AGU100 ADVANCING EARTH AND SPACE SCIENCE

Earth and Space Science

RESEARCH ARTICLE
10.1029/2018EA000426

High-Resolution Climate Projections for the Northeastern United States Using Dynamical Downscaling at Convection-Permitting Scales

M. Komurcu[1,2], K. A. Emanuel[2,3], M. Huber[4,5], and R. P. Acosta[4]

[1]Joint Program on the Science and Policy of Global Change, Center for Global Change Science, Massachusetts Institute of Technology, Cambridge, MA, USA, [2]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA, [3]Lorenz Center, Massachusetts Institute of Technology, Cambridge, MA, USA, [4]Department of Earth, Atmospheric and Planetary Sciences, Purdue University, Lafayette, IN, USA, [5]Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, NH, USA

Key Points:
- High-resolution dynamically downscaled projections are produced for the northeastern United States
- High-resolution dynamical downscaling with realistic boundary forcing can improve climate means and extremes
- Using detailed parameterizations and convection-permitting simulations can improve climate means and extremes

Correspondence to:
M. Komurcu,
muge@mit.edu

Citation:
Komurcu, M., Emanuel, K. A., Huber, M., & Acosta, R. P. (2018). High-resolution climate projections for the northeastern United States using dynamical downscaling at convection-permitting scales. *Earth and Space Science*, *5*, 801–826. https://doi.org/10.1029/2018EA000426

Received 22 JUN 2018
Accepted 2 NOV 2018
Accepted article online 8 NOV 2018
Published online 27 NOV 2018

Abstract To paraphrase former Speaker of the House Tip O'Neill, "All climate change is local"—that is, society reacts most immediately to changes in local weather such as regional heat waves and heavy rainstorms. Such phenomena are not well resolved by the current generation of coupled climate models. Here it is shown that dynamical downscaling of climate reanalyses using a high-resolution regional model can reproduce both the means and extremes of temperature and precipitation as observed in the well-measured northeastern United States. Given this result, the downscaling is applied to climate projections for the middle and end of the 21st century under Representative Concentration Pathway (RCP) 8.5 as well as for the historical time period to help assess regional climate impacts in the northeastern United States. The resulting high-resolution projections are intended to support regional sustainability studies for the northeastern United States and are made publicly available.

1. Introduction

Earth system models (ESM), comprised of coupled ocean, land, atmosphere, and ice sheet components, are able to simulate bulk properties of large-scale processes, global climate, and climate change and the response of the climate system to different forcings for past, present, and future time periods (e.g., Heavens et al., 2013; Prinn, 2013). However, the coarse resolution of these models prevents detailed analysis of climate change at regional and local scales. Comprehensive analysis of regional impacts of climate change, such as changes in climate extremes, water resources, and various other elements crucial for future planning, requires local, high-resolution climate variables that cannot be obtained directly from coarse-resolution ESM projections. High-resolution global ESM simulations are currently too expensive to be practical. Therefore, downscaling methodologies are proposed to produce the high-resolution climate variables that are needed to assess climate change impacts at regional scales.

One such methodology is dynamical downscaling using a regional climate model, where regional climate models are run with lateral and initial boundary conditions from ESM projections to produce higher resolution climate variables (e.g., Caldwell et al., 2009; Done et al., 2015; Dosio et al., 2014; Gao et al., 2012; Mearns et al., 2012; Pierce et al., 2012; Shaaf et al., 2017; Sun et al., 2016). Similarly, reanalysis products can be downscaled to assess regional phenomena and to understand historical changes in regional scales (e.g., Caldwell et al., 2009). Once high-resolution variables are generated, they are scale-appropriate to be used in further models and analysis to assess changes in a variety of regional elements. These subsequent models and analyses aim to project not only changes in climate means but also in extremes, such as heat, extreme rain, and flooding events. Physically based downscaling methods can potentially capture significant changes in extreme events in a more physically consistent way than statistical downscaling tuned to the present climate.

Several dynamical downscaling studies have been performed over the years (e.g., Caldwell et al., 2009; Done et al., 2015; Mearns et al., 2012; Mearns et al., 2017). Some of the key lessons learned from dynamical downscaling studies are summarized in the 2013 Intergovernmental Panel on Climate Change report (Flato et al., 2013). The report hints to some evidence that dynamical downscaling adds value by resolving convective precipitation, through detailed representation of topographical features and coastlines and

allowing simulation of smaller scale processes. It recommends caution in choice of model setup such as domain size, location, and application of boundary conditions. For example, large domain sizes, which allow the regional model more freedom to generate small-scale features, and applying no nudging at boundaries to avoid deterioration of extremes and two-way nesting, which allows for the higher resolution nested grids to communicate with the coarser grids, are more favored techniques that yielded results closer to observations (e.g., Alexandru et al., 2009; Leduc & Laprise, 2009; Lorenz & Jacob, 2005). Furthermore, as with all high-resolution modeling practices, to obtain climate variables more representative of the region simulated, synoptic events affecting the region and geographic location of boundaries should be taken into consideration upon model domain choice (e.g., Giorgi & Mearns, 1999).

Both global and regional models usually do a reasonable job simulating temperatures, but simulating precipitation rates have been problematic (Flato et al., 2013). Furthermore, simulating extreme precipitation events at high resolution is essential for planning and adaptation efforts. Going to a higher resolution can allow for better extremes and improved diurnal cycles of precipitation (e.g., Seneviratne et al., 2012; Walther et al., 2013). Higher resolution simulations, however, do not always guarantee better results. For example, simulations of precipitation for Southwest UK did not improve between 50, 12, and 1.5 km (Chan et al., 2013). Sun et al. (2016), on the other hand, used the Weather Research and Forecasting model (WRF) to simulate summer precipitation over Central Great Plains at 25- and 4-km horizontal resolution simulations with and without convection parameterization, respectively. They found that while the two simulations have comparable precipitation biases, 4-km simulation better replicated observed diurnal cycle and the magnitude of extreme precipitation. Furthermore, results of dynamical downscaling can be sensitive to the forcing data (e.g., Déqué et al., 2011; Dosio et al., 2014). For example, Dosio et al. (2014) used four climate models to drive regional climate modeling simulations over Africa to 0.4° resolution. They found that while the regional model was generally better at reproducing observed annual cycle of precipitation and precipitation statistics, boundary conditions strongly influence the spatial distribution of the downscaled temperatures, precipitation, and sea level pressure. Physics scheme used in the regional model also affects downscaled results. Most recently, Hu et al. (2018) performed downscaling simulations with reanalysis data using WRF at 20-km horizontal resolution over Southern Great Plains and found that model skill for simulating of warm season precipitation is influenced by the choice of physical parameterization. Furthermore, convection-permitting simulations using WRF over the continental United States showed that downscaling can reproduce the frequency and intensity of observed precipitation extremes for most regions (Prein et al., 2017). These studies suggest that the improvement obtained with higher resolution dynamical downscaling is dependent on the region simulated and the choice of parameterizations and model setup used in the regional model.

In this study, we use convection-permitting high-resolution regional climate modeling using WRF (Skamarock et al., 2008) to dynamically downscale global model projections to 3-km horizontal resolution and produce a high-resolution climate data set to support sustainable regional development efforts in the northeastern United States, specifically New Hampshire (NH). To reduce the bias in downscaled fields, particularly mean and extreme precipitation, we use a unique combination of model parameterizations and setup that we will detail in section 2. To show that our methodology works, meaning that downscaled results are in better agreement with observations compared to the driver data set, similar to Prein et al. (2017), we first perform ERA-Interim-driven historical WRF simulations (10 years) and compare the driver (ERA-Interim) and our dynamically downscaled results with gridded-observations, other reanalysis products, and station observations. This analysis is presented in section 3. Then, we apply the same downscaling methodology to historical and future projections from the Community Earth System Model (CESM) to determine projected regional changes in climate in high resolution. In section 4, we compare projected changes in regional climate obtained from our high-resolution simulations with those obtained from the CESM. Climate data produced at this fine scale are now scale appropriate to be used in smaller scale models and analysis (e.g., ecosystem, land surface hydrology, disease vector population, storm surge, and economy models) to help guide planning, mitigation, and adaptation efforts in the state of NH. We discuss our results in section 5. To our knowledge, this is the first and only study that has downscaled global model projections to such high resolution (3 km) for a long (55-year) time period for this region. As such, all model input, output, and restart fields from this study are available for public use. The details of how to access the data and the computational cost of this study are summarized in section 6.

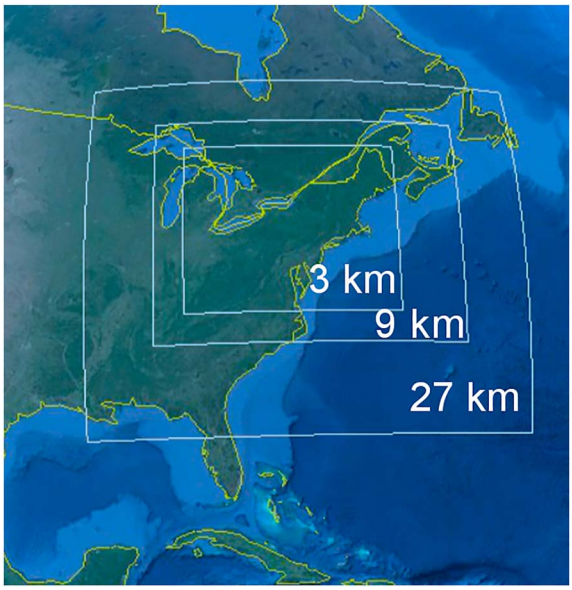

Figure 1. WRF simulation domain, three nested grids of 27-, 9-, and 3-km horizontal resolution, used in this study.

2. Methods

We use WRF v3.6.1 to dynamically downscale historical and future projections (55 years) of CESM. WRF is the U.S. *industry standard* high-resolution weather model, which is used both for research and is the basis of operational weather prediction (for example, see the National Center for Environmental Prediction [NCEP] 3-km High Resolution Rapid Update [HRRR] simulations http://ruc.noaa.gov/hrrr/). WRF has been shown to hindcast modern climate statistics over the continental United States and produce results in agreement with modern observations: For example, Harding and Snyder (2014) found that WRF captured the diurnal cycle, extremes, and averages of precipitation in historical simulations of the warm season in the central United States, and d'Orgeville et al. (2014) found that WRF compared well against observations of means and extremes of rainfall in the difficult Great Lakes regime.

CESM-driven dynamical downscaling using WRF has been used numerous times to simulate regional climate change (e.g., Caldwell et al., 2009; Pierce et al., 2012; Gao et al., 2012; Done et al., 2015). Gao et al. (2012), using WRF at 4-km resolution driven by CESM, were able to reproduce climate and extremes over the eastern United States and show statistically significant improvement in reproducing extreme weather events with WRF compared to CESM. Furthermore, over the past several years, WRF has become a standard part of the climate change modeling toolbox (Oleson et al., 2013; Wu et al., 2014). In fact, the combination of the CESM and WRF forms the core modeling capability of several major national efforts, including the Platform for Regional Integrated Modeling and Analysis (PRIMA) project (Kraucunas et al., 2014), the Department of Energy Pacific Northwest National Laboratory, and the nested regional climate modeling effort of the National Center for Atmospheric Research (NCAR; https://www2.cgd.ucar.edu/research/nested-regional-climate-model).

In this study, we dynamically downscale projections from a single ESM to 3-km horizontal resolution using convection-permitting WRF simulations. Due to the large computational cost associated with these simulations (see section 6), we are only able to perform downscaling experiments for a single ESM under one emission scenario. Studies such as the North American- Coordinated Regional Climate Downscaling Experiment (NA-CORDEX) and the North American Regional Climate Assessment Program (NARCCAP) are, on the other hand, able to downscale more ESM projections and evaluate changes under different climate change scenarios (Mearns et al., 2009, 2017). Hence, they can be used to study the structural uncertainties in climate change in regional climate modeling experiments in detail. These studies, however, are at much coarser resolution (25 or 50 km) compared to our products (3 km). While we cannot assess similar uncertainties using our experiments, what we create instead is a potential future scenario in high resolution. Our downscaled products are aimed to (1) provide high-resolution climate variables needed for regional impact models/analysis to support sustainable development efforts for the region (sustainability of water resources, ecosystems, resilient infrastructures, etc.) and (2) are intended to complement statistical downscaling efforts in NH. In fact, our study was partly supported through the National Science Foundation (NSF) NH Experimental Program to Stimulate Competitive Research (EPSCoR), Ecosystems and Society project, which aims to provide sustainable management of NH's natural resources.

2.1. Model Setup and Methodology

The WRF model domain used in this study consists of three nested grids of 27-, 9-, and 3-km horizontal grid spacing, with the highest resolution innermost nest focusing on the northeastern United States, as shown in Figure 1. The numbers of horizontal and vertical grids are $101 \times 121 \times 40$, $247 \times 187 \times 40$, and $508 \times 415 \times 40$ for domains 1, 2, and 3, respectively. To maintain computational stability, our simulations have a 20-s time step within the finest resolution innermost nest. We use an ~45-daily reinitialization of WRF from the driver data, meaning that we perform ~ 45-day-long WRF simulations, remove the first 15 days (e.g., 15 + 31, 15 + 28, and 15 + 30) to account for model initialization and spin-up, and obtain one simulation month. Each month is therefore individually simulated independent of the previous or subsequent months. We then repeat the same procedure for each of the 12 months for all years.

For downscaling simulations, as with all numerical simulations, domains should be large enough to incorporate regional features affecting the area (e.g., Giorgi & Mearns, 1999). Hence, the extent of our innermost, highest resolution domain is designed to ensure that we incorporate the effects of weather systems and processes known to affect the region, such as Nor'easters, landfalling hurricanes, and moist effects of the Great Lakes and the Atlantic Ocean. Furthermore, to account for the spatial spin-up and reduce boundary artifacts, simulation domains should be kept large and the region of interest should be away from the boundaries (Warner et al., 1997). Therefore, our model domain and nests are kept large, the regions of interest (primarily NH and MA) are kept as far away from the boundaries as is possible, and no nudging is applied at the boundaries to minimize numerical artifacts introduced from the boundaries to the innermost domain. Shaaf et al. (2017) investigated the effects of spectral nudging using a regional climate model with a horizontal resolution and model time step similar to ours and found no differences compared to simulations that did not include nudging. There is still debate on nudging-related numerical artifacts entering inner domains. The extents of our domain and nests, given the high-resolution nature of our simulations, are limited by computational cost.

Our simulations take advantage of two-way nesting to ensure that the CESM-driven lower resolution domain is interacting with the small-scale features simulated in the innermost, highest resolution nest (e.g., Wang et al., 2009). While computationally more expensive compared to one-way parasitic, nesting, in idealized studies two-way nesting has been shown to improve coarse grid fields feeding into nested grids and reduce boundary reflections (e.g., Harris & Durran, 2010).

