WeatherBENCH: A Benchmark Dataset For Data-driven Weather Forecasting

Soukayna Mouatadid

University of Toronto

Joint work with **Stephan Rasp & Nils Thuerey** (Technical University of Munich), **Peter D. Dueben** (European Centre for Medium-range Weather Forecasts), **Sebastian Scher** (Stockholm University), **Jonathan A. Weyn** (University of Washington)

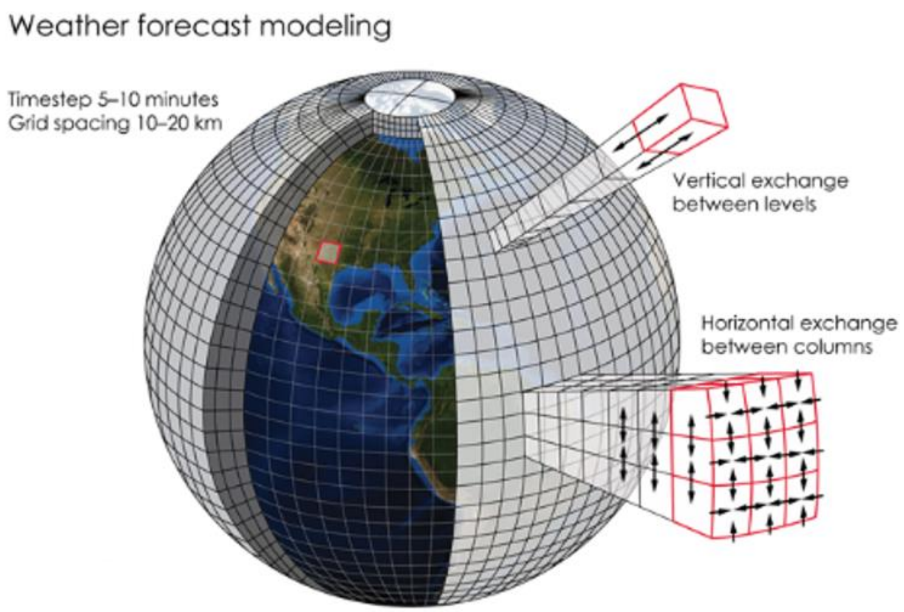
Goals

- **Awareness:** Inter-comparability of machine learning weather forecasting studies
- **Crowdsourced science:** WeatherBench dataset
- **Physics / Machine learning baselines:** numerical weather prediction models, neural network models, etc

How weather forecasting is done today

Traditional weather forecasting involves:

- Observation gathering
- Data assimilation
- Numerical weather prediction
- Forecast post-processing and evaluation

