Neural networks for post-processing ensemble weather forecasts

Stephan Rasp and Sebastian Lerch

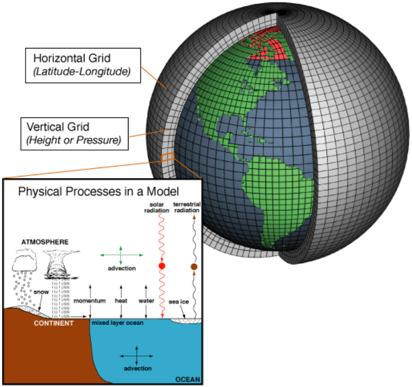
NOAA Workshop on Leveraging AI in Environmental Sciences
October 2020

Karlsruhe Institute of Technology

WAVES TO WEATHER

Numerical weather prediction

Modern weather forecasts rely on physical numerical weather prediction (NWP) models of atmospheric processes.



Source: NOAA[1]

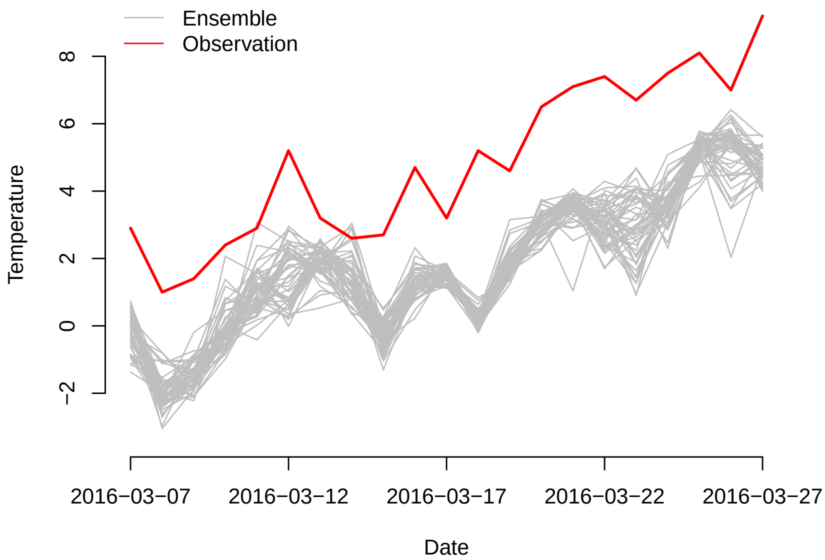
However, there are major sources of uncertainty (initial conditions, physical models).

Ensemble simulations seek to quantify uncertainty and provide probabilistic forecasts.

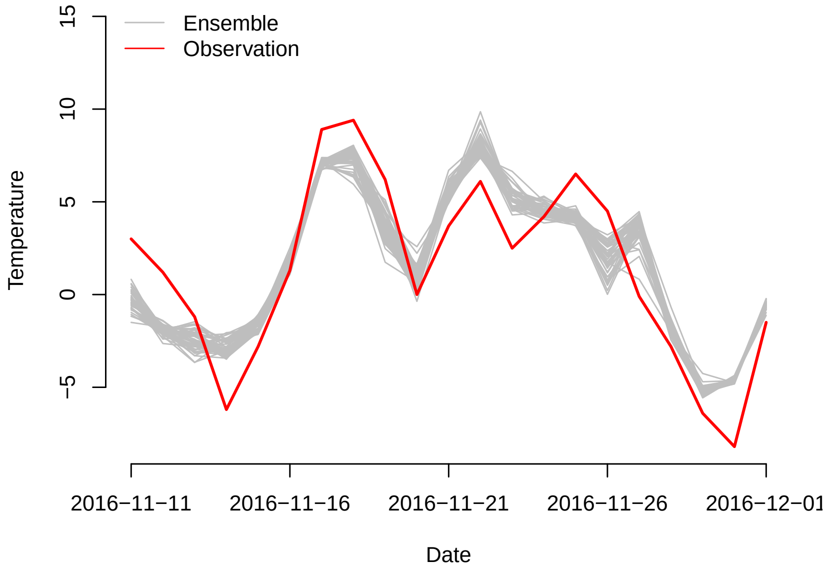
Despite continued improvements ensemble forecasts are subject to systematic errors.

[1] https://celebrating200years.noaa.gov/breakthroughs/climate_model/AtmosphericModelSchematic.png

Example: Ensemble forecasts of temperature



Example: Ensemble forecasts of temperature

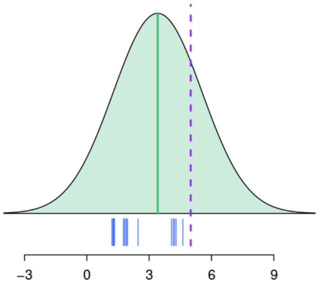


Post-processing with distributional regression models

NWP ensemble forecasts exhibit systematic errors (biases, lack of calibration, ...) that require correction via post-processing.

