`

SERDP Seed Project Final Report

Project Title:

Determining the Temporal and Spatial Scales of Nonstationarity in Temperature and Precipitation across the Continental United States for a Given Emissions Scenario

SERDP Project Number:

RC18-1577

Performing Organization:

Lawrence Berkeley National Laboratory

Lead Principal Investigator:

Daniel Feldman

Date:

20 December 2019

Version Number:

2.0

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
12/20/2019	SERDP Final Report	

4. TITLE AND SUBTITLE
Determining the Temporal and Spatial Scales of Nonstationarity in Temperature and Precipitation across the Continental United States for a Given Emissions Scenario

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Daniel Feldman

5d. PROJECT NUMBER
RC18-1577

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lawrence Berkeley National Laboratory
1 Cyclotron Road, MS74R-316C
Berkeley, CA 94720

8. PERFORMING ORGANIZATION REPORT NUMBER
RC18-1577

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Strategic Environmental Research and Development Program
4800 Mark Center Drive, Suite 17D03
Alexandria, VA 22350-3605

10. SPONSOR/MONITOR'S ACRONYM(S)
SERDP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
RC18-1577

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The means and extremes of surface air temperature and precipitation have changed significantly in the last 50 years, and may change more in the 21st Century. These changes may be without an historical analogue, and this nonstationarity presents a challenge for DoD infrastructure planning and operations, wherein historical distributions are used as the basis for managing infrastructure and operations risk from flooding, drought, heat waves, and other weather phenomena. The objective of this proposal is to use a combination of climate models, observations, and analysis to develop robust,defensible projections of the range of distributions of temperature and precipitation that DoD infrastructure and operations will face.

15. SUBJECT TERMS
Downscaling, quantile mapping, CMIP5, CMIP6, CORDEX, HighResMIP, WRF

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Daniel Feldman
					19b. TELEPHONE NUMBER *(Include area code)*
UNCLASS	UNCLASS	UNCLASS	UNCLASS	83	510-495-2171

Table of Contents

Front Matter iii
- **List of Tables** iii
- **List of Figures** iv
- **List of Acronyms** vi
- **Keywords** vii
- **Acknowledgements** vii

Abstract 8
- **Introduction and Objectives** 8
- **Technical Approach** 8
- **Results** 8
- **Benefits** 8

Executive Summary 9
- **Introduction** 9
- **Objectives** 11
- **Technical Approach** 12
- **Results and Discussion** 14
- **Implications for Future Research and Benefits** 17

Full Report 18
- **Objective** 18
- **Background** 18
- **Materials and Methods** 28
  - Conceptual Framework 28
  - Observational Datasets and Gridding 29
  - Statistical Analysis Approach 31
  - Dynamical Downscaling: WRF 35
  - Statistical Downscaling: LOCA-WRF and LOCA-Livneh 36
- **Results and Discussion** 37
  - Focus on TAFB and FBR (Tadić and Biraud, JAMC, Submitted) 37
  - Model Projections for California (Feldman et al, JWRPM, Submitted) 48
  - A Deeper Exploration into Nonstationarity (Walton et al, JAMC, In Review) 65
- **Conclusions and Implications for Future Research** 70
- **Literature Cited** 73

Front Matter

List of Tables

1. Exec. Summary: Mean projected changes of temperature and precipitation at TAFB
2. Exec. Summary: Mean projected changes of temperature and precipitation at FBR
3. Exec. Summary: Changes in extremes of temperature and precipitation at TAFB and FBR
4. Full Report: Mean projected changes of temperature and precipitation at TAFB
5. Full Report: Mean projected changes of temperature and precipitation at FBR
6. Full Report: Changes in extremes of temperature and precipitation at TAFB and FBR
7. Full Report: Exceedance probabilities of temperature and precipitation events at TAFB for RCP4.5 and RCP8.5 for different 30-year time windows.
8. Full Report: Exceedance probabilities of temperature and precipitation events at FBR for RCP4.5 and RCP8.5 for different 30-year time windows.
9. Full Report: Differences between QM and QDM downscaled projections relative to parent GCM differences for regions of California.

List of Figures

1. Exec. Summary: Flow-chart for translating CMIP projections to local risk.
2. Exec. Summary: Statistical analysis framework for developing local projections.
3. Exec. Summary: Statistically-based change in 20-year daily maximum temperature and precipitation return values across the CONUS for RCP4.5 and RCP8.5.
4. Exec. Summary: Divergence in statistical and dynamical downscaling solutions over California between historical record and future projections of the 21st Century.
5. Full Report: Observed changes in mean temperature and precipitation over the United States between the middle of the 20th Century and the early part of the 21st Century.
6. Full Report: Drivers of model projection uncertainty in the 21st Century.
7. Full Report: Regional changes in temperature over the United States for the 21st Century for the RCP4.5 and RCP8.5 scenarios.
8. Full Report: Regional changes in precipitation over the United States for the 21st Century by season for the RCP8.5 scenario.
9. Full Report: Contrast in precipitation from the GCM scale to downscaled projections.
10. Full Report: Observed changes in 20-year daily temperature return values over the CONUS and biases in 8 CMIP5 models in capturing this change.
11. Full Report: Observed changes in 20-year daily precipitation return values over the CONUS and biases in 8 CMIP5 models in capturing this change.
12. Full Report: Scatterplot of model skill vs. independence for North American hydrometeorology.
13. Full Report: Flow-chart for translating CMIP projections to local risk.
14. Full Report: Distribution of long-duration weather stations to develop historical observations.
15. Full Report: Directional correlations of extreme precipitation in historical observational record as a function of season.
16. Full Report: Scatterplot of extreme precipitation estimates in the historical record based on gridding that preserves extremes and traditional bilinear interpolation gridding.
17. Full Report: Spatial distribution of differences in historical precipitation extremes based on gridding that preserves extremes and traditional bilinear interpolation gridding.
18. Full Report: Statistical analysis framework for developing local projections.
19. Full Report: CDFs and PDFs of precipitation and temperature at TAFB built from statistical analysis framework.
20. Full Report: MLR β coefficients for CMIP5 models.
21. Full Report: MLR reconstruction of historical distributions vs actual historical distributions of temperature and precipitation.
22. Full Report: GEV PDF over TAFB of temperature and precipitation during the 21st Century.
23. Full Report: Marginal and joint distribution of observations of daily maximum temperature at TAFB and FBR during 1985–2011 period.
24. Full Report: Marginal and joint distribution of daily precipitation observations at TAFB and FBR during 1985–2011 period.
25. Full Report: Marginal and joint distribution of observations of non-zero precipitation and corresponding daily maximum temperature at TAFB, during 1985–2011 period.

26. Full Report: Marginal and joint distribution of observations of non-zero precipitation and corresponding daily maximum temperature at FBR, during 1985–2011 period.
27. Full Report: Statistically-based change in 20-year daily maximum temperature and precipitation return values across the CONUS for RCP4.5 and RCP8.5.
28. Full Report: Fractional change in projected precipitation over California for the DEW and WMW cases in the 21st Century for several different quantiles.
29. Full Report: Historical biases over California in DEW and WMW simulations as a function of quantile.
30. Full Report: Difference in projected precipitation in the 21st Century relative to the historical record over California for a set of quantiles for the DEW case.
31. Full Report: Difference in projected precipitation in the 21st Century relative to the historical record over California for a set of quantiles for the WMW case.
32. Full Report: Changes in daily return values of precipitation for DEW case over California.
33. Full Report: Numerical instabilities over California in developing downscaled projections using QDM methods.
34. Full Report: QDM exceedence and sensitivity to quantile resolution of QDM methods.
35. Full Report: VIC output based on downscaled DEW and WMW projections: Sacramento River Basin Unimpaired Streamflow.
36. Full Report: VIC output based on downscaled DEW and WMW projections: San Joaquin River Basin Unimpaired Streamflow.
37. Full Report: VIC output based on downscaled DEW and WMW projections: Sacramento River Basin Cumulative Winter-Season Precipitation.
38. Full Report: VIC output based on downscaled DEW and WMW projections: San Joaquin River Basin Cumulative Winter-Season Precipitation.
39. Full Report: VIC output based on downscaled DEW and WMW projections: Sacramento River Basin April 1 (Peak-Season) Snow-Water Equivalent.
40. Full Report: VIC output based on downscaled DEW and WMW projections: San Joaquin River Basin April 1 (Peak-Season) Snow-Water Equivalent.
41. Full Report: Divergence in statistical and dynamical downscaling solutions over California between historical record and future projections of the 21st Century.
42. Full Report: Differences in downscaling solutions of temperature over California for the 21st Century.
43. Full Report: Differences in downscaling solutions of precipitation over California for the 21st Century.
44. Full Report: Diagram depicting how influence of MBL and topography can violate stationarity assumptions for GCMs and downscaling.
45. Full Report: Observed spatial patterns in trends of coastal and near-coastal temperatures to be used as a basis for process-based metrics of GCMs and downscaling solutions.