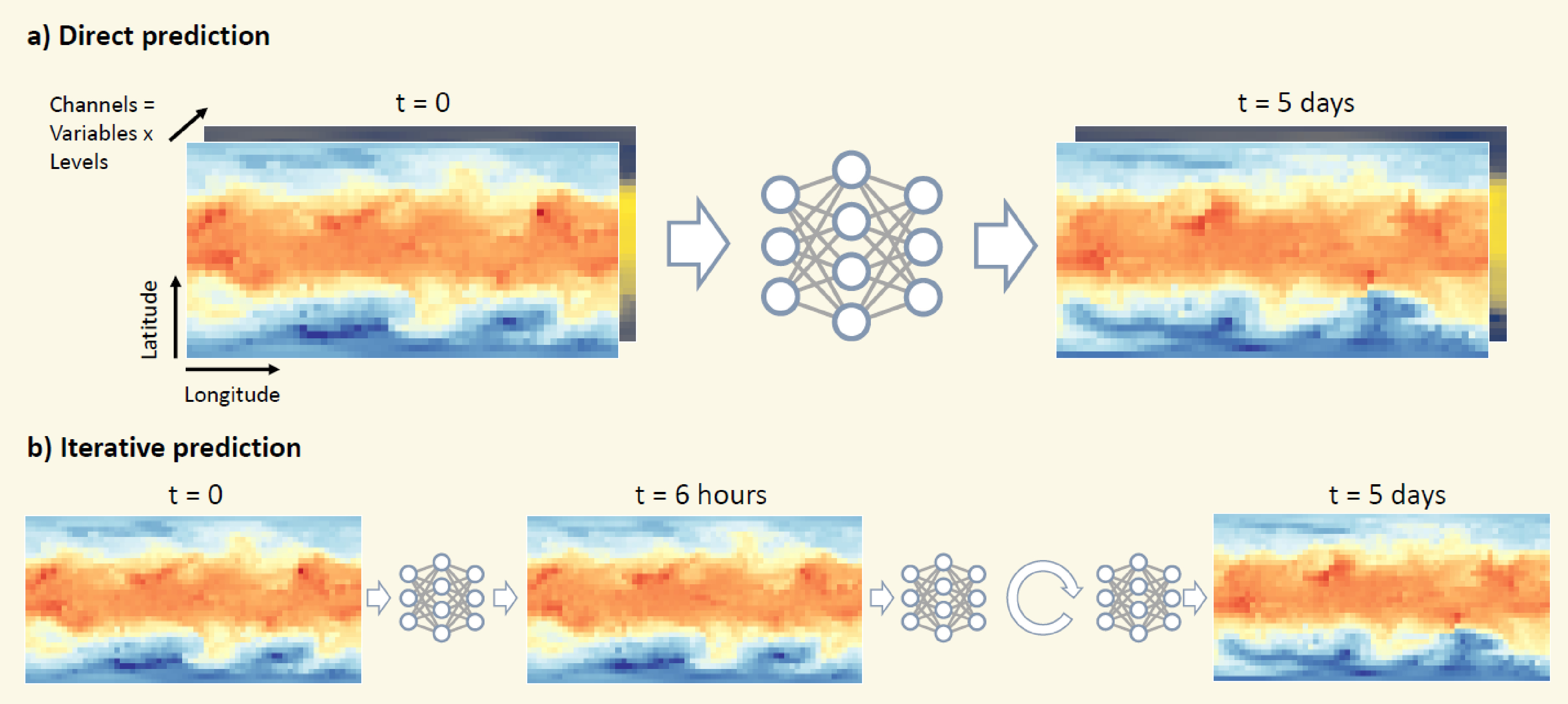


Credit: K. Cantner, AGI.

Concern:

- computationally **expensive**
- Poor performance on **extreme events**

Data-driven weather forecasting



Data-driven weather forecasting: SOTA?

Recent studies:

- NNs to predict 500 hPa geopotential 1 hour ahead (*Dueben and Bauer, 2018*)
- CNNs to predict GCM outputs 14 days ahead (*Scher, 2018; Scher & Messori, 2019*)
- CNNs to predict reanalysis derived Z500 at different lead times (*Weyn et al., 2019*)

Concern:

- different settings of general circulation models as ground truth
- different spatial and temporal resolutions
- different neural network architectures evaluated using different metrics

WeatherBENCH dataset

Goal:	Evaluate deep learning models for global medium range weather forecasting
Data:	ERA5 reanalysis dataset for training and evaluation
Spatial resolution:	40 years of hourly data (1979-2018)
Temporal resolution:	Data re-gridded to 5.625°, 2.8125° and 1.40525° Selected 10 vertical levels between 1 and 1000 hPa

WeatherBENCH dataset

3-D fields	2-D fields	Time-invariant fields
Geopotential	2-meter temperature	Land-sea mask
Temperature	10-meter wind	Soil type
Humidity	Total cloud cover	Orography
Wind	Precipitation	Latitude, longitude
	Top-of-atmosphere incoming solar radiation	

WeatherBENCH evaluation

Target fields: 500 hPa geopotential and 850 hPa temperature

Years: 2017-2018

Resolution: 5.625°

Metric:

$$RMSE = \frac{1}{N_{forecasts}} \sum_{i}^{N_{forecasts}} \sqrt{\frac{1}{N_{lat} N_{lon}} \sum_{j}^{N_{lat}} \sum_{k}^{N_{lon}} L(j)\,(\hat{y}_{i,j,k} - y_{i,j,k})^2}$$

