A BENCHMARK DATASET FOR METEOROLOGICAL DOWNSCALING

Michael Langguth & Ankit Patnala
Forschungszentrum Jülich GmbH (FZJ)
Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)
Research Group Earth System Data Exploration (ESDE)
Wilhelm-Johnen-Straße
52428 Jülich, Germany
{m.langguth,a.patnala}@fz-juelich.de

Sebastian Lehner & Markus Dabernig
& Konrad Mayer & Irene Schicker
GeoSphere Austria
Hohe Warte 38, A-1190 Vienna, Austria
{sebastian.lehner,markus.dabernig,
konrad.mayer,irene.schicker}
@geosphere.at

Paula Harder
Mila Quebec AI Institute
6650 Saint Urbain
Montreal, Canada
paula.harder@mila.quebec

ABSTRACT

High spatial resolution in atmospheric representations is crucial across Earth science domains, but global reanalysis datasets like ERA5 often lack the detail to capture local phenomena due to their coarse resolution. Recent efforts have leveraged deep neural networks from computer vision to enhance the spatial resolution of meteorological data, showing promise for statistical downscaling. However, methodological diversity and insufficient comparisons with traditional downscaling techniques challenge these advancements. Our study introduces a benchmark dataset for statistical downscaling, utilizing ERA5 and the finer-resolution COSMO-REA6, to facilitate direct comparisons of downscaling methods for 2m temperature, global (solar) irradiance and 100m wind fields. Accompanying U-Net, GAN, and transformer models with a suite of evaluation metrics aim to standardize assessments and promote transparency and confidence in applying deep learning to meteorological downscaling.

1 INTRODUCTION

Addressing climate change effectively requires accurate observation and modeling of weather and climate, including high-resolution data that is crucial for both understanding global climate patterns and supporting the transition towards renewable energy. The goal of achieving a fossil fuel-free status by 2040 emphasizes the urgency of shifting to renewable sources such as wind and photovoltaic (PV) production. This transition is not only a response to the imperative of climate change mitigation but also a strategic move to ensure energy security and sustainability. As such, localized and regional forecasts play a pivotal role in guiding adaptation strategies, influencing decisions across various sectors including agriculture, energy, and transportation, and specifically in planning the construction of new wind and PV power plant sites. Here, downscaling in climate science, similar to super-resolution in computer vision, offers an efficient method to infer local, high-resolution quantities from the coarser scale variables. This technique is of utmost importance for assessing renewable energy resources, like wind speed at a currently prominent hub height of 100 m and global solar radiation, and their possible changes with climate change.

There exists a rapidly growing number of deep learning (DL) approaches tackling downscaling of a variety of meteorological variables, across different scales and regions. Following the state-of-the-art in super-resolution, there are impressive results from methods like CNNs [33], GANs [30], diffusion-based models [21], or foundation models [23] in downscaling. Intercomparison though is hard, given that the metrics, baselines, and datasets used vary almost among all approaches.

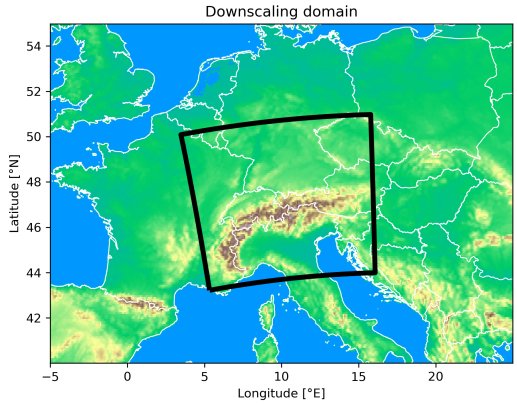

Figure 1: Target domain for downscaling depicted in black in the COSMO-REA6 projection.

