Assessment of NA-CORDEX regional climate models, reanalysis, and in-situ gridded-observational against U.S. Climate Reference Network datasets

Souleymane SY[1], Fabio Madona[2], Federico Serva[3], Ismaila Diallo[4], and Benjamin Quesada[5]

[1]University of Augsburg
[2]Consiglio Nazionale delle Ricerche - Istituto di Metodologie per l'Analisi Ambientale
[3]European Space Agency
[4]University of California Los Angeles
[5]Universidad del Rosario

December 7, 2022

Abstract

Climate models still need to be improved in their capability of reproducing the present climate at both global and regional scale. The assessment of their performance depends on the datasets used as comparators. Reanalysis and gridded (homogenized or not homogenized) observational datasets have been frequently used for this purpose. However, none of these can be considered a reference dataset. Here, for the first time, using in-situ measurements from NOAA U.S. Climate Reference Network (USCRN), a network of 139 stations with high-quality instruments deployed across the continental U.S, daily temperature, and precipitation from a suite of dynamically downscaled regional climate models (RCMs; driven by ERA-Interim) involved in NA-CORDEX are assessed. The assessment is extended also to the most recent and modern widely used reanalysis (ERA5, ERA-Interim, MERRA2, NARR) and gridded observational datasets (Daymet, PRISM, Livneh, CPC). Results show that biases for the different datasets are mainly seasonal and subregional dependent. On average, reanalysis and in-situ-based datasets are generally warmer than USCRN year-round, while models are colder (warmer) in winter (summer). In-situ-based datasets provide the best performance in most of the CONUS regions compared to reanalysis and models, but still have biases in regions such as the Midwest mountains and the Northwestern Pacific. Results also highlight that reanalysis does not outperform RCMs in most of the U.S. subregions. Likewise, for both reanalysis and models, temperature and precipitation biases are also significantly depending on the orography, with larger temperature biases for coarser model resolutions and precipitation biases for reanalysis.

Assessment of NA-CORDEX regional climate models, reanalysis, and in-situ gridded-observational against U.S. Climate Reference Network datasets

Souleymane SY[1#], Fabio Madonna[1], Federico Serva[2+], Ismaila Diallo[3] and Benjamin Quesada[4]

[1]Consiglio Nazionale delle Ricerche – Istituto di Metodologie per l'Analisi Ambientale, C. da S. Loja - Zona Industriale, I-85050 Tito Scalo (Potenza), Italy

[#]Now at Institute of Geography, University of Augsburg, 86159 Augsburg, Germany

[2]Consiglio Nazionale delle Ricerche-Istituto di Scienze Marine (CNR-ISMAR), Rome, Italy

[+]Now at Italian Space Agency, Rome, Italy, and European Space Agency, Frascati, Italy

[3]Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA 16802, United States

[4]Universidad del Rosario, Faculty of Natural Sciences and Earth System Science Program, 'Interactions Climate-Ecosystems (ICE)' Research Group, Kr 26No 63B-48 (Bogotá D.C.), Colombia

Corresponding author: Dr. Souleymane SY (souleymane.sy@geo.uni-augsburg.de)

Key Points:

- Reanalysis and in-situ-based datasets are generally warmer than the USCRN year-round, while models are colder (warmer) in winter (summer).
- Reanalysis does not perform better than RCMs in most of the CONUS subregions.
- Temperature and precipitation biases are significantly depending on the orography.

Abstract:

Climate models still need to be improved in their capability of reproducing the present climate at both global and regional scale. The assessment of their performance depends on the datasets used as comparators. Reanalysis and gridded (homogenized or not homogenized) observational datasets have been frequently used for this purpose. However, none of these can be considered a reference dataset. Here, for the first time, using in-situ measurements from NOAA U.S. Climate Reference Network (USCRN), a network of 139 stations with high-quality instruments deployed across the continental U.S, daily temperature, and precipitation from a suite of dynamically downscaled regional climate models (RCMs; driven by ERA-Interim) involved in NA-CORDEX are assessed. The assessment is extended also to the most recent and modern widely used reanalysis (ERA5, ERA-Interim, MERRA2, NARR) and gridded observational datasets (Daymet, PRISM, Livneh, CPC). Results show that biases for the different datasets are mainly seasonal and subregional dependent. On average, reanalysis and in-situ-based datasets are generally warmer than USCRN year-round, while models are colder (warmer) in winter (summer). In-situ-based datasets provide the best performance in most of the CONUS regions compared to reanalysis and models, but still have biases in regions such as the Midwest mountains and the Northwestern Pacific. Results also highlight that reanalysis does not outperform RCMs in most of the U.S. subregions. Likewise, for both reanalysis and models, temperature and precipitation biases are also significantly depending on the orography, with larger temperature biases for coarser model resolutions and precipitation biases for reanalysis.

Plain Language Summary:

Climate models are developed for understanding past, present, and future climate variability and change. The assessment of their performance through their capacity to reproduce present climate conditions has been typically carried out using reanalysis and/or in-situ gridded observational datasets affected by several uncertainties as reference comparators. This paper, for the first time, evaluates the performance of NA-CORDEX regional climate models, reanalysis (ERA5, ERA-Interim, MERRA2, NARR) and gridded observational datasets (Daymet, PRISM, Livneh, CPC) in the continental U.S using temperature and precipitation observations from NOAA U.S. Climate Reference Network (USCRN). Overall, results indicate gridded observational datasets provide the best performance in most of the CONUS regions compared to reanalysis and models, but still have biases in the Midwest mountains and the Northwestern Pacific that need to be considered before their use as a reference for model evaluation. Conversely, reanalysis datasets do not outperform models in most of subregions. Likewise, for both reanalysis and models, results highlight that orography misrepresentation significantly contributes to the simulated temperature and precipitation biases. Finally, this study highlights the importance of evaluating performance of regional climate models using reference measurements.

Keywords: reference measurements, temperature, precipitation, climate models performance, high-resolution, NA-CORDEX, USCRN

1- Introduction:

Near surface temperature and precipitation are the most frequently used variables in climate studies. Their trends in observations and climate models have been a subject of extensive study over the past decades due to their impact on health, food security, ecosystems, and water supply. Both are highly variable in space and time and the related time series are often prone to inhomogeneities due to changes in the instrumentation, calculation algorithms, station re-locations and other factors, which must be adjusted to enable the identification of climate signals (Essa et al., 2022; Hausfather et al., 2016; Madonna et al., 2022). Moreover, observations over certain regions are sparse and implies large sampling uncertainties (Sy et al., 2021). Nevertheless, near surface observations are still one of the main data sources for climate modeling evaluation.

Climate models are developed for understanding past, present, and future climate variability and change. However, one the key challenge in their performance assessment is the spatial mismatch between observations and climate models (e.g., Zhang et al., 2011), the latter with resolutions ranging from ~10 to 50 km for the Regional Climate Models (RCMs) and from ~50 to 300 km for the Global Climate Models (GCMs) (Eyring et al., 2016; Taylor et al., 2012). Consequences of scale mismatch are more evident for models and reanalysis with coarse resolution. Therefore, different interpolation methods have been proposed in the literature to aggregate the near surface observations (Comber & Zeng, 2019; Herrera et al., 2019; Militino et al., 2015). Nevertheless, the use of these methods requires a dense spatial coverage of observations. An alternative solution for evaluating climate model, widely used in literature, is to use reanalysis or gridded surface observational datasets (e.g., Gibson et al., 2019; Srivastava et al., 2020, 2022). An increasing number of studies evaluated regional climate models in the presence of observational uncertainties by considering different gridded observational datasets as references to rank their reliability (e.g., Gibson et al., 2019; Herold et al., 2016; Kotlarski et al., 2019). Consequently, different conclusions are often drawn altering model performance in nontrivial ways due to the contribution of observational uncertainties (Gibson et al., 2019; Gómez-Navarro et al., 2012; Kotlarski et al., 2019; Prein & Gobiet, 2017; Srivastava et al., 2022). A debate related to the uncertainties related to the usage of reanalysis and gridded surface observational datasets in the model evaluation is ongoing since several years, and it is a topic of several papers in the literature (e.g., Diaconescu et al., 2018; Dieng et al., 2022; Dosio et al., 2022; Flaounas et al., 2012; Gibson et al., 2019; Srivastava et al., 2020, 2022). The uncertainties

are influenced by several factors, such as: i) the use of interpolation methods/technics on sparse station networks (e.g. Contractor et al., 2015; Herrera et al., 2019; Hofstra et al., 2010), which may largely affect the representation of precipitation data (Avila et al., 2015; Contractor et al., 2015); ii) uncertainties related to the representation of complex terrain at high elevation, for example, when observations are sparse and do not fully capture the spatial elevation-precipitation dependence; and iii) the way climate variables are adjusted on the orography (Sandu et al., 2019).

Under these premises, the quality of observational records is key when assessing climate models. Reference dataset and homogenization of historical time series is required because non-climatic discontinuities can alter the interpretation of decadal climate variability and change. In the last two decades, many measurements program around the world, such as the GCOS (Global Climate Observing System), designed "reference networks" to monitor climate with the objective to fill an important gap in the global observing system (https://gcos.wmo.int/en/home). Reference networks can provide long-term, high-quality climate data records, traceable to SI standards, and quantified uncertainties (Thorne et al., 2017, 2018).

The US Climate Reference Network (USCRN) is one of the brightest examples of reference network measuring near-surface air temperature and precipitation and measuring at the same time several quantify of influence (Diamond et al., 2013). However, to assess homogenization effects of the U.S. Historical Climatology Network (USHCN) datasets, Hausfather et al., (2016) compared the nearby pairs of USHCN and USCRN stations and found that adjustments make both trends and monthly anomalies from USHCN stations much more similar to those of neighboring USCRN stations over the period from 2004 to 2015, especially when the network stations overlap. They also concluded that USCRN datasets were useful for empirically testing of USHCN station adjustments. Bell et al., (2013) provide technical description of USCRN soil moisture observations in the context of U.S. soil-climate–measurement efforts and discuss the advantage of the triple-redundancy approach applied by USCRN. Further, to better understand the nature of soil moisture measurement variability in USCRN, Palecki & Bell, (2013) pointed out that deploying triplicate configurations of soil probes were useful to detect faulty sensors.

In this work, the ability of regional climate models (RCMs) participating to the North American Coordinated Regional climate Downscaling Experiment (NA-CORDEX, Mearns et al., 2017) to simulate the daily mean temperature and precipitation over the Contiguous United States (CONUS) is evaluated, along with performances of atmospheric reanalysis and gridded surface observational datasets. To address these issues, differently from previous studies available in the literature (e.g., Gibson et al., 2019; Srivastava et al., 2020, 2022, among others), observations from the USCRN network are used as reference for comparison. This paper focuses on the three main scientific questions:

i) How well do NA-CORDEX climate models, gridded-surface observational and reanalysis datasets represent CONUS daily temperature and precipitation against local reference in-situ observations?

ii) Are recent gridded-observational datasets and reanalysis products credible for climate model evaluation?

iii) How the improved resolution in NA-CORDEX regional climate models bring value to in-situ reference observations?

The remainder of the paper is structured as follows. Section 2 describes the USCRN data, RCMs, gridded surface observational and atmospheric reanalysis data used, while the statistical metrics and the subregional evaluations are presented in Section 3. Section 4 assesses the performance of individual and multi-model ensemble mean in simulating the CONUS local climate characteristics, along with a subregional assessment of model biases and reanalysis as well as of gridded surface observational datasets uncertainties. Finally, a discussion of the most relevant results together with the main conclusions and recommendations are provided in Section 5.

2- Meteorological Data

a. U.S. Climate Reference Network (USCRN)

U.S. Climate Reference Network (USCRN, (Diamond et al., 2013) is a systematic and sustained network of 139 stations deployed across the conterminous U.S. (CONUS), Alaska, and Hawaii. Stations are managed and maintained by the National Oceanic and Atmospheric Administration's (NOAA) National Centers for Environmental Information (NCEI). The primary goal of USCRN is to provide long-term homogeneous observations of temperature, precipitation, and soil moisture/soil temperature that can be used for current climate

applications while also being coupled to past long-term observations for the detection and attribution of climate change (Diamond et al., 2013). USCRN stations use high-quality instruments to measure temperature, precipitation, wind speed, soil conditions, and other ancillary variables (https://www.ncdc.noaa.gov/crn/). For both temperature and precipitation, the concept of triple measurements redundancy (i.e., three collocated thermometers) is adopted in the data processing to improve the quality of the temperature estimation. The quality of individual observations and the continuity of the records at each site is monitored on a routine basis. Specific information regarding USCRN instrumentation can be also found at www.ncdc.noaa.gov/crn/instrdoc.html.

This is a major reason why USCRN is also considered at the international level a reference network as also assessed in the frame of the EU H2020 Research Project GAIA-CLIM (Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring) using a maturity matrix approach (Thorne et al., 2017). Among USCRN stations, only those with at least 12 valid years (i.e., without missing data) over the CONUS in the 2006-2020 period (black circles) is used in this paper (Figure 1a). USCRN stations cover quite uniformly U.S and measure at different altitude and under different climate regimes (Figure 1a-b). They are placed in as far away from trees and other vegetation as possible, with rural environments expected to be free of human activities and land-use/land cover change effects. Observations conform to standards that meet or exceed those established by the World Meteorological Organization (WMO), as well as U.S. requirements for the variables being observed (WMO, 2008)

In Figure 2, the USCRN climatological daily mean temperatures and precipitations (2006-2020) for annual, winter and summer seasons, are shown. For temperature (Figure 2, left column), annual patterns reveal strong spatial gradients over the CONUS with maximum values exceeding 25.45°C observed over the south, desert, and southwest pacific areas, while the smallest values of 2.04°C are observed over the northern parts of the Midwest and central areas. Similar spatial gradients are also observed in winter and summer, while the coldest temperature value of −13.36°C is recorded in the north part of the Midwest and Central areas, mainly during winter cold air incursions from the Arctic. By contrast, the highest temperature values exceeding 37.90°C are observed over the desert in summertime. Precipitation (Figure 2, right column) patterns are highly dependent on locations and seasons with highest values over the southeast and northwestern Pacific. The high precipitation patterns seen over southeastern CONUS mainly occurs in summer and may likely be generated by tropical systems (Mitchell

et al., 2019) or other mesoscale atmospheric circulations such as the convective activities of the North American Monsoon (Barlow et al., 2019). In contrast, the high precipitation over the northwestern CONUS (with values exceeding 20 mm/day in Quinault and 10 mm/day in Darrington stations) occurs in wintertime and can be controlled by large synoptic-scale atmospheric fronts (e.g., Castro et al., 2012; Yu et al., 2022).

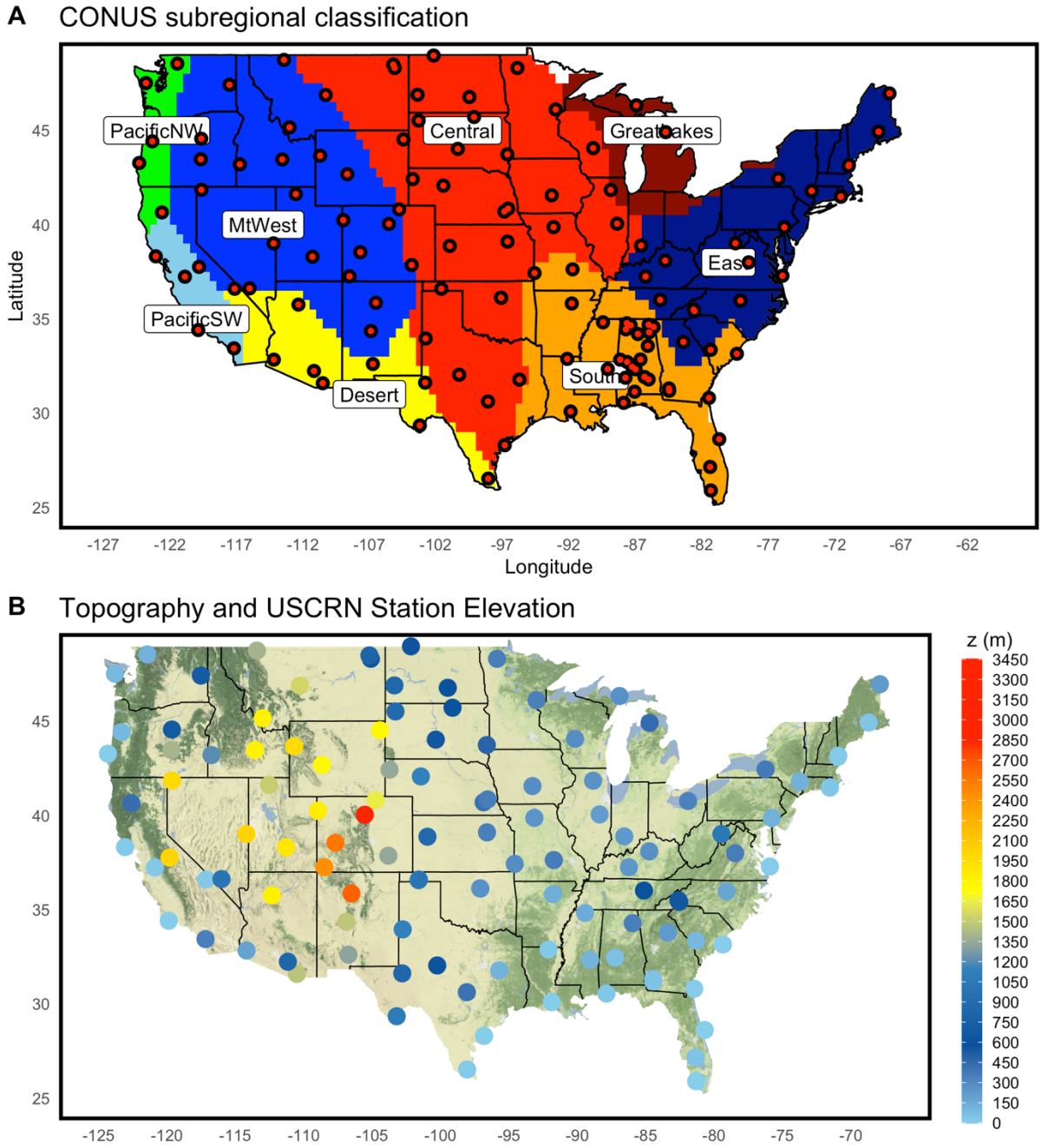

Figure 1: U.S. Climate Reference Network (USCRN) stations distribution along with the different station locations representing stations with at least 12 valid years (i.e. without missing data) over the 2006–2020 period (black circles). A) displays the eight large CONUS climatic subregional subdomains adapted from (Bukovsky, 2011): Desert, PacificSW, PacificNW,

MidWest, Central, East, South and Great Lakes (Top panel). B) displays the CONUS topographic maps along with the different station elevation values (unit: m) indicated by the different colors.

