EVALUATIONS OF HISTORICAL AND PROJECTED HIGH-RESOLUTION DYNAMICALLY DOWNSCALED ENSEMBLE OVER THE CONTINENTAL UNITED STATES

BY

ZACHARY ZOBEL

DISsertATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Atmospheric Sciences in the Graduate College of the University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Donald J. Wuebbles
Professor Robert Trapp
Associate Professor Francina Dominguez
Dr. V. Rao Kotamarthi
ABSTRACT

There has been extensive research in the science community to quantify the impact that anthropogenic greenhouse gas emissions will have on regional climate change. Earth system models (ESM) have been shown to accurately project changes in the climate system at a continental scale, but lack the spatial resolution needed to represent mesoscale processes that affect regional climate extremes. To rectify this limitation, a technique known as dynamically downscaling was introduced to better resolve these small-scale processes and provide society with quantitative evaluation of regional climate risks associated with a warming climate.

This dissertation uses an ensemble of dynamically downscaled model simulations with varying boundary conditions. The Weather Research and Forecast (WRF) model is used to evaluate the performance of five 12-km spatial resolution decadal historical and future simulations with a domain that covers most of North America. The initial and boundary conditions are from three ESMs (GFDL-ESM2G, CCSM4, and HadGEM2) with varying climate sensitivities. The future projections will use two greenhouse gas (GHG) concentration scenarios and two decadal-length time slices (2045-2054 and 2085-2094), which is compared to a historical decade (1995-2004).

Chapter 2 quantifies the uncertainty associated with bias correction, spectral nudging, and the lateral boundary conditions when comparing historical simulations to observations. In addition to showing the “added value” of the dynamical downscaling technique over the ESM data, this section evaluates the model performance for the ensemble. The results indicate that the simulation’s performance depends on both location and the features/variable being tested. The use of an ensemble mean and median leads to a better performance in measuring the climatology, but is significantly biased for the extremes when compared to the individual RCM simulations.
Chapter 3 of this dissertation examines projections of extreme temperatures. Probability density functions of daily maximum/minimum temperatures are analyzed. The uncertainties associated with using different boundary conditions as well as future GHG concentrations on extreme events, such as heat waves and days with temperature higher than 95°F, are investigated. The distribution of summer daily maximum temperature experiences a significant warm-side shift and increased variability, while the distribution of winter daily minimum temperature is projected to have a less significant warm-side shift with decreased variability.

Chapter 4 examines the projections of extreme daily precipitation over the U.S. to quantify the effects a warming climate on precipitation distribution and intensity. There is a large increase in the projected frequency of extreme precipitation events over the entire CONUS and a decrease in median precipitation days. Moreover, most regions show an increase in the number of dry days for the future scenarios. The magnitude of extreme precipitation events is projected to increase at all temperatures above freezing in the CONUS. The strongest precipitation events are increasing mostly as a result of a shift in the precipitation distribution due to the increase of temperature (Clausius-Claperyon relationship), but are also affected by changes in some dynamical factors.

Chapter 5 points to some of the important additions to the climate change literature this dissertation has made. By quantifying the uncertainties associated with climate extremes using this ensemble, there is a better understanding of how model setup, emission scenarios, and climate sensitivity in the ESM data will affect climate projections when considering extreme temperature and precipitation events.
# TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION ........................................................................................................... 1

CHAPTER 2: EVALUATIONS OF HIGH-RESOLUTION DYNAMICALLY DOWNSCALED ENSEMBLES OVER THE CONTIGUOUS UNITED STATES .............................................. 12

CHAPTER 3: HIGH RESOLUTION DYNAMICAL DOWNSCALING ENSEMBLE PROJECTIONS OF FUTURE EXTREME TEMPERATURE DISTRIBUTIONS FOR THE UNITED STATES ..................................................................................................................... 51

CHAPTER 4: ANALYSES FOR HIGH RESOLUTION PROJECTIONS THROUGH THE END OF THE 21ST CENTURY FOR PRECIPITATION EXTREMES OVER THE UNITED STATES ......................................................................................................................... 80

CHAPTER 5: SUMMARY AND CONCLUSIONS .............................................................................. 114

REFERENCES .................................................................................................................................. 119

APPENDIX A: SUPPLEMENTAL INFORMATION FOR CHAPTER 2 ............................................. 141

APPENDIX B: SUPPLEMENTAL INFORMATION FOR CHAPTER 3 ............................................. 143

APPENDIX C: SUPPLEMENTAL INFORMATION FOR CHAPTER 4 ............................................. 144
CHAPTER 1: INTRODUCTION

Global climate models (GCMs), also commonly referred to as Earth System Models (ESMs), show significant skill when simulating climate processes at the continental scale and incorporate a large proportion of the complexity of the global system (Sheffield et al. 2013a, 2013b; IPCC AR5). However, their ability to capture local or small scale features (e.g., topography) is limited due to their coarse spatial resolution (Wigley et al. 1990; Carter et al. 1994; Wang et al. 2015). When considering the impacts of global climate change, the focus is primarily on the physical effects occurring at regional, local, and smaller scales. In order to provide well-informed projections of future changes in climate for planners and the public, analyses of these effects for spatial scales on the order of tens of kilometers are necessary. Therefore, the concept for dynamical downscaling (DD) was developed. DD embeds a higher resolution regional climate model (RCM) within the boundary conditions from low resolution ESMs with the results producing a better understanding of the complexity of the climate system with significantly lower horizontal grid spacing (Giorgi 1990; Mearns et al. 1995; von Storch et al. 2000). This dissertation presents results from a 5-member ensemble of dynamically downscaled simulations with a 12-km spatial resolution to quantify the impacts of climate change on future extreme precipitation and temperature events. The results are presented using several climatologically cohesive regions of the Contiguous United States (CONUS) with an ensemble of simulations that make use of multiple decadal time slices throughout the 21st century, three separate ESM boundary conditions, and two representative concentration pathways for future emissions.
While there are some limitations to the DD approach, such as potential biases introduced by the RCM used to perform the downscaling, their ability to more accurately capture regional climate extremes are significantly improved over the low resolution GCM data (e.g. Zobel et al. 2018a; Wang et al. 2015; Racherla et al. 2012; Hayhoe et al. 2008; Liang et al. 2006). For example, mesoscale convective systems result in as much as 60% of observed seasonal rainfall in the central United States during the Spring and Summer, but existing ESM results lack the ability, including the high spatial resolution needed, to simulate these events accurately, leading to large uncertainty when making future projections (Kooperman et al. 2013). Increased resolution improves the ability to capture mesoscale processes that are vital for accurately projecting future extreme events (Bacmeister et al. 2013). However, running the model at high resolution (on order of a few tens of kilometers) over the entire globe is still not readily feasible for multi-decadal simulations due to the limitations in available computing resources. Regional climate models (RCMs) with considerably higher resolution are therefore constructed for limited areas to describe regional-scale climate variability and change. The RCMs are constrained at the boundaries and partially over the inner domain by global- or relatively large-scale meteorological driving data. Moreover, if the RCM model grid size is reduced to less than a few kilometers, the parameterization of the sub-grid scale process (e.g., convection) can be eliminated and the performance of RCM can be significantly further improved (Chang et al. 2018). Thus, dynamical downscaling has proven to be useful for generating climate projections and especially promising for understanding potential future changes in precipitation and temperature extremes. Several dynamical downscaling studies have been conducted over the North American domain for historical and/or future climate such as the North American Regional Climate Change
Assessment Program (NARCCAP). NARCCAP employs a domain similar to that used in this study over most North America with a 50-km spatial resolution for 30 years in historical and future simulations and a relatively large ensemble (Mearns et al. 2012). Many other studies have used downscaling to discuss regional precipitation and temperature changes, with a spatial resolution similar to this study, but over a smaller single region domain (e.g., Hayhoe et al. 2008; Bachmeister et al. 2013; d’Orgeville et al. 2014). Because of the computational expense of the climate models, some concessions must currently be made in terms of spatial resolution, temporal length, and/or number of unique boundary conditions used to perform downscaling (i.e., ensemble size). Increasing any of these features within a dataset can improve the understanding of future climate projections. The ensemble presented in this study is unique because few studies maintains a large domain similar to NARCCAP while increasing spatial resolution to a much higher resolution, albeit with a smaller ensemble size and shorter temporal length than NARCCAP.

1.1 Evaluations of historical biases and model setup

An important feature of any downscaling study aimed at making future projections of climate extremes is to first understand the historical biases and the uncertainties present in the ensemble when using different boundary conditions. Zobel et al. (2018a) aims to accomplish this by evaluating the differences between the ESMs used in this study compared to its dynamically downscaled counterparts. In addition, Zobel et al. (2017) [Chapter 3 of this dissertation] and Zobel et al. (2018b) [Chapter 4 of this dissertation] provide additional statistical measures to show how the downscaled simulations perform historically in terms of temperature probability density functions and extreme precipitation events when compared to observations.
Di Luca et al. (2012) showed that Regional Climate Models (RCMs) add high value for warm-season precipitation over short temporal scales, especially over regions of complex topography. Pryor et al. (2012) noted that an increase in RCM resolution from 50 to 6 km better captures extreme values of wind speeds. Vautard et al. (2013) found that heat extremes in Europe were generally better simulated in RCMs with resolution of 12 km versus 50 km. Tripathi and Dominguez (2013) found that 10 km simulation captures individual extreme summer precipitation events better than 50 km simulation. Similarly, Wang et al. (2015) found that a RCM at 12 km spatial resolution captures significantly more details of the spatial and temporal variations of precipitation (especially over mountainous regions) than does the 2.5-degree National Centers for Environmental Prediction-U.S. Department of Energy Reanalysis II (NCEP-R2), even when the RCM is aggregated to the grid resolution of NCEP-R2. In addition to analyzing the historical biases of this ensemble, Chapter 2 as well as Chapter 4 discuss the “added value” of the downscaled ensemble over the ESM counterpart for the historical period.

Uncertainties in the RCM simulations can come from a number of sources, including physics parameterization, model representation of internal variability, the choice of emission scenarios for projecting future changes in climate, and the differences in the global climate models used to drive the RCM (Giorgi and Bi 2000; Mearns et al. 2012). Development of multi-RCM and multi-ESM ensembles have been identified as one of the current research needs (IPCC4, Doherty et al. 2009) to reduce uncertainties in projections of climate. For example, in order to explore the uncertainties due to various ESMs and RCMs applied to investigate regional climate change, the North American Regional Climate Change Assessment Program (NARCCAP) simulated 30 years in historical and future periods respectively by using six RCMs driven by different GCMs
over North America (Mearns et al. 2012); the Prediction of Regional Scenarios and Uncertainties for Defining European Climate Change Risks and Effects (PRUDENCE) (Christensen et al. 2007) and ensembles (van der Linden and Mitchell 2009; Christensen et al. 2010) simulated historical and future periods over Europe. The Coordinated Regional Climate Downscaling Experiment (CORDEX) has built a climate projection framework for different regions over the entire globe (Giorgi et al. 2009) and aims to explore the maximum extent of the contribution of different sources of uncertainty such as different boundary conditions from various GCMs, different initializations (internal variability), and different emission/concentration scenarios. Future projections for the CMIP5 models simulated four separate future climate scenarios and these scenarios are defined by using different Representative Concentration Pathways (RCPs) (IPCC AR5). Here, we use two RCP scenarios: RCP 8.5 assumes continued heavy use of fossil fuels at a similar, or greater, rate as current emissions of CO$_2$ and other GHGs through the end of the century leading to a radiative forcing of 8.5 W/m$^2$ by 2100 (Riahi et al. 2011); RCP 4.5 scenario is a pathway to stabilize radiative forcing at 4.5 W/m$^2$ by 2100 and implies significant reduction in emissions from fossil fuel use by the end of the century. With this scenario, GHG emissions in the atmosphere peak in the 2040s before stabilizing toward the end of the century (Thomson et al. 2011).

Regarding the uncertainty that is induced by different lateral boundary conditions, Knutti et al. (2013) found that there is no single GCM that stands out as being particularly better or worse across all analyzed variables over the entire model domain. Sanderson et al. (2017) quantify the skill of all CMIP5 ESMs in attempt to provide a better a framework for weighting future climate projections, but that careful consideration of relevant processes and domains will be needed
when applying these metrics. Different simulations using the same RCM driven with multiple boundary conditions show varying skill over particular regions. For example, Halmstad et al. (2012) investigate the NARCCAP six simulations for the historical period over the Villamette River Basin, Oregon, and find that the Weather Research and Forecasting (WRF) regional climate model performs better in extreme precipitation than its boundary conditions when driven by the Community Climate System Model, version 3, but performs worse than the boundary conditions when driven by the Canadian Climate Centre Coupled General Circulation Model version 3. Pryor et al. (2012) analyze NARCCAP simulations over six subregions of CONUS, and find that every single RCM’s performance in wind climate (mean and extreme) can be very different when they are driven by varying boundary conditions. We choose three different ESMs, they are Geophysical Fluid Dynamics Laboratory Earth System Model with Generalized Ocean Layer Dynamics component (GFDL-ESM2G), Community Climate System Model, version 4 (CCSM4), and the Hadley Centre Global Environment Model, version 2-Earth System (HadGEM2-ES). According to the results presented by Sanderson et al. (2017), the CCSM4 and HadGEM2-ES simulations exhibit considerable skill across North America when compared to the rest of the CMIP5 simulations with the GFDL-ESM2G performing slightly below average. In addition, the three ESMs chosen for this study should provide a range of values for several regional climate variables due to differences in model setup when making future projections. This is evident when we consider their different climate sensitivities when forced with a doubling of CO₂; where GFDL-ESM2G yields the lowest increase in mean surface temperature at 2.38 K, CCSM4 yields 2.92 K warming, and HadGEM2-ES yields the highest warming of 4.55 K (Sherwood et al. 2014). It is important to understand the uncertainty of climate extremes
using this range as the IPCC 5\textsuperscript{th} Assessment considers 2 to 4.5 K as the best theoretical range for climate sensitivity

\subsection*{1.2 Temperature Distributions and Extreme Events}

The ensemble of dynamically downscaled RCM projections is used in Chapter 3 to quantify the effects and uncertainties of projected changes in climate on regional and seasonal temperature distributions. Results from this chapter have been published in Earth’s Future (citation for this paper provided on Page 49 of this dissertation).

From 1986-2015, the number one weather related cause of fatalities in the United States occurred as a result of excessive heat – more than hurricanes and tornados combined during this 30-year time span (http://www.nws.noaa.gov/om/hazstats.shtml). Depending on the GCM climate sensitivity, GCMs forecast a globally averaged temperature change of 2-4.5 °C when atmospheric CO\textsubscript{2} is doubled (Sherwood et al. 2014). However, temperature extremes are not projected to increase linearly with mean temperature changes (Hegerl et al. 2004). For example, if the mean temperature is increased by 2 degree Celsius, the 95\textsuperscript{th} percentile temperature could increase by as much as 2-3 times of that. Griffiths et al. (2004) and Sardeshmukh et al. (2015) has shown that when considering projection in temperature for mid-latitude locations, such as the United States, changes in mean temperature are not good indicators for projected changes in temperature extremes and that variance in temperature must also be considered. Therefore, it is vital to use temperature distributions when making future projections to better understand how the mean climatology and extremes are evolving under a changing climate.
In addition to understanding how the extremes and mean are evolving a warming climate, we aim to use this ensemble to better quantify the potential impacts of climate change on human health and agriculture. Heat waves have been shown to significantly increase the mortality rate in the United States, especially cities within the Midwest and the Northeast because of the high population centers and the infrequent nature of excessive heat events (Patz et al. 2005).

Similarly, Hajat et al. (2014) found a sizable increase in mortalities due to excessive heat-related events in a future climate projection in the United Kingdom, which is a location that typically experiences a much cooler climate compared to most of the United States. In addition to heat waves, extreme warm season temperatures can lead to reduction in crop yields. According to the third National Climate Assessment (NCA3), temperatures that exceed 95 °F increase frequency of negative health impacts and decrease agricultural crop yields (Melillo et al. 2014). Schlenker and Roberts (2009) found that yield rates for corn, soybeans, and cotton increase as temperatures approach 84 °F, 86 °F, and 90 °F respectively. However, temperatures greater than these thresholds act to drastically decrease crop yields. The agriculture sector must also plan to adapt to a longer growing season. According to the NCA3, it is likely that the length of the growing season (i.e. frost free season) will be increasing in length as the climate changes. These three metrics will all be considered for this study in Chapter 3 of the dissertation.

1.3 Precipitation Extremes

Like temperature extremes in Chapter 3, this ensemble will be utilized in Chapter 4 to discuss the impacts of a warming climate on precipitation extremes in terms of both distribution and magnitude. The work presented in Chapter 4 is currently under review for publication in Earth’s Future (citation for this paper provided on Page 78 of this dissertation).
From 1980-2017, flooding, excluding tropical cyclones, accounted for 12.8% of the billion-dollar weather related disasters in the United States (https://www.ncdc.noaa.gov/billions/). While not all flooding events are caused by extreme daily rainfall, there is generally a strong link between the two. The recent assessment of climate change across the United States, the Climate Science Special Report (USGCRP, 2017), found that the frequency and intensity of heavy precipitation has been increasing since 1901 and there is high confidence that this trend will continue through the end of the 21st century, but that there are important regional differences in these changes. Several previous studies have also shown that both hourly and daily extreme precipitation events have been increasing in intensity and spatial coverage throughout most of North America over the past several decades (Min et al. 2011; Peterson et al. 2013; Janssen et al. 2014, 2016; Anderson et al. 2015; Prein et al. 2017); these studies also find that these changes are primarily due to anthropogenic forcings, and not natural forcings, in the climate system (e.g., Zhang et al. 2013). Chang et al. (2016) found that by the end of 21st century, the size of individual storms is projected to shrink, while the intensity of these storms increase; the warming climate particularly promotes more convective precipitation for the future compared to the historical period.

