

# EMERGING AI-BASED WEATHER PREDICTION MODELS AS DOWNSCALING TOOLS

A PREPRINT

 **Nikolay Koldunov\***

Alfred Wegener Institute  
Helmholtz Centre for Polar  
and Marine Research (AWI)  
nikolay.koldunov@awi.de

 **Thomas Rackow**

European Centre for Medium-Range  
Weather Forecasts (ECMWF)  
thomas.rackow@ecmwf.int

 **Christian Lessig**

European Centre for Medium-Range  
Weather Forecasts (ECMWF)  
christian.lessig@ecmwf.int

 **Sergey Danilov**

Alfred Wegener Institute  
Helmholtz Centre for Polar  
and Marine Research (AWI)  
sergey.danilov@ecmwf.int

 **Suvarchal K. Cheedela**

Alfred Wegener Institute  
Helmholtz Centre for Polar  
and Marine Research (AWI)  
suvarchal.cheedela@awi.de

 **Dmitry Sidorenko**

Alfred Wegener Institute  
Helmholtz Centre for Polar  
and Marine Research (AWI)  
dmitry.sidorenko@ecmwf.int

 **Irina Sandu**

European Centre for Medium-Range  
Weather Forecasts (ECMWF)  
irina.sandu@ecmwf.int

 **Thomas Jung**

Alfred Wegener Institute  
Helmholtz Centre for Polar  
and Marine Research (AWI)  
thomas.jung@awi.de

June 27, 2024

## ABSTRACT

The demand for high-resolution information on climate change is critical for accurate projections and decision-making. Presently, this need is addressed through high-resolution climate models or downscaling. High-resolution models are computationally demanding and creating ensemble simulations with them is typically prohibitively expensive. Downscaling methods are more affordable but are typically limited to small regions. This study proposes the use of existing AI-based numerical weather prediction systems (AI-NWP) to perform global downscaling of climate information from low-resolution climate models.

Our results demonstrate that AI-NWP initialized from low-resolution initial conditions can develop detailed forecasts closely resembling the resolution of the training data using a one day lead time. We constructed year-long atmospheric fields using AI-NWP forecasts initialized from smoothed ERA5 and low-resolution CMIP6 models. Our analysis for 2-metre temperature indicates that AI-NWP can generate high-quality, long-term datasets and potentially perform bias correction, bringing climate model outputs closer to observed data.

The study highlights the potential for off-the-shelf AI-NWP to enhance climate data downscaling, offering a simple and computationally efficient alternative to traditional downscaling techniques. The downscaled data can be used either directly for localized climate information or as boundary conditions for further dynamical downscaling.

**Keywords** climate · climate modelling · downscaling · machine learning

---

\*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

## 1 Introduction

In the last two years, several data-driven AI-based numerical weather prediction systems (AI-NWP) for medium-range global weather forecasting have been developed. Notable examples include Pangu-Weather (Bi et al., 2023), GraphCast (Lam et al., 2023), FourCastNet (Pathak et al., 2022), FuXi (Chen et al., 2023) and AIFS (Lang et al., 2024). These models today outperform the best conventional, equation-based models in deterministic skill for most variables and metrics (Ben-Bouallegue et al., 2023). The models are trained on the ERA5 reanalysis (Hersbach et al., 2020) that combines a large number of observations into a best state estimate of the atmosphere. It has resolution of about 31 km. When used to generate medium-range forecasts, the AI-NWP models use either ERA5 data, operational ECMWF forecasts, or analysis data as initial conditions.

In this work, we exploit that AI-NWP models are trained to provide predictions that match ERA5 both in terms of its resolution and its data distribution and that they have demonstrated a surprising robustness to novel initial conditions (Hakim and Masanam, 2023). Based on this, we hypothesize that AI-NWP models initialised with low-resolution initial conditions will produce output with a higher level of detail matching their training data. Through this, they will effectively perform automatic downscaling without being trained for it, an approach known as zero-shot. Since the ERA5 reanalysis is observationally well-constrained, they potentially can also bias correct model output, e.g. from low resolution biased climate simulations.

Downscaling of data from low-resolution climate models is a well-developed field (Giorgi, 2019) and currently forms the basis for regional climate assessments, such as those in the IPCC reports (Ranasinghe et al., 2021). This is usually achieved with regional climate models (RCMs), which use low-resolution global climate models as boundary conditions and run simulations for smaller regions, often in atmosphere-only mode, although coupled regional climate simulations are also well developed (e.g. Ruti et al. (2016)). There are also statistical downscaling techniques, which employ statistical relationships between low-resolution inputs and high-resolution models or observations to downscale model results (Fowler et al. (2007); Ekström et al. (2015)). Downscaled climate data are fundamental for climate service products, as users require information on a local scale (e.g. van den Hurk et al. (2018); Koldunov et al. (2016); Damm et al. (2020)). The most prominent coordinated initiative in regional downscaling is the Coordinated Regional Climate Downscaling EXperiment (CORDEX, Giorgi and Gutowski Jr (2015)).

Neural networks have been used extensively for climate and weather data downscaling in recent years (e.g. Baño-Medina et al., 2020; Rampal et al., 2022; Höhlein et al., 2020). Most of the applications target weather prediction use cases, regional domains and use low/high resolution pairs to train the neural networks (Stengel et al., 2020; Harris et al., 2022). An interesting approach is to emulate dynamical downscaling done with RCMs, using low resolution climate simulations (Tomasi et al., 2024; Mardani et al., 2023) or reanalysis (Wang et al., 2021) as low-resolution data, and RCM simulations as the target high-resolution data. Foundation models for weather and climate can also be employed for downscaling tasks, and there are examples for regional (Lessig et al., 2023) and global climate (Nguyen et al., 2023).

In this work we will first evaluate the approach of downscaling with AI-NWP by applying it to low-resolution data given by smoothed ERA5 fields. We then study whether the AI-NWP models produce physically plausible output fields that have more fine-scale detail than the input. Next, we will assess whether it is feasible to construct a high-quality, long-term (year-long) record of atmospheric fields by combining the results from a sequence of forecasts initiated from low-resolution data. Finally, we will use model output of low-resolution CMIP6 models as initial conditions for a AI-NWP moddel for several years of historical and future scenario simulations.

The aim of this paper is to demonstrate that this approach is viable in principle. Therefore, we will focus primarily on surface temperature, one of the main climate parameters with significant impact, and use a limited number of years as examples. More detailed experiments are planned for the near future.

## 2 Methods

### 2.1 Pangu-Weather

We use Pangu-Weather (Bi et al., 2023) as AI-NWP model in this study. Pangu-Weather uses a 3D Earth-specific transformer architecture, designed to capture complex atmospheric interactions across multiple pressure levels and surface variables. The model is trained on 39 years of ERA5 reanalysis data (1979 to 2017), encompassing global atmospheric and surface variables at high temporal and spatial resolution, corresponding approximately to 31 km. The input variables for the Pangu-Weather model include geopotential, specific humidity, temperature, and the u- and v-components of wind on 13 pressure levels (50 hPa, 100 hPa, 150 hPa, 200 hPa, 250 hPa, 300 hPa, 400 hPa, 500 hPa, 600 hPa, 700 hPa, 850 hPa, 925 hPa, and 1,000 hPa), as well as surface variables such as 2-metre temperature, the u- and v-component of 10-meter wind, and mean sea level pressure. Pangu-Weather outperforms the operational Integrated

Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) in deterministic forecasts for many relevant metrics. It also demonstrates superior performance with respect to tropical cyclone tracks providing more accurate and computationally efficient predictions.

Pangu-Weather is used as distributed through the `ai-models` package ([Raoult et al., 2024](#)), created by ECMWF to run different AI-based weather forecasting models from a similar interface. Runs were performed at the German Climate Supercomputing Centre (DKRZ), on the Levante HPC’s Graphical Processing Unit (GPU) nodes equipped with 4 NVIDIA A100-SXM4 GPUs. One 10-day forecast takes about 2 minutes of wall clock time on a single GPU and produces a 5.7GB output file (6-hourly output frequency).

## 2.2 ERA5 and climate data preparation

As mentioned before, Pangu-Weather is called through the `ai-models` package, which requires data in GRIB format, interpolated on a regular  $0.25^\circ$  degree regular grid. ERA5 data were interpolated from the original reduced Gaussian grid to a regular grid using the `cdo` package ([Schulzweida, 2023](#)) with bi-linear interpolation (based on version implemented through YAC ([Hanke et al., 2016](#))). As the GRIB files produced by `cdo` are not compatible with the input expected by `ai-models`, the interpolated data were converted to GRIB format using the `climetlab` package ([ECMWF, 2024](#)). To smooth the initial conditions, we applied Gaussian smoothing using the `scipy.ndimage.gaussian_filter` function from the `scipy` package ([Virtanen et al., 2020](#)), with various `sigma` values.

As low-resolution climate simulations we use CMIP6 data, namely AWI-ESM-1-1-LR ([Danek et al., 2020](#)) based on the AWI-CM model ([Sidorenko et al., 2015; Rackow et al., 2018; Semmler et al., 2020](#)), and MPI-ESM1-2-LR ([Wieners et al., 2019b,a](#)) based on MPI-ESM ([Giorgetta et al., 2013](#)). Both models share the same atmospheric component ECHAM6.3 ([Stevens et al., 2013](#)), configured with a T63 Gaussian grid with 192 longitudes, 96 latitudes, and 47 vertical levels. The models however use different ocean components: FESOM1.4 ([Wang et al., 2014](#)) and MPI-OM ([Jungclaus et al., 2013](#)). This leads to different climate sensitivities in their higher-resolution versions ([Semmler et al., 2021](#)). The atmospheric resolution of those models is approximately 250 km. This resolution is typical for CMIP3, whereas CMIP6 generally features resolutions of around 100 km ([Haarsma et al., 2016](#)). We will explore low-resolution models with the expectation that results for higher-resolution models will be similar or better, since those are closer to the target resolution of AI models.

Scenario simulations for AWI-ESM-1-1-LR are not available, so we only use projections from MPI-ESM1-2-LR SSP585 scenario ([Wieners et al., 2019a](#)). Data were accessed from the ESGF node at DKRZ . We interpolated these data to a  $0.25^\circ$  regular grid from the original CMORised netCDF files using `cdo`’s bi-linear interpolation. The  $u$  and  $v$  fields on pressure levels required an additional step because the necessary pressure levels were not available in the original CMIP6 output. Therefore, we interpolated the data from model levels, which was also done using `cdo`.

## 2.3 Processing Software

In addition to already mentioned software, data processing, analysis, and visualisation were performed with the following python packages: `xarray` ([Hoyer and Hamman, 2017](#)), `dask` ([Rocklin, 2015](#)), `zarr` ([Miles et al., 2024](#)), `matplotlib` ([Hunter, 2007](#)), `cartopy` ([Met Office, 2015](#)), `regionmask` ([Hauser and Developers, 2024](#)).

## 3 Results

### 3.1 Initialisation with smoothed ERA5 fields

In this section, we assess whether Pangu-Weather produces meaningful outputs when initialized from smoothed, i.e., coarsened, data, whether we can combine the resulting forecasts into a long-term time series of meteorological fields, and how the statistics of these fields compare to the original ERA5 data.

We initialize the Pangu-Weather using the original ERA5 data, interpolated to a regular grid (AI orig experiment), and data smoothed with different  $\sigma$  values (standard deviations for the Gaussian kernel):  $\sigma = 2$  (AI  $2\sigma$ ),  $\sigma = 5$  (AI  $5\sigma$ ), and  $\sigma = 10$  (AI  $10\sigma$ ). The approximation of the equivalent resolution after smoothing can be done by using full width at half maximum of the Gaussian function by multiplying the initial resolution by  $2.355 \times \sigma$ . Given that the  $0.25^\circ$  regular grid corresponds to approximately 25 km resolution at the equator, AI  $2\sigma$  approximates a 100 km resolution, AI  $5\sigma$  corresponds to about 300 km resolution, and AI  $10\sigma$  to 500 km resolution. These estimations are approximate and do not account for the original ERA5 resolution of approximately 31 km, the variation in horizontal resolution of regular grids with latitude, or the inability of a 500 km resolution model to resolve certain features, such as atmospheric front positions, that are present in the smoothed data.

