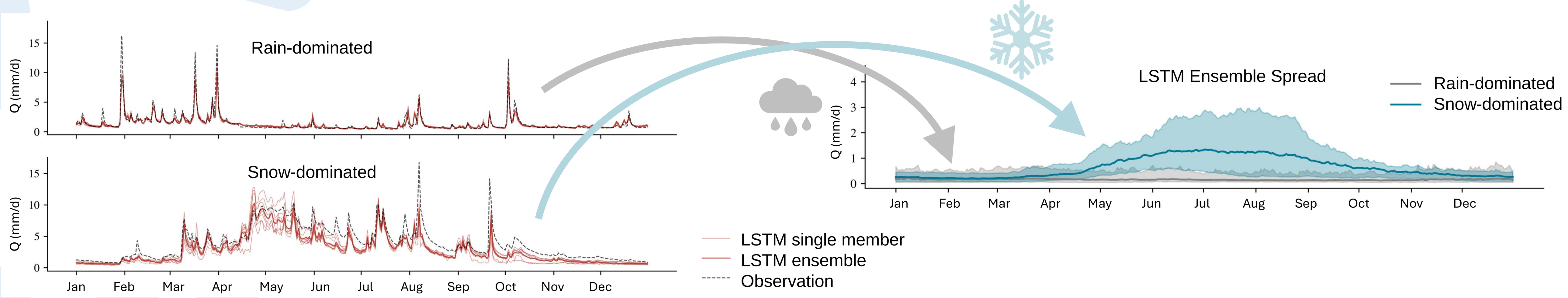
Cryosphere Data and Its Value for Deep Learning Hydrological Simulations

Corinna Frank[1,2,3], Jan Philipp Bohl[4,2], Manuela Brunner[1,2,3], Martin Gauch[5], and Marvin Höge[6]

[1] Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[2] WSL Institute for Snow and Avalanche Research SLF, Davos Dorf, Switzerland
[3] Climate Change, Extremes and Natural Hazards in Alpine Regions Research Center CERC, Davos Dorf, Switzerland
[4] Department of Mathematics, ETH Zurich, Zurich, Switzerland
[5] Google Research, Zurich, Switzerland
[6] Sustainserv, Zurich, Switzerland

1. BACKGROUND

While regionally-trained LSTM show good performance over mountain areas, the model uncertainty is still largest during the melting period. We observe an increased ensemble spread and larger errors of a single member during snow and glacier melt (Apr-Oct depending on elevation) [1].



How can we improve the simulation of melting processes in deep learning hydrological models?

2. HYPOTHESES

A Simulation of snowmelt can be enhanced by addition of (remotely sensed) snow cover data.

B Lumped models benefit from an improved description of the heterogeneity of topography and snow cover within mountain catchments.

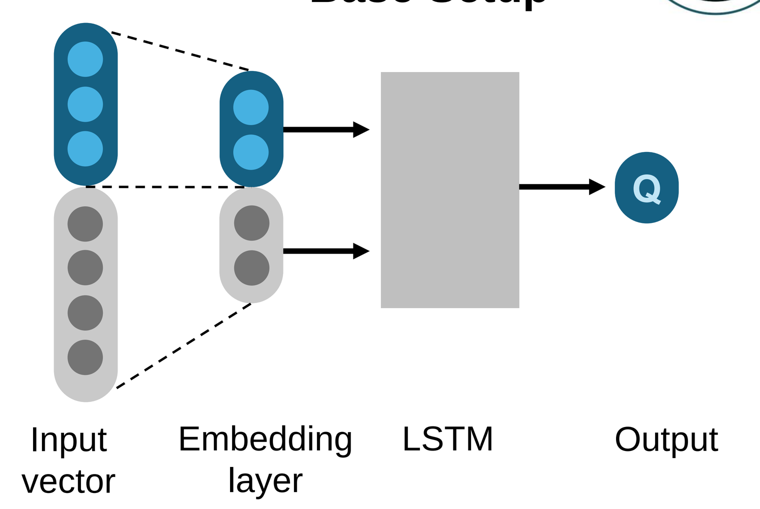
3. SETUP

- Lumped, daily
- Single model for 5224 catchments, with ≥ 30 years of streamflow observations
- 10-fold cross-validation (train/validation/test : 4178 / 523 / 523) out-of-space and out-of-time
- 1985-2020 (5y validation / 20y train / 10y test)