Benchmark datasets like WeatherBench [26] or ClimateBench [35] have helped to improve and unify fields like DL for weather forecasts or climate projections. The field of deep learning for downscaling though is still lacking a dataset and framework to compare and benchmark promising architectures. A downscaling benchmark dataset will aid both research and deployment. It provides an useful and carefully designed test setup for ML researchers and developers aiming to craft novel neural network architectures. This will give an insight of the most promising models to the researchers working on real-world applications of downscaling.

The repository can be found at `https://github.com/mlangguth89/downscaling_benchmark`.

2 The downscaling benchmark dataset

In the following, the design and principles of the benchmark dataset for statistical downscaling are outlined. With the ambition to provide a benchmark dataset that is usable and practical for the scientific community, we closely follow the recipe to build proper benchmark datasets for Atmospheric Sciences as provided by [7].

2.1 Downscaling tasks and data

The benchmark dataset comprises three downscaling tasks for the generation of high-resolved data fields of 2m temperature, 100m-wind, and global horizontal irradiance generated from a coarse-grained set of predictor variables.

Reanalysis datasets are well suited for this purpose since they provide a comprehensive, consistent, and quality-controlled database. The global ERA5 reanalysis dataset [14], is chosen as the coarse-grained input for our downscaling task. Its spatio-temporal coverage (global, 1940 to near present), however, restricts its spatial resolution to a 0.25°-grid and thus barely represents the spatial variability over complex terrain. The regional COSMO-REA6 reanalysis dataset [3], in contrast, can depict the spatial variability in meteorological fields in more detail.

The added value due to the 4-5 times finer grid ($\Delta x \simeq 6\,\mathrm{km}$) of the COSMO-REA6 data has been verified in various studies, in particular for temperature [27], solar irradiance [24; 25] and wind fields [24; 17; 4; 28; 25] over complex terrain. While the original COSMO-REA6 reanalysis provides high-quality (target) data for 2m temperature and 100m wind, we make use of a postprocessed COSMO-REA6 product for global horizontal irradiance [9]. As demonstrated in [9], the postprocessing reduces biases in the existing original reanalysis dataset, significantly enhancing its accuracy for PV power generation applications.

Since the added value of COSMO-REA6 is most pronounced over complex terrain, the target domain for the downscaling tasks is confined to Central Europe including the Alpine region. The domain comprises 144x128 grid points in (rotated) latitude and longitude direction is depicted in Fig. 1. In addition to the complex terrain of the Alpine region, the target domain covers relevant regions for the production of renewable energy in Central Europe. In particular, Southern Germany provides the major contribution to the national PV power generation (see Fig. 2 in [6]). While PV power generation is also a relevant driver for Austria's transition to renewable energy, several wind power farms are included in the target domain within complex terrain. Fostering wind energy production in Southern Germany and Austria is furthermore considered to be a necessary prerequisite to realize the zero-net target in energy production of both countries by 2045 [10; 11; 16; 22; 8].

The set of predictor variables as listed in the appendix (see Section A.1) for the three downscaling tasks have been chosen with the help of domain knowledge. To ease the reconstruction of high-resolution features from the coarse-grained ERA5-data, temperature and wind data from several model levels are included for the respective downscaling task. Further predictor variables are chosen to encode relevant processes in the planetary boundary layer that ultimately drive the spatial variability or to provide relevant information on the attenuance of solar radiation within the atmosphere (i.e. cloud-related predictors).

Easy access to the prepared datasets is realized with the help of a `CliMetLab` plugin, a Python package to enable easy downloading of meteorological data developed by ECMWF. As illustrated in Fig. 2, downloading the dataset for a specific downscaling task from the s3 bucket system of the European Weather Cloud is realized with a few lines of Python code. Noteworthy, the provided datasets are ready-to-use and no expert knowledge of data handling is required for its usage. All necessary steps of preprocessing (e.g. re-projecting the ERA5-data onto the rotated pole grid of COSMO-REA6) and data pairing have been accomplished with the provision of the data via the CliMetLab plugin.