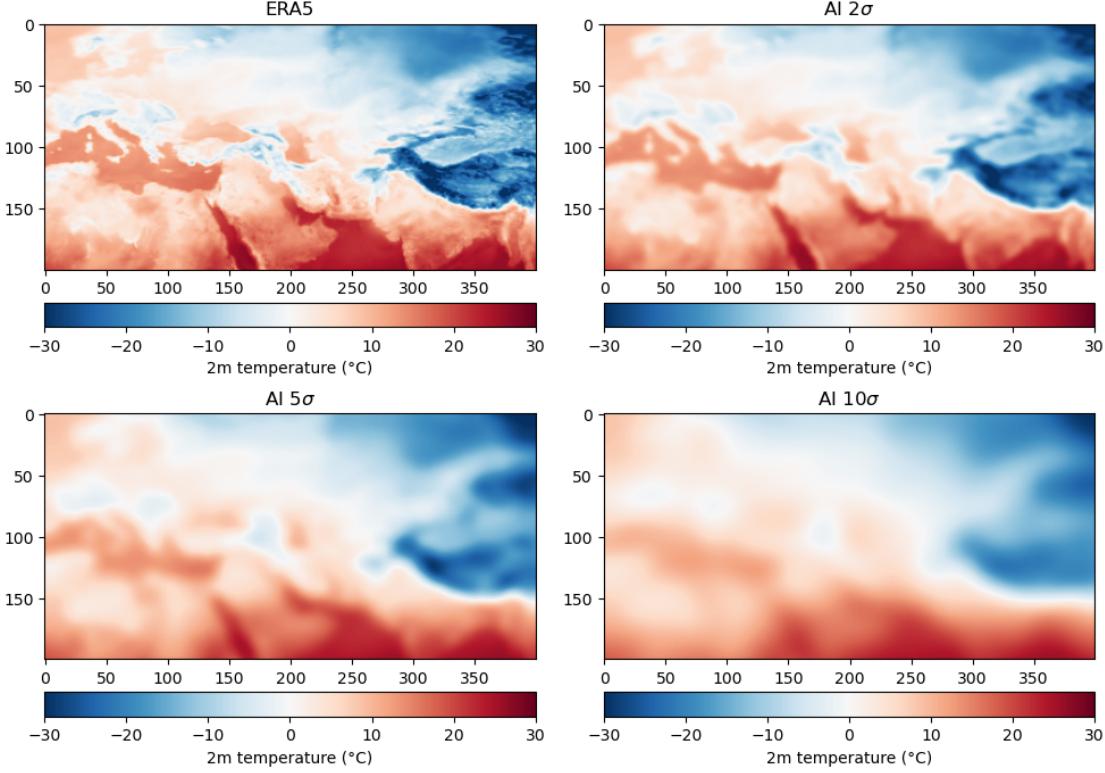


Figure 1: The 2-metre temperature used as part of the initial conditions (2020-01-01T00:00:00) with several levels of Gaussian smoothing. All other initial fields were smoothed analogously to emulate the fields generated by atmospheric models with different horizontal resolutions.

The initial 2-metre temperature fields provided to AI-NWP are shown in Figure 1. This figure displays parts of Europe, Asia, and Africa for the original ERA5 data (used to initialize the AI orig experiment) and data with varying levels of smoothing. For AI  $2\sigma$  and AI  $5\sigma$ , a considerable amount of detail remains visible. However, AI  $10\sigma$  retains primarily large-scale features. Figure A1 presents a global view of the same fields.

After one day of forecast, the 2-metre temperature fields initialized with different levels of smoothing converge significantly in terms of detail (Figure 2). The contrast between initial fields of AI  $5\sigma$  and  $10\sigma$  and the AI-NWP forecast is noticeable. Numerous small-scale details emerge, distinctions between land and ocean areas become evident, and mountain ridges and large valleys become visible in the temperature fields. Overall, the level of detail resembles that of the original ERA5 fields. Relatively high level of detail appears as early as the first forecast step (Figure A2).

Next we examine how initialization from smoothed initial conditions affects the statistics of the resulting temperature fields. We conducted a 10-day forecast for each day of the year 2020 using AI-NWP, initialized (at 00:00) from both the original (interpolated to a regular grid) and smoothed ERA5 data. We then constructed two annual datasets of atmospheric 2-metre temperature fields: one consisting of all first days (1-day), and the second dataset concatenating all second days (2-day) of the individual forecasts. For the 1-day dataset, the first time step is the initial condition, and the lead hours 6, 12, and 18 are from AI-NWP forecasts. For the 2-day dataset, all time steps are from AI-NWP forecasts (lead times of 24, 30, 36, and 42 hours).

Smoothing	1 day, °C	2 day, °C
AI orig.	-0.004	-0.032
AI $2\sigma$	-0.023	-0.069
AI $5\sigma$	-0.065	-0.129
AI $10\sigma$	-0.196	-0.177

Table 1: Mean differences in global mean 2-metre temperature time series relative to ERA5 (refer to Fig A3 for the time series of the differences).

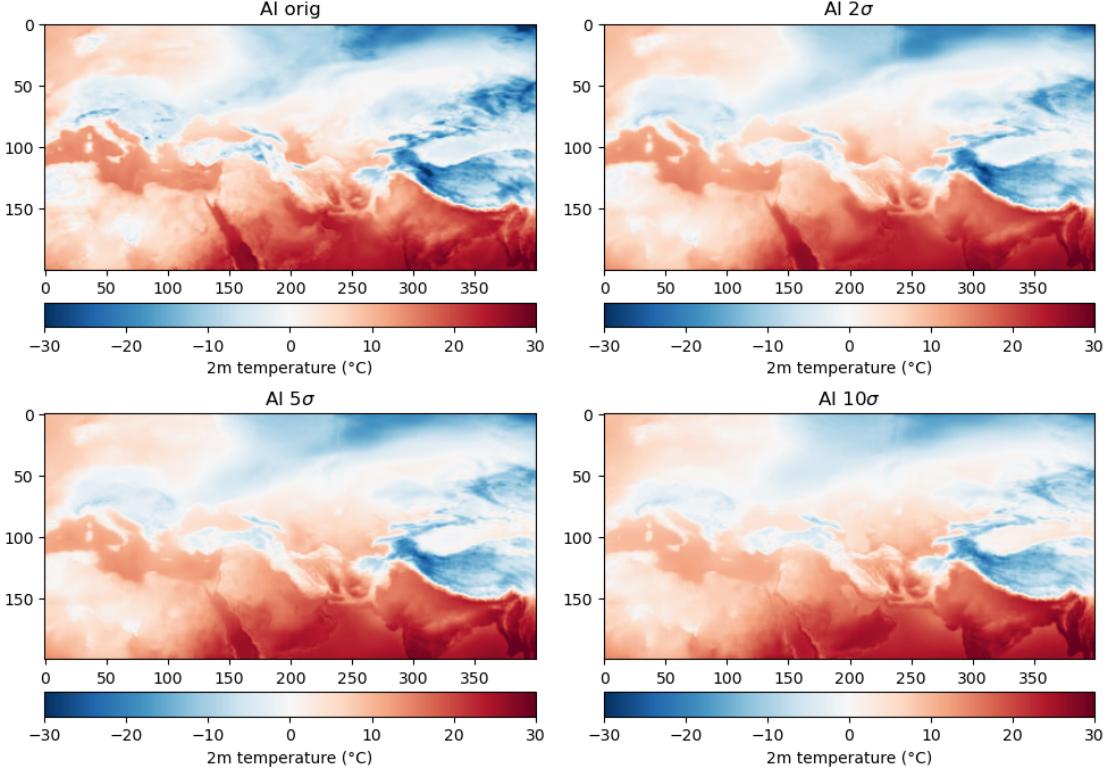


Figure 2: The 2-metre temperature for a day forecast (2020-01-02T00:00:00) using the AI-NWP for simulations with varying smoothing of initial conditions. After one day, the AI-NWP enhances the level of detail to a rate comparable to the fields it was originally trained on (ERA5).

The global mean 2-metre temperature time series computed from the 1-day and 2-day datasets closely match the ERA5 data for AI-NWP forecasts initialized from unsmoothed initial conditions (Figure 3, first row, right). The 1-day line is nearly indistinguishable from the original ERA5 line, with a mean difference of  $-0.004^{\circ}\text{C}$  (Table 1). This indicates that the approach of creating datasets of meteorological parameters by "filling" the gaps between ERA5 daily time steps with AI-NWP-generated data performs well. The 2-day line, although still close to the ERA5, exhibits larger differences, indicating some degradation of the results, with the mean difference becoming  $-0.032^{\circ}\text{C}$  (Table 1). The spatial distribution of temperature differences for the AI orig. experiment (Figure A3, first row, left and middle) shows relatively small biases for both the 1-day and 2-day datasets, with predominantly negative bias over the ocean and predominantly positive bias over the land.

The differences between ERA5 data and AI-NWP experiments with smoothed initial conditions increase with the level of smoothing (Figure 3). Notably, for AI  $5\sigma$  and AI  $10\sigma$ , some spatial biases over land for 2-day data are smaller than those for 1-day data, particularly over regions such as Africa and Australia. This suggests an adjustment from the smoothed initial conditions, which is evident in the mean 2-metre temperature RMSE values across different experiments (Figure A4). This adjustment is also reflected in the larger mean differences in global 2-metre temperature for the 2-day compared to the 1-day forecast in the AI  $10\sigma$  experiment (Table 1). The overall cool bias is primarily due to a uniform cooling effect over the ocean.

The results presented in this section indicate that AI-NWP can be initialized from smoothed initial conditions, and the resulting forecasts will have increased levels of detail, matching the resolution of the data on which AI-NWP was trained after a few forecasting steps. Year-long 2-metre temperature datasets, constructed from the first and second days of AI-NWP forecasts, represent well the temporal evolution and spatial distribution of the original data. The quality of the results decreases with increased smoothing of initial conditions, but for AI  $5\sigma$ , which is approximately equivalent to a 300 km resolution, the quality remains acceptable (for example well below typical biases of CMIP6 models (Fan et al., 2020)). The data for the second day of the forecast (2-day) exhibit smaller biases over land in the AI  $5\sigma$  experiment. Therefore, for further experiments with climate models, we will use the 2-day data.

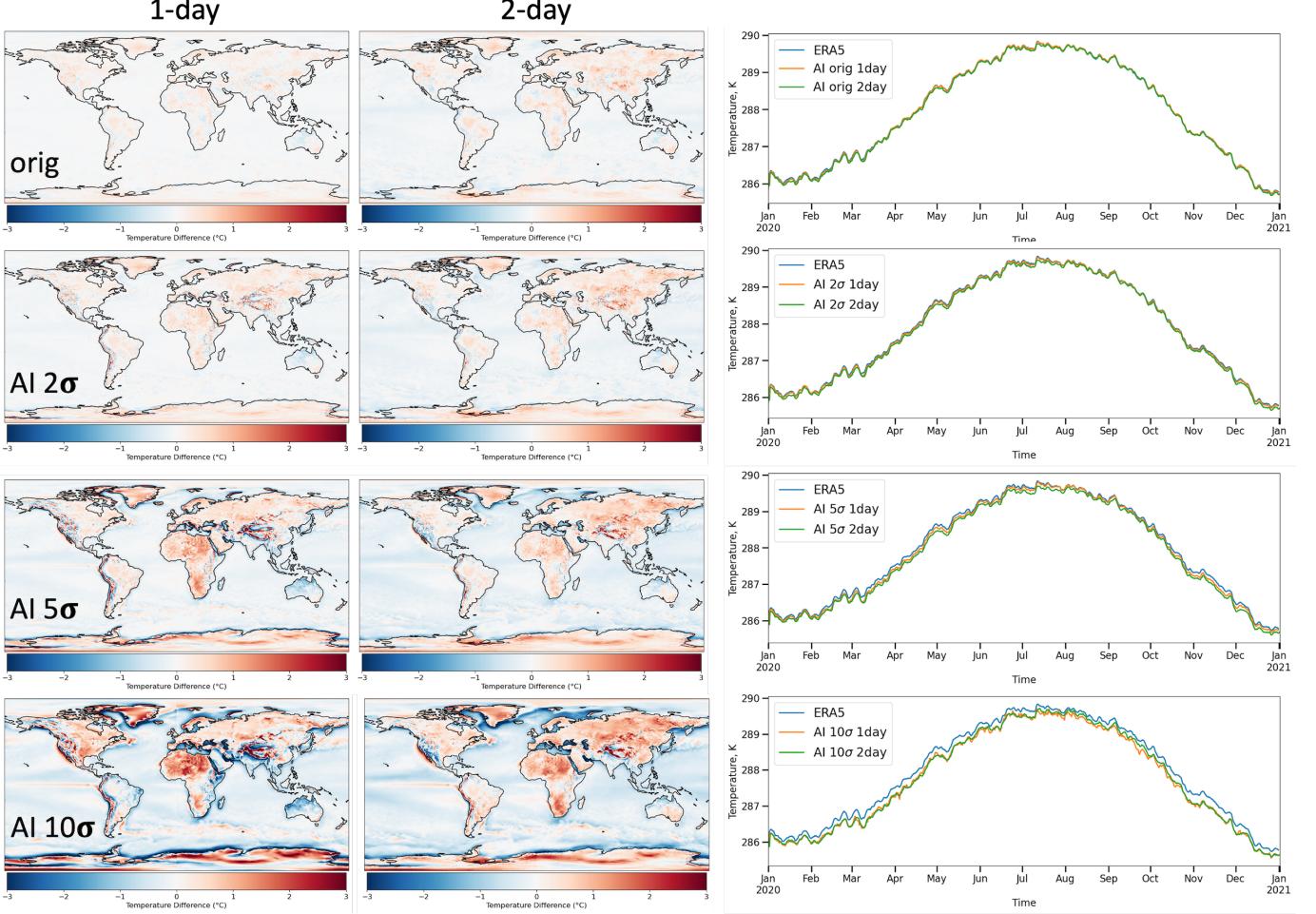


Figure 3: Annual mean 2-metre temperature difference relative to ERA5 (maps) and time series of global mean 2-metre temperature (graphs, 6-hour values, daily rolling mean). Values are compared for time series constructed from the first (1-day) and second (2-day) days of the AI-NWP forecast with varying levels of initial condition smoothing (original and multiple values of  $\sigma$ , the standard deviation for the Gaussian kernel). The level of  $\sigma = 5$  corresponds to fields from 300 km resolution models and can serve as a reference for low-resolution climate model fields. Constructing 6-hour time series from the first and second days of the forecast is effective, reproducing ERA5 properties well when initialized from original ERA5 data. Smoothing initial conditions tends to push the model towards a colder state. The main source of the cold bias appears to be the temperature over the ocean.

