Non-stationary Process Mixtures for Extreme Streamflow Forecasting

NC STATE UNIVERSITY

Reetam Majumder   Brian J. Reich

Overview

Motivation: The **central US (CUS)** corresponds to 2 **HUC-02 regions (10L and 11)**, and is characterized by severe convective storms [5], and precipitation trends that could potentially influence flooding. Extreme streamflow is a key indicator of flood risk; quantifying the changes in its distribution under non-stationary climate conditions is key to mitigating the impact of flooding events.

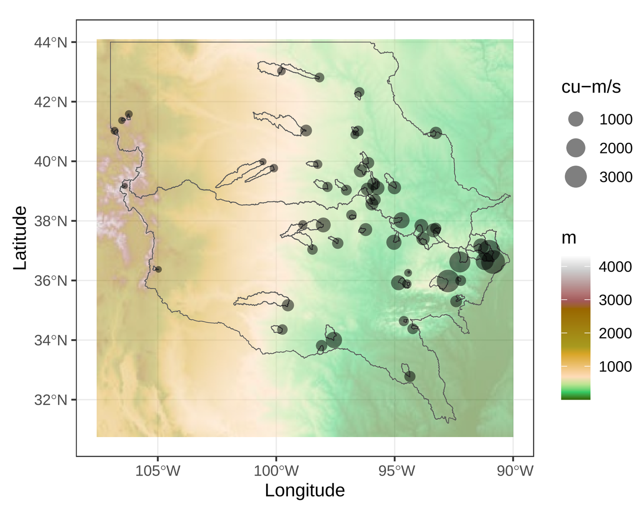

Figure 1. 0.99 quantiles of annual streamflow maxima at 55 CUS HCDN stations.

Challenges: Expressive spatial extremes processes often have intractable likelihoods, making computation challenging.

Goal: Develop a flexible and tractable spatial extremes model for climate-informed estimation of annual streamflow maxima.

Datasets: Annual streamflow maxima from **HCDN** (1972–2021), annual/seasonal precipitation from **NClimGrid** (1972–2021) as covariates, downscaled/bias-corrected precipitation from **MACA** (1972–2035) for projections.

The non-stationary process mixture model (NPMM)

Let $Y_t(\mathbf{s})$ be the extreme observation at time t and spatial location $\mathbf{s}$:

$$Y_t(\mathbf{s}) \sim \text{GEV}\{\mu_t(\mathbf{s}), \sigma_t(\mathbf{s}), \xi_t(\mathbf{s})\}.$$

GEV parameters vary spatially and depend on precipitation:

$$\mu_t(\mathbf{s}) = \mu_0(\mathbf{s}) + \sum_{i=1}^{5} \mu_i(\mathbf{s}) X_{it}(\mathbf{s}), \qquad \sigma_t(\mathbf{s}) = \sigma(\mathbf{s}), \qquad \xi_t(\mathbf{s}) = \xi(\mathbf{s}). \tag{1}$$

$X_{it}(\mathbf{s})$ includes seasonal (location-specific) and annual precipitation (regional).

Given streamflow data ($y_{1:n}$), marginal parameters ($\boldsymbol{\theta}_1$), and spatial parameters ($\boldsymbol{\theta}_2$), our Bayesian hierarchical model is:

Prior model: $\boldsymbol{\theta}_1 \sim p(\boldsymbol{\theta}_1) \perp \boldsymbol{\theta}_2 \sim p(\boldsymbol{\theta}_2)$,

Data model: $f_y(y_1, ..., y_n | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \underbrace{f_u(u_1, ..., u_n | \boldsymbol{\theta}_2)}_{\text{spatial dependence}} \underbrace{\prod_{i=1}^{n} \left| \frac{dF(y_i|\boldsymbol{\theta}_1)}{dy_i} \right|}_{\text{marginal GEV likelihoods}}$.

The likelihood decomposition can be viewed as a change-of-variables, or a copula.

