Regional Variability in the Performance of **Annual Maxima** (**AM**) vs. **Peaks-Over-Threshold** (**POT**) Methods for Predicting Frequent Floods

Francesco Dell'Aira[1] **(fdllaira@memphis.edu)**,

Antonio Cancelliere[2],

and **Claudio I. Meier**[1]

[1]: University of Memphis, USA. Department of Civil Engineering.

[2]: University of Catania, Italy. Department of Civil Engineering and Architecture.

Introduction

- In the United States, flood frequency analyses performed by government agencies (e.g., Southard 2010; Virginia Department of Transportation 2021) are typically carried out using annual maxima time series, which are available at more sites and for longer time periods than continuous time series (required for peak-over-threshold (POT) analysis).
- The distribution fitted to the annual maxima (AM) observed for a given watershed is used to predict floods for that basin for a range of probabilities of exceedance, including both frequent (e.g., with 2-year return period) and exceptional (e.g., 100-year return period) floods (Southard 2010; Virginia Department of Transportation 2021).
- However, POT is recommended for frequent flood estimation, because low (or zero) peaks occurring in dry years may have an undue influence on the shape of the distribution fitted on AM (Ball et al. 2016). In particular, the prediction of frequent floods is affected, since they tend to be underestimated.
- Frequent floods are of particular interest in frameworks such as river geomorphology (Hu et al. 2017).

This presentation participates in OSPP

EGU

Outstanding Student & PhD candidate Presentation contest

Introduction

- In this work, we show that the degree of underestimation of frequent floods resulting from using AM, as compared to POT, presents **geographical variability**, with a quite clear spatial pattern and smooth transition across regions from high to low degree of underestimation.
- We found that it is possible to **predict** the degree of underestimation (given as the ratio of AM to POT floods) from climatic indices. Marginal improvements in the predictions can be achieved by also considering basin characteristics.

Data

- We considered a subset of basins from the **CAMELS dataset** (Addor et al. 2017).
- **CAMELS dataset** provides a wide range of information for 671 basins in the contiguous United States with minimal human impact, including **topographic characteristics**, **climatic indices**, **hydrological signatures**, **land-cover** and **soil characteristics**.
- **CAMELS dataset** provides **daily precipitation time series** for all its watersheds, but hourly or 15-minutes time series are preferable to perform Annual Maxima (AM) and Peak-over-threshold (POT) flood frequency analyses (in order to consider instantaneous peaks), therefore we selected the subset of CAMELS basins for which **continuous flow observations** from the U.S. Geological Survey (USGS) are available for at least 25 years.
- The resulting subset is made of **497 basins.**

Methodology

- We performed **Peak-over-threshold** (**POT**) and **annual maxima** (**AM**) frequency analyses on each flow time series observed at each basin. More details are given below.
- We fit the **Generalized Pareto** (**GP**) (Solari et al. 2017) and **Generalized Extreme Value** (**GEV**) (Bezak et al. 2014) distributions to the peaks over threshold and annual maxima, respectively, using L-moments method (Hosking 1990). L-moments, in contrast with other popular distribution-fitting techniques (e.g., maximum likelihood and method of moments) is computationally cheap (Solari et al. 2017), this being an advantage in this context, where hundreds of distributions were fitted on as many time series.
- Known the GEV and GP distributions, we predicted **frequent floods** (i.e., with small **average recurrence intervals** (**ARIs**)) from both AM and POT, respectively.
- We used the **ratio** of AM to POT floods (Q_{AM}/Q_{POT}) as **indicator of underestimation,** for a given average recurrence interval (ARI). The smaller the ratio, the greater the degree of underestimation is; when the ratio is close to 1, the two approaches tend to provide the same results.

