Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

UNIVERSITÄT BONN

FOR 2131

International Max Planck Research School on Earth System Modelling

Max-Planck-Institut für Meteorologie

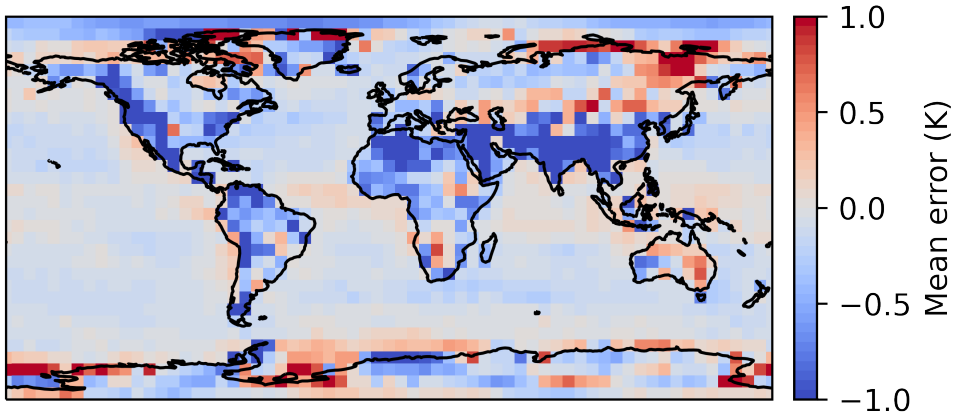
Self-Attentive Ensemble Transformer: Representing Ensemble Interactions in Neural Networks for Earth System Models

Tobias Sebastian Finn

Tackling Climate Change with Machine Learning Workshop at ICML 2021

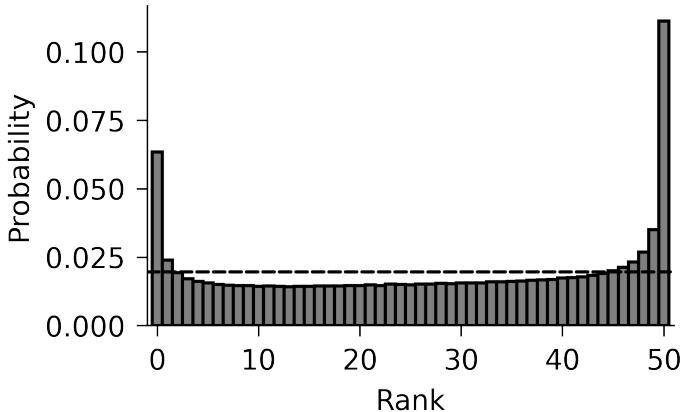
Why do we need to post-process ensemble simulations?

Model bias



Mean error in 2-metre-temperature of the IFS-EPS ensemble mean to the ERA5-reanalysis on a bilinearly regridded grid for 2019

Uncalibrated forecast



Rank histogram in 2-metre-temperature of the IFS-EPS compared to the ERA5-reanalysis for all grid points and the year 2019

+ Forecast non-modelled variables

How to post-process the ensemble data?

Let's use neural networks to correct the model bias

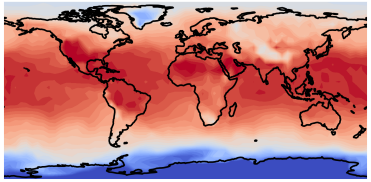
How to incorporate ensemble information?

Use self-attention similar to ensemble data assimilation

→ **Ensemble transformer**

	Raw Ensemble	PPNN (Rasp & Lerch, 2013)	W/o self-attention	**Transfromer** (This talk)
Bias correction	✗	✓	✓	✓
Calibration	✗	✓	✗	✓
Non-parametric output	✓	✗	✓	✓
Correlated output	✓	✗	?	✓