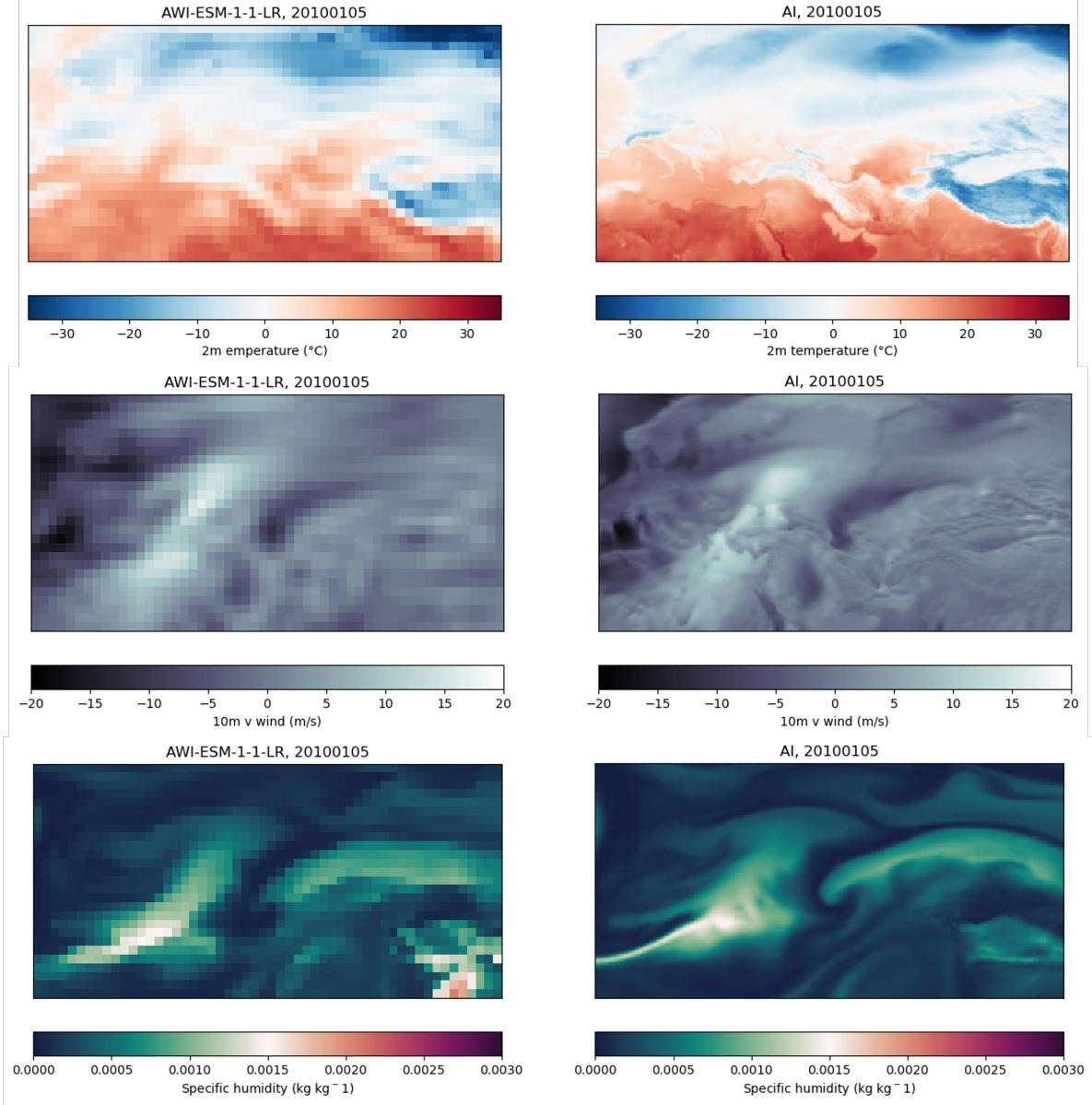


Figure 4: Snapshot from 2010-01-05 comparing original fields from a low-resolution climate model (left) and the result of an AI-NWP forecast initialized 2 days prior from climate model data (right). The region displayed includes parts of Europe, Africa, and Asia.

### 3.2 Initialisation with historical climate model data

Next, we use low-resolution climate models (AWI-CM-LR and MPI-ESM1-2-LR) to initialize AI-NWP forecasts at 00:00 UTC each day for one model year (2010) from historical experiments. The AI-NWP forecasts rapidly developed a level of detail comparable to the ERA5 data, as demonstrated for 2-metre temperature, 10-meter v wind, and specific humidity fields over regions in Europe, Africa, and Asia for AI-ESM-1-1-LR (Fig. 4). Similarly to the ERA5 case, there is an improved contrast between ocean and land, as well as more detailed representation of orographic features. Global fields also indicate that the large-scale patterns from the input are largely reproduced in AI-NWP, but with a greater level of detail, which appears plausible from a dynamical perspective (Figure A5).

Utilizing the same methodology as in the previous section for ERA5, we constructed one-year-long datasets from the second days of AI-NWP forecasts performed daily over the course of a year. The annual mean temperature difference between these datasets (AI) and the original climate model data from the corresponding time period (interpolated to

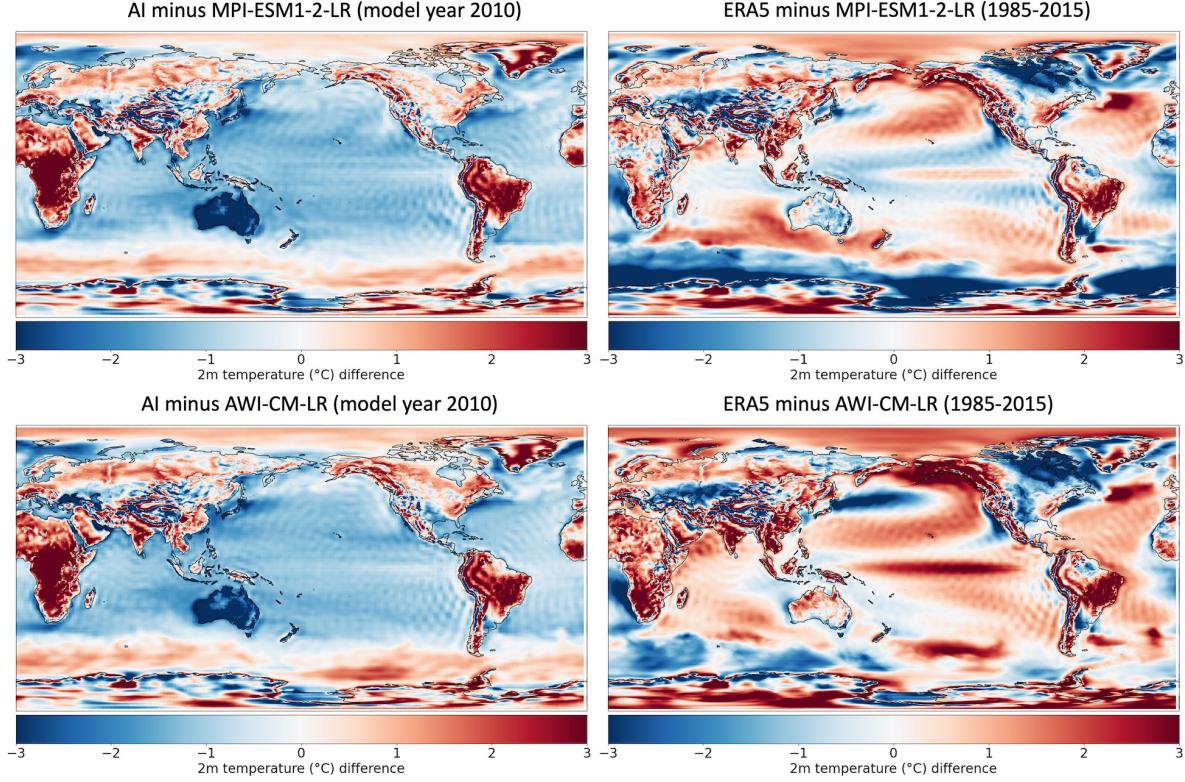


Figure 5: Difference in annual mean 2-metre temperature. (left) Between AI-NWP generated data and two CMIP6 climate models for model year 2010 (historical simulations). (right) Between ERA5 and climate models for the years 1985-2015 (historical simulations). The AI-NWP time series is derived from lead hours 24, 30, 36, and 42 (2-day dataset) of AI-NWP forecasts initiated from CMIP6 model fields. Two low-resolution (250km) models are used: AWI-CM-LR and MPI-ESM1-2-LR. Both models share the same ECHAM atmosphere but use different ocean components. This figure illustrates the bias of climate models relative to ERA5 and assesses whether the AI-NWP adjusts the climate model towards more realistic values. The AI-NWP model improves realism over large land areas, but problematic regions include the oceans, Australia, and northeastern North America.

a  $0.25^\circ$  grid) indicates the average direction of change that AI-NWP does for the temperature fields (Figure 5, left column). The differences for both climate models are similar in terms of magnitude and spatial distribution. Over most ocean areas, the temperature is lower in AI-NWP, except in the Southern Ocean, where it is warmer. Over land areas, AI-NWP data generally show higher temperatures, except for a significant negative difference over Australia.

To contextualize these differences, we compare the mean 2-metre temperature from 30 years of ERA5 data with climate model simulations (Figure 5, right column). Note, that we subtract model data from ERA5 rather than the reverse, to ensure comparability with the plot in the left column. In this way, similar colors indicate that the AI-NWP based dataset becomes closer to ERA5, and vice versa. The spatial distribution of differences relative to ERA5 for both climate models over land is similar, suggesting comparable model biases over land in historical simulations. When compared to the differences between AI-NWP data and climate models for one year, the spatial distribution, sign, and magnitude of the differences over land are quite similar. This suggests that the AI-NWP forecast moves the solution closer to ERA5, effectively performing a bias correction. Notable exceptions are Australia and northeastern North America, where the AI-NWP forecast is too cold. Additionally, temperatures over the ocean in AI-NWP-based data are too cold. Note that despite obvious and rather strong differences over the ocean in right plots for Figure 5, the difference in the left plot is remarkably similar, which is one more indication that solutions are moved in similar way. It should be also noted that for the AI-NWP based data, only one year is analyzed here as a demonstrator of the AI-NWP downscaling capability. A fairer comparison with the 30-year dataset will involve more AI-generated annual datasets in the future.