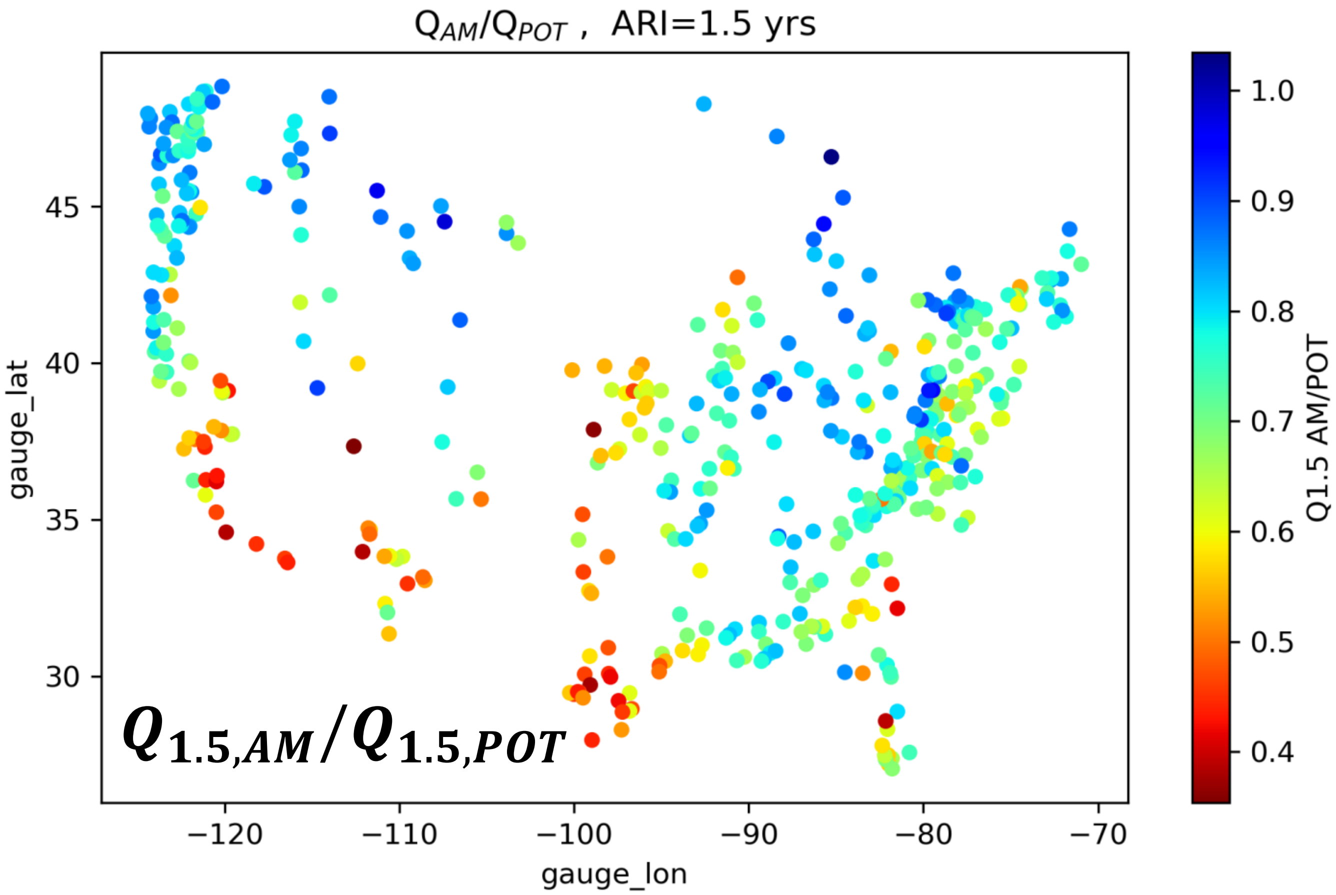
Methodology: operative details for POT

- In POT analysis, we chose the **threshold** u_{opt} for each flow time series following the approach proposed by Solari et al. (2017), based on the optimization of the goodness of fit p-value related to the fitted GP distribution.

- We also studied the **sensitivity** to **threshold selection**, by considering other thresholds – u_{2N}, u_{3N}, and u_{4N}, for which $2N$, $3N$, and $4N$ peaks were extracted from the time series, respectively, where N is the number of years spanned by the time series. We found that floods estimated from the GP distributions with threshold u_{2N}, u_{3N}, u_{4N}, and u_{opt} differ from each other for only a few percentage points, indicating very low sensitivity to the threshold selection criterion and great stability of the fitted distribution with respect to varying thresholds for extracting peaks from the continuous hydrograph.

- **Independent peaks** in the hydrographs were identified by a conservative criterion, that combines two literature methods mentioned in Pan et al. (2022): two consecutive peaks are independent if and only if they are separated in time by at least $\left(5+\sqrt{A}\right)$ days (A is the drainage area in square miles), and if the trough in between has dropped by at least 1/3 of the larger peak.

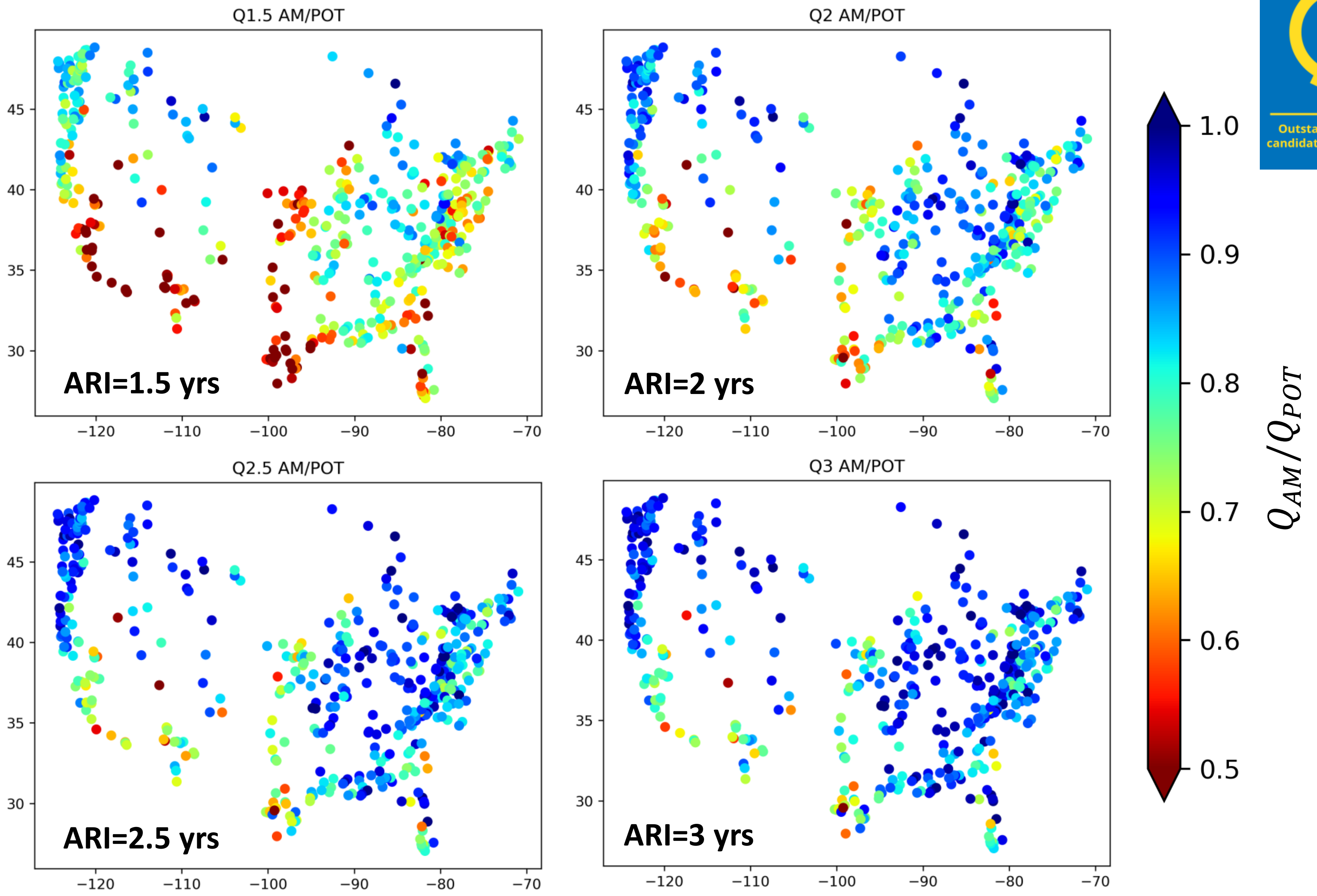
Results: geographical variability



- **Geographical variability** across the contiguous U.S. of Q_{AM}/Q_{POT} for an average recurrence interval (ARI) of 1.5 years.
- There is a **spatial pattern** in the way AM underestimates frequent peaks as compared to POT.