Because saturation vapor pressure in the atmosphere is a function of atmospheric temperature via the Clausius-Clapeyron (C-C) relationship, the root cause of increasing precipitation extremes is largely related to the increase in global temperatures as the atmospheric concentrations of greenhouse gases (GHGs) increase (e.g., Allen and Ingram, 2002). When projecting future
changes in climate, Earth Systems Models (ESMs) are the primary tools used by scientists to evaluate potential changes in temperature and precipitation (e.g. Hanssen et al. 2006; IPCC, 2013). For example, Sherwood et al. (2014) found that when atmospheric CO$_2$ is doubled, ESMs increase the globally averaged surface temperature by 2-4.5 °C depending on the differences in physical parameterizations in different ESMs that have been developed independently by different institutes. Accordingly, the magnitude of extreme precipitation events should be correspondingly increasing by about 5-10% per degree Celsius change based on the C-C relationship (e.g., Allen and Ingram, 2002; G. Wang et al. 2017), but the magnitude will depend on the dynamical features associated with precipitation events within specific regions (e.g., G. Wang et al. 2017; d’Orgeville et al. 2014; Bukovsky and Karoly, 2009; Hayhoe et al. 2008). Therefore, dynamical downscaling is a valuable tool in projecting extreme precipitation and most likely superior to statistical downscaling because of its ability accurately project changes in daily extreme precipitation through the capabilities to resolve projected changes in dynamics of the atmosphere (Hayhoe et al. 2007).

In the chapters to follow, this dissertation will present evidence that these downscaled simulations show significant added value over their ESM counterpart. In addition, Chapter 2 will also highlight some of the uncertainty in the historical RCM simulations by quantifying the biases in the models. Chapters 3 and 4 will continue to highlight the uniqueness of this ensemble by presenting the future projections for extreme events within all regions across the contiguous United States at a higher spatial resolution than that typically presented in climate change research. In addition, Chapter 3 quantifies the uncertainties associated with choice of emission scenario when projecting future extreme temperature events. Chapter 4 adds to understanding
precipitation extremes in the literature by describing the differences in both frequency and magnitude of the downscaled simulations in this ensemble with their ESM counterpart. Chapter 5 will provide further discussion on the findings and summarize the key conclusions from the entirety of this dissertation.
2.1 Introduction

Our study is the first analysis of the North American continent at a very high spatial resolution (12 km) using boundary conditions derived from a range of different Coupled Model Intercomparison Project Phase 5 (CMIP5) (Taylor et al. 2012) models for decadal length scale simulations for diagnostic studies and projections of time slices in the future (Wang and Kotamarthi 2015). We choose three different GCMs, they are Geophysical Fluid Dynamics Laboratory Earth System Model with Generalized Ocean Layer Dynamics component (GFDL-ESM2G), Community Climate System Model, version 4 (CCSM4), and the Hadley Centre Global Environment Model, version 2-Earth System (HadGEM2-ES). As presented by Sherwood et al. (2014), the ultimate change in global mean temperature in response to a doubling of atmospheric CO2 in CMIP5 models, span roughly 2.1 to 4.6 °C. Among more than 30 GCMs, GFDL-ESM2G is one of those models which have very low response to the increase of CO2, with global mean temperature increasing by 2.38 °C, while HadGEM2-ES is one out of two models that have the highest response to the doubling of CO2, with global mean temperature increasing by 4.55 °C. CCSM4 shows a response which is in between above two models, with global mean temperature increasing by 2.92 °C. Therefore, these three GCMs capture the range of responses of the climate models to RCP scenarios without doing every GCM in between.

Chapter 2 aims to rank the 12-km model performance based on different meteorological fields through an evaluation of a six-member ensemble of RCMs over seven subregions of CONUS. We seek to address questions that are raised by the user community, for example, which model has the smallest bias over an area of interest? We provide a possible path for a user by ranking the various sets of model runs that are performed at a high spatial resolution. The laborious process followed to carefully perform this task will give the user community an option of carefully selecting the outputs for specific uses. There has been extensive analysis on the effectiveness of downscaling to evaluate regional climate (e.g. Fowler et al. 2007; Wang et al. 2016; Xue et al. 2014), but this study’s use of a 12-km and multi-GCMs over a large domain is unique among other downscaling research. This project uses a similar domain as the NARCCAP projections covering most of North American continent, but with approximately a four times higher resolution. The increased resolution by this factor allows for more applications over small spatial scale such as watershed scale. The value of using three separate GCMs to create an ensemble over North America will allow for an improved ability to capture the future uncertainties. Section 2.2 describes the regional climate model and GCMs applied in this study as well as the reference data. Section 2.3 evaluates the “added value” of using downscaled models as well as ranks the model performance in terms of relative error and extremes. Discussion and summary follow in section 2.4.

2.2 Model description and reference data

2.2.1 Regional climate models

The WRF model version 3.3.1 (Skamarock et al. 2008) is applied at a horizontal resolution of 12 km, with 600 west-east × 515 south-north grid points over most of North America (Figure 1 in
Wang and Kotamarthi, 2014) (WK14 hereafter). The lateral boundary conditions are specified in two different ways. As shown in Table 2.1, in the first set of the experiments, the WRF model is driven by the reanalysis of the NCEP-R2 (Kalney et al. 1996) over the period 1980–2010. In the second to the sixth sets of the experiments, the WRF models are driven by datasets from three fully coupled GCMs. The evaluation of model performance focuses not only on the climatology, but also on the extreme climate events. The atmospheric fields of interest include near surface variables as well as above surface (e.g., the most common studied 850 hPa, 500 hPa, and 200 hPa) fields. The second to sixth sets of experiments span three different time periods: 11 years over historical period (1994-2004). The name of each GCM dataset is listed in Table 2.1.

The six WRF model runs listed in Table 2.1 are the same in horizontal resolution. They are also the same in most of the physical parameterizations, which includes the Grell-Devenyi convective parameterization (Grell and Devenyi 2002), the Yonsei University planetary boundary layer scheme (Hong et al. (2006); Noh et al. 2003), the Noah land surface model (Chen and Dudhia 2001), as well as the longwave and shortwave radiative schemes of the Rapid Radiation Transfer Model for GCM applications (http://rtweb.aer.com) (Iacono et al. 2008). However, as shown in Table 2.1, the first WRF run, driven by NCEP-R2, use WSM6 (Hong and Lim 2006) microphysics and it applies spectral nudging with a nudging coefficient of $3 \times 10^{-4}$ s$^{-1}$ (the strength of nudging) Moreover, it only allows one day as spin-up time and is re-initialized every year. Among the six WRF simulations, the NCEP-R2 driven run was conducted first, along with which we conducted sensitivity experiments considering different nudging strength, microphysics, convective parameterizations and spin-up time. The details are referred to WK14. The sensitivity study showed that using weaker nudging, Morrison microphysics, and longer
spin-up time helps reducing the model bias in several different aspects, respectively. Therefore, we adjust these settings for the GCM driven runs using the Morrison microphysics scheme (Morrison et al. 2009) and one-year spin-up time for each 10-yr continuous run. For those runs who apply spectral nudging, the nudging strength is $3 \times 10^{-5}$ s$^{-1}$. We apply weak spectral nudging to air temperature, geopotential height, and wind for levels above 850 hPa. See Wang and Kotamarthi (2013) for more details about spectral nudging applied in this study. It is worth mentioning that, the comparison between WRF simulation driven by NCEP-R2 and that driven by GCMs does not aim to investigate the effect of any one of the different model setup or physics (e.g., microphysics). As we stated in the introduction, one goal of this study is to provide the users a possible path to rank the various sets of model runs that fit their unique needs. While the study by Wang and Kotamarthi (2015) compare the impacts of bias correction using CCSM4 to drive the WRF model, they only focus on precipitation over different regions of North America. This study compares not only the effect of bias correction, but also the effect of spectral nudging and different lateral boundary conditions on the model performance. The GCMs include CCSM4 developed by National Center for Atmospheric Research, United States (Gent et al. 2011), GFDL-ESG2G developed by NOAA/Geophysical Fluid Dynamics Laboratory, United States (Donner et al. 2011), and HadGEM2-ES developed by Met Office Hadley Centre, United Kingdom (Jones et al. 2011). To explore the impacts of spectral nudging on model performance when bias correction is applied, we conducted two WRF runs driven by GFDL-ESG2G, with spectral nudging turned on in one of the simulations and turned off in the other simulation. In addition to these six simulations, two more individual datasets are incorporated into this ensemble — the mean and median of the six simulations at each grid point for the 10-year period.
2.2.2 Reference data

We employ North American Regional Reanalysis (NARR) data (Mesinger et al. 2006; Bukovsky and Karoly 2007) to evaluate model performance in near surface relative humidity, wind, and high level fields, such as geopotential height, humidity, and wind. The NARR is on a spatial resolution of 32 km and covers more than 30 years from 1979 to present. The NARR assimilates observed information from multiple sources (aircraft, satellite, stations, etc.) (Tables 1 and 2 in Mesinger et al. 2006), and has been used widely as reference data by the climate downscaling community (e.g., Bowden et al. 2012; Otte et al. 2012; Liu et al. 2011; Loikith et al. 2013), although inaccuracies remain in some regions. For example, Bukovsky and Karoly (2007) found that, while the NARR provides a fairly good representation of observed precipitation over much of CONUS, some inaccuracies are found over Canada because of the relatively poor data quality that NARR assimilates. Wang et al. (2016) found that NARR overestimates (underestimates) the warming trend of January temperature over southeastern CONUS (over most of western CONUS).

For other near surface fields, such as daily maximum and minimum temperature and precipitation, we use a gridded dataset based on observations (Maurer et al. 2002). This gridded dataset is on a spatial resolution of 1/8 degree and covers 66 years from 1950 to 2015. It has been applied extensively as meteorological references for evaluating dynamical and/or statistical downscaled results (e.g., Wood et al. 2004; Christensen et al. 2004; Maurer and Hidalgo 2008; Gutowski et al. 2010; Wehner 2013). The gridded precipitation within the CONUS is from the
National Oceanic and Atmospheric Administration Cooperative Observer (Co-op) stations. The precipitation gauge data are first gridded to the one eighth degree resolution using the synergraphic mapping system algorithm of Shepard (1984) as implemented by Widmann and Bretherton (2000). The gridded daily precipitation data are then scaled to match the long-term average of the parameter-elevation regressions on independent slopes model (PRISM) precipitation climatology (Daly et al. 1994, 1997), which is a comprehensive dataset that is statistically adjusted to capture local variations due to complex terrain. The minimum and maximum daily temperature data over CONUS, also obtained from Co-op stations, are gridded using the same algorithm as for precipitation, and are lapsed to the grid cell mean elevation. We also use PRISM monthly precipitation data set as the reference data to evaluate the model and understanding the uncertainty of model’s performance to different reference data.

For our analysis, the CONUS is broken into seven subregions that are consistent with those used in the U.S. National Climate Assessment (Melillo et al. 2014). They are Northwest (NW), North Great Plains (NGP), South Great Plains (SGP), Midwest (MW), Northeast (NE), Southwest (SW), and Southeast (SE) (see Figure 2 in Janssen et al. 2014).

2.3. Results

2.3.1 Added value by dynamical downscaling

One of the key questions in downscaling research is whether the high-resolution simulation adds value against the driving GCM data. Wang et al. (2015) developed spatial and spatiotemporal
correlations considering dynamics features of precipitation. They found the improvements are apparent not only at resolutions finer than that of GCMs, but also when the RCM and observational data are aggregated to the resolution of GCMs (not illustrated). In this study, we calculate the probability density functions (PDF) of precipitation and compare the bias of GCM and RCM at the tails of the PDF distribution to investigate the potential value added by downscaling. Figure 2.1 shows differences in certain percentiles (75th, 95th, and 99th percentile) between the model and observed PDFs over seven subregions. The 10-year precipitation PDF is calculated taking only grid points where daily precipitation is greater than 1 mm. We find that there is clear advantage by using the downscaled simulations over the raw GCM counterparts, especially in mountainous and convection dependent regions and for higher percentiles in the PDF distribution. This is an important aspect of these downscaled simulations because the ability to forecast these events has major economical and societal impacts. While the RCM data still has some shortcomings at forecasting the frequency and intensity of high impact precipitation (see section 2.3.3 and section 4.3.1 for more details), Figure 2.1 shows that these simulations are a significant improvement over using raw GCM data, except the raw CCSM4 data shows slightly smaller bias than the WCNB run for the 99th percentile over Southern Plains region. Overall, the improvements seen in typical DD simulation over raw GCM data is higher than 90% for most of the regions, which is similar to the results described in Gao et al. (2012).

Figure 2.2 shows the spatially dependent differences in the 95th percentile for maximum daily temperature when compared to observations for all 5 simulations in the RCM ensemble as well as for GCM data. For comparative purposes, the GCM data has been regridded to match the grid spacing for the observations and RCM data. It is evident that there are large improvements for
most grid points in the western regions. While the magnitude of error between the GCM and the observations is similar throughout the Rocky Mountains, West coast, and Southwest, the sign of the simulated 95th percentile is different in most locations from the GFDL-ESM2G compared to observations than when the HadGEM2 and CCSM4 is compared to the observed values. This was not the case in the RCM simulations. For most locations in the western regions, the RCM simulations simulated a slight negative bias in extreme temperatures. Since the RCM simulations yield similar biases in both magnitude and sign in these regions, it yields more confidence for post simulation bias correction than in the three GCMs.

In the regions that are east of the continental divide, the biggest difference between the GCM data and RCM simulations takes place over the Northern Plains and Midwest regions. To start, when no nudging is applied in the WGNN simulation, a large positive (i.e. warm) bias is produced over the Midwest that is not simulated in the GCM. This is not the case in the WGN simulation which yielded a bias of less than +/- 1 °C in most of these same locations. Conversely, the HadGEM2 and CCSM4 GCM datasets produce a greater than 5 °C in the majority of the Midwest as well as the eastern Plains. The RCM simulations that employ these boundary conditions in the downscaling process yield significant improvements with the WCB simulation reducing the bias in these regions to about 1 °C.

2.3.2 Relative error
In this and the next section, we evaluate model performance based on metrics that describe relative error of daily mean and PDF that drawing distribution tails. To describe relative error, we employ the performance metrics developed by Gleckler et al. (2008). To begin, root-mean-square-error (RMSE) is calculated for each variable and NCA subregion for all six model runs as well as their mean and median. The reference data set depends on the type of variable being analyzed. NARR is used to evaluate above surface variables (e.g., Liu et al. 2012) while the gridded observations are used to evaluate the appropriate surface variables (e.g., temperature and precipitation). Once we determine the RMSE values, we calculate relative error (i.e. error relative to the median of the members of this ensemble) for each variable. As shown in eq. (1), to calculate relative error for a field \( f \) and model \( m \) \( (\bar{E}_{mf}) \), we define a typical model error \( (\bar{E}_f) \), which is the median of the eight RMSE values (six simulations plus median and mean) for that region and variable. We use median of RMSEs rather than mean as the typical model error to prevent models with unusually large errors from influencing the results (Gleckler et al. 2008). \( E_{mf} \) is the RMSE of one particular simulation out of six simulations plus the mean and median. The relative error \( (\bar{E}_{mf}) \) is a measure of how well a particular model performs compared to the typical model error in the ensemble. For example, if a model has a negative \( \bar{E}_{mf} \) this means it has a lower RMSE than the simulations with positive \( \bar{E}_{mf} \).

\[
\bar{E}_{mf} = \frac{E_{mf} - \bar{E}_f}{\bar{E}_f} \tag{1}
\]

Figure 2.3 shows the relative error for daily precipitation (upper left), mean temperature (upper right), and daily maximum/minimum temperature (lower left and lower right) over seven NCA regions and CONUS from the WRF simulations comparing with the gridded observation data set.
described in section 2.2.2. In general, the WRF simulations driven by GCMs score worse for all four variables than the ensemble mean and median. For precipitation, the WH and WGNN show less RMSE than other WRF simulations driven by GCMs in the MW and that includes the NCEP driven simulation for the NGP and SGP regions. The WN and WCNB predict lower RMSE than other WRF simulations driven by GCMs in NE and SE. There are noticeable differences between the models with and without bias correction. The relative error between WCNB and WCB has the greatest differences for precipitation in NGP, SGP, and MW. Using bias correction for these regions caused larger error than when no bias correction is applied to the boundary conditions. A similar trend is observed for models with and without spectral nudging. For example, WRF_GFDLN shows larger error in precipitation relative error than does WRF_GFDLNN run for all regions except for the NE.

It is worth mentioning that over the Great Plains, the WN shows positive relative errors, which means it has larger RMSEs than the typical model error. This is because, although we are using the “perfect” boundary conditions, the physics and the model setup are somewhat different from the other WRF simulations driven by GCMs. WN run is the first run that we have conducted for the project, aiming to understand the model performance and the model sensitivity to different physics and setup. In this run, we only allowed one day as spin-up time, and we re-initialized the model every year. These are two of the reasons that model shows wet bias over Great Plains. In addition, the microphysics scheme that is applied for the run also induces wet bias over Great Plain in cold seasons (WK14). Thus, we modify the model setup and microphysics for WRF simulations driven by GCMs to reduce the bias that generated by those factors.
The preferred GCM and the model setup for mean temperature (Fig. 2.3 upper right) and maximum temperature (Fig. 2.3 lower left) is regionally and simulation dependent. For minimum temperature (Figure 2.3d), the WGNN shows smaller relative errors than does WGN for all the regions. WH shows the lowest relative error in comparison with other WRF simulations for all but two regions—MW and NE. There is not much difference between WCB and WCNB, but they both show far less error than the WGN run. WCB is significantly more accurate for all eight regions than the WGN for minimum temperatures. Since both simulations employ bias correction and nudging, much of the error in the WGN runs for minimum temperature is likely due to the biases in the boundary conditions of that GCM. Nudging does not necessarily improve the model performance in minimum temperature for the WGN.

Figure 2.4 shows relative errors of geopotential height at 500 hPa (Fig. 2.4 upper left), specific humidity at 850 hPa (Fig. 2.4 upper right), and zonal and meridional wind at 850 hPa (Fig. 2.4 lower left and Fig. 2.4 lower right) based on comparisons between WRF simulations and NARR. In general, the WN performs better than all GCM driven simulations for above surface variables. For the rest of the ensemble, the ranks are different depending on region and variable. WGNN and WGN outperform the other GCM driven simulations for many of the regions for low-level winds at the 850 hPa (Figs. 2.4 lower left and lower right), but rank in the bottom two for most of the regions for 500 hPa geopotential height and 850 hPa specific humidity (Figs. 2.4 upper left and upper right)—where WH and WCB/WCNB runs are superior. Depending on the region and variable, there are several instances where nudging has a significant difference between WGN and WGNN in Figure 2.4. For example, nudging helps reducing the relative error in all eight regions for 500 hPa geopotential height (Fig. 2.4 upper left), but yields higher RMSE values for
specific humidity with the exception of the NW (Fig. 2.4 upper right). The use of bias correction does not cause large differences between the two CCSM4 driven WRF runs in relative errors, which have ranks that are mostly near the middle of pack.

