of information criterion adopted.

Here we also show the effective number of parameters computed in the simulation study for a sample size of 10 and 20 years, in Figures ?? and ?? respectively. The same result for a sample size of $M_{train} = 50$ years is featured in the main text.

To complete the results in the manuscript, here we include results of the simulation study obtained for a sample size of $M_{train} = 20$ years. These include fractional square error ??, bias S6, and average width of credibility intervals S7.

Spatial distribution of the results for LPPD and FSE

Here we provide a spatially-explicit representation of model performances, by mapping, in Figures S8 and S9, the best model for each station as evaluated through the $lppd$ or FSE measures respectively. Each metric is

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

Parameter	Value		
$\alpha_{\mu\gamma 0} =$	$i_{\mu\gamma 0}$		
$\beta_{\mu\gamma 0} =$	$i_{\mu\gamma 0} \cdot e_{\mu\gamma 0}$		
$\alpha_{\mu\delta 0} =$	$i_{\mu\delta 0}$		
$\beta_{\mu\delta 0} =$	$i_{\mu\delta 0} \cdot e_{\mu\delta 0}$		
$\alpha_{\sigma\gamma 0} =$	$i_{\sigma\gamma 0}$		
$\beta_{\sigma\gamma 0} =$	$i_{\sigma\gamma 0} \cdot e_{\sigma\gamma 0} \cdot v_{\sigma\gamma 0}$		
$\alpha_{\sigma\delta 0} =$	$i_{\sigma\delta 0}$		
$\beta_{\sigma\delta 0} =$	$i_{\sigma\delta 0} \cdot e_{\sigma\delta 0} \cdot v_{\sigma\delta 0}$		
Constant	Value	Meaning	Empirical Prior
$i_{\mu\gamma 0}$	10	shape informativeness	no
$i_{\mu\delta 0}$	10	scale informativeness	no
$i_{\sigma\gamma 0}$	10	shape informativeness	no
$i_{\sigma\delta 0}$	10	scale informativeness	no
$e_{\mu\gamma 0}$	0.7	expected value shape	no
$e_{\mu\delta 0}$	$10/\Gamma(1 + 1/0.7)$	expected value scale	no (possible)
$v_{\sigma\gamma 0}$	0.05	expected variability shape	no
$v_{\sigma\delta 0}$	0.25	expected variability scale	no (possible)

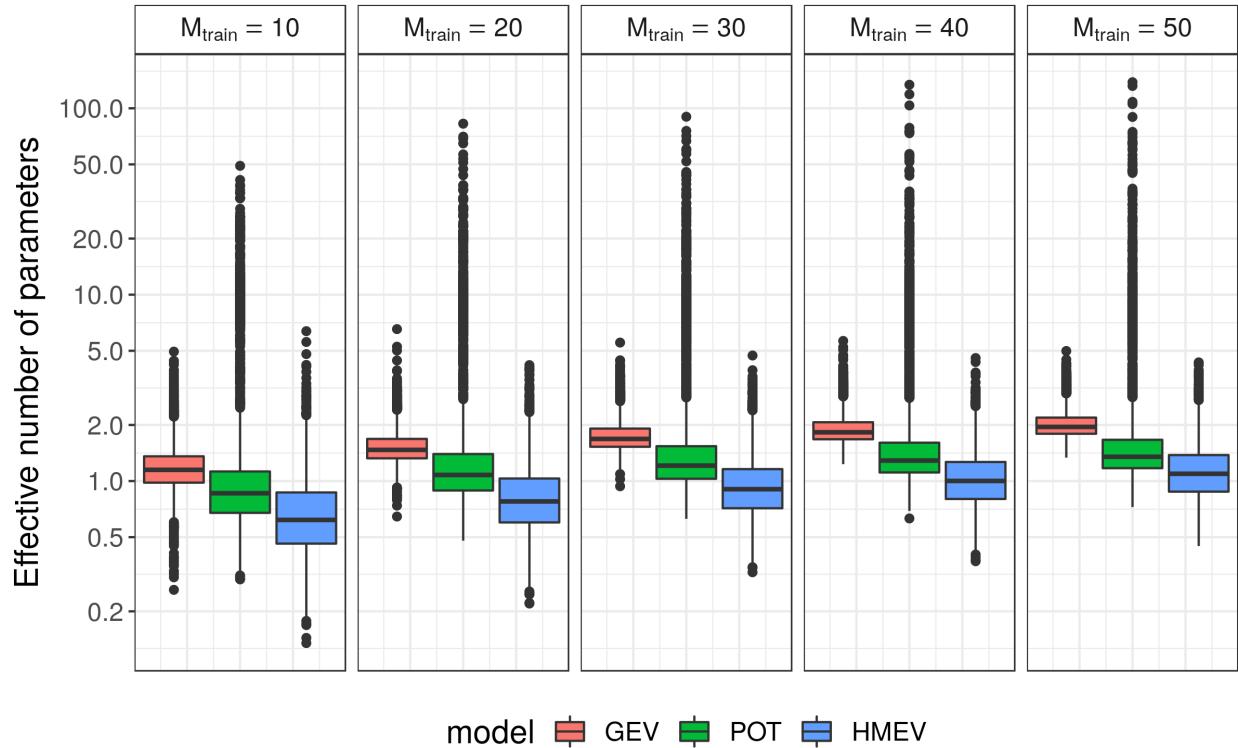


Figure S1: Effective number of parameters for the stations in the USHCN dataset.