Results

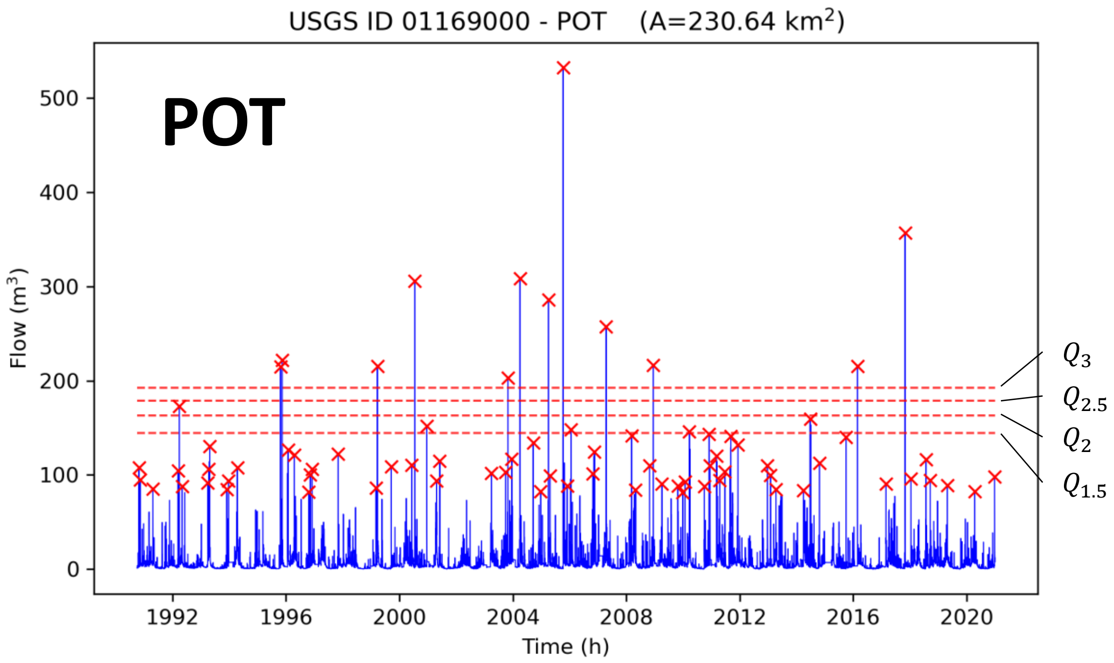
- For increasing ARIs, predictions from AM converge to those from POT, as expected, but the same spatial pattern is still visible.

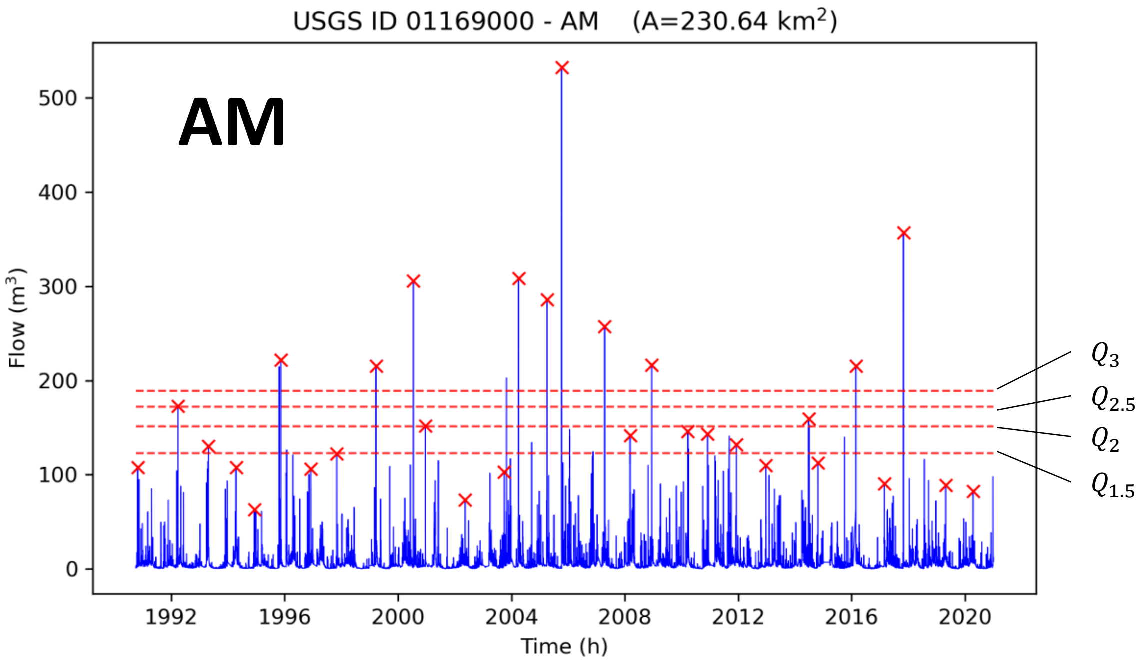


This presentation participates in OSPP

Outstanding Student & PhD candidate Presentation contest

Why AM and POT estimations differ





- **AM** tends to **underestimate** frequent peaks because it considers **low peaks** occurred in **dry years** and ignores other large peaks, besides large annual maxima, occurred in rainy years (Ball et al. 2016).

This presentation participates in OSPP

Outstanding Student & PhD candidate Presentation contest

Why AM and POT estimations differ

We speculate that **climate** is the **main responsible** for the variability in Q_{AM}/Q_{POT}, since it affects the occurrence of dry years and the frequency of large floods, and **basin effects** represent a **secondary factor**, by influencing the persistence of wet soil conditions after rainstorms and the damping of rainfall input, which in turn affect the magnitude and frequency of peaks.

- **Climate** effects (primary cause)
  - → Occurrence of **dry years** → Number of **small floods** considered by AM only
  - → Average **number of storms** per year → **Frequency** of floods ignored by AM
- **Basin** effects (secondary cause)
  - → Permanence of **wet soil conditions** after a storm → **Frequency** of floods ignored by AM; **Magnitude** of floods ignored by AM
  - → **Damping** and delaying of the rainfall input → **Magnitude** of floods ignored by AM

Predicting spatial variability of Q_{AM}/Q_{POT}

- To verify our hypothesis, we tried to predict the variability in the degree of underestimation of frequent floods, first from **climatic indices** only (first experiment), and then from **climatic indices** and **basin characteristics** (second experiment), using an artificial neural network (ANN) model.
- We chose the climatic indices and basin characteristics from the CAMELS dataset (Addor et al. 2017) that had an **influence** in the model performance and ignored those for which the model showed no sensitivity.