Recently, it was shown that simulations of the track of Hurricane Sandy depend on the cumulus scheme used in the model (Bassill, 2014). In order to eliminate this dependency in our WRF simulations, we choose to at least partially resolve convection by simulating a larger area, where synoptic activity has been known to affect the region, with higher resolution. Hence, domain 3 is kept large and only domains 1 and 2 use cumulus parameterization. In the coarsest two domains, the Kain-Fritsch cumulus parameterization (Kain, 2004) is used. Convection-permitting simulations provide a stronger physical basis for interpreting modeled changes, especially in fields such as extreme precipitation and heat conditions, hence potentially add significant value to further modeling efforts of regional impacts of climate change. While convection-permitting modeling reduces the uncertainty from convection parameterization (e.g., Bassill, 2014), detailed numerical studies suggest a grid spacing of 1 km or smaller is required for adequate simulations of moist convection (e.g., Bryan et al., 2003). Nevertheless, at 3-km horizontal resolution, we are able to partially resolve convection and produce the high-resolution climate data the impact models need to assess climate change in local and regional scales.

Many ESMs use simple cloud microphysics parameterizations, and even the ones with more sophisticated parameterizations have been unable to capture the observed vertical distribution of cloud phase (e.g., Komurcu et al., 2014). This problem is due partly to the coarse horizontal and vertical resolution of ESMs and partly to a lack of detailed understanding of cloud microphysical processes. The driving ESM (CESM 1.0) of our WRF simulations uses Community Atmospheric Model version 4 (CAM 4), which uses a bulk microphysics parameterization with temperature-based phase partitioning of the total condensate. A prognostic approach for both cloud water and ice, however, is essential for simulating better precipitation, cloud ice, and cloud liquid water and to prevent amplified ice concentrations in models (e.g., Avramov & Harrington, 2010; Gettelman et al., 2015; Komurcu, 2016). Hence, for high-resolution, convection-permitting simulations, cloud microphysics parameterization becomes very important, especially when downscaling projections from ESMs using simple representations of cloud processes and cloud phase. While the next version of CESM to our driver ESM, CAM 5.1, similarly lacks prognostic treatment of cloud ice, more detailed microphysics options are available in the latest versions of the CESM.

To improve regional projections of precipitation compared to our driver ESM, we are using high-resolution WRF simulations with more detailed cloud microphysics parameterizations. Furthermore, the time step used in our WRF simulations is much smaller compared to the driver ESM (30 min vs. 20 s), which we expect to help improve simulated fields. While single-moment (predicting mass mixing ratio of the condensate) bulk microphysics schemes available for WRF are more efficient to run, double-moment schemes, which can predict both number and mass mixing ratio of hydrometeor categories, may lead to better simulations of ice and liquid water for winter precipitation (e.g., Reisner et al., 1998). The Morrison et al. (2009) double-moment

Table 1
List of Parameterizations Used in Our WRF Simulations

Parameterization	Parameterization choice
Land Surface Model	CLM (Olesson et al., 2010)
Microphysics	Morrison Double Moment (Morrison et al., 2009)
Planetary Boundary Layer	Yonsei University Scheme (Hong et al., 2006)
Longwave and Shortwave Radiation	RRTMG (Iacono et al., 2008)

cloud microphysics parameterization is used in our WRF simulations to improve simulations of cloud water and precipitation fields compared to our driver ESM. It provides prognostic treatment of number concentration and mass mixing ratios of liquid water and ice categories and has been shown to yield better winter precipitation, particularly snow amounts (Liu et al., 2011). Furthermore, along with the Thompson scheme (Thompson et al., 2008), it is known to simulate more realistic convection compared to other schemes (Thompson, 2012).

We use the Rapid Radiative Transfer Model for Global climate models (RRTMG; Iacono et al., 2008) for both long and shortwave radiation schemes for several reasons: (1) RRTMG provides detailed physical representations of radiative processes, (2) it is the same parameterization used in the driver model CESM (Iacono et al., 2008), and (3) in WRF it has an option to use increasing greenhouse gas concentrations with time following Representative Concentration Pathway (RCP) 8.5 emissions. Hence, in our WRF simulations driven with CESM, greenhouse gas emissions follow RCP 8.5 emissions in time and their effects are reflected in radiation calculations.

We use the Community Land Model (CLM) version 4.0 (Oleson et al., 2010) as the land surface model (LSM) in WRF. CLM includes detailed representations of land surface and snow processes (Oleson et al., 2010), and it is the land surface parameterization in the driving ESM (CESM). These detailed representations, combined with the high-resolution nature of our simulations, are expected to improve simulated fields. For example, by using higher resolution soil properties and modifying the convection parameterization in the Regional Atmospheric Modeling System (RAMS), Miguez-Macho et al. (2005) were able to improve the biases in temperature in regional climate simulations over North America. Their study used larger grid spacing (50 km) compared to our setup.

For the planetary boundary layer (PBL) parameterization, we use the Yonsei University scheme (Hong et al., 2006) because of its explicit treatment of entrainment at the PBL top and because it has been shown to match PBL height as observed from lidar and balloon observations (e.g., Balzarini et al., 2014).

Using detailed cloud microphysics, PBL, LSM, and radiation schemes in high-resolution convection-permitting simulations over a large domain with no nudging at boundaries and two-way feedback between grids (as opposed to one way pass of information from the coarse to high-resolution grids), we are essentially aiming to improve regional projections (particularly precipitation) through (1) alleviating the effects of some of the less detailed parameterizations used in the driving ESM, (2) lessening the effects of ESM's coarse grid size and time step, and (3) allowing for detailed representation of spatial features, hence interactions between small-scale processes (surface-atmosphere-clouds).

Key parameterizations used in our WRF model setup are summarized in Table 1.

To evaluate our downscaling methodology, we perform historical WRF simulations using the European Centre for Medium Range Weather Forecasting (ECMWF) Interim (ERA-Interim) Reanalysis data set (10 years) and compare the results with ground-based stations as well as gridded observations and reanalysis products. We describe the data sets we used to drive WRF in section 2.2, and all WRF simulations we performed in section 2.4.

2.2. Data Sets Used to Drive WRF

We use both reanalysis (ERA-Interim) and CESM to drive WRF simulations. Similar to Prein et al. (2017), the aim of ERA-Interim-driven WRF simulations is to help assess the ability of our downscaling methodology in simulating observed fields during the historically observed time period.

2.2.1. The ERA-Interim Reanalysis

We use the ERA-Interim Reanalysis to perform historical WRF simulations to evaluate our model downscaling methodology. ERA-Interim is one of the best reanalyses of its generation (e.g., Lorenz & Kunstmann, 2012) and has been extensively used for downscaling with WRF (e.g., Bieniek et al., 2015; Soares et al., 2012). A six hourly ERA-Interim data set (Dee et al., 2011) to run WRF simulations is available from the Research Data Archive (RDA) at NCAR (data set 627.2; European Centre for Medium-Range Weather Forecasts, 2012).

2.2.2. CESM

CESM is a coupled Earth system model that has atmospheric, land, ocean, and ice components. We use the bias-corrected CESM v1.0 projections (~1° horizontal resolution) under a high impact emissions scenario

(RCP 8.5) available at NCAR's Computational Information Systems Laboratory (CISL) data archive (Monaghan et al., 2014). The data set is intended to support regional climate modeling studies and has been bias corrected (Bruyère et al., 2014, 2015) using ERA-Interim (Dee et al., 2011). Bias correction has been applied to correct the mean bias in 3-D fields of geopotential height, wind, temperature, humidity, and surface fields of surface pressure, sea level pressure, sea surface temperatures (SSTs), skin temperature, soil temperature, and moisture (Bruyère et al., 2014). We chose this data set because at the time, it was the only ESM data set that provided 6 hourly input fields needed for WRF simulations. The non-bias-corrected original CESM simulation is ensemble number 6 of the CESM runs, the Mother of All Runs (MOAR; b40.rcp8_5.1deg.007 (RCP8.5)). These non-bias-corrected CESM fields are also available at the RDA at NCAR (data set 316.0; Research Data Archive, 2011) and from the Earth System Grid.

2.2.3. Implementation of Driver Data in WRF

The initial and lateral boundary conditions come from the driver data (ERA-Interim or CESM); lateral boundary conditions are fed at 6 hourly intervals. Our high-resolution WRF domain includes the Great Lakes, which are a significant source of moisture and an unresolved geographical feature in the coarse resolution driver data. The default option to treat lakes in WRF is to use SSTs from the driver data and to interpolate them into the lakes, which has a tendency to lead to spurious hot or cold lake temperatures (e.g., Mallard et al., 2015). To avoid this problem, following Bruyère et al. (2015) for each of the simulated dates, we use monthly mean skin temperature from the driver model and feed that into the WRF model over the grid points containing lakes. This is now a commonly applied method in studies of climate simulations using WRF. Hence, in our simulations, both SSTs and lake temperatures change monthly, are obtained from the driving model, and fed into WRF simulations at lateral boundaries at 6 hourly intervals.

2.3. Data Sets Used for Model Evaluation

We use several data products including station observations, gridded observations, and renalyses to evaluate our model performance:

2.3.1. Modern-Era Retrospective Analysis for Research Applications, Version 2

Modern-Era Retrospective Analysis for Research Applications, version 2 (MERRA-2) is a reanalysis product we use to compare against ERA-Interim Reanalysis and downscaled ERA-Interim results. MERRA was developed by the National Aeronautics and Space Administration (NASA) using the Earth Observing System Data Assimilation System Version 5 (GEOS-5; Rienecker et al., 2011). MERRA-2 is an improvement over MERRA due to the more advanced assimilation of data, such as incorporation of modern hyperspectral radiance and microwave observations, along with Global Positioning System (GPS)-Radio Occultation data sets (Gelaro et al., 2017). We use the daily mean, maximum, and minimum 2-m air temperature products (M2SDNXSLV) and monthly total surface precipitation (M2TMNXFLX) product, versions 5.12.4. The resolution of MERRA-2 products is 0.5° × 0.625°. MERRA-2 data sets are available for download from NASA at https://giovanni.gsfc.nasa.gov/giovanni/.

2.3.2. North America Land Data Assimilation System

North America Land Data Assimilation System (NLDAS) is a data set consisting of a combination of best available observations and model output to aid modeling practices (Mitchell et al., 2004; Xia et al., 2012). We use the hourly maximum, minimum, and mean temperature at 2 m and total precipitation products (NLDAS_FORAH0125) at 0.125° horizontal resolution, which are the forcing data sets derived for the NLDAS land-surface model simulations. NLDAS data sets are available for download from NASA at https://giovanni.gsfc.nasa.gov/giovanni/.

2.3.3. Global Precipitation Climatology Project

The Global Precipitation Climatology Project (GPCP) provides a global precipitation analysis combining various satellite retrieval products, surface gauge observations, and sounding observations (Adler et al., 2003; Huffman et al., 1997; Huffman et al., 2001; Huffman et al., 2009). We use the version 2.3 combined monthly precipitation product at 2.5° resolution. GPCP precipitation data are provided by National Oceanic and Atmospheric Administration/Oceanic and Atmospheric Research/Earth System Research Laboratory Physical Sciences Division (NOAA/OAR/ESRL PSD), Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/.

2.3.4. Global Precipitation Climatology Center

The Global Precipitation Climatology Center (GPCC) provides full data reanalysis of global precipitation from 75,000 stations worldwide. For validation of our model output, we use the 1° resolution version 7

Table 2
List of Ground-Based Stations Used in This Study With Their Station ID, Location, and Altitude

Station name	Station ID	Latitude, longitude	Altitude
Boston, MA (Boston Logan International Airport)	USW00014739	42.3606°, −71.0097°	3.7 m
Hyannis, MA (Barnstable Municipal Airport)	USW00094720	41.66861°, −70.28°	16.8 m
Mount Washington, NH	USW00014755	44.2669°, −71.2997°	1911.4 m
JFK, NY (JFK International Airport)	USW00094789	40.6386°, −73.7622°	3.4 m
Hartford, CT (Bradley International Airport)	USW00014740	41.9375°, −72.6819°	53.3 m
Kennebunkport, ME	USC00174193	43.3605°, −70.4697°	6.1 m
Portland, ME (Jetport International Airport)	USW00014764	43.64222°, −70.30444°	13.7 m

monthly total precipitation product, because it is the most accurate GPCC in situ precipitation reanalysis product and it covers the entire timespan of our historical simulations (Schneider et al., 2015). GPCC Precipitation data are provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA (available at https://www.esrl.noaa.gov/psd/).

2.3.5. Station Observations

We use ground-based station observations from seven stations to compare with our model results and evaluate our downscaling methodology. The stations we chose are representative of different geographical features and areas of interest in the northeastern United States. Table 2 lists the station ID, geographical location, and altitude of each station used in this study. The station data are available from the online portal of the National Climatic Data Center (NCDC; https://www.ncdc.noaa.gov/cdo-web/).

2.3.6. ERA-Interim

Aside from the use of ERA-Interim as the driver of our WRF simulations, we also use 6-hourly ERA-Interim maximum, minimum, and mean temperatures at 2 m and total precipitation to evaluate our downscaling methodology (NCAR ds 627.2, available at: https://rda.ucar.edu/datasets/ds627.2/).

The reasons for including a variety of products in our analysis are as follows: (1) Ground-based observational network is too sparse and most stations do not have complete record of the historical period. (2) Available gridded observational and reanalysis products are too coarse in horizontal resolution compared to our downscaled product. (3) Available gridded observational and reanalysis products have different horizontal resolutions, including different types of observations, and have inherent uncertainties arising from retrieval methods, interpolation, and/or data assimilation techniques employed while creating them (e.g., Hartmann et al., 2013). For example, GPCC includes precipitation from global rain gauge stations only, GPCP, while of coarser resolution compared to GPCC, combines satellite and gauge observations of precipitation. NLDAS, on the other hand, includes Climate Prediction Center (CPC) gauge observations, stage II Dopler radar data, satellite microwave observations obtained through the CPC morphing method (CMORPH; Joyce et al., 2004), and NCEP North American Regional Reanalysis (NARR) data (Mesinger et al., 2006). Furthermore, reanalysis products do not necessarily assimilate precipitation. ERA-Interim forecasts precipitation given humidity and temperature from assimilated observations (Dee et al., 2011), while MERRA-II assimilates precipitation observations into model simulated precipitation (e.g., Reichle et al., 2017). Therefore, using different gridded observational and reanalysis products, which are of different horizontal resolution and obtained through different methods, is important to frame the variability in our downscaled fields in the historical time period. A detailed comparison of different observational and reanalysis products for precipitation is summarized in Sun et al. (2018).

2.4. Summary of Simulations

For model evaluation, similar to Prein et al. (2017), we perform 10 years (2006 to 2015) of ERA-Interim-driven WRF simulations using the model setup explained in section 2.1. We will refer to these simulations as WRF-ERA. The results of these simulations will be used to assess the ability of our downscaling method to replicate observed fields and to determine whether our downscaled fields are better in line with observed compared to driver ERA-Interim. Next, using the same WRF model setup, we perform dynamical downscaling of CESM projections over three time slices representative of present day (PD; 2006–2020), midcentury (MC; 2041–2060), and end of century (EC; 2081–2100) climate conditions, which will be referred to from now on as WRF-PD, WRF-MC, and WRF-EC, respectively. For ease of comparison with

Table 3
List of All Simulations Performed in This Study and Their Short Names

Simulation	Description	Simulation period
WRF-ERA	ERA-Interim Driven WRF Simulations	2006–2015
WRF-HIST	CESM RCP 8.5 driven WRF Historical Simulations (A subset of WRF-PD)	2006–2015
WRF-PD	CESM RCP 8.5 driven WRF Present Day Simulations	2006–2020
WRF-MC	CESM RCP 8.5 driven WRF Mid-Century (MC) Simulations	2041–2060
WRF-EC	CESM RCP 8.5 driven WRF End of Century (EC) Simulations	2081–2100

WRF-ERA, a subset of WRF-PD simulations (2006–2015) will be named CESM-driven historical simulations (WRF-Hist). Table 3 lists all WRF simulations performed and their short names that we will refer to for the remainder of this manuscript.

