Vecchia approximated density regression for intractable spatial extremes models

Reetam Majumder

Joint work with Brian Reich (NCSU) & Ben Shaby (CSU)

Workshop on Neural Simulation-Based Inference

STAMPS@CMU | Oct 5, 2025

UNIVERSITY OF ARKANSAS

Flash floods in Northeastern US (2023)

Vermont's rivers to continue rising overnight

Posted July 10, 2023 at 9:36 PM EDT

As the sun goes down in Vermont, attention is shifting to major rivers that are forecast to flood overnight and in the morning.

In central and northern Vermont, sections of the Winooski River, the Lamoille River and the Mad River are all expected to continue rising overnight and cause major flooding.

Jeff McDonough / Courtesy

- Extreme streamflow is a key indicator of flood risk

Source: vermontpublic.org

River flooding in Connecticut (2023–24)

Partial dam break on Yantic River in CT causes evacuation due to potential 'life threatening' floods

Connecticut Public Radio | authorBy Staff Report
publishedDateHeading January 10, 2024 at 9:11 AM EST
modifiedDateHeading January 11, 2024 at 10:38 AM EST

More than 27,000 in CT have no power in aftermath of storm; Connecticut River flood warning remains in place

Clean up work from a fallen tree at the intersection of Paddock Rd and Randolph Rd near Xavier High School in Middletown. (Aaron Flaum/Hartford Courant)

- Quantifying the probability and magnitude of extreme flooding events is key to mitigating their impacts

Source: Hartford Courant, Connecticut Public Radio

Case study: extreme streamflow data in the US

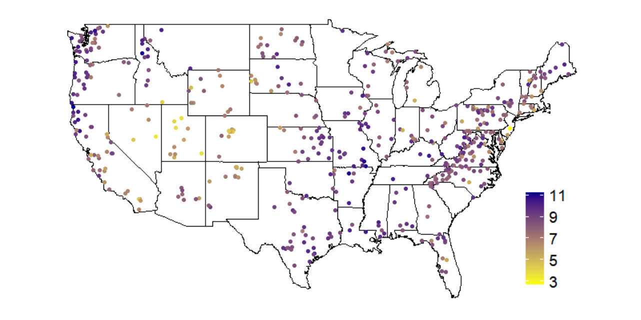


Figure 1: Sample 0.9 quantile of log annual streamflow maxima (m^3/s)

- The USGS Hydro Climatic Data Network (HCDN) provides streamflow data for watersheds which are minimally impacted by anthropogenic activity
- 487 stations with data from 1972–2021

Scientific objectives

1. Estimate return levels at each location using spatial modeling to borrow information across sites

2. Estimate joint exceedances probabilities to quantify regional risk

3. Test for changes over time in flood risk (as defined by 1. and 2.)

Spatial extreme value analysis

- Gaussian processes (GPs) are the workhorses of spatial statistics
- GPs are 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

Max-stable processes (MSP)

- **Non-spatial extremes**: Let $X_1, X_2, \ldots$ be *iid* variables, and suppose there exists normalizing functions $a_n > 0, b_n \in \mathbb{R}$ such that

$$R \stackrel{d}{=} \frac{\max_{i=1:n} X_i - b_n}{a_n},$$

where R is in the generalized extreme value (GEV) family

- **Spatial extremes**: If $\{X_i(\mathbf{s}) : i \in \mathbb{N}\}$ are independent copies of a process $\{X(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$, and if there exist $c_n(\mathbf{s}) > 0, d_n(\mathbf{s}) \in \mathbb{R}$, such that as $n \to \infty$,

$$\frac{\max_{i=1:n} X_i(\mathbf{s}) - d_n(\mathbf{s})}{c_n(\mathbf{s})} \to R(\mathbf{s}),$$

then $R(\mathbf{s})$ is either degenerate or an MSP (*de Haan, 1984*)

Methodological contributions

For large spatial extremes datasets, we want:

- Expressive and flexible spatial processes
  New model: Process mixture model specified as a convex combination of a GP and an MSP

- Computational strategies for intractable likelihoods
  New computational algorithm: we use a Vecchia approximation and, following recent trends[1], use machine learning to approximate the intractable likelihood

[1] See Polson and Sokolov (2023) for a review

Bayesian modeling framework for streamflow

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

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

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

- The CDF transformed variables $U(\mathbf{s}) := P\big(Y(\mathbf{s}) < y | \boldsymbol{\theta}_1\big)$ share common uniform marginal distributions but are spatially dependent

- If we want a Bayesian framework, we *do* need to somehow evaluate the likelihood

Marginal distribution of extremal streamflow

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

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

whose CDF is

$$P\big(Y(\mathbf{s}) < y | \boldsymbol{\theta}_1\big) = \exp\left\{-\left[1 + \xi(\mathbf{s})\left(\frac{\mathbf{y} - \mu(\mathbf{s})}{\sigma(\mathbf{s})}\right)\right]^{-1/\xi(\mathbf{s})}\right\} \tag{1}$$

- The GEV parameters $\boldsymbol{\theta}_1 = \{\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})\}$ model spatial variability in the marginal distribution

The process mixture model for spatial extremes

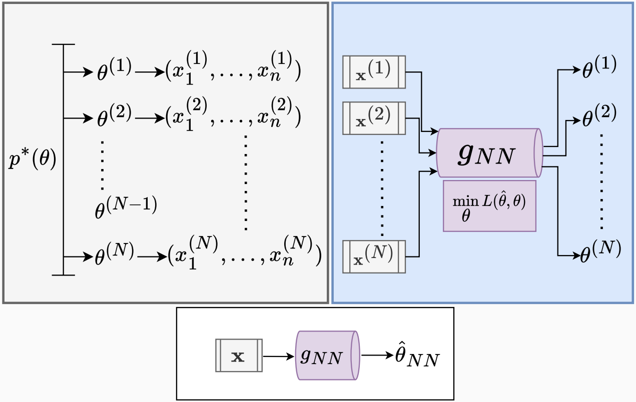
Our process mixture model (PMM) is

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

- $R(\mathbf{s})$ is an MSP transformed to have exponential margins
- $W(\mathbf{s})$ is a GP transformed to have exponential margins
- R and W are independent of each other (and over time)
- The parameter $\delta \in [0, 1]$ controls the dependence regime
- The transformation $U(\mathbf{s}) = G\{V(\mathbf{s}); \mathbf{s}\}$ gets us back to the copula
- This generalizes Huser and Wadsworth (2019) to have spatially-varying R

- 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

Likelihood-free inference using deep learning



Common forms of the functional of **x** include covariance matrices (*Gerber and Nychka, 2020*), or spatial maps (*Lenzi et al., 2021, Richards et al., 2023*)

Uncertainty for θ usually through bootstrapping

Bayesian modeling framework for streamflow

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

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

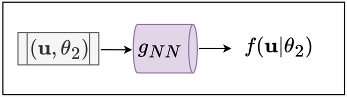
$$\text{Data model: } f_y(y_1, \ldots, y_n | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \underbrace{\prod_{i=1}^{n} \left| \frac{dF(y_i|\boldsymbol{\theta}_1)}{dy_i} \right|}_{\text{marginal GEV likelihoods}} \times \underbrace{f_u(u_1, \ldots, u_n | \boldsymbol{\theta}_2)}_{\text{spatial dependence}}.$$

- The CDF transformed variables $U(\mathbf{s}) := P\big(Y(\mathbf{s}) < y | \boldsymbol{\theta}_1\big)$ share common uniform marginal distributions but are spatially dependent

- If we want a Bayesian framework, we *do* need to somehow evaluate the likelihood