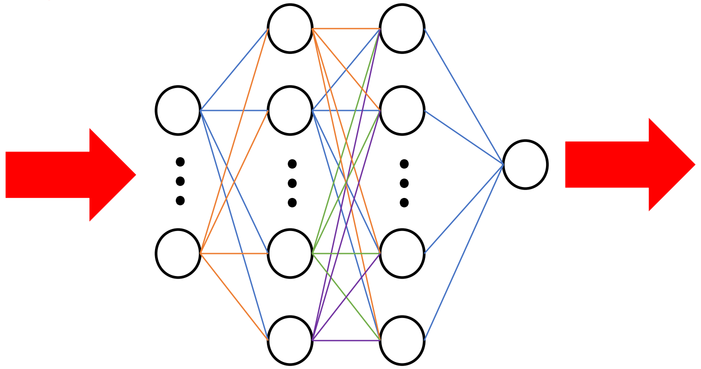
Predicting spatial variability of Q_{AM}/Q_{POT}

This presentation participates in OSPP

Outstanding Student & PhD candidate Presentation contest

- **Climatic indices** [1]:
  - ❑ **Frequency** of **high precipitation** days (days/yr)
  - ❑ Average **duration** of **high precipitation** days (days)
  - ❑ **Frequency** of **dry periods** (days/yr)
  - ❑ Average **duration** of **dry periods** (days)
  - ❑ Fraction of snow (-)
  - ❑ Aridity (PET/Prec.) (-)
  - ❑ Seasonality of precipitation (-)
- **Topographic characteristics** [1]
  - ❑ Area, Mean elevation, Mean slope
- **Hydrologic signatures** [1]
  - ❑ Mean Q, Stream-prec. Elasticity, Runoff ratio
- **Land Cover characteristics** [1]
  - ❑ Fraction of forest
- **Soil characteristics** [1]
  - ❑ Sand fraction, Silt fraction

1st experiment



Q_{AM}/Q_{POT} at the given basin

Artificial Neural Network (ANN)
(Multi-layer Perceptron, 2 hidden layers with 30 neurons each)

[1]: from Addor et al. 2017

This presentation participates in OSPP

Outstanding Student & PhD candidate Presentation contest

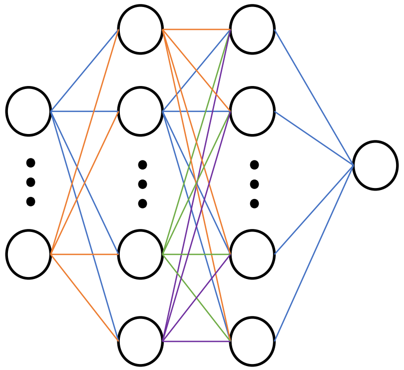
Predicting spatial variability of Q_{AM}/Q_{POT}

2nd experiment

- **Climatic indices** [1]:
  - ❑ **Frequency** of **high precipitation** days (days/yr)
  - ❑ Average **duration** of **high precipitation** days (days)
  - ❑ **Frequency** of **dry periods** (days/yr)
  - ❑ Average **duration** of **dry periods** (days)
  - ❑ Fraction of snow (-)
  - ❑ Aridity (PET/Prec.) (-)
  - ❑ Seasonality of precipitation (-)

Basin characteristics

- **Topographic characteristics** [1]
  - ❑ Area, Mean elevation, Mean slope
- **Hydrologic signatures** [1]
  - ❑ Mean Q, Stream-prec. Elasticity, Runoff ratio
- **Land Cover characteristics** [1]
  - ❑ Fraction of forest
- **Soil characteristics** [1]
  - ❑ Sand fraction, Silt fraction



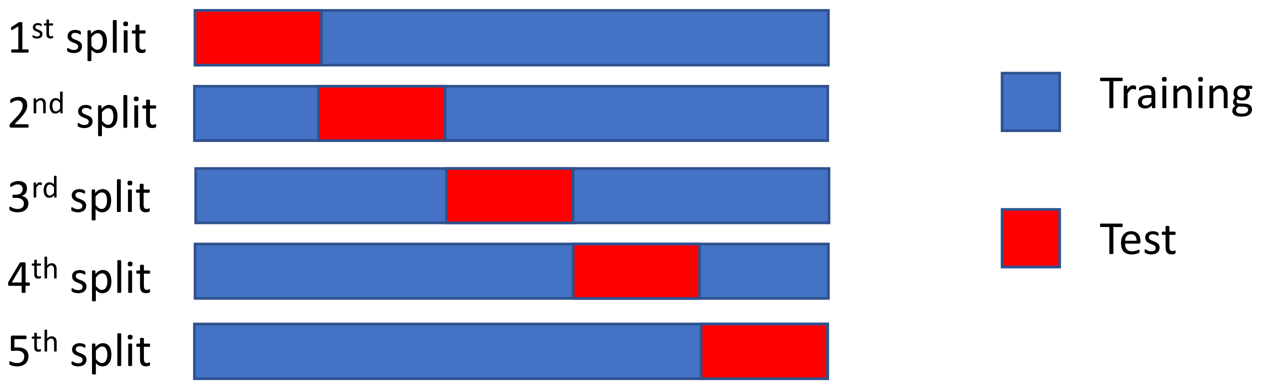
Q_{AM}/Q_{POT} at the given basin

Artificial Neural Network (ANN)
(Multi-layer Perceptron, 2 hidden layers with 30 neurons each)

