Uncertainty Quantification for Data-driven Weather Models

Large-scale Deep Learning for the Earth System Workshop

Sebastian Lerch, based on joint work with Christopher Bülte, Nina Horat and Julian Quinting | September 2023

From deterministic to probabilistic forecasts

Motivation: Efforts in data-driven weather models have mostly focused on deterministic forecasts, but:

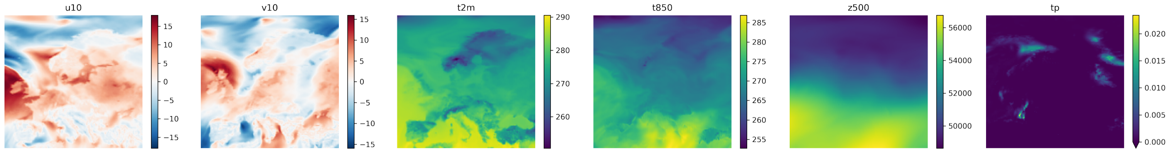
> "Deterministic forecasts are useless."

Uncertainty quantification approaches:

- (probabilistic (e.g. generative) data-driven weather models)
- generation of an ensemble of predictions by running a deterministic data-driven model based on
  - randomly perturbed initial conditions
  - initial conditions of an NWP ensemble model
  - suitably generated initial conditions based on past data
- statistical/AI methods for post-hoc uncertainty quantification to generate probabilistic forecasts from deterministic predictions
  - isotonic distributional regression / EasyUQ
  - neural network-based probabilistic post-processing

Setup

- FourCastNet (+ PrecipNet for precipitation) forecasts, up to a lead time of 192h in 6h steps, 0.25° grid
- 51-member ensemble with initial condition based on random (Gaussian) perturbations as in the original publication; and the initial conditions of the IFS ensemble
- operational ECMWF ensemble as a benchmark
- restriction to sub-domain (ECMWF European Grid) due to limited disk space and computing resources
- 2018–2022 period, with 2018–2021 as training data for the (post-hoc) UQ methods and 2022 as test set
- six target variables



Random field perturbation-based initial conditions

An ensemble of "data-driven" initial conditions is generated by adding random perturbations based on the difference between two randomly chosen atmospheric states (i.e. analyses):

$$\alpha \frac{a_{d_1} - a_{d_2}}{|a_{d_1} - a_{d_2}|_{\text{Etot}}},$$

where a_{d_i} is the state vector of the analysis from the date d_i, $|\cdot|_{\text{Etot}}$ is the total energy norm, and α is a tuning parameter.

Here, we set $\alpha = 0.5 \cdot 10^7$ and select d_1, d_2 randomly from the same month within the 2018–2021 period.

Magnusson, L., Nycander, J. and Källen, E. (2009). **Flow-dependent versus flow-independent initial perturbations for ensemble prediction**, *Tellus A*, 61A, 194–209. DOI:`10.1111/j.1600-0870.2008.00385.x`.

Isotonic distributional regression (IDR) / EasyUQ

EasyUQ, a special case of isotonic distributional regression (IDR; Henzi et al. 2019, arXiv:190903725), aims at transforming real-valued model output into calibrated statistical distributions, based solely on training data of model output-outcome pairs $(x_i, y_i), i = 1, ..., n$.

Key assumption: conditional distributions of the outcome $F_x(y) = P(Y \leq y|X = x)$ are increasing in stochastic order, i.e., $F_x(y) \geq F_{x'}(y)$ if $x \leq x'$.

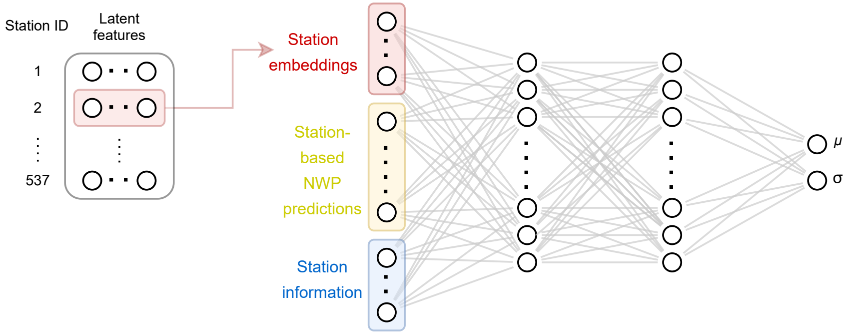
Then, statistically optimal (in final samples) and calibrated predictive distributions are obtained via

$$\hat{F}_{x_j}(y) = \min_{k=1,\ldots,j} \max_{l=j,\ldots,n} \frac{1}{l-k+1} \sum_{i=k}^{l} \mathbb{I}\{y_i \leq y\},$$

without requiring the choice of any tuning parameters.

