

**A framework for investigating the impacts of climate change on droughts over
Northeastern US**

By

Zeyu Xue
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ATMOSPHERIC SCIENCE

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Professor Paul Ullrich, Chair

Professor Shu-Hua Chen

Professor Erwan Monier

Committee in Charge

2022

To All My Loves...

May wellness and happiness always be with us.

Contents

1	Introduction	1
2	A Comprehensive Intermediate-Term Drought Evaluation System and Evaluation of Climate Data Products over the Conterminous United States	3
2.1	Introduction	3
2.2	Datasets	7
2.2.1	Climate Prediction Center (CPC) unified gauge-based analysis precipitation data	7
2.2.2	The fifth and sixth phase of Coupled Model Intercomparison Project (CMIP5/6) model data	7
2.2.3	The North American Coordinated Regional Climate Downscaling Experiment (NA-CORDEX) dynamical downscaled data	8
2.2.4	Localized Constructed Analogs (LOCA) statistical downscaling data	8
2.2.5	The Watershed Boundary Dataset (WBD)	8
2.2.6	Data preprocessing	8
2.2.7	Hydrologic regions	9
2.3	Statistical Methods	11
2.3.1	Scoring individual metrics	11
2.3.2	Principal Feature Analysis (PFA)	11
2.3.3	Total Score	12
2.4	Drought Metrics	13
2.4.1	Monthly means	13
2.4.2	Seasonality	14
2.4.3	Spatial character	15
2.4.4	Drought frequency	16
2.4.5	Drought intensity	16

2.4.6	Probability of drought	16
2.4.7	Taylor Diagram and Taylor Score	18
2.5	Results	18
2.5.1	Principal Metrics Employed	18
2.5.2	The validation of our evaluation system	19
2.5.3	CMIP6 Performance	20
2.5.4	CMIP5 Performance	22
2.5.5	CORDEX Performance	23
2.5.6	LOCA Performance	26
2.6	Conclusions	27
2.7	Acknowledgements	32
2.8	Supplements	33
2.8.1	Statistical hypothesis testing	33
2.8.2	Stanardized Precipitation Index (SPI) Calculation	35
2.8.3	Sensitivity analysis of total score to study period	37
2.8.4	Sensitivity analysis of model performance to common grid resolution	37
2.8.5	CORDEX over the California Region	39
3	Changing trends in drought patterns over the Northeastern U.S. using multiple large ensemble datasets	69
3.1	Introduction	69
3.2	Data and methods	73
3.3	Results	79
3.3.1	Result 1: A general wetting trend is undeniable in the mean	79
3.3.2	Result 2: More extremely wet conditions, with greater magnitude	83
3.3.3	Result 3: More short-term extreme droughts and intensified evapotranspiration	83
3.3.4	Result 4: More flash drought with faster initiation and greater intensity . . .	85
3.3.5	Result 5: Significant trends towards more extreme flash drought, particularly in the spring	85
3.3.6	Result 6: No clear consensus on intermediate-term drought	86
3.4	Discussion	88
3.4.1	More precipitation and more variable precipitation in the NEUS	88

3.4.2	Evapotranspiration is energy-limited within the humid NEUS	90
3.4.3	Differences in the trends of precipitation and evapotranspiration variability .	91
3.4.4	Increasing anti-correlation of evapotranspiration and precipitation	92
3.4.5	Modified evapotranspiration partitioning brought by the extension of growing season	93
3.5	Conclusions	97
3.6	Supplement	98
3.6.1	The study region and grid cells included	99
3.6.2	A case study of rapidly developing drought based on ERA5 over Northeastern US	99
3.6.3	Case studies of rapidly developing drought based on CESM1-CAM5 over Northeastern US	101
3.6.4	A case study of infamous 2012 US Midwest flash drought based on ERA5 . .	102
3.6.5	Justification of SNPI calculation	103
3.6.6	The soil moisture analysis	104
4	A retrospective and prospective examination of the 1960s U.S. Northeast Drought	138
4.1	Introduction	138
4.2	The record-setting 1960s drought	140
4.3	Uncertainty of a return to drought conditions under climate change	141
4.4	A simulation of present and future analogues of the 1960s drought	144
4.4.1	Methodology	144
4.4.2	Overview of simulations	148
4.4.3	Temperature	149
4.4.4	Precipitation	155
4.4.5	Evapotranspiration	158
4.4.6	Snowpack	160
4.4.7	Soil moisture and runoff	163
4.4.8	Drought indices	167
4.5	Conclusions	173
4.6	Supplement	175
4.6.1	Soil degradation in WRF-CLM4	175

4.6.2	Historical reanalysis precipitation data performance	176
4.6.3	Heat Index equation	179
4.6.4	Simulation and Spin-up Validation	182
5	Conclusion	189

Abstract

The significant impacts of climate change on extreme weather events have been widely demonstrated in past decades. As the most expensive climate disaster, the drought and its risks under the warming climate always attract people’s attention. However, huge uncertainties exist in climate projections, mainly consisting of scenario uncertainty, model uncertainty, and internal variability. Moreover, drought is more difficult to assess in climate datasets, due to its long duration per event, relative to the length of a typical simulation. Meanwhile, since the infamous extreme drought of the 1960s, the climate of the Northeastern United States (NEUS) has generally trended towards warmer and wetter conditions. Nonetheless, there is mounting evidence that short-term droughts will continue to pose a significant risk for this region. Therefore, there is a growing need for a comprehensive framework with the most advanced climate techniques and datasets to investigate the impacts of climate change on drought over NEUS.

In this thesis, firstly, a comprehensive drought feature-based evaluation system, equipped with statistical hypothesis testing and Principal Feature Analysis, is put forward and is designed to be easily used for any climate datasets. With its help, people can choose climate models with fewer model uncertainties and biases in capturing droughts in the regions of interest. This system is applied to three characteristically distinct regions in the conterminous US and across several commonly employed climate datasets (CMIP5/6, LOCA, and CORDEX). As a result, insights emerge into the underlying drivers of model bias in global climate models, regional climate models, and statistically downscaled models.

Then to reduce the impacts of internal variability on drought projections, 7 large ensemble (LE) models and a novel drought index are employed to examine the changing trends of drought at different temporal scales. We find most LE models indicate the NEUS will experience a long-term wetting trend with more “extremely wet” months, but also more frequent short-term extreme droughts. These changes are associated with increasing precipitation, atmospheric water demand, and climate variability. We also conclude that discrepant trends in precipitation and evapotranspiration variability will lead to increasing anti-correlation of these variables, which is relevant to the

intensification of rapidly developing drought – particularly in the spring season. These changes are associated with an increase in evapotranspiration from plants, brought by an earlier emergence of the growing season and denser vegetation.

Finally, by using the pseudo-global warming technology, we project the returned 1960s drought under future warming scenarios to illustrate the potential risks if the similar dynamical conditions of the historical 1960s drought re-occured under global warming. We find that in the returned 1960s droughts, although wet months get much wetter, dry months may become drier, meaning that wetting trends are most significant in wet months but are essentially negligible for extremely dry months with negative monthly mean net precipitation. For these months, the trend towards wetting conditions provides little relief from the effects of extreme dry months. These conditions may even aggravate water shortages due to an increasingly rapid transition from wet to dry conditions. Other challenges emerge for residents and stakeholders in this region, including more extreme hot days, record-low snow pack, frozen ground degradation, and subsequent decreases in surface runoff.

Acknowledgments

Time fleets, it feels like a dream that I'm about to complete my Ph.D. thesis. I never expect that I can get a Ph.D. degree in Atmospheric Science successfully in three years. It was a hard experience but finally, I made it. The help and support from numerous people contribute a lot to this thesis.

First and foremost, I would like to express my deep gratitude to my advisor – Professor Paul Ullrich for his trust, support, and help. Before I joined his group, I had majored in Hydrology and Water Resource and knew not much about climate science. All I have is an interest in studying the impacts of climate change on water resources and management. Thanks a lot to him for his trust and for giving me a chance to implement the research based on my interest. Also, He is always generous to all his students and provides support to satisfy all we need. In completing the papers, his enlightening suggestions and detailed modifications are the keys to my success. I have enjoyed the time we spent working together and expect further cooperation in the future.

I would like to thank all my committee members (Dr. Michael Wehner, Professor Shu-Hua Chen, Professor Erwan Monier, Professor Matthew Igel, and Professor Da Yang) for their feedback and suggestions that make this thesis more comprehensive. Especially, for the work in Chapter 3, I would like to thank Dr. Ruby Leung for her review and suggestions. For the work in Chapter 4, I would like to thank Professor Shu-Hua Chen for teaching me to conduct WRF simulations and explaining the underlying mechanisms. I thank Dr. Zexuan Xu a lot for helping me to set up WRF on Cori. Besides, I would like to thank Dr. Jimmy Dudhia, Dr. Kelly Werner, and Dr. Ming Chen in WRF Team for answering my questions about the applications and mechanisms of WRF. I would like to thank Dr. Melissa Bukovsky and Dr. David Lawrence for their guidance to modify land-use data in the WRF simulation. Thanks to all the friends I met at UC Davis. More importantly, I would like to thank my girlfriend for all her understanding and encouragement that give me the motivation to work hard.

Of course, the most sincere gratitude to my family for long-term and selfless support and encouragement. Their love always give me the power to face the challenges in my life and equipped

me with beliefs. Specially, I would like to thank my mother who always tried her best to give me all the support I need and respect all the choices I have made. Even when I was a problem student in high school, she still trusts me and believes I can accomplish achievements. I would like to thank my father who always serves as a role model for me and always tells me that I can make the thing he made. Without his encouragement, I would not study abroad and decided to pursue a Ph.D. degree. I would like to thank my sister who always believes in me and let me know I'm important and needed. May wellness and happiness always be with us.

Finally, thank destiny for letting me have such a memorable and special experience, in which I learned a lot.

Chapter 1 Introduction

Droughts are among the most disastrous extreme weather events in human society, especially for the agriculture and ecosystem, with the most severe droughts having economic impacts greater than \$10 billion dollars (Andreadis and Lettenmaier, 2006). Both historical observations and climate predictions indicate that climate change is likely to increase the intensity and frequency of extreme weather events such as droughts, floods, wildfires and heatwaves (Christian et al., 2021; Frumhoff et al., 2007; Pendergrass et al., 2020; Wilhite, 2000). However, significant uncertainties persist regarding droughts' frequency and magnitude in a warming climate especially over mid-latitude regions like Northeastern US (NEUS) (Hayhoe et al., 2007; Kharin et al., 2007; Pfahl et al., 2017). Given the advances in climate modeling over the past two decades, there's an increasing and unmet need for understanding how will droughts especially extreme droughts respond to climate change.

The Northeastern U.S. (NEUS) is not only well-known for its dense population and thriving economy (US Bureau of Economic Analysis, 2016), but also for its humid and mild climate (Frumhoff et al., 2007). Unlike drier regions of the country that often suffer from drought, the abundant precipitation of the NEUS generally enables it to avoid prolonged and severe drought. This has been especially true over the relatively wet period since the unprecedented and infamous 1960s drought, which has essential influence on local water management (Ford and Labosier, 2017; Seager et al., 2012). But water resource managers will be expected to deal with several impending challenges brought by the warming climate such as the increased risk of wildfire, heatwaves, and drought (Frumhoff et al., 2007; Hayhoe et al., 2007). Also, historical data may not be sufficiently reliable for future projections due to the loss of stationarity, more intense extreme weather events, and enhanced climate variability (Armal et al., 2018; Milly et al., 2008; Stryker et al., 2018; Yu et al., 2018). Consequently, there remains a need for high-quality climate information to inform regional adaptation strategy (AghaKouchak et al., 2015; Milly et al., 2008).

At present, climate models are the primary methods for decision-makers to project climate

change and its consequences (Frumhoff et al., 2007; Kharin et al., 2007; Wagener et al., 2010). Thanks to their rapid development in past decades, both the resolution and accuracy of climate models have been improved significantly; however, huge uncertainties still exist, primarily coming from scenario uncertainty, model uncertainty and internal variability (Deser et al., 2012; Hawkins and Sutton, 2009; Xie et al., 2015). The last two uncertainties account for most of the total uncertainty in projections in the next few decades and are potentially reducible by advances in climate science (Deser et al., 2012, 2020). This study aims at reducing the climate projection uncertainties and analyzing the impacts of climate change on drought and hydro-meteorological conditions over NEUS to provide extensive projections for residents and stakeholders. A corresponding framework will be put forward for regional climate impact analysis. Namely, this study proposes to answer the following three questions by solving respective tasks:

1. How well do earth system models capture the drought features?
2. How are the drought features, especially the flash drought, projected to change in the future?
3. What are the potential climate risks of the returned 1960s extreme drought under a warming climate?

Therefore, aiming at providing reliable regional climate projections, this thesis consists of three major case studies. In Chapter 2, we put forward a comprehensive drought feature-based evaluation system to select models good enough for drought projection to reduce the model uncertainties and understand their biases. In Chapter 3, to sample both internal variability and structural uncertainty, 7 large ensemble (LE) models are employed to answer the outstanding question – how do the frequency and character of drought and flash drought in the NEUS change under a warming climate? More importantly, answer the unsolved hot topic – what are the underlying drivers of flash drought? In Chapter 4, we conducted the pseudo-global warming (PGW) experiment to simulate the returned historical 1960s under climate warming and its potential climate risks.

Chapter 2 A Comprehensive Intermediate-Term Drought Evaluation System and Eval- uation of Climate Data Products over the Conterminous United States

2.1 Introduction

Currently, climate models are the primary methods to project climate change and its impacts on extreme weather events like drought (Frumhoff et al., 2007; Kharin et al., 2007; Wagener et al., 2010). However, all models are known to possess certain biases which can contaminate simulation results (Moon et al., 2018; Nasrollahi et al., 2015). Biases present in a single modeling system can be quantified and mitigated through the use of an ensemble of global climate models (GCMs), such as the multi-model ensemble featured in the Coupled Model Intercomparison Project phases 5 and 6 (CMIP5/6), and corresponding downscaled climate data products, such as the dynamical downscaled simulations from the Coordinated Regional Climate Downscaling Experiment (CORDEX) or the statistically downscaled Localized Analogues (LOCA) product. Nonetheless, there is significant value in understanding these biases in singular products, or across an ensemble, particularly before projections are used for climate adaptation planning (Collier et al., 2018; Eyring et al., 2019; Gleckler et al., 2008, 2016; Lee et al., 2018; Nguyen et al., 2017; Ukkola et al., 2018; Wagener et al., 2010). However, repetitive and redundant evaluations across modeling centers can incur unnecessary efforts and increased costs. Thus comprehensive and robust evaluation systems (Eyring et al., 2019; Gleckler et al., 2008, 2016; Lee et al., 2018) can reduce research costs and provide greater cer-

tainty in the employ of climate models for understanding drought character and impacts, especially when it comes to the large ensemble context (Deser et al., 2020).

With a patchwork of evaluation studies focusing on different products and different metrics, there remains significant uncertainty as to which model data can be trusted for correctly representing drought intensity, frequency, and duration for their region of interest (Jagannathan et al., 2020; Moon et al., 2018; Nasrollahi et al., 2015). In general, our capacity to evaluate drought depends on the duration of the drought and the length of the available dataset. While large ensembles (Kay et al., 2015) permit the assessment of multi-year droughts, individual CMIP5/6 models often only provide single realizations. In this study, our evaluation suite is applied to datasets of 60 years duration, but the model performance scores are shown to be relatively stable with as little as 30 years of data. With this in mind, this study proposes a system with four key novelties:

First, most past research has only focused on evaluating models using meteorological mean fields such as mean precipitation and temperature (Koutroulis et al., 2016; Nguyen et al., 2017); however, mean performance often does not necessarily correlate with model performance in the extreme (Wehner et al., 2020). In the case of drought, impacts from drought are related to temporal continuity features like consecutive duration which can't be captured by mean fields. Thus holistic metrics that are more comprehensive in capturing the features of each event are needed. For this, drought indices like standardized precipitation index (SPI) are proposed to calculate and straightforwardly employed for quantifying the character of droughts (Hayes et al., 2002; Svoboda and Fuchs, 2016; Ukkola et al., 2018; Yihdego et al., 2019). Second, most drought evaluation systems (Darand and Sohrabi, 2018; Orłowsky and Seneviratne, 2013; Teshome and Zhang, 2019) rely on a handful of Expert Team on Climate Change Detection and Indices (ETCCDI) metrics established by the World Meteorological Organization (WMO) and the World Climate Research Program (WCRP). Given that ETCCDI metrics use daily-mean precipitation, their metrics (e.g. consecutive dry days) capture features on a much shorter timescale than traditionally associated with drought (Mishra and Singh, 2010; Wilhite and Glantz, 1985). Third, to determine a score for a model's overall performance, most evaluation systems have normalized these metrics using a standard score or Min-Max Feature scaling (Collier et al., 2018; Koutroulis et al., 2016; Yang

et al., 2014). Others have evaluated models’ performance by assessing relative performance for each kind of metric (Chen and Sun, 2015; Meher et al., 2017; Ukkola et al., 2018). However, a key problem that arises in this case is that the scoring does not have an absolute optimum – the score cannot convey whether a group of models are all terrible or all indistinguishable from observations. So, herein we propose to use a normalization that emerges naturally from the theory of statistical hypothesis testing. Namely, we normalize the value of each metric relative to the statistical significance of its null hypothesis (that is, the hypothesis that the model and observations come from the same distribution or have the same mean value) at the 95% confidence level. This implies that a score of 1 has the same meaning across metrics. Finally, to build a comprehensive evaluation system which scores climate models’ consistency with historical simulations, we evaluate multiple feature-specific metrics related to drought, including: monthly precipitation, Standardized Precipitation Index (SPI), seasonality, drought spatial coverage, drought duration, intensity, frequency and probability of initiation/termination. And the overall performance of each model is illustrated by the assembling of principal metrics detected by selected by Principal Feature Analysis (PFA).

This study focuses on meteorological droughts with temporal scale between seasonal and annual (herein referred to as intermediate-term droughts). Although the term “intermediate-term drought” is widely used in numerous studies of droughts (Bhuyan et al., 2017; Kolb et al., 2016; Thomas et al., 2015; Vicente-Serrano et al., 2014; Wilhite, 2005), it has no strict definition. But generally, it refers to the droughts that are quantified and detected by drought indices that operate at time scales shorter than a year but longer than a month, especially the 6-month Standardized Precipitation Index (SPI6) (Bhuyan et al., 2017; M. Svoboda and Wood, 2012; Thomas et al., 2015; Vicente-Serrano et al., 2014; Wilhite, 2005). Intermediate-term drought tends to be a focus of water managers for several reasons: First, following its original definition (McKee et al., 1993; Wilhite, 2005), intermediate-term SPI (e.g. SPI6) retains advantages of both short-term and long-term indices. Namely, it not only responds quickly to emerging drought conditions but also provides greater stability and persistence for longer-term droughts; therefore, SPI6 is commonly employed in the drought monitor and associated applications (Center, 2021; M. Svoboda and Wood, 2012; Wilhite, 2005). Secondly, although SPI is designed to measure only meteorological drought conditions

(Guttman, 1999; McKee et al., 1993), some studies (Center, 2021; M. Svoboda and Wood, 2012; McKee et al., 1993; Svoboda and Fuchs, 2016) have shown that SPI can capture drought impacts on water resource availability because it is closely related to accumulated precipitation. Namely, soil moisture conditions (agriculture drought) respond to precipitation anomalies on a relatively short scale, while surface runoff and reservoir storage (hydrologic drought) are more so determined by longer-term precipitation anomalies. Therefore, at intermediate timescales, SPI6 can potentially capture the conditions related to agriculture and hydrologic droughts. This motivates its employ in characterizing “model droughts” for water management agencies (Center, 2021; M. Svoboda and Wood, 2012; Zargar et al., 2011). Thirdly, some previous studies have illustrated that there exists a general wetting trend at longer timescales and increased frequency of short-term drought conditions over many of the CONUS regions (Andreadis and Lettenmaier, 2006; Orlowsky and Seneviratne, 2013; Otkin et al., 2018; Seager et al., 2012; Taylor et al., 2013). However, large uncertainties still persist when it comes to intermediate-term drought conditions (Frumhoff et al., 2007; Strzepek et al., 2010). Therefore, in our present evaluation we focus on drought-feature metrics based on intermediate-term drought (SPI6), although we note that shorter droughts (as could be captured with SPI1 or SPI3) as well as multiyear droughts (SPI24 or SPI36) could also be employed in our package.

In this study we propose an evaluation system for drought and apply this system to evaluate simulated droughts in three characteristically distinct regions of the US in the CMIP5/6, CORDEX, and LOCA datasets. The evaluation system aims to select excellent models with less model uncertainties to project the features of drought and is subsequently employed to answer our motivating questions: First, what are the regional and model-specific characteristics that are most relevant for determining the quality of model-simulated drought? Second, when it comes to drought, does the CMIP6 ensemble outperform its predecessor CMIP5? Third, does dynamical or statistical downscaling provided robust added value in the representation of drought? And if so, how?

The study is structured as follows. Assessed datasets are described in section 2.2. The statistical methods employed across the metrics suite are described briefly in section 2.3. Our proposed suite of monthly-precipitation-based drought metrics are defined in section 2.4. Results and conclusions

are then provided in sections 2.5 and 2.6, respectively.

2.2 Datasets

Following the requirements of SPI, monthly mean precipitation from four datasets below will be used as input data and evaluated by our system.

2.2.1 Climate Prediction Center (CPC) unified gauge-based analysis precipitation data

The CPC Unified Gauge-Based Analysis precipitation data is a component of the CPC Unified Precipitation Project underway at NOAA Climate Prediction Center (CPC). The dataset covers the period 1948-2014 with a spatial resolution of 0.25° (Chen et al., 2008; NOAA Physical Sciences Laboratory, 2020; Xie et al., 2007, 2010). For this study, CPC is used as our observational dataset and serves as the “real” values – hereafter whenever the text refers to “observational data” we are referring to the CPC data.

2.2.2 The fifth and sixth phase of Coupled Model Intercomparison Project (CMIP5/6) model data

The Coupled Model Intercomparison Project (CMIP) provides a framework for GCMs to produce simulations for understanding climate change (Balaji et al., 2018; Eyring et al., 2016). As the most advanced presently-available GCM ensemble, CMIP6 is expected to be frequently employed for understanding future climate impacts. As such, it’s particularly important to understand the biases and performance of its component models in capturing features of drought. The resolution of each model simulation is determined by individual modeling centers, varying between 0.5° and 2.8° . Here 33 CMIP6 r1 ensembles are evaluated, which all have more than 60 years overlap with CPC observed data. To understand CMIP6 performance relative to its predecessor (Knutti and Sedláček, 2013; Taylor et al., 2012), 33 CMIP5 r1 ensembles are also evaluated.

2.2.3 The North American Coordinated Regional Climate Downscaling Experiment (NA-CORDEX) dynamical downscaled data

The NA-CORDEX dataset is produced using boundary conditions from global climate model simulations (CMIP5) to drive several regional climate models (RCMs) over North America. The simulations are from 1950-2100 with a spatial resolution of $0.22^\circ/25\text{km}$ or $0.44^\circ/50\text{km}$ (Giorgi et al., 2012; Mearns, 2020). Here we use the raw CORDEX data, which doesn't include bias correction, to ascertain the added value from dynamical downscaling relative to the original CMIP5/6 data.

2.2.4 Localized Constructed Analogs (LOCA) statistical downscaling data

LOCA is a statistical downscaling technique that uses historical analogues to add fine-scale details to global climate model simulations. The LOCA dataset includes 28 downscaled CMIP5 models from 1950-2005 at a resolution of 0.0625° (Pierce et al., 2014). Bias correction is applied in LOCA based on the Livneh observationally-based gridded product (Livneh et al., 2015).

2.2.5 The Watershed Boundary Dataset (WBD)

WBD is a highly-organized and seamless US national hydrologic units dataset describing watershed boundaries across the Continental US (Survey and US Department of Agriculture, 2013). Generally, decisions on water management and policy are made at the watershed level. With this in mind, our evaluation regions are based on the hydrological unit 2-digit level (HUC2) of the WBD. These boundaries are used to subset models' data and evaluate their performance in some typical hydrological units with different meteorological and hydrologic conditions.

2.2.6 Data preprocessing

In our present study, the four climate model datasets described above are evaluated and compared over three select hydrologic regions. In the context of simulated drought, we aim to determine if CMIP6 outperforms its precursor, CMIP5, and ascertain whether or not dynamical downscaled products perform better than their driver GCMs. Because one of our metrics (namely, fractional

drought coverage) requires all compared datasets to have the same number of grid points and hydrologic boundaries, in the evaluation performed in this study we interpolate all datasets onto a $1^\circ \times 1^\circ$ latitude-longitude grid using conservative interpolation (Zhuang et al., 2020). While a common grid is required for the employ of our drought metrics package, the actual grid employed is an input parameter. This particular grid is chosen as it represents an intermediate resolution among available products. As most metrics employed here are based on regional fields, we expect that the interpolation doesn't actually have a significant impact on the evaluation results for any metric except fractional drought coverage. To confirm this is the case we have conducted a sensitivity analysis to common grid resolution in Supplement. Indeed we find that if we interpolate all datasets onto a 0.22° or 0.44° grid (the resolution of the CORDEX ensemble), the evaluation produces total scores with correlation between 0.90 to 0.94 (Fig. 2.24, 2.25, 2.26 and 2.27). Fractional drought coverage does indeed produce worse scores at higher resolution, but scores that are in essentially the same rank-order as the coarser evaluation.

2.2.7 Hydrologic regions

Although our drought evaluation could be easily conducted over all hydrologic regions from WBD, in the present study we focus on three characteristically distinct regions – the California Region (CA), the Lower Mississippi Region (LM) and the New England Region (NE) (Fig. 2.1). These regions are chosen because of their relatively distinct climatological and topographical characteristics, as well as their potential vulnerability to drought. Specifically, the California Region is well known for its pronounced intra- and inter-annual climatological variability (a consequence of its Mediterranean climate), rough topography and substantial spatial heterogeneity, with hot deserts, humid coastline, rugged highlands, and a flat central valley (Fig. 2.1). California has often suffered from droughts and is projected to experience even worse drought conditions in the future (Ullrich et al., 2018; Williams et al., 2015). As a part of the southeastern US, the Lower Mississippi Region includes parts of Arkansas, Kentucky, Louisiana, Mississippi, Missouri, and Tennessee (Fig. 2.1) and is one of the most productive and diverse agricultural regions in the US. It accounts for a quarter of the total US cotton and two-thirds of the total US rice production (Committee et al., 2014; USDA, 2012) and

has a humid subtropical climate with ample rainfall and a spatially and temporally homogeneous climatology. However, it is also one of the most heavily irrigated land areas (7.1 million irrigated acres) over the CONUS (USDA, 2020) and often experiences droughts that result in substantial crop loss, typically driven by lower rainfall and higher temperatures occurring simultaneously over the growing season (Melillo et al., 2014; Mo, 2011; US Army Corps of Engineers, 2013; USDA, 2012, 2020). The most well-known example of droughts in this region is the 1988-89 drought, which even led to a stoppage of barge traffic over the lower Mississippi River (Changnon, 1989; Trenberth and Guillemot, 1996; Trenberth et al., 1988). Finally, the New England Region is one of the most populated and developed HUC2 regions in the US, accounting for about 20% of GDP and population with only 5% of the land area (Hobbs, 2008; US Bureau of Economic Analysis, 2016) (Fig. 2.1). It has a humid continental climate with abundant precipitation evenly spread throughout the year. However, there is significant spatial variation in precipitation between the inland and coastal areas, driven by its rugged topography. Although the Northeastern US (NEUS) is presently experiencing a wet period that has continued since the unprecedented 1960s drought (Seager et al., 2012), it’s largely acknowledged that droughts are not things of the past for NEUS (Hayhoe et al., 2007; Krakauer et al., 2019; Seager et al., 2012). In fact, several studies conclude the risks of severe droughts for the NEUS remain, with increased likelihood of short-term droughts under global warming (Frumhoff et al., 2007; Hayhoe et al., 2007; Krakauer et al., 2019). Overall, we anticipate the performance and biases of climate models to depend on each region’s temporal and spatial variability of precipitation, as well as their local geography. Therefore, these three regions are selected to sample a spread of regional climatologies – large temporal and spatial variability over mountainous topography (CA), small temporal and spatial variability over flat and even topography (LM), and large spatial but small temporal variability over complex topography (NE).

2.3 Statistical Methods

2.3.1 Scoring individual metrics

Standard statistical tests, including Kolmogorov-Smirnov (K-S) Test and Z Test (described in Supplement) are used to derive quantitative measures of performance for various characteristics of drought. These statistical tests also provide a means for absolute normalization of our performance metrics: Our “score” for each metric is defined as the absolute ratio of each test’s statistic to the critical value at 95% confidence level. As such, a score less than or equal to 1 indicates that we cannot reject the null hypothesis at this level. On the other hand, if a score is larger than 1 there is evidence that this particular metric is significantly different from observations. Given that scores are of approximately the same magnitude, this also means that scores convey comparable information across metrics. In general, for continuous metrics we will use K-S Test to examine if the model data has the same distribution as observed data. For discrete variables, we will use the Z test to test if the model has the same mean value as observed data. And for those variables involving proportion or probability, we will conduct one proportion Z test to see if model and observed data have the same proportion/probability, with the assumption that the proportion or probability derived from observational data is the real value.

2.3.2 Principal Feature Analysis (PFA)

Because of the multifaceted character of drought, multiple metrics are needed to capture all aspects of model performance. However, metrics within the same category are highly correlated, and so it’s necessary to select a subset of principal metrics so as to reduce redundancy in the total score calculation. In the context of dimension reduction, Principal Components Analysis (PCA) is commonly used (Bryant and Yarnold, 1995); however, PCA yields a set of orthogonal vectors consisting of the combination of original features instead of several selected original features. Therefore, here we employ Principal Feature Analysis (PFA) to select a subset of metrics used to calculate the total score.

PFA is a novel feature selection method based on PCA and the k -means unsupervised clustering

algorithm used to select the p original features that are most independent and can best represent the first q principal components produced by PCA (Lu et al., 2007; Song et al., 2010). To apply this technique, we firstly standardize all the metrics' scores and apply a PCA to see how many principal components (q) are needed to explain 95% variance. By applying the k -means algorithm with $k = p$, we cluster the first q principal components that explain 95% variance into p clusters. Then we choose the original feature which is closest to the mean of each selected cluster as the principal features. As suggested (Lu et al., 2007; Song et al., 2010), we set $p = q + 1$ because the number of features should be slightly higher than the number of principal components to explain the same variability.

2.3.3 Total Score

To better compare each model's performance within the evaluation region, we also introduce a Total Score which is defined as the sum of all principal metrics. By the reasonably performant model will have metrics with individual scores between 0 to 2 (with 1 again corresponding to the 95% confidence level). However, sometimes particularly strong disagreement arises and some metrics produce scores that are larger than 2. Left unattended, such large metrics would dominate the total score, and so we set an upper limit on the score of each metric of 2 in calculating the total score and define the *Total Score* as:

$$\langle \text{Total Score} \rangle = \sum_{\text{all metrics } i} \min(\langle \text{Score} \rangle_i, 2). \quad (2.1)$$

Under this modification, models with a significant deficiency in one metric can still achieve decent performance in the total score if they have smaller scores in the other metrics; nonetheless, care should be taken not to ignore such deficiencies as they may still be indicative of more severe issues in the model's treatment of drought. Also note that although a total score consisting of principal metrics is recommended, users can still customize the total score by self-selecting features of drought that they care about the most.

2.4 Drought Metrics

The metrics discussed here are selected to capture a handful of meaningful quantities from data that spans two dimensions in space and one in time. We expect the chosen metrics should only be expected to produce reasonable scores with at least 30 years monthly precipitation as input to meet the requirements of SPI and Central Limit Theorem (Supplement and (Guttman, 1999; Svoboda and Fuchs, 2016; Svoboda et al., 2012)). These metrics do rely on the specification of a region over which the evaluation takes place, which can (in essence) be any arbitrary shapefile of sufficient size to ensure our statistics are robust. In the text that follows we refer to this as the *evaluation region*.

Generally, drought refers to a prolonged time period with abnormally low precipitation (meteorological drought) and a shortage of water (hydrological drought), usually varies from months to years (Mishra and Singh, 2010; Wilhite and Glantz, 1985). In this study we focus specifically on meteorological drought, since hydrological drought indicators such as soil moisture and runoff are less frequently available and often inconsistently defined across models. For meteorological droughts, the *Standardized Precipitation Index (SPI)* has been widely used as a drought indicator to capture the drought’s beginning, end, frequency, intensity and probability at various time scales (Hayes et al., 2002; McKee et al., 1993; Svoboda and Fuchs, 2016; Ukkola et al., 2018). Using the SPI classification table (Table 2.1), drought magnitude can be easily assessed. *Regional SPI* and *gridpoint SPI* are, respectively, calculated from regional mean precipitation and each gridpoint’s precipitation via the Climate Indices Package (Adams, 2017). Because of the normalization, dry months ($SPI \leq -1$) refer to approximately 16% coverage over the whole time period (Hayes et al., 2002; McKee et al., 1993; Yihdego et al., 2019).

Our 11 metrics have been placed into 6 categories associated with different characteristics of drought. These categories and metrics are discussed in the following sections.

2.4.1 Monthly means

Monthly mean regional precipitation (A1. Mean Precip.) As precipitation is the most relevant upstream driver of drought, it is important to determine if models capture the distribution

Table 2.1: SPI Classification following Guttman (1999).

SPI Value	Conditions	Expected Frequency (Historical)
$SPI = 3.09$	Exceptionally Wet	Once every 83.3 years
$SPI \geq 2$	Extremely Wet	Once every 3.66 years
$2 \geq SPI \geq 1.5$	Very Wet	Once every 1.24 years
$1.5 \geq SPI \geq 1$	Moderately Wet	Once every 6.3 months
$1 \geq SPI \geq -1$	Nearly Normal	-
$-1.5 \leq SPI \leq -1$	Moderately Dry	Once every 6.3 months
$-2 \leq SPI \leq -1.5$	Very Dry	Once every 1.24 years
$SPI \leq -2$	Extremely Dry	Once every 3.66 years
$SPI = -3.09$	Exceptionally Dry	Once every 83.3 years

of monthly mean precipitation correctly. To evaluate this, we apply the K-S test (Supplement 2.8.1) to monthly regional precipitation (namely, the monthly mean precipitation of all grids' within the evaluation region) to evaluate if the comparative distributions of model data and observational data.

Regional SPI6 (A2. Mean SPI6) Regional SPI6 is the SPI6 calculated by the regional monthly mean 6 months' accumulative precipitation in the evaluation region, and is representative of the nature of intermediate term droughts. The model performance for this metric is again scored using the K-S test (Supplement 2.8.1).

Regional SPI36 (A3. Mean SPI36) Regional SPI36 is also evaluated analogous to SPI6, using the regional monthly mean 3-year accumulated precipitation in the evaluation region. This metric is selected to evaluate the model's performance at capturing longer-term droughts. Evaluation is also conducted by K-S test (Supplement 2.8.1).

2.4.2 Seasonality

Regional precipitation at each month (B1. Season Precip.) An issue with the metrics in category A is that they group all months into the same statistical sample and so cannot capture the seasonality of precipitation. However, drought seasonality determines the intra-annual distribution of drought, which is particularly important in regions of high climatic variability such as California.

Here drought seasonality is assessed by applying the K-S Test to each month separately and using the mean of the normalized score from each month to measure the model's performance.

Regional long term monthly mean (accumulated seasonality) (B2. LTMM) The long term monthly mean, a 12-month time series consisting of average precipitation in each month of the year, is commonly used to describe the seasonality of precipitation and droughts. By normalizing the LTMM by the total precipitation over the entire year, we obtain an accumulated fractional contribution through the given month, analogous to a cumulative distribution function. This then provides another means to test seasonality through calculation of the K-S test statistic for the difference between the long term monthly mean of model and observational data. Unlike B1, this metric focuses on the accumulated seasonality, which tends to emphasize the precipitation in the middle of year (the growing season).

2.4.3 Spatial character

Fractional Drought Coverage (C1. Frac. Cover.) Each evaluation region typically contains more than one grid point. Because of the variation in spatial characteristics of the domain, including topography and land surface type, different grid points will also tend to have differences in precipitation and susceptibility to dry conditions. Therefore, the fractional drought coverage metric is intended to evaluate if the model can simulate the spatial distribution of drought within a given region. Here, each grid point is deemed as dry when the SPI6 for that month is no larger than -1 (as Table 2.1). We define the ratio of dry grids to total number of grids at each month as the *Fractional Drought Coverage*:

$$\langle \text{Fractional Drought Coverage} \rangle = \frac{\langle \text{Number of grid points with SPI6} \leq -1 \rangle}{\langle \text{Total number of grid points within the evaluation region} \rangle} \quad (2.2)$$

The monthly dry grid ratio is then evaluated by the K-S Test statistic (Supplement 2.8.1) to assess if model's distribution is the same as observational data.

2.4.4 Drought frequency

Proportion of dry months (D1. Dry Frac.). Using the definition of SPI (Table 2.1), we define a dryness indicator via

$$\langle \text{Dry month indicator} \rangle = \begin{cases} 1, & \text{if regional } SPI6 \leq -1 \\ 0, & \text{if regional } SPI6 > -1 \end{cases} \quad (2.3)$$

The *proportion of dry months* is then the number of dry months divided by the total number of months:

$$\langle \text{Proportion of dry months over all years} \rangle = \frac{\langle \text{Number of months with regional } SPI6 \leq -1 \rangle}{\langle \text{Total number of months} \rangle} \quad (2.4)$$

To determine the score for this metric, we use the one proportion Z test (Supplement 2.8.1).

Annual number of dry months (D2. Dry Count) In order to assess the interannual variability of drought frequency, we define the annual number of dry months as the number of dry months (equation 2.3) from each year. Because all models have more than 60 years data, and so meets the criteria of the Central Limit Theorem, we calculate the Z test statistic (Supplement 2.8.1) to evaluate and normalize this metric.

2.4.5 Drought intensity

Intensity from SPI6 (E1. Intensity) The drought intensity is defined as the regional SPI6 values over all months when SPI6 is no larger than -1. To test if model can produce the same drought intensity distribution, we employ the K-S Test statistic (Supplement 2.8.1) to obtain the normalized score for this metric.

2.4.6 Probability of drought

Probability of drought initiation (and average non-dry period duration) (F1. Prob. Init.) The probability of drought initiation is defined as the probability that the following month is

dry given the current month is not dry. This is connected to the average non-dry period duration via

$$\begin{aligned}
& \langle \text{Probability of drought initiation} \rangle \\
&= P(\langle \text{dry month} \rangle | \langle \text{previous month is non-dry} \rangle) \\
&= \frac{\langle \text{Number of dry months when the last month is non-dry} \rangle}{\langle \text{Total number of months when previous month is non-dry} \rangle} \\
&= \langle \text{Average non-dry period duration in months} \rangle^{-1}.
\end{aligned} \tag{2.5}$$

This connection between drought initiation and the duration of the non-dry period emerges naturally if one treats each month as a sample from a geometric distribution. As is typically done with proportions, to normalize the metric we apply the one proportion Z test to the non-drought duration data.

Probability of drought termination (and average dry period duration) (F2. Prob. Term.) We define the probability of drought termination as the probability that the following month is not dry given the current month is dry:

$$\begin{aligned}
& \langle \text{Probability of drought termination} \rangle \\
&= P(\langle \text{non-dry month} \rangle | \langle \text{previous month is dry} \rangle) \\
&= \frac{\langle \text{Number of non-dry months when the last month is dry} \rangle}{\langle \text{Total number of months when previous month is dry} \rangle} \\
&= \langle \text{Average drought duration in months} \rangle^{-1}.
\end{aligned} \tag{2.6}$$

This equation is analogous to the probability of drought initiation, again emerging naturally from the geometric distribution. The metrics is also used to assess models' ability in capturing average drought duration. As with probability of drought initiation, we use the one proportion Z test to normalize the model's score.

Notably, the two metrics F1 and F2 are in essence derived from the transition probabilities of a two-state Markov chain, with states corresponding to non-dry and dry conditions.

2.4.7 Taylor Diagram and Taylor Score

In addition to the statistical metrics described above, we also compare our evaluation system with a more traditional evaluation method. The Taylor diagram is a widely used mathematical plot in earth model assessments that depicts correspondence between the model and observational data using a mathematical property that relates the standard deviation, the Pearson correlation coefficient and the root-mean-square error (RMSE) (Taylor, 2001). One way to quantify models' performance on the Taylor diagram is by the Taylor score,

$$\langle \text{Taylor Score} \rangle = \exp(-\alpha(1 - R) - \beta(\sigma + 1/\sigma - 2)) \quad (2.7)$$

where α and β are scaling factors that are set to 1 (the default value) in this study, R is the Pearson correlation between each gridpoint's mean precipitation, and σ is the model's normalized standard deviation of each gridpoint's mean precipitation (normalization is conducted by dividing this by the standard deviation from the observational data). Notably, the Taylor score uses the mean precipitation to measure performance but ignores temporal continuity (as we will see later). Unlike our statistical scores above, a higher Taylor Score indicates better performance. In this study, we rank all models from the best to the worst in the Taylor diagram and compare the results with our evaluation system.

2.5 Results

2.5.1 Principal Metrics Employed

To maintain consistency among study regions, we apply PFA to the CMIP6 dataset over the entirety of Continental US to select a set of principal metrics to apply analysis for all case evaluations. These metrics and their total scores will then be our primary means to evaluate each dataset's performance.

PCA shows 6 principal components are needed to explain 95% variance (Fig. 2.15). To match this level of variance, the PFA procedure gives a set of 7 principal metrics; as apparent from Fig.

2.2, most principal metrics, have less than 0.6 correlations with each other which confirms the reliability of PFA to select distinguish and significant metrics. Note that in Fig. 2.2, metrics within one category are highly correlated partially because it's over CONUS. And within each evaluation region (like the California Region), metrics in one category are less correlated as shown by Fig. 2.16.

PFA's identification of at least one metric from each category also confirms our natural intuition on the independence of these metrics. In the following sections, we will evaluate the performance of CMIP6, CORDEX and LOCA datasets over the California, the Lower Mississippi, and the New England Region based on these 7 principal metrics and their total score. We conclude a model "performs well" when its total score is less than the number of principal metrics, with a lower total score generally indicative of better performance. In the result heatmaps, individual metric scores less than 1, and total scores less than 7 are indicated in black font (and white font otherwise).

2.5.2 The validation of our evaluation system

In order to examine if our evaluation system can really distinguish the high-performance climate models, here we evaluate the ECMWF Reanalysis v5 (ERA5) monthly precipitation (European Centre for Medium-Range Weather Forecasts, 2020), which is the most advanced fifth generation ECMWF atmospheric reanalysis dataset, and processes numerous improvements and advances (Hersbach, 2016; Hersbach et al., 2020; Hoffmann et al., 2019). From Fig. 2.3, it's obvious that the highly performant reanalysis data is also identified to be excellent to capture droughts' features over most CONUS hydrologic regions (14 out 18). Even if for those regions where ERA5 has total scores larger than 7, the total scores are still relatively small (ranging from 7.07 to 8.32). This confirms that our system works well in capturing the drought features. Also, for these four hydrologic regions, most CMIP6 models (32/33 for Great Basin Region and Upper Colorado Region, 30/33 for Souris-Red-Rainy Region, and 24/33 for Great Lakes Region) cannot get good total scores (less than 7) too, which might be caused by the biases in CPC observed data, the deficiency to simulate certain atmospheric circulation in climate models over these regions. More studies are needed to explore the reasons why such a huge systematic discrepancy exists between CPC observed data and

most climate models and reanalysis data.

2.5.3 CMIP6 Performance

Our first application of the metrics package focuses on CMIP6 model performance.

CMIP6 Performance over the California Region Performance for the CMIP6 suite of models in the California Region is tabulated in Fig. 2.4. From the principal metrics alone, 24 out of 33 CMIP6 models produce a total score less than 7, implying that more than a half CMIP6 models perform well within this region. HadGEM3-GC31-LL is the best model, with the lowest total score and with only one principal metric slightly larger than 1. CESM2-WACCM is also an impressive model as it ranks 4 out of 33 and has all principal metrics less than 1.

Although overall CMIP6 performance is good, most models have a fractional drought coverage score larger than 1. There are two main reasons why CMIP6 models tend to perform so poorly on this metric over the California: First, the California Region is a region of significant spatial variability due to its complex mountainous topography with steep altitude gradient (particularly in the east of the state) which makes it hard to capture the spatial coverage of drought. As we can see from Fig. 2.5 (g), CMIP6 tends to obtain much better performance at C1. Frac. Cover. over low-lying regions without complex topography significantly (with p-value less than 0.03). Second, compared with the resolution of the CPC observations, CMIP6 models have much coarser spatial resolution that makes it particularly difficult to simulate the finer-scale geographic features present in the CPC data. Examining CMIP6 models' original resolution (defined by $\sqrt{dlon \times dlat}$ where $dlon$ is the model's longitudinal grid spacing and $dlat$ is the latitudinal grid spacing) versus fractional drought coverage scores, a weak but positive correlation emerges ranging from 0.14 to 0.28 (with p-value 0.12 to 0.44) in all 3 evaluation regions (Fig. 2.5 (a), (c), (e)). This indicates the models with finer resolution tend to better capture the fractional drought coverage and agrees with our hypothesis that coarse resolution is a limiting factor behind why CMIP6 cannot simulate drought coverage well. Moreover, nearly all the CMIP6 model have scores less than 1 for their proportion of dry months, probability of drought initiation and termination. One potential reason

why CMIP6 models tend to simulate these temporally-related metrics well is because the significant intra-annual and inter-annual climatological variability in California leads to a large normalization factor in the Z test. Also the region’s overall dryness and strong seasonal contrast makes it easy for models to correctly capture the probability of drought’s duration, initiation and termination. Therefore, CMIP6 models are generally trustworthy when it comes to simulating these drought features in the California Region, as opposed to spatial metrics like fractional drought coverage.

If we compare our results with the Taylor diagram which is based on each grid point’s mean precipitation (Fig. 2.6 (a)), we can see there exists obvious differences between conclusions obtained from the Taylor diagram and our system’s assessment. For example, BCC-ESM1 has a total score of 5.16 and is ranked as 5 out of 33 in our system; however, it is assessed as a low-performing model with rank 27 out of 33 in the Taylor diagram. This is mainly due to the low spatial standard deviation of each grid point’s mean precipitation, which is captured by the precipitation mean magnitude metrics in our framework (A1. Mean Precip., B1. Season Precip. and D1. Dry Frac.). In Fig. 2.4, although BCC-ESM1 exhibits poor performance for two mean magnitude metrics (A1. Mean Precip., B1. Season Precip.) and further shows poor performance at C1. Frac. Cover., it shows better performance at the temporally related metrics (especially F1. Prob. Init. and F2. Prob. Term.). This difference in scoring also confirms the value of our system, which can assess models’ performance on drought’s temporal features. On the other hand, some models like GISS-E2-1-H perform the worst in terms of both the Taylor diagram (Fig. 2.6 (a)) and under our metrics (Fig. 2.4). The Taylor diagram metrics, namely spatial standard deviation (precipitation mean magnitude) and spatial correlation of mean precipitation at each grid point, are not represented in our system and so it makes sense that there is a not strong correlation (-0.392) between our total score and the Taylor score (Fig. 2.5 (b)).

CMIP6 Performance over the Lower Mississippi Region Performance of the CMIP6 suite of models in the Lower Mississippi region is tabulated in Fig. 2.7. As we mentioned, this region is the flattest one over CONUS with small spatial variability of precipitation, thus it is perhaps not surprising that the CMIP6 models capture fractional drought coverage much better over this

relatively flat region than over California; 16 out of 33 models have a score less than 1 at fractional drought coverage in this region, which is the best among all the hydrologic regions of the Continental US. On the other hand, the Lower Mississippi Region has the most intense rainfall within the Continental US, with much less climatological variability. Such a smooth precipitation seasonality means it is difficult for CMIP6 to simulate drought frequency, duration and probability well. Thus most CMIP6 models exhibit poor performance at D1. Dry Frac., F1. Prob. Init. and F2. Prob. Term.. The poor performance among these metrics thus results in slightly worse performance across the CMIP6 ensemble as compared with the California region, with 15 out of 33 models have total scores less than 7. Comparing our total score to the Taylor diagram for this region, there still only exists a weak correlation (-0.418) between our total score and the Taylor score (Fig 2.5 (d)).

CMIP6 Performance over the New England Region Performance for the CMIP6 suite of models in the New England region is tabulated in Fig. 2.8. This weak temporal variability and precipitation seasonality leads to poor performance for temporal features in the CMIP6 models compared with the California Region (i.e. E1. Intensity and F2. Prob. Term.). The large spatial variability of precipitation explains why CMIP6 models tend to perform poorly with respect to spatial coverage (D1. Fractional Drought Coverage) as well. As a result, only 6 out of 33 models have total scores less than 7. Again, there is only a weak correlation (-0.225) between our evaluation results and the Taylor score (Fig. 2.5 (f)). We note that the poor performance for drought intensity and fractional drought coverage may be affected by poor representation of tropical cyclones in GCMs due their coarse resolution (Henderson-Sellers et al., 1998). Metrics to evaluate the representation of tropical cyclones and their associated precipitation in model data remain under development (Stansfield et al., 2020; Zarzycki et al., 2021), and can be employed to understand specific model deficiencies in this regard.

2.5.4 CMIP5 Performance

To quantify any improvements in CMIP6 in capturing the drought, we evaluate the same number (33) of CMIP5 models over all regions. As depicted in Fig. 2.17, 19 out of the 33 CMIP5 models

perform well over the California Region, which is less than CMIP6 (with 24 out of 33). In general, the improvements in CMIP6 are present in almost all principal metrics and total scores, with the multi-model average total score decreasing from 7.20 to 6.66. These improvements are perhaps not surprising given CMIP6 represents a later generation of models (Ahn et al., 2020; Eyring et al., 2016; Li et al., 2020; Priestley et al., 2020). However, others have shown only relatively modest improvement between CMIP5 and CMIP6 (Srivastava et al., 2020). The same process was also applied to the Lower Mississippi Region (Fig. 2.18) and the New England Region (Fig. 2.19), with the ratio of performant models increasing from 12/33 to 15/33 in LM but staying steady at 6/33 in NE. However, the means of total scores decreased from 8.08 to 7.59 over the LM Region and 9.25 to 8.76 over the NE Region. In fact, such improvements are present in nearly all hydrologic regions, suggesting that CMIP6 is the preferred ensemble for study of drought features.

2.5.5 CORDEX Performance

We now turn our attention to the CORDEX dataset, consisting of downscaled simulations from several regional climate models. In particular, we focus on using our evaluation system to understand whether or not dynamical downscaling improves the quality of simulated drought compared its driving model (CMIP5) and how these simulations compare to the CMIP6 ensemble.

CORDEX Performance over the California Region CORDEX performance is evaluated with the same principal metrics mentioned above (Fig. 2.9). The most obvious finding is that CORDEX achieves much better performance on fractional drought coverage within the California Region compared with CMIP6. There are 26/31 CORDEX models that perform well at C1. Frac. Cover. (compared with CMIP6’s 7 out of 33 and CMIP5’s 5 out of 33). This is perhaps unsurprising as the low CMIP5/6 model resolution is one main causes of poor performance when it comes to spatial coverage of drought. Namely, the higher model resolution improves how the model captures the spatial variability of precipitation over complex topography. Evidence for this result also emerges from CORDEX products with different resolutions. For example, MPI-M-MPI-ESM-MR.UQAM-CRCM5.22 is the downscaling product based on MPI-M-MPI-

ESM-MR, downscaled by UQAM-CRCM5 with a spatial resolution of 0.22° . Its performance when it comes to fractional drought coverage (0.25) is better than MPI-M-MPI-ESM-MR.UQAM-CRCM5.44's (0.47), which uses the same two models but with a coarser resolution. Nonetheless, 3 products don't follow this trend (CCCma-CanESM2.UQAM-CRCM5, UQAM-GEMatm-CanESMsea.UQAM-CRCM5 and HadGEM2-ES.NCAR-WRF), because we interpolate all CORDEX models onto the same 1° grid which masks the benefits of finer resolution products in capturing spatial features. In the sensitivity analysis with different resolutions, if we interpolate all models into a 0.22 or 0.44° grid (the original grid used in CORDEX), finer products always get better performance in fractional drought coverage than their coarser analogues (see detailed information at Supplement).

At first glance, there appears to be clear added benefit in using higher resolution products for assessing the spatial character of drought/precipitation. However, we also find that for the other principal metrics, CORDEX doesn't provide improvement compared with CMIP5/6 – even yielding worse performance for average drought intensity. For example, the mean average drought intensity score from the CORDEX dataset is 2.20, which is larger than CMIP5's (1.80) and CMIP6's (1.48). Further, while the difference in overall performance is much smaller, CMIP6 still tends to produce the best overall performance with an average total score of 6.66, beating CMIP5 (7.20) and CORDEX (6.83). Therefore, it seems that dynamical downscaling doesn't provide significant benefit in representation of drought features (and may even degrade some features' representation). The exception is in the fractional drought coverage, which does show some benefit from finer model resolution (Supplement). Nonetheless the degradation of performance is perhaps unsurprising, as droughts are primarily a product of synoptic-scale meteorological drivers which are largely prescribed by the driving model. That is, this appears to be a combination of “garbage in, garbage out” and systematic biases from regional downscaling at play (Giorgi and Mearns, 1999; Hall, 2014; Pontoppidan et al., 2018; Rummukainen, 2010; Switanek et al., 2017). Notably, RCMs are known to possess significant biases when it comes to precipitation over mountainous topography (Caldwell et al., 2009; Maraun and Widmann, 2015); consequently we will see later that CORDEX performance is much better over homogeneous and flat regions (e.g. the Lower Mississippi Region). Our

sensitivity analysis using a common resolution grid at 0.44° and 0.22° resolution further supports these findings (Supplement).

We do acknowledge that the mediocre overall performance of CORDEX compared with CMIP5 over the California Region may also be due to selection bias. That is, only a small subset of the 33 CMIP5 models were chosen as CORDEX drivers, and these may not be representative of the larger ensemble. However, according to Mearns (2020), the CMIP5 drivers used in CORDEX were chosen because they were highly performant. To better understand if this is the case we selected the 6 CMIP5 models used as drivers in CORDEX and tabulated their performance in Fig. 2.9. Compared to their CMIP5 drivers we find that CORDEX products all have better performance at fractional drought coverage. However, in most cases, CORDEX is actually worse for the other principal metrics and total scores. For example, the average total score of the 6 CMIP5 drivers is 6.05, far smaller than the CORDEX's (6.83) and the whole set of 33 CMIP5 models' (7.20). This verifies that CORDEX CMIP5 drivers are performant models and supports our claim that dynamical downscaling does not always improve the quality of simulated drought in these models (Giorgi and Mearns, 1999; Hall, 2014; Pontoppidan et al., 2018; Rummukainen, 2010; Switanek et al., 2017). Additionally, these results also appear to confirm that systematic biases from dynamical downscaling do degrade the fidelity of simulated drought in regions of complex terrain (Caldwell et al., 2009; Maraun and Widmann, 2015).

Using the Taylor diagram and score to evaluate CORDEX performance (Fig. 2.10 (a)) versus its CMIP5 drivers' performance (Fig. 2.10 (a)), we find analogous results to our evaluation. For example, CORDEX models achieve better performance for spatial metrics (i.e. spatial correlation) than their CMIP5 drivers as a result of their finer resolution. However, the mean magnitude metrics (e.g. mean precipitation standard deviation) is not obviously improved and even degraded due to systematic biases in the RCMs.

CORDEX Performance over the Lower Mississippi Region As in the California Region, CORDEX simulations in the Lower Mississippi also provide significant improvement in the fractional drought coverage; however, unlike California, here CORDEX also shows improvement for

other metrics (Fig. 2.11). In this flat region the systematic biases from the RCMs are smaller than the improvements that come from higher resolution, which also follows previous studies that systematic biases in regional precipitation simulations are much larger over complex topography (Caldwell et al., 2009; Maraun and Widmann, 2015). In total, 21 out of 31 CORDEX models perform well here with a total score less than 7, compared with 15 out of 33 CMIP6 models and 12 out of 33 CMIP5 models. The improvement from the CORDEX models also holds against their CMIP5 drivers (Fig. 2.11); except HadGEM2-ES, all other CORDEX models perform better than their associated CMIP5 drivers. As CORDEX improves on the CMIP models for essentially all of our metrics, it's perhaps no surprise that these models also show improvements using the Taylor diagram and score (Fig. 2.10 (b)) compared with CMIP6's (Fig. 2.6 (b)) and their CMIP5 drivers (Fig. 2.10 (b)).

CORDEX Performance over the New England Region The New England Region has high spatial precipitation variability and complex terrain (mountains and coasts). In line with the other two regions analyzed, we expect CORDEX to have better performance on fractional drought coverage but worse performance for other metrics (Fig. 2.12). In general, CORDEX produces worse performance for nearly all other metrics, with only 4 out of 31 CORDEX models having a total score less than 7 (compared with 6 out of 33 CMIP5 and CMIP6 models). But perhaps surprisingly, even for the fractional drought coverage, CORDEX doesn't exhibit any obvious improvement. This is likely because the New England Region has a more complex topography and precipitation distribution, as well as coastal effects contributing to precipitation variability (Agel et al., 2015). In the Taylor diagram (Fig. 2.10 (c)) we also see although CORDEX is evaluated to be better than its CMIP5 drivers (Fig. 2.10 (c)), it still has clearly worse performance compared with CMIP6 (Fig. 2.6 (c)); CORDEX models generally have larger standard deviation and spatial correlation.

2.5.6 LOCA Performance

Although LOCA is also a downscaled product of CMIP5, LOCA has much finer resolutions (1/16°) compared with CORDEX, and uses a statistical downscaling method based on local analogues

combined with bias correction (Pierce et al., 2014). As such it is not surprising that LOCA’s climatology tends to match the historical climate better than direct model simulations – nonetheless, this does not necessarily mean that LOCA does a better job of capturing future change as it relies on a functional relationship between coarse GCM and regional climate data that may not hold in the future (i.e., the stationarity problem). Further, unlike dynamical downscaling, this method doesn’t attempt to preserve relationships among different variables as variables are corrected independently (Prasanna, 2018).

Over the California Region (Fig. 2.13), LOCA achieves much better performance than both CMIP6 and CORDEX in not only the fractional drought coverage but across all principal metrics. All LOCA models actually produce total scores below 7, and the best LOCA model (IPSL-CM5A-LR with 2.70 total score) also outperforms the best models from CORDEX (CCCma-CanESM2.UQAM-CRCM5.22 with 4.67 total score), CMIP5 (MPI-ESM-MR with 4.35) and CMIP6 (HadGEM3-GC31-LL with 4.07 total score). There’s little doubt that LOCA captures the historical drought climatology better than these other products. Of course, its performance is strongly related to the use of bias correction (Pierce et al., 2014), finer resolution and avoidance of biases from direct simulation of physical processes.

From Fig. 2.14 (a), it’s apparent that when evaluated with the Taylor Diagram all models are compressed to the same point and it’s impossible to distinguish performance. This problem persists in other evaluation regions as well, and arises because LOCA uses statistical downscaling and correction based on the observational data’s mean fields of precipitation; however, the Taylor diagram only evaluates precipitation mean fields. Therefore, a Taylor diagram cannot be used for distinguishing LOCA products – the results are effectively identical to comparing CPC with Livneh.

Results from LOCA are similar in the Lower Mississippi Region (Fig. 2.21 and Fig. 2.14 (b)) and the New England Region (Fig. 2.22 and Fig. 2.14 (c)).

2.6 Conclusions

To help with selecting datasets that are more likely to capture the features and process drivers of intermediate-term drought, we have put forward a comprehensive evaluation framework that can

be easily used within an arbitrary evaluation region, given more than 30 years monthly precipitation data. In total, 11 metrics have been defined encapsulating drought’s monthly mean, seasonality, spatial coverage, frequency, intensity, duration and probability. Principal Feature Analysis (PFA) is suggested to select principal metrics which are significant and relatively independent; however users can customize the metrics used to meet their needs. A framework based on statistical hypothesis testing is used to normalize each metric and determine model performance. By employing this framework to four climate datasets – CMIP5/6, CORDEX and LOCA over three characteristically distinct hydrologic regions in CONUS, we have developed a better understanding of the sensitivity of model performance to region and dataset. This enables us to provide guidance to stakeholders and data users on the selection of suitable models, and furthers our understanding of how various factors impact model biases in capturing the drought features.

Our evaluation system complements evaluation using ETCCDI drought indices and spatial analysis of precipitation using the Taylor diagram and score. We evaluate a variety of drought characteristics that depend only on monthly precipitation fields, including temporal characteristics such as duration and occurrence/termination probability. We argue that these characteristics are highly relevant for climate adaptation planning, and so should be included in any evaluation system. Notably, the total scores from our proposed system are only weakly correlated with evaluation using Taylor scores, with correlations between -0.12 to -0.42 across the three evaluation regions.

The system proposed here is a unique combination of statistical hypothesis testing, drought indices, and a comprehensive multi-metric framework addressing various essential characters of drought. It has several significant advantages:

1. The system can evaluate models’ performance on temporal continuity and probabilistic features of drought which are often ignored in traditional evaluation. It can also be used on arbitrarily large or small regions (data permitting), accounting for the fact that drought is often widespread.
2. By using statistical hypothesis testing, the system builds a standard criteria to assess absolute performance of models for each metric. Consequently, we can easily determine if one model

produces drought that is similar to observational data and compare models' performance across multiple categories of models.

3. The system provides a method to analyze the performance of statistical downscaling products (like LOCA) while a traditional method like a Taylor score shows essentially identical performance. This is because traditional methods usually only evaluate precipitation means which are modified by the bias correction process.
4. The system is flexible depending on user need. Users can either use the principal metrics selected by PFA, which can explain most variance of all metrics, or customize the metrics used according to which drought features they need. And it's easy to conduct evaluations based on whatever regions (shapefile) or datasets (NetCDF) the users are interested in.

The major conclusions of this study are as follows:

- We have confirmed our system indeed assesses different drought features, and that the metrics selected via PFA represent distinct measures of model performance. Using ERA5 data we have validated that it can successfully identify that a high quality reanalysis product (ERA5) captures drought features well over most hydrologic regions (14 out of 18). Among the four remaining regions, the total scores are still reasonable (from 7.07 to 8.32).
- Evaluations of CMIP5/6 and CORDEX confirm our hypothesis that the performance of climate models in capturing regional drought features depends highly on the characteristics of the evaluation regions. For example, 32 out of 33 CMIP6 models, and even the most advanced reanalysis dataset (ERA5) cannot capture drought features well over the Great Basin Region and Upper Colorado Region (that is, they get a total score no larger than 7).
- Climate models do an excellent job in simulating the temporal continuity features like drought duration and probability over regions with a clear precipitation seasonality and strong precipitation variability like the California Region, but performance on temporal features is poor over regions with weak temporal variability. Certainly, it's intuitive that initiation and termination of drought is easier to simulate if precipitation follows a strong seasonality. Also, the

significant intra-annual and inter-annual climatological variability can lead to a large normalization factor in the statistical tests employed. On the other hand, CMIP5/6 models generally better simulate the spatial features of droughts over flat regions with less spatial variability since their typically coarse resolution (between 0.5° and 2.8° for CMIP6) prevents them from capturing the finer spatial features represented in the CPC observational data (0.25°) over the regions with complex topography and significant altitude gradients.

- Following recent advancements in climate modeling systems (Balaji et al., 2018; Eyring et al., 2016), we also find CMIP6 outperforms its precursor – CMIP5 – in all regions and in nearly all metrics employed. EC-Earth3 demonstrates particularly high performance and is identified as the best CMIP6 model, with strong performance in all three evaluation regions. Overall, even compared with CMIP5-based dynamical downscaling products (CORDEX), CMIP6 still produced comparable or better performance in temporal continuity of droughts. With that said, the coarser resolution of CMIP6 models produces worse performance in the fractional drought coverage.
- For the uncorrected CORDEX runs we find that, compared with CMIP5/6, both our evaluation system and the Taylor diagram indicates that dynamical downscaling significantly improves models’ performance on spatial metrics (i.e., spatial correlation in Taylor diagram and fractional drought coverage). However, CORDEX models do not always exhibit better performance on magnitude mean and temporal continuity metrics. And the degradation of performance mainly exist over regions with complex topography and large latitude gradient. Given that droughts are largely driven by synoptic meteorology, this problem appears to be, at least in part, because of the “garbage in, garbage out” problem (Hall, 2014). CORDEX models often produce worse performance on these metrics in regions with complex topography and precipitation inhomogeneity due to systematic errors common in dynamical downscaling systems over mountainous and coastal regions (Caldwell et al., 2009; Maraun and Widmann, 2015; Velasquez et al., 2019). But in the flattest hydrologic region examined – namely, the Lower Mississippi region – the RCMs consistently exhibited overall improvements in their

representation of droughts.

- The LOCA dataset produces far lower scores (better performance) across essentially all metrics. This is unsurprising as bias correction has been employed to match the historical climatology. However, care still needs to be taken in the employ of LOCA for future projections, as the statistical relationships employed in LOCA may not hold in the future and LOCA doesn't attempt to preserve physical relationship among different variables, as variables are corrected independently (Prasanna, 2018).
- By examining the sensitivity of our results to different evaluation periods and common grid spacing (Supplement), we find our conclusions are largely independent of these choices. As long as the study period is larger than 30 years (the minimum requirement of Central Limit Theorem and Standardized Precipitation Index), modifying the study period from 1948-2014 to 1970-1999 and 1970-2009 yields total scores that are highly correlated (0.88 and 0.93, Supplement). Further, when interpolating CMIP6 and CORDEX to finer resolutions (0.22° or 0.44°) the total score correlations are similarly strong (0.90 to 0.94, Supplement).

Notably, the evaluation system proposed here only takes monthly precipitation as input data, which may not account for the impacts of temperature and evapotranspiration on drought; therefore, future work will focus on expanding the suite of available metrics, and examining how this system can be employed for long-term (multi-year) droughts.

The evaluation system proposed in this study is available as an open source software package implemented in Python (https://github.com/Zeyu88/Drought_Metrics). Users can easily conduct evaluations by inputting monthly precipitation and observational data fields (NetCDF files) along with the precipitation variable name, evaluation regions (shapefiles), principal metrics to be calculated and the evaluation period. We recommended PFA be performed first over all evaluation regions to select the principal metrics used. This package also provides a function to draw Taylor diagrams for comparison. Prior to evaluation, both model and observational data must be interpolated to the same resolution (provided by our package).

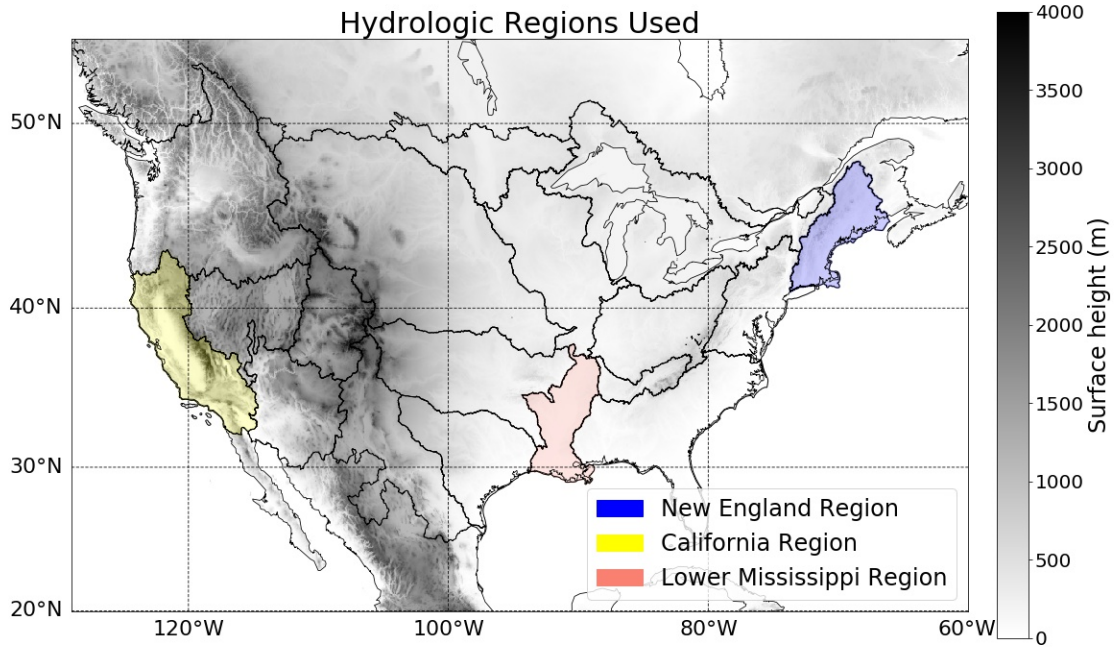


Figure 2.1: Hydrologic regions used in this study

2.7 Acknowledgements

This research was supported by the RGMA program area(s) in the U.S. Department of Energy’s Office of Biological and Environmental Research as part of the multi-program, collaborative Integrated Coastal Modeling (ICoM) project. PAU was supported by Department of Energy Office of Science award number DE-SC0016605, “A Framework for Improving Analysis and Modeling of Earth System and Intersectoral Dynamics at Regional Scales (HyperFACETS).” This project is also supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, hatch project under California Agricultural Experiment Station project accession no. 1016611. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for

archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

2.8 Supplements

Fig. 2.15 is the number of principal components versus variance explained. In this paper, we try to use principal metrics selected by PFA that explain 95% of the variance. Therefore, as indicated by Fig. 2.15, we need 6 principal components to capture this level of variance. Following Lu et al. (2007); Song et al. (2010), we would use 7 principal features because the number of principal features should be slightly higher than the number of principal components needed to explain the same variability. Fig. 2.16 is the cross-correlation heatmap of all metrics over the California Region from CMIP6 models. We can see that within each category, metrics are highly correlated. But across different categories, metrics are relatively independent. Notably, the principal metrics selected by PFA are from different unique categories, which further confirms that our metrics indeed capture different features of drought. Fig. 2.17 - 2.19 depict the principal metrics and total score from the CMIP5 datasets over the three evaluation regions. Fig. 2.20 is the Taylor diagram of CMIP5 datasets over three hydrologic regions. Fig. 2.21 and 2.22 are the principal metrics and total score from LOCA datasets over the Lower Mississippi Region and the New England Region.

2.8.1 Statistical hypothesis testing

Kolmogorov-Smirnov (K-S) Test

The two-sample K-S test is employed in our analysis for assessing differences between two cumulative distribution functions (CDFs). The K-S Test is one of the most useful and commonly employed non-parameteric hypothesis tests to determine if two samples (here one is from model and the other from observations) have the same empirical distribution (Brands et al., 2013). Its statistic is

defined by

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|, \quad (2.8)$$

where $D_{n,m}$ is the K-S Test statistic, \sup is the supremum function, and $F_{1,n}(x)$ and $F_{2,m}(x)$ are the empirical distribution functions of the model and the observed data, respectively.

The K-S test indicates that there is evidence against the null hypothesis (namely, that the two samples are drawn from the same distribution) if

$$D_{n,m} > c(\alpha) \sqrt{\frac{n+m}{n \cdot m}}, \quad (2.9)$$

where n and m are the number of samples from the model and observational data, and $c(\alpha)$ is 1.358 at the 95% confidence level. Thus we calculate the normalized score by equation

$$\text{Score} = \frac{D_{n,m}}{c(\alpha) \sqrt{\frac{n+m}{n \cdot m}}}. \quad (2.10)$$

Thus if the score is greater than 1, this is the same as saying there is evidence that the null hypothesis does not hold, which may indicate poor model performance.

Z test

The Z test is a hypothesis test used to determine if two normally distributed sample populations share the same mean. Although many of our metrics are not normally distributed, the Z test may nonetheless be employed if the sample size is much larger than 30 (as per the Central Limit Theorem) (Islam, 2018; Pollard, 1982). Due to the relatively long time coverage (more than 60 years) of climate models and CPC observed data, this is generally not an issue for the datasets examined in this paper. Nonetheless, care should be taken if the metrics in this paper are employed on datasets with fewer samples.

Here we use the one proportion Z test (equation 2.11) and two sample Z test (equation 2.12) separately for different situations. The one proportion test is used to normalize the probability related metrics such as the probability of drought initiation and termination with the null hypothesis

that model and observational data have the identical proportions or probability. The two sample Z test is used with the null hypothesis that observe and test data have the same mean values.

