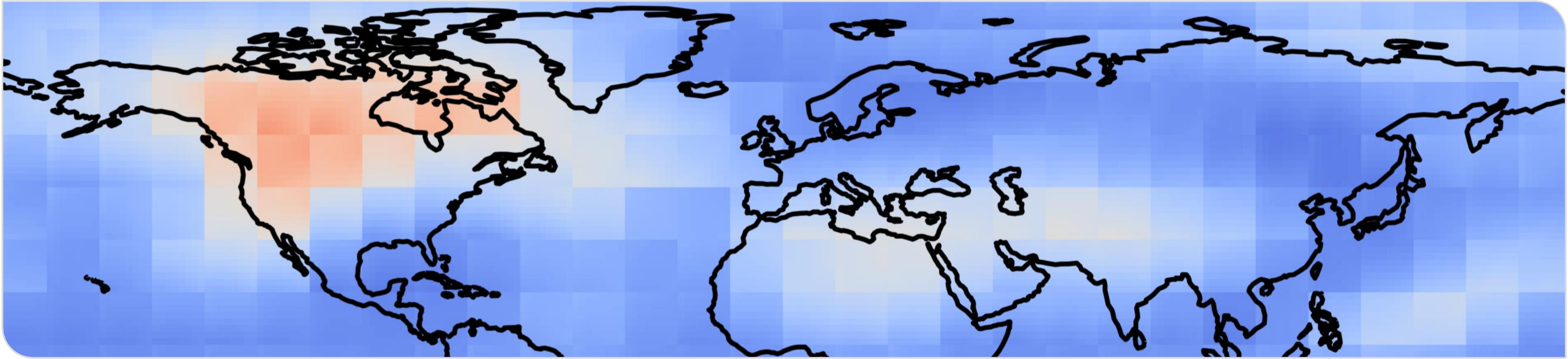
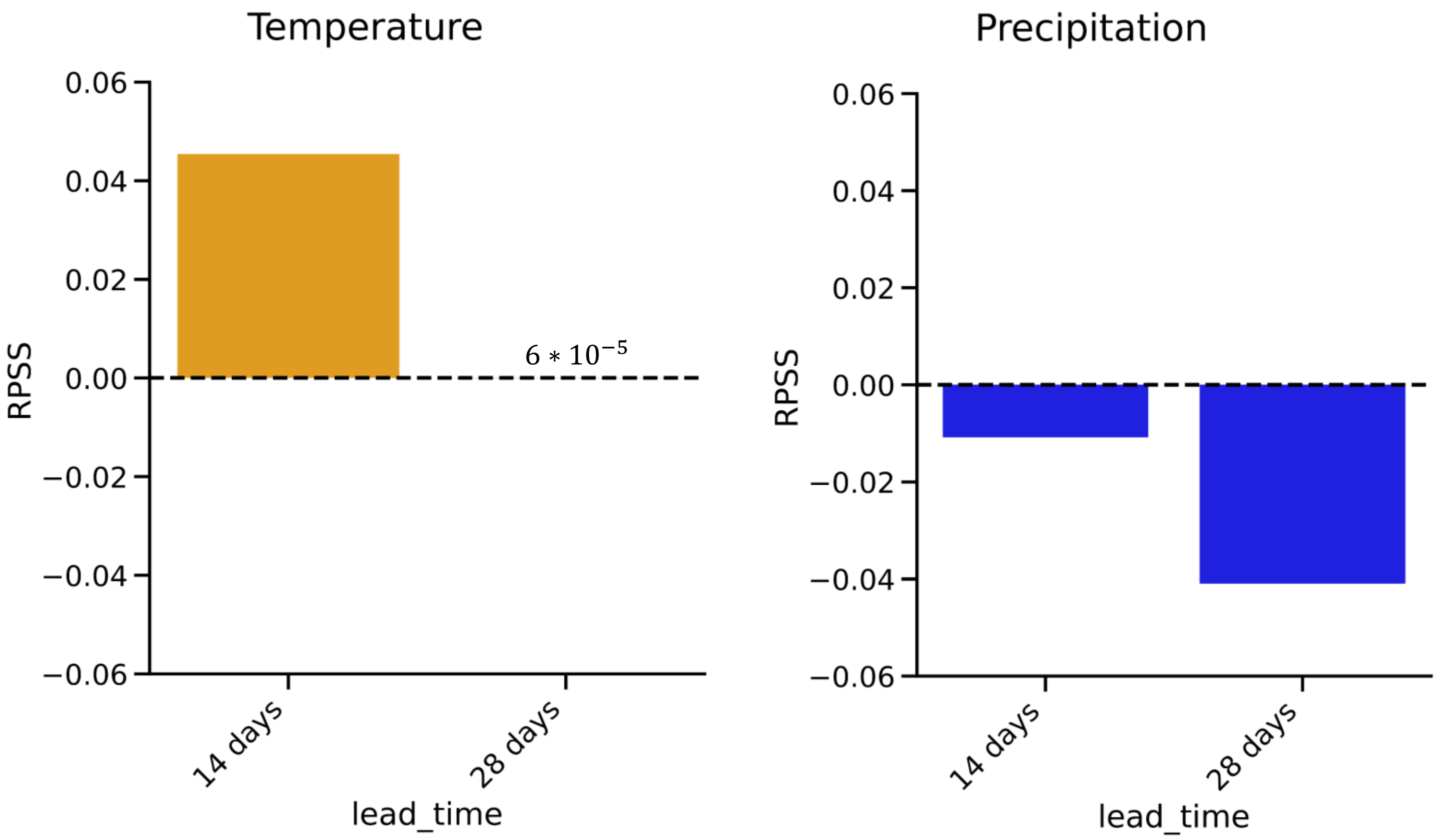
Karlsruhe Institute of Technology