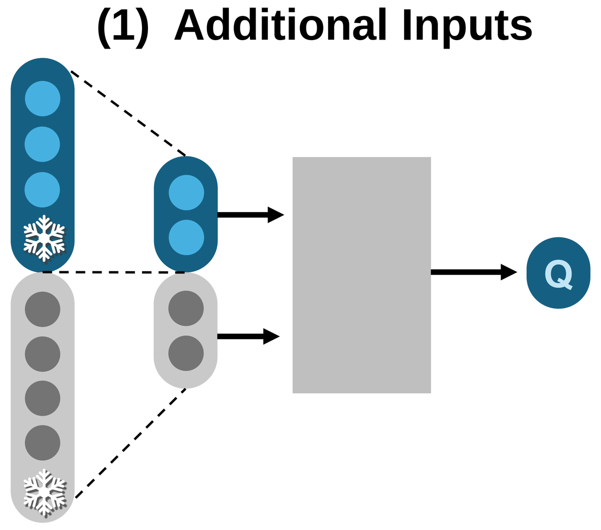
Base Setup

CERRA daily time-series:
- P
- T_min
- T_max
- ISWR

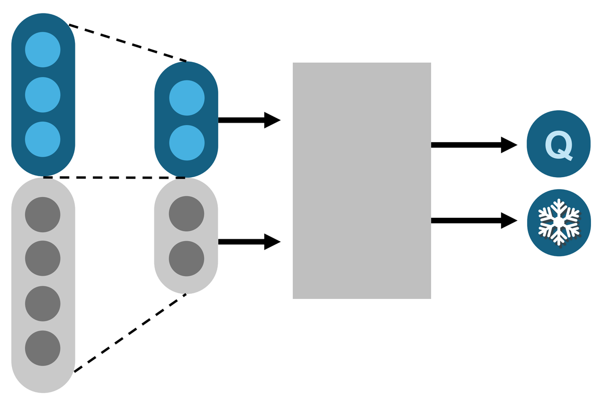
41 catchment attributes retrieved from EStreams [2]