Inference for the PMM

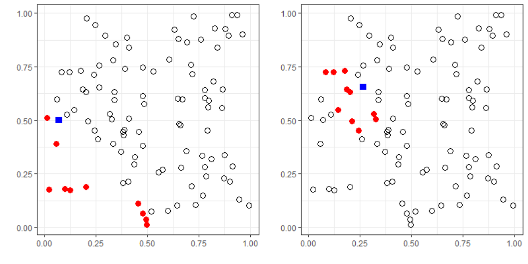
- Ideally, we want a pipeline like this:



- θ_2 is low dimensional and $\mathbf{u}$ is high dimensional, so this isn't straightforward

- We approximate and estimate the (surrogate) likelihood using a Vecchia approximated density regression (VADR) approach

Vecchia approximation



Use a Vecchia approximation[2] to approximate the joint likelihood as:

$$f_u(u_1, \dots, u_n|\boldsymbol{\theta}_2) = \prod_{i=1}^{n} f_i(u_i|\boldsymbol{\theta}_2, u_1, \dots, u_{i-1}) \approx \prod_{i=1}^{n} f_i(u_i|\boldsymbol{\theta}_2, u_{(i)}), \qquad (2)$$

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

[2] see, e.g., *Vecchia (1988)*, and *Stein, Chi, and Welty (2004)*

Density regression

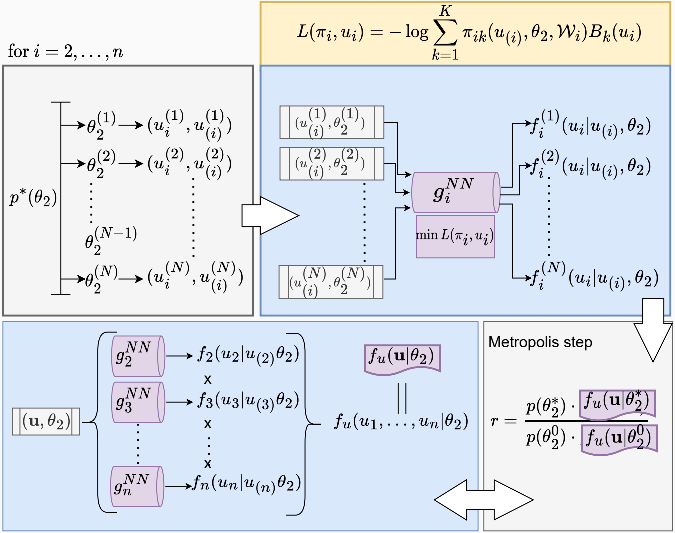
- Obtain density estimates of each term $f_i(u_i|\boldsymbol{\theta}_2, u_{(i)})$ using a semi-parametric quantile regression (SPQR) model (*Xu and Reich, 2021*):

$$f_i(u_i|\mathbf{x}_i, \mathcal{W}) = \sum_{k=1}^{K} \pi_{ik}(\mathbf{x}_i, \mathcal{W}_i) B_k(u_i), \tag{3}$$

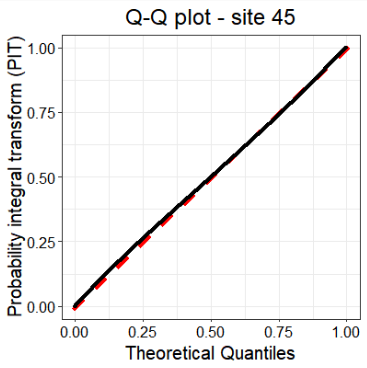
$$\pi_{ik}(\mathbf{x}_i, \mathcal{W}_i) = g_i^{NN}(\mathbf{x}_i, \mathcal{W}_i), \text{ for } i = 2, \ldots, n. \tag{4}$$

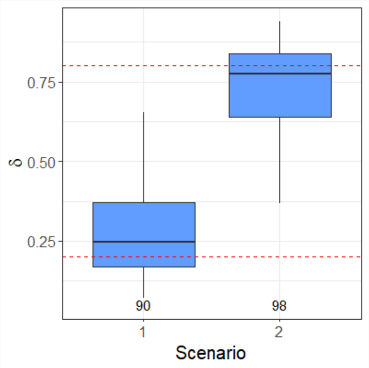
- π_{ik} are probabilities, $\sum_k \pi_{ik} = 1$
- $B_k(u_i)$ are M-spline basis functions, each of which is a PDF

