Widespread floods, do they occur everywhere?

Applied Bayesian Analysis (ST 540)

Sneha Rani
Jinge Xu
Junjian Wang
Kichul Bae

Project Presentation Outline

1. Introduction
2. Data collection & Data pre-processing

Research Question 1 models:

3. Model 1: The Bayesian hierarchical model with random intercepts and random slopes
4. Model 2: Linear Regression with uninformative priors and fixed slopes

Research Question 2 models:

5. Model 3: Linear Regression for Intercept Analysis with Multiple Covariates
6. Model 4: Variable selection with SSVS
7. Findings, Conclusions & Implications
8. Limitations and further research
9. References

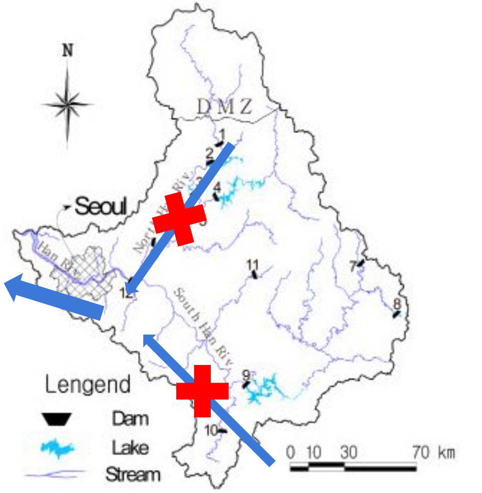
Introduction: Research Questions and Motivation

[Research questions]

1. Are floods spatially correlated, and do they differ between regions?
2. If so, which factor would drive this difference?

[Motivation]

We can avoid over- and under-estimate the risk of floods if we can properly understand their spatial correlation.



Lee, Sangho & Kang, Taeuk & Lee, Kil. (2015).

Kichul Bae

Data collection & Data preprocessing

[Stream flow data]

Annual peak flow of 702 stations from HCDN-2009

year	A	B
1910	9	9
1911	2	1
1912	5	7
1913	4	3
1914	7	2
….	…	…
2018	5	2
2019	4	8
2020	3	6
2021	1	2
2022	2	7

[Joint Probability]

- Measures how jointly the flood occurred for various magnitude

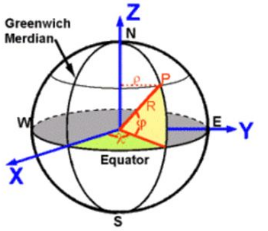
$$P_{RP}(A,B) = \frac{x}{n}$$

$n = total\ number\ of\ years$

$the\ flood\ simultaneously$

Kichul Bae

Data collection & Data preprocessing



[Distance between stations]

Distance between two points on a spheroid

- (Latitude(ϕ), Longitude(λ)) to (X, Y, Z)
- $X=R\cos(\phi)\cos(\lambda)$, $Y=R\cos(\phi)\sin(\lambda)$, $Z=R\sin(\phi)$, R = radius of the earth (6371 km)
- Distance = $\sqrt{(\Delta X^2+\Delta Y^2+\Delta Z^2)}$

[Hydrologic characteristics of HUC2 regions]

Annual mean precipitation, daily maximum precipitation, elevation(mean),

annual mean streamflow, number of basins, mean area of basins,

temperature(min, max), vapor pressure

* Bakhshaee, Alireza. (2018). The CAMELS dataset: catchment attributes and meteorology for large sample studies 1980 -2014.

Sneha Rani

Model 1: The Bayesian hierarchical model with random intercepts and random slopes

RQ1: Are floods spatially correlated, and do they differ between regions?