The AI-NWP-based 2-metre temperature time series over land closely follows the variability of climate model data (Figure 6, top). Differences (Figure 6, bottom) are mostly positive (indicating that AI-NWP data are warmer over land), with two pronounced maxima (January, August) and minima (March, November). These differences are remarkably

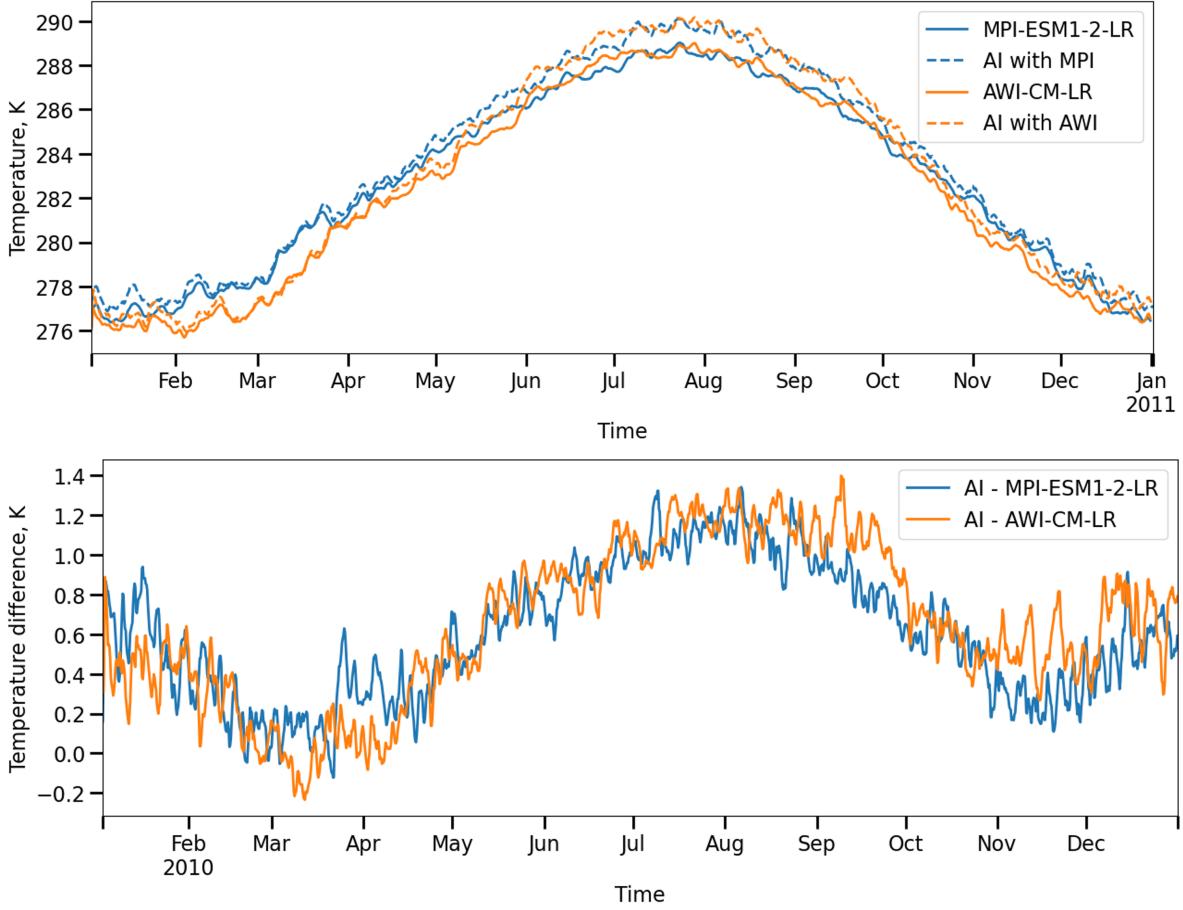


Figure 6: Top: Global mean 2-metre temperature time series for land areas only (6-hour values, 1-day rolling mean) for CMIP6 historical simulations (year 2010) and AI-NWP generated daily forecasts initialized from those simulations. The AI-NWP time series is derived from lead hours 24, 30, 36, and 42 (2-day dataset) of AI-NWP forecasts initiated from CMIP6 model fields. Bottom: Difference between CMIP6 and AI-NWP generated data. Two low-resolution (250km) models are used for initial conditions: AWI-CM-LR and MPI-ESM1-2-LR. The differences between AI-generated data and both models are of similar amplitude.

similar between the models. As the considered climate models exhibit mostly positive biases over land compared to ERA5, AI-NWP data over land are closer to realistic temperatures. The range of differences lies between about -0.2 K and 1.3 K and falls within the range of CMIP6 model biases. The time series that include ocean data show a negative bias, which is also within the 1 K range (Figure A6).

Initializing AI-NWP with low-resolution climate model data produces forecast fields with improved detail. These fields can be combined into a dataset that compares well with the original data in terms of global mean 2-metre temperature variability and spatial distribution of temperature over land, indicating a correction towards a state closer to ERA5, which is constrained by observations. Although our conclusions are based on two year-long simulations of two low-resolution climate models, the results demonstrate that downscaling of reasonable quality with current AI-NWP is feasible.

### 3.3 Initialisation with scenario simulation data

To evaluate the applicability of our approach to scenario climate simulations, we generated AI-NWP data for the years 2050 and 2099 using the MPI-ESM2-2-LR model (scenario SPP585). Similarly to the historical simulations, the second forecast days were used. The differences in yearly mean 2-metre temperature between AI-NWP generated data and original climate model data (Figure 9) show patterns and magnitudes similar to those observed in historical simulations (Figure 5). This similarity suggests that the AI-NWP attempts to correct systematic differences related to fast (daily)

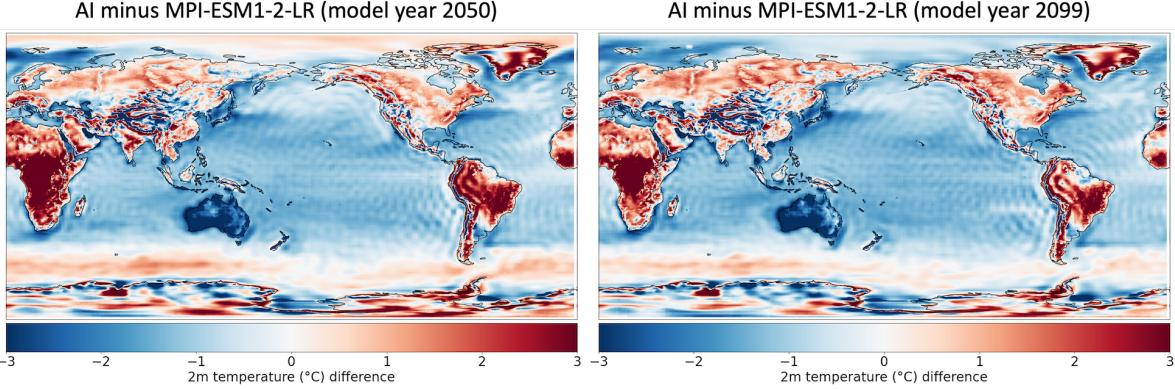


Figure 7: Difference in annual mean 2-metre temperature between AI-NWP generated data and MPI-ESM1-2-LR scenario simulations for model year 2050 (left) and 2099 (right). The AI-NWP time series is derived from lead hours 24, 30, 36, and 42 (2-day dataset) of AI-NWP forecasts initiated from MPI-ESM1-2-LR model fields.

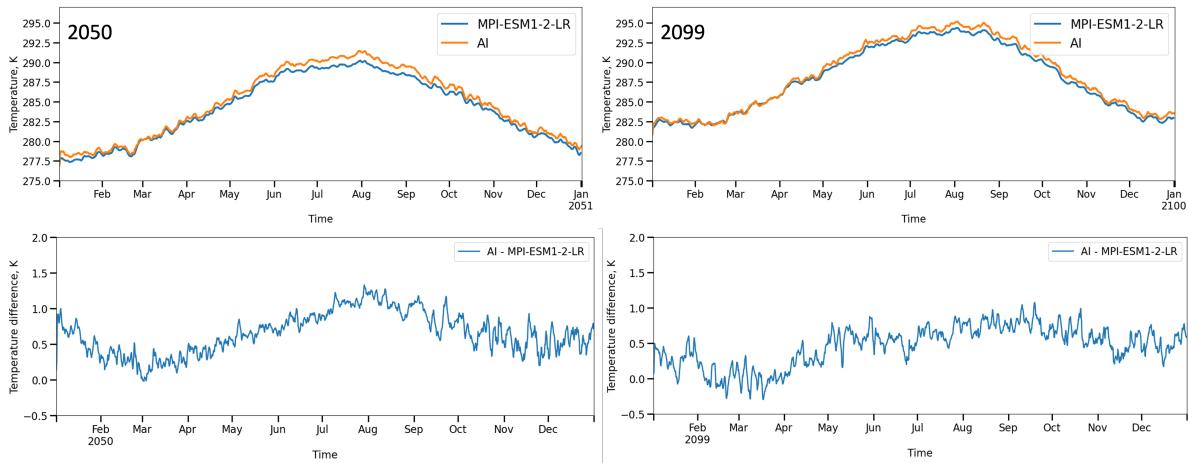


Figure 8: Top: Global mean 2-metre temperature time series for land areas only (6-hour values, 1-day rolling mean) for MPI-ESM1-2-LR scenario simulations (years 2050 and 2099, SSP585) and AI-NWP generated daily forecasts initialized from those simulations. The AI-NWP time series is derived from lead hours 24, 30, 36, and 42 (2-day dataset) of AI-NWP forecasts initiated from CMIP6 model fields. Bottom: Difference between MPI-ESM1-2-LR scenario simulations and AI-NWP generated data. The difference for both years are the same order of magnitude as for historical simulations.

atmospheric dynamics, such as the generation of low clouds in regions like central Africa and the northern part of South America. Over the oceans, the temperature is generally cooler, with the exception of the Southern Ocean.

The differences in the time series of 2-metre temperature between AI-NWP generated data and the climate model (Figure 8) for the years 2050 and 2099 are consistent with the results from the historical period (Figure 6). These differences are predominantly positive over land, as shown in Figure 8, and generally fall within the 1 K range for most of the time. The difference for the year 2099 exhibits a less pronounced seasonal cycle. The discrepancies between AI-based and original data are significantly smaller than the climate change signal.

To illustrate the potential spatial distribution of the climate change signal following downscaling with AI-NWP, we subtracted the data for the year 2010 from the year 2099 of the MPI-ESM-LR model (Figure 9). For an accurate climate signal, it would be necessary to average over more extended periods from both historical and scenario simulations and account for model drift. However, this figure serves as a representative demonstration. The figure indicates an improved localization of temperature differences, incorporating orographic details. The magnitude of differences between past and future years in the AI-NWP data is smaller and more diffuse compared to the more localized warming hotspots observed in the original MPI-ESM-LR data.

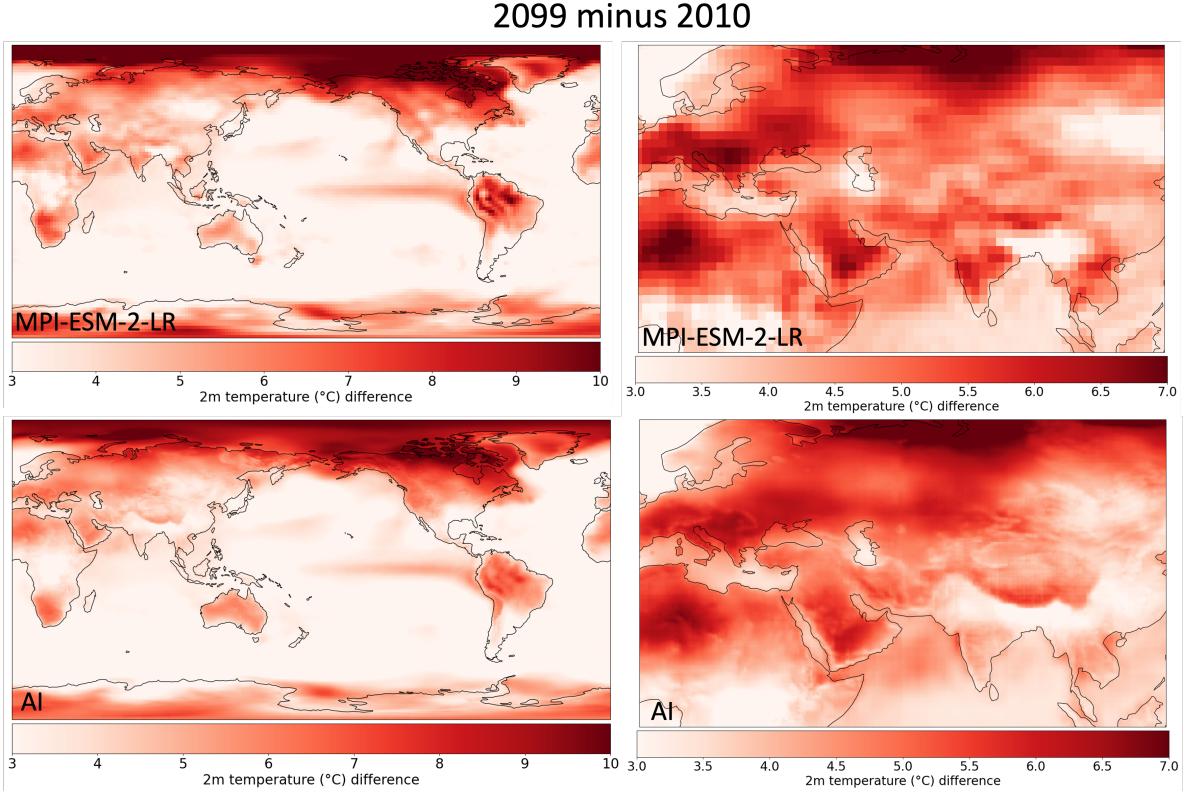


Figure 9: Global mean 2-metre temperature difference between model years of 2010 (historical) and 2099 (scenario, SSP585) of MPI-ESM1-2-LR simulations (top) and AI-NWP data generated from MPI-ESM1-2-LR data. Global view (left) and zoom to regions over Europe, Africa and Asia (right) to demonstrate level of details in the differences. Large scale spatial patterns are mostly the same.