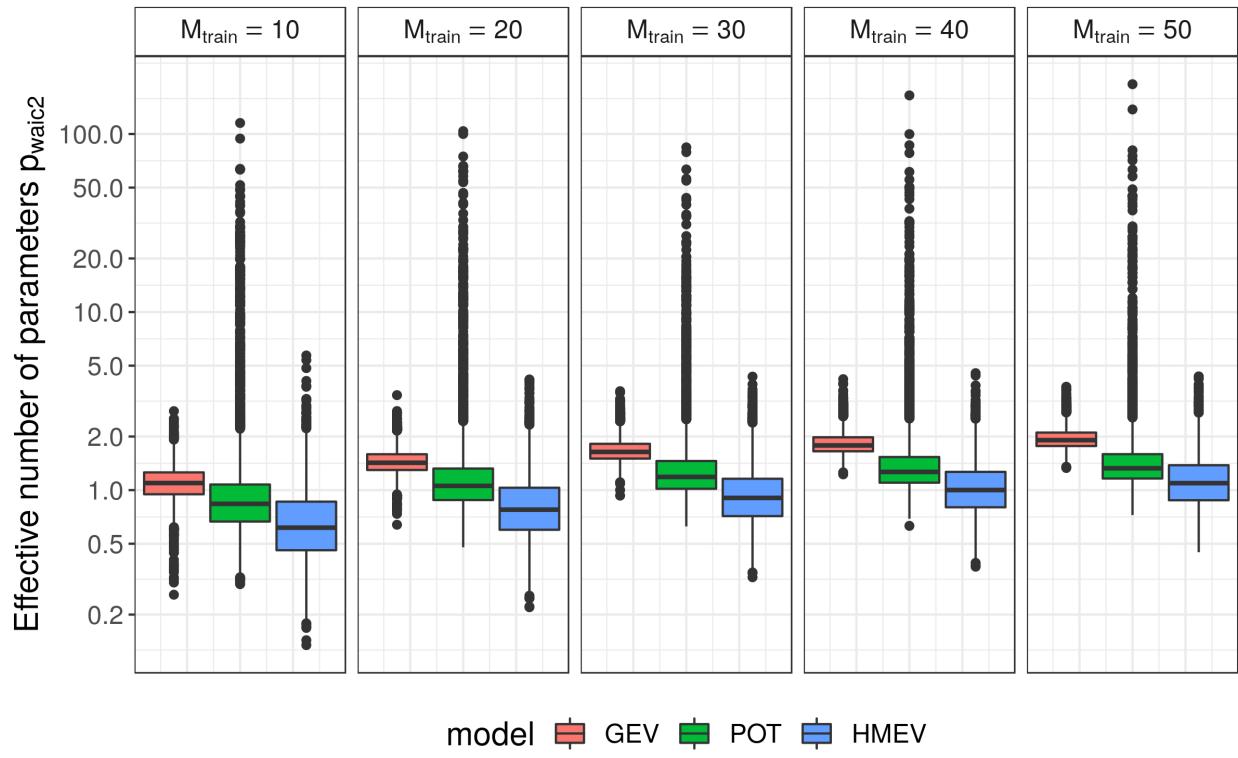


Figure S2: Effective number of parameters for the stations in the USHCN dataset, computed using the Watanabe-Aikake information criterion (WAIC).