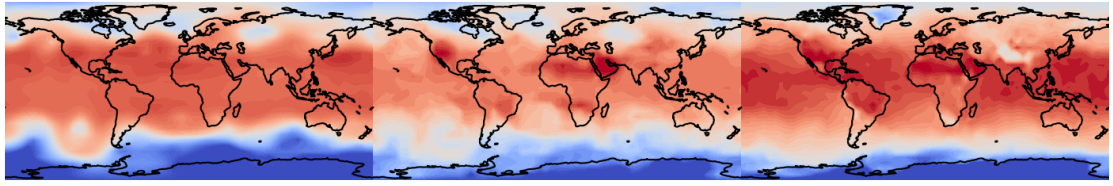
Global post-processing case



ERA5 reanalysis

2-metre-temperature

Global regression with
convolution neural networks



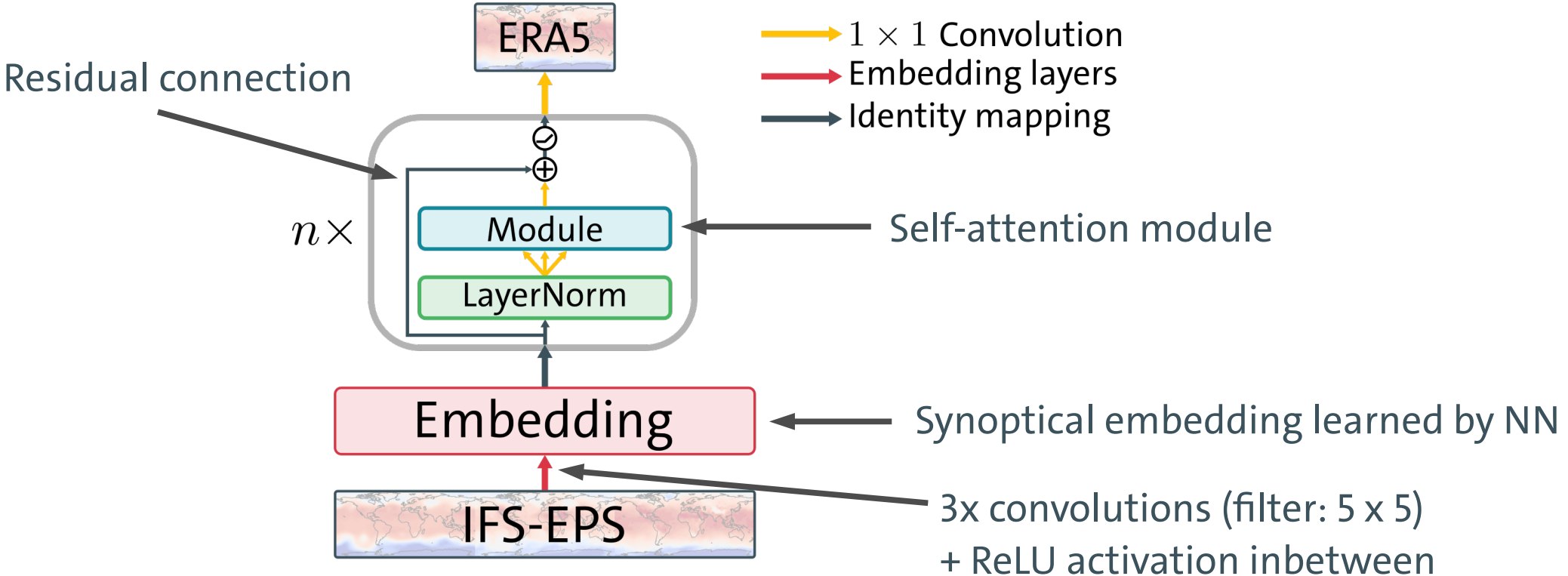
Geopotential
on 500 hPa

Temperature
on 850 hPa

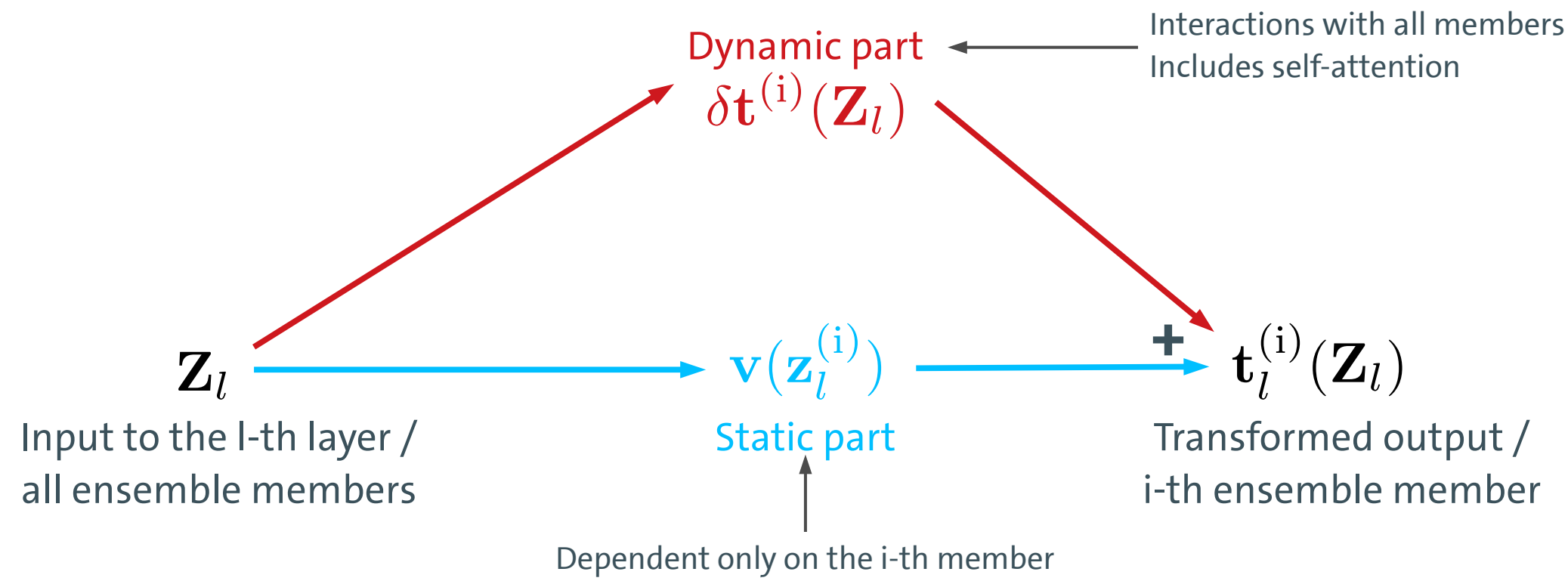
2-metre-temperature