Figure 2.5 shows relative errors for more near surface variables (10-m wind, Figs. 2.5 upper left and upper right) and mean sea level pressure (SLP) [Fig. 2.5 lower left]. The GFDL driven runs perform slightly better than the other WRF runs for 10m-winds in several of the regions, especially the meridional wind (v component of the wind). The WGNN generally performs better than WRF_GFDLN for both the U and V component. For SLP, nudging significantly reduces the overall error and WGN shows less relative error than WGNN in all the regions except for the NW—where the difference is minor. Bias correction in the WRF simulations driven by CCSM4 does not result in lower RMSEs in all of the regions for SLP. Where WH ranks in terms of the other GCM simulations for Figures 2.3, 2.4, and 2.5 is different depending on the variable and the region, but there is no discernable trend where/when it consistently outperforms its GCM counterparts for any region or variable. It would provide significant value if one could develop a single index to evaluate individual model performance considering all the variables of interest (Gleckler et al. 2008). This way one could consider more weight on the “better” model than the “worse” model when considering future climate projection. However, different model output is related to different aspect of model physics and/or model setup. Subjectively ranking the model performance based on an average score of all the variables of interest would substantially hide model errors for some aspects (Gleckler et al. 2008). The following section will show that performing well in simulating overall climatology (i.e. relative error) does not necessarily mean the simulation is accurate in terms of model extremes.
2.3.3 Extreme temperature

Another primary motivation for dynamically downscaling climate models is to gain a more comprehensive idea of regional extreme events, and eventually, make more accurate predictions of frequency and intensity of future anomalous climate events. In Sections 2.3.3 – 2.3.5, we discuss the model reliability at forecasting frequency and/or intensity of extreme warm/cold temperature, heat stress, and both single and multi-day heavy precipitation events.

Temperature values that are located in the left/right tail of the PDF curve provides valuable information as to how the model simulation captures the extreme minimum/maximum temperature for a given location. In this study, we calculate the 95% threshold of summer (June, July, and August) maximum and the 5% threshold of winter (December, January, and February) minimum near surface temperature in the reference data and WRF simulations to judge the model’s ability in capturing the extreme high and low temperatures. Figure 2.6 shows the differences in extreme high temperature between model simulations and the observations based on NCA subregion averages. To calculate these values, first, the 95% thresholds for each grid point in both observations and the simulations are calculated, second, the differences are found for each grid point (e.g. Figure 2.2), last, regional average of the differences are displayed. Comparing with the six individual model outputs, the mean and median show the largest bias (cold bias) in extreme high temperature from the reference data set in most of the regions because mean and median filters out the day-to-day variability at each location, which reduces the variance of the PDF curve and acts to smooth out the real extremes. The two CCSM4 driven
WRF simulations (with and without bias correction) also underestimate the extreme maximum temperature for all seven climate regions, but show smaller cold bias than do mean and median. The use of bias correction (WCB) reduces the cold bias over the Northwest, Southwest as well as Southeast regions by 0.5–2°C, but increase the bias over the Midwest and Northeast regions by 1°C in comparison with the run without bias correction (WCNB). Overall, the GFDL driven simulations have warm bias over the Great Plains and Midwest regions and a smaller cold bias than the CCSM4 driven runs. In the two runs where WGN and WCB use both spectral nudging and bias correction, there are large differences in all seven regions, indicating that the GCM used to force the WRF makes a larger difference than the use of bias correction and nudging does for maximum temperature. This is especially true for the Midwest and the Southern Plains where the two runs not only disagree on sign, but the difference in magnitude in the 95% threshold is greater than a 3°C between WGN and WCB. Spectral nudging does improve the model performance in extreme high temperature over most of the regions. For example, the regions NGP, SGP, MW, SE, and NE all have smaller bias in the WGN run by 0.36 – 1.53°C than in the WGNN.

The HadGEM2 driven simulation underestimates the extreme maximum temperature for all of the regions other than the Midwest. WH’s proximity to the observed value for the Northwest and Northern Plains is closer than both WCB and WCNB, but this is not the case when compared to the WGN and WGNN. In the Northeast and Southeast, WH performs much worse than the other GCM driven runs with the threshold being missed by at least 3.5°C in these regions. For the Southeast, WH is actually closer to the error for the mean and median runs than the other GCM simulations. Overall, in the central part of the country (MW, NGP, and SGP), there are
ambiguous signs between the five GCM runs. The GCM driven runs underestimate the intensity of the extreme maximum temperature events for both the eastern (NE and SE) and western regions (NW and SW). The NCEP-R2 driven WRF also underestimate the NW and SW regions thresholds by as much as 2.8°C.

Figure 2.7 shows the differences in extreme low temperature between model simulations and the observations based on NCA subregion averages. Similar to the maximum temperature, the tail in the PDF curve for GCM driven simulations are too close to the mean and underestimate the intensity of extreme cold temperature (with warm bias) for many of the regions. The only region where all the models (except WN) are consistently too cold is the Northwest region. The WH model was the closest to the reference data set for this region when compared to the other GCM runs. Spectral nudging and bias correction affect the GFDL and CCSM4 driven runs differently. Spectral nudging shifts the threshold value in WGN to the right for all the regions, making the extreme cold temperature closer to the observation than WGNN over Northwest, but further from the observation than WGNN over other regions. Different from the effects of nudging, bias correction reduces the bias of extreme cold temperature by as much 2°C in most of the NCA subregions, with the exception of NW and SW. The two simulations (WGN and WCB) that use both of bias correction and nudging have the same sign and similar magnitudes with the exception for the Northeast. This is different from Figure 2.6, where the GCM boundary conditions play a much more significant role in the biases for extreme maximum temperatures between these two runs.
2.3.4 Heat index

In addition to temperature, relative humidity (RH) plays an equally important role on the amount of stress the human body can endure in hot conditions. Thus, heat index (HI) was developed world-wide (Buzan et al. 2014). In this study, we apply one of the heat indices, which is developed by Rothfusz (1990) and using temperature and RH and is applied primarily by the National Weather Service in the United States. Rothfusz (1990) performed a multiple regression analysis on the original table of HI that was computed by Steadman (1979). However, the equation calculated by Rothfusz (1990) is not applicable for all ranges of RH and temperatures. An adjustment of the HI equation is needed (http://www.wpc.ncep.noaa.gov/html/heatindex_equation.shtml). Using the equation of Rothfusz (1990) and the adjustment mentioned above, we are able to compare how well the models capture the right tail of the PDF for HI. The HI value for each location is calculated for both the simulations and the observations, so that the difference can be plotted in gridded form. The maximum temperatures and near surface RH values are used to calculate the HI. Maximum temperature is used instead of mean temperature to help determine if the errors in HI are a result from the known extreme maximum temperature biases discussed for Figure 2.6 or if RH biases could affect the results more significantly in some regions.

Figure 2.8 shows the difference in HI for each simulation’s 95% threshold and the observations. Generally, the WRF_NCEP shows the smallest bias over the entire CONUS, followed by WCB. There is large positive bias for HI in the southern Plains and western CONUS for WGNN, WGN, and WCB that is not evident in the maximum temperature, indicating that the RH is overestimated in those regions. In comparison with WGNN, nudging reduce the bias for HI in the Northwest. In the Southeast, where HI values tend to be the highest during the summer, the
models without bias correction underestimate the 95% threshold. In contrast, the models that use bias correction have a slight overestimation of HI and perform better over Southeast. Overall, there are significant differences in extreme HI values between the two GFDL runs when nudging is applied and the two CCSM4 runs when bias correction is used, especially over Great Plains and western part of the CONUS. When comparing the GFDL and CCSM4 runs that use both bias correction and spectral nudging, there are still a couple important differences. For example, in parts of the Midwest and central Plains, the differences in HI threshold are as high ~6 °C in some locations, indicating the biases in the boundary conditions are still significant in those areas.

2.3.5 Extreme precipitation

Figure 2.9 is the difference between model and observed 95th percentile of daily precipitation. All the precipitation data is filtered to only include precipitation days that record at least 0.01 inches or 0.254 mm to shield from minimum unrealistic values. In comparison with GCM or reanalysis driven WRF runs, the mean and median of the six WRF simulations show significant dry bias in extreme precipitation. WN shows wet bias over Great Plains for not only the daily mean precipitation (as shown in Figure 2.3 upper left), but also the extreme precipitation. The reason can be due to the short spin-up time and/or the strong nudging strength applied in this simulation (WK14), which has been modified in the GCM driven runs. The five GCM driven simulations underestimate the extreme precipitation by 3.5-8.3 mm over Southeast. This is likely because these simulations lack the ability to capture small scale convection that takes place regularly in this region that cannot be fully resolved with a 12-km horizontal resolution. In comparison with WGNN, WGN significantly reduces the model bias in extreme precipitation
over Great Plains and Midwest. In comparison with WCNB, WCB significantly reduce the bias in extreme precipitation over all the regions except NW and NE. The WH run, which does not use bias correction or spectral nudging, shows much larger dry bias over most of the regions with the exception of the Northeast. It is worth mentioning that, knowing how the model performs in terms of relative error does not effectively forecast how the model predicts the extremes from the reference data. For example, the RMSE for the two models (WGN and WCBC) that use both bias correction and nudging is higher for the North Plains, South Plains, and Midwest regions, but they both show smaller bias in the same region for extreme precipitation than the other simulations.

Extreme precipitation events occur frequently when daily precipitation values are to the right of the 95% threshold in the PDF curve for multiple consecutive days (Janssen et al. 2014). In many cases, the heaviest precipitation events occur because a storm system is stagnant over similar areas for consecutive days (e.g. Francis and Vavrus 2012). While many other environmental factors determine the extent and magnitude of flash floods (Montz and Gruntfest 2002), the best these models can do is attempt to improve on forecasting frequency of long-term extreme precipitation events. For this reason, in addition to daily precipitation extremes, this study also analyzes the model’s ability to simulate major precipitation events for 2 and 3-day storm totals. Figure 2.10 shows the differences in frequency of 99% threshold for 2-consecutive day precipitation. By finding the 99% average regional threshold for 2-day precipitation events from the observations, the difference in number of times the model predicts this occurring shows how well the simulation handles storm system movement across the U.S. This is calculated by ranking all the total 2-day precipitation events at each location that experienced at least a trace of
precipitation and calculating the number of occurrences greater than the regional observed threshold for the whole decade in each of the six simulations. Figure 2.10 shows the number of times the model output was greater than the regionally averaged 99% threshold in the reference data and is standardized by subtracting the number of events in the reference data at each grid point that were greater than the 99% threshold. The reason difference is calculated is because, depending on the location or region, there may be a high frequency of precipitation days meaning it is expected that there would be more 99% events for these locations over the course of a decade. The 2-day and 3-day results for this metric are similar enough that only the difference in 2-day precipitation extremes is presented in this study.

The GCM driven simulations tend to underestimate the frequency of 99% events along the Gulf of Mexico in the Southeast and along the West Coast. Other regions, such as the Midwest, have differences in regional signs for each of the 6 simulations. Bias correction and nudging used together tend to slow storm system movement across the Plains and Midwest indicated by the positive 2-day precipitation positive anomalies. Without nudging or bias correcting the boundary conditions, the WRF simulations move storm systems across the central U.S. faster leading to fewer events that meet the observed 99% threshold criteria for that location. The addition of nudging in the GFDL runs enhances a strong positive bias in a large area of the Southwest as well as through most of the Northern Plain states that is not present in the no nudging run. The WGN run does reduce high negative bias in the WGNN in much of the Midwest. To a lesser extent, bias correction also reduces this same negative anomaly for the CCSM runs in most of the Midwest.
2.3.6 Large-scale circulations

To further understand the impacts of bias correction and spectral nudging on the simulated precipitation, in this section, we investigate the large-scale circulation to figure out the dominant physical reasons for the differences between WCB and WCNB as well as between WGN and WGNN. The regions that have the largest differences in extreme precipitation occurs in the Plains, Midwest, and Southeast. Since the warm season is when most of the extreme events take place in these regions because of the Great Plains Low-Level Jet (GPLLJ) and large scale flow (Hitchens et al. 2013), we focus on summer average 500 hpa geopotential height and 850 hpa V wind, or the north-south component of the wind vector, as shown in Figures 2.11 and 2.12. The north-south winds that are usually important for large daily precipitation events in these regions because it not only allows MCSs to grow upscale during the summer due to the southerly flow, but also provides an overall source of moisture to these regions. The figures show the average anomaly when WRF simulation is subtracted from the observations (here we use NARR) at each grid point. As shown in Figure 2.11, the 500 hPa geopotential height anomalies between the WGNN and WGN runs are quite different. WGNN produces significant ridging over much of the western third of the country and a trough centered most likely in southern Canada, but extending into the Northeast. The WGN run has an overall positive geopotential height bias over the country, but the strongest positive anomalies are located in the Southeast along the Georgia and Carolina coast. This strong positive anomaly in the WGN over the Southeast leads to the drier bias over Southeast as shown in Figure 2.9. In the three central regions, the key difference is the large area of strong positive anomalies over the West Coast, Rocky Mountains, and extending into the western Plain states in the WGNN that is not as strong as that in WGN run. As shown in
Figure 2.11, in the WGNN simulation, the large negative anomaly, indicative of an overestimation of north wind strength, that persist in these regions is most certainly the primary driver of large negative anomalies for extreme precipitation events in the Plains and Midwest for this run. If the simulation has winds that move from north to south more than what is observed, like in the WGNN run, the air mass in place will tend to be drier Canadian air as opposed to the typical moist Gulf of Mexico air that would be the product of a southerly wind. In contrast, WGN captures the low-level moisture transport across these regions more realistically than WGNN, with the exception of a positive anomaly across the Texas coast, and as a result is much close to the observed values.

Figure 2.12 shows a similar dependency of atmospheric circulation on extreme precipitation for CCSM4 driven WRF runs. While both runs have positive height anomalies across the Northwest and NPlains, WCNB extends the ridge further south and east into the SPlains. These positive height anomalies are indicative of low-level anticyclonic circulation that results in a north wind on the eastern side of this circulation in both simulations across the Plain states on the lee side of the Rockies. However, since the positive anomaly is shifted to the southwest in the WCNB, the northerly winds extend further to east. Equally important to those anomalies, the WCB has a positive height anomaly centered in the Gulf of Mexico that brings southerly flow of moist air into Texas and eventually into parts of the Midwest and Southeast. This is the reason that the SPlains and Southeast have larger dry bias for WCNB but smaller dry bias in WCB. In addition, because of the southerly flow which brings the moisture from Gulf of Mexico, WCB shows smaller bias than does WCNB which has significant dry bias over the Midwest and NPlains.
2.4 Discussion and Summary

This study provides analysis for six WRF simulations by ranking their performance when evaluated based on measures of relative error and extreme climate events. Ranking the models based on relative error (Figures 2.3-2.5) allows future researchers to make informed decisions on which type of boundary conditions and regional model settings are needed to achieve the most ideal results. Few downscaling projects have compared CMIP5 GCM-driven dynamical downscaled model performance for variables other than surface temperature and precipitation, especially as an ensemble (Fowler et al. 2007; Lee et al. 2014). This study evaluates both lower and upper atmospheric variables and the dominant physics for extreme events, which can better inform researchers and users of the model results. The variables we study can aid in reconstructing dynamical profiles for the atmosphere to better understand their precipitation and temperature regional biases (see Supplementary Figures A1 and A2 for the regional table of relative error rankings). Results show that when modeling climate extremes, the use of DD creates substantial “added value” or improved ability compared to low resolution GCM data. Model setup (i.e. bias correction and nudging) can be more important to predictability and biases than the GCM boundary conditions; a simulation that has low RMSE does not necessarily mean it is efficient at modeling extreme event frequency or day-to-day fluctuations in these variables. In addition, our results show that many variables have the largest errors for surface variables in the wettest and driest regions of the continental United States. As mentioned above, high precipitation regions, such as the Southeast, yield higher errors because of the dominance of convective processes in these regions, which is challenging to predict at this resolution (Bryan...
Similarly, drier regions have been shown to have greater errors or biases due to small scale processes that are hard to capture using downscaling techniques (Fowler et al. 2007).

One of the challenges in a study like this is to compare the model output to best reference data set available, but in reality, the “ground truth” for variables, such as precipitation, often have sources of biases and error themselves (Cosgrove et al. 2003). We compare the relative error when using PRSIM (Fig. 2.13 left panel) and NARR (Fig. 2.13 right panel) as the reference data sets for monthly precipitation. Overall, many of the regional ranks of the models are mostly similar between the two, but there are several cases where the difference in the magnitude of relative error is as high as 25% (e.g. Northeast region for GFDL using nudging and Southwest region for HadGEM2). Therefore, using multiple reference data sets that yield multiple results for errors for both relative error and extremes for a historical period, provides a more comprehensive understanding of the model performance. Understanding where the simulations fail, or do not closely match the observations, is the most important feature of this research and is vital to understanding future projections of climate extremes (Ekström et al. 2005). Similarly, the WRF model itself and the physical schemes used introduces an additional set of regional biases (e.g. Ruiz et al. 2010; Jankov et al. 2005; Ries and Schlünzen 2009; Cheng and Steenburgh 2005; Aligo et al. 2007). All of these studies discuss the importance of WRF configuration and that the ideal settings will have high temporal and geographical dependence based on the test variables. The large domain of this study evaluates regions with varying topography and climates making the choice in physical parameterizations for the WRF difficult despite the sensitivity experiments tested in WK14.
If there is a known overall bias in the dynamically downscaled method for a specific region in all members of the ensemble, that can now be accounted for when making projections of future climate change. As we mentioned above, each of these GCM’s raw data have a different climate sensitivity. This is why the use of our ensemble could prove valuable at making analyses of uncertainties in projected extreme values. Since most of the uncertainty in future climate comes from choices such as the climate model used and the emission scenario (Déqué et al. 2007), our multi-climate model ensemble, while employing bias correction and spectral nudging, can prove valuable at analyzing the uncertainties in future climate extremes. In this study, we find the regions each metric most accurately is represented in the models. We show where both bias correction and spectral nudging can be beneficial using dynamical downscaling as well as which of the three GCMs tested have model biases. The primary focus was on the shortcomings of the historical period in the model, but the information gathered from this analysis will be most useful in our future work on climate projections.