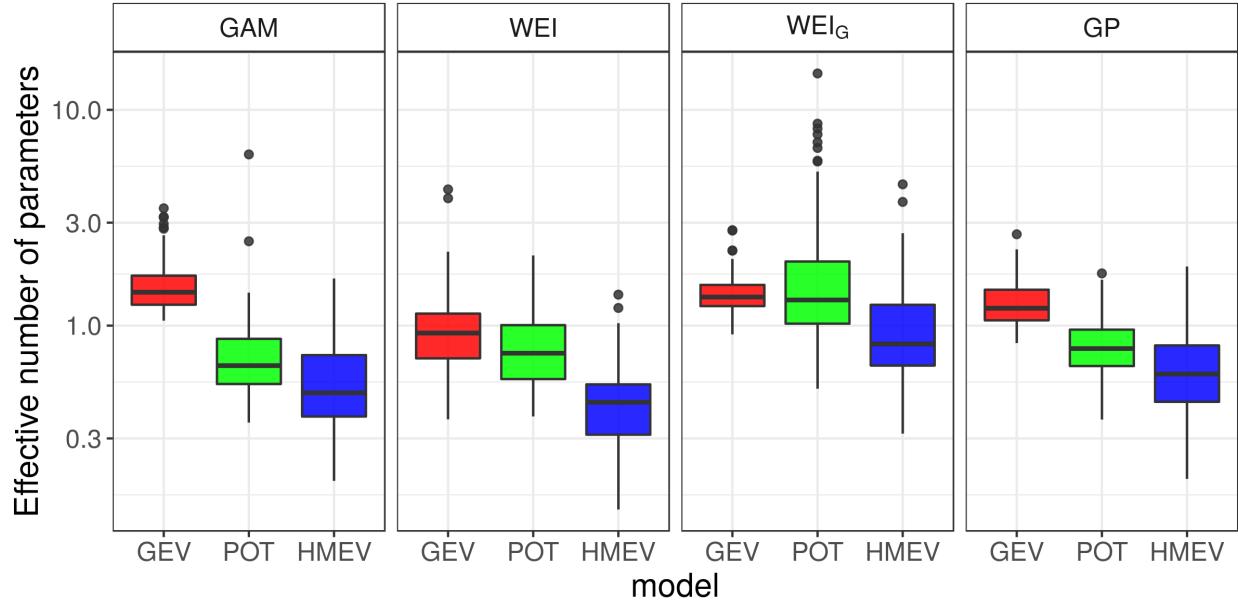


Figure S3: Effective number of parameters computed in the simulation study for a sample size of 10 years.

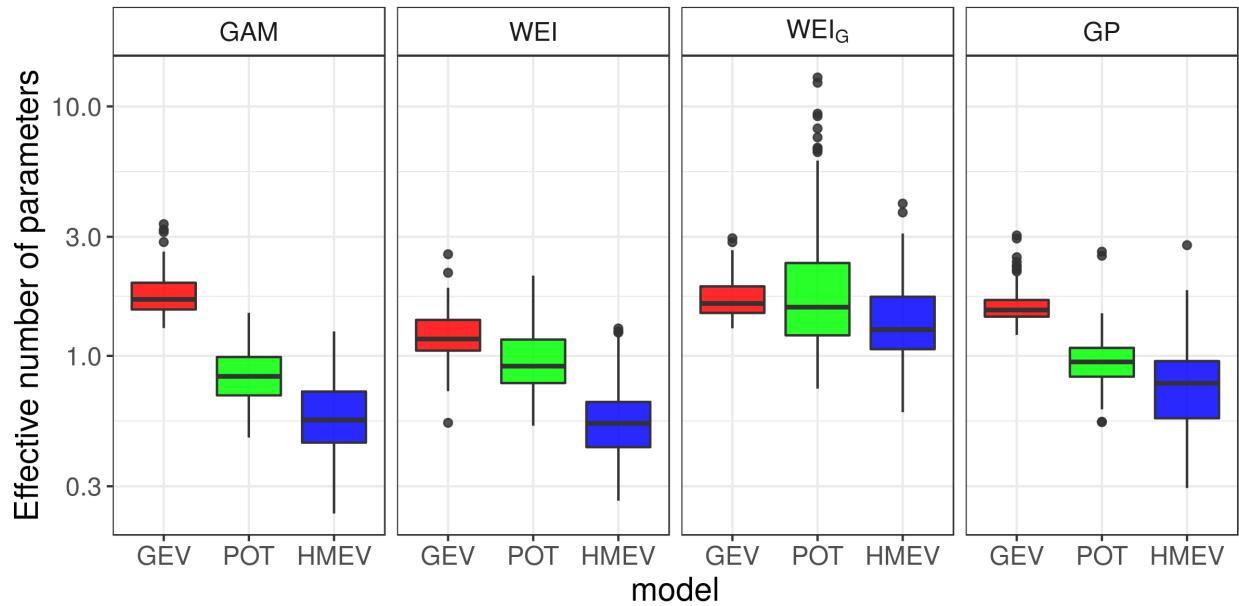


Figure S4: Effective number of parameters computed in the simulation study for a sample size of 20 years.