The VADR workflow



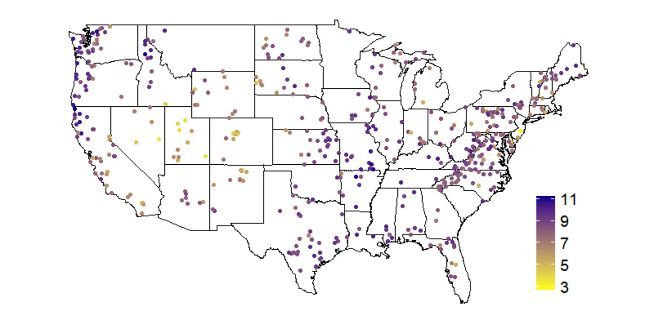
Simulation studies





- Simulation studies for a GP as well as for a PMM
- 2 hidden layers in the NN, 15 output knots
- MCMC for parameter estimation with Metropolis updates

Case study: extreme streamflow data in the US



- HCDN has 487 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 [3]

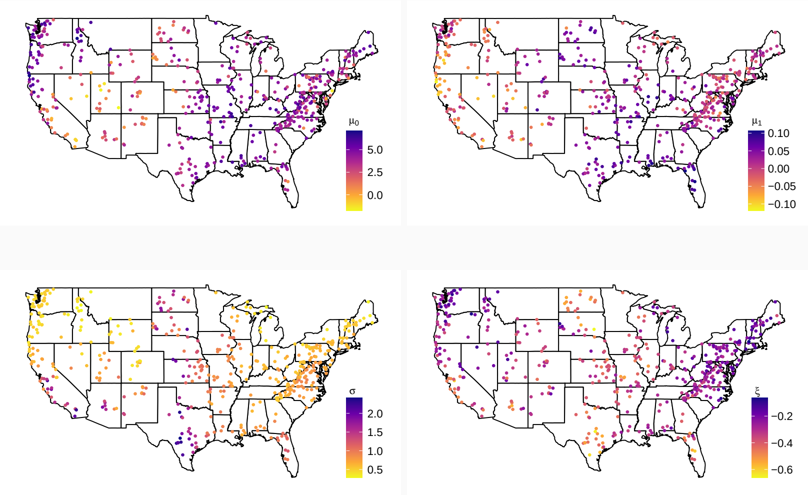
$$Y_t(\mathbf{s}) \sim \text{GEV}\left\{\mu_0(\mathbf{s}) + \mu_1(\mathbf{s}) \cdot X_t, \sigma(\mathbf{s}), \xi(\mathbf{s})\right\}, \tag{5}$$

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

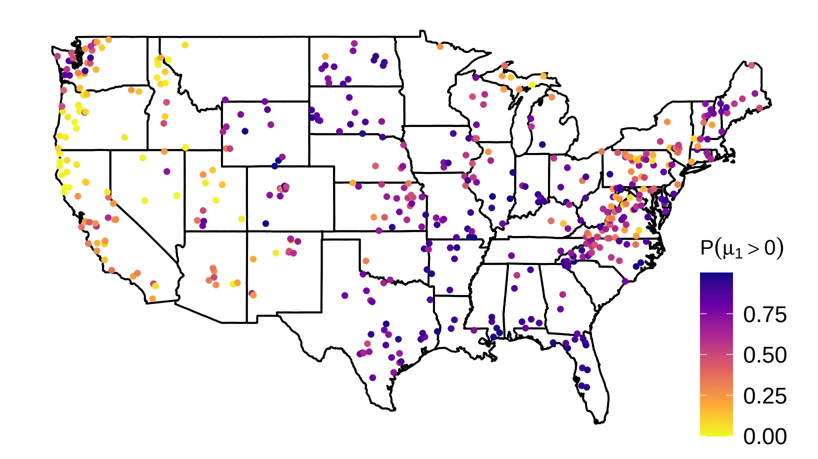
- The trend parameter $\mu_1(\mathbf{s})$ captures change per decade at site $\mathbf{s}$
- Given GEV parameters, obs. are independent over time but dependent in space