Model

Likelihood: $Y_i = N(\mu_i + 0.25, \tau e_i)$

where $\mu = \text{expit}(\beta_0 j + \beta_1 j * X)$

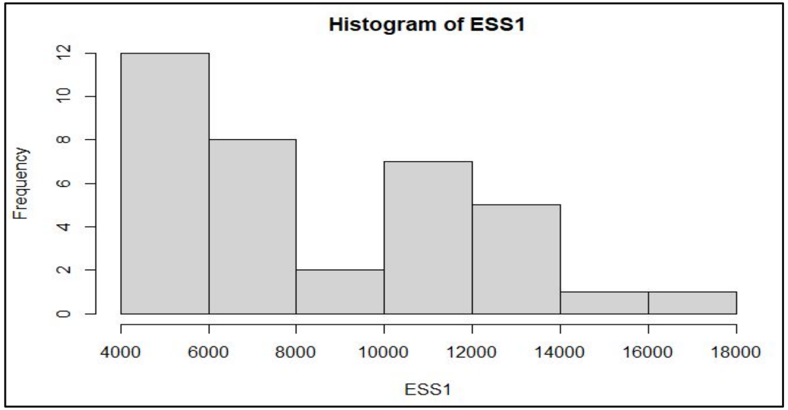
Priors: $\beta j \sim N(\mu b, \tau b)$, $\mu b \sim N(0, 0.01)$; $\tau b \sim Gamma(0.01, 0.01)$; $\tau e \sim Gamma(0.1, 0.1)$

DIC:

Mean deviance: -7253
Penalty: 40.7
Penalized deviance: -7213

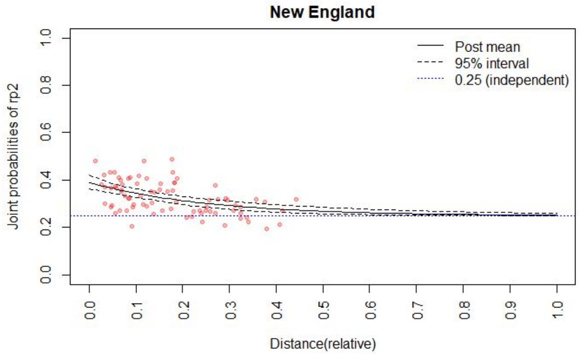
Convergence diagnostics:

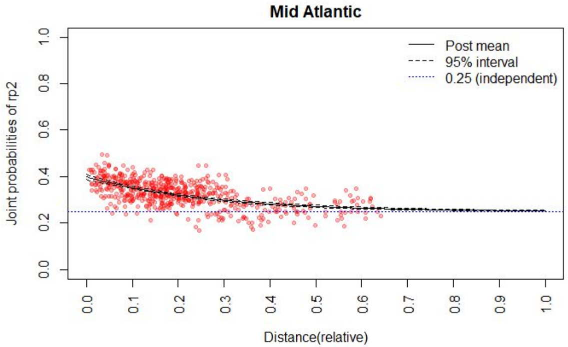
max(Gelman-Rubin) = 1.01 (Good convergence among all Betas shown by trace plots as well)

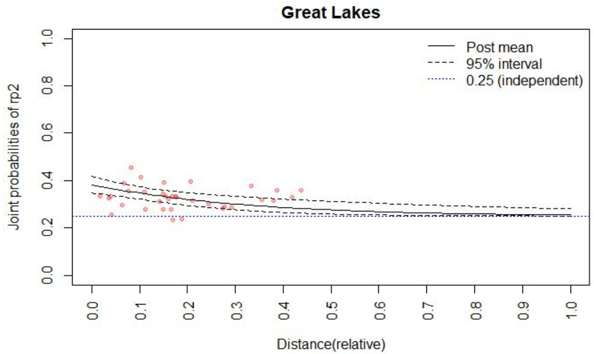


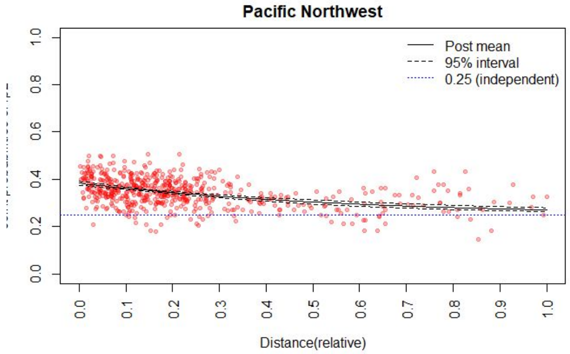
Sneha Rani

Model 1 fitting (4 out of total 18):









Findings:

At small distance, joint probability of flood RP2 is high.