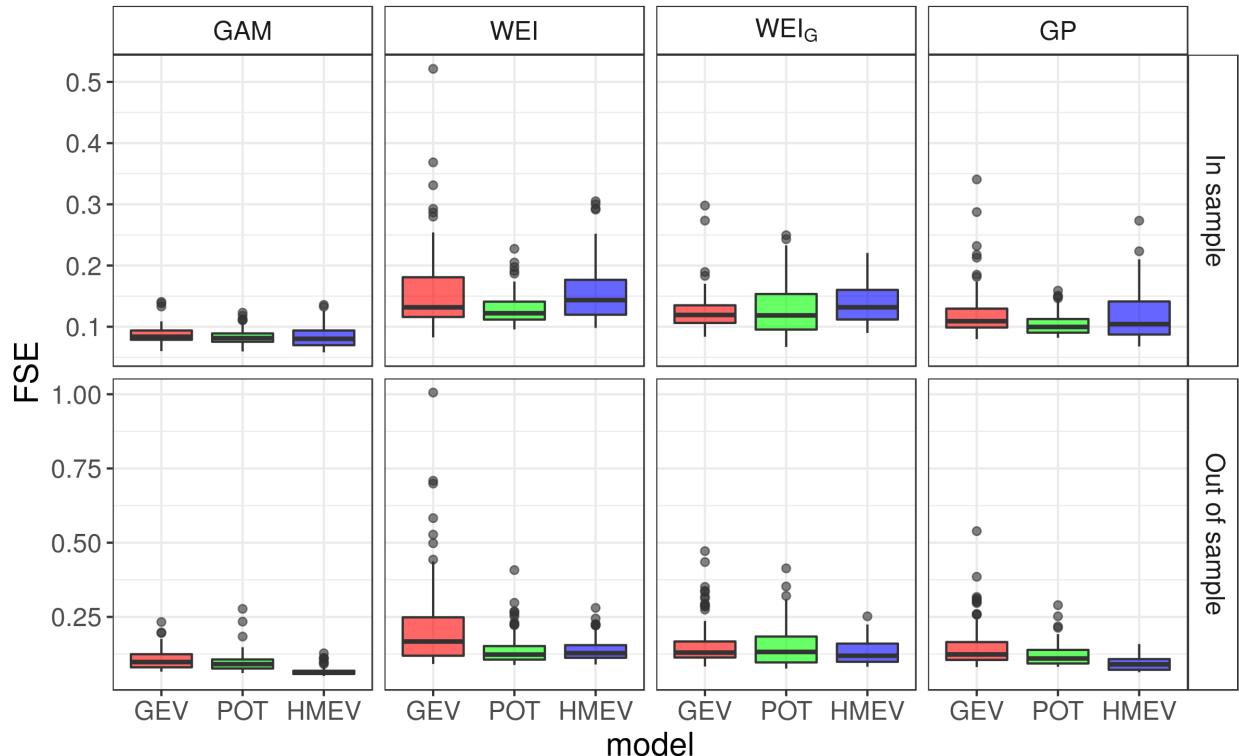


Figure S5: Fractional square error computed for the 4 different model specifications for in-sample data (upper panels) and for out-of-sample data (lower panels), computed for sample size of 20 years.

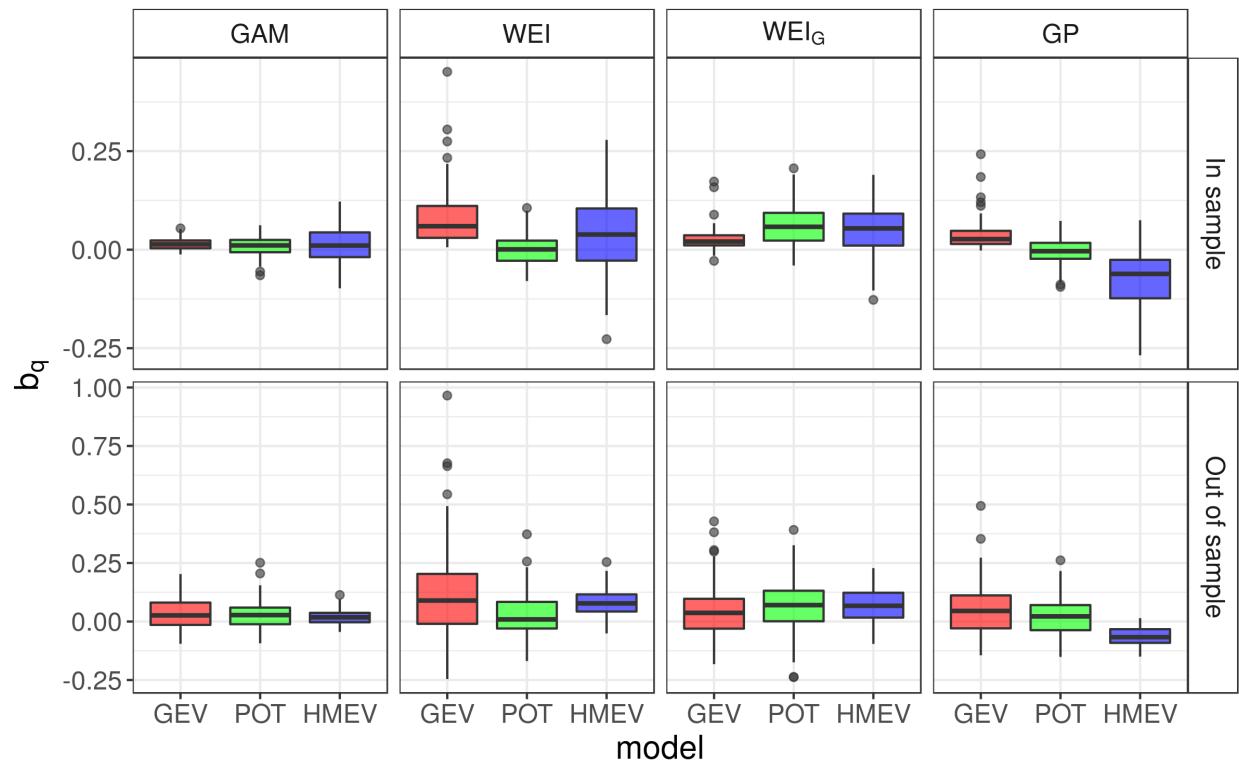


Figure S6: Mean bias computed for the 4 different model specifications for in-sample data (upper panels) and for out-of-sample data (lower panels), computed for sample size of 20 years.

