

Modeling extremal streamflow using deep learning approximations and a flexible spatial process

Brian Reich, NCSU

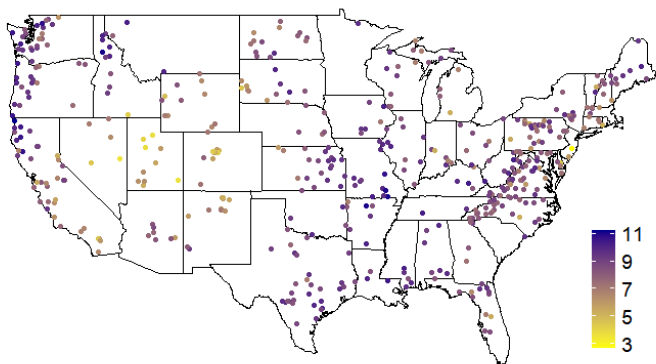
Joint work with Reetam Majumder (NCSU) and Benjamin A. Shaby (CSU)

July, 2023

NC STATE UNIVERSITY

Motivation

- Extremal streamflow is a key measure of flood risk
- Quantifying how the probability and magnitude of extreme flooding events are changing is key to mitigating their impacts under changing climate
- We use data from the USGS Hydro Climatic Data Network (HCDN)
- These stations are selected because they are relatively unaffected by human activity
- We use data from 1972–2021 at 487 stations with complete data
- The extreme value at a spatial location for given a year is the annual maximum of daily streamflow observations



Sample 0.9 quantile of log annual streamflow (m^3/s) by station

1. Estimate return levels at each location using spatial modelling to borrow information across sites
2. Estimate joint exceedences probabilities to quantify regional risk
3. Test for changes over time in flood risk (as defined by 1. and 2.)

- Gaussian processes (GPs) are the workhorses of spatial statistics
- GPs inadequate for modeling extremes because they focus on deviations around the mean
- Max-stable processes (MSP) are a natural model for extremes, however:
 - Restrictive in the class of dependence types they can incorporate
 - Intractable likelihood for even moderately large problems

Methodological contributions

- For large spatial extremes datasets, we want:
 - Expressive and flexible spatial processes
 - Computational strategies for intractable likelihoods
- New model: Process mixture model specified as a convex combination of a GP and an MSP
- New computational algorithm: we use a Vecchia approximation and, following recent trends¹, use machine learning to approximate the intractable likelihood

¹See Polson and Sokolov (2023) for a review

The process mixture model (PMM)

- We model the extreme observation at spatial location \mathbf{s} and year t , $Y_t(\mathbf{s})$, with a generalized extreme value (GEV) distribution:

$$Y_t(\mathbf{s}) \sim \text{GEV}\{\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})\}$$

- The GEV parameters μ , σ and ξ model spatial variability in the marginal distribution
- The (copula) model is $Y_t(\mathbf{s}) = G\{V_t(\mathbf{s}); \mathbf{s}\}$ where G is the function so that $Y_t(\mathbf{s})$ has the desired marginal distribution
- Spatial dependence is capture by $V(\mathbf{s})$ which is modelled as a mixture of a GP and a MSP to allow for spatial dependence

The process mixture model is

$$V_t(\mathbf{s}) = \delta R_t(\mathbf{s}) + (1 - \delta)W_t(\mathbf{s})$$

- $R_t(\mathbf{s})$ is an MSP transformed to have exponential margins
- $W_t(\mathbf{s})$ is a GP transformed to have exponential margins
- R_t and W_t are independent with each other and over t
- The parameter $\delta \in [0, 1]$ controls the dependence regime
- This generalizes Huser and Wadsworth (2019) to have spatially-varying R

Asymptotic dependence

- Extremal spatial dependence is quantified by

$$\chi_u(\mathbf{s}_1, \mathbf{s}_2) = \text{Prob}\{V_t(\mathbf{s}_1) > q_u | V_t(\mathbf{s}_2) > q_u\}$$