List of Acronyms

AR5:	Fifth Assessment Report
AR6:	Sixth Assessment Report
BCSD:	Bias Correction with Spatial Disaggregation
BCCA:	Bias Correction with Constructed Analogues
CDF:	Cumulative Distribution Function
CMIP3:	Coupled Model Intercomparison Project – Phase 3
CMIP5:	Coupled Model Intercomparison Project – Phase 5
CMIP6:	Coupled Model Intercomparison Project – Phase 6
CMIP7:	Coupled Model Intercomparison Project – Phase 7
CONUS:	Conterminous United States
DECK:	Diagnostics, Evaluation, and Characterization of Klima
DEW:	Drier Extreme Warming
DOD:	Department of Defense
ENSO:	El Niño Southern Oscillation
ERA:	European Centre for Medium-Range Weather Forecasting ReAnalysis
ESGF:	Earth System Grid Federation
FACETS:	Framework for Assessing Climate's Energy-Water-Land Nexus Using Targeted Simulations
FBR:	Fort Bragg
GCM:	Global Circulation Model
GDO:	Greed Data Oasis
GEV:	Generalized Extreme Value
GHCN:	Global Historical Climatology Network
GP:	Gaussian Process
HighResMIP:	High-Resolution Model Intercomparison Project
IPCC:	Intergovernmental Panel on Climate Change
LBNL:	Lawrence Berkeley National Laboratory
LLNL:	Lawrence Livermore National Laboratory
LOCA:	Localized Constructed Analogs
MAF:	Million Acre-Feet
MBL:	Marine Boundary Layer
MCMC:	Markov-Chain Monte Carlo
MIP:	Model Intercomparison Project
MLR:	Multi-Linear Regression
NA-CORDEX:	North American Coordinated Regional Downscaling Experiment
NARR:	North American Regional Reanalysis
NCA4:	Fourth National Climate Assessment
NCAR:	National Center for Atmospheric Research
PCMDI:	Program for Climate Modeling Diagnostics and Intercomparison
PDF:	Probability Distribution Function
RAWS:	Remote Automated Weather Station
RCM:	Regional Climate Model
RCP:	Representative Concentration Pathway
RCP4.5:	Representative Concentration Pathway, version 4.5

RCP8.5:	Representative Concentration Pathway, version 8.5
QDM:	Quantile Delta Mapping
QM:	Quantile Mapping
SSP:	Shared Socioeconomic Pathway
SSP1:	Shared Socioeconomic Pathway, version 1
SSP5:	Shared Socioeconomic Pathway, version 5
SERDP:	Strategic Environmental Research and Development Program
SNOTEL:	Snow Telemetry
SWE:	Snow Water Equivalent
TAFB:	Travis Air Force Base
USACE:	United States Army Corps of Engineers
VIACS:	Vulnerability, Impacts, Adaptation, and Climate Services
VIC:	Variable Infiltration Capacity
WMW:	Wetter Moderate Warming
WRF:	Weather Research and Forecasting
WSIP:	Water Storage and Investment Program

Keywords

Downscaling, quantile mapping, CMIP5, CMIP6, CORDEX, HighResMIP, WRF

Acknowledgements

The investigating team would like first and foremost to acknowledge the SERDP program for providing the funding to perform this research, supporting the 2018 Symposium which enabled scientific discussions and presentations, for inviting this team to present and discuss this team's research at the SERDP webinar, and for inviting the team to participate in the Resilient Aviation Infrastructure Workshop. This team would also like to recognize the detailed discussions with hydrologists from the State of California Department of Water Resources including Dr. Michael Anderson, Elissa Lynn, Wyatt Arnold, and Kevin He. Dr. Michael Wehner of the Lawrence Berkeley National Laboratory provided valuable feedback on the Coupled Model Intercomparison Project (CMIP) landscape. The team would also like to recognize the efforts of potential collaborators most especially Dr. Travis O'Brien of the Lawrence Berkeley National Laboratory for attending, discussing, and contributing to the Resilient Aviation Workshop. Additional discussions with Neil Berg of UCLA and Dave Pierce and Dan Cayan of UCSD were very useful in developing detailed perspectives on the strengths and weaknesses of different downscaling approaches. The team would further like to recognize the guidance of Dr. Jeffrey Arnold of the U.S. Army Corps of Engineers Climate Preparedness and Resiliency Program, since he graciously took time out of his schedule to provide information on how downscaled CMIP data are used to characterize the risk in specific watersheds around potentially vulnerable DoD installations. Finally, the team would like to thank collaborators including Ashray Manepalli of the University of California-Berkeley and Adrian Albert, Alan Rhoades, and Prabhat of the Lawrence Berkeley National Laboratory for developing discussions and leading efforts to use state-of-the-art machine-learning methods for developing new ways to probe models for their sensitivity to different processes.

Abstract

Introduction and Objectives

The means and extremes of surface air temperature and precipitation have changed significantly in the last 50 years, and may change more in the 21st Century. These changes may be without an historical analogue, and this nonstationarity presents a challenge for DoD infrastructure planning and operations, wherein historical distributions are used as the basis for managing infrastructure and operations risk from flooding, drought, heat waves, and other weather phenomena. The objective of this proposal is to use a combination of climate models, observations, and analysis to develop robust, defensible projections of the range of distributions of temperature and precipitation that DoD infrastructure and operations will face.

Technical Approach

There are several important tools for the development of these projections: Global Circulation Models (GCMs) and downscaling. This investigation advanced the understanding of a framework to manage these challenges and then determined the proper approach(es) to translating the coarse and biased climate model projections into the fine-scale, calibrated projections. This work used ~30 models from the Coupled Model Intercomparison Project – Phase 5 (CMIP5) downscaled with the Localized Constructed Analogues (LOCA) technique and developed regressed projections of 21st Century climate across the Conterminous United States.

Results

We found that a regression-based statistical framework can be employed to develop these projections but we also found that there are significant pitfalls that must be addressed before these projections can be confidently utilized by end-users. There are a number of potentially significant physical mechanisms, including changes in snow and coast weather pattern shifts, which can violate stationarity assumptions, and a regression-based approach using statistically downscaled models would not capture these effects. Therefore, we find that it will be necessary to advance research that addresses these pitfalls by (1) specifically testing stationarity assumptions inherent in statistical downscaling techniques, and (2) undertaking a set of intercomparisons between statistical and dynamical downscaling solutions to uncover and quantify the physical mechanisms that violate stationarity assumptions. The results of these expanded research efforts can then serve as a basis projections from the upcoming CMIP exercises (e.g., CMIP6) to ensure that they are appropriate for the development of best available estimates for how temperature and precipitation projections may change over the 21st Century.

Benefits

The benefit of this research is identifying what additional, targeted information is needed to ensure that local risks to DoD infrastructure and operations are not underestimated due to known physical mechanisms that could be un- or under-represented in climate models.

Executive Summary

Introduction

According to Milly et al. (2008) and as quoted in the Climate Change Installation Adaptation and Resilience 2017 Report, "stationarity is dead," and "In the new reality of non-stationarity, projections about climate must assume a constantly evolving basis." But how does this translate practically to making critical infrastructure decisions?

Large infrastructure decisions are made based on average temperature and precipitation, their variability, and extreme events, and such decisions may require many years of lead-time. General Circulation Models (GCMs), sometimes referred to as global climate models, can be a tool to assess changes in the large-scale distribution of temperature and precipitation as a function of future human activities, which are conventionally divided into discrete greenhouse gas emissions scenarios (van Vuuren et al, 2011). The range of global-scale temperature response to the forcing associated with a doubling of carbon dioxide is likely not more than 4.5 °C and has been established through a sustained focus on climate sensitivity (Charney et al, 1979; Knutti and Hegerl, 2008; Forster et al, 2016; Knutti et al, 2017). The response of the Earth's hydrological cycle as it adjusts to warming has also been studied, and gross features, such as the global hydrological cycle response to warming (Jeevanjee and Romps, 2018), the expansion of the troposphere (Lu et al, 2007; Hu et al, 2013), and shifts in the storm tracks (Yin, 2005; Bengtsson et al, 2006; Ulbrich et al, 2008), have been observed and have a strong theoretical basis.

However, these models are also blunt tools for finer-scale questions: the numerical integration of the intricate computer program that represents a GCM must balance a number of competing factors including, but not limited to, top-of-atmosphere energy balance, large-scale circulation, cloud formation and dissipation, ocean dynamics, and earth system response (Mauritsen et al, 2012; Hourdin et al, 2017; Schmidt et al, 2017). Furthermore, model development often focuses on achieving unbiased assessments of mean temperature and precipitation at larger spatial-scales and long time-scales. Because regional variability in precipitation and precipitation change across models is quite broad (Pendergrass and Hartmann, 2014), it is not known *a priori* that a given model will produce skillful temperature and precipitation projections as they relate to DoD facilities.

Nominally, model intercomparison projects serve to characterize the spread of model results associated with a given emissions scenario, but it has been recently recognized that navigating through the multi-model ensemble to understand the information contained in the ensemble's projections must consider both model errors with respect to observations and the interdependence of models (Sanderson et al, 2015; Knutti et al, 2017; Sanderson et al, 2017; Eyring et al, 2019). This navigation depends on the specific question being asked. Model democracy, where each simulation is treated as independent and unbiased, has been found to be suboptimal, primarily owing to the fact that models have not been developed independently and may even have substantial similarities in their code bases (Sanderson et al, 2015). Rather, it is essential that models being considered exhibit a level of skill with respect to regional metrics (Rupp et al, 2013). Taking California as an example, only models that reproduce the correlation and variance of mean seasonal spatial patterns, amplitude of seasonal cycle, diurnal temperature range, annual- to decadal-scale variance, long-term persistence, and Western United States

regional precipitation teleconnections to El Niño Southern Oscillation (ENSO) are appropriate for hydroclimate studies, and only 10 of the 33 CMIP5 models exhibited skill with respect to this metric (Pierce et al, 2018; Thorne et al, 2018). The 4th National Climate Assessment (NCA4) (USGCRP, 2017) adopted a model-weighting approach to produce, from the CMIP5 ensemble, regional-scale estimates of change in average temperature and precipitation. The results broadly indicate high confidence in predicting more warming with the Representative Concentration Pathway 8.5 (RCP8.5) scenario than the RCP4.5 scenario across the Conterminous United States (CONUS), and, with less confidence, more variability in regional precipitation for both of these scenarios.

