Approximating Likelihoods for Spatial Extremes with Deep Learning

Reetam Majumder, NCSU

Joint work with Brian J. Reich (NCSU) and Benjamin A. Shaby (CSU)

CMStatistics 2022

NC STATE UNIVERSITY

Background

Motivation

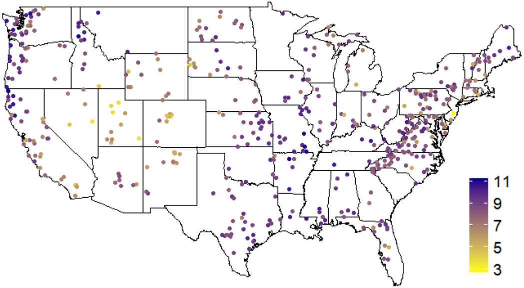


Figure 1: Sample 0.9 quantile of log annual streamflow maxima at 489 locations.
Source: USGS Hydro Climatic Data Network (HCDN).

- 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

Max-stable processes for spatial extremes

- Gaussian processes (GP) are inadequate for modeling extremes
- Max-stable processes (MSP) are a natural model for block maxima, *however*:
  - Intractable likelihood for even moderately large problems
  - Restrictive in the class of dependence types they can incorporate
- Approximation - Composite Likelihood[1]
  - Inefficient, finite sample bias, computational challenges for large *n*
- Approximation - Vecchia approximation
  - Simplifies likelihoods for spatial processes including MSPs[2]

[1] Padoan *et al.* (2010)
[2] Huser *et al.* (2022)

Objectives

- For large spatial extremes datasets, we want:
  - Expressive and flexible spatial processes
  - Computational strategies for intractable likelihoods
- **Our approach** - Process mixture model specified as a convex combination of a GP and an MSP
- Vecchia approximation simplifies likelihood as a product of univariate (intractable) PDFs
- Deep learning to approximate the intractable PDFs

The Process Mixture Model

The process mixture model (PMM)

- Let $Y(\mathbf{s})$ be the extreme observation at spatial location $\mathbf{s}$ with a generalized extreme value (GEV) distribution:

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

- $Y(\mathbf{s}) \sim F_{\mathbf{s}}$, $U(\mathbf{s}) = F_{\mathbf{s}}(Y(\mathbf{s}))$, and express the joint likelihood as

$$f_y(y_1, \dots, y_n; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = f_u(u_1, \dots, u_n; \boldsymbol{\theta}_2) \prod_{i=1}^{n} \left| \frac{dF_{\mathbf{s}}(y_i; \boldsymbol{\theta}_1)}{dy_i} \right|, \qquad (1)$$

  where $y_i \equiv y(\mathbf{s}_i)$ and $u_i = F_{\mathbf{s}}(y_i; \boldsymbol{\theta}_1)$
- Take $U(\mathbf{s}) = G\{V(\mathbf{s})\}$ to get spatial dependence model on $U(\mathbf{s})$

$$V(\mathbf{s}) = \delta \cdot g_R\{R(\mathbf{s})\} + (1 - \delta) \cdot g_W\{W(\mathbf{s})\} \qquad (2)$$

Spatial dependence in the PMM

- Take $U(\mathbf{s}) = G\{V(\mathbf{s})\}$ to get spatial dependence model on $U(\mathbf{s})$

$$V(\mathbf{s}) = \delta \cdot g_R\{R(\mathbf{s})\} + (1 - \delta) \cdot g_W\{W(\mathbf{s})\}$$

- $R(\mathbf{s})$ is an MSP, $W(\mathbf{s})$ is a GP; $\delta \in [0, 1]$
- Conditional exceedance probability defined as:

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

- $\chi(\mathbf{s}_1, \mathbf{s}_2) = \lim_{u \to 1} \chi_u(\mathbf{s}_1, \mathbf{s}_2) > 0$ iff $\delta > 0.5 \implies$ asymptotic dependence
- $g_R\{R(\mathbf{s})\}, g_W\{W(\mathbf{s})\} \overset{iid}{\sim}$ Exponential(1)
- Process mixture $V(\mathbf{s})$ - hypoexponential distribution marginally
- Generalization of *Huser and Wadsworth (2019)*.