It is important to note that the driving CESM model was initialized in 1950 without additional data assimilation; as a result, the PD in the model is not expected to exactly replicate the real-world PD climate. Hence, when analyzing future changes in section 4, we will use the PD simulations as a baseline. For the same reasons, resulting fields from WRF-Hist simulations are also not expected to replicate historical observations. When presenting results in section 3, we will include WRF-Hist and CESM-HIST to show the influence of our downscaling on historical CESM fields and to show how fields from WRF-Hist compare with historical gridded observations and reanalyses.

3. Evaluation of the Downscaling Methodology

To evaluate our model setup and downscaling methodology, we perform ERA-Interim-driven historical WRF runs (WRF-ERA) for 10 years (2006–2015). We first compare results from WRF-ERA and driver data ERA-Interim against each other and with gridded observations and reanalysis products. This analysis is (1) to show that our downscaling of ERA-Interim produces comparable results to available observational products and (2) to highlight the improvement in downscaled fields (WRF-ERA) compared to the driver data ERA-Interim. Similarly, we also compare CESM-driven historical WRF simulations with driver CESM data and compare results with reanalysis and gridded observational products. This analysis is done to highlight how the historical CESM- and WRF-simulated fields compare to those observed in the historical period. Because bias-corrected CESM fields for 2-m temperatures and precipitation are not available, we use raw CESM fields to compare with bias-corrected and WRF-downscaled CESM output in the following analysis. Next, to further illustrate that our downscaling methodology works successfully, we compare results from WRF-ERA and the driver data from ERA-Interim with data obtained from ground-based station observations and NLDAS. We will focus on the simulation of mean and extreme fields, respectively.

3.1. Simulation of Historical Mean Fields

To assess our downscaling methodology, we first compare simulated mean fields from WRF-ERA and the driver ERA-Interim with historical gridded observations and other reanalysis products. We use several gridded observational products and reanalysis with different horizontal resolution in our comparison to showcase the similarities and differences among them.

General Circulation Models (GCMs) tend to simulate temperatures for the historical time period well (Flato et al., 2013); thus, we expect this to be the case with our more detailed high-resolution regional modeling methodology. We present annual mean, maximum, and minimum temperatures at 2 m averaged over 2006 to 2015 from ERA-Interim, WRF-ERA, CESM, and WRF-Hist in Figure 2 along with MERRA-2 and NLDAS for comparison.

WRF-ERA- and ERA-Interim-simulated temperatures at 2 m share similar spatial distribution and magnitude with increased granularity in WRF-ERA simulations due to increased resolution. Compared to MERRA-2 and NLDAS, mean temperatures at 2 m are well represented in both WRF-ERA and WRF-Hist simulations. This result is not surprising because temperatures from the driver CESM for WRF-Hist simulations are bias corrected using ERA-Interim. CESM tends to overestimate the magnitudes of mean and maximum temperatures at 2 m compared to downscaled CESM, downscaled ERA-Interim, and the reanalysis products. Minimum temperatures, however, are coldest in downscaled simulations (WRF-Hist and WRF-ERA). Colder minimum temperatures simulated from downscaled simulations (WRF-ERA and WRF-Hist) are still in line with minimum temperatures from MERRA-2 (Figure 2). The spatial distribution and magnitudes of mean and maximum temperatures at 2 m from downscaled simulations (WRF-Hist and WRF-ERA) are closer to NLDAS compared to their driver data (Figure 2). Compared to CESM, bias-corrected and downscaled CESM (WRF-Hist) yields 2-m temperatures more in line with reanalysis (Figure 2). In the case of CESM-driven simulations it is difficult to assess the contribution of downscaling over bias correction for improved temperatures at 2 m, due to

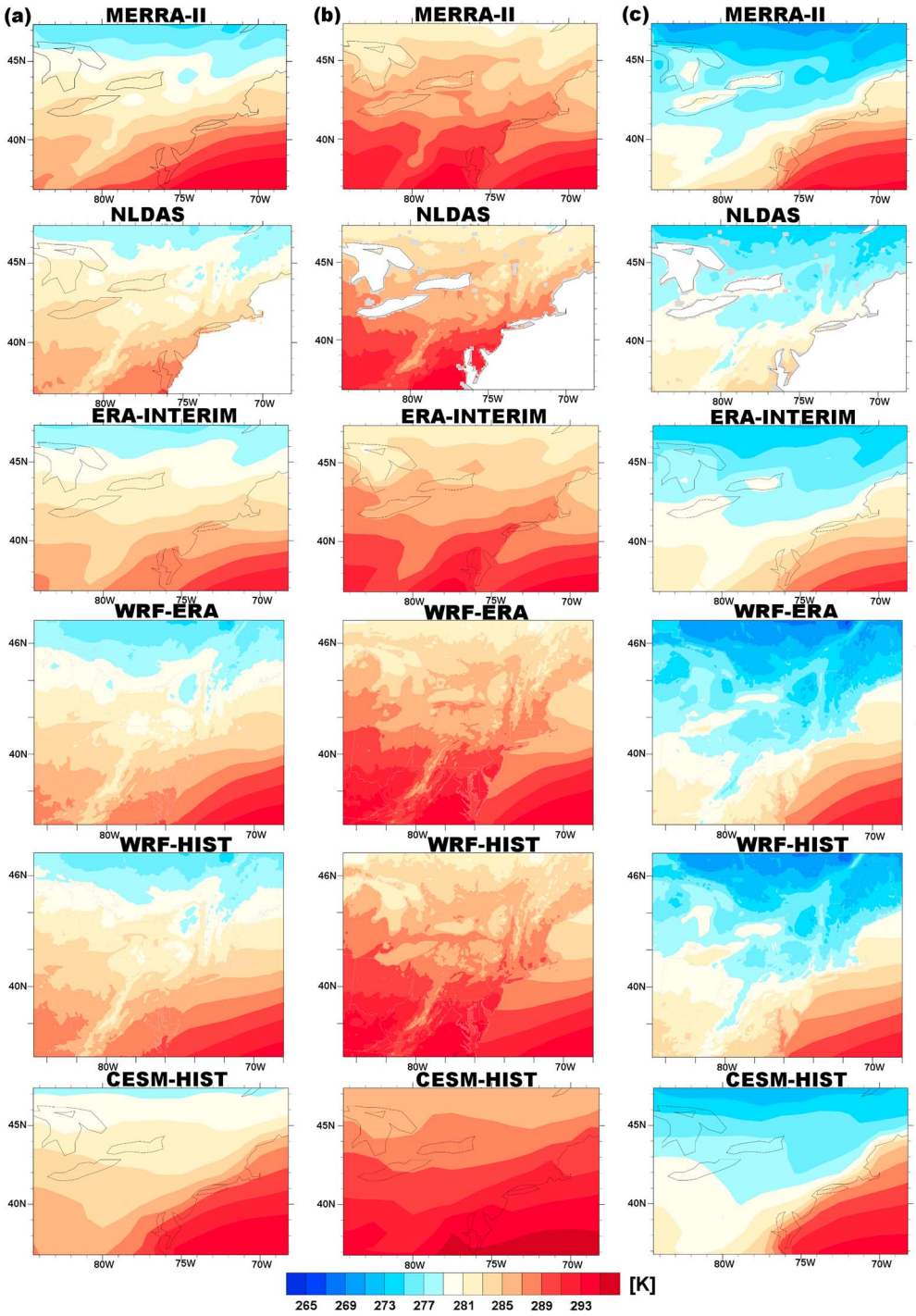

Figure 2. (a) Mean, (b) maximum, and (c) minimum temperatures at 2 m (K) from MERRA-2; NLDAS, ERA-Interim, WRF-ERA, WRF-HIST, and CESM-HIST.

the bias correction applied to CESM-simulated temperatures before downscaling. Nonetheless, the downscaled WRF simulations (WRF-ERA and WRF-Hist) are able to reproduce historical mean, maximum, and minimum of temperatures at 2 m with some improvement compared to their driver data (Figure 2).

To further compare WRF-ERA with historical fields, we present scatter plots of annual mean, maximum, and minimum daily temperatures at 2 m between NLDAS and WRF-ERA in Figures 3a–3c. Linear

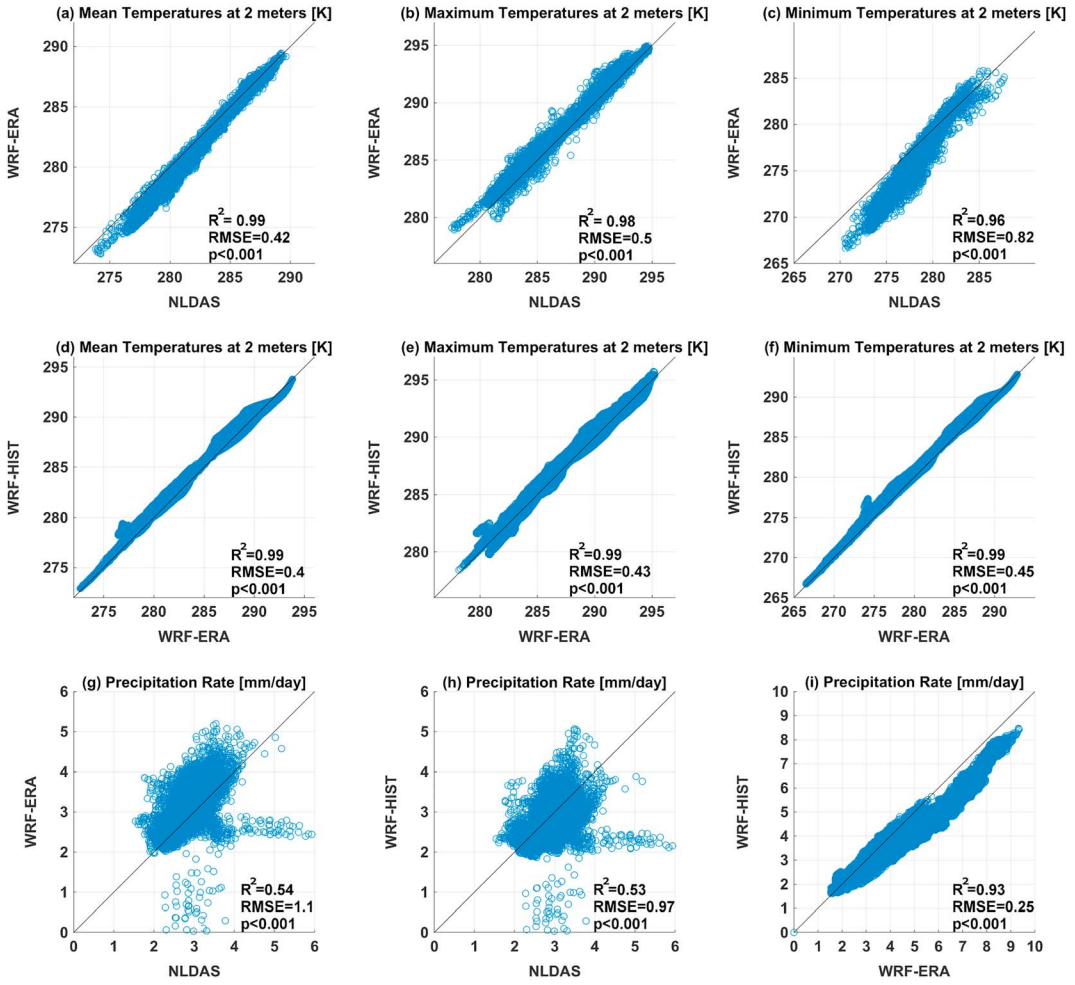

Figure 3. Scatter plots of annual mean, maximum and minimum temperatures (K) between (a–c) NLDAS and WRF-ERA and (d–f) WRF-ERA and WRF-HIST and annual mean precipitation rates (mm/day) between (g) WRF-ERA-NLDAS, (h) WRF-HIST-NLDAS, and (i) WRF-HIST-WRF-ERA along with linear regression coefficient (R^2), root-mean-square error (RMSE), and *p* values.

regression coefficients and the root-mean-square errors are also provided in Figure 3. We use NLDAS in this analysis because NLDAS data set we use incorporates best available surface and radar observations with NARR reanalysis and it is the highest resolution gridded data set in our comparison. Mean, maximum, and minimum temperatures at 2 m from WRF-ERA and NLDAS are highly correlated (Figure 3). Furthermore, to show the influence of boundary conditions in our WRF simulations, we also present scatter plots of mean, maximum, and minimum temperatures of WRF-ERA versus WRF-Hist in Figures 3d–3f, respectively. We find that mean, maximum, and minimum temperatures at 2 m from WRF-ERA and WRF-Hist are highly correlated (Figure 3). This finding is not surprising due to the bias correction (using ERA-Interim fields) applied to CESM temperature fields before downscaling. Nevertheless, these results suggest that our downscaling method is able to produce results in line with historical observations.

Precipitation is more difficult to simulate and evaluate compared to temperatures. Part of this difficulty arises from the lack of a comprehensive observational network to help improve model parameterizations and from the uncertainties in retrieval methodologies, assimilation, and interpolation methods (e.g., Flato et al., 2013; Hartman et al., 2013; Sun et al., 2018).