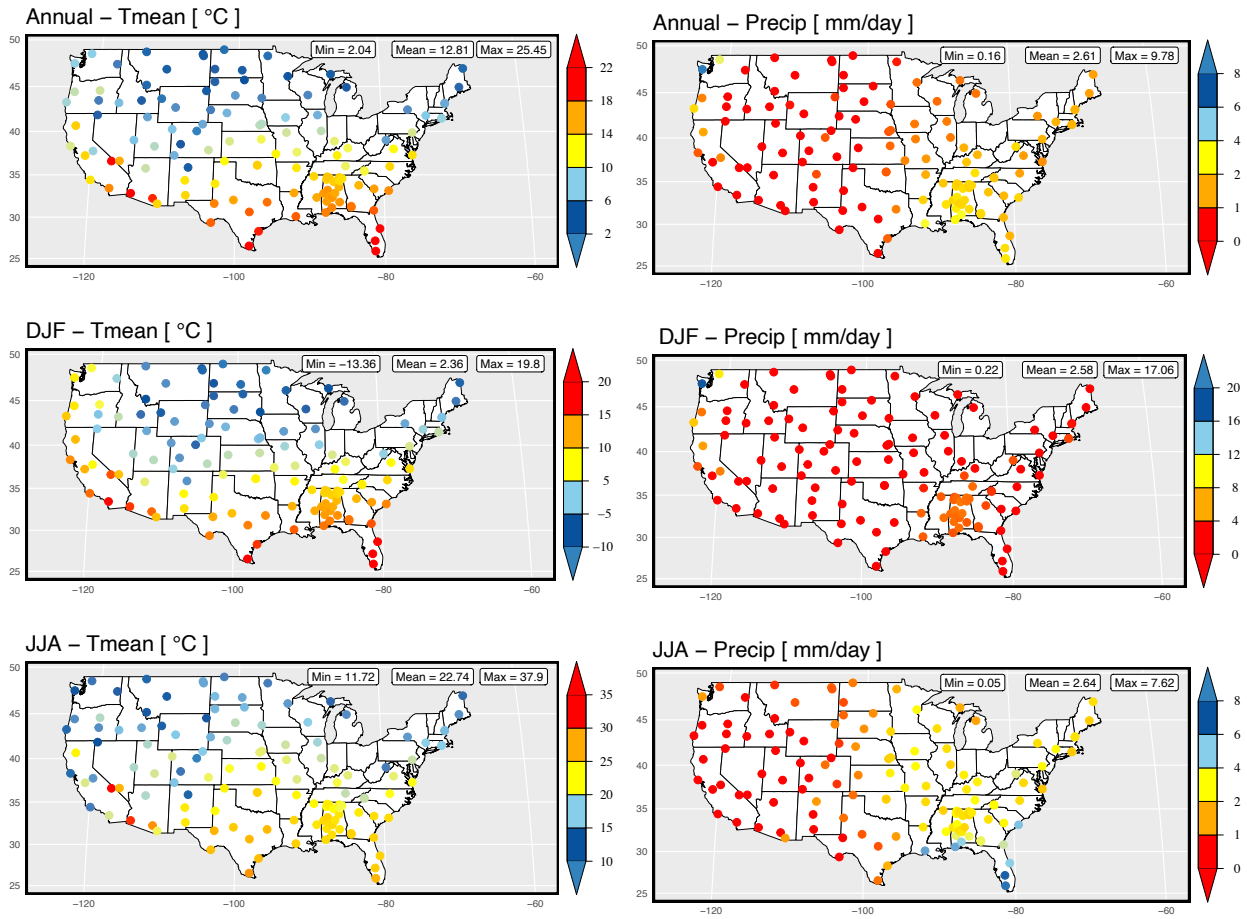

Figure 2: Climatological daily mean temperature (left panels) and precipitation (right panels) over the 2006–2020 period at USCRN stations.

b. Atmospheric reanalysis datasets

The list of datasets assessed in this study includes four widely used reanalysis and four gridded observational products. The atmospheric reanalysis products used in this paper include the global ERA-Interim (hereafter, ERAI) (Dee et al., 2011), ERA5 (Hersbach et al., 2020), MERRA-2 (Gelaro et al., 2017) and the North American Regional Reanalysis (NARR; Mesinger et al., 2006). ERA5 is the latest climate reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), providing hourly data on a regular latitude–longitude grid at 0.25° x 0.25° resolution (Hersbach et al., 2020). It replaces the ERA-Interim reanalysis (used to force NA-CORDEX simulations, see next section) and is based on the

Integrated Forecasting System (IFS) Cy41r2 which was operational in 2016. It thus benefits from a decade of developments in model physics, core dynamics and data assimilation compared to ERAI. In addition to a significantly enhanced horizontal resolution of 31 km, compared to 80 km for ERAI, it has hourly output throughout, and an uncertainty estimate from an ensemble. The general set-up of ERA5, as well as a basic evaluation of characteristics, are provided by (Hersbach et al., 2020) and data is publicly available through the Copernicus Climate Data Store (CDS, https://cds.climate.copernicus.eu).

Beyond the two ECMWF products, one of the most recent global atmospheric reanalysis product is MERRA-2 (Gelaro et al., 2017). MERRA-2 is the latest atmospheric global reanalysis of the modern satellite era produced by NASA's Global Modeling and Assimilation Office (GMAO). It also assimilates observation types not available to the earlier generation MERRA-1 reanalysis (Rienecker et al., 2011) and includes updates to the Goddard Earth Observing System (GEOS) model and analysis scheme to provide a viable ongoing climate analysis beyond MERRA-1. Overall, MERRA-2 system has most of the same basic features as the MERRA-1 system but includes several important updates (Gelaro et al., 2017). The regional reanalysis product used in this study is the high-resolution North American Regional Reanalysis (NARR; Mesinger et al., 2006). NARR consists of a long-term, consistent, high-resolution climate dataset for the North American domain, as a major improvement upon the earlier global reanalysis datasets in both resolution and accuracy. It notably differs from the other reanalysis products described above because it does assimilate rain gauge networks into its latent heating scheme (Bukovsky and Karoly, 2007). The direct assimilation from observed precipitation datasets makes it more of a 'hybrid'' product compared to other products (Mesinger et al., 2006).

c. In-situ gridded observational datasets

Along with the reanalysis products used, four recent gridded-observational datasets used in this paper include the Climate Prediction Center (CPC) Unified CONUS dataset (hereafter, CPC) (Higgins et al., 2000), Livneh (Livneh et al., 2013, 2015), the Oregon State University Parameter-Elevation Regressions on Independent Slopes Model (PRISM) (Daly et al., 2008), and the Daily Surface Weather Data on a 1-km Grid for North America, Version 4 (hereafter Daymet-V4; Thornton et al., 2021). CPC (Higgins et al., 2000) is from the National Oceanic and Atmospheric Administration (NOAA). It is provided by the NOAA/OAR/ESRL PSD,

Boulder, Colorado, USA and is obtained from the website https://psl.noaa.gov/data/gridded/data.unified.daily.conus.html.

CPC uses station data from the U.S. unified rain gauge dataset, composed of multiple sources (Higgins et al., 2000). Based on the inverse-distance weighting interpolation algorithms of (Cressman, 1959), CPC was developed with the aim to create regional analyses over the CONUS-Mexico and the South America domains (Higgins et al., 2000). The primary goal consists of developing a US Precipitation Quality Control (QC) system and analysis that improves the QC of rain gauge data used in precipitation analyses for the United States improving precipitation products and applications in support of climate monitoring, climate prediction, and applied research. Livneh (Livneh et al., 2013, 2015) is a station-based 1⁄16°(~6 km) resolution gridded data. It was developed based on the meshing procedure of (Maurer et al., 2002). With the effort to create regional analyses over Mexico, the United States, and southern Canada, it accounts for orographic effects using the elevation-scaling procedure for precipitation climatology from 1961 to 1990. PRISM dataset (Daly et al., 2008) primarily uses station data from Cooperative Observer Program (COOP) stations and snowpack telemetry (SNOTEL), as well as several others smaller networks. The daily product also incorporates radar observations a 4 km-resolution from the Advanced Hydro-Weather Prediction System (AHPS) over central and eastern CONUS. To correct the precipitation-elevation dependence, a linear method, using weights at each grid-point based on elevation and location characteristics, is used. Finally, Daymet-4 (Thornton et al., 2021) daily surface weather data on a 1-km grid for North America is based on COOP and SNOTEL station networks like PRISM. The precipitation-elevation dependence is also corrected using a weighted local linear regression. The comparison between gridded-observational datasets also allows to discuss uncertainties due to the potential impact of interpolating data from sparse surface observations. Dataset's main specification and related references for further details are summarized in Table 1.

Table 1. Characteristics of gridded data sets and reanalysis products used in this study.

Reanalysis/In-situ-based	**Considered period**	**Original horizontal resolution**	**References**	**Type**
ERA−Interim	2006−2018	0.75°x 0.75°	(Dee et al., 2011)	Reanalysis

ERA5	2006–2020	0.25°x0.25°	(Hersbach et al., 2020)	Reanalysis
MERRA–2	2006–2020	0.5°x 0.62°	(Gelaro et al., 2017)	Reanalysis
NARR	2006–2020	0.3°x0.3°	(Mesinger et al., 2006)	Reanalysis
CPC	2006–2020	0.5°x0.5°	(Higgins et al., 2000)	In-situ-based
Daymet-V4	2006–2020	0.01°x0.01°	(Thornton et al., 2021)	In-situ-based
PRISM	2006–2020	0.04°x0.04°	(Daly et al., 2008)	In-situ-based
Livneh	2006–2013	0.06° x 0.06	(Livneh et al., 2013, 2015)	In-situ-based

d. NA-CORDEX model ensemble

The study utilized the NA-CORDEX model ensemble (Mearns et al., 2017) composed of seven Regional Climate Models (RCMs): CRCM5–OUR (Martynov et al., 2013; Šeparović et al., 2013),CRCM5–UQAM (Martynov et al., 2013; Šeparović et al., 2013), RCA4 (Samuelsson et al., 2011), RegCM4 (Giorgi et al., 2012), WRF , CanRCM4 (Scinocca et al., 2016), and HIRHAM5 (Christensen et al., 2007). The NA-CORDEX datasets are obtained/retrieved from https://na–cordex.org/index.html data archive. The NA-CORDEX experiment aims to add value to the existing body of regional climate models by using multiple simulations with high spatial-resolutions to facilitate regional climate model intercomparison studies and ultimately serve the impact and adaptation communities (Giorgi et al., 2009). Note, for further details the main information about the model datasets, horizontal resolutions and related references are summarized in Table 2. As the main purpose of this study is the model evaluation, historical experiment data only, not subject to any form of bias correction (labelled as ‘‘Eval’’ driven by ERAI reanalysis), are assessed using USCRN in-situ reference measurement. The RCMs are run at either 0.44° (~50 km) or 0.22° (~25 km), with a single higher-resolution run of the

CRCM5−OUR model at 0.11° (~12.5km) to allow direct assessment of potential added-value from increased resolution (https://na−cordex.org/rcm−characteristics.html).

Table 2: NA-CORDEX climate models employed in the study, the horizontal resolutions, and references.

NA-CORDEX Models	Spatial resolutions	References
CRCM5−OUR	0.44°/50km 0.22°/25km	(Martynov et al., 2013; Šeparović et al., 2013)
CRCM5−UQAM	0.44°/50km 0.22°/25km 0.11°/12.5km	(Martynov et al., 2013; Šeparović et al., 2013)
CanRCM4	0.44°/50km 0.22°/25km	(Scinocca et al., 2016)
HIRHAM5	0.44°/50km	(Christensen et al., 2007)
RCA4	0.44°/50km	(Samuelsson et al., 2011)
RegCM4	0.44°/50km 0.22°/25km	(Giorgi et al., 2012)
WRF	0.44°/50km 0.22°/25km	(Skamarock and Klemp, 2008)

3- Methodology

a. Subregional assessment

The comparison between models, reanalysis and in-situ-based datasets refers to daily mean temperatures and precipitations and is carried out by selecting the nearest grid-points matching USCRN stations. In fact, for a fixed value of the horizontal resolution, the representativeness uncertainty associated with the use of the nearest grid-point is assumed to be smaller or comparable to that affecting other interpolation methods (Madonna et al., 2022). The datasets

are aggregated annually and seasonally for summer (June, July, and August; JJA) and winter (December, January, and February; DJF). Although the NA-CORDEX simulations covers the period from 1979 to 2014/2015 (model dependent), part of this study covers the recent period 2006−2014 relative to the time frame of the NA-CORDEX simulations and observation datasets. Biases for reanalysis and in-situ-based datasets are also assessed over the extended period 2006−2020. For a consistent inter-comparison, reanalysis and in-situ-based products were interpolated/remapped to a common resolution grid of 0.44° CONUS land−only grid using a first-order conservative remapping procedure (Jones, 1999).

Part of the analysis presented in this study is focused on different CONUS subregions, according to the climate classification of (Bukovsky, 2011) (Figure 1a). Among the 29 subregions created in the North American Regional Climate Change Assessment Program (NARCCAP) domain (Mearns et al., 2017), grouping regions (Figure 1a) have been selected using the climate classification by (Ricketts et al., 1999), dividing the CONUS into eight climatic subregions: Desert, PacificSW, PacificNW, Midwest, Central, East, South, and Great Lakes. This division also accounts for the different orography, since while the Eastern US has a relatively flat topography and low elevation, the Western CONUS, especially the Midwest, features a highly fractured relief rising gradually from sea level to 3000 m, with magnitude gradually decreasing towards the West coast (see Figure 1b).

b. Evaluation metrics

The performance of the RCMs, reanalysis, in-situ-based datasets was evaluated against USCRN in each selected subregion. Different metrics were chosen to evaluate different aspects of the datasets. The metrics were computed for both temperature and precipitation and for each dataset, covering all seasons and subregions. The different datasets were first evaluated in their ability to reproduce the observed temperature and precipitation climatology over the entire CONUS and for each subregion. Then, biases in the seasonal cycle of the rainfall and temperature distributions have been quantified. As a third step, the spatial-temporal variability throughout the subregions was assessed using Taylor diagrams (Taylor, 2001). Please, note that the Taylor diagram summarizes the main scores skills: correlation coefficient, standard deviation, and root mean square deviation and has been already employed in the ranking of models and reanalysis products in many studies over the CONUS regions (e.g., Gibson et al., 2019; Srivastava et al., 2022). Furthermore, to estimate correlations, a modified Taylor diagram

using robust non-parametric Kendall rank correlation test τ (Croux & Dehon, 2010) is used. Kendall rank correlation test has demonstrated to be less sensitive to errors and discrepancies in data compared to Pearson or Spearman tests (e.g., Diouf et al., 2022). Finally, the orography-dependent temperature/precipitation biases are also examined. To estimate the slopes, Theil's Sen regression method (Sen, 1968) is used. Theil's Sen slope estimator is a resistant and non-parametric regression method based on the median of pairwise slopes and can be significantly more accurate than simple linear regression for skewed and heteroskedastic data (Sy et al., 2021).

