Hierarchical Bayesian modelling of hydrological extremes

Enrico Zorzetto[1] Antonio Canale[2] Marco Marani[3]

5/6/2020

[1]Division of Earth and Ocean Sciences, Duke University, Durham, NC
enrico.zorzetto@duke.edu
[2]Department of Statistical Sciences, University of Padova, Padova, Italy
[3]Department of Civil, Environmental and Architectural Engineering, University of Padova, Padova, Italy

Scope of the work

- We present an extension of the Metastatistical Extreme Value Distribution (Marani and Ignaccolo 2015; Zorzetto, Botter, and Marani 2016) aimed at quantifying estimation uncertainty in extreme value modelling of hydrological time series.
- The interannual variability of hydrological variables is described through a '*slow*' latent variable level varying with yearly time scale.
- Elicitation of informative prior distributions allows for the inclusion of physical information aimed at reducing estimation uncertainty.
- Here we present an application to daily rainfall extremes over the continental United States, and report a benchmark comparison to other extreme value models.

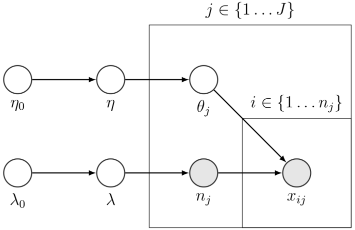
Model Structure

Observed quantities:

- $x_{ij} = i - th$ event magnitude in block j
- $n_j =$ number of events in block j

We specify parametric models for event magnitudes θ_j and occurrence λ. Magnitudes parameters θ_j are in turn generated by a latent yearly parametric model η

Schematic structure of the hierarchical model. Observed quantities are grey-shaded, parameters in white:



We model the non exceedance probability of annual maxima y as

$$P(Y \leq y \mid \lambda, \eta) = \sum_{n=0}^{N_t} \int_\Theta F(y;\theta)^n g(\theta;\eta) p(n;\lambda) d\theta.$$

Prior elicitation

We need to specify prior distributions for the model of event magnitudes and frequency of occurrence:

- Weakly informative prior for the occurrences' model (n_j)
- Informative priors for the magnitudes (x_{ij}) model. These can be based on:
  - Shape parameter priors based on the analysis by Wilson and Toumi (2005), e.g., prior belief of sub-exponential tails.
  - Scale parameter priors can be based on information on climatic variability, or empirical, based on observations of the local rainfall characteristic intensity.

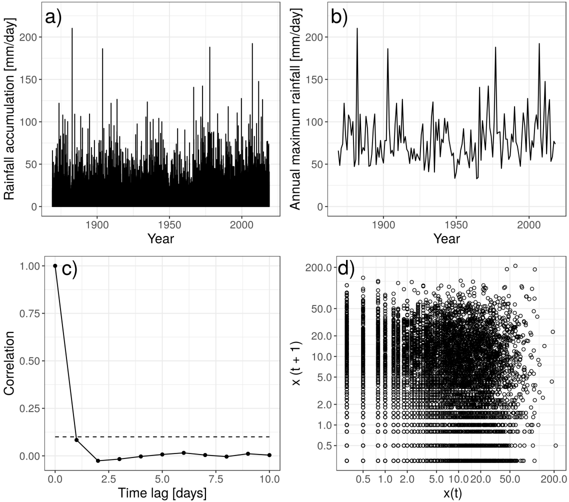
Posterior computation

- Analytical expression for the posterior distribution is not available, so we use an approximation obtained by a Markov Chain Monte Carlo (MCMC) approach. Based on the posterior distribution of the model's parameters, functionals of interest such as the cumulative probability $\zeta(y)$ of a value y can be approximated by the expectation over B MCMC posterior draws:

$$\hat{\zeta}_{MC}(y) = \frac{1}{B} \sum_{b=1}^{B} \hat{\zeta}^{(b)}(y). \tag{1}$$