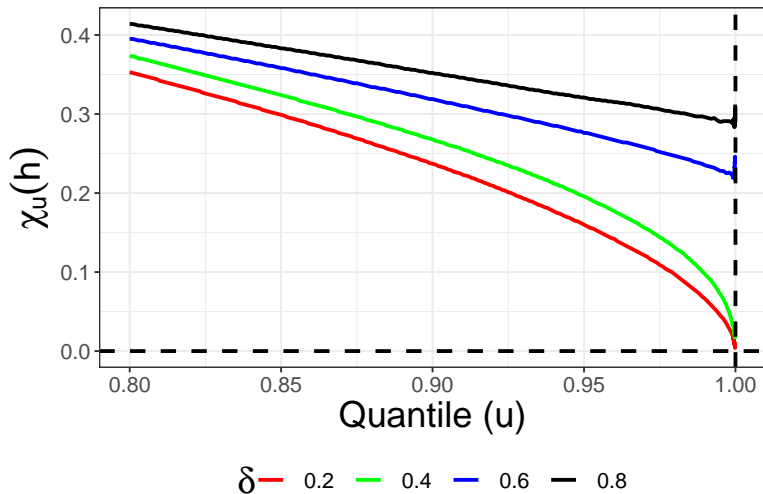
where q_u is the u quantile

- If the process is isotropic, $\chi_u(\mathbf{s}_1, \mathbf{s}_2) \equiv \chi_u(h)$ for $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$
- The process is asymptotically dependent if

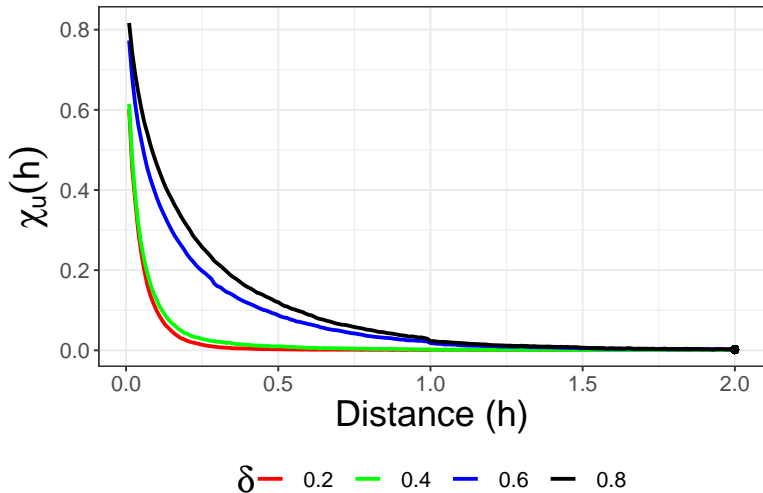
$$\lim_{u \rightarrow 1} \chi_u(h) > 0$$

- For any fixed h , the PMM is asymptotically dependent iff $\delta > 0.5$
- Unlike the common R model of H&W, the PMM model is asymptotically independent as $h \rightarrow \infty$

Asymptotic dependence: $\lim_{u \rightarrow 1} \chi_u(h) > 0$ iff $\delta > 0.5$



Spatial dependence: $\lim_{h \rightarrow \infty} \chi_u(h) = 0$ for any δ



- The PMM has desirable properties, but the likelihood function is intractable even for two observations
- However, it is straightforward to draw samples from the model
- We propose to approximate the likelihood by training deep learning models to data simulated from the PMM
- Approximating a high-dimensional density is difficult, so we use a Vecchia approximation reduce the approximation to a sequence of a univariate problems

- The model has two sets of parameters
 - θ_1 are the GEV parameters in μ , σ and ξ
 - θ_2 are the PMM parameter, i.e., δ , the spatial range of $W_t(\mathbf{s})$, etc
- The copula likelihood separates as

$$f_y(y_1, \dots, y_n; \theta_1, \theta_2) = f_u(u_1, \dots, u_n; \theta_2) \prod_{i=1}^n \left| \frac{dF_{\mathbf{S}}(y_i; \theta_1)}{dy_i} \right|$$