[3] see e.g., Gelfand et al., 2003

Posterior mean of the GEV coefficients



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 $\hat{\rho}_{\xi} = 20320\,(11481)$

Joint exceedance probabilities in CO

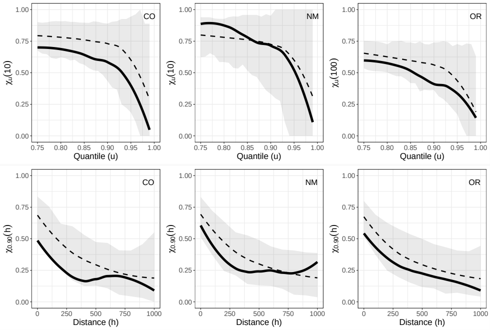


Figure 2: Empirical estimates of the upper tail coefficient in bold, compared against estimates based on the posterior distribution.

- We compute the joint posterior probability that all CO sites exceed their observed 0.90 quantile
- The joint exceedance probability is 0.075 (0.040) in 1972 and 0.170 (0.046) in 2021
- Probability that the joint exceedance in 2021 is higher than in 1972 is 0.90

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 (PMM), tractable (Vecchia approximation), can take advantage of GPU acceleration (SPQR)
- Strengths:
  - Main idea of VADR can be applied to virtually any spatial process
  - Compatible with a wide array of classical inference methods
  - A form of explainable AI
- Weaknesses:
  - Predictions at new locations is... complicated
  - The best SPQR fits don't always translate to the best parameter estimates

Thousands of U.S. homes have flooded over and over again. Here's where.

'What we are seeing is flooding is increasing faster than we are mitigating our risk,' one analyst says of data from the National Flood Insurance Program

CLIMATE

America underwater

Extreme floods expose the flaws in FEMA's risk maps

By Samuel Oakford, John Muyskens, Sarah Cahlan and Joyce Sohyun Lee

Dec. 6 at 11:59 a.m.

Source: Washington Post

References

- **R. Majumder**, B. J. Reich, and B. A. Shaby (2024). Modeling extremal streamflow using deep learning approximations and a flexible spatial process. *Annals of Applied Statistics*, 18(2): 1519–1542.
- **R. Majumder** and B. J. Reich (2023). A deep learning synthetic likelihood approximation of a non-stationary spatial model for extreme streamflow forecasting. *Spatial Statistics*, 55:100755.
- S.G. Xu and **R. Majumder** (2022). SPQR: Semi-Parametric Quantile Regression. R package version 0.1.0. `https://github.com/stevengxu/SPQR`

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

Appendix

Asymptotic properties of the PMM

- Extremal spatial dependence can be studied for pairs of variables by:

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

where $u \in (0, 1)$ is a threshold.

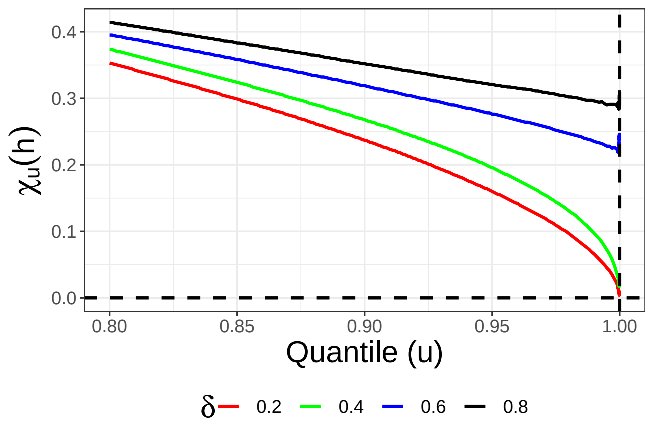
- If the process is isotropic, $\chi_u(\mathbf{s}_1, \mathbf{s}_2) \equiv \chi_u(h)$ for $h = ||\mathbf{s}_1 - \mathbf{s}_2||$