In Figure 4, we present annual mean daily precipitation rates from our simulations, driver data of our simulations (ERA-Interim, CESM, WRF-ERA, and WRF-Hist), two gridded observational products (GPCP and GPCC),

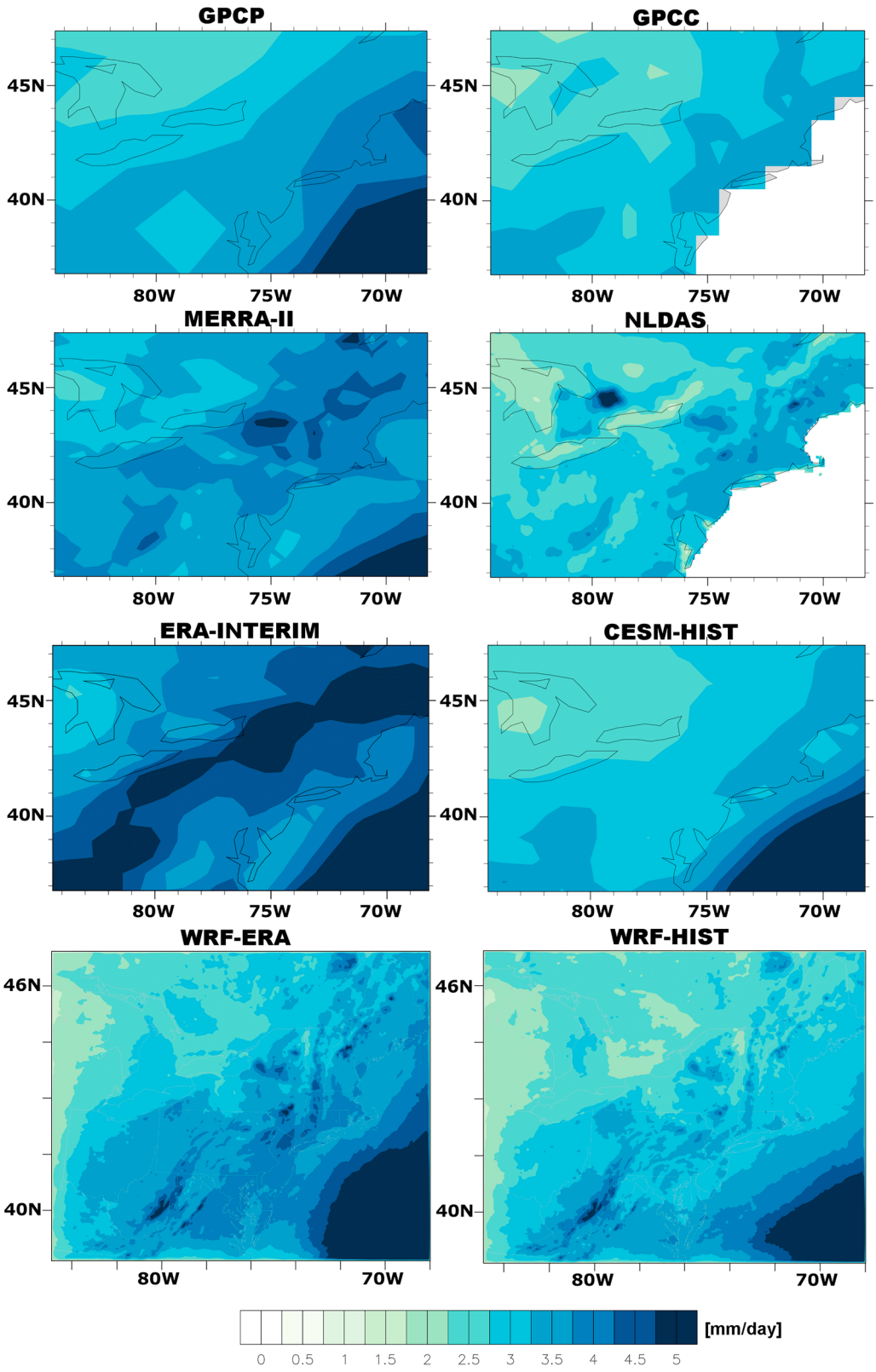

Figure 4. Annual mean daily accumulated precipitation rate (mm/day) averaged over 2006 to 2015 from GPCP, GPCC, MERRA-2, NLDAS, ERA-Interim, CESM-Hist, WRF-ERA, and WRF-Hist.

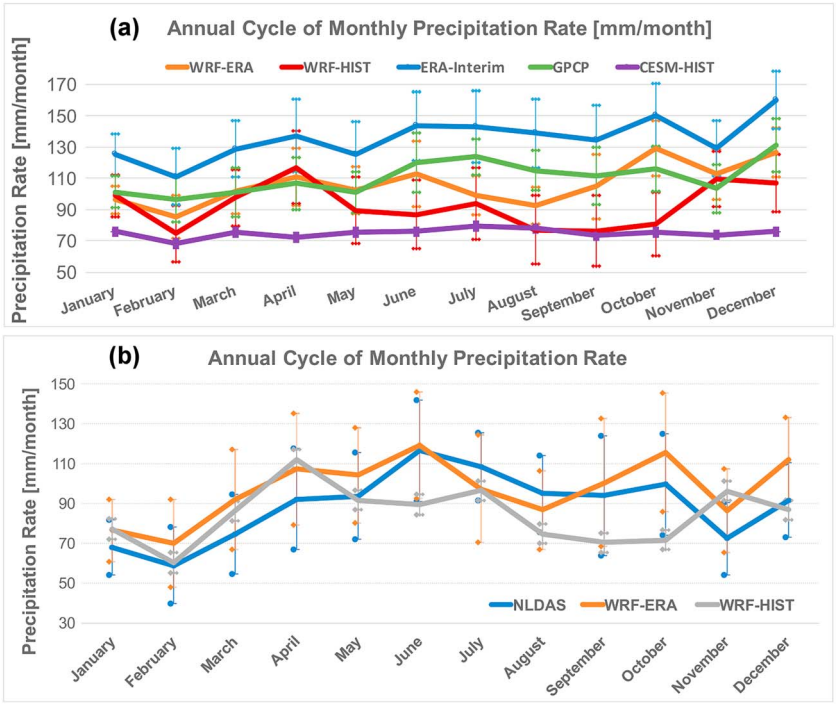

Figure 5. Annual cycles of monthly accumulated precipitation rate (mm/month) (a) averaged over the highest resolution domain (domain 3) from 2006 to 2015 from WRF-ERA (orange line), WRF-Hist (red line), ERA-Interim (blue line), CESM-Hist (purple line), and GPCP (green line) and (b) averaged over land points only from WRF-ERA, WRF-Hist, and NLDAS.

and two reanalysis products (MERRA-2 and NLDAS). WRF-ERA is able to capture some of the spatial detail and locations of maximum precipitation as observed from NLDAS and GPCC with some overestimation over the southeastern part of the domain. For all data sets, the overall structure and magnitude of the annual mean precipitation rates are similar overland. However, ERA-Interim tends to produce an unrealistic band of precipitation that is not reflected in gridded observations, reanalysis products or WRF simulations. Hence, precipitation rates from WRF-ERA are better in line with available observations and reanalysis products compared to ERA-Interim. Similarly, comparing results from CESM and WRF-Hist with NLDAS, spatial distribution and locations of increased precipitation are generally well represented in WRF-Hist. WRF-Hist underestimates the precipitation rates over coastal NY, CT, and MA compared to NLDAS and CESM, which can capture some of the increased precipitation rates. WRF-ERA on the other hand is able to simulate the coastal precipitation and regions of increase in precipitation closer to NLDAS. Differences in spatial distribution and magnitude of precipitation between WRF-ERA and WRF-Hist highlight the influence of the boundary conditions (Figure 4).

We find that annual mean daily precipitation rates from WRF-ERA are moderately correlated with NLDAS (Figures 3g). Similarly, daily precipitation rates from NLDAS and WRF-Hist are also moderately correlated (Figure 3h). To show the influence of boundary conditions in our WRF simulations, we present a scatter plot of WRF-ERA versus WRF-Hist in Figure 3i. We find that precipitation rates from WRF-ERA and WRF-Hist are highly correlated with WRF-Hist, while the magnitudes of precipitation obtained from WRF-Hist are generally less than WRF-ERA. This finding is somewhat expected due to the bias correction that was applied to CESM fields prior to downscaling with WRF.

Annual cycles of monthly precipitation rates averaged over the highest resolution domain (including both land and ocean points) from models and gridded observations from GPCP are presented in Figure 5a along with 95% confidence intervals. Here we are using GPCP because it is the coarsest resolution observational data set we have (comparable to coarse output from CESM) that includes ocean points. Interannual variability for both downscaled WRF simulations, GPCP, and ERA-Interim are large, partly because our comparison is limited to only 10 years of data due to the expensive nature of our simulations. Comparing

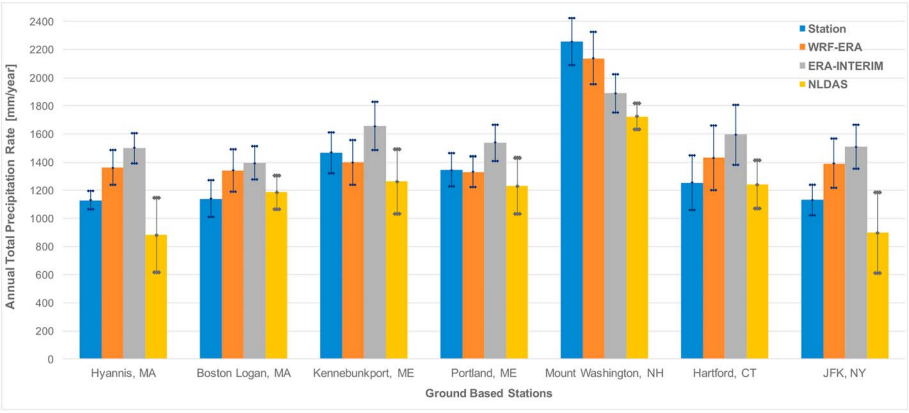

Figure 6. Annual total precipitation rate (mm/year) from seven ground-based stations (blue), WRF-ERA (orange), ERA-Interim (gray), and NLDAS (yellow) for Hyannis, MA; Boston, MA; Kennebunkport, ME; Portland, ME; Mount Washington, NH; Hartford, CT; and JFK, NY, shown with 95% confidence intervals.

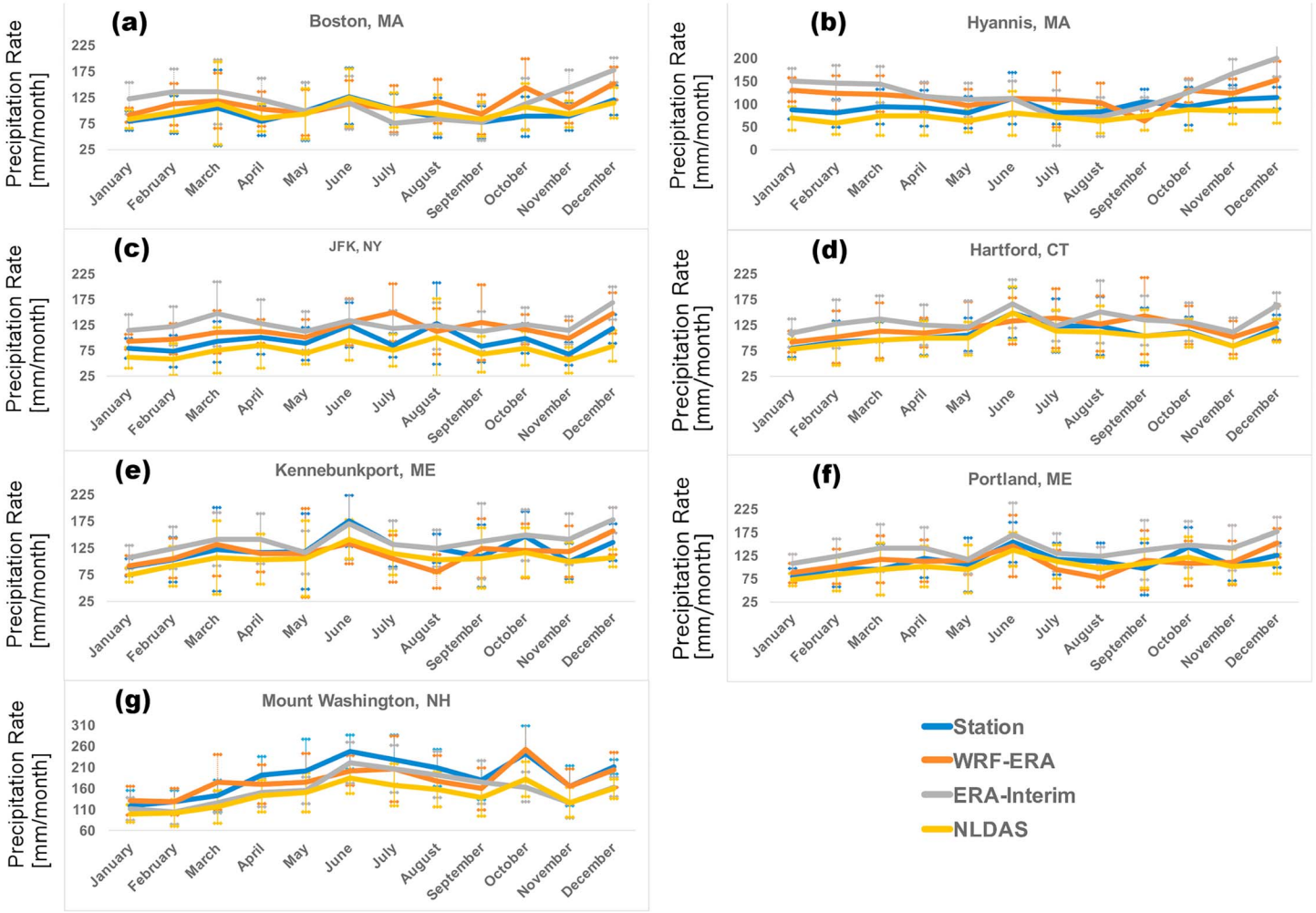

Figure 7. Annual cycles of monthly accumulated precipitation rate (mm/month) at seven station locations as observed from stations (blue) and simulated from WRF-ERA (orange), ERA-Interim (gray), and NLDAS (yellow): (a) Boston, MA; (b) Hyannis, MA; (c) JFK, NY; (d) Hartford, CT; (e) Kennebunkport, ME; (f) Portland, ME; and (g) Mount Washington, NH.

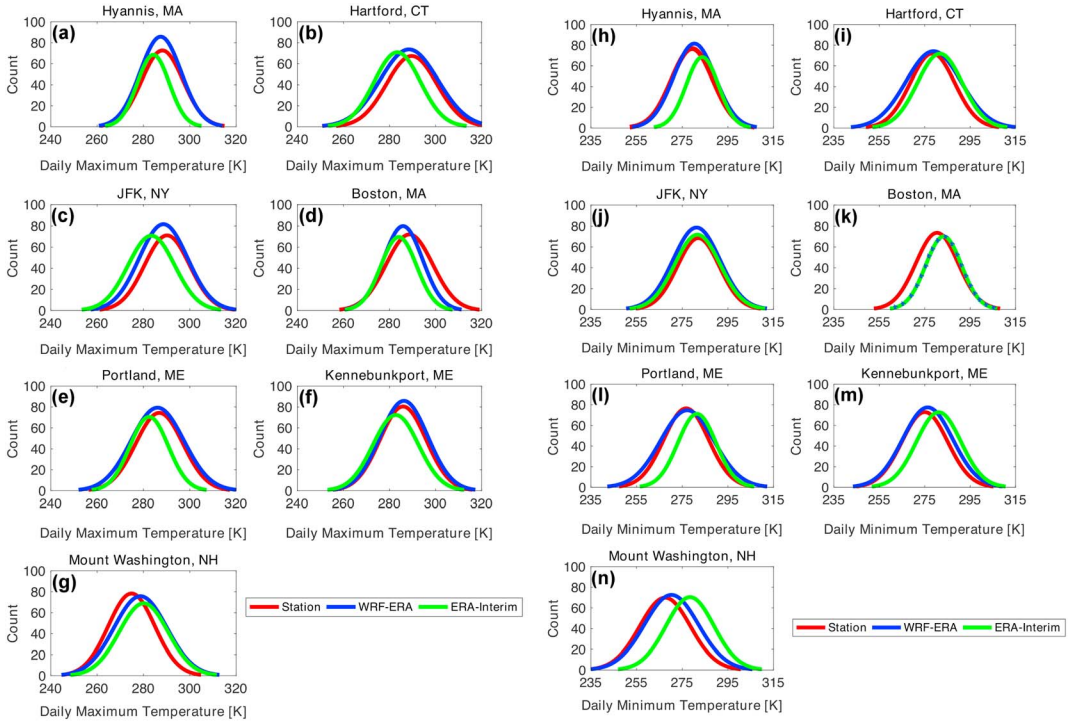

Figure 8. Probability density distributions of daily maximum temperatures at 2 meters (K) as observed from stations and simulated from WRF-ERA and ERA-Interim at seven stations: (a) Hyannis, MA; (b) Hartford, CT; (c) JFK, NY; (d) Boston, MA; (e) Portland, ME; Kennebunkport, ME; and (g) Mount Washington, NH. Probability density distributions of daily minimum temperatures at 2 meters (K) as observed from stations and simulated from WRF-ERA and ERA-Interim at seven stations: (h) Hyannis, MA; (i) Hartford, CT; (j) JFK, NY; (k) Boston, MA; (l) Portland, ME; (m) Kennebunkport, ME; and (n) Mount Washington, NH.