The high-resolution modeling studies provide stakeholders and the public with a knowledge of the uncertainties on a range of climate indicators, including assessing effect on local hydrological processes; surface temperature changes and heat stress on humans in a warmer climate (Fowler et al. 2007 and Buzan et al. 2014). Understanding these strengths and weakness of using dynamic downscaling methods is an important step in finding a way to access the risks of future climate. These types of ensemble downscaling studies can provide an evaluation of future uncertainties in societal impacts at spatial scales of interest to the impact assessment and adaptation community.
(Fowler et al. 2007). Wang and Kotamarthi (2015) used two members of this ensemble to discuss precipitation changes over the continental United States by first comparing the model’s historical accuracy. Chapters 3 and 4 of this dissertation build on those projections by analyzing temperature and precipitation extremes with this ensemble as well as potential societal impacts associated with those changes.
Figure 2.1. Difference between model and observed precipitation percentiles (model-observed) in the PDF curve for WRF driven GCM simulations and GCM raw data. Regions shown from top to bottom and left to right are Midwest, Southeast, Northwest, and SPlains (SGP).
Figure 2.2: Difference between the observed 95th percentile for daily maximum temperature (°C) and each of the 5 RCM simulations as well as the 3 GCMs used as boundary conditions to perform the downscaling. In addition, the RCM and GCM average is provided in the bottom row.
Figure 2.3: Relative error of the RMSE for surface variables compared to observed gridded values: Daily precipitation (upper left); Daily mean temperature (upper right); Daily maximum temperature (lower left); Daily minimum temperature (lower right)
Figure 2.4: Relative error of the RMSE for above surface variables compared to North American Regional Reanalysis (NARR): 500 hPa geopotential height (upper left); 850 hPa specific humidity (upper right); 850 hPa U component (lower left); 850 hPa V component (lower right).
Figure 2.5: Relative error of the RMSE values for: 10-m U wind (upper left), and 10-m V wind (upper right); and sea-level pressure (lower left),
Figure 2.6: Average subregional difference (model-observations) in 95% threshold of daily maximum summer (June, July, and August) temperature (°C).
Figure 2.7: Average subregional difference (model-observations) in 5% threshold of daily minimum winter (December, January, and February) temperature (°C)
Figure 2.8: Difference (model – observations) in 95% heat index threshold (in °C) between the 6 model simulations and observation.
Figure 2.9: Average regional difference (model – observations) in 95% threshold extreme precipitation events between the models and observations. Values for each region are in millimeters.
Figure 2.10: Differences (model - observations) in the frequency of 99% threshold events for 2-day precipitation events. In order to be categorized as an “event” the grid point must experience at least a trace of precipitation for 2 consecutive days. To standardize these values, the difference between the number of 99% events in the observations is subtracted from the model values.
Figure 2.11: Figure 10: Difference between model and observed 500 hPa geopotential height (m) averaged across June, July, and August (JJA).
Figure 2.12: the same as Figure 10, expect for WRF runs driven by CCSM4 with and without bias correction.
Figure 2.13: RMSE figures for total monthly rainfall when compared to two reference data sets:

(a) PRISM (left) (b) NARR (right)
<table>
<thead>
<tr>
<th>Lateral boundary conditions</th>
<th>WRF Simulation</th>
<th>Microphysics</th>
<th>Spin-up</th>
<th>Spectral nudging strength</th>
<th>Bias correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCEP-R2</td>
<td>WN</td>
<td>WSM6</td>
<td>1 day</td>
<td>Yes ($3 \times 10^{-4}$ s$^{-1}$)</td>
<td>No</td>
</tr>
<tr>
<td>CCSM4</td>
<td>WCNB</td>
<td>Morrison</td>
<td>1 year</td>
<td>Yes ($3 \times 10^{-5}$ s$^{-1}$)</td>
<td>No</td>
</tr>
<tr>
<td>CCSM4</td>
<td>WCB</td>
<td>Morrison</td>
<td>1 year</td>
<td>Yes ($3 \times 10^{-5}$ s$^{-1}$)</td>
<td>Yes</td>
</tr>
<tr>
<td>GFDL-ESM2G</td>
<td>WGNN</td>
<td>Morrison</td>
<td>1 year</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>GFDL-ESM2G</td>
<td>WGN</td>
<td>Morrison</td>
<td>1 year</td>
<td>Yes ($3 \times 10^{-5}$ s$^{-1}$)</td>
<td>Yes</td>
</tr>
<tr>
<td>HadGEM2-ES</td>
<td>WH</td>
<td>Morrison</td>
<td>1 year</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2.1 The lateral boundary conditions that drive the six WRF model runs as well as information about microphysics, spin-up time, spectral nudging, and bias correction that are applied.
CHAPTER 3: HIGH RESOLUTION DYNAMICAL DOWNSCALING ENSEMBLE
PROJECTIONS OF FUTURE EXTREME TEMPERATURE DISTRIBUTIONS FOR
THE UNITED STATES

3.1 Introduction

In Chapter 3 of this dissertation, an ensemble of five dynamical downscaled RCM projections that incorporate initial and boundary conditions from three different GCMs are employed to analyze changes to regional temperature distributions across the United States. Three key questions are investigated in Chapter 3: (1) How is the overall morphology of temperature distributions affected by a changing climate on a seasonal temporal scale and on a regional spatial scale? (2) How are the tails of the temperature distributions affected by a changing climate? (3) How do the changes in morphology and tails of temperature distribution affect metrics related to agricultural demands and oppressive heat waves?

3.2 Methodology

3.2.1 Data and model description

Building off of the analysis from Wang and Kotamarthi (2014), Wang and Kotamarthi (2015), and Zobel et al. (2018) [Chapter 2], this study discusses the future projections for our dynamically downscaled ensemble. The WRF model version 3.3.1 (Skamarock et al. 2008) is applied at a horizontal resolution of 12 km.

(600 west-east × 515 south-north grid points over most of North America) (Figure 1 in Wang and Kotamarthi, 2014). Wang and Kotamarthi (2014) also apply the same physical parameterizations as this study, which includes the Grell-Devenyi convective parameterization (Grell and Devenyi 2002), the Yonsei University planetary boundary layer scheme (Hong et al. 2006; Noh et al. 2003), the Noah land surface model (Chen and Dudhia 2001), the Morrison microphysics scheme (Morrison et al. 2009), as well as the longwave and shortwave radiative schemes of the Rapid Radiation Transfer Model for GCM applications (http://rtweb.aer.com) (Iacono et al. 2008). More details can be found in Section 2.2 of this dissertation.

The three fully coupled GCMs used in this study as boundary conditions are the Community Climate System Model 4 (CCSM4) developed by National Center for Atmospheric Research, United States (Gent et al. 2011), the Geophysical Fluid Dynamics Laboratory Earth System Model 2 (GFDL-ESM2G) developed by NOAA/Geophysical Fluid Dynamics Laboratory, United States (Donner et al. 2011), and the Hadley Centre Global Environment Model version 2 (HadGEM2-ES) developed by Met Office Hadley Centre, United Kingdom (Jones et al. 2011). These three GCMs represent a range of climate sensitivities that encompasses most of the CMIP5 GCMs when projecting future temperature changes. For example, GFDL-ESM2G yields a global mean temperature change of 2.38 °C with a doubling of CO₂. HadGEM2-ES has the highest sensitivity to a doubling of CO₂ with a projected temperature increase of 4.5 °C. CCSM4 responds near average among more than 30 GCMs with an increase of about 2.9 °C (Sherwood et al. 2014).
Future projections for the CMIP5 models simulated four separate future climate scenarios and these scenarios are defined by using different Representative Concentration Pathways (RCPs) (IPCC AR5). Here, we use two RCP scenarios: RCP 8.5 assumes the continued heavy use of fossil fuels at a similar, or greater, rate as current concentrations of CO₂ and other GHGs through the end of the century leading to a radiative forcing of 8.5 W/m² by 2100 (Riahi et al. 2011); RCP 4.5 scenario is a pathway to stabilize radiative forcing at 4.5 W/m² by 2100 and implies significant reduction in concentrations from fossil fuel use by the end of the century. With this scenario, GHG concentrations in the atmosphere peak in the 2040s before stabilizing toward the end of the century (Thomson et al. 2011). We analyze these two scenarios for a mid-century decade (2045-2054) and a late-century decade (2085-2094) and allow one year spin-up time for each of these decades (2044 and 2084 respectively). Additionally, to investigate the impacts of spectral nudging on future projections of temperature distribution, we run two simulations using GFDL-ESM2G as boundary conditions – one with spectral nudging and one without. To understand the effects of bias-correction on temperature distribution in the future, we also run two simulations using CCSM4 as boundary conditions – one with bias correction and one without (see Zobel et al. 2018 for more details). Table 3.1 lists the sets of simulations that are conducted for this analysis. For the remainder of this analysis, each scenario is characterized by the RCP scenario and the decade during which it is being analyzed. For example, the late-century RCP 8.5 and the mid-century RCP 4.5 scenarios are abbreviated as R8Y8 and R4Y4, respectively. The regional analysis uses the National Climate Assessment (NCA) definition for 7 climate subregions throughout the contiguous United States (CONUS) (Melillo et al. 2014). These regions and their abbreviations are Northwest (NW), Northern Great Plains (NGP),
3.2.2 Seasonal PDFs

Probability density functions (PDFs) for daily maximum temperature (TMAX) and minimum temperature (TMIN) are built for each of the five simulations, four scenarios, four seasons, and all seven regions. TMAX (TMIN is used in the Winter) is modeled daily by the WRF simulations based on a time-step of 40 seconds. The calculation for this metric is motivated by Hansen et al. (2006). PDF curves are calculated using 3 steps. First, the decadal average and standard deviation (SD) of TMAX at each grid point is calculated for the historical decade. We do this by grid point because the historical values for mean and SD vary throughout even the smaller regions. Using the Midwest and the historical WCB simulation as an example, we found that there was a regional North-South gradient for average summer temperature of about 10 °F and a North-South SD gradient of 3 with the Northern states having the lowest average temperature and highest SD. Second, we calculate the normalized TMAX value (\( \tilde{T}_{i,j,t} \)) by taking the difference between each TMAX value at a given longitude (x), latitude (y), and time (t) (\( T_{i,j,t} \)) and the historical mean at that grid point (\( \bar{T}_{i,j} \)). This change in temperature is then normalized to the modeled temperature variability by dividing with the historical standard deviation of each grid point (\( \frac{1}{N} \sum_{t=1}^{N} (T_{i,j,t} - \bar{T}_{i,j})^2 \)). For each of the five sets of boundary conditions, we calculate this metric for all four scenarios and the historical simulations as well as observed TMAX values (see Equation 1). In theory, this will generate a Gaussian, or quasi-Gaussian, PDF curve around 0 for both historical simulations and observations of regional maximum temperature values. Finally, PDF curves for all the decadal simulations including
historical and future periods are plotted. This process was repeated for the Summer (JJA), Spring (MAM), and Fall (SON) seasons. The same calculations are performed for TMIN values in Winter (DJF). While there were noticeable changes to the projected temperature distributions in Fall and Spring when compared to their historical counterparts, this paper focus primarily on extreme warm temperature in summer and extreme cold temperature in winter.

**Equation 1:**

\[
T_{i,j,t} = \frac{T_{i,j,t} - T_{i,j}}{\sqrt{\frac{1}{N} \sum_{t=1}^{N} (T_{i,j,t} - T_{i,j})^2}}
\]

It should also be noted that these PDFs are built encompassing all the grid points from each season for the entire decade. To quantify some of the differences in the PDF curves, especially within the tails of the PDF curve, we calculate the historical percentile of interest (e.g. 95\(^{th}\) percentile is used for TMAX and the 25\(^{th}\) percentile is used for Winter TMIN) at each grid point. We then find the percentile this value falls on the PDF curves for each of the future scenarios. For example, if we compare the historical 95\(^{th}\) percentile to the future simulations, it tells us how much more likely (or unlikely) an extreme TMAX value, which only occur 5% of the time historically, is for the future simulations. We calculate this change at each grid point, but the values given for this metric in Section 3 are regionally averaged, so we can present one concise value for each region. We chose to use the 25\(^{th}\) percentile for Winter TMIN because most of the future scenarios would have \(~0\ mass in the PDF curve less than historical 5\(^{th}\) percentile.
3.2.3 Statistical Analysis

To understand the robustness of the changes in extreme temperature from historical to future period, we conduct statistical significance tests on the means of the PDF curves. As Ruff and Neelin (2012) and Loikith and Neelin (2015) suggested, the PDF curves of daily temperature are not always normally distributed. In particular, Ruff and Neelin (2012) found as the climate warms, individual station data projects modest or large departures from Gaussianity. While our regional analysis does not yield as large of a departure as the data presented in Ruff and Neelin (2012), we do have several simulations that produce PDF curves that are not normally distributed. To remedy this, as stated by the central limit theorem, we employ bootstrap technique to resample the means of all the non-normally distribution, which results in a normal distribution and allows for student t-test to be performed on the new data that retains the same mean as the original data (Wilks, 2011). To avoid possible false significance when conducting student t-tests on very large data sets, we test the significance of the resampled data with varying degrees of freedom by using 20, 50, and 100 (Wilks, 2011). This procedure gives us reliable indications of the statistical test in the PDF curves.

3.2.4 Exceedance Rate of 95 °F Days

We define exceedance rate of 95°F days as days where TMAX is equal to or higher than 95°F. To calculate the changes of this metric in future, we take the difference between future and historical simulations at each latitude and longitude. According to the third National Climate Assessment (NCA3), temperatures that exceed 95 °F increase frequency of negative health impacts and decrease agricultural crop yields (Melillo et al. 2014). Schlenker and Roberts (2009) found that yield rates for corn, soybeans, and cotton increase as temperatures approach 84 °F, 86
°F, and 90 °F respectively. However, temperatures greater than these thresholds act to drastically decrease crop yields. Using days greater than 95 °F as the threshold for this analysis, we can calculate the increase in days that will diminish the potential crop yields for corn, soybeans, and cotton based on the results from Schlenker and Roberts (2009).

3.2.5 Frequency of Heat Waves
Heat waves have been shown to significantly increase the mortality rate in the United States, especially cities within the Midwest and the Northeast because of the high population centers and the infrequent nature of excessive heat events (Patz et al. 2005). Similarly, Hajat et al. (2014) found a sizable increase in mortalities due to excessive heat-related events in a future climate projection in the United Kingdom, which is a location that typically experiences a much cooler climate compared to most of the United States. Our analysis aims to quantify the difference in frequencies of heat waves between the future and historical simulations. The definition of what constitutes a “heat wave” does vary in the literature. Here, we define a heat wave using a percentile based approach similar to the metric used by Meehl and Tabaldí (2004) and Kunkel et al. (2010), but we consider not only temperature but also humidity as indices of a heat wave. To start, we define a heat wave as any 3-day event where the maximum heat index (HI) is greater than the 95th percentile of HI at each grid point. To calculate HI, we used daily TMAX and daily average relative humidity (Zobel et al. 2018). The 95th percentile is calculated based on the HI values from April through September. Once we calculate the historical 95th percentile for HI, we count the number of times a heat wave occurred at a given location. One caveat of this calculation is that, for example, if a location experienced a 6-day heat wave event, then the metric would count four 3-day-heat-wave events. This would slightly inflate the frequencies of
heat waves in our analysis compared to reality when considering shorter temporal thresholds. To investigate this inflation, we use different temporal thresholds (3, 5, and 10 days) to determine if the differences in heat waves occurs primarily because of longer heat waves, more frequent heat waves, or a combination of the two.

3.2.6 Length of the Frost-Free Season

Similar to the heat wave metric, the calculation used to quantify the length of the growing season is geographically dependent. The NCA3 defines the length of the growing season by calculating the number of days between the last below 32 °F TMIN value each spring until the first day where TMIN is below 32 °F in the following fall/winter (i.e. “Frost free season”) (Melillo et al. 2014). Like NCA3, for our analysis, we use the frost-free season to define the length of the growing season. To calculate the difference in length of the growing season between historical and future simulations, the longest period of consecutive days where TMIN values are above 32 °F is found at each grid point and then averaged over the decade to generate the difference in growing season per year. In some locations, the end of this period takes place within the Winter during the following year and those days are counted toward the previous year.

Research that focuses primarily on this metric finds that the location and type of crop that is grown in that region will determine how the growing season is defined. Studies show that there are many other factors that need to be considered when addressing regional agriculture growing season. Additionally, these studies find that regions with more temperate climates or multiple planting season must be considered differently than using temperature thresholds (e.g. Waha et al. 2012 and Sacks et al. 2010). Therefore, in the southern regions of the U.S. using the length of
the frost-free season may not adequately define the true length of the growing season, but for the Midwest and other northern regions this metric is an appropriate approximation.

3.3 Results

3.3.1 Summer PDF curves

Figure 3.1 displays the summer TMAX PDF curves for 3 of the 7 NCA regions: Southeast, Northeast, and Midwest. Southwest and Southern Great Plains have similar changes to Southeast (Figure 3.1a) under a warmer climate. Northwest and Northern Great Plains experience similar changes to the projections for Northeast (Figure 3.1b) and Midwest (Figure 3.1c), respectively.

To start, there is a clear difference between the R8Y8 scenario and the other 3 scenarios for both changes to the model mean and the extreme TMAX distributions. Here we focus on discussing R8Y8 (blue curves in Figure 3.1) and R4Y8 (purple curves in Figure 3.1) to understand how different RCP scenarios affect climate change by the end of 21st century. Other scenarios are also investigated and their statistics are presented in Figure 3.1. For R8Y8, the shift in the mean is statistically significant at a 0.001 significance level for all 7 regions, all 5 simulations, and for the 3 different sample sizes tested, according to the statistical analysis described in Section 3.2.3.

