A BENCHMARK DATASET FOR STATISTICAL DOWNSCALING OF METEOROLOGICAL FIELDS WITH DEEP NEURAL NETWORKS

AMS 2024 - 104th Annual Meeting | Baltimore

2024-01-29 I **MICHAEL LANGGUTH**[1], SCARLET STADTLER[1], BING GONG[2], MARTIN G. SCHULTZ[1] (ET AL..)

[1] Juelich Supercomputing Centre (JSC), [2] JSC, now at Shanghai Normal University (SHNU)

"The MAELSTROM project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955513. The JU receives support from the European Union's Horizon 2020 research and innovation programme and United Kingdom, Germany, Italy, Luxembourg, Switzerland, Norway".

EuroHPC Joint Undertaking

Mitglied der Helmholtz-Gemeinschaft

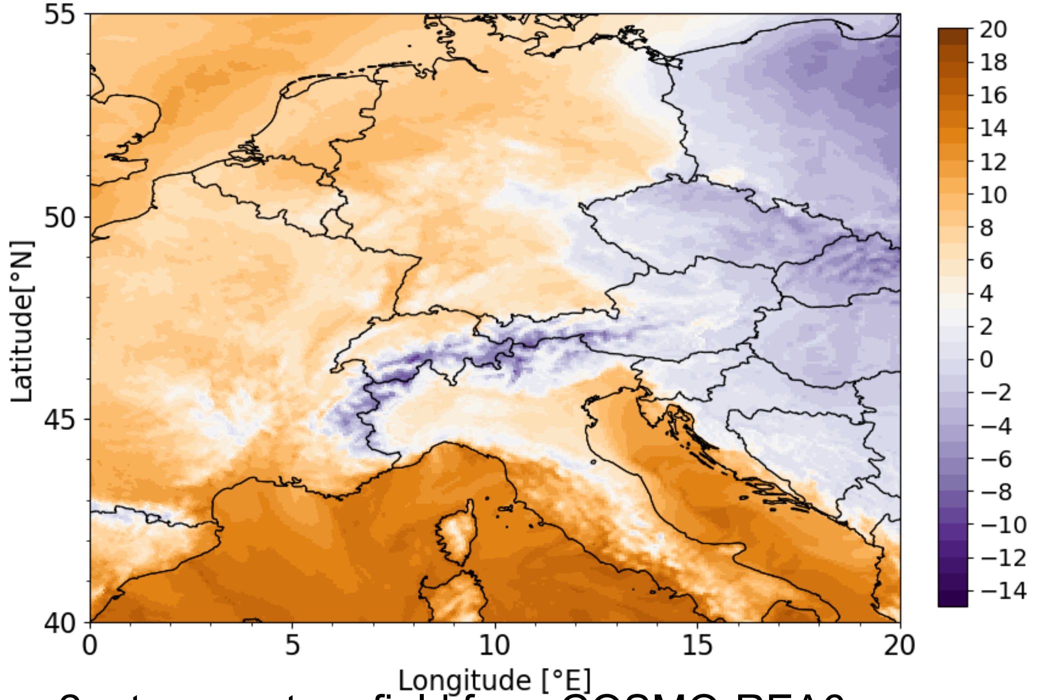
JÜLICH Forschungszentrum

MOTIVATION

Why do we need a benchmark dataset?

- High-resolved weather data is crucially demanded
- Recent success of deep learning for statistical downscaling (e.g. *Mardani et al., 2023*, *Harder et al., 2023*, *Harris et al., 2022*)
- Intercomparison difficult:
  - Variety of downscaling t
  - Different dataset
  - Different eva
- Benchmark datas in AI, e.g. ImageNet (*Deng et* GLUE (*Wang et al., 2018*)
- For meteorological a ications, benchmarks such as WeatherBench 2 (*Rasp et al., 2023*) are rare

A benchmark dataset to steer progress in statistical downscaling



2m temperature field from COSMO-REA6 reanalysis ($\Delta x = 0.055°$) on 2018-12-01, 13:00 CET.

THE BENCHMARK DATASET

Design choices and components of the benchmark dataset

- Benchmark dataset closely follows requirements listed in *Dueben et al., 2022*
  1) Clear problem statement for real-life task
  2) Open data provision in high-level programming language
  3) Results and code of baseline competitor models
  4) Evaluation metrics defined
  5) Visualization and diagnostics in code

1) Downscaling task(-s):

Emulate a highly-resolved reanalysis (COSMO REA6) of i) ***2m temperature***, ii) ***100m wind*** and iii) ***global horizontal irradiance*** from the ERA5-reanalysis