When applied to scenario climate simulations, downscaling with AI-NWP functions similarly to its application to historical simulations. The spatial distribution of 2-metre temperature differences between AI-NWP-based data and climate model data is comparable, as are the differences in global mean 2-metre temperature. This suggests that the differences are more systematic in nature and are not significantly influenced by the application of AI-NWP, which was trained on ERA5 reanalysis, to conditions outside the training envelope.

## 4 Discussion and conclusions

Our study demonstrates that AI-based numerical weather prediction systems (AI-NWP) can effectively downscale low-resolution climate data to generate high-resolution atmospheric fields. By initializing AI-NWP with smoothed ERA5 data and low-resolution CMIP6 climate model data, we found that AI-NWP can develop detailed outputs within one forecast day with a resolution that closely resembles those of the training data. Our approach shows promise for creating high-resolution, long-term datasets from CMIP-type models. Since the AI-NWP models are trained on observationally well-constrained ERA5 reanalysis, they can potentially perform bias correction, moving climate model outputs closer to observed data. More work is, however, required to better understand the distributional shift that the models generate.

The dynamics of atmospheric fields generated by AI-NWP closely resembles that of the original climate models. This is particularly evident in the absence of additional mesoscale dynamics, such as the formation of new cyclones. Unlike regional climate models (RCMs), which can develop such dynamics when applied to sufficiently large regions, our method aligns more closely with statistical downscaling techniques rather than dynamical downscaling. And yet, a significant amount of detail in both cases is triggered by the responses to the details of the underlying surface which were absent in coarse simulations. It is likely that this part of small scale dynamics is rendered correctly in both cases.

Moreover the lack of mesoscale dynamics might be a consequence of the limited lead time in our experiments and further work is required to understand if this is an intrinsic limitation.

In this work, we have applied an AI-NWP downscaling method to individual years of climate model simulations. Although we can generate more detailed spatial fields (at higher resolution of  $\approx 30$  km), the AI results follow the climate trajectory realised by the coarse-resolution model. Hence, relevant feedbacks from small scale processes to the long-term trajectory of the simulation are still missing or only parameterised, such as feedbacks from mesoscale ocean eddies, or the response of clouds to warming. The method presented here can thus not replace kilometre-scale resolution (e.g. 2–5 km), eddy- and cloud-resolving climate simulations (Hohenegger et al., 2023; Rackow et al., 2024) that potentially resolve very different responses of the coupled atmosphere-ocean system to warming than typical CMIP models.

The resulting high-resolution data from AI-NWP can be used for several applications. They can serve as boundary conditions for further downscaling with regional climate models (RCMs). This eliminates the need for multiple nested downscaling steps, simplifying the process and reducing computational costs. Additionally, if AI-NWP models are enhanced to output hourly or even finer data, they can provide the temporal resolution required by sectors like wind energy for accurate statistical analyses. Beyond boundary conditions, the detailed spatial information generated by AI-NWP is valuable for climate adaptation and mitigation, enabling more precise risk assessments, infrastructure planning, and policy-making at local and regional scales. Aurora, a large-scale foundational model, already produces global output at  $0.1^\circ$  resolution (Bodnar et al., 2024), and other AI-NWP models will most probably catch up in the near future.

The fact that initial conditions can be smoothed with only a minor reduction in model skill indicates the potential for distributing lower-resolution initial conditions for initiatives like ECMWF's "forecast in the box", that would allow users to run forecast models locally to recreate full-resolution data. Current  $1/4$  degree initial conditions contain only few fields, and require just about 140MB of disk space. However, if tens of thousands of users are supposed to be served daily, even small optimizations in the volume of distributed data will count, especially if models use more fields for initial conditions and higher target resolutions.

This work represents a first step, leveraging AI-NWP tools that are not specifically designed for climate downscaling purposes. A natural next step would be to fine-tune the models for the downscaling task. Alternative approaches, such as foundation models trained on extensive climate datasets and subsequently fine-tuned for global or regional downscaling, may yield better results in the future. However, a compelling aspect of our study is the use of AI-based weather prediction systems, which are readily available and likely to continue receiving substantial resources for development from both scientific and industrial sectors. These systems will likely incorporate the most advanced atmospheric data for training, optimized for various hardware configurations and have user-friendly interfaces. Consequently, directly applying AI-NWP to climate data could prove more straightforward and lead to better skill than developing specialized solutions. Moreover, our approach offers potential for identifying deficiencies in AI-NWP models, such as the too cold temperature over the ocean for Pangu-Weather, and incorporates a built-in bias correction mechanism.

Future research could focus on testing other AI-NWP models and expanding the analysis to include a wider range of variables. We plan to downscale complete historical and scenario simulations for various climate models, particularly those representing different ends of the spectrum in terms of climate sensitivity and the quality of historical simulations. This will help refine the methodology, enhance the accuracy of downscaled data, and help to understand the applicability of AI-NWP in diverse climatic contexts, in particular if ERA5-trained models can be used for different scenario runs. It will also be interesting to compare results of the downscaling with RCMs and with AI-NWP, as well as to try driving RCMs with AI-NWP generated data.

## 5 Acknowledgements

The computing time and storage for Pangu-Weather simulations and data processing is provided by Deutsches Klimarechenzentrum (DKRZ). Nikolay Koldunov and Sergey Danilov are supported by projects S1: Diagnosis and Metrics in Climate Models and M3: Towards Consistent Subgrid Momentum Closures of the Collaborative Research Centre TRR 181 "Energy Transfer in Atmosphere and Ocean", funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project no. 274762653). Thomas Rackow has been supported by the European Commission Horizon 2020 Framework Programme nextGEMS (grant no. 101003470). Suvarchal K. Cheedela is supported by BMBF-funded project WarmWorld. Thomas Jung is supported by the EERIE project (Grant Agreement No 101081383) funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Climate Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

## References

- Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M. (2020). Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, 13(4):2109–2124.
- Ben-Bouallegue, Z., Clare, M. C. A., Magnusson, L., Gascon, E., Maier-Gerber, M., Janousek, M., Rodwell, M., Pinault, F., Dramsch, J. S., Lang, S. T. K., Raoult, B., Rabier, F., Chevallier, M., Sandu, I., Dueben, P., Chantry, M., and Pappenberger, F. (2023). The rise of data-driven weather forecasting.
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970):533–538.
- Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J., Dong, H., Vaughan, A., Gupta, J. K., Tambiratnam, K., Archibald, A., Heider, E., Welling, M., Turner, R. E., and Perdikaris, P. (2024). Aurora: A foundation model of the atmosphere.
- Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., and Li, H. (2023). FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, 6(1):190.
- Damm, A., Köberl, J., Stegmaier, P., Alonso, E. J., and Harjanne, A. (2020). The market for climate services in the tourism sector—An analysis of Austrian stakeholders' perceptions. *Climate Services*, 17:100094.
- Danek, C., Shi, X., Stepanek, C., Yang, H., Barbi, D., Hegewald, J., and Lohmann, G. (2020). AWI AWI-ESM1.1LR model output prepared for CMIP6 CMIP historical.
- ECMWF (2024). climetlab, <https://github.com/ecmwf/climetlab>. Accessed: 2024-06-23.
- Ekström, M., Grose, M. R., and Whetton, P. H. (2015). An appraisal of downscaling methods used in climate change research. *Wiley Interdisciplinary Reviews: Climate Change*, 6(3):301–319.
- Fan, X., Duan, Q., Shen, C., Wu, Y., and Xing, C. (2020). Global surface air temperatures in CMIP6: Historical performance and future changes. *Environmental Research Letters*, 15(10):104056.
- Fowler, H. J., Blenkinsop, S., and Tebaldi, C. (2007). Linking climate change modelling to impacts studies.
- Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., et al. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, 5(3):572–597.
- Giorgi, F. (2019). Thirty years of regional climate modeling: where are we and where are we going next? *Journal of Geophysical Research: Atmospheres*, 124(11):5696–5723.
- Giorgi, F. and Gutowski Jr, W. J. (2015). Regional dynamical downscaling and the CORDEX initiative. *Annual review of environment and resources*, 40:467–490.
- Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., et al. (2016). High resolution model intercomparison project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, 9(11):4185–4208.
- Hakim, G. J. and Masanam, S. (2023). Dynamical tests of a deep-learning weather prediction model.
- Hanke, M., Redler, R., Holfeld, T., and Yastremsky, M. (2016). YAC 1.2. 0: new aspects for coupling software in Earth system modelling. *Geoscientific Model Development*, 9(8):2755–2769.
- Harris, L., McRae, A. T., Chantry, M., Dueben, P. D., and Palmer, T. N. (2022). A generative deep learning approach to stochastic downscaling of precipitation forecasts. *Journal of Advances in Modeling Earth Systems*, 14(10):e2022MS003120.
- Hauser, M. and Developers (2024). regionmask, <https://github.com/regionmask/regionmask/>. Accessed: 2024-06-23.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.
- Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., Bao, J., Bastin, S., Behravesh, M., Bergemann, M., Biercamp, J., Bockelmann, H., Brokopf, R., Brüggemann, N., Casaroli, L., Chegini, F., Datseris, G., Esch, M., George, G., Giorgetta, M., Gutjahr, O., Haak, H., Hanke, M., Ilyina, T., Jahns, T., Jungclaus, J., Kern, M., Klocke, D., Kluft, L., Kölling, T., Kornblueh, L., Kosukhin, S., Kroll, C., Lee, J., Mauritsen, T., Mehlmann, C., Mieslinger, T., Naumann, A. K., Paccini, L., Peinado, A., Praturi, D. S., Putrasahan, D., Rast, S., Riddick, T., Roeber, N., Schmidt, H., Schulzweida, U., Schütte, F., Segura, H., Shevchenko, R., Singh, V., Specht, M., Stephan, C. C., von Storch, J.-S., Vogel, R., Wengel, C., Winkler, M., Ziemen, F., Marotzke, J., and Stevens, B. (2023). ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales. *Geoscientific Model Development*, 16(2):779–811.