4- Results

4.1 Climatological biases

In Figure 3, the spatial patterns of the temperature daily mean bias estimated from the ensemble -mean of models (hereinafter, NAM) at 0.44°, 0.22°, and 0.11° resolutions (for the latter one model only), reanalysis (hereinafter, Rean) and gridded datasets (hereinafter, in-situ-based) subsampled at USCRN station locations are shown. On average, Rean and in-situ-based datasets are generally warmer in both seasons with biases of 0.41°C in winter (and 0.94 °C in summer) for Rean and biases exceeding 0.31°C in summer for the in-situ-based datasets. In contrast, NAMs are generally colder (warmer) in wintertime (summertime) with biases ranging from -0.51 (0.94) °C at NAM-44 to -0.27 (1.44) °C at NAM-11 (Figure 3). Overall, the spatial variability is well reproduced in both seasons with correlation values up to 0.90 for Rean, 0.92 for in-situ-based datasets and often exceeding 0.88 for NAMs. Regarding the skills (RMSE values), in-situ-based datasets provides the best performance compared to Rean and NAMs, as summarized in Table 3. Further, across the subregions, all datasets display cold (warm) biases in the western part (eastern part) in summer and wintertime with values ranging ±8.00 °C in some stations over the Midwest mountains, PacificNW and PacificSW, while warm biases up to 8.00 °C are found in summer for stations in central U.S. Conversely, over the east and south, a closer agreement with USCRN is found in wintertime. Regarding the sensitivity to the increased resolution, overall, the effects are quite small and do not show a clear improvement of biases.

In Figure 4, biases for precipitation are shown. On average, Rean are drier over the entire CONUS, with mean bias exceeding -0.69 mm/day in winter (and -0.37 mm/day in summer). Across subregions, the larger biases are simulated in winter with values exceeding -4.0 mm/day

at the PacificNW and PacificSW regions (especially in Quinault and Darrington stations). For the in-situ-based datasets, biases are generally negligible over the entire CONUS in both seasons, although biases exceeding 2.0 mm/day can be observed in the Darrington and Bodega stations over the PacificNW and PacificSW regions. For NAMs, biases are in general larger in both seasons with a mean bias estimated around 0.01mm/day (0.23mm/day) for NAM-44, 0.20mm/day (0.47 mm/day) for NAM-22 and 0.6mm/day (0.34 mm/day) for NAM-11 in winter (summer). Considering both seasons, Rean and in-situ-based datasets show quite similar rainfall biases over the entire CONUS, while for NAMs (except for NAM-11) opposite biases are simulated over the southeast part during both seasons, i.e., dry in winter and wet in summer. In other words, NAMs generally overestimate the heaviest precipitation patterns over the southeastern occurring mainly in summer while underestimating them in winter. This poor performance can be related to the misrepresentation of the tropical cyclones and/or of the mesoscale circulation near the surface such as the local convective activities of the monsoon (Hsu et al., 2019). Considering correlations, values up to r= 0.83 for Rean, r= 0.86 for in-situ-based datasets and r= 0.83 for NAMs are obtained. Regarding the skills, overall, the in-situ-based datasets display the best performance in both seasons (Table 3).

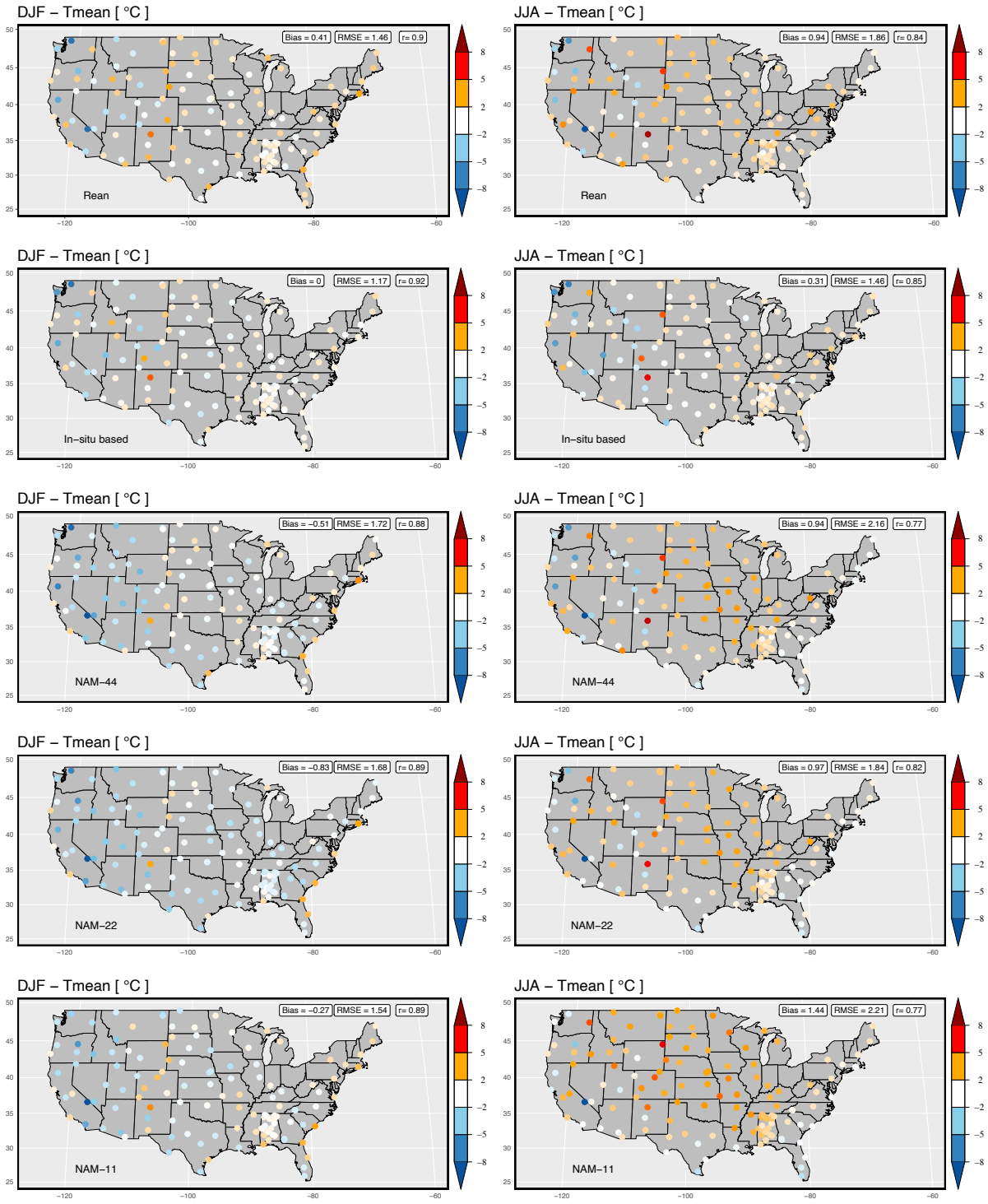

Figure 3: Spatial distribution of the daily mean temperature bias (°C), using USCRN as the reference, estimated for ensemble means of reanalyses products (Rean), models at 0.44°, 0.22°, and 0.11° resolutions (NAM-44, NAM-22, and NAM-11 respectively) and gridded observational datasets (In-situ-based) for both winter (DJF, left panel) and summer (JJA, right panel). The skills scores (the spatial mean bias, Kendall rank-correlations and RMSE values) are provided at top left of each panel.

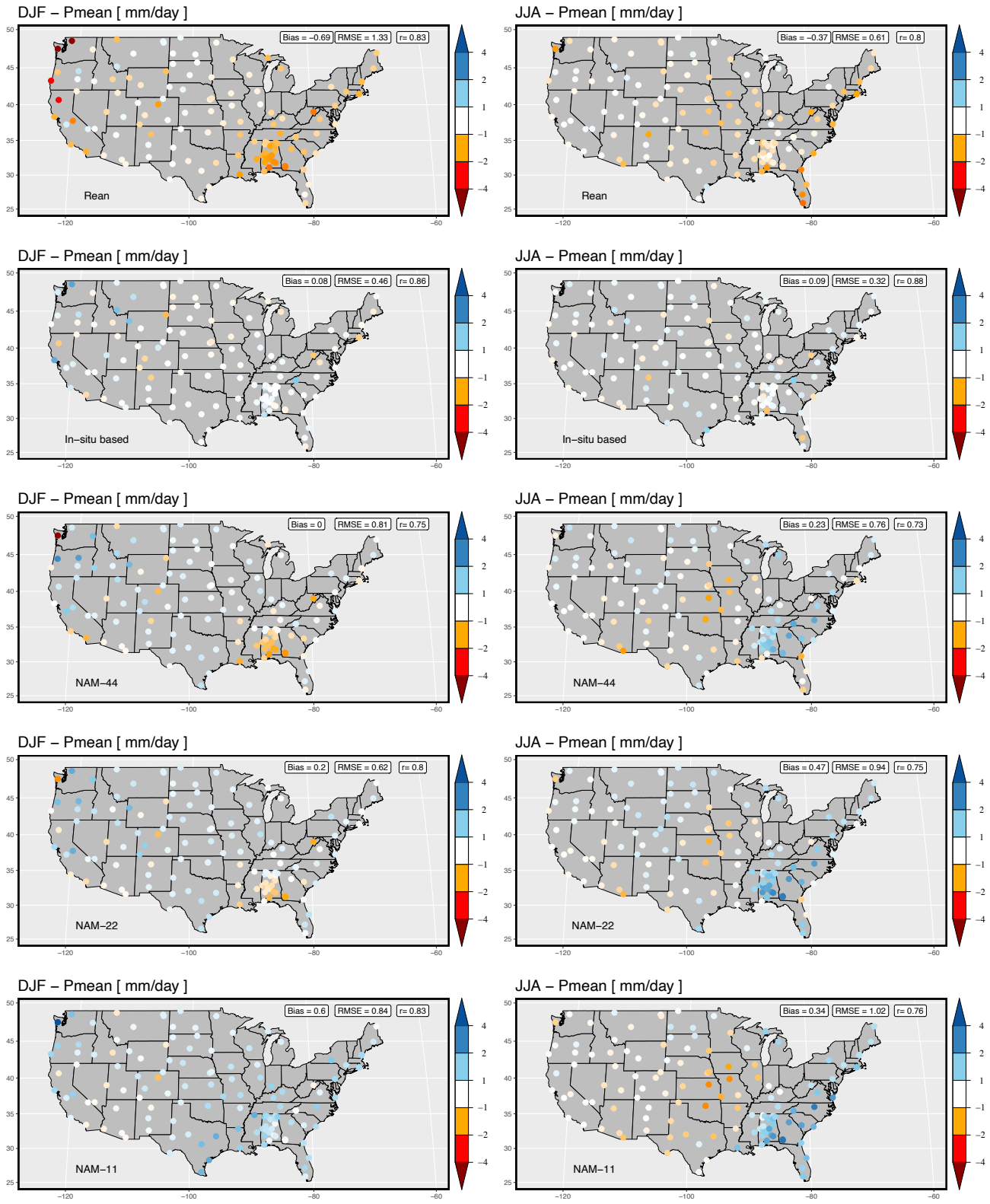

Figure 4: Same as in Figure 3 but for the spatial patterns of the daily mean precipitation biases (mm/day)

Table 3: An overview of the daily temperature and precipitation spatial mean bias, the Kendall rank-correlation and the root mean square errors provided in Figures 3 and 4.

	Temperature (°C)			Precipitation (mm/day)		
	DJF					
	Bias	RMSE	Corr	Bias	RMSE	Corr
Rean	0.41	1.46	0.90	−0.69	1.33	0.83
In-situ-based	0.00	1.17	0.92	0.08	0.46	0.86
NAM-44	−0.51	1.72	0.88	0.00	0.81	0.75
NAM-22	−0.83	1.68	0.89	0.20	0.62	0.80
NAM-11	−0.27	1.54	0.89	0.60	0.84	0.83
	JJA					
Rean	0.94	1.86	0.84	−0.37	0.61	0.80
In-situ-based	0.31	1.46	0.85	0.09	0.32	0.88
NAM-44	0.94	2.16	0.77	0.23	0.76	0.73
NAM-22	0.97	1.84	0.82	0.47	0.94	0.75
NAM-11	1.44	2.21	0.77	0.34	1.02	0.76

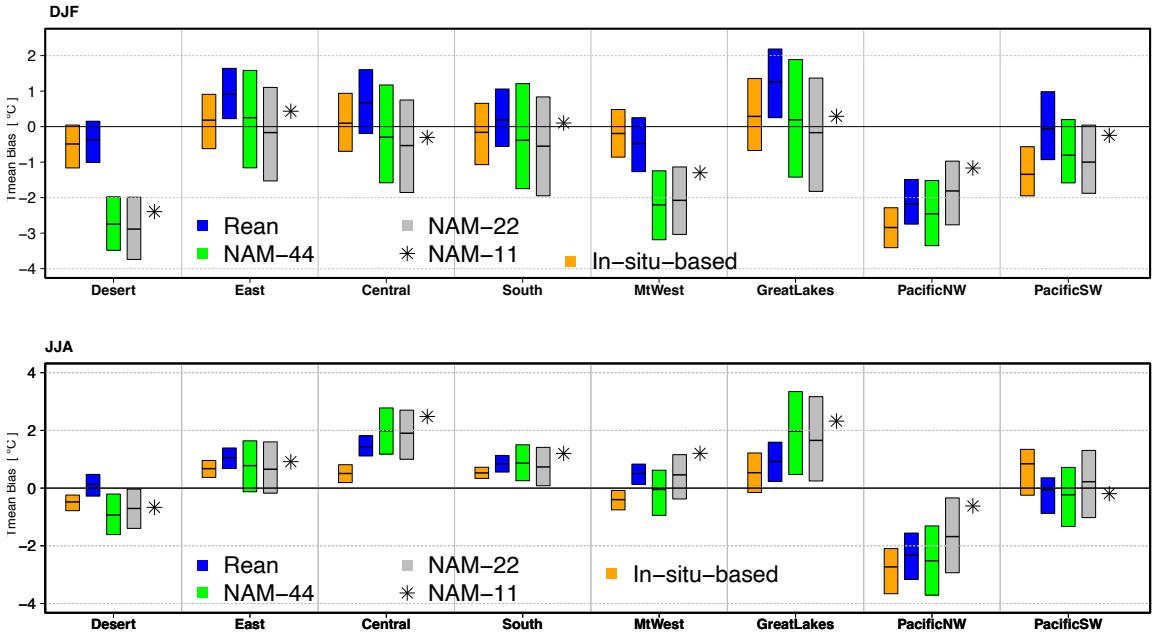

Figure 5: Daily mean temperature bias (°C) estimated from the reanalysis (Rean), models at 0.44°, 0.22°, and 0.11° resolutions (NAM-44, NAM-22 and NAM-11 respectively), and gridded observational datasets (in-situ-based), in each study subregion against USCRN for the period 2006-2014 and for both winter (DJF, top panel) and summer (JJA, bottom panel). The median value is indicated with a black line while the lower hinge of each box is Q1 quartile (25th), and the upper hinge for Q3 quartile (75th). The subregional daily mean distribution of

each dataset is obtained by spatially averaging all values within the subregion highlighted in Figure 1. The evaluation results for the individual RCMs at all resolutions, reanalysis and gridded-observational datasets are shown in Figure S1.

A comparison of seasonal biases for the daily mean temperature within each subregion is shown in Figure 5. NAMs at all resolutions show significant cold biases of -3.0 °C in winter over the deserts, Midwest mountains and the PacificNW regions and may be attributed to the misrepresentation of surface albedo in wintertime (Bonan, 1998; Li et al., 2016). Instead, Rean and in-situ-based datasets show large cold biases (-2.5°C for Rean and -3.0°C for in-situ-based datasets) over the Pacific Northwest. Nevertheless, over the Midwest mountains, where the topography is relatively complex, the Rean and in-situ-based datasets show a better agreement with USCRN compared to NAMs. In summer, warm biases are generally simulated by all datasets in most of the CONUS subregions, except in the Pacific Northwest and in the Desert, where the large cold biases of -3.0 °C are still persistent for all datasets.

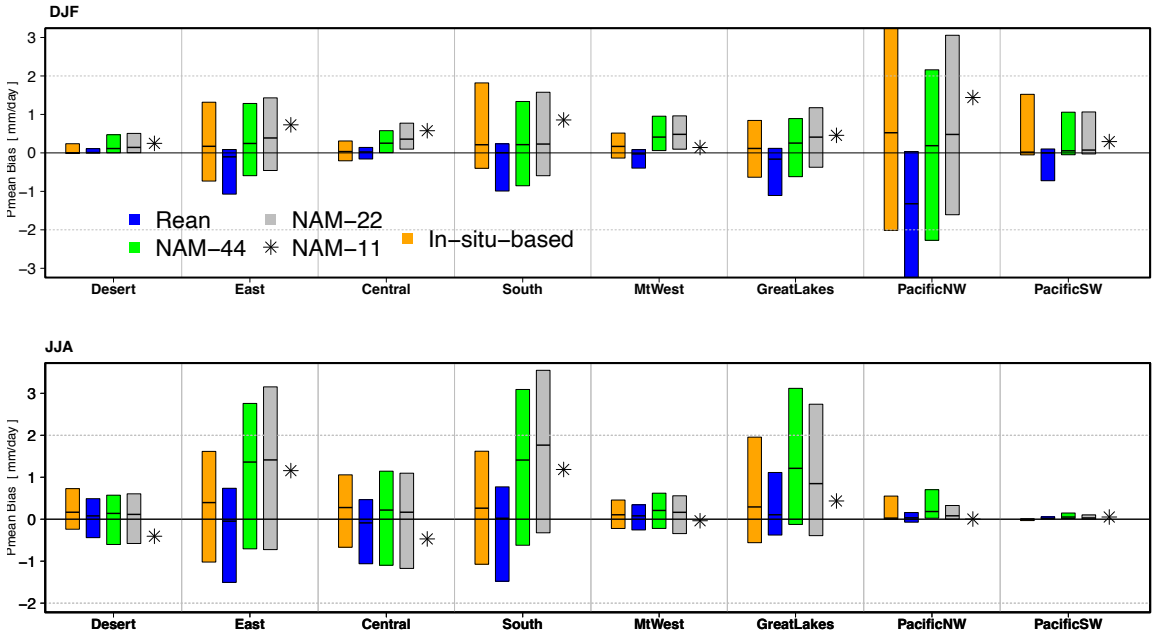

Figure 6: Same as for Figure 5 but for the daily mean precipitation bias (mm/day). The evaluation results for the individual RCMs, reanalysis and gridded-observational datasets are shown in Figure S2.