In the Southeast, Southwest, and Southern Great Plain regions, there is a projected shift in the median of at least 1.5 standard deviations to the right and as high as ~2 standard deviations for the WCB and WCNB simulations. Using the average standard deviation value for the region, this shift in the median equates to 4.2 to 4.5 °F of overall warming. The historical 95th percentile for each of the three regions and five simulations are denoted with a vertical black line. For the R8Y8 scenario over Southeast, the historical 95th percentile value ranges between the 39th (WCB, WCNB, and WH) to the 51st (WGN) percentiles for R8Y8 scenario, and 73rd (WCNB and WH)
to the 86\textsuperscript{th} (WGN) percentile for R4Y8 scenarios. In other words, under R8Y8 (R4Y8) scenario, for the WCB, WCNB, WH and WGN simulations, 49-61\% (14-27\%) of the summer days will be hotter than the historical 95\textsuperscript{th} percentile which is only 5\% of the summer days. Southwest is projected to have similar experience to the Southeast. In Northeast, the historical 95\textsuperscript{th} percentile ranges between 48\textsuperscript{th} (WCB and WCNB) and 62\textsuperscript{nd} (WGNN) percentile for R8Y8 scenario, and between 74\textsuperscript{th} (WCNB and WH) and 89\textsuperscript{th} (WGN and WGNN) percentile for R4Y8 scenario, meaning that 38-52\% (11-26\%) of the summer days are projected by R8Y8 (R4Y8) to be hotter than the historical 95\textsuperscript{th} percentile. In the Midwest, the historical 95\textsuperscript{th} percentile ranges between 56\textsuperscript{th} (WCB and WCNB) and 87\textsuperscript{th} (WGN and WH) percentile under R8Y8 scenario, indicating that 44\% (13\%) of the summer days are projected by the WCB and WCNB (the WGN and WH) to be hotter than the historical 95\textsuperscript{th} percentile. For scenario R4Y8, the WCNB simulation projects that 32\% of the summer days will be hotter than the historical 95\textsuperscript{th} percentile, while other simulations do not find substantial changes.

There are some regional differences in the magnitude in the overall warming and how frequent extreme temperature events are projected to occur, but overall the range of regional difference between the “coldest” and “warmest” model is similar. For example, in the Southeast regions, the ensemble agreement is high and the simulations projected large right shift in the median and increasing frequency of extreme TMAX days. In addition to the significant warm shift, these regions also project increased variability in the PDFs (i.e. less days near the median and more mass in the tails), which further increase the frequency of extreme events. One way to quantify the variability is by measuring the interquartile range (IQR, Wilks, 2011), defined by the difference between 75th and 25th percentiles. If the IQR in a future period is larger (smaller)
than that in the historical period, then the temperature distribution projected by the future scenarios has increased (decreased) variability in the PDF curve. For the Southeast, the IQR increases by 13-23% for the R8Y8 scenario. Despite the model agreement that takes place in the projections for the southern regions, in the Midwest, the right-moving shift in the median for the WCB (warmest) is 2 times greater than the WGN (coolest). This is largely due to the difference in shape of the PDF curve. There is a discrepancy in how the variability of the PDF change with significant background warming for the R8Y8 scenario. The WCB, WCNB, and WGNN models all simulate morphology similar to that in the in the Southeast, but the WH and WGN increase the frequency of days concentrated around the median in the future scenario (i.e. decreasing variability). In the Midwest, the interquartile range decreases for the WH and WGN when compared to its historical counterpart, but the other 3 simulations increase this variability by at least 12%.

It should be noted that the WGN model was the most accurate at simulating the Midwest observed historical variability and subsequently should be given more weight when considering these future projections (e.g. Bukovsky, 2012). The decreased variability noted above also occurred in the Northeast for the WGN and WH simulations, but the background warming for this region caused the median to shift enough that the historical 95th percentile still occurred close to the 50th percentile. This result is in good agreement with Loikith and Neelin (2015) where they concluded that shorter-than-Gaussian high-side tails have the potential to greatly increase in the frequency of extreme values if the shift in the PDF mean is large enough. For the Midwest and Northeast, the WGN and WH simulations experience a shorter-than-Gaussian high-side tail, but the background warming in the mean is large enough in the Northeast to still
yielded greater increases in the TMAX extremes. In addition, Huang et al. (2016) conducted an experiment using GCM output to project changes in the extremes under a future climate. Their results suggest that the difference in summer extremes are due largely to a shifting mean and they find very little change in the shape and variance of the curve. These conclusions do not apply to Summer TMAX over all the regions we present here. It is likely the increased horizontal resolution resulted in additional small-scale processes to be observed in our simulations when compared to those run for Huang et al. (2016) and thus created larger differences in how the simulations represented the variance in the PDF curves presented here.

### 3.3.2 Exceedance Rate of 95 °F Days

In this section and in section 3.3.3, we discuss additional metrics to describe in more depth what types of extreme climate events can be expected with the PDF curves shown in Figures 3.1. Figure 3.2 display the difference in occurrences per year between 1995-2004 and 2045-2054 under RCP4.5 and RCP8.5 for the exceedance rate of the 95°F threshold. While these events are not exclusive to the summer months, the historical frequency of these events occurring within a summer month was at least 85% for 6 out of the 7 regions (Southwest being the exception). It is apparent that as the magnitude of the GHG concentration increases, the coverage of this type of extreme event expands throughout the country. As expected, the changes in these events are primary confined to regions that already experience the hottest overall climate in the CONUS. As shown in Figure 3.2, even with the two mid-century scenarios, there is high confidence there will be close to an additional ~30 days per year of these events for the Southern Plain states and parts of Southwest and Southeast. There are some small areas of decreasing frequency of these events in the upper Midwest with the WH simulation and GFDL driven simulations. It should be
noted that with all four scenarios the grid point with the largest decrease is 4 days per year, so despite some simulations having a minor decline in 95°F days, the overwhelming trend is in the positive direction.

Similar to the summer PDF plots, the R8Y8 scenario with TMAX equal to and higher than 95 °F (Figure 3.3 right column) is significantly different from the other three future scenarios. For R4Y8 scenario (Figure 3.3 left), there are only small areas in the most southern parts of the United States where the increase in this threshold is projected to be at least 50 days per year. For R8Y8 scenario, this same increase is as far north as the Dakotas for all five simulations. Additionally, on average, our ensemble brings the 30 day/year contour into the southern Ohio Valley. Therefore, the reduction of GHG concentration in RCP 4.5 scenario is projected to decrease this type of event by at least 300 days throughout the 10-year period (i.e. 30 days per year). For R8Y8 scenario, there is fairly good agreement with all 5 simulations with only minor differences in the magnitude of the changes in occurrence of ≥95°F days. Areas that the RCM simulations predicted ~30-day increase in the other 3 scenarios are predicted to experience at least 60 more days per year with TMAX higher than 95°F. The most significant change in this scenario is the overall coverage of red and dark red (i.e. 24–54+ day increases) that now takes up most of the CONUS, especially for areas at relatively low elevations. Most of Northwest, Northern Great Plains, Midwest, and southern parts of Northeast that experienced this type of event less than 3-5 times per year historically, have TMAX values that exceed 95°F at least 20-30 days per year under R8Y8 scenario. In these four regions, as well as other locations, 95°F days are no longer completely exclusive to June-August months. They occur more frequently in late spring and early fall for this scenario.
3.3.3 Frequency of Heat Waves

The primary focus for this section is on R4Y8 and R8Y8 scenarios. These two scenarios had significantly higher frequencies of heat wave events than the mid-century scenarios. Figure 3.4 (right column) shows the increase in 3-day and 5-day heat wave frequency for the R4Y8 and R8Y8 scenario, respectively. We use a 5-day threshold for this scenario to minimize some of the inflated numbers of a 3-day values. On average, the southern regions experience 30-40 more 5-day heat waves under R8Y8 when compared to 3-day heat waves under R4Y8 scenario. For reference, no region, other than parts of the Southwest, has an average increase of more than four 5-day heat waves under the R4Y8 scenario (figure not shown). Overall, in the northern regions, the historical decade for the WH, WCNB, WGNN simulations has less than five total 5-day heat wave events occurring throughout the entire decade (Southwest excluded), while it occurs at least a few times every year under R8Y8 scenario. Additionally, the frequency of 10-day heat wave events is as high as 30 in Southeast and western regions under R8Y8 scenario, while there is no occurrence of heat wave that lasted 10 days in the historical simulations using this metric except over desert Southwest, extreme southern Texas and southern Florida. Therefore, under R8Y8 scenario, an entirely new event begins to occur in the Midwest, Northern Great Plains, Northwest, and Northeast regions. We also find that the spatial coverage for the largest increases in heat wave frequency in our simulations is consistent to the results presented in Kunkel et al. (2010), although there are slight differences in the definition of future climate scenario and heat wave metrics which makes an exact comparison challenging.
The R4Y4 and R8Y4 scenarios (not shown) experienced an increase of at least 2 or 4 events per year on average for most of the country. Despite the R8Y4 and R4Y8 scenarios having similar concentrations of GHGs and comparable PDF curves during the warm season, there is increased probability of a 3-day heat wave occurring in the R4Y8 simulation (Figure 3.4 left column). The south and western regions are where the most significant changes are expected to take place. Overall, in the R4Y8 simulation, there were about 20+ additional 3-day heat wave events for nearly the entire Southwest compared to historical simulations for the WH and WCNB simulations when compared to the R8Y4 scenario. The Northeast and Midwest are projected to experience similar heat wave frequencies compared to R4Y4 based on GFDL driven simulation; however, the WH and WCNB simulations increase this to 10 or more in many locations throughout these regions. Overall, the R4Y4, R8Y4, and R8Y4 scenarios simulate a moderate increase in 3-day heat wave frequency with the R4Y8 scenario projecting the greatest increase.

We found that the sensitivity to the temporal threshold for heat waves affect the values for 3-day heat waves for the mid-century scenarios, but not as significantly in the R4Y8 scenario. This result indicates that more prolonged heat waves are projected for the R4Y8 scenario than in the two mid-century scenarios and possibly is why we see a larger increase in 3-day heat waves mentioned above.

To put the values discussed above into context, the Chicago heat wave in 1995, when over 200 people died due to heat related issues in Chicago alone, was a 4-day event and the heat index values ranged about 100-115°F (Semenza, 1996). The heat index range of the historical 95th percentile for Chicago area ranges between 100-106 °F in our 3 simulations. That means that a 5-day heat wave in our models is within an acceptable approximation of this catastrophe Midwest...
event. For the other 3 scenarios (R4Y4, R8Y4, and R4Y8), there is a greater probability of a 5-day heat wave in Chicago than there was in the historical period. For the R8Y8 scenario, this 5-day heat wave event is greater than 10 times more likely in this area based on the historical frequency of heat waves. The R8Y8 simulations also project 1-4 10-day heat waves throughout the decade for most of the Midwest, including Chicago. As we mentioned above, this event did not take place during the entire historical decade.

3.3.4 Winter PDF curves

Different from the summer months, in which TMAX is the primary focus of the analysis, for Winter months, we choose to show the TMIN PDFs for Southeast (Figure 3.5a), Midwest (Figure 3.5b), and Northwest (Figure 3.5c). The dominant trend for Midwest, Northeast, Northern Great Plains, Northwest, and Southwest is a substantial decrease in anomalous cold temperature days, a slight shift in the median to the right in the PDF curve (i.e. warmer), and not as large of an increase in extreme warm temperatures as projected in the summer. Here we compare the historical 25th percentile (denoted by a black vertical line in Figure 3.5) to the future scenarios. For Midwest under R8Y8 scenario, there are only 3% (WH and WCNB) and 6% (WGNN) of winter days that are colder than the historical 25th percentile. In Southeast under R8Y8 scenario, there are only 1% (WCNB) and 10-14% (WGNN, WGN, WH, WCNB) of winter days that are colder than the historical 25th percentile. Note that the WCNB simulation project a much smaller change than the other 4 simulations, suggesting that, in order to make credible assessments of the climate extremes, the models must produce an accurate representation of mean values, the variance in the PDF curve, as well as the shape parameters (Sardeshmukh et al. 2015). In the case of the winter TMIN PDF plots the WGNN and WCNB
simulations did the best at capturing these historical parameters in Southeast and Midwest with
the WGN model best reproducing these parameters for Northwest. This illustrates one of the
primary advantages of this ensemble. By choosing GCMs with different climate sensitivities and
model setup, we not only can test the future uncertainties associated with the changing climate,
but also, we can assess the regional and seasonal strengths and weakness of each ensemble
member.

The other difference between the future simulations and historical/observed curves is the
increased frequency of mild temperatures (i.e. near the median of the PDF curve), which can be
presented using IQR change. This is most evident in Northwest (Figure 3.5c) where the
variability decreases by 15-38% (17-28%) for all 5 simulations when measuring the interquartile
range for 2085-2094 under both scenarios (R8Y4) scenario. The proximately to the Pacific
Ocean is likely a moderating factor in these PDF curves as the cool ocean water likely limit large
increases to the warm side during the winter months (Lokith and Neelin, 2015). The Midwest
experience similar morphology change in the PDF curves with a 16-29% decrease in IQR change
under R8Y8.

For the other three moderate scenarios, there are still projected large decreases in the frequency
of historical 25th percentile. For example, as shown by the statistics in Figure 3.5, for R4Y4 over
Midwest, the historical 25th percentile is projected to range between the 10th (WH) and 19th
(WCB) percentile. A similar range was projected for the Northern Great Plains, Northeast, and
Northwest for this scenario. There is only 0-5% change to the historical 25th percentile with the
R4Y4 scenario in Southeast, Southwest, and Southern Great Plains. For these three regions, only
the WH simulations yield statistically significant changes in the mean in all three scenarios at a
0.001 significance level when using 50 degrees of freedom. Overall, the winter months experience more changes manifested in a large decrease in frequency of extreme cold days, but the difference in mass is redistributed near the historical median which causes a decrease in overall variability measured using IQR. For example, over Northwest under the three moderate scenarios, the IQR changes are as large as those under R8Y8. This may have beneficial impacts on energy demand (i.e. less cold outbreaks mean less demand for natural gas to stay warm), but more mild winters will also disrupt the ecological and hydrological processes in many regions. For example, mild winters have been shown to increase the ragweed coverage as well as other allergy inducing plants (e.g. Zista et al. 2011). Also, invasive species, such as some types vines, can survive and thrive as well as expand northward with less severe winters. This has impacts on both local ecology and infrastructure in these regions (e.g. Bradley et al. 2010). There are also hydrological issues associated with more mild Winters. Snowfall is vital for the water budget for many regions, especially in the western CONUS, because as it melts in the drier Spring and Summer months it produces a water source year-round (Segal, 2013; Barnett et al. 2005). Approximately one-sixth of the world’s population relies on the seasonal melt of snow and glaciers (Barnett et al. 2005). Therefore, if the probability of winter precipitation falling as a liquid and not solid increases or if the freezing level occurs at a higher altitude in the mountainous regions, Spring and Summer water shortages could become more frequent in these regions (Barnett et al. 2005). More discussion on Western precipitation is discussed in Chapter 4.

3.3.5 Length of the Frost-Free Season

Figure 3.6 shows the average difference in days for length of the growing season per year between the future and historical simulations for R4Y8 (left) and R8Y8 (right) scenarios. There
are several caveats when it comes to defining the length of the growing season that are mentioned in Section 3.2.6, but here we are simply looking at the difference in length of frost free season.

With R4Y8 scenario, the areas in the southern part of the CONUS that experienced a 12-month frost free season historically are denoted by the white color (Figure 3.6 left). As for the rest of the country, the WH model has the greatest increase to the growing season with at least 42 day per year increases to almost the entire country. Comparing WH to WGNN, the increase in growing season is only about half as long as WH. There is considerable model spread between the “longest” and “shortest” simulations and therefore, greater uncertainty for this scenario. The 5 simulations indicate at least 30 days increase in the growing season for much of the country (excluding Northern Great Plains and parts of the Rockies) and as many as 35-45-day increase for much of the Southeast and Northwest. This could have the greatest negative impact in the mountainous parts of the Northwest, where the season that can support snowfall will be diminished by at least a month (Bradley et al. 2010) and in agriculturally dependent regions (Melillo et al. 2014).

With R8Y8 scenario, there is an increase in the frost-free season of at least 35 days for the entire country. The areas in white (that were not also in white for the R4Y8 scenario) represent regions that are simulated to increase its growing season by at least 2 months or in some cases, now have 12-month long growing seasons. For most of the simulations, this 2-month increase encompasses the majority or at least half of Southeast, Midwest, Northeast, Northwest, and Southern Great Plains as well as the low altitude parts in Southwest.
With the R4Y4 scenario (not shown), for most of the Midwest and Northeast, there is good agreement in the simulations that there will be about 3 additional weeks in the growing season on average. In some parts of these two regions, along with much of the Southeast and Northwest, an increase of near a month is projected. There is overall good agreement in the 4 simulations for this scenario. Historically, the most recent two decades experienced an overall increase of 1-3 weeks depending on the region (Melillo et al. 2014). This scenario projects this trend to continue to increase linearly with time by mid-century. With the R8Y8 scenario, there is no longer a linear increase to the frost-free season in most regions when compared to the historical trends presented by Melillo et al. (2014). Based on this figure and other results presented above, the historical definition for seasons is projected to shift towards much shorter winters and decreasing frequency of days that can support frost and snowfall.

3.4. Summary and Conclusions

We used five dynamically downscaled RCM simulations with 12-km horizontal resolution to quantify future projections of extreme temperature changes. The initial and boundary conditions used to perform the downscaling came from three GCMs that are representative of the CMIP5 range of sensitivity when CO₂ is doubled. We analyze several metrics that measure the spatial patterns and the PDFs of temperature distributions for two future decades and two RCP scenarios across seven climate regions throughout the CONUS. We found that changes in the temperature distribution is not projected to be linear for all four seasons. During summer, temperature distributions in the future are projected to shift significantly in the mean to the right. For most regions, the temperature extremes typically increase at a greater rate than the climatology mean
(Figure 3.1). This was also accompanied with a large decrease in frequency of days that occurred within the historical lower quartile. This is projected to occur especially in the late-century projections, but there are still significant increases in temperature extremes for the two mid-century scenarios. For most regions, during the winter months, the primary difference between the historical and future projections occurred in the left the side of the curve (Figure 3.5). We do not see as significant of an increase in the extreme warm days as we did during the Summer. The most significant changes in the PDF curves took place within the Northern regions (i.e. Northeast, Midwest, and Northwest), which was in direct opposition to the summer months, where the most significant differences between future and historical PDF curves took place in the southern regions.