Deep Learning Vecchia approximation for the PMM

- Joint likelihood:
  $f_y(y_1, ..., y_n; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = f_u(u_1, ..., u_n; \boldsymbol{\theta}_2) \prod_{i=1}^n \left| \frac{dF_{\mathbf{S}}(y_i; \boldsymbol{\theta}_1)}{dy_i} \right|$
- Approximate the first term of the likelihood as[3]

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

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

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

Deep Learning Vecchia approximation for the PMM

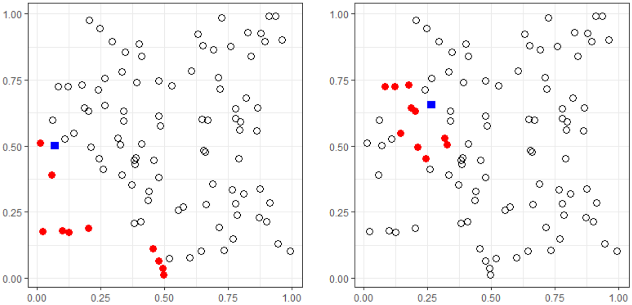


Figure 2: Vecchia neighboring sets when locations are ordered by distance from origin

- The Vecchia neighboring set has up to 10 locations in this example
- No analytical form for $f_i(u_i|\boldsymbol{\theta}_2, u_{(i)})$

Deep Learning Vecchia approximation for the PMM

- Model $f_i(u_i|\theta_2, u_{(i)})$ using semi parametric quantile regression (SPQR)[4] as:

$$f(u_i|\mathbf{x}_i, \mathcal{W}_i) = \sum_{k=1}^{K} \pi_k(\mathbf{x}_i, \mathcal{W}_i) \cdot B_k(u_i) \tag{4}$$

- M-spline basis functions $B_k(u) \geq 0$: satisfy $\int B_k(u)du = 1$ for all k
- Probability weights $\pi_k(\mathbf{x}_i, \mathcal{W})$: softmax outputs from a feed-forward neural network (FFNN)
- Can approximate conditional densities smooth in its arguments[5]

[4] Xu and Reich (2021)
[5] Chui *et al.* (1980), Hornik *et al.* (1989)

SPQR implementation

- Each $f(u_i|\mathrm{x}_i, \mathcal{W}_i)$ is modeled using its own FFNN; $\mathrm{x}_i := (\boldsymbol{\theta}_2, u_{(i)})$
- FFNN weights $\mathcal{W}_i$ for location i estimated using synthetic data generated using plausible parameter values
- Parameter estimation carried out afterwards using MCMC

Numerical Results

Simulation Study - Process Mixture Model

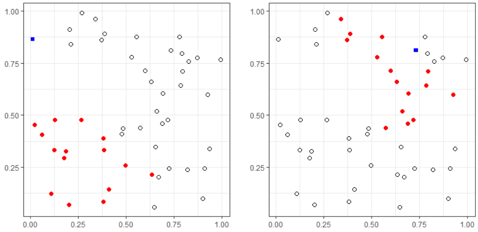
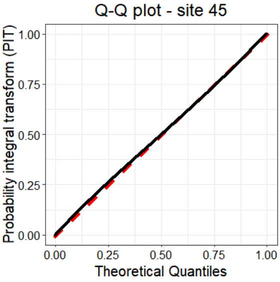


Figure 3: Locations used in the EVP simulation studies: 50 locations, and nearest neighbor assignments for locations 16 (left) and 45 (right).

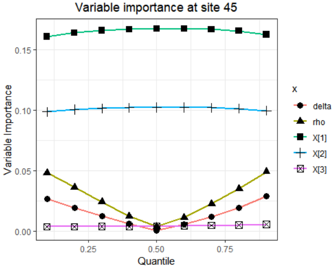
- Common smoothness parameter $\alpha_R = \alpha_W = \alpha = 1$
- Range $\rho = \rho_W$, $\rho_R = 0.19\rho$
- Range chosen such that distance at which GP correlation reaches 0.05 = distance at which $\chi_u(\mathbf{s}_1, \mathbf{s}_2)$ for MSP is 0.05, where

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

SPQR model fit diagnostics - PMM



(a) Q-Q plots for checking goodness of fit.