450 In analogy to temperature, the comparison of seasonal daily mean biases for precipitation is
451 shown in Figure 6. Rean bias in winter is limited to -1.0 mm/day expect over the PacificNW,
452 where the bias is larger. Instead, in-situ-based datasets and NAMs show positive biases, while
453 NAMs remain particularly dry over the Greatlakes, southern, eastern parts in summer.
454 Regarding the resolutions, for both temperature and precipitation, no substantial improvement
455 in the simulated biases is obtained if the model resolutions increase.
456

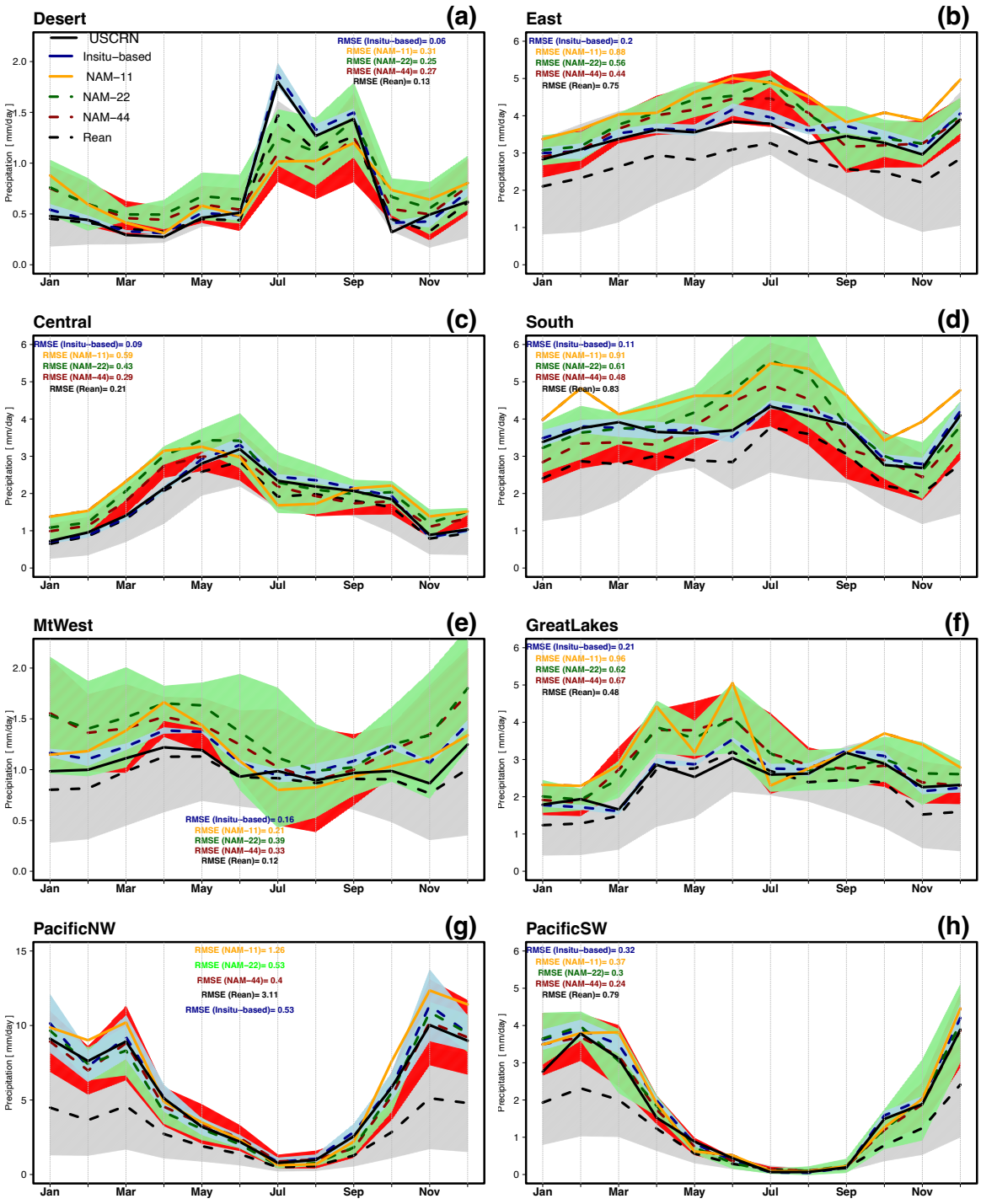

Figure 7: Seasonal monthly mean rainfall distribution (mm/day) over the different subregions obtained from the ensemble-mean of reanalysis (Rean, black dashed line), models at 0.44° (NAM-44, dark red dashed line), at 0.22° (NAM-22, dark-green dashed line) and at 0.11° (NAM-11, solid orange line) resolutions and the in-situ-based datasets (dark-blue dashed line). The values are obtained by spatially averaging all values within each subregion over the period 2006-2014. Red (light green) shaded area shows the range within ±1 sigma of NAM-44 (NAM-22) grid spacing models, while gray (light-blue) shaded area shows the reanalysis (in-situ-based datasets) ensemble uncertainty (Hersbach et al., 2020). Results for the individual RCMs (NAM-

44, NAM-22 and NAM-11), reanalysis and gridded-observational datasets are shown in Figure S3.

Figure 7 explores the annual cycle of precipitation over all subregions in USCRN stations and other datasets during the period 2006-2014. The annual cycle is computed from monthly average of station-based daily precipitation over each subregion. The annual cycle of observed precipitation (solid black curve) reflects the wet summer season observed over most subregions (East, Central, GreatLakes, and South). On the other hand, the wet winter season over the PacificNW and PacificSW regions, driven by fronts coming from the northwest, is also well characterized by USCRN. As is clear from Figure 7, the in-situ-based datasets capture well both the phase and amplitude of the observed annual cycle compared to Rean and NAMs. This is also clear from the RSME values estimated over the different subregions, except over the Pacific NW and the Midwest mountains where in-situ-based datasets are generally drier throughout the entire year (Figure 7e and 7g). This effect is likely due to the orographic bias. Rean (black dashed line) can capture the different phase of the annual cycle but underestimate the magnitude in most subregions with a large ensemble uncertainty, especially in wintertime. By contrast, as already observed, NAMs (e.g., NAM-22 and NAM-44) are generally wetter than USCRN, although they can properly reproduce the seasonal cycle. NAM-11 (solid yellow line) is not able to properly reproduce the phase and amplitude of the observed annual cycle (with large RSME values) in most subregions and this may be due to usage of one model only (NAM-11). Regarding temperature, both the phase and amplitude of the observed annual cycle are well captured by the datasets (Figure S4 and S5).

In summary, the observed annual cycle for in-situ-based datasets is close to USCRN than both the Rean and NAMs, while Rean datasets do not perform better than NAMs in most of subregions. In fact, the large variability in the simulated annual are particularly evident in the MidWest mountain region (Figure 7e) and can be due to the difference in topographic elevation in Rean, NAMs and USCRN stations (see section 4.3 above), the latter typically located in the valleys. Consequently, Rean and NAMs can differ from USCRN where stations are closer to orographic barriers, e.g., on the side of the arrival of the air masses (upwind, more rain) or behind reliefs (rain shadowed, less rain).

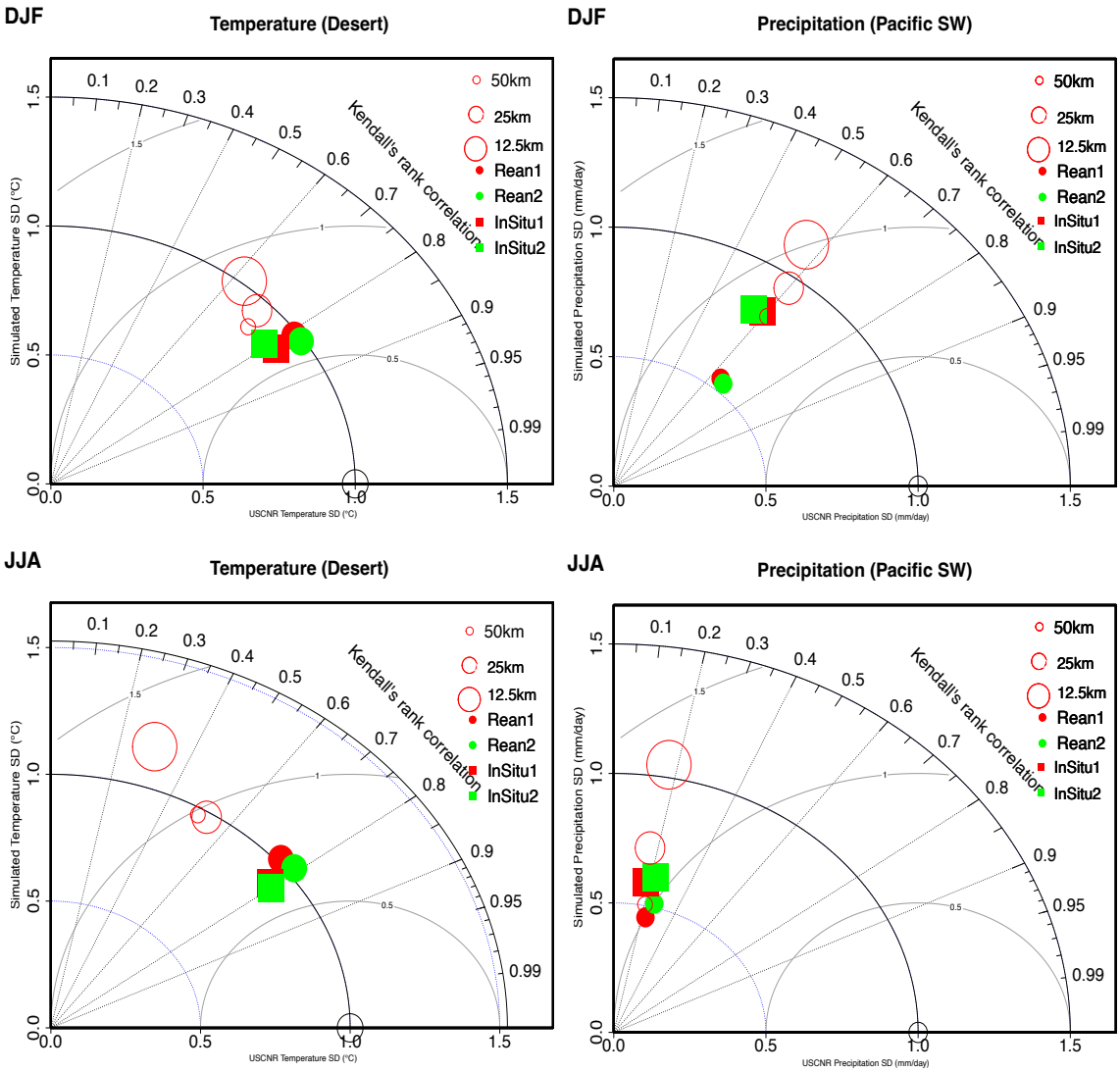

Figure 8: DJF (top panels) and JJA (bottom panels) Taylor diagram for temperature (left panels, over the Desert) and precipitation (right panels, over the PacificSW) of the models, reanalysis, and in-situ-based datasets versus USCRN. The reference point (USCRN) is marked as a solid black circle. Symbols indicate the position of each dataset: red circles for models, red dots and squares for reanalysis and in-situ-based datasets over 2006-2014 (Rean1/Insitu1). Green dots and squares symbols are also for reanalysis and in-situ-based datasets (Rean2/Insitu2), respectively, but for the extended period 2006-2020. The dashed black lines on the outermost semicircle indicate Kendall rank correlations between USCRN and each dataset. The blue dashed curves indicate the normalized standard deviations. The grey dashed curves show the centered normalized root mean squared error (NRMSE).

To rank the reliability of Rean, in-situ-based datasets, and NAMs in reproducing patterns of temperature and precipitation against USCRN, Taylor diagrams are also used. Figure 8 shows

512 evaluation results for one subregion per variable (for temperature, Desert, left panels and for
513 precipitation, PacificSW, right panels). Results for other subregions are shown in Figure S6 and
514 Figure S7. Regarding temperature across the various subregions, Rean and the in-situ-based
515 datasets provide quite similar skills, with the best skill scores obtained in the Central U.S (with
516 correlation values up to r=0.9), while the worst skill scores are obtained in the coastal regions
517 (i.e., PacificNW and PacificSW regions). NAMs have the best performance in the PacificSW
518 in wintertime with correlation values up to 0.7, while the poorest performance (with a
519 correlation of r= 0.50 for NAM-44 and NAM-22 and r=0.30 for NAM-11) over the Desert
520 (Figure 8, left panels). Nevertheless, the performance is relatively better in wintertime
521 compared to summertime (Figure S6 vs. S7). Regarding precipitations across the various
522 subregions, the score skills are low in all datasets (Figure 8 right panels for the PacificSW
523 region, and Figure S8 and Figure S9 for the other subregions). The worst skill scores (with
524 correlation values of r = 0.20 associated with large variability) are given over the PacificSW in
525 summertime, with a slight improvement especially for the in-situ-based datasets in wintertime.
526 On the other hand, even though the datasets better agree with USCRN values of daily mean
527 precipitation (Figure 6), biases remain persistent in reproducing the local climate variability.
528 Regarding the sensitivity of the simulated skill scores for both Rean and In-situ-based datasets
529 in the extended period, no relevant skill difference is found for both temperature and
530 precipitation (Figure 8).
531
532 **4.2 Relationship to orography**
533 Figure 9 shows the spatial patterns of the elevation mean bias estimated from Rean and NAMs.
534 It can be pointed out that models and reanalysis still suffer from uncertainties in the
535 representation of a complex orography. From USCRN stations, both Rean and NAMs show
536 quite similar orography biases over the entire CONUS with values varying ±600 m at stations
537 in the western part of US, especially in the MidWest, where topography is relatively complex.
538 Moreover, comparing the simulated biases between Rean and NAMs, it can be also noticed that
539 Rean performs no better than NAMs in the representation of orography. For models, the biases
540 remain even for high-resolution models (NAM-44 vs. NAM-22/NAM-11), suggesting that the
541 resolutions are still too coarse to capture important fine-scale features of the complex
542 orography. For all datasets, the largest biases (with values exceeding 1000 m) are found at the
543 Darrington station. As consequence, temperatures in both reanalysis and models are usually
544 colder than temperatures recorded at USCRN stations (see Figure 3) likely because of the

545 difference between simulated mean grid-point and station elevations, with the latter lower than
546 the former.
547
548 Temperature and precipitation biases for reanalysis and models are significantly linked to
549 orography biases (Figure 10a-d). A sensitivity of temperature biases to orography higher is
550 found at coarser resolutions (50km, -7.9 °C/km, r=-0.67, $p<0.05$; Figure 10b) than finer
551 resolutions (12.5km, -5.1°C/km, r=-0.61, $p<0.05$; Figure 10d). Across reanalysis and models at
552 all resolutions, the sensitivity of precipitation biases to orography is between 0.2 and 1
553 mm/day/km (Figure 10a-d), though reanalysis is more affected by precipitation biases than
554 models (slopes with red dots: 1 mm/day/km for reanalysis vs. 0.2-0.4 mm/day/km for models;
555 Figure 10a vs. 10b-d). The elevation-dependent biases follow approximately the value of the
556 adiabatic lapse rate of -6.5°C/km, which is consistent with hydrostatic equilibrium and
557 thermodynamic principles (Dutra et al., 2020).
558
559 Precipitation-dependent temperature biases are also investigated in Figure S10. The negative
560 relationship between temperature and precipitation biases is critical for understanding the
561 impacts of changing climate on snowpack, drought, and heat stress. In summer, drier models
562 are in general hotter (Figure 5) which can cause more heat waves due to less moisture available
563 (summer dry soils heat up faster than wet soils; (Miralles et al., 2019)), affecting the land
564 surface energy balance, with implications for local and downwind precipitation (Schumacher
565 et al., 2022). By contrast, in winter, wet models are usually too cold because of the snow's
566 presence (Li et al., 2016). On the other hand, in winter, models have typical problems in
567 reproducing snow-related processes in regions of complex orography (Bordoy & Burlando,
568 2013), mainly because of the poor representation of the orography which can further increase
569 the cold waves and their duration (Dutra et al., 2012).
570