Machine Learning for global probabilistic prediction of temperature and precipitation on sub-seasonal time-scales

Nina Horat, Sebastian Lerch



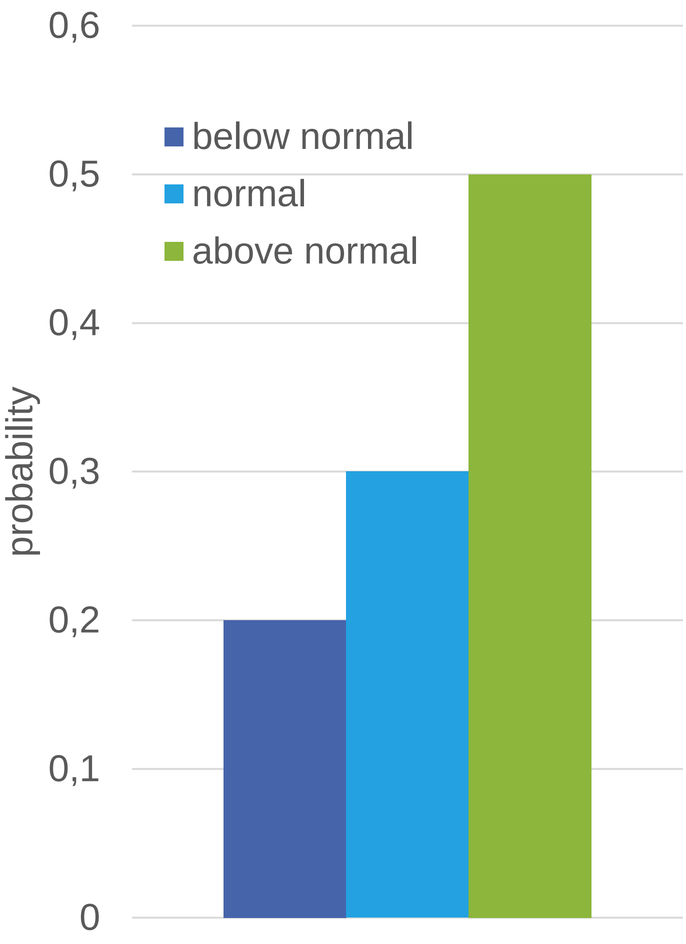
Probabilistic extended-range forecasts of ECMWF



- Globally aggregated RPSS for all Thursday forecasts in 2020.
- RPSS > 0: forecast is better than climatology (random guessing).

Challenge to improve Sub-seasonal to Seasonal Predictions using Artificial Intelligence (2021)