This is achieved via distributional regression models for statistical post-processing which produce forecast distributions.

Example: EMOS for temperature forecasting



Using ensemble predictions of temperature as input the post-processed forecast takes the form of a Gaussian distribution.

$$y|\mathbf{X}^{\text{t2m}} \sim \mathcal{N}_{(\mu,\sigma)},$$

$$\mu = a + b \cdot \text{mean}(\mathbf{X}^{\text{t2m}})$$
$$\sigma = c + d \cdot \text{sd}(\mathbf{X}^{\text{t2m}})$$

Parametric distributional regression models

Model probability distribution of target variable y given input predictors $\mathbf{X}$ by a parametric distribution $F_{\boldsymbol{\theta}}$,

$$y|\mathbf{X} \sim F_{\boldsymbol{\theta}}, \qquad \text{where} \qquad \boldsymbol{\theta} = g(\mathbf{X})$$

with a link function g connecting predictors $\mathbf{X}$ and distribution parameters $\boldsymbol{\theta}$.

Limitations of fully parametric approaches:

- requires choice of link function g
  - difficult to specify functional form of dependencies if many possible predictors are available
- requires estimation of parameters of g
  - global (using all training data) or local (location-specific) models?
- requires choice of parametric model F_θ

Neural networks for distributional regression

Novel semi-parametric approach: Estimate distribution parameters θ directly as output of a neural network designed to

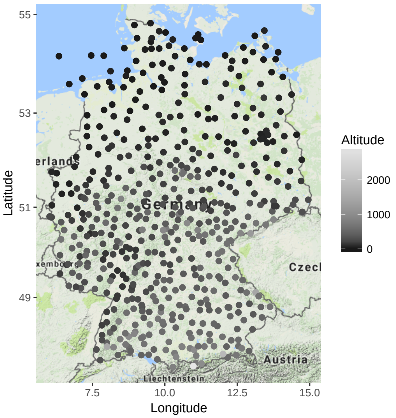
- learn arbitrary nonlinear relations between predictors and distribution parameters in an automated, data-driven manner,
- generate local adaptivity in globally estimated models,
- gain meteorological insight from trained models.

Rasp, S. and Lerch, S. (2018)
Neural networks for post-processing ensemble weather forecasts,
Monthly Weather Review, 146, 3885–3900.

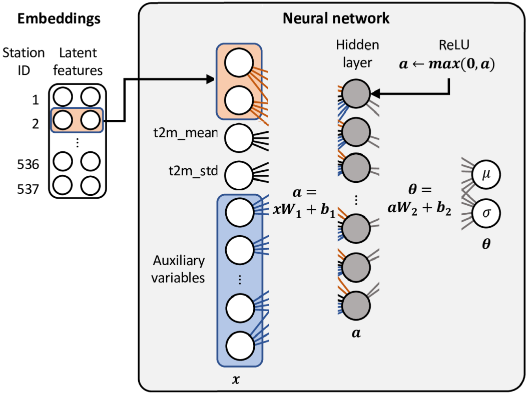
Python/R code available at https://github.com/slerch/ppnn.

Data

- 10 years of forecasts and observations (2007–2016)
- 48 hours-ahead ECMWF 50-member ensemble forecasts of temperature (and 17 other variables)
- station observations at 537 locations
- data from 2016 used as evaluation set
- two training datasets: 2015 and 2007–2015



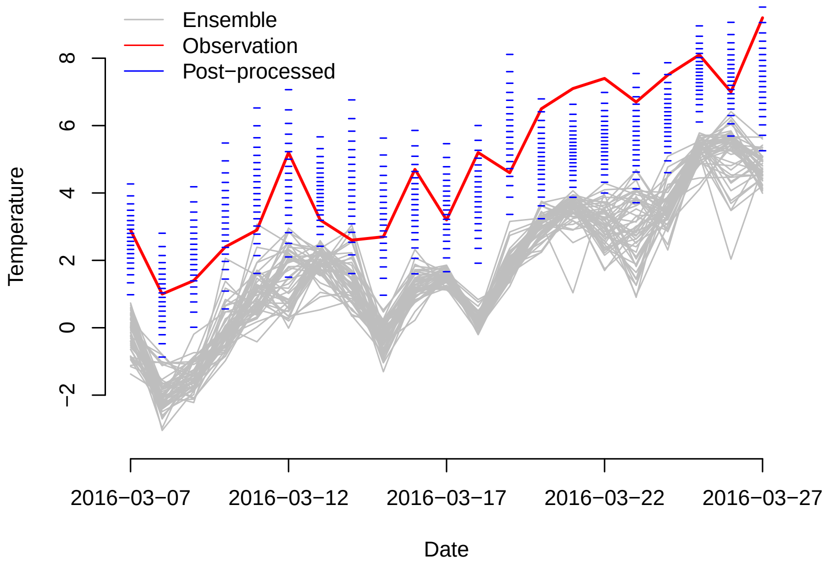
Neural networks for distributional regression