(1) Additional Inputs



(2) Additional Target

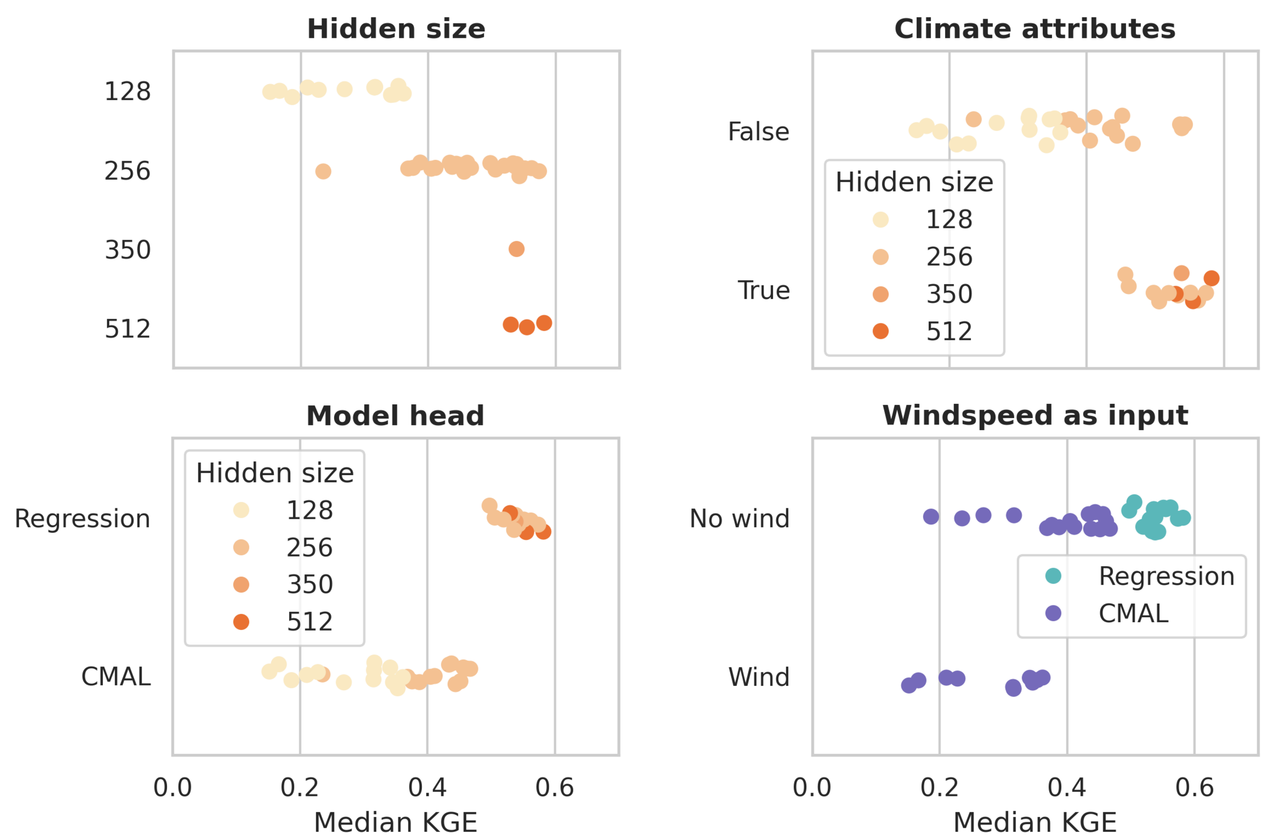


- Dynamic input (P, T_min, T_max, ISWR)
- Static input (topographic, climatologic, soil, landcover, lakes + reservoirs)
- Snow-related variable (SWE, fSCA, %south-facing slope, …)

