Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

AI4ES Journal Club

Presenter: Carlos Gomez Gonzalez

Group: CES

2020/03/03

WeatherBench

WeatherBench: A benchmark dataset for data-driven weather forecasting

Stephan Rasp[1], Peter D. Dueben[2], Sebastian Scher[3], Jonathan A. Weyn[4], Soukayna Mouatadid[5], and Nils Thuerey[1]

[1]Technical University of Munich, Germany
[2]European Centre for Medium-range Weather Forecasts, Reading, UK
[3]Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, Sweden
[4]Department of Atmospheric Sciences, University of Washington, Seattle, USA
[5]Department of Computer Science, University of Toronto, Canada

BSC
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Context

- DL is a family of algorithms driving the recent revolution in many fields (computer vision, language understanding, etc)
- Data is a big part of the success of DL. Physical sciences lack training/benchmark datasets
- First effort to create a community dataset for medium-range weather forecasting, basic metrics and baseline algorithms

Methods: data

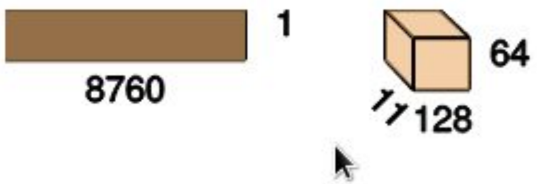
- Variables from ERA5 archive, preprocessed for the direct application of ML (3.81 TB in total)

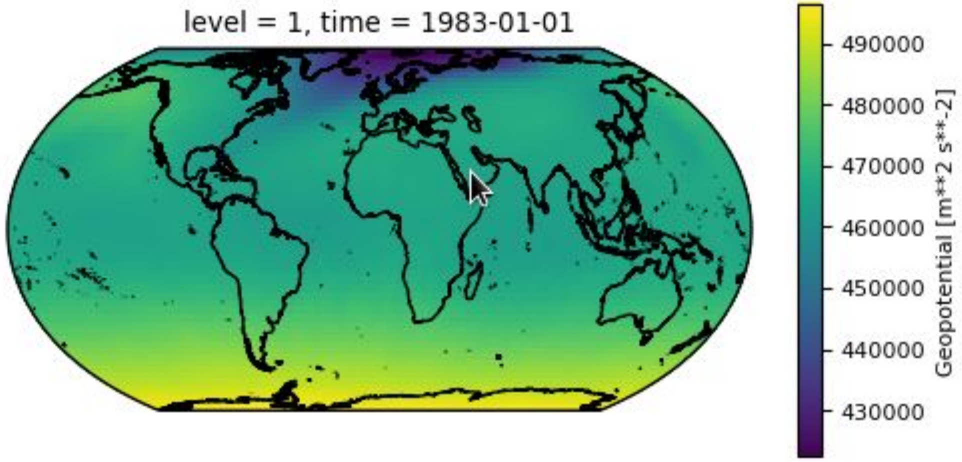
    - 40 years of 1hourly data
    - 3 spatial resolutions: 5.625° (32×64 grid points), 2.8125° (64×128 grid points) and 1.40525° (128×256 grid points)
    - 3D fields with 11 selected vertical levels (1, 10, 100, 200, 300, 400, 500, 600, 700, 850 and 1000 hPa)