IFS-EPS (50 members, lead time: 48 h)

Ensemble transformer architecture



Main idea of a single self-attention module



Self-attention can explain itself and create informative maps

Attention maps for 2019-09-01 12:00 UTC

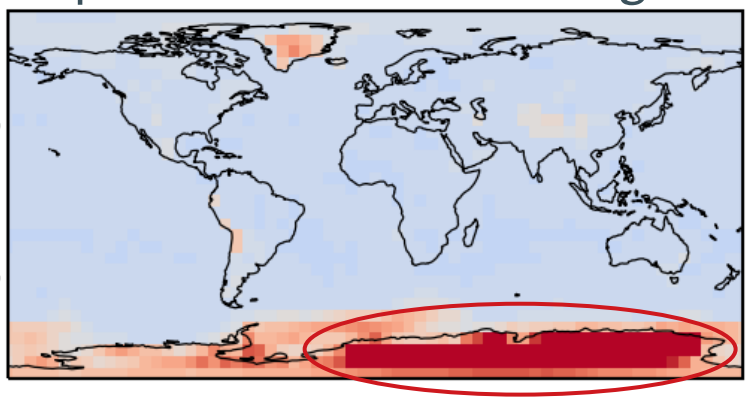
Temperatures below freezing level?

Positive temperature anomalies?