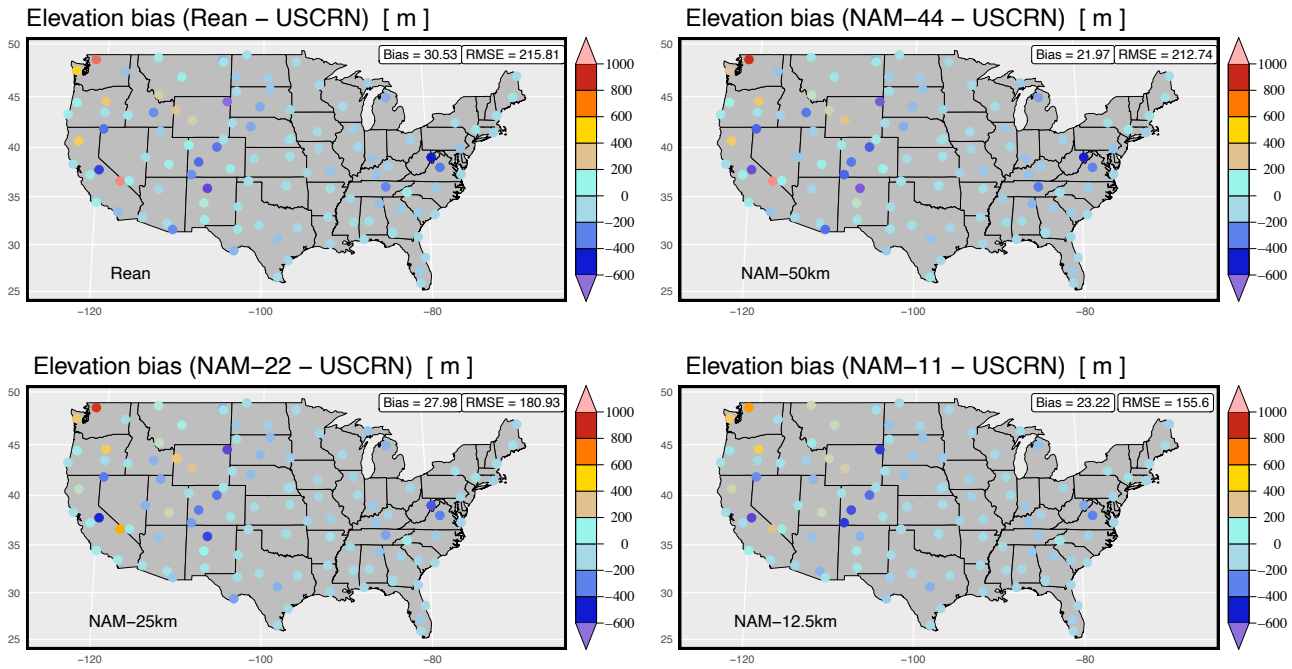

Figure 9: Spatial patterns of the elevation mean biases (m. above sea level) estimated from ensemble mean of reanalyses (Rean) and models at 0.44°, 0.22°, and 0.11° resolutions (NAM-44, NAM-22, and NAM-11) in comparison to USCRN values. The skills scores (the spatial mean bias and RMSE values) are provided at top left of each panel.

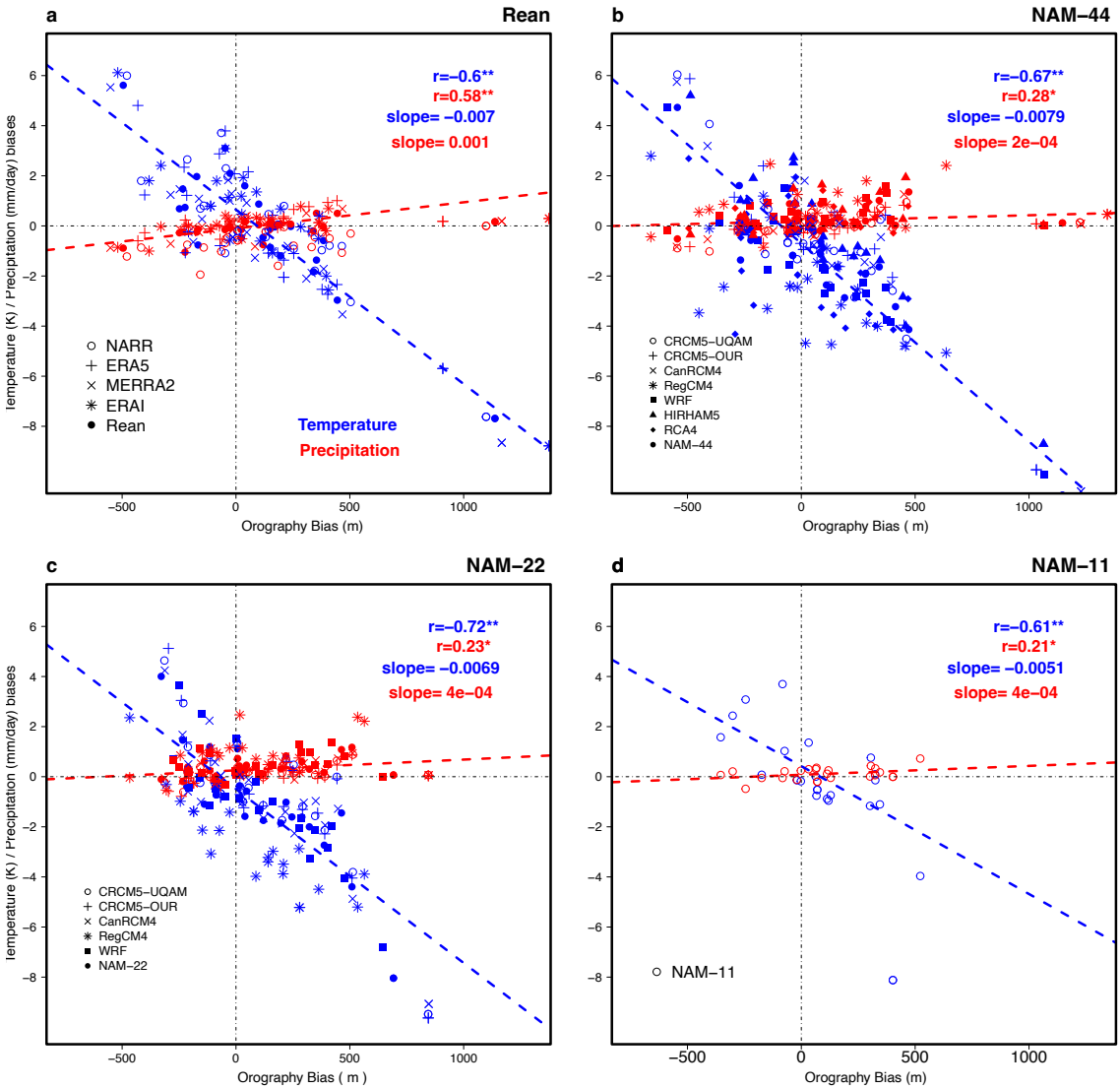

Figure 10: Contributions of orography to the simulated temperature (°C, in blue) and precipitation (mm/day, in red) mean biases. The relations between orography biases (x-axis) versus temperature and precipitation biases (y-axis) are drawn for Midwest region. Note that the Midwest region is considered due to its complex topography. Correlations are calculated using the Kendall non-linear rank (τ) test between the simulated elevation bias and temperature/precipitation mean biases for all Rean (a), NAM-44 (b), NAM-22 (c), and NAM-11 (d). Two-star symbols (**) are added when the correlation is significant at 99% confidence intervals (i.e., $p < 0.01$; 99% C.I), while one-star symbol (*) is added when the correlation is significant at 95% (i.e., $p < 0.05$; 99% C.I). The slopes indicated at the top right of each panel are estimated using median of pairwise non-parametric linear regression method (Sy et al., 2021).

5- Discussion and Conclusion

This paper, for the first time, assess the performance of NA-CORDEX regional climate models, reanalysis (ERA5, ERA-Interim, MERRA2, NARR) and gridded observational datasets (Daymet, PRISM, Livneh, CPC) in the CONUS using reference temperature and precipitation observations from USCRN. Assessment of climate model's performance through their capacity to reproduce present climate conditions has been typically carried out using reanalysis and/or gridded surface observational datasets as comparators (Gibson et al., 2019; Gómez-Navarro et al., 2012; Kotlarski et al., 2019; Srivastava et al., 2020, 2022). However, these datasets are affected by several uncertainties, mainly due to: i) interpolation methods implemented in gridded products based on sparse station networks; ii) the choice of interpolation scheme, and its application; ii) the representation of complex terrain at high elevation, as well as the elevation-precipitation/temperature spatial dependence at different altitude ranges; and iii) the approaches used to correct climate variables due to orographic misrepresentation (Napoli et al., 2019; Velasquez et al., 2019). Furthermore, it is worth pointing out that several existing papers (e.g., Gibson et al., 2019; Srivastava et al., 2020, 2022) used daily precipitation and/or temperature-based indices to rank model skills. However, the indices-based analyses do not fully capture the model performance (Alexander et al., 2020) and can hence compromise their assessment.

Being a reference comparator (Thorne et al, 2017), the use of USCRN measurements for the assessment of the datasets listed above allows to conclude that:

- The simulated mean biases among the datasets are primarily seasonal and subregional dependent.
- For temperature, reanalysis and in-situ-based datasets are generally warmer in both seasons, while models are generally colder in wintertime and warmer in summertime.
- Across the subregions, spatial patterns of the temperature mean biases are quite similar between datasets with cold (warm) biases in the western (eastern) U.S. with large values exceeding ±8.00°C in stations over Midwest mountains, Northwestern and Southwestern Pacific likely due to complex orography.
- Overall, results for precipitation show that in-situ-based datasets can capture the daily mean patterns and have the best skills in reproducing the phase and amplitude of the observed rainfall annual cycle compared to reanalysis and models.

- The worst skills for precipitation for all datasets are particularly obtained in the PacificNW, where the largest rainfall is typically recorded at stations and on Midwest mountains likely due to orographic enhancement (Napoli et al., 2019; Velasquez et al., 2019; Xie et al., 2007).
- Reanalysis are generally drier in both seasons and show large biases exceeding -4.00mm/day in the Northwestern Pacific. However, they are typically able to capture the phase of the precipitation patterns annual cycle although they underestimate the magnitude in most of the subregions with large uncertainty bounds (Figure 7-e and 7-g). Such findings are consistent with previous studies available in the literature (e.g., Alexander et al., 2020; Bador et al., 2020; Gibson et al., 2019; Srivastava et al., 2022).
- Models are generally wetter than USCRN over the CONUS but show opposite winter-summer patterns in the southeast U.S. – i.e., dry in winter and wet in summer. They typically capture the different phases of the seasonal cycle but overestimate the amplitudes in most regions.
- The poorest model performance is found over the desert, Midwest mountains and the Northwestern Pacific regions, with more pronounced biases in wintertime likely because of the surface albedo and the complex orography (Li et al., 2016).

The present study also reveals that models and reanalysis still suffer from uncertainties due to the inaccurate altitude representation in the most complex orographic areas (Figure 9) such as Midwest and Northwestern Pacific in the U.S. These uncertainties are mainly due to: i) the model grid-scale and the sub-grid scale orography representation; ii) the parameterization scheme and representation of surface processes (Diallo et al., 2019; Sy et al., 2017) iii) the orographic source datasets as well as the methodologies applied for deriving orography fields (Elvidge et al., 2019).

Our results also show that temperature and precipitation biases are significantly linked to orography biases, both in reanalysis and models, and temperature biases are higher at coarser resolutions (50 km, -7.9 °C/km, $r=-0.67$, $p<0.05$; Figure 10b) than at finer resolutions (12.5 km, -5.1°C/km, $r=-0.61$, $p<0.05$; Figure 10d). For precipitations, reanalysis are more affected than models. The positive elevation-dependent precipitation biases found for all the investigated datasets over the CONUS are a novel finding. In fact, very few publications in literature discusses the change of elevational precipitation gradients using near surface observations (Kuhn and Olefs, 2020). These biases show a more variable sensitivity than the corresponding

for temperature because it depends on latitude, seasons, the shape of the mountain and the relative station position respect to the arrival direction of the air masses (upwind, more rain) and the orography (rain shadowed, less rain). It also depends on several other mechanisms such as snow presence, vegetation cover, clouds, water vapor, aerosols, and soil moisture which vary with orographic characteristics (e.g., Pepin et al., 2015; Pepin et al., 2022; Rangwala and Miller, 2012).

Summarizing, this study highlights the importance of evaluating performance of regional climate models for historical data using reference measurements. Overall, our results indicate that in-situ-based datasets provide the best performance in most of the CONUS regions compared to reanalysis and models, but still have biases in Midwest mountains and the Northwestern Pacific that need to be considered before their use as a reference for model evaluation. Conversely, reanalysis datasets do not outperform models in most of subregions. Hence, we recommend caution when using them depending on their intended application, especially when assessing model performance in mountainous and coastal regions.

Finally, our study suggests that the mountainous regions will remain sensitive to the projected warming during the 21st century. To improve our understanding of how different climate drivers influence changes in high mountain regions, it requires a densification of climate reference network in the regions with most complex orography and the measurement of broader set of climate parameters related to energy fluxes. This will enhance the assessment of regional climate models and, consequently, climate mitigation and adaptation strategies.

Acknowledgments: The authors thank the European Union's Copernicus Climate Change Service (C3S) implemented by ECMWF for supporting this research. We also thank the modelers and the USCRN and NA-CORDEX contributors for their efforts and for access to their datasets. Souleymane SY's current research is supported by the WASCAL-CONCERT project at the University of Augsburg. Ismaila Diallo's research is supported by the center for Earth System Modeling, Analysis, and Data (ESMAD) at The Pennsylvania State University.

Contributions: Souleymane SY conceived the study, carried out the analyses and wrote the first draft. Fabio Madonna contributed significantly to the discussions and improved the manuscript. Federico Serva, Ismaila Diallo and Benjamin Quesada contributed to ideas and provided valuable feedbacks.

Competing interests: The authors declare that they have no conflict of interest.

Data availability statement: The USCRN data are available from https://www.ncei.noaa.gov/pub/data/uscrn/products/daily01/. The NA-CORDEX datasets are retrieved from https://na−cordex.org/index.html data archive. All the reanalysis and gridded-observational datasets used in this study are publicly available.

ORCIDs:

Souleymane Sy: https://orcid.org/0000-0001-8834-7620

Fabio Madonna: https://orcid.org/0000-0001-7628-8870

Federico Serva: https://orcid.org/0000-0002-7118-0817

Ismaila Diallo: https://orcid.org/0000-0003-1209-4107

Benjamin Quesada: https://orcid.org/0000-0002-8827-4801

References

Alexander, L. V., Bador, M., Roca, R., Contractor, S., Donat, M. G., & Nguyen, P. L. (2020). Intercomparison of annual precipitation indices and extremes over global land areas from in situ, space-based and reanalysis products. *Environmental Research Letters*, *15*(5), 055002. https://doi.org/10.1088/1748-9326/ab79e2

Avila, F. B., Dong, S., Menang, K. P., Rajczak, J., Renom, M., Donat, M. G., & Alexander, L. V. (2015). Systematic investigation of gridding-related scaling effects on annual statistics of daily temperature and precipitation maxima: A case study for south-east Australia. *Weather and Climate Extremes*, *9*, 6–16. https://doi.org/10.1016/j.wace.2015.06.003

Bador, M., Boé, J., Terray, L., Alexander, L. V., Baker, A., Bellucci, A., et al. (2020). Impact of Higher Spatial Atmospheric Resolution on Precipitation Extremes Over Land in Global Climate Models. *Journal of Geophysical Research: Atmospheres*, *125*(13), e2019JD032184. https://doi.org/10.1029/2019JD032184

Barlow, M., Gutowski, W. J., Gyakum, J. R., Katz, R. W., Lim, Y.-K., Schumacher, R. S., et al. (2019). North American extreme precipitation events and related large-scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics*, *53*(11), 6835–6875. https://doi.org/10.1007/s00382-019-04958-z

Bell, J. E., Palecki, M. A., Baker, C. B., Collins, W. G., Lawrimore, J. H., Leeper, R. D., et al. (2013). U.S. Climate Reference Network Soil Moisture and Temperature Observations. *Journal of Hydrometeorology*, *14*(3), 977–988. https://doi.org/10.1175/JHM-D-12-0146.1

Bonan, G. B. (1998). The Land Surface Climatology of the NCAR Land Surface Model Coupled to the NCAR Community Climate Model. *Journal of Climate*, *11*(6), 1307–1326. https://doi.org/10.1175/1520-0442(1998)011<1307:TLSCOT>2.0.CO;2

Bordoy, R., & Burlando, P. (2013). Bias Correction of Regional Climate Model Simulations in a Region of Complex Orography. *Journal of Applied Meteorology and Climatology*, *52*(1), 82–101. https://doi.org/10.1175/JAMC-D-11-0149.1

Bukovsky, M. (2011). Masks for the Bukovsky regionalization of North America, Regional Integrated Sciences Collective, Institute for Mathematics Applied to Geosciences, National Center for

Atmospheric Research, Boulder, CO. Retrieved from [http://www.narccap.ucar.edu/contrib/bukovsky/]

Bukovsky, M. S., & Karoly, D. J. (2007). A Brief Evaluation of Precipitation from the North American Regional Reanalysis. *Journal of Hydrometeorology*, *8*(4), 837–846. https://doi.org/10.1175/JHM595.1

Castro, C. L., Chang, H.-I., Dominguez, F., Carrillo, C., Schemm, J.-K., & Juang, H.-M. H. (2012). Can a Regional Climate Model Improve the Ability to Forecast the North American Monsoon? *Journal of Climate*, *25*(23), 8212–8237. https://doi.org/10.1175/JCLI-D-11-00441.1

Christensen, O., Drews, M., Christensen, J., Dethloff, K., Hebestadt, I., Ketelsen, K., & Rinke, A. (2007). The HIRHAM regional climate model version 5 (beta). *DMI Technical Report 06-17*.