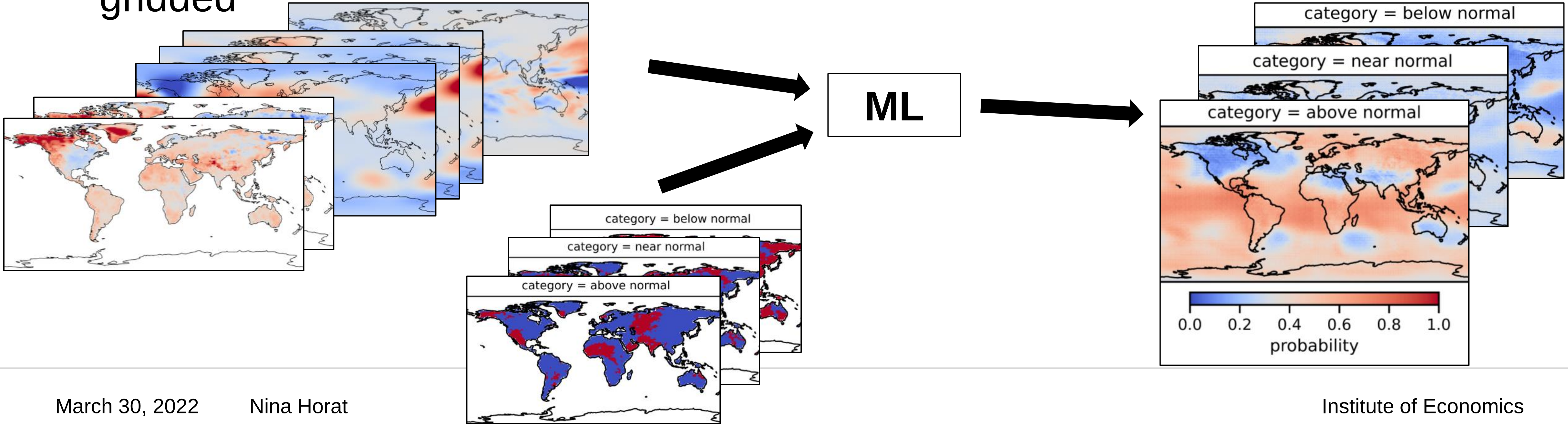
- Task:
  - Predict biweekly aggregates of temperature and precipitation for two lead times (3-4 & 5-6 weeks) for the year 2020 over land (globally)
- To beat:
  - extended-range probabilistic prediction of ECMWF (based on terciles)
  - random guessing climatology: [1/3, 1/3, 1/3]
- All information: https://s2s-ai-challenge.github.io/



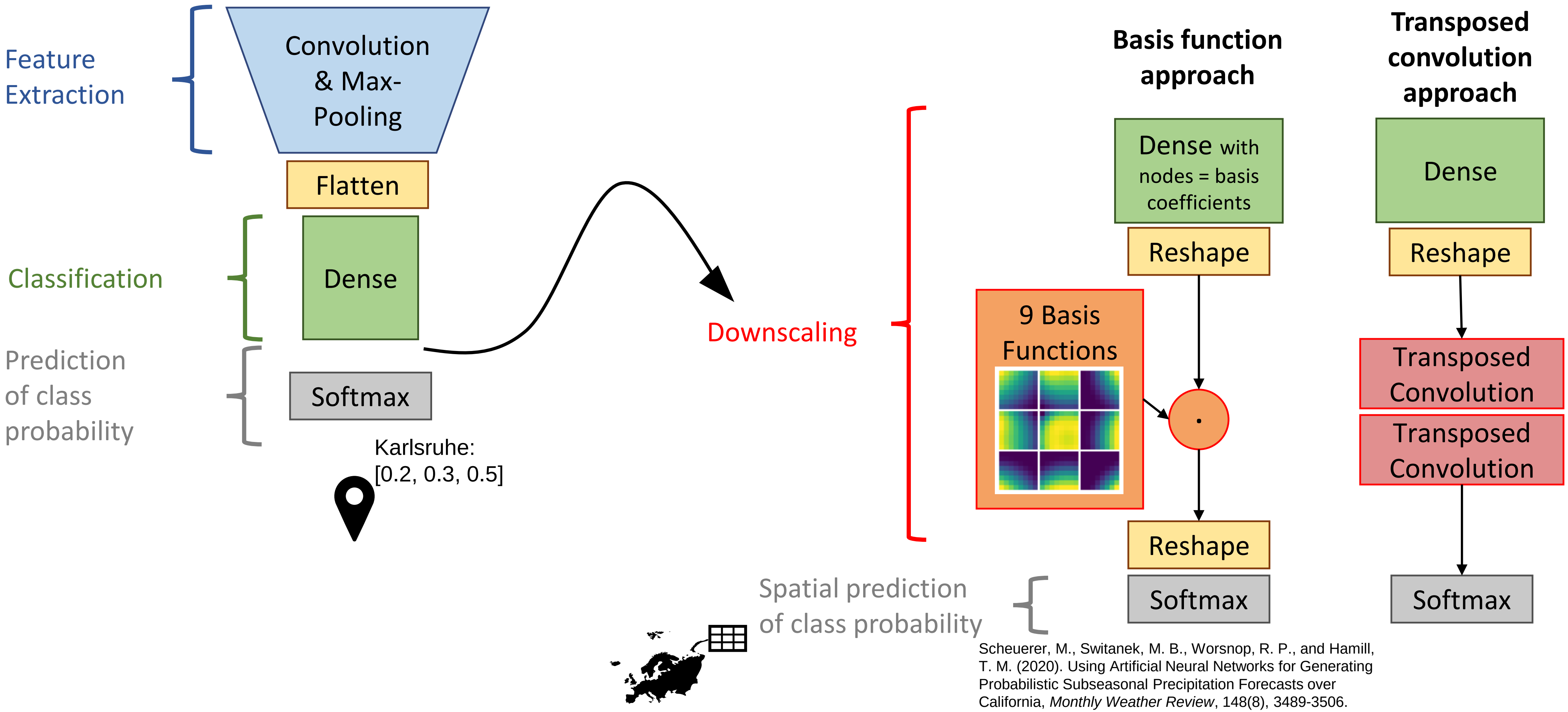
Challenge to improve Sub-seasonal to Seasonal Predictions using Artificial Intelligence