(b) Variable importance of δ, ρ, and nearest neighbors.

Figure 4: **Model diagnostics for process mixture model**: Q-Q plot and VI plot.

SPQR settings: 50 epochs, batch size 100, learning rate 0.001, 2 hidden layers (30, 15 neurons), 15 output knots, 10^6 obs.

SPQR model fit diagnostics

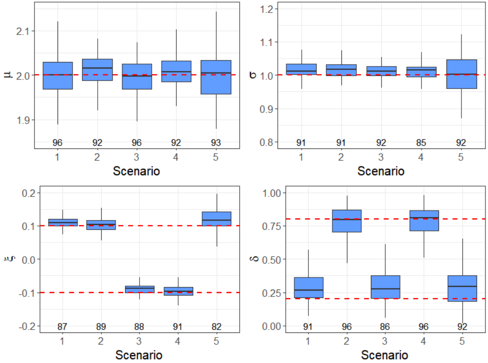


Figure 5: Sampling distribution of posterior means: Horizontal dashed lines are true values with empirical coverage of the 95% intervals at the bottom.

- Scenario 5: MCAR with probability $\pi_M = 0.05$ and censored below the threshold $T = \hat{q}_{0.5}$ (over space and time)

Case Study: Extreme Streamflow

Case study: extreme streamflow data

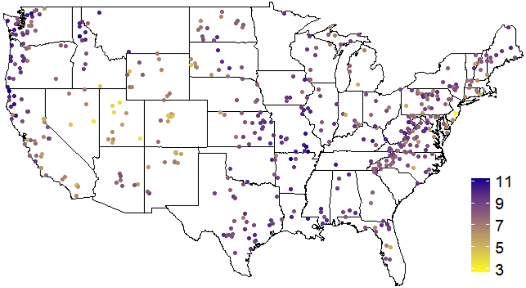


Figure 6: Sample 0.9 quantile of log annual streamflow maxima $Y_t(s)$ at 489 locations.

- **489 locations** across the US part of the USGS Hydro-Climatic Data Network (HCDN)
- **50 years** of complete data from 1972–2021 - annual streamflow maxima

Spatio-temporally varying coefficients model for the marginals

- $Y_t(\mathbf{s})$: log annual maxima for year t, location $\mathbf{s}$
- GEV marginals with spatio-temporally varying coefficients (STVC):

$$Y_t(\mathbf{s}) \sim \text{GEV}\left[\mu_0(\mathbf{s}) + \mu_1(\mathbf{s})X_t, \exp\{\sigma(\mathbf{s})\}, \xi(\mathbf{s})\right], \tag{5}$$

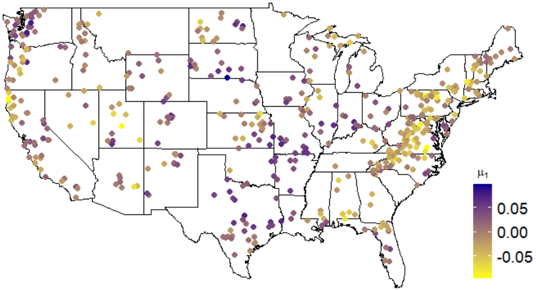
$X_t = (\text{year}_t - 1996.5)/10$ for $\text{year}_t = 1972 + t - 1$

- X_t captures changes in the location due to changing climate
- $(\mu_0(\mathbf{s}), \mu_1(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})) \sim$ GPs with common range parameter ρ^*
- FFNN architecture: 15 neighbors, 2 hidden layers (30, 20 neurons), 15 output knots, batch size 1000, learning rate 0.01, 50 epochs

Posterior estimates

Posterior means and SD of spatial parameter estimates:

- $\hat{\delta}$: 0.47 (0.02); $\hat{\rho}$: 1004 km (80); $\hat{r}$: 0.56 (0.07); $\hat{\rho}^*$: 17907 km (1806)
- Asymptotic independence regime with high probability



•

Figure 7: Posterior mean of $\mu_1(\mathbf{s})$ at 489 gauges for log annual streamflow maxima.