THE BENCHMARK DATASET

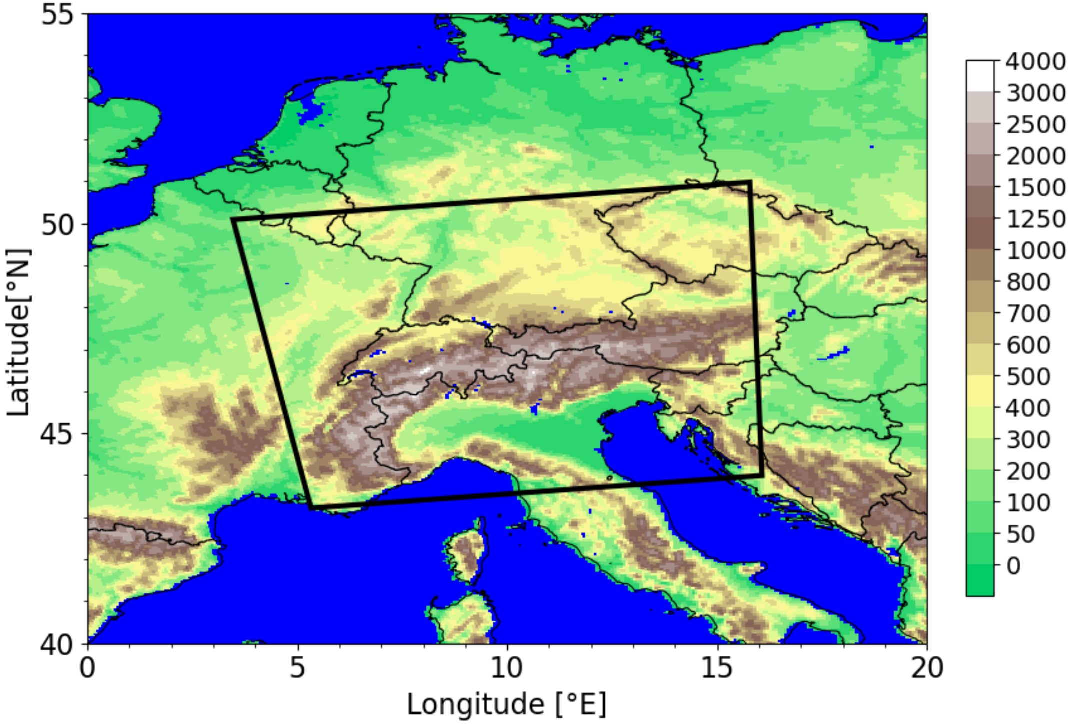
Design choices and components of the benchmark dataset

2) The data:

- Predictands from COSMO-REA6 data ($\Delta x^{rot}_{CREA6} = 0.055°$)
- Task-specific set of predictor variables from the ERA5-reanalysis dataset ($\Delta x_{ERA5} = 0.25°$)
- Data pairing after re-projection of ERA5-data onto rotated pole grid ($\Delta x^{rot}_{ERA5} = 4\Delta x^{rot}_{CREA6} = 0.225°$)
- Provision via climetlab-plugin → conversion to xarray.Dataset and netCDF

```
!pip install climetlab climetlab_downscaling_benchmark
import climetlab as cml
cml_ds = cml.load_dataset("t2m_downscaling", dataset="validation")

ds = cml_ds.to_xarray()
ds.to_netcdf("downscaling_benchmark_t2m_val.nc")
```



Surface topography form COSMO REA6. The target domain of the downscaling benchmark comprises 144x128 grid points and is rendered.

THE BENCHMARK DATASET

Design choices and components of the benchmark dataset

3) Baseline competitor models

a) Deep neural networks:
- U-Net by *Sha et al., 2020* (tuned)
- DeepRU (U-Net variant) by *Höhlein et al., 2020*
- WGAN with U-Net by Sha as generator
- WGAN by *Harris et al., 2022*
- SwinIR by *Liang et al., 2021*

b) Classical statistical model:
- Standardized Anomaly MOS (SAMOS) by *Dabernig, 2017*