- Höhlein, K., Kern, M., Hewson, T., and Westermann, R. (2020). A comparative study of convolutional neural network models for wind field downscaling. *Meteorological Applications*, 27(6):e1961.
- Hoyer, S. and Hamman, J. (2017). xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, 5(1).
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95.
- Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and Von Storch, J. (2013). Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *Journal of Advances in Modeling Earth Systems*, 5(2):422–446.
- Koldunov, N. V., Kumar, P., Rasmussen, R., Ramanathan, A., Nesje, A., Engelhardt, M., Tewari, M., Haensler, A., and Jacob, D. (2016). Identifying climate change information needs for the himalayan region: Results from the GLACINDIA stakeholder workshop and training program. *Bulletin of the American Meteorological Society*, 97(2):ES37–ES40.
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al. (2023). Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421.
- Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., Clare, M. C., Lessig, C., Maier-Gerber, M., Magnusson, L., et al. (2024). Aifs-ecmwf's data-driven forecasting system. *arXiv preprint arXiv:2406.01465*.
- Lessig, C., Luise, I., Gong, B., Langguth, M., Stadtler, S., and Schultz, M. (2023). AtmoRep: A stochastic model of atmosphere dynamics using large scale representation learning.
- Mardani, M., Brenowitz, N., Cohen, Y., Pathak, J., Chen, C.-Y., Liu, C.-C., Vahdat, A., Kashinath, K., Kautz, J., and Pritchard, M. (2023). Residual diffusion modeling for km-scale atmospheric downscaling.
- Met Office (2010 - 2015). *Cartopy: a cartographic python library with a Matplotlib interface*. Exeter, Devon.
- Miles, A., jakirkham, Bussonnier, M., Moore, J., Orfanos, D. P., Bennett, D., Stansby, D., Hamman, J., Bourbeau, J., Fulton, A., Lee, G., Abernathey, R., Rzepka, N., Patel, Z., Kristensen, M. R. B., Verma, S., Chopra, S., Rocklin, M., AWA, A. B., Jones, M., Durant, M., de Andrade, E. S., Schut, V., raphael dussin, Chaudhary, S., Barnes, C., Nunez-Iglesias, J., and shikharsg (2024). zarr-developers/zarr-python: v3.0.0-alpha.
- Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K., and Grover, A. (2023). ClimaX: A foundation model for weather and climate.
- Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K., et al. (2022). Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*.
- Rackow, T., Goessling, H. F., Jung, T., Sidorenko, D., Semmler, T., Barbi, D., and Handorf, D. (2018). Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part II: climate variability. *Climate Dynamics*, 50:2369–2394.
- Rackow, T., Pedruzo-Bagazgoitia, X., Becker, T., Milinski, S., Sandu, I., Aguridan, R., Bechtold, P., Beyer, S., Bidlot, J., Boussetta, S., Diamantakis, M., Dueben, P., Dutra, E., Forbes, R., Goessling, H. F., Hadade, I., Hegewald, J., Keeley, S., Kluft, L., Koldunov, N., Koldunov, A., Kölling, T., Kousal, J., Mogensen, K., Quintino, T., Polichtchouk, I., Sármány, D., Sidorenko, D., Streffing, J., Sützl, B., Takasuka, D., Tietsche, S., Valentini, M., Vannière, B., Wedi, N., Zampieri, L., and Ziemen, F. (2024). Multi-year simulations at kilometre scale with the Integrated Forecasting System coupled to FESOM2.5/NEMOv3.4. *EGUphere*, 2024:1–59.
- Rampal, N., Gibson, P. B., Sood, A., Stuart, S., Fauchereau, N. C., Brandolini, C., Noll, B., and Meyers, T. (2022). High-resolution downscaling with interpretable deep learning: Rainfall extremes over New Zealand. *Weather and Climate Extremes*, 38:100525.
- Ranasinghe, R., Ruane, A. C., Vautard, R., Arnell, N., Coppola, E., Cruz, F. A., Dessai, S., Saiful Islam, A., Rahimi, M., Carrascal, D. R., et al. (2021). Climate change information for regional impact and for risk assessment.
- Raoult, B., Mertes, G., and Pinault, F. (2024). ai-models, <https://github.com/ecmwf-lab/ai-models>. Accessed: 2024-06-23.
- Rocklin, M. (2015). Dask: Parallel computation with blocked algorithms and task scheduling. In Huff, K. and Bergstra, J., editors, *Proceedings of the 14th Python in Science Conference*, pages 130 – 136.
- Ruti, P. M., Somot, S., Giorgi, F., Dubois, C., Flaounas, E., Obermann, A., Dell'Aquila, A., Pisacane, G., Harzallah, A., Lombardi, E., et al. (2016). MED-CORDEX initiative for mediterranean climate studies. *Bulletin of the American Meteorological Society*, 97(7):1187–1208.

- Schulzweida, U. (2023). Cdo user guide.
- Semmler, T., Danilov, S., Gierz, P., Goessling, H. F., Hegewald, J., Hinrichs, C., Koldunov, N., Khosravi, N., Mu, L., Rackow, T., et al. (2020). Simulations for CMIP6 with the AWI climate model AWI-CM-1-1. *Journal of Advances in Modeling Earth Systems*, 12(9):e2019MS002009.
- Semmler, T., Jungclaus, J., Danek, C., Goessling, H. F., Koldunov, N. V., Rackow, T., and Sidorenko, D. (2021). Ocean model formulation influences transient climate response. *Journal of Geophysical Research: Oceans*, 126(12):e2021JC017633.
- Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., Dethloff, K., Dorn, W., Fieg, K., Gößling, H. F., et al. (2015). Towards multi-resolution global climate modeling with ECHAM6–FESOM. Part I: model formulation and mean climate. *Climate Dynamics*, 44:757–780.
- Stengel, K., Glaws, A., Hettinger, D., and King, R. N. (2020). Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29):16805–16815.
- Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., et al. (2013). Atmospheric component of the MPI-M Earth system model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, 5(2):146–172.
- Tomasi, E., Franch, G., and Cristoforetti, M. (2024). Can AI be enabled to dynamical downscaling? training a latent diffusion model to mimic km-scale COSMO-CLM downscaling of ERA5 over Italy.
- van den Hurk, B., Hewitt, C., Jacob, D., Bessembinder, J., Doblas-Reyes, F., and Döscher, R. (2018). The match between climate services demands and Earth System Models supplies. *Climate services*, 12:59–63.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.
- Wang, J., Liu, Z., Foster, I., Chang, W., Kettimuthu, R., and Kotamarthi, V. R. (2021). Fast and accurate learned multiresolution dynamical downscaling for precipitation. *Geoscientific Model Development*, 14(10):6355–6372.
- Wang, Q., Danilov, S., Sidorenko, D., Timmermann, R., Wekerle, C., Wang, X., Jung, T., and Schröter, J. (2014). The Finite Element Sea Ice-Ocean Model (FESOM) v. 1.4: formulation of an ocean general circulation model. *Geoscientific Model Development*, 7(2):663–693.
- Wieners, K.-H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., Bittner, M., Gayler, V., Haak, H., de Vrese, P., Raddatz, T., Mauritsen, T., von Storch, J.-S., Behrens, J., Brovkin, V., Claussen, M., Crueger, T., Fast, I., Fiedler, S., Hagemann, S., Hohenegger, C., Jahns, T., Kloster, S., Kinne, S., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Müller, W., Nabel, J., Notz, D., Peters-von Gehlen, K., Pincus, R., Pohlmann, H., Pongratz, J., Rast, S., Schmidt, H., Schnur, R., Schulzweida, U., Six, K., Stevens, B., Voigt, A., and Roeckner, E. (2019a). MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 ScenarioMIP SSP585.
- Wieners, K.-H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., Bittner, M., Legutke, S., Schupfner, M., Wachsmann, F., Gayler, V., Haak, H., de Vrese, P., Raddatz, T., Mauritsen, T., von Storch, J.-S., Behrens, J., Brovkin, V., Claussen, M., Crueger, T., Fast, I., Fiedler, S., Hagemann, S., Hohenegger, C., Jahns, T., Kloster, S., Kinne, S., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Müller, W., Nabel, J., Notz, D., Peters-von Gehlen, K., Pincus, R., Pohlmann, H., Pongratz, J., Rast, S., Schmidt, H., Schnur, R., Schulzweida, U., Six, K., Stevens, B., Voigt, A., and Roeckner, E. (2019b). MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 CMIP historical.

## A Appendix A: Additional Information

This appendix contains supplementary figures.

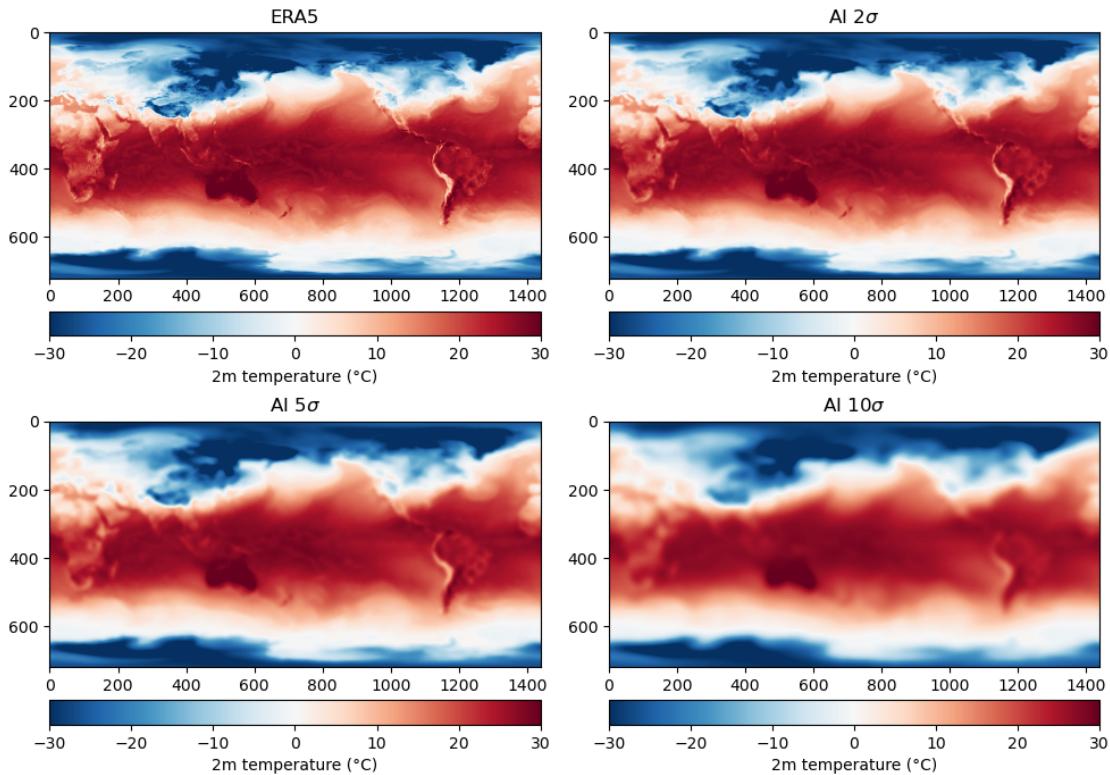


Figure A1: The 2-metre temperature used as part of the initial conditions (2020-01-01T00:00:00) with several levels of Gaussian smoothing. All other initial fields were smoothed similarly to emulate the fields generated by atmospheric models with different horizontal resolutions. Global view.

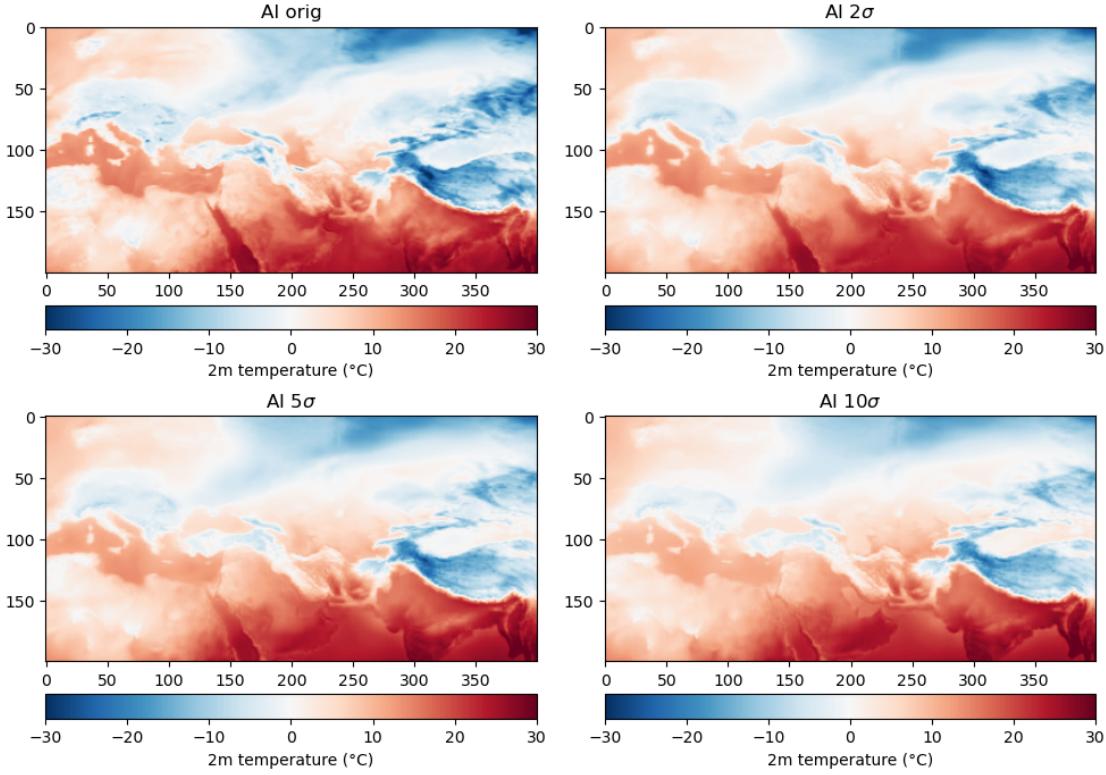


Figure A2: The 2-metre temperature after 6 hours of forecast (2020-01-01T06:00:00) using the AI-WPS for simulations with varying smoothing of initial conditions. Already after 6 hours the AI-WPS shows good level of detail comparable to the fields it was originally trained on (ERA5).