ERA-Interim and WRF-ERA with GPCP, our downscaling methodology can produce results closer to observed (Figure 5a). Precipitation rates obtained from WRF-Hist are increased compared to CESM and are closer to observations from GPCP. However, WRF-Hist tends to underestimate precipitation rates compared to GPCP. On the other hand, while they share similar spatial distributions of precipitation, GPCP, GPCC, and NLDAS do not agree (Figure 4) in terms of magnitudes of precipitation and locations of maxima, suggesting observational uncertainty as well as uncertainties in methodologies used for creating of these data sets. In Figure 5b, we compare annual cycle of precipitation averaged over land points only from WRF-ERA, WRF-Hist, and NLDAS presented with 95% confidence intervals. We choose NLDAS in our analysis here due to its higher horizontal resolution compared to other products (closest in horizontal resolution to WRF-ERA and WRF-Hist simulations). While interannual variability is large (partly because we are only comparing a 10 points time series), WRF-ERA is generally able to replicate annual cycle of precipitation close to the highest resolution reanalysis (NLDAS; with some overestimation over all months except for July and August where it underestimates). Even though magnitudes of precipitation rates are not well replicated and predicted trend in precipitation is reversed (decrease as opposed to increase) in May and June and (increase as opposed to decrease) in November, month-to-month trend of annual cycle in WRF-Hist closely follows NLDAS. It is important to remember that precipitation fields in WRF-Hist are not bias corrected and driver ESM run is not nudged to historical observations. Hence, neither is expected to match historical observations. The reason we are comparing WRF-Hist and CESM with gridded observational products is to show how the climate of the historical time period as simulated in WRF and CESM compares with the available observations and reanalysis.

To further investigate the skill of our downscaling methodology in simulating precipitation, we compare results from WRF-ERA and ERA-Interim with NLDAS and station observations from seven select ground-based stations. These stations, while not covering our entire domain, have complete records for the time period of

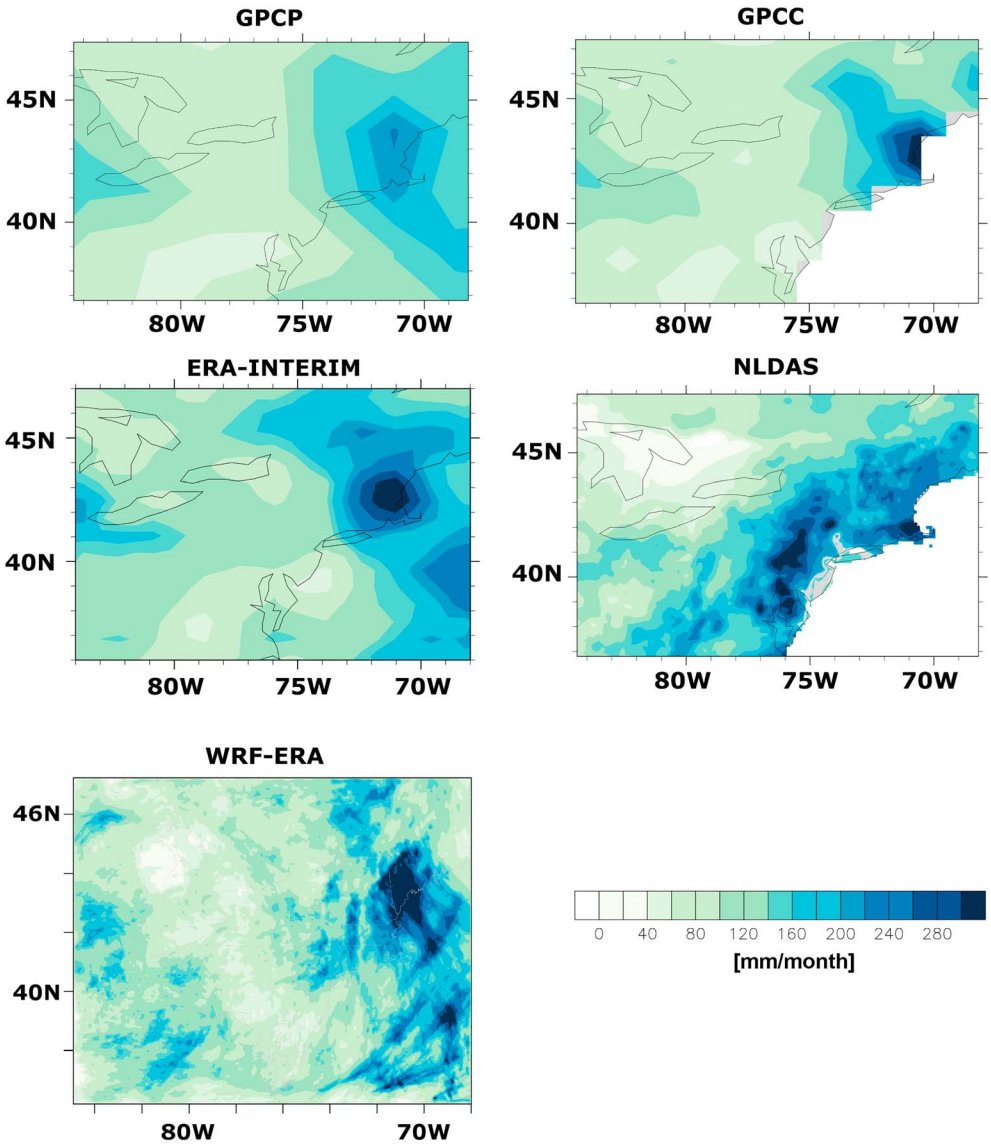

Figure 9. Monthly accumulated precipitation rate (mm/month) for May 2006 from GPCP, GPCC, ERA-Interim, NLDAS, and WRF-ERA.

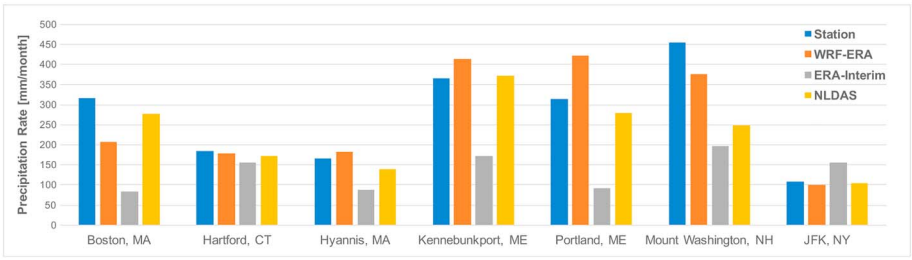

Figure 10. May 2006 monthly total precipitation (mm/month) rates from ground-based stations (blue), WRF-ERA (orange), ERA-Interim (gray), and NLDAS (yellow) at seven station locations: Boston, MA; Hartford, CT; Hyannis, MA; Kennebunkport, ME; Portland, ME; Mount Washington, NH; and JFK, NY.

Table 4
Mean, Standard Deviation, 90th Percentile of Daily Precipitation Rates From Station Data, WRF-ERA, and ERA-Interim, and Correlation Coefficients of WRF-ERA Time Series With Station Observations and ERA-Interim With Station Observations for All Stations

Station	Mean	Standard deviation	90th percentile	Correlation with station
Hyannis, MA				
Station	3.1 ± 0.3	8.2	10.1	
WRF-ERA	3.7 ± 0.3	9.4	12.1	0.35
ERA-Interim	4.1 ± 0.3	10.2	13.1	0.006
Boston, MA				
Station	3.1 ± 0.3	8.4	10.3	
WRF-ERA	3.7 ± 0.3	9.9	11.1	0.49
ERA-Interim	3.8 ± 0.3	9.6	12	0.007
New York, NY				
Station	3.1 ± 0.3	9.1	10.2	
WRF-ERA	3.8 ± 0.3	10.3	12.5	0.43
ERA-Interim	4.1 ± 0.3	10	13.4	0.01
Hartford, CT				
Station	3.5 ± 0.3	9.1	11.7	
WRF-ERA	3.9 ± 0.3	10.6	12.6	0.52
ERA-Interim	4.4 ± 0.3	9.7	14	0.02
Kennebunkport, ME				
Station	4. ± 0.4	10.6	11.2	
WRF-ERA	3.8 ± 0.3	10.2	11.5	0.44
ERA-Interim	4.5 ± 0.3	10.5	14	−0.015
Portland, ME				
Station	3.7 ± 0.4	11	12.7	
WRF-ERA	3.7 ± 0.3	9.8	11	0.59
ERA-Interim	4.2 ± 0.3	10.4	13.2	−0.013
Mount Washington, NH				
Station	6.2 ± 0.4	11.7	18.5	
WRF-ERA	5.9 ± 0.4	12.9	17.3	0.55
ERA-Interim	5.2 ± 0.3	9.5	15.7	−0.02

interest and are representative of both low and high topography areas as well as unique coastal features. When extracting results from the models at stations, the grid point location closest to the station is chosen. Figure 6 shows the annual total precipitation averaged over the historical (2006–2015) time period at each station and from WRF-ERA, ERA-Interim, and NLDAS. For all stations, WRF-ERA produces results more in line with station observations compared to its driver reanalysis ERA-Interim (Figure 6). Annual precipitation rates from NLDAS are smaller compared to WRF-ERA and ERA-Interim and are usually underpredicted compared to station observations except for Boston Logan, MA. For some stations WRF-ERA is a better match to observed total precipitation rates, while for others NLDAS is a better match. It is, however, important to highlight that WRF-ERA produces the closest precipitation rates to observed at a region of high topography, Mount Washington compared to both reanalyses (Figure 6).

Figure 7 shows annual cycles of monthly precipitation rates averaged over the historical (2006–2015) time period from WRF-ERA, ERA-Interim, and NLDAS for all stations along with station observations and 95% confidence intervals. WRF-ERA is able to capture the observed month-to-month trend in the annual cycle of precipitation (Figure 7). While both WRF-ERA and ERA-Interim produce precipitation rates and annual cycle in line with observations, simulated precipitation rates from WRF-ERA are in general closer to observed precipitation rates at all stations. Similarly, NLDAS and WRF-ERA produce similar annual cycles of precipitation rates that are comparable to observations at all stations, while precipitation rates from NLDAS are underpredicted in some stations (Figures 7b, 7c, 7e, and 7g). During the cold season (November to April), WRF-ERA-simulated precipitation rates are improved compared to ERA-Interim for all stations. ERA-Interim tends to produce comparable precipitation rates to station observations in May and June for some stations, but WRF-ERA is usually a better match except for Kennebunkport, ME. Simulations of precipitation rates at Mount Washington NH, a region of complex, high topography, from WRF-ERA are a nearly perfect fit to observations during the cold months (October–March). Even though month-to-month trend in NLDAS precipitation rates closely follows those observed at stations, precipitation rates from NLDAS are underpredicted at Mount Washington, NH.

Improved cold season precipitation obtained in WRF-ERA could be suggestive of the detailed microphysics parameterization used in our WRF simulations as well as increased resolution over complex topography and is an interesting finding that warrants further studies to be conclusive. In general, our analysis suggests that WRF-ERA produces precipitation rates comparable with station observations and NLDAS and shows improvement over the driver ERA-Interim data set (and NLDAS over the region of high topography). This analysis only contains 10 years of data (due to the computational expense associated with these simulations) and interannual variability is large. A time series of at least 30 years and including additional stations would yield more robust results and provide a better characterization of inter-annual variability.

3.2. Simulation of Climate Extremes

One of the key purposes of downscaling is to assess sustainability of regional structures and determine potential impacts on the regional environment and economy in a changing climate. Compared to changes in the mean climate, the severity of such impacts is more sensitive to changes in climate extremes. While the ability of a downscaling methodology to reproduce mean observed fields and improve upon reanalysis forecasts is significant, a worthwhile downscaling methodology should also have the ability to simulate climate extremes well. We will therefore assess the ability of our downscaled results to simulate extremes. When presenting results and comparisons, we will focus on temperatures at 2 m and precipitation rates.

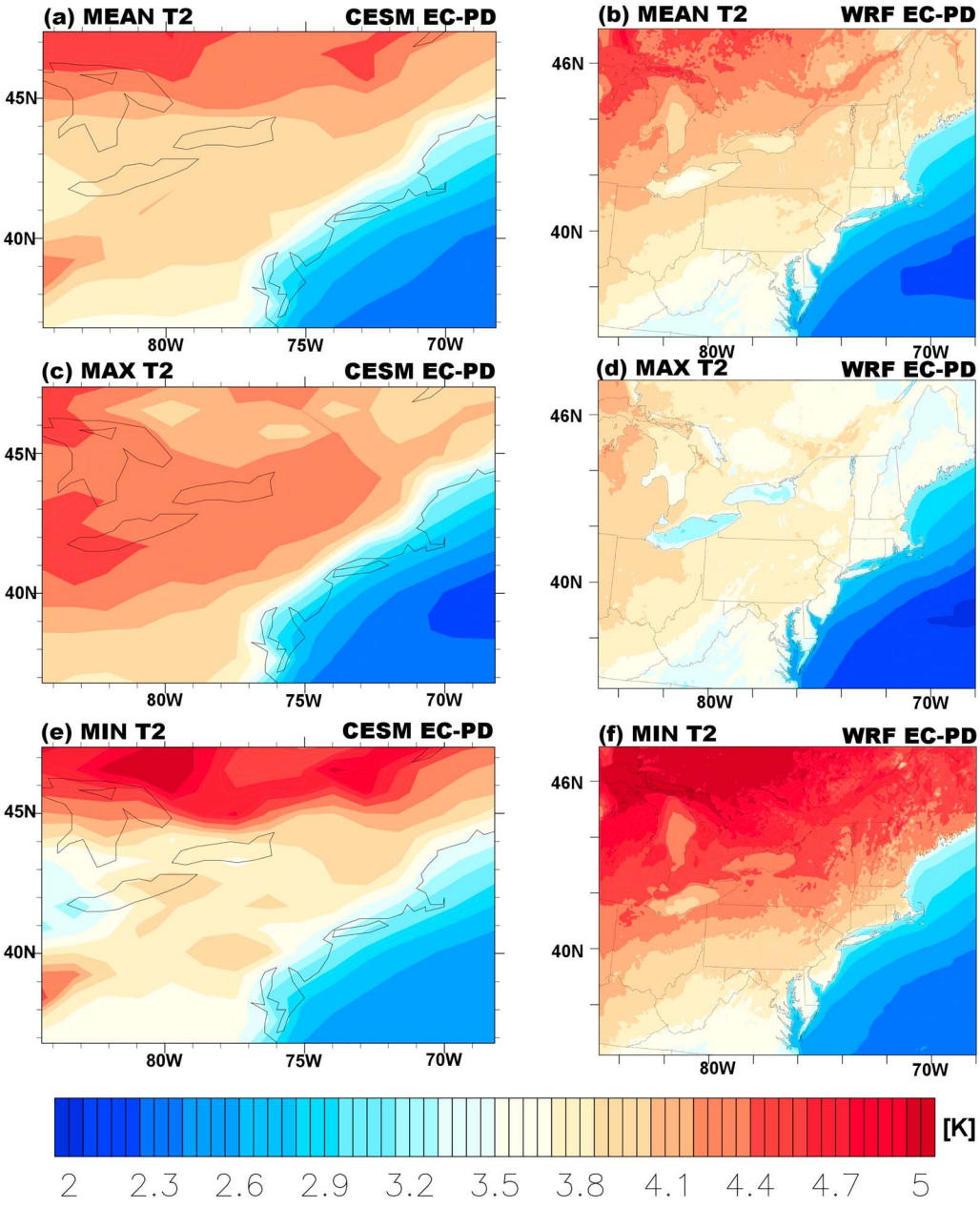

Figure 11. Difference in temperatures at 2 m (K) (a and b) means, (c and d) maximums, and (e and f) minimums between end of century and present-day CESM projections (CESM-EC–CESM-PD) (left column) and WRF projections (WRF-EC–WRF-PD) (right column).

