Effect of merging large datasets on prediction accuracy of low flow estimation by random forest

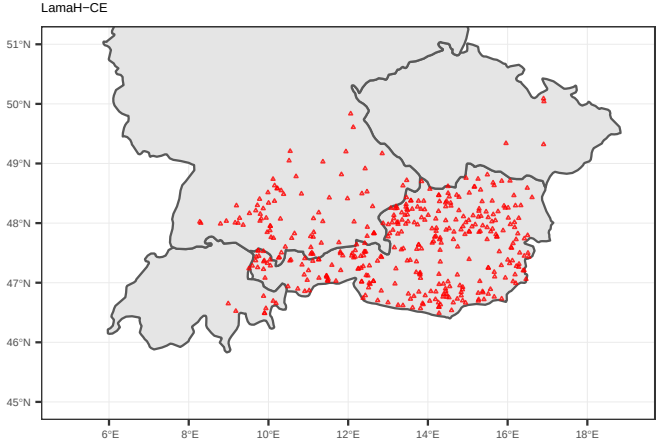
Johannes Laimighofer Gregor Laaha

University of Natural Resources and Life Sciences, Vienna, Department of Landscape, Spatial and Infrastructure Sciences, Institute of Statistics, Peter-Jordan-Straße 82/I, 1190 Vienna, Austria

EGU General Assembly 2022

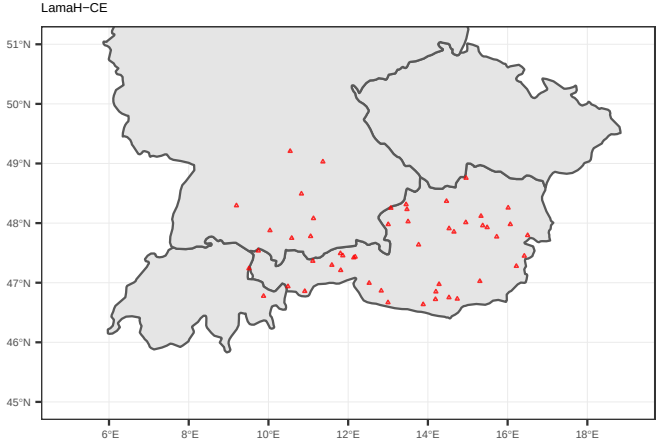
PUB in data-scarce regions

What is the potential of large international datasets to compensate missing data in data-poor regions?



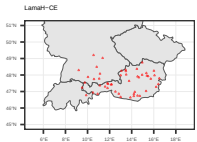
PUB in data-scarce regions

What is the potential of large international datasets to compensate missing data in data-poor regions?

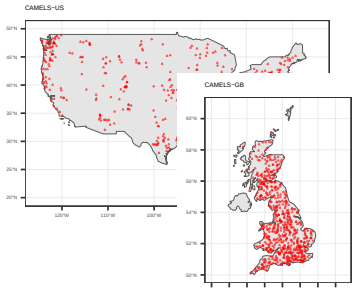


PUB in data-scarce regions

What is the potential of large international datasets to compensate missing data in data-poor regions?