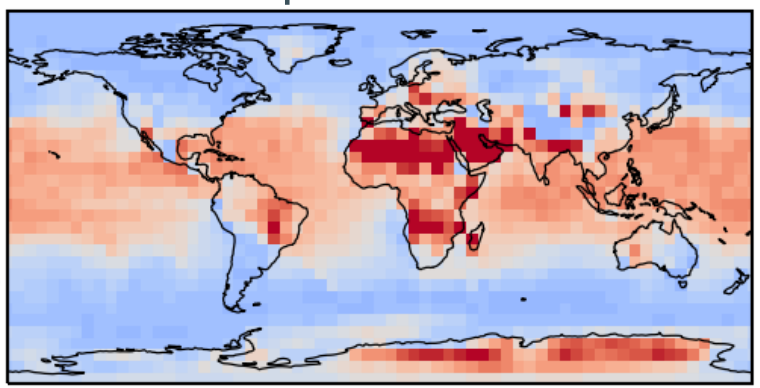
Head →

Layer →

(1,56)

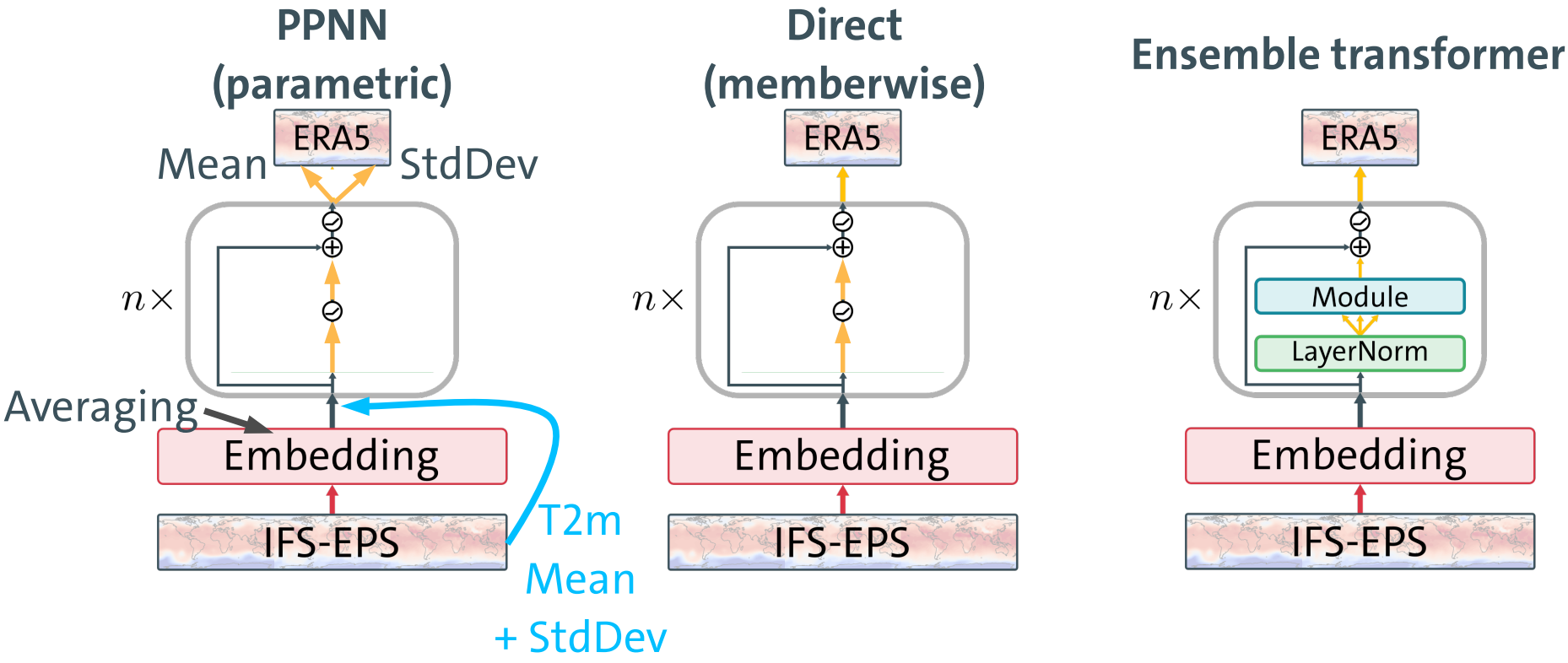


(5,28)



Important areas

Experiment: Used methods



Experiment: Training

2017 & 2013: Training and validation (10 %)
2019: Testing

Optimizer: Adam
Optimized: Univariate continous ranked probability score (CRPS) for Gaussians

Subsampled ensemble members for each training step:

	RMSE (K)	Spread (K)
10 samples	0.91	0.91
20 samples	**0.92**	**0.90**
50 samples	0.92	0.89