Comber, A., & Zeng, W. (2019). Spatial interpolation using areal features: A review of methods and opportunities using new forms of data with coded illustrations. *Geography Compass*, *13*(10), e12465. https://doi.org/10.1111/gec3.12465

Contractor, S., Alexander, L. V., Donat, M. G., & Herold, N. (2015). How Well Do Gridded Datasets of Observed Daily Precipitation Compare over Australia? *Advances in Meteorology*, *2015*, e325718. https://doi.org/10.1155/2015/325718

Cressman, G. P. (1959). AN OPERATIONAL OBJECTIVE ANALYSIS SYSTEM. *Monthly Weather Review*, *87*(10), 367–374. https://doi.org/10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2

Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods & Applications*, *19*(4), 497–515. https://doi.org/10.1007/s10260-010-0142-z

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., et al. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Diaconescu, E. P., Mailhot, A., Brown, R., & Chaumont, D. (2018). Evaluation of CORDEX-Arctic daily precipitation and temperature-based climate indices over Canadian Arctic land areas. *Climate Dynamics*, *50*(5), 2061–2085. https://doi.org/10.1007/s00382-017-3736-4

Diallo, I., Xue, Y., Li, Q., De Sales, F., & Li, W. (2019). Dynamical downscaling the impact of spring Western US land surface temperature on the 2015 flood extremes at the Southern Great Plains: effect of domain choice, dynamic cores and land surface parameterization. *Climate Dynamics*, *53*(1), 1039–1061. https://doi.org/10.1007/s00382-019-04630-6

Diamond, H. J., Karl, T. R., Palecki, M. A., Baker, C. B., Bell, J. E., Leeper, R. D., et al. (2013). U.S. Climate Reference Network after One Decade of Operations: Status and Assessment. *Bulletin of the American Meteorological Society*, *94*(4), 485–498. https://doi.org/10.1175/BAMS-D-12-00170.1

Dieng, D., Cannon, A. J., Laux, P., Hald, C., Adeyeri, O., Rahimi, J., et al. (2022). Multivariate Bias-Correction of High-Resolution Regional Climate Change Simulations for West Africa: Performance and Climate Change Implications. *Journal of Geophysical Research: Atmospheres*, *127*(5), e2021JD034836. https://doi.org/10.1029/2021JD034836

Diouf, I., Sy, S., Senghor, H., Fall, P., Diouf, D., Diakhaté, M., et al. (2022). Potential Contribution of Climate Conditions on COVID-19 Pandemic Transmission over West and North African Countries. *Atmosphere*, *13*(1), 34. https://doi.org/10.3390/atmos13010034

Dosio, A., Lennard, C., & Spinoni, J. (2022). Projections of indices of daily temperature and precipitation based on bias-adjusted CORDEX-Africa regional climate model simulations. *Climatic Change*, *170*(1), 13. https://doi.org/10.1007/s10584-022-03307-0

Dutra, E., Viterbo, P., Miranda, P. M. A., & Balsamo, G. (2012). Complexity of Snow Schemes in a Climate Model and Its Impact on Surface Energy and Hydrology. *Journal of Hydrometeorology*, *13*(2), 521–538. https://doi.org/10.1175/JHM-D-11-072.1

Dutra, E., Muñoz-Sabater, J., Boussetta, S., Komori, T., Hirahara, S., & Balsamo, G. (2020). Environmental Lapse Rate for High-Resolution Land Surface Downscaling: An Application to ERA5. *Earth and Space Science*, *7*(5), e2019EA000984. https://doi.org/10.1029/2019EA000984

Elvidge, A. D., Sandu, I., Wedi, N., Vosper, S. B., Zadra, A., Boussetta, S., et al. (2019). Uncertainty in the Representation of Orography in Weather and Climate Models and Implications for Parameterized Drag. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2567–2585. https://doi.org/10.1029/2019MS001661

Essa, Y. H., Cagnazzo, C., Madonna, F., Cristofanelli, P., Yang, C., Serva, F., et al. (2022). Intercomparison of Atmospheric Upper-Air Temperature From Recent Global Reanalysis Datasets. *Frontiers in Earth Science*, *10*. Retrieved from https://www.frontiersin.org/articles/10.3389/feart.2022.935139

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Flaounas, E., Drobinski, P., Borga, M., Calvet, J.-C., Delrieu, G., Morin, E., et al. (2012). Assessment of gridded observations used for climate model validation in the Mediterranean region: the HyMeX and MED-CORDEX framework. *Environmental Research Letters*, *7*(2), 024017. https://doi.org/10.1088/1748-9326/7/2/024017

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Gibson, P. B., Waliser, D. E., Lee, H., Tian, B., & Massoud, E. (2019). Climate Model Evaluation in the Presence of Observational Uncertainty: Precipitation Indices over the Contiguous United States. *Journal of Hydrometeorology*, *20*(7), 1339–1357. https://doi.org/10.1175/JHM-D-18-0230.1

Giorgi, F., Coppola, E., Solmon, F., Mariotti, L., Sylla, M. B., Bi, X., et al. (2012). RegCM4: model description and preliminary tests over multiple CORDEX domains. *Climate Research*, *52*, 7–29. https://doi.org/10.3354/cr01018

Giorgi, Filippo, Jones, C., & Asrar, G. R. (2009). *Addressing climate information needs at the regional level: the CORDEX framework* (p. 9).

Gómez-Navarro, J. J., Montávez, J. P., Jerez, S., Jiménez-Guerrero, P., & Zorita, E. (2012). What is the role of the observational dataset in the evaluation and scoring of climate models? *Geophysical Research Letters*, *39*(24). https://doi.org/10.1029/2012GL054206

Hausfather, Z., Cowtan, K., Menne, M. J., & Williams Jr., C. N. (2016). Evaluating the impact of U.S. Historical Climatology Network homogenization using the U.S. Climate Reference Network. *Geophysical Research Letters*, *43*(4), 1695–1701. https://doi.org/10.1002/2015GL067640

Herold, N., Alexander, L. V., Donat, M. G., Contractor, S., & Becker, A. (2016). How much does it rain over land? *Geophysical Research Letters*, *43*(1), 341–348. https://doi.org/10.1002/2015GL066615

Herrera, S., Kotlarski, S., Soares, P. M. M., Cardoso, R. M., Jaczewski, A., Gutiérrez, J. M., & Maraun, D. (2019). Uncertainty in gridded precipitation products: Influence of station density, interpolation method and grid resolution. *International Journal of Climatology*, *39*(9), 3717–3729. https://doi.org/10.1002/joc.5878

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Higgins, R. W., Shi, W., Yarosh, E., & joyce, R. (2000). Improved United States precipitation quality control system and analysis. NCEP/Climate Prediction Center ATLAS No. 7. *National Oceanic and Atmospheric Administration*, 40.

Hofstra, N., New, M., & McSweeney, C. (2010). The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dynamics*, *35*(5), 841–858. https://doi.org/10.1007/s00382-009-0698-1

Hsu, W.-C., Patricola, C. M., & Chang, P. (2019). The impact of climate model sea surface temperature biases on tropical cyclone simulations. *Climate Dynamics*, *53*(1), 173–192. https://doi.org/10.1007/s00382-018-4577-5

Kotlarski, S., Szabó, P., Herrera, S., Räty, O., Keuler, K., Soares, P. M., et al. (2019). Observational uncertainty and regional climate model evaluation: A pan-European perspective. *International Journal of Climatology*, *39*(9), 3730–3749. https://doi.org/10.1002/joc.5249

Kuhn, M., & Olefs, M. (2020). Elevation-Dependent Climate Change in the European Alps. *Oxford Research Encyclopedia of Climate Science*. https://doi.org/10.1093/acrefore/9780190228620.013.762

Li, Y., Wang, T., Zeng, Z., Peng, S., Lian, X., & Piao, S. (2016). Evaluating biases in simulated land surface albedo from CMIP5 global climate models. *Journal of Geophysical Research: Atmospheres*, *121*(11), 6178–6190. https://doi.org/10.1002/2016JD024774

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., et al. (2013). A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions. *Journal of Climate*, *26*(23), 9384–9392. https://doi.org/10.1175/JCLI-D-12-00508.1

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., et al. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Scientific Data*, *2*(1), 150042. https://doi.org/10.1038/sdata.2015.42

Madonna, F., Tramutola, E., Sy, S., Serva, F., Proto, M., Rosoldi, M., et al. (2022). The New Radiosounding HARMonization (RHARM) Data Set of Homogenized Radiosounding Temperature, Humidity, and Wind Profiles With Uncertainties. *Journal of Geophysical Research: Atmospheres*, *127*(2), e2021JD035220. https://doi.org/10.1029/2021JD035220

Martynov, A., Laprise, R., Sushama, L., Winger, K., Šeparović, L., & Dugas, B. (2013). Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation. *Climate Dynamics*, *41*(11), 2973–3005. https://doi.org/10.1007/s00382-013-1778-9

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., & Nijssen, B. (2002). A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States. *Journal of Climate*, *15*(22), 3237–3251. https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2

Mearns, L., McGinnis, S., Korytina, D., Arritt, R., Biner, S., Bukovsky, M., et al. (2017). The NA-CORDEX dataset [NetCDF]. https://doi.org/10.5065/D6SJ1JCH

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., et al. (2006). North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, *87*(3), 343–360. https://doi.org/10.1175/BAMS-87-3-343

Militino, A. F., Ugarte, M. D., Goicoa, T., & Genton, M. (2015). Interpolation of daily rainfall using spatiotemporal models and clustering. *International Journal of Climatology*, *35*(7), 1453–1464. https://doi.org/10.1002/joc.4068

Miralles, D. G., Gentine, P., Seneviratne, S. I., & Teuling, A. J. (2019). Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. *Annals of the New York Academy of Sciences*, *1436*(1), 19–35. https://doi.org/10.1111/nyas.13912

Mitchell, T. J., Knapp, P. A., & Patterson, T. W. (2019). Changes in southeastern USA summer precipitation event types using instrumental (1940–2018) and tree-ring (1790–2018) data. *Environmental Research Communications*, *1*(11), 111005. https://doi.org/10.1088/2515-7620/ab4cd6

Napoli, A., Crespi, A., Ragone, F., Maugeri, M., & Pasquero, C. (2019). Variability of orographic enhancement of precipitation in the Alpine region. *Scientific Reports*, *9*(1), 13352. https://doi.org/10.1038/s41598-019-49974-5

Palecki, M. A., & Bell, J. E. (2013). U.S. Climate Reference Network Soil Moisture Observations with Triple Redundancy: Measurement Variability. *Vadose Zone Journal*, *12*(2), vzj2012.0158. https://doi.org/10.2136/vzj2012.0158

Pepin, N., Bradley, R. S., Diaz, H. F., Baraer, M., Caceres, E. B., Forsythe, N., et al. (2015). Elevation-dependent warming in mountain regions of the world. *Nature Climate Change*, *5*(5), 424–430. https://doi.org/10.1038/nclimate2563

Pepin, N. C., Arnone, E., Gobiet, A., Haslinger, K., Kotlarski, S., Notarnicola, C., et al. (2022). Climate Changes and Their Elevational Patterns in the Mountains of the World. *Reviews of Geophysics*, *60*(1), e2020RG000730. https://doi.org/10.1029/2020RG000730

Prein, A. F., & Gobiet, A. (2017). Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *International Journal of Climatology*, *37*(1), 305–327. https://doi.org/10.1002/joc.4706

Rangwala, I., & Miller, J. R. (2012). Climate change in mountains: a review of elevation-dependent warming and its possible causes. *Climatic Change*, *114*(3–4), 527–547. https://doi.org/10.1007/s10584-012-0419-3

Ricketts, T. H., Dinerstein, E., Olson, D. M., Eichbaum, W., Loucks, C. J., DellaSala, D. A., et al. (1999). *Terrestrial Ecoregions of North America: A Conservation Assessment*. Island Press.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., et al. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, *24*(14), 3624–3648. https://doi.org/10.1175/JCLI-D-11-00015.1

Samuelsson, P., Jones, C. G., Willén, U., Ullerstig, A., Gollvik, S., Hansson, U., et al. (2011). The Rossby Centre Regional Climate model RCA3: model description and performance. *Tellus A*, *63*(1), 4–23. https://doi.org/10.1111/j.1600-0870.2010.00478.x

Sandu, I., van Niekerk, A., Shepherd, T. G., Vosper, S. B., Zadra, A., Bacmeister, J., et al. (2019). Impacts of orography on large-scale atmospheric circulation. *Npj Climate and Atmospheric Science*, *2*(1), 1–8. https://doi.org/10.1038/s41612-019-0065-9

Schumacher, D. L., Keune, J., Dirmeyer, P., & Miralles, D. G. (2022). Drought self-propagation in drylands due to land–atmosphere feedbacks. *Nature Geoscience*, *15*(4), 262–268. https://doi.org/10.1038/s41561-022-00912-7

Scinocca, J. F., Kharin, V. V., Jiao, Y., Qian, M. W., Lazare, M., Solheim, L., et al. (2016). Coordinated Global and Regional Climate Modeling. *Journal of Climate*, *29*(1), 17–35. https://doi.org/10.1175/JCLI-D-15-0161.1

Šeparović, L., Alexandru, A., Laprise, R., Martynov, A., Sushama, L., Winger, K., et al. (2013). Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model. *Climate Dynamics*, *41*(11), 3167–3201. https://doi.org/10.1007/s00382-013-1737-5

Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of Computational Physics*, *227*(7), 3465–3485. https://doi.org/10.1016/j.jcp.2007.01.037

Srivastava, A., Grotjahn, R., & Ullrich, P. A. (2020). Evaluation of historical CMIP6 model simulations of extreme precipitation over contiguous US regions. *Weather and Climate Extremes*, *29*, 100268. https://doi.org/10.1016/j.wace.2020.100268

Srivastava, A., Grotjahn, R., Ullrich, P. A., & Zarzycki, C. (2022). Evaluation of precipitation indices in suites of dynamically and statistically downscaled regional climate models over Florida. *Climate Dynamics*, *58*(5), 1587–1611. https://doi.org/10.1007/s00382-021-05980-w

Sy, S., Noblet-Ducoudré, N. D., Quesada, B., Sy, I., Dieye, A. M., Gaye, A. T., & Sultan, B. (2017). Land-Surface Characteristics and Climate in West Africa: Models' Biases and Impacts of Historical Anthropogenically-Induced Deforestation. *Sustainability*, *9*(10), 1917. https://doi.org/10.3390/su9101917

Sy, S., Madonna, F., Rosoldi, M., Tramutola, E., Gagliardi, S., Proto, M., & Pappalardo, G. (2021). Sensitivity of trends to estimation methods and quantification of subsampling effects in global radiosounding temperature and humidity time series. *International Journal of Climatology*, *41*(S1), E1992–E2014. https://doi.org/10.1002/joc.6827

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. https://doi.org/10.1029/2000JD900719

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Thorne, P. W., Madonna, F., Schulz, J., Oakley, T., Ingleby, B., Rosoldi, M., et al. (2017). Making better sense of the mosaic of environmental measurement networks: a system-of-systems approach and quantitative assessment. *Geoscientific Instrumentation, Methods and Data Systems*, *6*(2), 453–472. https://doi.org/10.5194/gi-6-453-2017

Thorne, P. W., Diamond, H. J., Goodison, B., Harrigan, S., Hausfather, Z., Ingleby, N. B., et al. (2018). Towards a global land surface climate fiducial reference measurements network. *International Journal of Climatology*, *38*(6), 2760–2774. https://doi.org/10.1002/joc.5458

Thornton, P. E., Shrestha, R., Thornton, M., Kao, S.-C., Wei, Y., & Wilson, B. E. (2021). Gridded daily weather data for North America with comprehensive uncertainty quantification. *Scientific Data*, *8*(1), 190. https://doi.org/10.1038/s41597-021-00973-0

Velasquez, P., Messmer, M., & Raible, C. C. (2019). *A new bias-correction method for precipitation over complex terrain suitable for different climate states* (preprint). Climate and Earth System Modeling. https://doi.org/10.5194/gmd-2019-131

WMO. (2008). *Guide to Meteorological Instruments and Methods of Observation (7th edition), Geneva: World MeteorologicalOrganization*.