The equation governing the one proportion Z test is

$$Z = \frac{|p_1 - p_2|}{\sqrt{p_1 \times (1 - p_1) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \quad (2.11)$$

where p_1 is the proportion of samples in model data (i.e. number of events divided by number of samples), p_2 is the proportion of samples in the observational data, and n_1 and n_2 are the sample sizes of the model and observational data. The equation governing the two sample Z test is

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}. \quad (2.12)$$

where \bar{x}_1 and \bar{x}_2 are the mean value of the metrics from model and observations, and σ_1 and σ_2 are the computed standard deviations of the metrics from model and observations. In both cases the score is calculated as

$$\text{Score} = \frac{Z}{Z_{crit}}. \quad (2.13)$$

where Z_{crit} is the critical value of Z test and its value at 95% confidence level is 1.96. Therefore, if the Z Score is large than 1.96, there is evidence to suggest the model and observations have different means or probability, indicating that the model is in disagreement with observations. On the other hand, if the normalized score is less than 1, we can't reject the possibility that model and observations have the same means or probability, and so the model is scored more highly.

2.8.2 Stanardized Precipitation Index (SPI) Calculation

SPI is calculated in accordance with Guttman (1999); McKee et al. (1993). The parameters of gamma distribution are calculated separately for each month and then the resulting distribution is employed to transform the accumulated precipitation to be normally distributed. To do so we first compute i^{th} accumulated precipitation ($P_{n,i,k}$) at month k using the i^{th} monthly mean precipitation

$(P_{i,k})$ via

$$P_{n,i,k} = \sum_{i=i-n}^i P_{i,k}. \quad (2.14)$$

Since the non-zero $P_{n,i,k}$ is nearly Gamma distributed at each month, we will transform month k 's $P_{n,i,k}$ to be normally distributed using the transform as following and calculate the probability of non-zero values ($P_{n,i,k,nonzero}$) and zero values ($P_{n,i,k,zero}$) separately

$$\langle \text{Probability}_{zero,k} \rangle = \frac{\langle \text{Number of samples with } P_{n,i,k} \leq 0 \rangle}{\langle \text{Total number of samples of } P_{n,i,k} \rangle} \quad (2.15)$$

$$\langle P_{n,i,nonzero,k} \rangle = \begin{cases} P_{n,i,k}, & \text{if } P_{n,i,k} > 0 \\ NaN, & \text{if } P_{n,i,k} \leq 0 \end{cases} \quad (2.16)$$

$$a_k = \log \mu_k - \overline{\log P_{n,i,nonzero,k}}, \quad \alpha = (1 + \sqrt{1 + 4 \times a_k/3})/(4 \times a_k) \quad \beta_k = \mu_k/\alpha_k \quad (2.17)$$

$$\text{Probability}_{n,i,k,gamma} = \text{cdf}_{gamma}(P_{n,i,k,nonzero}, \alpha_k, \beta_k) \quad (2.18)$$

$$\begin{aligned} & \langle \text{Probability}_{n,i,k} \rangle \\ &= \begin{cases} \text{Probability}_{k,zero} + ((1 - \text{Probability}_{k,zero}) \times \text{Probability}_{n,i,k,gamma}), & \text{if } P_{n,i,k} > 0 \\ \text{Probability}_{k,zero}, & \text{if } P_{n,i,k} \leq 0 \end{cases} \end{aligned} \quad (2.19)$$

$$SPI_{n,i,k} = \text{ppf}_{normal}(\text{Probability}_{n,i,k}) \quad (2.20)$$

where μ_k and σ_k are the mean value and standard deviation of $P_{n,i,k,nonzero}$ at month k from the model or observational data, α_k and β_k are factors in Gamma distribution at month k ,

$\overline{\log P_{n,i,k,nonzero}}$ is the mean of $\log P_{n,i,k,nonzero}$, $Probability_{n,i,k,gamma}$ is the probability of $P_{n,i,k,nonzero}$ under the Gamma distribution, $Probability_{k,zero}$ is the probability of zero values, $Probability_{n,i,k}$ is the probability of all samples (including non-zero and zero values) at month k , cdf_{gamma} is the cumulative Gamma distribution function, ppf_{normal} is percent point function (inverse cdf) of the Normal distribution, $SPI_{n,i,k}$ is the i^{th} SPI_n value at k month. In this paper, we use SPI6 and SPI36 for capturing intermediate- and long-term meteorological droughts.

2.8.3 Sensitivity analysis of total score to study period

To confirm the period used here (1948-2014) is sufficient to evaluate models' quality and performance on simulating drought, we evaluated model performance using two shorter time periods: the 30-year period from 1970-1999, the 40-year period from 1970-2009, and the 50-year period from 1960-2009. By comparing all models' total scores within California Region for each period (Fig. 2.23) we see that our evaluation scores are largely stable with more than 30 years of data. Model scores are highly correlated among evaluation periods (between 0.88 and 0.93), with correlations that increase with the duration of overlap. However, we also see that models tend to worsen uniformly over longer time periods, attributed to the short period (1970-1999) being relatively wet in California (1.77 mm/day on average). However, both extended periods 2000-2014 and 1948-1969 contain extreme droughts with much lower mean precipitation (1.57 and 1.68 mm/day). These prolonged dry periods skew the statistics towards dry conditions, which is less frequently simulated in the GCMs. Nonetheless, 30 years of data appears to be sufficient to roughly assess the model's ability to simulate the multi-month droughts modeled with SPI6.

2.8.4 Sensitivity analysis of model performance to common grid resolution

By construction, the fractional drought coverage metric requires that all compared datasets must have the same number of grid points and the same hydrologic boundaries. To enable comparison between data products across scales, we interpolate all datasets onto a 1° grid. Considering all of our metrics are computed at the regional scale, except for fractional drought coverage, we anticipate that interpolation will not significantly impact the evaluation results and our conclusions. To confirm this

is the case we conduct a sensitivity analysis of common interpolation grids and model performance. Here both CMIP6 and CORDEX models are interpolated onto 0.44° and 0.22° grids (the resolutions employed in the original CORDEX datasets) to assess the effect on model performance. The results presented here are from the California Region, but the results of other hydrologic regions are analogous.

CMIP6 over the California Region

CMIP6 interpolated to a 0.44° grid From Fig. 2.24 (as compared with Fig. 4), it's apparent that the different common grids have very small impacts on the total scores and most principal metrics. The correlation between each model's total score is 0.94 with only small deviations in the magnitude of the score and resultant ranking. However, all CMIP6 models' performance for fractional drought coverage does universally deteriorate (with all models producing scores larger than 2) if we interpolate CMIP6 into a 0.44° grid. This is perhaps not unexpected because of the crude nature of the interpolation performed; an interpolation method that respected the underlying spatial features of the region should be employed if the model is taken to high resolution.

CMIP6 interpolated to a 0.22° grid Analogous results are found if we instead interpolate CMIP6 to a 0.22° grid (Fig. 2.25). Notably, the correlation between each model's total score is still high (0.92) which again demonstrates our evaluation is stable to the resolution of the common grid. Notably, the fractional drought coverage scores are poor for all CMIP6 models, which indicates this simple interpolation scheme applied to CMIP6 data does not capture the spatial features at fine spatial scale.

CMIP6 performance with different interpolation resolution

From Fig. 2.26, we can directly see that the total scores are not sensitive to the interpolation resolution used but fractional drought coverage will degrade because of poor representation of the underlying spatial features.

2.8.5 CORDEX over the California Region

The other concern related to the employ of a common 1° grid is that such a relatively coarse grid might lead to degraded performance for high-resolution dynamical downscaling products (CORDEX). Therefore we also interpolate CORDEX data onto 0.44° and 0.22° grids (the two resolutions employed in the original CORDEX datasets) to examine if the finer common grids will impact our conclusions regarding CORDEX.

CORDEX interpolated to a 0.44° grid

From Fig. 2.27, we can see that, compared with Fig. 9, the impacts from different common grids are small for most principal metrics and the total score, which is unsurprising since most principal metrics describe features at the region scale. When CORDEX models are interpolated into a 0.44° grid, the range of total scores only changes from $[4.67, 9.68]$ to $[4.67, 9.69]$. For each model, the total score also does not change significantly, and the correlation between each model's total score is 0.93. Thus we claim our evaluation methodology is stable and model performance of CORDEX models is largely unchanged at coarser resolution. We also find that if we interpolate CORDEX towards a finer resolution, some advantage is seen in fractional drought coverage for CORDEX models with finer resolution (0.22°). Again, this doesn't mean that CORDEX model with finer original resolution always outperform the coarser ones in other metrics and their total scores (Giorgi and Mearns, 1999; Hall, 2014; Pontoppidan et al., 2018; Rummukainen, 2010; Switanek et al., 2017). Regardless of the resolution, these models are further subject to well-documented biases in RCMs when it comes to precipitation over mountainous topography (Caldwell et al., 2009; Maraun and Widmann, 2015).

CORDEX interpolated to a 0.22° grid

The same results still emerge if we interpolate CORDEX to a 0.22° grid (Fig. 2.28). Namely, total score changes little overall. Again the correlation between each model's total score is still very large (0.90), which again provides evidence supporting the choice of 1° grid for our combined analysis.

As with the CMIP6 models, we notice that when interpolating to finer grids, CORDEX models tend to produce a worse score in fractional drought coverage.

CORDEX performance with different interpolation resolution

Fig. 2.29 supports our findings that model performance is stable to interpolation to different common grids, and the observation that fractional drought coverage scores do improve when models are interpolated into a coarser resolution. This supports the robustness of our conclusions in this study, and suggests that poor CORDEX model performance is more likely attributed to their driving models and internal biases (Giorgi and Mearns, 1999; Hall, 2014; Pontoppidan et al., 2018; Rummukainen, 2010; Switanek et al., 2017).

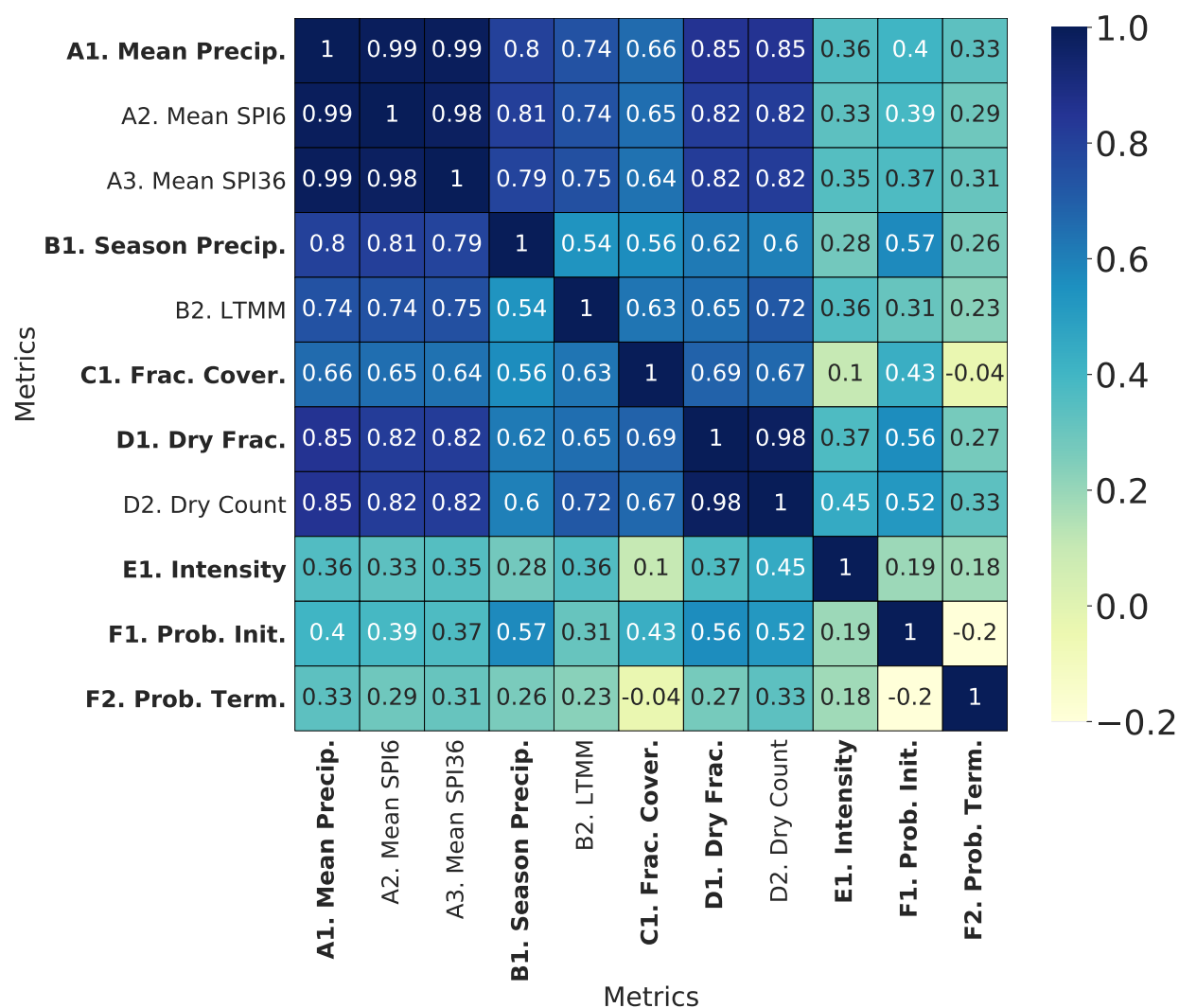


Figure 2.2: Cross-correlation heatmap of all metrics calculated over the continental United States from CMIP6 models (bold metrics are principal metrics selected by Principal Feature Analysis).

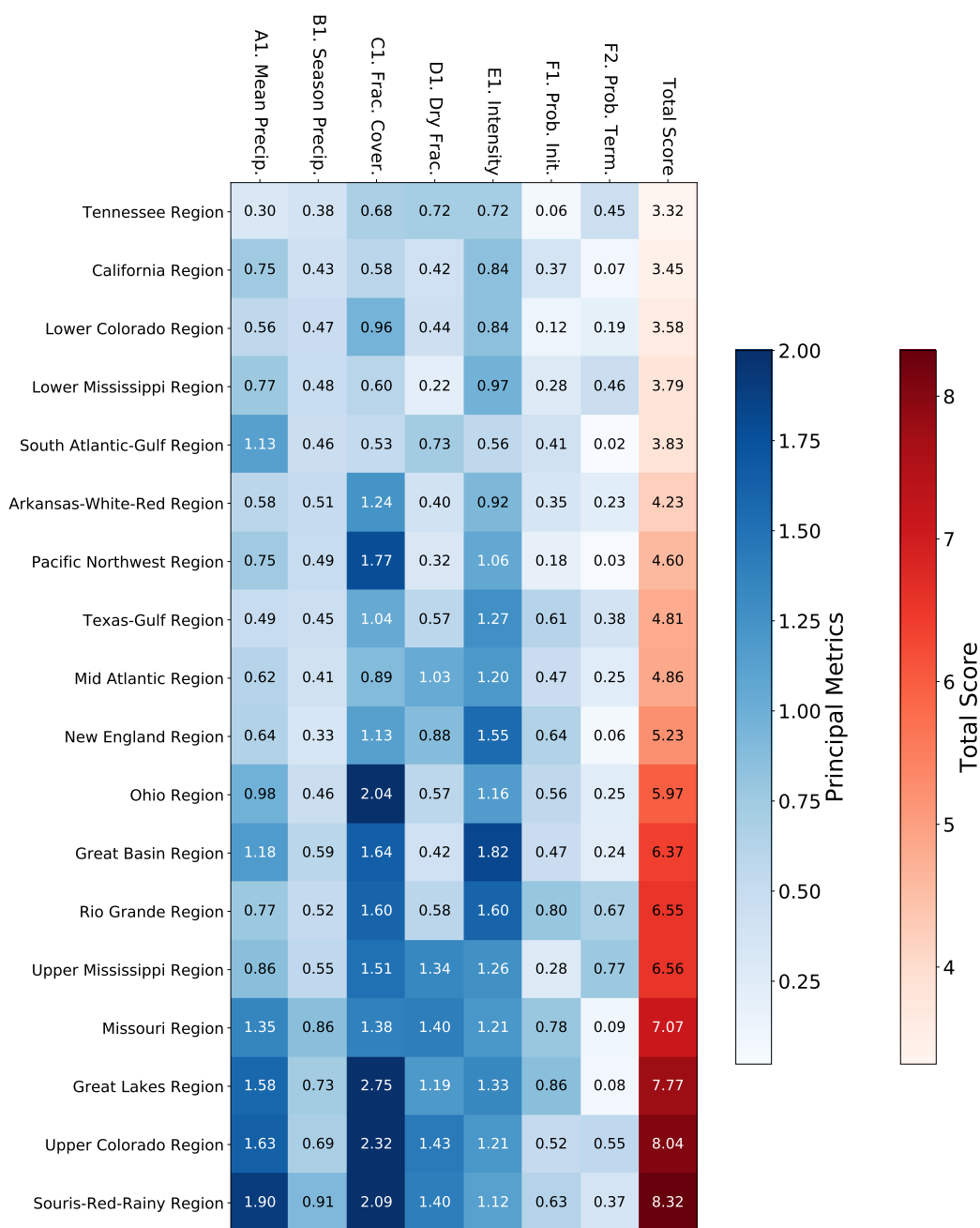


Figure 2.3: Principal metrics and total score of ERA5 reanalysis data over all eighteen hydrologic regions over CONUS (the performance is identified as good when individual metric scores less than 1, and total scores less than 7 are indicated in black font (and white font otherwise)).

2.8. SUPPLEMENTS

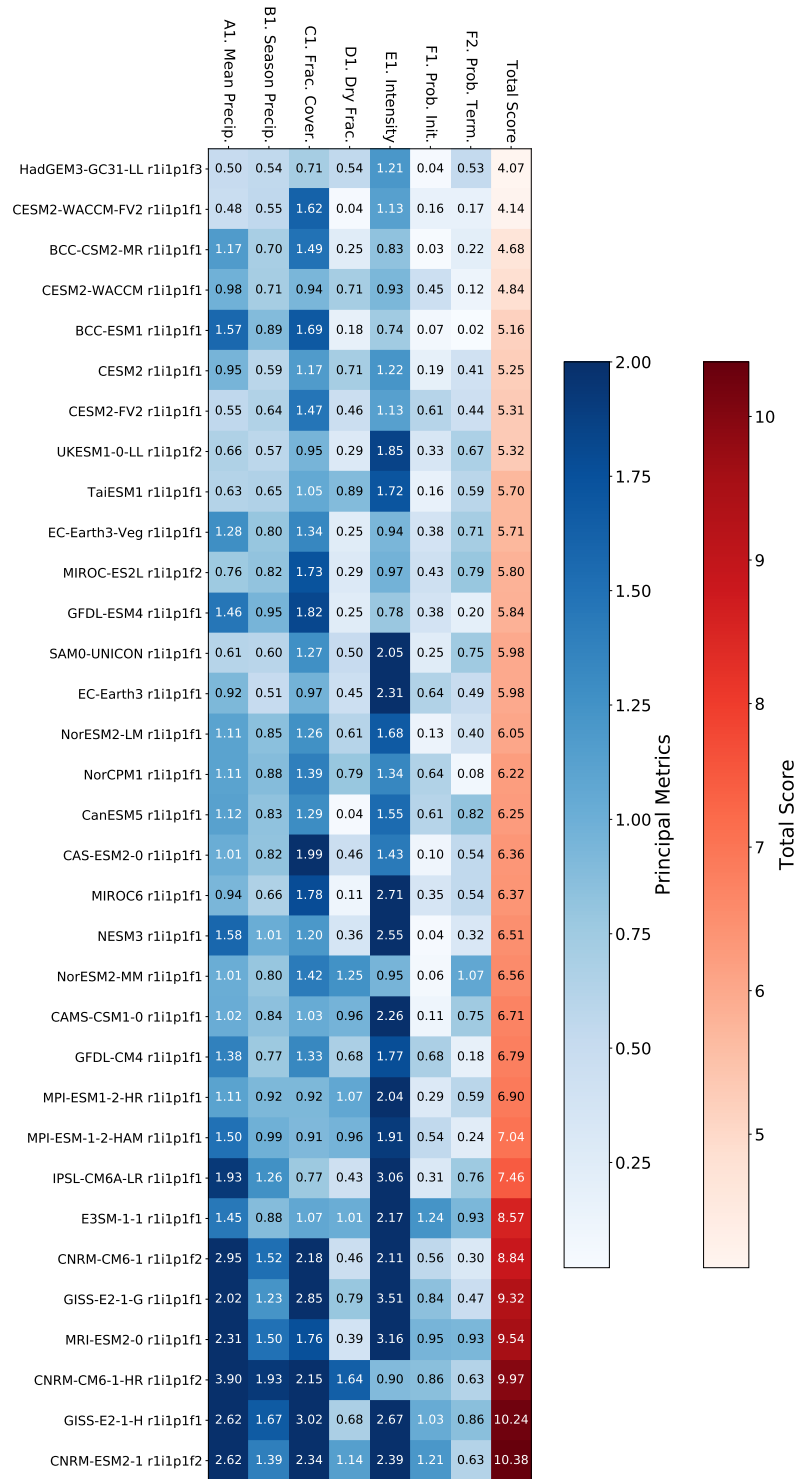


Figure 2.4: Principal metrics and total score of CMIP6 datasets over the California Region (the performance is identified as good when individual metric scores less than 1, and total scores less than 7 are indicated in black font (and white font otherwise)).

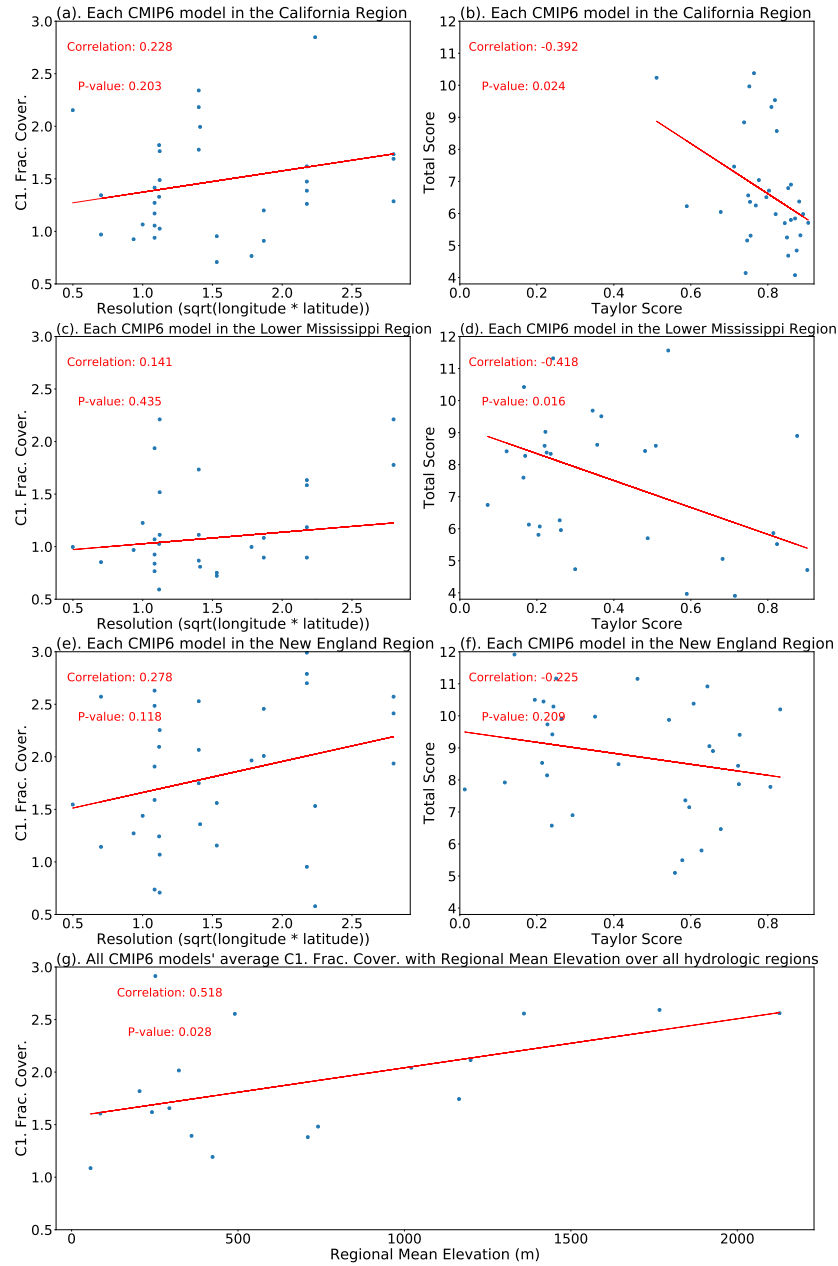


Figure 2.5: Scatter plots where correlations appear between model's and region's characteristics and evaluation results. (a). Each CMIP6 model's C1. Frac. Cover. score versus its original resolution in the California Region, (b). Each CMIP6 model's total score versus Taylor score in the California Region, (c). Each CMIP6 model's C1. Frac. Cover. versus its original resolution in the Lower Mississippi Region, (d). Each CMIP6 model's total score versus Taylor score in the Lower Mississippi Region, (e). Each CMIP6 model's C1. Frac. Cover. score versus its original resolution in the New England Region, (f). Each CMIP6 model's total score versus Taylor score in the New England Region, (g). The average scores of all CMIP6 models' C1. Frac. Cover. scores versus evaluation region's elevation mean over all eighteen CONUS hydrologic regions

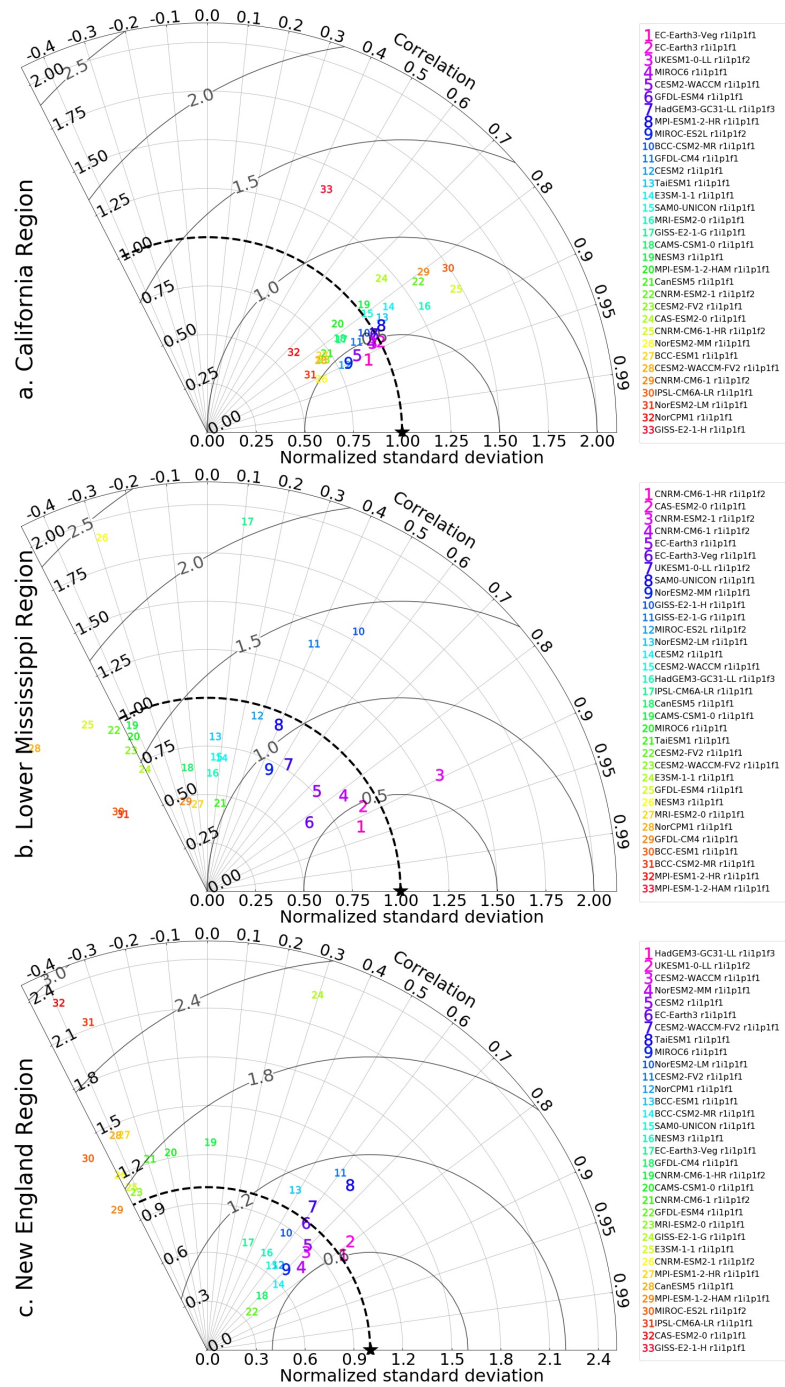


Figure 2.6: Taylor diagram of CMIP6 datasets over three hydrologic regions, which is based on each grid point's monthly mean precipitation. Models are labeled starting from 1 based on their Taylor scores from the highest to the lowest; so the index 1 corresponds to the model with highest Taylor score and best performance identified by Taylor diagram. (a). Over the California Region. (b). Over the Lower Mississippi Region. (c). Over the New England Region.

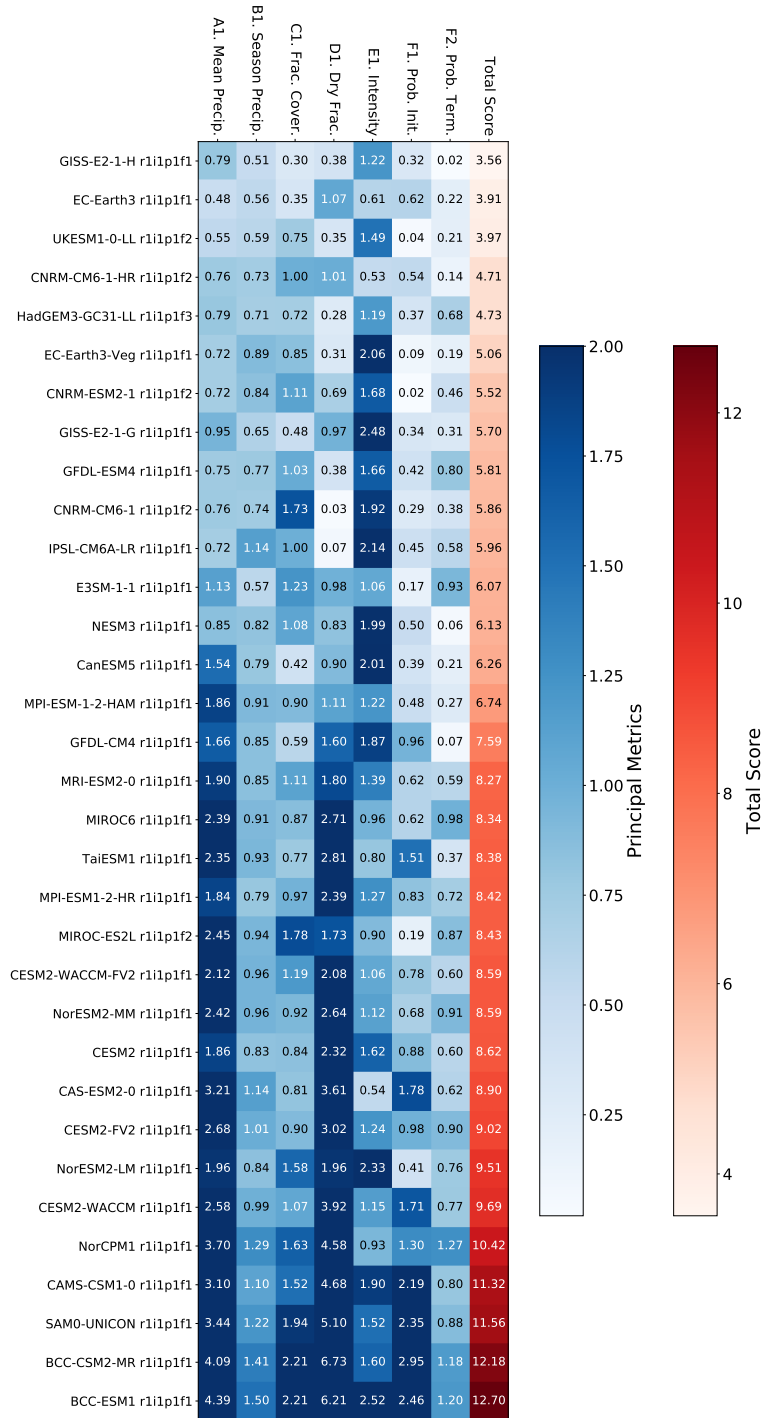


Figure 2.7: Principal metrics and total score of CMIP6 datasets over the Lower Mississippi Region (the performance is identified as good when individual metric scores less than 1, and total scores less than 7 are indicated in black font (and white font otherwise)).

2.8. SUPPLEMENTS

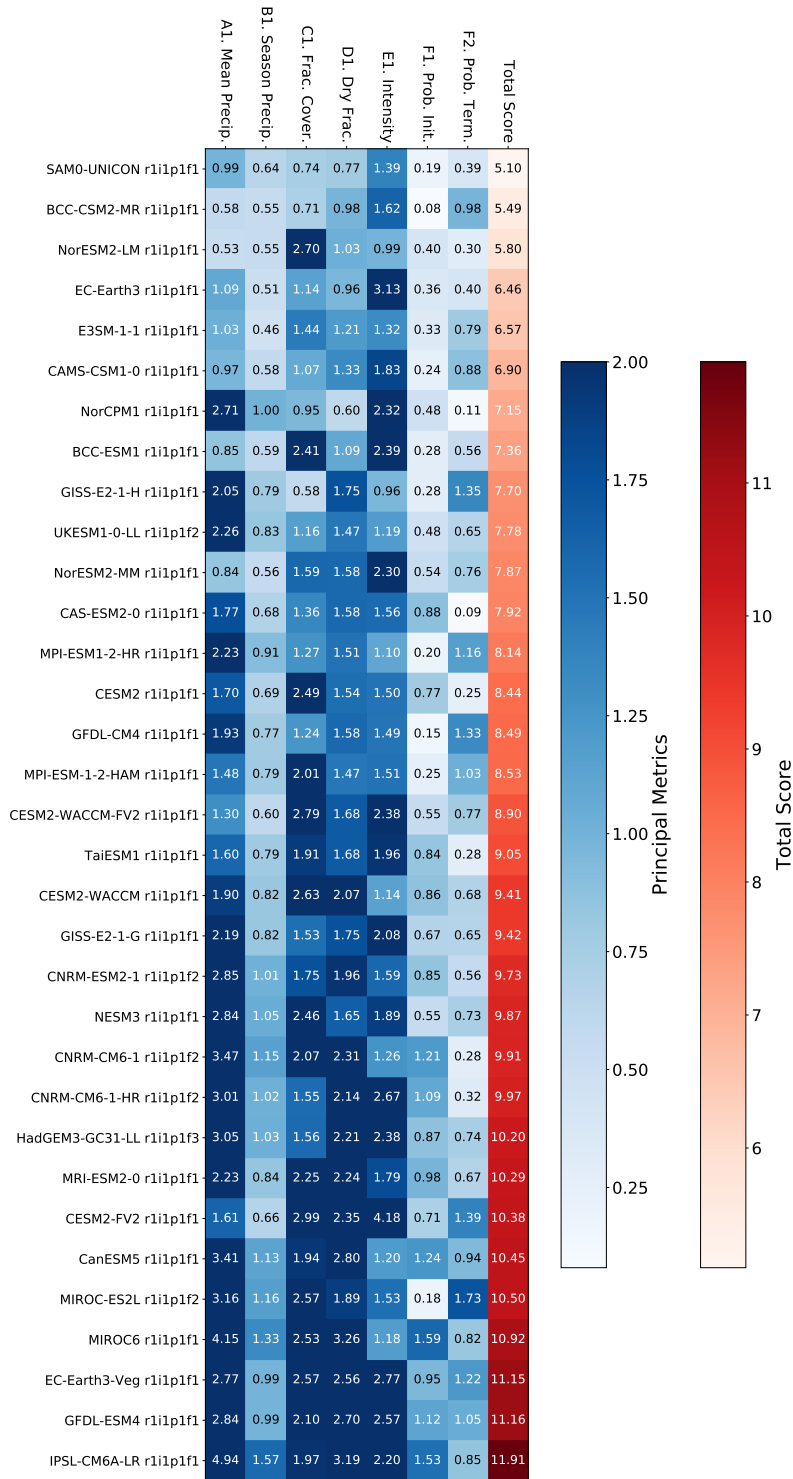


Figure 2.8: Principal metrics and total score of CMIP6 datasets over the New England Region (the performance is identified as good when individual metric scores less than 1, and total scores less than 7 are indicated in black font (and white font otherwise)).

2.8. SUPPLEMENTS

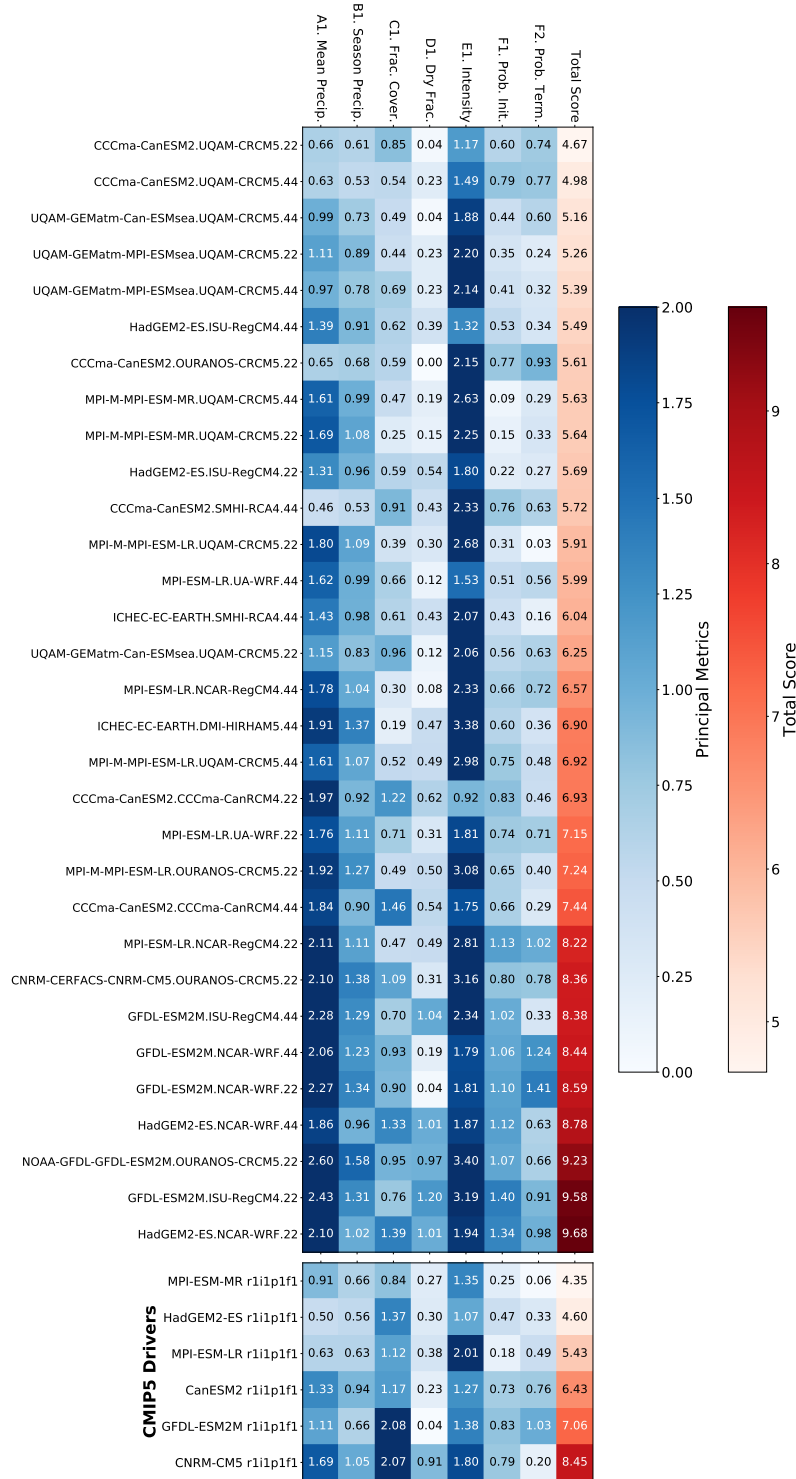


Figure 2.9: Principal metrics and total score of CORDEX datasets and CMIP5 drivers over the California Region (the performance is identified as good when individual metric scores less than 1, and total scores less than 7 are indicated in black font (and white font otherwise)).

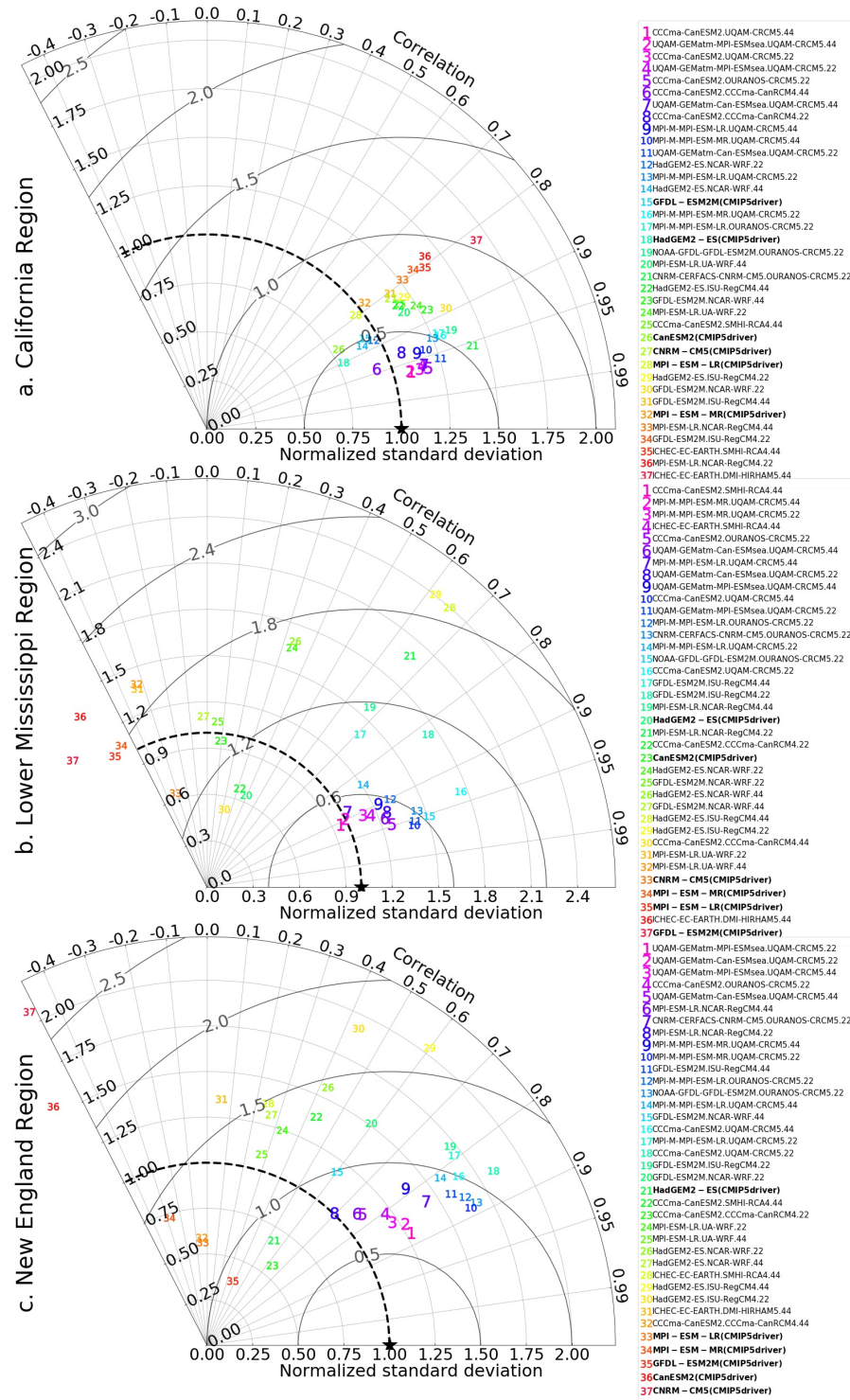


Figure 2.10: Taylor diagram of CORDEX datasets and their drivers over the three hydrologic regions (CMIP5 drivers are bolder). (a). Over the California Region. (b). Over the Lower Mississippi Region. (c). Over the New England Region.

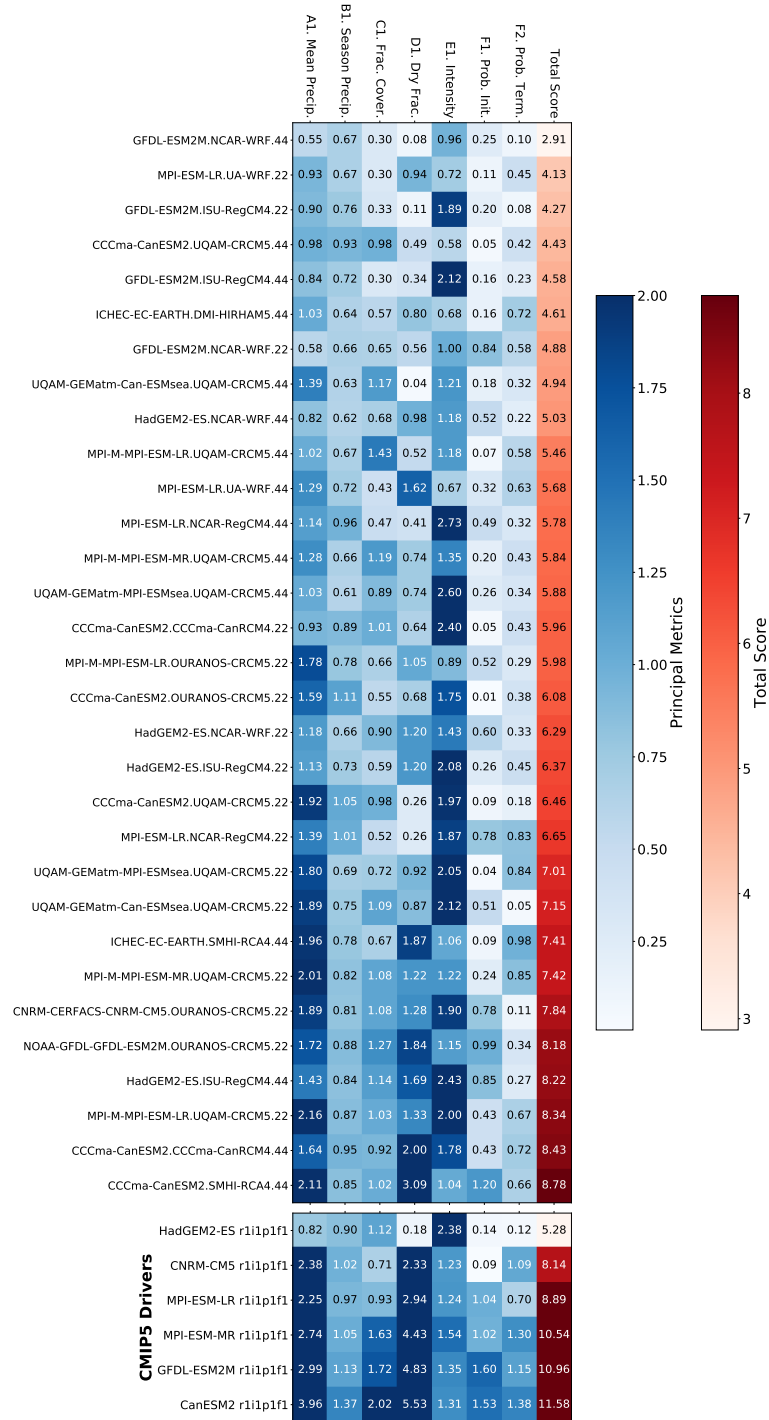


Figure 2.11: Principal metrics and total score of CORDEX datasets and CMIP5 drivers over the Lower Mississippi Region (the performance is identified as good when individual metric scores less than 1, and total scores less than 7 are indicated in black font (and white font otherwise)).

2.8. SUPPLEMENTS

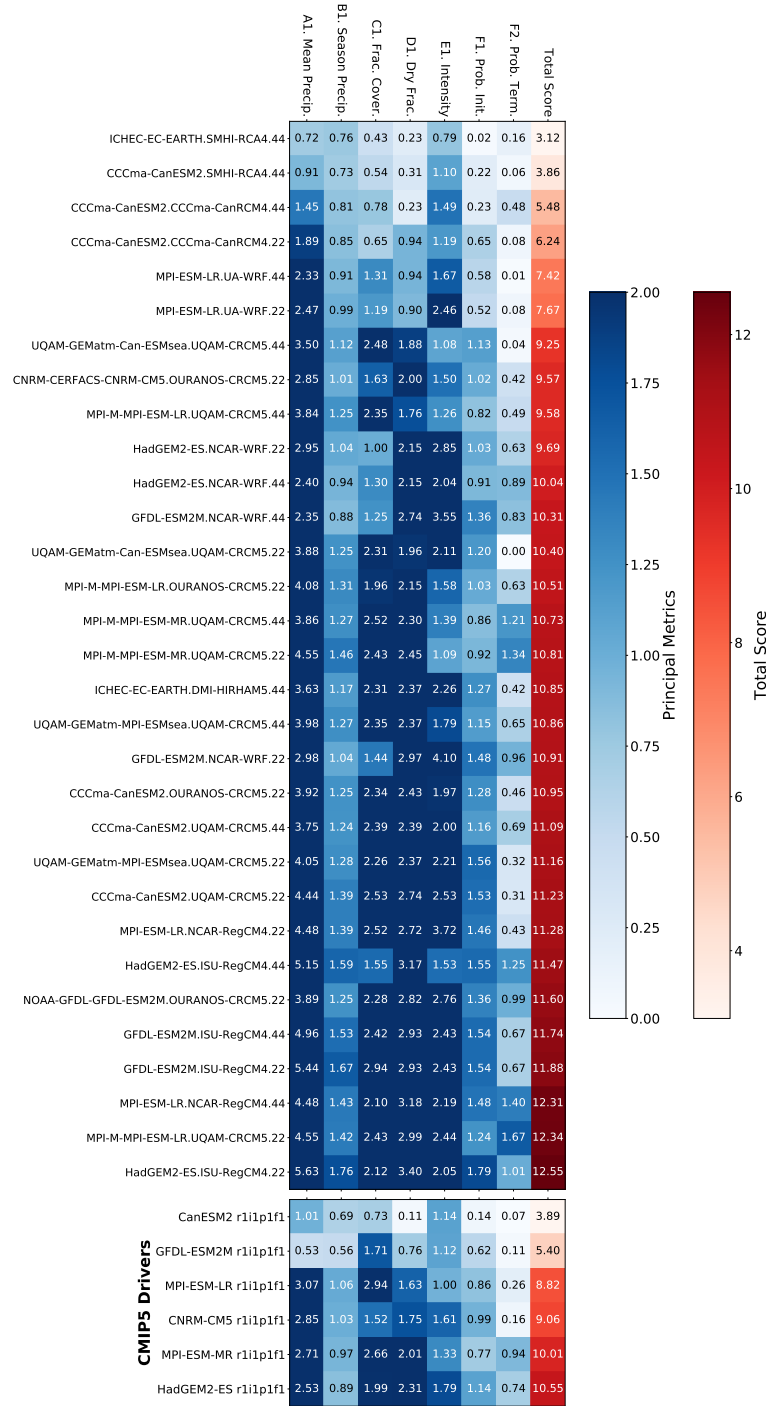


Figure 2.12: Principal metrics and total score of datasets and CMIP5 drivers over the New England Region (the performance is identified as good when individual metric scores less than 1, and total scores less than 7 are indicated in black font (and white font otherwise)).

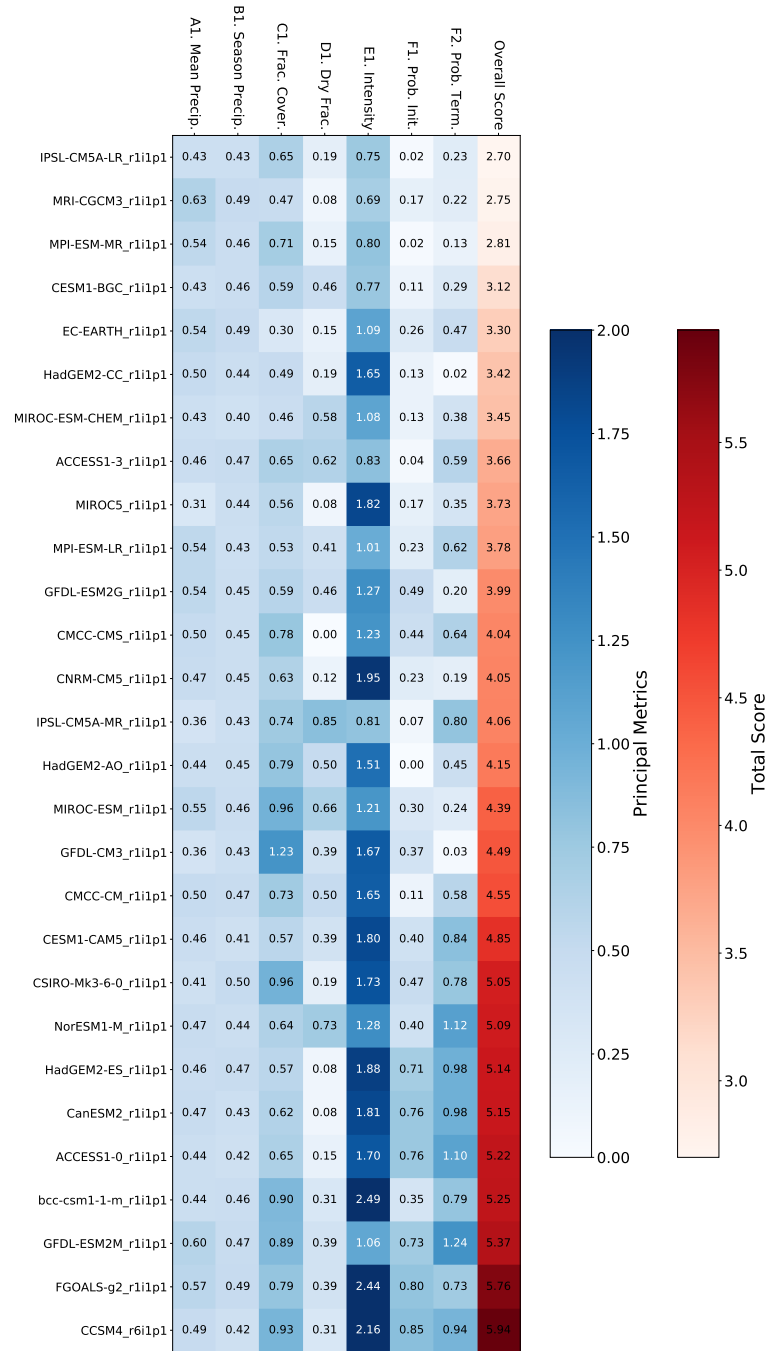


Figure 2.13: Principal metrics and total score of LOCA datasets over the California Region (the performance is identified as good when individual metric scores less than 1, and total scores less than 7 are indicated in black font (and white font otherwise)).

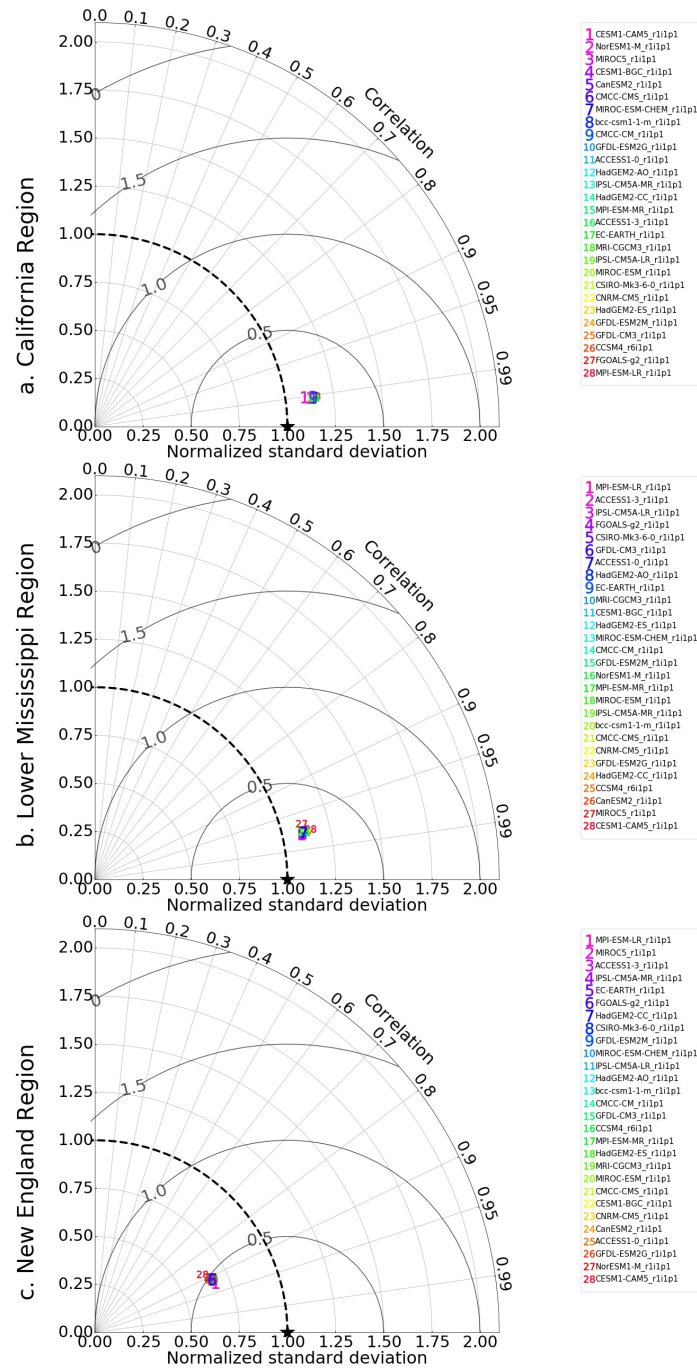


Figure 2.14: Taylor diagram of LOCA datasets over three hydrologic regions. (a). Over the California Region. (b). Over the Lower Mississippi Region. (c). Over the New England Region.

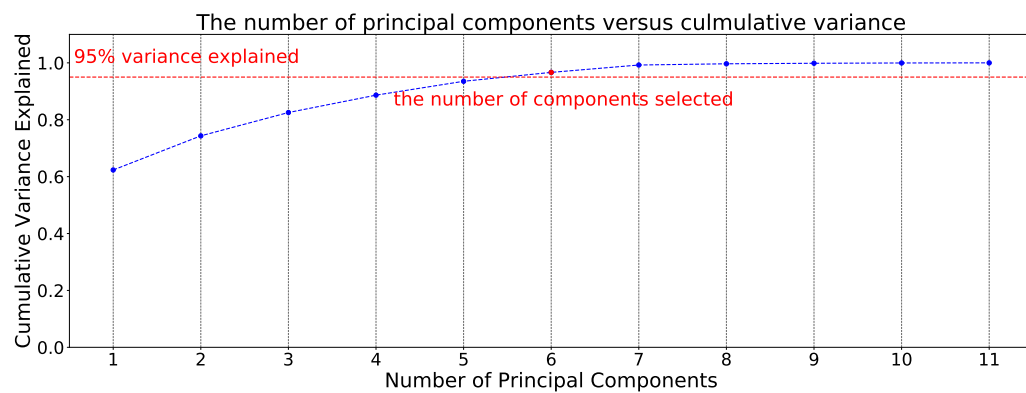


Figure 2.15: The number of principal components versus variance explained (the red dash line indicates 95% cumulative variance explained).

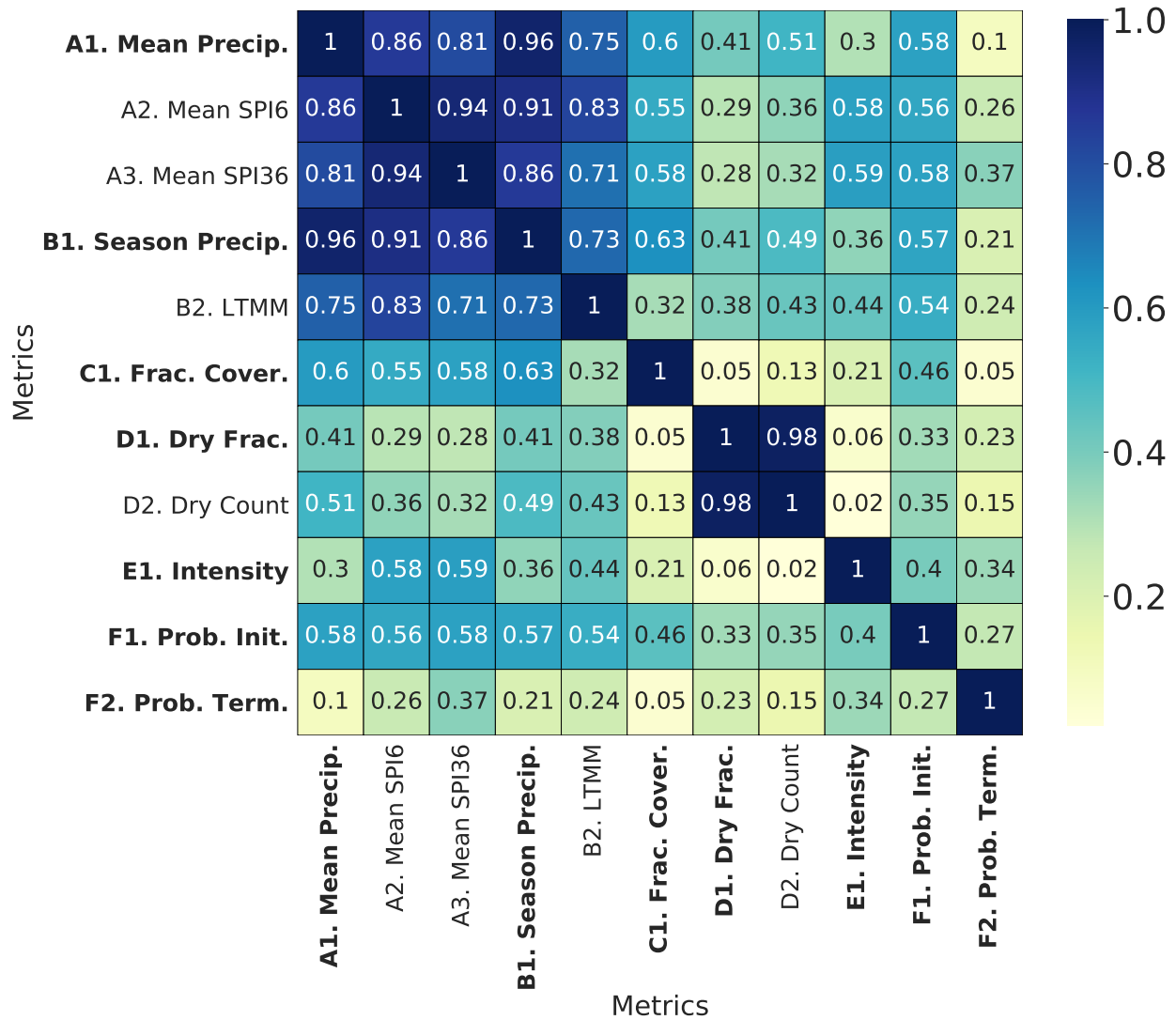


Figure 2.16: Cross-correlation heatmap of all metrics over the California Region from CMIP6 models (bold metrics are principal metrics selected by Principal Feature Analysis).

2.8. SUPPLEMENTS

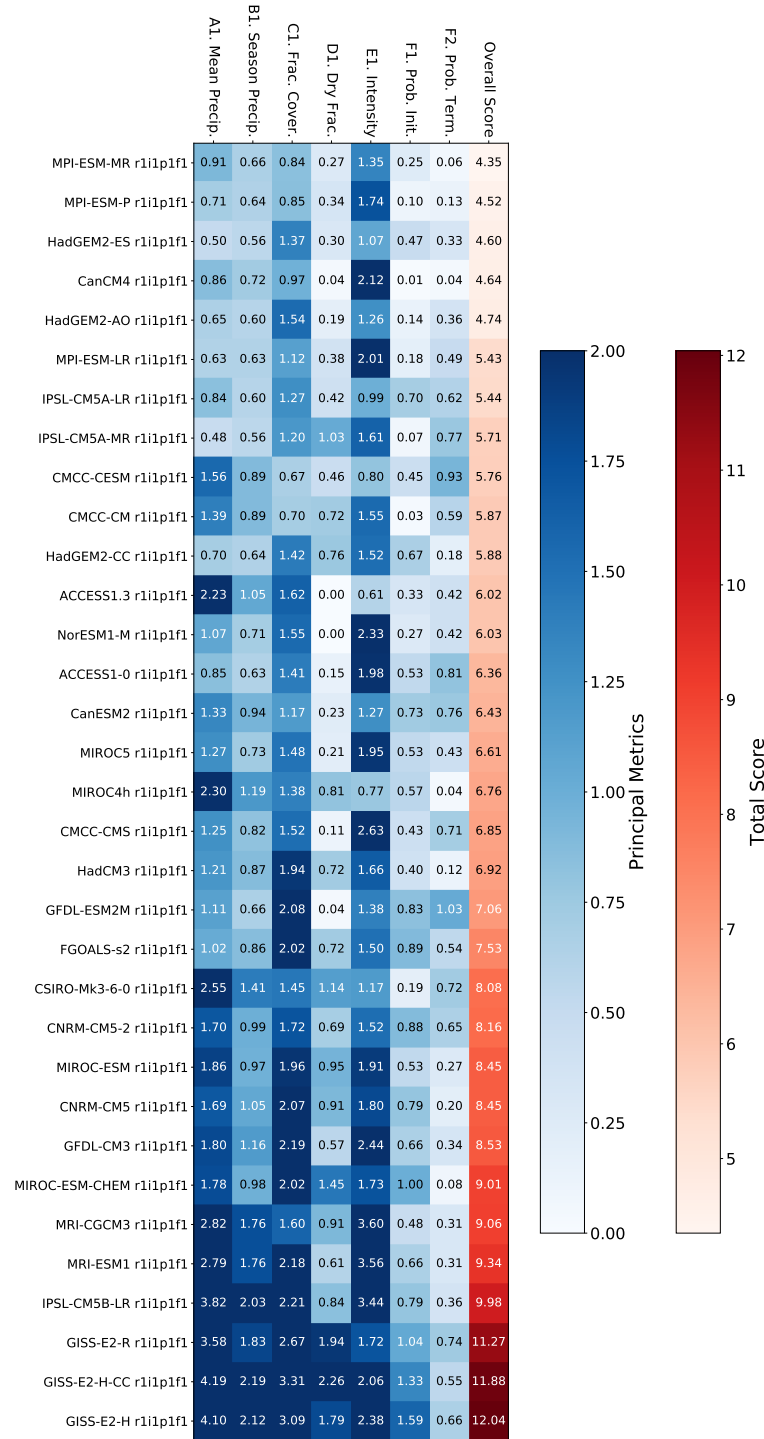


Figure 2.17: Principal metrics and total score from the CMIP5 dataset over the California Region (individual metric scores less than 1, and total scores less than 7 are indicated in black font, while scores not meeting this criteria are indicated in white).

2.8. SUPPLEMENTS

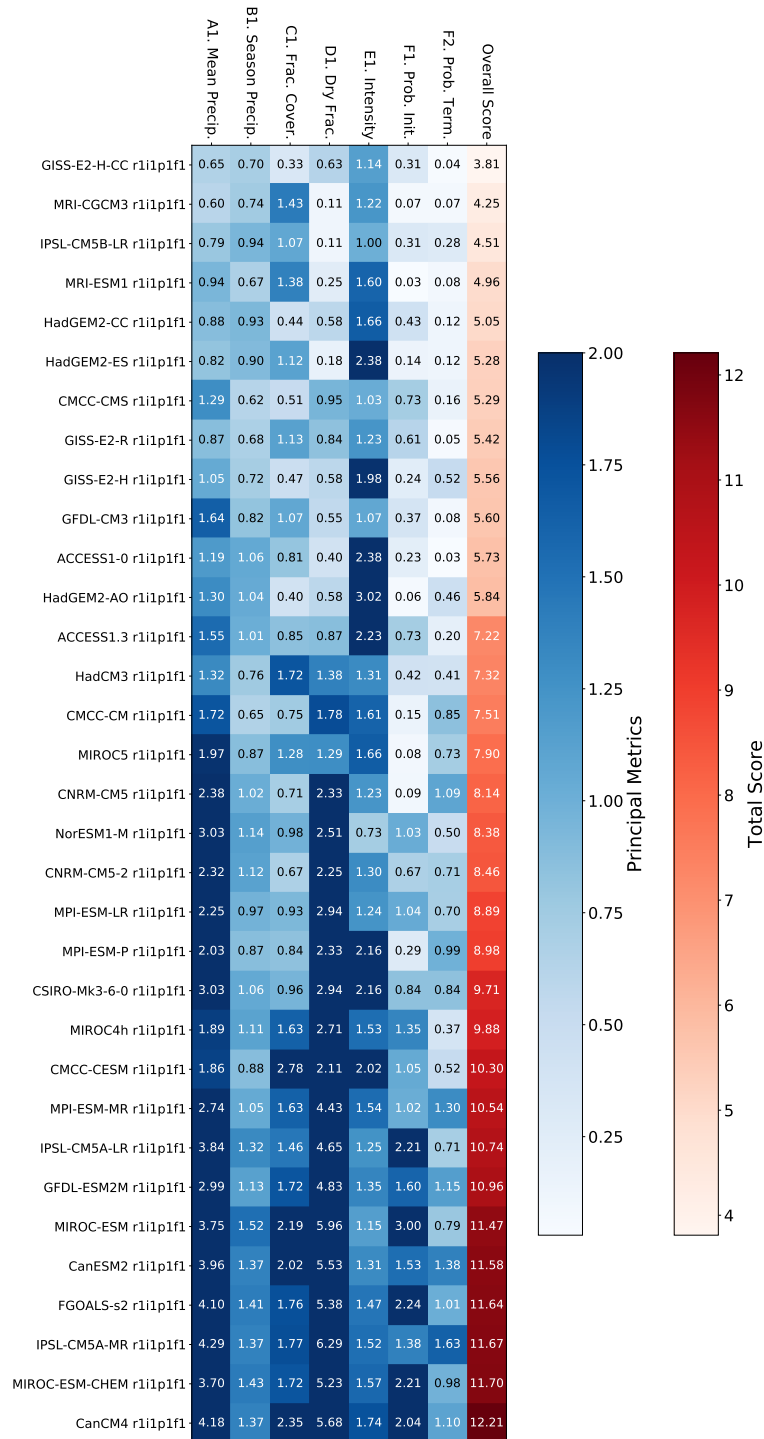


Figure 2.18: Principal metrics and total score from the CMIP5 dataset over the Lower Mississippi Region.