Karlsruhe Institute of Technology

- Data (provided by the challenge):
  - S2S forecasts:
    - Train data: weekly ECMWF hindcasts for the years 2000-2019 (11 members)
    - Test data: weekly ECMWF forecasts for the year 2020 (51 members)
  - Observations: NOAA CPC temperature and total precipitation from IRIDL, gridded

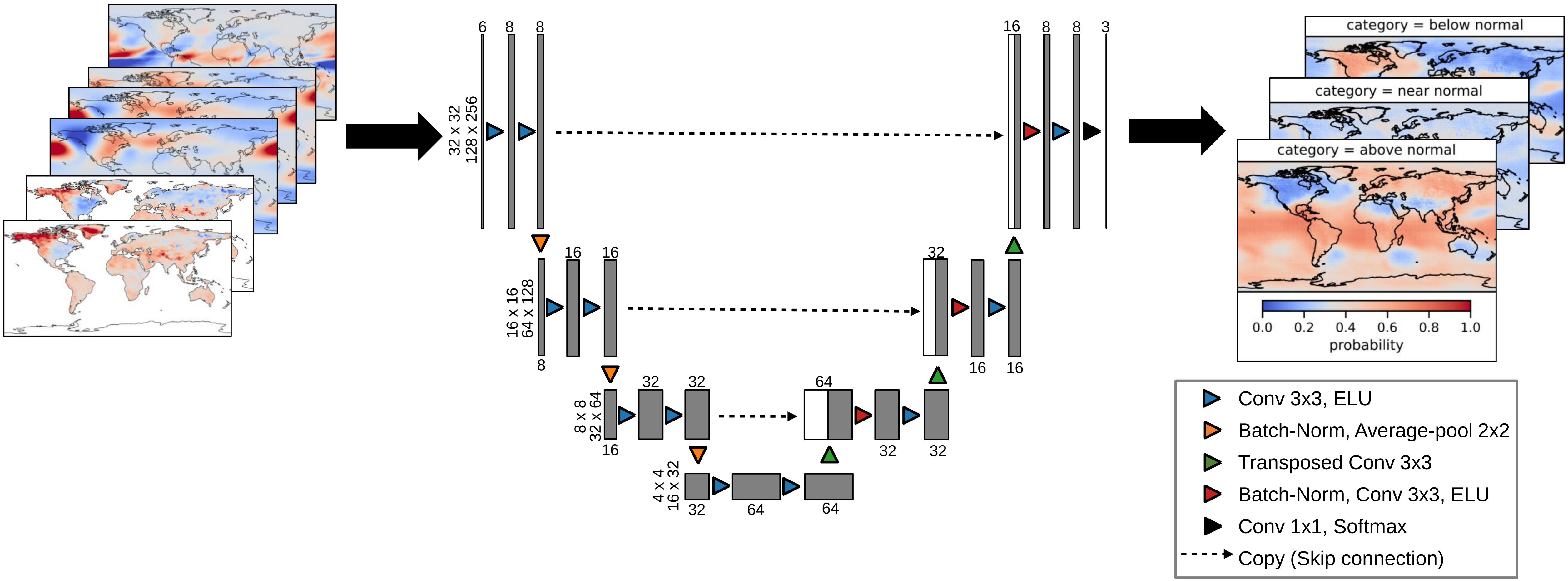


CNNs for spatial prediction



Scheuerer, M., Switanek, M. B., Worsnop, R. P., and Hamill, T. M. (2020). Using Artificial Neural Networks for Generating Probabilistic Subseasonal Precipitation Forecasts over California, *Monthly Weather Review*, 148(8), 3489-3506.