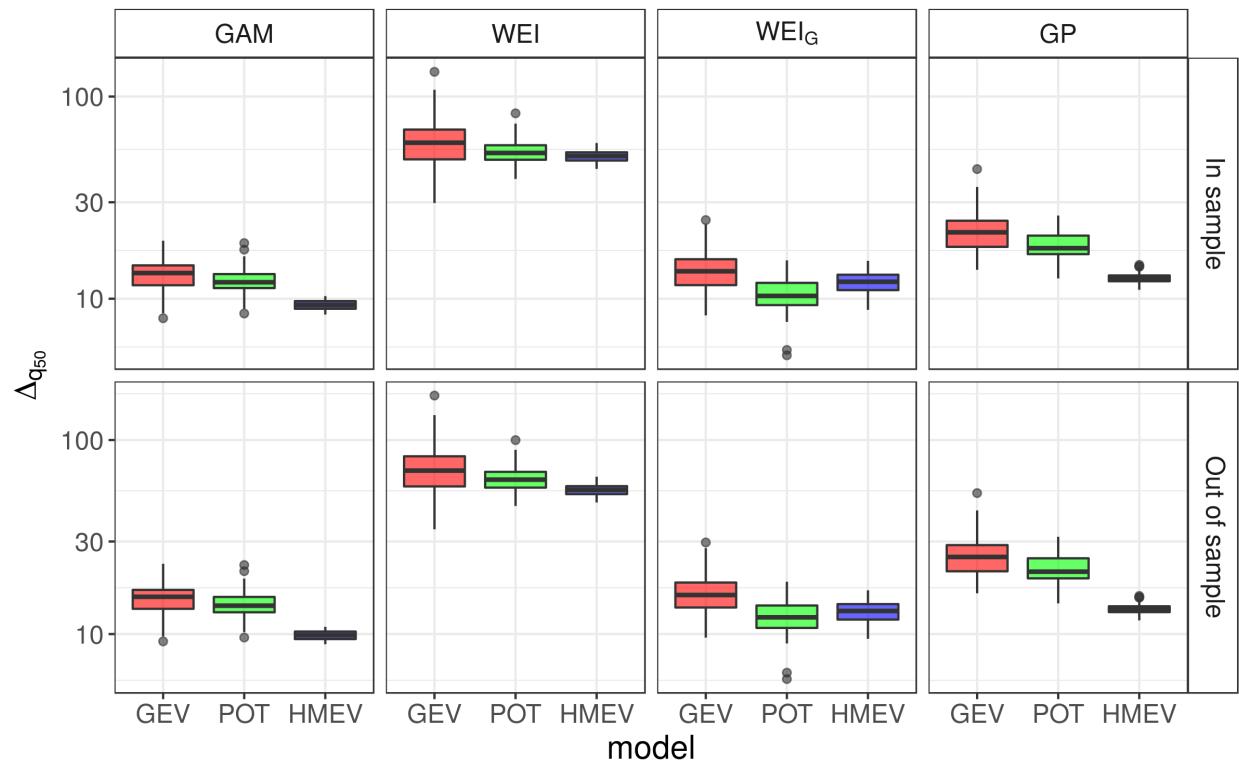


Figure S7: Mean credibility interval width for the 4 different model specifications for in-sample data (upper panels) and for out-of-sample data (lower panels), computed for sample size of 20 years.

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

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

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	$n_j \sim Bin(\pi_0)$	$\pi_0 \sim Beta(\alpha_n, \beta_n)$	$\mu_n \sim \Gamma(10, 0.1)$ $\omega_n \sim \Gamma(0.2, 0.2)$ $\alpha_n = \left(1 - \frac{(\omega_n+1)}{(N_t - \mu_n)}\right) / \left(\frac{(\omega_n+1)}{(N_t - \mu_n)} N_t - 1\right)$ $\beta_n = \frac{\alpha_n}{\mu_n} (N_t - \mu_n)$

evaluated both for in sample and out of sample data, and the result is averaged over the R_g resampling of the original data, according to the procedure described in the main manuscript. 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.

Examples of fit to simulated samples

in Figures S10, S12, and S11 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.
- 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.