- Positive values of μ_1 indicate increasing streamflow maxima

Posterior estimates

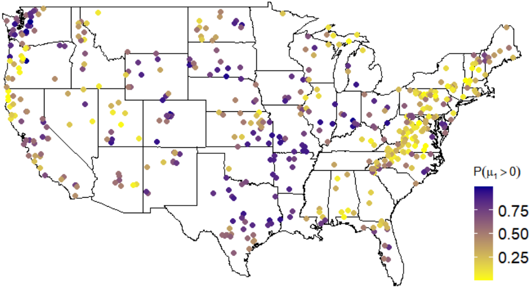


Figure 8: Estimates of $Pr[\mu_1(\mathbf{s}) > 0]$ for the GEV location parameters.

- Higher values indicate stronger evidence of increased streamflow magnitude between 1972 and 2021
- Joint exceedances can be studied for clusters; e.g. in CO, posterior probability that 0.9 quantile has gone up is 0.975

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

SPQR R package

- Majumder, R. and Reich, B. J. (2022) A deep learning synthetic likelihood approximation of a non-stationary spatial model for extreme streamflow forecasting. *arXiv preprint*, arXiv:2212.07267.
- 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. G. and Reich, B. J. (2021) Bayesian non-parametric quantile process regression and estimation of marginal quantile effects. *Biometrics*, 00, 1–14.

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

Related References

- 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.
- Vecchia, A. V. (1988) Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50, 297–312.
- 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.
- 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.

Appendix

Global SPQR

Algorithm 1 Global SPQR approximation

Require: Locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ with neighbor locations $\mathbf{s}_{(1)}, \ldots, \mathbf{s}_{(n)}$
Require: Design distribution p^*, sample size N
  $k \leftarrow 1$
  while $k \leq N$ **do**
    Draw sample location $\mathbf{s}_{l_k}$, where $l_k \in \{2, \ldots, n\}$
    Draw values of $\boldsymbol{\theta}_{2k} \sim p^*$, using (2)
    Generate $U(\mathbf{s}) = G\{V(\mathbf{s})\}$ at $\mathbf{s} \in \{\mathbf{s}_{l_k}, \mathbf{s}_{(l_k)}\}$, using (2)
    Define features $\mathbf{x}_{l_k} = (\boldsymbol{\theta}_{2k}, u_{(l_k)}, \mathbf{s}_{(l_k)} - \mathbf{s}_{l_k})$, where $u_{(l_k)} = \{U_{l_k}(\mathbf{s}); \mathbf{s} \in \mathbf{s}_{l_k}\}$
    $k \leftarrow k + 1$
  end while
  solve $\hat{\mathcal{W}} \leftarrow \arg_{\mathcal{W}} \max \prod_{k=1}^{N} f(u_{l_k} | \mathbf{x}_{l_k})$, for $f(u | \mathbf{x}, \mathcal{W})$ defined in (8), using SPQR

Local SPQR

Algorithm 2 Local SPQR approximation

Require: Locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ with neighbor locations $\mathbf{s}_{(1)}, \ldots, \mathbf{s}_{(n)}$
Require: Design distribution p^*, training sample size N
  $i \leftarrow 2$
  while $i \leq n$ **do**
    $k \leftarrow 1$
    while $k \leq N$ **do**
      Draw values of $\boldsymbol{\theta}_{2k} \sim p^*$
      Generate $U_k(\mathbf{s})$ at $\mathbf{s} \in \{\mathbf{s}_i, \mathbf{s}_{(i)}\}$ given $\boldsymbol{\theta}_{2k}$ using (2)
      Define features $\mathbf{x}_{ik} = (\boldsymbol{\theta}_{2k}, u_{(i)k})$, where $u_{(i)k} = \{U_k(\mathbf{s}); \mathbf{s} \in \mathbf{s}_{(i)}\}$
      $k \leftarrow k + 1$
    end while
    solve $\hat{\mathcal{W}}_i \leftarrow \underset{\mathcal{W}}{\text{argmax}} \prod_{k=1}^{N} f(u_{ik}|\mathbf{x}_{ik}, \mathcal{W})$ for $f(u|\mathbf{x}, \mathcal{W})$ defined in (8) using SPQR
    $i \leftarrow i + 1$
  end while