Different joint probabilities of flood RP2 across 18 HUC2 regions

Jinge Xu

Model 2: Linear Regression with uninformative priors and fixed slopes

Model

Likelihood:

$Y_i \sim N(\mu_i + 0.25, \tau e_i)$

$\mu_i = \mathrm{expit}(\beta_{0,i} + \beta_{1,i} * X_i)$

Priors:

$\beta_i \sim N(0, 0.01)$

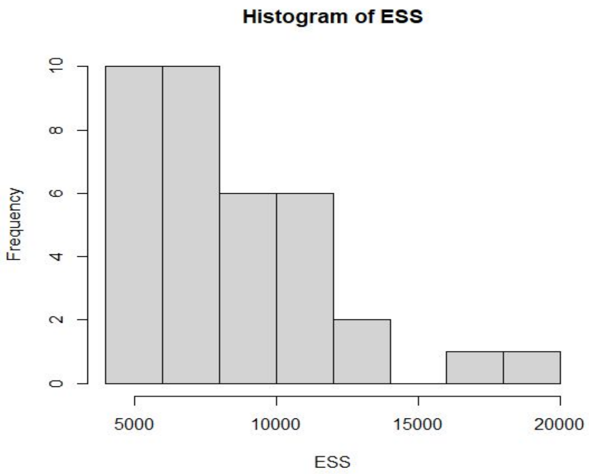
$\tau e \sim \mathrm{Gamma}(0.1, 0.1)$

DIC:

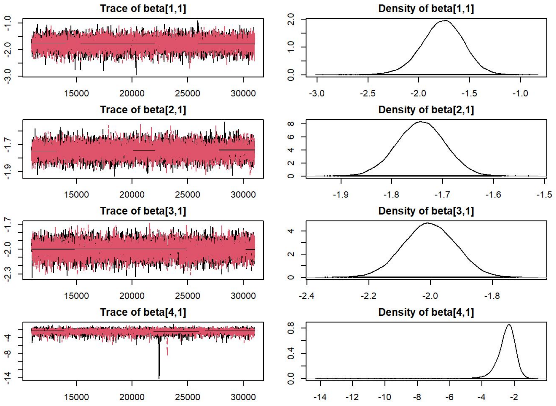
Mean deviance: -7250
Penalty: 55.01
Penalized deviance: -7195

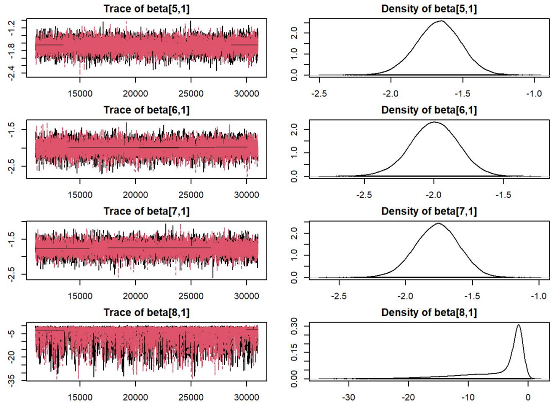
Convergence diagnostics:

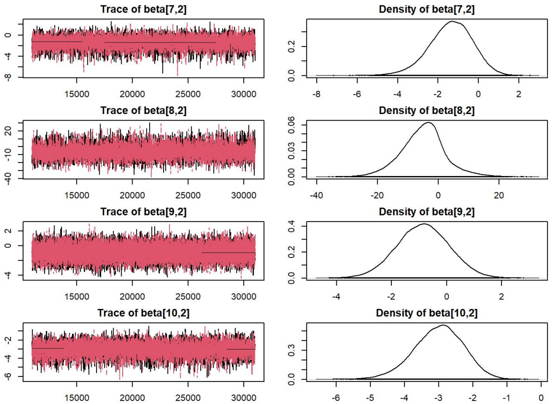
max(Gelman-Rubin) = 1.05
Multivariate psrf = 1.01