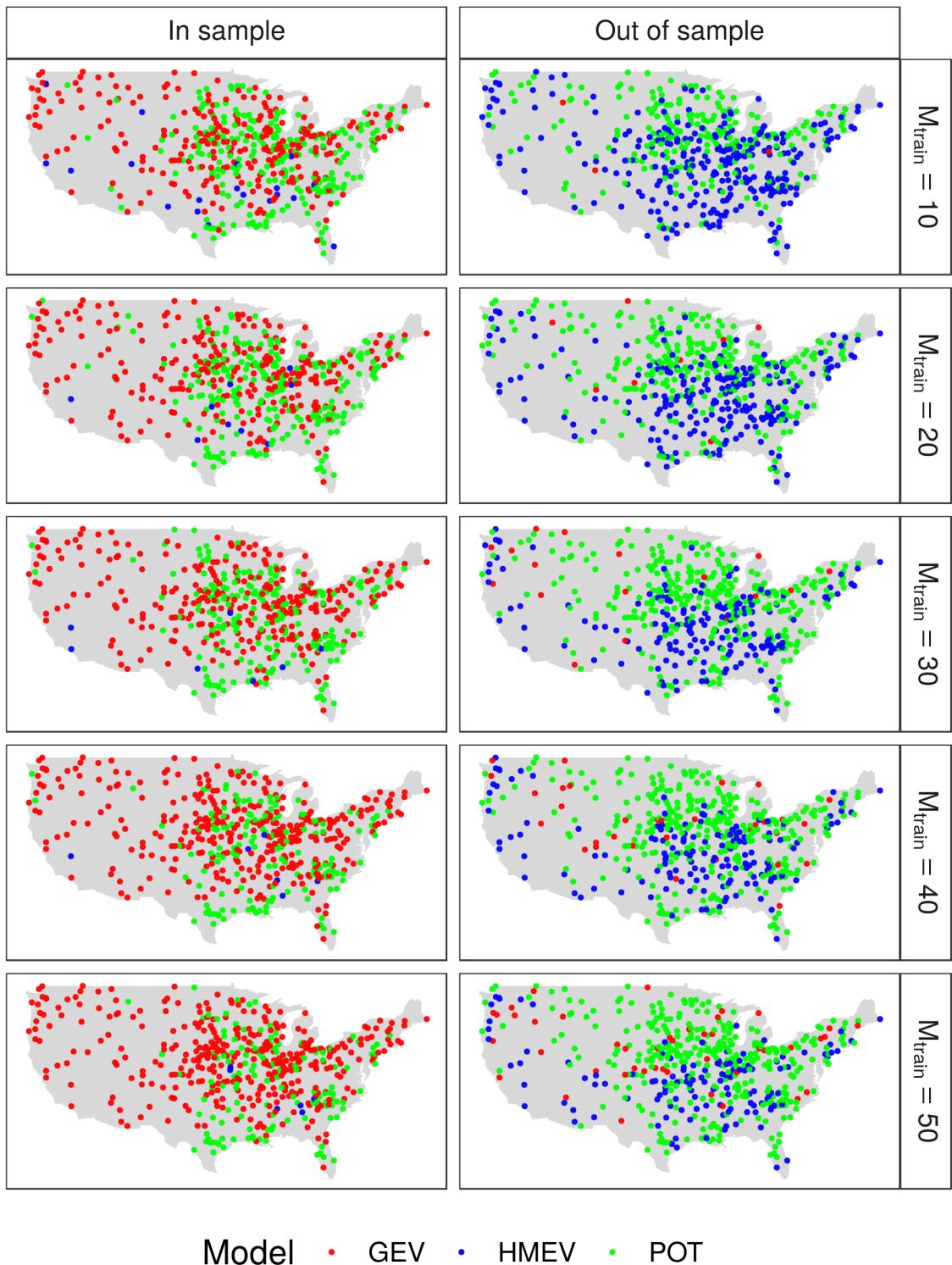


Figure S8: Best model for each station, as evaluated through the lppd measure for in-sample and out-of-sample data.