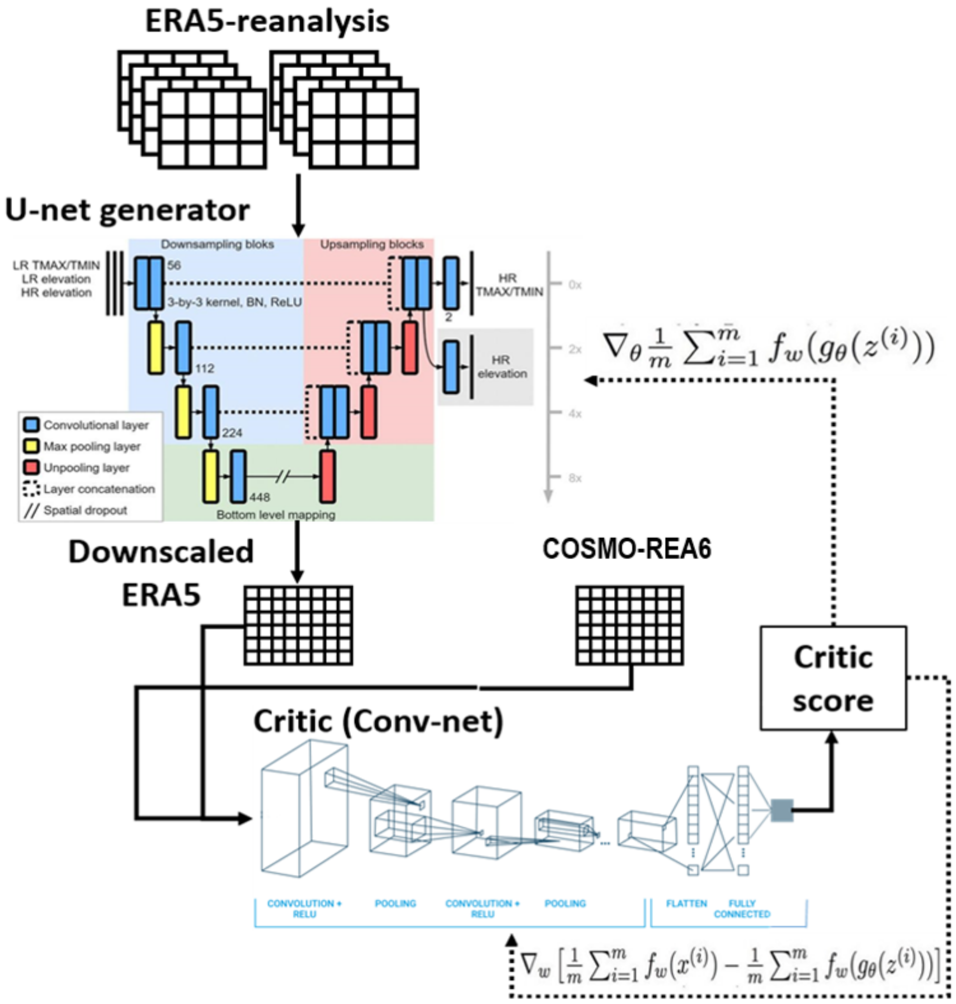


Illustration of the WGAN with the U-Net by Sha et al. (2020) as generator.

THE BENCHMARK DATASET

Design choices and components of the benchmark dataset

4&5) Evaluation metrics and diagnostics

- Task-specific set of evaluation metrics
- Diagnostics for marginal distribution, e.g. power spectra and histograms
- Various plot products
- Two postprocessing steps:
  1) Single model evaluation
  2) Intercomparison (in terms of skill scores)

SOME PRELIMINARY RESULTS

Examples from the 2m temperature downscaling task

Dynamic predictors (ERA5 only):

- 2m temperature
- Temperature at model levels 137, 135, 131, 127 and 122
- 10m (u,v)-wind
- Surface pressure
- Surface latent and sensible heat fluxes
- Boundary layer height

Static predictors (ERA5 & COSMO REA6):

- Surface topography
- Land-sea mask

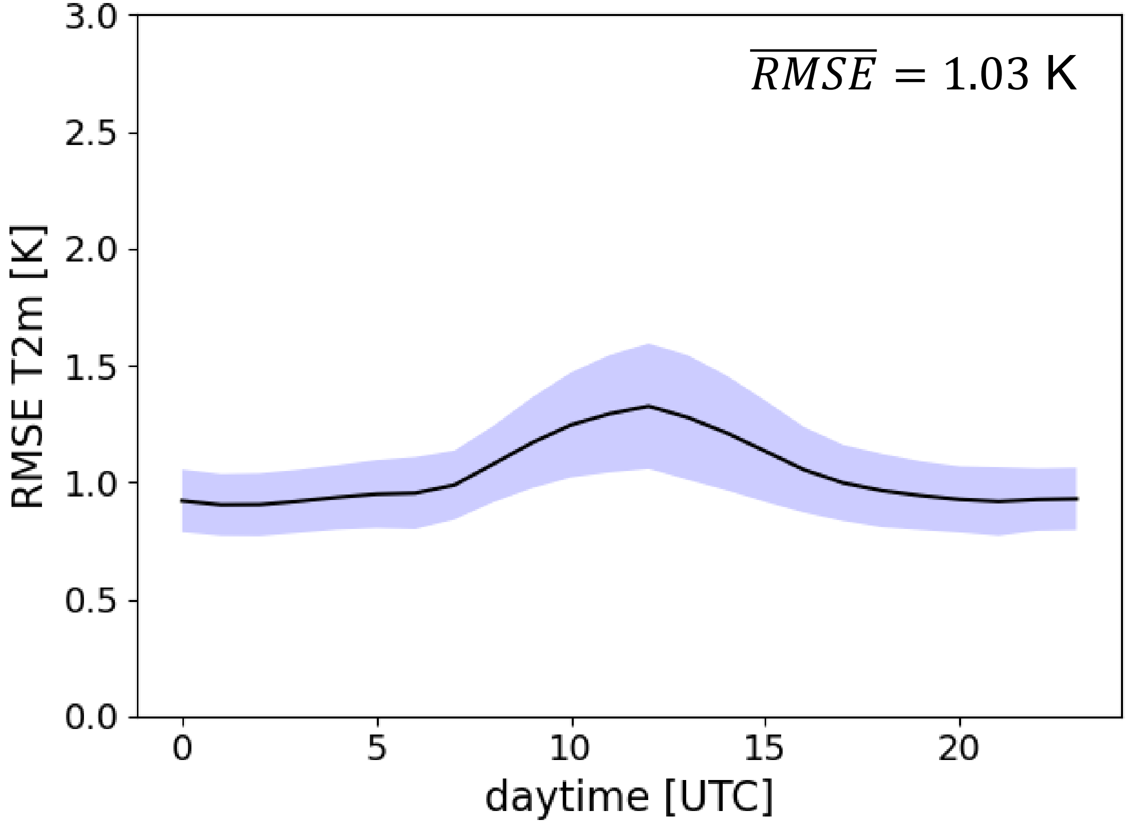
Evaluation metrics and diagnostics:

- RMSE
- Bias
- Mean Error of local standard deviation
- Average gradient amplitude error
- Energy spectra analysis
- Conditional Quantile Plots

SOME PRELIMINARY RESULTS

Examples from the 2m temperature downscaling task

Results from
Sha WGAN



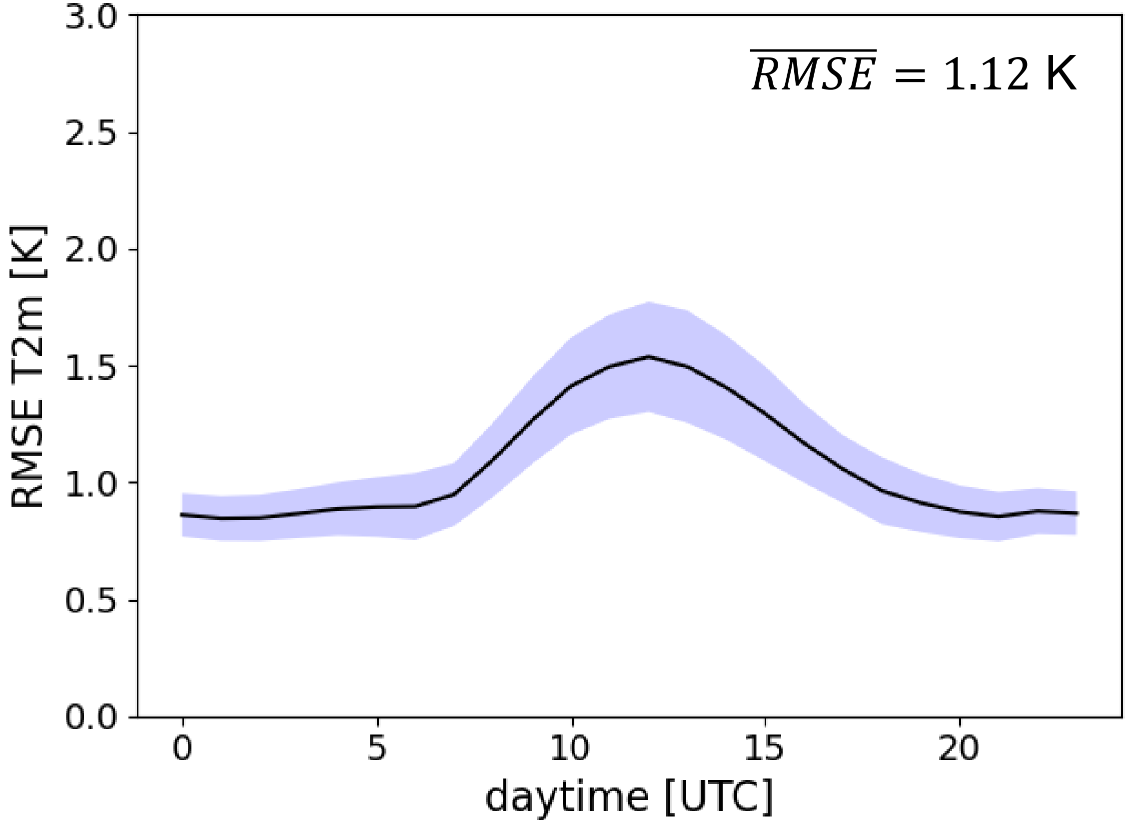
Better

Averaged metrics
plotted against daytime