[1]: from Addor et al. 2017

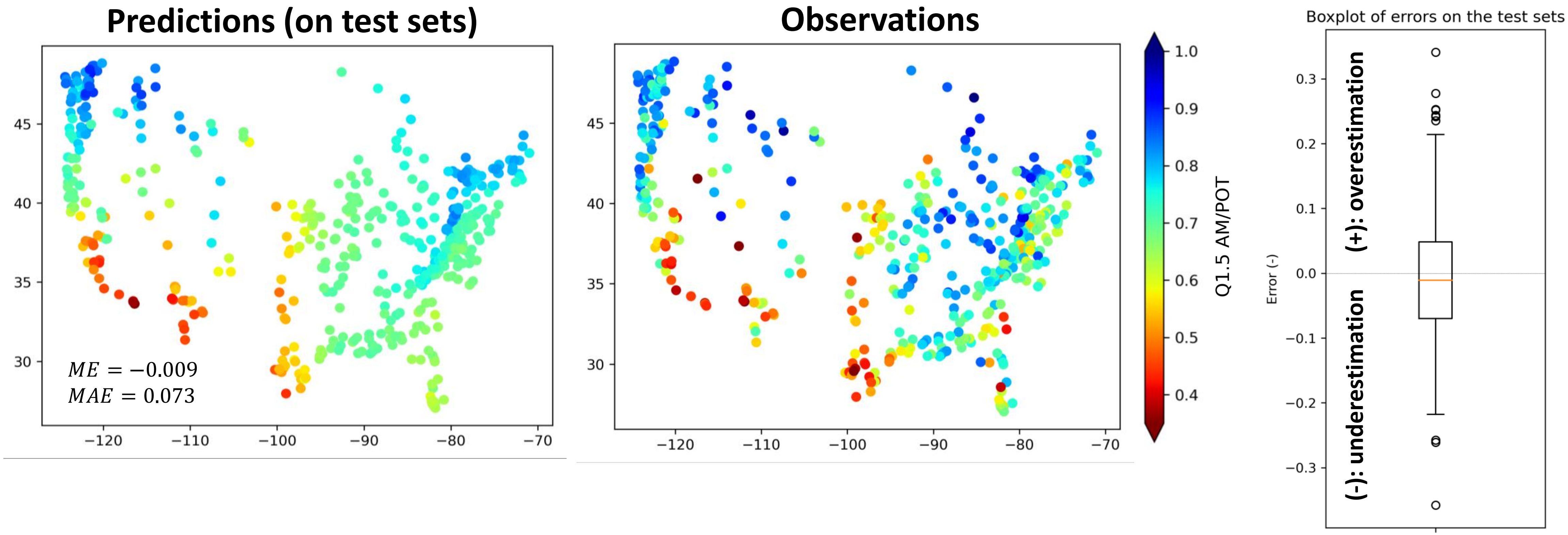
Training of ANN model

- We followed **K-fold** method for training and testing the network, considering, one at a time, 5 alternative splits into **training (80%)** and **test (20%)** sets of the full dataset. Therefore, we trained and tested the model 5 times, each time on the current training and test set, respectively.



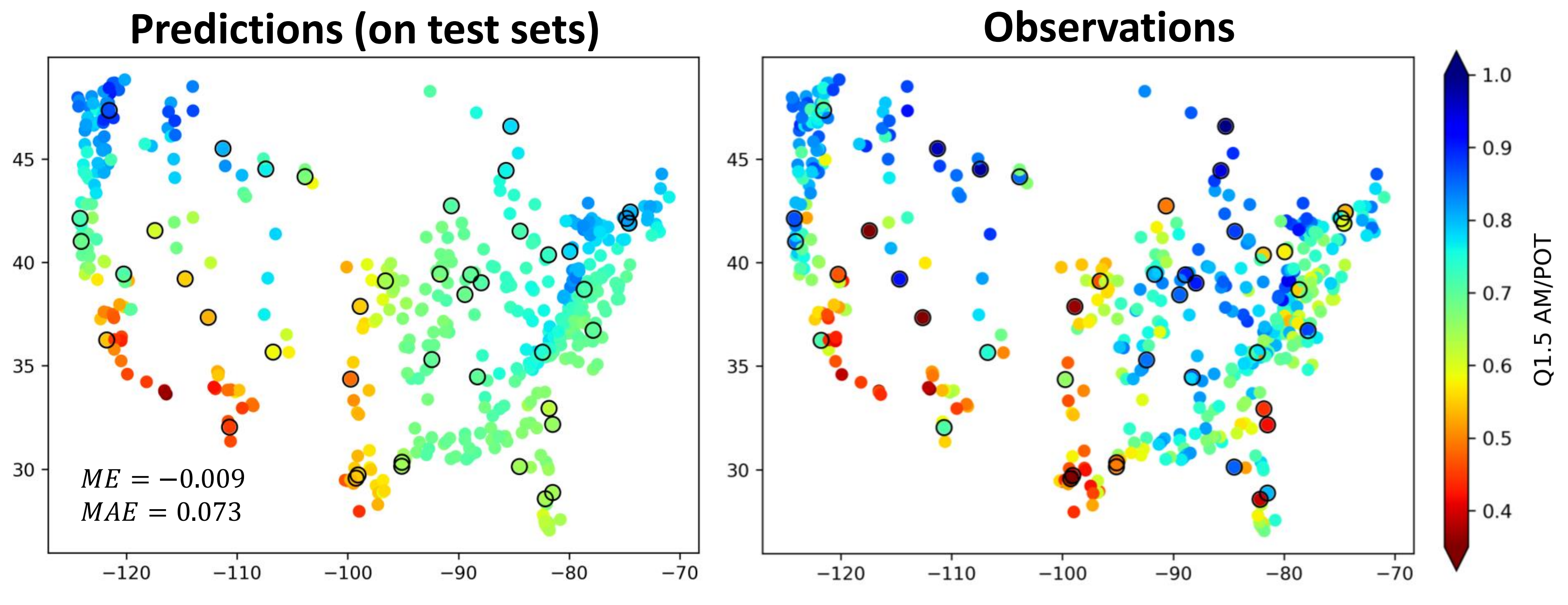
- This allowed to perform a **blind-test** over the **whole** dataset, obtained as the union of test sets from the five alternative splits.
- To account for the stochastic nature of the training process, each time we used an **ensemble** of 5 models, instead of just one model (Kratzert et al. 2019), by training 5 ANNs separately and **averaging out** their predictions to obtain the **ensemble predictions**.

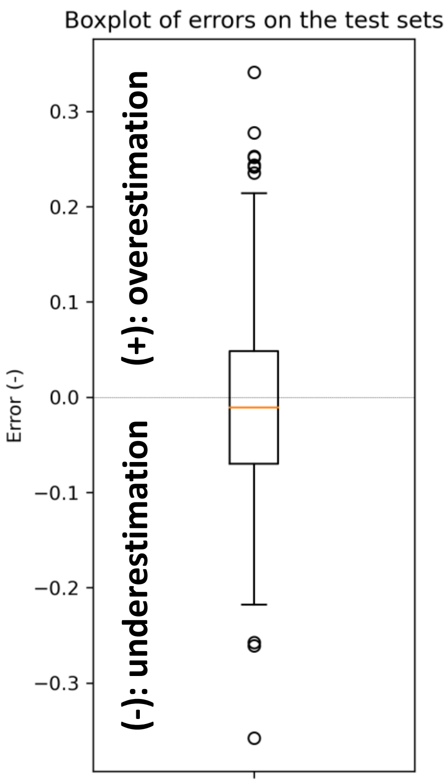
Predictions from climatic indices (1st experiment)



- The predictions obtained from climatic indices only are shown above (for ARI=1.5 years). The model was able to depict a good extent of geographical variability.
- The interquartile range in the boxplot is smaller than the range between -0.1 and 0.1.
- Mean error (ME) and mean absolute error (MAE) are small.