We also present metrics that illustrate implications that the evolving regional temperature distributions could have on agricultural demands as well as human health. The first of these metrics was the increasing spatial coverage and increased frequency of 95 °F days (Figures 3.2-3.3). Schlenker and Roberts (2009) found that crop yields for soybeans, corn, and cotton all decrease in yield dramatically with temperature greater than 90 °F. We find that the 95 °F degree days is more than doubled under the R8Y8 scenario in nearly every region. This would cause significant disruptions in the agriculture yields based on the results from Schlenker and Roberts (2009). In addition to the agricultural effects a warming climate will have, the length of the frost-free season is projected to increase by late-century (Figures 3.6), which would require further adaptation for the agriculture community as well as for hydrological demands in some of the western regions. The average ensemble increase for the growing season will be anywhere from 2 weeks to more than 2 months per year for all regions across all regions and all 4 scenarios.
Considering heat stress on humans, we find that potentially deadly heat waves will occur yearly in several locations for all four scenarios (Figure 3.4). More significantly, the R8Y8 scenario projects an increase in events that would be unprecedented within the current climate time line (e.g. 10-day heat waves in the northern regions).

Finally, we discuss the difference in extreme temperature events that can be affected by the two late-century GHG concentration scenarios. We find that the difference in RCP 4.5 and RCP 8.5 scenarios is significant, especially when considering late-century temperatures extremes. Supplementary Table B1 present quantitatively the ensemble average for some of the results presented in this study including only the two late-century scenarios. These estimates are based on the regional and ensemble average of our five simulations. RCP 8.5 is projected to increase each of these metrics by several standard deviations when compared to RCP 4.5. The true ensemble range between the 5 individual simulations is much greater than the values indicated in the table due to the differences in climate sensitivity of the GCMs used as boundary conditions.

These results add to the already large body of literature on temperature change and dynamical downscaling studies. Due to the large domain of this analysis coupled with high spatial resolution, we are able to compile a quantitative analysis across all regions of the CONUS that is rarely presented in the downscaling community. The models presented in this study shows considerable skill at representing the observed PDFs in measures of variance, mean, and extreme values and thus can be considered trustworthy simulations at making future temperature projections (Sardeshmukh et al. 2015). Future projects in this field should be dedicated to increasing the temporal length of this analysis in order to better understand how decadal
oscillations and other long-term climate features are playing a role in temperature distribution evolution in several regions presented in the analysis. Loikith and Broccoli (2013) found that the Pacific-North American and northern annular mode play a key part when it comes the extreme warm days during the winter months. Much like temperature, GCMs also project increasing frequencies of extreme precipitation events over the next several decades (e.g. Janssen et al. 2014). From Figure 2.1 in Zobel et al. (2018) [Chapter 2], this ensemble also shows considerably more skill at simulated extreme rainfall events when compared to its GCM counterpart. Future research will conduct analyses for regional precipitation extremes over the CONUS.
3.5 Figures

Figure 3.1: Summer TMAX distributions over Southeast (a), Northeast (b), and Midwest (c) for four future scenarios, the corresponding historical period, as well as gridded observations. Each daily TMAX value for each of the future projections are calculated using Equation 1 from section 2.2.1. The historical 95th percentile is denoted with a black vertical line. The changes in the historical 95th percentile (numbers on the left) and in the interquartile range (numbers on the right) in future scenarios are also shown using the same color as for the temperature distributions.
Figure 3.2: Changes in number of days greater than 95 °F at each grid point in period of 2045-2054 under RCP4.5 (left) and RCP8.5 (right). Changes are shown as number of days per year.
Figure 3.3: The same as Figure 3.2, but for period of 2085-2094.
Figure 3.4: Changes in number of heat waves per year for period of 2085-2094. Left: changes in frequency of 3-day heat waves under RCP4.5; Right: changes in frequency of 5-day heat waves under RCP8.5 compared to the historical period. In order to properly interpret the results, the color bar for the right column had to be adjusted to account for more events.
Figure 3.5: The same as Figure 3.1, but for Winter TMIN distributions. The regions provided are the Southeast (a), Midwest (b), and Northwest (c). The historical 25th percentile is denoted with a black vertical line. The changes in the historical 25th percentile (numbers on the left) and in the interquartile range (numbers on the right) in future scenarios are also shown using the same color as for the temperature distributions.
Figure 3.6: Changes in the length of Growing Season (i.e. “frost free season”) for period of 2085-2094 under RCP4.5 (left) and RCP8.5 (right) scenarios. Changes are shown in days per year. The white areas in the left column are areas that are warm enough to experience a 12-month growing season in the historical simulations.
CHAPTER 4: ANALYSES FOR HIGH RESOLUTION PROJECTIONS THROUGH THE END OF THE 21ST CENTURY FOR PRECIPITATION EXTREMES OVER THE UNITED STATES

4.1 Introduction

Chapter 4 presents results from a 5-member ensemble of dynamically downscaled simulations with a 12-km spatial resolution to quantify the impacts of climate change on future extreme precipitation within several climatologically cohesive regions of the Contiguous United States (CONUS) by the end of the 21st century. The primary objective of this research is to gain a better quantitative understanding of precipitation extremes using high spatial resolution climate models and provide regional information for potential hydrology and agriculture impacts. The precipitation data presented here build on the analysis of historical accuracy of this ensemble from Zobel et al. (2018) [Chapter 2] as well as the analysis on the evolution of temperature distributions through the 21st century in Zobel et al. (2017) [Chapter 3]. The initial objective of this study is to quantify the “added value” – or lack thereof – dynamical downscaling has over their ESM counterpart for both extreme events and seasonal mean precipitation (Section 4.3.1). Confidence in making future projections is derived from analyses of the ability of the model to represent historical events accurately (Loikith et al. 2015; Christensen et al. 2010; Lorenz and Jacob, 2010; Bukovsky, 2012; Wang and Kotamarthi, 2015). Therefore, in addition to spatially quantifying historical accuracy, we use this ensemble as well as raw ESM data to examine future projections for regional extreme precipitation events measuring the difference between its historical counterpart.

Section 4.3.2 discusses the change in frequency for future event within several regions across the continental United States (CONUS) when compared to the historical simulation. Section 4.3.3 focuses on the impact of increasing regional temperatures discussed in Zobel et al. (2017) [Chapter 3] on precipitation magnitude and distribution. Section 4.4 provides discussion on the potential impacts evolving precipitation events presented in Section 4.3.1-4.3.3 could have on several regions throughout the United States.

### 4.2 Model and Validation Data

This study uses an ensemble of dynamically downscaled simulations to discuss the impact a warming climate has on extreme precipitation events in the future. The RCM used to perform the downscaling is the Weather Research and Forecasting Model (WRF) version 3.3.1 with a 12-km spatial resolution over most of North America (see Figure 1 in Wang and Kotamarthi, 2014 for domain used). For subgrid scale treatments, WRF uses the Yonsei University planetary boundary layer scheme (Noh et al. 2003; Hong et al., 2006;), the Grell-Devenyi convective parameterization (Grell and Devenyi, 2002), the Noah land surface model (Chen and Dudhia, 2001), the longwave and shortwave radiative schemes of the Rapid Radiation Transfer Model ([http://rtweb.aer.com](http://rtweb.aer.com)) (Iacono et al., 2008), and the Morrison microphysics scheme (Morrison et al. 2009). For those simulations that apply the spectral nudging technique (see model description below), nudging is applied above planetary boundary layer height to wavelengths around 1200 km. The nudging coefficient is $3 \times 10^{-5} \text{ s}^{-1}$. A one-year spin-up period is allowed for both historical and future simulations to reach equilibrium. These physics schemes and model setup were selected by Wang and Kotamarthi (2014) after conducting sensitivity tests of the WRF.
model output to different convective parameterizations, microphysics, spatial resolutions, nudging factors, as well as the length of spin-up time.

The ensemble we use has five members with initial and boundary conditions from three ESMs. The ESMs used are the Geophysical Fluid Dynamics Laboratory Earth System Model 2 (GFDL-ESM2G) developed by the NOAA/Geophysical Fluid Dynamics Laboratory (Donner et al. 2011), the Hadley Centre Global Environment Model Version 2 (HadGEM2-ES) developed by the Meteorological Office Hadley Centre, United Kingdom (Jones et al. 2011), and the Community Climate System Model 4 (CCSM4) developed by the National Center of Atmospheric Research (Gent et al. 2011). While there are many more ESMs that participated in the CMIP5 project, these three ESMs well represent the range of the sensitivities of all ESMs response to doubled CO$_2$. The three ESMs in this study range from a climate sensitivity range of 2.38 °C (GFDL-ESM2G) to 4.5 °C increase (HadGEM2-ES), with CCSM4 projecting 2.9 °C, about the average of all CMIP5 models (Sherwood et al. 2014).

This ensemble is made up of one simulation using HadGEM2-ES boundary conditions (abbreviated WH); two simulations using GFDL-ESM2G boundary conditions, one with nudging (abbr. WGN) and one without (abbr. WGNN); and two simulations using CCSM4 boundary conditions, one with bias correction (abbr. WCB) and one without (abbr. WCNB) (see Zobel et al. 2018 for more details). WCB and WCNB also apply the spectral nudging technique, and WGN and WGNN apply the bias correction as well. Applying bias correction and a weak nudging to one simulation and not the other with the same boundary conditions allows for a benefit comparison when using nudging and bias correction. CMIP5 simulations have been
developed using potential pathways for future greenhouse gas (GHG) emissions; these emissions scenarios are known as Representative Concentration Pathways (RCP) (IPCC AR5). For the future emissions scenarios used in this study, we employ two RCP scenarios: RCP 8.5, the Business-as-usual scenario, which implies continued heavy emissions of GHGs at or greater than current emissions through the end of the century leading to radiative forcing of 8.5 W/m² by 2100 (Riahi et al., 2011); and the RCP 4.5 or mitigation scenario which implies stabilization of GHG emissions by the end of the century and peaks by mid-century (Thomson et al., 2011). For the most part, each of the lateral boundary conditions are run with two scenarios for two future time periods each: 1) Historical simulations (1995-2005); 2) RCP 4.5 (2045-2054); 3) RCP 4.5 (2085-2094); 4) RCP 8.5 (2045-2054); 5) RCP 8.5 (2085-2095) (see Table 1 from Zobel et al. 2017). The results presented in this study focus mostly on the RCP8.5 scenario for 2085-2094 (abbr. R8Y8).

For the validation of historical simulations, we use a set of gridded observed precipitation data that comes from the National Oceanic and Atmospheric Administration Cooperative Observer (Co-op) stations. Precipitation data from these gauges are gridded to one eighth degree resolution using a mapping system algorithm from Shepard (1984) and was first implemented by Widmann and Bretherton (2000) (See Zobel et al. 2018 for more details). Finally, the original ESM data is regridded to match the spatial resolution of the observations and RCM simulations for comparison purposes in Sections 3.1-3.3. In this study, we conduct regional analyses using the 7 climatologically cohesive regions set by the National Climate Assessment (NCA) (Melillo et al, 2014; USGCRP, 2017). These regions are abbreviated as followed: Northwest (NW), Southwest
(SW), Southern Great Plains (SGP), Northern Great Plains (NGP), Midwest (MW), Northeast (NE), and Southeast (SE) (see Figure 2 in Janssen et al. 2014).

4.3 Results

4.3.1 Spatial Added Value

Extensive past research in dynamical downscaling has focused on the added value relative to their ESM counterparts (e.g., Hu et al., 2018; Chang et al. 2018). In this section, we discuss how the ESM and RCM simulations in this study capture extreme and seasonal precipitation compared to observations and in which U.S. region the most significant added value by RCMs can be found. To better illustrate the spatially added value of RCMs, Figure 4.1 shows how the RCM and ESM simulations perform at representing the 95th percentile at each grid point, and Table 1 shows regionally averaged absolute errors of RCM and ESM simulated 95th percentile precipitation over seven subregions that are used in the US. National Climate Assessment (Melillo et al. 2014) and Zobel et al. (2018). It is evident that there is a major dry bias for extreme events in all regions of all three ESMs especially east of the Rocky Mountains (bottom two rows of Figure 4.1). In the Midwest and the Great Plains, most areas are 32% drier than the observed 95th percentile with many areas having as high as 60% drier than observations closer to the Gulf of Mexico, although the HadGEM2-ES model performs better than the other two ESMs in all seven subregions (Table 4.1). The RCM simulations significantly reduce the dry bias for extreme precipitation in all seven subregions (Table 4.1), especially in the eastern two-thirds of the United States (top three rows of Figure 4.1). In the Plains and Midwest, WGN and WCB produce less than 8% bias in the majority of grid points. Near the Gulf of Mexico and Southeast U.S., the dry bias of extreme precipitation events is reduced from ~60% from their ESM
counterparts to ~16-32%, although the bias is still large over this region compared to the rest of the U.S. This is most likely due to the difficulty of convective parameterization to capture the extreme events that are caused by tropical cyclone and mesoscale convective events. Chang et al. (2018), using the same model setup as this study for WRF driven by NCEP reanalysis data, found that if the model grid size is reduced to 4 km with convection permitted, the model bias in heavy precipitation over the Southeast can be significantly reduced.

While the HadGEM2-ES performs the best at representing the historical extremes compared to the other two ESMs for the locations east of the Rockies, when it was used as boundary conditions for the downscaling, it did not outperform WGN or WCB. This is likely because of not using spectral nudging and bias correction in these simulations. Spectral nudging is especially prudent in these regions because it emphasizes the large-scale features in the boundary conditions and Kunkel et al. (2011) show that historically, frontal systems and extra-tropical cyclones produce the majority of extreme events in these regions. The results from the two WRF-GFDL and WRF-CCSM simulations reveal that the use of both spectral nudging and bias correction is important over the Midwest and Plains in order to yield greater improvements on representing the extreme values in the RCM compared to their ESM counterparts. Similar conclusion about the improvements by both spectral nudging and bias correction were also found by Xu and Yang (2015). In the western United States, all RCM simulations improve the accuracy of precipitation events compared to the ESMs due to the reality that high spatial resolution results in increased accuracy to simulate orographic lift in these regions, which is important for both seasonally averaged precipitation and extreme events (e.g. Bacmeister et al., 2014). There is less dependency on the effects from the use of bias correction or spectral nudging in the western
United States compared to the central regions of the U.S. All five RCM simulations yield similar results when compared to the observed values with the exception of the WGN simulation in the southern part of Southwest where a wet bias is present. Overall, the RCMs produce much smaller absolute errors compared to the ESMs in extreme precipitation for all seven subregions. For example, as shown in Table 1, the percent difference for absolute error in extreme precipitation when compared to the observations for the Midwest generated by a multiple RCM mean is 10.78%, while that by a multiple ESM mean is 26.01%.

However, the added value of RCMs relative to their ESM counterparts is not always present in seasonal precipitation. We use spring seasonal precipitation as an example to illustrate this finding because most of the extreme precipitation events over the central and eastern parts of the country typically occur from March through August (e.g., Kane and Chelius, 1986; Bukovsky and Karoly, 2011). Figure 4.2 shows the difference in MAM precipitation (mm/day) between each of the simulations and observations. Again, the high-resolution RCMs better capture seasonal precipitation in the topographically diverse western regions (top four rows of Figure 4.2). However, this is not the case for spring precipitation over the central United States. The use of bias correction and spectral nudging appears to be essential for both seasonal precipitation and extreme events within the Midwest and Plains as the WGNN, WCNB, and WH produce a large negative bias in precipitation over these regions that is not present in the ESMs (bottom two rows of Figure 4.2). With that being said, WGN and WCB increase the magnitude of the wet bias from the ESMs in the Southern Great Plains, but overall are still much more accurate than their ESM counterpart. Previous regional modeling studies also showed that spectral nudging is crucial to alleviating the dry bias over the Plains and Midwest regions, specifically during the
warm season (Hu et al., 2018). We found this is not only the case in the Spring, but also in the Summer (not shown). During Summer, spectral nudging improves the simulated seasonal precipitation in the Central U.S., but bias correction can introduce a wet bias over the Plains states that is not present in the ESM counterpart. For Winter seasonal precipitation, WGN and WCB significantly increases a slight wet bias from the ESM over the Northern Plains and Midwest, but WH, WGNN, and WCNB out perform their ESM counterpart in these regions. The RCM seasonal precipitation in the Western regions of the U.S. is improved for all four seasons in a similar magnitude as that shown for Spring in Figure 4.2. Overall, with the exception of the western United States, it appears that the greatest added value of using high resolution downscaled ensemble is its ability to capture extreme precipitation more accurately. Therefore, the sections below focus on future projections of the frequency and intensity of extreme precipitation and the impacts of increasing surface temperature on extreme precipitation.

**4.3.2 Frequency of Extreme Precipitation**

In this section, we aim to understand how climate change affects the distribution of precipitation. Different from the temperature distribution, the precipitation distribution is not Gaussian in nature but is heavily skewed towards low precipitation events. In order to get the best idea of how future projections represent precipitation events, we apply a percentile interval method, similar to that used in Chapter 7 of USGCRP (2017), the 1st Volume of the Fourth National Climate assessment, also called the Climate Science Special Report (CSSR), for our RCM ensemble as well as the original ESM data. First, we calculate the number of dry days for both historical and future simulations. A dry day was defined as any grid point with less than 2 millimeters (mm) precipitation. For the remaining precipitation days, we calculate the historical
percentile thresholds in intervals of 10 from 0 (2 mm) to the 90\textsuperscript{th} percentile and in intervals of 5 for the upper 10\% precipitation days. Once we calculate these values for the historical period, we find the difference in percentages for future non-dry days that fall between each of these intervals. For example, if 8\% of the future non-dry days occur between the 30-40\textsuperscript{th} historical percentiles, which contained 10\% of the non-dry days historically, then the change would be -2\% for those events. We perform this same procedure using the original ESM data as well.