2.2 Baseline downscaling models

A set of deep neural network architectures described in the literature is chosen to enable systematic benchmarking against baseline solutions. Since statistical downscaling has a long tradition in statistics for atmospheric science [20], this set is complemented by an advanced approach based on classical model output statistics. A brief summary of the baseline models is provided subsequently:

Deep Learning Models Convolutional neural networks (ConvNets) have been applied in numerous studies for statistical downscaling [32; 2; 1; 34]. In particular, the U-Net architecture, a variant of hierarchical ConvNets, is known for its ability to effectively combine high-level context with detailed local information which is considered to be highly relevant for downscaling tasks. Here, we choose the DeepRU-architecture [15] and the U-Net suggested in [29] (Sha U-Net) as two representative models for the family of ConvNets. More recently, conditional Generative Adversarial Networks (GANs) have become popular due to their capability to reconstruct small-scale features. GANs combine two neural networks, a generator and a discriminator model, to efficiently learn the statistical properties of the ground truth with the help of adversarial optimization. Two so-called Wasserstein GANs (WGANs), where the discriminator is replaced by a critic model, are chosen as baseline models for this benchmark, that are the WGAN suggested by [13] and a WGAN model that deploys the Sha U-Net as the generator (Sha WGAN). Furthermore, we adapt the SwinIR as a novel vision transformer based on Swin transformer[19], that showed promise in super-resolution tasks of computer vision [18].

SAMOS - Standardized Anomaly Model Output Statistics SAMOS [5] is a statistical post-processing method based on the idea of ensemble model output statistics [12]. Instead of using real/normalized input and output pairs, SAMOS uses a mixture of classical statistics with machine learning through gradient boosting providing both deterministic and if wanted, probabilistic forecasts. By removing local information and fitting towards a climatology, standardized anomalies are attained that allow global optimization for all data points of the downscaling domain.

2.3 Performance evaluation

Choosing the right evaluation setup for a benchmark dataset is crucial. Here, we include a variety of downscaling metrics such as the RMSE, the Mean Error of local Standard Deviation [37], and the Integrated Quadratic Distance [31], alongside advanced visual analyses like conditional quantile plots. Depending on the variable evaluated, there exist different sets of metrics, e.g. for wind, we additionally include the cosine dissimilarity score, the vector RMSE, and magnitude difference as suggested in [15]. In addition to standard metrics, we provide diagnostics for the marginal distribution of the target variable, e.g. power spectra analysis and histograms. The evaluation of global irradiance mirrors that of 2-meter temperature but emphasizes relative errors to account for diurnal and seasonal cycles, incorporating daily sums and hit-miss rates as well as the proposed verification scores by [36].

Our evaluation setup consists of two steps: single model evaluation for all task-specific metrics followed by a comparison to a reference model using skill scores. Fig. 3a-c in the appendix exemplifies preliminary results of the Sha WGAN and additionally compares three baseline models against a simple bilinear interpolation as a baseline solution (Fig. 3d).

3 Conclusion

With this work, we introduce a benchmark dataset tailored for statistical downscaling of meteorological fields. Our dataset is readily accessible, eliminating the need for extensive preprocessing (e.g. data pairing) by application developers. Moreover, the benchmark enables comparison against existing baseline solutions including both standard deep learning solutions, such as U-Nets, WGANs and vision transformers, and the classical competitor method SAMOS. By providing a well-defined evaluation framework, developed solutions can be analyzed in detail to identify optimal strategies for statistical downscaling. Such a framework does not only foster advancements in research, but can also be used to facilitate the operational deployment of statistical downscaling models. We anticipate that the availability of benchmark datasets will accelerate progress statistical downscaling of meteorological fields, leading to improved accuracy, reliability and trustworthiness in deep learning-based solutions.