Walz, E.-M., Henzi, A., Ziegel, J. and Gneiting, T. (2022). **Easy Uncertainty Quantification (EasyUQ): Generating Predictive Distributions from Single-valued Model Output**, Working paper, arXiv:2212.08376.

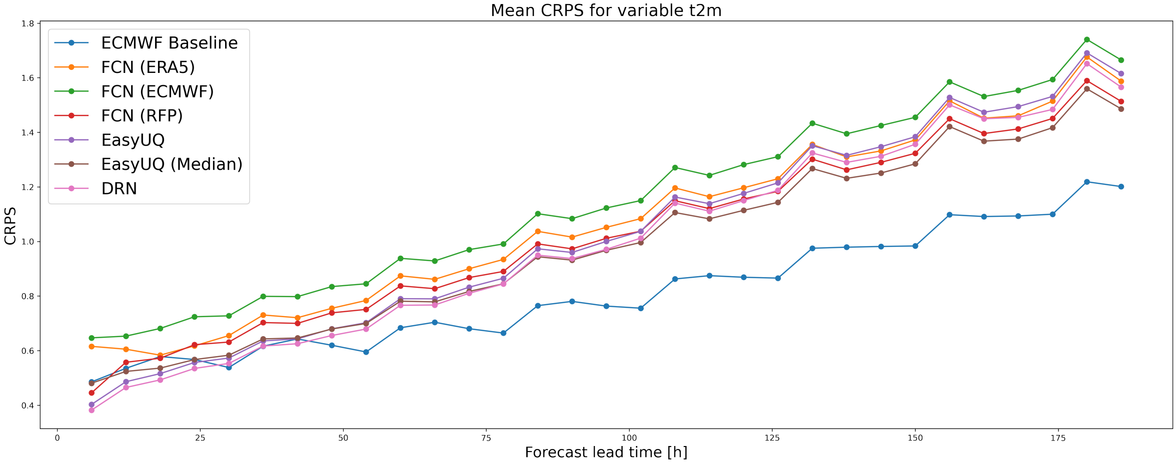
Distributional regression networks (DRN) for probabilistic post-processing



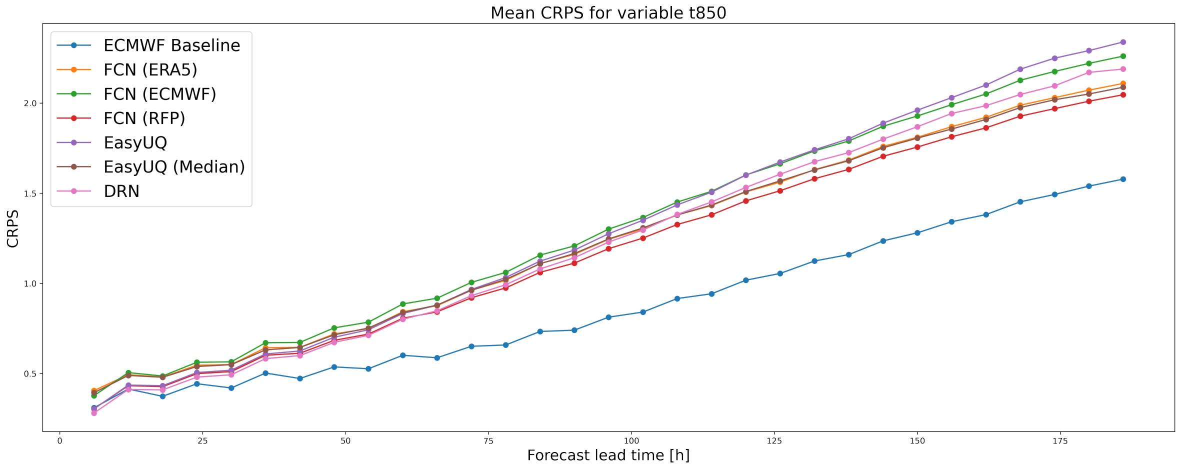
- Input: deterministic FourCastNet predictions, grid point characteristics
- Output: Distribution parameters μ, σ
- Embeddings generate local adaptivity
- CRPS as a mathematically principled loss function

Rasp, S. and Lerch, S. (2018). **Neural networks for post-processing ensemble weather forecasts**, *Monthly Weather Review*, 146, 3885–3900. DOI:10.1175/MWR-D-18-0187.1.