SOME PRELIMINARY RESULTS

Examples from the 2m temperature downscaling task

Results from
Sha WGAN



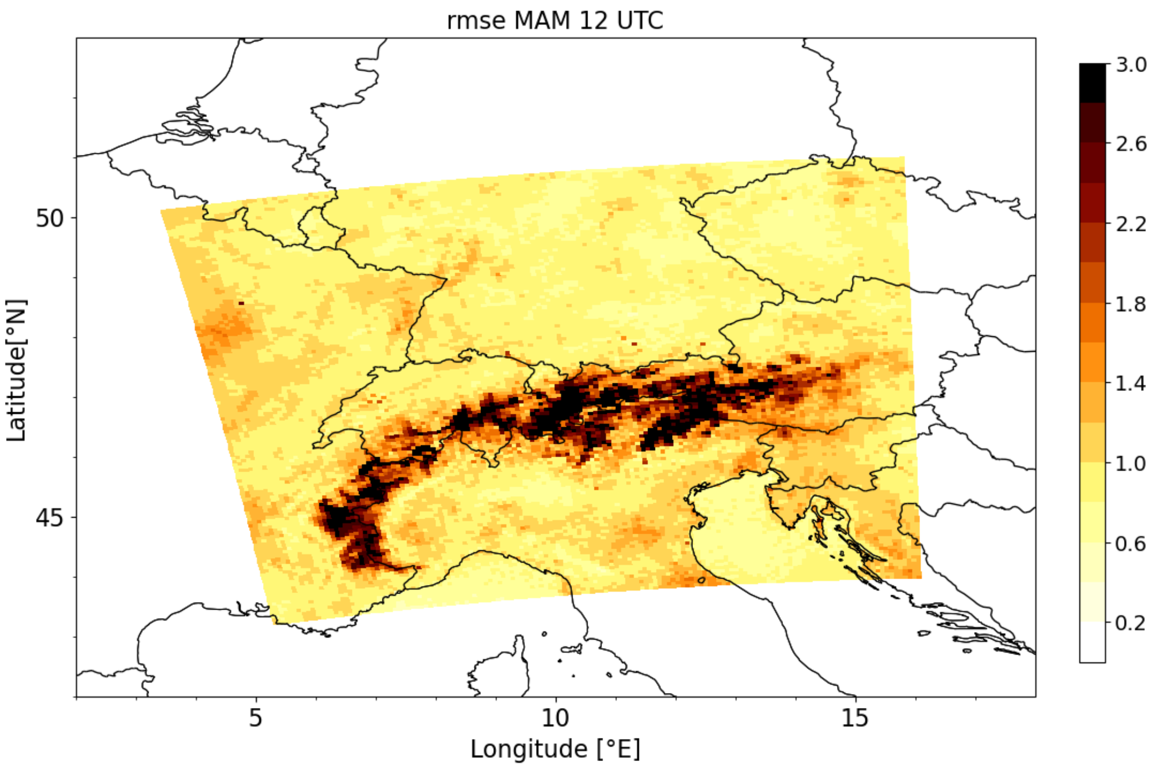
Better

Season-wise evaluation, e.g. RMSE for spring months MAM, or …

SOME PRELIMINARY RESULTS

Examples from the 2m temperature downscaling task

Results from
Sha WGAN