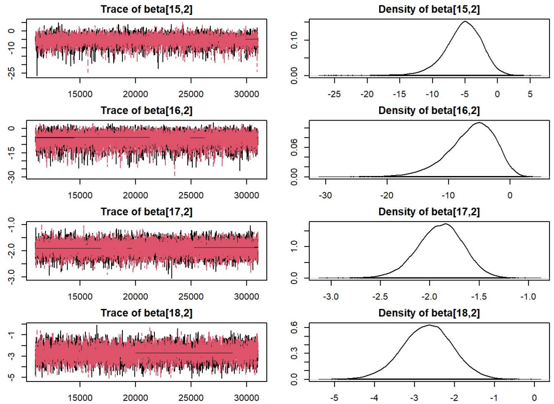


Convergence



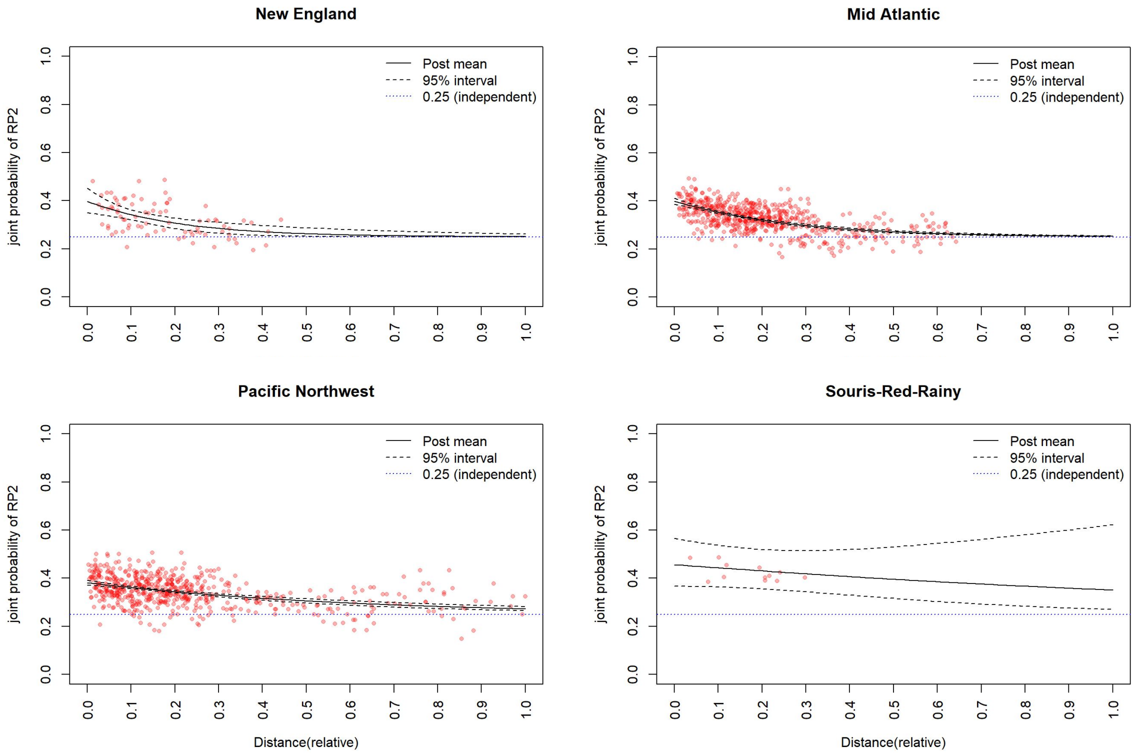






Jinge Xu

Fitting



Junjian Wang

Model 3: Lasso penalized Regression for Intercept Analysis with Multiple Covariates