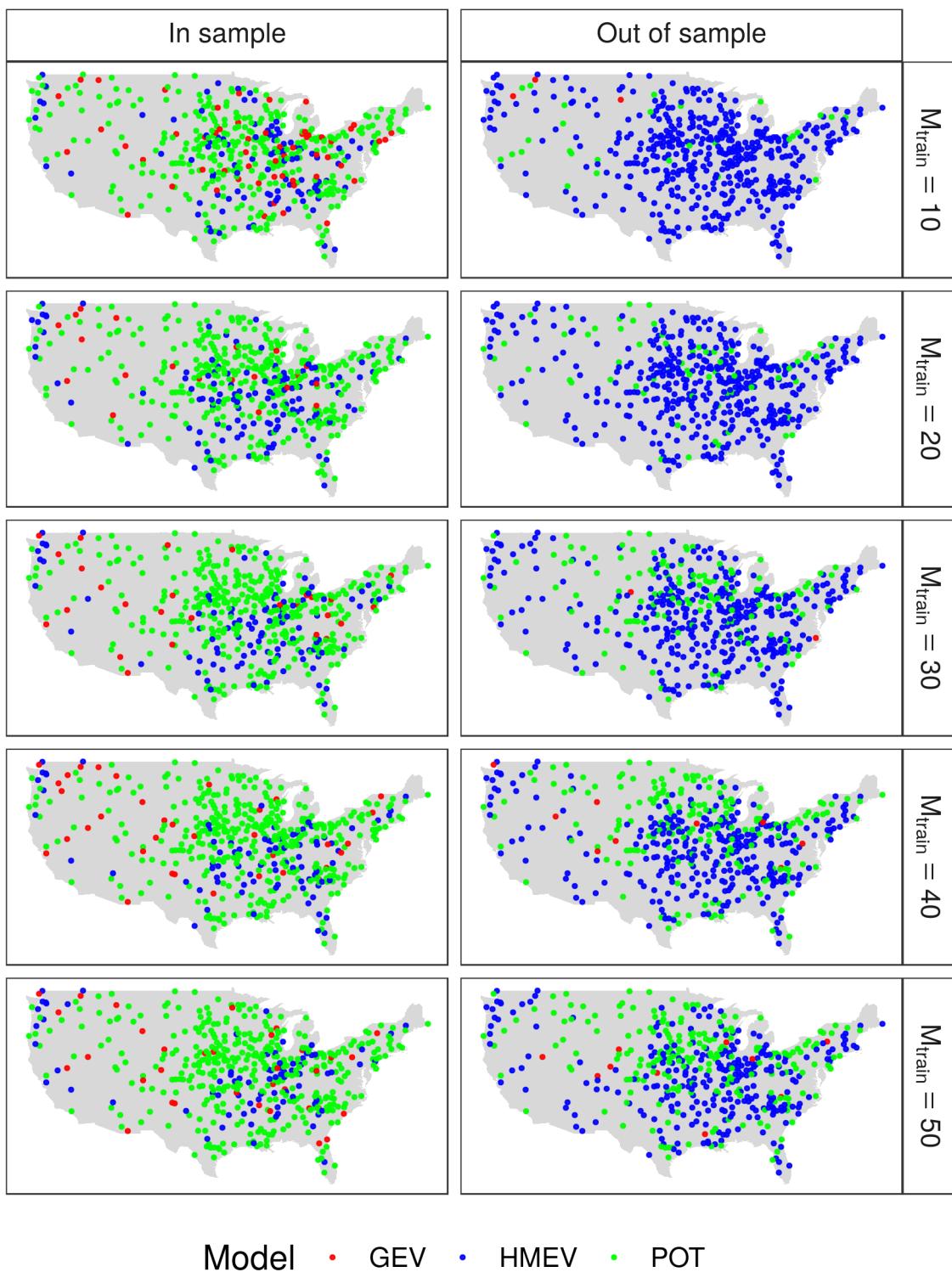


Figure S9: Best model for each station, as evaluated through the fractional square error (FSE), evaluated for in-sample and out-of-sample data.

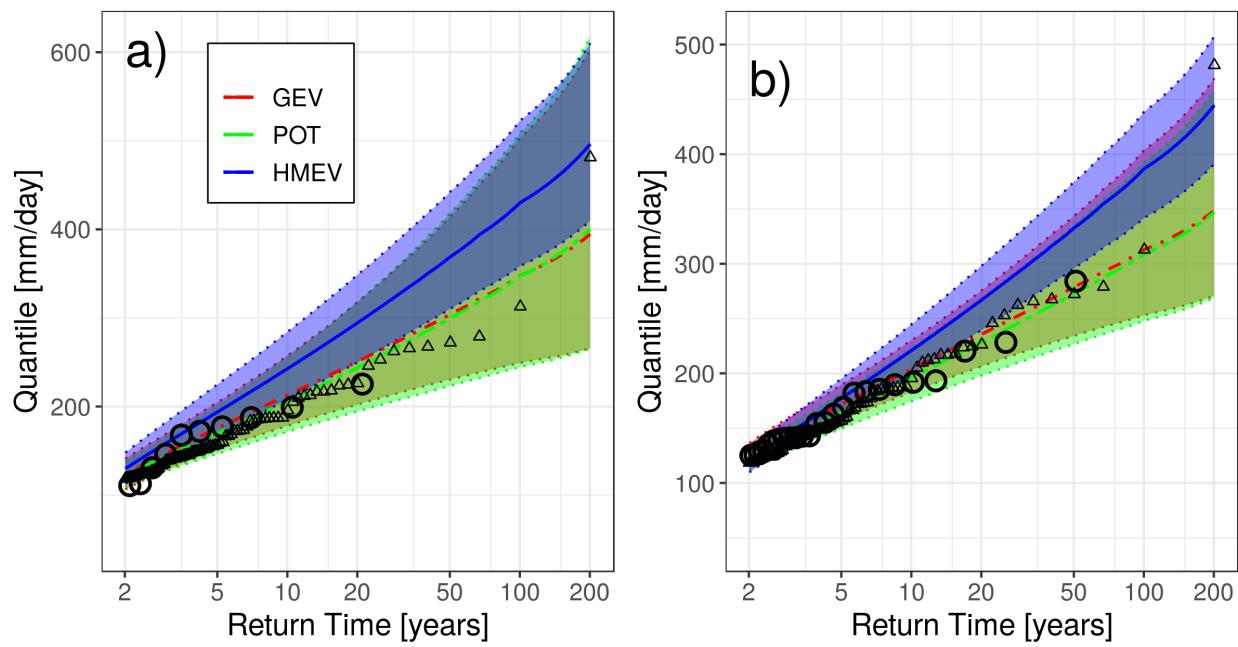


Figure S10: Example of fit to samples of 20 (left panel) and 50 (right panel) yearly blocks of data generated according to the Weibull specification.