At the same time, the Coupled Model Intercomparison Project - Phase 6 (CMIP6) is, as of this writing, in progress, with preliminary results coming online. At present, 151 modeling activities (one model running one set of experiments for one model intercomparison project constitutes an activity) have been reported to the Earth System Grid Federation (Cinquini et al, 2014) The contributing modeling centers are expected to produce between 20 and 40 Petabytes of climate model simulations as part of the CMIP6 exercise (Eyring et al, 2016). This is far too voluminous for planning purposes, and therefore, there is a strong need to curate these results to determine ranges of conditions that are likely to be faced locally. A number of different experiments are contained within CMIP6, including the Shared Socioeconomic Pathways (O'Neill et al, 2014; O'Neill et al, 2017) that have close heritage with the Representative Concentration Pathways (van Vuuren et al, 2011) of CMIP5 (Taylor et al, 2012). Other experiments including the North-American Coordinated Regional Downscaling Experiment (NA-CORDEX) (Giorgi et al, 2009; Jones et al, 2011; Mearns et al, 2012), the High Resolution Model Intercomparison Project (HighResMIP) (Haarsma et al, 2016), and the Vulnerability, Impacts, Adaptation, and Climate Services (VIACS) (Ruane et al, 2016) will yield simulations at the spatial resolution of interest to DoD facilities. But again, it is expected that there will be many participating modeling centers in each of these intercomparison projects, and there is a need to determine how to interpret the range of results that arise thereof specifically in terms of DoD facility-level risk.

It is critical that the climate change signals that are being captured by state-of-the-art climate models are translated to the local level and that the process of doing so especially does not underestimate the risks that climate change effects may have on local infrastructure.
Unfortunately, raw climate model output can exhibit biases in temperature and precipitation at regional scales, and these biases will propagate in complex and potentially nonlinear ways into subsequent models and decision-making efforts (Christensen et al, 2008; Mearns et al, 2012; Sillmann et al, 2013). To address this issue, one approach, while not without controversy (Ehret et al, 2012; Maraun 2012; Maraun 2013) is to bias-correct the raw GCM output based on historical observations (Gudmundsson et al, 2012; Teutschbein and Seibert, 2012; Teutschbein and Seibert, 2013; Pierce et al, 2015). An additional and critical step involves downscaling to address the gross mismatch in the spatial resolution on which GCMs operate and the processes that impact the local-level.

In November 2017, the SERDP Program released a Workshop Report entitled "Nonstationary Weather Patterns and Extreme Events" and found that "[improving] modeling and tools to develop, apply, and evaluate scenarios of multiple interacting stresses, including training in underlying assumptions and application of scenarios" and "[validating] current [downscaling]

methods, [developing] and [testing] new approaches, and [providing] guidance for different situations, locations, and variables" were among the greatest needs and opportunities for future research (Moss et al, 2018).

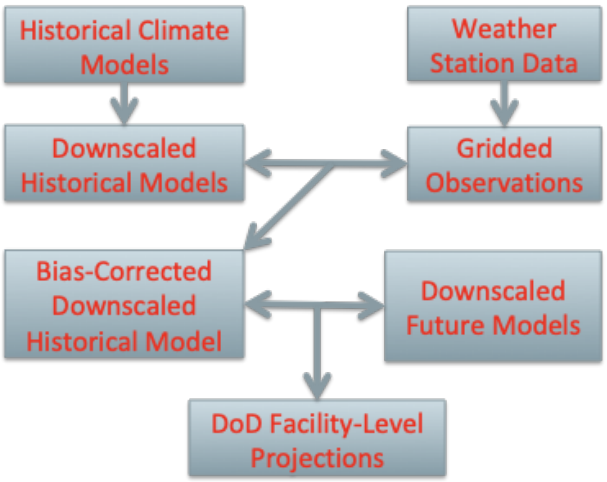


Figure 1: Flow-chart indicating how CMIP model projections are translated into model facility-level risk. Adapted from Feldman et al, (Submitted 2019).

Figure 1 shows a flow-chart that qualitatively produces DoD facility-level projections that balance historical observations with the relative change in the downscaled, bias-corrected model between future and historical simulations.

Distributions of temperature and precipitation can be constructed based on one or more bias-corrected, downscaled models results, but each step of this process must be critically evaluated, and several questions arise:

- How should individual models and the ensemble of model results be handled to develop confidence intervals?
- Can certain model results be trusted more than others?
- When and where is it appropriate to use, in the course of developing future projections, downscaling techniques and bias-correction based on historical observations?

Objectives

This investigation, arising in response to RCSON-18-L2 "Quantifying the Effects of Environmental Nonstationarity at DoD Relevant Scales," seeks to uncover the heterogeneous performance of the General Circulation Models (GCMs) with respect to nonstationarity of temperature and precipitation. The primary objective of this research is to build the foundation for a comprehensive determination of this risk by using historical observations to inform model projections of the statistical properties of temperature and precipitation, including nonstationarity, at regional scales in the Continental United States (CONUS).

The research is therefore directly responsive to the research objective in RCSON-18-L2 of "Discernment of relationships (e.g., mean temperature and heat wave frequency or severity) as revealed in both historical observations and climate models with an emphasis on describing model heterogeneity.

This research has been designed to lay the groundwork for tangible benefits to DoD by establishing a physical basis for clarity in (1) the risk that future changes in the distribution of temperature and precipitation that a single DoD facility may face for a set of plausible emissions scenarios and (2) the risk that multiple DoD facilities may face simultaneously in the future due to these scenarios. The research first focused on the use of traditional geostatistical techniques including spatiotemporal variography and kriging to blend historical data and GCM projections to examine the spatiotemporal behavior of temperature and precipitation over the CONUS in future emissions scenarios.

As a corollary to this primary objective, we pursued a secondary objective: as per the needs expressed in RCSON-18-L2: we explored the assumptions inherent in pursuing a statistical approach to understanding the risks associated with nonstationarity, and whether physical modeling and understanding are needed to bringing clarity to understanding these risks under future climate change.

Technical Approach

The null hypothesis of this investigation was that the statistics of temperature and precipitation and their covariability in the CONUS on regional scales have been and will continue to be stationary. This technical approach describes the methods that we used to test this hypothesis. At the start of the investigation, we pursued a purely statistical approach to this question. That is, we sought to develop projections of changes in the distributions of temperature and precipitation based on a combination of model projections where the projections were weighted according to the model's historical performance relative to observations.

Pursuant to this strategy, we developed a multi-linear regression based on the three-dozen models from the Couple Model Intercomparison Project – Phase 5 (CMIP5) (Taylor et al 2012) archive that was downscaled using the Localized Constructed Analogues (LOCA) technique (Pierce et al, 2014). We developed a regression approach of the historical model simulations by comparing those against a gridded hydrometeorological product (Livneh et al, 2013). We derived biases as a function of quantile for each model and used this as a basis to develop a multi-linear regression. As part of this process, we were able to develop exceedance probabilities for the projections based on a generalized extreme value (GEV) distribution. We also developed a number of techniques to manage potential numerical instabilities associated with the construction of these projections. We present a flow-chart of this approach in Figure 2 and describe them in detail in Feldman et al (2019, Submitted) and Tadić and Biraud (2019, Submitted). We also developed a set of more advanced statistical techniques to manage potential changes in the spatiotemporal correlation in temperature and precipitation, through a copula-based technique (Tadić and Biraud, 2019, Submitted).

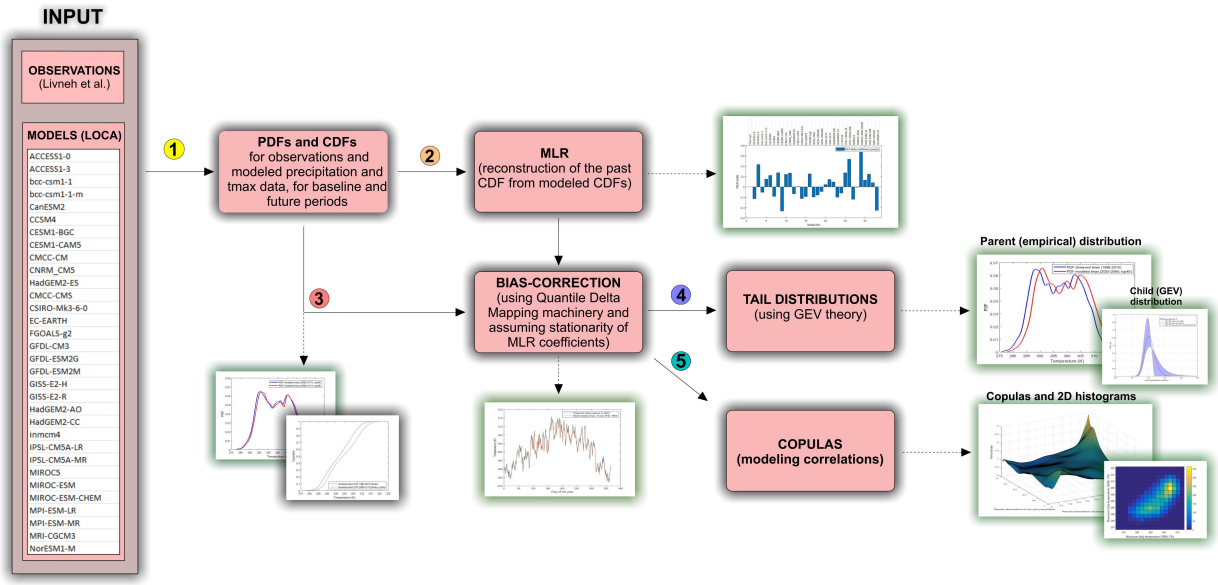


Figure 2: Flow diagram of a regression-based analysis. (1) Input data (observational and model data) are first converted into empirical Probability Distribution Functions (PDFs) and Cumulative Distribution Functions (CDFs). (2) Multi-linear regression (MLR) coefficients are constructed to reconstruct observational CDFs using model CDFs. (3) Quantile Delta Mapping (Cannon et al, 2015) QDM is applied to develop future CDFs. (4) We apply Generalized Extreme Value (GEV) theory and examine change in the likelihood of extreme events. Finally, (5) copula theory is applied to assess whether climate change increases risk of simultaneous or collocated extreme events. From Tadić and Biraud (2019, Submitted).