Model

$Y_i \sim N(\mu_i, \tau e_i)$

$\mu_i = \beta_0 + \sum \beta_j X_{ij}$

where $\beta_0 \sim N(0, 0.01)$,

$\beta_j \sim DE(0, \tau e * \tau b)$

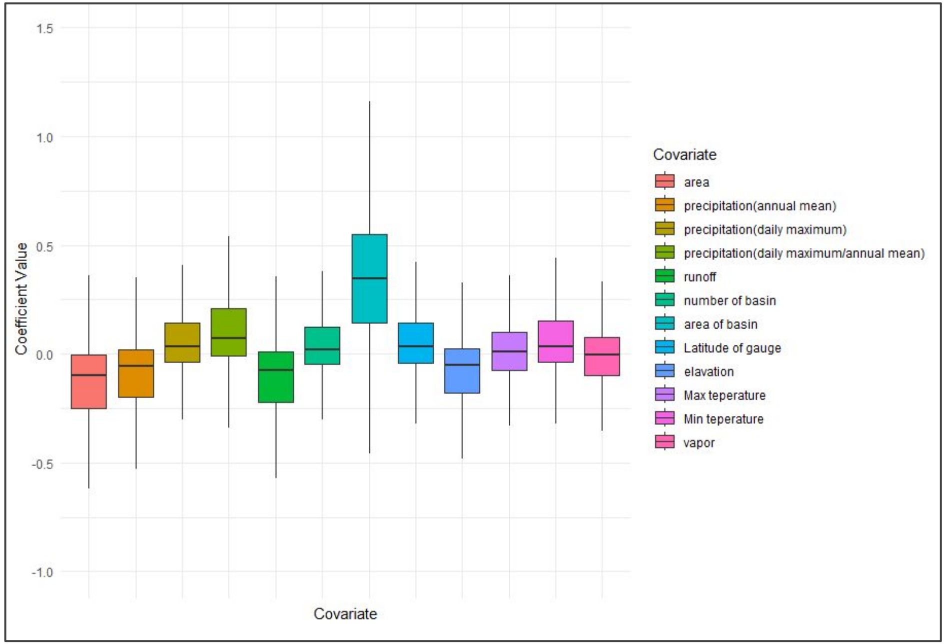
Priors: $\tau e, \tau b \sim Gamma(0.1, 0.1)$

DIC:

Mean deviance: 43.74
Penalty: 8.249
Penalized deviance: 51.99

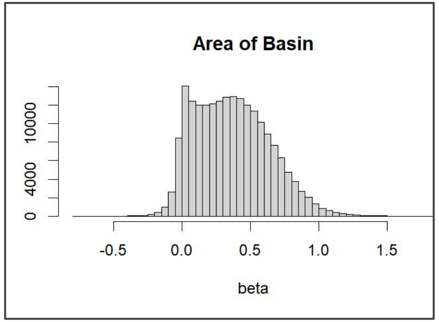
Convergence diagnostics:

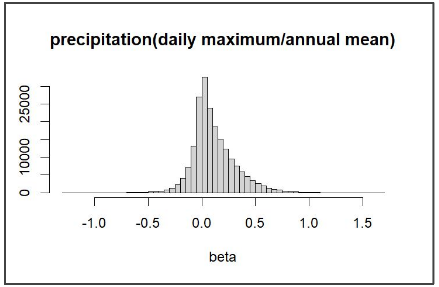
max(Gelman-Rubin) = 1
min(ESS) >1000



- Beta values boxplot for Model 4. (Intercept excluded)

- Posterior distribution





Junjian Wang

Model 4: Linear Regression with Stochastic search variable selection

Model

$Y_i \sim N(\mu_i, \tau e_i)$

$\mu_i = \beta_0 + \sum \beta_j X_{ij}$

where $\beta_j = \gamma_j \delta_j$.