2.8. SUPPLEMENTS

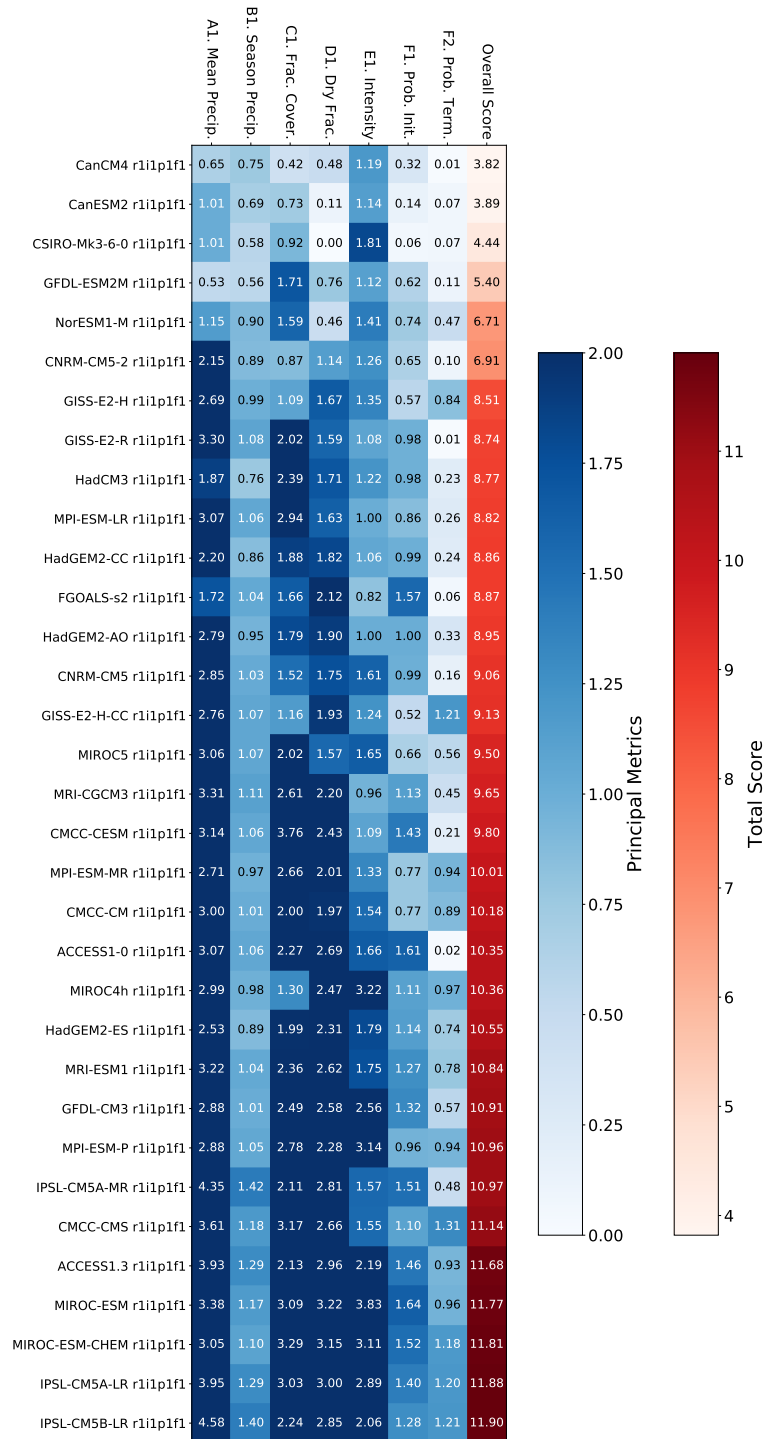


Figure 2.19: Principal metrics and total score from the CMIP5 dataset over the New England Region.

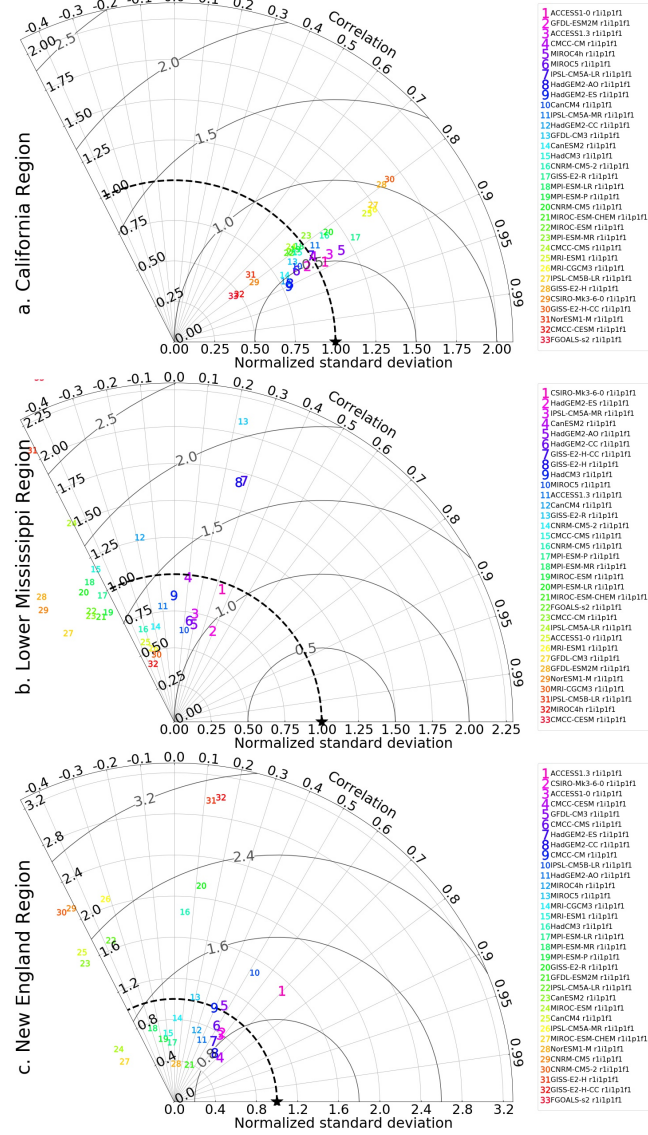


Figure 2.20: Taylor diagram of CMIP5 datasets over three hydrologic regions. (a). Over the California Region. (b). Over the Lower Mississippi Region. (c). Over the New England Region.

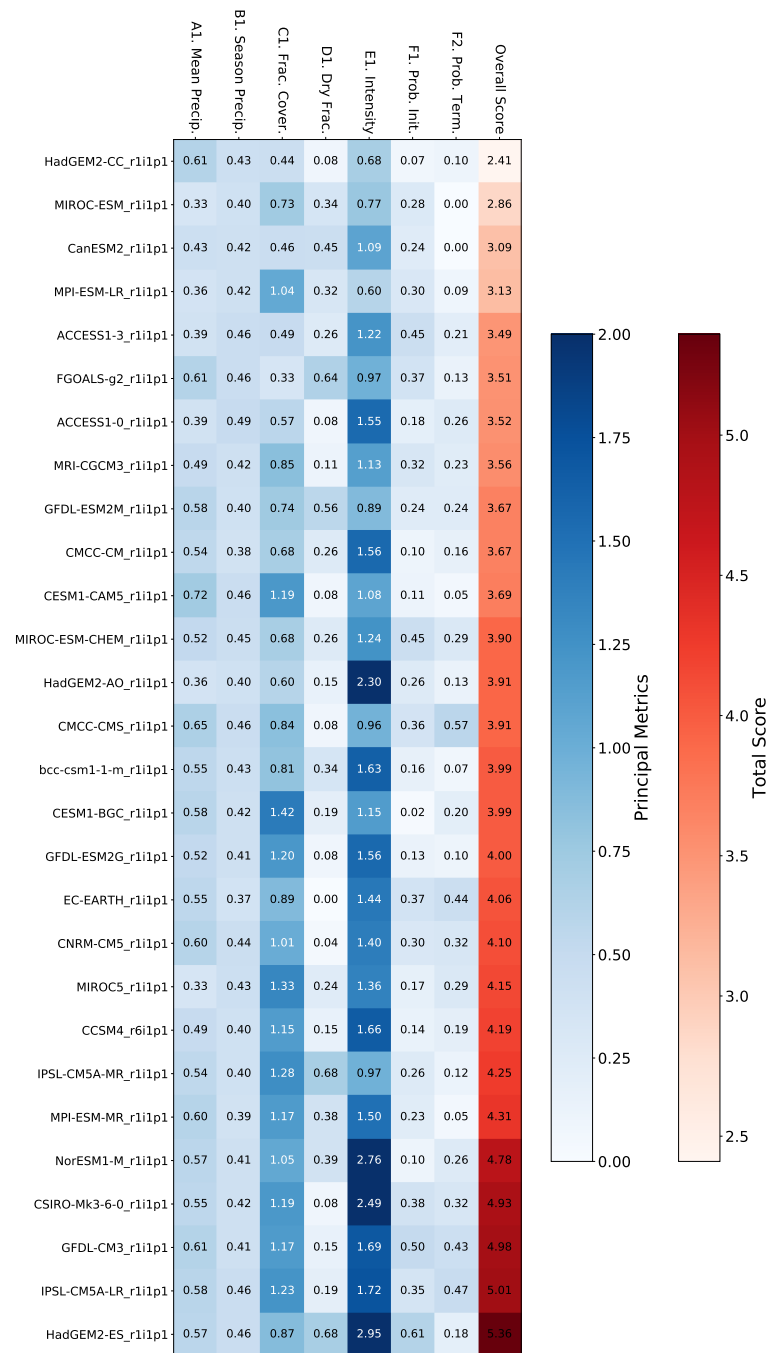


Figure 2.21: Principal metrics and total score from LOCA datasets over the Lower Mississippi Region.

2.8. SUPPLEMENTS

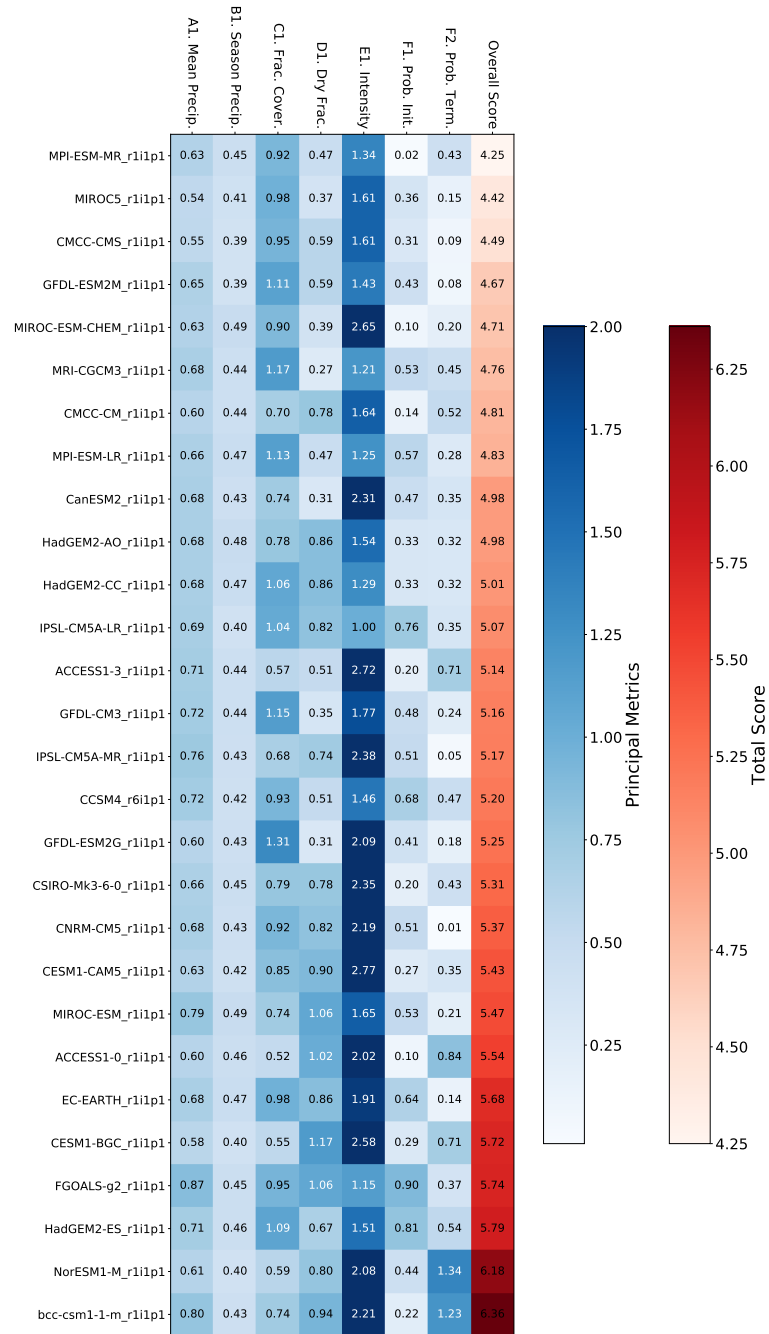


Figure 2.22: Principal metrics and total score from LOCA datasets over the New England Region.

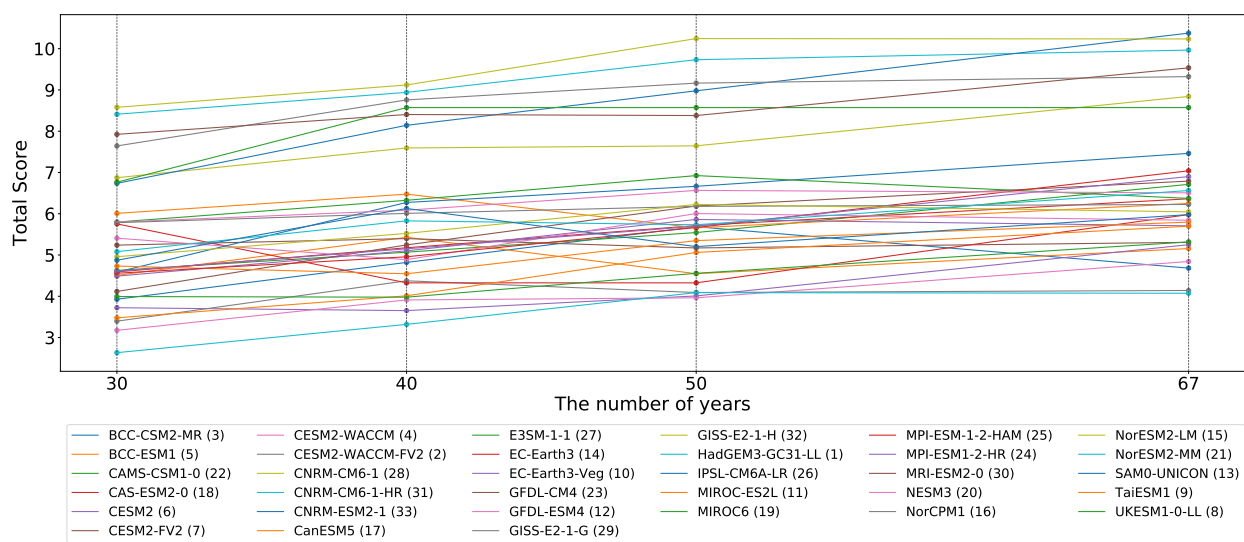


Figure 2.23: Total score of CMIP6 over the California Region during different time periods (the number in bracket indicates each model's rank of the total score, 1 represents the best one).

2.8. SUPPLEMENTS

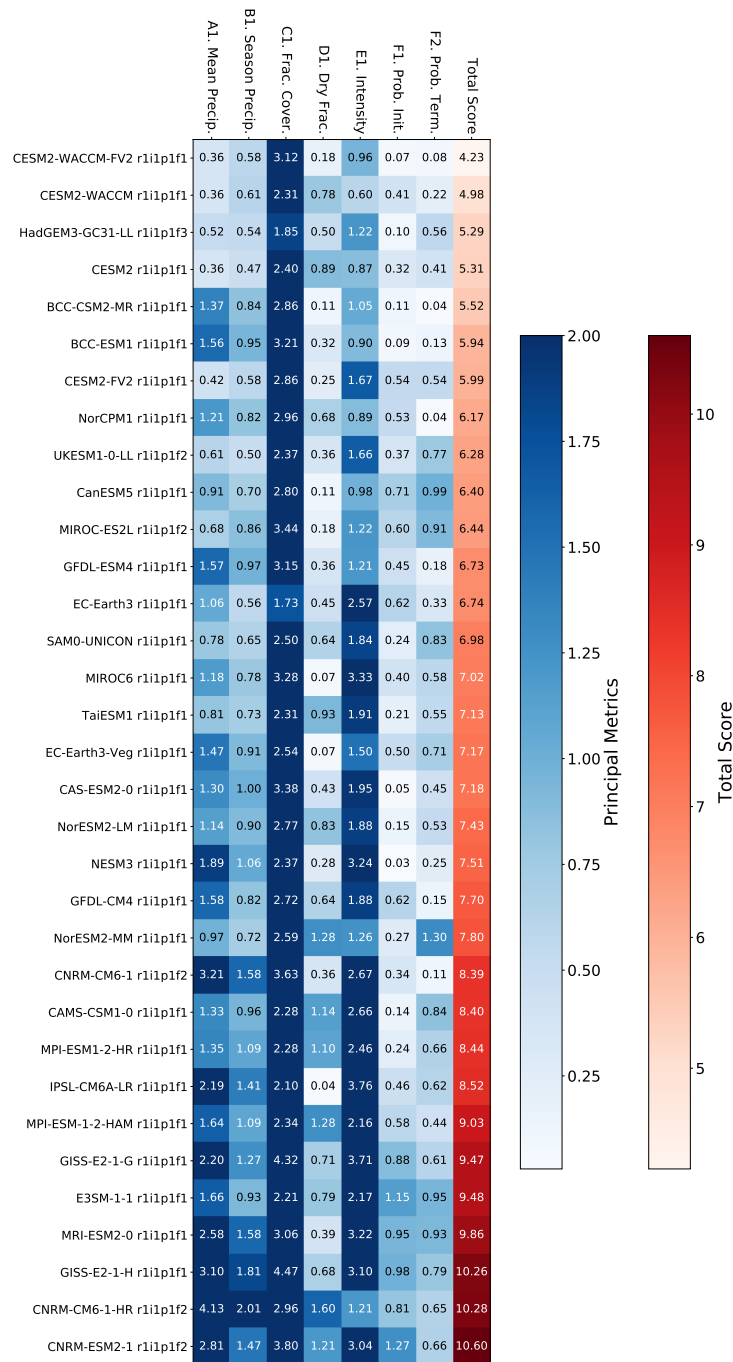


Figure 2.24: Principal metrics and total score of CMIP6 datasets (interpolated to a 0.44° grid) over the California Region.

2.8. SUPPLEMENTS

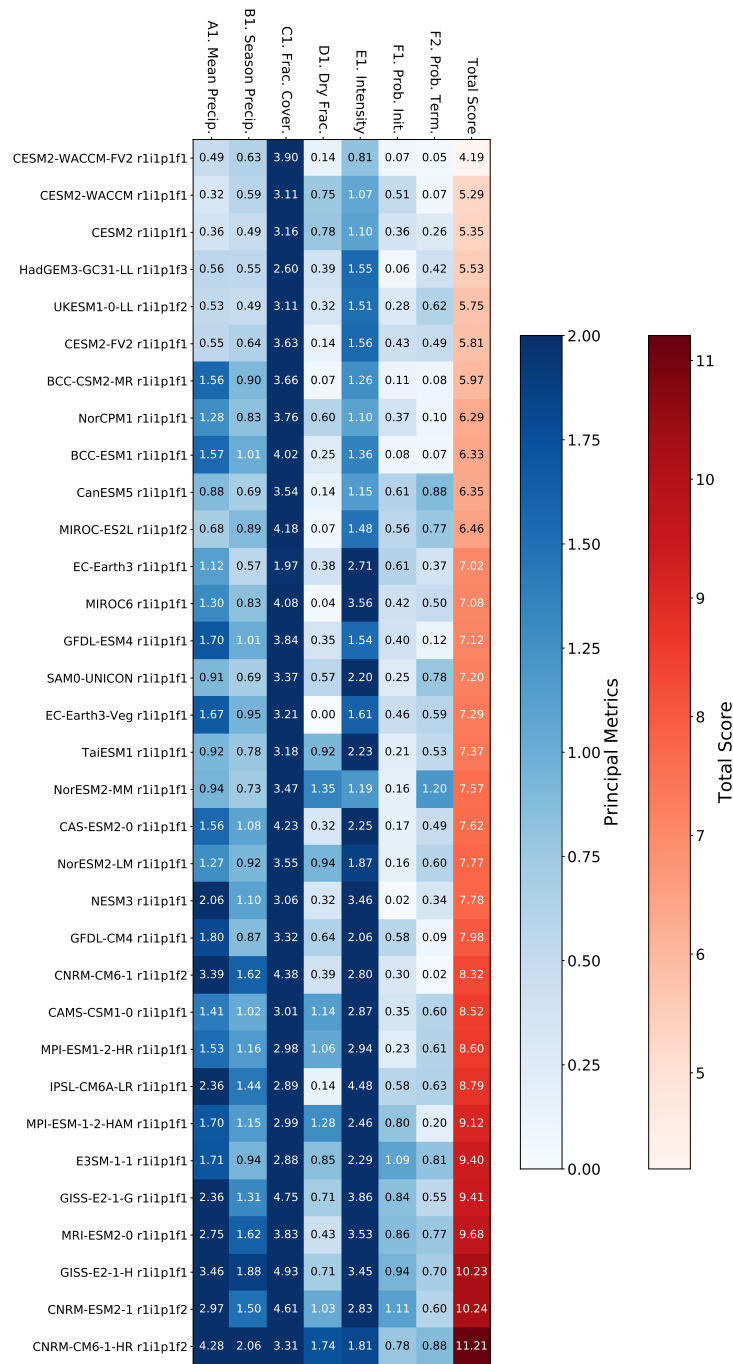


Figure 2.25: Principal metrics and total score of CMIP6 datasets (interpolated to a 0.22° grid) over the California Region.

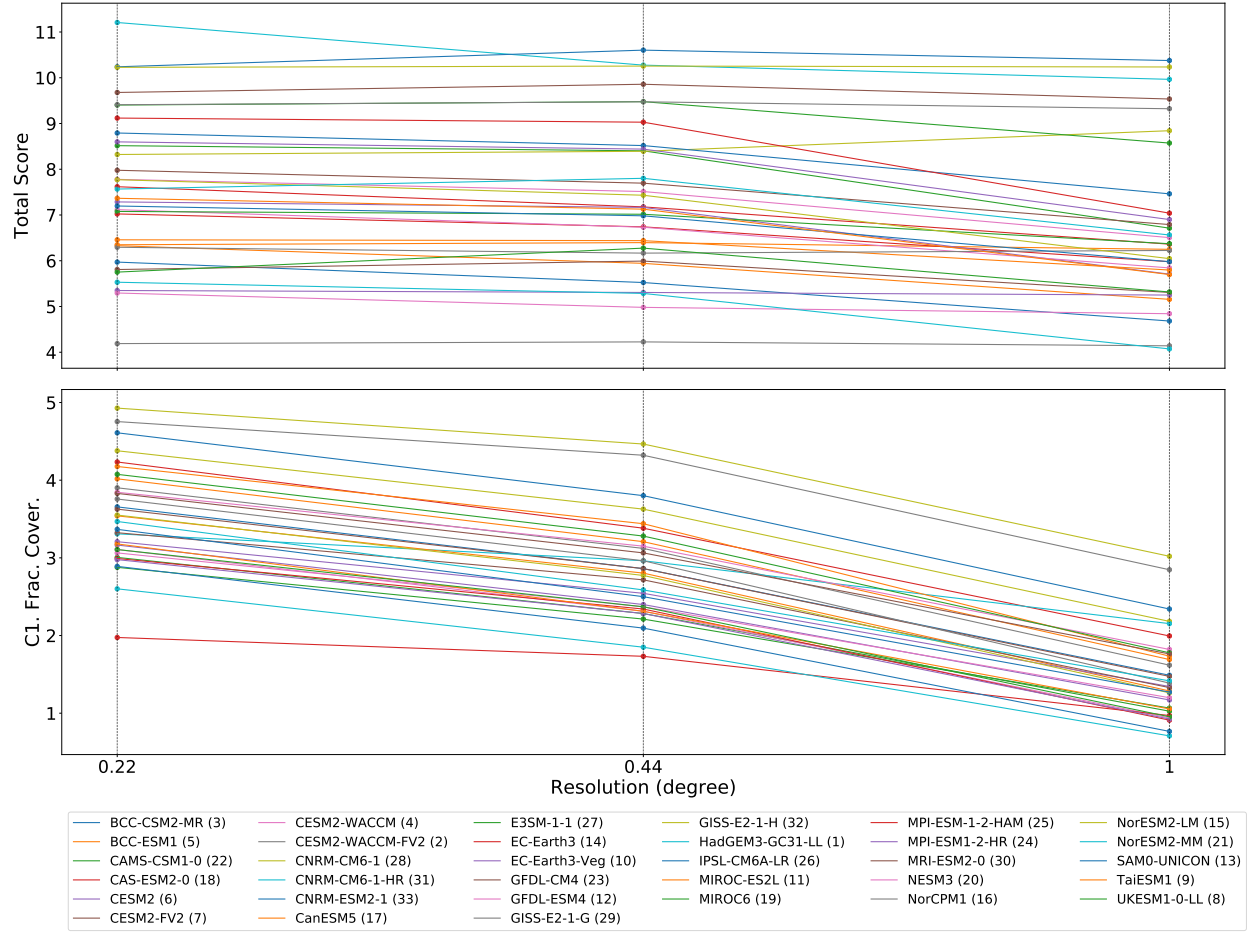


Figure 2.26: Total score and fractional drought coverage of CMIP6 datasets with different interpolation resolution over the California Region (the number in bracket indicates each model's rank of the total score, 1 represents the best one).

2.8. SUPPLEMENTS

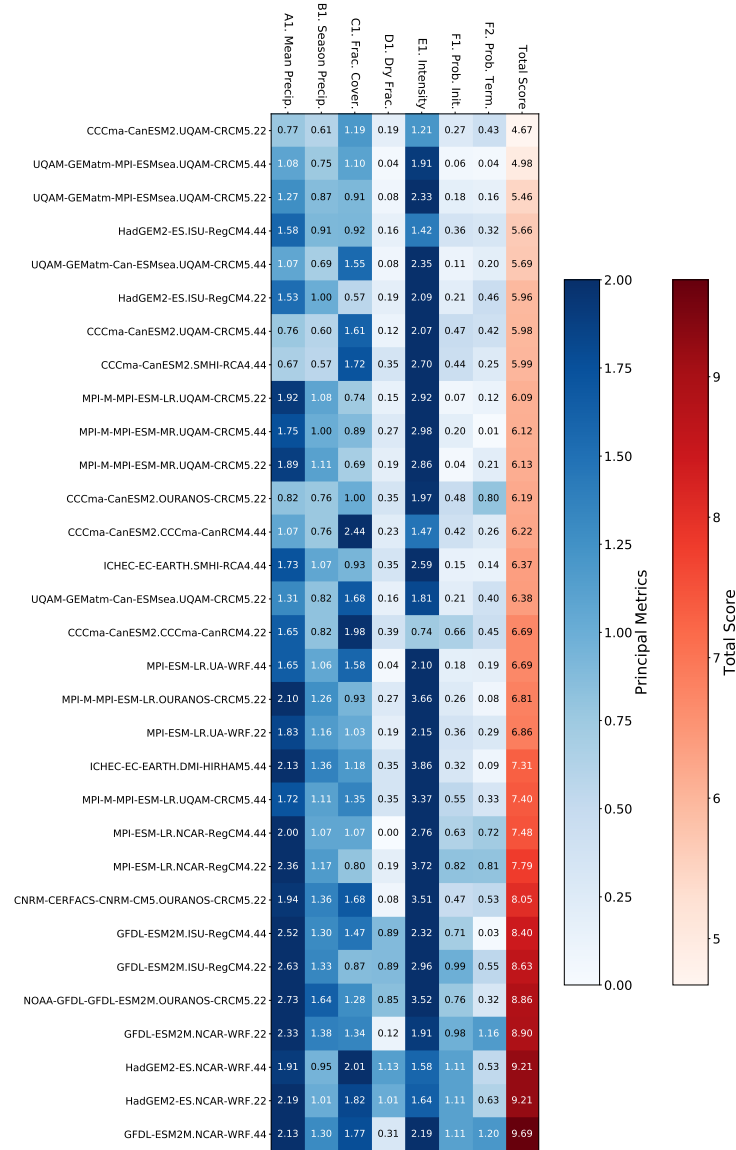


Figure 2.27: Principal metrics and total score of CORDEX datasets (interpolated to a 0.44° grid) over the California Region.

2.8. SUPPLEMENTS

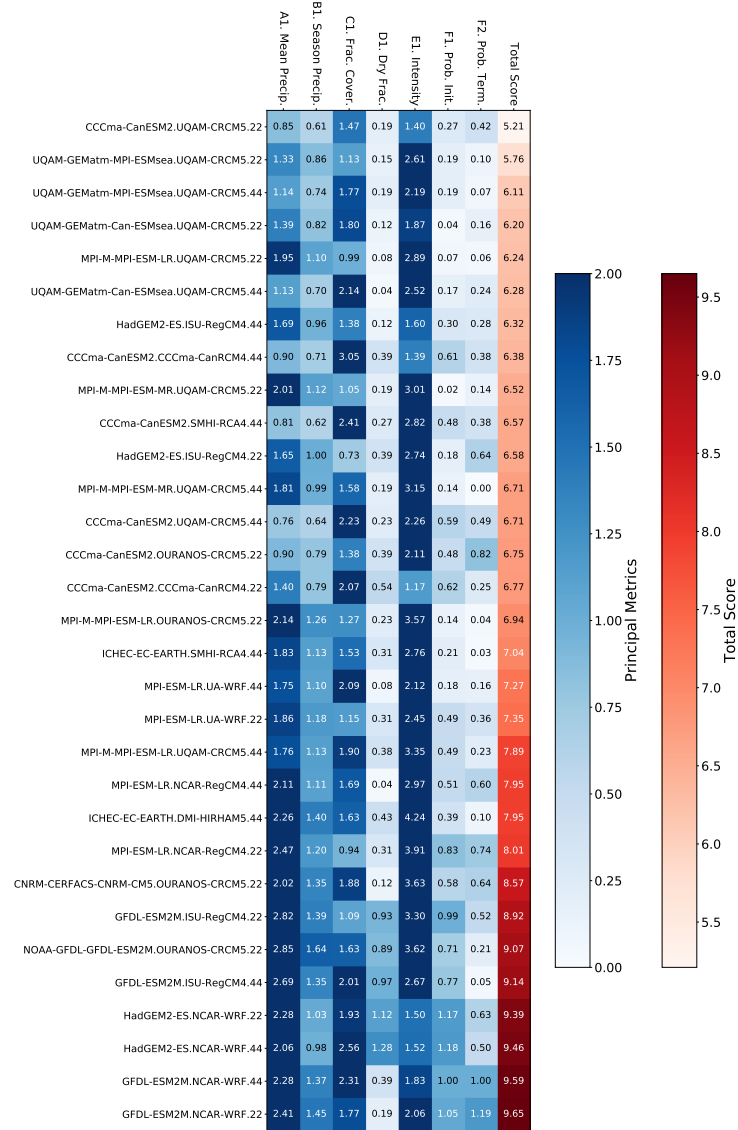


Figure 2.28: Principal metrics and total score of CORDEX datasets (interpolated to a 0.22° grid) over the California Region.

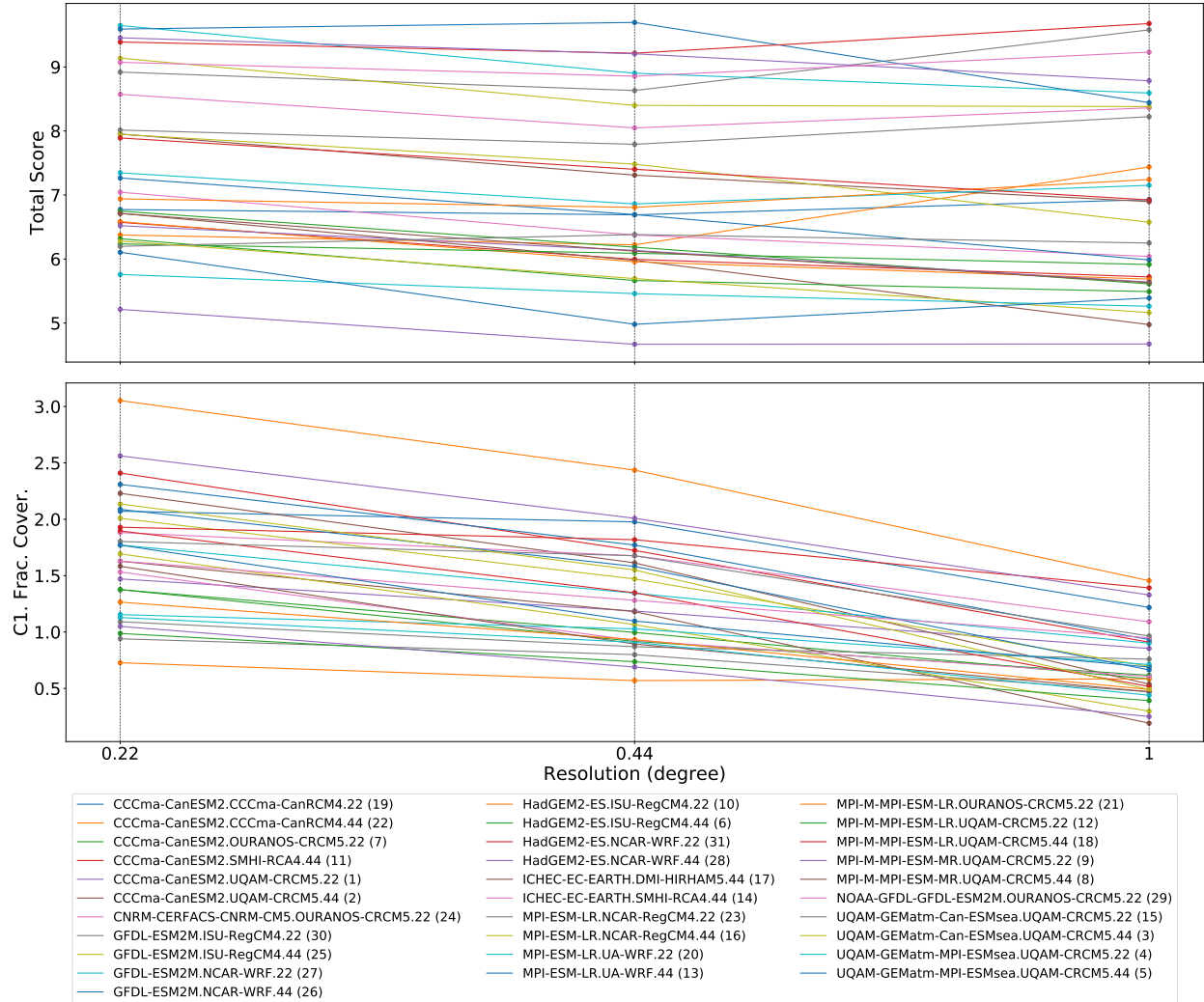


Figure 2.29: Total score and fractional drought coverage of CORDEX datasets with different interpolation resolution over the California Region. The number in bracket indicates each model's rank of the total score with 1 representing the best model under this metric.

Chapter 3 Changing trends in drought patterns over the Northeastern U.S. using multiple large ensemble datasets

3.1 Introduction

Although a general wetting trend is projected to continue over NEUS due to increases in precipitation, risk from extremely dry conditions does not disappear and short-term extreme droughts are even projected to become more frequent in the future (Frumhoff et al., 2007; Hayhoe et al., 2007; Krakauer et al., 2019). Evidence has also emerged that flash droughts, characterized by short-term lack of rainfall and abnormal evaporative demand associated with high temperature, are an increasing threat for this region (Moser et al., 2008; NIDIS/NOAA, 2021; Otkin et al., 2018; Pendergrass et al., 2020; Trenberth et al., 2014). Recent examples of such rapid-onset droughts like the 2016 and 1999 droughts, having occurred during the growing season and leading to severe impacts for local agriculture and related sectors (Lombard et al., 2020; NIDIS/NOAA, 2017, 2021). Unlike traditional drought, which is common in dry regions (Kogan, 1997; Mishra and Singh, 2010), flash droughts can often occur in areas with dense vegetation, as plants can enhance evapotranspiration by drawing on water that is deeper in the soil (Chen et al., 2021; Mo and Lettenmaier, 2016; Pendergrass et al., 2020). In examining hydroclimatic trends in this region, it is clear that water resource managers will be expected to deal with several impending challenges brought by the warming climate such as increased risk of wildfire, heatwaves, and drought (Frumhoff et al., 2007; Hayhoe et al., 2007). Historical data also may not be sufficiently reliable for future projections

due to loss of stationarity, more intense extreme weather events, and enhanced climate variability (Armal et al., 2018; Milly et al., 2008; Stryker et al., 2018; Yu et al., 2018). Consequently, there remains a need for high-quality climate information to inform regional adaptation strategy (AghaKouchak et al., 2015; Milly et al., 2008).

Earth-system models (ESMs) are commonly employed for projecting the impacts of climate change on water resources and the hydrologic cycle (Frumhoff et al., 2007; Kharin et al., 2007; Wagener et al., 2010). In particular, these models allow us to better understand the dynamic nature of drought under climate change (Christian et al., 2019; Joetzjer et al., 2012; Strzepek et al., 2010; Yuan et al., 2019). Although climate models have improved largely over the past several decades, in the process incorporating more complex representations of the atmosphere–land–ocean–sea ice system and its interactions (Bonan and Doney, 2018; Rodgers et al., 2015), large and persistent uncertainties limit our ability to produce reliable projections (Xie et al., 2015; Xue and Ullrich, 2021a). As one of the three main sources of such climate projection uncertainty, internal variability arises from the unforced natural variability of the climate system and is magnified by coupled model processes (Deser et al., 2012; Hawkins and Sutton, 2009; Xie et al., 2015). It potentially accounts for half of the inter-model spread in near-surface variables such as precipitation and temperature over North America over the next 50 years (Deser et al., 2014, 2020). Due to drought’s low frequency and its relatively long duration, internal variability has meant that traditional studies attempting to characterize drought from only one or several future realizations have had difficulty drawing statistically significant conclusions (Deser et al., 2012; Hawkins and Sutton, 2009; National Academies of Sciences, Engineering, and Medicine and others, 2016; Xie et al., 2015). Consequently, projections of drought have largely remained ambiguous (Taylor et al., 2013; Trenberth et al., 2014; Van Loon, 2015), suggesting a need for a large number of realizations to extract a signal from the noise (Taylor et al., 2013). Fortunately, this awareness has led to the development of datasets such as the Multi-Model Large Ensemble Archive (MMLEA), which presently includes 7 CMIP5-class climate models with at least 16 (and up to 100) ensembles with slightly different initial conditions. This resource and analogous datasets have proven immensely useful for quantifying internal variability, especially at regional scales (Deser et al., 2020; Kay et al., 2015).

All 7 Large Ensemble (LE) models presently available through the MMLEA are used in this study, so as to sample both internal variability and structural uncertainty. Meteorological droughts are identified using both the Standardized Precipitation Index (SPI) and Standardized Net Precipitation Index (SNPI). SNPI is analogous to SPI but uses net precipitation (i.e., precipitation minus evapotranspiration) instead of precipitation, and so represents the net moisture input to the land surface. These indices are employed at multiple temporal scales over the historical (1950-2000) and future (2050-2100) periods, with the latter using the RCP 8.5 emission scenario. Several recent studies have shown that plant response to increases in carbon dioxide need to be accounted for to properly understand future droughts, as vegetation plays an essential role in modulating evapotranspiration, surface hydrological conditions and the development of the flash droughts (Bonfils et al., 2017; Dai et al., 2018; Knauer et al., 2017; Pendergrass et al., 2017; Swann et al., 2016; Weiss et al., 2012). Consequently, indices such as SPI, SPEI and PDSI, which do not directly account for plant response, may not reliably capture the nature of future droughts (Bonfils et al., 2017; Dai et al., 2018; Mankin et al., 2019; Swann et al., 2016; Yang et al., 2019). Therefore, in this study, we principally rely on SNPI to project the change of water availability and its underlying factors in this region, whereas SPI is used to project dry conditions without considering water demand. Short-term, intermediate-term and long-term droughts are analyzed using three-month (SPI3/SNPI3), six-month (SPI6/SNPI6) and 24-month (SPI24/SNPI24) drought indices (M. Svoboda and Wood, 2012; Svoboda and Fuchs, 2016; Thomas et al., 2015; Wilhite, 2005). Additionally, our focused investigation of flash droughts uses SNPI1 (which is based on 1-month temporal scale).

With the MMLEA at our disposal, in this study we address the following questions: How are droughts of different temporal scales changing over the NEUS? Will flash droughts become more intense and frequent over the NEUS? What are the underlying factors that affect flash drought in this region?

Table 3.1: Models of the multi-model large ensemble archive (MMLEA) and data repository from Deser et al. (2020).

Modelling Centre \ Model Version	Resolution (atmosphere)	Years	No. of Members	Forcing	Plant Response \ Vegetation Change	Reference
CCCma \ CanESM2	$\sim 2.8^\circ \times 2.8^\circ$	1950-2100	50	Historical, RCP 8.5	Yes \ Yes	Kirchmeier-Young et al. (2017)
CSIRO \ MK3.6	$\sim 1.9^\circ \times 1.9^\circ$	1850-2100	30	Historical, RCP 8.5	No \ No	Jeffrey et al. (2013)
GFDL \ ESM2M	$2.0^\circ \times 2.5^\circ$	1950-2100	30	Historical, RCP 8.5	Yes \ Yes	Rodgers et al. (2015)
GFDL \ CM3	$2.0^\circ \times 2.5^\circ$	1920-2100	20	Historical, RCP 8.5	Yes \ Yes	Sun et al. (2018)
MPI \ MPI-ESM-LR	$\sim 1.9^\circ \times 1.9^\circ$	1850-2100	100	Historical, RCP 2.6, RCP 4.5, RCP 8.5	Yes \ Yes	Maher et al. (2019)
NCAR \ CESM1	$\sim 1.3^\circ \times 0.9^\circ$	1920-2100	40	Historical, RCP 8.5	Yes \ Yes	Kay et al. (2015)
SMHI or KNMI \ EC-Earth	$\sim 1.1^\circ \times 1.1^\circ$	1860-2100	16	Historical, RCP 8.5	No \ Yes	Hazeleger et al. (2010)

3.2 Data and methods

Data

In this study, our analysis uses data from the historical period (1950-2000) and future period (2050-2100) under the RCP 8.5 emission scenario (note that MPI-ESM only has data through 2099, so its future period is 2050-2099). All 7 Large Ensemble (LE) CMIP5 models provided by the US CLIVAR Working Group on Large Ensembles are employed (National Center for Atmospheric Research, 2020). As shown in Table 3.1, the LE models have at least 16 and up to 100 ensemble members. Among the 7 LE models, 5 of them capture plant response to increasing CO₂, and 6 of them account for future increases in total vegetation (Hazeleger et al., 2010; Jeffrey et al., 2013; Kay et al., 2015; Kirchmeier-Young et al., 2017; Maher et al., 2019; Reick et al., 2021; Rodgers et al., 2015; Sun et al., 2018). We further use the aggregated set or the multi-ensemble mean of all ensembles of each model to make projections and estimate uncertainty for our projections. Note that without specific annotation, analysis is based on each models' aggregated set of all ensembles. By using large ensembles from multiple models, we are able to incorporate structural uncertainties in our sample. Further, enough realizations are employed to compensate for the internal variability, while the 51-year study period provides a long enough window to capture longer droughts. Drought and its impacts often occur at the watershed scale, and water management decisions are usually also made at this level as well (Diaz, 1983; Mishra and Singh, 2010; Wilhite, 2000; Wilhite and Glantz, 1985). Furthermore, surface water balance is primarily valid at the watershed scale (Arnell, 1999; Gleick, 1987; Peel et al., 2010; Wang and Dickinson, 2012), and so our analysis focuses on regional monthly mean data over the New England Region defined by the Watershed Boundary Dataset (WBD) (Survey and US Department of Agriculture, 2013). All datasets are first interpolated to $1^\circ \times 1^\circ$ by conservative interpolation (Schulzweida, 2019) before masking to ensure each dataset has the same boundary.

Methods

Standardized Precipitation Index (SPI). As one of the most widely used drought indices, the Standardized Precipitation Index (SPI) is a flexible drought indicator designed to capture the magnitude of meteorological drought conditions at different temporal scales (Guttman, 1999; McKee et al., 1993; Svoboda and Fuchs, 2016). Intuitively, the value of SPI_n represents how much the accumulation of the past n months' precipitation departs from average conditions. The value of the index is then used to classify droughts following Table 2.1. Here, we specifically use SPI3, SPI6, and SPI24 to quantify short-term, intermediate-term, and long-term droughts, following a number of past studies (M. Svoboda and Wood, 2012; Svoboda et al., 2012; Thomas et al., 2015; Wilhite, 2005).

SPI must first be calibrated before use, which entails estimation of the arguments of the gamma distribution for a given region and realization. This procedure allows us to use the inverse CDF to transform precipitation data to be normally distributed. For each ensemble member, the historical period is used for this calibration. SPI values given here will then use the same calibration for the historical period (1950-2000) and the future period (2050-2100) to illustrate the impacts of climate change on drought conditions. If the future climate has the same drought statistics as the historical period, future SPI should follow the standard normal distribution, as in the historical period. On the other hand, any departure from normality is evidence of changing drought patterns. Note that for SPI, values are clipped to the range of -3.09 to 3.09 (corresponding to normal probabilities of 0.001 and 0.999) as suggested by Guttman (1999).

Although SPI is widely used in practice and research (Hayes et al., 2002; Svoboda and Fuchs, 2016; Wilhite, 2005), and is recommended by World Meteorological Organization (WMO) (M. Svoboda and Wood, 2012; Svoboda et al., 2012) to assess meteorological droughts, it does have some known shortcomings (Dai, 2011; Vicente-Serrano et al., 2010, 2012). For instance, SPI only relies on precipitation, under the argument that precipitation is the primary driver of drought and is more variable than other relevant quantities (Hayes et al., 2002; McKee, 1995; McKee et al., 1993). As such, it does not account for intensification in evapotranspiration brought on by a warming climate

(Frumhoff et al., 2007; Lyon et al., 2005; Seager et al., 2012) and associated enhancement in drought conditions (Dai, 2011; Vicente-Serrano et al., 2010, 2012). As a result, SPI is unable to detect flash droughts primarily driven by enhanced evaporative demand (Otkin et al., 2013; Pendergrass et al., 2020). This has led to the use of alternate drought indices, such as the Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010) and Palmer Drought Severity Index (PDSI) (Palmer, 1965), which rely on the potential evapotranspiration (PET) derived from offline empirical equations like the Penman–Monteith equation. Although in the Penman–Monteith equation, the plant resistance is considered, assuming over a reference plant, the PET equations do not consider the change of physiological plant response to increasing CO₂, a potentially essential factor in estimating future surface water balance: It is known that increases to atmospheric CO₂ will enhance photosynthesis efficiency – namely, by reducing their stomatal conductance, plants can decrease transpiration (water losses) per unit of carbon gain, then further mitigate plant water stress and increase soil moisture (Field et al., 1995; Milly and Dunne, 2016; Swann, 2018; Swann et al., 2016). Consequently, several studies have argued that capturing the transpiration response from plants to rising CO₂ enables more reliable drought projections (Bonfils et al., 2017; Dai et al., 2018; Milly and Dunne, 2016; Swann et al., 2016). This suggests that indices that employ PET will overestimate drought intensification under a warming climate (Betts et al., 2007; Milly and Dunne, 2016; Roderick et al., 2015; Sellers et al., 1996; Yang et al., 2019), and so may not be appropriate for investigations of drought under climate change. Moreover, vegetation has been shown to be an essential factor for determining surface hydrology and estimating evapotranspiration, and so increases in evapotranspiration driven by more vegetation in the growing season can potentially trigger flash drought (Knauer et al., 2017; Otkin et al., 2018; Pendergrass et al., 2017, 2020; Weiss et al., 2012; Zhang et al., 2015). This emphasizes the importance of a correct representation of vegetation in climate projections, which is otherwise absent in traditional PET calculations (Frank et al., 2015; Knauer et al., 2017; Weiss et al., 2012, 2014; Zhang et al., 2015).

Standardized Net Precipitation Index (SNPI). As the actual evapotranspiration from ESMs is internally consistent with other meteorological and hydrologic variables, we argue that

it can also be used to better quantify surface water balance in a warming climate. Additionally, the actual evapotranspiration in ESMs almost always incorporates the physiological effect of CO₂ on plants and increases in vegetation due to GHG (Table. 3.1), which is important for correctly modeling surface moisture fluxes (Bonan and Doney, 2018; Milly and Dunne, 2016; Swann, 2018; Swann et al., 2016). Therefore, in this study, we will employ a modified SPI (referred to as SNPI) based on net precipitation: that is, precipitation minus actual evapotranspiration, with both quantities taken directly from the calculation used in the ESMs. As such, SNPI is primarily for use with ESMs where actual evapotranspiration is computed from a comprehensive land component model. The physical interpretation of SNPI is also analogous to Table 2.1. Calibration and calculation of SNPI is otherwise performed in a similar manner to SPI except for a constant adjustment to ensure all net precipitation data is larger than zero (See detailed calculation below). Specifically, the net precipitation will be adjusted to be positive via

$$\text{Adjusted Net Precipitation}_{n,i} = \text{Net Precipitation}_{n,i} - \min(\text{Net Precipitation}) + 0.01. \quad (3.1)$$

This adjustment is necessary since the gamma distribution cannot be used for negative quantities. As with the calculation of SPI (Guttman, 1999; McKee, 1995; McKee et al., 1993), the adjusted net precipitation is fitted to a gamma distribution and subsequently transformed to be normally distributed. Since in this study we will compare drought conditions defined by SPI/SNPI, fitting the data to the same distribution makes the comparison more straightforward. However, we have also confirmed that the fit of the gamma distribution to the the adjusted net precipitation is as good as its fit to precipitation, and superior to the log-logistic distribution (see Supplement 3.6.4).

Because the calibration of SPI is sensitive to the length of data employed (e.g., Guttman (1999); McKee et al. (1993)), more than 50 years data is recommended. In our study, we calibrate both historical and future SNPI/SPI based on the historical period which includes 51 years (1950-2000). That is, the gamma distribution parameters used in transforming the precipitation and adjusted net precipitation of both periods from a gamma distribution to a normal distribution are all calculated based on the historical data. Only the historical period is used as calibration to avoid

future projections impacting the SPI/SNPI of the historical droughts, and so that differences in the future can be identified departures from normality.

SNPI has several notable advantages which make it a desirable index for our study. First, net precipitation is a robust and interpretable quantity representing the net moisture input from atmosphere and reflecting the actual water balance locally; in similar terms, drought can be considered as a long-term imbalance between precipitation and evapotranspiration (Wilhite, 2000). Water balance can be modeled by the equation

$$\frac{dW(t)}{dt} + R(t) = P(t) - E(t), \quad (3.2)$$

where $\frac{dW(t)}{dt}$ is the change of land water storage, $P(t)$ is the regional precipitation, $E(t)$ is the regional evapotranspiration, and $R(t)$ is the regional runoff (Crowley et al., 2006; Hayhoe et al., 2008; Wang and Dickinson, 2012). In this context, precipitation minus evapotranspiration represents fluctuations in the local water availability for agriculture or other purposes (Arnell, 1999; Gleick, 1987; Mintz and Serafini, 1992). Consequently, net precipitation can be a rough indicator for the sum of soil moisture and runoff. Second, at the regional scale, the strong correlation between net precipitation and runoff or soil moisture has been confirmed from observational and modeled data (Cassano et al., 2007; Crowley et al., 2006; Dai et al., 2018; Gleick, 1987; Hayhoe et al., 2008; Teuling et al., 2009; Van Loon, 2015; Wilhite, 2005). Finally, since surges in water demand from crops and vegetation during the growing season serve as a primary driver of flash drought (Christian et al., 2019; Otkin et al., 2018; Pendergrass et al., 2020), the fact that SNPI incorporates the change of physiological plant response to increasing GHGs makes it a particularly useful tool for identifying potential trends in flash drought under global warming. We do admit that because of using the actual evapotranspiration which can also be influenced by the land-use changes, the SNPI might not distinguish the impacts of land-use changes and climate change. Further study will be done to distinguish these two impacts on drought.

Drought duration and magnitude. Drought intensity in a certain month can be easily quantified using the value of SPI/SNPI in that month, with more negative values indicative of more

intense drought conditions. Although dryness exists on a continuum, and so the exact start and end dates of a drought are difficult to ascertain, here we define a drought as initiating when SPI/SNPI drops below -1 and terminating when SPI/SNPI is next above 0 (Guttman, 1999; McKee et al., 1993). The duration of the drought is then the number of months between its start and end points (including the start month, but not the end month). The accumulated *drought magnitude* (DM) of each drought event is defined as the absolute value of accumulated SPI/SNPI during a drought event:

$$DM = - \sum_{\text{month } i} SPI_i. \quad (3.3)$$

where the index i is taken over all months of the drought (Guttman, 1999; McKee et al., 1993).

Defining flash drought and extreme flash drought Although flash drought is an increasingly popular topic in the literature, there is not yet consensus on the best methods for characterizing these events. However, studies generally agree that flash drought should be characterized by its rate of intensification (Ford and Labosier, 2017; Otkin et al., 2013, 2018) and that the variables used to represent the magnitude of flash drought should reflect the short-term water demand from evapotranspiration, e.g. soil moisture, evaporative stress, and potential evapotranspiration (Christian et al., 2019; Osman et al., 2021; Otkin et al., 2016, 2018; Pendergrass et al., 2020; Zhang and Yuan, 2020). Given its relationship with water demand, in this study we propose to use SNPI1 to identify and measure the monthly flash drought. The flash drought is generally defined as sub-seasonal drought event (Christian et al., 2021; Pendergrass et al., 2020). Although weekly flash droughts may occur, most infamous flash drought events over CONUS are monthly flash droughts such as the 2012 US Midwest flash drought and 2017 US Northern Great Plains flash drought (Hoell et al., 2020; Hoerling et al., 2012; Otkin et al., 2021). Also, only monthly data is currently available for calculation of the SNPI from the MMLEA and there is some danger in considering drought on time scales shorter than a month: the weekly time scale is often insufficient for meteorological drought to drive significant drying of soil (Tallaksen and Van Lanen, 2004; Van Loon, 2015).

Therefore, we represent the development speed of drought (DSD) at month i via

$$\text{DSD}_i = -(SNPI1_i - SNPI1_{i-1}). \quad (3.4)$$

Since flash drought is characterized by rapid development and extremely dry conditions after development (Otkin et al., 2018; Pendergrass et al., 2020), herein we define flash drought in month i as $\text{DSD}_i \geq 2$ and $SNPI1_i \leq -2$. Intuitively, this means that, at the very least, “extremely dry conditions” ($SNPI1 \leq -2$) should develop from conditions that are at least neutral ($SNPI \geq 0$) in only one month. We also define a notion of “extreme flash drought”, since extreme events of this nature have the potential to result in enormous damage, particularly to agriculture (Otkin et al., 2018; Pendergrass et al., 2020). Herein, extreme flash drought is defined as drought with $\text{DSD}_i \geq 4$ and $SNPI1_i \leq -2$, indicating extremely dry conditions develop from from extremely wet conditions in only one month. Because flash drought is strongly associated with enhanced evaporative demand, previous studies of flash droughts usually exclude the winter season (Ford and Labosier, 2017; Otkin et al., 2018; Pendergrass et al., 2020). Herein we will also ignore the winter season when we analyze trends for flash droughts.

3.3 Results

3.3.1 Result 1: A general wetting trend is undeniable in the mean

The significant wetting trend expected for the NEUS shows up clearly in Fig. 3.1 as a rightward shift of the mean of each model’s probability distribution. This trend is overwhelming in the SPI at all temporal scales, and for all LE models. However, the shift in SPI is unsurprising as this index only takes precipitation as input, and there is high confidence from both historical observations and climate models that annual mean precipitation is presently increasing (and will increase) over the NEUS (Demaria et al., 2016; Frumhoff et al., 2007; Guilbert et al., 2015; Hayhoe et al., 2008; Huntington et al., 2004). As for the SNPI, even when increases in evapotranspiration are taken into account, a long-term wetting trend remains in most models (5 out of 7), confirming a clear tendency towards wetting of the NEUS. Moreover, for both SPI and SNPI, this wetting trend is

3.3. RESULTS

more obvious at longer temporal scales. Examining the ratio of the frequency density in Fig. 3.2, we see that long-term dry conditions ($\text{SPI}_{24} < -1$) are projected to become clearly less frequent (occurring $\sim 90\%$ less often), while long-term pluvial conditions ($\text{SPI}_{24} \geq 2$) are between 2 and 100 times more likely.

3.3. RESULTS

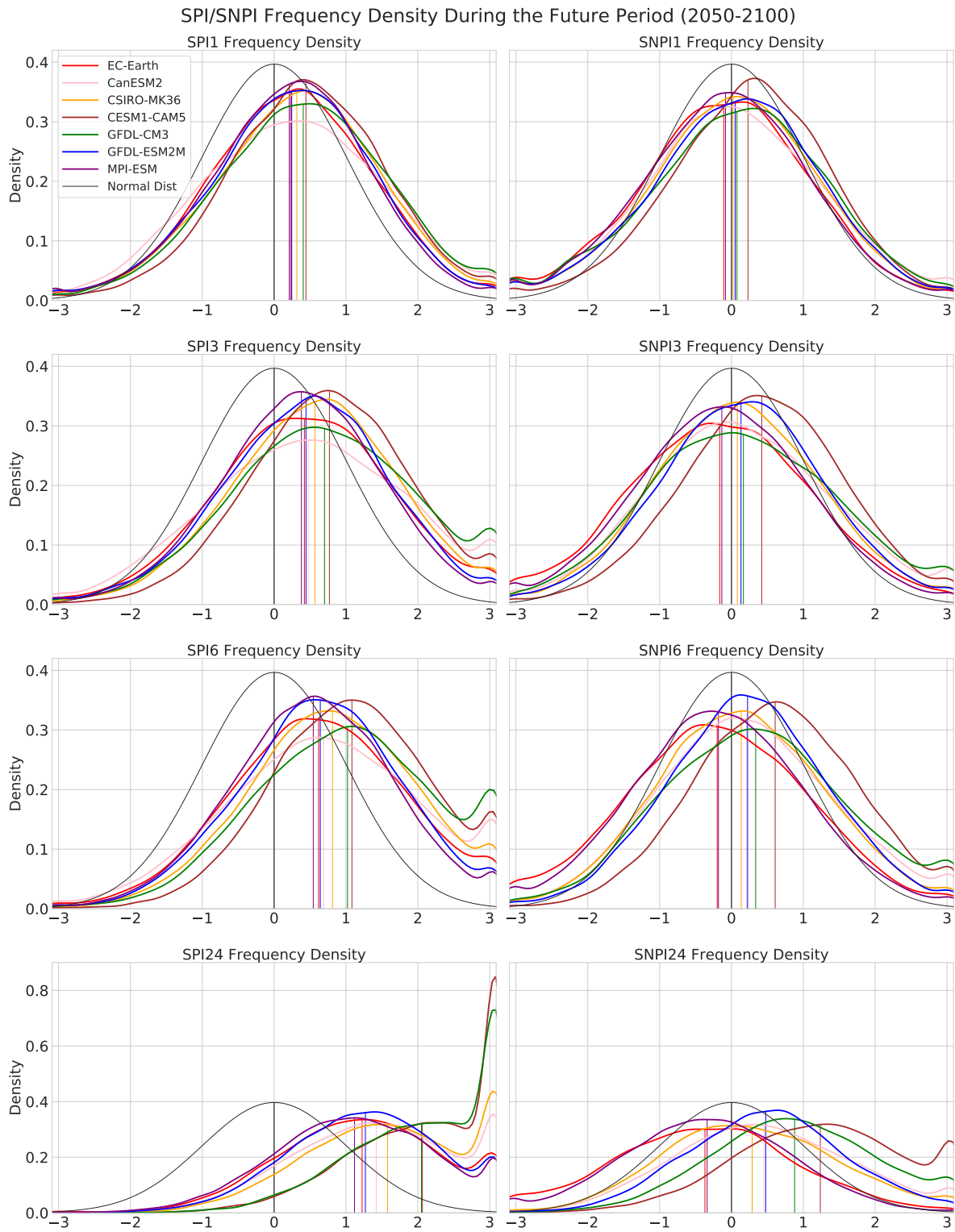


Figure 3.1: SPI/SNPI density of the future period (2050-2100) for all 7 LE models. Each vertical line indicates the mean value from that model.

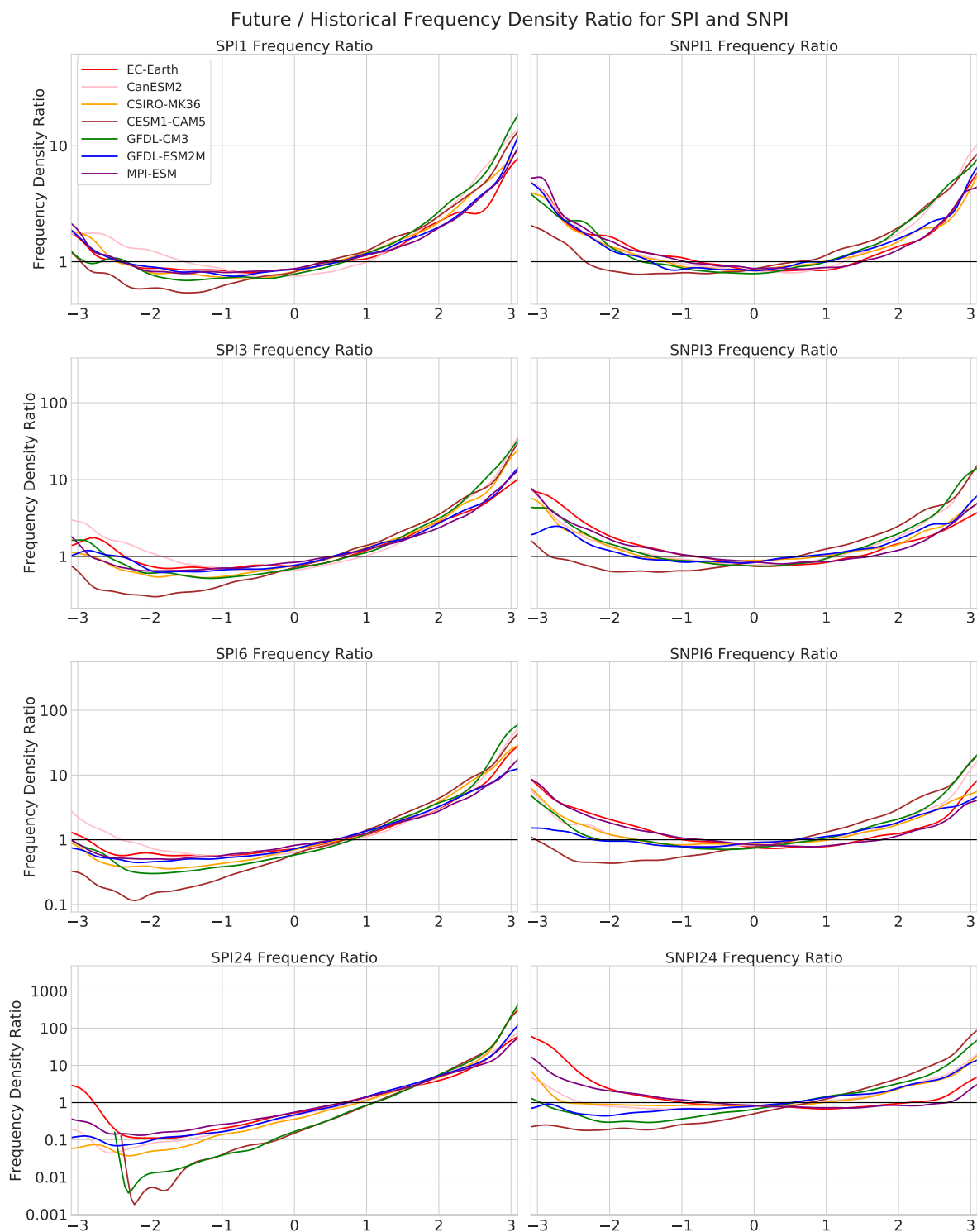


Figure 3.2: SPI/SNPI frequency density ratio for all 7 LE models, defined as the frequency density of the future period divided by the frequency density of the historical period.

3.3.2 Result 2: More extremely wet conditions, with greater magnitude

The tendency towards wetter conditions is most obvious at the extreme right tail of Fig. 3.1 and 3.2 (namely, “extremely wet” months), with both more frequent conditions with $\text{SPI}/\text{SNPI} \geq 2$, and larger average SPI/SNPI for months in this category (as well as more months at the right boundary 3.09). This accords with the continued intensification of extremely wet conditions over the NEUS (Donat et al., 2016; Melillo et al., 2014; Pfahl et al., 2017). Using SNPI, all LE models favor an increase in the frequency of extremely wet conditions and greater SNPI values within this category. For example, the multi-model mean probability of extremely wet months surges from 0.018 to 0.069 (for SPI1) and from 0.019 to 0.048 (for SNPI1). More significantly, exceptionally wet conditions ($\text{SPI1}/\text{SNPI1}$ equal to 3.09) will become 19 and 12 times more frequent, using SPI1 and SNPI1 respectively. These changes suggests greater challenges for water management in this region, as well as increased risk and intensity of flooding.

3.3.3 Result 3: More short-term extreme droughts and intensified evapotranspiration

While these two prior results may seem to suggest abundant water resources in the future, this general wetting trend does not imply that drought will disappear entirely. Although short-term extreme precipitation droughts (months with $\text{SPI3} \leq -2$) are projected to be less frequent in all LE models, nearly all models (6 out of 7) agree that short-term net precipitation droughts ($\text{SNPI3} \leq -2$) are projected to increase largely in frequency (Fig. 3.2). Further, short-term exceptional droughts (SNPI3 is equal to -3.09) are expected to be more frequent in all models. This increase in short-term extreme drought appears to be driven by concordant increases in two factors: evapotranspiration and precipitation variability (Pendergrass et al., 2017). Because of evapotranspiration’s essential role in the water cycle (as the second-largest component in the water balance formula after precipitation (Melillo et al., 2014)) and its significance in producing drying over land (A. Dai and Chen, 2018; Wilhite, 2000; Zhao and Dai, 2015), increases of evapotranspiration need to be accounted for when projecting the changing trends of droughts under a warming climate. Increasing precipitation

3.3. RESULTS

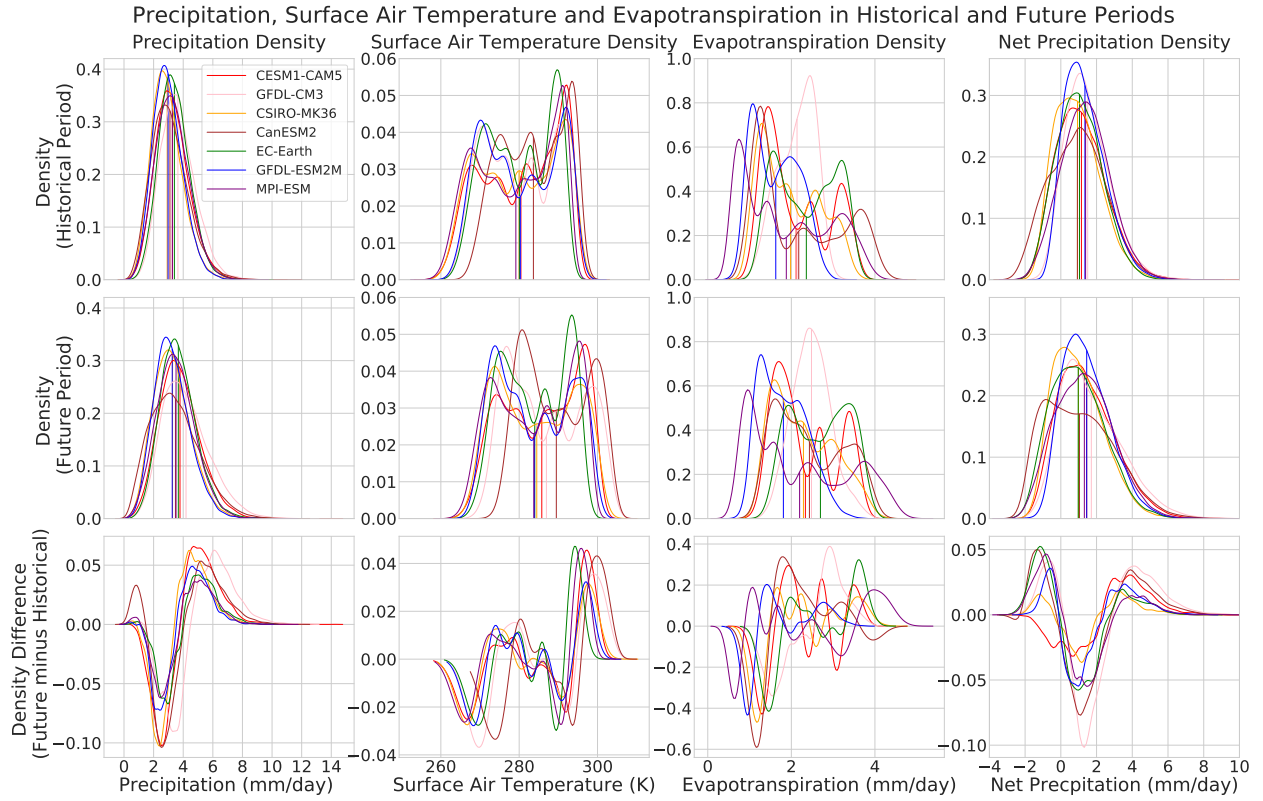


Figure 3.3: Each LE model's density plot and density difference of precipitation, surface air temperature, evapotranspiration and net precipitation.

variability, on the other hand, is associated with an increase in the frequency of both months with extremely low precipitation and high precipitation (and corresponding signals in streamflow), as has been documented in several past studies (Demaria et al., 2016; Frumhoff et al., 2007; Hayhoe et al., 2007; Van Loon, 2015). As discussed later, increases in the frequency of extremely low values of SPI/SNPI in Figs. 3.1 and 3.2 are strongly related to a flattening of the probability density of precipitation / net precipitation intensity around the mean in Fig. 3.3 (Zhao and Dai, 2015). This flattening suggests less frequent “normal conditions” and more frequent extremes. It is quantified by the departure from unity (the standard deviation of SPI/SNPI under a normal distribution) of the multi-model mean standard deviations of SPI1/SNPI1 to 1.16/1.20 in the future.

3.3.4 Result 4: More flash drought with faster initiation and greater intensity

Because of the strong seasonality of flash drought (Christian et al., 2019; Otkin et al., 2018; Pendergrass et al., 2020), each LE model’s projections for the development speed of flash drought in each season except the winter season is shown in Fig. 3.4 (here MAM indicates the spring season, JJA indicates the summer season and SON indicates the autumn season). Although there are discrepancies among seasons and models, several significant trends are apparent. The most concrete among these being that nearly all LE models agree that the frequency of flash drought is projected increase significantly in all seasons (except for CanESM2 and CESM1-CAM5 in JJA). The average increase in the probability of flash drought is about 106%, indicating that flash drought could occur with more than twice as frequently in the future period. Further, flash drought is also projected to initiate faster and with greater severity, as suggested by the smaller values of the median, third quartile, and lower whiskers of the DSD. Note that the p-values in Fig. 3.4 indicates the statistical significance of the t-test about whether the future and historical periods have the same average Delta SNPI1 instead of if they have the same average flash drought frequency.

3.3.5 Result 5: Significant trends towards more extreme flash drought, particularly in the spring

Comparing Fig. 3.4 to the corresponding plot for extreme flash drought (Fig. 3.5), it’s apparent that into the future extreme flash drought frequency tends increase even more than flash drought frequency. The average return period of extreme flash drought (defined as the multi-model mean years needed to occur once) drops from 65.9 years historically to 16.2 years in the future. By contrast, the return period for all flash droughts decreases from 6.3 years to 2.9 years. This historical flash drought return period agrees with previous work by Ford and Labosier (2017), which is based on an examination of surface soil water content. Extreme flash drought also has a strong seasonality, with intensification being much more significant during the growing season, particularly in the spring (MAM). This poses substantial risk for agricultural productivity, and suggests the possibility of more frequent crop failure (Otkin et al., 2018; Pendergrass et al., 2020).