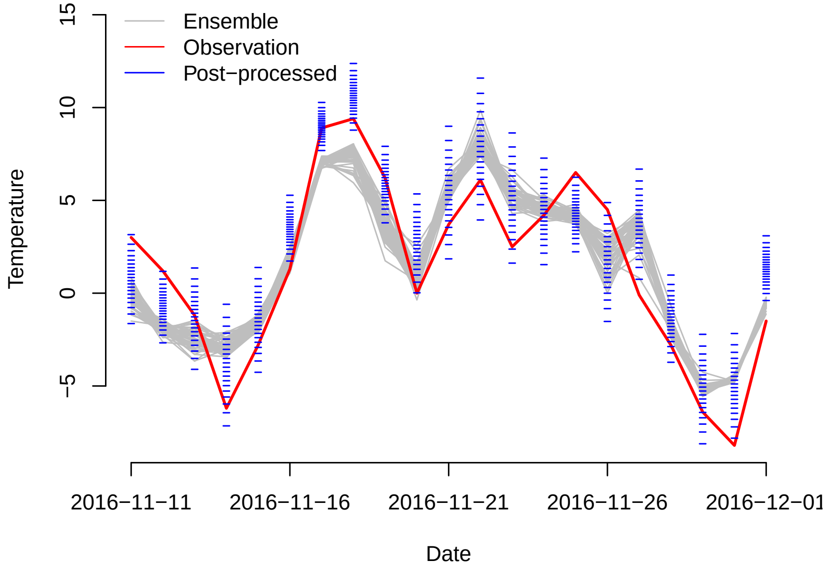
- Input: Predictor variables (NWP quantities, station characteristics).
- Output: Distribution parameters $\boldsymbol{\theta}$
- Embeddings generate local adaptivity.

Training via CRPS minimization (mathematically principled non-standard choice).

Example: Ensemble forecasts of temperature



Example: Ensemble forecasts of temperature



Advanced benchmark methods

- Gradient boosting for EMOS (Messner et al., 2017):
  Let $F_{\boldsymbol{\theta}} = \mathcal{N}_{(\mu,\sigma)}$ and

  $$(\mu, \sigma) = \left(\mathbf{X}^{\mathsf{T}}\boldsymbol{\beta},\ \exp(\mathbf{X}^{\mathsf{T}}\boldsymbol{\gamma})\right),$$

  and iteratively update coefficient vector entries improving the current model fit most.

- Quantile regression forest (Meinshausen, 2006; Taillardat et al., 2016): Nonparametric quantile regression based on random forests. Quantile estimates are obtained from an ensemble of decision trees.

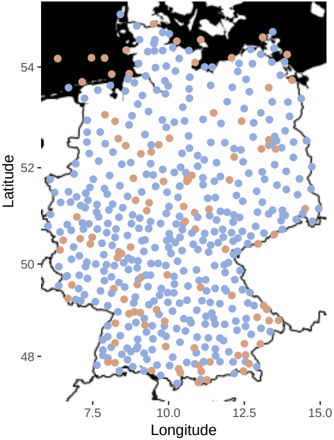
Have to be implemented as local models to achieve good forecasts.

Overview of results

CRPS: Continuous ranked probability score, lower is better

Model	Mean CRPS for training period 2015	Mean CRPS for training period 2007–2015
Raw ensemble	1.16	1.16
Benchmark post-processing methods		
Global EMOS	1.01	1.00
Local EMOS	0.90	0.90
Local EMOS with boosting	0.85	0.80
Local quantile regression forest	0.95	0.81
Neural network models		
Neural network with auxiliary predictors and station embeddings	**0.82**	**0.78**