3.2.1. Station-Level Temperature Analysis

To illustrate the ability of our downscaling methodology to simulate both climate means and extremes, we compare probability density distributions for maximum and minimum temperatures at 2 m as simulated at each station location from stations with WRF-ERA and the driver ERA-Interim data set in Figure 8. For all locations we investigated, WRF-ERA provides a better match with station observations compared to ERA-Interim Reanalysis in terms of reproducing both the means and tails of the distribution. These results illustrate the ability of our downscaling methodology to simulate both means and extremes of 2-m temperatures rather well compared to the driver data.

3.2.2. May 2006 Extreme Precipitation Event

We focus on an extreme precipitation event observed in New England in May 2006, when a Nor'easter event led to significant flooding and damage in the region. Figure 9 shows the accumulated precipitation

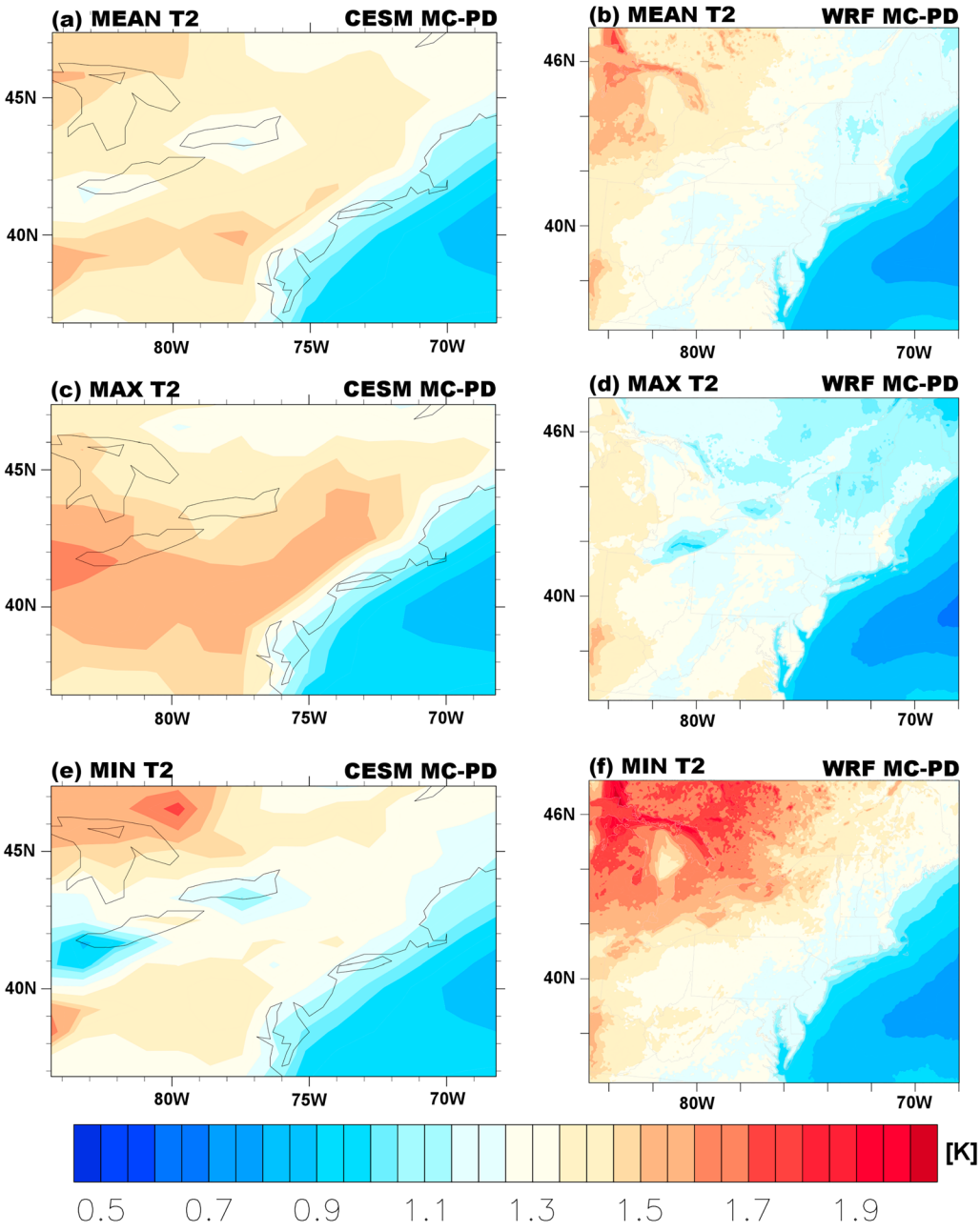

Figure 12. Difference in temperatures at 2 m (K) (a and b) means, (c and d) maximums, and (e and f) minimums between midcentury and present-day CESM projections (CESM-MC–CESM-PD) (left column) and WRF projections WRF-MC–WRF-PD) (right column).

from WRF-ERA, ERA-Interim, gridded precipitation products (GPCP and GPCC), and NLDAS. Our methodology is able to produce the location of precipitation over New England with its magnitude slightly over predicted compared to GPCC, GPCP, and NLDAS. Compared to NLDAS, WRF-ERA produces similar spatial distribution over the maxima around coastal MA and Cape Cod, while precipitation rate increase over coastal southeastern United States (the region between the coast line and the Great Lakes) is underpredicted in WRF-ERA and the other gridded observational products. Nevertheless, our downscaling methodology (WRF-ERA) is capable of reproducing the observed precipitation structure during this extreme precipitation event that led to significant economic loss and damage in the region. Furthermore, to further evaluate the ability of our downscaling method to reproduce extreme precipitation events, we compared the total precipitation observed in May 2006 from station observations with those simulated from

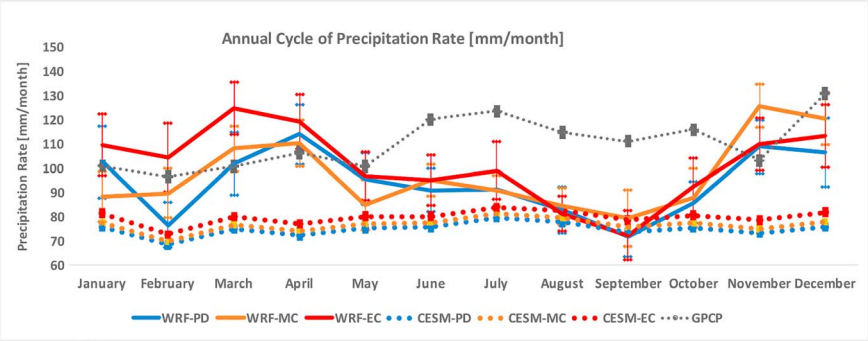

Figure 13. Annual cycle of monthly precipitation rate (mm/month) averaged over the highest resolution domain (domain 3) from WRF-PD (blue line), WRF-MC (orange line), WRF-EC (red line), CESM-PD (dotted blue line), CESM-MC (dotted orange line), CESM-EC (dotted red line), and GPCP gridded observations (dotted black line) along with 95% confidence intervals.

WRF-ERA, ERA-Interim, and NLDAS (Figure 10). WRF-ERA successfully produces results very close to those observed at stations and NLDAS, while precipitation totals from NLDAS are less than those observed at stations (except for Kennebunkport, ME). Of all three products compared, WRF-ERA produces the closest result to observed at Mount Washington, NH. Hence, though our comparison is limited to a single extreme event and limited number of stations, our results suggest that our downscaling methodology can produce results close to observations with improvement over ERA-Interim for all stations and NLDAS for some stations (Hartford, CT; Hyannis, MA, and Mount Washington, NH; Figure 10) during an extreme precipitation event.

3.2.3. Station-Level Precipitation Analysis

To further illustrate the ability of our downscaling to simulate precipitation extremes, we present the means, 95% confidence intervals, standard deviations, and 90th percentiles of time series of daily precipitation rates between 2006 and 2015 at different station locations against WRF-ERA simulations and the ERA-Interim data set in Table 4. Correlations of ERA-Interim and WRF-ERA time series with time series of station observations are also presented in Table 4. For all stations, means, deviations, and correlation with observed time series significantly improve with downscaling. Mount Washington, a region of high topography, is especially well represented in WRF-ERA compared to ERA-Interim. Simulated precipitation extremes are noticeably improved in WRF-ERA for all stations, and even for Mount Washington. This result is very important because it has been difficult for models to simulate precipitation well in high topography areas and it has been suggested that higher resolution can allow for the simulation of orographic effects and yield better precipitation (e.g., Jang et al., 2017). Similarly, for all stations, the extreme 90th percentile values are comparable to station observations in WRF-ERA. Our results suggest that our downscaling methodology can reproduce mean and extreme precipitation statistics and is closer to observations and NLDAS at all stations investigated here compared to driver ERA-Interim data set. Hence, our model setup is able to produce better results compared to reanalysis without the use of any special boundary conditions, nudging, or tuning. These results suggest that our methodology works well and therefore can be used to downscale ESM projections.

4. Comparison of Projected Future Changes Between CESM and WRF Output

In the previous sections, we revealed that our downscaling methodology can recreate historical observations of temperature and precipitation in the Northeast. In this section, we will compare simulated changes into the future between CESM projections and WRF-downscaled CESM projections.

Figure 11 (Figure 12) shows the differences in mean, maximum, and minimum temperatures at 2 m between EC and PD (MC and PD) simulations for CESM and WRF. All differences are statistically significant at 99% level. While mean temperatures increase into the future in both models, there is less pronounced overall increase in downscaled results by both the MC and EC. Compared to CESM projections, WRF provides granularity over Great Lakes, where mean temperatures increase more in WRF projections for both MC and EC. These differences in the maximum of mean temperatures between CESM and WRF projections over Great Lakes suggest the effects of lake treatments in regional climate simulations. While our 3-km horizontal resolution WRF simulations allow for the lakes in the domain to interact with and influence the

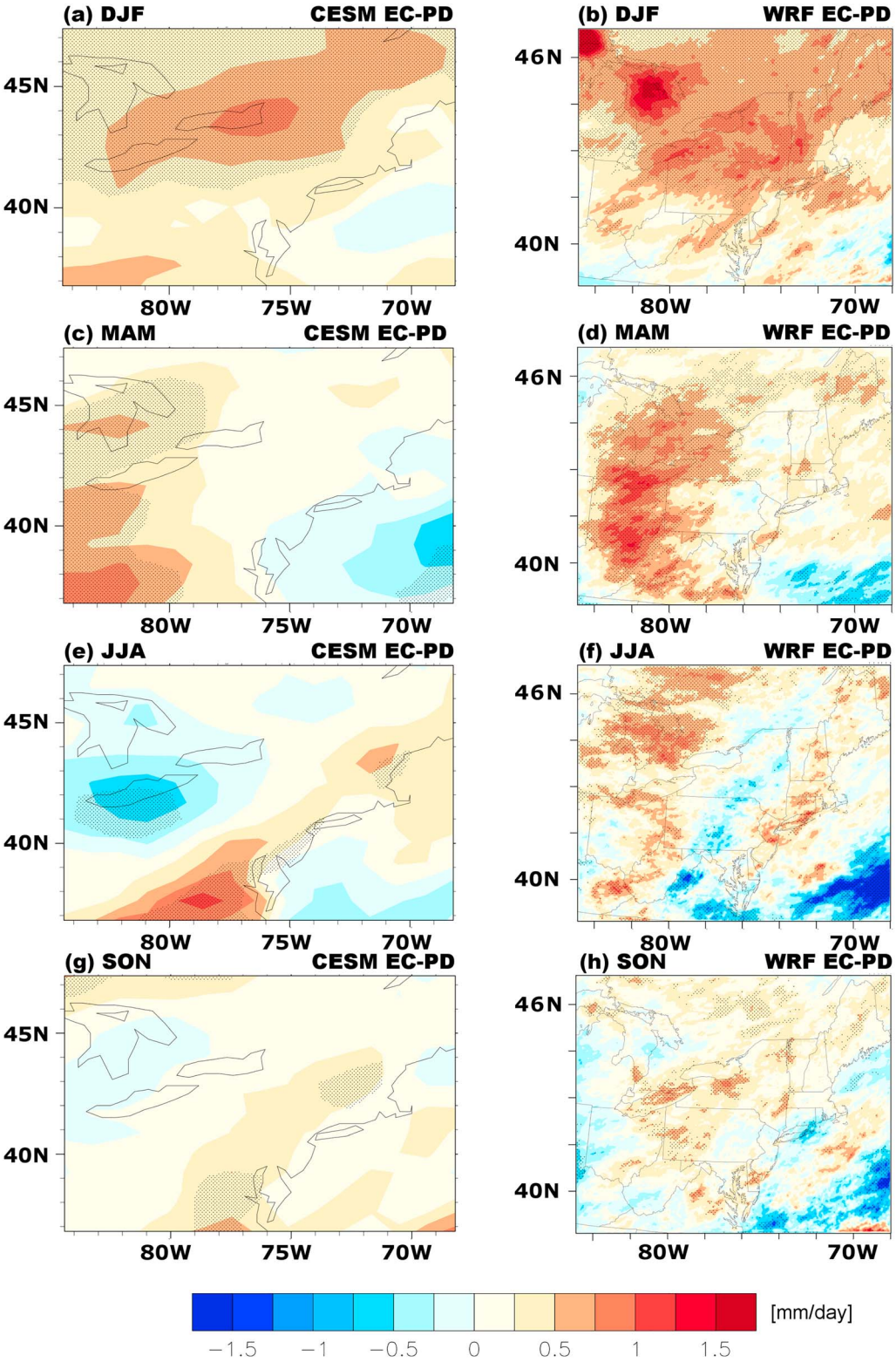

Figure 14. Seasonal differences (DJF, MAM, JJA, and SON) in daily precipitation rates [mm/day] between end of century and present day Community Earth System Model (CESM) projections (CESM-EC–CESM-PD) (left column) and WRF projections (WRF-EC–WRF-PD) (right column).