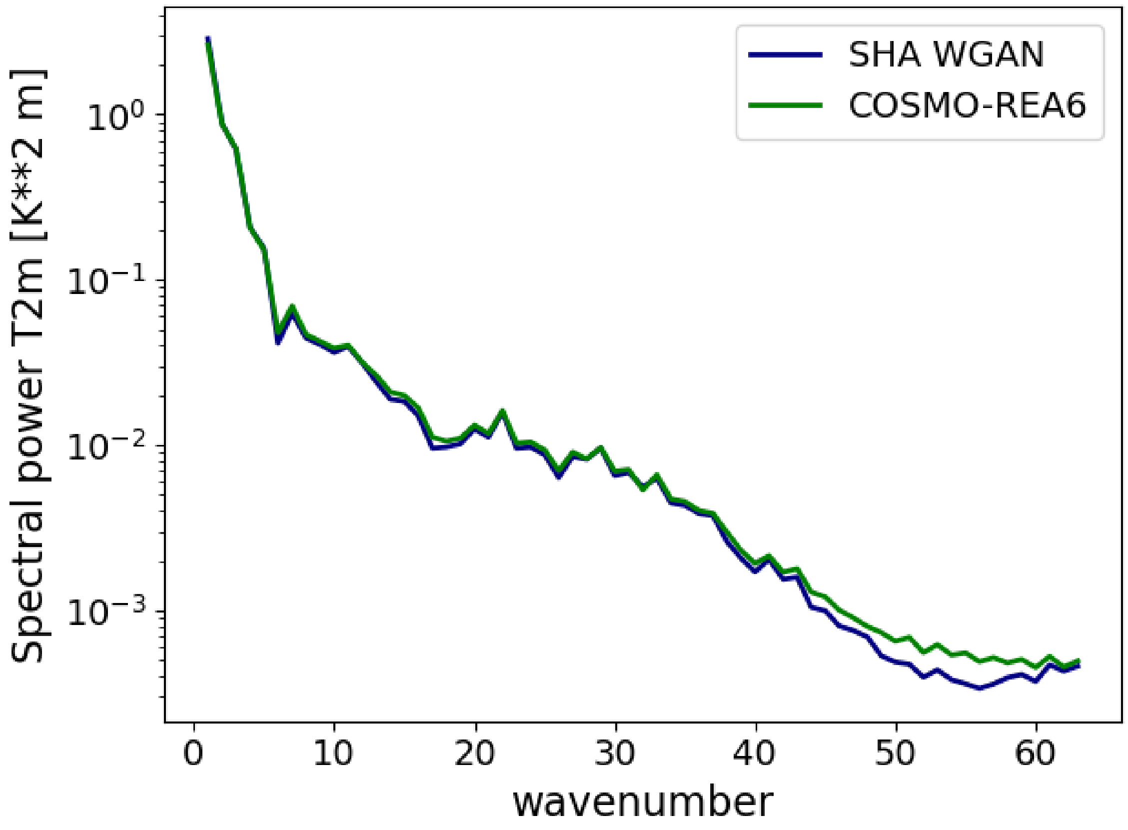


Season-wise evaluation, e.g. RMSE for spring months MAM, or …

SOME PRELIMINARY RESULTS

Examples from the 2m temperature downscaling task

Results from
Sha WGAN

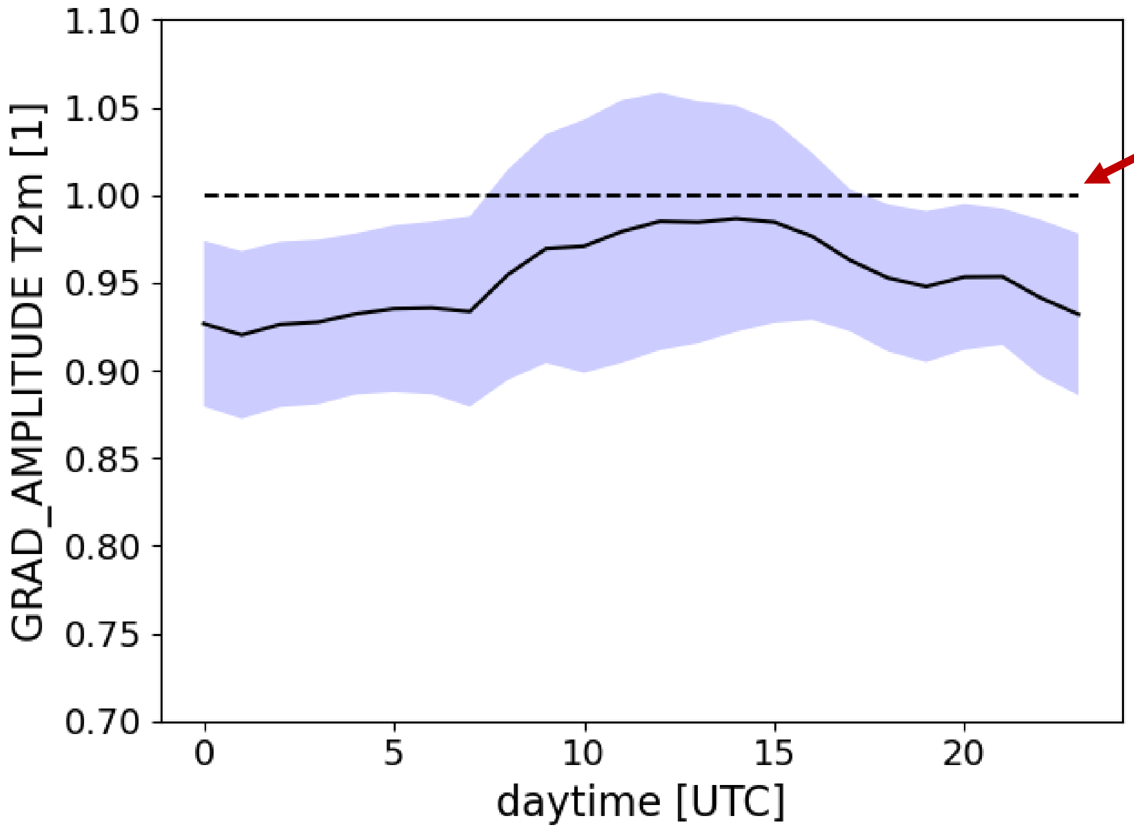


Various options to investigate the spatial variability, e.g. with power spectra …