With the mechanics of what amounts to a pure statistically-based approach to developing projections to assess future changes for a given location and correlated risks for multiple locations, we sought to revisit key assumptions that were made in the course of constructing this statistical model.

Based on key feedback at the SERDP Symposium from Dr. Linda Mearns of NCAR, and following the suggestions of Dr. Michael Wehner of LBNL, we recognized the importance of adopting a more cautious approach to performing analysis on a downscaled multi-model ensemble. Specifically, we looked into the key assumptions underlying the approach that we took, including that models are independent and that the stationarity assumptions in downscaling techniques are tenable. Through discussions with the co-performers on this proposal, it became clear that a careful intercomparison of different downscaling solutions provides information not just on the downscaling technique, but on the parent model as well.

A deeper understanding of the strengths and weaknesses of the parent models and downscaling techniques, which can be informed by the intercomparison of different downscaling solutions, is needed, along with a robust test of stationarity assumptions in downscaling. With a multi-pronged approach to testing parent models and downscaling techniques, we can develop a rational basis for weighting or deweighting climate model projections and downscaling techniques. Our original approach to developing projections represents a default set of assumptions regarding the *a priori* equality in the independence, uniqueness, and plausibility of

model projections and stationarity in statistical downscaling relationships between the 20th and 21st Century. We can test these assumptions and present an overview for doing so at the conclusion of this Executive Summary and present a detailed approach for accomplishing this at the conclusion of this Report.

Results and Discussion

Using the approach outlined in Figure 2, we first focused on the development of projections of daily precipitation and surface air temperature around two selected DoD facilities: Travis Air Force Base (TAFB) near Fairfield, California, and Fort Bragg (FBR) in North Carolina. Below is a summary of average and extreme changes for these two locations. The changes of mean temperature are not necessarily monotonic with time and with emissions scenario, though they are large, and the changes in mean precipitation are moderate.

	TAFB			
	RCP45		RCP85	
Year (centered around)	Precipitation (mm/day) (delta)	T_{max} (K) (delta)	Precipitation (mm/day) (delta)	T_{max} (K) (delta)
2020	+0.1	+1.1	+0.2	+1.1
2050	-0.0	+2.0	+0.0	+1.9
2100	-0.1	+2.2	+0.0	+2.7

Table 1: Changes in average daily precipitation and maximum daily surface temperature values (future-past) for TAFB for three future periods centered around years 2020, 2050 and 2100, derived from bias-corrected future projections using MLR and past observations, for two emission scenarios (RCP4.5 and RCP8.5).

	FBR			
	RCP45		RCP85	
Year (centered around)	Precipitation (mm/day) (delta)	T_{max} (K) (delta)	Precipitation (mm/day) (delta)	T_{max} (K) (delta)
2020	-0.5	+0.9	-0.3	+0.1
2050	-0.4	+1.2	-0.3	+1.5
2100	-0.5	+1.6	-0.4	+2.2

Table 2: Changes in average daily precipitation and maximum daily surface temperature values (future-past) for FBR for three future periods centered around years 2020, 2050 and 2100, derived from bias-corrected future projections using MLR and past observations, for two emission scenarios (RCP4.5 and RCP8.5).

Additionally, we can use this approach to develop estimates of changes in extreme values of precipitation and daily maximum temperature in the 21st Century and associated confidence

intervals using GEV parameter estimates. These changes are a function of emissions scenario, which becomes a large driver of the divergence in estimates by the end of the 21st Century (Hawkins and Sutton, 2009).

	TAFB				FBR			
	Precipitation (mm/day) (95% CI)		T_{max} (K) (95% CI)		Precipitation (mm/day) (95% CI)		T_{max} (K) (95% CI)	
	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5
2005-2035	1.09 (0-16.5)	2.77 (0-20)	4.35 (0-21.4)	4.63 (0-21.5)	0.30 (0-10)	0.47 (0-8)	7.58 (0-35.5)	1.86 (0-26)
2035–2065	0.20 (0-11)	1.87 (0-17)	11.35 (0-31.1)	11.63 (0-31.5)	0.23 (0-9)	0.33 (0-8)	11.83 (0-38)	17.65 (0-41.5)
2085–2115	0.06 (0-10)	2.97 (0-19)	13.60 (0-33.4)	22.45 (0.7-42.5)	0.21 (0-9)	0.25 (0-9)	19.98 (0-45)	31.04 (0-52.5)

Table 3: The ratio of probability of the extreme event in the future that corresponded to p<0.01 in the past, and its probability for three future periods (centered around years 2020, 2050 and 2100), for both TAFB and FBR locations under two emission scenarios, based on GEV distribution. The uncertainty is reported as 95% Confidence Intervals (CI).

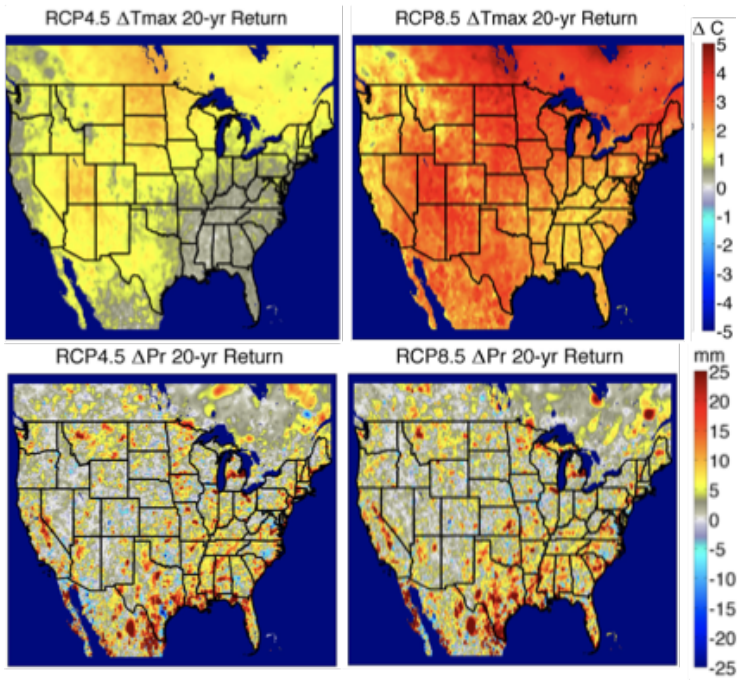


Figure 3: Change in 20-year daily return value of maximum surface air temperature and precipitation based on the MLR approach across the CONUS for two emissions scenarios from 1985-2015 to 2070-2100 .

This analysis can be extended to develop distributions across the CONUS. We summarize these results by showing the change in the 20-year daily return value of maximum daily temperature and precipitation in Figure 3.

This approach provides projections, but the steps shown in Figure 1 need to be approached cautiously. In point of fact, seemingly straightforward tasks, such as the development of gridded historical data from a discrete set of weather station observations, can introduce biases. Co-Performers Risser and O'Brien recently published a paper showing that gridding observations will tend to underestimate extreme precipitation because that precipitation varies spatially, and averaging high and low precipitation data will dampen extremes (Risser et al, 2019a). A better approach is to use station data to estimate extreme precipitation statistics and develop spatial modeling to produce gridded observational products.

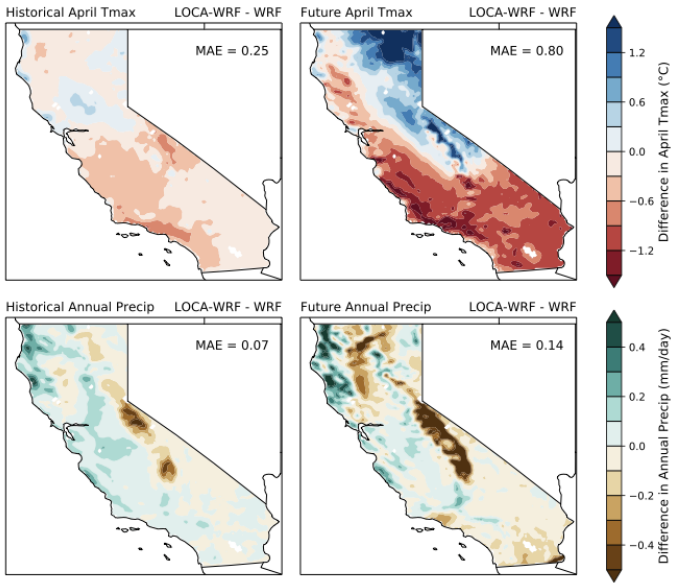


Figure 4: Differences between LOCA-WRF and WRF projections for historical and future April maximum temperature (panels a and b, respectively) and for historical and future annual mean precipitation (panels c and d, respectively). Unit for (a) and (b) is degrees C. Unit for (c) and (d) is millimeters per day. The mean absolute error (MAE) is reported in the upper right margin of each panel. From Walton et al, JAMC, In Review.