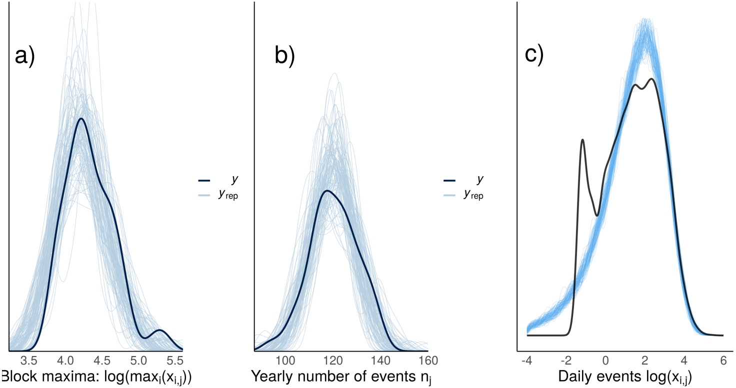
- In the case of time series with significant correlation, we decluster the time series following the procedure proposed by (Marra et al. 2018).

Example of application to the New York Central Park time series



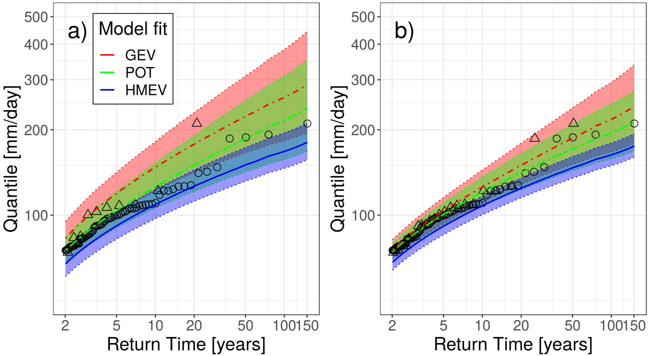
Data source: NOAA USHCN Station ID USW00094728 (Menne et al. 2012)

Posterior predictive checks



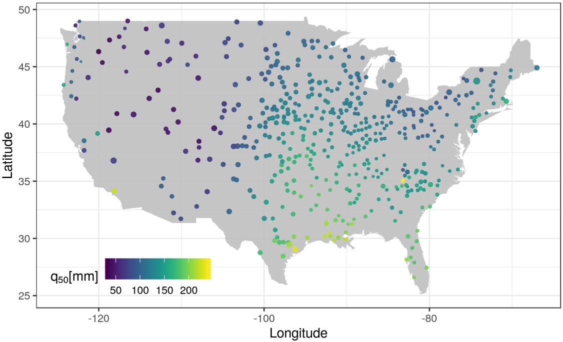
Posterior predictive checks for annual maxima (a), yearly number of events (b), and all daily values (c). Observed densities are in black, and MCMC replicates in blue.

Estimation of quantiles from small samples



Example of quantile estimates from the model described here (HMEV) and comparison with other extreme value models for a sample time series (POT, GEV). Estimates are compared to in-sample (triangles) and out-of-sample (circles) data points

Mapping return levels and relative uncertainty



50-years return level and relative uncertainty. The color indicates the magnitude of the expected 50-years daily rainfall, while the marker size is proportional to the relative width of the posterior 90% probability intervals.

Measuring predictive accuracy

To measure the predictive accuracy of the model (with posterior predictive distribution approximated by B posterior draws) for n annual maxima values $y_i, \ldots y_n$ we use:

1. The *log pointwise predictive density*, or *lppd*:

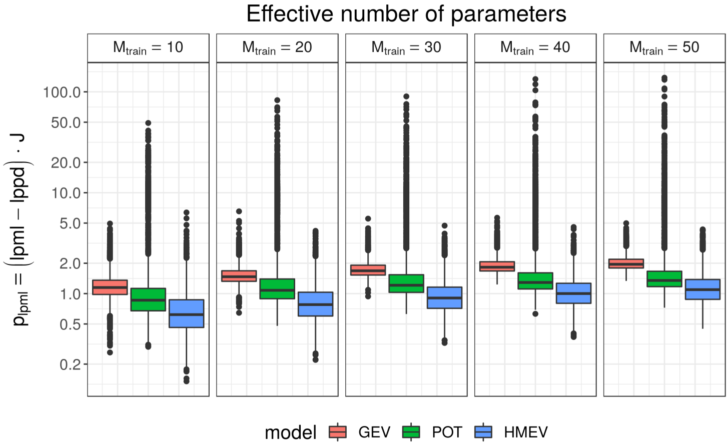
$$lppd = -\frac{1}{n}\sum_{i=1}^{n} log\left(\frac{1}{B}\sum_{b=1}^{B} p(y_i|\theta^{(b)})\right) \tag{2}$$