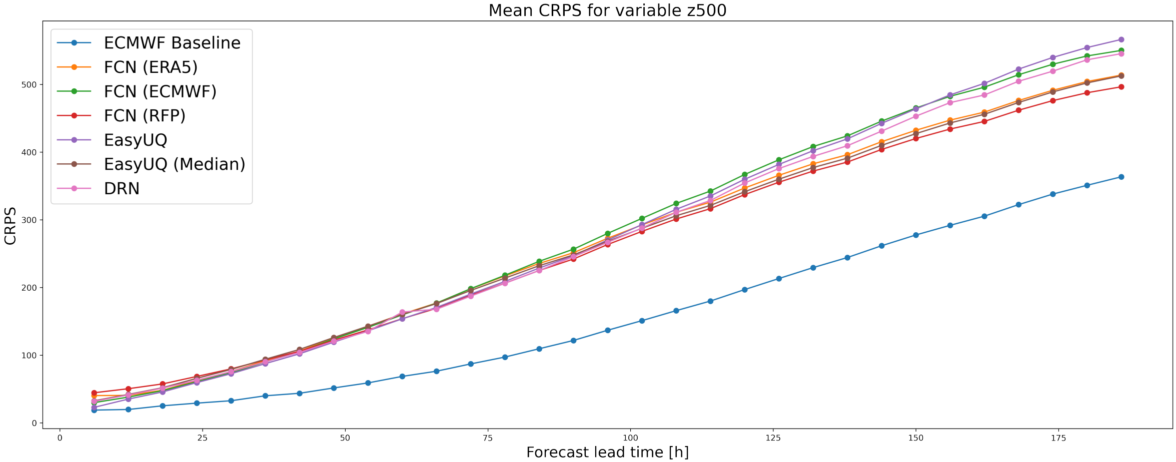
(Preliminary) Results: T2M



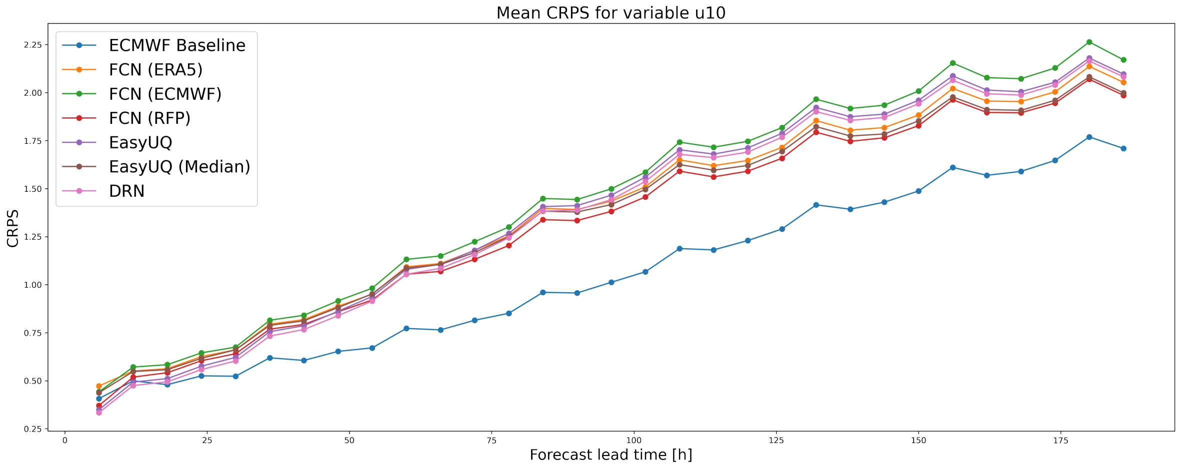
(Preliminary) Results: T850



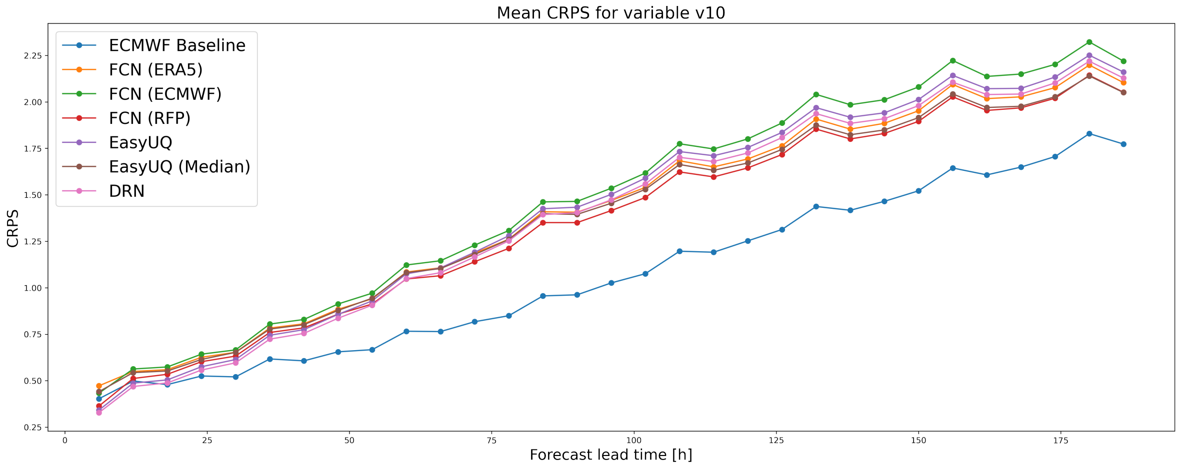
(Preliminary) Results: Z500



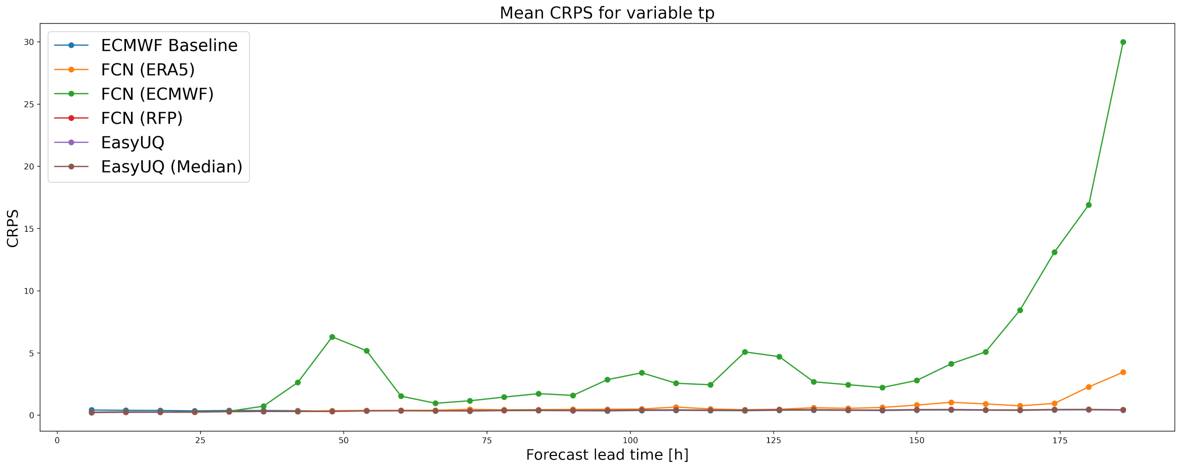
(Preliminary) Results: U10



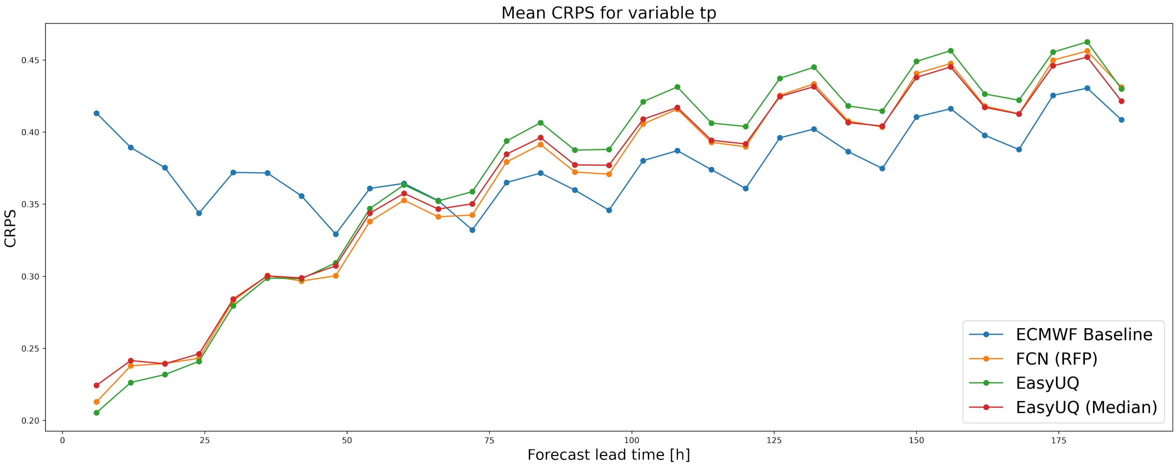
(Preliminary) Results: V10



(Preliminary) Results: TP



(Preliminary) Results: TP



Summary and conclusions

- comparison of various UQ approaches for obtaining probabilistic forecasts from deterministic data-driven weather models
- no single best approach across variables and lead times
- stability issues for precipitation

Next steps and open questions:

- fine-tuning of post-hoc UQ approaches and random field perturbations
- systematic evaluation of forecast quality and reliability
- application for other data-driven weather models, e.g. within WeatherBench 2
- extension/comparison to station observations[1]

Thank you for your attention.

[1] see also related work by Bremnes, Nipen & Seierstad, which will soon be available on arXiv