Acknowledgments

Part of the work is funded by the European Union's Horizon 2020 research and innovation programme and the EuroHPC Joint Undertaking under Grant Agreement 955513 and co-funded by the German Ministry of Education and Research (BMBF) under funding reference 16HPC029. The authors have used computation resources of the supercomputers JUWELS and Juwels Booster at JSC.

References

[1] Adewoyin, R. A., Dueben, P., Watson, P., He, Y., and Dutta, R. TRU-NET: a deep learning approach to high resolution prediction of rainfall. *Machine Learning 110* (2021), 2035–2062.

[2] Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M. Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development 13*, 4 (2020), 2109–2124.

[3] Bollmeyer, C., Keller, J. D., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., Hense, A., Keune, J., Kneifel, S., Pscheidt, I., Redl, S., and Steinke, S. Towards a high-resolution regional reanalysis for the european cordex domain. *Quarterly Journal of the Royal Meteorological Society 141*, 686 (2015), 1–15.

[4] Brune, S., Keller, J. D., and Wahl, S. Evaluation of wind speed estimates in reanalyses for wind energy applications. *Advances in Science and Research 18* (2021), 115–126.

[5] Dabernig, M., Schicker, I., Kann, A., Wang, Y., and Lang, M. N. Statistical postprocessing with standardized anomalies based on a 1 km gridded analysis. *Meteorologische Zeitschrift 29*, 4 (10 2020), 265–275.

[6] Drücke, J., Borsche, M., James, P., Kaspar, F., Pfeifroth, U., Ahrens, B., and Trentmann, J. Climatological analysis of solar and wind energy in Germany using the Grosswetterlagen classification. *Renewable Energy 164* (2021), 1254–1266.

[7] DUEBEN, P. D., SCHULTZ, M. G., CHANTRY, M., GAGNE, D. J., HALL, D. M., AND MCGOVERN, A. Challenges and benchmark datasets for machine learning in the atmospheric sciences: Definition, status, and outlook. *Artificial Intelligence for the Earth Systems 1*, 3 (2022), e210002.

[8] FEDERAL MINISTRY FOR THE ENVIRONMENT, NATURE CONSERVATION, N. S., AND PROTECTION, C. *Climate ACtion Plan 2015*. Federal Ministry for the Environment, Nature conservation, Nuclear Strategy and Consumer Protection, 2016.

[9] FRANK, C. W., WAHL, S., KELLER, J. D., POSPICHAL, B., HENSE, A., AND CREWELL, S. Bias correction of a novel european reanalysis data set for solar energy applications. *Solar Energy 164* (2018), 12–24.

[10] FÜR KLIMASCHUTZ, B. *Integrierter nationaler Energie- und Klimaplan für Österreich*. Bundesministerium für Nachhaltigkeit und Tourismus, 2023.

[11] GAUGL, R., KLATZER, T., BACHHIESL, U., WOGRIN, S., AND JODL, S. Gis-based optimization – achieving austria's 2030 wind energy target. *e & i Elektrotechnik und Informationstechnik* (09 2021).

[12] GNEITING, T., RAFTERY, A. E., WESTVELD, A. H., AND GOLDMAN, T. Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review 133*, 5 (2005), 1098 – 1118.

[13] HARRIS, L., MCRAE, A. T. T., CHANTRY, M., DUEBEN, P. D., AND PALMER, T. N. A generative deep learning approach to stochastic downscaling of precipitation forecasts. *Journal of Advances in Modeling Earth Systems 14*, 10 (2022), e2022MS003120. e2022MS003120 2022MS003120.