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., & Liu, C. (2007). A Gauge-Based Analysis of Daily Precipitation over East Asia. *Journal of Hydrometeorology*, *8*(3), 607–626. https://doi.org/10.1175/JHM583.1

959 Yu, G., Wright, D. B., & Davenport, F. V. (2022). Diverse Physical Processes Drive Upper-Tail Flood
960 Quantiles in the US Mountain West. *Geophysical Research Letters*, *49*(10), e2022GL098855.
961 https://doi.org/10.1029/2022GL098855
962 Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., et al. (2011). Indices
963 for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley*
964 *Interdisciplinary Reviews: Climate Change*, *2*(6), 851–870.
965

966

Journal of Geophysical Research, Atmospheres

Supporting Information for

Assessment of NA-CORDEX regional climate models, reanalysis, and in-situ gridded-observational against U.S. Climate Reference Network datasets

Souleymane SY[1#], Fabio Madonna[1], Federico Serva[2+], Ismaila Diallo[3] and Benjamin Quesada[4]

[1]Consiglio Nazionale delle Ricerche – Istituto di Metodologie per l'Analisi Ambientale, C. da S. Loja - Zona Industriale, I-85050 Tito Scalo (Potenza), Italy

[#]Now at Institute of Geography, University of Augsburg, 86159 Augsburg, Germany

[2]Consiglio Nazionale delle Ricerche-Istituto di Scienze Marine (CNR-ISMAR), Rome, Italy

[+]Now at Italian Space Agency, Rome, Italy, and European Space Agency, Frascati, Italy

[3]Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA 16802, United States

[4]Universidad del Rosario, Faculty of Natural Sciences and Earth System Science Program, 'Interactions Climate-Ecosystems (ICE)' Research Group, Kr 26No 63B-48 (Bogotá D.C.), Colombia

Corresponding author: Dr. Souleymane SY (souleymane.sy@geo.uni-augsburg.de)

Contents of this file

Figures S1 to S10

Supplementary Information

Supplementary Figure Captions:

Figure S1: Box-and-whisker plots showing the seasonal daily mean temperature bias (K) for each of the selected reanalysis (NARR, ERAI, ERA5 and MERRA2), gridded observational products (Livneh, CPC, Daymet and PRISM), and individual NA-CORDEX models listed in Table 2 (at 50km, 25km and 12.5km resolutions) in each study subregion with respect to USCRN for summer (JJA) and winter (DJF) during the period 2006−2014. The median is indicated with a black line while the lower hinge of each box is Q1 quartile (25th), and the upper hinge for Q3 quartile (75th). The upper bar represents maximal value whereas the lower bar represents minimal value. The outliers are not shown in the plot. The subregional daily mean distribution of each simulation is obtained by spatially averaging all values within the subregion highlighted in Figure 1.

Figure S2: As Figure S1 but for the daily mean precipitation bias (mm/day).

Figure S3: Seasonal monthly mean rainfall distribution (mm/day) over the different subregions obtained from each of the individual models at 0.44° (blue lines), at 0.22° (red lines) and at 0.11° (solid dark-red line), for each of the four reanalysis (NARR, ERAI, ERA5, and MERRA-2) (green-lines), the gridded-observational products (Livneh, CPC, Daymet, PRISM) (in yellow lines) against USCRN datasets for the period 2006–2014 represented by the solid black line.

Figure S4: As Figure S3 for the seasonal monthly mean temperature distribution (K).

Figure S5: As Figure 7 but for seasonal monthly mean temperature distribution (K) over the different subregions.

Figure S6: As figure 8 but for temperature seasonal cycle in all subregions in summer.

Figure S7: As figure S6 but for temperature seasonal cycle in all subregions in winter.

Figure S8: As figure S6 but for precipitation seasonal cycle in all subregions in summer

Figure S9: As figure S6 but for precipitation seasonal cycle in all subregions in winter.

Figure S10: Precipitation-dependent temperature biases. The relations between precipitation biases (x-axis) and temperature biases (y-axis) are drawn for MidWest region. Note that the MidWest region is considered due to its complex topography. Correlations are calculated using the Kendall non-linear rank (τ) test between the simulated temperature mean biases and precipitation mean biases for all Rean (a) and NAMs at all resolutions. Two-star symbols (**) are added when the correlation is significant at 99% confidence intervals (i.e., $p < 0.01$; 99% C.I), while one-star symbol (*) is added when the correlation is significant at 95% (i.e., $p < 0.05$; 99% C.I).

Figure S1: Box-and-whisker plots showing the seasonal daily mean temperature bias (K) for each of the selected reanalysis (NARR, ERAI, ERA5 and MERRA2), gridded observational products (Livneh, CPC, Daymet and PRISM), and individual NA-CORDEX models listed in Table 2 (at 50km, 25km and 12.5km resolutions) in each study subregion with respect to USCRN for summer (JJA) and winter (DJF) during the period 2006−2014. The median is indicated with a black line while the lower hinge of each box is Q1 quartile (25th), and the upper hinge for Q3 quartile (75th). The upper bar represents maximal value whereas the lower bar represents minimal value. The outliers are not shown in the plot. The subregional daily mean distribution of each simulation is obtained by spatially averaging all values within the subregion highlighted in Figure 1.

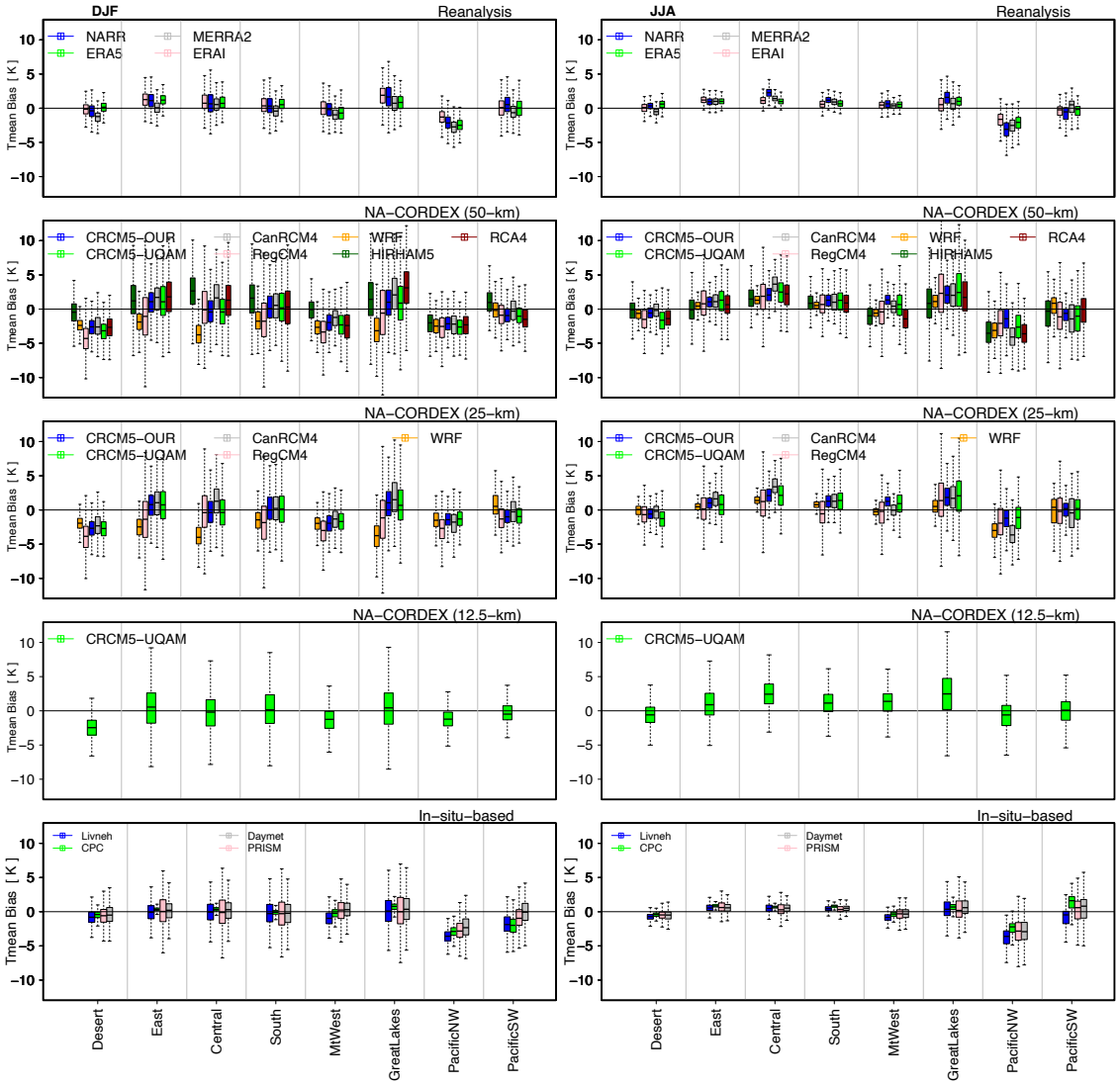

Figure S2: As Figure S1 but for the daily mean precipitation bias (mm/day)

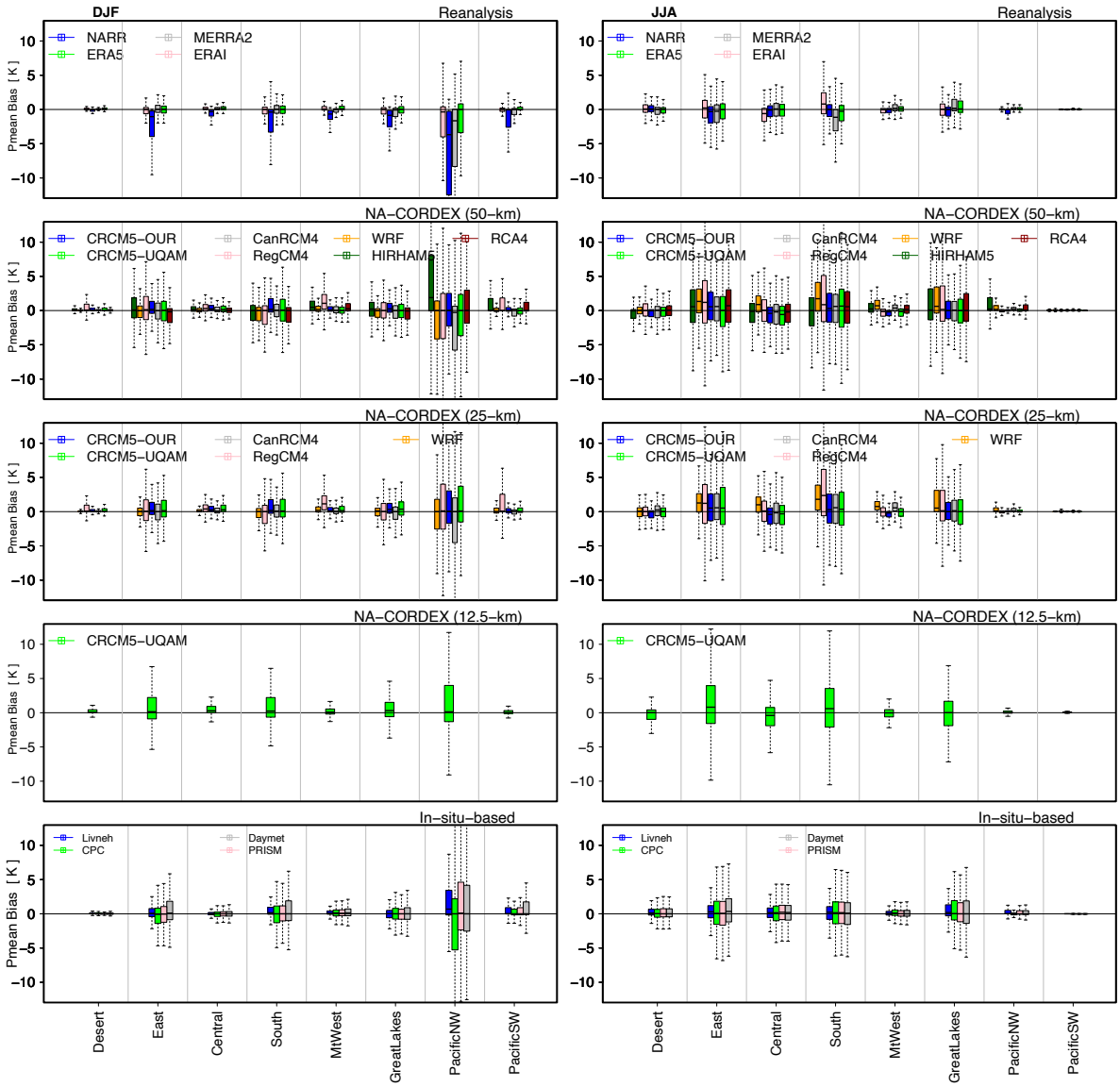

Figure S3: Seasonal monthly mean rainfall distribution (mm/day) over the different subregions obtained from the individual models at 0.44° (blue lines), at 0.22° (red lines) and at 0.11° (solid dark-red line), for each of the reanalysis (NARR, ERAI, ERA5, and MERRA-2) (green-lines), the gridded-observational products (Livneh, CPC, Daymet, PRISM) (in yellow lines) against USCRN (solid black line) for the period 2006–2014.

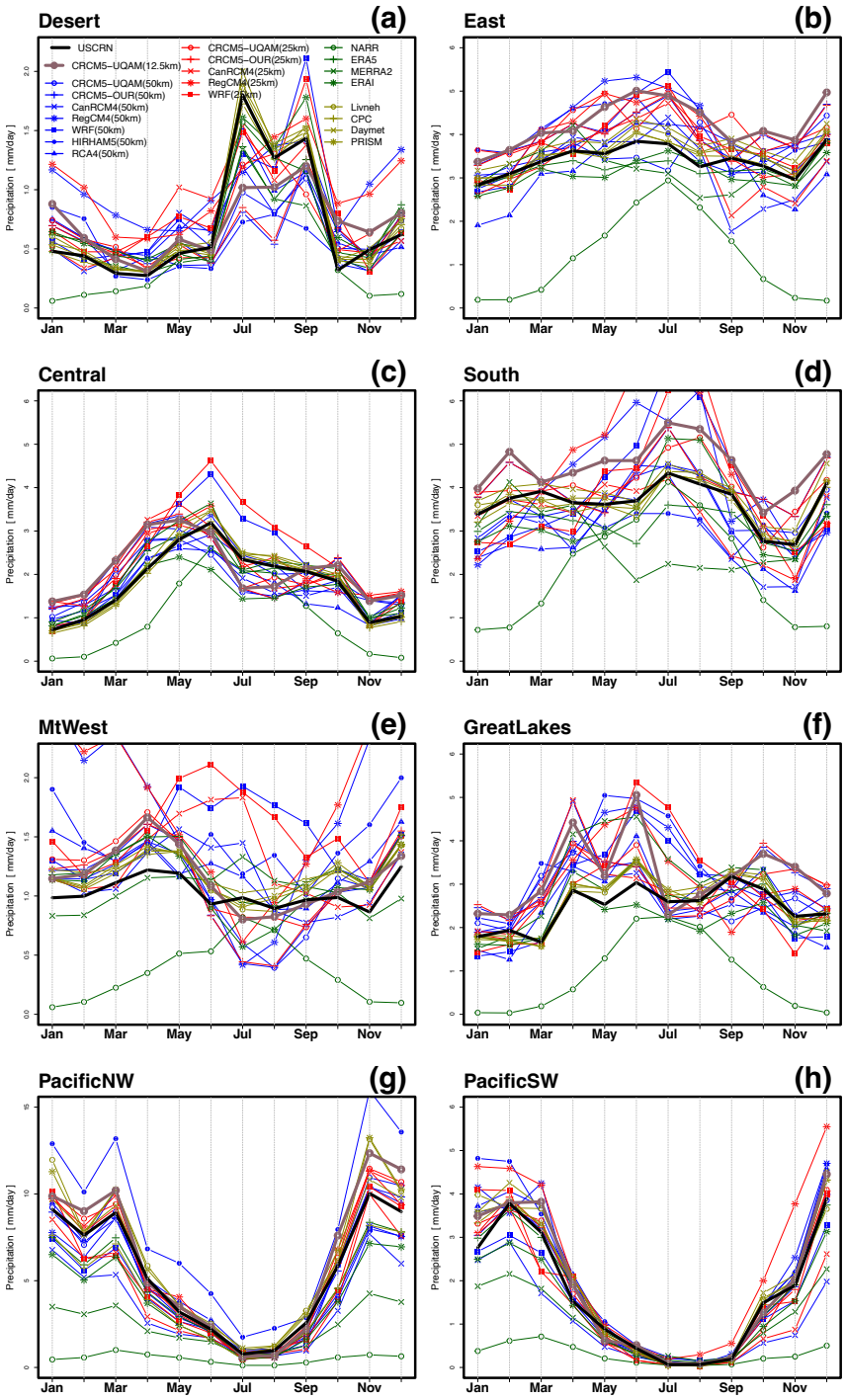

Figure S4: As Figure S3 for the seasonal monthly mean temperature distribution (K)