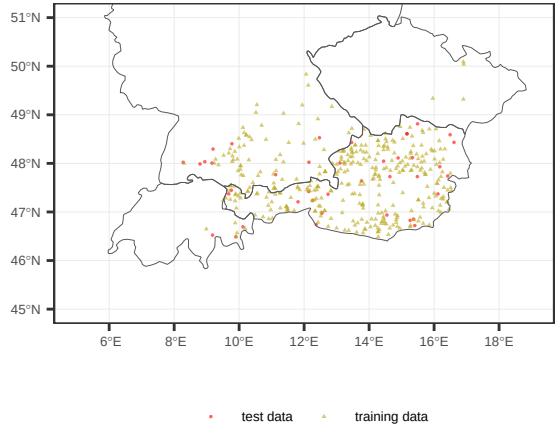


+ additional training data



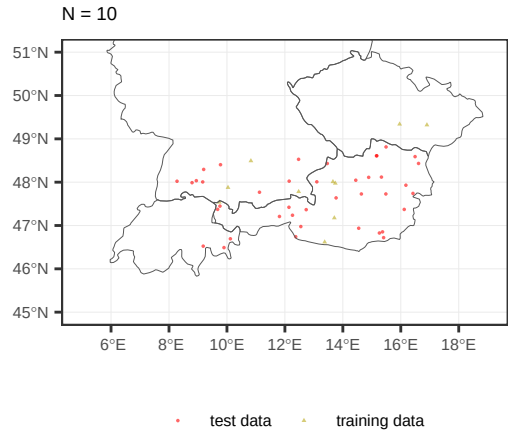
Research questions

1. Effect for **data-dense areas**?
2. Effect for **data-scarce areas**?
3. Added value of **catchment grouping**?



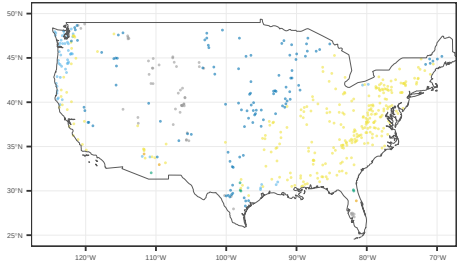
Research questions

1. Effect for **data-dense areas**?
2. Effect for **data-scarce areas**?
3. Added value of **catchment grouping**?



Research questions

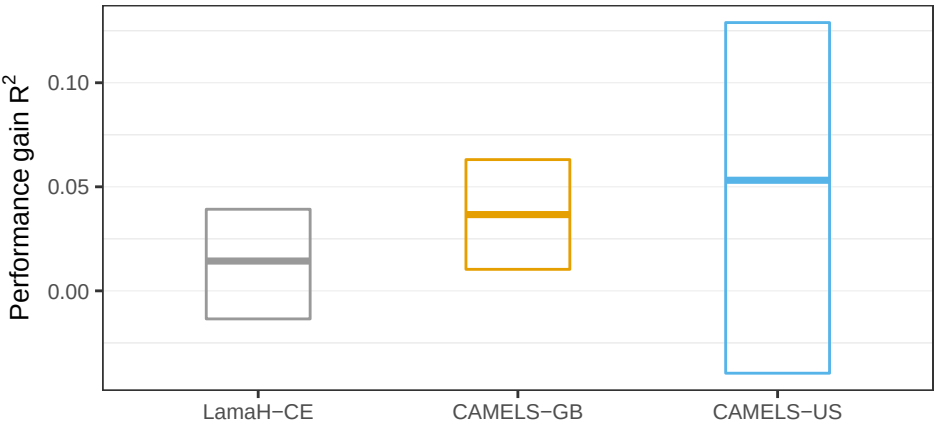
1. Effect for **data-dense areas**?
2. Effect for **data-scarce areas**?
3. Added value of **catchment grouping**?