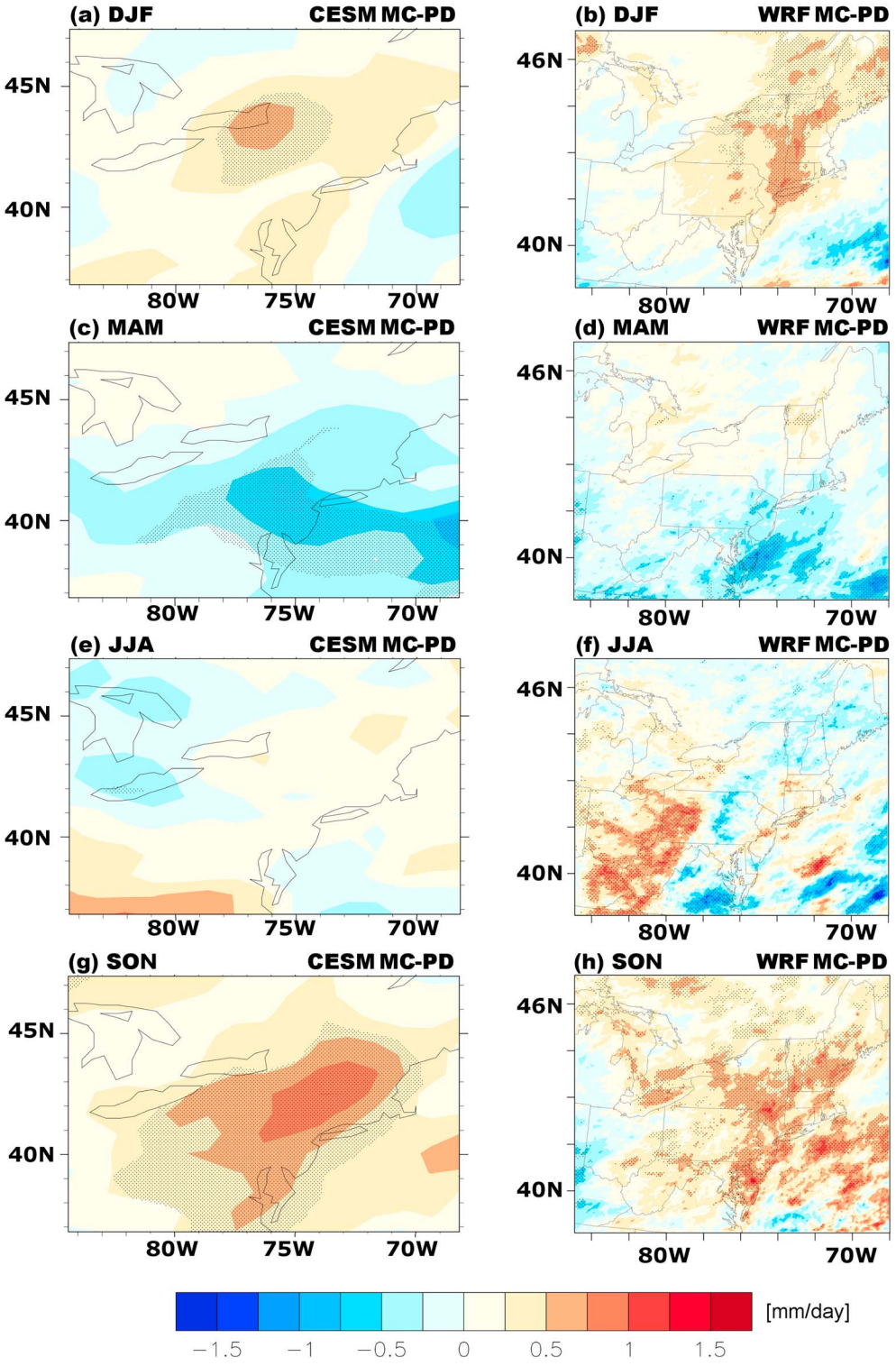

Figure 15. Seasonal differences (December-January-February [DJF], MAM, JJA, and SON) in daily precipitation rates [mm/day] between end of century and present day CESM projections CESM-MC–CESM-PD) (left column) and WRF projections (WRF-MC–WRF-PD) (right column).

regional climate compared to the coarser resolution driver ESM, further studies are needed to determine the effect of lakes on our downscaled results.

The differences between CESM- and WRF-simulated changes in temperatures are more pronounced for daily extreme temperatures especially toward the end of the century. Daily minimum and maximum temperatures at 2 m increase for both CESM and WRF-downscaled CESM simulations by MC and EC (Figures 11 and 12). Compared to CESM projections, the increase in maximum temperatures at 2 m is less in WRF-downscaled projections (Figures 11 and 12). On the other hand, PD temperature minimums are lower in WRF-downscaled projections (e.g., Figure 2; CESM-Hist vs. WRF-Hist) and go through a more pronounced increase from PD into the end of the century (Figure 11). In downscaled projections, PD daily mean temperatures become the daily temperature minimums by the end of the century (not shown). Due to the bias correction applied to CESM output prior to downscaling, it is hard to attribute the differences between WRF and CESM-simulated changes solely to downscaling.

Annual cycles of precipitation rate averaged over domain 3 (including both land and ocean grid points) for all WRF downscaled (WRF-PD, WRF-MC, and WRF-EC) and CESM (CESM-PD, CESM-MC, and CESM-EC) projections are plotted in Figure 13 along with GPCP estimations of historical precipitation rates from 2006 to 2015 for comparison and 95% confidence intervals. Precipitation rates from CESM for all time periods (PD, MC, and EC) are significantly underestimated especially during PD compared to GPCP estimations of historical precipitation rates. In CESM simulations, precipitation rates systematically increase from PD to EC for all months while following the same month-to-month trend as in CESM-PD (Figure 13). This continually increasing trend in precipitation rates is not replicated in WRF-downscaled simulations: In WRF-downscaled projections, precipitation rates prominently increase from November to April and in June, July, and October by the end of the century (WRF-EC) compared to PD (WRF-PD) projections. EC projections (WRF-EC) precipitation rates in May, August, September, and November, however, are nearly the same as PD rates (WRF-PD). For MC, precipitation rates (WRF-MC) increase in February, March, June, September, and December compared to PD (WRF-PD). Furthermore, MC precipitation rates (WRF-MC) are reduced in January, April, and May and the same in July and October compared to present-day projections (WRF-PD).

The nonsystematic changes in precipitation rates for WRF-downscaled projections compared to precipitation rates from CESM highlight the detailed nature of the high-resolution regional modeling with WRF. Though it would need further analysis and simulations to suggest what leads to these differences in precipitation rates between WRF-downscaled and CESM projections, nonsystematic changes indicate that the model is utilizing the spatial features of the domain, permitting convection, and resolving more small-scale processes as the climate changes through boundary conditions from CESM to create a climate that is aware of the regional features.

To better understand the difference in the annual cycle of monthly precipitation rates between WRF-MC and WRF-EC and between WRF and CESM results, we present spatial distributions of precipitation rates: Figure 14 (Figure 15) shows the differences in precipitation rates between EC and PD (MC and PD) from CESM- and WRF-downscaled simulations for December-January-February (DJF), June-July-August (JJA) March-April-May (MAM), and September-October-November (SON). In Figures 14 and 15, statistically significant differences at 95% level are shown via stippling. For EC projections, the increase in precipitation rates from PD is more pronounced over the winter (DJF) in WRF-downscaled projections. To further evaluate the differences in changes in DJF precipitation rates between MC and EC, we compared skin temperatures for WRF-MC and WRF-EC (not shown). Skin temperatures reveal that lakes are partially frozen for MC-downscaled projections, while they are warmer and unfrozen as climate warms toward the end of the century. Hence, the 3-km resolution of our WRF simulations, which allows for the simulation of moisture and heat fluxes from these areas, makes it possible to simulate the influences of lakes on regional climate. Previous downscaling studies using much coarser horizontal resolution with fully parameterized convection revealed that using a lake model and accounting for lake depth can produce better precipitation for some regions (e.g., Small et al., 1999), while for others 2-m temperatures and frozen fractions of lakes have improved at the cost of enhanced wet bias in precipitation (e.g., Mallard et al., 2014). Using the skin temperatures from the driver ESM may lead to erroneous lake temperatures depending on lake depth because the approach does not account for the warming and cooling time of the lakes (e.g., Mallard et al., 2015). Nevertheless, we are taking advantage of the much higher resolution and convection-permitting nature of our WRF simulations and the fact that we are using monthly

averaged skin temperatures from the driver data over lakes. These considerations allow us to at least partially mimic the seasonal changes each month and produce improved surface heat and moisture fluxes and hence precipitation in this manner. To establish the validity and consistency of these results, we also plotted 10 random years of precipitation differences between MC and PD and EC and PD and obtained similar results (not shown).

5. Discussion and Conclusions

In this study, we generated the first high horizontal resolution (3-km), convection-permitting WRF dynamically downscaled data set of 45 future years of CESM projections and 10 historical years of CESM and ERA-Interim simulations for the northeastern United States. In our unique downscaling methodology, we combined explicit convection, in a large domain with no nudging applied at lateral boundaries with carefully chosen parameterizations (e.g., microphysics and land surface parameterizations) to allow for the interaction of small scale regional features with simulated climate and a two-way feedback between nested grids. We find that downscaling using our methodology provides better match with observations for mean and extreme temperatures at 2 m compared to driver data ERA-Interim. Similarly, precipitation rates are better simulated in our WRF-downscaled simulations compared to ERA-Interim and improve at some stations compared to NLDAS. Finally, we conclude that applying our unique methodology in high-resolution (3 km) simulations produces better climate means and extremes compared to the driver data.

When utilizing our WRF-downscaled CESM projections for assessments of regional scale climate change impacts, it is important that users properly take into account the PD biases in precipitation rates. Furthermore, the improvements in climate variables obtained from our downscaling methodology are specific to our model setup, region of simulation, and driver models and may not be necessarily true for other cases. While with downscaled results from high-resolution regional models are expected to lead to improved climate variables in regional scales given suitable model setup, there are still uncertainties and unknowns in parameterizations of cloud process and land surface-boundary layer-cloud interactions (e.g., Flato et al., 2013). These uncertainties will continue to pose difficulties in simulating accurate regional climate variables. Our results also suggest that running ultra-high-resolution ESMs utilizing detailed parameterizations (e.g., microphysics, boundary layer, and land surface) with future advances in computational resources could improve simulated mean and extreme climates (e.g., Schneider et al., 2017).

6. Data Access and Computational Expense

Model output containing more than 200 climate variables saved at hourly intervals for domain 3 and six hourly for domains 1 and 2 are available for public use in netCDF format. Model restart (daily), boundary, and input files for all simulations are also available for users interested in continuing the simulations for longer time periods, for further downscaling and for reproducibility. The data set (including all input, output, and restart files totaling to ~2 PB of data) will be available through the Data Distribution Center at the University of New Hampshire; until then, authors can be contacted to obtain a subset of the data set.

We will briefly summarize the computational expense of these simulations to help other researchers who would be interested in a similar study for other regions or who would like to further downscale our results. We completed all simulations using our 12 million core hour allocation on NCAR Yellowstone High Performance Computer, utilizing 256 cores for each simulation. Each simulated day takes about an hour to complete excluding the preprocessing of input fields in the WRF Preprocessing System (WPS). Each simulated day takes about 30 GB of disk space including input, restart, and output files for all three domains.

Acknowledgments
Funding support for part of the modeling portion of this project was provided through NSF NH EPSCoR grant 1101245. Authors would like to thank NSF program manager Audrey D. Levine for her support on obtaining computational resources at NCAR, the 12 million core hour NCAR Yellowstone CISL allocation (CNHA001; as well as small allocations awarded to this project UNHA002 and UNHA003) that made this study possible, David L. Hart, and all high-performance computing staff at NCAR CISL. Authors also acknowledge Purdue University Research Applications Center for high-performance computing resources that helped complete the analysis stage. M. Komurcu would like to acknowledge Cindy Bruyère, Jimmy Dudhia, and Hugh Morrison of NCAR for treasured discussions during the initial stages of this study; Amin Dezfuli of NASA for his suggestions on NASA observational products; and Lodovica Illari of MIT for guidance on ground-based precipitation observations. M. Komurcu would like to thank Daniel J. Cziczo, Chien Wang, and David McGee of MIT for providing the material resources that helped complete this study and for valuable discussions during the analysis stage. Authors would like to thank the two anonymous reviewers whose comments greatly improved this work. All data sets used in the analysis and how to access them are explained in detail in sections 2.2 and 2.3.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., et al. (2003). The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, *4*(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004%3C1147:TVGPCP%3E2.0.CO;2

Alexandru, A., de Elia, R., Laprise, R., Separovic, L., & Biner, S. (2009). Sensitivity study of regional climate model simulations to large-scale nudging parameters. *Monthly Weather Review*, *137*(5), 1666–1686. https://doi.org/10.1175/2008MWR2620.1

Avramov, A., & Harrington, J. Y. (2010). Influence of parameterized ice habit on simulated mixed phase Arctic clouds. *Journal of Geophysical Research*, *115*, D03205. https://doi.org/10.1029/2009JD012108

Balzarini, A., Angelini, F., Ferrero, L., Moscatelli, M., Perrone, M. G., Pirovano, G., et al. (2014). Sensitivity analysis of PBL schemes by comparing WRF model and experimental data. *Geoscientific Model Development Discussion*, *7*(5), 6133–6171. https://doi.org/10.5194/gmdd-7-6133-2014

Bassill, N. P. (2014). Accuracy of early GFS and ECMWF Sandy (2012) track forecasts: Evidence for a dependence on cumulus parameterization. *Geophysical Research Letters*, *41*, 3274–3281. https://doi.org/10.1002/2014GL059839

Bieniek, P. A., Bhatt, U. S., Walsh, J. E., Rupp, T. S., Zhang, J., Krieger, J. R., et al. (2015). Dynamical downscaling of ERA-Interim temperature and precipitation for Alaska. *Journal of Applied Meteorology and Climatology*, *55*(3), 635–654. https://doi.org/10.1175/JAMC-D-15-0153.1

Bruyère, C. L., Done, J. M., Holland, G. J., & Fredrick, S. (2014). Bias corrections of global models for regional climate simulations of high-impact weather. *Climate Dynamics*, *43*(7-8), 1847–1856. https://doi.org/10.1007/s00382-013-2011-6

Bruyère, C. L., Monaghan, A. J., Steinhoff, D. F. & Yates, D. (2015). *Bias-corrected CMIP5 CESM data in WRF/MPAS intermediate file format* (technical report TN-515+STR, 27). Boulder CO: National Center for Atmospheric Research. doi:10.5065/D6445JJ7

Bryan, G. H., Wyngaard, J. C., & Fritsch, J. M. (2003). Resolution requirements for the simulation of deep moist updrafts. *Monthly Weather Review*, *131*(10), 2394–2416. https://doi.org/10.1175/1520-0493(2003)131<2394:RRFTSO>2.0.CO;2

Caldwell, P., Chin, H.-N. S., Bader, D. C., & Bala, G. (2009). Evaluation of WRF dynamical downscaling simulation over California. *Climate Change*, *95*(3–4), 499–521. https://doi.org/10.1007/s10584-009-9583-5

Chan, S. C., Kendon, E. J., Fowler, H. J., Blenkinsop, S., Ferro, C. A. T., & Stephenson, D. B. (2013). *Climate Dynamics*, *41*(5-6), 1475–1495. https://doi.org/10.1007/s00382-012-1568-9

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

Déqué, M., Somot, S., Sanchez-Gomez, E., Goodess, C. M., Jacob, D., Lenderink, G., et al. (2011). The spread amongst ENSEMBLES regional scenarios: Regional climate models, driving general circulation models and interannual variability. *Climate Dynamics*, *38*(5-6), 951–964. https://doi.org/10.1007/s00382-011-1053-x

Done, J. M., Holland, G. J., Bruyère, C. L., Leung, L. R., & Suzuki-Parker, A. (2015). Modeling high-impact weather and climate: Lessons from a tropical cyclone perspective. *Climatic Change*, *129*(3–4), 381–395. https://doi.org/10.1007/s10584-013-0954-6

d'Orgeville, M., Peltier, W. R., Erler, A. R., & Gula, J. (2014). Climate change impacts on Great Lakes Basin precipitation extremes. *Journal of Geophysical Research: Atmospheres*, *119*, 10,799–10,812. https://doi.org/10.1002/2014JD021855

Dosio, A., Panitz, H. J., Schubert-Frisius, M., & Luthi, D. (2014). Dynamic downscaling of CMPI5 global circulation models over CORDEX-Africa with COSMO-CLM: Evaluation over the present climate and analysis of the added value. *Climate Dynamics*, *44*(9-10), 2637–2661. https://doi.org/10.1007/s00382-014-2262-x

European Centre for Medium-Range Weather Forecasts (2012). ERA-Interim project, monthly means. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D68050NT

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., et al. (2013). Evaluation of climate models. In T. F. Stocker, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Vol. 9, pp. 741–866). Cambridge, UK and New York: Cambridge University press.