SPQR model fit diagnostics - GP

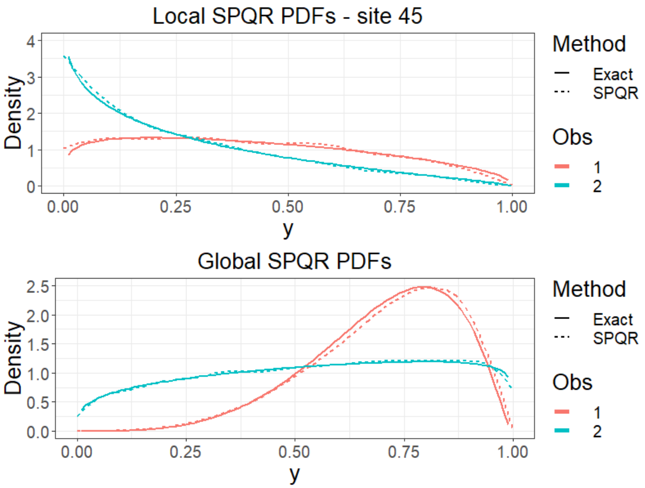
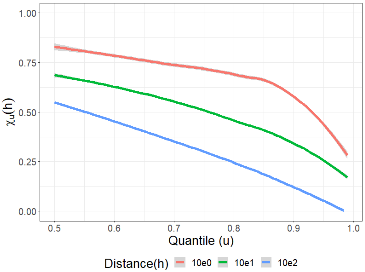
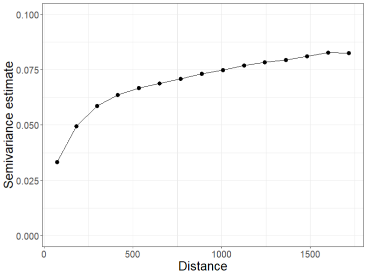


Figure 9: SPQR fit for simulated data: True and estimated PDFs for two out-of-sample observations fitted using local and global SPQR.



(a) Conditional exceedance $\chi_u(h)$ for log annual maximum streamflow computed for different distances.



(b) Sample variogram for log annual maximum streamflow, averaged over 50 years of data.

Figure 10: Spatial behaviour of log annual maximum streamflow.

Model priors

- $\mu_0(\mathbf{s}) = \tilde{\mu}_0(\mathbf{s}) + e(\mathbf{s})$
- $e(\mathbf{s}) \stackrel{iid}{\sim} \text{Normal}(0, v_{\mu_0})$, $\tilde{\mu}_0(\mathbf{s})$ is a GP
- $\text{E}\{\mu_0(\mathbf{s})\} = \beta_{\mu_0}$, variance $\text{V}\{\mu_0(\mathbf{s})\} = \tau^2_{\mu_0}$
- $\text{Cor}\{\mu_0(\mathbf{s}), \mu_0(\mathbf{s}')\} = \exp\{-||\mathbf{s} - \mathbf{s}'||/\rho^*\}$
- $\mu_1(\mathbf{s})$, the log scale $\sigma(\mathbf{s})$, and the shape $\xi(\mathbf{s})$ modeled similarly using GPs
- Common spatial range ρ^*
- $\beta_{\mu_0}, \beta_{\mu_1}, \beta_\sigma, \beta_\xi \stackrel{iid}{\sim} \text{Normal}(0, 100^2)$
- $\tau_{\mu_0}, \tau_{\mu_1}, \tau_\sigma, \tau^2_\xi \stackrel{iid}{\sim} \text{InvGamma}(0.1, 0.1)$
- $v_{\mu_0}, v_{\mu_1}, v_\sigma, v^2_\xi \stackrel{iid}{\sim} \text{InvGamma}(0.1, 0.1)$
- $\log(\rho^*) \sim \text{Normal}(9.74, 0.1^2)$
- $\delta \sim \text{Uniform}(0, 1)$ and $\rho \sim \text{Uniform}(0, 3126)$