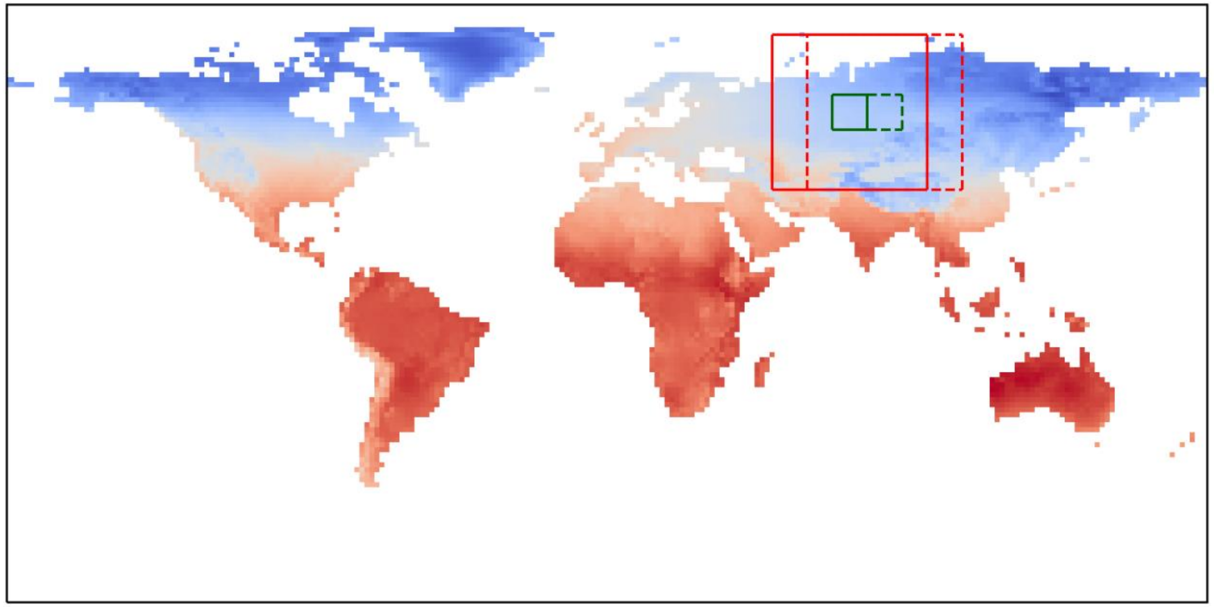
UNets



6
8
8
32 x 32
128 x 256
16
8
8
3
16
16
16 x 16
64 x 128
8
32
16
16
32
32
8 x 8
32 x 64
16
64
32
32
4 x 4
16 x 32
32
64
64
category = below normal
category = near normal
category = above normal
0.0
0.2
0.4
0.6
0.8
1.0
probability
Conv 3x3, ELU
Batch-Norm, Average-pool 2x2
Transposed Conv 3x3
Batch-Norm, Conv 3x3, ELU
Conv 1x1, Softmax
Copy (Skip connection)

Method overview

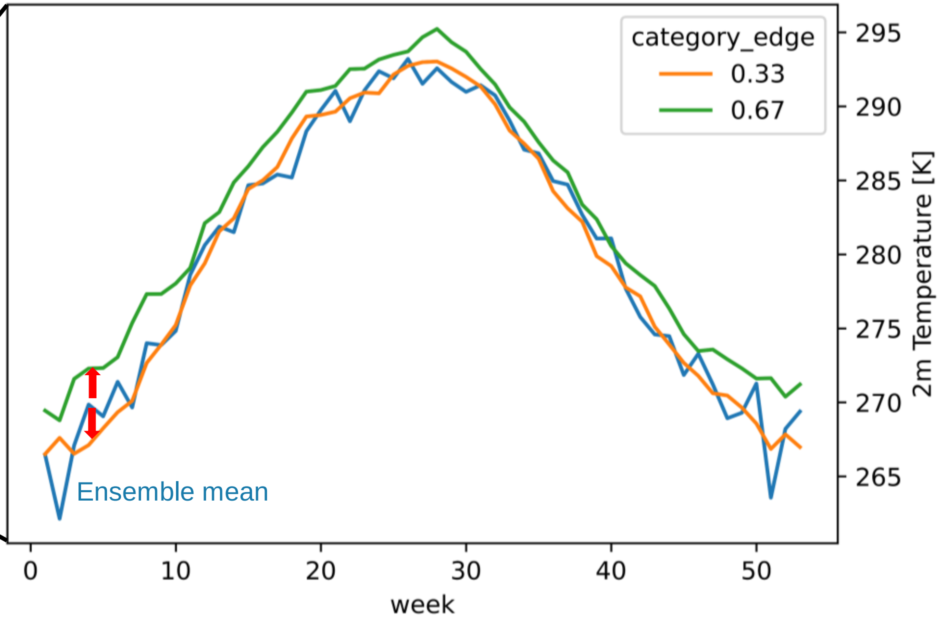
- 4 different ML models:
  - CNN
    - with basis functions
    - with transposed convolutions
  - UNet
    - trained on patches
    - trained on global input
- Patchwise training = data augmentation