2. The *log posterior marginal likelihood*, or *lpml* (Gelfand and Dey 1994):

$$lpml = -\frac{1}{n}\sum_{i=1}^{n} \log\left(\left[\frac{1}{B}\sum_{b=1}^{B}\frac{1}{p\left(y_i \mid \theta^{(b)}\right)}\right]^{-1}\right) \tag{3}$$

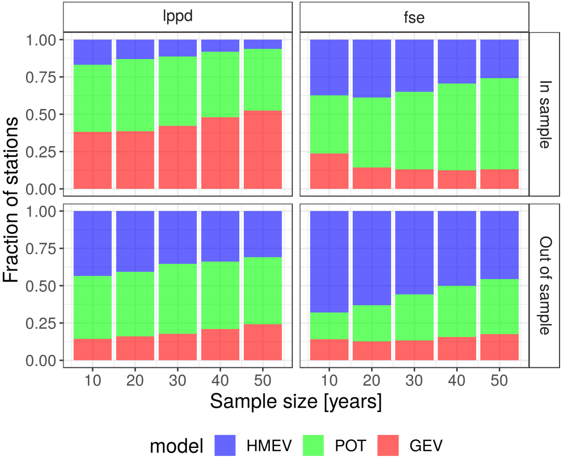
Here we use the difference between *lpml* and *lppd* to quantify a model's tendency to overfit a sample of annual maxima. Both measures are evaluated numerically from $B = 4000$ MCMC posterior draws.

Effective number of parameters



Effective number of parameters of different models, based on the difference between *lpml* and *lppd*. Including more data in the analysis tends to reduce overfit to the annual maxima.

Testing model performance



Benchmark with other extreme value models, including log posterior predictive density (lppd) and a fractional square error (fse) computed for return times $T_r > 2$ years.

Contributions

- We introduce a hierarchical model for describing extremes of environmental time series.
- Building on the Metastatistical Extreme Value Distribution, this approach now allows to quantify estimation uncertainty through a Bayesian approach.
- This method can reduce estimation uncertainty for small sample sizes, as evaluated using out-of-sample data for independent validation.
- Our analysis underlines the importance of out-of-sample validation for extreme value models. When long time series are not available for independent cross validation, in-sample techniques such as *lpml* can help inform model selection.

References I

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.

Marani, Marco, and Massimiliano Ignaccolo. 2015. "A Metastatistical Approach to Rainfall Extremes." *Advances in Water Resources* 79. Elsevier: 121–26.

Marra, Francesco, Efthymios I Nikolopoulos, Emmanouil N Anagnostou, and Efrat Morin. 2018. "Metastatistical Extreme Value Analysis of Hourly Rainfall from Short Records: Estimation of High Quantiles and Impact of Measurement Errors." *Advances in Water Resources* 117. Elsevier: 27–39.

References II

Menne, Matthew J, Imke Durre, Bryant Korzeniewski, Shelley McNeal, Kristy Thomas, Xungang Yin, Steven Anthony, et al. 2012. "Global Historical Climatology Network-Daily (Ghcn-Daily), Version 3." *NOAA National Climatic Data Center* 10. http://doi.org/10.7289/V5D21VHZ.

Wilson, PS, and R Toumi. 2005. "A Fundamental Probability Distribution for Heavy Rainfall." *Geophysical Research Letters* 32 (14). Wiley Online Library.

Zorzetto, Enrico, Gianluca Botter, and Marco Marani. 2016. "On the Emergence of Rainfall Extremes from Ordinary Events." *Geophysical Research Letters* 43 (15). Wiley Online Library: 8076–82.