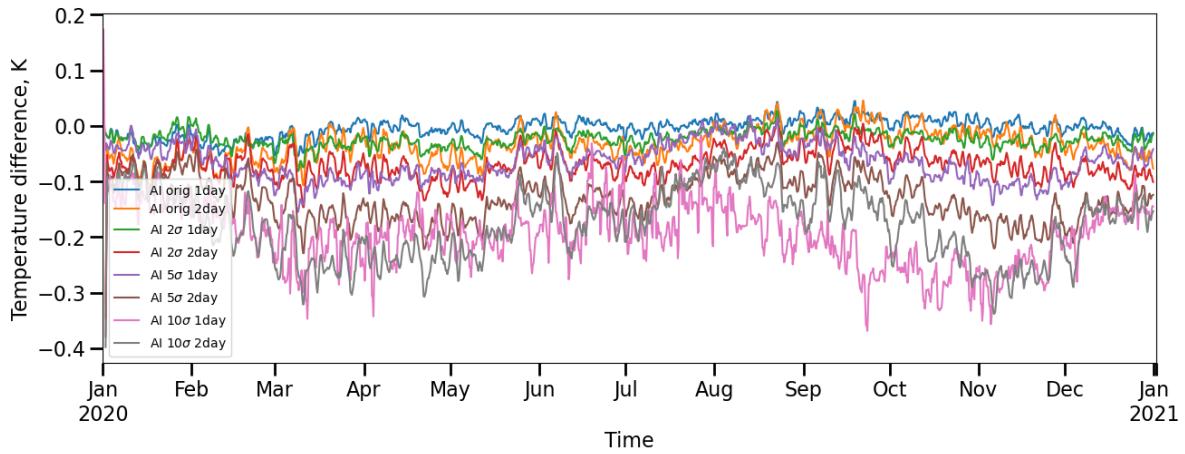


Figure A3: Difference in global mean 2-metre temperature time series, constructed from AI-NWP data with varying initial condition smoothing and lead times (1-day, 2-day), compared to ERA5 time series (6-hour values, daily rolling mean). These differences correspond to the graphs in Fig. 3. Mean values are listed in Table 1.

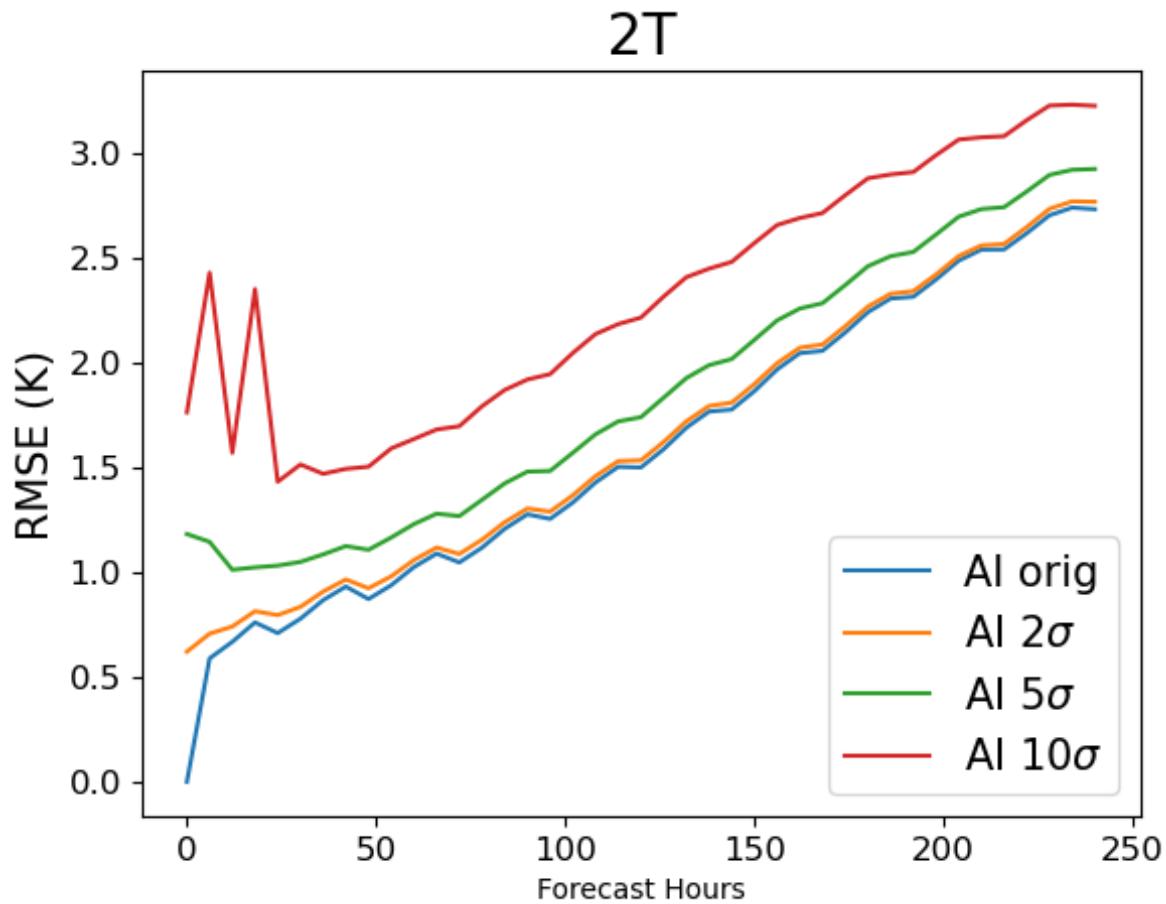


Figure A4: Root Mean Square Error (RMSE) of global 2-metre temperature at different levels of smoothing relative to ERA5. Mean values over one year (2020).

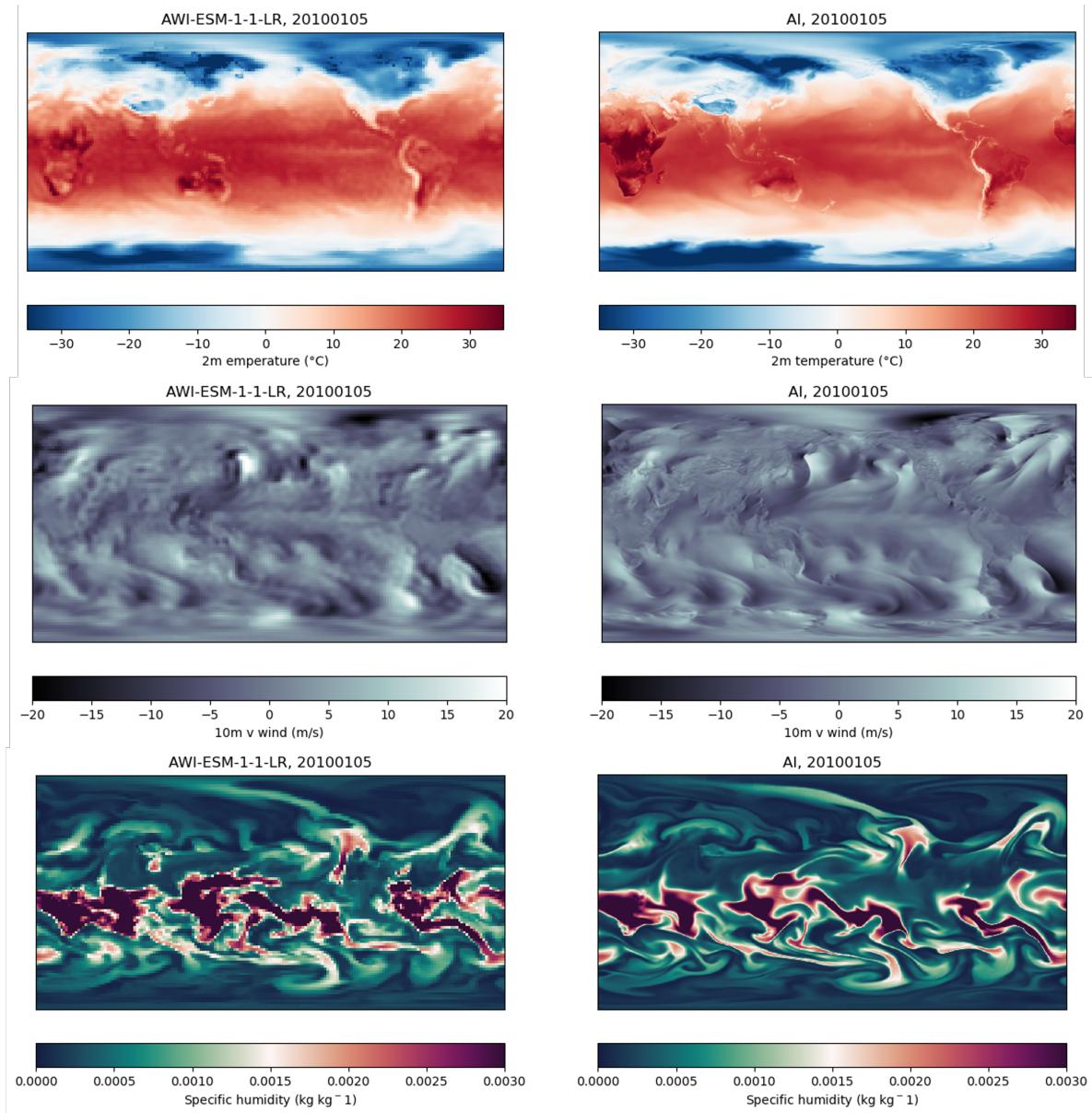


Figure A5: Snapshot from 2010-01-05 comparing original fields from a low-resolution climate model (left) and the result of an AI-NWP forecast initialized 2 days prior from climate model data (right). Same as Fig 4, but global.

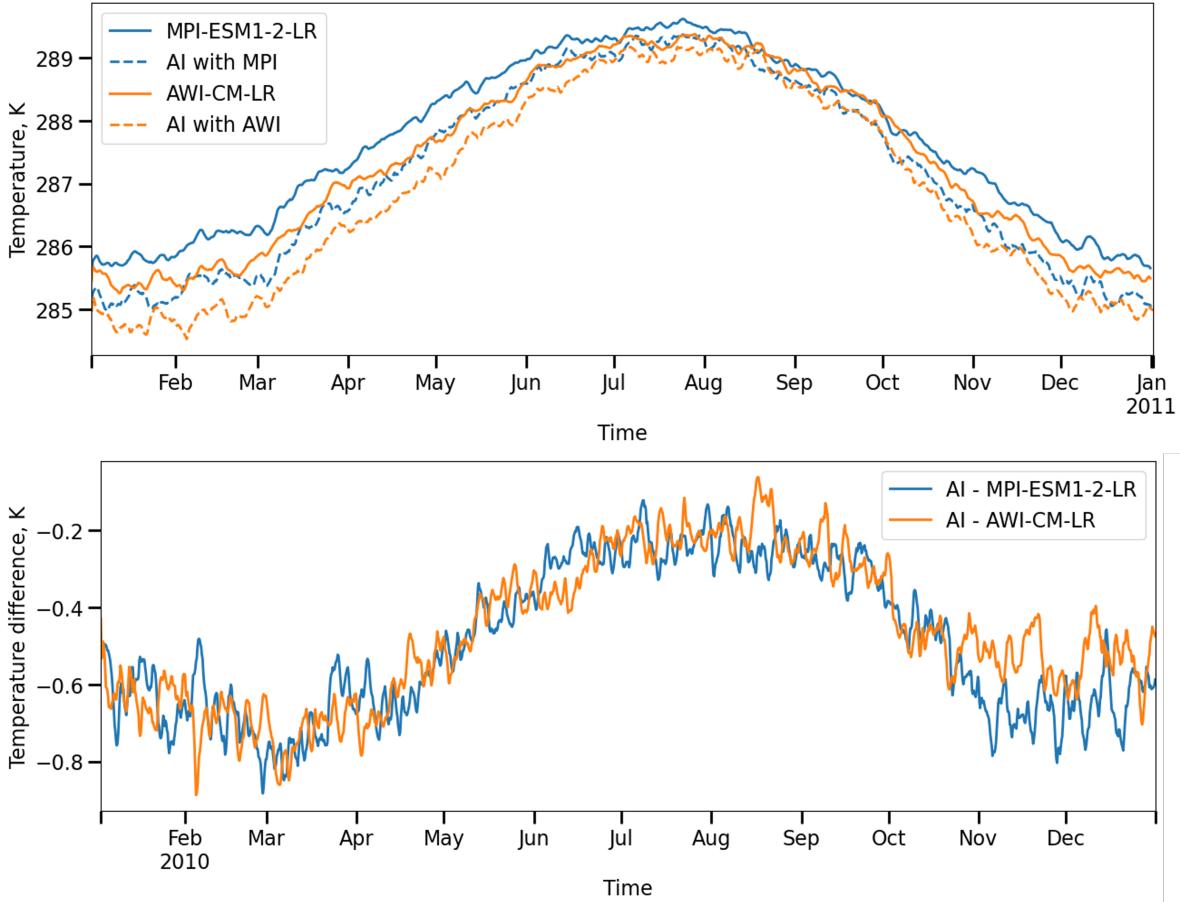


Figure A6: Top: Global mean 2-metre temperature time series (6-hour values, 1-day rolling mean) for CMIP6 historical simulations (year 2010) and AI-NWP generated daily forecasts initialized from those simulations. The AI-NWP time series is derived from lead hours 24, 30, 36, and 42 (2-day dataset) of AI-NWP forecasts initiated from CMIP6 model fields. Bottom: Difference between CMIP6 and AI-NWP generated data. Two low-resolution (250km) models are used: AWI-CM-LR and MPI-ESM1-2-LR. Both models share the same ECHAM atmosphere but use different ocean components. The differences between AI-generated data and both models are of similar amplitude.