Additionally, recent work by Walton et al, (JAMC, In Review), as shown in Figure 4, indicates that the regression approach is suboptimal. An intercomparison of different downscaling solutions across California for historical and future simulations reveals a divergence that only appears later in the 21st Century. A closer look reveals that the dynamically-downscaled solution is qualitatively and quantitatively different from the statistical downscaling solution. The reason that the historical performance of the statistical downscaling solution does not translate well into future performance requires careful analysis. Walton et al, (JAMC, In Review) conducted this analysis and found that, for California, the downscaling methodology itself is only a symptom of a larger problem. The primary driver for the divergence in downscaling solutions in future projections is that the parent model does not capture changes in snow cover well, which greatly

impact temperatures, and does not capture changes in North American Monsoon dynamics, which affects Southeastern California.

These findings show that the blind development of regression solutions based on historical biases is suboptimal, because the historical biases of some models are not directly translatable to future biases. Rather, it is the process representations of snow albedo and monsoons that must be considered in the development of projections. It is precisely this nonstationarity that RCSON-18-L2 seeks to address head on. The suboptimal performance of MLR-based projections, where the regression is based on the univariate performance of the models in the historical record, is only revealed with the intercomparison of dynamical and statistical downscaling solution.

Implications for Future Research and Benefits

This research has explored the landscape of developing projections of surface air temperature and precipitation at the DoD facility level from an ensemble of downscaled climate model projections. The research has shown that this very landscape is not easily navigated because the climate models have representation errors in the processes that can lead to nonstationarity and statistical downscaling solutions are constructed explicitly with a stationarity assumption that has not been tested.

The primary benefit to the DoD of this seed project's research is to detail the scale of the work that is needed to address in a comprehensive fashion specific research tasks that go beyond a purely statistical approach to developing climate model projections from an ensemble of model results to provide insight into the models with which the DoD should have increased or diminished confidence. The challenges of using downscaled projections for infrastructure and operations planning should not be underestimated, because they are generally completely opaque to the end-user. These end-users can easily have either too much or too little confidence in the model projections. For the former case, the provision of a set of numbers, and even error bars, that can be fed into end-user models is a straightforward exercise, but the danger here is that those numbers can be misleading if they do not consider the processes that break stationarity.

For the latter case, the esoteric nature of climate models and downscaling can preclude their adoption even though it is very likely that long-term shifts in surface air temperature and precipitation distributions will occur. In this situation, planning exercises either use historical data, that does not include nonstationarity, or use generic information on increased variability as a placeholder for a deeper understanding of the risks associated with climate change, and this generic information may or may not be relevant to a specific location.

Fortunately, there are tractable solutions to these challenges. Focused research efforts that develop a deeper understanding of the processes that "break" stationarity and evaluate model projections accordingly are needed. Additionally, it is critical that the researchers engaged in developing climate model projections bridge the communication divide between the modeling community and end-users so that they have the appropriate level of confidence in these projections. Ultimately, end-users need to understand the challenges and opportunities with using climate model projections for infrastructure and operations planning and so that they can utilize model projections judiciously. It is incumbent on those with knowledge and expertise in climate model projections to guide the utilization of these data for end-users.

Full Report

Objective

This investigation, arising in response to RCSON-18-L2 "Quantifying the Effects of Environmental Nonstationarity at DoD Relevant Scales," seeks to uncover the heterogeneous performance of the General Circulation Models (GCMs) with respect to nonstationarity of temperature and precipitation. The primary objective of this research is to build the foundation for a comprehensive determination of this risk by using historical observations to inform model projections of the statistical properties of temperature and precipitation, including nonstationarity, at regional scales in the Continental United States (CONUS).

The research is therefore directly responsive to the research objective in RCSON-18-L2 of "Discernment of relationships (e.g., mean temperature and heat wave frequency or severity) as revealed in both historical observations and climate models with an emphasis on describing model heterogeneity.

This research has been designed to lay the groundwork for tangible benefits to DoD by establishing a physical basis for clarity in (1) the risk that future changes in the distribution of temperature and precipitation that a single DoD facility may face for a set of plausible emissions scenarios and (2) the risk that multiple DoD facilities may face simultaneously in the future due to these scenarios. The research first focused on the use of traditional geostatistical techniques including spatiotemporal variography and kriging to blend historical data and GCM projections to examine the spatiotemporal behavior of temperature and precipitation over the CONUS in future emissions scenarios.

As a corollary to this primary objective, we pursued a secondary objective: as per the needs expressed in RCSON-18-L2: we explored the assumptions inherent in pursuing a statistical approach to understanding the risks associated with nonstationarity, and whether physical modeling and understanding are needed to bringing clarity to understanding these risks under future climate change.

Background

According to Milly et al. (2008) and as quoted in the Climate Change Installation Adaptation and Resilience 2017 Report, "stationarity is dead," and "In the new reality of non-stationarity, projections about climate must assume a constantly evolving basis." But how does this translate practically to making critical infrastructure decisions?

Large infrastructure decisions are made based on average temperature and precipitation, their variability, and extreme events, and such decisions may require many years of lead-time. Already, there are significant changes to the distributions of temperature and precipitation that have been observed. These have been recently summarized in the 4th National Climate Assessment and are shown in Figure 5.

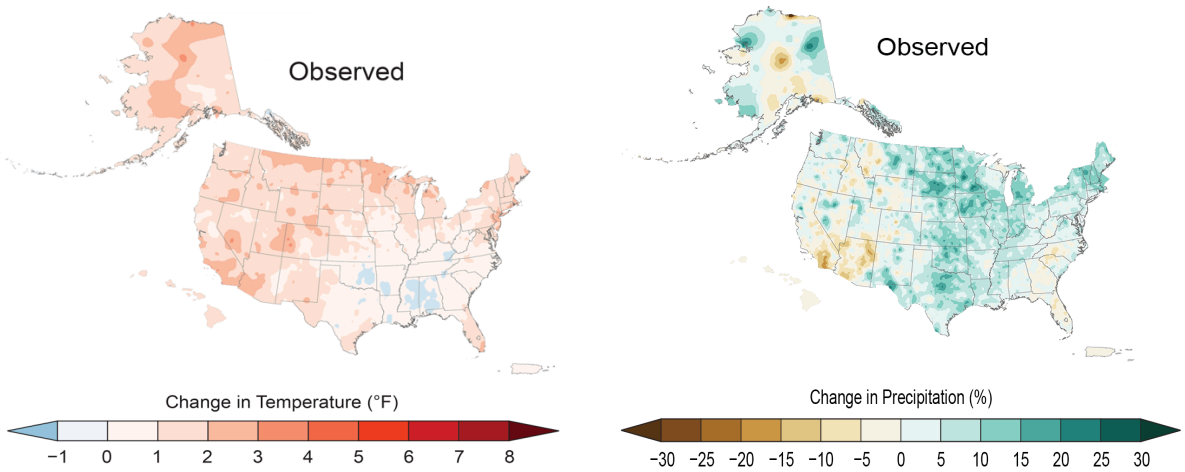


Figure 5: Left panel: Observed change for 1986–2016 (relative to 1901–1960 for the contiguous United States and 1925–1960 for Alaska, Hawai'i, Puerto Rico, and the U.S. Virgin Islands) as adapted from Vose et al, 2017 for the 4th National Climate Assessment. Right panel: same as left but for observed precipitation changes for the period 1986–2015 (relative to 1901–1960 for the contiguous United States and 1925–1960 for Alaska, Hawai'i, Puerto Rico, and the U.S. Virgin Islands as adapted from Easterling et al, 2017 for the 4th National Climate Assessment.

It is expected that these observed changes will not necessarily persist in the 21st Century, as long term shifts in the Earth's climate system become more pronounced and the effects that result across the United States, particularly for the critical societal variables of surface air temperature and precipitation, shift further. Projections for the 21st Century have uncertainty due to natural variability, model performance, and human activities. By the end of the 21st Century, that uncertainty in human contributions dominates the uncertainty in projections (Hawkins and Sutton, 2009).

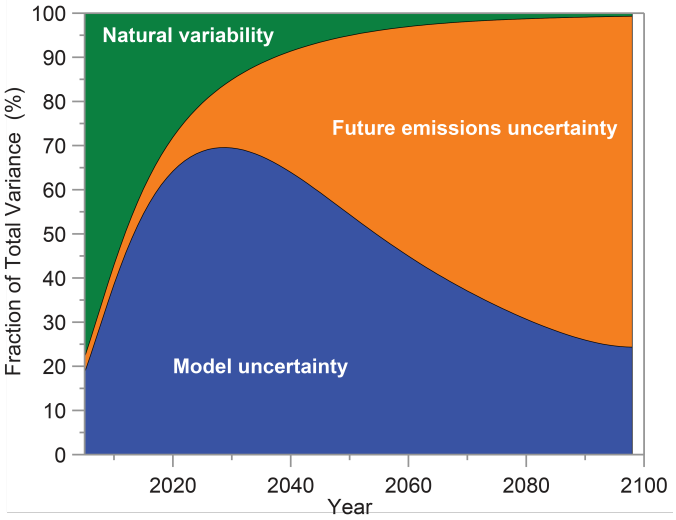


Figure 6: The fraction of total variance in decadal mean surface air temperature predictions explained by the three components of total uncertainty is shown for the lower 48 states (similar results are seen for Hawai'i and Alaska, not shown). Orange regions represent human or scenario uncertainty, blue regions represent model uncertainty, and green regions represent the internal variability component. Adapted from Hawkins and Sutton, (2009).

Despite this uncertainty, regional patterns of projected change in temperature and precipitation have been developed for the 21st Century.