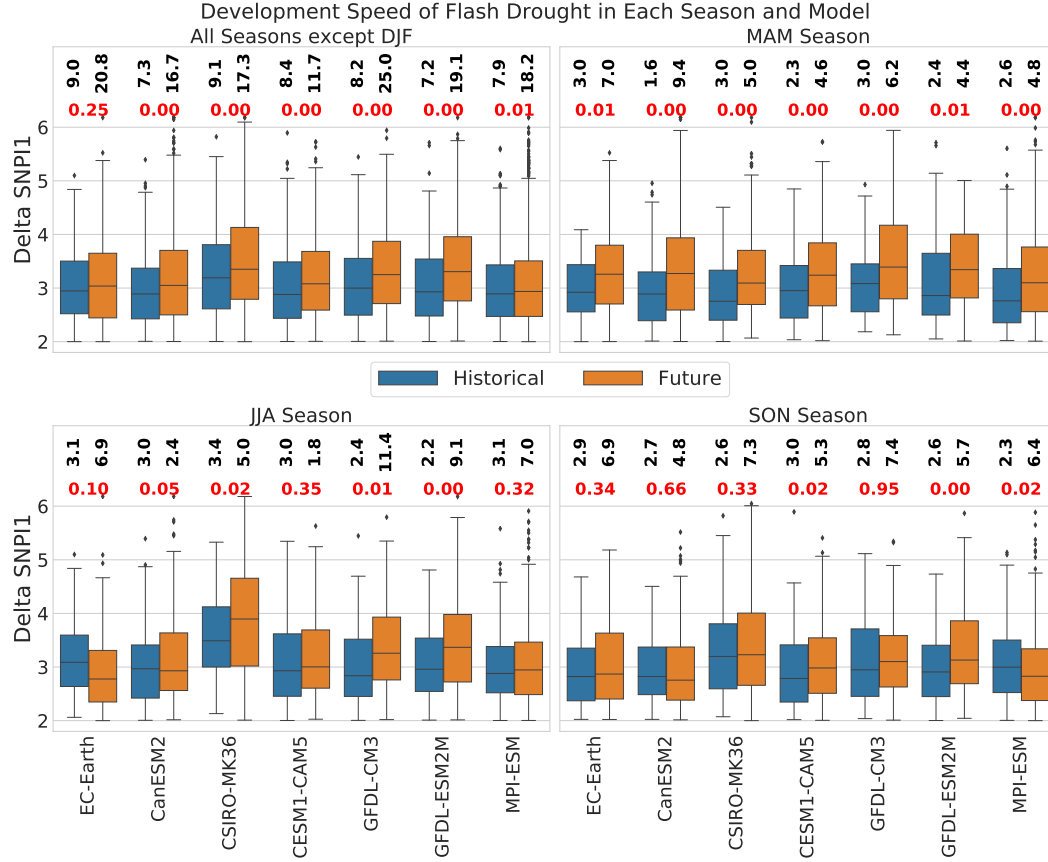


Figure 3.4: Development speed of flash drought for each season (except winter) and model. The black bold numbers above each box plot indicate the average frequency within each period (51 years). The red bold numbers indicate the p-values obtained from a t-test of the Delta SNPI1 mean between historical and future periods.

The development speed of extreme flash drought is also trending higher, producing the possibility that extremely dry conditions may immediately follow extremely wet conditions ($DSD_i \geq 4$). Such rapid development of intense drought would be unprecedented and hard to predict, and contrary to the traditional view of drought as a slowly evolving extreme weather event.

3.3.6 Result 6: No clear consensus on intermediate-term drought

As depicted by the Fig. 3.1, for intermediate-term drought, there are substantial disagreements of future change among LE models and between the two drought indexes (SPI6/SNPI6). Although nearly all models indicate that the magnitude of exceptional intermediate droughts (SPI6/SNPI6 at

3.3. RESULTS

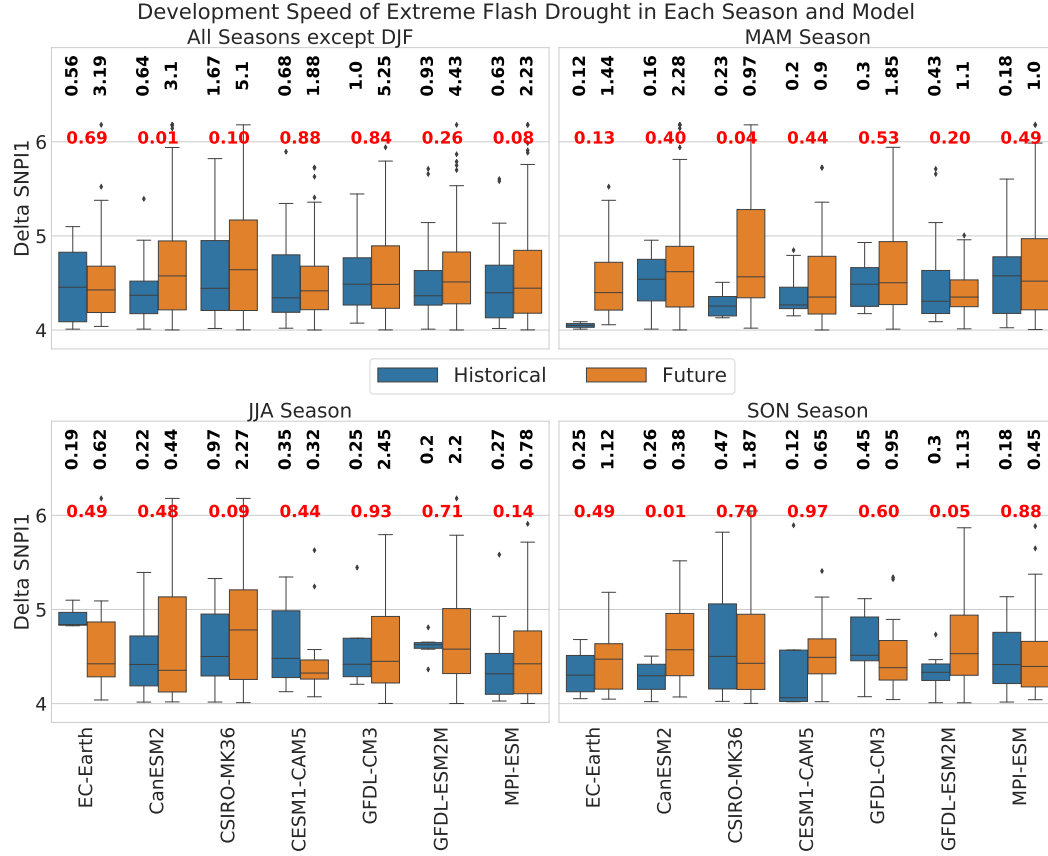


Figure 3.5: Development speed of extreme flash drought for each season (except winter) and model. The black bold numbers above each box plot indicate the average frequency within each period (51 years). The red bold numbers indicate the p-values obtained from a t-test of the Delta SNPI1 mean between historical and future periods.

the lowest boundary of -3.09) are projected to increase, SPI and SNPI show significant divergence at this timescale. All LE models agree that monthly intermediate-term dryness tends to become less frequent (months with $SPI6 \leq -1$) and general wetness (the average $SPI6$) tends to increase; however, when looking at $SNPI6$, only 4 out of 7 models project such increases. Therefore, to better describe the changing patterns of intermediate drought, we compare three essential drought features – drought duration, frequency, and magnitude.

From Fig. 3.6, we can see that, as a result of significant increases in precipitation, all LE models project decreases in $SPI6$ -derived intermediate drought frequency, average duration, maximum duration, average magnitude and maximum magnitude. On the other hand, for the $SNPI6$ -derived

intermediate drought, most LE models suggest that intermediate-term droughts tend to be more frequent (6 out of 7) but with shorter average duration (5 out of 7). However, the mean and maximum intermediate-term drought magnitude exhibits a wide spread between periods with little agreement between models, suggesting no statistically significant change.

3.4 Discussion

Flash drought and unprecedented extreme flash droughts are projected to be a major challenge into the future, with faster initiation and greater intensity; this result holds especially true in the spring season, and is certainly projected to threaten local agricultural production. At meanwhile, the underlying causes of flash drought intensification is still ambiguous thus we will further explore the potential drivers of increasing flash droughts. In this study, we have used SNPII and its change over subsequent months to quantify frequency and intensity of flash droughts, that is, rapidly developing water shortage related to precipitation, temperature, and evaporative demand (Taylor et al., 2013). By construction, changes to the frequency and character of flash drought are projected to be brought about through changes to net precipitation, i.e., precipitation inputs minus evapotranspiration outputs. In this section, we turn our attention to explaining future trends in drought and flash drought through a detailed examination of precipitation, temperature, evapotranspiration, leaf-area index, and their associated relationships. In general, we note that increasing monthly precipitation variability affects the inputs to the net precipitation, while P/E anti-correlation modifies the outputs. Consequently, these modifications to net precipitation drive a clear intensification of flash droughts over the NEUS, particularly in the spring season.

3.4.1 More precipitation and more variable precipitation in the NEUS

In Fig. 3.3 (left column), a clear positive shift in mean precipitation intensity is observed in all models, along with a flattening of the frequency density and subsequent increase in the frequency in both tails. This change is indicative of the well-known shift to increased precipitation variability under global warming, driven by increasing moisture and mitigated by weaker atmosphere circulation (Pendergrass et al., 2017). With that said, precipitation change over the NEUS has a

significant seasonal pattern (Frumhoff et al., 2007; Hayhoe et al., 2007) that necessitates a deeper examination of each season separately. To that end, Fig. 3.7 shows precipitation mean and standard deviation of each LE model. From here it's quickly apparent that nearly all LE models project that precipitation exhibit a significant increase in its mean and variance over all seasons, in agreement with historical observations of increasing annual mean precipitation and intensified extreme precipitation events (Demaria et al., 2016; Frumhoff et al., 2007; Pendergrass et al., 2017). Among the 7 LE models, the relative change between historical (1950-2000) and future periods (2050-2100) of the mean ranges from 7.57% to 15.70%, while the increase in the standard deviation ranges from 14.10% to 35.90%.

Increasing regional mean precipitation is certainly the main driver behind the general wetting trend in this region; however, the increase is not uniform over four seasons. As depicted in Fig. 3.7, winter (DJF) and spring (MAM) produce a greater precipitation increase, with multi-model mean relative change of 20.36% and 16.75%, significantly exceeding increases in the summer (4.38%) and autumn (3.31%). This seasonal dependency agrees with previous studies examining GCM projections (Frumhoff et al., 2007; Hayhoe et al., 2008; Xue and Ullrich, 2021b). Similarly, increases in precipitation variability in the winter and spring seasons are also larger than corresponding increases in summer and autumn. Notably, each LE model's mean and standard deviation are strongly correlated in almost all seasons: correlation in winter, spring, and autumn is 0.95, 0.96, and 0.74; however, this correlation is not significant in summer (0.04), suggesting this season is dominated by different processes. This synchronization is a topic for potential future exploration. In conjunction with warmer temperatures shifting snowfall to rainfall, particularly in the spring, it is clear that these two seasons may experience more extreme precipitation and flooding (Xue and Ullrich, 2021b). On the other hand, as shown in Fig. 3.3, 6 out of 7 models agree that months with nearly zero monthly precipitation (monthly precipitation less than 0.5 mm/day) tend to be more likely in the future, thus supporting an increase in the occurrence of flash droughts.

3.4.2 Evapotranspiration is energy-limited within the humid NEUS

Increased precipitation variability is one of the primary drivers behind possibility of emerging flash drought; however, this single factor is insufficient to explain why spring is at the greatest risk for increased flash drought, while it does not have the greatest increase in variability. To understand this discrepancy, it is necessary to additionally consider evapotranspiration's role in the development of flash droughts. Evapotranspiration depends on the water and energy availability, and so it is important to ascertain whether particular regions are energy-limited or water-limited (Roderick et al., 2009a,b; Teuling et al., 2009). Put simply, the amount of available energy can limit evapotranspiration in regions where water is abundant, whereas water supply limits it in regions of abundant energy. As not all LE models provide radiation data, here we employ surface air temperature and precipitation to quantify the supply of energy and water, respectively. While other factors are important for estimating evapotranspiration, such as surface wind, relative humidity and vegetation, changes in these factors under global warming are modest compared with changes in precipitation and temperature (Held and Soden, 2006; Laîné et al., 2014; Ma et al., 2016; Trenberth et al., 2014).

In the NEUS, all LE models exhibit very strong correlations between the multi-ensemble mean surface air temperature and evapotranspiration. Correlations range from 0.92 to 0.99 (with p-values much less than 0.05 which indicates that the correlation is statistically significant at 95% confidence level), with the mean of 0.95 during the historical period (1950-2000). This trend is even more significant in the spring season with a mean correlation of 0.97 (with p-values much less than 0.05). Consequently, it is safe to say that evapotranspiration over the NEUS is primarily energy-limited. On the other hand, correlations between evapotranspiration and precipitation vary among models, with a mean correlation of 0.04 (with all p-values less than 0.05). To further confirm that surface air temperature increases are the primary driver of evapotranspiration differences, we calculate the correlation of each ensemble's average deltas of precipitation, evapotranspiration, and surface air temperature between the historical and future periods over all LE members. As in the historical data, the correlations between the deltas of evapotranspiration and surface air temperature are

obvious, with the multi-ensemble mean ranging from 0.15 to 0.83 (5 out of 7 models have p-values less than 0.05). However, examining the correlations between deltas of evapotranspiration and precipitation does not produce robust relationship, with the multi-ensemble mean correlations varying from -0.24 to 0.65 with a multi-model mean of 0.07 (only 1 model has p-value less than 0.05).

3.4.3 Differences in the trends of precipitation and evapotranspiration variability

From Fig. 3.3, we see that the density plots of surface air temperature and evapotranspiration essentially all shift to the right while maintaining the same “shape” of the distribution. This is in contrast to precipitation, which exhibits both an increase in its mean and a widening of its distribution. Although evapotranspiration and precipitation show increases in the multi-model mean of 13.76% and 11.43% over the NEUS, the relative change of the standard deviation of surface air temperature (0.77%) and evapotranspiration (-0.96%) are essentially negligible compared when compared with that of precipitation (22.86%). Note that the small decrease in evapotranspiration variability appears to primarily emerge from a decrease in the frequency of low evapotranspiration months. Nonetheless, net precipitation is subject to both a shift in its mean (with a multi-model mean increase of 7.52%) and an increase to its variability (with a multi-model mean increase of 20.16%). Namely, while the regional climate becomes wetter, there will be an increase in the frequency of both the extremely dry and wet conditions that is clearly apparent in the fourth column of Fig. 3.3. Looking more closely into each season, the variability increase is much more significant in the spring as a result of greater precipitation variability in these seasons; however, MAM and SON evapotranspiration variability actually decreases slightly (with multi-model mean relative change of -3.11% and -2.43%). MAM is projected to produce the most discrepant variability trends, with reduced evapotranspiration variability and simultaneously greater precipitation variability. Consequently, the spring tends to have the largest increases in the variability of net precipitation. The disparate behavior of precipitation and evapotranspiration leads to increasing anti-correlation of these processes, which we investigate next.

3.4.4 Increasing anti-correlation of evapotranspiration and precipitation

In the NEUS, where evapotranspiration is energy-limited, there is a very strong positive correlation between monthly mean surface air temperature and evapotranspiration that occurs in all 7 LE models over the historical period; however, correlations between precipitation and evapotranspiration vary largely across the 7 LE models, ranging from -0.29 to 0.65 (with p-values less than 0.05). Under future global warming, the positive correlations between surface air temperature and evapotranspiration continue to be evident, with a multi-model mean correlation of 0.92 (with p-values less than 0.05). In the spring season, these correlations even reach a multi-model mean value of 0.98 (with p-values less than 0.05). On the other hand, the correlations between precipitation and evapotranspiration become increasingly negative into the future over all seasons (with an average decrease of 0.24). This trend is even more significant in the spring (with an average decrease of 0.46). Annually, 6 out of 7 LE models project that these correlations will be far more negative in the future, compared with the historical period. For the only model that shows an increased correlation, CSIRO-MK36, precipitation and evapotranspiration in the future spring season remains negative (-0.34) (with p-value less than 0.05). Across all models, multi-model mean historical correlations between precipitation and evapotranspiration are roughly zero annually (0.04) and in the spring season (-0.04); in the future they decrease to -0.21 and -0.50 respectively. Such a negative shift in correlation indicates that, during future periods of low precipitation, evapotranspiration will be more likely to be greater than under analogous historical circumstances, in turn magnifying the local moisture deficit. Similarly, when precipitation is high, evapotranspiration is not expected to be obviously larger than its mean state as what it would have been historically. These two effects, in turn, exacerbate extreme events such as flash drought and the extremely wet conditions, as apparent in both tails of density plots of Fig. 3.8.

To better illustrate how P/E anti-correlation impacts flash drought, we define anti-correlated low moisture conditions as the months with precipitation no larger than 50th percentile and evapotranspiration larger than 50th percentile which experience the subnormal moisture input but above normal moisture output, and provide appropriate conditions for the flash drought. In Fig. 3.9, we

normalize each model’s multi-ensemble regional monthly mean precipitation and evapotranspiration within the range of 0 and 100 in historical and future periods via

$$\langle \text{Data} \rangle_i = \frac{\langle \text{Rank of Data} \rangle_i \times 100}{\langle \text{Total Number of Data} \rangle_i}. \quad (3.5)$$

The points of anti-correlated low moisture condition we are interested in are located at the upper left corner of the figure. Indeed, we see that 6 out of 7 models project a large increase of such anti-correlated low moisture conditions in the spring with a multi-model mean relative increase of 54.80%, with little doubt that this shift is attributed to P/E anti-correlation. As a result of increasing P/E anti-correlation, anti-correlated high moisture conditions (precipitation larger than 50th percentile and evapotranspiration no larger than 50th percentile) will also increase, providing favorable moisture conditions for floods.

In the next section we explore the processes underlying these shifts and argue that an earlier onset (spring) of growing seasons is a likely culprit (Backlund et al., 2008; Christiansen et al., 2011).

3.4.5 Modified evapotranspiration partitioning brought by the extension of growing season

Several previous studies have shown that changes in vegetation are a key factor behind trends in evapotranspiration (Peel et al., 2010; Zhang et al., 2015, 2001), while intensified water demand from plants in the growing season also plays an essential role in the development of flash drought (Otkin et al., 2018; Pendergrass et al., 2017, 2020). These processes have certain implications for the regional hydroclimate of the NEUS, as this region is predicted to experience a longer growing season with denser vegetation under a warming climate (Christiansen et al., 2011; Frumhoff et al., 2007; Xue and Ullrich, 2021b). Indeed, we now argue that this is the primary factor driving the increasing P/E anti-correlation in the spring (and to a lesser degree in the autumn). Specifically, we argue that the modified partitioning of evapotranspiration brought on by a prolonged growing season directly drives this anti-correlation and, consequently, flash droughts.

Total evapotranspiration is defined as the sum of evaporation from soil and surface water (E_{soil}),

the evaporation from water intercepted by plants (E_{plant}), and the plant transpiration (T_{plant}) (Ferguson and Veizer, 2007; Lee et al., 2010; Wang and Dickinson, 2012). Of these, transpiration is the largest contributor, although all three are captured in modern ESMs (Dirmeyer et al., 2005; Lawrence et al., 2007). A number of past studies have used observations and models to demonstrate the impacts of interannual shifts in vegetation on evapotranspiration that occur through modification of its partitioning into these categories (Gong et al., 2007; Jung et al., 2010; Lawrence et al., 2007; Lawrence and Chase, 2009; Wang and Dickinson, 2012; Wang et al., 2010). It has been noted that soil evaporation can only occur if it acquires water at the surface level, because of the disconnect between surface and deep soil layers, especially during the dry season and over bare soil (Heitman et al., 2008; Wang and Dickinson, 2012); however, transpiration from plants can extract water from deeper soil through their rooting systems. As a result, most soil evaporation occurs along with or shortly following precipitation, while transpiration can exhibit a lagged response to precipitation and is associated more with biological processes and energy available (solar radiation) (Wang and Dickinson, 2012; Williams et al., 2004). As a direct consequence of increased vegetation in the NEUS, particularly in the spring, there’s good reason to believe vegetation changes (might be caused by climate change induced land cover change or by land-use change) plays a greater role in driving surface evapotranspiration. Namely, evapotranspiration tends to draw more water from the rooting zones instead of surface soil, and so will be more weakly associated with recent or concurrent precipitation. In turn, there is an increasing anti-correlation of evapotranspiration and precipitation.

To further confirm this hypothesis, we employ the large ensemble CESM1 model, the only LE model that provides three evapotranspiration components (E_{soil} , E_{plant} and T_{plant}) and Total Leaf Area Index (TLAI), to examine changing trends in vegetation and evapotranspiration partitioning. We also introduce the Plant Transpiration Component Ratio (PTCR), which is defined as the ratio of plant transpiration to the soil evaporation. PTCR is employed as a metric to quantify the increasingly noteworthy role of plants in driving surface evapotranspiration. From Fig. 3.10, we observe a clear increase in ensemble mean TLAJ over the next century, which is particularly significant in the spring season. Namely, compared with the historical period (1950-2000), the

future period (2050-2100) TLAI is projected to be 33.42% larger over all seasons and 45.24% larger over the spring season only. Annual mean PTCR is highly associated with more vegetation (TLAI) (with a correlation of 0.75 (with a p-value less than 0.05) over all seasons and 0.91 (with a p-value less than 0.05) during the spring season). It has risen continuously since the 1950s, indicative of how transpiration from plants is becoming a greater component of surface total evapotranspiration. The relative change of PTCR in the spring (39.54%) is much larger than the relative change during all season (10.35%) and other individual seasons, supporting our conclusion that the extension of growing season is the main driver of P/E anti-correlation in the spring. We do recognize that the land-use change might be a driver of the change of TLAI; however, it is unclear if changes in vegetation types across the region could explain the seasonality in the TLAI changes we observed in this analysis. Note that the TLAI has a faster increase than PTCR, which is potentially caused by the increasing water use efficiency which has been illustrated by the observations (Keeling et al., 2017; Keenan et al., 2013; Swann et al., 2016).

Although changes to evapotranspiration partitioning in the growing season and its downstream effect on the increasing P/E anti-correlation is well-supported by past studies and CESM1-CAM5 data, one could hypothesize that the increasing P/E anti-correlation is instead related to the increasing soil moisture brought by earlier snowmelt in the growing season (Frumhoff et al., 2007; Xue and Ullrich, 2021b). To demonstrate this is not the case, we employ all LE models which provide the soil moisture data (CESM1-CAM5, CSIRO-MK36, GFDL-CM3, and MPI-ESM) and examine if soil moisture is higher in the spring season into the future. While this would, in turn, provide more water for evapotranspiration (see Supplement 3.6.5), it is notable that evapotranspiration is primarily energy-limited in this region (as discussed earlier). However, even if this was the case, these models actually exhibit soil drying in the spring season: although there are some discrepancies among the four models, they all agree that the topsoil moisture (defined as the average soil moisture of the first 0.1-meter soil) will decrease into the future. Indeed, 3 out of 4 models agree that such a decrease also occur within all soil layers. This indeed supports the claim that snowmelt’s influence on springtime soil moisture is not the primary driver of the observed P/E anti-correlation. Importantly, we find that all models agree that the correlation between the

ensemble mean soil moisture and net precipitation is much larger than the correlation of ensemble mean soil moisture and precipitation, again illustrating the role of net precipitation in driving soil moisture.

We note that CSIRO-MK36 is the only model among the LE models which does not produce an increased P/E anti-correlation into the future. This is also the only model among those analyzed which does not account for changes in vegetation into the future. This lends credence to our theory of vegetation being the primary driver of increasing anti-correlation. Also, we need to clarify that although some models project that there exists a negative correlation between precipitation and evapotranspiration in the future MAM, it does not mean that relatively high precipitation directly determines the relatively low evapotranspiration and also does not suggest that the region is water-limited. Instead, this is caused by different monthly changing trends brought by various factors. For example, the precipitation will increase more in March (0.81 mm/day) instead of May (0.26 mm/day) in the future, which we believe is mainly driven by more warming in cold months (also shown above) due to the snow-albedo feedback (Xue and Ullrich, 2021b). Meanwhile, compared with March, May will experience a stronger growing season extension because the length of growing season is usually defined by the threshold based on daily minimum temperature (Christiansen et al., 2011). Therefore, a stronger growing season extension makes May have a larger transpiration increase compared with March, as we can see the CESM1 multi-ensemble plant transpiration increase is much larger in the May (0.28 mm/day) compared with March (0.06 mm/day). More warming in March and larger growing season extension in May make evapotranspiration have an even increase in the March, April and May with multi-model means of 0.44, 0.39 and 0.35 mm/day. Therefore, the evapotranspiration will not have the same monthly changing trend as the precipitation which partially induces their negative correlations in the future. Note that the projection of flash drought is based on monthly scale in this study and may ignore the flash drought and its development within sub-monthly scale. And the examine of SNPI1 as a flash drought index is only examined over NEUS and US Midwest at monthly scale (see Supplement). A comprehensive examination of SNPI1 in capturing development of flash drought need to and will be put forward before its wide and practical usage. Also, the LENS models used are imperfect and process uncertainties especially

in land surface processes thus corresponding study of projected changes in actual soil moisture are needed.

3.5 Conclusions

To better understand trends in drought character and frequency over the NEUS, we have applied both the Standardized Precipitation Index (SPI) and Standardized Net Precipitation Index (SNPI) to 7 large ensemble model datasets. These two indices have subsequently enabled insight into shifts in this region’s precipitation characteristics and atmospheric water demand characteristics, respectively. Short-term, intermediate-term, and long-term droughts are explored with SPI3/SNPI3, SPI6/SNPI6, and SPI24/SNPI24, while our study of flash drought employs SNPI1, which captures drought events that are characterized by rapid development. A clear rightward shift in the SPI probability distributions indicates that the NEUS will experience significant wetting as a result of precipitation increases, particularly at longer time scales. Even in light of similarly increasing mean evapotranspiration, most models (5 out of 7) still project a positive trend in overall water availability. Flattening of the frequency distribution of SPI/SNPI among all LE models indicates more frequent extremes, particularly wet extremes, at the expense of more moderate periods. For example, the multi-model mean frequency of exceptionally wet months (SPI1/SNPI1 equal to 3.09) is predicted to increase by 19 times (using SPI1) and 12 times (using SNPI1). Consequently, more frequent and intense flooding is expected to become a growing concern for local water management. This wetting trend does not imply that the drought is a purely historical concern, however. In light of surging evapotranspiration and increasing precipitation variability, 6 out of 7 models suggest that short-term extreme droughts ($SNPI3 \leq -2$) is also projected to be much more frequent in the future. Moreover, flash droughts (as indicated by SNPI1) will see a 106% increase in frequency, dropping the average return time from 6.3 years to 2.9 years, with faster development. Extreme flash drought, where SNPI1 drops by at least 4 points to below -2, is projected to exhibit a drop in return period from 65.9 years to 16.2 years. This increase in frequency for both flash droughts and extreme flash droughts is most pronounced in the spring season, and will likely drive significant challenges to the agriculture and ecosystem. Although the LE models don’t provide a clear con-

sensus on intermediate-term drought, most models agree that intermediate-term drought tends to be more frequent, with a lower average duration. The projected changes in flash drought, particularly during the spring season, are attributed to an increase in precipitation variability that is not matched by evapotranspiration; consequently, these two fields become increasingly anti-correlated. Increasing precipitation variability has been well-studied, and is generally attributed to increased atmospheric moisture counteracted by a weakening circulation. We confirm that the increasing anti-correlation, which is strongest in the spring, is largely a result of the extension of the growing season and attributed to increase in leaf area index (i.e., more vegetation). Namely, direct soil evaporation is typically highly correlated with precipitation because it draws from moisture in the surface soil; however, rooting enables plants to derive soil moisture from deeper in the soil. As a result, we argue that increased vegetation in the NEUS, which is associated with the shift to a warmer, moister climate is expected to increase the risk of the sudden drying episodes explored in this work.

3.6 Supplement

The contents of the supplement are as follows. In subsection 3.6.1, the map of study region is provided. In subsection 3.6.2 and 3.6.3, we conduct case studies of flash drought based on ERA5, a high quality reanalysis dataset, and all ensembles of CESM1-CAM5. In subsection 3.6.4, we justify the calculation of SNPI. In subsection 3.6.5, we provide an analysis of total soil moisture (i.e., the total all phases water over all soil layers) and topsoil moisture (i.e., the top 0.1 meter of soil) over the NEUS based on the ensemble mean of the four LE models that provide the necessary soil moisture data (CESM1-CAM5, CSIRO-MK36, GFDL-CM3, and MPI-ESM). Additionally, we examine the correlation of soil moisture with precipitation and net precipitation to illustrate the strength of this latter relationship. This result confirms that the net precipitation can capture soil moisture trends and thus validates that SNPI is a good metric for examining drought. The supporting figures show the map of study regions (Fig. 3.11), the evolving process of flash droughts in the ERA5 over NEUS (Fig. 3.12-3.15) and the CESM1-CAM5 (Fig. 3.16-3.20), the evolving process of flash droughts in the ERA5 over Kansas (Fig. 3.21-3.23), Q-Q plots of all LE models

in each month (Fig. 3.24-3.35), and each model’s ensemble average long-term monthly mean soil moisture over the NEUS over historical and future periods (Fig. 3.36-3.39). The supporting table shows correlations between monthly mean soil moisture and topsoil moisture with precipitation, net precipitation and evapotranspiration in the historical and future periods (Table 3.2-3.3). Although the SNPI1 is designed as a metric in capturing rapidly developing droughts (flash droughts at one-month interval), we can still see that it performs well in the flash drought case studies conducted.

3.6.1 The study region and grid cells included

As we mentioned, this study is conducted over the New England Region, defined by the Watershed Boundary Dataset (Survey and US Department of Agriculture, 2013). All LE models have been interpolated onto a 1° grid by conservative interpolation to make sure all models have the same boundaries. This process has been shown to have a little impact on climate models’ projection on the drought in our previous study (Xue and Ullrich, 2021a). As shown in Fig. 3.11, the New England Region is located in the northeastern US and covers 23 grid cells.

3.6.2 A case study of rapidly developing drought based on ERA5 over Northeastern US

In order to examine if the net precipitation and SNPI1 are good indicators for detection of rapidly developing drought and rapid decrease in soil moisture, here we firstly employ the 6-hourly ECMWF Reanalysis 5th Generation (ERA5) from 1950 to 2000 (the historical period in the main paper), which is a next-generation reanalysis product that replaces the ERA-Interim reanalysis (European Centre for Medium-Range Weather Forecasts, 2020; Hersbach et al., 2020). As the most advanced reanalysis data to date, ERA5 has improved data assimilation, core dynamics, model physics, temporal and spatial resolution. Consequently, the ERA5 has demonstrable performance at representing low-frequency variability and producing more accurate meteorological variables. It has also been shown to be better at simulating extreme weather events like droughts (Dullaart et al., 2020; Hersbach et al., 2020; Tarek et al., 2020). We do not directly use observational dataset since there does not exist a comprehensive observational dataset providing all variables needed. Also, the

focus of this study is not to confirm that SNPI is successful when applied to observational datasets because the “observed” actual evapotranspiration itself usually comes from estimations based on the satellite remote sensing data (which relies on algorithms) or empirical equations (García et al., 2013; Li et al., 2016; Wu et al., 2021). Therefore, here we will examine the evolving processes of SNPI1, precipitation, actual evapotranspiration, net precipitation and root zone soil moisture (the top 40 cm soil moisture according to previous study (Ford and Labosier, 2017)) when the significant root zone soil moisture drops quickly.

From Fig. 3.12 we can see that, from the annual mean SNPI6, the 1960s drought is the most extreme drought event over the historical period, and is deemed as the most extreme one over the NE in the past century (Janes and Brumbach, 1965; Palmer, 1965). Also, the annual mean SNPI6 fits the root zone soil moisture well and it is highly correlated with annual mean root zone soil moisture (correlation is 0.70 and p-value is 1.92×10^{-8}). Note that to represent each year’s drought condition, we use the annual mean SNPI6 and because SNPI itself already consider the accumulation of drought conditions so that we choose an intermediate-term SNPI (SNPI6).

Following Ford and Labosier (2017), here we simply define rapid root zone soil moisture decrease as when the monthly root zone soil moisture decrease larger than the 20th percentile (e.g., the root zone soil moisture drops from its 40th percentile to 20th percentile). Since SNPI/SPI is standardized within each month to eliminate seasonality, the soil moisture percentile here is also calculated within each month so that the fast root soil moisture drops will not be limited to soil moisture’s strong seasonality (i.e., soil moisture tends to hugely decrease in the summer and then get recharged). We examine whether the rapidly developing drought events detected by our definition have such a large decrease in root zone soil moisture. Overall, there are 7 rapidly developing drought events detected using SNPI1, and 6 of these months possess rapid root zone soil moisture decrease larger than 20th percentile in that month or the following month. Among these 7 months, the 1999 April experienced both the largest SNPI1 decrease (DSD) and root zone soil moisture drop (in percentile). From Fig. 3.13, we can see that the large DSD and negative net precipitation directly causes the rapid drop in root zone soil moisture, and even in the dry months with subnormal soil moisture, the net precipitation is still negative. The soil moisture is recharged when the net precipitation

becomes positive. It also confirms the ability of SNPI1 in capturing the flash drought over NEUS.

This historical 1999 NEUS drought itself is not only deemed as the considerable one after the 1960s NEUS drought but also has an unusually fast development (Blaine Friedlander , 2000; National Center for Environmental Information, 2022a,b) and is one of the most severe short-term droughts in the past century (Paulachok et al., 2000). At the beginning of 1999, the precipitation was abnormal in January and March, which caused the about normal groundwater levels over the NEUS (Paulachok et al., 2000); however, the rainfall deficits and intensified evapotranspiration made the drought initiated rapidly (Dupigny-Giroux, 2001). As its fast development during the spring and summer, the impacts of water shortage became evident in shallow-rooted vegetation like flowers and grasses (Dupigny-Giroux, 2001). And the extreme heatwave in the summer further aggravated the drought conditions (Blaine Friedlander , 2000; National Center for Environmental Information, 2022a) with the emergence of record low streamflow and groundwater levels over the NEUS (Paulachok et al., 2000). In a word, although the year 1999 started with above-average rainfall, fast drought development made the April-July period the second driest such period in the past century (Blaine Friedlander , 2000; Dupigny-Giroux, 2001; National Center for Environmental Information, 2022a). Consequently, the agriculture losses were more than 1 billion over NEUS (Morehart et al., 1999; National Center for Environmental Information, 2022b). In Fig. 3.14 and 3.15, the development of flash drought over NEUS from 1999 March to April is well observed.

3.6.3 Case studies of rapidly developing drought based on CESM1-CAM5 over Northeastern US

To further examine if net precipitation and SNPI1 are good indicators to detect rapidly developing drought and fast decreases in soil moisture, here we employ all forty CESM1-CAM5 ensembles during the historical period (1950-2000), as CESM1-CAM5 is one of four LE models providing the soil moisture data.

Here we simply follow the analysis in the section above. The only difference is that here we use the surface soil (the top 0.1-meter soil) moisture instead of root zone soil moisture because LE models can only provide the average surface soil moisture or the average total soil moisture (the

total soil usually has thickness larger than 2 meter which makes it have a longer response time). Overall, there are 369 rapidly developing drought events in forty CESM1 ensembles, as indicated by our rapidly developing drought definition based on SNPI1, and 311 of them possess the rapid surface soil moisture decrease larger than 20th percentile in the identified month or the following month. Also, we find that the Development Speed of Drought (DSD, which is used to define rapidly developing drought in this paper) is a good indicator of large surface soil moisture drop. Among the top 40 months with the largest surface soil moisture percentile decrease, 38 of them have DSD larger than 2 (the threshold of DSD we used to define rapidly developing drought).

To better illustrate how an evolving drought manifests in relevant variables, we look into the SNPI1, precipitation, actual evapotranspiration, net precipitation and surface soil moisture (in percentile) in the years having months with the largest surface soil moisture decrease (Fig. 3.16 to 3.20). It's obvious that the significant surface soil moisture decreases occur in months with large DSD (which is much larger than 2), subnormal precipitation, abnormal evapotranspiration and negative net precipitation. On the other hand, increasing SNPI1 and abnormal net precipitation indicate recharging of surface soil moisture. Therefore, we further confirm that the SNPI1 and net precipitation are good indicators of soil moisture conditions even during rapid drying events.

3.6.4 A case study of infamous 2012 US Midwest flash drought based on ERA5

As we already demonstrated the ability of SNPI1 in capturing the soil moisture and rapidly developing drought development in ERA5 and CESM1, in this section we aim to examine the performance of SNPI1 in real-world flash drought events. A particularly positive example is the most infamous flash drought event to occur over the US – the 2012 US Midwest flash drought which happened over the predominant agricultural lands of the central US and brought losses more than \$30 billion (Hoell et al., 2020; Hoerling et al., 2012; National Center for Environmental Information, 2021; Otkin et al., 2021; Pendergrass et al., 2020). The 2012 US Midwest flash drought started in April with north-central Kansas as its epicenter and then developed across large portions of the Central US with considerable impacts to agriculture and water resources (Basara et al., 2019; DeAngelis et al., 2020; Otkin et al., 2016).

Note that, to cover this flash drought event, here we employ the ERA5 from 1950 to 2020. The 2012 flash drought evolving process over the epicenter (Kansas) is shown in Fig. 3.21. We can see that SNPI1 dropped from below 0 in April to below -2 in May with Delta SNPI1 larger than 2, which exactly satisfies our definition of flash drought. Then SNPI1 remained below 0.05 in the rest months except April, which indicated subnormal net precipitation and further aggravated drought conditions. Correspondingly, the regional mean root zone soil moisture dropped from 54th in April to 10th percentile in May with a decrease much larger than 20th percentile (the threshold we used to define rapid root zone soil moisture decrease) and then kept below the 10th percentile over the whole growing season (until October). This explains why the 2012 flash drought brought such extensive agricultural losses. Moreover, in Fig. 3.22 and 3.23, we can see that over the Kansas, the SNPI1 indicated neutral or wet conditions in April but a sudden drop below 2 in the central Kansas, which indicates the initiation of 2012 US Midwest flash drought. We further find that the flash drought development from April to May was brought by the considerable decrease in precipitation of May but, more importantly, the exceptionally intensified evapotranspiration from April to May. Namely, in Fig. 3.16 to 3.20, the annual evapotranspiration peak is usually from June to August; however, the peak lasted between April and May in 2012 over Kansas, which not only demonstrates the importance of evapotranspiration in the development of the 2012 flash drought but also illustrates the advantage of SNPI1 as a metric considering both precipitation and evapotranspiration in capturing flash drought.

3.6.5 Justification of SNPI calculation

This section illustrates our reasons for not using the log-logistic distribution (which is used in calculating Standardized Precipitation Evapotranspiration (SPEI)), and demonstrates that the adjusted net precipitation is fitted well by the gamma distribution. Although the three-parameter log-logistic distribution used in calculating SPEI (Vicente-Serrano et al., 2010) can be employed for negative values, there are several arguments to support the use of the gamma distribution for transforming SNPI. Firstly, this paper projects the changing trends of SPI and SNPI droughts. As the most widely used drought index, the SPI is suggested to be fitted to gamma distribution. There-

fore, to compare droughts with and without actual evapotranspiration, fitting the net precipitation to the same distribution can make the comparison more straightforward. More importantly, we conduct a comprehensive examination of the goodness-of-fit of precipitation, adjusted net precipitation and net precipitation to the gamma distribution and three parameters log-logistic distribution respectively. From Fig. 3.24 to 3.35, we provide Quantile-Quantile (Q-Q) plots to demonstrate goodness-of-fit for a certain distribution. Namely, we plot the quantiles of samples and prescribed distribution against each other, if the points approximately lie on the line $y = x$ then we can claim the samples fit the distribution well. We can see that for all LE models in all months, both the precipitation and adjusted net precipitation fits well to the gamma distribution and are much better than the fit of net precipitation to log-logistic distribution. To further quantify the fit, we conduct the Kolmogorov–Smirnov Test (K-S Test) in each case and provide corresponding p-values in Fig. 3.24 to 3.35. In all cases, the K-S Test p-values of precipitation and adjusted net precipitation, when compared to the gamma distribution, are much larger than 0.05, which means that at more than 95% confidence level we cannot reject the null hypothesis that the samples come from the prescribed distribution; however, the K-S Test p-values of net precipitation to the log-logistic distribution are less than 0.05 in some cases. Therefore, regardless of the ability of the log-logistic distribution to capture negative values, the better fit for adjusted net precipitation by the gamma distribution justifies its use in this study.

3.6.6 The soil moisture analysis

In order to determine if snowmelt-driven increases in soil moisture are responsible for increases in P/E anti-correlation in the spring season, here we analyze the soil moisture from all four LE models that provide the needed data. First, recall from the analysis in the main paper that all LE models except CSIRO-MK36 detected a P/E anti-correlation in the spring season. From Fig. 3.36-3.39, we find that three of the four models also agree that both the total soil column and topsoil actually contain less moisture in the spring in the future period. The exception, GFDL-CM3, projects an increase in total soil moisture in both winter and spring seasons; however, this result appears to be the exception, and so it appears that there is insufficient evidence to support snowmelt-driven soil

Table 3.2: Correlations between each model’s ensemble mean total soil moisture and precipitation, net precipitation and evapotranspiration.

Model	Precipitation	Net precipitation	Evapotranspiration	Period
CESM1-CAM5	0.30	0.67	-0.58	Historical
CESM1-CAM5	0.43	0.81	-0.71	Future
CSIRO-MK36	0.60	0.78	-0.67	Historical
CSIRO-MK36	0.53	0.86	-0.75	Future
GFDL-CM3	0.48	0.70	-0.62	Historical
GFDL-CM3	0.71	0.67	-0.32	Future
MPI-ESM	-0.04	0.85	-0.70	Historical
MPI-ESM	0.05	0.83	-0.56	Future

Table 3.3: Correlations between each model’s ensemble mean topsoil (top-0.1m) moisture with precipitation, net precipitation and evapotranspiration.

Model	Precipitation	Net precipitation	Evapotranspiration	Period
CESM1-CAM5	0.32	0.96	-0.87	Historical
CESM1-CAM5	0.52	0.98	-0.85	Future
CSIRO-MK36	0.38	0.89	-0.89	Historical
CSIRO-MK36	0.27	0.85	-0.88	Future
GFDL-CM3	0.47	0.89	-0.95	Historical
GFDL-CM3	0.87	0.95	-0.50	Future
MPI-ESM	-0.28	0.93	-0.84	Historical
MPI-ESM	0.23	0.93	-0.76	Future

moisture as the reason for P/E anti-correlation.

We further examine the correlations between each model’s multi-ensemble mean total soil moisture and topsoil moisture with precipitation, net precipitation and evapotranspiration (Table 3.2-3.3). From these results it’s evident that the correlations between net precipitation and both total soil and topsoil moisture are far larger than precipitation’s, which confirms the strong linkage between net precipitation and soil moisture. This result validates that SNPI is a good metric for assessing agricultural and rapidly developing drought within our region, as both of these types of drought are associated with deficient soil moisture over our study region. Similarly, the evapotranspiration shows a significant negative correlation with soil moisture, illustrating that evapotranspiration can represent the evaporative demand within our study region.

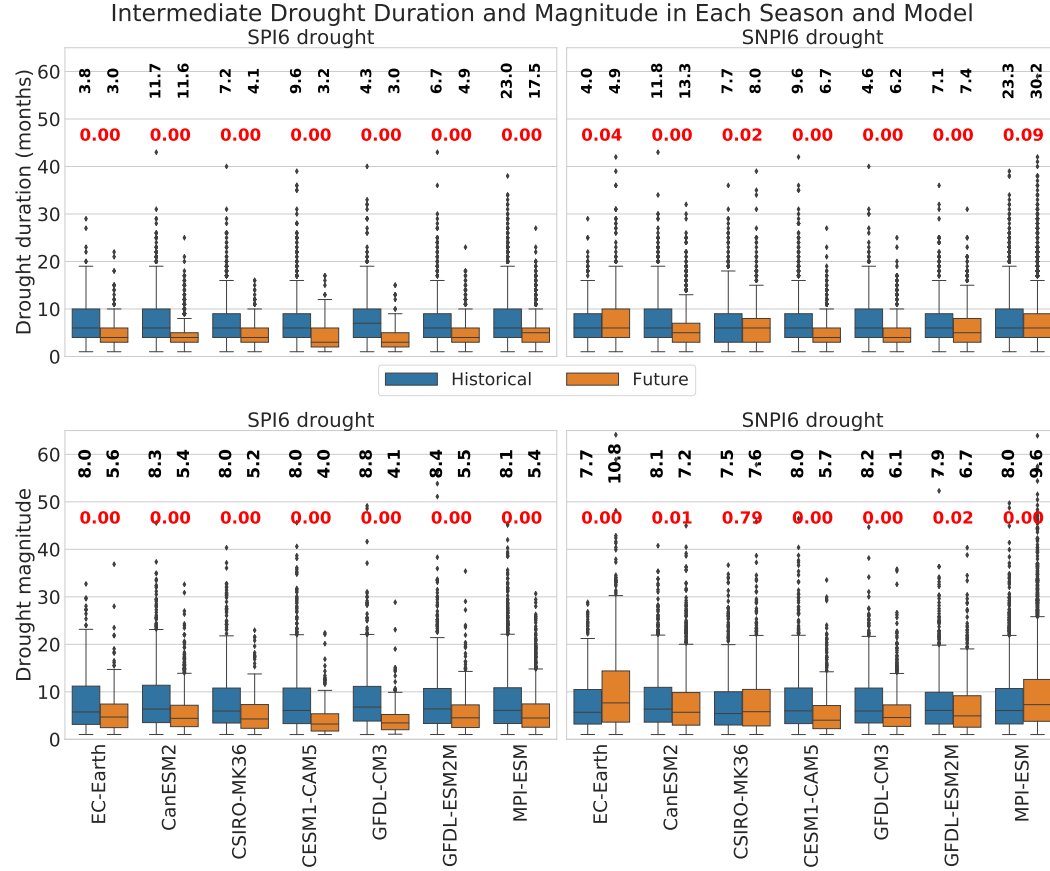


Figure 3.6: Duration and magnitude of intermediate-term droughts derived from (left) SPI6 and (right) SNPI6. For duration, the black bold numbers above each box plot indicate the average frequency (the number of dry months per 51 years). The red bold numbers indicate the p-values obtained from a t-test of the drought duration mean between historical and future periods. For drought magnitude (the accumulated SPI6/SNPI6 during each drought event), the black bold numbers above each box plot indicate the mean drought magnitude. The red bold numbers indicate the p-values obtained from a t-test of the drought magnitude mean between historical and future periods.

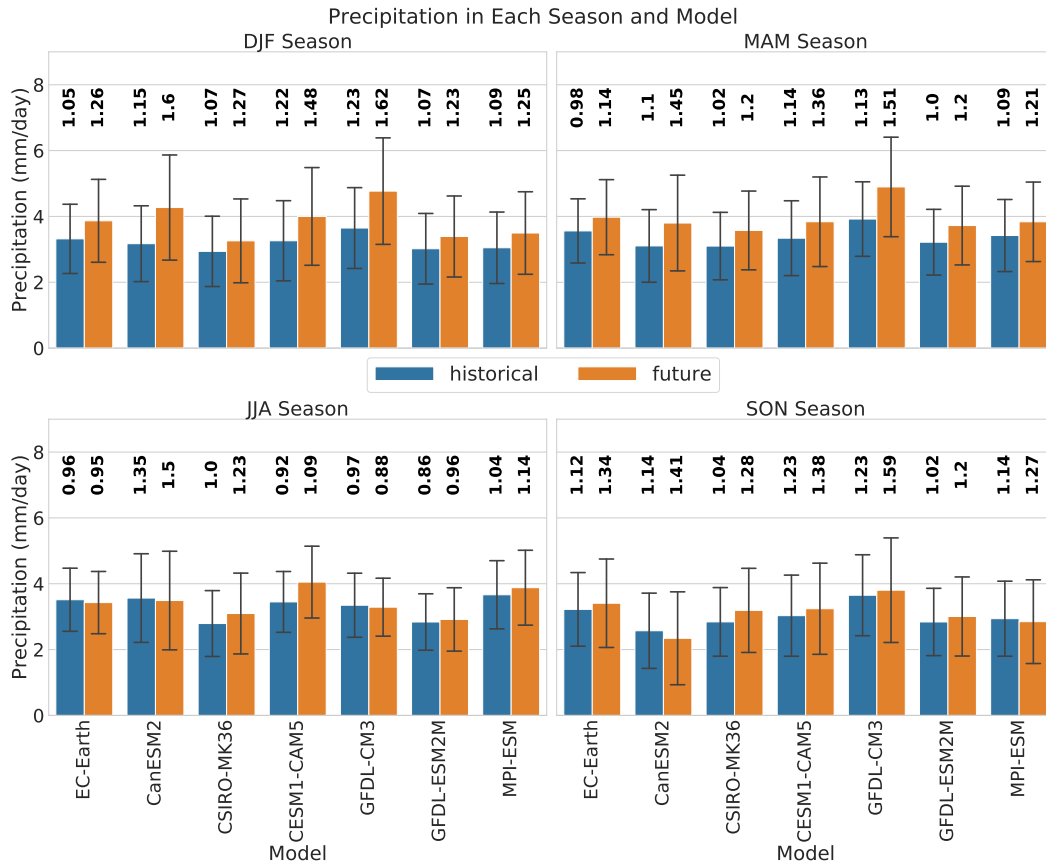


Figure 3.7: Mean and standard deviation of each LE model's precipitation during historical and future periods. The bold numbers above each bar indicate the value of the standard deviation.

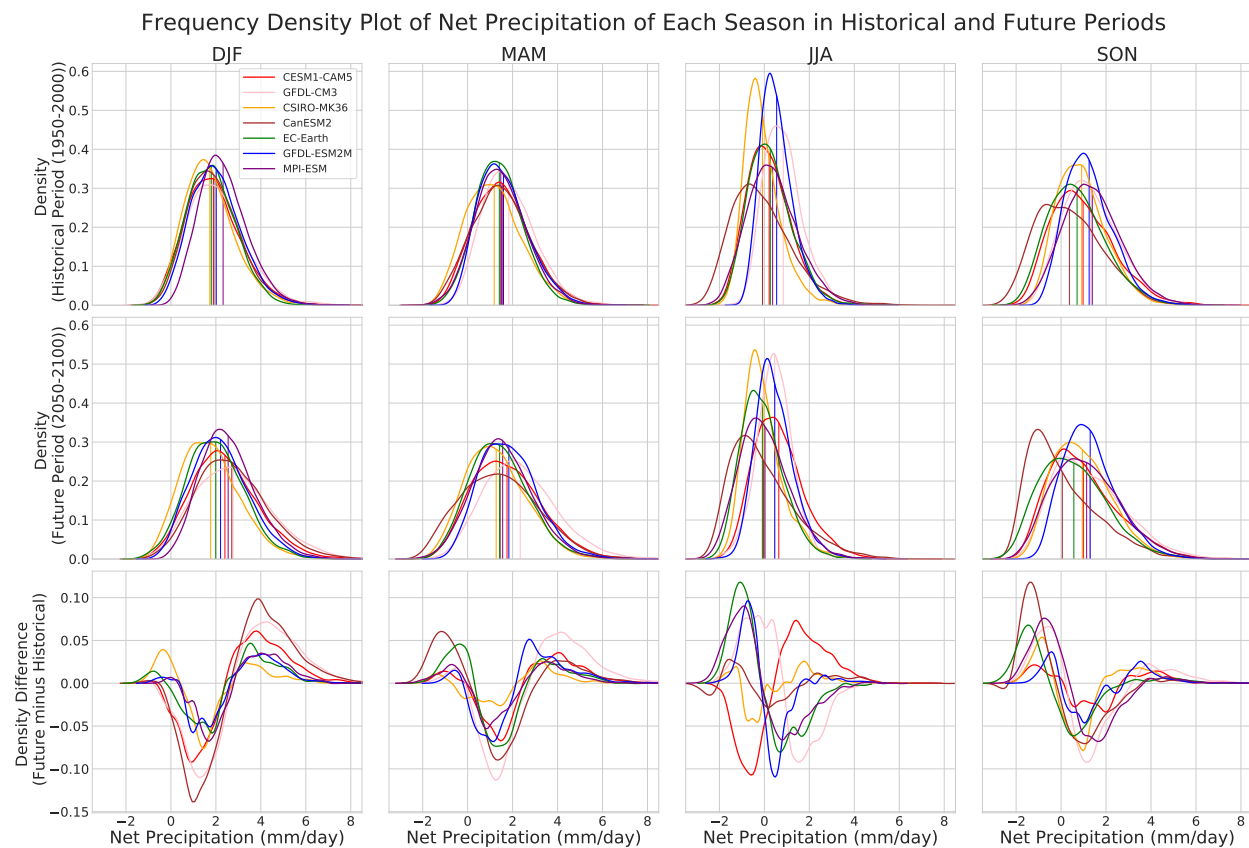


Figure 3.8: Frequency plots of net precipitation in each season for (top) the historical period and (middle) the future period, along with (bottom) their difference.

Multi-ensemble Annually Average Monthly Mean Plant Transpiration Component Ratio and Leaf Area Index

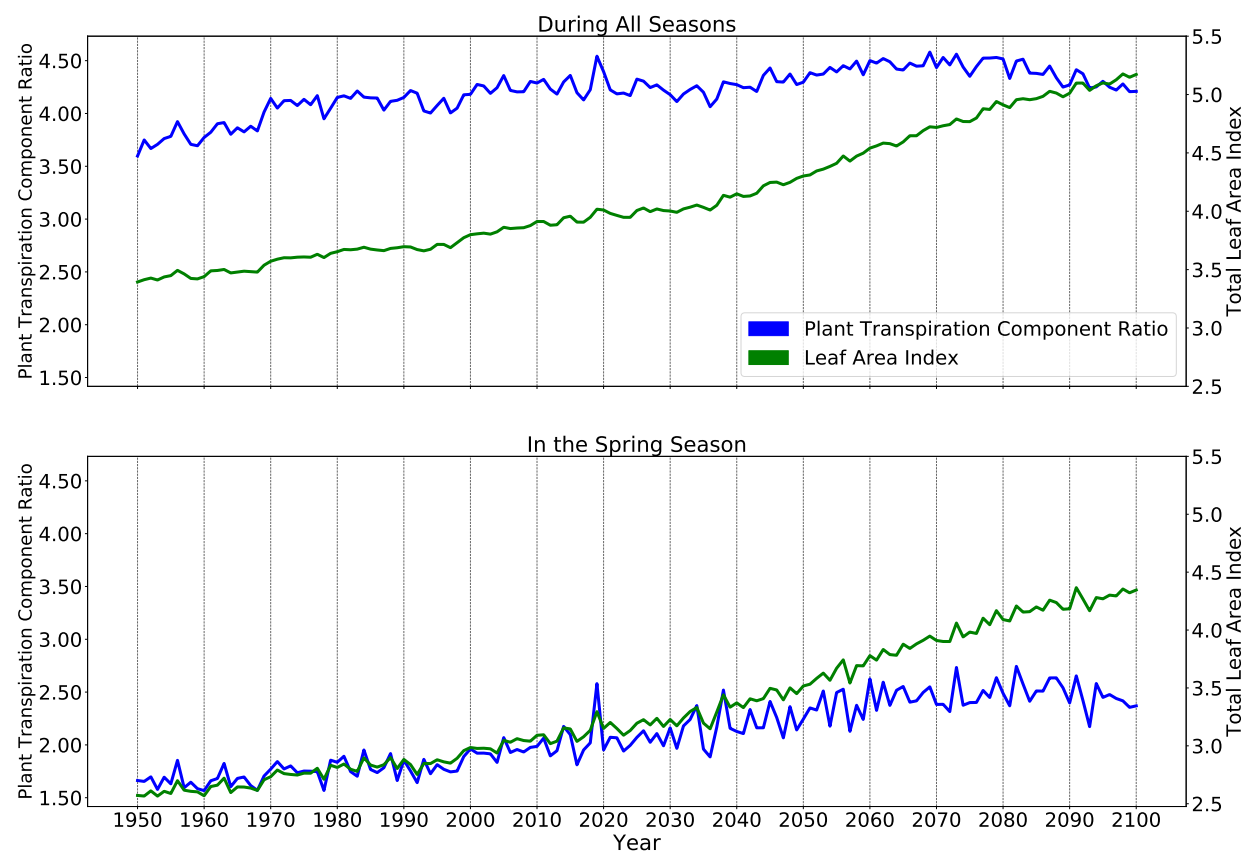


Figure 3.10: Ensemble mean annually averaged monthly Plant Evapotranspiration Component Ratio and Leaf Area Index from CESM1-CAM5 during all seasons and the spring season.

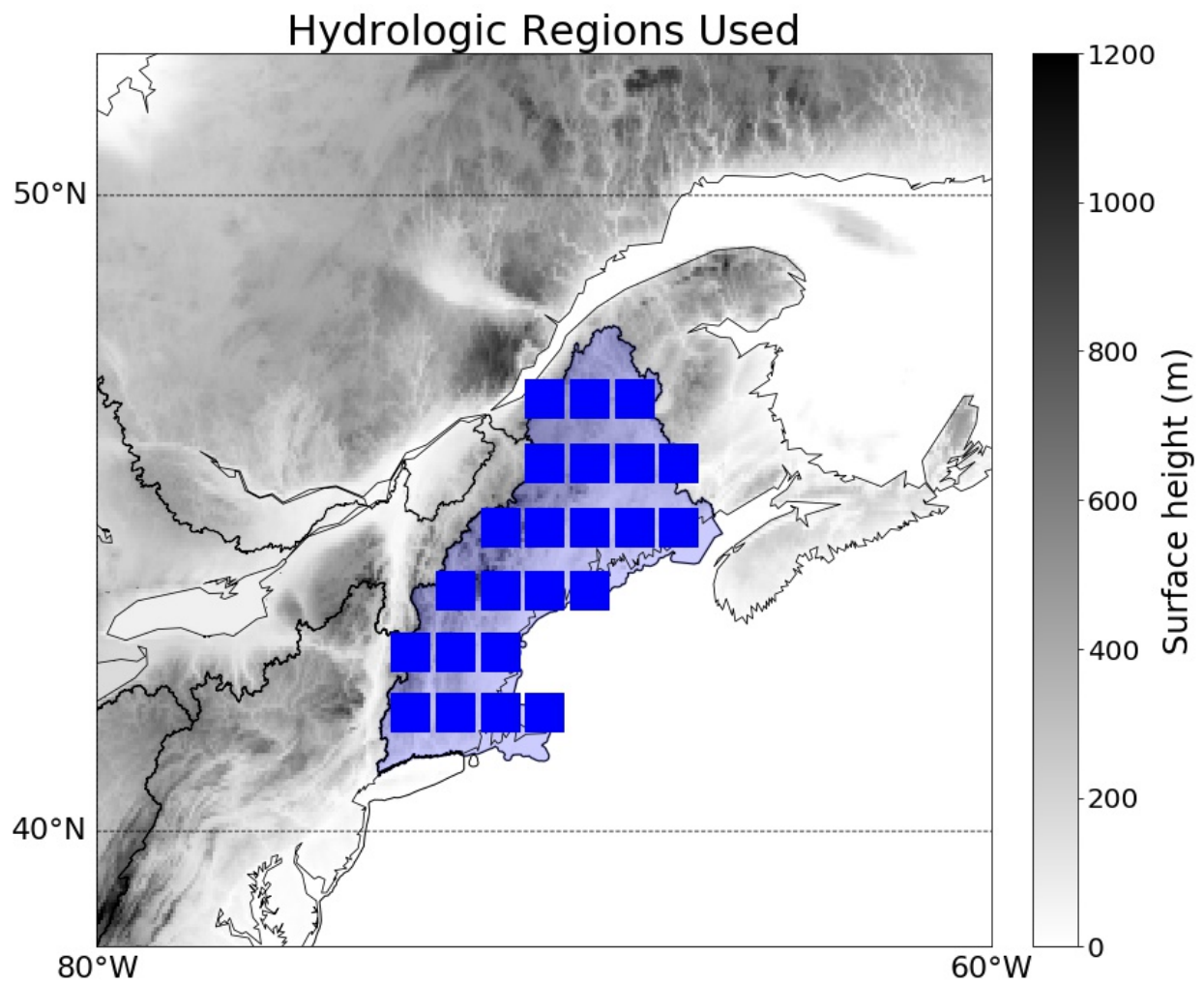


Figure 3.11: The study region (New England Region) and grid cells within it.

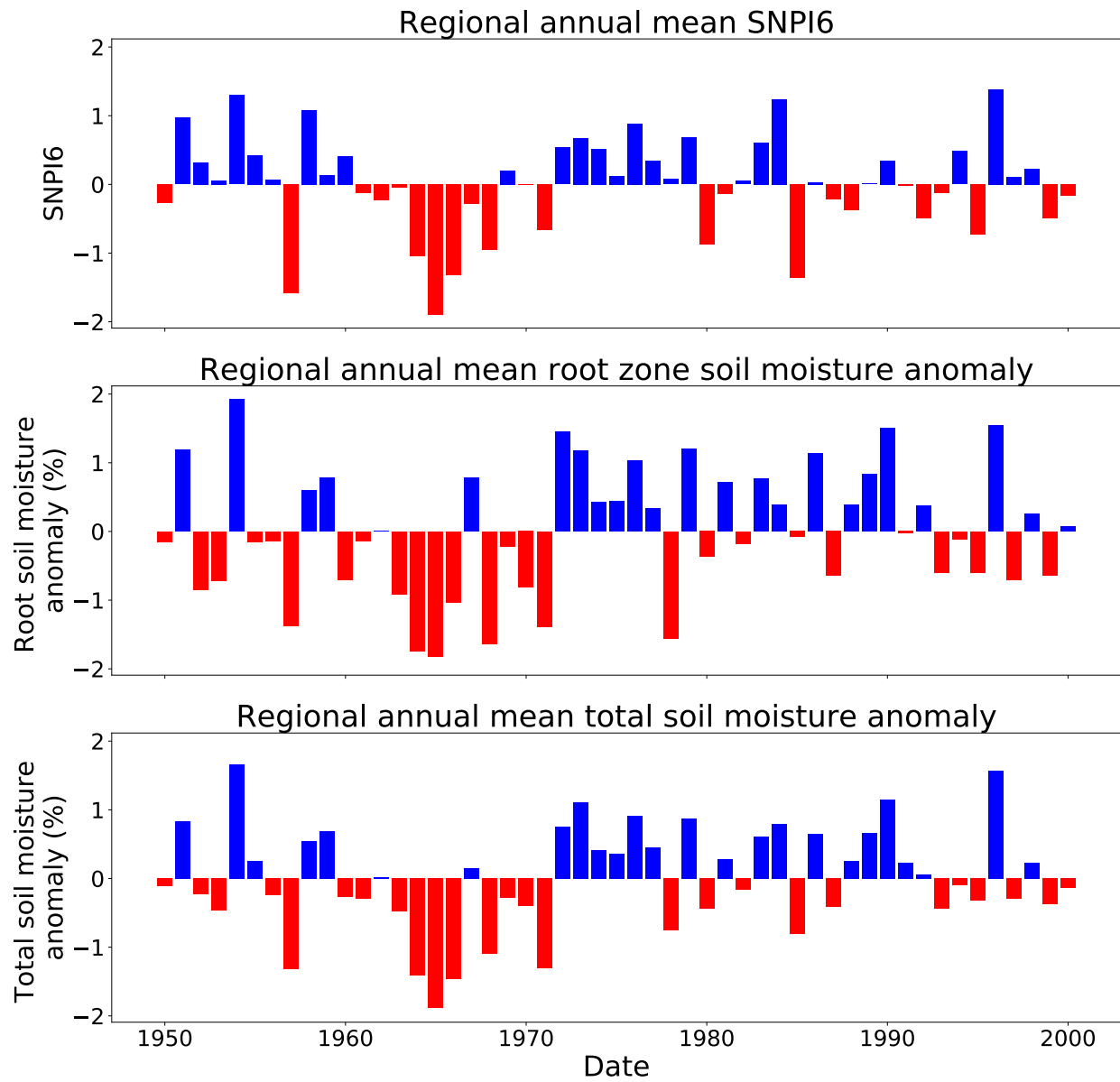


Figure 3.12: The annual average SNPI6, root zone and total soil moisture in ERA5 from 1950-2000.

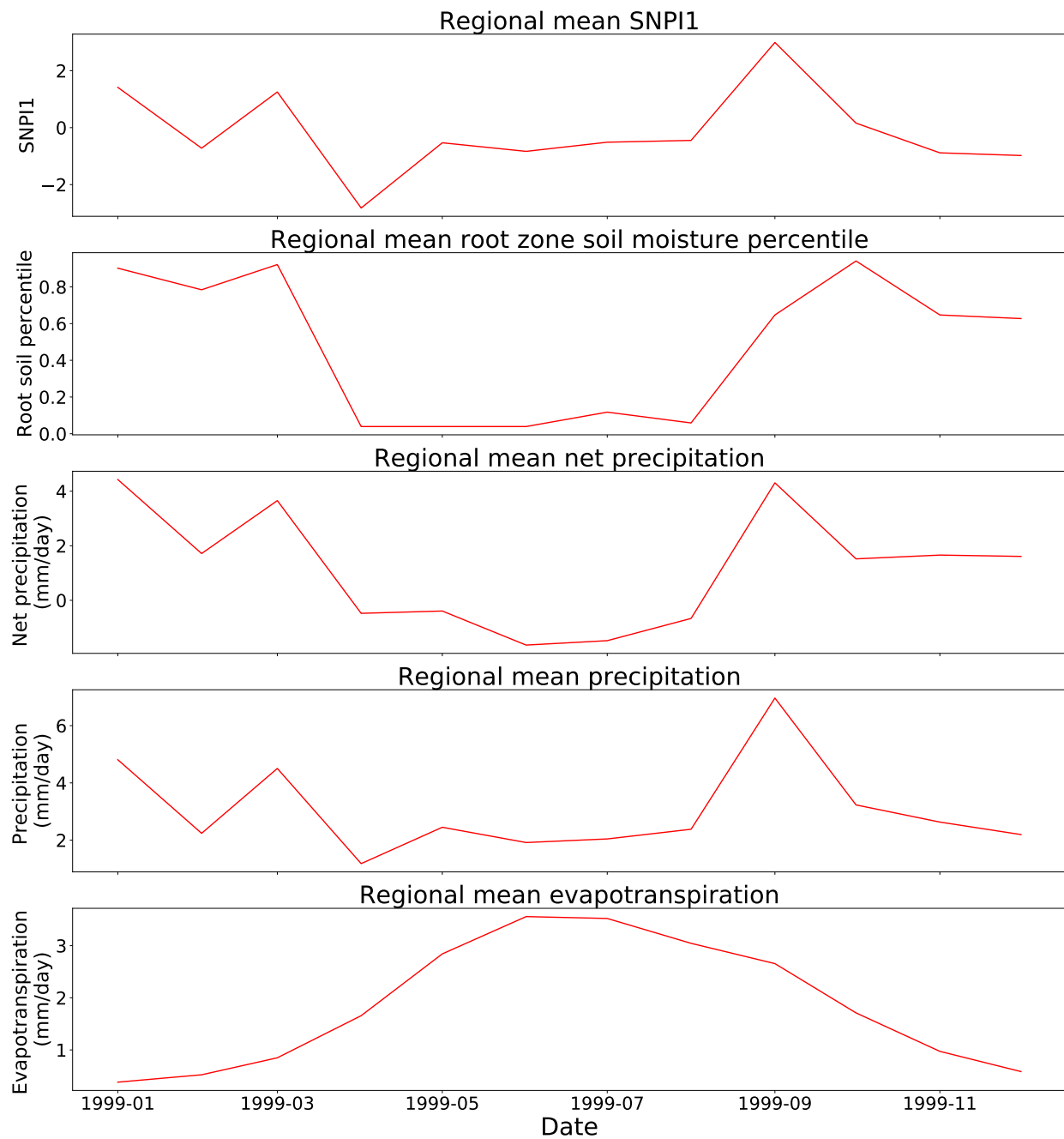


Figure 3.13: Evolving drought conditions in 1999 from ERA5.

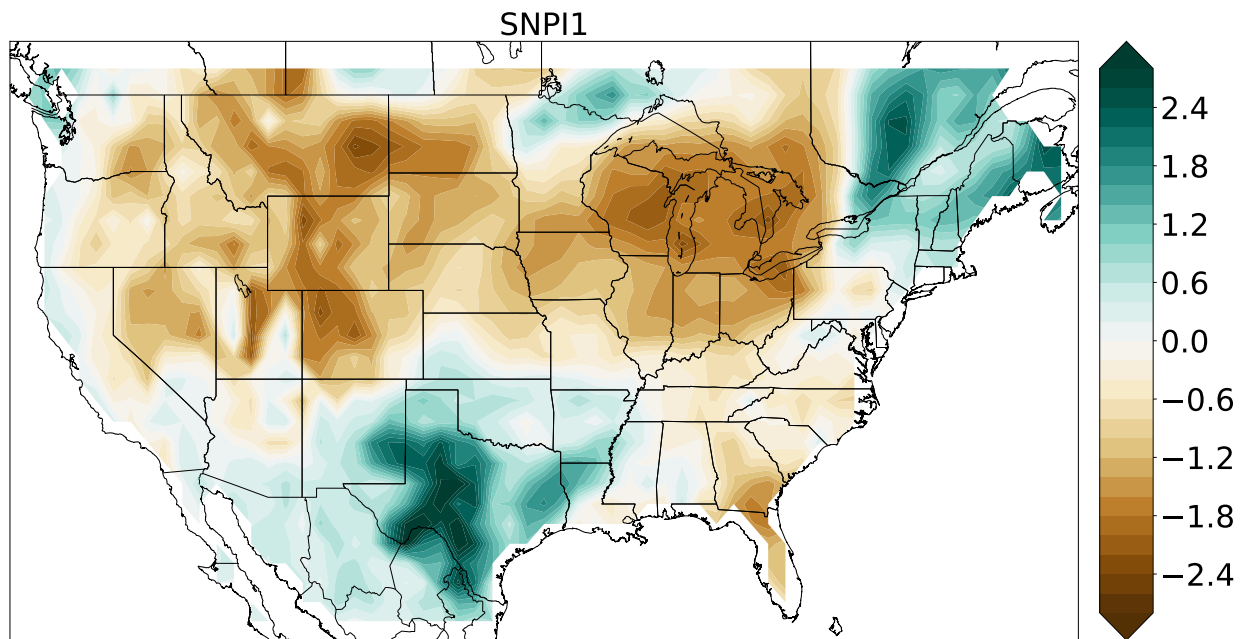


Figure 3.14: SNPI1 in 1999 March from ERA5.

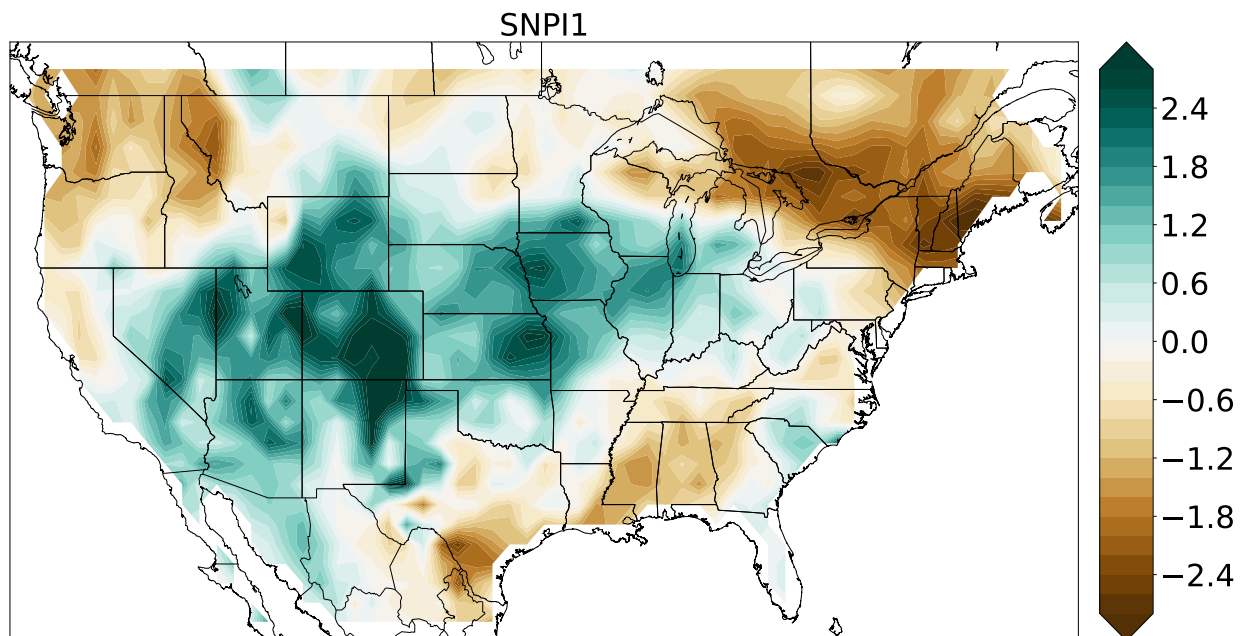


Figure 3.15: SNPI1 in 1999 April from ERA5.