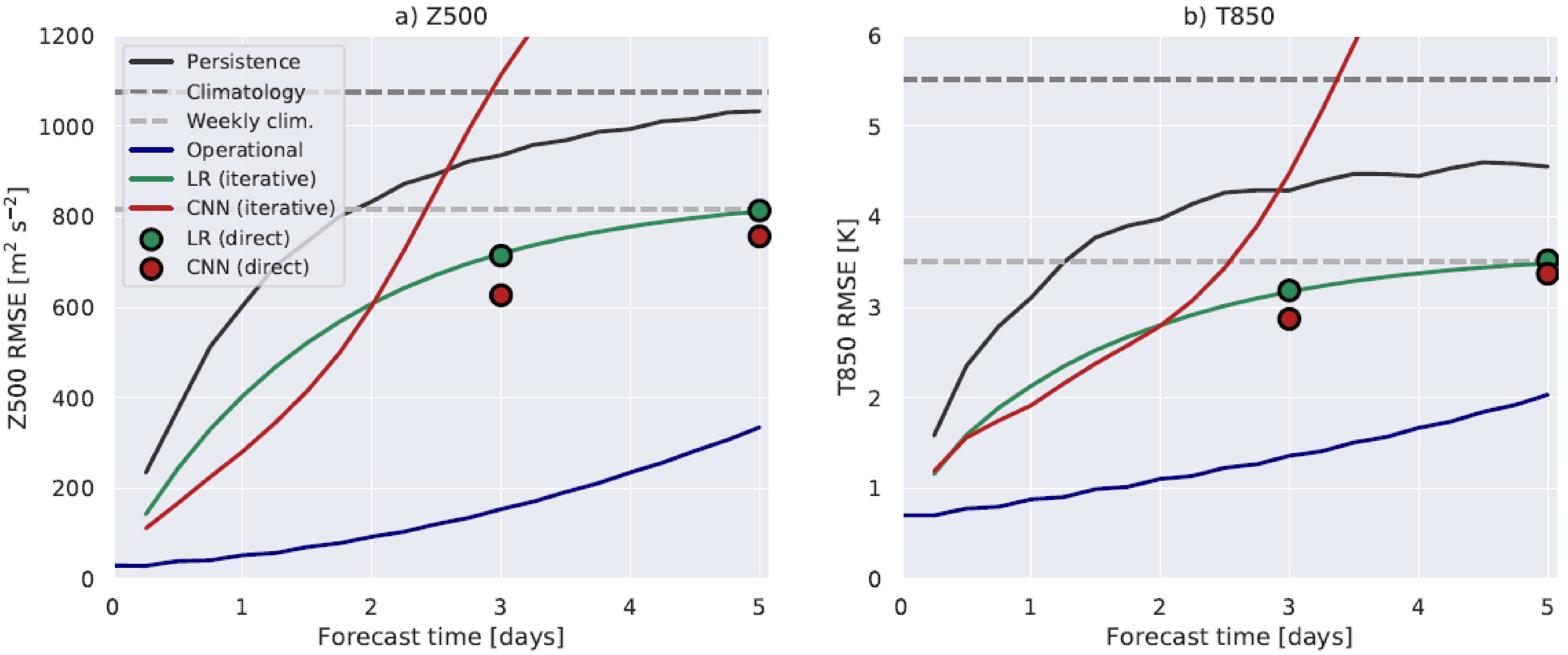
with $L(j)$, the latitude weighting factor for the latitude at the j^{th} latitude index

$$L(j) = \frac{\cos(lat(j))}{\frac{1}{N_{lat}} \sum_{j}^{N_{lat}} \cos(lat(j))}$$