[14] HERSBACH, H., BELL, B., BERRISFORD, P., HIRAHARA, S., HORÁNYI, A., MUÑOZ-SABATER, J., NICOLAS, J., PEUBEY, C., RADU, R., SCHEPERS, D., SIMMONS, A., SOCI, C., ABDALLA, S., ABELLAN, X., BALSAMO, G., BECHTOLD, P., BIAVATI, G., BIDLOT, J., BONAVITA, M., DE CHIARA, G., DAHLGREN, P., DEE, D., DIAMANTAKIS, M., DRAGANI, R., FLEMMING, J., FORBES, R., FUENTES, M., GEER, A., HAIMBERGER, L., HEALY, S., HOGAN, R. J., HÓLM, E., JANISKOVÁ, M., KEELEY, S., LALOYAUX, P., LOPEZ, P., LUPU, C., RADNOTI, G., DE ROSNAY, P., ROZUM, I., VAMBORG, F., VILLAUME, S., AND THÉPAUT, J.-N. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*, 730 (2020), 1999–2049.

[15] HÖHLEIN, K., KERN, M., HEWSON, T., AND WESTERMANN, R. A comparative study of convolutional neural network models for wind field downscaling. *Meteorological Applications 27*, 6 (2020), e1961.

[16] HÖLTINGER, S., SALAK, B., SCHAUPPENLEHNER, T., SCHERHAUFER, P., AND SCHMIDT, J. Austria's wind energy potential – a participatory modeling approach to assess socio-political and market acceptance. *Energy Policy 98* (2016), 49–61.

[17] JOURDIER, B. Evaluation of era5, merra-2, cosmo-rea6, newa and arome to simulate wind power production over france. *Advances in Science and Research 17* (2020), 63–77.

[18] LIANG, J., CAO, J., SUN, G., ZHANG, K., VAN GOOL, L., AND TIMOFTE, R. Swinir: Image restoration using swin transformer. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (2021), pp. 1833–1844.

[19] LIU, Z., LIN, Y., CAO, Y., HU, H., WEI, Y., ZHANG, Z., LIN, S., AND GUO, B. Swin transformer: Hierarchical vision transformer using shifted windows. *CoRR abs/2103.14030* (2021).

[20] MARAUN, D., AND WIDMANN, M. *Statistical downscaling and bias correction for climate research*. Cambridge University Press, 2018.

[21] MARDANI, M., BRENOWITZ, N., COHEN, Y., PATHAK, J., CHEN, C.-Y., LIU, C.-C., VAHDAT, A., KASHINATH, K., KAUTZ, J., AND PRITCHARD, M. S. Residual diffusion modeling for km-scale atmospheric downscaling.

[22] MOEMKEN, J., REYERS, M., FELDMANN, H., AND PINTO, J. G. Future changes of wind speed and wind energy potentials in euro-cordex ensemble simulations. *Journal of Geophysical Research: Atmospheres 123*, 12 (2018), 6373–6389.

[23] NGUYEN, T., BRANDSTETTER, J., KAPOOR, A., GUPTA, J. K., AND GROVER, A. Climax: A foundation model for weather and climate. In *International Conference on Machine Learning* (2023).

[24] NIERMANN, D., BORSCHE, M., KAISER?WEISS, A., AND KASPAR, F. Evaluating renewable?energy?relevant parameters of cosmo?rea6 by comparison with satellite data, station observations and other reanalyses. *Meteorologische Zeitschrift 28*, 4 (11 2019), 347–360.

[25] RAMIREZ CAMARGO, L., GRUBER, K., AND NITSCH, F. Assessing variables of regional reanalysis data sets relevant for modelling small-scale renewable energy systems. *Renewable Energy 133* (2019), 1468–1478.

[26] RASP, S., HOYER, S., MEROSE, A., LANGMORE, I., BATTAGLIA, P. W., RUSSEL, T., SANCHEZ-GONZALEZ, A., YANG, V. Q., CARVER, R., AGRAWAL, S., CHANTRY, M., BOUALLÈGUE, Z. B., DUEBEN, P., BROMBERG, C., SISK, J., BARRINGTON, L., BELL, A., AND SHA, F. Weatherbench 2: A benchmark for the next generation of data-driven global weather models. *ArXiv abs/2308.15560* (2023).

[27] SCHERRER, S. C. Temperature monitoring in mountain regions using reanalyses: lessons from the Alps. *Environmental Research Letters 15*, 4 (2020), 044005.