Figure 7 shows a summary of these changes for surface air temperature and highlights how warming becomes more severe as the 21st Century progresses, and how the higher emissions scenario, as detailed in the Representative Concentration Pathway 8.5 (RCP8.5) (Van Vuuren et al, 2011), which roughly represents business-as-usual conditions, leads to greater warming than the Representation Concentration Pathway 4.5 (RCP4.5) emissions scenario.

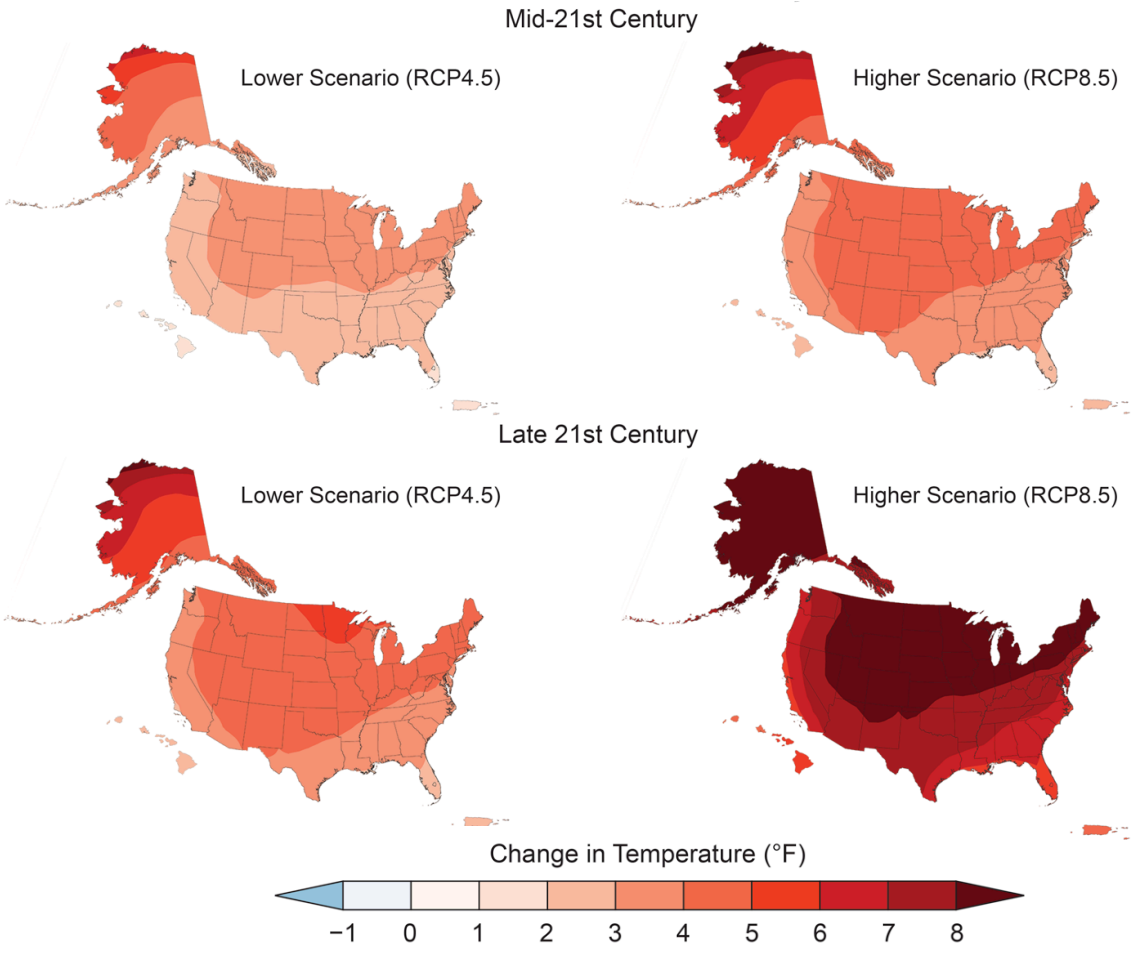


Figure 7: Annual average temperatures across North America are projected to increase, with proportionally greater changes at higher as compared to lower latitudes, and under a higher scenario (RCP8.5, right) as compared to a lower one (RCP4.5, left). This figure shows projected differences in annual average temperature for mid-century (2036–2065, middle) and end-of-century (2070–2099, bottom) relative to the near-present (1986–2015). Adapted from Vose et al. 2017 for the for the 4th National Climate Assessment.

The patterns of projected precipitation change across the United States are expected to be more spatially variable and, because they arise due to very different processes from surface air temperature, there is greater uncertainty in these projections.

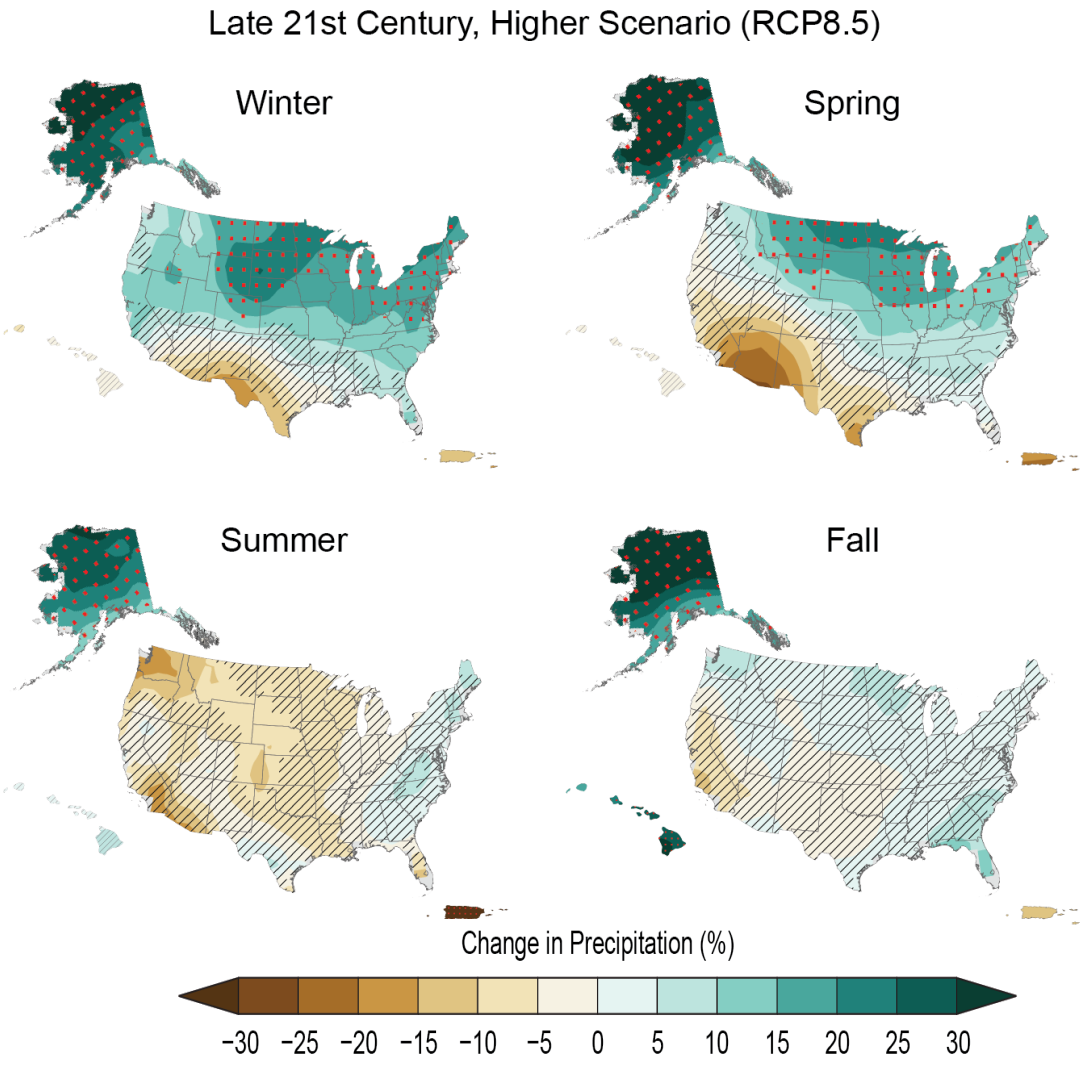


Figure 8: Projected precipitation changes vary by region and season. In the future, under the higher scenario (RCP8.5), the northern United States, including Alaska, is projected to receive more precipitation, especially in the winter and spring by the period 2070–2099 (relative to 1986–2015). Parts of the southwestern United States are projected to receive less precipitation in the winter and spring. Areas with red dots show where projected changes are large compared to natural variations; areas that are hatched show where changes are small and relatively insignificant. As adapted from Easterling et al. (2017) for the 4th National Climate Assessment.

These changes can be difficult to predict, most especially since they are without an historical analogue. Therefore, the scientific community must rely on detailed computer models for developing predictions of change. General Circulation Models (GCMs), sometimes referred to

as global climate models, are one such widely-utilized tool for developing changes in the large-scale distribution of temperature and precipitation as a function of future human activities, which are conventionally divided into discrete greenhouse gas emissions scenarios (van Vuuren et al, 2011). The primary rationale for using GCMs is that they are physics-based, and therefore describe long term shifts in the Earth's climate system that are without an historical analogue. From these models, we can deduce the range of global-scale temperature response to the forcing associated with a doubling of carbon dioxide is likely not more than 4.5 °C, which has been established through a sustained focus on climate sensitivity (Charney et al, 1979; Knutti and Hegerl, 2008; Forster et al, 2016; Knutti et al, 2017). The response of the Earth's hydrological cycle as it adjusts to warming has also been studied with GCMs, and gross features, such as the global hydrological cycle response to warming (Jeevanjee and Romps, 2018), the expansion of the troposphere (Lu et al, 2007; Hu et al, 2013), and shifts in the storm tracks (Yin, 2005; Bengtsson et al, 2006; Ulbrich et al, 2008), have been observed and have a strong theoretical basis. GCMs physically model these features and so confidence in large-scale projections can be achieved.