- The Vecchia approximation² for the first term is

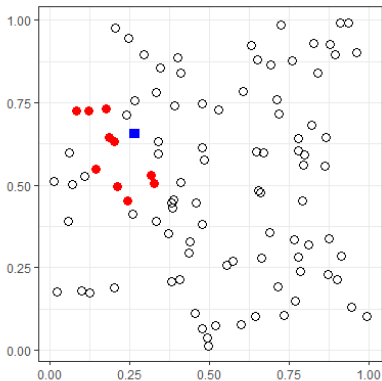
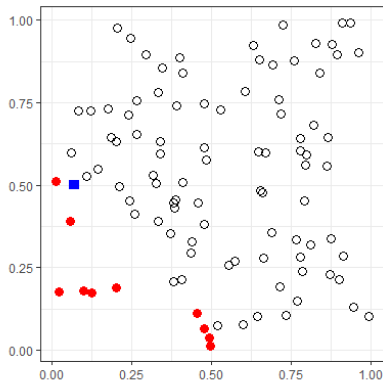
$$f_u(u_1, \dots, u_n; \theta_2) = \prod_{i=1}^n f(u_i | \theta_2, u_1, \dots, u_{i-1}) \approx \prod_{i=1}^n f_i(u_i | \theta_2, u_{(i)}), \quad (1)$$

for $u_{(i)} = \{u_j; j \in \mathcal{N}_i\}$ and neighboring set $\mathcal{N}_i \subseteq \{1, \dots, i-1\}$

- $u_{(i)}$ is the Vecchia neighboring set

²Vecchia (1988), Stein *et al.* (2004), Datta *et al.* (2016), Katzfuss and Guinness (2021)

Deep Learning Vecchia approximation for the PMM



The Vecchia neighboring set has 10 locations in this example

- There is no analytical form for $f_i(u_i|\boldsymbol{\theta}_2, u_{(i)})$
- It is a function of $\mathbf{x}_i = (\boldsymbol{\theta}_2, u_{(i)})$
- It is approximated using semi parametric quantile regression (SPQR)³ as:

$$f_i(u_i|\mathbf{x}_i, \mathcal{W}_i) = \sum_{k=1}^K \pi_k(\mathbf{x}_i, \mathcal{W}_i) \cdot B_k(u_i) \quad (2)$$

- M-spline basis functions $B_k(u) \geq 0$: satisfy $\int B_k(u)du = 1$ for all k
- The probability weights $\pi_k(\mathbf{x}_i, \mathcal{W}_i)$ are modeled as softmax outputs from a feed-forward neural network (FFNN)
- The FFNN weights to be learned are \mathcal{W}_i

³Xu and Reich (2023)

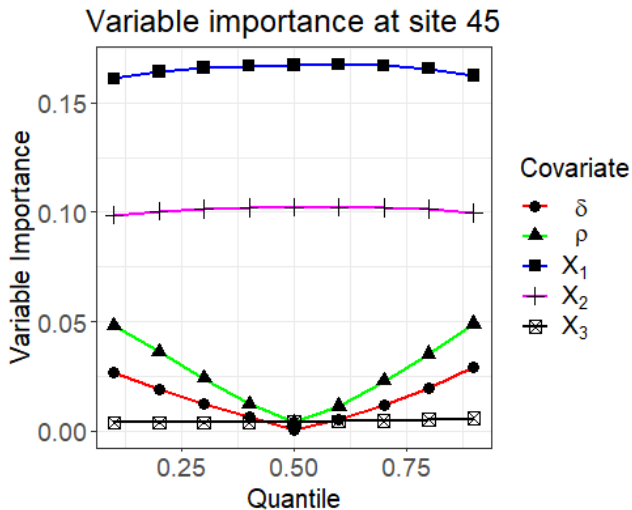
- The approximation of f_i is not contingent on real data
- We approximate the density by simulating data from the model and training SPQR on these synthetic data
- This approximation can be arbitrarily precise because
 1. SPQR spans all conditional densities smooth in its arguments
 2. The target f_i is univariate and \mathbf{x}_i is low-dimensional
 3. We can simulate a massive number of samples to train the model
- Given the FFNN weights/approximate likelihood, parameter estimation is carried out using standard MCMC

1. Generate plausible values of $\boldsymbol{\theta}_2$ from a design distribution p^*
2. Generate $U_k(\mathbf{s})$ at $\mathbf{s} \in \{\mathbf{s}_i, \mathbf{s}_{(i)}\}$ given $\boldsymbol{\theta}_{2k}$ for $k = 1, \dots, N$
3. Define features $\mathbf{x}_{ik} = (\boldsymbol{\theta}_{2k}, u_{(i)k})$, where $u_{(i)k} = \{u_k(\mathbf{s}); \mathbf{s} \in \mathbf{s}_{(i)}\}$
4. Solve $\hat{\mathcal{W}}_i \leftarrow \underset{\mathcal{W}}{\operatorname{argmax}} \prod_{k=1}^N f_i(u_{ik} | \mathbf{x}_{ik}, \mathcal{W})$ for $f_i(u | \mathbf{x}, \mathcal{W})$ using *SPQR*