[28] SCHICKER, I., GANGLBAUER, J., DABERNIG, M., AND NACHT, T. Wind power estimation on local scale—a case study of representativeness of reanalysis data and data-driven analysis. *Frontiers in Climate 5* (08 2023).

[29] SHA, Y., GAGNE, D. J., WEST, G., AND STULL, R. Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. part i: Daily maximum and minimum 2-m temperature. *Journal of Applied Meteorology and Climatology* (2020).

[30] STENGEL, K., GLAWS, A., HETTINGER, D., AND KING, R. N. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences 117* (2020), 16805 – 16815.

[31] THORARINSDOTTIR, T. L., GNEITING, T., AND GISSIBL, N. Using proper divergence functions to evaluate climate models. *SIAM/ASA Journal on Uncertainty Quantification 1*, 1 (2013), 522–534.

[32] VANDAL, T., KODRA, E., GANGULY, S., MICHAELIS, A., NEMANI, R., AND GANGULY, A. R. Generating high resolution climate change projections through single image super-resolution: An abridged version. In *International Joint Conferences on Artificial Intelligence Organization* (2018).

[33] VANDAL, T. J., KODRA, E., GANGULY, S., MICHAELIS, A. R., NEMANI, R. R., AND GANGULY, A. R. Deepsd: Generating high resolution climate change projections through single image super-resolution. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2017).

[34] WANG, F., TIAN, D., LOWE, L., KALIN, L., AND LEHRTER, J. Deep learning for daily precipitation and temperature downscaling. *Water Resources Research 57*, 4 (2021), e2020WR029308.

[35] WATSON-PARRIS, D., RAO, Y., OLIVIÈ, D. J. L., SELAND, Ø., NOWACK, P. J., CAMPS-VALLS, G., STIER, P., BOUABID, S., DEWEY, M., FONS, E., GONZALEZ, J., HARDER, P., JEGGLE, K., LENHARDT, J., MANSHAUSEN, P., NOVITASARI, M., RICARD, L., AND ROESCH, C. E. Climatebench v1.0: A benchmark for data-driven climate projections. *Journal of Advances in Modeling Earth Systems 14* (2022).

[36] Yang, D., Alessandrini, S., Antonanzas, J., Antonanzas-Torres, F., Badescu, V., Beyer, H. G., Blaga, R., Boland, J., Bright, J. M., Coimbra, C. F., David, M., Âzeddine Frimane, Gueymard, C. A., Hong, T., Kay, M. J., Killinger, S., Kleissl, J., Lauret, P., Lorenz, E., van der Meer, D., Paulescu, M., Perez, R., Perpiñán-Lamigueiro, O., Peters, I. M., Reikard, G., Renné, D., Saint-Drenan, Y.-M., Shuai, Y., Urraca, R., Verbois, H., Vignola, F., Voyant, C., and Zhang, J. Verification of deterministic solar forecasts. *Solar Energy 210* (2020), 20–37. Special Issue on Grid Integration.

[37] Zerenner, T., Venema, V., Friederichs, P., and Simmer, C. Downscaling near-surface atmospheric fields with multi-objective genetic programming. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion* (2017), pp. 11–12.

A Appendix

A.1 Predictor Variables

2m temperature downscaling task :

- 2m temperature
- temperature at model levels [137, 135, 131, 127, 122, 115]
- 10m (u,v)-wind components
- surface latent and sensible heat flux
- surface pressure
- static predictors: coarse and high-resolved surface topography and land-sea mask

100m wind downscaling task :

- 100m (u,v)-wind components
- (u,v)-wind components at model levels [137, 135, 131, 127, 122, 115]
- boundary layer height
- geopotential at 500 hPa
- static predictors: coarse and high-resolved surface topography and land-sea mask

Solar Irradiance: :

- Surface net solar radiation
- Top net solar radiation
- Low, medium and high cloud cover
- Total column cloud liquid water
- Total precipitation
- Convective available potential energy
- Surface pressure
- Slope of sub-gridscale orography
- Cloud base height
- Evaporation
- Static predictors: topography, land-sea mask

A.2 Figures

```
!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")
```

Figure 2: Exemplary Python code snippet to download the validation dataset of the 2m temperature downscaling task with the ClimMetLab plugin. The data is stored in a netCDF-file on disk with the help of `xarray`.