However, these models are also blunt tools for finer-scale questions. These models cannot span the spatial and temporal scales of the processes that interact to control the Earth's climate system, which, in the former case, range from the micrometer-scale for cloud droplet formation to scales of tens of thousands of kilometers for global-scale circulation patterns, and in the latter case, range from sub-second for cloud formation, precipitation, and radiation processes, to at least one thousand years for the equilibration of the climate system. Furthermore, the numerical integration of the intricate computer program that represents a GCM must balance a number of competing factors including, but not limited to, top-of-atmosphere energy balance, large-scale circulation, cloud formation and dissipation, ocean dynamics, and earth system response (Mauritsen et al, 2012; Hourdin et al, 2017; Schmidt et al, 2017). The balance that the developers must achieve is often referred to as model calibration or tuning. Model development and calibration often focus on achieving unbiased assessments of mean temperature and precipitation at larger spatial-scales and long time-scales. Because regional variability in precipitation and precipitation change across models is quite broad (Pendergrass and Hartmann, 2014), it is not known *a priori* that a given model will produce skillful temperature and precipitation projections as they relate to DoD facilities. Additionally, as we will show, it is also not known *a priori* whether models have the right process representations such that skillful model performance over the historical record relative to observations is necessarily indicative of skillful model performance for future projections.

In spite of these challenges, the climate modeling community has undertaken the development first-order estimates of model uncertainty, developed an ongoing series of model intercomparison projects, which fall under the umbrella of the Coupled Model Intercomparison Project (CMIP). CMIP phases recognize the heavy level of parameterization that is necessary to construct global models that span the range of temporal and spatial scales that are climate-relevant. The original international, widely-utilized CMIP phase was CMIP3 (Meehl et al, 2007), which informed the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC). CMIP5 represented the next major international effort and was completed around 2014 (Taylor et al, 2012), with simulations supporting the Fifth Assessment Report (AR5) of the IPCC. Most recently, CMIP6 (Eyring et al, 2016) represents the current

major international effort that provides modeling support to the Sixth Assessment Report (AR6) of the IPCC. These exercises are designed to simultaneously serve several different purposes including (1) establishing a range of emergent properties such as globally-averaged surface air temperature for participating models, (2) testing model sensitivities to different configurations, and (3) understanding earth system processes. A number of different experiments are contained within CMIP6, including the Shared Socioeconomic Pathways (O'Neill et al, 2014; O'Neill et al, 2017) that have close heritage with the Representative Concentration Pathways (van Vuuren et al, 2011) of CMIP5 (Taylor et al, 2012) and represent efforts to develop projections for the 21st Century, rather than testing model sensitivities. For the purposes of this investigation, there are three key experiments in CMIP6 that seek a better understanding of model resolution, downscaling, vulnerabilities and impacts. These include the North-American Coordinated Regional Downscaling Experiment (NA-CORDEX) (Giorgi et al, 2009; Jones et al, 2011; Mearns et al, 2012), the High Resolution Model Intercomparison Project (HighResMIP) (Haarsma et al, 2016), and the Vulnerability, Impacts, Adaptation, and Climate Services (VIACS) (Ruane et al, 2016).

At present, 151 modeling activities (one model running one set of experiments for one model intercomparison project constitutes an activity) have been reported to the Earth System Grid Federation (Cinquini et al, 2014) for CMIP6. The contributing modeling centers are expected to produce between 20 and 40 Petabytes of climate model simulations as part of the CMIP6 exercise (Eyring et al, 2016). This is far too voluminous for planning purposes, and therefore, there is a strong need to curate these results to determine ranges of conditions that are likely to be faced locally. it is expected that there will be many participating modeling centers in each of these intercomparison projects, and there is a need to determine how to interpret the range of results that arise thereof specifically in terms of DoD facility-level risk.

Nominally, model intercomparison projects serve to characterize the spread of model results associated with a given emissions scenario, but it has been recently recognized that navigating through the multi-model ensemble to understand the information contained in the ensemble's projections must consider both model errors with respect to observations and the interdependence of models (Sanderson et al, 2015; Knutti et al, 2017; Sanderson et al, 2017; Eyring et al, 2019). This navigation depends on the specific question being asked. Model democracy, where each simulation is treated as independent and unbiased, has been found to be suboptimal, primarily owing to the fact that models have not been developed independently and may even have substantial similarities in their code bases (Sanderson et al, 2015). Rather, it is essential that models being considered exhibit a level of skill with respect to regional metrics (Rupp et al, 2013). Taking California as an example, only models that reproduce the correlation and variance of mean seasonal spatial patterns, amplitude of seasonal cycle, diurnal temperature range, annual- to decadal-scale variance, long-term persistence, and Western United States regional precipitation teleconnections to El Niño Southern Oscillation (ENSO) are appropriate for hydroclimate studies, and only 10 of the 33 CMIP5 models exhibited skill with respect to this metric (Pierce et al, 2018; Thorne et al, 2018). The 4th National Climate Assessment (NCA4) (USGCRP, 2017) adopted a model-weighting approach to produce, from the CMIP5 ensemble, regional-scale estimates of change in average temperature and precipitation. The results broadly indicate high confidence in predicting more warming with the Representative Concentration Pathway 8.5 (RCP8.5) scenario than the RCP4.5 scenario across the Conterminous United States

(CONUS), and, with less confidence, more variability in regional precipitation for both of these scenarios.

It is critical that the climate change signals that are being captured by state-of-the-art climate models are translated to the local level and that the process of doing so especially does not underestimate the risks that climate change effects may have on local infrastructure. Unfortunately, raw climate model output can exhibit biases in temperature and precipitation at regional scales, and these biases will propagate in complex and potentially nonlinear ways into subsequent models and decision-making efforts (Christensen et al, 2008; Mearns et al, 2012; Sillmann et al, 2013). To address this issue, one approach, while not without controversy (Ehret et al, 2012; Maraun 2012; Maraun 2013) is to bias-correct the raw GCM output based on historical observations (Gudmundsson et al, 2012; Teutschbein and Seibert, 2012; Teutschbein and Seibert, 2013; Pierce et al, 2015).

An additional and critical step involves downscaling to address the gross mismatch in the spatial resolution on which GCMs operate and the processes that impact the local-level. In order to manage the computational expense of running a GCM over the temporal and spatial range of experiments in the phases of CMIP (e.g., CMIP3, CMIP5, CMIP6), the spatial resolution of these models is on the order of 100 km. As shown in Figure 9, significant information may be destroyed with coarse resolution simulations, particularly over complex terrain.

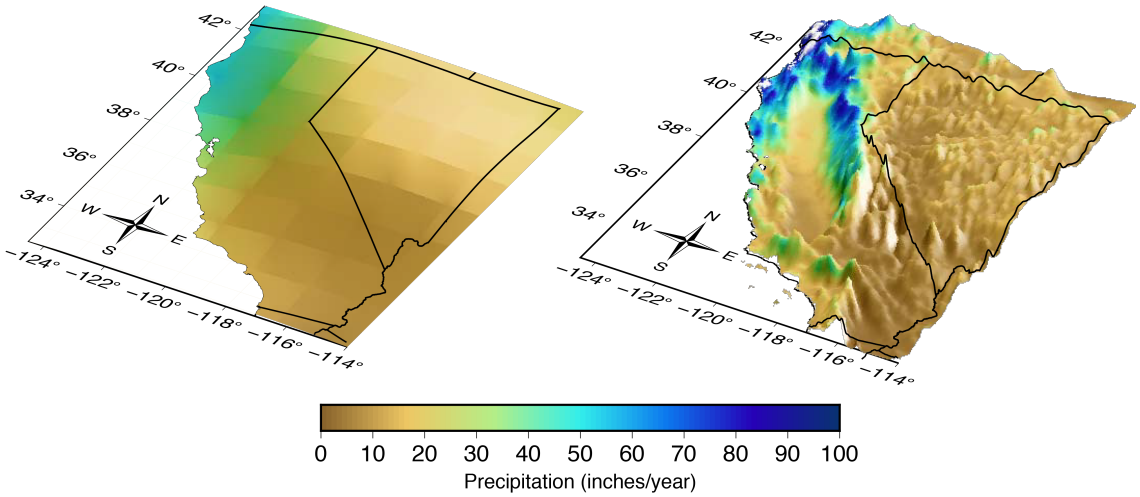


Figure 9: Annual precipitation in California and Nevada (inches) in the CNRM-CM5 model with a resolution of 100 mile (left), and after using a statistical model to account for the effects of topography at a 3.6 mile resolution (right). The global model only has a few grid cells over the entire state of California, so does not resolve the coastal mountain ranges, interior valley, or Sierra Nevada Mountains on the border with Nevada. The precipitation field in the right panel, by contrast, captures the wet conditions on the west slopes of the mountains, and the dry, rain shadow region to the east of the mountains. The vertical scale has been exaggerated for clarity, and by the same amount in both panels. From the 4th National Climate Assessment.

Downscaling techniques have been developed to provide spatial details of how climate differences and variability impact, in particular, hydrology, since the coarse data from regional projections of the climate model provide little actionable information for impacts

on watersheds and rivers. There are a wide variety of downscaling techniques and they can be divided conceptually into two categories: dynamical and statistical. With the former method, detailed, high-resolution process modeling is used to represent the influence of topography, atmospheric processes, and land-atmosphere interactions at a level of detail that is much greater than is represented in GCMs. These nominally incorporate physics and translate the regional features of the parent GCM to the local level. Dynamical methods have two primary disadvantages, however: (1) they are quite computationally expensive, to the point of being generally infeasible for multi-model ensembles, and (2) the process models still do not resolve all of the relevant physics and themselves have parameterizations that, while testable, are generally used without validation. They can easily exhibit biases relative to observations that may or may not have bearing on future downscaled projections.