SOME PRELIMINARY RESULTS

Examples from the 2m temperature downscaling task

Results from
Sha WGAN



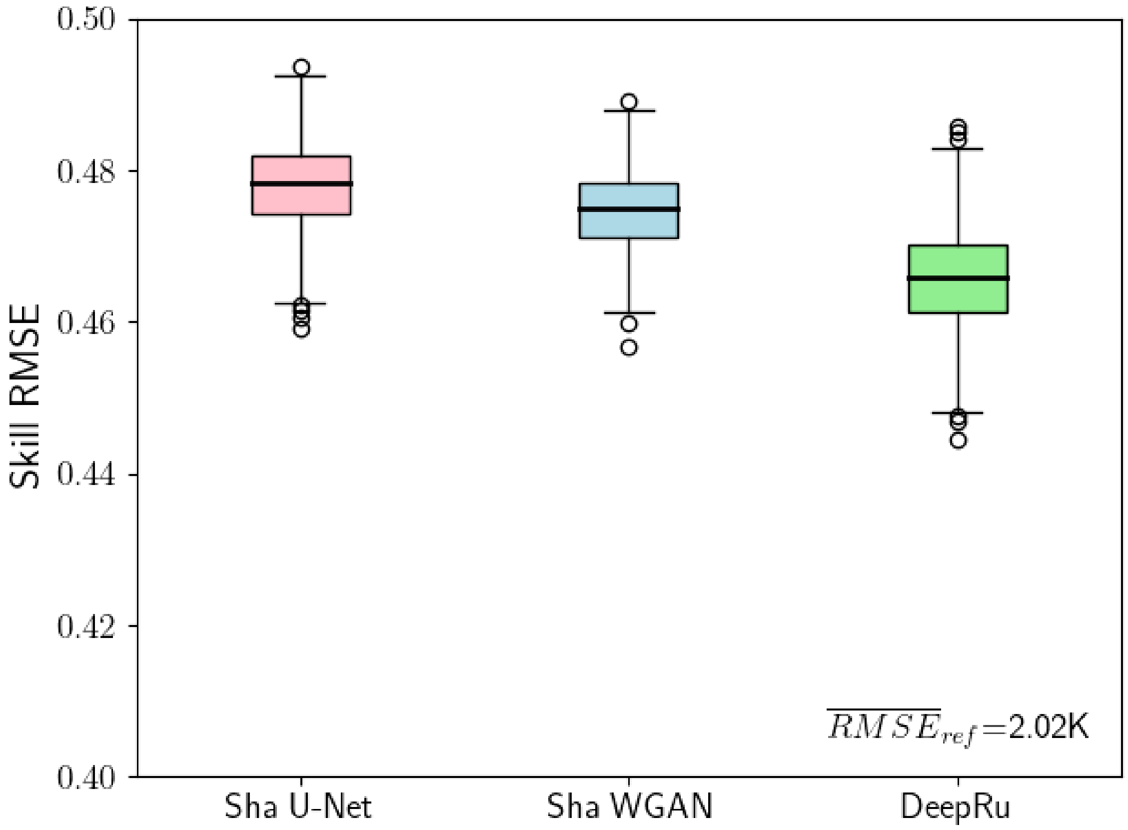
Optimal value

… or in terms of the
gradient amplitude ratio

SOME PRELIMINARY RESULTS

Examples from the 2m temperature downscaling task

Intercomparison



Better

Bilinear interpolation as (simple) reference
→ to be replaced by SAMOS

OUTLOOK

What’s coming next

- Finalize model implementations and training
- Full implementation of metrics and diagnostic tools
- Publication of data via climetlab plug-in
- Publication of code on github
- Accompanying paper

Planned for April’ 24

Future steps:

- Extension to probabilistic downscaling
- Include precipitation downscaling task

Work initiated by the
MAELSTROM project

Check this out

THE DOWNSCALING BENCHMARK TEAM

Hard-working collaboration

- Working in a small team is good, working in a motivated collaboration is even better!

Michael Langguth (JSC)

Ankit Patnala (JSC)

Bing Gong (JSC,now: SHNU)

Martin Schultz (JSC)

Scarlet Stadtler (JSC)

Sebastian Lehner (GSA)

Irene Schicker (GSA)

Markus Dabernig (GSA)

Konrad Mayer (GSA)

Paula Harder (Mila)

GSA: GeosphereAustria
Mila: Mila - Quebec AI Institute

REFERENCES

- Dabernig, Markus, et al. "Spatial ensemble post-processing with standardized anomalies." *QJRMS* 143.703 (2017): 909-916.
- Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." *2009* IEEE Inter. Conf. Comp. Vis., IEEE, 2009.
- Dueben, Peter D., et al. "Challenges and benchmark datasets for machine learning in the atmospheric sciences: Definition, status, and outlook." Art. Intell. Earth Sc., 1.3 (2022): e210002.
- Harder, Paula, et al. "Hard-Constrained Deep Learning for Climate Downscaling." *JMLR*, 24.365 (2023): 1-40.
- Harris, Lucy, et al. "A generative deep learning approach to stochastic downscaling of precipitation forecasts." *J. Adv. Model. Earth Sy.*, 14.10 (2022):
- Höhlein, Kevin, et al. "A comparative study of convolutional neural network models for wind field downscaling." *Meteorol. Appl., 27.6* (2020): e1961.
- Liang, Jingyun, et al. "Swinir: Image restoration using swin transformer." *IEEE Inter. Conf. Comp. Vis.* (2021).
- Mardani, Morteza, et al. "Generative residual diffusion modeling for km-scale atmospheric downscaling." *arXiv preprint arXiv:2309.15214* (2023).
- Rasp, Stephan, et al. "WeatherBench 2: A benchmark for the next generation of data-driven global weather models." *arXiv preprint arXiv:2308.15560* (2023).
- Sha, Yingkai, et al. "Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part I: Daily maximum and minimum 2-m temperature." *J. Appl. Meteorol. Clim.,* 59.12 (2020): 2057-2073.
- Wang, Alex, et al. "GLUE: A multi-task benchmark and analysis platform for natural language understanding." *arXiv preprint arXiv:1804.07461* (2018).