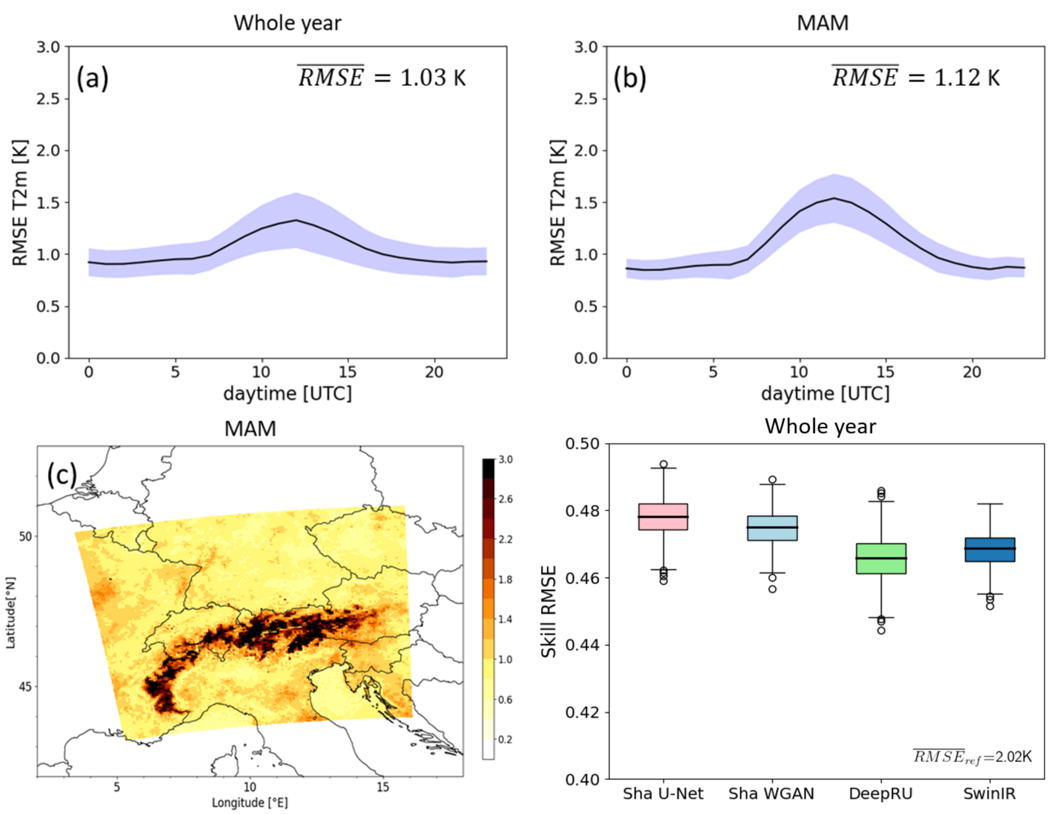

Figure 3: Example evaluation results from the Sha WGAN on the 2m temperature downscaling task. The diurnal cycle of RMSE is evaluated for the whole test year 2018 (a), for a specific season such as spring (b), or the RMSE can be analysed spatially for a specific season and daytime (c). (d) provides an intercomparison in terms of the Skill RMSE between the Sha U-Net, the Sha WGAN, the DeepRU and the SwinIR with bilinear interpolation as the reference downscaling approach. The bilinear reference will be replaced by SAMOS in the future.