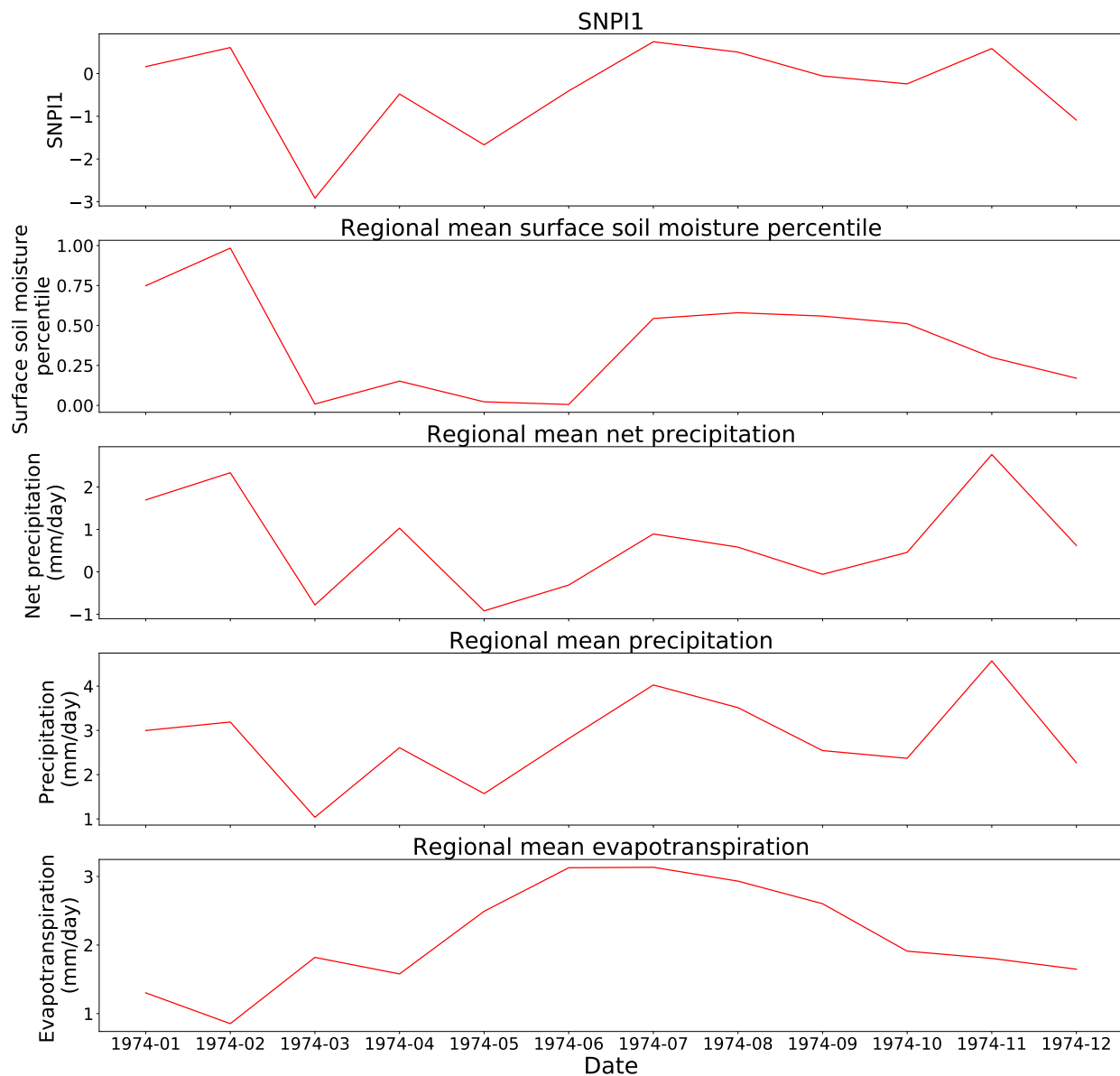


Figure 3.16: Relevant fields during an evolving drought from CESM1 r20i1p1 in 1974 (which experiences the largest surface soil moisture drop).

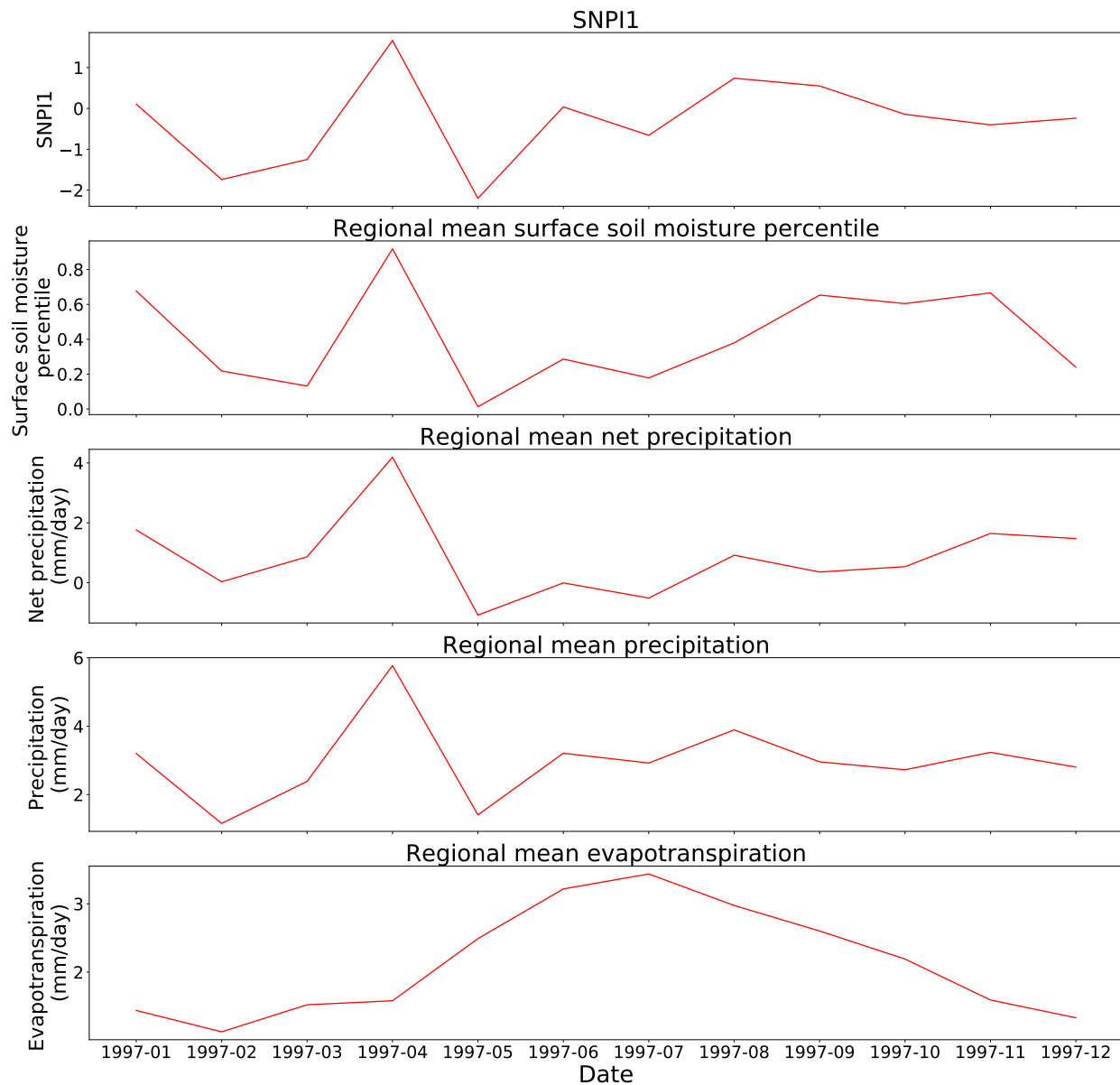


Figure 3.17: Relevant fields during an evolving drought from CESM1 r19i1p1 in 1997 (which experiences the second largest surface soil moisture drop).

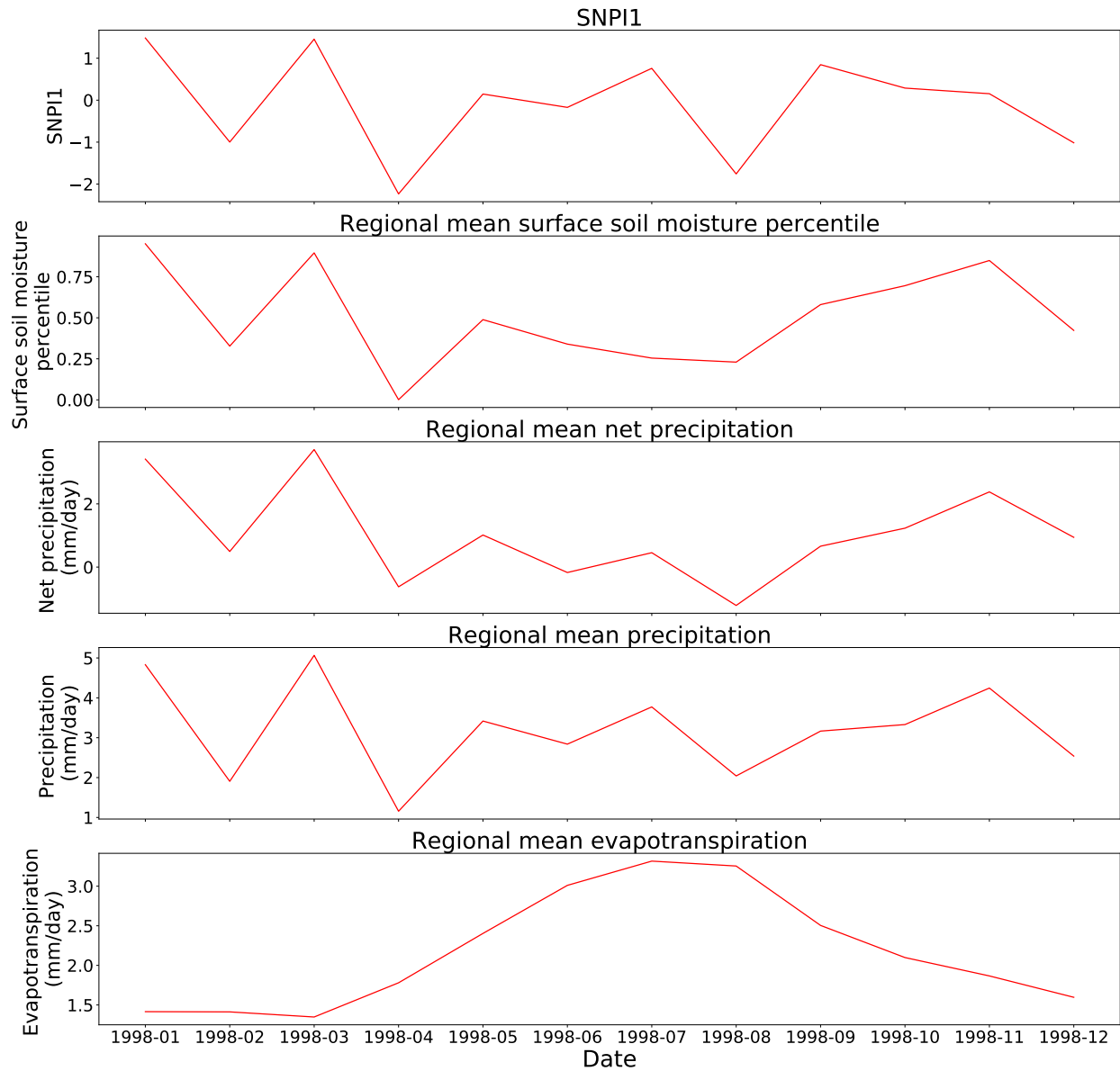


Figure 3.18: Relevant fields during an evolving drought from CESM1 r29i1p1 in 1998 (which experiences the third largest surface soil moisture drop).

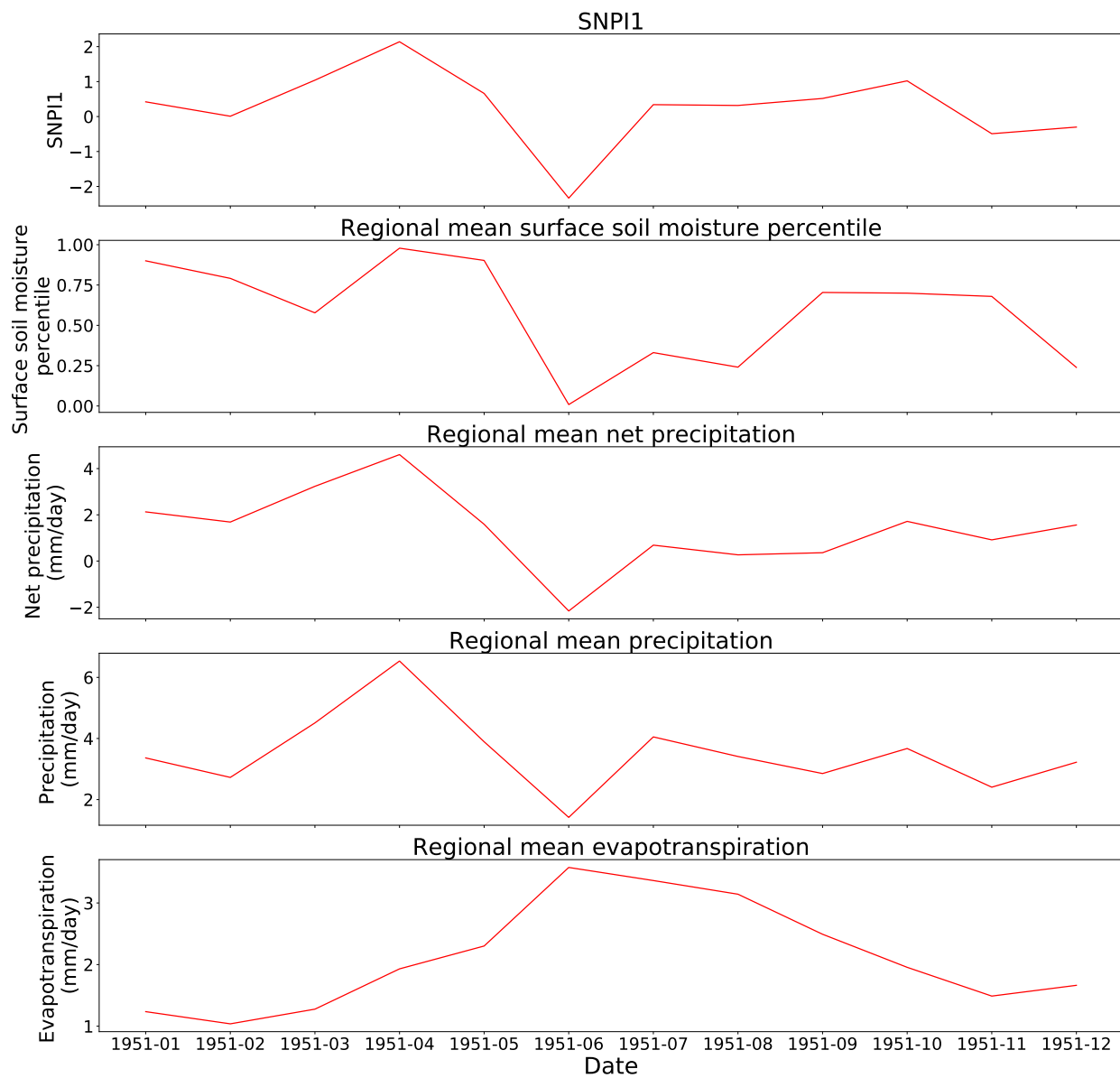


Figure 3.19: Relevant fields during an evolving drought from CESM1 r34i1p1 in 1951 (which experiences the fourth largest surface soil moisture drop).

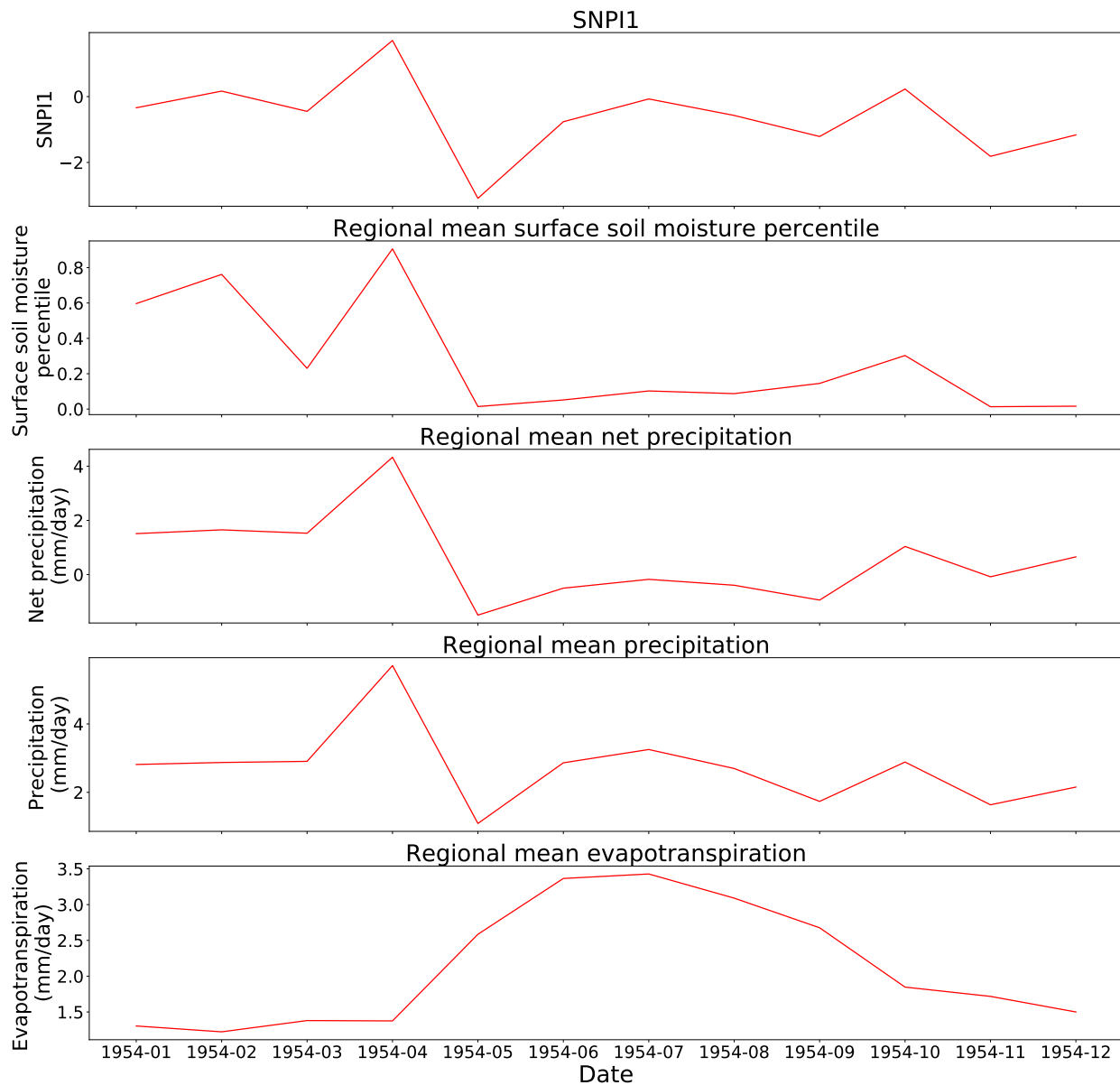


Figure 3.20: Relevant fields during an evolving drought from CESM1 r38i1p1 in 1954 (which experiences the fifth largest surface soil moisture drop).

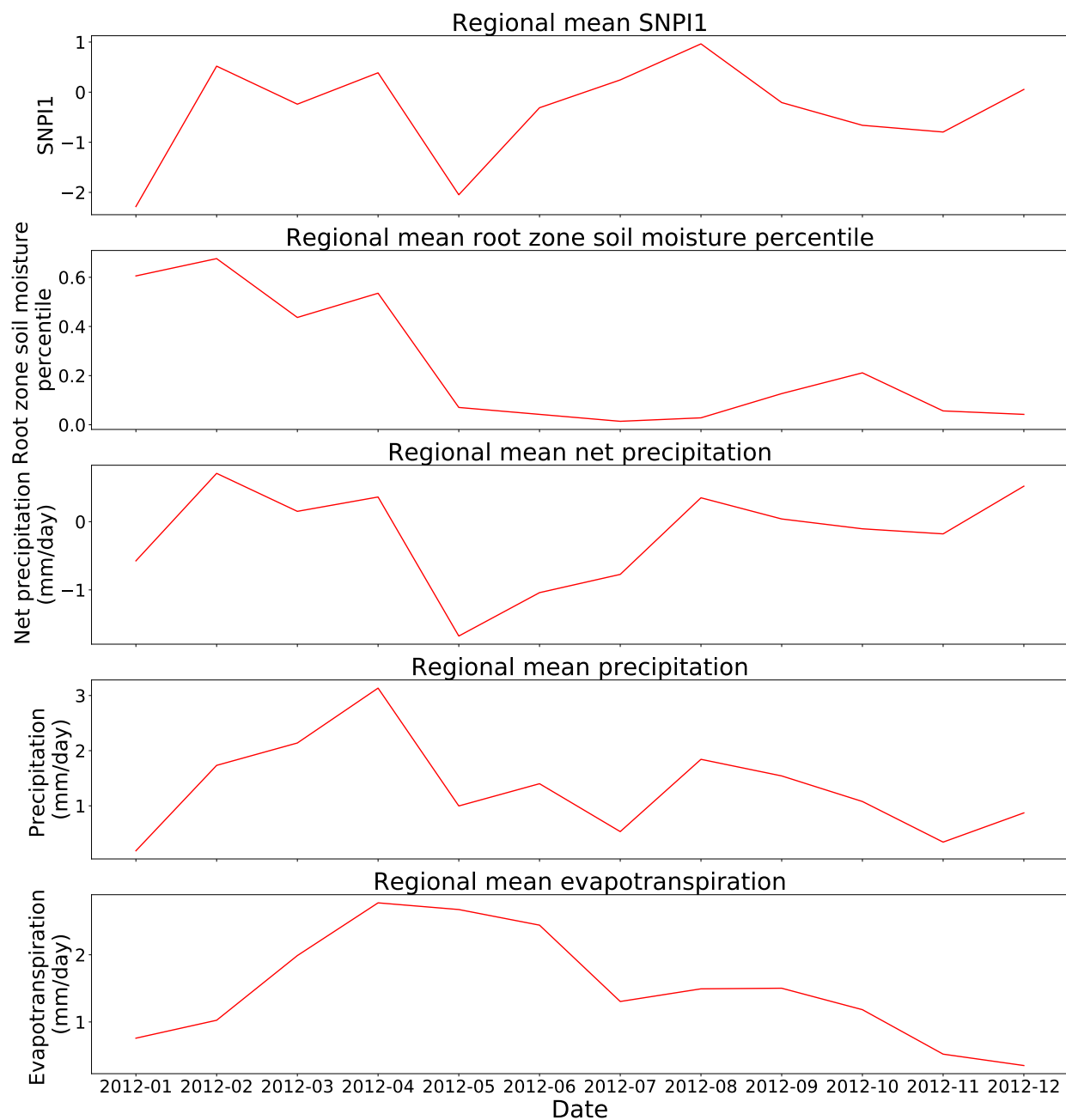


Figure 3.21: Relevant fields during the evolving process of 2012 US Midwest flash drought over Kansas from ERA5 (which is the most infamous flash drought over CONUS).

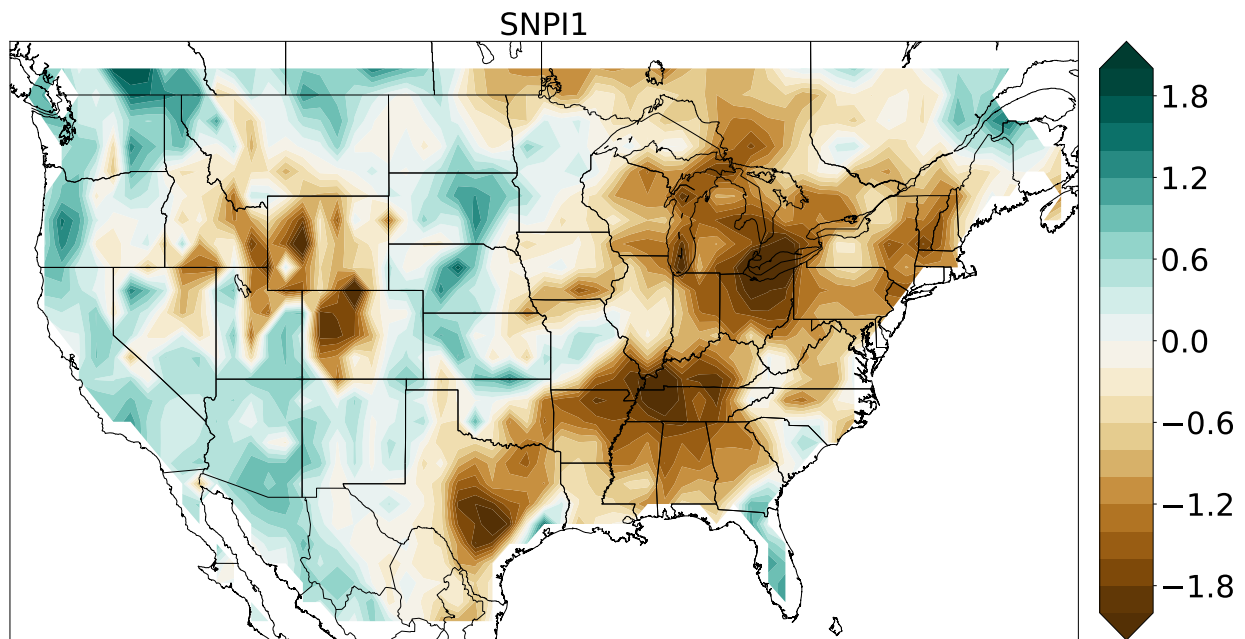


Figure 3.22: SNPI1 in 2012 April from ERA5.

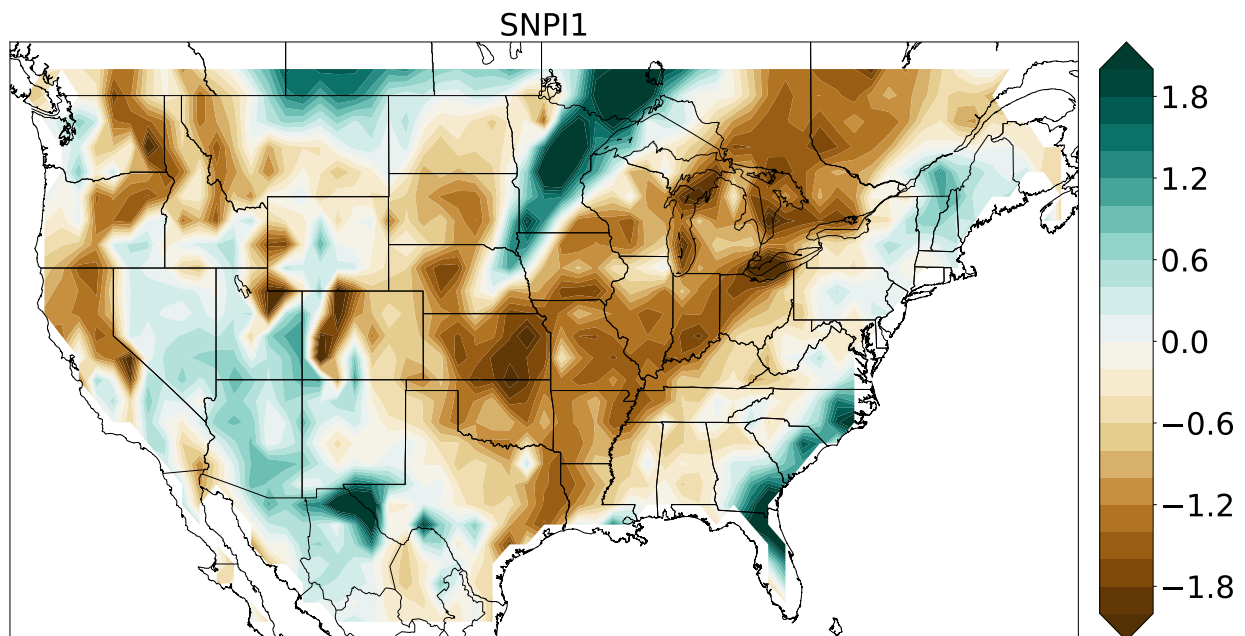


Figure 3.23: SNPI1 in 2012 May from ERA5.

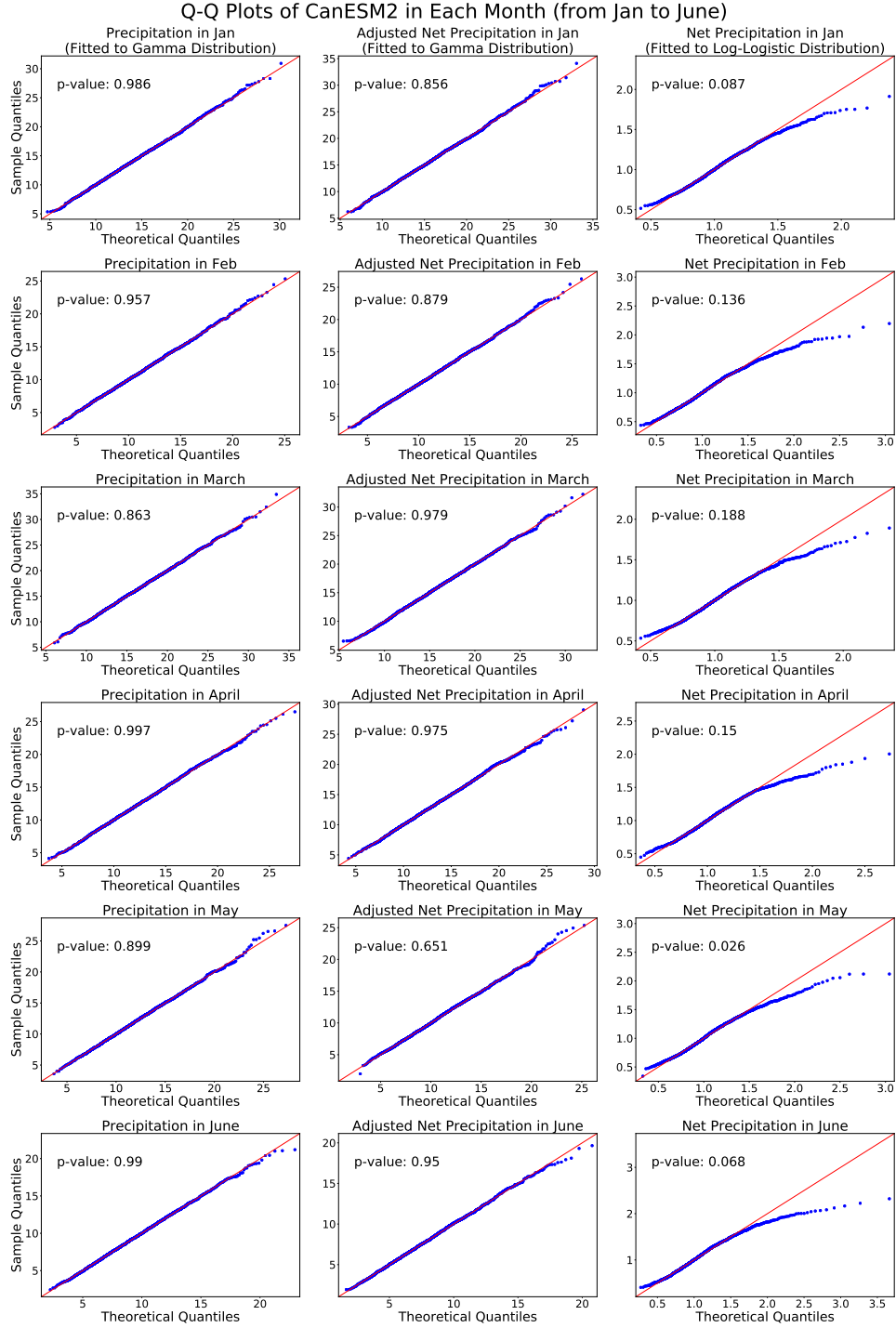


Figure 3.24: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of CanESM2 (January to June).

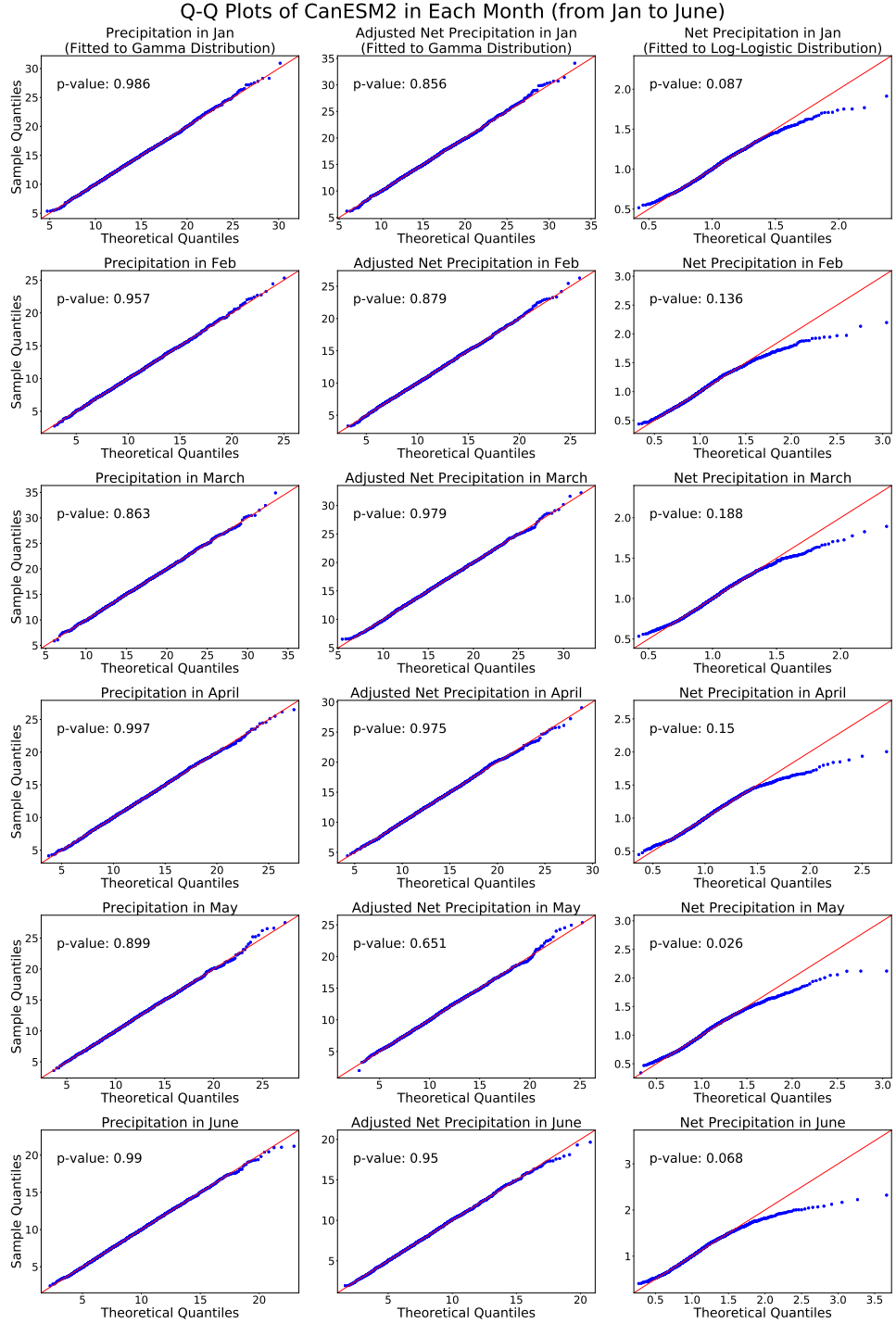


Figure 3.25: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of CanESM2 (July to December).

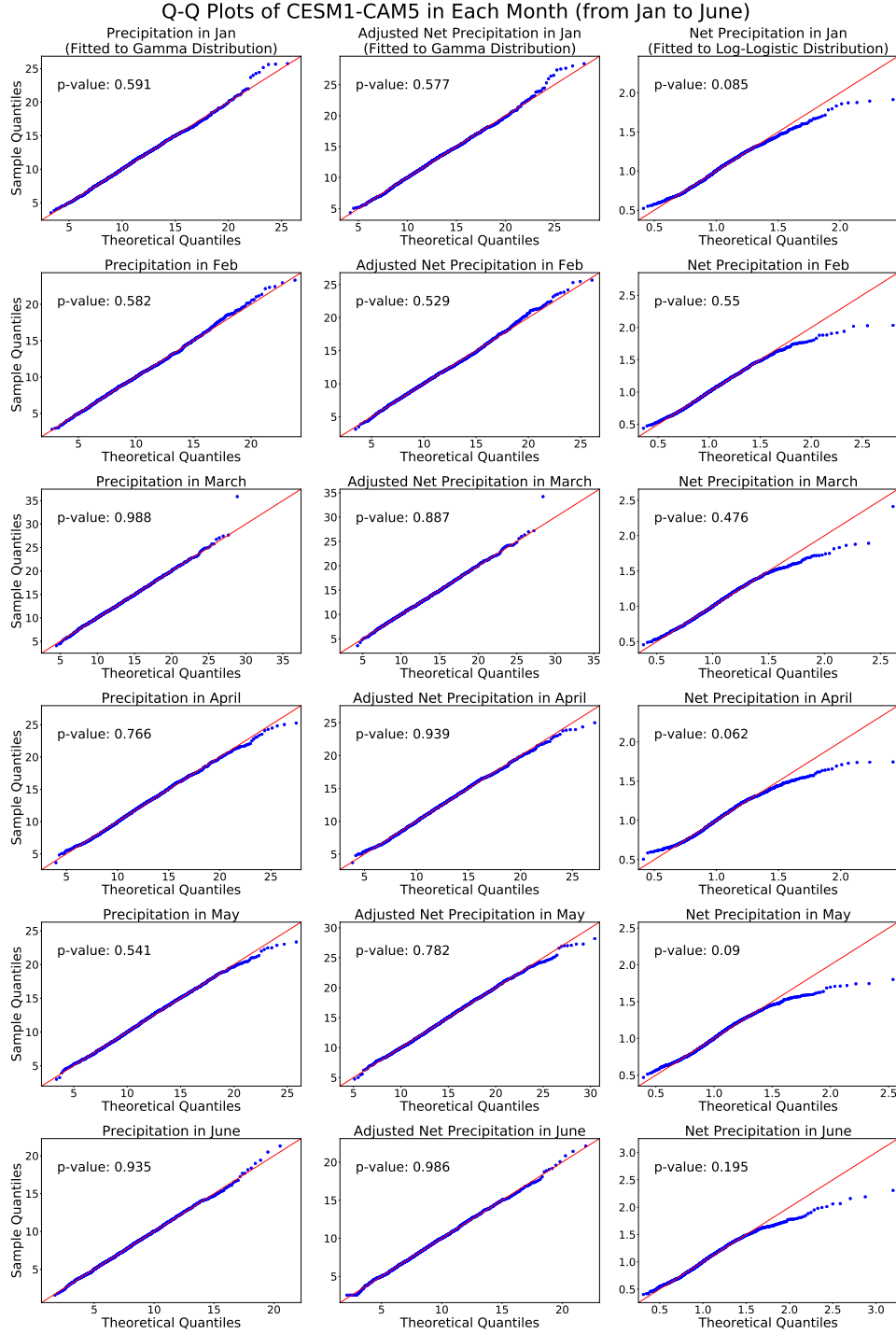


Figure 3.26: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of CESM1-CAM5 (January to June).

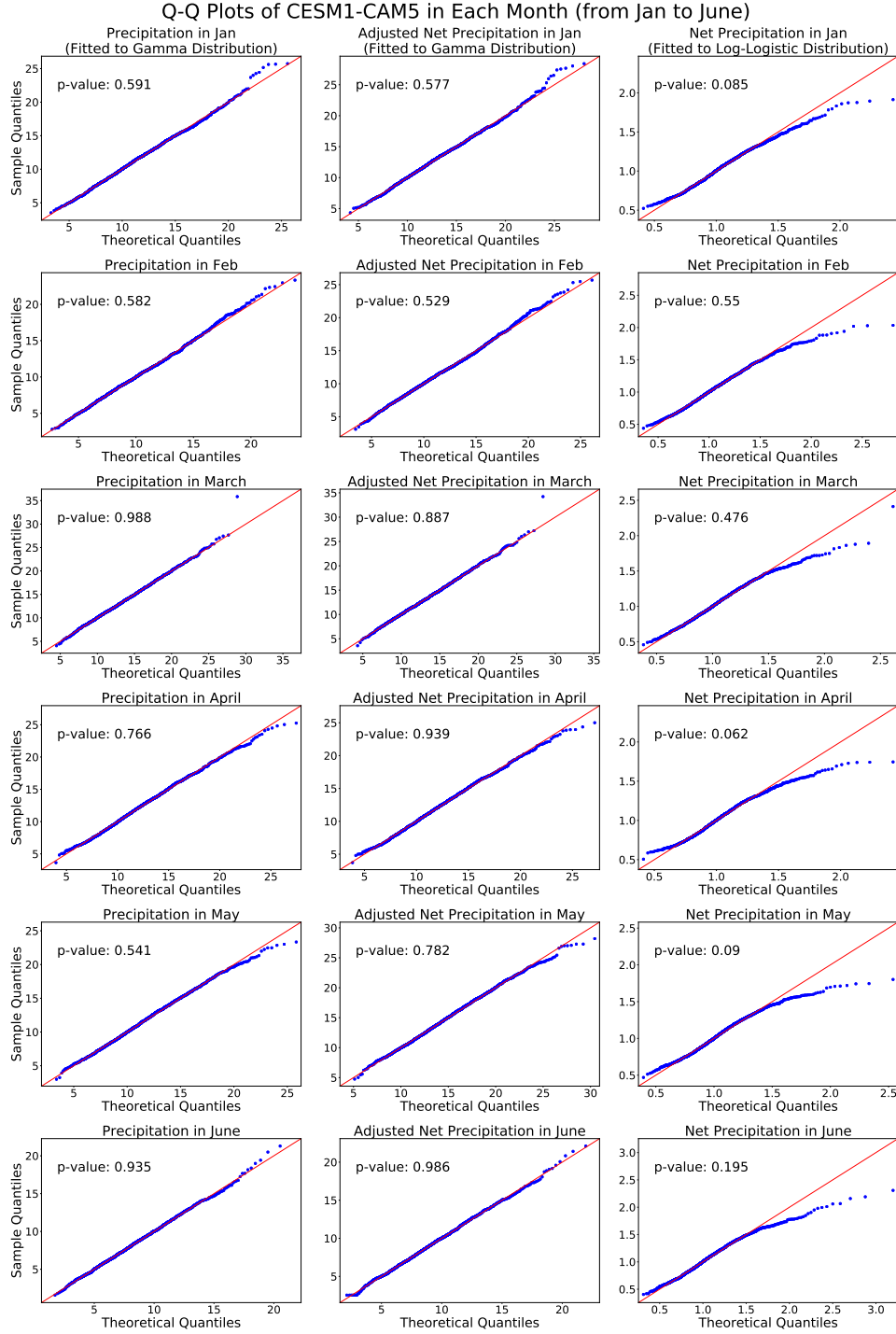


Figure 3.27: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of CESM1-CAM5 (July to December).

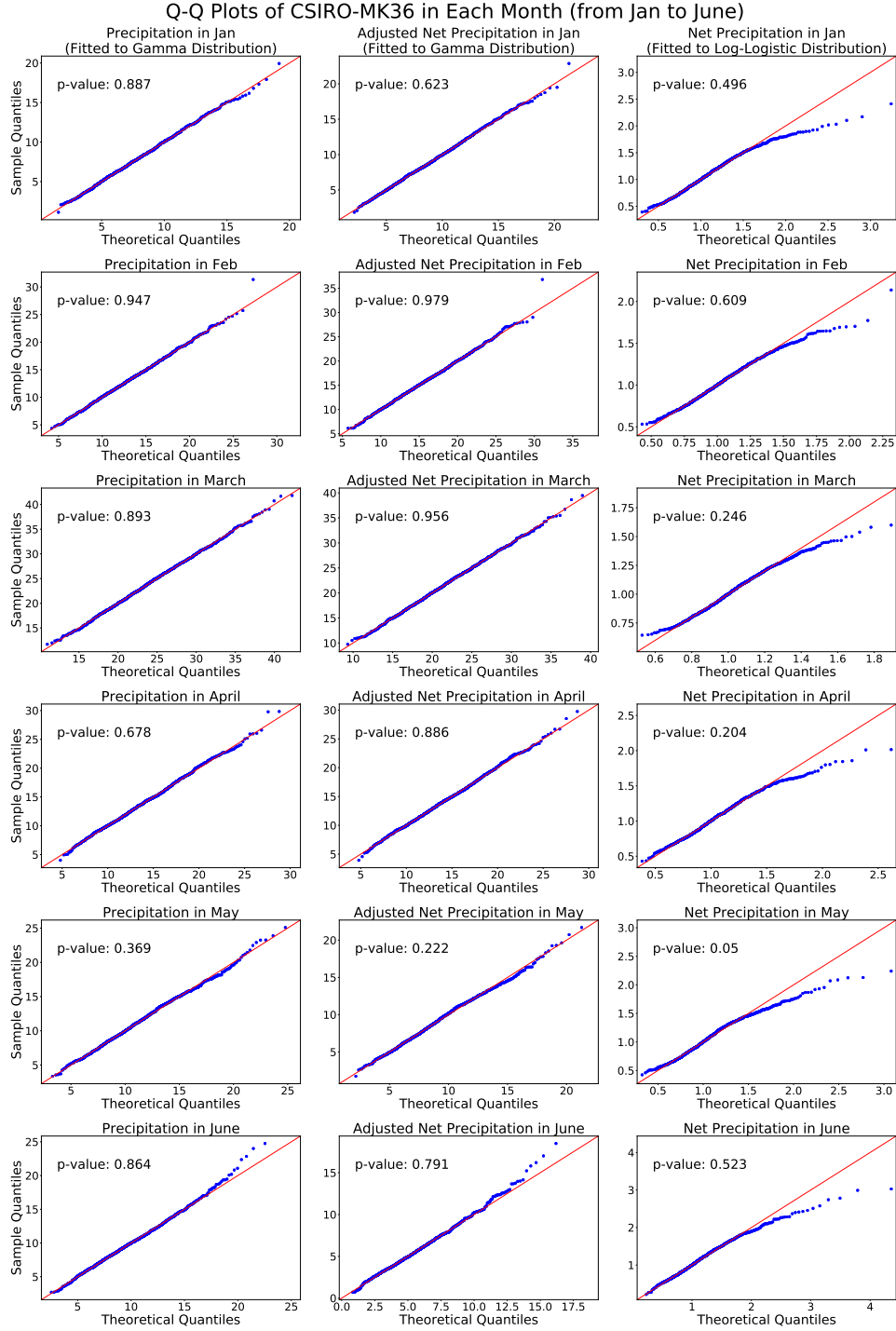


Figure 3.28: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of CSIRO-MK36 (January to June).

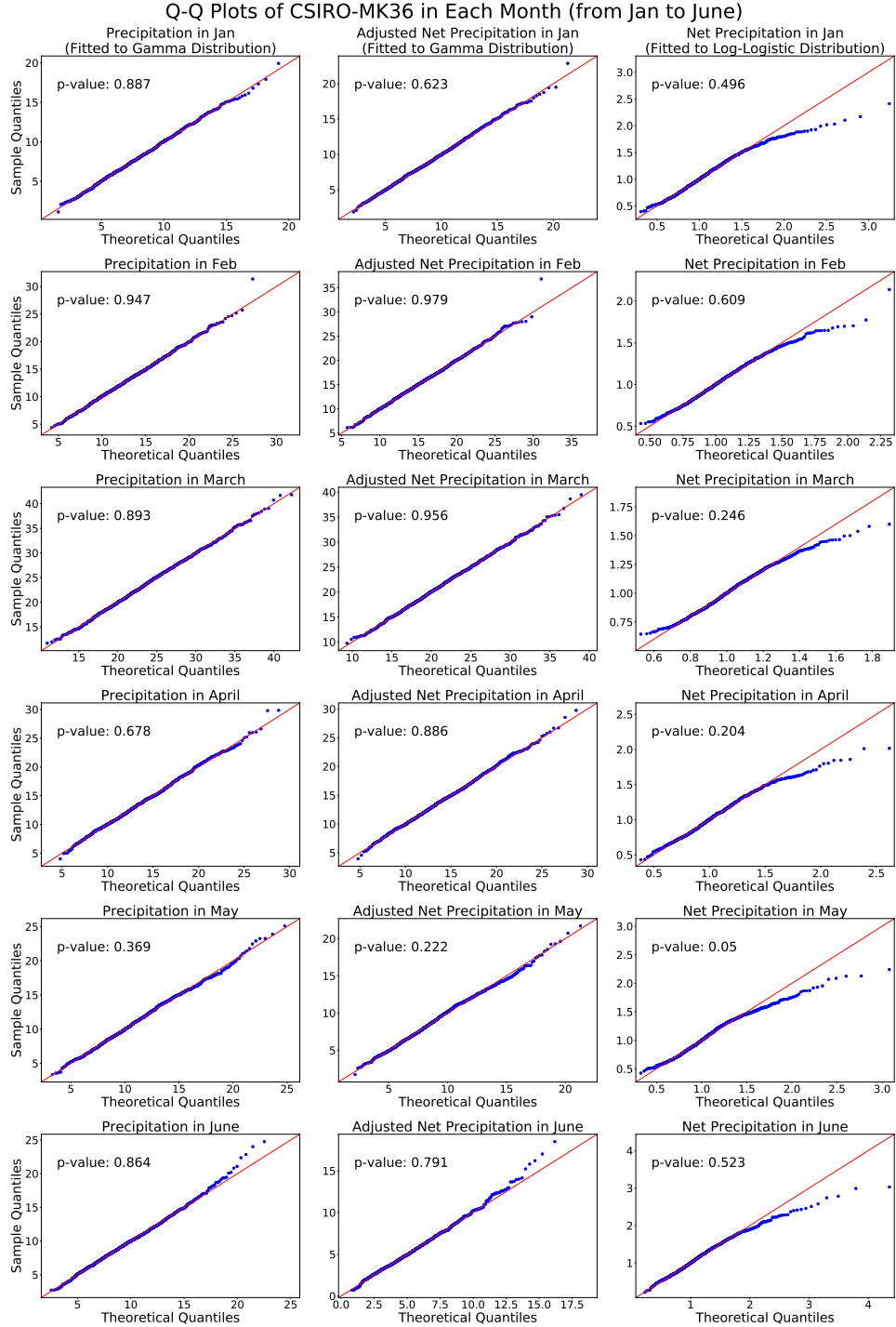


Figure 3.29: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of CSIRO-MK36 (July to December).

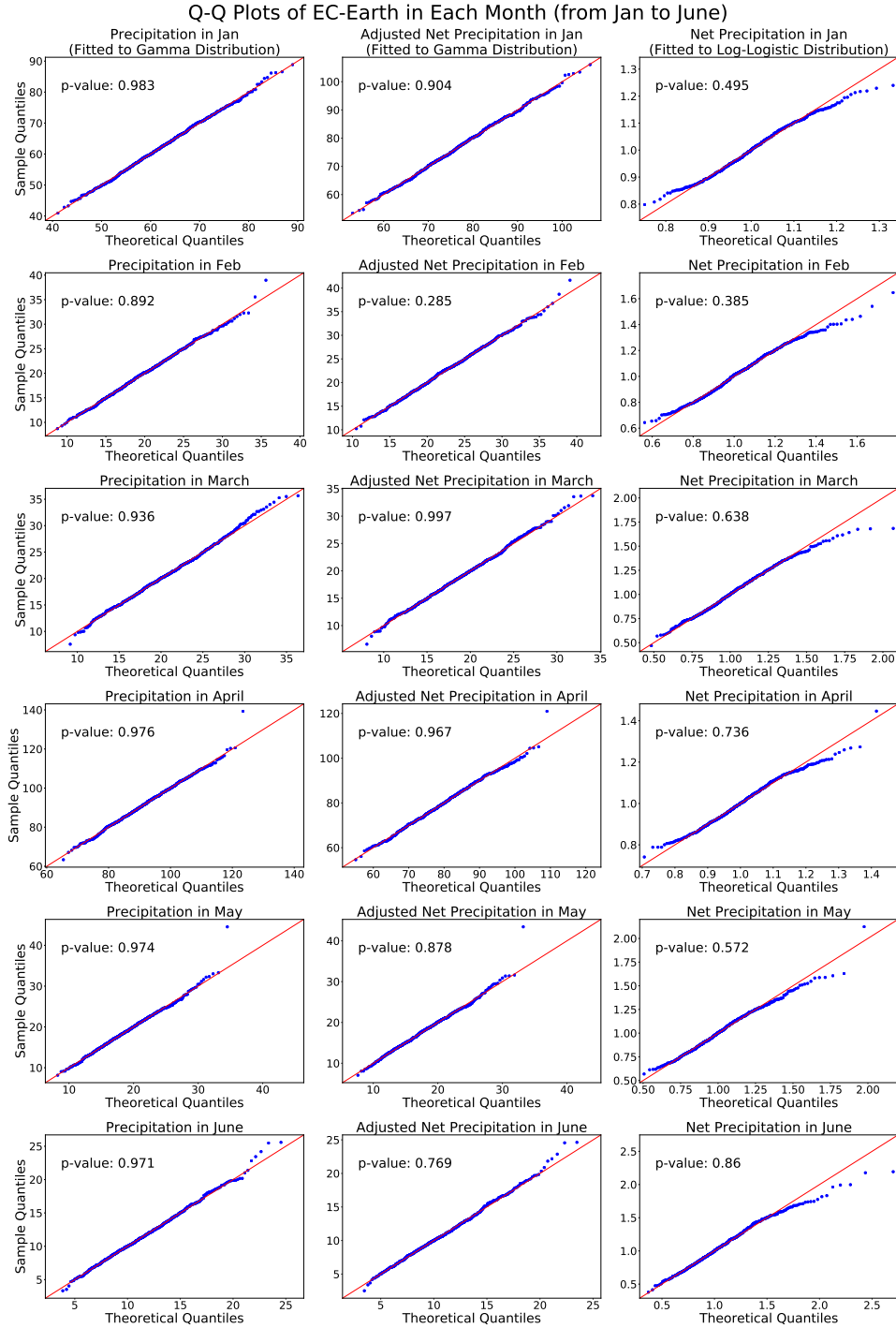


Figure 3.30: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of EC-Earth (January to June).

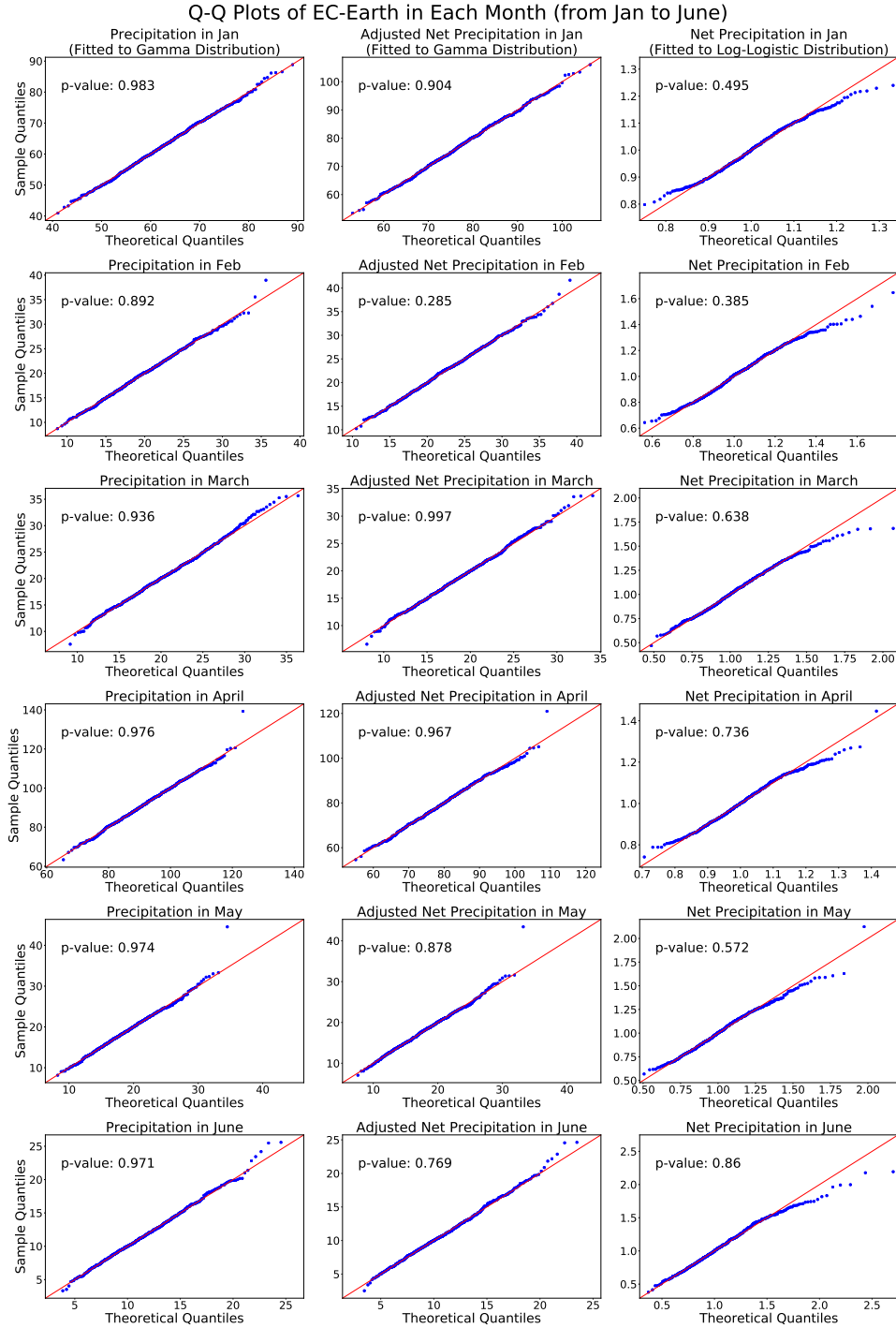


Figure 3.31: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of EC-Earth (July to December).

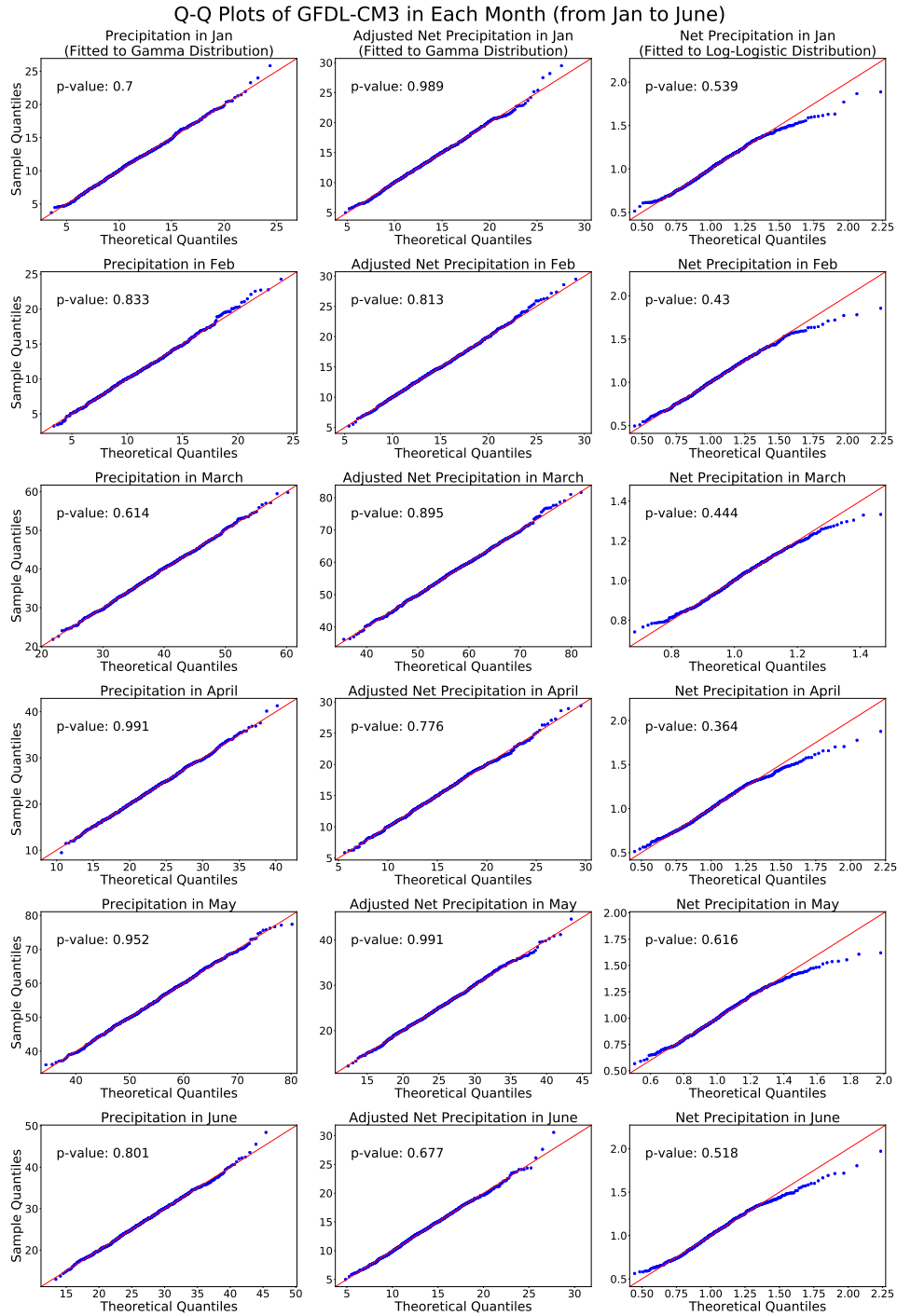


Figure 3.32: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of GFDL-CM3 (January to June).

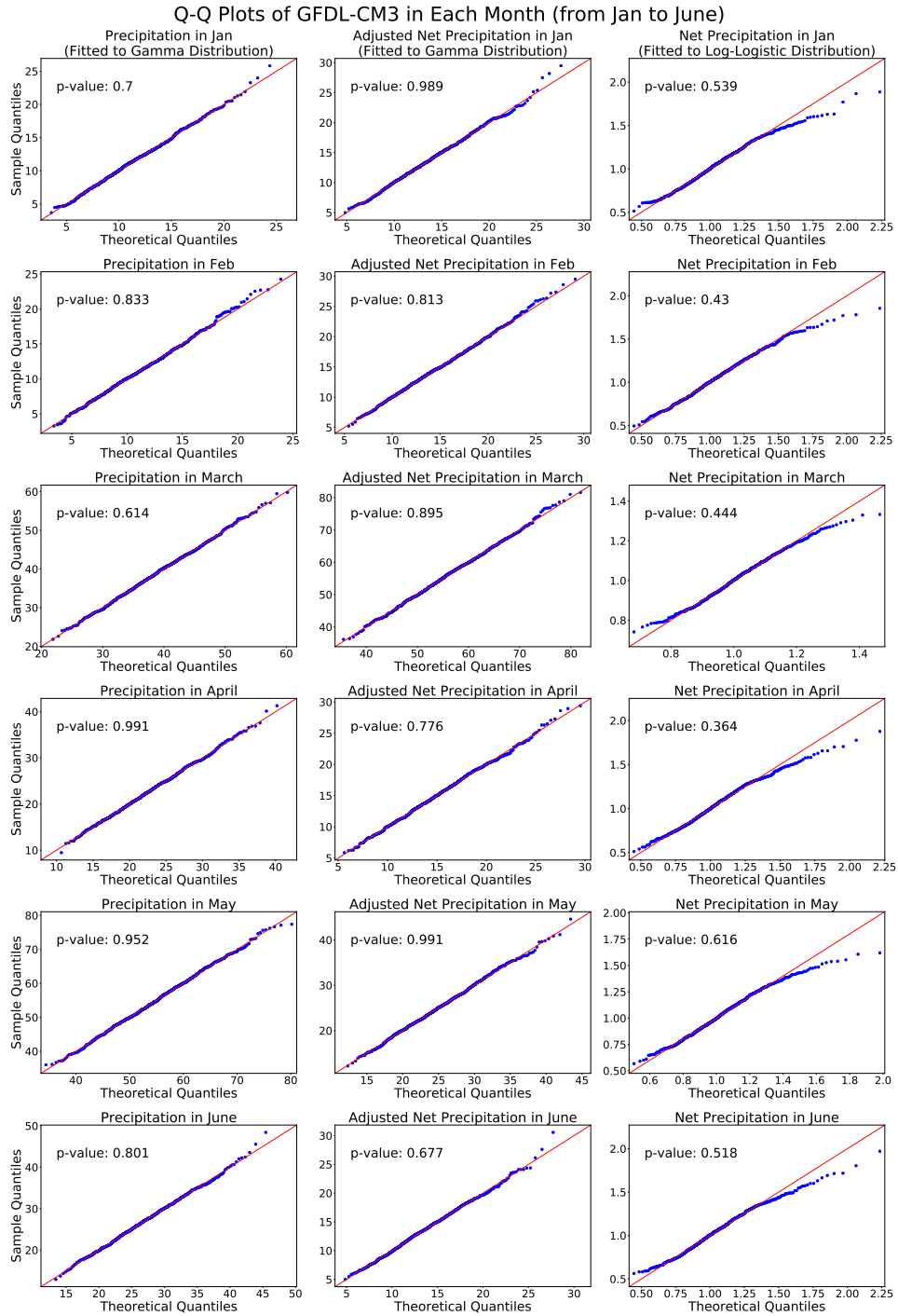


Figure 3.33: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of GFDL-CM3 (July to December).

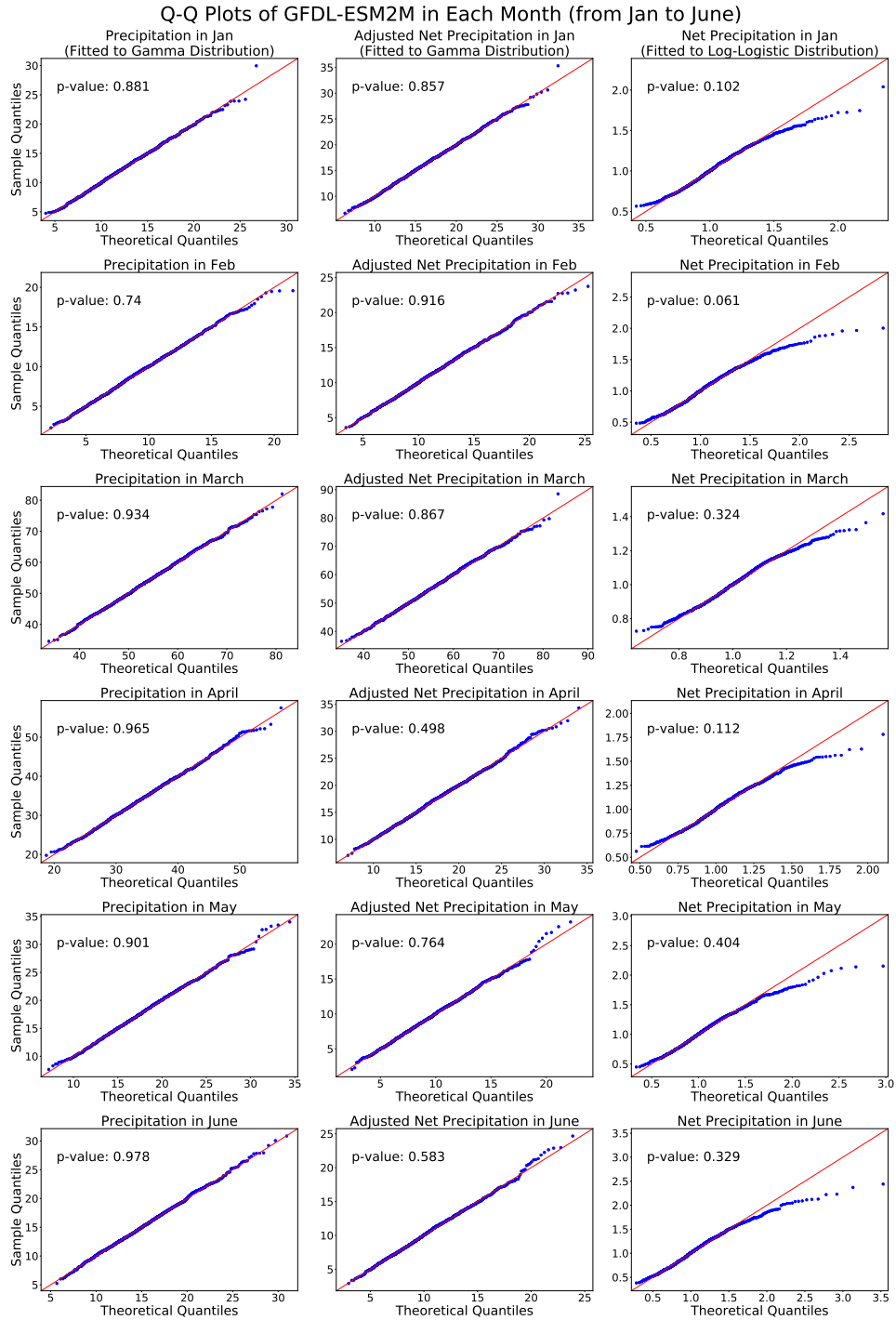


Figure 3.34: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of GFDL-ESM2M (January to June).

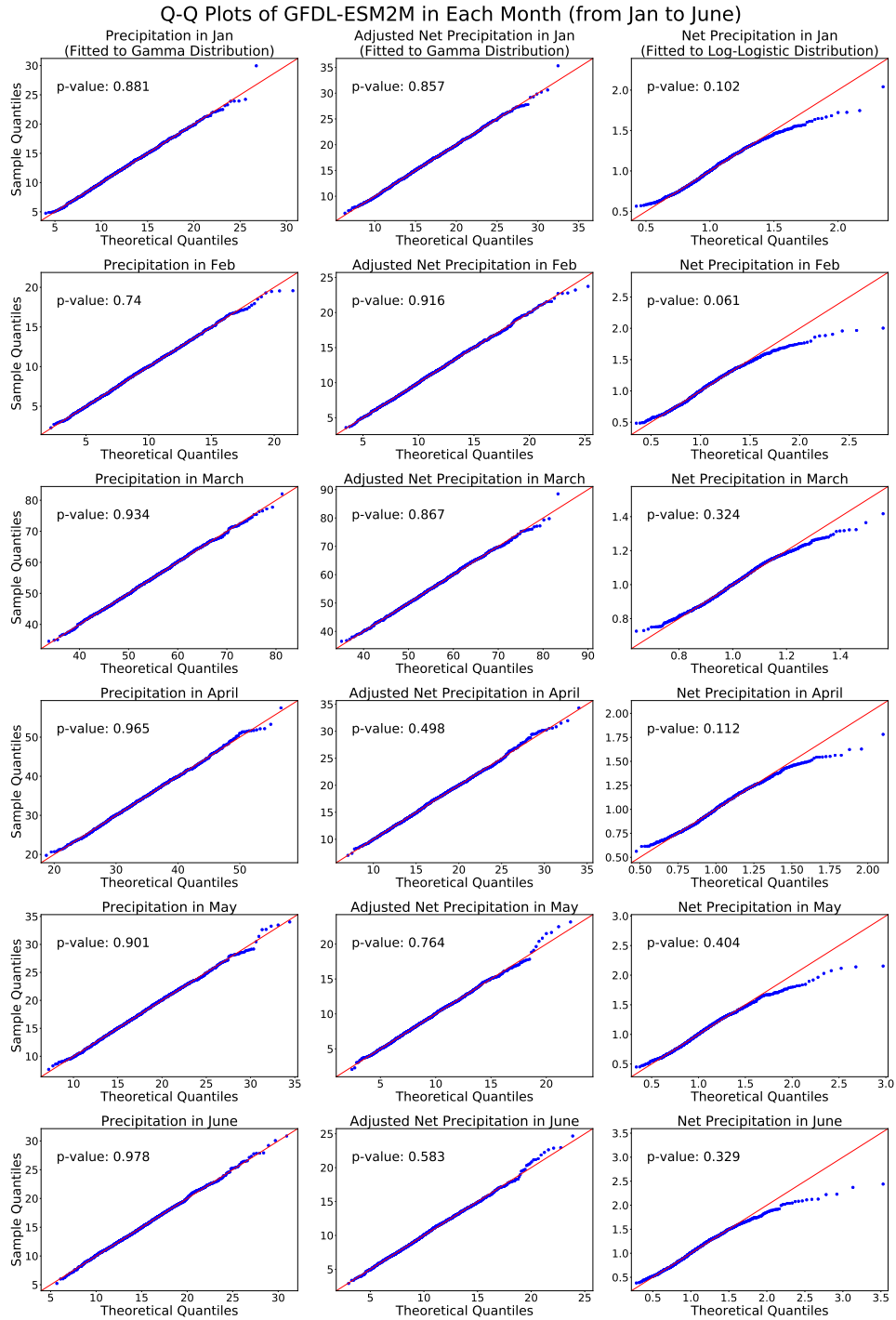


Figure 3.35: Q-Q plots of precipitation, adjusted net precipitation and net precipitation fitted to corresponding distributions, and their p-values of K-S Test within all ensembles of GFDL-ESM2M (July to December).

