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

Concern:

- computationally expensive
- Poor performance on extreme events
Data-driven weather forecasting

a) Direct prediction

Channels = Variables x Levels

b) Iterative prediction

Soukayna Mouatadid / University of Toronto
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

<table>
<thead>
<tr>
<th>3-D fields</th>
<th>2-D fields</th>
<th>Time-invariant fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geopotential</td>
<td>2-meter temperature</td>
<td>Land-sea mask</td>
</tr>
<tr>
<td>Temperature</td>
<td>10-meter wind</td>
<td>Soil type</td>
</tr>
<tr>
<td>Humidity</td>
<td>Total cloud cover</td>
<td>Orography</td>
</tr>
<tr>
<td>Wind</td>
<td>Precipitation</td>
<td>Latitude, longitude</td>
</tr>
<tr>
<td></td>
<td>Top-of-atmosphere incoming solar radiation</td>
<td></td>
</tr>
</tbody>
</table>
WeatherBENCH evaluation

Target fields: 500 hPa geopotential and 850 hPa temperature
Years: 2017-2018
Resolution: 5.625°

Metric:

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

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

Graph a) Z500:
- Persistence
- Climatology
- Weekly clim.
- Operational
- LR (iterative)
- CNN (iterative)
- LR (direct)
- CNN (direct)

Graph b) T850:
- Z500 RMSE [m²/s²]
- T850 RMSE [K]

Forecast time [days]:
0 1 2 3 4 5

April 26, 2020
Soukayna Mouatadid / University of Toronto
Climate forecasts

Initial condition

Truth +5d

CNN +5d

Physical model +5d
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

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