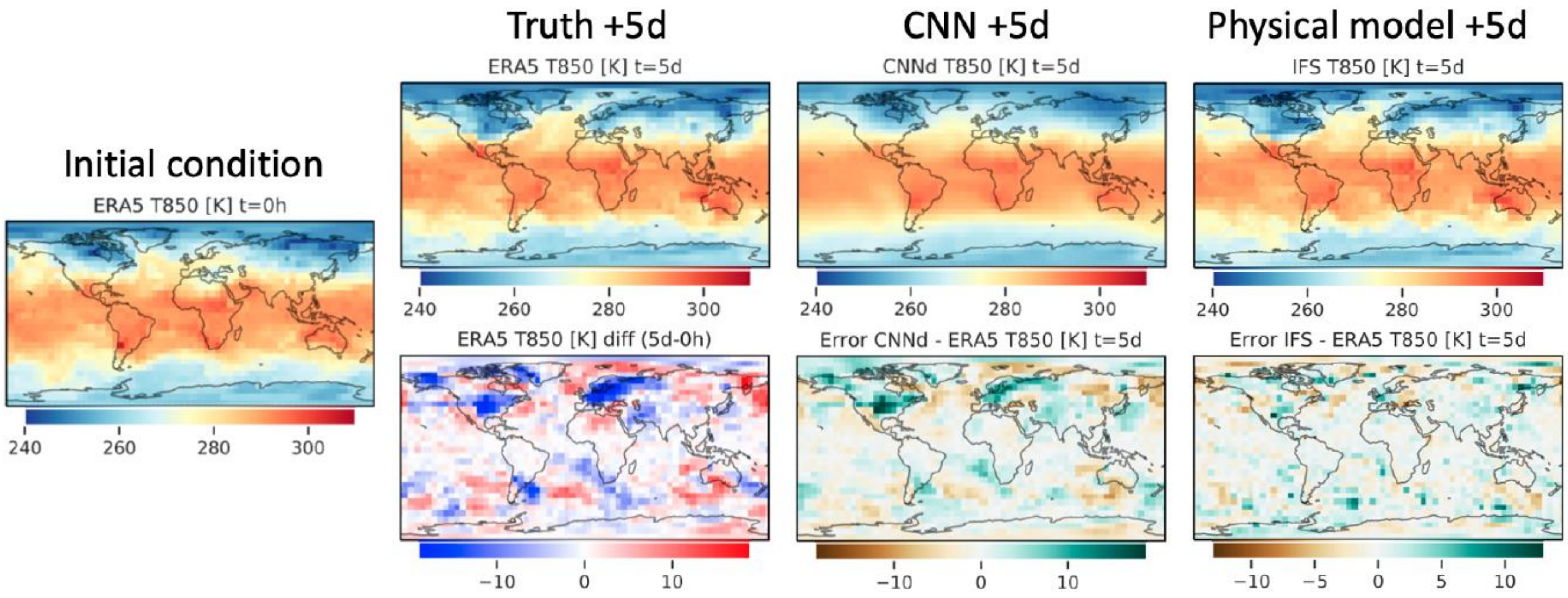
Meaningful baselines

Persistence:	Tomorrow's weather is today's weather
Climatology:	Mean over 1979 – 2016
Operational NWP model:	Operational IFS (Integrated Forecast System) from the ECMWF
Linear regression	
Convolutional neural network:	Five layer CNN with a filter size of 5

Meaningful baselines



Climate forecasts



Conclusion

We hope the benchmark can provide a starting point for:

- Scientific understanding
- Challenge for data science
- Clear metric for success
- Quick start
- Reproducibility and citability
- Communication platform

The end

For more details, see:

WeatherBENCH: A benchmark dataset for data-driven weather forecasting

https://arxiv.org/abs/2002.00469

The benchmark development is ongoing and we encourage you to develop and evaluate your own solutions!

https://mediatum.ub.tum.de/1524895

https://github.com/pangeo-data/WeatherBench

Sources

- Cohen, T. S., Geiger, M., Köhler, J., and Welling, M.: Spherical CNNs, in: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, International Conference on Learning Representations, ICLR, 2018.
- Dueben, P. D. and Bauer, P.: Challenges and design choices for global weather and climate models based on machine learning, Geosci. Model Dev., https://doi.org/10.5194/gmd-2018-148, https://www.geosci-model-dev-discuss.net/gmd-2018-148/gmd-2018-148.pdf, 2018.
- Scher, S.: Toward Data-Driven Weather and Climate Forecasting: Approximating a Simple General Circulation Model With Deep Learning, Geophysical Research Letters, 45, 616–12, https://doi.org/10.1029/2018GL080704, https://onlinelibrary.wiley.com/doi/abs/10.1029/2018GL080704, 2018.
- Scher, S. and Messori, G.: Generalization properties of neural networks trained on Lorenzsystems, Nonlinear Processes in Geophysics Discussions, pp. 1–19, https://doi.org/10.5194/npg-2019-23, https://www.nonlin-processes-geophys-discuss.net/npg-2019-23/, 2019a.
- Scher, S. and Messori, G.: Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground, Geoscientific Model Development, 12, 2797–2809, https://doi.org/10.5194/gmd-12-2797-2019, https://www.geosci-model-dev.net/12/2797/2019/, 2019b.
- Weyn, J. A., Durran, D. R., and Caruana, R.: Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data, Journal of Advances in Modeling Earth Systems, p. 2019MS001705, https://doi.org/10.1029/2019MS001705, https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001705, 2019.