$U_t(\mathbf{s}) := F_{t,\mathbf{s}}(Y_t(\mathbf{s}))$ are spatially-correlated uniform variables. A spatial dependence model on $U_t(\mathbf{s})$ is obtained via the transformation $U_t(\mathbf{s}) = G_{t,\mathbf{s}}(V_t(\mathbf{s}))$:

$$V_t(\mathbf{s}) = \delta_t(\mathbf{s}) R_t(\mathbf{s}) + (1 - \delta_t(\mathbf{s})) W_t(\mathbf{s}), \tag{2}$$

where $R_t(\mathbf{s})$ is a max-stable process, $W_t(\mathbf{s})$ is a Gaussian process,

$\delta_t(\mathbf{s}) \in [0, 1]$ are weight parameters depending on regional annual precipitation. We use δ_1 and δ_2 for the 2 regions within the CUS.

Eqn. (2) defines the NPMM. No closed form available for the likelihood.

Extreme streamflow distribution and projections for the CUS

- δ_1, δ_2, corresponding to HUC-02 regions 10L and 11, do not change with changes in basin-wide annual precipitation
- δ_1, δ_2 have posterior means of 0.53 and 0.71 (**asymptotically dependent**)
- **Precipitation is a significant predictor** of streamflow maxima
- Summer (AMJ) precipitation is the most significant predictor at most locations.

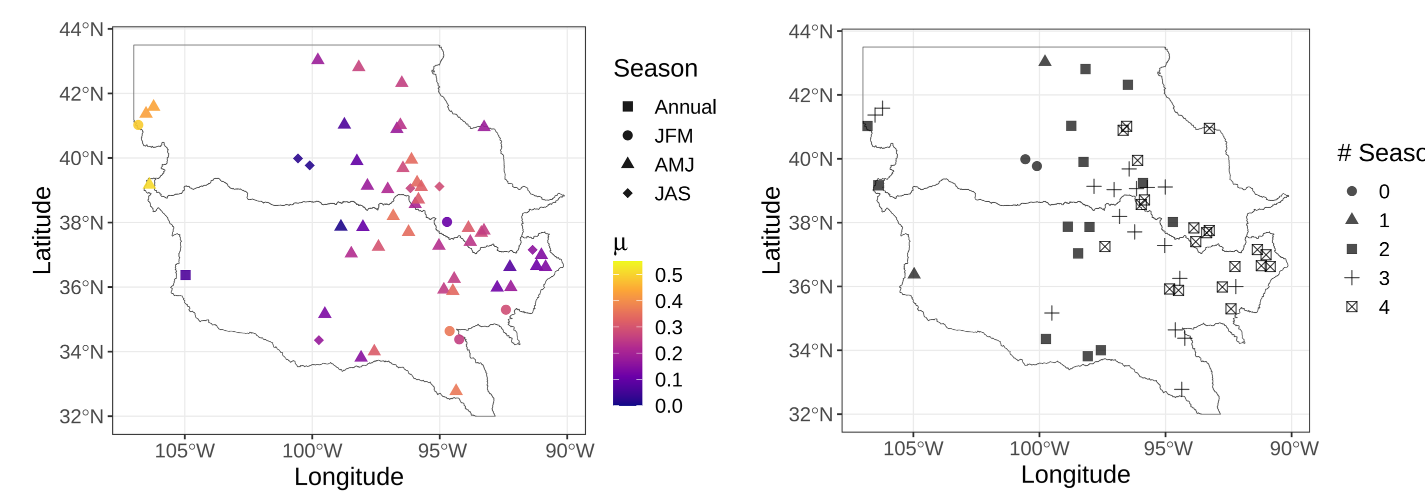

Figure 2. Estimates of $\mu(\mathbf{s}) = \max(\mu_j(\mathbf{s}))$ for $j = 2:5$ corresponding to the 4 seasons with shapes denoting the season with the highest slope value (left), and number of seasons (excluding annual) where $\mathbb{P}[\mu(\mathbf{s}) > 0] > 0.90$ (right).

6 CMIP5 models chosen for each representative climate pathway (RCP) scenario (3 wet + 3 dry). Baseline period of 1972–2005, projection period of 2006–2035.