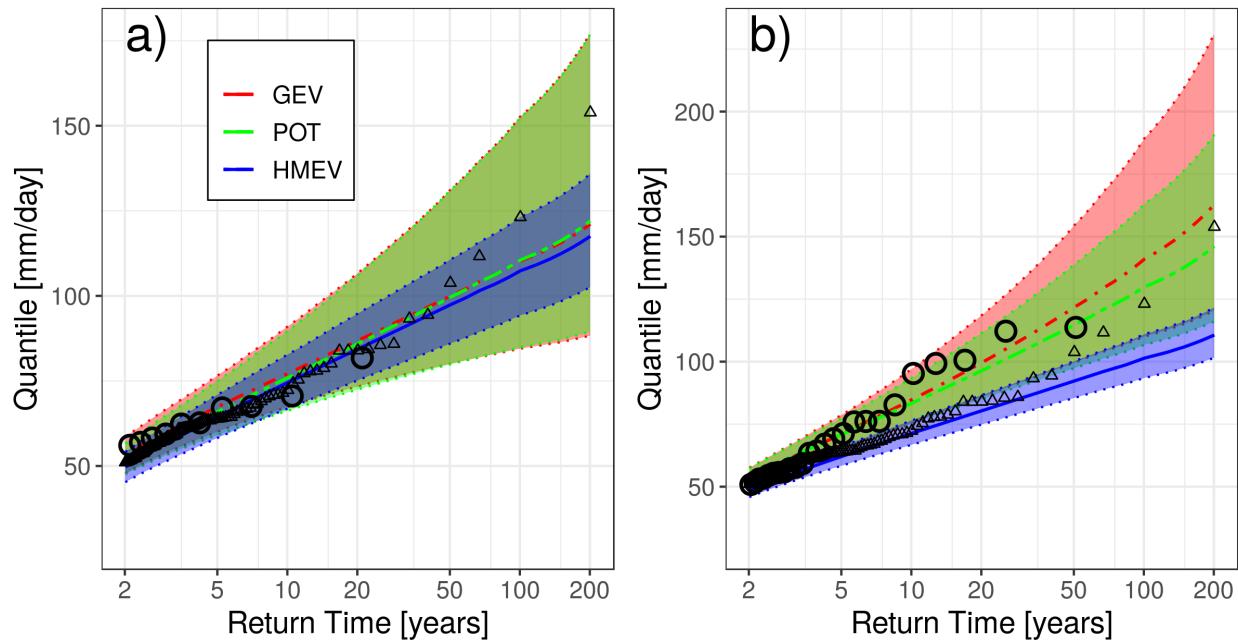


Figure S11: Example of fit to samples of 20 (left panel) and 50 (right panel) yearly blocks of data generated according to the Generalized Pareto specification.

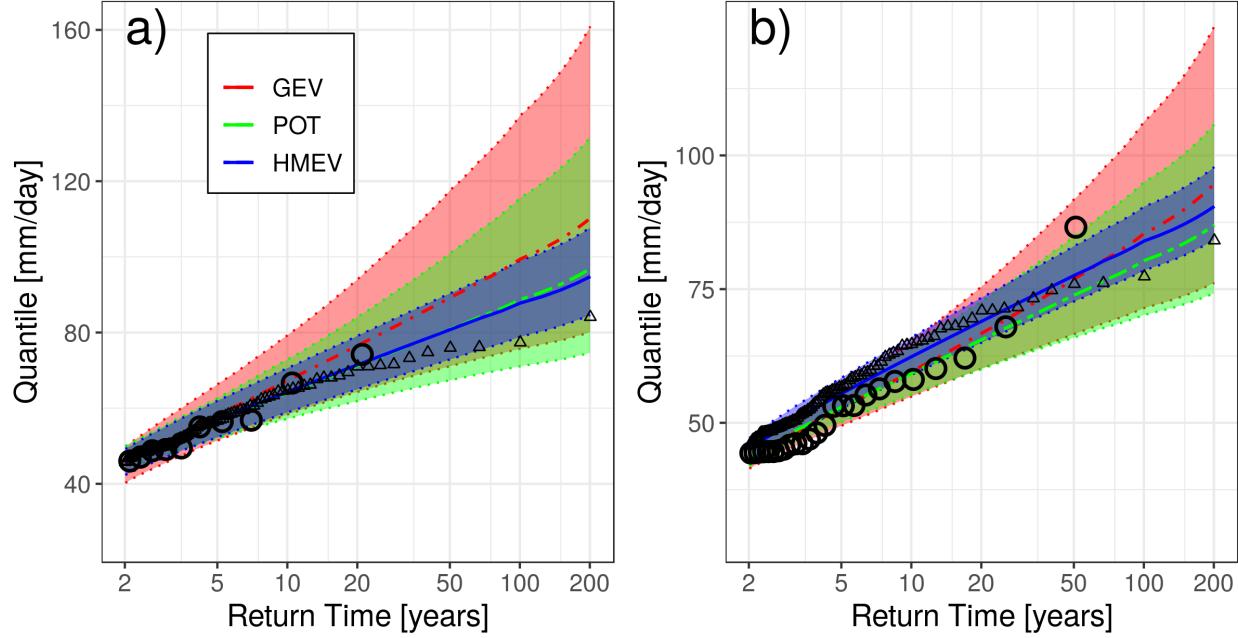


Figure S12: Example of fit to samples of 20 (left panel) and 50 (right panel) yearly blocks of data generated according to the Gamma specification.

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).