- $U(\mathbf{s}_1)$ and $U(\mathbf{s}_2)$ are defined to be asymptotically dependent (AD) if

$$\chi(h) = \lim_{u \to 1} \chi_u(h) \tag{7}$$

is positive, and independent (AI) if $\chi(h) = 0$

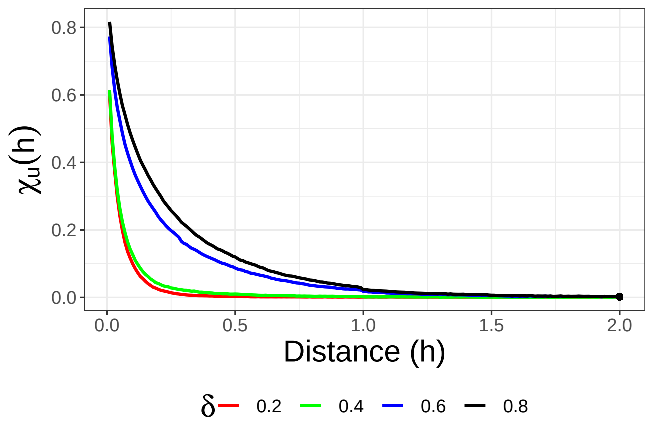
Asymptotic dependence: $\lim_{u \to 1} \chi_u(h) > 0$ iff $\delta > 0.5$



For any fixed h, the PMM is AD iff $\delta > 0.5$, because

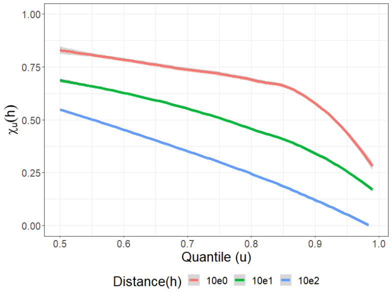
$$V(\mathbf{s}) = \delta R(\mathbf{s}) + (1 - \delta) W(\mathbf{s})$$

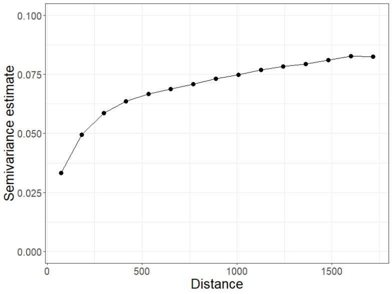
Spatial dependence: $\lim_{h\to\infty} \chi_u(h) = 0$ for any δ



Unlike the common R model of HW2019, the PMM model has long-range independence as $h \to \infty$

Exploratory data analysis - spatial dependence





Empirical conditional exceedance, $\chi_u(h)$, and variogram of MLE residuals

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 $R(\mathbf{s}) \equiv R$)
- MSP is the max-stable process (PMM with $\delta = 1$)
- GP is the Gaussian process (PMM with $\delta = 0$)

Simulation studies

- The first simulation study uses Gaussian data
- The full conditional distributions have a closed form so we can asses the fit of the approximate density
- Using Gaussian data also allows us to separate the effects of the Vecchia and deep learning approximations
- We also compare local and global[4] SPQR fits
- The second simulation uses data simulated from the PMM

[4] The same SPQR fit to all sites

Simulation Study - GP

The simulated data are Gaussian with

- Marginal distribution

$$Y_t(\mathbf{s}) \sim \text{Normal}(\mu, \sigma^2)$$

- Spatial correlation

$$\text{Cor}\{Y_t(\mathbf{s}_i), Y_t(\mathbf{s}_j)\} = r\exp(-||\mathbf{s}_i - \mathbf{s}_j||/\rho)$$