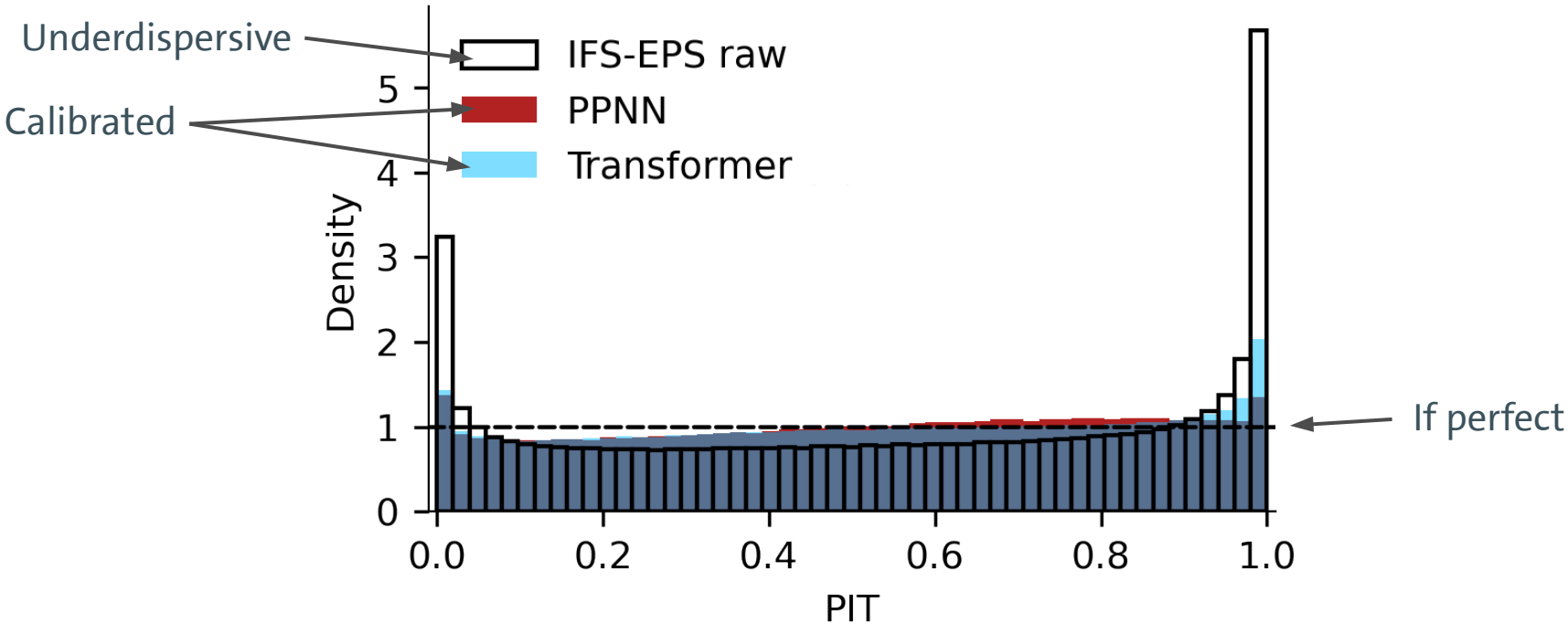
Increased noise during training
+
Increased trainig speed

Transformer has lowest error and best spread-skill ratio

	RMSE (K)	Spread (K)
IFS-EPS	1.12	0.73
“State-of-the-art” Parametric approach → PPNN (Rasp & Lerch, 2013)	0.93	0.87
Apply NN to each member independently → w/o Self-Attention	0.95	0.70
Transformer	**0.90**	**0.90**

Self-attention can extract additional information from ensemble data
+ helps to calibrate the ensemble