Priors:

$\delta_j \sim N(0, \tau)$ $\gamma_j \sim$ Bernoulli(0.5)

$\tau, \tau e \sim$ Gamma(0.1, 0.1)

DIC:

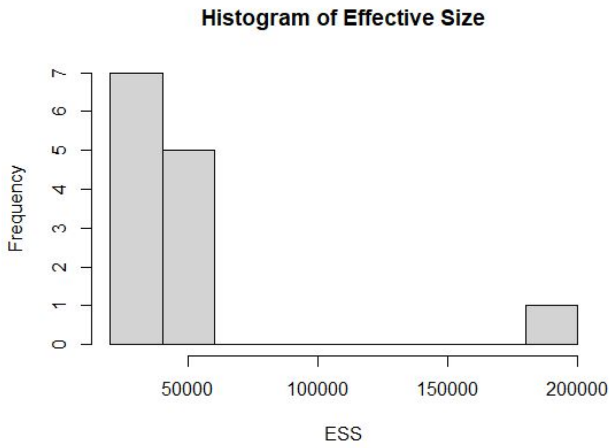
Mean deviance: 43.13

Penalty: 8.572

Penalized deviance: 51.68

Convergence diagnostics:

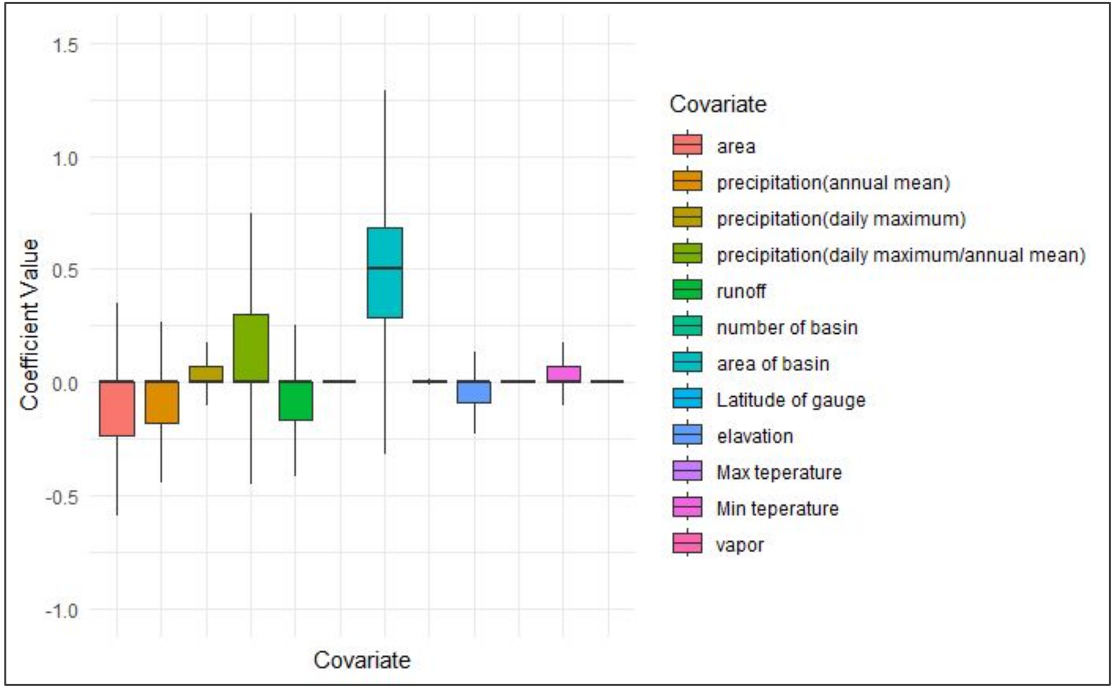
max(Gelman-Rubin) = 1



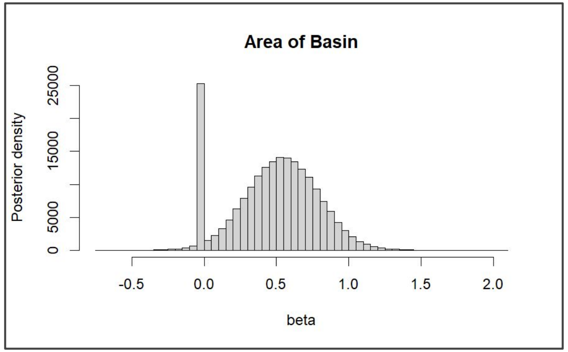
Junjian Wang

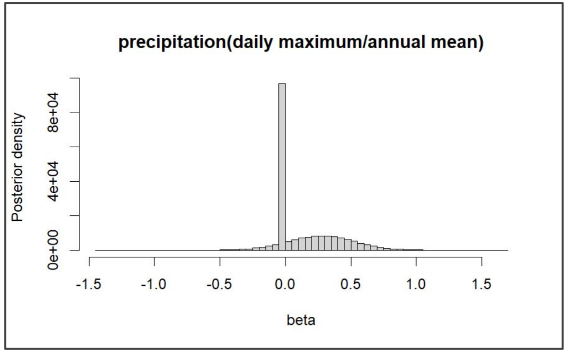
Model 4 Plots

- Beta values boxplot for Model 4. (Intercept excluded)



- Posterior distribution for the SSVS analysis





Findings & Conclusions

[Answers to our research questions]

1. **Are floods spatially correlated, and do they differ between regions?**
   - The closer, the strongly correlated floods are
   - For some regions, joint probability is significantly different from other regions

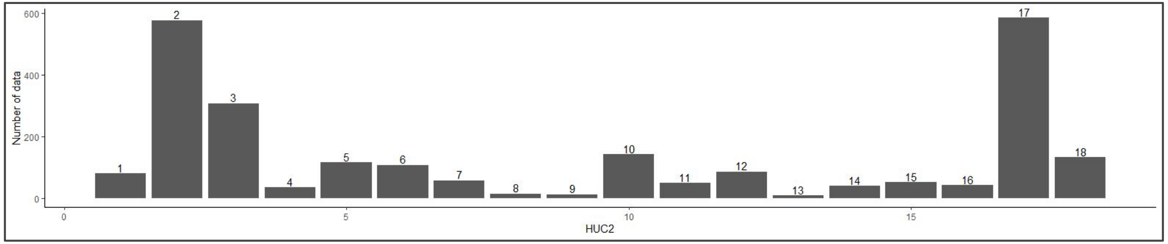
2. **If so, which factor would drive this difference?**
   - Larger the mean area of basin is, higher the joint probability is

Implications of this study: The findings can inform the development of targeted flood risk management strategies that account for the unique characteristics and challenges of different regions, ultimately contributing to more effective flood prevention and mitigation efforts.

Limitations & further research