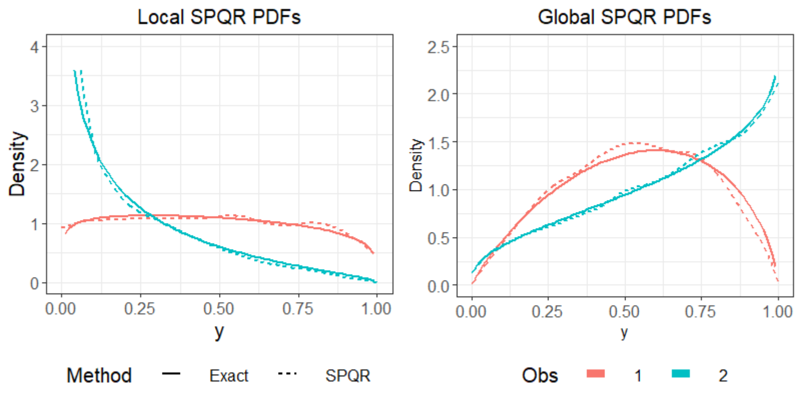
- The parameters are $\boldsymbol{\theta}_1 = (\mu, \sigma^2)$ and $\boldsymbol{\theta}_2 = (\delta, r)$
- Data are simulated at 100 location in $[0, 1]^2$ for 5 independent time replications
- The Vecchia approximation uses 10 neighbors and we select uninformative priors

Simulation study - GP

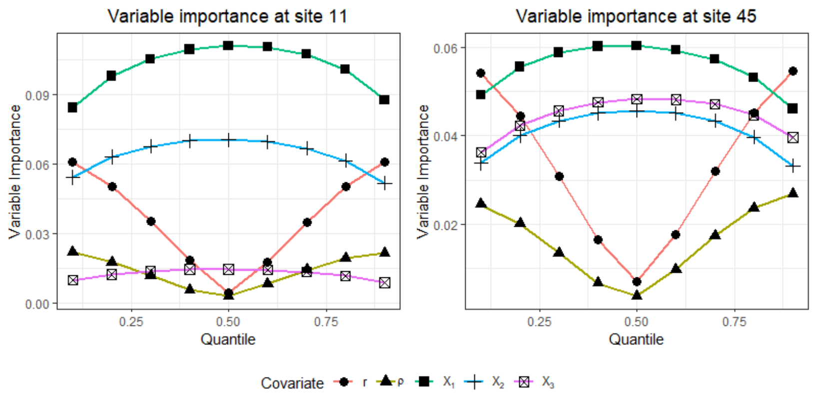
Table 1: Design distribution p^* (top) and FFNN tuning parameters (bottom).

Hyperparameter	Global SPQR	Local SPQR
r	Uniform$(0, 1)$	Uniform$(0, 1)$
ρ	Uniform$(0.1, 2)$	Uniform$(0.1, 1.23)$
Number of features	31	12
Hidden layers	3	2
Output knots	20	10
Learning Rate	0.001	0.01
Batch size	100	1000
Epochs	20	20
Observations	10^8	10^6

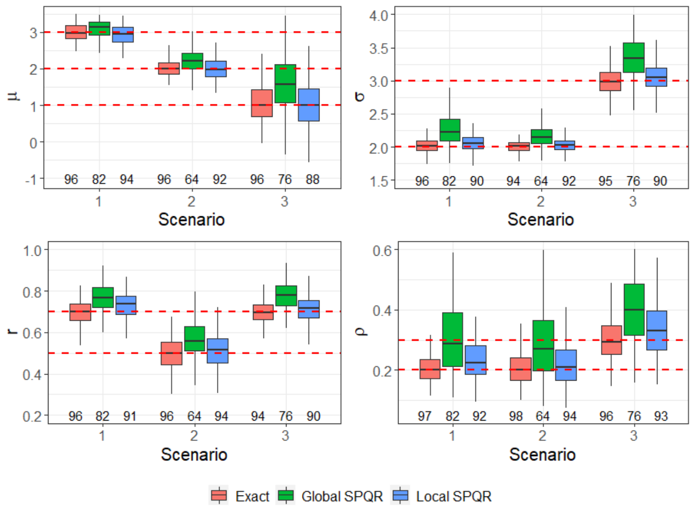
Models fitted using the *SPQR* package in *R*



True and estimated PDFs for two out-of-sample observations fitted using local and global SPQR.