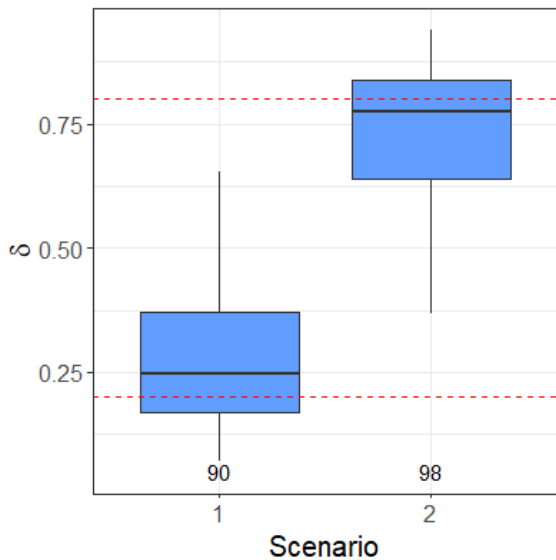
This strategy is not specific to the GEV/PMM model

Simulation Study - Process Mixture Model

- Marginal distribution is $Y_t(\mathbf{s}) \sim \text{GEV}\{\mu_0(\mathbf{s}) + \mu_1(\mathbf{s}) \cdot t, \sigma(\mathbf{s}), \xi(\mathbf{s})\}$
- GEV parameters set to be smooth functions of space
- R_t and W_t both have powered exponential dependence functions with power set to one
- Their effective spatial ranges are the same
- Data generated at 50 locations and 50 time points per location
- We fit the model with GP priors for the GEV parameters and uninformative priors for other parameters
- SPQR settings: 50 epochs, batch size 100, learning rate 0.001, 2 hidden layers (30, 15 neurons), 15 output knots, 10^6 obs.



Sampling distribution of the posterior mean of δ

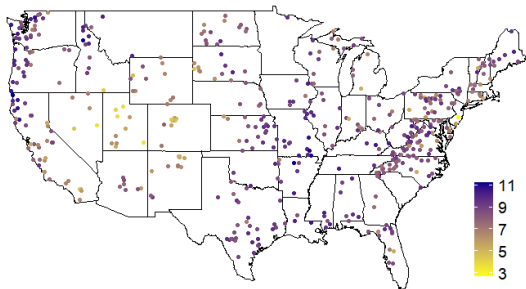


Coverage of 95% intervals

	μ_0	μ_1	σ	ξ
$\delta = 0.2$	93	96	92	94
$\delta = 0.8$	94	96	92	96

Coverage (%) for marginal GEV parameters under 2 scenarios based on MCMC simulations over 100 datasets.

Case study: extreme streamflow data in the US



- HCDN has 489 locations across the US
- We use annual maximum streamflow from 1972–2021
- The plot above is the sample 0.9 quantile at each station

Spatiotemporally varying coefficients model for the marginals

- $Y_t(\mathbf{s})$ is the log annual maxima for year t and location \mathbf{s}
- GEV marginals with spatiotemporally varying coefficients

$$Y_t(\mathbf{s}) \sim \text{GEV} \{ \mu_0(\mathbf{s}) + \mu_1(\mathbf{s}) \cdot X_t, \sigma(\mathbf{s}), \xi(\mathbf{s}) \}, \quad (3)$$

$$X_t = (\text{year}_t - 1996.5)/10 \text{ for } \text{year}_t = 1972 + t - 1$$

- The trend parameter $\mu_1(\mathbf{s})$ captures change per decade at site \mathbf{s}