A BENCHMARK DATASET FOR STATISTICAL DOWNSCALING OF METEOROLOGICAL FIELDS WITH DEEP NEURAL NETWORKS

AMS 2024 - 104th Annual Meeting | Baltimore

2024-01-29 I **MICHAEL LANGGUTH**[1], SCARLET STADTLER[1], BING GONG[2], MARTIN G. SCHULTZ[1] (ET AL..)

[1] Juelich Supercomputing Centre (JSC), [2] JSC, now at Shanghai Normal University (SHNU)

EuroHPC Joint Undertaking

"The MAELSTROM project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955513. The JU receives support from the European Union's Horizon 2020 research and innovation programme and United Kingdom, Germany, Italy, Luxembourg, Switzerland, Norway".

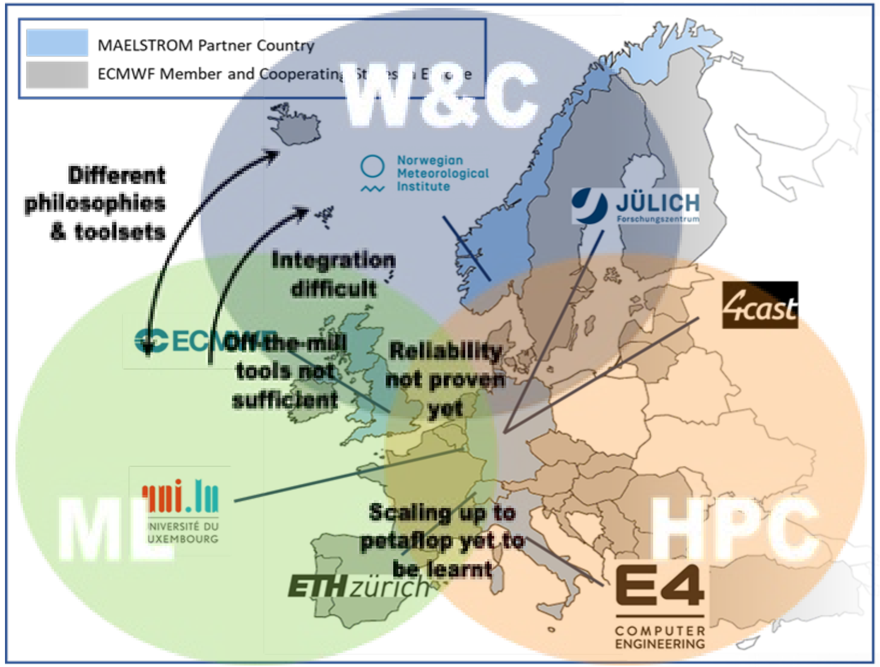
JÜLICH Forschungszentrum

Mitglied der Helmholtz-Gemeinschaft

THE MAELSTROM PROJECT

A EURO-HPC project to foster ML for meteorological applications

- **MA**chin**E** Learning for **S**calable me**T**eo**RO**logy and cli**M**ate
- Euro HPC project coordinated by ECMWF (Apr’21 – Apr’24)
- Main objectives:
  - Develop ML solutions for meteorological applications
  - Enable efficient use of new capacities on supercomputers for the Weather and Climate community
- Collaboration between meteorologists, software developers and HPC specialists
- Six machine learning applications under development
- Benchmark initiative from Application 5



Headquarter location of all partners of the MAELSTROM consortium.
Co-design cycle in MAELSTROM.

Check this out

DATASETS OF THE DOWNSCALING TASKS

The 100m (u,v)-wind downscaling task

Dynamic predictors (ERA5 only):

- 100m (u,v)-wind
- (u,v)-wind at model levels 135, 133, 131, 127 and 122
- Boundary layer height
- Geopotential height at 500 hPa
- Surface pressure

Static predictors (ERA5 & COSMO REA6):

- Surface topography
- Land-sea mask

Evaluation metrics and diagnostics:

- MSE and absolute relative error
- Cosine dissimilarity
- Magnitude difference
- Mean Error of local standard deviation
- Kinetic energy spectra

DATASETS OF THE DOWNSCALING TASKS

The horizontal global radiation downscaling task

- Post-processed global horizontal irradinace (GHI) from *Frank et al., 2018* as target data (rather than raw COSMO REA6)

Dynamic predictors (ERA5 only):

- Surface net solar radiation
- Top net solar radiation
- High, medium and low cloud cover
- Cloud base height
- Total column liquid water
- Surface pressure
- CAPE
- Evaporation

Static predictors (ERA5 & COSMO REA6):

- Surface topography
- Land-sea mask
- Slope of sub-grid scale orography

Evaluation metrics and diagnostics:

- RMSE
- MAE
- Bias
- Mean Error of local standard deviation
- Conditional Quantile Plots