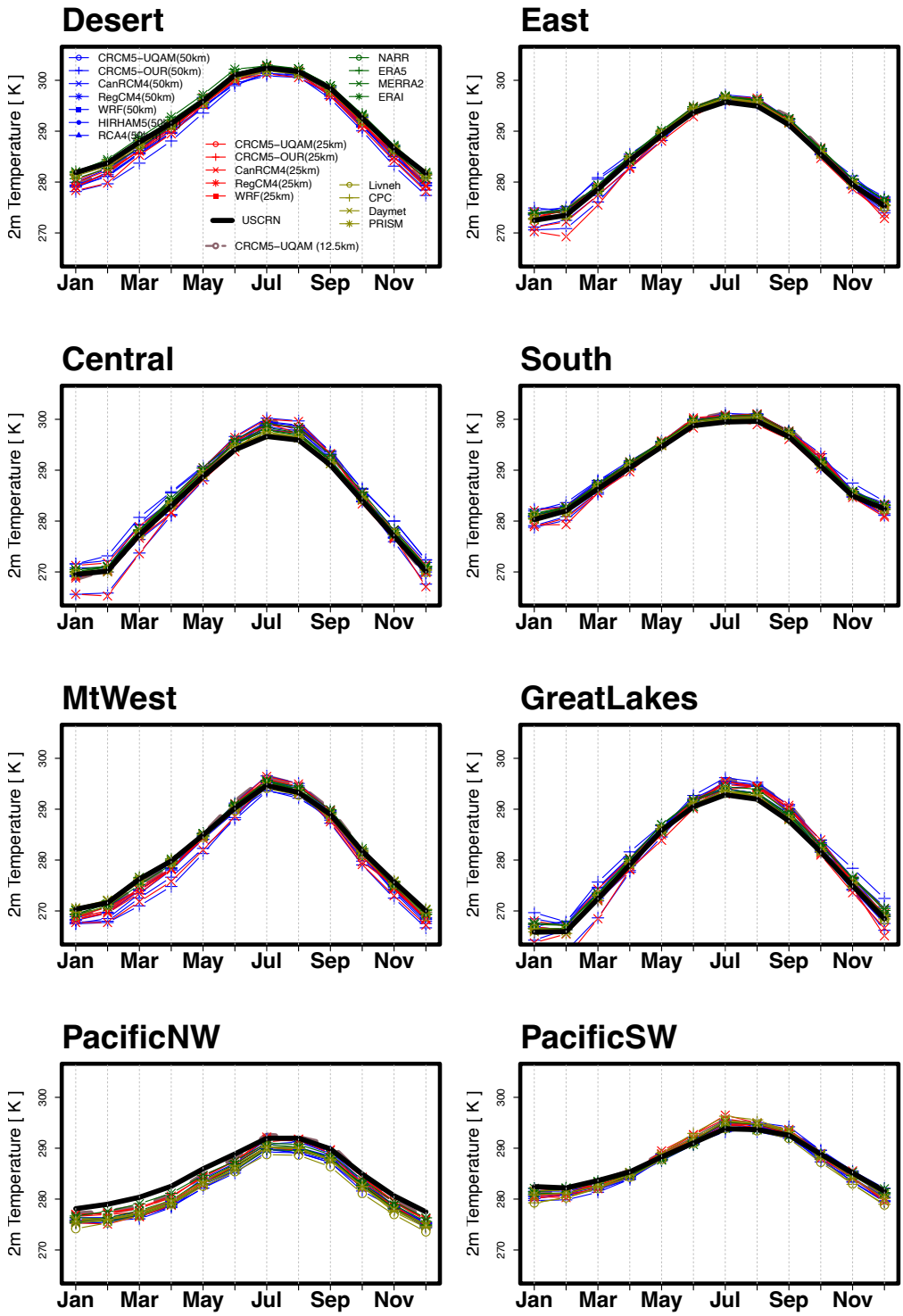

Figure S5: As Figure 7 as for seasonal monthly mean temperature distribution (K) over the different subregions.

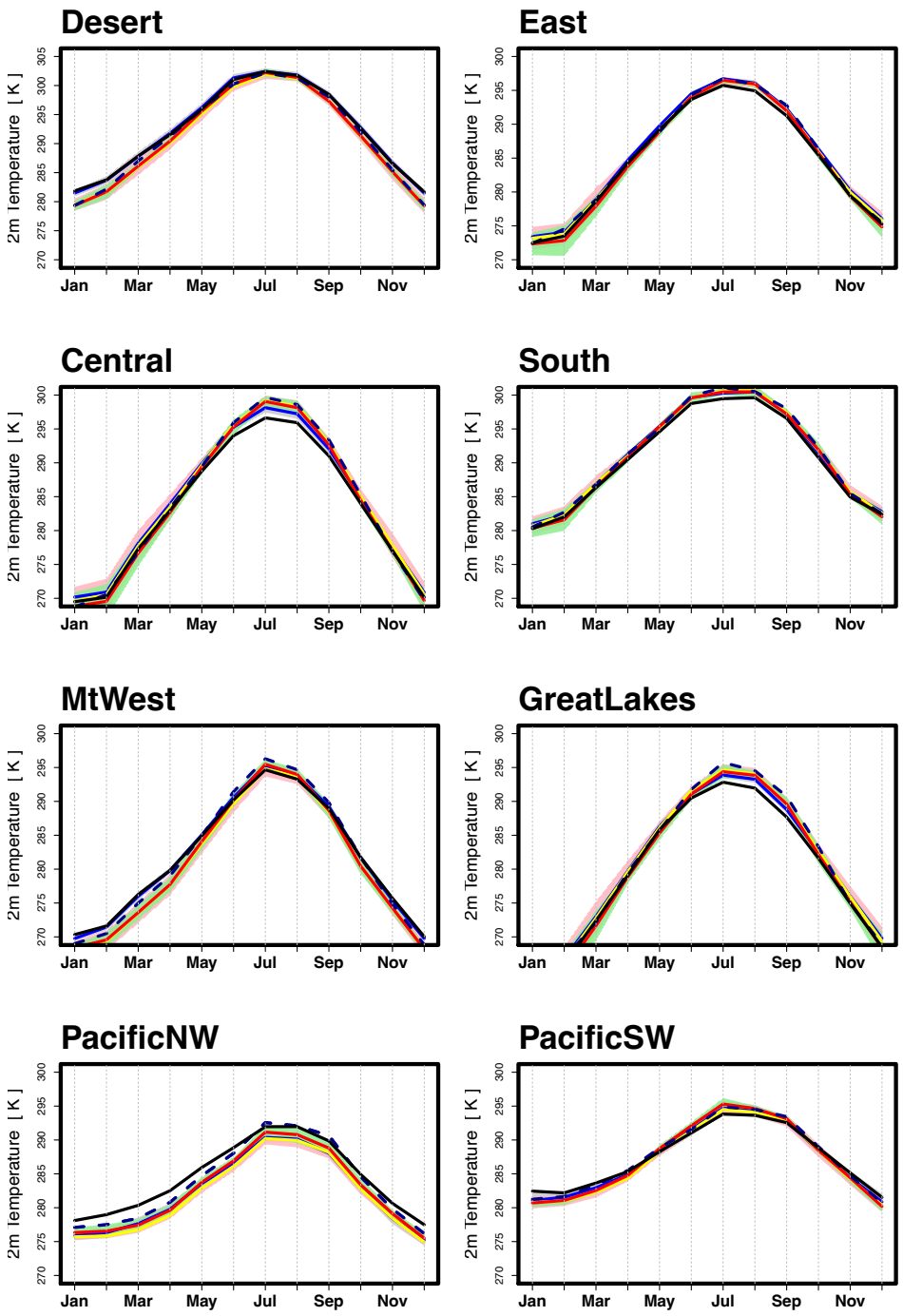

Figure S6: As figure 8 but for temperature seasonal cycle in all subregions in summer

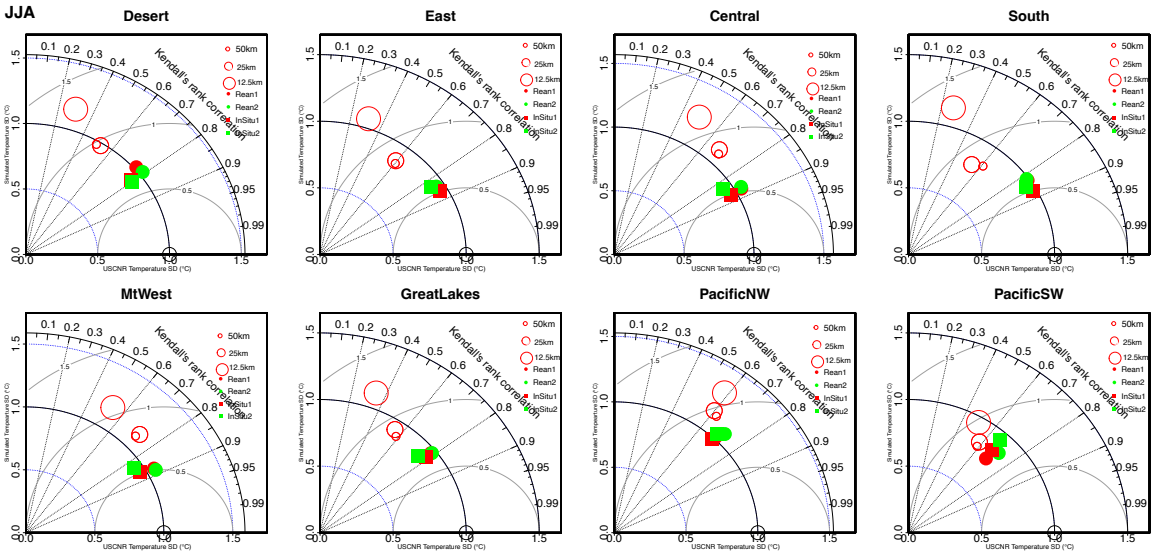

Figure S7: As figure S6 but for temperature seasonal cycle in all subregions in winter

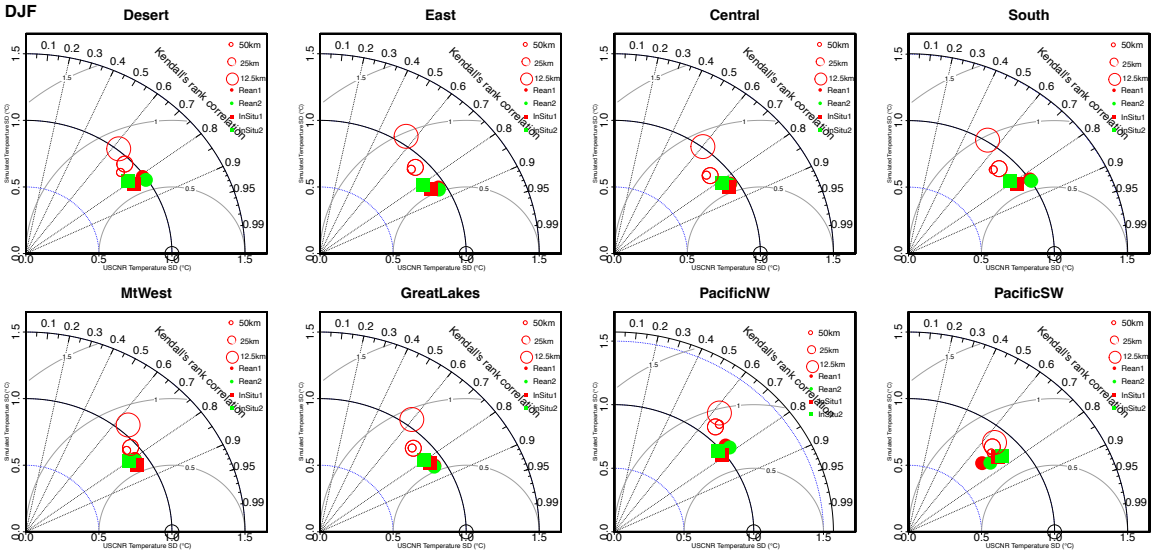

Figure S8: As figure S6 but for precipitation seasonal cycle in all subregions in summer

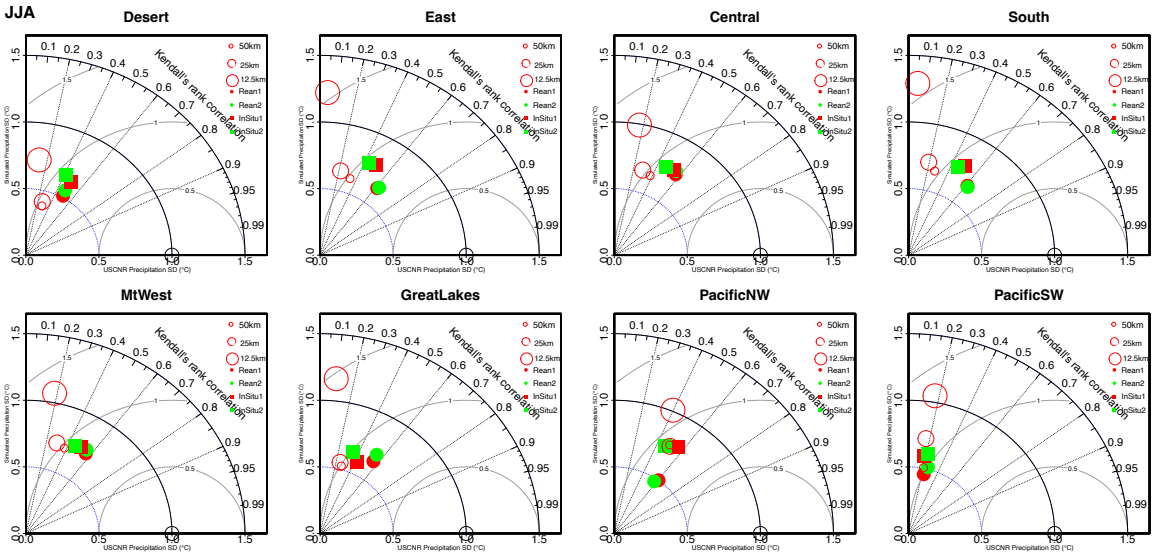

Figure S9: As figure S6 but for precipitation seasonal cycle in all subregions in winter

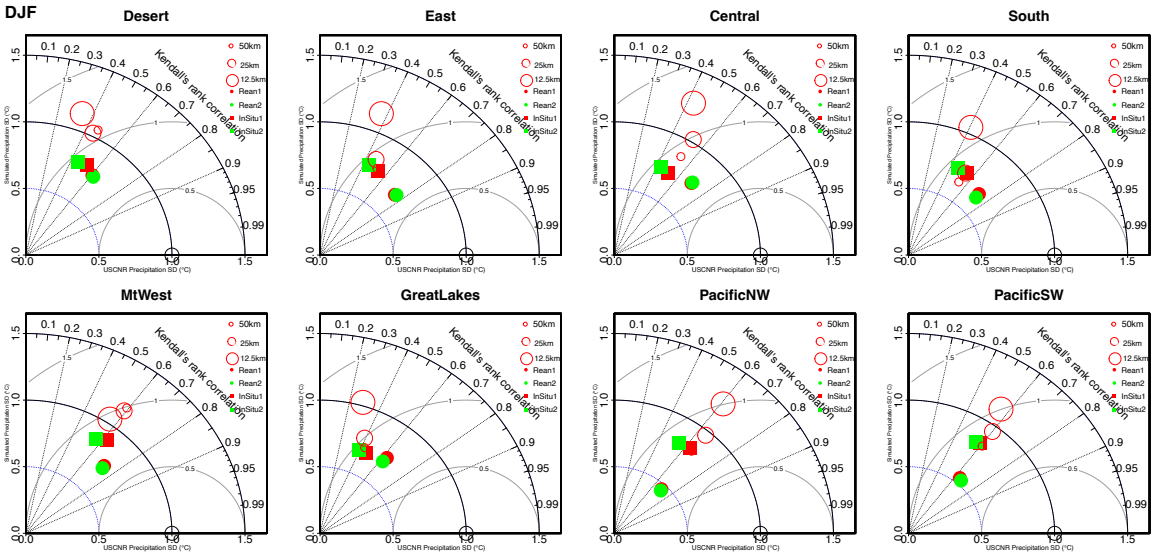

Figure S10: Precipitation-dependent temperature biases. The relations between precipitation biases (x-axis) and temperature biases (y-axis) are drawn for MidWest region. Note that the MidWest region is considered due to its complex topography. Correlations are calculated using the Kendall non-linear rank (τ) test between the simulated temperature mean biases and precipitation mean biases for all Rean (a) and NAMs at all resolutions. Two-star symbols (**) are added when the correlation is significant at 99% confidence intervals (i.e., $p < 0.01$; 99% C.I), while one-star symbol (*) is added when the correlation is significant at 95% (i.e., $p < 0.05$; 99% C.I).