Predictions from climatic indices (1st experiment)





○ $Absolute\ Error\ (AE) > 0.15$

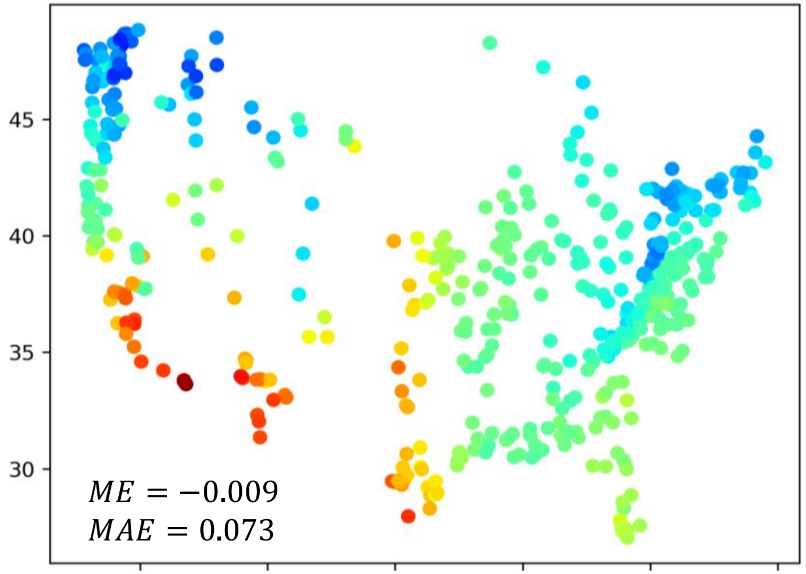
- Here we highlighted the sites characterized by large errors (absolute error > 0.15) with a circle.
- Errors in the interquartile range in the boxplot lay in an interval narrower than]-0.10 ; 0.10[, indicating that most absolute errors are <0.10

This presentation participates in OSPP

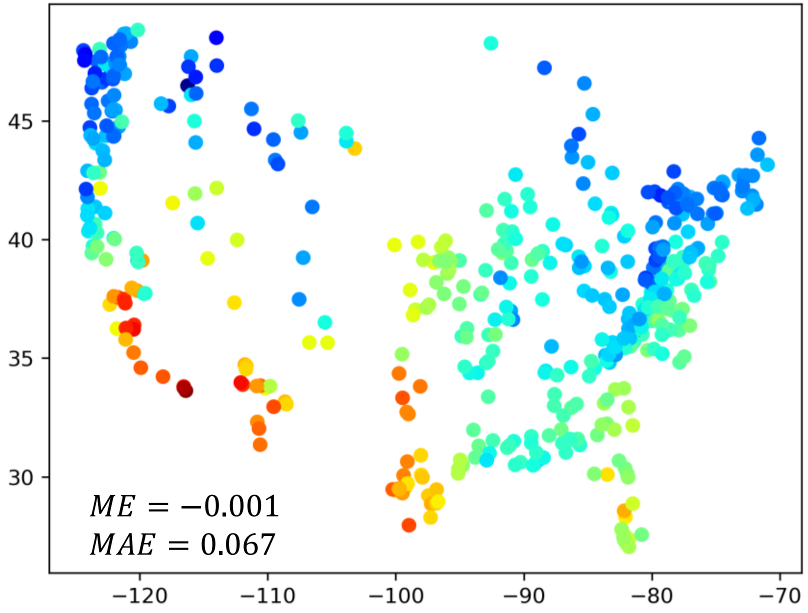
Outstanding Student & PhD candidate Presentation contest

Predictions (on test sets)

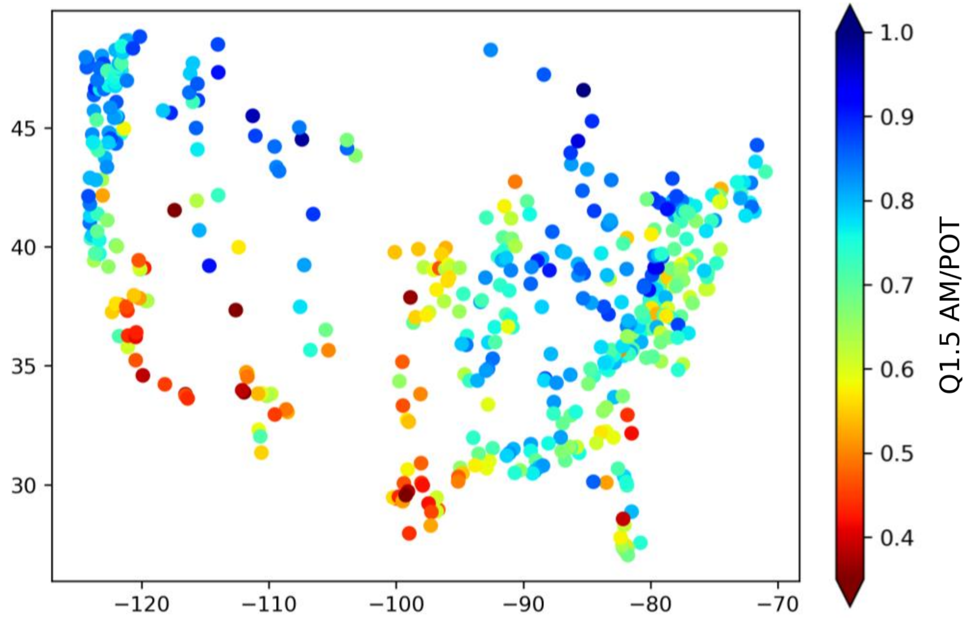
Predictions from **Climatic indices** (1st experiment)



Predictions from **Climatic indices** **+** **Basin characteristics** (2nd experiment)