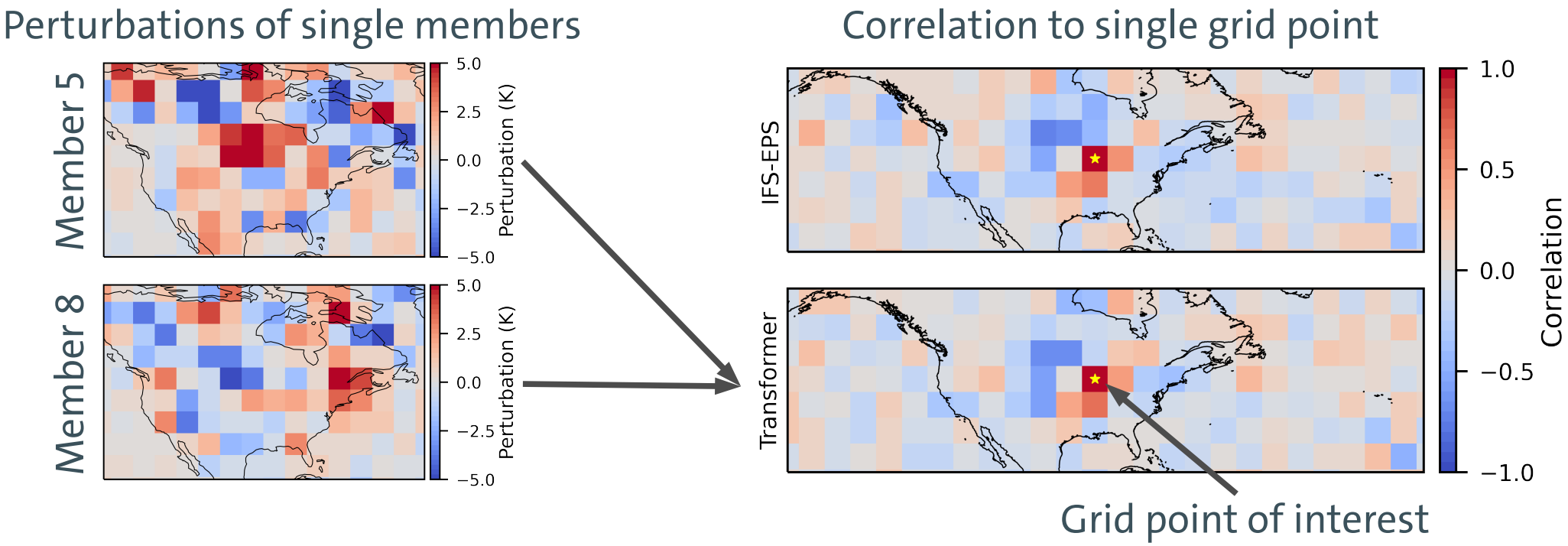
PPNN and Transformer are similarly good calibrated



Similar to a rank histogram – Probability integral transform

Transformer can represent spatial correlations

Cold wave in North America – 2019-01-26 12:00 UTC

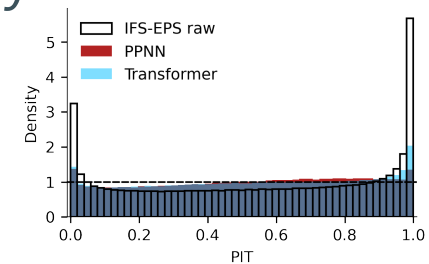


Conclusions of this study

Self-attention can be used to improve post-processing for Earth system models

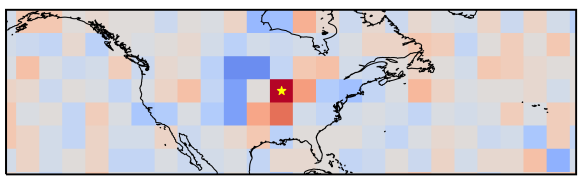
→ Extraction of additional information → Reduced error

→ Calibration of ensemble for improved uncertainty estimation



The ensemble transformer enables member-by-member post-processing with neural networks

→ Non-parametric processing of ensemble members without aggregated statistics



→ Output of spatially and multivariately correlated forecasts

If you have questions

Take a look into the paper:

Self-Attentive Ensemble Transformer: Representing Ensemble Interactions in Neural Networks for Earth System Models

Tobias Sebastian Finn [1 2]

and/or take a look into the official code:
https://github.com/tobifinn/ensemble_transformer

and/or write me an e-mail:
tobias.sebastian.finn@uni-hamburg.de

and/or follow me on twitter:
@tobias_finn

Thank you for your attention!