Variable importance at site 11
Variable importance at site 45
Variable Importance
Quantile
0.00
0.03
0.06
0.09
0.02
0.04
0.06
0.25
0.50
0.75
Covariate r ρ X1 X2 X3

SPQR model fit diagnostics - GP

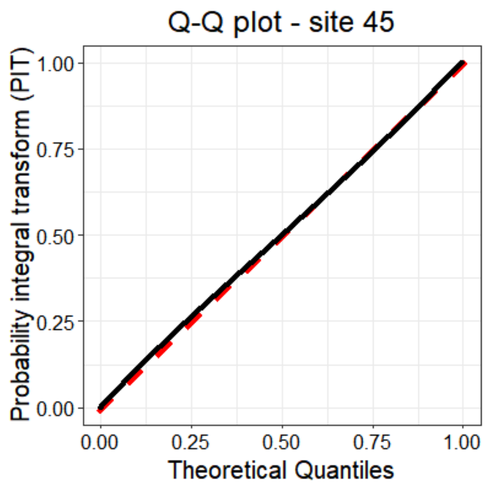


Local SPQR has less bias and lower variability than global SPQR

Simulation Study - Process Mixture Model

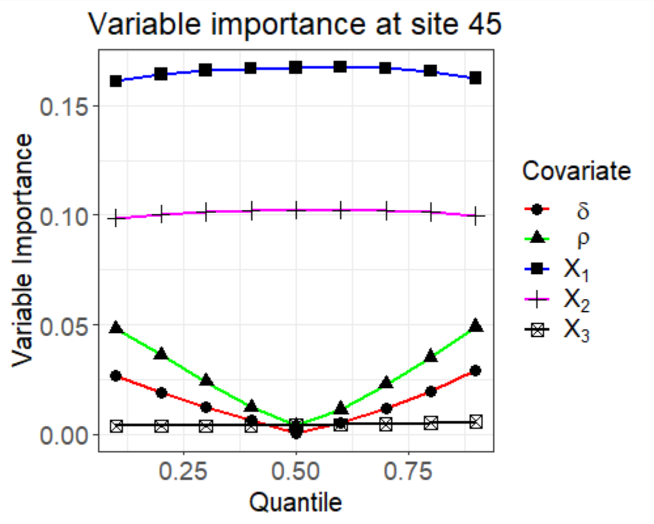
- Marginal distribution is $Y_t(\mathbf{s}) \sim \text{GEV}\{\mu_0(\mathbf{s}) + \mu_1(\mathbf{s}) \cdot t, \sigma(\mathbf{s}), \xi(\mathbf{s})\}$
- GEV parameters set to be smooth functions of space
- R_t and W_t both have powered exponential dependence functions with power set to one
- Their effective 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.

SPQR model fit diagnostics - PMM

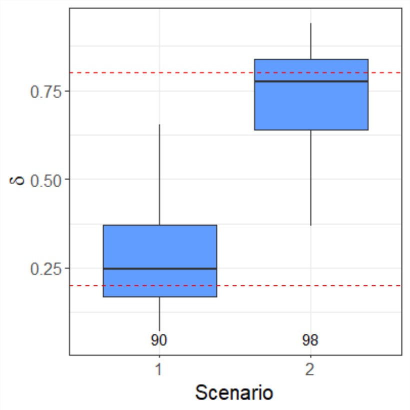


The Q-Q plot shows the SPQR model fits well

SPQR model fit diagnostics - PMM



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.