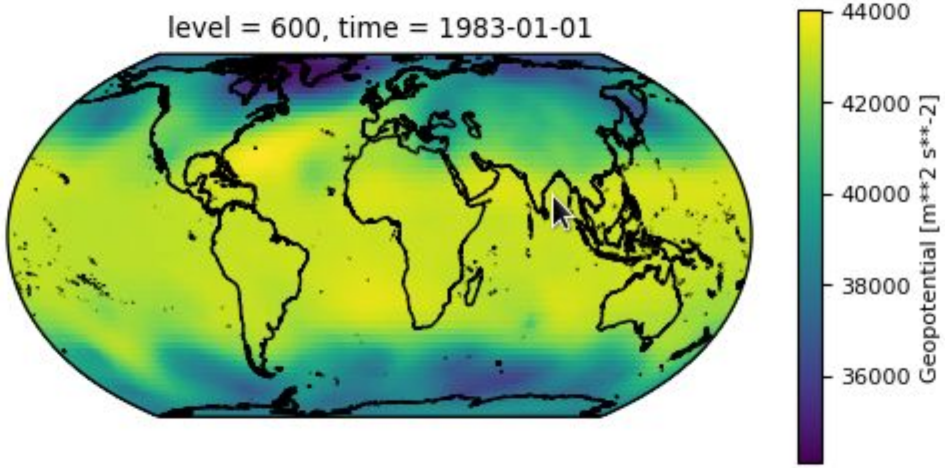
Methods: data

1y NetCDF file

	Array	Chunk
Bytes	3.16 GB	360.45 kB
Shape	(8760, 11, 64, 128)	(1, 11, 64, 128)
Count	8761 Tasks	8760 Chunks
Type	float32	numpy.ndarray







Methods: metric and baselines

- Metric: mean latitude-averaged RMSE over all forecasts. The anomaly correlation coefficient might be more adequate for other variables

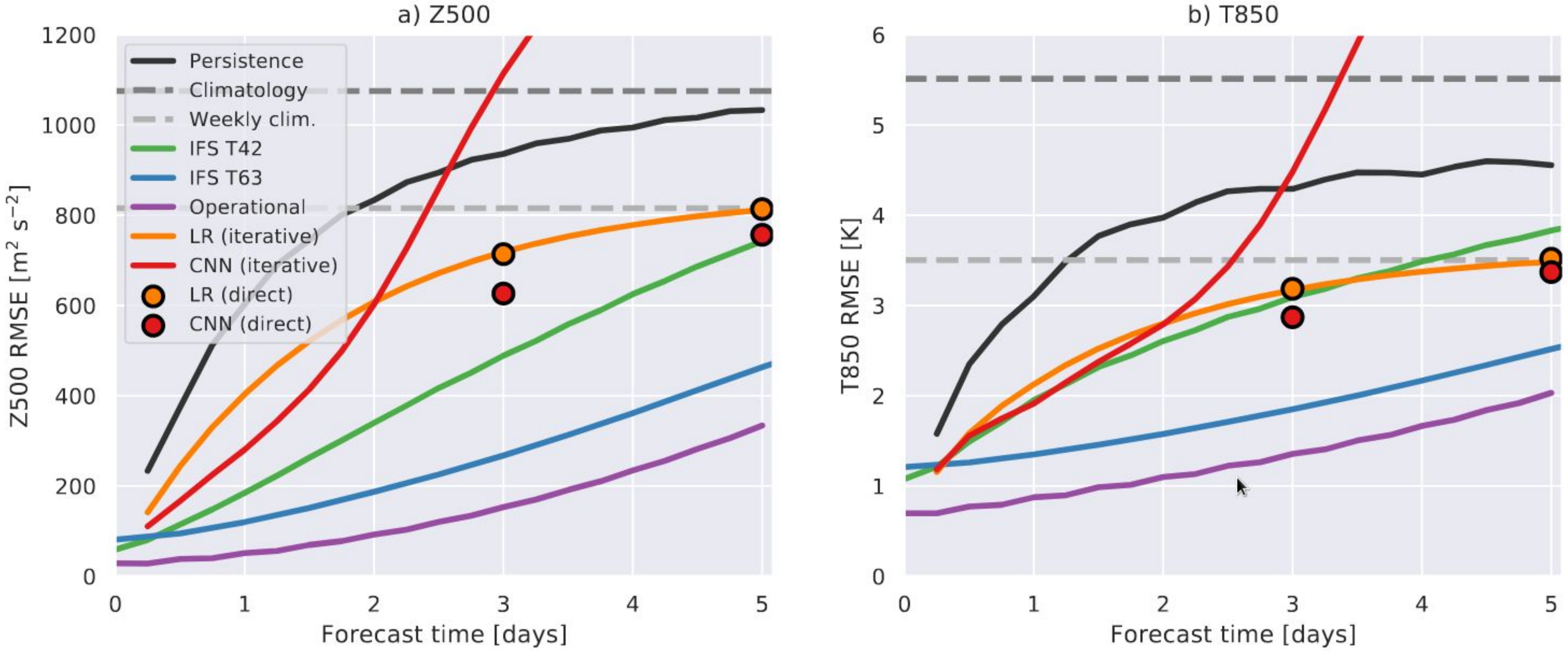
$$\text{RMSE} = \frac{1}{N_{\text{forecasts}}} \sum_{i}^{N_{\text{forecasts}}} \sqrt{\frac{1}{N_{\text{lat}} N_{\text{lon}}} \sum_{j}^{N_{\text{lat}}} \sum_{k}^{N_{\text{lon}}} L(j) (f_{i,j,k} - t_{i,j,k})^2} \qquad L(j) = \frac{\cos(\text{lat}(j))}{\frac{1}{N_{\text{lat}}} \sum_{j}^{N_{\text{lat}}} \cos(\text{lat}(j))}$$