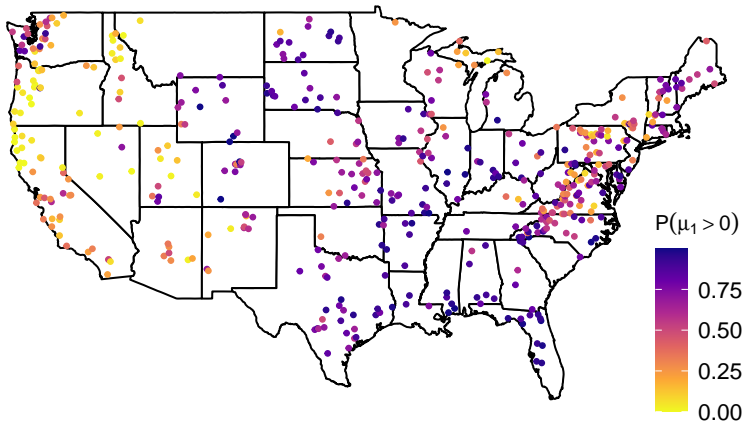
Model comparisons

Estimates (standard errors) from leave-one-out cross validation and the Watanabe-Akaike information criterion

	PMM	HW	MSP	GP
LOO-CV	29108 (540)	29708 (544)	32058 (583)	33842 (561)
WAIC	29559 (549)	30193 (565)	33441 (552)	34440 (585)

- PMM is the proposed model
- HW is the Huser and Wadsworth model (PMM with $R_t(\mathbf{s}) \equiv R_t$)
- MSP is the max-stable process (PMM with $\delta = 1$)
- GP is the Gaussian process (PMM with $\delta = 0$)

Posterior probability $\mu_1(s) > 0$



Spatial correlation parameters

- The posterior means (sd) of the PMM spatial parameters are $\hat{\delta} = 0.45$ (0.02), $\hat{\rho} = 807$ (45) km, and $\hat{r} = 0.92$ (0.004).
- The posterior of δ has a 95% interval of (0.40, 0.49), indicating the asymptotic independence regime with high probability
- The GEV Matérn smoothness parameter estimate is $\hat{\kappa} = 0.60$ (0.03)
- The four range parameters (km) are $\hat{\rho}_{\mu_0} = 12435$ (10645), $\hat{\rho}_{\mu_1} = 27605$ (10689), $\hat{\rho}_{\sigma} = 20311$ (11232) and $\hat{\rho}_{\xi} = 20320$ (11481)

- We compute at the joint posterior probability that all 10 CO sites exceed their observed 0.90 quantile
- The joint exceedance probability is 0.075 (0.04) in 1972 and 0.170 (0.046) in 2021
- The probability that the joint exceedance in 2021 is higher than in 1972 is 0.90
- Under independence, the joint exceedance probabilities are all approximately 10^{-10}

Summary and Discussion

- Extreme value analysis of climate signals is of growing importance, but existing methods are often intractable
- The process mixture model identifies patterns of increasing streamflow due to changing climate within the US
- Flexible, tractable, parallelizable, can take advantage of GPU acceleration
- Main idea can be applied to virtually any spatial process

References

- Majumder, R., Reich, B. J., and Shaby, B. A. (2022) Modeling extremal streamflow using Deep learning approximations and a flexible spatial process. *arXiv preprint*, arXiv:2208.03344.
- Xu, S. and Majumder, R. (2022) SPQR: Semi-Parametric Quantile Regression. R package version 0.1.0.
<https://cran.r-project.org/package=SPQR>
- Xu, S. G. and Reich, B. J. (2023) Bayesian non-parametric quantile process regression and estimation of marginal quantile effects. *Biometrics*, 79, 151–164

Acknowledgments: SE National Synthesis Wildfire, USGS National Climate Adaptation Science Center (G21AC10045), NSF (CBET2151651, DMS2152887), NIH (R01ES031651-01)

Related References

- Chui, C., Smith, P. and Ward, J. (1980) Degree of L_p Approximation by Monotone Splines. *SIAM Journal on Mathematical Analysis*, 11:436–447.
- Hornik, K., Stinchcombe, M. and White, H. (1989) Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366.
- Huser, R. and Wadsworth, J. L. (2019) Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114:434–444.
- Huser, R., Stein, M. L. and Zhong, P. (2022) Vecchia likelihood approximation for accurate and fast inference in intractable spatial extremes models. *arXiv preprint arXiv:2203.05626*.
- Polson and Sokolov (2023), Generative AI for Bayesian computation, arxiv:2305.14972.
- Stein, M. L., Chi, Z. and Welty, L. J. (2004) Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66:275–296.
- Vecchia, A. V. (1988) Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50, 297–312.