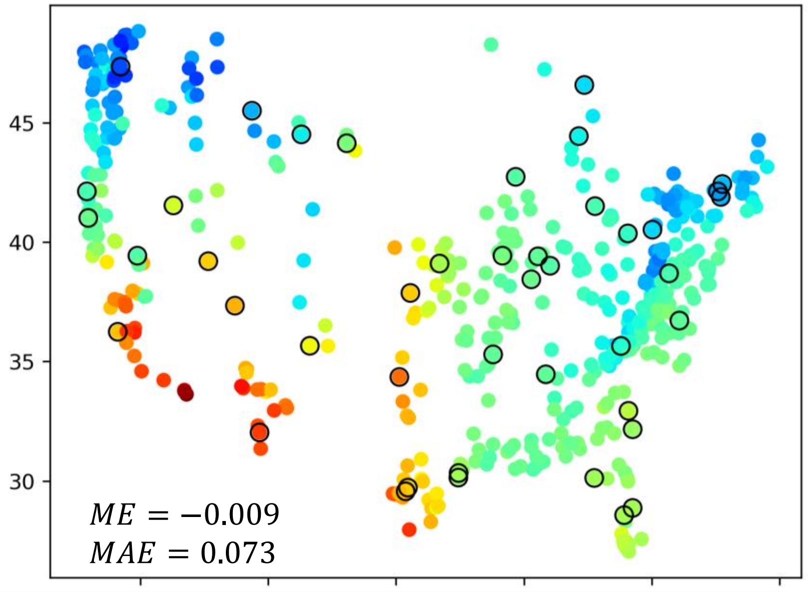


Observations

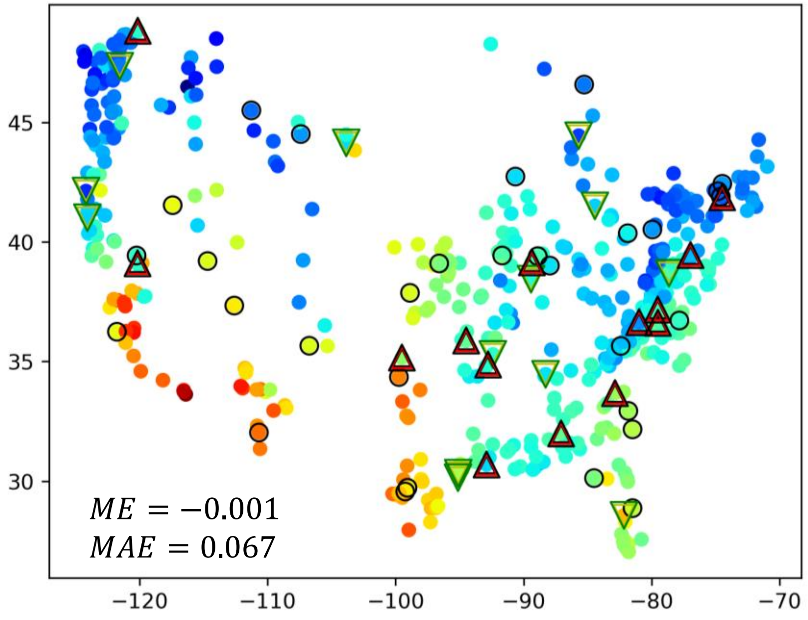


Predictions (on test sets)

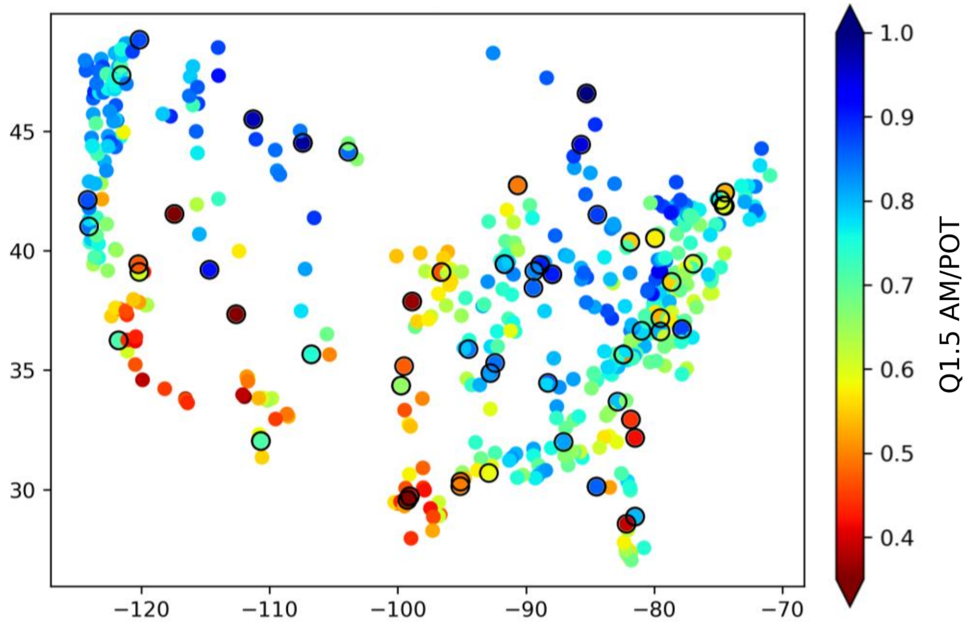
Predictions from **Climatic indices** (1st experiment)



Predictions from **Climatic indices** **+** **Basin characteristics** (2nd experiment)



Observations



○ $AE > 0.15$

▽ AE reduced (<0.15)

△ AE increased (>0.15)