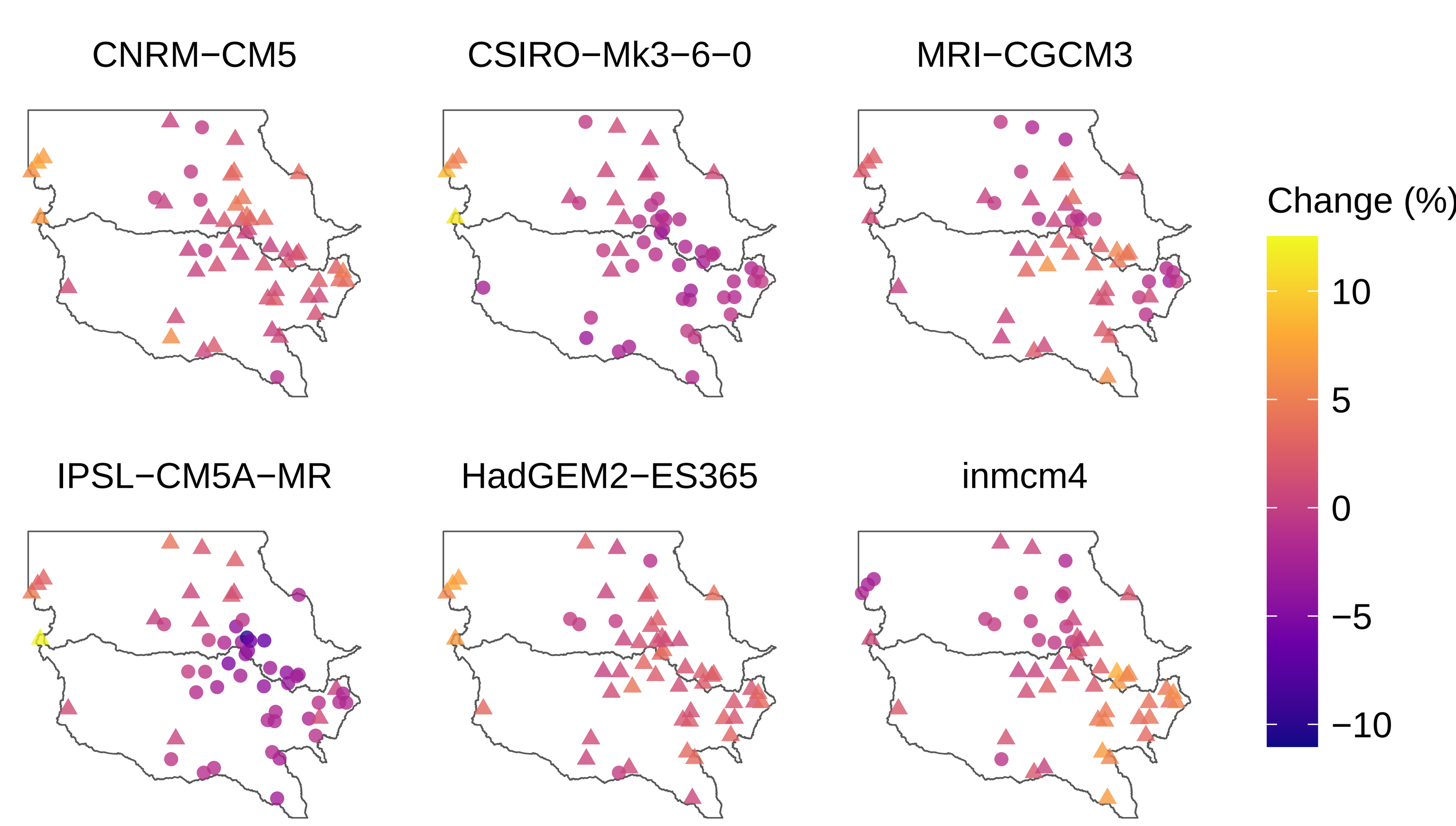

Figure 3. Percentage change in observed 0.90 quantile under RCP 8.5 between baseline and projection periods.

Changes from -10.3%–12.3% for the 0.90 quantile of annual streamflow maxima.

All 6 models under RCP 4.5 and 4 models under RCP 8.5 estimate that **more than 50% locations have increased streamflow in the projection period**.

Less pronounced but similar results for 0.99 quantile.

Future work: Additional covariates (e.g., temperature), long term forecasts.

Inference for the NPMM

Problem: Evaluate the intractable NPMM likelihood for different values of $\boldsymbol{\theta}_2$ and $(y_1, \ldots, y_n)$ without knowing its functional form.