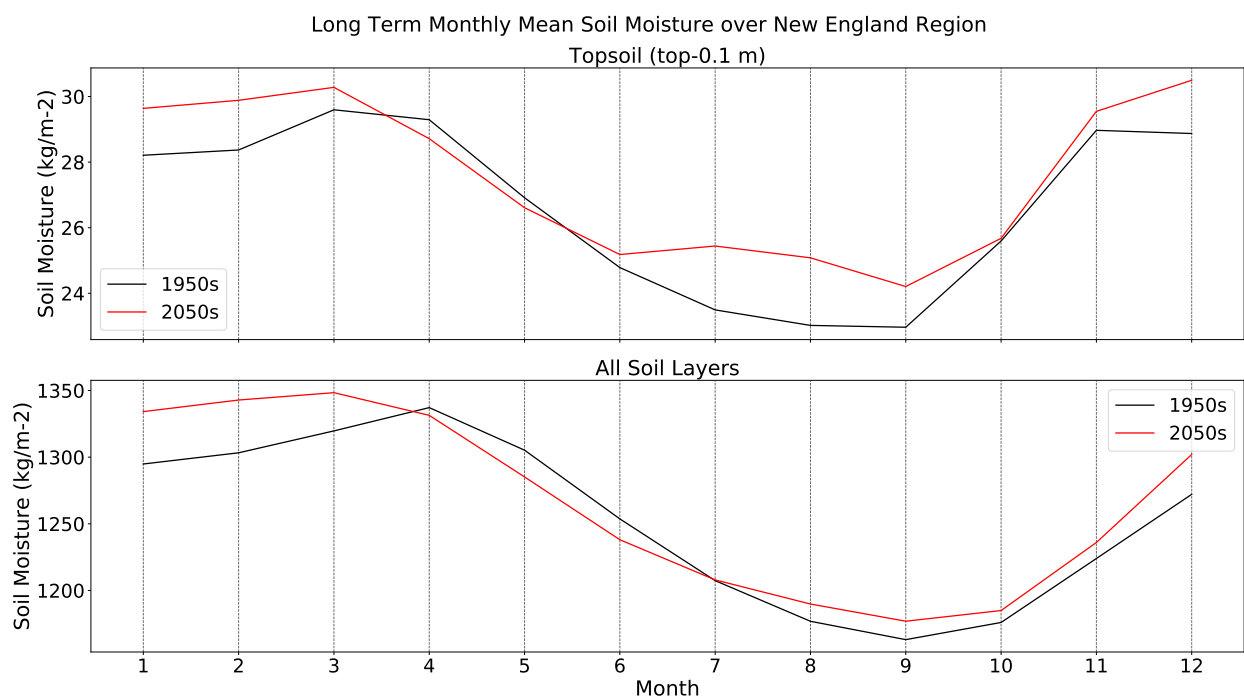


Figure 3.36: CESM1-CAM5's ensemble average long-term monthly mean total soil moisture and topsoil moisture over New England region in the historical (1950-2000) and future (2050-2100) period.

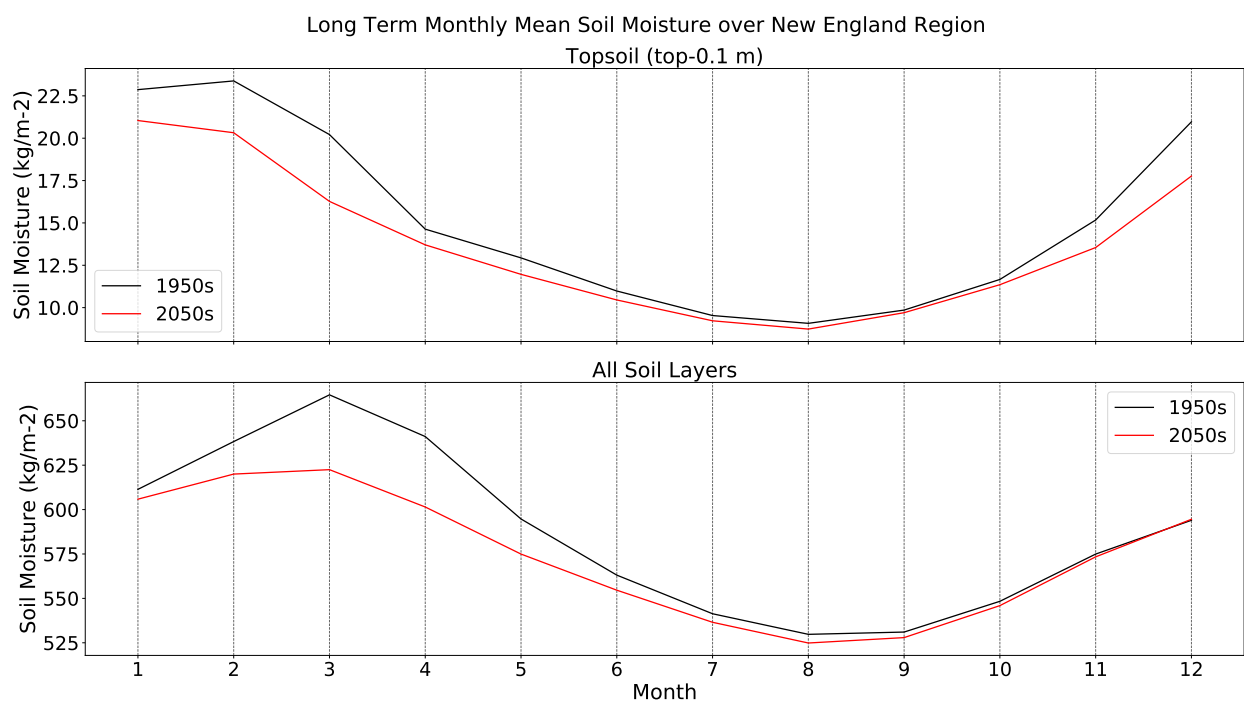


Figure 3.37: CSIRO-MK36's ensemble average long-term monthly mean total soil moisture and topsoil moisture over New England region in the historical (1950-2000) and future (2050-2100) period.

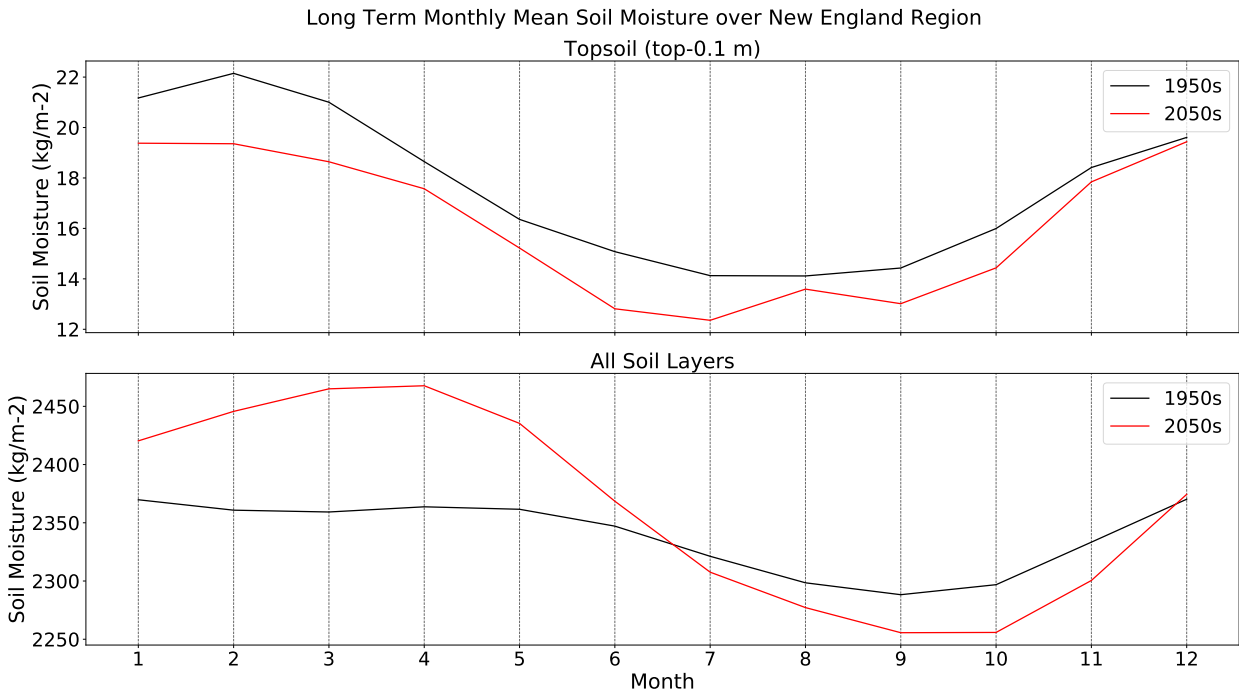


Figure 3.38: GFDL-CM3's ensemble average long-term monthly mean total soil moisture and topsoil moisture over New England region in the historical (1950-2000) and future (2050-2100) period.

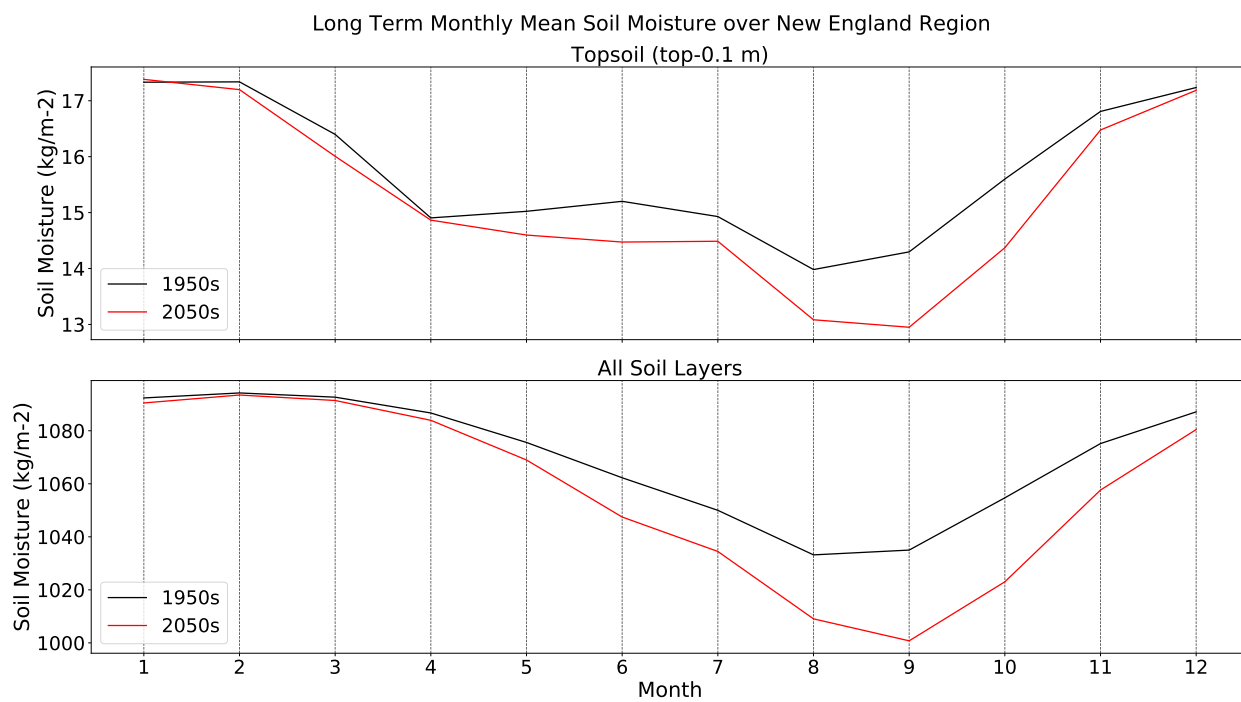


Figure 3.39: MPI-ESM's ensemble average long-term monthly mean total soil moisture and topsoil moisture over New England region in the historical (1950-2000) and future (2050-2100) period.

Chapter 4 A retrospective and prospective examination of the 1960s U.S. Northeast Drought

4.1 Introduction

The Northeastern United States (NEUS) is the most economically developed and populated region in the US, accounting for about 20% of US GDP and population but only 5% of its land area (Hobbs, 2008; US Bureau of Economic Analysis, 2016). Here, extreme weather events – primarily floods, droughts and snowstorms – result in disproportionate socioeconomic damage. One of the most well known examples of extreme weather in this region was the 1962-66 drought, which had pronounced implications for agriculture and water management practices (Barksdale, 1968; Janes and Brumbach, 1965; Namias, 1966). Although the direct economic damage was not extensive (DeGaetano, 1999), this event has since framed water resource planning in the NEUS. Consequently, a return of the water stresses from this period would have enormous implications. To this end, it’s important to understand how would such an extreme drought’s characters change under future climatological conditions? Notably, the unprecedented 1960s drought was followed by a long wet period that continued through today (Seager et al., 2012). Both historical observations and climate models show continued increase in precipitation over NEUS (Frumhoff et al., 2007); however, this should not imply that droughts here are things of the past. In fact, there is evidence that the risk of potentially even more severe droughts remains (Burns et al., 2007; Frumhoff et al., 2007; Hayhoe et al., 2007). Advances in climate models have made it possible to improve our confidence in these projections, and so it is timely to revisit the nature of drought in this region.

Pseudo-global warming (PGW) is a demonstrably effective method for simulating the effects

of global warming. This method not only reduces large-scale model biases and ensures that dynamical conditions are consistent with a historical analogue, but also allows us to directly estimate differences between current and future climatological conditions (Kimura et al., 2007; Ullrich et al., 2018). Using PGW, global climate model (GCM) projections are used to modify the meteorological boundary conditions of the historical 1961-1967 period to reflect the impact of climate change on dry and moderate periods, and speculate on the characteristics of such an extreme drought at the beginning-of-century (2021-2027), mid-century (2041-2047) and end-of-century (2091-2097). This study thus focuses on how the dynamical conditions of this period would manifest in a warming climate.

This study focuses on trends in hydroclimatic variables and the consequences for society and agriculture. Perhaps the most obvious trend being that there will be significant warming, which is observed to be particularly strong over the wintertime at higher latitudes. This causes a decrease in the number of freezing days, early spring snowpack melt and areas of seasonally frozen ground essentially disappearing by the end-of-century. Further, this warming drives a surge in the number of extreme hot days (those with regional mean heat index larger than 41°C). Even with subsequent increases in evapotranspiration, net precipitation increases over most of the NEUS. Using 24-month long-term standardized precipitation index (SPI24) of net precipitation we project mean meteorological conditions of these future drought analogues to be nearly normal, wet and extremely wet at the beginning, middle and end-of-century. However, the short-term SPI (SPI1) of net precipitation indicates this general wetting trend is primarily manifest during moderate months, and so net precipitation variability increases and is responsible for exacerbating the discrepancy between dry and moderate periods. By end-of-century, an extreme drought could potentially develop in only one month from extremely wet conditions.

Other risks also emerge that threaten water resources in this region. For instance, unprecedented extreme precipitation events will emerge during moderate periods. The 99th percentile of precipitation will increase by more than 50% by the end-of-century compared with analogous years in the 1960s. Widespread flood events are expected to become more frequent, and are likely to impact aging infrastructure. Further, early melt of snowpack will lead to less runoff recharge in

the early spring. And degradation of frozen ground will lead to more infiltration of water from surface runoff to soil. In conjunction, by end-of-century these factors will reduce March surface runoff to below half of historical levels, with impacts for the growing season. These changes have likely consequences for both ecosystems and agriculture in the region.

4.2 The record-setting 1960s drought

The 1960s drought, which occurred from late 1962 to 1966, has been deemed as the most severe drought in the Northeastern US over last century. Its prominence in the region’s water resource planning emphasizes that drought is not only limited to commonly dry regions (Barksdale, 1968; Cook and Jacoby, 1977; Janes and Brumbach, 1965; Lyon et al., 2005; Seager et al., 2012). The drought affected millions of people, and covered an area from New England to Virginia and from the Atlantic Coast to Ohio (Barksdale, 1968). As seen in Fig. 4.1, meteorological dryness was the primary driver of the drought, as temperatures were anomalously low over this period (Namias, 1966). These low temperatures spared the region from potentially more severe impacts (Janes and Brumbach, 1965; Namias, 1966). In the New England Region, negative Palmer Drought Severity Index (PDSI) values, associated with drought conditions, began in 1962 and ended at 1966; however, 1962’s annual average PDSI was nearly 0 as the drought’s effects only manifested in the latter half of the year. Therefore, in this study, we refer to the years 1963-1966 as “dry” years and 1961, 1962 and 1967 as “moderate” years. This distinction is important as we will contrast future impacts for dry and moderate periods. The most negative PDSI and lowest soil moisture level of the climatological record occurred in 1965, exemplifying the intensity of the drought and the importance of 1965 as the year with the most pronounced impacts. Therefore, our study uses 1965 as the exemplar dry year and 1961, a year with the largest positive precipitation anomaly of the 1960s, as the exemplar moderate year. Notably, the 1960s drought was at its most severe in the growing season (Barksdale, 1968; Barksdale et al., 1966; Namias, 1966). While the exact dynamical drivers of the drought is still an open question, some suggest the 1960s was driven by precipitation suppression from a low pressure anomaly over the North Atlantic Ocean and a descending, northerly flow over the NEUS (Namias, 1966; Seager et al., 2012), and some demonstrate it was closely linked to the North Pacific

mode (Barlow et al., 2001).

At the beginning of 1962, there was little indication that the NE was descending into a drought state. Precipitation in the early spring of this year was nominal, but after the 6 months of below-average precipitation that followed, a water shortage gradually began to emerge that depleted the soil being used for irrigation (Barksdale, 1968). By late 1962, most observations of runoff and groundwater were below normal levels, and pronounced impacts to agricultural productivity were being felt in states like New York (Barksdale et al., 1966).

Dryness persisted beyond 1962, and although heavy precipitation occurred in late 1963 and early/late 1964, outside of the growing season this did little to prevent the spread of drought (Barksdale, 1968; Barksdale et al., 1966). Consequently, the growing season of 1964 was recorded as the driest of the last century (Janes and Brumbach, 1965). The drought intensified further in 1965 and spread over a wider swath of the northeast. Besides limiting water use, the drought also had an impact on water quality (Barksdale, 1968), as previously unused and polluted water sources began being used to counter water shortages. Rivers' pollutant concentration increased due to insufficient dilution, and sea water intrusion threatened coastal freshwater quality.

Strict rules on water conservation and better management helped greatly in managing the water shortage from 1965 to 1966 such as limitations of air conditioning, swimming pools, domestic water usage, automobile washing and temporary shutdown of some industries and commercial establishments (Barksdale, 1968; Barksdale et al., 1966). At last, the drought ended with abundant precipitation in September 1966 (Barksdale, 1968), with regional mean PDSI rising above zero for the first time in four years.

4.3 Uncertainty of a return to drought conditions under climate change

Large uncertainty remains for future trends of drought in the NEUS. Studies generally conclude that although the region is becoming wetter, drought – especially short-term drought – will occur more frequently and intensely under climate change (Demaria et al., 2016; Frumhoff et al., 2007; Hayhoe et al., 2008). Nonetheless, drought is an emergent feature that is affected by changes in

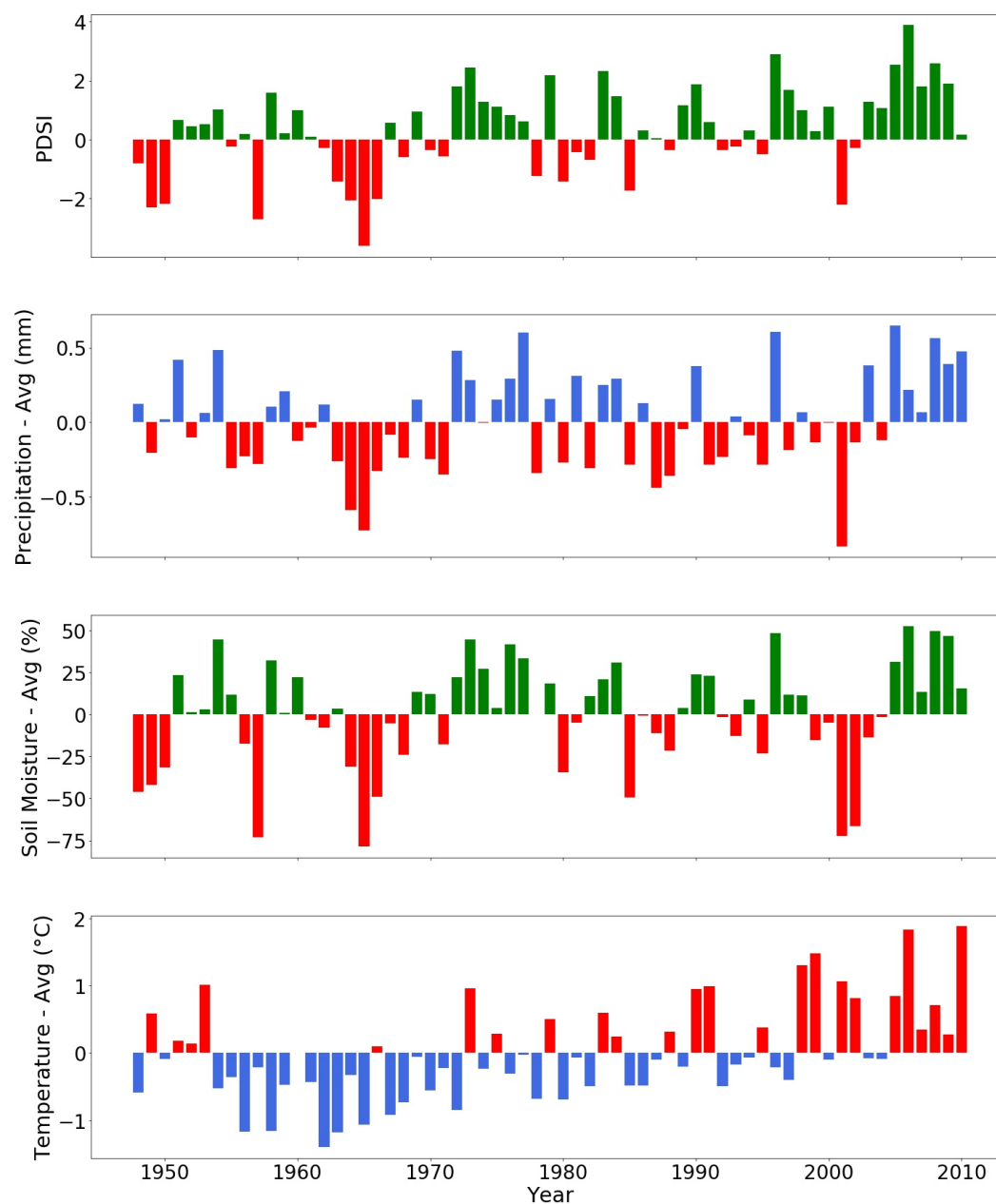


Figure 4.1: Palmer Drought Severity Index (PDSI) and the anomalies of precipitation, temperature and soil moisture from 1948 to 2010 over the New England Region. PDSI data from NCAR (Dai et al., 2004). Soil moisture data from CPC Soil Moisture (Van den Dool et al., 2003). Precipitation and temperature from CERA-20C R7 (European Centre for Medium-Range Weather Forecasts, 2016).

thermodynamics (e.g. increases in atmospheric water vapor), hydrology (e.g. precipitation phase, runoff, related land surface variables, and surface-atmosphere fluxes) and dynamical conditions (shifts in the frequency, intensity or duration of meteorological patterns). Overall, both historical observations and model projections indicate an upward trend in average temperature ($0.3^{\circ}\text{C}/0.5^{\circ}\text{F}$ per decade since 1970, with wintertime warming of $0.7^{\circ}\text{C}/1.3^{\circ}\text{F}$ per decade) and a slight increase in average runoff and evapotranspiration (Frumhoff et al., 2007; Hayhoe et al., 2007; Seager et al., 2012). This has meant more extreme heat days, early melt dates, a lower snowfall-rainfall ratio, and a longer growing season along with more water demand (Burns et al., 2007; Frumhoff et al., 2007; Hayhoe et al., 2007; Seager et al., 2012). Although increased precipitation has meant that drought indices such as SPEI and SPI are shifting towards more positive values, indicative of generally wetter conditions, the spread of these indices is also increasing; consequently, the probability of extreme drought is largely unchanged in both observational data and models (Krakauer et al., 2019).

Water resource planning in the NEUS is highly reliant on a model drought based off of the 1960s drought period (1962-1966). Given subsequent climatic shifts (and foreseeable climatic shifts), there are concerns with the use of a model drought from more than a half century ago (Moser et al., 2008). Consequently, NEUS water management agencies agree that this model drought should be revisited in light of climate change. In the future, earlier snowmelt dates and reduced wintertime snowpack will certainly impact seasonal availability of water (Burns et al., 2007; Frumhoff et al., 2007; Huntington et al., 2004). Future warming will lead to a longer growing season and enhanced evaporation, thus enhancing consumption of available freshwater, particularly in spring and summer (Frumhoff et al., 2007; Lyon et al., 2005; Seager et al., 2012). While the 1960s drought is notable for its severe water shortage in these seasons, its socioeconomic impacts were also tempered by low temperatures. Capturing these factors under climate change motivates the use of a comprehensive model-based study of this period.

4.4 A simulation of present and future analogues of the 1960s drought

Having motivated the purpose of our study, we now present our methodology and results from our simulations using pseudo-global warming, including temperatures, precipitation, evapotranspiration, snowpack, soil moisture, runoff, and drought indices.

4.4.1 Methodology

In this study the Weather Research and Forecasting (WRF) Model (Powers et al., 2017; Skamarock et al., 2008) is used for simulating the regional hydrometeorological and drought conditions of the NEUS during the returned historical 1960s drought periods. WRF is one of the most commonly-employed regional climate modeling systems currently available, incorporating many widely-recognized physical parameterizations. Thousands of research studies have been conducted with WRF worldwide, demonstrating WRF’s utility for robust simulation of regional climate. With an appropriate choice of parameterizations, WRF has been shown in past studies to accurately reproduce the hydroclimatology of the NEUS (Ganetis and Colle, 2015). In this study WRF 3.9.1 is used with the parameterization suite given in Table 4.1. The land surface model employed is the Community Land Model 4 (CLM 4) (Oleson et al., 2010), which is the most complicated and expensive of the available options in WRF, and one that shows reasonable performance across a variety of geographies (Case et al., 2008; Jin et al., 2010; Ullrich et al., 2018). Extensive validation was performed on our simulation to ensure good model performance (see Supplement 4.6.4).

The parameterization sets all come from the Community Earth System Model (CESM) (Hurrell et al., 2013; NCAR, 2019) and are chosen because, in this model, they have been shown to be robust and work well across regions and scales. These parameterization sets were also employed in our previous study on the 2012-2016 California drought (Ullrich et al., 2018) and have been demonstrated to exhibit good performance in other studies (Bretherton and Park, 2009; Jiménez et al., 2012; Neale et al., 2010; Zhang and McFarlane, 1995).

Table 4.1: Physical parameterizations used in our WRF simulations.

Process	Parameterization
Microphysics	CAM V5.1 two-moment five-class (Neale et al., 2010)
Radiation	RRTMG (Iacono et al., 2008)
Surface layer	Revised MM5 similarity theory (Jiménez et al., 2012)
Land surface model	CLM4 (Oleson et al., 2010)
Planetary boundary layer	UW (Bretherton and Park, 2009)
Cumulus parameterization	ZM (Zhang and McFarlane, 1995)

Simulation period and domain

Our simulations in this study cover four time periods: historical (1960s) (1960-1967), present-day (2020s) (2020-2027), mid 21st century (2040s) (2040-2047) and late 21st century (2090s) (2090-2097). In each simulation the first year serves as the spin-up period to ensure hydrologic and meteorological conditions have stabilized. Two nested domains are used (Fig. 4.2). The outer and inner domains have 105×89 and 187×133 grid points, with resolutions of 18 and 6 km, respectively. Due to the long duration of the simulation, spectral nudging is employed (with the default relaxation timescale) so as to reduce internal model drift. In our simulations the 30-arc second (~ 1 km) resolution United States Geological Survey-based land use and land cover and topography datasets are interpolated to the model grids as geographical input.

Although most of our analysis focuses on the inner domain (Fig. 4.2), some detailed analyses are conducted within the southern New England subregion (defined as 41N to 43N latitude and 74W to 70W longitude). This location comprises the most populated and developed areas of the NEUS.

Modified forcing from pseudo-global warming

Lateral forcing data for this historical period is from the 6-hourly Coupled ECMWF Re-Analysis system of the 20th-century (CERA-20C) R7 interpolated to 0.5° resolution. CERA-20C is a coupled reanalysis dataset with global coverage from 1901-2010, designed to capture low-frequency climate variability (European Centre for Medium-Range Weather Forecasts, 2016). This dataset is chosen because of its relatively high spatial and temporal resolution, and because its precipitation amounts

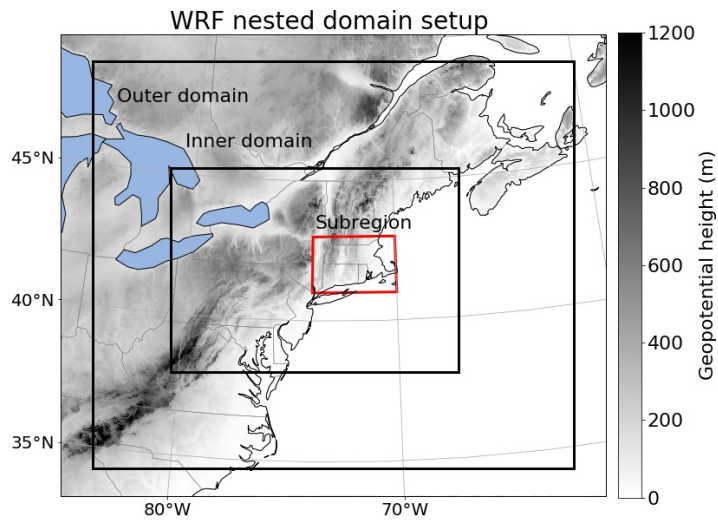


Figure 4.2: Our WRF domain setup for all simulations in this study. Shading indicates the surface height. Grid spacing in the outer (inner) domain is 18 km (6 km).

best match observations of mean precipitation over the NEUS. After comparing the performance across all 10 CERA-20C ensembles, we selected the CERA-20C R7 ensemble as it again provided the highest performance among ensembles. More details on the evaluation protocol are provided in the Supplement section 4.6.2.

Anticipated future changes to lateral forcing under climate change are derived from Coupled Model Intercomparison Project phase 6 (CMIP6) projections under SSP585 with the socioeconomic assumption of fast economical and technical development and low population (O'Neill et al., 2014, 2017). This scenario is widely used and assumes high greenhouse gas emissions, illustrating the high risks and challenges under a fossil-fuelled rapid development (O'Neill et al., 2017; Riahi et al., 2017). End-of-century conditions under lower emission scenarios can be approximated with our early or mid-century PGW simulation results. In this study we use data from the multi-model mean of four CMIP6 models with demonstrably good performance in the NEUS region (namely, CESM2, MRI-ESM2-0, CNRM-ESM2-1 and GFDL-CM4), as identified by Srivastava et al. (2020). Following Ullrich et al. (2018), the spatially averaged monthly mean projections are used to calculate the difference between each of the 2020s, 2040s, and 2090s periods against the 1960s period. Both temperature and relative humidity are assessed in this manner. The resulting temperature differences as a function of month and altitude are depicted in Fig. 4.21. We observe a positive temperature delta throughout the troposphere (up to around 100 hPa altitude), with a local maximum occurring around 350 hPa over the summer, and at the surface over winter. There is a negative temperature delta in the stratosphere, as anticipated under climate change. The magnitude of the temperature delta (both positive and negative) clearly increases from the 2020s to the 2040s and the 2090s, although the patterns are consistent. Relative humidity differences are small and so are not shown.

Lateral forcing data for the future simulations are the same as historical, except with the temperature delta (Fig. 4.21) added over the entire domain, on constant pressure surfaces. Based on our observation of essentially negligible changes in relative humidity, relative humidity is held fixed (resulting in enhancement to specific humidity). Sea surface temperatures are analogously modified using the multi-model mean of the selected CMIP6 models to accord with the change to air temper-

atures. Finally, greenhouse gas concentrations are modified in WRF’s radiation parameterization in accordance with the RCP8.5 emission scenario.

4.4.2 Overview of simulations

By comparing our historical simulation with its driver (CERA-20C) and other reliable, finer-resolution datasets – in this case, the Climate Prediction Center unified gauge-based analysis precipitation data (CPC) (NOAA Physical Sciences Laboratory, 2020) and the ECMWF Reanalysis v5 (ERA5) (European Centre for Medium-Range Weather Forecasts, 2020) – we conclude that our simulation generally captures surface temperature, precipitation, and soil moisture over the historical period. From Fig. 4.3, we see historical regional monthly mean precipitation is similar to these products over both the inner domain and the subregion. Correlations between simulated and CERA-20C regional monthly mean precipitation are 0.91 over the inner domain and 0.84 over the subregion. Correlations between simulated and CPC/ERA5 precipitation over the inner domain are 0.81 and 0.77. However, our simulation does tend to underestimate wintertime temperatures in regions of rugged topography and generally overestimates summertime precipitation over coastal regions (Fig. 4.28 and 4.29). With regards to the temperature underestimation, this can be partially explained because of better resolution of topographic effects (Heikkilä et al., 2011; Zhu et al., 2020). Inaccuracy in summertime precipitation is likely because our grid spacing within the inner domain does not permit resolution of some convective features (Li et al., 2014). Examining the spatial structure of the bias, we see that precipitation is overestimated over coastal regions and underestimated over the inland regions and the Great Lakes (Fig. 4.30).

Soil moisture appears to be captured well after initial spin-up; since there is little change in the soil moisture structure thereafter, the 9 month spin-up period appears sufficient. This is likely because our domain is relatively small and the simulation is initialized from reanalysis data (Cosgrove et al., 2003; Leduc and Laprise, 2009). (See Supplement 4.6.4 for the detailed information and more figures). In general, there is good agreement between the historical simulation and observational datasets; while some differences are apparent, we assume that these biases are largely independent of the simulated time period (that is, dominated by structural error) and so will not

significantly affect conclusions related to future change.

Simulated annual mean temperature and precipitation over our subregion for each year of the drought period is depicted in Fig. 4.4. Historical data comes from CERA-20C R7 and simulations are corrected by the regional mean difference between historical data (1961-1967) and WRF 1960s simulation. This plot also gives us a quick glimpse of how the climate of this region is projected to change: Whereas all years of the 1960s were below the 50th percentile of precipitation and most were below the 50th percentile of temperature, each simulated year of the 2040s and 2090s is above the 99.9th percentile of temperature, and all years of the 2090s are well above the 95th percentile of precipitation. This figure clearly highlights the significant regional shift towards a future warmer and wetter climate.

4.4.3 Temperature

From the CMIP model ensemble, the average warming rate over land from the 1960s period to 2090s period is 0.052°C per year, which is higher than the observed global warming rate over land and ocean since 1981 (0.018°C per year) (Lindsey and Dahlman, 2020). Simulated warming of this magnitude is not unreasonable, as warming is expected to be much stronger over land and at higher latitudes (Hoegh-Guldberg et al., 2018). Fig. 4.5 shows the spatial pattern of 2m temperature from historical and its corresponding change. In general the magnitude of warming intensifies from the 2020s drought to the 2040s and the 2090s droughts, in accordance with expectations from the CMIP6 models under SSP585. However, the spatial and seasonal distributions of warming are uneven, with a stronger warming trend in winter (DJF) and at higher latitudes, where historical temperatures are lower; for example, regional mean change over land in 2045 DJF (4.14°C) is 1.52°C larger than 2045 JJA (2.62°C). From Fig. 4.6, a clear correlation between future 2m temperature change and historical mean temperature at each grid point emerges, with enhancement of the change in the winter season and under increased forcing; for example, the correlation over the 2095 winter (-0.62) is much larger than over the 2095 summer (-0.53), the winter of 2045 (-0.50), and the winter of 2025 (-0.43). This trend suggests that cold regions will be warming faster than warm regions, indicative of some homogenization of temperatures over seasons and regions. Thus

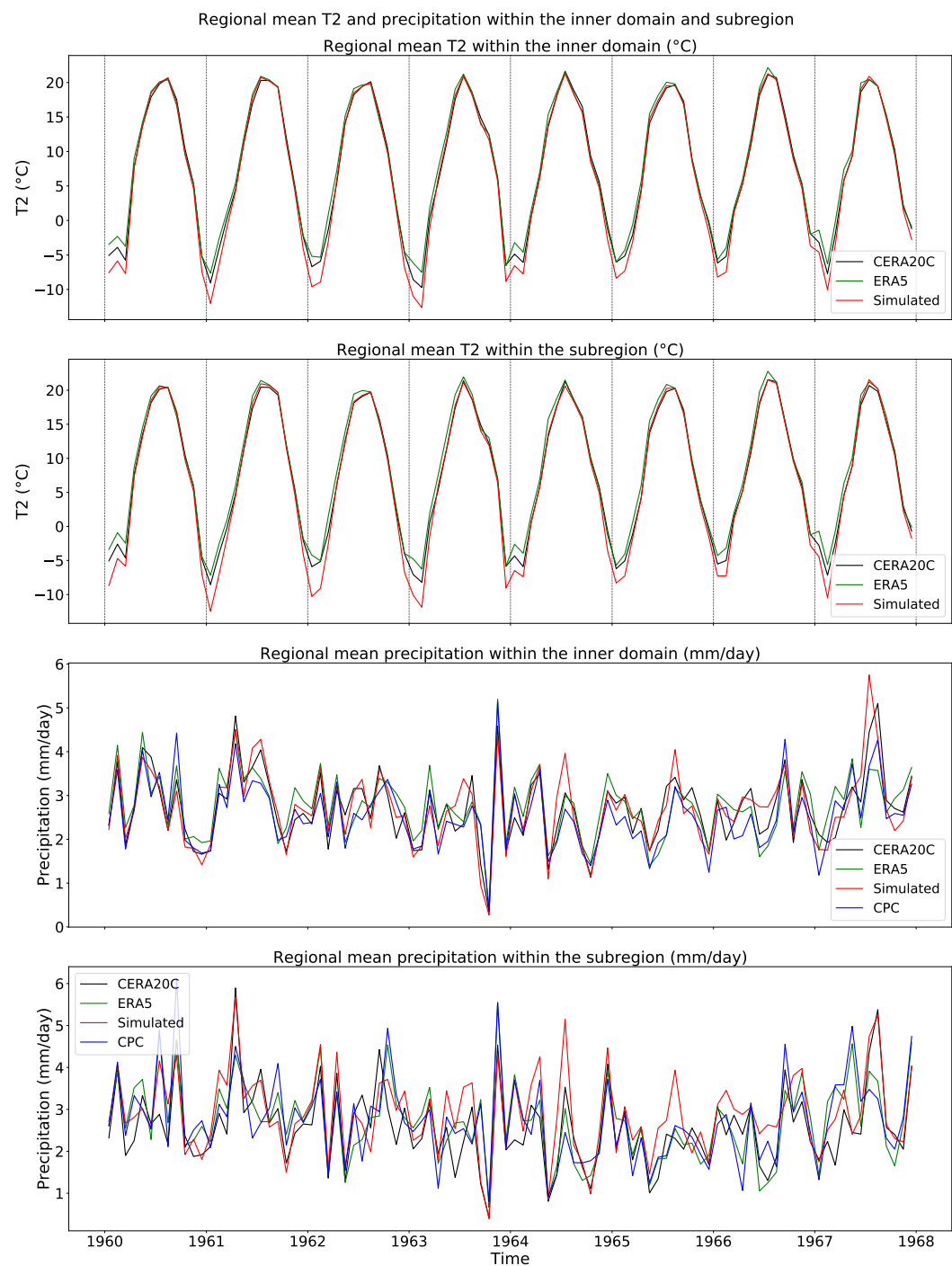


Figure 4.3: Simulated regional monthly mean 2-meter temperature (°C) and precipitation (mm/day), and reference from CERA-20C R7, ERA5 and CPC within the inner domain and subregion.

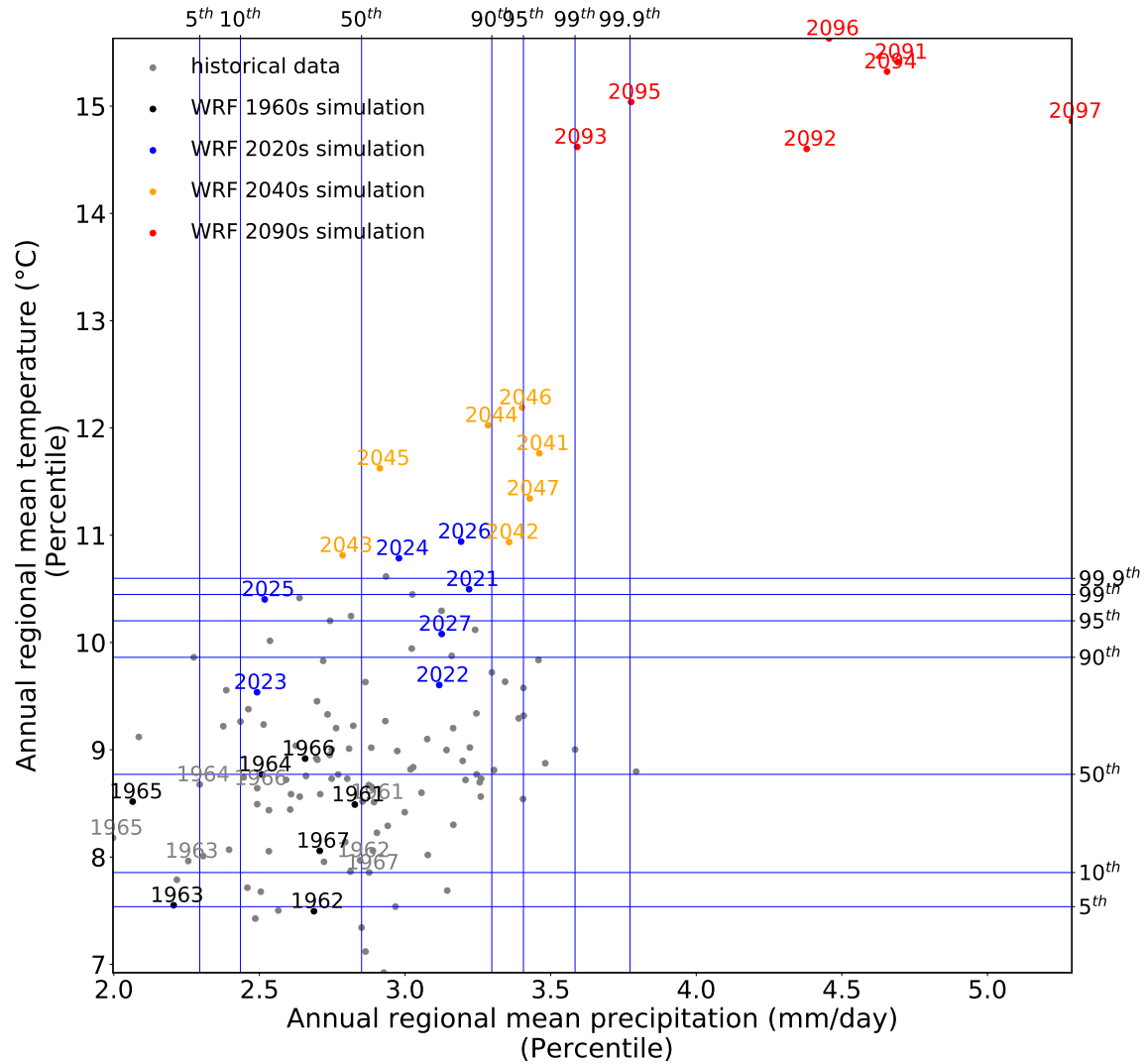


Figure 4.4: Regional annual mean precipitation and 2 meter temperature within the subregion during historical and future periods, as compared with historical percentiles over the period from 1910 to 2010.

temperature spatial and temporal variability are reduced, in turn driving earlier snowmelt and intensified evapotranspiration.

Do these trends also hold for moderate periods? Although the lateral temperature deltas of both the dry periods (1963-1966) and moderate periods (1961, 1962 and 1967) are the same, the simulations produce greater regional warming during moderate periods than dry periods (although the difference is small). Further, some differences in spatial distribution of temperature change persist: From Fig. 4.5 (fourth row), we can see that during the moderate wintertime period, the regions with highest temperature change are along the southern extent of New England; however, dry years have the greatest warming along northern extent of the New England (Fig. 4.5, second row). These wintertime spatial differences have consequences for dry years and non-dry years, such as shifts in the number of freezing days and snowmelt (touched on later).

Extreme temperatures

It is well known that shifts in mean temperatures will have a disproportionate influence on the frequency of extreme temperatures. From Fig. 4.7, there is a clear increase in the mean annual maximum 2m temperature at all grid points, with more extremely hot days in the future; however, there is essentially no change in the annual variance of temperatures. With that said, both the mean and outliers of annual maximum daily temperature increase more in dry years rather than moderate years, which consequently drives an increase in evaporation and risk of flash drought. Frequency of extreme heat days are assessed using the Heat Index (HI) (Rothfus and Headquarters, 1990) to better distinguish extremely hot days with potential for significant socioeconomic impact (see Supplement 4.6.3 for the detailed definition of the Heat Index).

As defined by NOAA, values of HI larger than 41°C indicate dangerously hot conditions which may trigger sunstroke, heat cramps and heat exhaustion (NOAA, 2020). Fig. 4.8 shows changes in the number of extreme heat days for each period for the subregion. Compared with historical conditions (noting that this was a relatively cool period), the number of extreme hot days increases from 6 in 1965 to 27 in 2045 and 56 at 2095. It's clearly the case that extreme heat will be a major public health concern in the NEUS going forward.

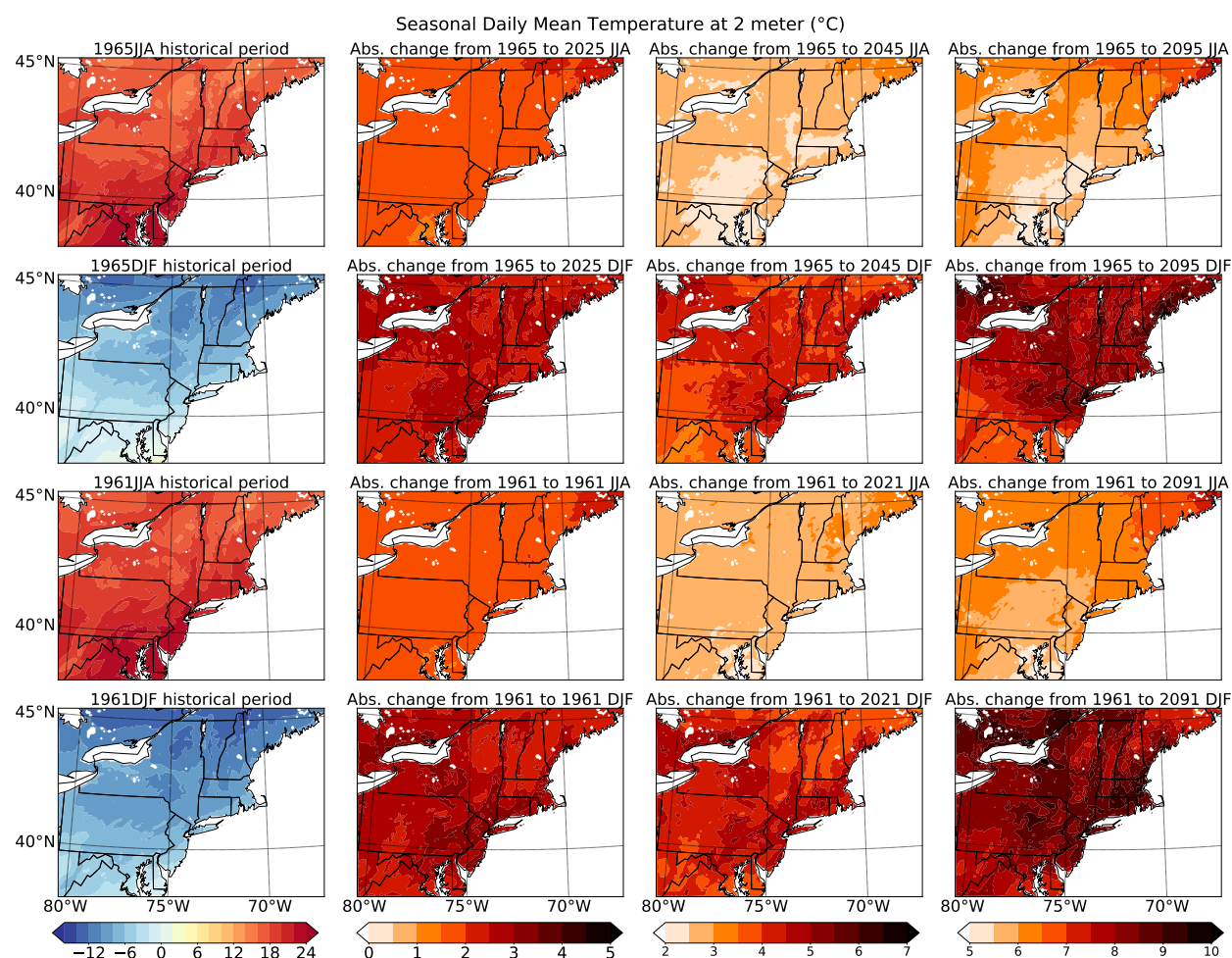


Figure 4.5: Average daily 2m temperatures (in degrees Celsius) over June-July-August (JJA) and December-January-February (DJF) in 1965 and 1961 (and their future analogues), exemplary of dry and moderate years.

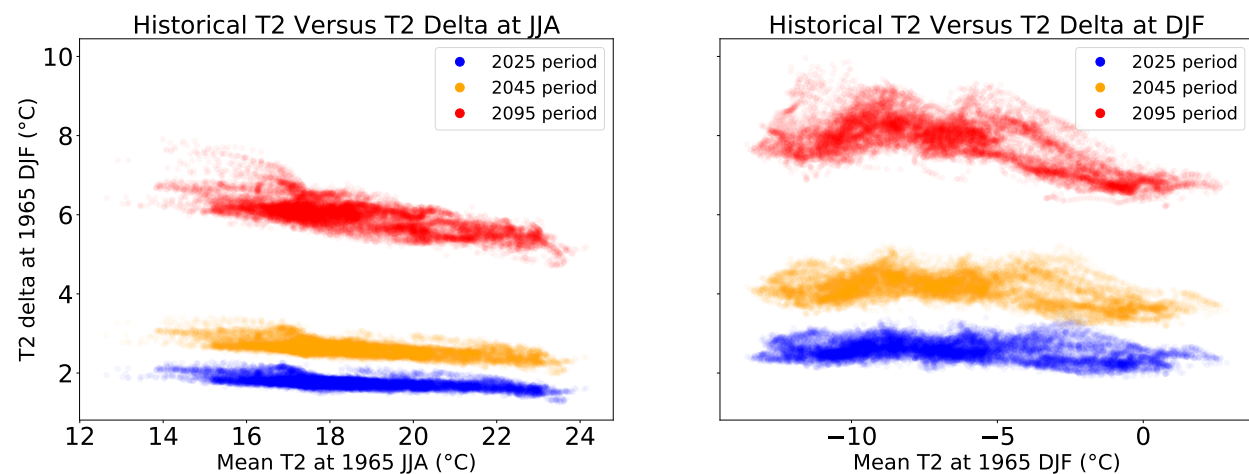


Figure 4.6: The relationship between historical mean daily 2m temperatures and future 2m temperatures deltas over JJA and DJF in 1965 drought conditions.

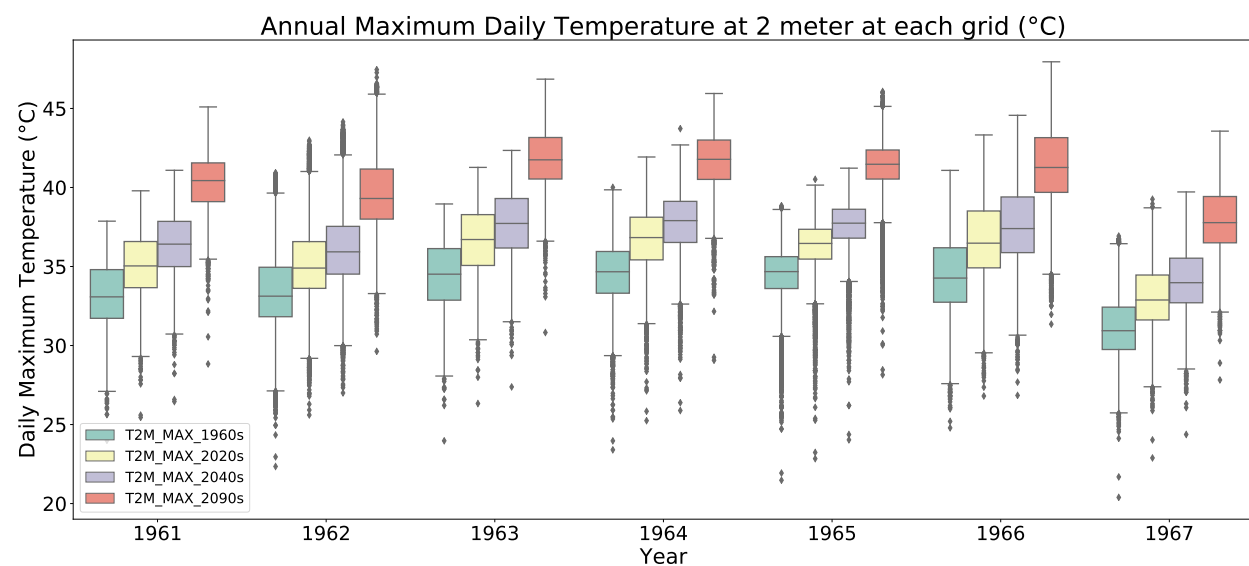


Figure 4.7: Annual maximum daily 2m temperature (in degrees Celsius) at each grid point.

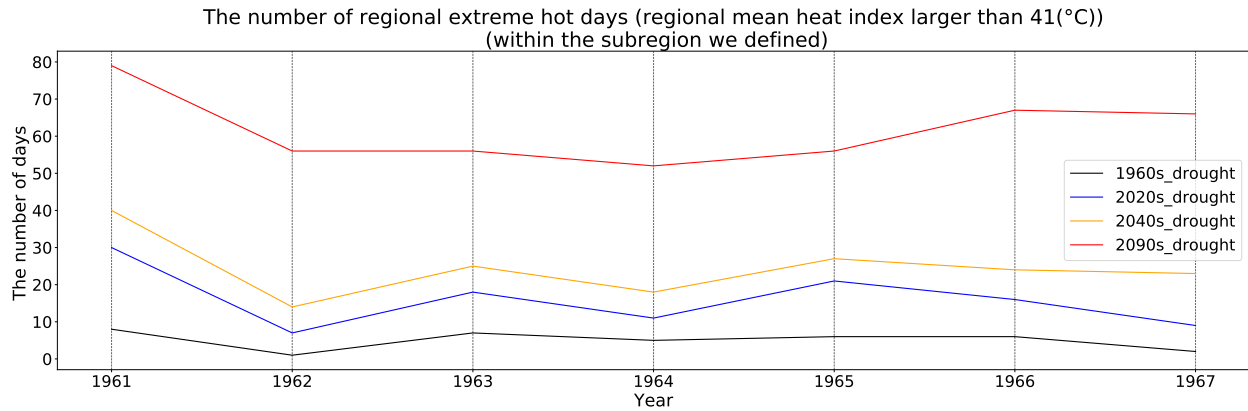


Figure 4.8: The number of regional mean extreme heat days (regional mean Heat Index larger than 41°C) for the historical period and each future analogue.

The warming climate will also reduce the number of freezing days (days with daily 2m temperature minimum less than 0°C) significantly in both dry year and moderate year (Fig. 4.9). The change in freezing days is highly correlated with change in wintertime T2 (Fig. 4.5). Further, the spatial distributions of the change in freezing day count differs significantly between our exemplar dry year (1965) and moderate year (1961), in accord with their associated temperature deltas and historical number of freezing days. Higher latitudes produce greater decreases of freezing days, where historical freezing days are more common and warming is larger. In these regions, we thus expect degradation of frozen ground (Zhang et al., 2003), which we will revisit later.

4.4.4 Precipitation

Fig. 4.10 depicts seasonal mean daily precipitation over the historical and projected periods. Increasing precipitation is apparent in most regions, especially along the southeastern coasts during winter and in the southwest during summer. These increases are expected because of an intensified hydrological cycle in the warming climate (Huntington et al., 2004; Pfahl et al., 2017). In the literature, the rule-of-thumb of warming climates ‘wet becomes wetter’ (Chou and Neelin, 2004; Donat et al., 2016; Seager et al., 2010) has been often employed to explain precipitation change over the ocean (Byrne and O’Gorman, 2015); nonetheless, it also applies here (especially in the winter season). In Fig. 4.10 regions with greater precipitation increase coincide with regions of

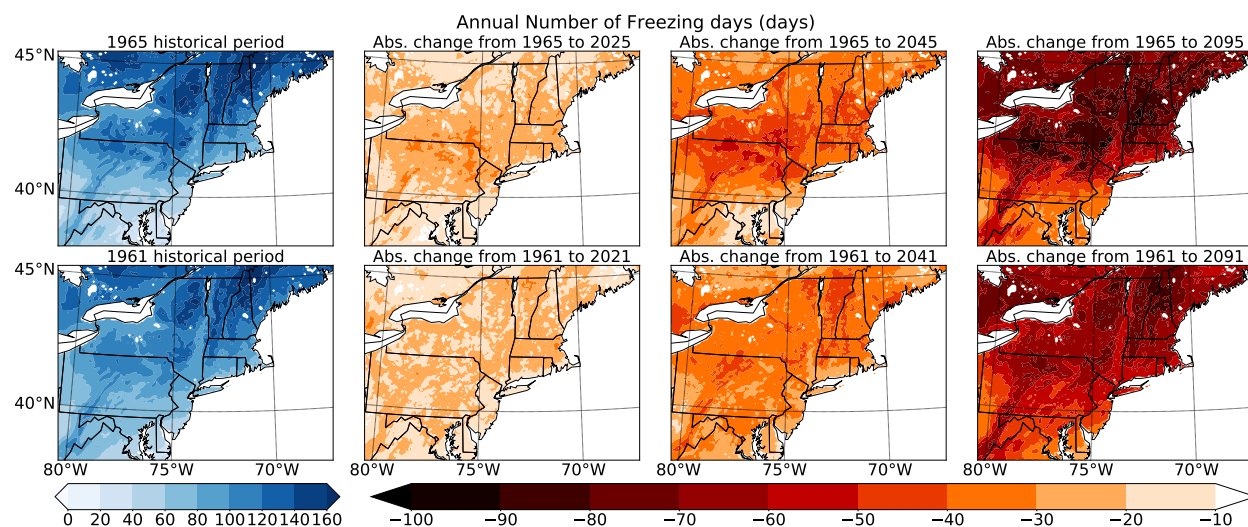


Figure 4.9: The number of freezing days in 1965 and 1961 (and their future analogues), defined by daily minimum 2m temperature less than 0°C.

larger historical mean wintertime precipitation, with pattern correlation in 2025 DJF, 2045 DJF and 2095 DJF of 0.51, 0.55 and 0.50, respectively (FIG. 4.26). This result also applies for all other dry and moderate years, with even higher correlations of 0.8 in some cases (e.g. 2043 and 2093 DJF). The applicability of this rule of thumb to the inland NEUS is likely a consequence of a relative abundance of water vapor in the region from the Atlantic Ocean and Gulf of Mexico. What’s more, unlike the dry period, the moderate period doesn’t experience more precipitation in the northeastern part of the inner domain; however, as we will discuss later, this region experiences a significant soil moisture increase during the moderate period.

Although this study is focused on drought, the dramatic increase in future precipitation deserves some discussion. Extreme precipitation is notorious for its disastrous impacts on society, and has been increasing in frequency across the continental US. This increase is particularly pronounced over the NEUS (Hayhoe et al., 2007; Huntington et al., 2004), where the most intense daily precipitation events (99th percentile daily precipitation) have increased by more than 70% from 1958 to 2012 (Melillo et al., 2014). Our simulations also indicate that more extreme precipitation events will occur here in the future. Fig. 4.11 shows that both absolute and relative precipitation percentiles will increase, with greater increases from the 2020s to the 2090s. In particular, the 99th percentile

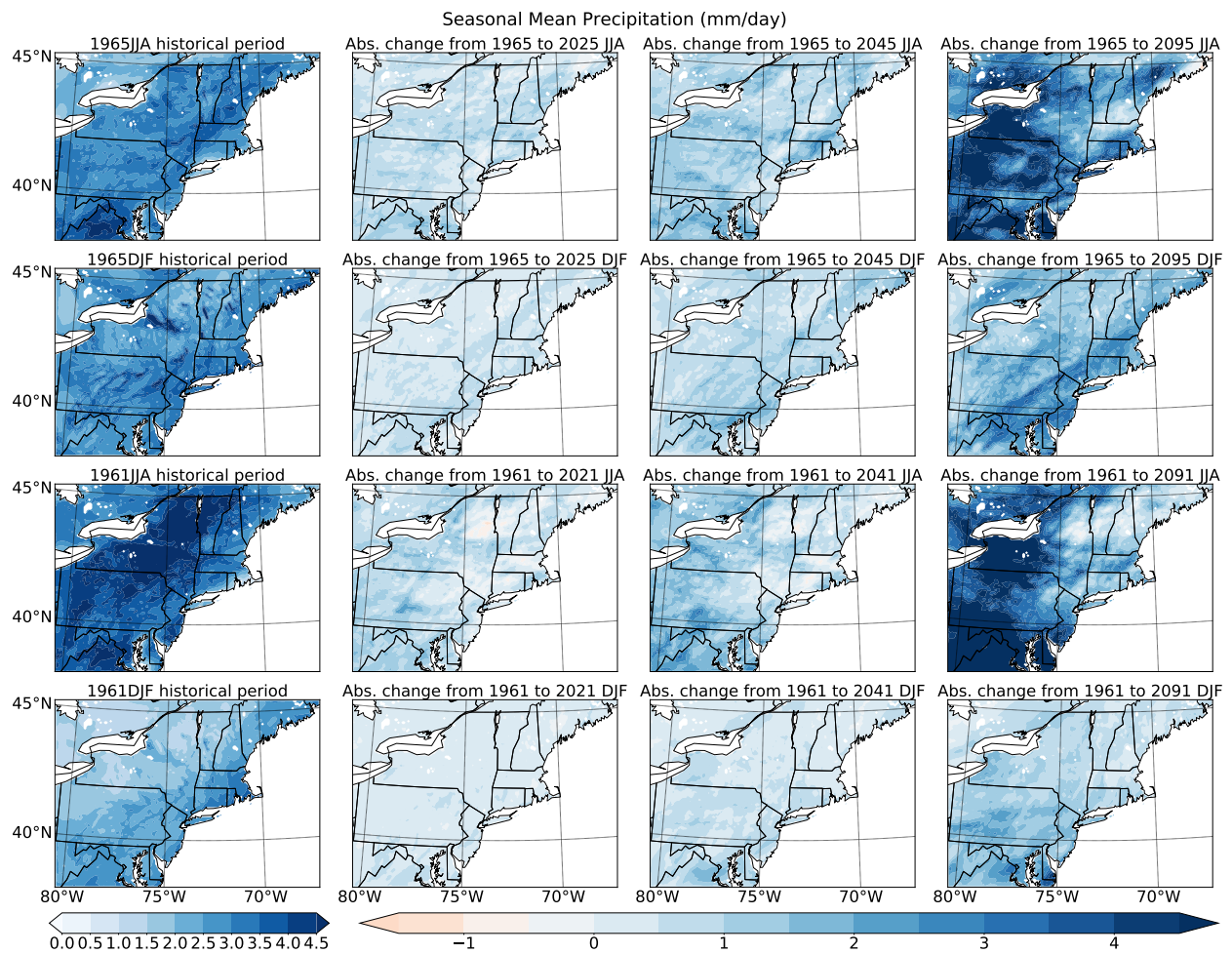


Figure 4.10: Seasonal mean precipitation distribution and change in the exemplar 1965 dry year and 1961 moderate year (mm/day).

of precipitation will increase more than 50% in both dry and moderate periods in 2090s versus the 1960s. Examining inner domain grid points' annual maximum precipitation (Fig. 4.12), the mean and upper tail of the annual maximum precipitation distribution both increase into the future. We expect unprecedented extreme precipitation (daily precipitation larger than 160 mm/day) may occur (especially in non-dry years) that will challenge the capacity of flood control equipment in NEUS.

4.4.5 Evapotranspiration

Enhanced evapotranspiration can directly reduce the net input of water from atmosphere to land, decrease runoff and soil moisture, and increase water demands for agriculture and ecosystems. Fig. 4.13) shows that our future analogues exhibit greater summertime evapotranspiration compared to winter, in accord with the spatial and temporal distribution of historical precipitation (Fig. 4.10). Of course, this is unsurprising as evapotranspiration amounts are closely related to water available. Evapotranspiration also increases more toward higher latitudes because of the warming effect, especially in wintertime. These trends hold for all dry and moderate years (Fig. 4.10). Moderate periods producing stronger evapotranspiration intensity are likely caused by more significant warming and more abundant precipitation.

As noted earlier, precipitation increases correlate with historical precipitation. Considering the strong relationship between the evapotranspiration and precipitation change, this inspires the question “how does a net precipitation (precipitation minus evapotranspiration) change emerge?” Fig. 4.14 shows that although evapotranspiration increases, precipitation increases more rapidly, thus producing an overall increase of net precipitation. Consequently, our earlier use of “wet becomes wetter” also applies to net precipitation, especially in winter months where correlations are more than 0.6 (and up to 0.87) between historical net precipitation and its change during both dry and moderate periods. It's further clear that the dry period summertime has much less net precipitation than the moderate period, suggesting that net precipitation is valid for indicating drought conditions. Note that the wet conditions of 1965 DJF was caused by a short-term abundant historical precipitation event.

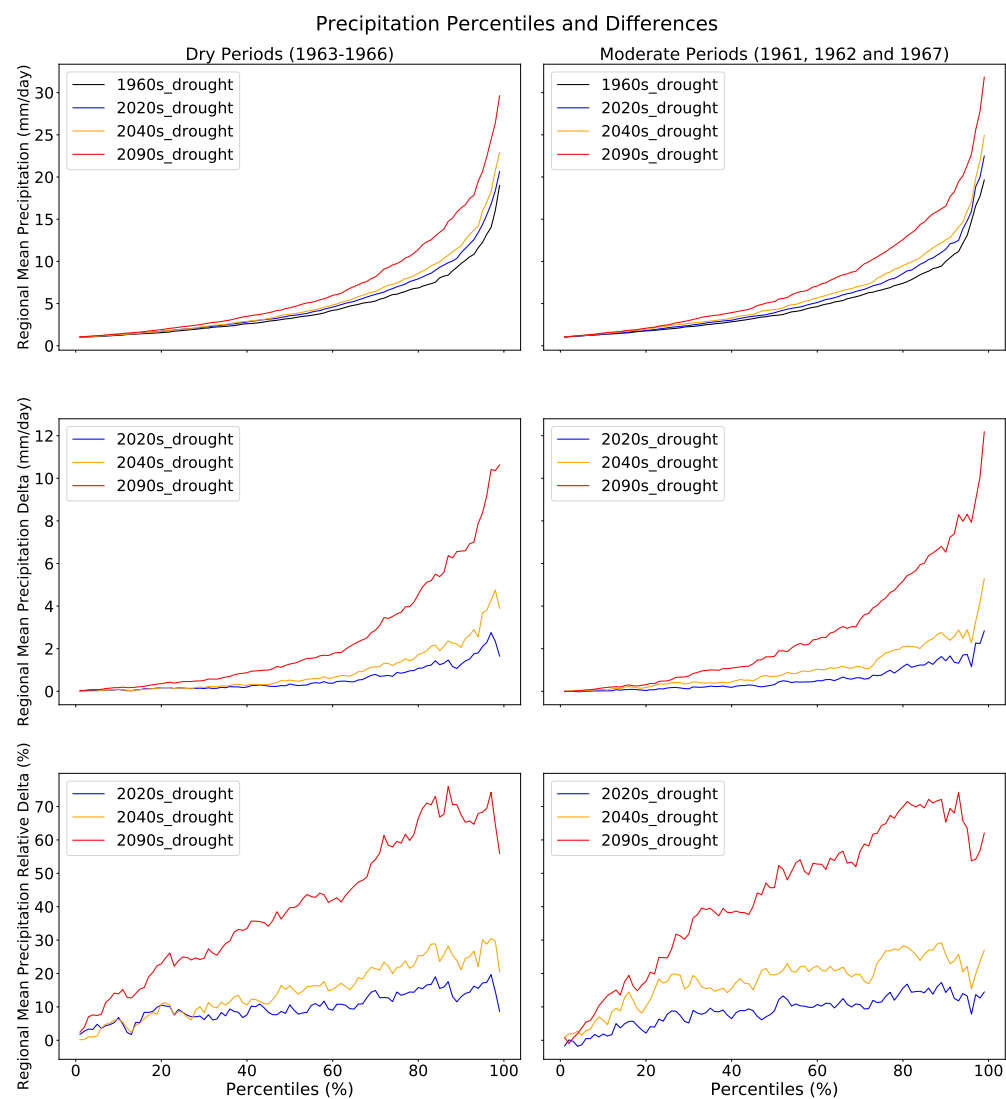


Figure 4.11: Regional mean precipitation change at different percentiles. Only daily precipitation events larger than 1 mm/day are included.