Station-specific comparison of NN and benchmark models

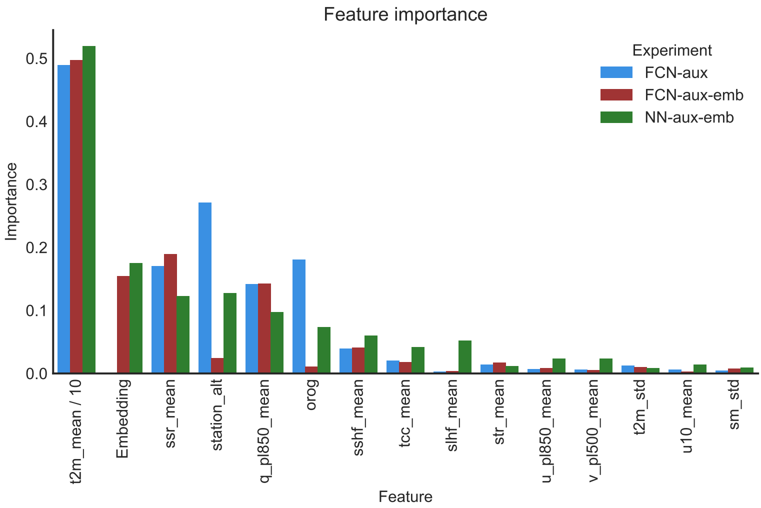


Station-specific best model is a NN model / benchmark model

NN models perform best at more than 80% of the stations.

Differences are statistically significant at a large fraction of stations.

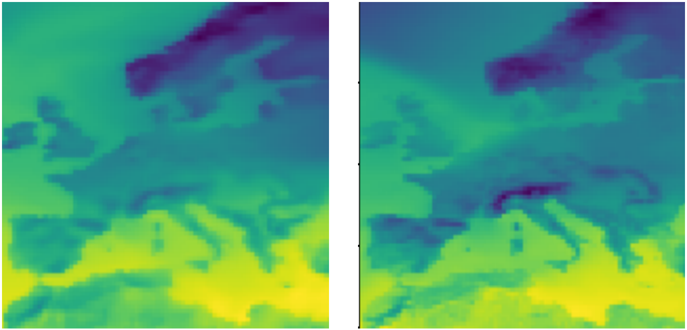
Meteorological interpreation of neural network models



Change in mean CRPS after permuting a single input variable according to a random permutation across stations and dates.

Challenges: Incorporating spatial information

Ensemble forecasts are gridded 2D fields of forecasts of weather variables. Thus far, those were interpolated to station locations.

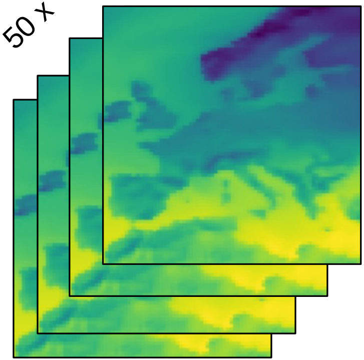


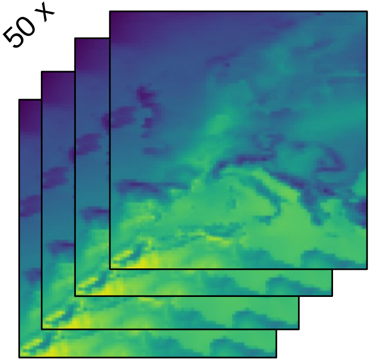
Gridded ECMWF forecasts over Europe (0.5° resolution, 81×81 pixels)

However, large-scale spatial structure and predictability information (e.g., ‘weather regimes’) get lost in the interpolation step.

Challenges: Incorporating ensemble information

Ensemble members provide 50 physically coherent forecasts of weather variables. Thus far, only mean and standard deviation of (interpolated) ensemble forecasts were used.





Possibly important uncertainty information might get lost by the use of summary statistics.

Job advertisements

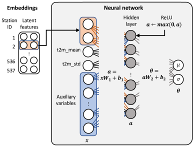
I am looking for PhD students to develop AI methods for post-processing and probabilistic weather forecasting.

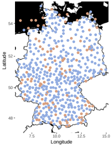
Backgrounds in mathematics, computer science and/or atmospheric sciences welcome!

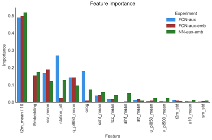
Starting dates around mid-2021.

For details, contact me at Sebastian.Lerch@kit.edu.

Summary







- flexible, automated and data-driven modelling of nonlinear relations between predictors and distribution parameters
- perform better than state of the art approaches
- surprisingly computationally efficient and scale well
- gain meteorological insight from trained models

Rasp, S. and Lerch, S. (2018)
Neural networks for post-processing ensemble weather forecasts,
Monthly Weather Review, 146, 3885–3900.

Python/R code available at https://github.com/slerch/ppnn.