Approach: Density estimation of a surrogate likelihood based on a Vecchia decomposition [3] of the joint distribution $f_u(u_1, ..., u_n | \boldsymbol{\theta}_2)$:

$$f_u(u_1, ..., u_n | \boldsymbol{\theta}_2) = \prod_{i=1}^{n} f_i(u_i | \boldsymbol{\theta}_2, u_1, ..., u_{i-1}) \approx \prod_{i=1}^{n} f_i(u_i | \boldsymbol{\theta}_2, u_{(i)}), \tag{3}$$

$u_{(i)} \subseteq \{u_1, \ldots, u_{i-1}\}$. The subset of locations $\mathbf{s}_{(i)}$ are often the nearest neighbors.

Density regression is carried out for each of the $n-1$ terms separately using neural networks in a **semi-parametric quantile regression (SPQR)** model [4]:

$$f_i(u_i | \mathbf{x}_i, \mathcal{W}) = \sum_{k=1}^{K} \pi_{ik}(\mathbf{x}_i, \mathcal{W}_i) B_k(u_i), \tag{4}$$

$$\pi_{ik}(\mathbf{x}_i, \mathcal{W}_i) = f_i^{NN}(\mathbf{x}_i, \mathcal{W}_i), \text{ for } i = 2, \ldots, n. \tag{5}$$

- $\mathbf{x}_i = (u_{(i)}, \boldsymbol{\theta}_2)$ are treated as **covariates**, with u_i as the **response variable**
- Each NN maximizes the log-likelihood of a univariate conditional (RHS of (4))
- NNs are trained using synthetic data
- Given a value of $\boldsymbol{\theta}_2$ and $u_{(i)} = F(y_{(i)})$, we can then evaluate $f_u(u_1, ..., u_n | \boldsymbol{\theta}_2)$ as a product of surrogate conditional distributions
- Can be used in an MCMC to estimate $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$.

Extremal spatial dependence often measured in terms of the upper-tail coefficient:

$$\chi_u(\mathbf{s}_1, \mathbf{s}_2) := \text{Prob}\{U(\mathbf{s}_1) > u | U(\mathbf{s}_2) > u\}, \tag{6}$$

where $u \in (0, 1)$ is a threshold. $U(\mathbf{s}_1)$ and $U(\mathbf{s}_2)$ are defined as asymptotically dependent if

$$\chi(\mathbf{s}_1, \mathbf{s}_2) = \lim_{u \to 1} \chi_u(\mathbf{s}_1, \mathbf{s}_2) \tag{7}$$

is positive and independent if $\chi(\mathbf{s}_1, \mathbf{s}_2) = 0$. For the PMM/NPMM [1, 2],

$\delta < 0.5 \implies$ asymptotic independence, and

$\delta > 0.5 \implies$ asymptotic dependence.

The NPMM is flexible (desirable asymptotic properties), and tractable (computational cost increases linearly in number of locations).

Funding and Acknowledgments

This work was supported by grants from the Southeast National Synthesis Wildfire and the USGS NCASC (G21AC10045), and the NSF (DMS2152887, CBET2151651, DMS1929348). The authors thank Prof. Sankarasubramanian Arumugam of NC State University for discussion of the data and scope of the project.

References

[1] R. Majumder, B. J. Reich, and B. A. Shaby. Modeling extremal streamflow using deep learning approximations and a flexible spatial process. 2022. doi:10.48550/arXiv.2208.03344.

[2] Reetam Majumder and Brian J. Reich. A deep learning synthetic likelihood approximation of a non-stationary spatial model for extreme streamflow forecasting. *Spatial Statistics*, page 100755, 2023. doi:https://doi.org/10.1016/j.spasta.2023.100755.

[3] M. L. Stein, Z. Chi, and L. J. Welty. Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2):275–296, 2004.

[4] S. G. Xu and B. J. Reich. Bayesian nonparametric quantile process regression and estimation of marginal quantile effects. *Biometrics*, 00:1–14, 2021. doi:https://doi.org/10.1111/biom.13576.

[5] L. Zhang, M. D. Risser, E. M. Molter, M. F. Wehner, and T. A. O'Brien. Accounting for the spatial structure of weather systems in detected changes in precipitation extremes. *Weather and Climate Extremes*, 38:100499, 2022.