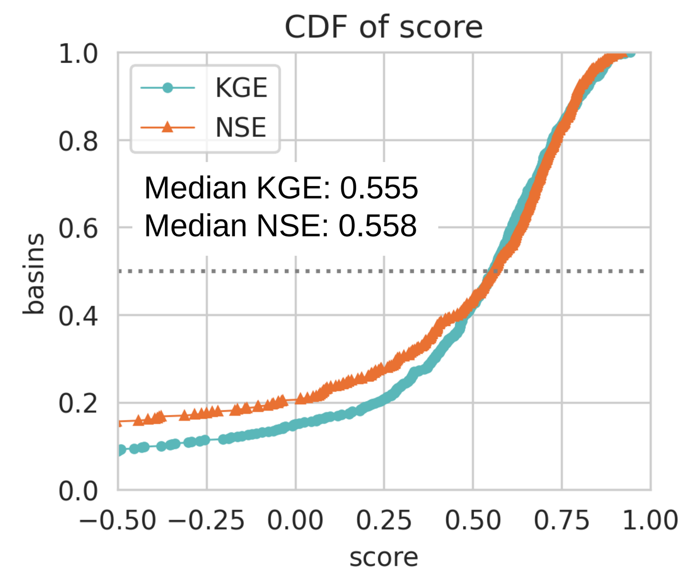
4. PRELIMINARY RESULTS

Base Setup

Optimizing the Architecture



Test Results



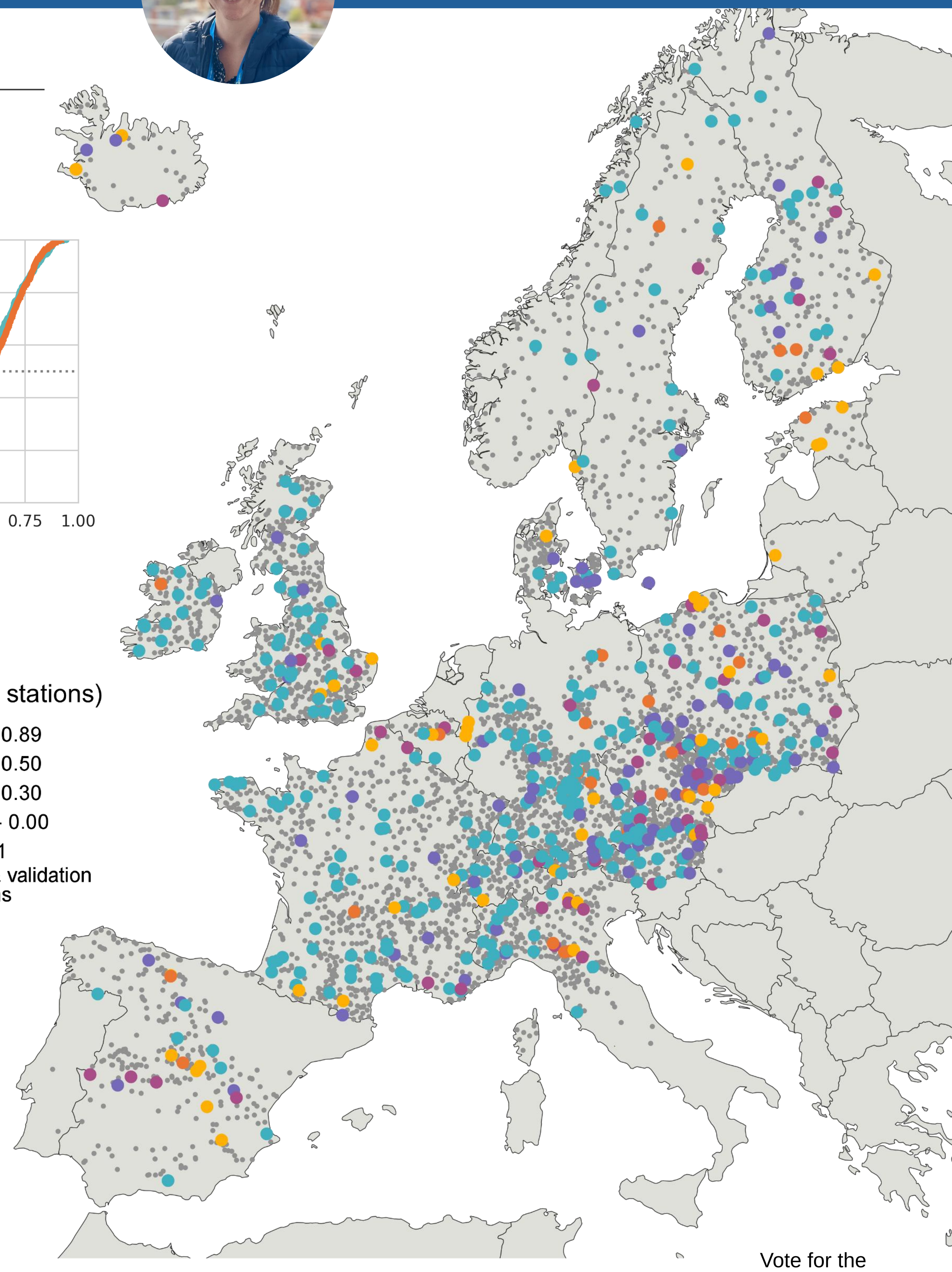
(1) Additional Inputs

Table: Model test performance over the snow melt season (Apr-Jun, 160 basins in the Central Alps). Taken from [1].

Model	Median KGE		Median NSE	
	Ensemble mean	SD	Ensemble mean	SD
Base Setup	0.65	σ = 0.046	0.70	σ = 0.039
(1) Additional Inputs *	0.64	σ = 0.023	0.72	σ = 0.016

* simulated SWE from distributed model PCR-GLOBWB 2.0 [3]

- Increasing the hidden size helps performance.
- The model is sensitive to the model head (output layer).
- **Adding simulated SWE** from a distributed model [3] as additional input to the model **reduces the ensemble spread** for the snow melt season (Apr-Jun).
- Absolute Performance does not improve by addition of catchment-average simulated SWE.



5. OUTLOOK

In future work, we will further investigate the integration of additional data describing the snow cover:

- Comparison of different integration methods
- Comparison of different data levels (local vs. Europe-wide snow products, time-series vs. attributes)

REFERENCES

[1] Master Thesis Jan Philipp Bohl, March 2025
[2] do Nascimento, T. V. M., et al. EStreams: An integrated dataset and catalogue of streamflow, hydro-climatic variables and landscape descriptors for Europe. 1.2, Zenodo, doi:10.5281/zenodo.14778580, 2025.
[3] Janzing, J., et al.. Hyper-resolution large-scale hydrological modelling benefits from improved process representation in mountain regions, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-3072, 2024.
[4] Kratzert et al. NeuralHydrology --- A Python library for Deep Learning research in hydrology. Journal of Open Source Software, 7(71), 4050, https://doi.org/10.21105/joss.04050, 2022.

Vote for the OSPP contest

Sharing not permitted