- Comparison of baselines, wrt a persistence and a climatological forecast and the gold standard of medium-range NWP (IFS, 2017/2018 forecasts from TIGGE)
  - Operational forecasting: ~12k cores and 1h (10 day forecast at 10km)
  - Coarser resolution NWP much faster, only a few minutes. This should be the target to beat

Methods: baselines and dd models

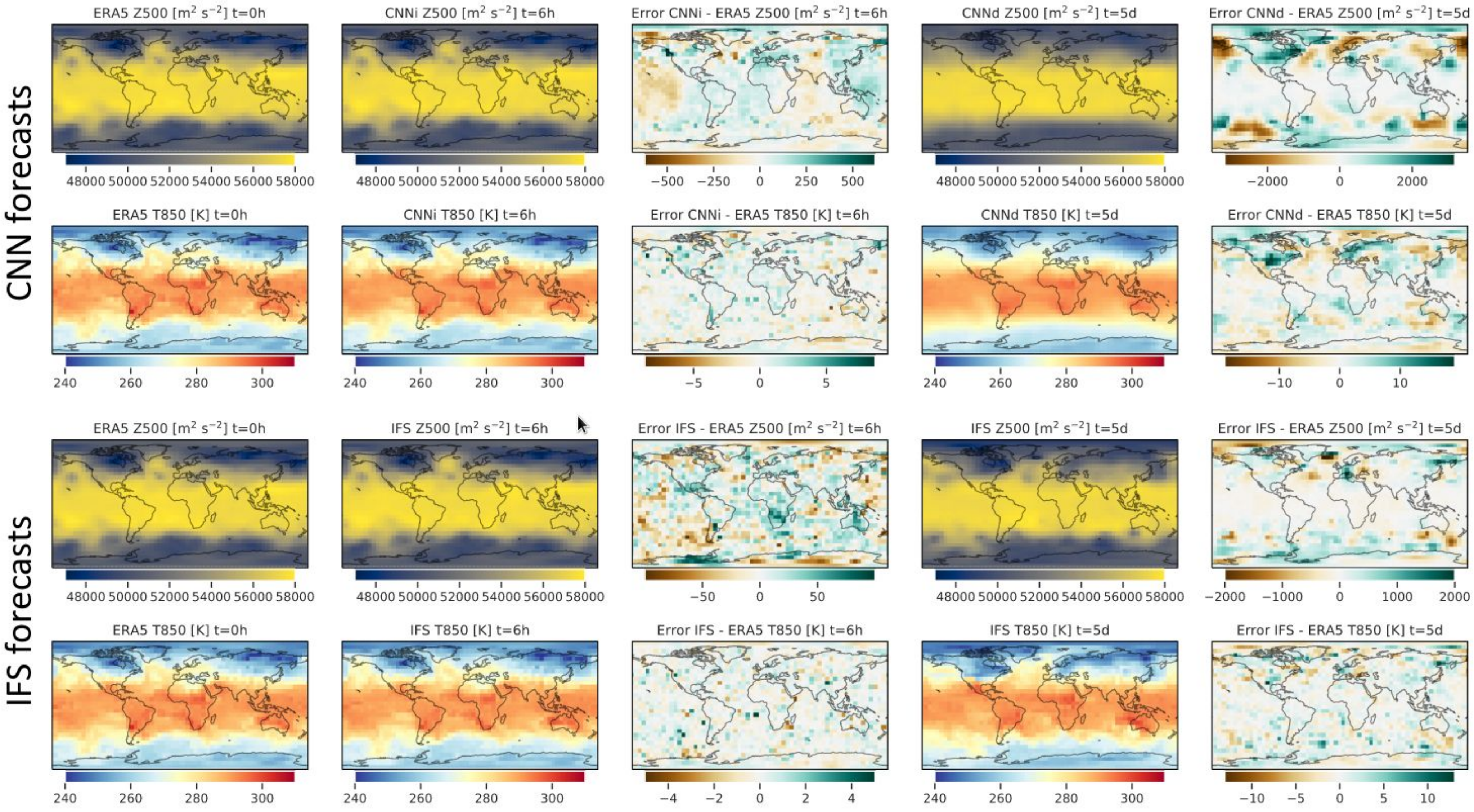
- Two variables chosen for the tests: Z500 and T850
- Persistence forecast: fields at initialization time are used as forecasts (tomorrows weather is todays weather)
- Climatology: two were computed for the training data (1979-2016): a mean over all data, and a weekly one
    - *A useful forecast must beat the weekly climatology and the persistence forecast*
- Linear Regression: data flattened and concatenated. Iterative forecasts (single model takes its previous output as input for the next step)
- CNN: simple model exploiting the translational invariances in images/fields (two channels t850 and z500). 5 Conv layers, with 64 channels, ksize=5, ELU activations, Adam optimizer, MSE loss, 313k trainable params, periodic convolutions in longitude