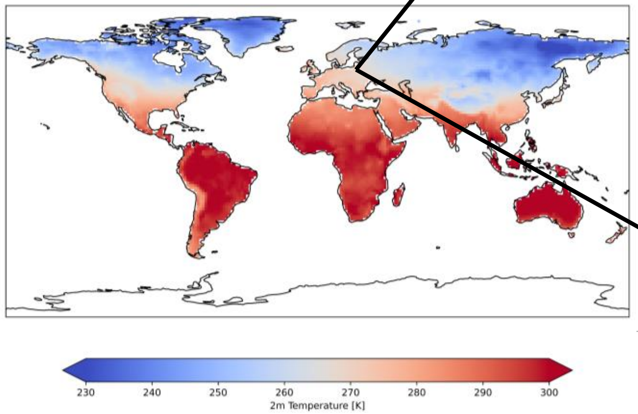


Feature Engineering

- Features:
  - Ensemble mean
  - For temperature:
    - t2m
    - gh500, gh850, msl
  - For precipitation:
    - tp
    - gh500, gh850, msl
    - tcw

- Target variable features: distance to tercile edges

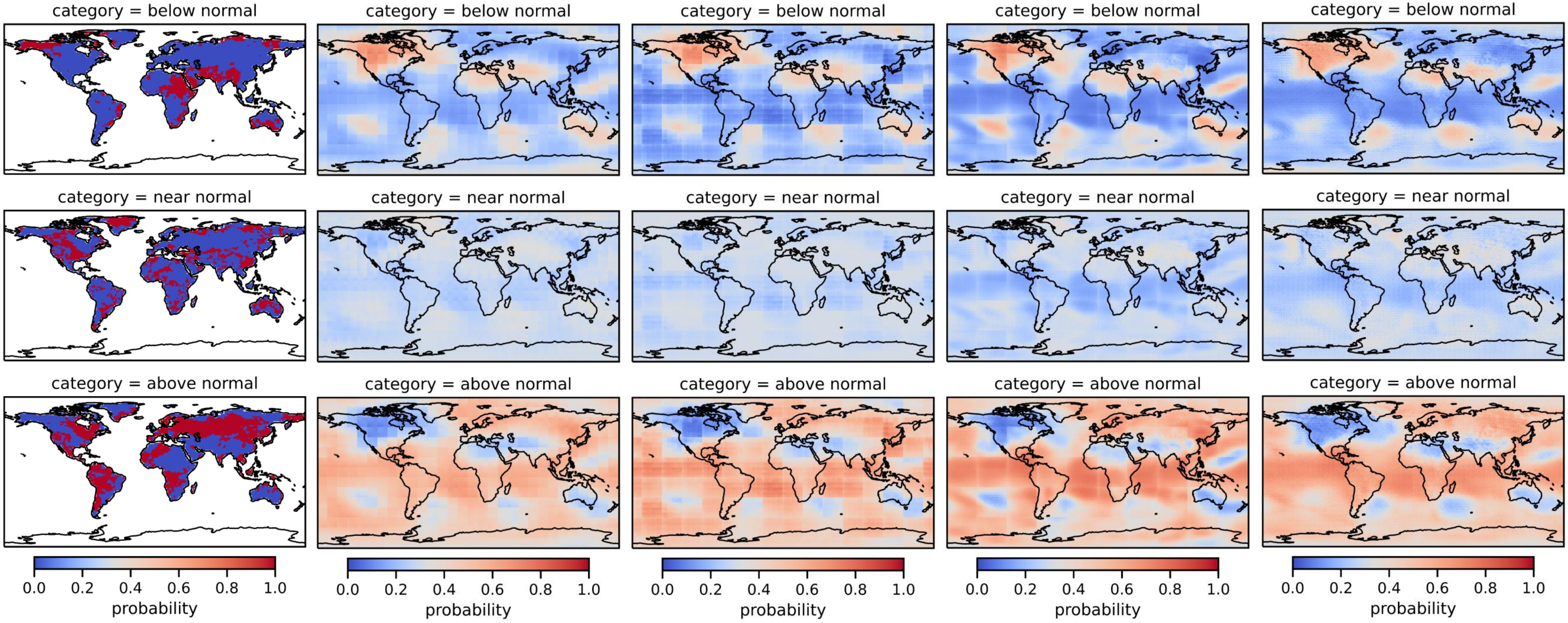




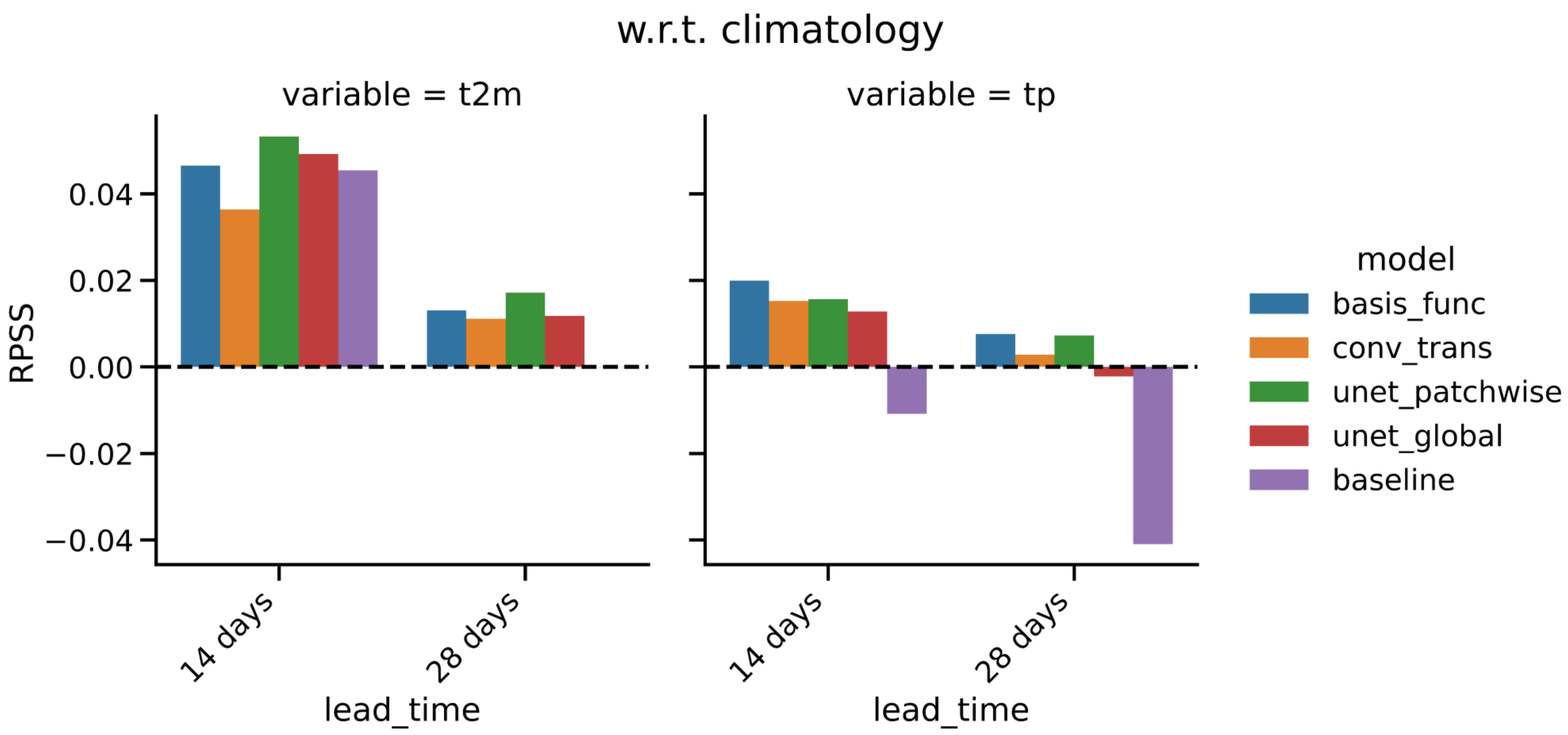
Lower tercile edge t2m

Temperature: lead time = 3 – 4 weeks

Observation	CNN with basis functions	CNN with transposed convolutions	UNet trained on patches	UNet trained on global input