Gao, Y., Fu, J. S., Drake, J. B., Liu, Y., & Lamarque, J.-F. (2012). Projected changes in extreme events in the eastern United States based on a high-resolution climate modeling system. *Environmental Research Letters*, *7*(4), 044025. https://doi.org/10.1088/1748-9326/7/4/044025

Gelaro, R., McCarty, W., Suarez, 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

Gettelman, A., Morrison, H., Santos, S., Bogenschutz, P., & Caldwell, P. (2015). Advanced two-moment bulk microphysics for global models. Part II: Global model solutions and aerosol–cloud interactions. *Journal of Climate*, *28*(3), 1288–1307. https://doi.org/10.1175/JCLI-D-14-00103.1

Giorgi, F., & Mearns, L. O. (1999). Introduction to special section: Regional climate modeling revisited. *Journal of Geophysical Research*, *104*(D6), 6335–6352. https://doi.org/10.1029/98JD02072

Harding, K. J., & Snyder, P. K. (2014). Examining future changes in the character of Central U.S. warm-season precipitation using dynamical downscaling. *Journal of Geophysical Research: Atmospheres*, *119*, 23. https://doi.org/10.1002/2014JD022575

Harris, L. M., & Durran, D. R. (2010). An idealized comparison of one-way two-way grid nesting. *Monthly Weather Review*, *138*(6), 2174–2187. https://doi.org/10.1175/2010MWR3080.1

Hartmann, D. L., Klein Tank, A. M. G., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y., et al. (2013). Observations: Atmosphere and surface. In T. F. Stocker, et al. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Vol. 2, pp. 159–254). Cambridge, UK and New York: Cambridge University Press.

Heavens, N. G., Ward, D. S., & Natalie, M. M. (2013). Studying and projecting climate change with Earth system models. *Nature Education Knowledge*, *4*(5), 4.

Hong, S. Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*(9), 2318–2341. https://doi.org/10.1175/MWR3199.1

Hu, X.-M., Xue, M., McPherson, R. A., Martin, E., Rosendahl, D. H., & Qiao, L. (2018). Precipitation dynamical downscaling over the Great Plains. *Journal of Advances in Modeling Earth Systems*, *10*(2), 421–447. https://doi.org/10.1002/2017MS001154

Huffman, G. J., Adler, R. F., Arkin, P., Chang, A., Ferraro, R., & Gruber, et al. (1997). The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bulletin of the American Meteorological Society*, *78*(1), 5–20. https://doi.org/10.1175/1520-0477(1997)078%3C0005:TGPCPG%3E2.0.CO;2

Huffman, G. J., Adler, R. F., Bolvin, D. T., & Gu, G. (2009). Improving the global precipitation record: GPCP version 2.1. *Geophysical Research Letters*, *36*, L17808. https://doi.org/10.1029/2009GL040000

Huffman, G. J., Adler, R. F., Morrissey, M., Bolvin, D. T., Curtis, S., Joyce, R., et al. (2001). Global precipitation at one-degree daily resolution from multi-satellite observations. *Journal of Hydrometeorology*, *2*(1), 36–50. https://doi.org/10.1175/1525-7541(2001)002%3C0036:GPAODD%3E2.0.CO;2

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long–lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*, D13103. https://doi.org/10.1029/2008JD009944

Jang, S., Kavvas, M. L., Ishida, K., Trinh, T., Ohara, N., Kure, S., et al. (2017). A performance evaluation of dynamical downscaling of precipitation over northern California. *Sustainability*, *9*(8), 1457. https://doi.org/10.3390/su9081457

Joyce, R., Janowiak, J., Arkin, P. A., & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, *5*(3), 487–503. https://doi.org/10.1175/1525-7541(2004)005%3C0487:CAMTPG%3E2.0.CO;2

Kain, J. S. (2004). The Kain–Fritsch convective parameterization: An update. *Journal of Applied Meteorology and Climatology*, *43*(1), 170–181. https://doi.org/10.1175/1520-0450(2004)043%3C0170:TKCPAU%3E2.0.CO;2

Komurcu, M. (2016). Influences of ice crystal number concentrations and habits on arctic mixed-phase cloud dynamics. *Pure and Applied Geophysics*, *173*(9), 3125–3140. https://doi.org/10.1007/s00024-015-1132-8

Komurcu, M., Storelvmo, T., Tan, I., Lohmann, U., Yun, Y., Penner, J. E., et al. (2014). Intercomparison of the cloud water phase among global climate models. *Journal of Geophysical Research: Atmospheres*, *119*, 3372–3400. https://doi.org/10.1002/2013JD021119

Kraucunas, I., Clarke, L., Dirks, J., Hathaway, J., Hejazi, M., Hibbard, K., et al. (2014). Investigating the Nexus of climate, energy, water, and land at decision-relevant scales: The Platform for Regional Integrated Modeling and Analysis (PRIMA). *Climatic Change, Special Issue on Regional Earth System Modeling*, *129*(3–4), 573–588. https://doi.org/10.1007/s10584-014-1064-9

Leduc, M., & Laprise, R. (2009). Regional climate model sensitivity to domain size. *Climate Dynamics*, *32*(6), 833–854. https://doi.org/10.1007/s00382-008-0400-z

Liu, C., Ikeda, K., Thompson, G., Rasmussen, R. M., & Dudhia, J. (2011). High-resolution simulations of wintertime precipitation in the Colorado headwaters region: Sensitivity to physics parameterizations. *Monthly Weather Review*, *139*(11), 3533–3553. https://doi.org/10.1175/MWR-D-11-00009.1

Lorenz, C., & Kunstmann, H. (2012). The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *Journal of Hydrometeorology*, *13*(5), 1397–1420. https://doi.org/10.1175/JHM-D-11-088.1

Lorenz, P., & Jacob, D. (2005). Influence of regional scale information on the global circulation: A two-way nesting climate simulation. *Geophysical Research Letters*, *32*, L18706. https://doi.org/10.1029/2005GL023351

Mallard, M. S., Nolte, C. G., Bullock, O. R., Spero, T. L., & Gula, J. (2014). Using a coupled lake model with WRF for dynamical downscaling. *Journal of Geophysical Research: Atmospheres*, *119*, 7193–7208. https://doi.org/10.1002/2014JD021785

Mallard, M. S., Nolte, C. G., Spero, T. L., Bullock, O. R., Alapaty, K., Herwehe, J. A., et al. (2015). Technical challenges and solutions in representing lakes when using WRF in downscaling applications. *Geoscientific Model Development*, *8*(4), 1085–1096. https://doi.org/10.5194/gmd-8-1085-2015

Mearns, L. O., Arritt, R., Biner, S., Bukovsky, M. S., McGinnis, S., Sain, S., et al. (2012). The North American regional climate change assessment program. *Bulletin of the American Meteorological Society*, *93*(9), 1337–1362. https://doi.org/10.1175/BAMS-D-11-00223.2

Mearns, L. O., Gutowski, W. J., Jones, R., Leung, L.-Y., McGinnis, S., Nunes, A. M. B., et al. (2009). A regional climate change assessment program for North America. *Eos, Transactions of the American Geophysical Union*, *90*(36), 311–312. https://doi.org/10.1029/2009EO360002

Mearns, L. O., McGinnis, S., Korytina, D., Arritt, R., Biner, S., Bukovsky, M., et al. (2017). The NA-CORDEX dataset, version 1.0. NCAR Climate Data Gateway, Boulder CO, accessed [September, 22, 2018], doi: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

Miguez-Macho, G., Stenchikov, G. L., & Robock, A. (2005). Regional climate simulations over North America: Interaction of local processes with improved large-scale flow. *Journal of Climate*, *18*(8), 1227–1246. https://doi.org/10.1175/JCLI3369.1

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., et al. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research*, *109*, D07S90. https://doi.org/10.1029/2003JD003823

Monaghan, A. J., Steinhoff, D. F., Bruyère, C. L. & Yates D. (2014). NCAR CESM global Bias-corrected CMIP5 output to support WRF/MPAS research. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D6DJ5CN4

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-Moment schemes. *Monthly Weather Review*, *137*(3), 991–1007. https://doi.org/10.1175/2008MWR2556.1

Oleson, K. W., Monaghan, A., Wilhelmi, O., Barlage, M., Brunsell, N., Feddema, J., et al. (2013). Interactions between urbanization, heat stress, and climate change. *Climatic Change, Special Issue on Regional Earth System Modeling*, *129*(3–4), 525–541. https://doi.org/10.1007/s10584-013-0936-8

Oleson, K. W., Lawrence, D. M., Gordon, B. B., Flanner, M. G., Kluzer, E., Lawrence, P. J., et al. (2010). *Technical description of version 4 of the Community Land Model (CLM)* (Technical Note NCAR/TN–478+STR). Boulder CO: National Center for Atmospheric Research. https://doi.org/10.5065/D6FB50WZ

Pierce, D. W., Tapash, D., Cayan, D. R., Maurer, E. P., Miller, N. L., Bao, Y., et al. (2012). Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling. *Climate Dynamics*, *40*(3–4), 839–856. https://doi.org/10.1007/s00382-012-1337-9

Prein, A., Rasmussen, R. M., Ikeda, K., Liu, C., Klark, M. P., & Holland, G. J. (2017). The future intensification of hourly extremes. *Nature Climate Change*, *7*(1), 48–52. https://doi.org/10.1038/nclimate3168

Prinn, R. G. (2013). Development and application of earth system models. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(Supplement 1), 3673–3680. https://doi.org/10.1073/pnas.1107470109

Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P. P., & Partyka, G. S. (2017). Land surface precipitation in MERRA-2. *Journal of Climate*, *30*(5), 1643–1664. https://doi.org/10.1175/JCLI-D-16-0570.1

Reisner, J. R., Rasmussen, R. M., & Bruintjes, R. T. (1998). Explicit forecasting of super cooled liquid water in winter storms using the MM5 mesoscale model. *Quarterly Journal of the Royal Meteorological Society*, *124*(548), 1071–1107. https://doi.org/10.1002/qj.49712454804

Research Data Archive/Computational and Information Systems Laboratory/National Center for Atmospheric Research/University Corporation for Atmospheric Research, and Community Climate System Model/Climate and Global Dynamics Division/National Center for Atmospheric Research/University Corporation for Atmospheric Research (2011). NCAR Community Earth System Model, EaSM project dataset. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D6TH8JP5

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

Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, *44*, 12,396–12,417. https://doi.org/10.1002/2017GL076101

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., & Ziese, M. (2015). GPCC full data reanalysis version 7.0 at 1.0°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. https://doi.org/10.5676/DWD_GPCC/FD_M_V7_100

Seneviratne, S., et al. (2012). Changes in climate extremes and their impacts on the natural physical environment. In C. B. Field, et al. (Eds.), *IPCC WGI/WGII special report on managing the risks of extreme events and disasters to advance climate change adaptation* (pp. 109–230). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139177245.006

Shaaf, B., von Storch, H., & Fesser, F. (2017). Does spectral nudging have an effect on dynamical downscaling applied in small regional model domains? *Monthly Weather Review*, *145*(10), 4303–4311. https://doi.org/10.1175/MWR-D-17-0087.1

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M., et al. (2008). *A Description of the Advanced Research WRF Version 3* (technical note NCAR/TN-475+STR). Boulder CO: National Center for Atmospheric Research. https://doi.org/10.5065/D68S4MVH

Small, E. E., Sloan, L. C., Hostetler, S., & Giorgi, F. (1999). Simulating the water balance of the Aral sea with a coupled regional climate-lake model. *Journal of Geophysical Research*, *104*(D6), 6583–6602. https://doi.org/10.1029/98JD02348

Soares, P. M. M., Cardoso, R. M., Miranda, P. M. A., de Medeiros, J., Belo-Pereira, M., & Espirito-Santo, F. (2012). WRF high resolution dynamical downscaling of ERA-Interim for Portugal. *Climate Dynamics*, *39*(9–10), 2497–2522. https://doi.org/10.1007/s00382-012-1315-2

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K.-L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, *56*, 79–107. https://doi.org/10.1002/2017RG000574

Sun, X., Xue, M., Brotzge, J., McPherson, R. A., Hu, X.-M., & Yang, X.-Q. (2016). An evaluation of dynamical downscaling of Central Plains summer precipitation using a WRF-based regional climate model at a convection permitting 4 km resolution. *Journal of Geophysical Research: Atmospheres*, *121*, 13,801–13,825. https://doi.org/10.1002/2016JD024796

Thompson, G. (2012). High-resolution simulations of winter precipitation over the Colorado Rockies. Paper presented at ECMWF Workshop on Parametrization of Clouds and Precipitation, Shinfield Park, Reading, UK.

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, *136*(12), 5095–5115. https://doi.org/10.1175/2008MWR2387.1

Walther, A., Jeong, J.-H., Nikulin, G., Jones, C., & Chen, D. (2013). Evaluation of the warm season diurnal cycle of precipitation over Sweden simulated by the Rossby Centre regional climate model RCA3. *Atmosheric Research*, *119*, 131–139. https://doi.org/10.1016/j.atmosres.2011.10.012

Wang, Y., Long, C. N., Leung, L., Dudhia, J., McFarlane, S. A., & Mather, et al. (2009). Evaluating regional cloud-permitting simulations of the WRF model for the Tropical Warm Pool International Cloud Experiment (TWP-ICE), Darwin, 2006. *Journal of Geophysical Research*, *114*, D21203. https://doi.org/10.1029/2009JD012729

Warner, T. T., Peterson, R. A., & Treadon, R. E. (1997). A tutorial on lateral boundary conditions as a basic and potential serious limitation to regional numerical weather prediction. *Bulletin of the American Meteorological Society*, *78*(11), 2599–2617. https://doi.org/10.1175/1520-0477(1997)078%3C2599:ATOLBC%3E2.0.CO;2

Wu, J., Zhou, Y., Gao, Y., Fu, J. S., Johnson, B. A., Huang, C., et al. (2014). Estimation and uncertainty analysis of impacts of future heat waves on mortality in the eastern United States. *Environmental Health Perspectives*, *122*(1), 10–16. https://doi.org/10.1175/MWR-D-11-00215.110.1289/ehp.1306670

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., et al. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research*, *117*, D03109. https://doi.org/10.1029/2011JD016048