Figure 4.3 shows how the distributions change from historical to R8Y8 throughout the CONUS and four subregions for three ESMs (left) used to perform the downscaling as well as three of the five RCM simulations (right). Over the CONUS, the WGNN and WCNB simulations yielded similar results as WGN and WCB, respectively, but we show WGN and WCB 3 because of their historical accuracy at simulating precipitation extremes. The overall trend for the RCM simulations is an increase in both dry days and heavy precipitation events in the upper quartile, but a decrease in light and moderate precipitation events between 0-70\textsuperscript{th} percentiles. For the R8Y8 scenario over the CONUS, dry days increase in all the RCM simulations with the exception of WH, which simulated approximately zero change in dry days from the historical period. The frequency of events that fell within the upper quartile increases in all 5 RCM simulations compared to the frequency of historical events. This is not the case for the ESM simulations. While the dry days increase in the HadGEM2 and GFDL-ESM2G models, the CCSM4 projected a modest decrease in the number of dry days. Across the CONUS, both the RCM simulations and ESM data project an increase in events that occur in the upper 95\textsuperscript{th} percentile, but all 5 RCMs project a greater increase in these events compared to its ESM counterpart. In the ESMs, it is projected that there will be a slight decrease or no change in the
frequency of the events occurring between the historical 0-50\textsuperscript{th} percentile range and minimal increases in the frequency of events in the 80-95\textsuperscript{th} percentile range. The RCMs project similar sign changes for these percentiles by late-century, but more significant decreases to all events below the 60\textsuperscript{th} percentile and more significant increases to the events above the 60\textsuperscript{th} percentile. There is also better agreement in the RCM simulations between the 0-95\textsuperscript{th} percentile than in the ESMs, leading to more confidence in the RCM simulations as there is less model spread.

Although there are many important differences in the changes of precipitation distribution between various regions, as highlighted in Figure 4.2, overall, the regional distribution of precipitation events project an increase in extreme wet days and dry days while a decrease of median precipitation days. For example, the distribution changes in the Midwest and the Northeast are quite similar so we chose to present only the Midwest in Figure 4.2 to remain concise. For the Midwest (second row of Figure 4.3), there is a good consensus among the ESM simulations that dry days will decrease considerably. However, only the WH and WGNN (not shown) reflect the same result of decreasing dry day frequency. In addition, the increase in frequency of events in the brackets above the 80\textsuperscript{th} percentile increases at a greater rate than the CONUS rate for all 5 RCM simulations and, to a lesser extent, the ESM data.

The southeast region projections between the ESM data and RCM simulations are drastically different. Again, the events greater than the 95\textsuperscript{th} percentile increase in all of the simulations, but the ESM data is much different than the RCMs in all of the other percentile brackets. For example, the magnitude at which the number of dry-days change in the 5 RCM simulations is much less, but the sign change in dry days is the same to its ESM counterpart. The median and
lower quartile percentile brackets are projected to increase or have minimal change in the ESM simulations whereas the RCM data indicate significant decreases to the frequency of these events. In addition, all RCM simulations increase the frequency of 70-95th percentile events and the ESM data deviates significantly from this projection. While the ESM simulations at times project similar increasing frequency in the 95+ percentile as their RCM counterparts, when considering the upper quartile (i.e. ~70-100th percentiles), there is a much more significant increase in the RCM projections for the Southeast and Midwest.

The western regions yielded slightly different results when comparing the ESMs to the RCMs. In the Northwest region, the ESMs project more extreme events (i.e. 95+ percentile) than its RCM counterpart with the exception of the CCSM4. The RCM CCSM4 runs are the only simulations that project a slight decrease in the dry days and the rest of the models all forecast large increases to the number of dry days for this region. There is good agreement between all the models in projection of the events greater than the 80th+ percentiles. The ESMs forecast a more significant increase in the frequency of these events compared to the RCM simulations with the CCSM4 again being the only one that doesn’t. Even though the most significant added value between the ESM and RCM simulations occurred in the Western regions, it appears that in the Northwest they project similar overall changes in terms of precipitation distribution for future projections. In the Southwest, the difference in frequency between the historical 0-95th percentiles and the future projections was similar in the ESMs and RCMs. For the dry days, the RCM simulations all project a large increase in dry days. The WGN projects 2-3 times more dry days than the GFDL-ESM2G, which was the only ESM that projected an increase in dry days by the end of the century.
4.3.3 Intensity of extreme precipitation versus Temperature

Previous research found that precipitation intensities are likely changing due to more moisture in the atmosphere owing to the increased air temperature (e.g. Allen and Ingram, 2002). For example, in the Midwest, a general estimate in the literature for the C-C relationship is that seasonal precipitation will increase 5-7% per degree Celsius warming and extreme events greater than the 95\textsuperscript{th} percentile will increase in intensity by 7-10% (e.g. d’Orgeville et al., 2014; Wang et al., 2017). This section is motivated to determine if daily extreme precipitation events increase in intensity because of the warmer surface temperature and by what magnitude extreme events will increase. We first analyze the percentage changes from historical top 20 heavy precipitation events to future top 20 events in R8Y4 (left) and R8Y8 (right) scenarios (Figure C1 in Supplementary Information). By mid 21\textsuperscript{st} century, for the RCP8.5 scenario, there is a 4-16\% increase in intensity for these events over much of the eastern CONUS. By late 21\textsuperscript{st} century for the RCP8.5 scenario, the magnitude of top 20 events in areas across the Midwest, the Great Plains, and central California are projected to increase by 16-20\% compared to the historical top events. In addition, the spatial coverage of the 4-16\% contours encompass all grid points east of the Rockies. In the following analysis, we will show that, there is typically a small temperature range where precipitation is the strongest in each of these regions. Analyzing the top 20 events likely only yields events from this range and thus, does not adequately describe precipitation extremes throughout all temperature values throughout the year.

In order to quantify changes of daily precipitation intensities due to the change of temperature, we employ a method similar to that used by Wang et al. (2017). We bin all daily precipitation
events greater than 10 mm by their corresponding daily maximum temperature values with bin sizes of 1°C. For each temperature bin, we calculate the 99th percentile for precipitation events during historical and future periods, respectively. We then find the average rainfall per event that occurs within the top 1% in that temperature bin using only bins with at least 100 events. In addition, we find the probability density functions (PDFs) for daily precipitation events versus temperature to determine the percentage of events, greater than 10 mm, that occur within each temperature bin.

Figures 4.4 and 4.5 show the average magnitude of top 1% precipitation events taking place within each temperature bin for the R8Y8 (dotted) and historical (solid) period across the CONUS and five subregions. We use the same three RCM simulations used in the previous section (Figure 4.3) for consistency. The distribution of rainfall events greater than 10 mm for both historical and future scenarios with all temperatures are indicated by black lines, with solid indicating historical and dashed indicating future. For CONUS (left column of Figure 4.4), all the simulations indicate that the extreme events become more intense for all temperatures greater than ~0°C. We choose to focus on the temperature bins between 0-30°C because this is where over 85% of the 10 mm or greater precipitation events take place in all the historical simulations. Overall, the biggest differences between the future and historical simulations occur around 10-14°C and 29-30°C with an average of a 30% and 40% increase in precipitation magnitude, respectively. Within the 10-14°C temperature range, the frequency of rainfall events is about 10-20% higher historically (i.e. black solid line versus dashed line), but for the 29-30°C range, the frequency of rainfall events is 30-60% less historically than in the future, meaning more events are taking place at this warmer temperature range. This is due to the overall shift towards warmer
temperatures in the PDF curves for all the simulations, which shifts the mode in the distributions about 2-3 °C to the right in future simulations. There is a localized peak in precipitation intensity when historical distributions are at a peak and future simulations still indicate an overall increase of 10-20% in intensity despite having 10-30% fewer 10+ mm events at these temperatures. In CONUS and in most subregions, the peaks of the precipitation intensity for both historical and future precipitation events occurred near the temperature bin at or near the mode of the distribution. When comparing the precipitation intensity at the mode for each of the individual RCM simulations and factoring in the 2-3 °C temperature shift, we see that over CONUS there is a C-C relationship of 10-15% increase per °C in precipitation intensity for these temperature bins.

Over the Midwest (Figure 4.4 middle column), extreme precipitation events have projected similar future intensities as the historical simulations at cooler temperatures and this remains the case until surface temperature reaches 15-18 °C, when the future simulations begin to increase the intensity of precipitation at a much greater rate than the historical period. The historical peak intensity of precipitation extremes, which is the temperature bin when precipitation intensity is the highest, occurs at 30-32 °C in all simulations and is actually warmer than the temperature for peak precipitation intensity of the future period (28 °C in all three simulations). In the WGN simulations, there are higher values in the historical period warmer than 32 °C, but they occur when the frequency of events is ~0.05% and is likely a product of small subset of values that do not reflect the overall trend in the models. The focus for this discussion remain at or near the mode of the distributions, where the difference between the historical and future precipitation magnitude is the greatest. There is a ~30% increase for the future scenarios at the mode of the
historical period (i.e. 30-32 °C) with at least 20% less events at this temperature. At 28°C, the magnitude of extreme events is projected to be 50-70% stronger on average per event, which is equivalent to 43-50 mm (1.7-2.0 inches). This temperature bin contains 5-6% of all yearly precipitation events in the future decades, which is about a 25% increase compared to the historical period. While the overall peak in precipitation intensity occurs at a cooler temperature in the future and not within the mode of the distribution, we still see that the overall warm shift in the distribution accounts for an 11-15% increase per °C warming in precipitation intensity for the mode.

In contrast to the Midwest, the Northeast (Figure 4.4 right column) region is projected to experience more intense precipitation events at all temperature bins in the RCM simulations. For the historical simulations, there does not seem to be a temperature at which precipitation is maximized as all temperatures greater than 10°C have similar intensities. By 30°C, all three historical simulations have frequencies less than 1%. With the future projections, precipitation intensities continue to increase with temperature range from 10-30°C. Because of this gradual increase with warmer temperatures, the magnitude of extreme events goes from 5-15% higher at 10°C to 35-40% stronger at 25°C where the distribution of historical events is large enough to produce a meaningful comparison. This trend continues at temperature warmer than 25°C as the frequency of historical events approaches zero. The Southeast (Figure 4.5 left column) yields similar results to the Northeast. The difference in magnitude between the historical and future projections at the mode of the historical distributions is 20-45% at 27°C when the historical frequency is much greater, but this difference increases to 45-60% just 2°C warmer when the frequencies are about the same between the historical and future simulations. This is equivalent
to an increase of 10-50 mm (0.4-2.0 inches) at 27°C and an increase of 40-50 mm (1.6-2.0 inches) per event for extreme precipitation at 29°C. The temperature bin for the peak precipitation intensity increases 3-6°C, but only yields about a 3% increase per °C for the WH and WCB simulations and a 15% increase per °C in the WGN simulation.

The Northwest (Figure 4.5 middle column) is different than the other regions in both the dynamical features associated with large precipitation events as well as how extreme precipitation events change in the future. First, the majority (~95%) of extreme precipitation events occur between 0-20°C in the Northwest region. Historically, the temperature for peak precipitation intensity occurs between 6-9°C in the three RCM simulations. For the future simulations, the peak precipitation intensity occurs in temperature range from 10-12°C. This shift towards warmer temperatures in the distribution is directly related to the shift in the peak precipitation intensity. The future projections produce similar precipitation intensity when compared to the historical simulations until the warmer temperatures bins when the frequency of precipitation events is greater in the future projections. This transition occurs at 7-9°C in the models, which is approximately the median (~7.8°C) for historical distributions. The percent difference in the peak precipitation bins between the future and historical projections ranges between 10-21% and the average warm shift in the peak precipitation values is about 3°C. This leads to the precipitation intensity increasing by 4-6% per degree Celsius increase in surface temperature which is relatively close to the shift in extreme precipitation events projected using the C-C relationship (e.g. Pall et al., 2007). The intensity increase of 10-21% results in 13-19 mm (0.5-0.75 inches) of precipitation per extreme event. In addition to the changes projected for the peak of the curve, the shift in the distribution curves causes precipitation events that occur
below freezing to decrease from 18-24\% in the historical projections to 5-6\% in the future projections. There is also an overall decrease in the precipitation intensity for events below freezing.

The Southwest (Figure 4.5 right column) has several different dynamical features that bring large rainfall events to this region. For example, in the historical simulations, the summer months contain the largest rainfall events to parts of Arizona, Colorado, and New Mexico, but most of Arizona, California, and Nevada, see the majority of their large precipitation events during the winter months (also see [https://www.usclimatedata.com/](https://www.usclimatedata.com/)). To start, the number of precipitation events that occur during days at or below freezing decrease from occurring 12-15\% of the time to only 3-4\% for the future projections. Also, the magnitude of events below freezing are significantly weaker in the WGN and WH models (27-33\% decrease) with only a slight decrease in intensity in the WCB model (5\% percent decrease). For the majority of rainfall events above freezing, the magnitude of future precipitation intensity increases for most all temperature bins. The largest increase in precipitation occurs between 6-12°C, which increase precipitation intensity by 12-50\%. In the WCB and WH, this increase is equivalent to 38-51 mm (1.5-2.0 inches) more per event during this temperature range. For the warmer temperatures, the future simulations all have a localized peak in precipitation between 20-30°C that occurs to the left (colder) of the historical period. With the exception of WGN (15\% increase), this “warm” side peak is projected to be less intense or minimal change compared to the historical peak at warmer temperatures. Overall, as shown in Table 2, multiple RCM mean projects 4°C shift to the warmer side of the temperature distribution over the entire CONUS, Midwest, Northeast, and Southeast, while 10°C warm shift over Northwest. The precipitation extremes are accordingly changed due
to the temperature increase across various regions, with an increase of 12.94% per degree Celsius for precipitation intensity over Midwest and of 2.61% per degree Celsius over Northwest. While Southwest experience the smallest temperature increase, the intensity of precipitation extremes is increased the most (17.73%/°C).

Figures 4.6 and 4.7 employ the same method as Figures 4.4 and 4.5, but with ESM data. With the exception of CCSM4, the two ESM simulations have some similar features in precipitation intensity evolution as RCMs. For example, we see a warm shift in the temperature bin with the highest precipitation intensity in all regions. However, precipitation extremes over CONUS and subregions are significantly lower than those simulated by the RCMs as well as a lower temperature shift in the distributions. Precipitation extremes increase in intensity due to a shift toward the warmer side in the distribution. Unlike the RCMs, the peak precipitation intensity does not occur near the mode in most regions and/or simulations, so we focus on the temperature bin with the peak extreme precipitation intensity instead. In the GFDL-ESM2G and HadGEM2, due to the warm shift, the peak precipitation intensity corresponds to an ~8% per °C for the CONUS and ~5% per °C for the Midwest (Figure 4.6 left and middle respectively). This finding is similar to the results presented by Wang et al. (2017) for the U.S. Midwest using multiple ESM data. It is noteworthy that, the warm shift in the mode of temperature distributions is much higher in RCMs than those in ESMs, leading to the potential for a greater increase in the precipitation extremes. Other than near the peak intensity value, most other temperature bins do not yield precipitation intensity increases at the same rate as their RCM counterpart.
There are large regional differences between the ESM and RCM projections in terms of magnitude of precipitation intensity. In some cases, the precipitation increases via the C-C relationship (i.e. shift in precipitation distributions to warmer temperatures) are similar between historical to future projections in the ESMs and RCMs. As mentioned above, the overall trend in each region is that there is a greater shift in the mode of temperature distributions for the RCM simulations leading to the potential for higher magnitude precipitation events. In addition, because the magnitude of extreme events in the ESMs is significantly less than the RCM simulations, a similar or smaller percent increase in intensity to RCM simulations does not yield the same change in intensity. For example, excluding the CCSM4 results, the temperature bin with the highest average precipitation total in the ESMs for the CONUS increase by 16-20 mm (0.63-0.79 inches). Despite experiencing slightly higher percent increase compared to the historical period as the ESMs, the temperature bin with the highest average extreme precipitation for CONUS increases by 30 mm (1.2 inches) in both the WGN and WH simulations and 26 mm for the WCB. If we consider the multi-model average of the values depicted in Figure 4.4-4.8, as shown in Table 2, the ESMs project 2-3°C temperature warming over CONUS and five subregions, smaller than the RCM projections with the exception of the Southwest. For example, over Midwest and Southwest, the intensities of precipitation extremes are increased by 7.38% and 5.32% per degree Celsius, much smaller than those projected by the multiple RCM mean. Over CONUS, the intensity of precipitation extremes is increased by 13.12% per degree Celsius, larger than that projected by the RCMs, but with the caveat that a multi-model average over a region as large as the CONUS may not yield optimal results.
Finally, the biggest difference between the RCM and ESM simulations occurs when the CCSM4 data is compared to the WCB. There are minimal changes to temperature dependent precipitation intensities for all regions in the CCSM4 data. With the exception of the Northwest (Figure 4.7 middle column), the peak precipitation intensity is about the same or slightly higher in all of the other regions shown in Figures 4.6-4.7. This is best illustrated in the Northeast region (Figure 4.6 right column). The WCB simulations (Figure 4.4 right column) projects increased precipitation intensity at all temperatures, but for the CCSM4 there are only a few temperature bins with higher precipitation intensities in the future projections and the peak precipitation intensity is less than the historical period.

4.4 Summary and Discussion

We used an ensemble of dynamically downscaled simulations with a 12-km horizontal spatial resolution to make future projections for precipitation extremes for several regions across the CONUS. We found that, compared to ESM data, dynamical downscaling shows significant improvements in simulating extreme precipitation, but in several instances, RCM simulations had lower accuracy at simulating seasonal precipitation compared to ESM data. Because models representing historical extreme events more accurately have been shown to also produce more reliable future projections (Loikith et al., 2015; Christensen et al. 2010; Lorenz and Jacob, 2010; Bukovsky, 2012; Wang and Kotamarthi, 2015), it is valuable to quantify the improvements made compared to ESM data for the ensemble used in this study. Using nudging and bias correction significantly reduces precipitation biases over the eastern two thirds of the country. There are also improvements seen in seasonal Spring and Fall precipitation for the Midwest and much of the Plains states when nudging and bias correction are employed. We also suggest that
employing bias correction and nudging have benefits when considering extreme precipitation events over the eastern two-thirds of the CONUS.