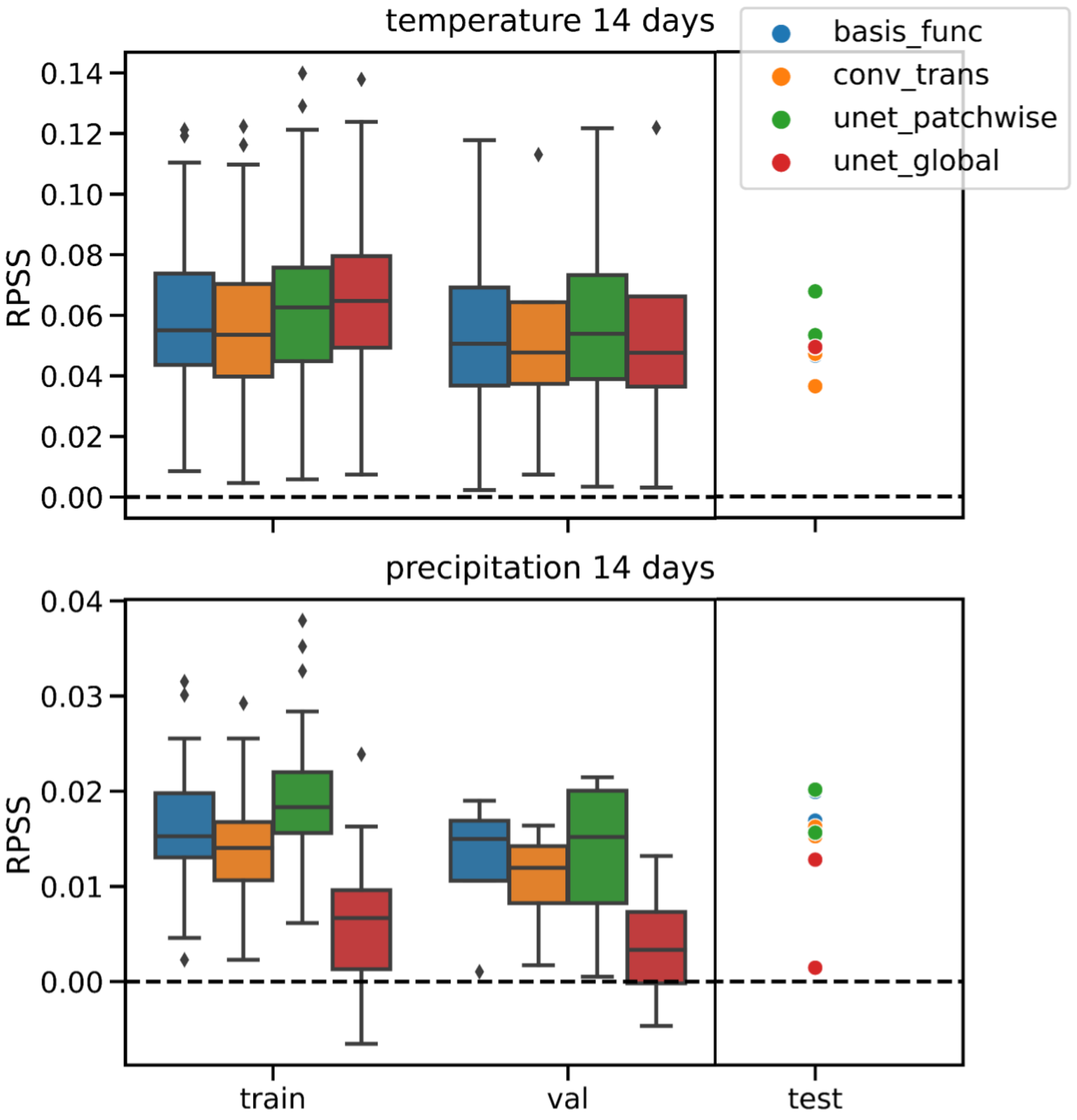


Performance on test data (2020):



How representative is the performance on the test data?

- No systematic differences between performance on training, validation and test data.
- No overfitting
- Performance on the test year 2020 is representative for the „true“ performance



Summary and Outlook

- Development of new ML architectures specifically designed for correcting global forecasts
  - Exploit spatial information
  - Create spatial prediction
  - Patchwise training as data augmentation technique
- ML methods outperform the operational probabilistic extended-range forecasts of ECMWF
- Outlook:
  - Include additional predictors
  - Combine spatial input with timeseries input