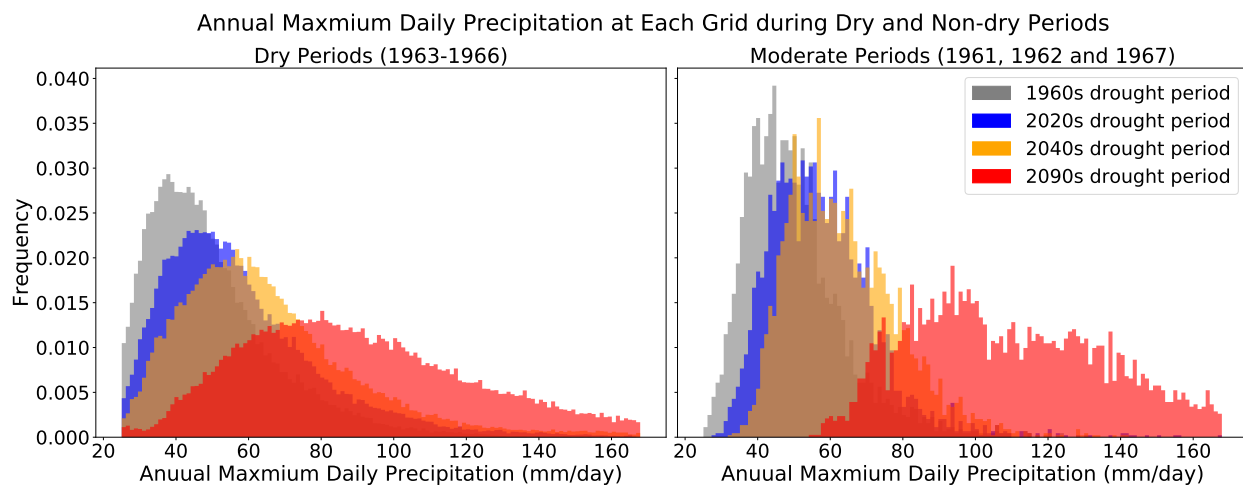


Figure 4.12: Annual maximum precipitation distribution of all grid points within the inner domain.

4.4.6 Snowpack

Due to its connection to the hydrologic cycle, water supply and ecosystems in the NEUS, an understanding of future snowpack is necessary for water resource planning. Fig. 4.15 shows a clear and rapid decrease in snowpack in this region in DJF and MAM in response to warming. Within the inner domain, seasonal regional mean snow water equivalent (SWE) was 20.56 kg/m^2 in 1965 DJF; however, 2095 DJF only produced 7.16 kg/m^2 of SWE (a 61% decline). This decrease is most pronounced in the spring season (MAM); 2095 MAM exhibits a 94% drop in SWE over 1965 MAM (Fig. 4.16). Lower latitudes are most strongly impacted as here snow is more sensitive to temperature increases. The result is a loss of spring snowmelt contribution to runoff (Fig. 4.15).

Although there is a greater absolute SWE loss over the moderate period, the relative change in the 1961 moderate year (Fig. 4.15) is still much smaller than in the 1965 dry year (Fig. 4.15). Notably, during the 1965 dry period there is practically no historical snow accumulation in spring over the northeastern states of the NEUS (Fig. 4.15), a result of low precipitation. The depleted wintertime snowpack was a major reason for the seriousness of the drought in springtime; without snowmelt the surface runoff reached record lows. Over the northeastern corner of the domain the absolute change in SWE between 1965 and its future analogue is thus fairly small, because there is essentially no snow to remove (Fig. 4.15). On the other hand, during the moderate periods this

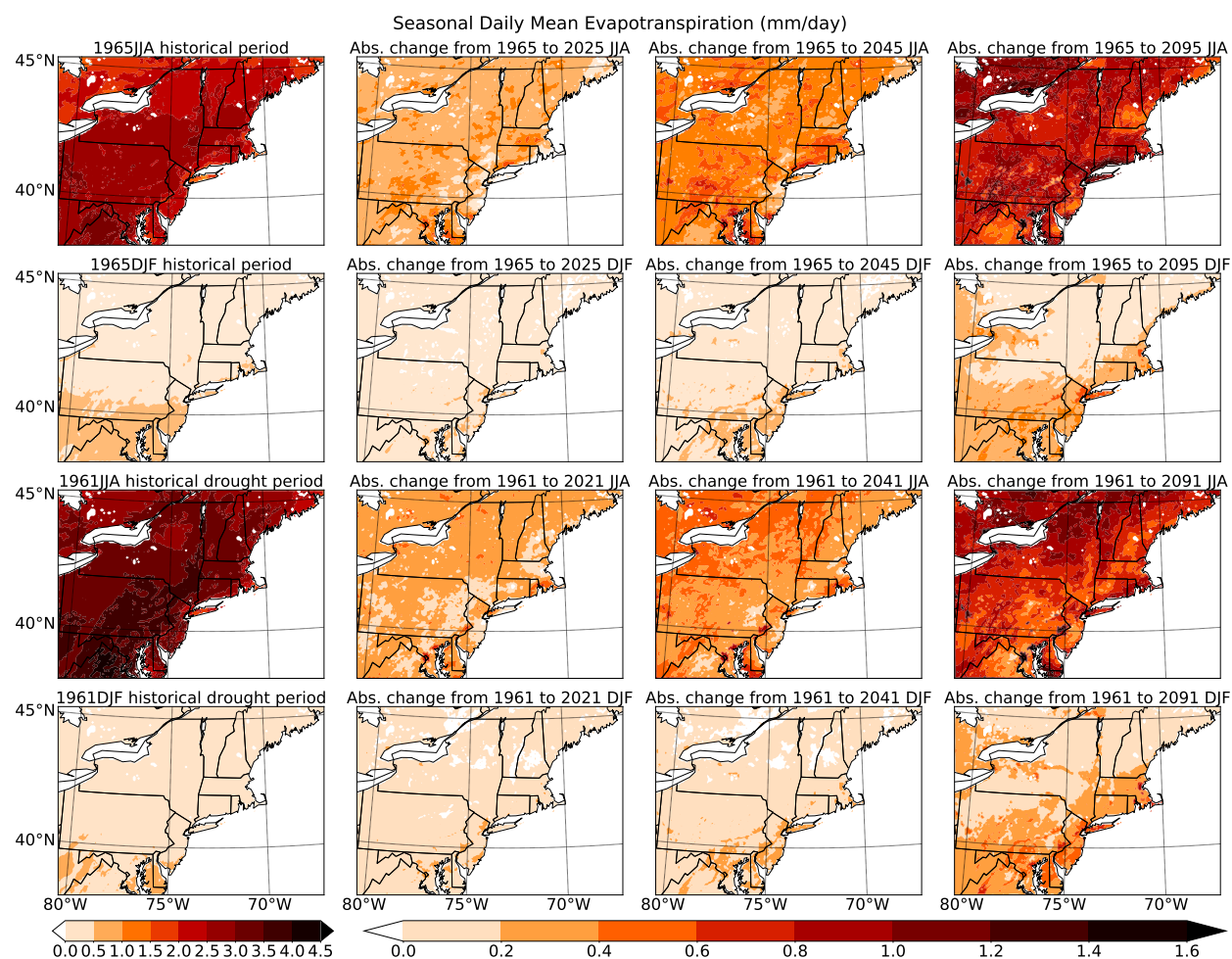


Figure 4.13: Seasonal mean evapotranspiration (mm/day) over the 1965 dry exemplar and 1961 moderate exemplar, and projected changes in their future analogues for the JJA and DJF seasons.

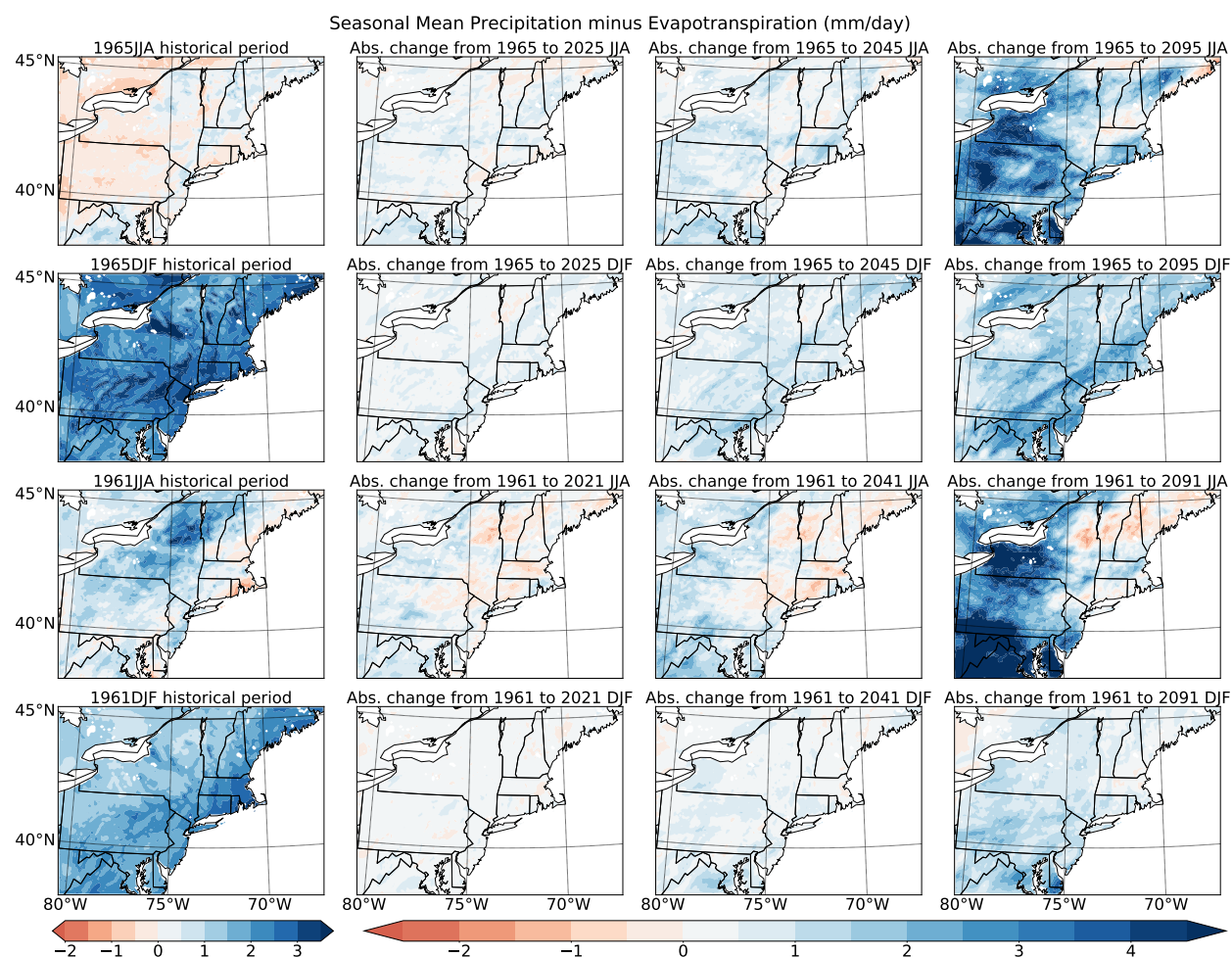


Figure 4.14: Seasonal mean net precipitation (mm/day) over the 1965 dry exemplar and 1961 moderate exemplar, and projected changes in their future analogues for the JJA and DJF seasons.

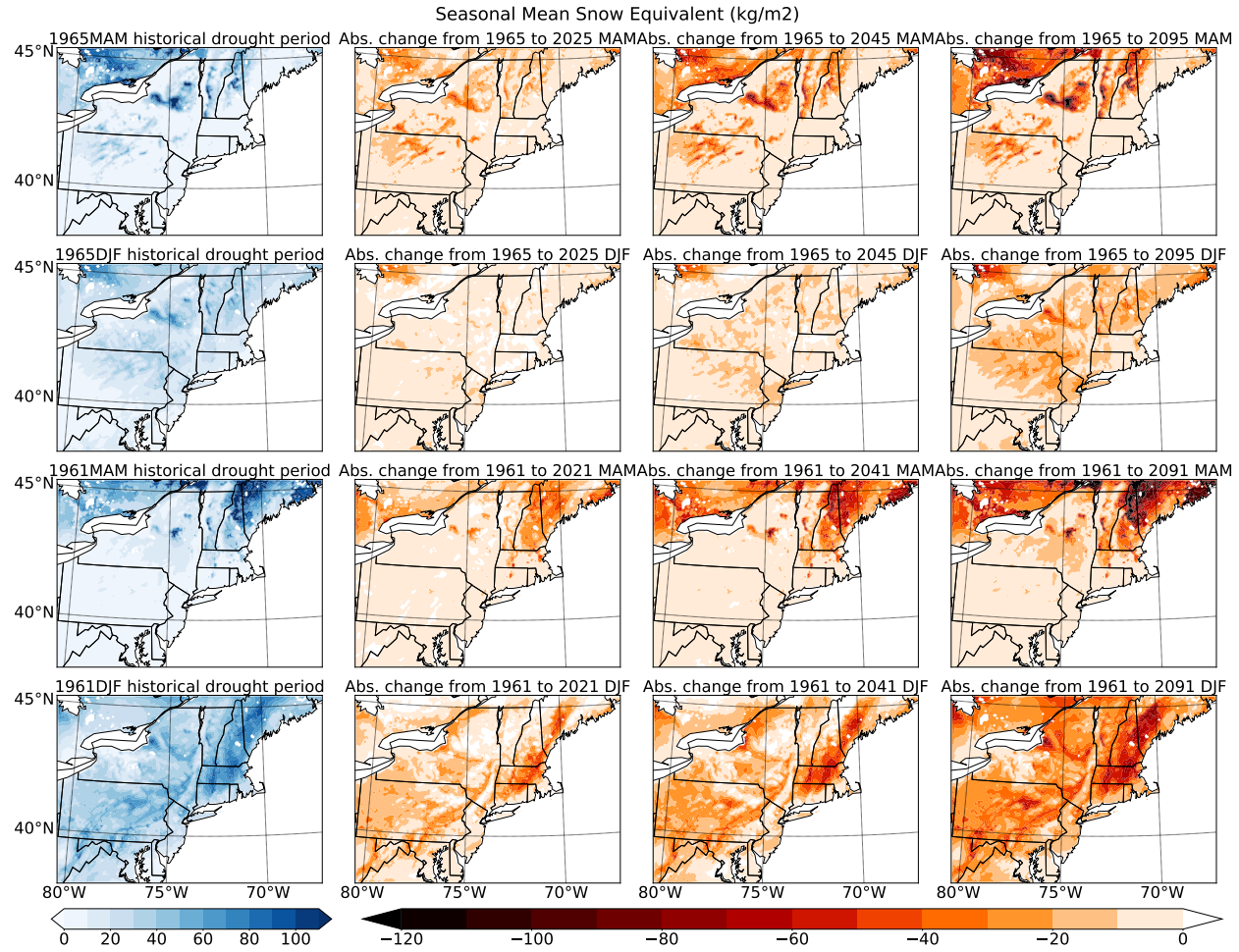


Figure 4.15: Seasonal mean snowpack absolute change (kg/m^2) over the 1965 dry exemplar and 1961 moderate exemplar, and projected changes in their future analogues for the JJA and DJF seasons.

region has a healthy snowpack, which is severely depleted in the future (Fig. 4.15).

4.4.7 Soil moisture and runoff

Soil moisture and runoff are two essential hydrologic variables and indicators of drought and water supply. In WRF-CLM4, soil moisture is accumulated over 10 layers; we focus on the average column soil moisture, which is the average soil moisture in each layer weighted by its thickness. Seasonal mean soil moisture over the 1965 and 1961 exemplar years (and differences in their future analogues) are depicted in Fig. 4.17. Simulated runoff is directly output by WRF and its seasonal means and

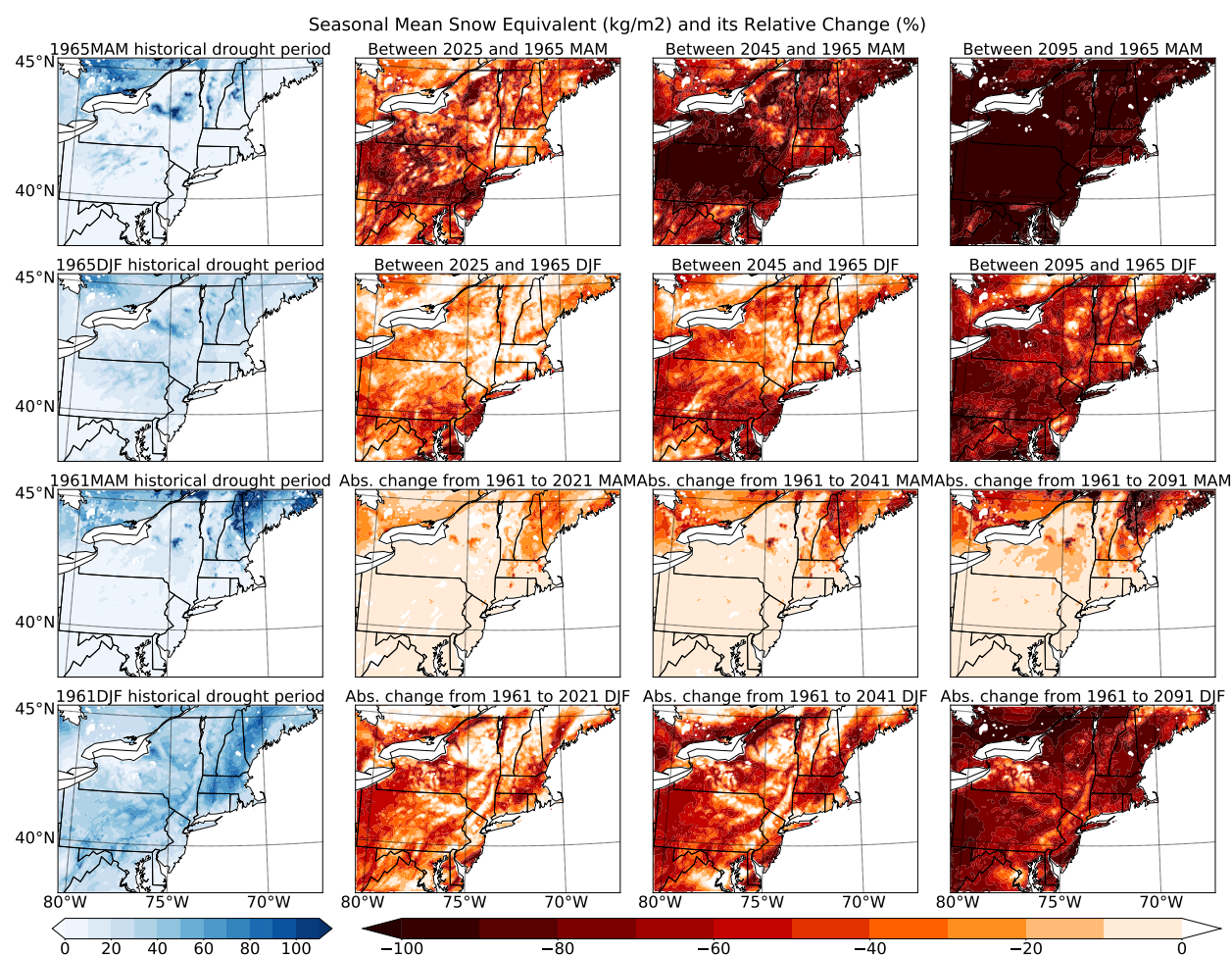


Figure 4.16: Seasonal mean snow water equivalent and its relative change over the 1965 dry exemplar and 1961 moderate exemplar, and projected changes in their future analogues for the JJA and DJF seasons.

future change shown in Fig. 4.18. Unsurprisingly, soil moisture trends upwards in accordance with net precipitation, and the long term monthly mean soil moisture will keep increasing for every month during both moderate and dry period from historical period to the end of this century (2090s). Both dry and moderate periods have more soil moisture near the coast, however a significant increase can also be found during the moderate periods to the northeast. Although both net precipitation and soil moisture are generally increasing, surface runoff exhibits a decreasing trend in some regions, particularly during the dry periods, which we attribute to increasing snowmelt and frozen ground degradation and not obviously increased net precipitation (as discussed in section 4.4.5).

Frozen ground degradation

Degradation of frozen ground is apparent for both the soil moisture and runoff fields regardless of time period. Frozen ground refers to permanently or seasonally frozen soil moisture, and can be assessed in terms of the number of freezing days (Zhang et al., 2003). Freezing of soil moisture drives up soil impermeability and reduces hydraulic conductivity, leading to a decline in soil infiltration and increase in surface runoff (more information on soil permeability in WRF-CLM4 is given in supplement 4.6.1). On the other hand, frozen ground degradation increases soil infiltration and reduces surface runoff. Frozen ground degradation is triggered by a loss of snowpack and reduction in freezing days, both of which are anticipated in a warmer climate. We argue that, particularly in DJF and MAM, frozen ground degradation is even more important for affecting soil moisture and runoff than net precipitation.

First, we observe that, compared with the dry years (Fig. 4.15 and 4.14), moderate years experience a significant summertime net precipitation decrease and soil moisture increase simultaneously in the northeast (particularly in Canada). Certainly this would appear contradictory if net precipitation was the only driver of soil moisture change. However, the discrepancy can instead be explained by increases in snowmelt accompanied by frozen ground degradation, leading to greater infiltration to soil. Because historical snowpack was essentially zero in this region during the dry period, absolute decreases in snowpack and their recharge to soil moisture are also low in the future periods (Fig. 4.15). But during the moderate period, abundant historical snowpack was

present over the same region (Fig. 4.15), resulting in far more snowmelt in spring and summer, and greater soil recharge and surface runoff under a warming climate. This extra recharge from snowmelt also explains why, during the moderate period, the surface runoff decrease is much smaller than during the dry period (Fig. 4.18). It also illustrates why over the northeast, dry periods have a greater net precipitation increase but lower soil moisture increase.

What's more, our hypothesis that frozen ground degradation has essential impacts on moistening of the soil is evinced with the fact that there exists a pretty strong negative correlation between regional mean soil moisture change and freezing days change (-0.88) that is even larger than its correlation with regional mean net precipitation change (0.79) during the winter season of dry periods over the inner domain. And in a multivariate linear regression model, freezing degree days and net precipitation change are together strong predictors of soil moisture delta ($R^2 > 0.85$).

Shifting runoff seasonality

In general, regions whose historical temperatures are just below 0°C are the most vulnerable to frozen ground degradation, as any enhancement in temperature would prevent freezing of soil moisture. As the soils of these regions then permit greater infiltration, they are also the regions in our simulations that experience the greatest decrease in surface runoff. As a result, frozen ground degradation leaves a clear seasonal signature in the runoff field: In Fig. 4.18 it is apparent that the regions with the most surface runoff decrease do overlap with regions of historically seasonally frozen soil, at lower latitudes in DJF and higher latitudes during MAM.

The most obvious decrease in runoff under dry conditions occurs in our New England subregion, also the most populated subregion of our domain (Fig. 4.18), and one with significant surface water demand. Loss of snowpack and frozen ground degradation here can be implicated in producing lower runoff and more infiltration, especially in the late winter and early spring. Examination of the long term monthly mean runoff change (Fig. 4.19) confirms our claim that the largest runoff decrease is in early spring when SWE and frost days are most reduced into the future. Over the dry period, the inner domain produces the largest historical regional monthly mean runoff in March due to abundant recharge from snowmelt; however, by end-of-century, March monthly mean runoff is

reduced from 0.330 to 0.155 $\mu\text{m}/\text{day}$ (more than a 50% loss). In fact, by end-of-century the surface regional runoff in the dry period peak moves from March to August in response to increasing summer precipitation. The springtime decrease in runoff is even more obvious within the New England subregion (Fig. 4.27), where reductions in frozen days and snow water equivalent are more pronounced (Fig. 4.16). Shifting of surface runoff away from spring has important consequences for agriculture – as discussed in section 4.2, the water shortage from the 1960s drought was at its most severe in the early spring due to agricultural demands.

4.4.8 Drought indices

From our earlier analysis, a generally warming climate with greater precipitation, evaporation and snowmelt are likely for a future analogue to the 1960s drought. However, overall wetter mean conditions doesn't necessarily imply that such a drought comes with fewer challenges. After all, the impacts of drought are complex and the product of multiple variables. Given wetter conditions are accompanied by increased temperatures and evapotranspiration, which in turn magnify the need for water, it's important to consider compound indices of drought as applied to historical and future conditions.

Standardized Precipitation Index (SPI) is a widely used family of drought indicators designed to capture the intensity of meteorological drought conditions (Hayes et al., 2002; Svoboda and Fuchs, 2016). Specifically, the metrics SPI_n quantify the accumulated departure from the mean of n consecutive months' accumulated precipitation. Smaller values of n are relevant for short-term droughts and larger values for long-term droughts. However, a key limitation of the basic SPI metric is that it cannot account for evapotranspiration, preventing it from capturing moisture demand, and making it unsuitable for detecting flash droughts. Therefore, here we examine a modified version of SPI which instead uses net precipitation in place of actual precipitation (hereafter referred to as standardized net precipitation index, SNPI). We choose not to employ Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010) since WRF-CLM4 provides an accurate and internally-consistent version of evapotranspiration directly (Lawrence et al., 2011; Xu et al., 2020), whereas SPEI would require an empirical calculation of potential evapotranspiration

4.4. A SIMULATION OF PRESENT AND FUTURE ANALOGUES OF THE 1960S DROUGHT

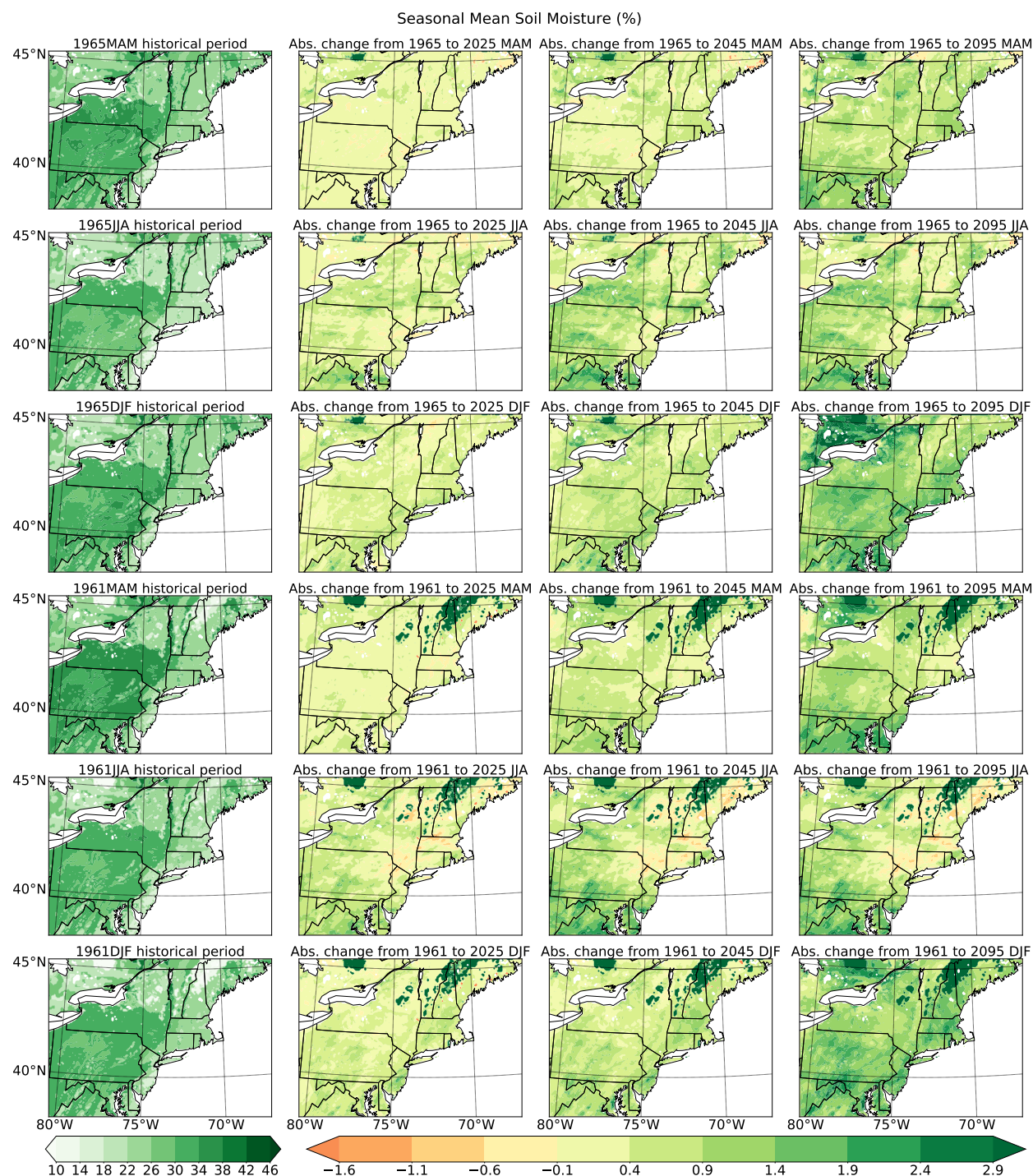


Figure 4.17: Seasonal mean soil moisture change (%) over the 1965 dry exemplar and 1961 moderate exemplar, and projected changes in their future analogues for the MAM, JJA, and DJF seasons.

4.4. A SIMULATION OF PRESENT AND FUTURE ANALOGUES OF THE 1960S DROUGHT

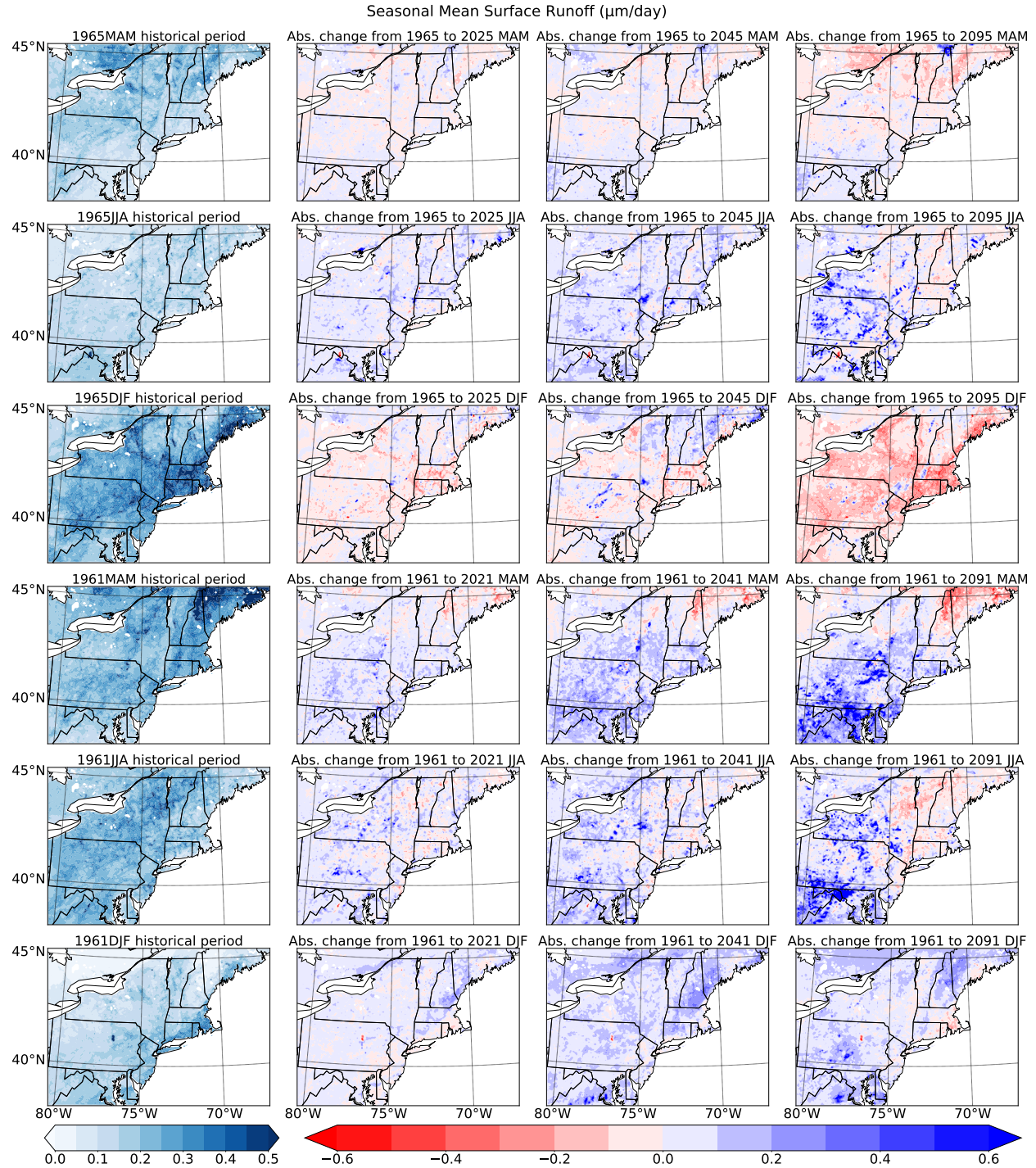


Figure 4.18: Seasonal mean runoff change (mm/day) over the 1965 dry year and 1961 moderate year, and projected changes in their future analogues for the MAM, JJA, and DJF seasons.

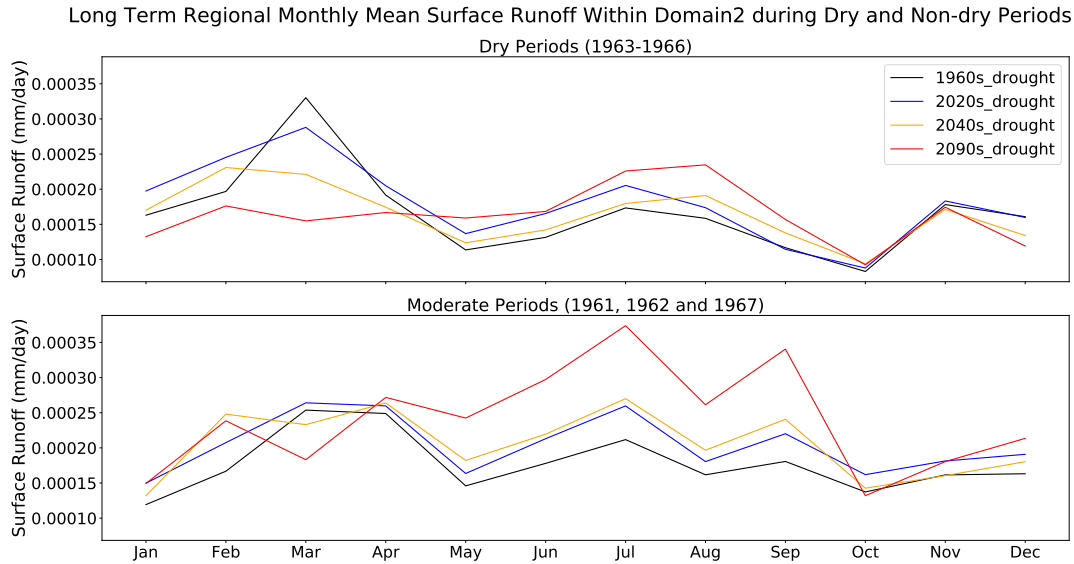


Figure 4.19: Regional long-term monthly mean runoff within the inner domain during dry years and moderate years.

(ET0). Past work has also illustrated that over sufficiently wet regions SNPI and SPEI are largely indistinguishable (Beguería et al., 2014; Joetzjer et al., 2012). Details on the calculation of SNPI are provided in section 3.2. Our interpretation of SNPI values is analogous to the interpretation of SPI given in Table 2.1 (Guttman, 1999). Note that in this study, the calibration of SNPI is based on the historical data from 1910 to 2010. Our rationale for only using the historical period to calibrate the SNPI is to prevent our future drought projections from impacting the SNPI of the historical drought. To ensure our simulation data are consistent with CERA-20C R7, WRF-CLM4 simulations are corrected by adding the regional mean differences between CERA-20C R7 and WRF over the historical period (1961-1967).

To begin, trends of long-term drought conditions are examined using SNPI24. The reason why we use SNPI24 here is to examine changes in long-term drought first, since the extended drought period is the most significant feature of the historical 1960s drought. Then we will then use SNPI1 to examine the short-term features and questions around the emergence of flash droughts. A similar analysis has been performed using SNPI6 and SNPI12 and conclusions analogous to those

of SNPI24. The 1960s drought is clearly visible in Fig. 4.20 (top), and appears as the driest period in the past 100 years. The year 1965 exhibits the lowest annual mean SNPI24 value, in accord with the claim that 1965 was the driest of the past century. These results validate the use of SNPI and its effectiveness for identifying drought conditions. Looking to the future, although both precipitation and evapotranspiration increase substantially, annual mean regional SNPI24 at 2025 is only about -1, barely classifying as a drought. Under further warming, 2045 actually becomes anomalously wet – with SNPI24 in 2041 actually surpassing any historical value of SNPI24. At the end of this century (2090s), SNPI24 in every year is larger than 2, indicative that even under the same dynamical conditions of the 1960s, the climate will be unprecedented compared to historical. These results generally suggest that the threat from long-term meteorological drought over the next century will be greatly diminished.

Although the climatological shift towards wetter conditions will mitigate long-term drought, we can still ask if extreme drought conditions are similarly mitigated on shorter time scales? In fact, our simulations suggest the answer is “probably not.” Specifically, we calculate the SNPI1 of 1960s historical drought period (1963-1966) with that of the three future drought scenarios. Fig. 4.20 (bottom) clearly shows that even as wetter months experience enhanced net precipitation, short-term extreme drought conditions persist. In fact, in the most extremely dry months (e.g. May 1964 and 1965), dryness is largely unchanged. Although the mean of SNPI1 during this period rises from -0.33 to 0.61, in accord with the general wetting trend, the standard deviation of SNPI1 also soars from 0.95 to 1.31, indicative of enhanced drought variability. This reflects enhanced climatological differences between dry and wet periods. More importantly, drought tends to happen more quickly – that is, a likely increase in the frequency *flash drought* (Christian et al., 2019). For example, April 2094 has a SNPI1 larger than 2 followed by a sudden drop to less than -3 in May; extremely dry conditions develop from extremely wet conditions in only one month! Suddenly adapting to dry conditions in such a short time would be an immense challenge for the region’s water managers and stakeholders.

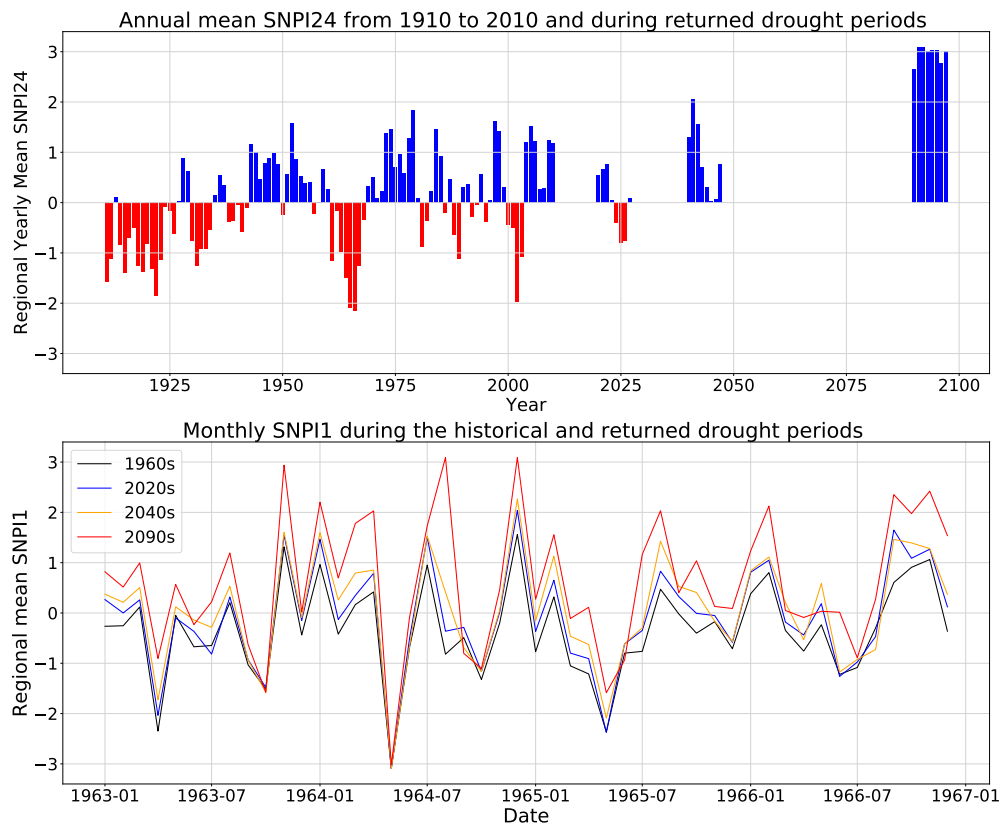


Figure 4.20: Regional annual mean SNPI24 and monthly mean SNPI1 within the New England subregion over historical and future periods.

4.5 Conclusions

In this study, the unprecedented 1960s NEUS drought is simulated as it occurred historically and subject to anticipated climate change from the early (2020s), middle (2040s) and late (2090s) 21st century. To do so, the pseudo-global warming methodology is employed in WRF-CLM4: dynamical boundary conditions are identical to the historical period, while thermodynamics (atmospheric temperature, sea surface temperature and greenhouse gas concentration) are modified using the mean of four highly performant CMIP6 models under SSP585. Overall, our simulations reveal that although there is a significant wetting trend due to the overall increase in net precipitation (precipitation minus evapotranspiration) and moistening of the soil, this wetting is only apparent during non-dry months, while dry months with negative net precipitation are generally unchanged. This enhanced hydrologic variability has the potential to accelerate the development of drought, and make it possible for an extreme flash drought to rapidly emerge from wet conditions. Further, additional socioeconomic challenges will arise because of surges in extreme hot days, unprecedented extreme heavy precipitation events, obvious shifts of climate patterns, and far less runoff in early spring as a result of frozen ground degradation and loss of snowpack. Our main findings are as follows:

- Compared to the 1960s period, the annual regional warming overland is 1.92-2.01°C in the 2020s, 3.16-3.27°C in the 2040s, and 6.74-6.87°C in the 2090s. Regional inhomogeneities in warming are primarily caused by snow-albedo feedback, which causes cold regions to warm faster than warm regions, particularly in the winter season. Extreme heat days surge in frequency in all years, from less than 10 historical, to 7-30 in the 2020s, 14-40 in the 2040s, and 52-79 in the 2090s. Further, drier years tend to experience greater increases in mean and extreme annual maximum daily temperature.
- A clear positive trend in annual mean precipitation emerges (Fig. 4.4), with increases of 15%, 27% and 70% per year in the 2020s, 2040s, and 2090s compared to historical. These precipitation increases are highly correlated to historical precipitation totals. Even accounting for increased evapotranspiration, most regions maintain a positive net precipitation change.

However, extremely dry conditions remain a problem in the future as precipitation increases are only apparent in wet months. Net precipitation variability increases also lead to greater risk of sudden onset or flash drought, which poses significant risks for agriculture and could require greater lead times when it comes to water planning.

- Increases in precipitation amount and variability are indicative of increased risk from extreme storm events. Probability of annual maximum precipitation in excess of 100 mm/day increases from 7.16% (6.28%) in the 1960s to 25.19% (36.45%) in the 2090s during dry (moderate) periods. We observe 99th percentile regional mean precipitation intensifies by more than 58% and 51% by end of century during moderate and dry periods, respectively. Extreme precipitation brings with it a high risk of flooding, and suggests a need for investments in protective infrastructure.
- Wintertime warming in the region leads to 60% fewer freezing days and 75% less snow water equivalent when comparing the 2090s to the 1960s. This in turn drives frozen ground degradation, which produces increased soil infiltration and reduced surface runoff. The transition from snow to rain further leads to spring snowpack essentially disappearing by the end of century, and further reduces spring surface runoff and water availability.
- Although net precipitation increases are also a culprit in increased soil moisture and surface runoff, we argue that frozen ground degradation plays a larger role here. Supporting this claim, summers of the moderate period feature a northeastern region with lower net precipitation in the future, but a significant increase in soil moisture. However, if only net precipitation were implicated in moister soil we would expect a decrease in soil moisture. Nonetheless it's clear that reduced snowpack and fewer freezing days permit greater infiltration from snowmelt – in fact, during the winter season of dry periods, a strong negative correlation (-0.88) emerges between the change in the regional mean number of freezing days and soil moisture.
- Surface runoff is projected to decrease in the winter and spring because of reduced snowmelt and greater soil infiltration from frozen ground degradation. March surface runoff is expected

to decrease more than 50% in the 2090s dry period, which may cause issues from insufficient early spring water supply. Our results suggest the growing season will be the most vulnerable to anticipated future changes, particularly if coupled with exceptionally warm temperatures.

This study primarily focuses on seasonal and regional scale changes, but ignores the consequences of particular weather events that occur on finer temporal and spatial scales. Given that the finest spatial and temporal resolution of our simulation is 9 km and 6 hours, our dataset could enable deeper exploration into specific events, along with their underlying process drivers. For example, this data could enable a better understanding of the strongest hurricane during 1960s period – Hurricane Donna – and its manifestation in the future in this region. Questions also remain about the potential for flash drought in this region under more general dynamical conditions, and how the 1960s drought compares to potentially more extreme droughts of the future for this region. Finally, given the simplifications made in CLM, there are substantial uncertainties in historical and projected surface and groundwater hydrology in these simulations; consequently it would be insightful to examine the response of a process-based hydrologic model to forcing data from these simulations.

4.6 Supplement

4.6.1 Soil degradation in WRF-CLM4

In the land model we used (WRF-CLM4), the soil infiltration factor is defined by equation (4.1). We can see that, when there are fewer freezing days, for each soil layer i , the ice contents ($w_{ice,i}$) will decrease and the liquid water contents ($w_{liq,i}$) will increase. Consequently, the impermeable fraction $f_{frz,i}$ will decrease as well, indicating that the soil layers are less impermeable and there will be more infiltration to recharge the soil moisture (Oleson et al., 2010).

$$f_{frz,i} = \frac{\exp[-\alpha(1 - \frac{w_{ice,i}}{w_{ice,i}+w_{liq,i}})] - \exp(-\alpha)}{1 - \exp(-\alpha)} \quad (4.1)$$

where $f_{frz,i}$ is the impermeable fraction which impacts the infiltration capacity, $w_{ice,i}$ and $w_{liq,i}$ ($kg \times m^{-2}$) are the ice and liquid water contents of soil layer i . $\alpha = 3$ is a scale-dependent parameter.

Table 4.2: Basic information for each reanalysis dataset investigated.

Product	Time coverage	Temporal resolution	Spatial resolution
The Japanese 55-year Reanalysis (JRA55)	1958-2012	3 hours	$0.5625^\circ \times 0.5625^\circ$
The Coupled ECMWF Re-Analysis system of the 20th-century (CERA-20C)	1901-2010	3 hours	$1.125^\circ \times 1.125^\circ$
NOAA-CIRES 20th Century Reanalysis version 2c (20CRv2c)	1851-2012	3 hours	$2^\circ \times 2^\circ$
NCEP/NCAR Reanalysis I	1948-present	6 hours	$2.5^\circ \times 2.5^\circ$

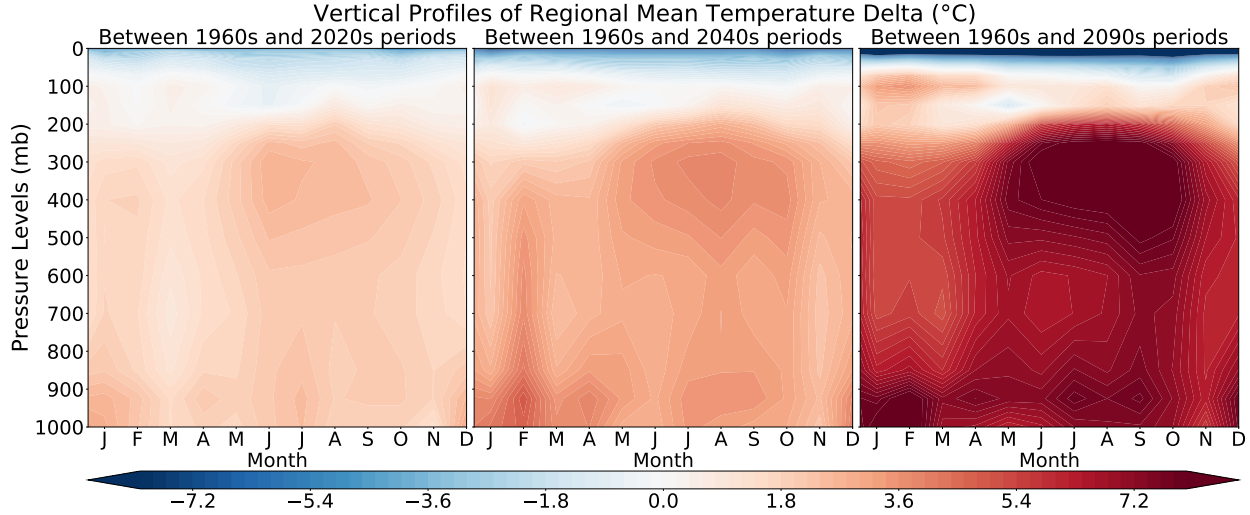


Figure 4.21: Vertical profiles of multi-model regional monthly mean temperature delta between 2020s (2020-2028), 2040s (2040-2048) and 2090s (2090-2098) periods versus the 1960s historical periods (1960-1968) based on 4 CMIP6 models with RCP8.5 emission scenario.

4.6.2 Historical reanalysis precipitation data performance

In order to provide reliable historical lateral conditions, here we evaluate the performance of 4 reanalysis datasets (shown as Table 4.2) which cover the 1960-1967 period in simulating the historical monthly precipitation compared with CPC observed data. Note that here we use the multi-ensemble mean of all 10 CERA-20C ensembles to represent CERA-20C and will evaluate the performance of each of CERA-20C's ensembles later. All analysis in this section is based on monthly regional mean precipitation over the NEUS region (here defined as the land region within $285 \leq \text{longitude} \leq 290$ and $40 \leq \text{latitude} \leq 45$).

From the histogram of monthly precipitation during the overlap period (1958-2005) of the CPC observed data and the four models (Fig. 4.22) and the monthly time series precipitation during the drought (1960-1967) (Fig. 4.23), it's obvious that CERA-20C and JRA55 have much better

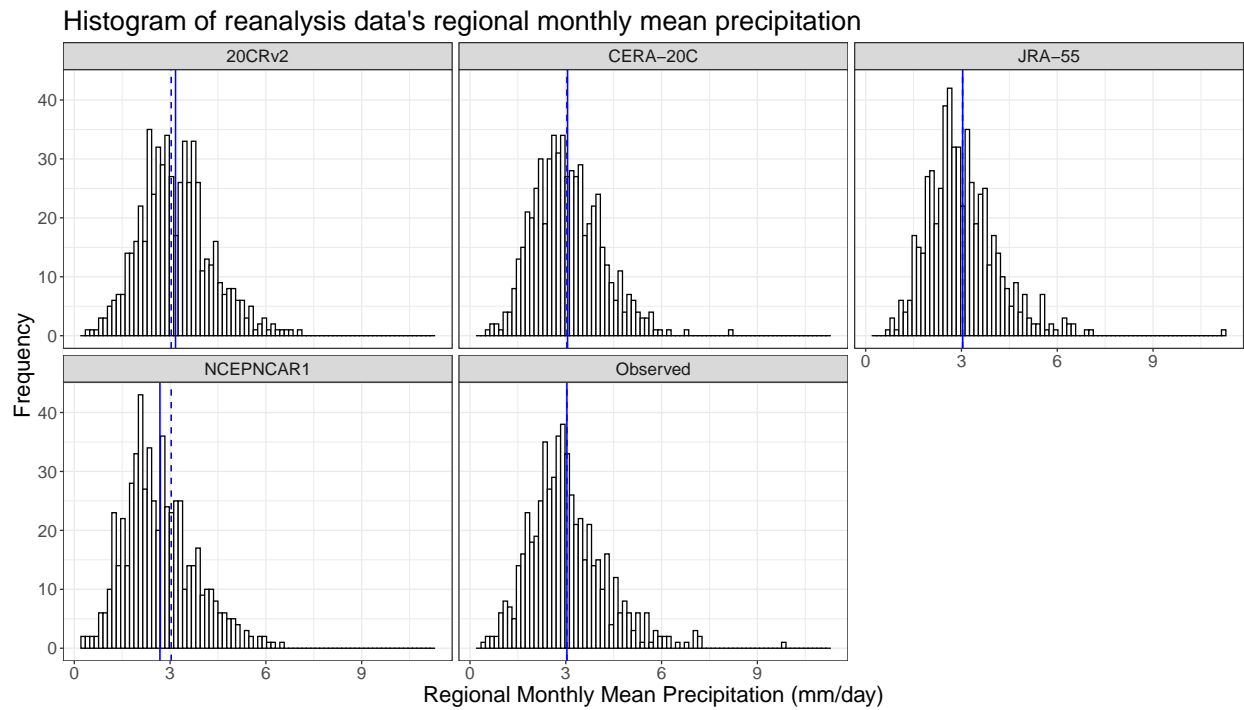


Figure 4.22: Histogram of regional monthly mean precipitation from reanalysis. The dashed line denotes the mean of the reanalysis, while the solid line denotes the mean of CPC observed data.

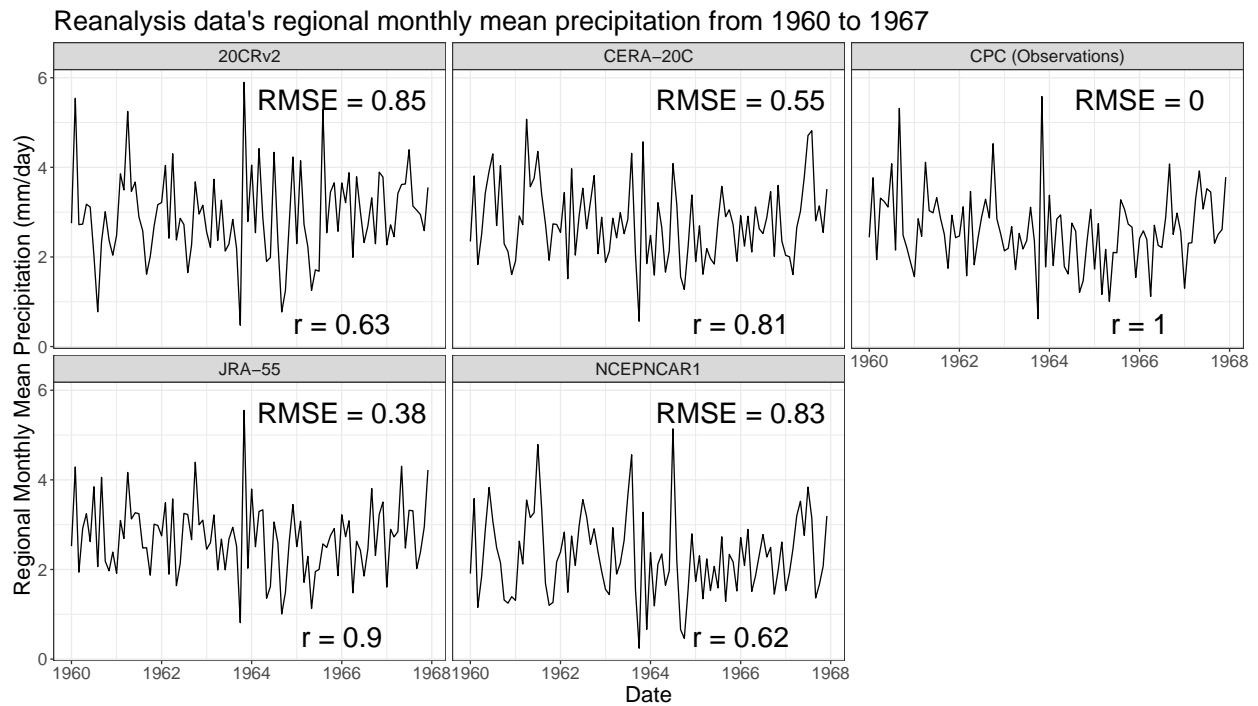


Figure 4.23: Regional monthly mean precipitation from 1960 to 1967 from various reanalysis products.

performance in simulating the mean of historical precipitation, the precipitation distribution and the historical precipitation during the 1960s period. Also, as indicated in Table 4.2, CERA-20C and JRA55 have higher spatial and temporal resolution among available products; however, because the time coverage of JRA55 is much shorter than CERA-20C, and CERA-20C is a more recent product, we choose to use CERA-20C instead of JRA55 to provide the boundary conditions of our simulations.

CERA-20C has in total 10 ensembles (R0 to R9) with different initial conditions. Because computational capacity limits us to only run simulations with a single set of lateral conditions, we downselect to a single CERA-20C ensemble. From Fig. 4.24, we can see that all ensembles have very similar and good performance compared with CPC observed data. In order to select one for use, we define dry months as those with regional mean precipitation less than the 10th percentile of regional monthly precipitation of CPC observed data. Then we calculate each ensemble's dry months during 1950-2005 period and calculate each ensemble's absolute difference with CPC observed data (Fig. 4.25), we find CERA-20C R7 has the most similar number of dry months as CPC observed data. Therefore, CERA-20C R7 is selected to provide historical boundary conditions in our simulations.

4.6.3 Heat Index equation

In the Heat Index equation (4.2), both Heat Index and T (nominally ambient dry-bulb temperature, although here we use regional mean daily maximum 2m temperature) are in degrees Celsius and R is relative humidity in percentage value (from 0 to 100).

$$\begin{aligned} \langle \text{Heat Index} \rangle = & 8.78469475556 + 1.61139411T + 2.33854883889R \\ & - 0.14611605TR - 0.012308094T^2 - 0.0164248277778R^2 \\ & + 0.002211732T^2R + 0.00072546TR^2 - 0.000003582T^2R^2 \end{aligned} \quad (4.2)$$

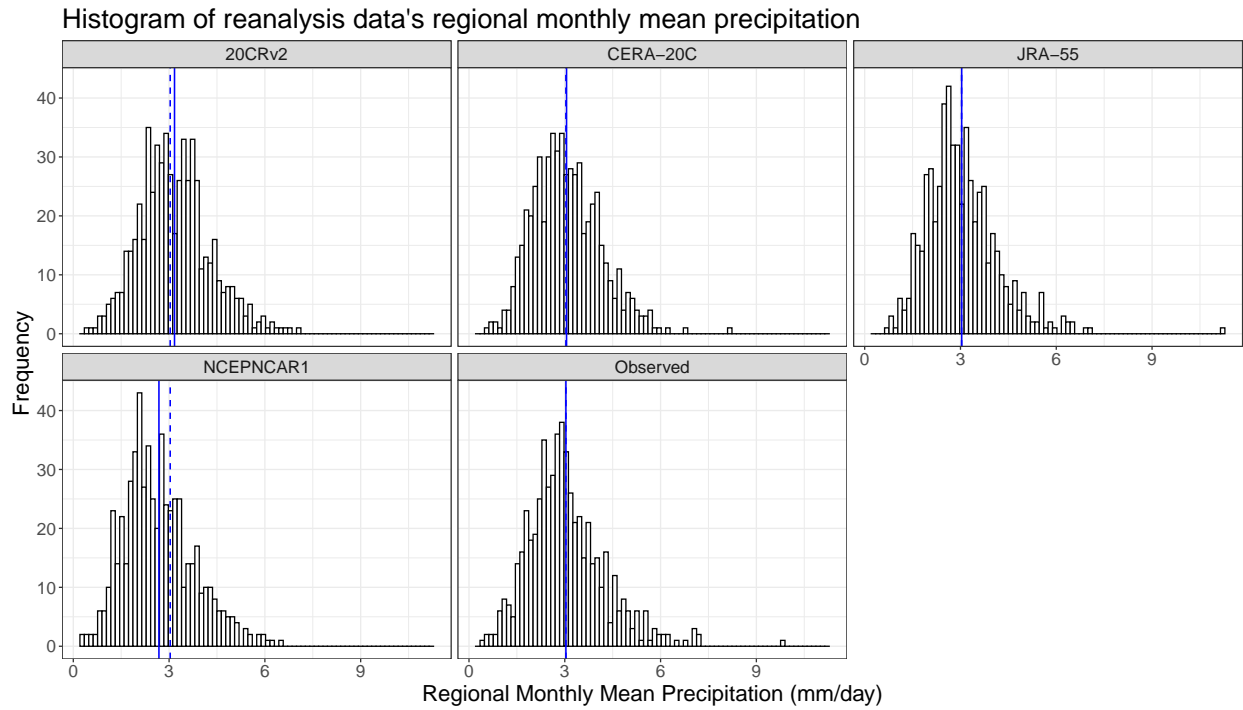


Figure 4.24: Histogram of each CERA-20C ensemble's regional monthly mean precipitation. The dashed line is the mean of each CERA-20C ensemble and the solid line the mean of the CPC observations.

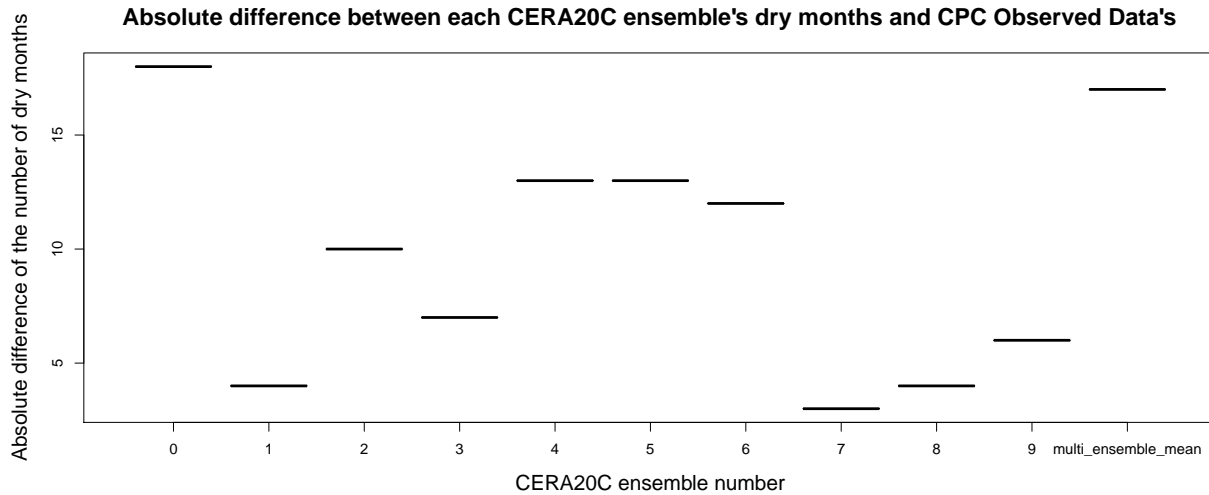


Figure 4.25: Absolute difference between each CERA-20C ensemble's dry months and CPC observed data's dry months. A dry month is defined by the 10th percentile of CPC observed regional monthly mean precipitation.

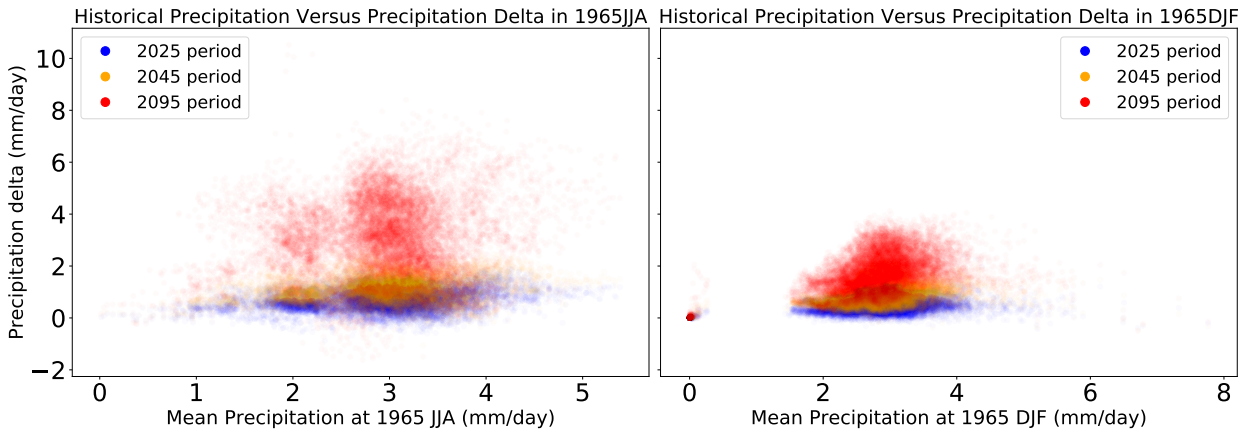


Figure 4.26: The relationship between historical precipitation and precipitation delta in the future at each grid point.

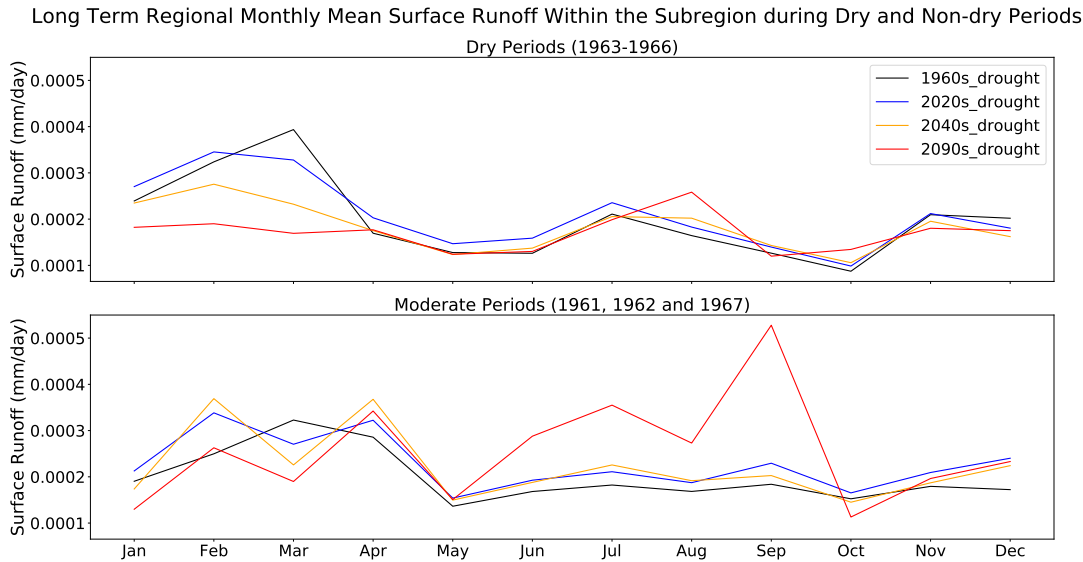


Figure 4.27: Regional long term monthly mean runoff within our subregion.

4.6.4 Simulation and Spin-up Validation

In this section, we briefly assess model performance versus historical reanalysis data to validate the historical simulation and verify sufficiency of 1-year spin-up period. Fig. 4.28 depicts moderate (1961, 1962 and 1967) and dry (1963 - 1966) periods' wintertime (DJF) and summertime (JJA) 2-meter temperatures from our historical simulation and corresponding reanalysis driver data (CERA-20C R7). Generally, our simulation produces a temperature climatology that matches closely with the reanalysis driver, but with a much higher resolution and detailed profiles. Due to our simulation's much finer resolution (6km) compared with the CERA-20C (125km), we notice that our simulation displays more detailed geographic profiles and better captures the spatial distribution of temperature. For example, we can see from Fig. 4.28 that our simulation produces significantly lower temperatures over mountainous regions like the Mt Marcy, Mt Washington and Mt Grandfather. Similarly, Fig. 4.29 depicts moderate and dry periods' seasonal average precipitation from model output and CERA-20C. Modeled precipitation tends to match closely with reanalysis data. Although some discrepancies exist, we believe they are largely a product of the higher spatial resolution of topographic features. If we look at the regional scale, we can find that over the inner domain and subregion, the 2-meter temperature and precipitation are both well simulated, and highly correlated with reanalysis data. The correlations of regional monthly mean temperatures are more than 0.99, while precipitation correlations are more than 0.91 and 0.84 within the inner domain and subregion. This demonstrates that the meteorological features of the historical period are well represented in our simulations and validates the spin-up period we used and our future projections.

As argued above, the meteorological fields of our simulations are well represented even during the spin-up period (1960) with the help of spectral nudging; however, the soil system generally develops slowly and requires longer time to get equilibrium. And that's why we use more than 1 year spin-up period (1959 December and 1960). As in several analogous studies (Huang et al., 2016; Ullrich et al., 2018; Xu and Yang, 2012; Yang et al., 2011; Zhuo et al., 2019), a one year spin-up period is typical in Community Land Model (CLM) and WRF simulations to make the soil

reach equilibrium. For the reanalysis data we used – CERA-20C – only one or two years’ spin-up period is used to initialize the land and ocean (Laloyaux et al., 2018). Moreover, several past studies have argued that small domain size and high-quality initialization data will significantly reduce the spin-up period (Cosgrove et al., 2003; Leduc and Laprise, 2009). Considering that our domain has a limited regional scale and our simulations are initialized from CERA-20C reanalysis data, a one year spin-up appears sufficient to initialize the soil moisture. Indeed, we can see from Fig. 4.31 that our simulation appears to capture a soil moisture distribution similar to that of CERA-20C. Note that because CERA-20C uses the proportion of land as the land-mask instead of 0 or 1 (ECMWF, 2019), the coastal regions are masked for having less than a 0.5 land proportion. Notably, some detailed geographic features are captured like the high soil moisture the downstream of the Lake Champlain. Although some biases are present, we argue that within both the inner domain and subregion, the regional mean soil moisture are well simulated compared with CERA-20C (Fig. 4.32). The correlations between our simulation and reanalysis data are 0.86 and 0.81 over these two regions. It needs to be said that the difference between simulation and CERA-20C over the inner domain is partially caused the different land-mask definitions of WRF and CERA-20C (ECMWF, 2019) so that lake and coastal regions’ data cannot be well identified, as shown in Fig. 4.31. However, our simulation can still capture the historical soil moisture trend after one year and one month spin-up. Notably, there is little change in the mean soil moisture over the duration of the simulation period, indicative that a longer spin-up period would likely have little effect on the quality of the simulation.

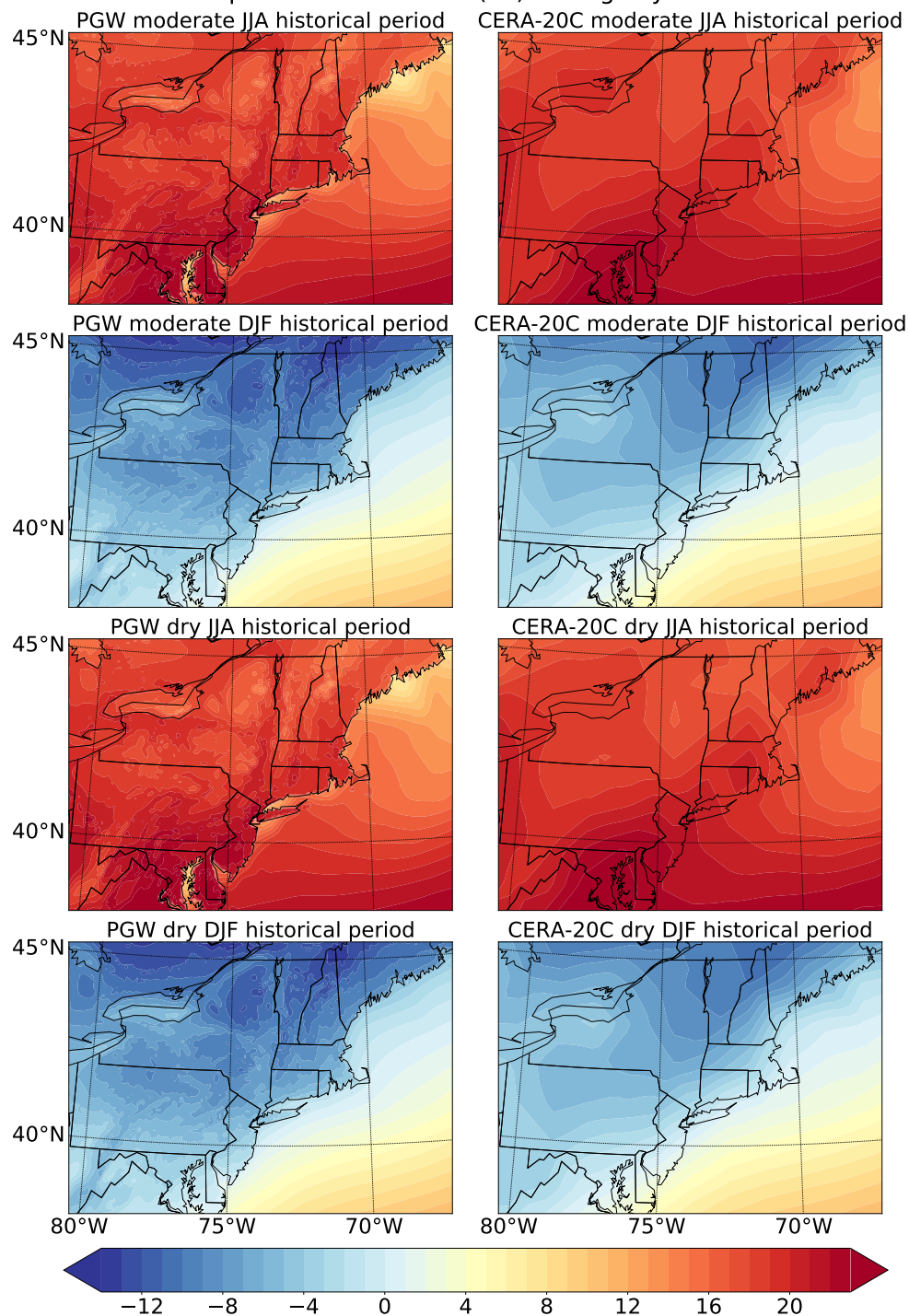
Seasonal Mean Temperature at 2 meter ($^{\circ}\text{C}$) During Dry And Moderate Periods

Figure 4.28: Simulated seasonal average 2-meter temperature ($^{\circ}\text{C}$) and reference from CERA-20C R7 during moderate and dry periods.