Results



Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Results



CNN forecasts
ERA5 Z500 [m² s⁻²] t=0h
CNNi Z500 [m² s⁻²] t=6h
Error CNNi - ERA5 Z500 [m² s⁻²] t=6h
CNNd Z500 [m² s⁻²] t=5d
Error CNNd - ERA5 Z500 [m² s⁻²] t=5d
48000 50000 52000 54000 56000 58000
−500 −250 0 250 500
−2000 0 2000
ERA5 T850 [K] t=0h
CNNi T850 [K] t=6h
Error CNNi - ERA5 T850 [K] t=6h
CNNd T850 [K] t=5d
Error CNNd - ERA5 T850 [K] t=5d
240 260 280 300
−5 0 5
−10 0 10
IFS forecasts
ERA5 Z500 [m² s⁻²] t=0h
IFS Z500 [m² s⁻²] t=6h
Error IFS - ERA5 Z500 [m² s⁻²] t=6h
IFS Z500 [m² s⁻²] t=5d
Error IFS - ERA5 Z500 [m² s⁻²] t=5d
48000 50000 52000 54000 56000 58000
−50 0 50
−2000 −1000 0 1000 2000
ERA5 T850 [K] t=0h
IFS T850 [K] t=6h
Error IFS - ERA5 T850 [K] t=6h
IFS T850 [K] t=5d
Error IFS - ERA5 T850 [K] t=5d
240 260 280 300
−4 −2 0 2 4
−10 −5 0 5 10

Conclusions

- State-to-state weather prediction similar to image-to-image translation, lots of recent developments in CV
- 3D structure needs to be exploited, data is extremely rich/complex
- More complex networks could be used, but few data points (overfitting)
    - Augmentation not easy
    - U-Nets, Resnets, CGANs, etc
- Technical challenges: GPU memory!
- Interesting venues:
    - Extreme events
    - Climate simulations

Critique

- Data in an standardized format
- Pangeo data repo as a communication platform (better a challenge platform?)
- Good comparison of the relevant literature
- Baseline implementations provided (simple algorithms)
    - 3D structure not taken into account
- Metric is convenient and simple (maybe too basic?)