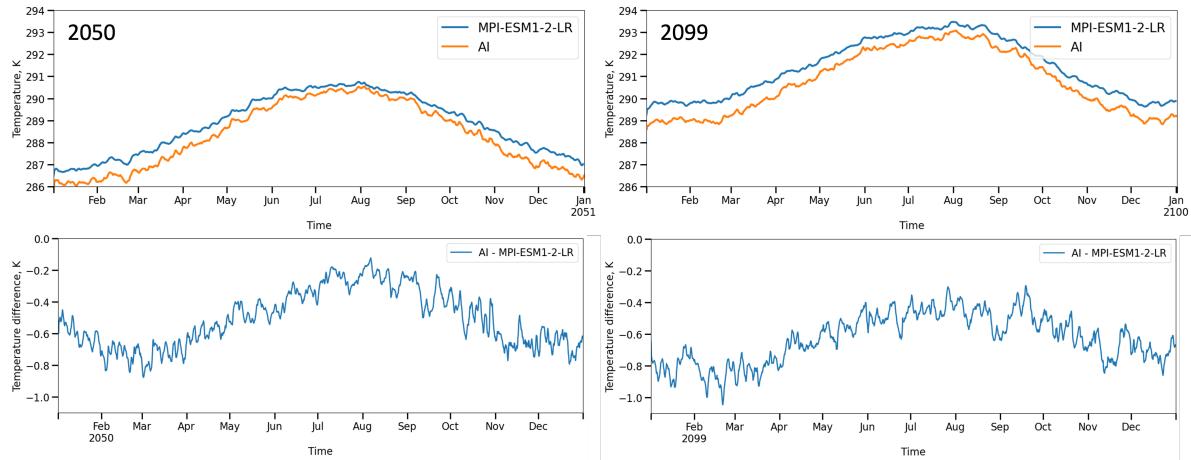


Figure A7: Top: Global mean 2-metre temperature time series (6-hour values, 1-day rolling mean) for MPI-ESM1-2-LR scenario simulations (years 2050 and 2099, SSP585) and AI-NWP generated daily forecasts initialized from those simulations. The AI-NWP time series is derived from lead hours 24, 30, 36, and 42 (2-day dataset) of AI-NWP forecasts initiated from CMIP6 model fields. Bottom: Difference between MPI-ESM1-2-LR scenario simulations and AI-NWP generated data.

## B Appendix B: Data locality use cases

We aim to demonstrate the behavior of downscaled climate variables by examining their time series around areas exhibiting potentially strong spatial gradients. This demonstration will underscore the advantages of creating high-resolution versions of low-resolution climate model datasets for adaptation and mitigation applications.

First, we analyze the wind speed data (Figure B1). Two points were selected, one on-land and the other offshore. Data were extracted from the grid points nearest to the selected locations for both a low-resolution climate model and the corresponding AI-NWP dataset. The selection ensured that, for the low-resolution model, both points fell within a single grid point, although the on-land grid point was situated several hundred kilometers inland.

The time series of wind speed plotted for the low-resolution climate model are naturally identical for both points, while they show significant differences in the AI-NWP data. As anticipated, high-resolution data indicate stronger winds offshore, with differences reaching several meters per second (Figure B1, middle row). Similarly, the on-land point shows lower wind speeds on average. High-resolution AI-NWP data demonstrate more frequent occurrences of stronger winds, reaching higher speeds (Figure B1, bottom row).

Even small differences in wind speed can lead to substantial differences in wind energy production, given that wind power increases approximately with the cube of the wind speed. Therefore, more accurate wind speed estimates are crucial for adaptation use cases related to wind energy.

Another regions with strong spatial gradients are mountainous areas. Similarly to the wind use case, we select points that in reality are located in the Alps and on flatter area next to them (Figure B2, top row). For low resolution data the nearest grid point is the same, while in AI-NWP generated data there is few degree difference between mountainous and plain areas (Figure B2, middle row). Histogram of temperature values shows that warmer temperatures occurs more often in AI-NWP generated data than anticipated by the low resolution climate model (Figure B2, bottom row).

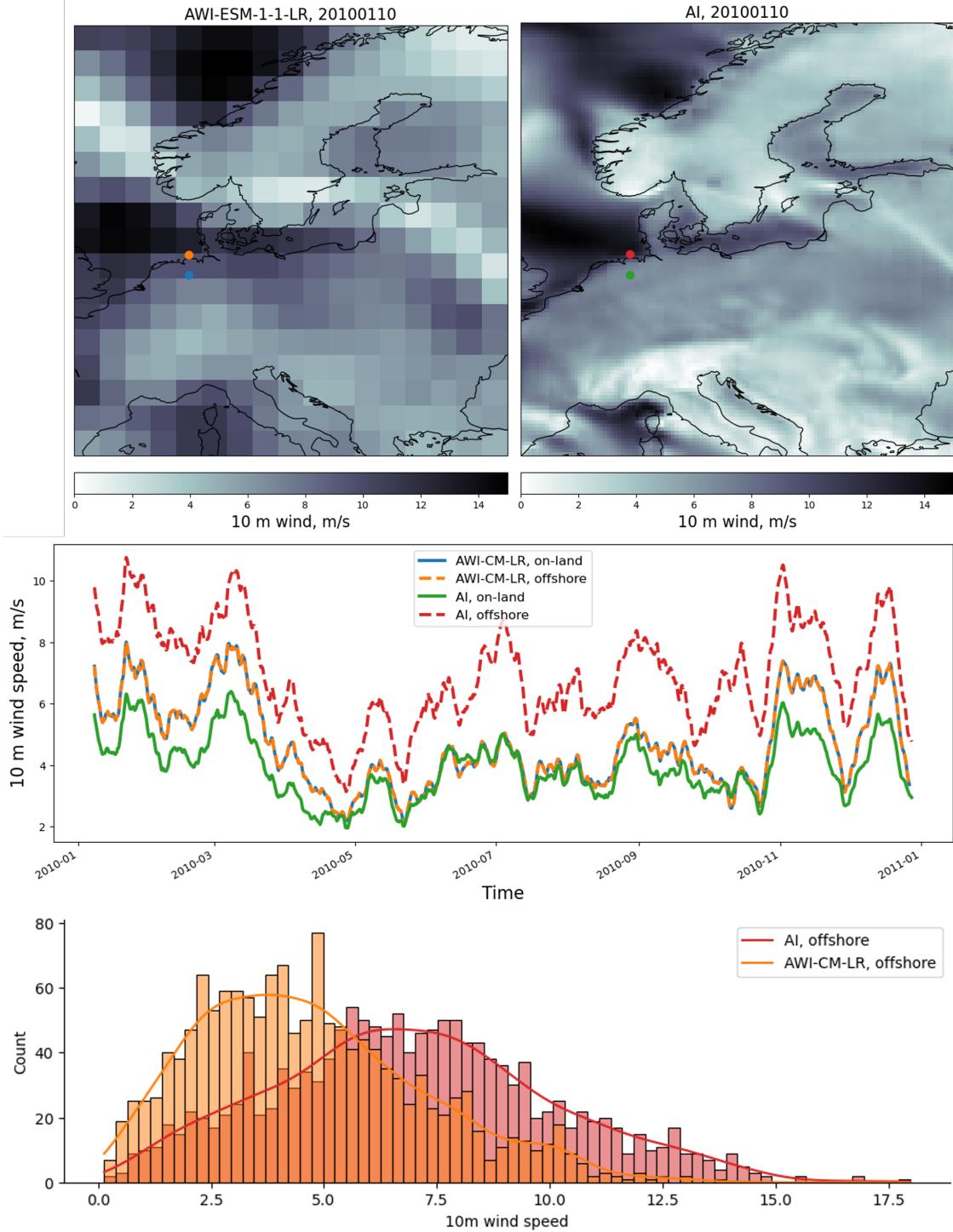


Figure B1: Top: Wind speed at 10m from AWI-ESM-LR (left) and AI-NWP generated data (right). The analyzed points are indicated. Data for the blue and orange points are extracted from the low-resolution AWI-ESM-LR data, falling into the same grid point, despite being on-land and offshore respectively. Data for the green and red points are extracted from AI-NWP generated data, falling into different grid points. Middle: Time series of 10m wind speed for the indicated points. The blue and dashed orange lines coincide exactly, as the points fall within the same grid box. A clear difference is observed between on-land and offshore points in the AI-generated data. Bottom: Histogram of 10m wind speeds for offshore points in AWI-ESM-LR and AI-NWP generated data. The downscaled information provides users with enhanced locality

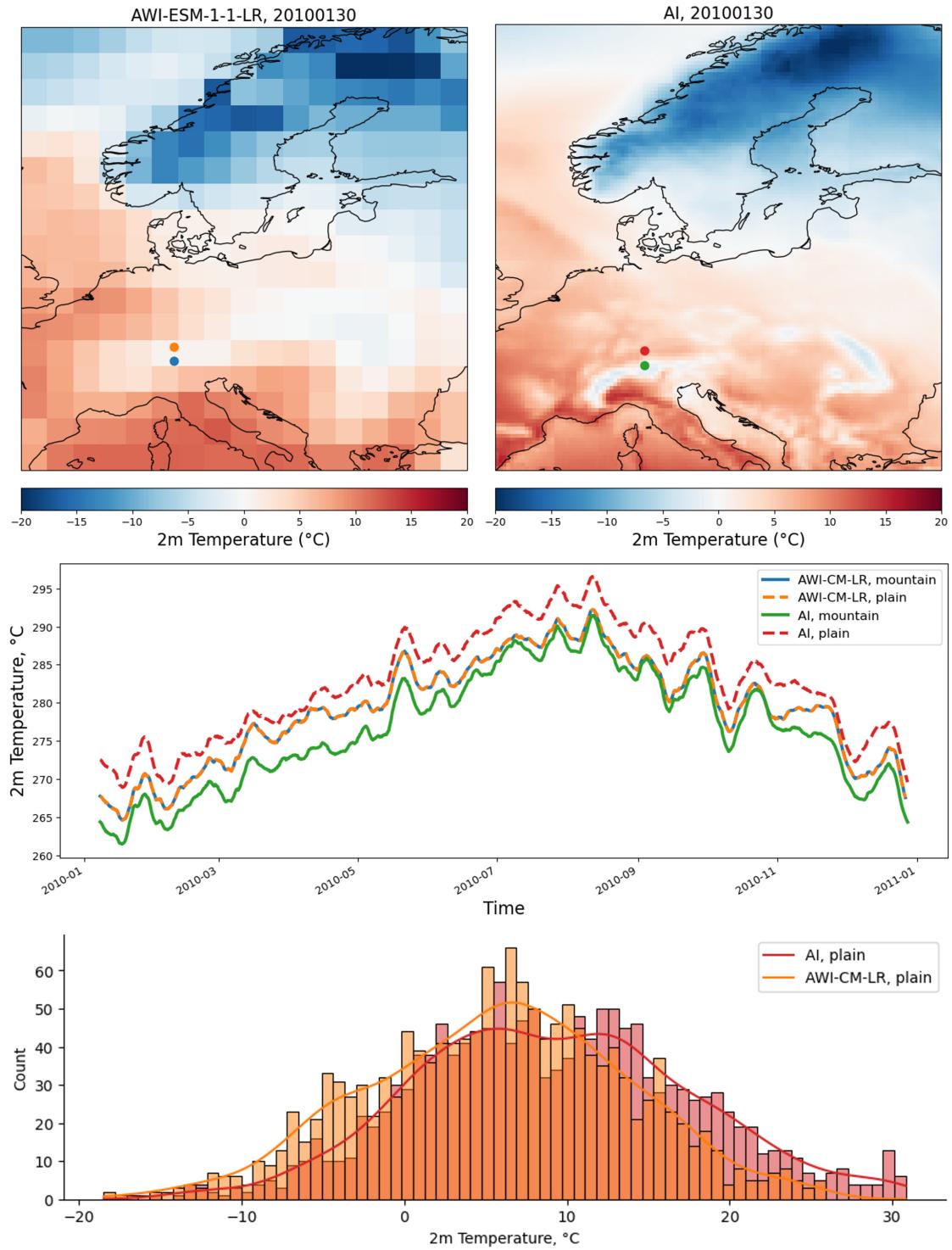


Figure B2: Same ad Fig B1, but for temperature, and points located on over the mountains, and on the plain.