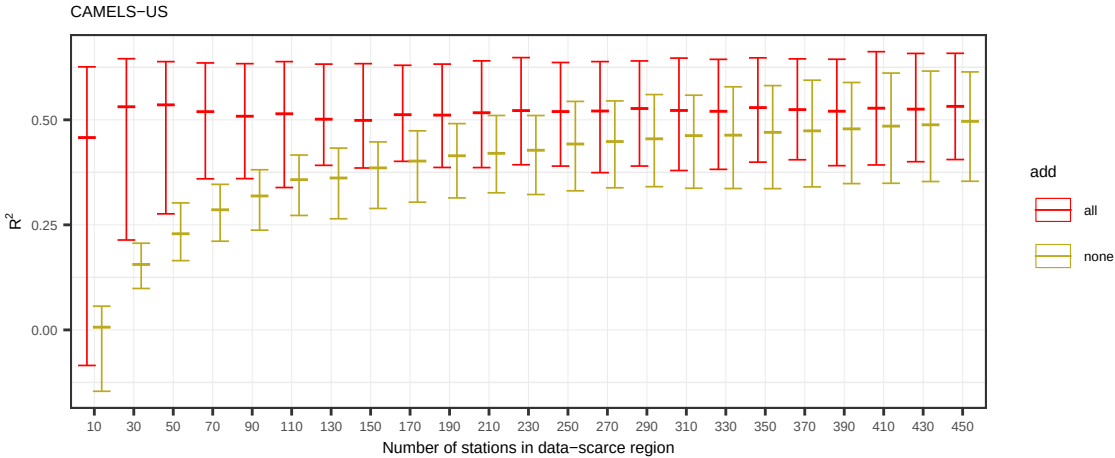
Methods

- Prediction of *Q*95 (0.05 quantile of daily discharge).
- Using random forest (RF) for prediction - no hyperparameter tuning.
- Evaluation by repeated cross-validation (250 times)
- Functional clustering of annual hydrographs for catchment grouping.