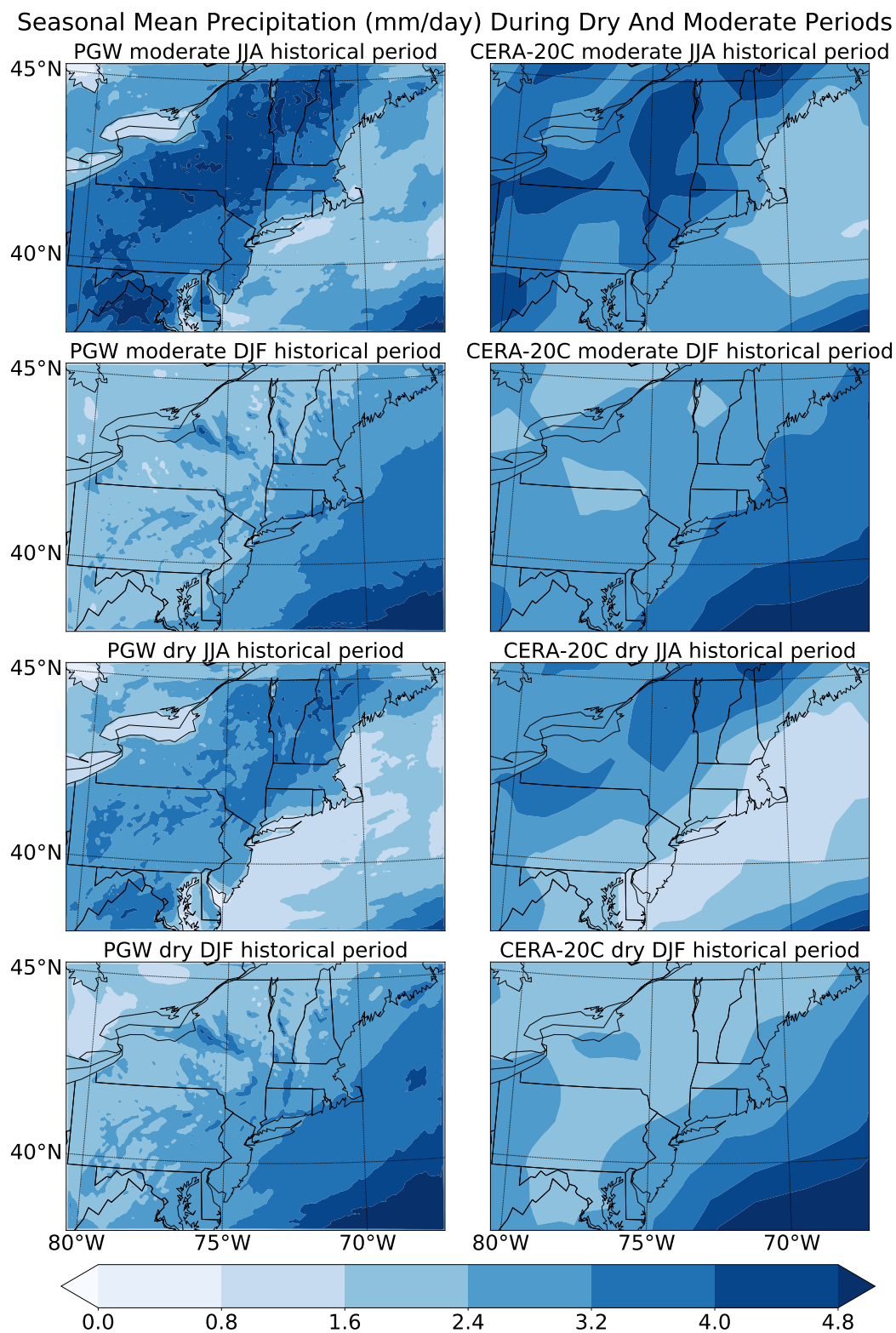


Figure 4.29: Simulated seasonal average precipitation (mm/day) and reference from CERA-20C R7 during moderate and dry periods.

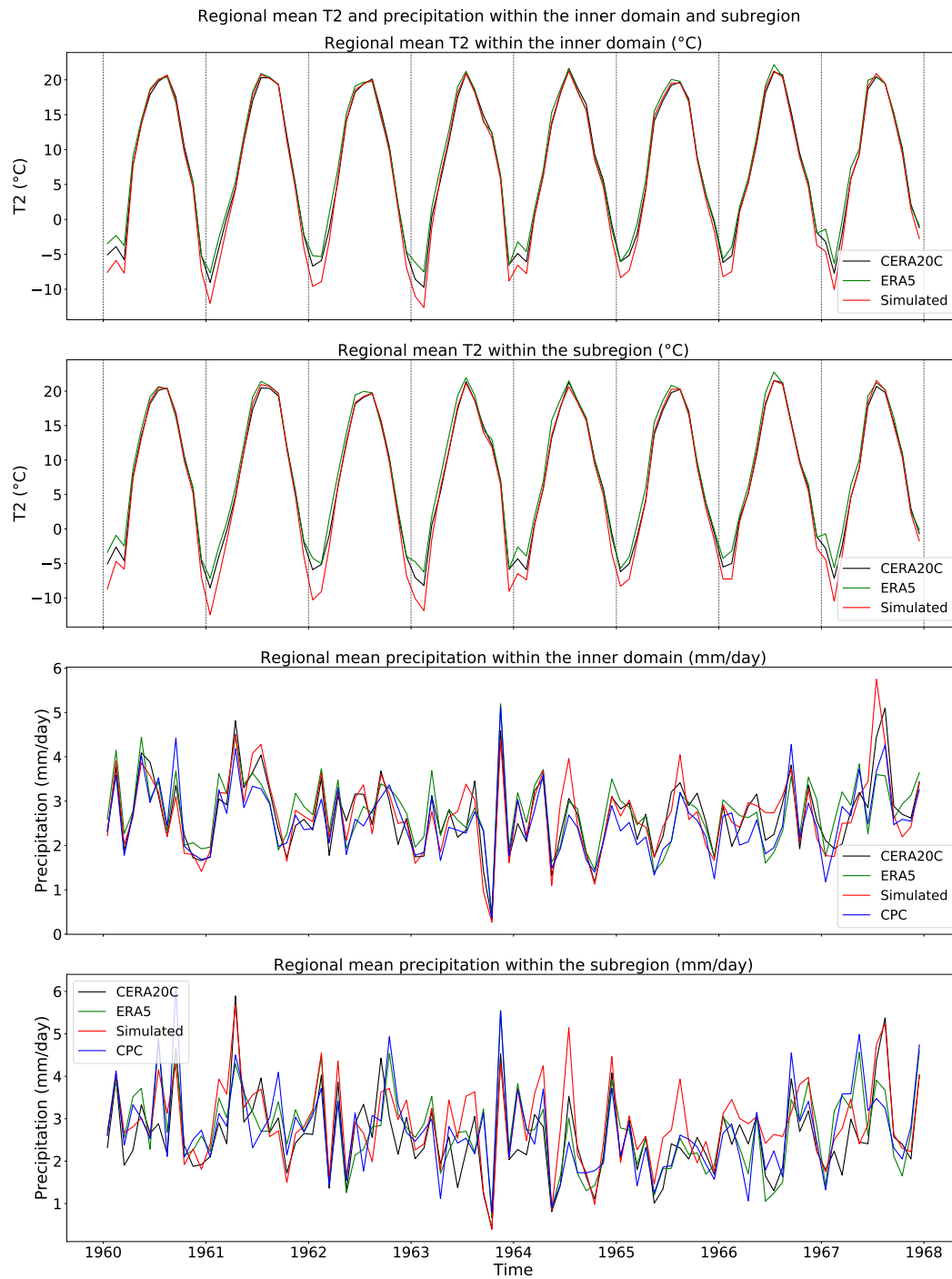


Figure 4.30: Simulated regional monthly mean 2-meter temperature (°C) and precipitation (mm/day) within the inner domain and subregion, along with reference data from CERA-20C R7, ERA5 and CPC.

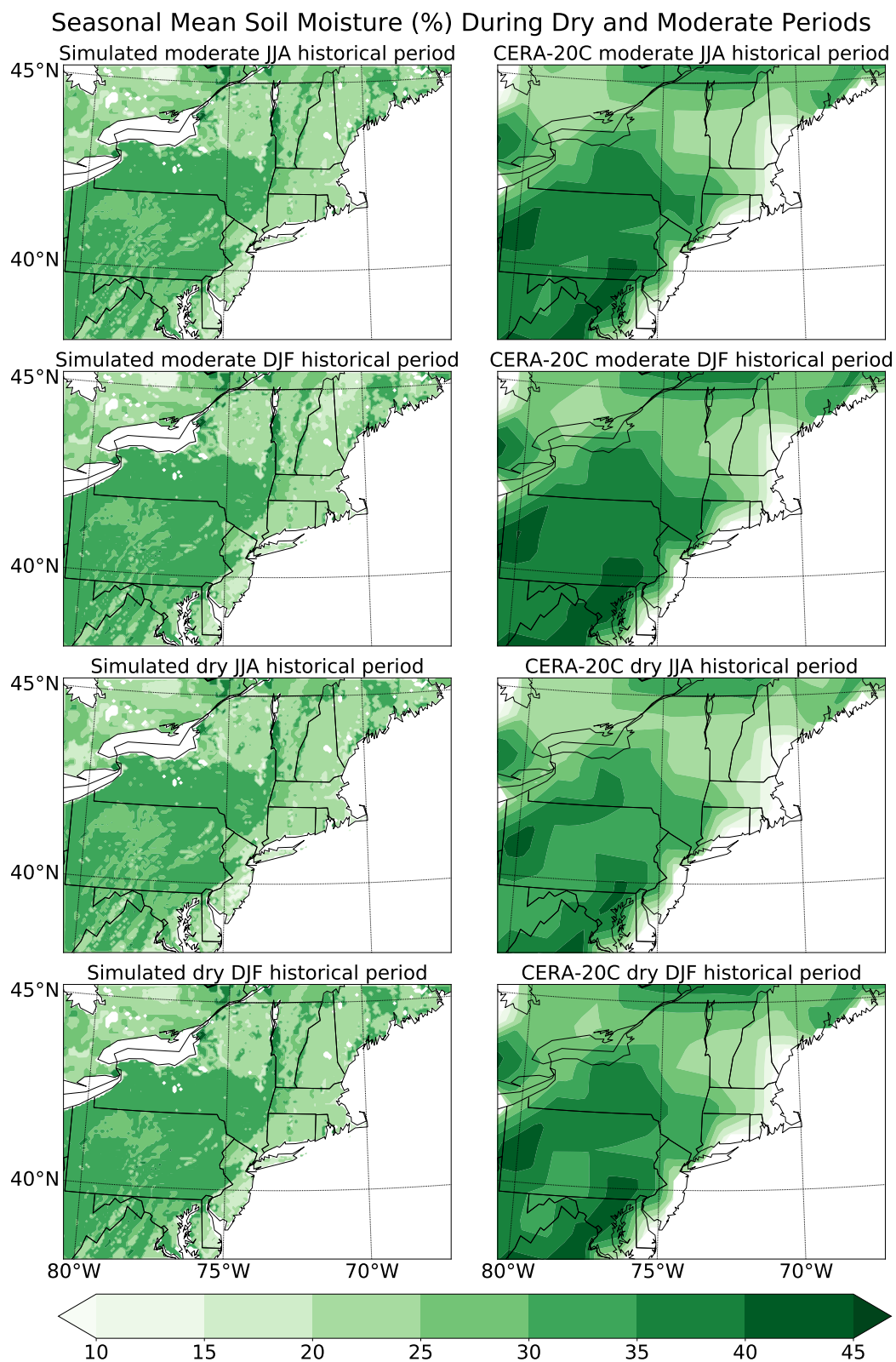


Figure 4.31: Simulated seasonal average soil moisture (%) and reference from CERA-20C R7 during moderate and dry periods.

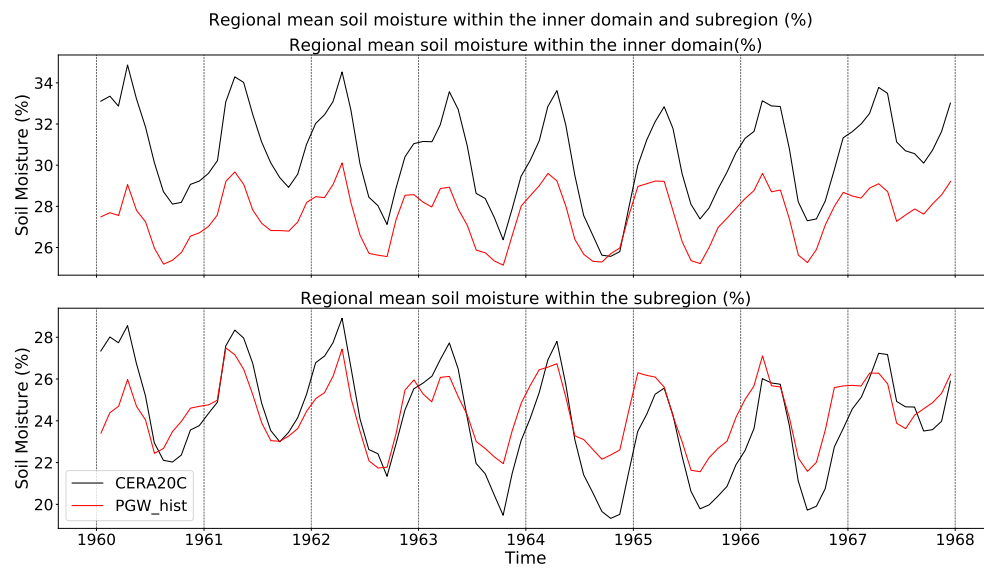


Figure 4.32: Simulated regional monthly soil moisture (%) and reference from CERA-20C R7 within the inner domain and subregion.

Chapter 5 Conclusion

This thesis aims to understand the drought nature under global warming and build a regional drought projection framework to provide extensive and more sound regional projections for policymakers and stakeholders. Overall, we put forward a drought feature-based evaluation system to help people easily examine the performance of climate models in capturing droughts over regions of interest. And we find the performance of climate models in simulating droughts varies significantly from climate datasets and study regions thus further illustrating the importance of choosing suitable climate models in drought projections. In exploring drought's changing trends over NEUS, although the current general wetting trend will continue, especially at a long-term scale, the short-term droughts and flash droughts are projected to be more frequent and magnified in the future, due to intensified evapotranspiration, increasing climate variability, and anti-correlation of precipitation and evapotranspiration. We also conclude that the extended growing season and consequent changing evapotranspiration partitioning are the main drivers of flash drought intensification. In simulating the returned 1960s drought under a warming climate, in contrast to the historical 1960s drought, similar dynamical conditions will generally produce more precipitation, increased soil moisture and evapotranspiration, and reduced snowpack. However, we also find that although wet months get much wetter, dry months may become drier, meaning that wetting trends are most significant in wet months but are essentially negligible for extremely dry months with negative monthly mean net precipitation. For these months, the trend towards wetting conditions provides little relief from the effects of extreme dry months. These conditions may even aggravate water shortages due to a rapid transition from wet to dry conditions. Other challenges emerge for residents and stakeholders in this region, including more extreme hot days, record-low snowpack, frozen ground degradation, and subsequent decreases in surface runoff.

Reference

- A. Dai, T. Z. and Chen, J.: Climate change and drought: A precipitation and evaporation perspective, *Current Climate Change Reports*, 3, 301–312, 2018.
- Adams, J.: climate-indices, an open source Python library providing reference implementations of commonly used climate indices, https://github.com/monocongo/climate_indices, 2017.
- Agel, L., Barlow, M., Qian, J.-H., Colby, F., Douglas, E., and Eichler, T.: Climatology of daily precipitation and extreme precipitation events in the northeast United States, *Journal of Hydrometeorology*, 16, 2537–2557, 2015.
- AghaKouchak, A., Feldman, D., Hoerling, M., Huxman, T., and Lund, J.: Water and climate: Recognize anthropogenic drought, *Nature News*, 524, 409, 2015.
- Ahn, M.-S., Kim, D., Kang, D., Lee, J., Sperber, K. R., Gleckler, P. J., Jiang, X., Ham, Y.-G., and Kim, H.: MJO propagation across the Maritime Continent: Are CMIP6 models better than CMIP5 models?, *Geophysical Research Letters*, 47, e2020GL087250, 2020.
- Andreadis, K. M. and Lettenmaier, D. P.: Trends in 20th century drought over the continental United States, *Geophysical Research Letters*, 33, 2006.
- Armal, S., Devineni, N., and Khanbilvardi, R.: Trends in extreme rainfall frequency in the contiguous United States: Attribution to climate change and climate variability modes, *Journal of Climate*, 31, 369–385, 2018.
- Arnell, N. W.: A simple water balance model for the simulation of streamflow over a large geographic domain, *Journal of Hydrology*, 217, 314–335, 1999.
- Backlund, P., Janetos, A., and Schimel, D.: The effects of climate change on agriculture, land resources, water resources, and biodiversity in the United States, *Synthesis and Assessment*

- Product 4.3. Washington, DC: US Environmental Protection Agency, Climate Change Science Program. 240 p., 2008.
- Balaji, V., Taylor, K. E., Jukes, M., Lawrence, B. N., Durack, P. J., Lautenschlager, M., Blanton, C., Cinquini, L., Denvil, S., Elkington, M., et al.: Requirements for a global data infrastructure in support of CMIP6, *Geoscientific Model Development*, 11, 3659–3680, 2018.
- Barksdale, H. C.: The northeast water supply crisis of the 1960’s, Tech. rep., US Government Printing Office, Reston, VA, 1968.
- Barksdale, H. C., O’Bryan, D., and Schneider, W. J.: Effect of drought on water resources in the Northeast, Tech. rep., US Government Printing Office, Reston, VA, 1966.
- Barlow, M., Nigam, S., and Berbery, E. H.: ENSO, Pacific decadal variability, and US summertime precipitation, drought, and stream flow, *Journal of climate*, 14, 2105–2128, 2001.
- Basara, J. B., Christian, J. I., Wakefield, R. A., Otkin, J. A., Hunt, E. H., and Brown, D. P.: The evolution, propagation, and spread of flash drought in the Central United States during 2012, *Environmental Research Letters*, 14, 084 025, 2019.
- Beguéría, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B.: Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring, *International Journal of Climatology*, 34, 3001–3023, 2014.
- Betts, R. A., Boucher, O., Collins, M., Cox, P. M., Falloon, P. D., Gedney, N., Hemming, D. L., Huntingford, C., Jones, C. D., Sexton, D. M., et al.: Projected increase in continental runoff due to plant responses to increasing carbon dioxide, *Nature*, 448, 1037–1041, 2007.
- Bhuyan, U., Zang, C., and Menzel, A.: Different responses of multispecies tree ring growth to various drought indices across Europe, *Dendrochronologia*, 44, 1–8, 2017.
- Blaine Friedlander : 1999 drought topped year’s most significant weather events, from hurricane to heat wave, in Northeast, <https://news.cornell.edu/stories/2000/01/1999-drought-topped-years-most-significant-weather-events-northeast>, 2000.

- Bonan, G. B. and Doney, S. C.: Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models, *Science*, 359, 2018.
- Bonfils, C., Anderson, G., Santer, B. D., Phillips, T. J., Taylor, K. E., Cuntz, M., Zelinka, M. D., Marvel, K., Cook, B. I., Cvijanovic, I., et al.: Competing influences of anthropogenic warming, ENSO, and plant physiology on future terrestrial aridity, *Journal of Climate*, 30, 6883–6904, 2017.
- Brands, S., Herrera, S., Fernández, J., and Gutiérrez, J. M.: How well do CMIP5 Earth System Models simulate present climate conditions in Europe and Africa?, *Climate Dynamics*, 41, 803–817, 2013.
- Bretherton, C. S. and Park, S.: A new moist turbulence parameterization in the Community Atmosphere Model, *Journal of Climate*, 22, 3422–3448, 2009.
- Bryant, F. B. and Yarnold, P. R.: Principal-components analysis and exploratory and confirmatory factor analysis., *Reading and Understanding Multivariate Statistics*, 1995.
- Burns, D. A., Klaus, J., and McHale, M. R.: Recent climate trends and implications for water resources in the Catskill Mountain region, New York, USA, *Journal of Hydrology*, 336, 155–170, 2007.
- Byrne, M. P. and O’Gorman, P. A.: The response of precipitation minus evapotranspiration to climate warming: Why the “wet-get-wetter, dry-get-drier” scaling does not hold over land, *Journal of Climate*, 28, 8078–8092, 2015.
- Caldwell, P., Chin, H.-N. S., Bader, D. C., and Bala, G.: Evaluation of a WRF dynamical downscaling simulation over California, *Climatic Change*, 95, 499–521, 2009.
- Case, J. L., Crosson, W. L., Kumar, S. V., Lapenta, W. M., and Peters-Lidard, C. D.: Impacts of high-resolution land surface initialization on regional sensible weather forecasts from the WRF model, *Journal of Hydrometeorology*, 9, 1249–1266, 2008.

- Cassano, J. J., Uotila, P., Lynch, A. H., and Cassano, E. N.: Predicted changes in synoptic forcing of net precipitation in large Arctic river basins during the 21st century, *Journal of Geophysical Research: Biogeosciences*, 112, 2007.
- Center, T. N. D. M.: Drought Classification of United States Drought Monitor, <https://doi.org/https://droughtmonitor.unl.edu/About/AbouttheData/DroughtClassification.aspx>, 2021.
- Changnon, S. A.: The 1988 drought, barges, and diversion, *Bulletin of the American Meteorological Society*, 70, 1092–1104, 1989.
- Chen, H. and Sun, J.: Assessing model performance of climate extremes in China: an intercomparison between CMIP5 and CMIP3, *Climatic Change*, 129, 197–211, 2015.
- Chen, L., Ford, T. W., and Yadav, P.: The Role of Vegetation in Flash Drought Occurrence: A Sensitivity Study Using Community Earth System Model, Version 2, *Journal of Hydrometeorology*, 22, 845–857, 2021.
- Chen, M., Shi, W., Xie, P., Silva, V. B., Kousky, V. E., Wayne Higgins, R., and Janowiak, J. E.: Assessing objective techniques for gauge-based analyses of global daily precipitation, *Journal of Geophysical Research: Atmospheres*, 113, 2008.
- Chou, C. and Neelin, J. D.: Mechanisms of global warming impacts on regional tropical precipitation, *Journal of Climate*, 17, 2688–2701, 2004.
- Christian, J. I., Basara, J. B., Otkin, J. A., Hunt, E. D., Wakefield, R. A., Flanagan, P. X., and Xiao, X.: A methodology for flash drought identification: Application of flash drought frequency across the United States, *Journal of Hydrometeorology*, 20, 833–846, 2019.
- Christian, J. I., Basara, J. B., Hunt, E. D., Otkin, J. A., Furtado, J. C., Mishra, V., Xiao, X., and Randall, R. M.: Global distribution, trends, and drivers of flash drought occurrence, *Nature communications*, 12, 1–11, 2021.
- Christiansen, D. E., Markstrom, S. L., and Hay, L. E.: Impacts of climate change on the growing season in the United States, *Earth Interactions*, 15, 1–17, 2011.

- Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., Mu, M., and Randerson, J. T.: The international land model benchmarking (ILAMB) system: design, theory, and implementation, *Journal of Advances in Modeling Earth Systems*, 10, 2731–2754, 2018.
- Committee, L. M. R. C. et al.: The economic profile of the Lower Mississippi River: an update, 2014.
- Cook, E. R. and Jacoby, G. C.: Tree-ring-drought relationships in the Hudson Valley, New York, *Science*, 198, 399–401, 1977.
- Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Sheffield, J., Duan, Q., Luo, L., et al.: Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS), *Journal of Geophysical Research: Atmospheres*, 108, 2003.
- Crowley, J. W., Mitrovica, J. X., Bailey, R. C., Tamisiea, M. E., and Davis, J. L.: Land water storage within the Congo Basin inferred from GRACE satellite gravity data, *Geophysical Research Letters*, 33, 2006.
- Dai, A.: Drought under global warming: a review, *Wiley Interdisciplinary Reviews: Climate Change*, 2, 45–65, 2011.
- Dai, A., Trenberth, K. E., and Qian, T.: A global dataset of Palmer Drought Severity Index for 1870–2002: Relationship with soil moisture and effects of surface warming, *Journal of Hydrometeorology*, 5, 1117–1130, 2004.
- Dai, A., Zhao, T., and Chen, J.: Climate change and drought: A precipitation and evaporation perspective, *Current Climate Change Reports*, 4, 301–312, 2018.
- Darand, M. and Sohrabi, M. M.: Identifying drought-and flood-prone areas based on significant changes in daily precipitation over Iran, *Natural Hazards*, 90, 1427–1446, 2018.

- DeAngelis, A. M., Wang, H., Koster, R. D., Schubert, S. D., Chang, Y., and Marshak, J.: Prediction skill of the 2012 US Great Plains flash drought in subseasonal experiment (SubX) models, *Journal of Climate*, 33, 6229–6253, 2020.
- DeGaetano, A. T.: A temporal comparison of drought impacts and responses in the New York city metropolitan area, *Climatic Change*, 42, 539–560, 1999.
- Demaria, E. M., Palmer, R. N., and Roundy, J. K.: Regional climate change projections of stream-flow characteristics in the Northeast and Midwest US, *Journal of Hydrology: Regional Studies*, 5, 309–323, 2016.
- Deser, C., Knutti, R., Solomon, S., and Phillips, A. S.: Communication of the role of natural variability in future North American climate, *Nature Climate Change*, 2, 775–779, 2012.
- Deser, C., Phillips, A. S., Alexander, M. A., and Smoliak, B. V.: Projecting North American climate over the next 50 years: Uncertainty due to internal variability, *Journal of Climate*, 27, 2271–2296, 2014.
- Deser, C., Lehner, F., Rodgers, K., Ault, T., Delworth, T., DiNezio, P., Fiore, A., Frankignoul, C., Fyfe, J., Horton, D., et al.: Insights from Earth system model initial-condition large ensembles and future prospects, *Nature Climate Change*, pp. 1–10, 2020.
- Diaz, H. F.: Drought in the United States, *Journal of Applied Meteorology and Climatology*, 22, 3–16, 1983.
- Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., Hanasaki, N., et al.: The second Global Soil Wetness Project (GSWP-2): Multi-model analysis and implications for our perception of the land surface, *Bull. Amer. Meteor. Soc*, 2005.
- Donat, M. G., Lowry, A. L., Alexander, L. V., O’Gorman, P. A., and Maher, N.: More extreme precipitation in the world’s dry and wet regions, *Nature Climate Change*, 6, 508–513, 2016.
- Dullaart, J. C., Muis, S., Bloemendaal, N., and Aerts, J. C.: Advancing global storm surge modelling using the new ERA5 climate reanalysis, *Climate Dynamics*, 54, 1007–1021, 2020.

REFERENCE

- Dupigny-Giroux, L.-A.: Towards Characterizing And Planning For Drought In Vermont-Part I: A Climatological Perspectwe 1, *Jawra journal of the american water resources association*, 37, 505–525, 2001.
- ECMWF: ECMWF 128 Table parameter details of the land-mask, <https://apps.ecmwf.int/codes/grib/param-db?id=172>, 2019.
- European Centre for Medium-Range Weather Forecasts: Coupled ECMWF Re-Analysis system of the 20th-century, <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/cera-20c>, 2016.
- European Centre for Medium-Range Weather Forecasts: ECMWF Reanalysis 5th Generation, <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>, 2020.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development*, 9, 1937–1958, 2016.
- Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., et al.: Taking climate model evaluation to the next level, *Nature Climate Change*, 9, 102–110, 2019.
- Ferguson, P. R. and Veizer, J.: Coupling of water and carbon fluxes via the terrestrial biosphere and its significance to the Earth’s climate system, *Journal of Geophysical Research: Atmospheres*, 112, 2007.
- Field, C. B., Jackson, R. B., and Mooney, H. A.: Stomatal responses to increased CO₂: implications from the plant to the global scale, *Plant, Cell & Environment*, 18, 1214–1225, 1995.
- Ford, T. W. and Labosier, C. F.: Meteorological conditions associated with the onset of flash drought in the eastern United States, *Agricultural and Forest Meteorology*, 247, 414–423, 2017.

- Frank, D. C., Poulter, B., Saurer, M., Esper, J., Huntingford, C., Helle, G., Treydte, K., Zimmermann, N. E., Schleser, G., Ahlström, A., et al.: Water-use efficiency and transpiration across European forests during the Anthropocene, *Nature Climate Change*, 5, 579–583, 2015.
- Frumhoff, P. C., McCarthy, J. J., Melillo, J. M., Moser, S. C., and Wuebbles, D. J. e. a.: Confronting climate change in the US Northeast: Science, impacts, and solutions., 2007.
- Ganetis, S. A. and Colle, B. A.: The thermodynamic and microphysical evolution of an intense snowband during the northeast US blizzard of 8–9 February 2013, *Monthly Weather Review*, 143, 4104–4125, 2015.
- García, M., Sandholt, I., Ceccato, P., Ridler, M., Mougín, E., Kergoat, L., Morillas, L., Timouk, F., Fensholt, R., and Domingo, F.: Actual evapotranspiration in drylands derived from in-situ and satellite data: Assessing biophysical constraints, *Remote Sensing of Environment*, 131, 103–118, 2013.
- Giorgi, F. and Mearns, L. O.: Introduction to special section: Regional climate modeling revisited, *Journal of Geophysical Research: Atmospheres*, 104, 6335–6352, 1999.
- Giorgi, F., Coppola, E., Solmon, F., Mariotti, L., Sylla, M., Bi, X., Elguindi, N., Diro, G., Nair, V., Giuliani, G., et al.: RegCM4: model description and preliminary tests over multiple CORDEX domains, *Climate Research*, 52, 7–29, 2012.
- Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, *Journal of Geophysical Research: Atmospheres*, 113, 2008.
- Gleckler, P. J., Durack, P. J., Stouffer, R. J., Johnson, G. C., and Forest, C. E.: Industrial-era global ocean heat uptake doubles in recent decades, *Nature Climate Change*, 6, 394–398, 2016.
- Gleick, P. H.: The development and testing of a water balance model for climate impact assessment: modeling the Sacramento basin, *Water Resources Research*, 23, 1049–1061, 1987.
- Gong, D., Kang, S., Yao, L., and Zhang, L.: Estimation of evapotranspiration and its components

- from an apple orchard in northwest China using sap flow and water balance methods, *Hydrological Processes: An International Journal*, 21, 931–938, 2007.
- Guilbert, J., Betts, A. K., Rizzo, D. M., Beckage, B., and Bomblies, A.: Characterization of increased persistence and intensity of precipitation in the northeastern United States, *Geophysical Research Letters*, 42, 1888–1893, 2015.
- Guttman, N. B.: Accepting the standardized precipitation index: a calculation algorithm 1, *JAWRA Journal of the American Water Resources Association*, 35, 311–322, 1999.
- Hall, A.: Projecting regional change, *Science*, 346, 1461–1462, 2014.
- Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, *Bulletin of the American Meteorological Society*, 90, 1095–1108, 2009.
- Hayes, M. J., Alvord, C., and Lowrey, J.: Drought indices, National Drought Mitigation Center, University of Nebraska, 2002.
- Hayhoe, K., Wake, C. P., Huntington, T. G., Luo, L., Schwartz, M. D., Sheffield, J., Wood, E., Anderson, B., Bradbury, J., and DeGaetano, A. e. a.: Past and future changes in climate and hydrological indicators in the US Northeast, *Climate Dynamics*, 28, 381–407, 2007.
- Hayhoe, K., Wake, C., Anderson, B., Liang, X.-Z., Maurer, E., Zhu, J., Bradbury, J., DeGaetano, A., Stoner, A. M., and Wuebbles, D.: Regional climate change projections for the Northeast USA, *Mitigation and Adaptation Strategies for Global Change*, 13, 425–436, 2008.
- Hazeleger, W., Severijns, C., Semmler, T., Ștefănescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J. M., Bintanja, R., et al.: EC-Earth: a seamless earth-system prediction approach in action, *Bulletin of the American Meteorological Society*, 91, 1357–1364, 2010.
- Heikkilä, U., Sandvik, A., and Sorteberg, A.: Dynamical downscaling of ERA-40 in complex terrain using the WRF regional climate model, *Climate Dynamics*, 37, 1551–1564, 2011.
- Heitman, J., Horton, R., Sauer, T., and DeSutter, T.: Sensible heat observations reveal soil-water evaporation dynamics, *Journal of Hydrometeorology*, 9, 165–171, 2008.

- Held, I. M. and Soden, B. J.: Robust responses of the hydrological cycle to global warming, *Journal of climate*, 19, 5686–5699, 2006.
- Henderson-Sellers, A., Zhang, H., Berz, G., Emanuel, K., Gray, W., Landsea, C., Holland, G., Lighthill, J., Shieh, S.-L., Webster, P., et al.: Tropical cyclones and global climate change: A post-IPCC assessment, *Bulletin of the American Meteorological Society*, 79, 19–38, 1998.
- Hersbach, H.: The ERA5 Atmospheric Reanalysis., in: AGU fall meeting abstracts, vol. 2016, pp. NG33D–01, 2016.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, 2020.
- Hobbs, J. J.: *World regional geography*, Nelson Education, 2008.
- Hoegh-Guldberg, O., Jacob, D., Bindi, M., Brown, S., Camilloni, I., Diedhiou, A., Djalante, R., Ebi, K., Engelbrecht, F., and Guiot, J. e. a.: Impacts of 1.5 °C global warming on natural and human systems, *Global warming of 1.5° C. An IPCC Special Report*, 2018.
- Hoell, A., Parker, B.-A., Downey, M., Umphlett, N., Jencso, K., Akyuz, F. A., Peck, D., Hadwen, T., Fuchs, B., Kluck, D., et al.: Lessons learned from the 2017 flash drought across the US Northern Great Plains and Canadian Prairies, *Bulletin of the American Meteorological Society*, 101, E2171–E2185, 2020.
- Hoerling, M. P., Eischeid, J. K., Quan, X.-W., Diaz, H. F., Webb, R. S., Dole, R. M., and Easterling, D. R.: Is a transition to semipermanent drought conditions imminent in the US Great Plains?, *Journal of Climate*, 25, 8380–8386, 2012.
- Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., Heng, Y., Konopka, P., Müller, R., Vogel, B., et al.: From ERA-Interim to ERA5: the considerable impact of ECMWF’s next-generation reanalysis on Lagrangian transport simulations, *Atmospheric Chemistry and Physics*, 19, 3097–3124, 2019.

- Huang, X., Rhoades, A. M., Ullrich, P. A., and Zarzycki, C. M.: An evaluation of the variable-resolution CESM for modeling California’s climate, *Journal of Advances in Modeling Earth Systems*, 8, 345–369, 2016.
- Huntington, T. G., Hodgkins, G. A., Keim, B. D., and Dudley, R. W.: Changes in the proportion of precipitation occurring as snow in New England (1949–2000), *Journal of Climate*, 17, 2626–2636, 2004.
- Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., et al.: The community earth system model: a framework for collaborative research, *Bulletin of the American Meteorological Society*, 94, 1339–1360, 2013.
- Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, *Journal of Geophysical Research: Atmospheres*, 113, 2008.
- Islam, M. R.: Sample size and its role in Central Limit Theorem (CLT), *International Journal of Physics & Mathematics*, 1, 37–47, 2018.
- Jagannathan, K., Jones, A. D., and Ray, I.: The making of a metric: Co-producing decision-relevant climate science, *Bulletin of the American Meteorological Society*, 2020.
- Janes, B. E. and Brumbach, J. J.: The 1964 Agriculture Drought in Connecticut, University of Connecticut, 1965.
- Jeffrey, S., Rotstayn, L., Collier, M., Dravitzki, S., Hamalainen, C., Moeseneder, C., Wong, K., and Syktus, J.: Australia’s CMIP5 submission using the CSIRO-Mk3. 6 model, *Aust. Meteor. Oceanogr. J.*, 63, 1–13, 2013.
- Jiménez, P. A., Dudhia, J., González-Rouco, J. F., Navarro, J., Montávez, J. P., and García-Bustamante, E.: A revised scheme for the WRF surface layer formulation, *Monthly Weather Review*, 140, 898–918, 2012.

- Jin, J., Miller, N. L., and Schlegel, N.: Sensitivity study of four land surface schemes in the WRF model, *Advances in Meteorology*, 2010, 2010.
- Joetzjer, E., Douville, H., Delire, C., Ciais, P., Decharme, B., and Tyteca, S.: Evaluation of drought indices at interannual to climate change timescales: a case study over the Amazon and Mississippi river basins., *Hydrology & Earth System Sciences Discussions*, 9, 2012.
- Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., De Jeu, R., et al.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, *Nature*, 467, 951–954, 2010.
- Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S., Danabasoglu, G., Edwards, J., et al.: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability, *Bulletin of the American Meteorological Society*, 96, 1333–1349, 2015.
- Keeling, R. F., Graven, H. D., Welp, L. R., Resplandy, L., Bi, J., Piper, S. C., Sun, Y., Bollenbacher, A., and Meijer, H. A.: Atmospheric evidence for a global secular increase in carbon isotopic discrimination of land photosynthesis, "Proceedings of the National Academy of Sciences", 114, 10 361–10 366, 2017.
- Keenan, T. F., Hollinger, D. Y., Bohrer, G., Dragoni, D., Munger, J. W., Schmid, H. P., and Richardson, A. D.: Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise, *Nature*, 499, 324–327, 2013.
- Kharin, V. V., Zwiers, F. W., Zhang, X., and Hegerl, G. C.: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, *Journal of Climate*, 20, 1419–1444, 2007.
- Kimura, F., Kitoh, A., Sumi, A., Asanuma, J., and Yatagai, A.: Downscaling of the global warming projections to Turkey, *The Final Report of ICCAP*, 10, 2007.

- Kirchmeier-Young, M. C., Zwiers, F. W., and Gillett, N. P.: Attribution of extreme events in Arctic sea ice extent, *Journal of Climate*, 30, 553–571, 2017.
- Knauer, J., Zaehle, S., Reichstein, M., Medlyn, B. E., Forkel, M., Hagemann, S., and Werner, C.: The response of ecosystem water-use efficiency to rising atmospheric CO₂ concentrations: Sensitivity and large-scale biogeochemical implications, *New Phytologist*, 213, 1654–1666, 2017.
- Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, *Nature Climate Change*, 3, 369–373, 2013.
- Kogan, F. N.: Global drought watch from space, *Bulletin of the American Meteorological Society*, 78, 621–636, 1997.
- Kolb, T. E., Fettig, C. J., Ayres, M. P., Bentz, B. J., Hicke, J. A., Mathiasen, R., Stewart, J. E., and Weed, A. S.: Observed and anticipated impacts of drought on forest insects and diseases in the United States, *Forest Ecology and Management*, 380, 321–334, 2016.
- Koutroulis, A., Grillakis, M., Tsanis, I., and Papadimitriou, L.: Evaluation of precipitation and temperature simulation performance of the CMIP3 and CMIP5 historical experiments, *Climate Dynamics*, 47, 1881–1898, 2016.
- Krakauer, N. Y., Lakhankar, T., and Hudson, D.: Trends in Drought over the Northeast United States, *Water*, 11, 1834, 2019.
- Lainé, A., Nakamura, H., Nishii, K., and Miyasaka, T.: A diagnostic study of future evaporation changes projected in CMIP5 climate models, *Climate Dynamics*, 42, 2745–2761, 2014.
- Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J.-R., Broennimann, S., Buizza, R., Dalhgren, P., Dee, D., Haimberger, L., Hersbach, H., et al.: CERA-20C: A coupled reanalysis of the twentieth century, *Journal of Advances in Modeling Earth Systems*, 10, 1172–1195, 2018.
- Lawrence, D. M., Thornton, P. E., Oleson, K. W., and Bonan, G. B.: The partitioning of evapotranspiration into transpiration, soil evaporation, and canopy evaporation in a GCM: Impacts on land–atmosphere interaction, *Journal of Hydrometeorology*, 8, 862–880, 2007.

- Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., et al.: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model, *Journal of Advances in Modeling Earth Systems*, 3, 2011.
- Lawrence, P. J. and Chase, T. N.: Climate impacts of making evapotranspiration in the Community Land Model (CLM3) consistent with the Simple Biosphere Model (SiB), *Journal of Hydrometeorology*, 10, 374–394, 2009.
- Leduc, M. and Laprise, R.: Regional climate model sensitivity to domain size, *Climate Dynamics*, 32, 833–854, 2009.
- Lee, D., Kim, J., Lee, K.-S., and Kim, S.: Partitioning of catchment water budget and its implications for ecosystem carbon exchange, *Biogeosciences*, 7, 1903–1914, 2010.
- Lee, J., Sperber, K. R., Gleckler, P. J., Bonfils, C. J., and Taylor, K. E.: Quantifying the agreement between observed and simulated extratropical modes of interannual variability, *Climate Dynamics*, 52, 4057–4089, <https://doi.org/10.1007/s00382-018-4355-4>, 2018.
- Li, J.-L., Xu, K.-M., Jiang, J., Lee, W.-L., Wang, L.-C., Yu, J.-Y., Stephens, G., Fetzer, E., and Wang, Y.-H.: An Overview of CMIP5 and CMIP6 Simulated Cloud Ice, Radiation Fields, Surface Wind Stress, Sea Surface Temperatures, and Precipitation Over Tropical and Subtropical Oceans, *Journal of Geophysical Research: Atmospheres*, 125, e2020JD032848, 2020.
- Li, L., Li, W., and Jin, J.: Improvements in WRF simulation skills of southeastern United States summer rainfall: physical parameterization and horizontal resolution, *Climate Dynamics*, 43, 2077–2091, 2014.
- Li, S., Kang, S., Zhang, L., Zhang, J., Du, T., Tong, L., and Ding, R.: Evaluation of six potential evapotranspiration models for estimating crop potential and actual evapotranspiration in arid regions, *Journal of Hydrology*, 543, 450–461, 2016.
- Lindsey, R. and Dahlman, L.: *Climate Change: Global Temperature*, 2020.

REFERENCE

- Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D. R., and Brekke, L.: A spatially comprehensive, hydrometeorological data set for Mexico, the US, and Southern Canada 1950–2013, *Scientific Data*, 2, 1–12, 2015.
- Lombard, P. J., Barclay, J. R., and McCarthy, D.-A. E.: 2020 drought in New England, Tech. rep., US Geological Survey, 2020.
- Lu, Y., Cohen, I., Zhou, X. S., and Tian, Q.: Feature selection using principal feature analysis, in: *Proceedings of the 15th ACM International Conference on Multimedia*, pp. 301–304, 2007.
- Lyon, B., Christie-Blick, N., and Gluzberg, Y.: Water Shortages, Development, and Drought in Rockland County, New York, *JAWRA Journal of the American Water Resources Association*, 41, 1457–1469, 2005.
- M. Svoboda, M. H. and Wood, D.: Standardized Precipitation Index User, <https://doi.org/https://public.wmo.int/en/resources/library/standardized-precipitation-index-user-guide>, 2012.
- Ma, J., Foltz, G. R., Soden, B. J., Huang, G., He, J., and Dong, C.: Will surface winds weaken in response to global warming?, *Environmental Research Letters*, 11, 124 012, 2016.
- Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., Kröger, J., Takano, Y., Ghosh, R., Hedemann, C., et al.: The Max Planck institute grand ensemble: Enabling the exploration of climate system variability, *Journal of Advances in Modeling Earth Systems*, 11, 2050–2069, 2019.
- Mankin, J. S., Seager, R., Smerdon, J. E., Cook, B. I., and Williams, A. P.: Mid-latitude freshwater availability reduced by projected vegetation responses to climate change, *Nature Geoscience*, 12, 983–988, 2019.
- Maraun, D. and Widmann, M.: The representation of location by a regional climate model in complex terrain, *Hydrology and Earth System Sciences*, 19, 3449–3456, 2015.
- McKee, T. B.: Drought monitoring with multiple time scales, in: *Proceedings of 9th Conference on Applied Climatology*, Boston, 1995, 1995.

REFERENCE

- McKee, T. B., Doesken, N. J., Kleist, J., et al.: The relationship of drought frequency and duration to time scales, in: *Proceedings of the 8th Conference on Applied Climatology*, vol. 17, pp. 179–183, Boston, 1993.
- Mearns, L.O., e. a.: The NA-CORDEX dataset, version 1.0. NCAR Climate Data Gateway, Boulder CO, <https://doi.org/10.5065/D6SJ1JCH>, 2020.
- Meher, J. K., Das, L., Akhter, J., Benestad, R. E., and Mezghani, A.: Performance of CMIP3 and CMIP5 GCMs to simulate observed rainfall characteristics over the Western Himalayan region, *Journal of Climate*, 30, 7777–7799, 2017.
- Melillo, J. M., Richmond, T., and Yohe, G.: Climate change impacts in the United States: The third national climate assessment. US Global Change Research Program, 841 pp. DOI: 10.7930/J0Z31WJ2, Online at: nca2014.globalchange.gov, 2014.
- Milly, P., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Stationarity is dead: Whither water management?, *Earth*, 4, 20, 2008.
- Milly, P. C. and Dunne, K. A.: Potential evapotranspiration and continental drying, *Nature Climate Change*, 6, 946–949, 2016.
- Mintz, Y. and Serafini, Y.: A global monthly climatology of soil moisture and water balance, *Climate Dynamics*, 8, 13–27, 1992.
- Mishra, A. K. and Singh, V. P.: A review of drought concepts, *Journal of Hydrology*, 391, 202–216, 2010.
- Mo, K. C.: Drought onset and recovery over the United States, *Journal of Geophysical Research: Atmospheres*, 116, 2011.
- Mo, K. C. and Lettenmaier, D. P.: Precipitation deficit flash droughts over the United States, *Journal of Hydrometeorology*, 17, 1169–1184, 2016.
- Moon, H., Gudmundsson, L., and Seneviratne, S. I.: Drought persistence errors in global climate models, *Journal of Geophysical Research: Atmospheres*, 123, 3483–3496, 2018.

REFERENCE

- Morehart, M., Gollehon, N., Dismukes, R., Breneman, V., and Heimlich, R.: Agricultural impacts are severe locally, but limited nationally, An Economic Assessment of the 1999 Drought, 1999.
- Moser, S. C., Kasperson, R. E., Yohe, G., and Agyeman, J.: Adaptation to climate change in the Northeast United States: opportunities, processes, constraints, *Mitigation and Adaptation Strategies for Global Change*, 13, 643–659, 2008.
- Namias, J.: Nature and possible causes of the northeastern United States drought during 1962–65, *Mon. Wea. Rev.*, 94, 543–554, 1966.
- Nasrollahi, N., AghaKouchak, A., Cheng, L., Damberg, L., Phillips, T. J., Miao, C., Hsu, K., and Sorooshian, S.: How well do CMIP5 climate simulations replicate historical trends and patterns of meteorological droughts?, *Water Resources Research*, 51, 2847–2864, 2015.
- National Academies of Sciences, Engineering, and Medicine and others: Attribution of extreme weather events in the context of climate change, National Academies Press, 2016.
- National Center for Atmospheric Research: Multi-Model Large Ensemble Archive, <https://www.cesm.ucar.edu/projects/community-projects/MMLEA/>, 2020.
- National Center for Environmental Information: Billion-Dollar Weather and Climate Disasters, <https://www.ncdc.noaa.gov/billions/>, 2021.
- National Center for Environmental Information: NOAA Drought Report - July 1999, <https://www.ncdc.noaa.gov/sotc/drought/199907>, 2022a.
- National Center for Environmental Information: NOAA Drought Report - Summer (JJA) 1999, <https://www.ncdc.noaa.gov/sotc/drought/199916>, 2022b.
- NCAR: WRF Official User Guide Chapter 5 (parameterization information), https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_v4/v4.0/users_guide_chap5.html#restart, 2019.
- Neale, R. B., Chen, C.-C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., Conley, A. J., Garcia, R., Kinnison, D., and Lamarque, J.-F. e. a.: Description of the NCAR community atmosphere model (CAM 5.0), NCAR Tech. Note, 1, 1–12, 2010.

- Nguyen, P., Thorstensen, A., Sorooshian, S., Zhu, Q., Tran, H., Ashouri, H., Miao, C., Hsu, K., and Gao, X.: Evaluation of CMIP5 model precipitation using PERSIANN-CDR, *Journal of Hydrometeorology*, 18, 2313–2330, 2017.
- NIDIS/NOAA: Coping with drought and its aftermath in the Northeast, Tech. rep., National Oceanic and Atmospheric Administration, 2017.
- NIDIS/NOAA: 2018-2019 Northeast DEWS Strategic Plan, Tech. rep., National Oceanic and Atmospheric Administration, 2021.
- NOAA: Heat Index definition by NOAA, <https://www.weather.gov/safety/heat-index>, 2020.
- NOAA Physical Sciences Laboratory: CPC Unified Gauge-Based Analysis of Daily Precipitation over CONUS, <https://psl.noaa.gov/data/gridded/data.unified.daily.conus.html>, 2020.
- Oleson, K. W., Lawrence, D. M., Gordon, B., Flanner, M. G., Kluzek, E., Peter, J., Levis, S., Swenson, S. C., Thornton, E., Feddema, J., et al.: Technical description of version 4.0 of the Community Land Model (CLM), 2010.
- Orlowsky, B. and Seneviratne, S. I.: Elusive drought: uncertainty in observed trends and short-andlong-term CMIP5 projections, *Hydrology and Earth System Sciences*, 17, 1765–1781, 2013.
- Osman, M., Zaitchik, B. F., Badr, H. S., Christian, J. I., Tadesse, T., Otkin, J. A., and Anderson, M. C.: Flash drought onset over the contiguous United States: sensitivity of inventories and trends to quantitative definitions, *Hydrology and Earth System Sciences*, 25, 565–581, 2021.
- Otkin, J. A., Anderson, M. C., Hain, C., Mladenova, I. E., Basara, J. B., and Svoboda, M.: Examining rapid onset drought development using the thermal infrared–based evaporative stress index, *Journal of Hydrometeorology*, 14, 1057–1074, 2013.
- Otkin, J. A., Anderson, M. C., Hain, C., Svoboda, M., Johnson, D., Mueller, R., Tadesse, T., Wardlow, B., and Brown, J.: Assessing the evolution of soil moisture and vegetation conditions

- during the 2012 United States flash drought, *Agricultural and Forest Meteorology*, 218, 230–242, 2016.
- Otkin, J. A., Svoboda, M., Hunt, E. D., Ford, T. W., Anderson, M. C., Hain, C., and Basara, J. B.: Flash droughts: A review and assessment of the challenges imposed by rapid-onset droughts in the United States, *Bulletin of the American Meteorological Society*, 99, 911–919, 2018.
- Otkin, J. A., Zhong, Y., Hunt, E. D., Christian, J. I., Basara, J. B., Nguyen, H., Wheeler, M. C., Ford, T. W., Hoell, A., Svoboda, M., et al.: Development of a flash drought intensity index, *Atmosphere*, 12, 741, 2021.
- O’Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., and van Vuuren, D. P.: A new scenario framework for climate change research: the concept of shared socioeconomic pathways, *Climatic Change*, 122, 387–400, 2014.
- O’Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., et al.: The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century, *Global Environmental Change*, 42, 169–180, 2017.
- Palmer, W. C.: *Meteorological drought*, vol. 30, US Department of Commerce, Weather Bureau, 1965.
- Paulachok, G. N., Krejmas, B. E., and Soden, H. L.: *Hydrologic aspects of the 1998-99 drought in the Delaware River basin*, US Department of the Interior, US Geological Survey, 2000.
- Peel, M. C., McMahon, T. A., and Finlayson, B. L.: Vegetation impact on mean annual evapotranspiration at a global catchment scale, *Water Resources Research*, 46, 2010.
- Pendergrass, A. G., Knutti, R., Lehner, F., Deser, C., and Sanderson, B. M.: Precipitation variability increases in a warmer climate, *Scientific Reports*, 7, 1–9, 2017.
- Pendergrass, A. G., Meehl, G. A., Pulwarty, R., Hobbins, M., Hoell, A., AghaKouchak, A., Bonfils,

- C. J., Gallant, A. J., Hoerling, M., Hoffmann, D., et al.: Flash droughts present a new challenge for subseasonal-to-seasonal prediction, *Nature Climate Change*, 10, 191–199, 2020.
- Pfahl, S., O’Gorman, P. A., and Fischer, E. M.: Understanding the regional pattern of projected future changes in extreme precipitation, *Nature Climate Change*, 7, 423–427, 2017.
- Pierce, D. W., Cayan, D. R., and Thrasher, B. L.: Statistical downscaling using localized constructed analogs (LOCA), *Journal of Hydrometeorology*, 15, 2558–2585, 2014.
- Pollard, D.: A central limit theorem for empirical processes, *Journal of the Australian Mathematical Society*, 33, 235–248, 1982.
- Pontoppidan, M., Kolstad, E., Sobolowski, S., and King, M.: Improving the reliability and added value of dynamical downscaling via correction of large-scale errors: a Norwegian perspective, *Journal of Geophysical Research: Atmospheres*, 123, 11–875, 2018.
- Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., Coen, J. L., Gochis, D. J., Ahmadov, R., and Peckham, S. E. e. a.: The weather research and forecasting model: Overview, system efforts, and future directions, *Bulletin of the American Meteorological Society*, 98, 1717–1737, 2017.
- Prasanna, V.: Statistical bias correction method applied on CMIP5 datasets over the Indian region during the summer monsoon season for climate change applications, *Theoretical and Applied Climatology*, 131, 471–488, 2018.
- Priestley, M. D., Ackerley, D., Catto, J. L., Hodges, K. I., McDonald, R. E., and Lee, R. W.: An overview of the extratropical storm tracks in CMIP6 historical simulations, *Journal of Climate*, 33, 6315–6343, 2020.
- Reick, C. H., Gayler, V., Goll, D., Hagemann, S., Heidkamp, M., Nabel, J. E., Raddatz, T., Roeckner, E., Schnur, R., and Wilkenskjeld, S.: JSBACH 3-The land component of the MPI Earth System Model: documentation of version 3.2, Tech. rep., MPI für Meteorologie, 2021.

- Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O’neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., et al.: The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview, *Global Environmental Change*, 42, 153–168, 2017.
- Roderick, M., Hobbins, M., and Farquhar, G.: Pan Evaporation Trends and the Terrestrial Water Balance. I. Principles and Observations, *Geogr. Compass*, 3, 746–760, 2009a.
- Roderick, M. L., Hobbins, M. T., and Farquhar, G. D.: Pan evaporation trends and the terrestrial water balance. II. Energy balance and interpretation, *Geography Compass*, 3, 761–780, 2009b.
- Roderick, M. L., Greve, P., and Farquhar, G. D.: On the assessment of aridity with changes in atmospheric CO₂, *Water Resources Research*, 51, 5450–5463, 2015.
- Rodgers, K. B., Lin, J., and Frölicher, T. L.: Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model, *Biogeosciences*, 12, 3301–3320, 2015.
- Rothfus, L. P. and Headquarters, N. S. R.: The heat index equation (or, more than you ever wanted to know about heat index), Fort Worth, Texas: National Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology, 9023, 1990.
- Rummukainen, M.: State-of-the-art with regional climate models, *Wiley Interdisciplinary Reviews: Climate Change*, 1, 82–96, 2010.
- Schulzweida, U.: CDO User Guide, <https://doi.org/10.5281/zenodo.3539275>, 2019.
- Seager, R., Naik, N., and Vecchi, G. A.: Thermodynamic and dynamic mechanisms for large-scale changes in the hydrological cycle in response to global warming, *Journal of Climate*, 23, 4651–4668, 2010.
- Seager, R., Pederson, N., Kushnir, Y., Nakamura, J., and Jurburg, S.: The 1960s drought and the subsequent shift to a wetter climate in the Catskill Mountains region of the New York City watershed, *Journal of Climate*, 25, 6721–6742, 2012.

- Sellers, P., Bounoua, L., Collatz, G., Randall, D., Dazlich, D., Los, S., Berry, J., Fung, I., Tucker, C., Field, C., et al.: Comparison of radiative and physiological effects of doubled atmospheric CO₂ on climate, *Science*, 271, 1402–1406, 1996.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: 2005: A description of the advanced research WRF version 2, in: NCAR Tech. Note, Citeseer, 2008.
- Song, F., Guo, Z., and Mei, D.: Feature selection using principal component analysis, in: 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization, vol. 1, pp. 27–30, IEEE, 2010.
- Srivastava, A., Grotjahn, R., and Ullrich, P.: Evaluation of historical CMIP6 model simulations of extreme precipitation over contiguous US regions, *Weather and Climate Extremes*, p. 100268, 2020.
- Stansfield, A. M., Reed, K. A., Zarzycki, C. M., Ullrich, P. A., and Chavas, D. R.: Assessing Tropical Cyclones’ Contribution to Precipitation over the Eastern United States and Sensitivity to the Variable-Resolution Domain Extent, *Journal of Hydrometeorology*, 21, 1425–1445, 2020.
- Stryker, J., Wemple, B., and Bomblies, A.: Modeling the impacts of changing climatic extremes on streamflow and sediment yield in a northeastern US watershed, *Journal of Hydrology: Regional Studies*, 17, 83–94, 2018.
- Strzepek, K., Yohe, G., Neumann, J., and Boehlert, B.: Characterizing changes in drought risk for the United States from climate change, *Environmental Research Letters*, 5, 044 012, 2010.
- Sun, L., Alexander, M., and Deser, C.: Evolution of the global coupled climate response to Arctic sea ice loss during 1990–2090 and its contribution to climate change, *Journal of Climate*, 31, 7823–7843, 2018.
- Survey, U. G. and US Department of Agriculture, N. R. C. S.: Federal Standards and Procedures for the National Watershed Boundary Dataset (WBD), *Techniques and Methods*, 11, 63, 2013.

- Svoboda, M. and Fuchs, B.: Handbook of drought indicators and indices. Integrated drought management programme (IDMP), integrated drought management tools and guidelines series 2, World Meteorological Organization and Global Water Partnership, Geneva, Switzerland, 52, 2016.
- Svoboda, M., Hayes, M., and Wood, D.: Standardized precipitation index user guide, World Meteorological Organization Geneva, Switzerland, 2012.
- Swann, A. L.: Plants and drought in a changing climate, *Current Climate Change Reports*, 4, 192–201, 2018.
- Swann, A. L., Hoffman, F. M., Koven, C. D., and Randerson, J. T.: Plant responses to increasing CO₂ reduce estimates of climate impacts on drought severity, *Proceedings of the National Academy of Sciences*, 113, 10 019–10 024, 2016.
- Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R., and Demaria, E.: Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes, *Hydrology and Earth System Sciences*, 21, 2649–2666, 2017.
- Tallaksen, L. M. and Van Lanen, H. A.: Hydrological drought: processes and estimation methods for streamflow and groundwater, Elsevier, 2004.
- Tarek, M., Brissette, F. P., and Arsenault, R.: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America, *Hydrology and Earth System Sciences*, 24, 2527–2544, 2020.
- Taylor, I., Burke, E., McColl, L., Falloon, P., Harris, G., and McNeall, D.: The impact of climate mitigation on projections of future drought, *Hydrology and Earth System Sciences*, 17, 2339–2358, 2013.
- Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *Journal of Geophysical Research: Atmospheres*, 106, 7183–7192, 2001.

REFERENCE

- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, 93, 485–498, 2012.
- Teshome, A. and Zhang, J.: Increase of Extreme Drought over Ethiopia under Climate Warming, *Advances in Meteorology*, 2019, 2019.
- Teuling, A., Hirschi, M., Ohmura, A., Wild, M., Reichstein, M., Ciais, P., Buchmann, N., Ammann, C., Montagnani, L., Richardson, A., et al.: A regional perspective on trends in continental evaporation, *Geophysical Research Letters*, 36, 2009.
- Thomas, T., Nayak, P., and Ghosh, N. C.: Spatiotemporal analysis of drought characteristics in the bundelkhand region of central india using the standardized precipitation index, *Journal of Hydrologic Engineering*, 20, 05015 004, 2015.
- Trenberth, K. E. and Guillemot, C. J.: Physical processes involved in the 1988 drought and 1993 floods in North America, *Journal of Climate*, 9, 1288–1298, 1996.
- Trenberth, K. E., Branstator, G. W., and Arkin, P. A.: Origins of the 1988 North American drought, *Science*, 242, 1640–1645, 1988.
- Trenberth, K. E., Dai, A., Van Der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., and Sheffield, J.: Global warming and changes in drought, *Nature Climate Change*, 4, 17–22, 2014.
- Ukkola, A., Pitman, A., De Kauwe, M., Abramowitz, G., Herger, N., Evans, J., and Decker, M.: Evaluating CMIP5 model agreement for multiple drought metrics, *Journal of Hydrometeorology*, 19, 969–988, 2018.
- Ullrich, P., Xu, Z., Rhoades, A., Dettinger, M., Mount, J., Jones, A., and Vahmani, P.: California’s drought of the future: A midcentury recreation of the exceptional conditions of 2012–2017, *Earth’s Future*, 6, 1568–1587, 2018.
- US Army Corps of Engineers: Conservation Plan for the Interior Least Tern, Pallid Sturgeon, and Fat Pocketbook Mussel in the Lower Mississippi River, <https://doi.org/https://www.lmrcc.org/>

REFERENCE

- wp-content/uploads/2021/02/LMR-Conservation-Plan-Final-USACE-CIP-23-July-2013.pdf, 2013.
- US Bureau of Economic Analysis: Gross domestic product (GDP) by state (millions of current dollars), US Bureau of Economic Analysis, 2016.
- USDA: Support for Lower Mississippi River Basin Nomination for Long-Term Agroecosystem Research (LTAR) Network Membership, <https://doi.org/https://www.ars.usda.gov/ARSEUserFiles/np211/LMRBProposal.pdf>, 2012.
- USDA: Long-Term Agroecosystem Research (LTAR) Report of the Lower Mississippi River Basin, https://doi.org/https://ltar.ars.usda.gov/wp-content/uploads/2020/11/LMRB_Climate_Final_Edits.pdf, 2020.
- Van den Dool, H., Huang, J., and Fan, Y.: Performance and analysis of the constructed analogue method applied to US soil moisture over 1981–2001, *Journal of Geophysical Research: Atmospheres*, 108, 2003.
- Van Loon, A. F.: Hydrological drought explained, *Wiley Interdisciplinary Reviews: Water*, 2, 359–392, 2015.
- Velasquez, P., Messmer, M., and Raible, C. C.: A new bias-correction method for precipitation over complex terrain suitable for different climate states, *Geoscientific Model Development Discussions*, pp. 1–27, 2019.
- Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index, *Journal of Climate*, 23, 1696–1718, 2010.
- Vicente-Serrano, S. M., Beguería, S., Lorenzo-Lacruz, J., Camarero, J. J., López-Moreno, J. I., Azorin-Molina, C., Revuelto, J., Morán-Tejeda, E., and Sanchez-Lorenzo, A.: Performance of drought indices for ecological, agricultural, and hydrological applications, *Earth Interactions*, 16, 1–27, 2012.

- Vicente-Serrano, S. M., Camarero, J. J., and Azorin-Molina, C.: Diverse responses of forest growth to drought time-scales in the Northern Hemisphere, *Global Ecology and Biogeography*, 23, 1019–1030, 2014.
- Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., Kumar, P., Rao, P. S. C., Basu, N. B., and Wilson, J. S.: The future of hydrology: An evolving science for a changing world, *Water Resources Research*, 46, 2010.
- Wang, K. and Dickinson, R. E.: A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability, *Reviews of Geophysics*, 50, 2012.
- Wang, K., Dickinson, R. E., Wild, M., and Liang, S.: Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 1. Model development, *Journal of Geophysical Research: Atmospheres*, 115, 2010.
- Wehner, M., Gleckler, P., and Lee, J.: Characterization of long period return values of extreme daily temperature and precipitation in the CMIP6 models: Part 1, model evaluation, *Weather and Climate Extremes*, 30, 100283, 2020.
- Weiss, M., van den Hurk, B., Haarsma, R., and Hazeleger, W.: Impact of vegetation variability on potential predictability and skill of EC-Earth simulations, *Climate Dynamics*, 39, 2733–2746, 2012.
- Weiss, M., Miller, P. A., van den Hurk, B. J., van Noije, T., Ștefănescu, S., Haarsma, R., Van Uft, L. H., Hazeleger, W., Le Sager, P., Smith, B., et al.: Contribution of dynamic vegetation phenology to decadal climate predictability, *Journal of Climate*, 27, 8563–8577, 2014.
- Wilhite, D. A.: Drought: A Global Assessment, Chapter 1: Drought as a natural hazard: concepts and definitions, Routledge, 2000.
- Wilhite, D. A.: Drought and water crises: science, technology, and management issues, Crc Press, 2005.

- Wilhite, D. A. and Glantz, M. H.: Understanding: the drought phenomenon: the role of definitions, *Water International*, 10, 111–120, 1985.
- Williams, A. P., Seager, R., Abatzoglou, J. T., Cook, B. I., Smerdon, J. E., and Cook, E. R.: Contribution of anthropogenic warming to California drought during 2012–2014, *Geophysical Research Letters*, 42, 6819–6828, 2015.
- Williams, D., Cable, W., Hultine, K., Hoedjes, J., Yepez, E., Simonneaux, V., Er-Raki, S., Boulet, G., De Bruin, H., Chehbouni, A., et al.: Evapotranspiration components determined by stable isotope, sap flow and eddy covariance techniques, *Agricultural and Forest Meteorology*, 125, 241–258, 2004.
- Wu, F., Wu, B., Zhu, W., Yan, N., Ma, Z., Wang, L., Lu, Y., and Xu, J.: ETWatch Cloud: APIs for regional actual evapotranspiration data generation, *Environmental Modelling & Software*, 145, 105 174, 2021.
- Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., and Liu, C.: A gauge-based analysis of daily precipitation over East Asia, *Journal of Hydrometeorology*, 8, 607–626, 2007.
- Xie, P., Chen, M., and Shi, W.: CPC unified gauge-based analysis of global daily precipitation, in: *Preprints, 24th Conf. on Hydrology, Atlanta, GA, Amer. Meteor. Soc.*, vol. 2, 2010.
- Xie, S.-P., Deser, C., Vecchi, G. A., Collins, M., Delworth, T. L., Hall, A., Hawkins, E., Johnson, N. C., Cassou, C., Giannini, A., et al.: Towards predictive understanding of regional climate change, *Nature Climate Change*, 5, 921–930, 2015.
- Xu, X., Li, X., Wang, X., He, C., Tian, W., Tian, J., and Yang, L.: Estimating daily evapotranspiration in the agricultural-pastoral ecotone in Northwest China: A comparative analysis of the Complementary Relationship, WRF-CLM4. 0, and WRF-Noah methods, *Science of The Total Environment*, p. 138635, 2020.
- Xu, Z. and Yang, Z.-L.: An improved dynamical downscaling method with GCM bias corrections and its validation with 30 years of climate simulations, *Journal of Climate*, 25, 6271–6286, 2012.

- Xue, Z. and Ullrich, P.: A Comprehensive Intermediate-Term Drought Evaluation System and Evaluation of Climate Data Products over the Conterminous United States, *Journal of Hydrometeorology*, 2021a.
- Xue, Z. and Ullrich, P.: A Retrospective and Prospective Examination of the 1960s U.S. Northeast Drought, *Earth's Future*, 9, xxx, 2021b.
- Yang, S., Feng, J., Dong, W., and Chou, J.: Analyses of extreme climate events over China based on CMIP5 historical and future simulations, *Advances in Atmospheric Sciences*, 31, 1209–1220, 2014.
- Yang, Y., Uddstrom, M., and Duncan, M.: Effects of short spin-up periods on soil moisture simulation and the causes over New Zealand, *Journal of Geophysical Research: Atmospheres*, 116, 2011.
- Yang, Y., Roderick, M. L., Zhang, S., McVicar, T. R., and Donohue, R. J.: Hydrologic implications of vegetation response to elevated CO₂ in climate projections, *Nature Climate Change*, 9, 44–48, 2019.
- Yihdego, Y., Vaheddoost, B., and Al-Weshah, R. A.: Drought indices and indicators revisited, *Arabian Journal of Geosciences*, 12, 69, 2019.
- Yu, L., Zhong, S., Heilman, W. E., and Bian, X.: Trends in seasonal warm anomalies across the contiguous United States: contributions from natural climate variability, *Scientific Reports*, 8, 1–13, 2018.
- Yuan, X., Wang, L., Wu, P., Ji, P., Sheffield, J., and Zhang, M.: Anthropogenic shift towards higher risk of flash drought over China, *Nature Communications*, 10, 1–8, 2019.
- Zargar, A., Sadiq, R., Naser, B., and Khan, F. I.: A review of drought indices, *Environmental Reviews*, 19, 333–349, 2011.
- Zarzycki, C. M., Ullrich, P. A., and Reed, K. A.: Metrics for evaluating tropical cyclones in climate data, *Journal of Applied Meteorology and Climatology*, 60, 643–660, 2021.

- Zhang, G. J. and McFarlane, N. A.: Role of convective scale momentum transport in climate simulation, *Journal of Geophysical Research: Atmospheres*, 100, 1417–1426, 1995.
- Zhang, K., Kimball, J. S., Nemani, R. R., Running, S. W., Hong, Y., Gourley, J. J., and Yu, Z.: Vegetation greening and climate change promote multidecadal rises of global land evapotranspiration, *Scientific reports*, 5, 1–9, 2015.
- Zhang, L., Dawes, W., and Walker, G.: Response of mean annual evapotranspiration to vegetation changes at catchment scale, *Water Resources Research*, 37, 701–708, 2001.
- Zhang, M. and Yuan, X.: Rapid reduction in ecosystem productivity caused by flash droughts based on decade-long FLUXNET observations, *Hydrology and Earth System Sciences*, 24, 5579–5593, 2020.
- Zhang, T., Barry, R., Knowles, K., Ling, F., and Armstrong, R.: Distribution of seasonally and perennially frozen ground in the Northern Hemisphere, in: *Proceedings of the 8th International Conference on Permafrost*, vol. 2, pp. 1289–1294, AA Balkema Publishers Zürich, Switzerland, 2003.
- Zhao, T. and Dai, A.: The Magnitude and Causes of Global Drought Changes in the Twenty-First Century under a Low-Moderate Emissions Scenario, *Journal of Climate*, 28, 4490 – 4512, <https://doi.org/10.1175/JCLI-D-14-00363.1>, 2015.
- Zhu, X., Wei, Z., Dong, W., Ji, Z., Wen, X., Zheng, Z., Yan, D., and Chen, D.: Dynamical downscaling simulation and projection for mean and extreme temperature and precipitation over central Asia, *Climate Dynamics*, pp. 1–28, 2020.
- Zhuang, J., Dussin, R., Jüling, A., and Rasp, S.: JiaweiZhuang/xESMF: v0.3.0 Adding ESMF.LocStream capabilities, <https://doi.org/10.5281/zenodo.3700105>, 2020.
- Zhuo, L., Dai, Q., Han, D., Chen, N., and Zhao, B.: Assessment of simulated soil moisture from WRF Noah, Noah-MP, and CLM land surface schemes for landslide hazard application, *Hydrology and Earth System Sciences*, 23, 4199–4218, 2019.