RCM simulations project a large increase in the number of dry days as well as very heavy precipitation days (i.e. 95+ percentile) in most models and regions while a decrease in the number of moderate precipitation days (e.g. 0-70th percentile). While the magnitude of distribution changes is different, the projected evolution of precipitation distributions in this study are similar to statistical downscaling analyses presented in Chapter 7 of USGCRP (2017) as well as those found in previous dynamical downscaled studies (e.g., Bukovsky et al., 2011 and Harding and Snyder, 2014). It is also likely dynamical downscaling is superior to statistical downscaling in this regard because statistical downscaling cannot accurately project changes in daily extreme precipitation without the ability to resolve projected changes in dynamics (Hayhoe et al., 2007). When compared to the ESMs, the overall trend for most regions is that the frequency of increasing >95th percentile days are similar to the RCM projections. The biggest difference between the ESMs and RCMs occurs from the 0-95th percentiles as well as magnitude and sign of dry day change. The ESMs project minimal change to the distributions between the 0-95th percentiles. The analyses of the relationship between temperature distribution and corresponding precipitation extremes show that, the rightward shift in projected future temperature distributions likely contributes to the overall increasing precipitation intensity for most regions, including CONUS.

The dynamically downscaled simulations project similar changes in the frequencies of extreme events and dry days in the Northwest as the ESM data. With that being said, the magnitude of
extreme events is significantly higher in the RCM simulations. Most strong precipitation events in the Northwest are due to atmospheric rivers (e.g., Shields and Kiehl, 2016). Dettinger (2010) concludes that atmospheric rivers are not projected to increase in frequency in future projections, but they have the potential to be stronger and occur at warmer temperatures due to increasing surface temperatures. Our results would support this finding as the magnitude of precipitation extremes as well as the temperature in which precipitation extremes are maximized are both increasing.

For the Southwest, there were several discrepancies between the ESMs and RCMs in regards to the frequencies of precipitation events. For example, dry days are expected to increase in all RCM simulations at a much higher rate than its ESM counterpart. In both the Northwest and Southwest, the magnitude and frequency of extreme precipitation that occur when maximum daily temperature is below freezing is significantly less. Lute et al. (2014) used a downscaled ESM analysis to show a decreasing number of snowfall days that were of similar magnitude as the simulations presented in this study. In addition, declining large snowfall events, decreasing snowpack, and increasing wildfires were projected by Gergel et al. (2014). An overall increase in dry days in the Southwest along with decreasing snowfall magnitude and frequencies in both western regions projected by this ensemble could lead to significant need for water management adaptations in these regions.

The regional precipitation frequencies in Figure 4.3 yield the most significant differences compared to ESM data in convectively driven regions such as the Southeast. Overall, the RCM simulations in the Southeast all agree with the projection of increasing frequencies of events
greater than the 80\textsuperscript{th} percentile and that median days will decrease at a modest rate. The overall signal in the ESM data for the Southeast was not clear as each model yielded different results for how they projected the frequencies of median-high precipitation events in the future, but the RCM simulations yielded higher confidence in their projections because of strong agreement despite different boundary conditions.

Unlike the Northwest, there is projected to be a modest increase in precipitation intensity at most temperatures in the Northeast and to a lesser extent in the Southeast. The magnitude of precipitation intensity increases for most temperatures is larger in the RCM compared the ESMs for these two regions. In addition, the evidence suggests that more frequent precipitation events will occur near 30 °C. At this temperature, historical precipitation intensity has begun to drastically decrease along with frequency. Previous studies have alluded to the potential that higher resolution simulations decrease midlatitude cyclone activity along the East coast of the United States, but the strength of midlatitude cyclones are expected to increase (Collie et al., 2013; Pfahl et al., 2015; Chang et al., 2016). More research is needed on this ensemble to determine if that is what leads to large increases of the magnitude at most surface temperatures.

The Midwest presents an interesting case study in our ensemble. Different from regions, the results from Figure 4.4 show that shift in precipitation distributions toward the warm temperatures had little effect on the temperature where the peak precipitation intensity occurred. The peak in precipitation intensity occurred at a cooler temperature than the historical period in all of the simulations. This was not the case for the ESM data presented in Figure 4.6 where increasing precipitation extreme intensity peaks at a warmer temperature. Figure 4.6 supports the
overall conclusions of the study conducted by Wang et al. (2017). They used ESM results and a similar method to analyze the U.S. Midwest as one of its regions. It is likely that an improved ability to represent dynamical features in the RCMs plays a more significant role in the Midwest, as suggested by the contrasts between Figures 4.4 and 4.6.

Previous studies show that increasing extreme events tend to result from increased moisture convergence in the Midwest from the low-level jet during Spring (e.g. Cook et al. 2008). Figure 4.8 analyzes the role of the low-level jet played on warm-season precipitation events in the future projections. Figure 4.8a uses the same method presented in Figure 4.3, but for March-April-May in the Midwest. Unlike the yearly distribution in Figure 4.3, all models show a significant decrease the number of dry days and yield much larger increases to the upper quartile than the yearly average. In Figure 4.8b, we show the average 850 hPa meridional wind (i.e. north-south wind) anomaly during the top 10 historical events to get an idea of how the low-level jet (LLJ) plays a role in large Midwestern precipitation events. Comparing that to the average change in meridional winds in the projections shown in Figure 4.8c and Figure 4.8d, it can be concluded that April-May yields the most favorable changes in low level jet dynamics compared to June-July. It appears that Spring precipitation extremes will increase at a greater rate due to dynamical features in the RCMs. This increased Northward moving low level jet will likely increase the temperature gradient for frontal systems associated with mid-latitude extratropical cyclones by bring warm moist air from the Gulf of Mexico into the Midwest. This is important because Kunkel et al. (2011) show that frontal systems result in 90-93% of historical extreme precipitation events in the Midwest during the Spring season. Historically, the peak in the low-level jet is maximized during June-July (Higgins et al., 1996) however our results indicate that
the greatest change in LLJ will occur during the spring months owing to the peak precipitation events potentially occurring at cooler temperatures. These results are also supported by Cook et al. (2008) that found increasing low-level jet strength peaks earlier in the year for future simulations. The large increase in Springtime Midwest precipitation is also likely the primary driver leading to the Midwest experiencing the smallest regional shift in the extreme JJA temperature presented in Zobel et al. (2017). Changing atmospheric dynamics in a warmer climate likely played a more significant role in Midwestern intensity changes than increasing surface temperatures and led to the peak precipitation intensity occurring at a lower temperature in future simulations. Future research should be aimed at analyses of the dynamics associated with these changes in precipitation extremes for all regions.
### 4.5 Tables and Figures

<table>
<thead>
<tr>
<th>Average Absolute Error</th>
<th>Midwest</th>
<th>NGreatPlains</th>
<th>Northeast</th>
<th>Northwest</th>
<th>SGreatPlains</th>
<th>Southeast</th>
<th>Southwest</th>
</tr>
</thead>
<tbody>
<tr>
<td>WGN</td>
<td>17.87831</td>
<td>17.27172</td>
<td>18.57605</td>
<td>20.69206</td>
<td>24.29123</td>
<td>17.81328</td>
<td>18.85184</td>
</tr>
<tr>
<td>ESM GFDL</td>
<td>29.83719</td>
<td>28.28375</td>
<td>19.73166</td>
<td>33.16609</td>
<td>40.50405</td>
<td>40.4041</td>
<td>27.67067</td>
</tr>
<tr>
<td>ESM CCSM4</td>
<td>27.99889</td>
<td>32.00439</td>
<td>18.42086</td>
<td>28.86309</td>
<td>46.4872</td>
<td>41.73857</td>
<td>25.95526</td>
</tr>
</tbody>
</table>

**Table 4.1:**

The table above shows the regionally averaged absolute error for the 7 regions in the CONUS that is depicted in Figure 4.1. Regionally averaged absolute error is calculated by taking the grid point absolute value of the values shown in Figure 4.1 and then averaging regionally.

<table>
<thead>
<tr>
<th></th>
<th>CONUS</th>
<th>Midwest</th>
<th>Northeast</th>
<th>Northwest</th>
<th>Southeast</th>
<th>Southwest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg-RCM</td>
<td>4(5.11)</td>
<td>4(12.94)</td>
<td>4(5.11)</td>
<td>10(2.61)</td>
<td>4(4.29)</td>
<td>2(17.73)</td>
</tr>
<tr>
<td>Avg-ESM</td>
<td>2(13.12)</td>
<td>2(7.38)</td>
<td>3(7.42)</td>
<td>2(-0.83)</td>
<td>3(4.23)</td>
<td>3(5.32)</td>
</tr>
</tbody>
</table>

**Table 4.2:**

Table 4.2 takes the multi-model average of the values depicted in Figures 4.4-4.8 and displays the temperature shift in the mode and the percent difference in precipitation from historical to future from the mode. This is done for each of the regions in Figures 4-8 (i.e. temp change (precipitation change in %/°C)).
Figure 4.1: Percentage difference in the 95\textsuperscript{th} percentile of daily precipitation between observations and RCM as well as ESM simulations.
Figure 4.2: Difference (unit: mm/day) in seasonal precipitation (March, April, and May) between observations and RCM as well as ESM simulations.
Figure 4.3: Percentage difference in the frequency bins between historical and RCP 8.5 2085-2094 simulations. ESM simulations appear in the left column and RCM simulations in the right column.
Figure 4.4: Temperature versus RCM average precipitation intensity for top 1% events (red) as well as distributions of precipitation events with greater than 10 mm (black) for CONUS (left column), Midwest (middle), and Northeast (right). R8Y8 scenario is represented by dashed line and historical simulations are solid. Black temperature values indicate the temperature shift in the mode (i.e. peak) of the distributions and red values represent the percentage change of precipitation intensity in the mode of the distributions per °C temperature change.
Figure 4.5: Same as Figure 4.4, but for the Southeast (left column), Northwest (middle), and Southwest (right).
Figure 4.6: Same as Figure 4.4, but for ESM data. The top row of numbers represents the shift in the mode of the distribution and the corresponding precipitation intensity change per °C temperature change. The second row of values represents the shift in the temperature that experienced peak precipitation intensity as well as the change in precipitation intensity per °C change.
Figure 4.7: Same as Figure 4.6, but for the Southeast (left column), Northwest (middle), and Southwest (right).
Figure 4.8: a. MAM precipitation distributions for Midwest under the R8Y8 scenario compared to the historical simulations. b. Average 850 hPa meridional wind anomaly for top 10 historical Midwest events. c. R8Y8 anomaly for 850 hPa meridional wind anomaly compared to historical simulation for April-May (middle column) and d. June-July (right column).
Quantifying the uncertainty associated with climate projections is a vital process when informing
decision makers and the public of the potential risks associated with regional climate change. In
an effort to better quantify this uncertainty, this dissertation utilizes a dynamical downscaling
technique to produce an ensemble of simulations with differing boundary and initial conditions
(i.e. GFDL-ESM2G, HadGEM2, CCSM4), model setup (i.e. bias correction and spectral
nudging), concentration pathway scenarios (i.e. RCP 4.5 and 8.5), and time periods (i.e. mid and
late 21st century). To further quantify the uncertainty associated with regional climate change,
the ensemble presented here uses 3 ESM datasets with a large range of climate sensitivities that
encompasses the climate sensitivities of the majority of all CMIP5 ESMs. The dataset created in
this project yields regional climate simulations with high enough resolution to better capture
local and regional atmospheric processes and climate extremes better than their low-resolution
ESM counterparts while maintaining a domain large enough to provide a detailed quantitative
analysis for most of North America. This section summarizes the key findings from each of the
chapters in this dissertation as well as how these results can be beneficial to future researchers
and decision makers focused on adapting to future changes in the climate.

In Chapter 2, we present the “added value” of these RCM simulations compare to their ESM
counterparts as well as describe the uncertainties associated with the choice of boundary
conditions and model setup. The RCM simulations used in this study show large improvements
over the ESM used as boundary conditions for both extreme precipitation and temperature with
only sporadic instances to the contrary. For example, the WGNN historical simulation produces
a strong warm bias in extreme temperature over the Midwest that is not present in the GFDL-ESM2G simulation. With that being said, the simulations from this ensemble overwhelmingly outperform the ESM simulations. In terms of precipitation, it is shown that as the event becomes more infrequent, and thus has the potential to create the largest economic impact on society, these RCM simulations show a greater spread. For example, Figure 2.1 shows that the “added value” of ESM and RCM simulations increases from the 75th percentile to the 99th percentile. Overall, the RCM simulations have the greatest added value in regions of complex topography and regions where small scale processes play a large role in extreme events (e.g. convective events).

In addition, Chapter 2 provides information to future researchers by describing the regional dependency that model setup and choice of ESM boundary conditions effect the biases present in the historical simulations. Overall, there was not one RCM simulation that outperformed the others for all regions and atmospheric variables. It can be concluded that model setup (i.e. bias correction and spectral nudging) can be more important in several instances to the predictability and historical biases of the simulations than the boundary conditions from the ESMs. For example, the use of spectral nudging appears to be vital for the predictability of temperature and precipitation extremes over the Midwest and Plains due to the dependence of large scale atmospheric processes in these regions on extreme events. In other cases, the boundary conditions used when performing the downscaling had more of an effect on the predictability of the simulations. This was the case when considering the relative error for minimum temperature in the Northwest and Southwest where both simulations that used GFDL-ESM2G as boundary
conditions produced significant error and the WH simulation outperformed the other 4 RCM simulations by a significant margin (lower right panel in Figure 2.3).

In Chapter 3, this ensemble is employ to quantify the evolution of PDF curves for regional temperatures within the future projections. As shown in Figure 3.1 and 3.5, extreme events are not changing linearly for each season and region. In the Summer months, there is significant shift in the median of the PDF curve, but for most regions the extreme events are increasing a greater rate than the median. This is likely due to increase variability in the PDF curves for many of the simulations and scenarios during summer. The Winter months’ project less of a warm shift in the median compared to the Summer. The primary difference in the future projections compared to the historical simulations during Winter occurred in the left tail of the PDF curve. Decreasing extreme cold days as well as decreasing variability in most regions led to a large reduction of lower quartile events with those events now occurring near the median in the future projections.

These projected changes to the temperature distribution in the summer and winter seasons will have several negative impacts on society through increased health concerns and agriculture demands. Chapter 3 attempts to quantify the uncertainty associated with emission scenario by comparing the RCP 4.5 and RCP 8.5 late century scenarios using metrics such as increasing heat waves events, change in the growing season, and frequency of 95 °F days per year. Comparing these scenarios reveal that without significant reduction in greenhouse gas emissions needed to achieve the RCP 4.5 scenario, there will be a need for regional adaptation to combat a drastically different regional climate.
Similar to Chapter 3, Chapter 4 examines the projections for extreme daily precipitation events in terms of both frequency and magnitude. The RCM simulations show precipitation events in the upper quartile as well as dry days will increase in all regions while the median, or average, precipitation events will decrease. When analyzing precipitation intensity as a function of surface temperature, the ensemble shows that precipitation events will increase in intensity for most temperature bins, especially near the mode of the distributions, as the precipitation events predominately occur at warmer temperature.

In addition to showing the “added value” of these simulations compared to their ESM counterparts historically, Chapter 4 uses ESM projections to highlight the differences in projected precipitation extremes when performing downscaling. In most of the regions, the RCM simulations project significantly more intense and frequent extreme precipitation events when compared to the ESM results. This comparison to projected precipitation extremes along with the quantitative analysis of the uncertainty in emission scenarios from Chapter 3 showcase the importance of this dynamically downscaled ensemble. The RCM simulations used in this study advance the scientific understanding of future climate extremes. As previous mentioned, this multi-ESM ensemble pushes the boundary of downscaling research already present in the climate change research. The results presented in this dissertation provide valuable insight to policy makers of the potential risks associated with a warming climate. While this study provides a comprehensive and quantitative analysis of projected climate extremes, future work should use this dataset to examine more of the dynamical processes associated with these climate extremes. This research should also serve as guideline for future downscaling studies as future iterations of ESM models become available, such as CMIP6. In addition, future research should attempt to
increase the resolution to even finer resolution than this study. Going to spatial resolutions of 1-2 km or higher will increase our understanding of processes within clouds and convective storms as the climate system warms. Understanding how cloud coverage and convective events will change in the future through dynamical regional climate modeling is still an important need in the climate change research.
REFERENCES

doi:10.1175/WAF995.1

doi:10.1029/2005JD006290

https://doi.org/10.1002/2014JD022960


https://doi.org/10.1175/JCLI-D-13-00387.1


Casanueva, A., Herrera, S., Fernandez, J., Kotlarski, S., & Gutierrez, J.M. (2016). Added value of high resolution RCM simulations and comparison with Statistical Downscaling Methods within the EURO-CORDEX framework. WCRP CORDEX.


Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and


https://doi.org/10.1175/2008JCLI2355.1


Daly, C., Taylor, G.H., & Gibson, W.P. (1997). The PRISM approach to mapping precipitation and temperature. *In Proc. 10th AMS Conf. on Applied Climatology* (pp. 20-23).


Hong, S.Y. & Kanamitsu, M. (2014). Dynamical downscaling: Fundamental issues from an


Rothfusz, L. (1990). The heat index (or, more than you ever wanted to know about heat index), Technical Attachment SR 90-23. Fort Worth: Scientific Services Division, National Weather Service.


[https://doi.org/10.1175/JCLI-D-15-0020.1](https://doi.org/10.1175/JCLI-D-15-0020.1).


[http://dx.doi.org/10.1038/nclimate3239](http://dx.doi.org/10.1038/nclimate3239)


Yongkang, X. et al. (2014). A review on regional dynamical downscaling in intraseasonal to seasonal simulation/prediction and major factors that affect downscaling ability. Atmospheric Research 147:68-85. doi:10.1016/j.atmosres.2014.05.001


Figure A1: All variables ranked in terms of relative error for the CONUS, Northeast, Southeast, and Midwest regions.
Figure A2: All variables ranked in terms of relative error for the CONUS, Northeast, Southeast, and Midwest regions.
Table B1: These tables summarize the regional range for the difference in the ensemble average between the historical and future simulations discussed in section 3. We show only the results for R4Y8 and R8Y8. Rows 1-3 show the difference in days per year where TMAX is at least 95 °F. Rows 5-7 show the difference in the frequency of heat waves per year. Rows 9-11 shows how much longer the growing season will be with the two scenarios.
Figure C1: Percentage difference between the top 20 events of daily precipitation between historical and projected RCM simulations.