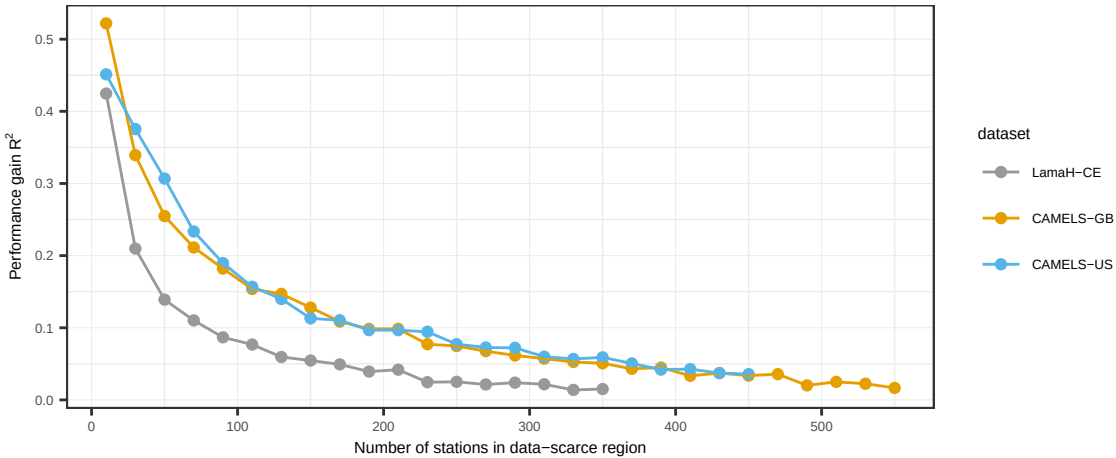
Effect in data-dense areas



Effect in data-scarce regions

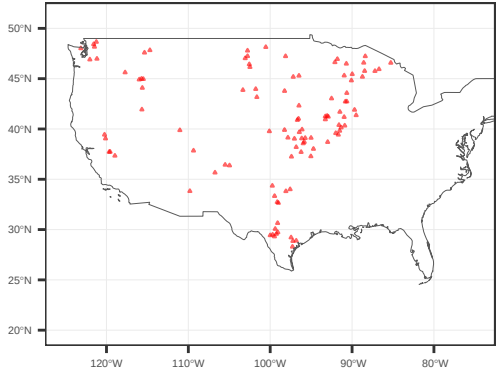


Effect in data-scarce regions

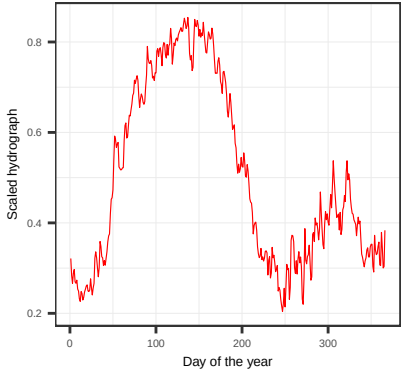


Effect of catchment grouping

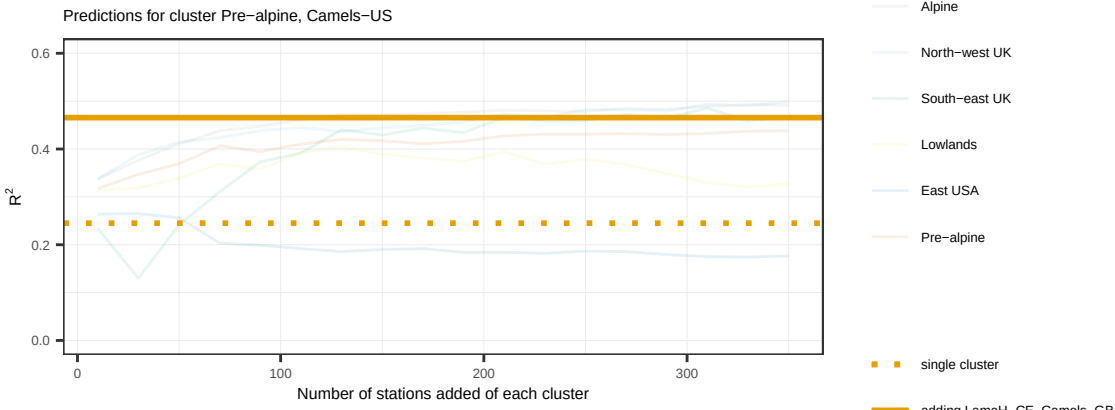
(a) Pre−alpine cluster USA



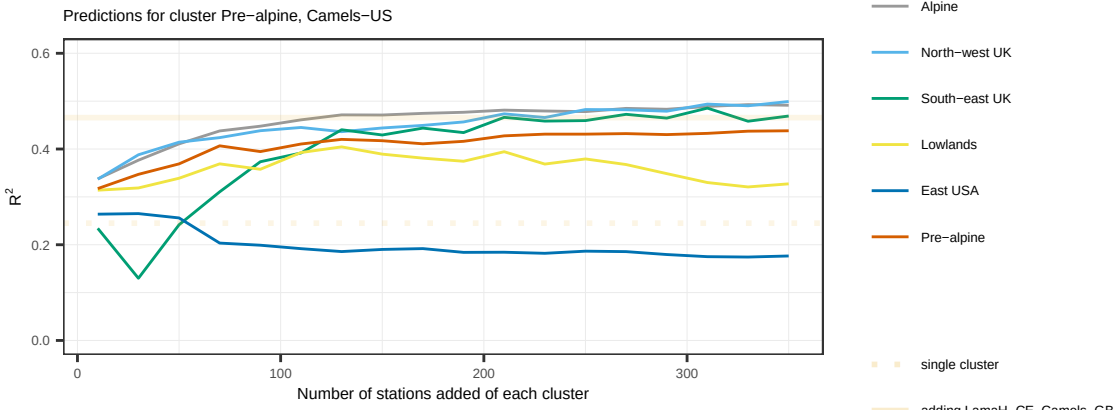
(b) Average hydrograph of pre−alpine cluster in the USA



Effect of catchment grouping



Effect of catchment grouping



Conclusions

- Effect in *data-dense* regions: **Slight improvement** of R^2.
- Effect in *data-scarce* regions: **Large improvement** for study area with **low station density**!
- Effect on *catchment grouping*: Catchment grouping has a **negative effect** on prediction performance.
- Not each cluster has a positive effect on prediction performance.

Acknowledgements

Thank you for your attention!

ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

VSC VIENNA SCIENTIFIC CLUSTER