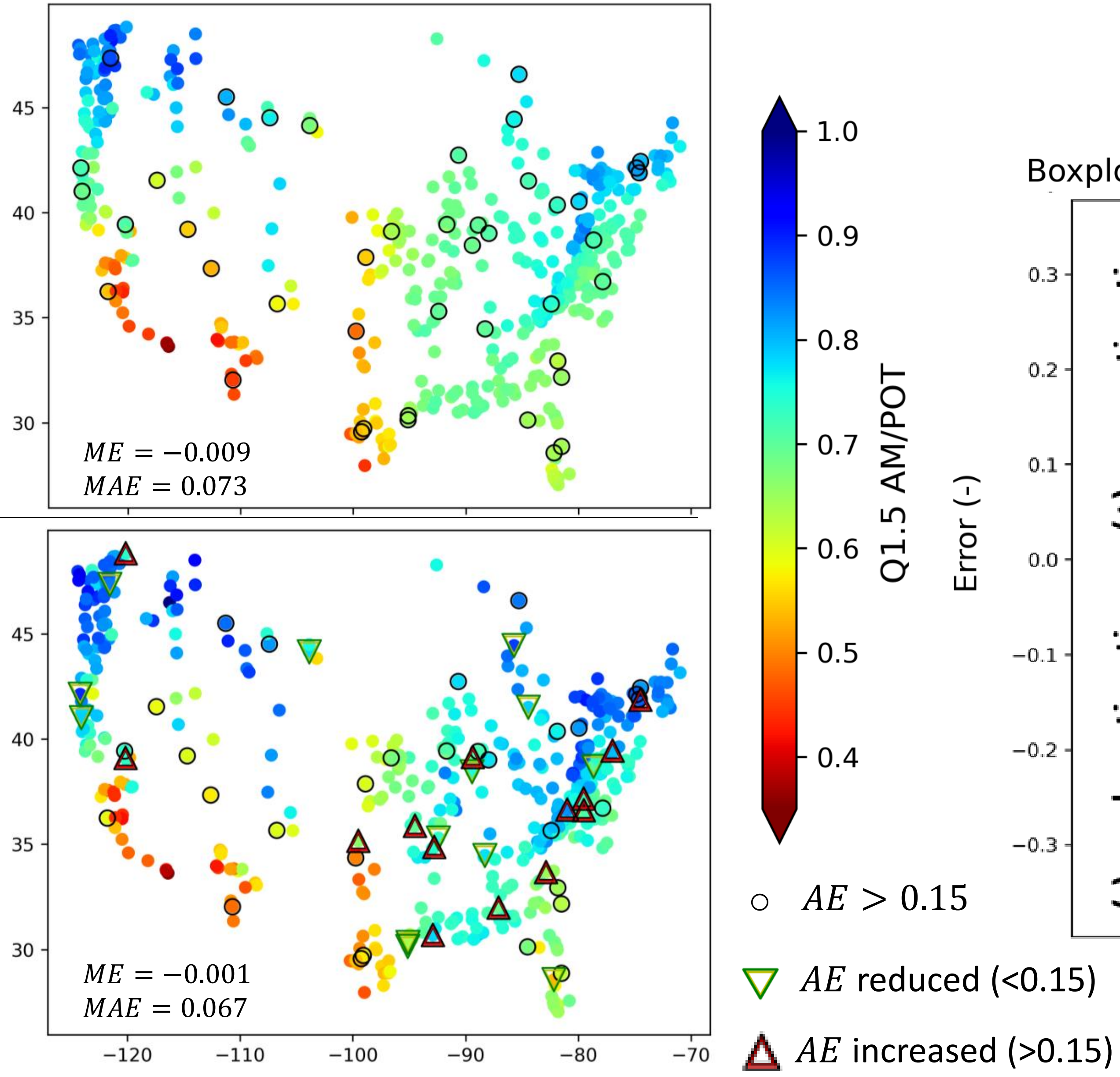
This presentation participates in OSPP

Outstanding Student & PhD candidate Presentation contest

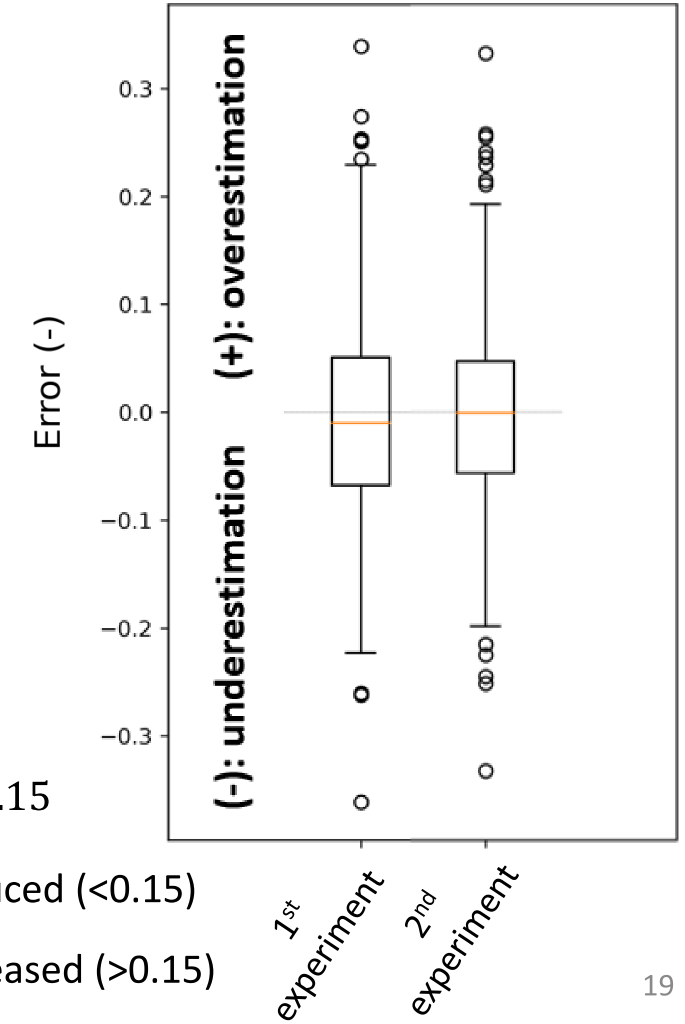
Predictions (on test sets)

Predictions from **Climatic indices** (1st experiment)

Predictions from **Climatic indices** + **Basin characteristics** (2nd experiment)



Boxplots of errors on the test sets



Comparison of the two experiments

This presentation participates in OSPP

Outstanding Student & PhD candidate Presentation contest

- When also basin characteristics are considered among the input features, large errors encountered in the first experiment (climatic indices only) are reduced at some sites, but other sites experience an increase in error.
- Overall, predictions from the model that considers both climatic indices and basin characteristics are closer to the observations, indicating that the latter helps depict more geographical variability in Q_{AM}/Q_{POT}.
- In agreement with that, the boxplot of errors obtained from the second model is narrower, because of the overall reduced errors.

Conclusions

- **Geographical pattern** in the **degree of underestimation of frequent floods** by AM as compared to POT.
- The geographical variability in Q_{AM}/Q_{POT} is mostly due to local **climatic characteristics.**
- The degree of underestimation can be **predicted** from **climatic indices.**
- Considering also **basin properties** to make predictions might lead to marginal improvements.
- When interested in frequent floods, for sites where **only annual maxima are available**, **AM** flood frequency analyses can be **corrected** based on **climatic indices** and **basin characteristics**.

References

- Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*(10), 5293-5313.
- Ball, J.; Babister, M., Nathan, R., Weeks,W., Weinmann, E., Retallick, M., Testoni, I. (2016). *Australian Rainfall and Runoff: A Guide to Flood Estimation*. Commonwealth of Australia: Canberra, Australia.
- Bezak, N., Brilly, M., & Šraj, M. (2014). Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal*, *59*(5), 959-977.
- Hosking, J. R. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(1), 105-124.
- Hu, G.-M.; Ding, R.-X.; Li, Y.-B.; Shan, J.-F.; Yu, X.-T.; Feng, W. (2017). Role of flood discharge in shaping stream geometry: Analysis of a small modern stream in the Uinta Basin, USA. *J. Palaeogeogr*., 6, 84–95.

References

- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences, 23*(12), 5089-5110.
- Pan, X., Rahman, A., Haddad, K., & Ouarda, T. B. (2022). Peaks-over-threshold model in flood frequency analysis: a scoping review. *Stochastic Environmental Research and Risk Assessment*, 1-17.
- Solari, S., Egüen, M., Polo, M. J., & Losada, M. A. (2017). Peaks Over Threshold (POT): A methodology for automatic threshold estimation using goodness of fit p-value. *Water Resources Research*, *53*(4), 2833-2849.
- Southard, R.E., 2010, Estimating the magnitude and frequency of floods in urban basins in Missouri. *Scientific Investigations Report 2010 5073, USGS Publication Warehouse* (https://doi.org/10.3133/sir20105073).
- Virginia Department of Transportation (2021). *Drainage Manual* (*https://www.virginiadot.org/business/locdes/hydra-drainage-manual.asp*)

Thank you for your attention!

Scan the QR code to open the link to the abstract and the judging option.

This presentation participates in OSPP

EGU

Outstanding Student & PhD candidate Presentation contest

Contacts:

fdllaira@memphis.edu