[Limitations]

Imbalance of quantity of data among HUC2 regions



[Future research]

1. What will happen if we set the different threshold other than 2-year return period
2. Is there a trend over time due to climate change?

References

- Brunner, M. I., Gilleland, E., Wood, A., Swain, D. L., & Clark, M. (2020). Spatial dependence of floods shaped by spatiotemporal variations in meteorological and land-surface processes. *Geophysical Research Letters*, 47, e2020GL088000. https://doi.org/10.1029/2020GL088000
- Brunner, M. I., Papalexiou, S., Clark, M. P., & Gilleland, E. (2020). How probable is widespread flooding in the United States? Water Resources Research, 56, e2020WR028096. https://doi.org/10.1029/2020WR028096
- Majumder, Reetam & Reich, Brian & Shaby, Benjamin. (2022). Modeling Extremal Streamflow using Deep Learning Approximations and a Flexible Spatial Process. 10.48550/arXiv.2208.03344.
- Sun, A. Y., Scanlon, B. R., Save, H., & Rateb, A. (2021). Reconstruction of GRACE total water storage through automated machine learning. *Water Resources Research*, 57, e2020WR028666. https://doi.org/10.1029/2020WR028666
- Bakhshaee, Alireza. (2018). The CAMELS dataset: catchment attributes and meteorology for large sample studies 1980 -2014.