Statistical methods are built from observed relationships between the coarse and fine scale. These are computationally efficient and, because they are based on observations, exhibit negligible error during the observational record. Their disadvantage lies in the assumption of stationarity and cannot directly incorporate climate change signals that may violate this assumption under climate change conditions.

There are a number of other strategies for achieving higher-resolution projections including variable-resolution modeling which has the advantage over dynamical methods of having consistency with the parent model and conservation of energy, mass, and momentum fluxes at the interfaces between the region of interest and its circumscribed area.

In addition to the need for downscaling of GCM projections, the role of local and regional biases must be directly addressed. As shown in Figures 10 and 11, downscaled models agree with observations only at spatial scales larger than the CONUS, and, as a corollary to this, exhibit first-order biases relative to observationally-derived datasets.

These patterns indicate that, to first order, biases in model calculations over the historical period could preclude meaningful future analysis. There are several different approaches to handling biases and they must be considered within the context of making future projections. First, model results for a given grid-box can be bias-corrected (i.e., adjusted) by a univariate value determined by that model's average historical performance relative to observational products. Second, model results can be adjusted by a multivariate value derived for a model's historical performance relative to observational products for a given quantile. Regardless of the historical bias-correction, it is an open question regarding an appropriate approach for translating the historical bias into future projections.

The result of these analysis indicate bias, but it is also quite critical to recognize that models are not necessarily independent of each other. They share many parameterizations, which is reasonable where they are physics-based, but is problematic where models essentially share errors. This can potentially challenge the construction of model weights, since weighting that does not consider model independence can produce biased results by inadvertently overweighting a given parameterization. The widely-utilized strategy of Sanderson et al (2015)

focuses on developing weights based on historical model performance and historical model independence (derived through estimates of error correlation). Figure 12 illustrates this point. At the same time, it is not known *a priori* (and therefore must be shown) that model independence, which is generally assessed during the historical record, has implications for model independence for future projections Furthermore, the nonlinearities of model parameterization interactions within a GCM can greatly complicate efforts to weight models by a combination of performance and independence (Francis Zwiers, Pers. Comm., 2019) and would suggest that less-presumptive, model-democratic approah is warranted.

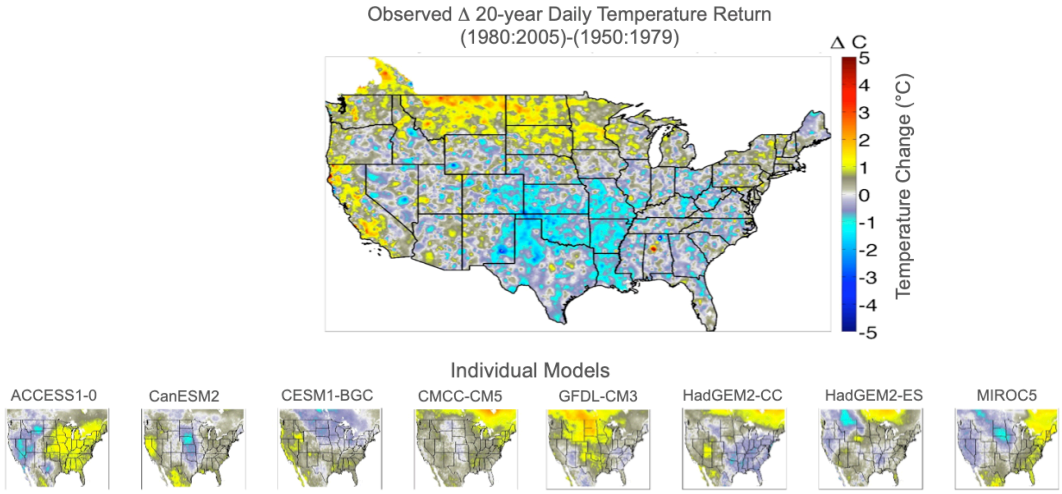


Figure 10: Top panel: historical performance of models in the change in 20-year daily maximum temperature return value between 1980-2005 and 1950-1979 as derived from the Livneh hydrometeorological datasets (Livneh et al, 2013) and return values derived from a Type-I GEV distribution. Bottom panel: corresponding changes in the 20-year daily maximum temperature return value from 8 CMIP5 models performing a historical simulation extending through 2005.

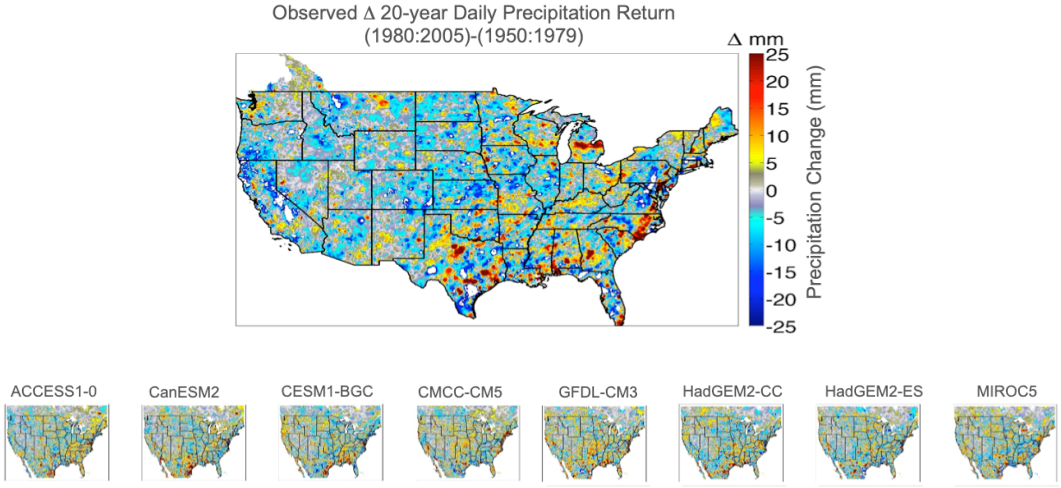


Figure 11: Top panel: historical performance of models in the change in 20-year daily precipitation return value between 1980-2005 and 1950-1979 as derived from the Livneh hydrometeorological datasets (Livneh et al, 2013) and return values derived from a Type-I GEV distribution. Bottom panel: corresponding changes in the 20-year daily maximum temperature return value from 8 CMIP5 models performing a historical simulation extending through 2005.

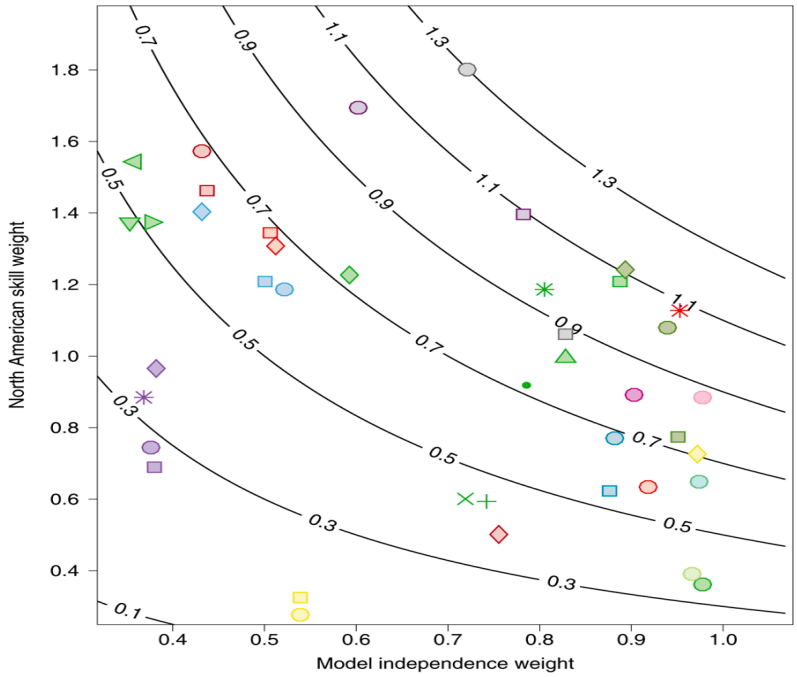


Figure 12: Scatterplot of model skill and independence. Note that the upper-right quadrant, indicating high-skill and high-independence, is unoccupied. Skillful models are correlated and not independent while unskilled models are more independent. By extension, there should be reduced confidence in the latter models. From Eyring et al (2019).

Given this fraught landscape, in November 2017, the SERDP Program released a Workshop Report entitled "Nonstationary Weather Patterns and Extreme Events" and found that "[improving] modeling and tools to develop, apply, and evaluate scenarios of multiple interacting stresses, including training in underlying assumptions and application of scenarios" and "[validating] current [downscaling] methods, [developing] and [testing] new approaches, and [providing] guidance for different situations, locations, and variables" were among the greatest needs and opportunities for future research (Moss et al, 2018).

Distributions of temperature and precipitation can be constructed based on one or more bias-corrected, downscaled models results, but each step of this process must be critically evaluated, and several questions arise:

- How should individual models and the ensemble of model results be handled to develop confidence intervals?
- Can certain model results be trusted more than others?
- When and where is it appropriate to use, in the course of developing future projections, downscaling techniques and bias-correction based on historical observations?

Through this investigation, we seek to shed light on optimal approaches to developing temperature and precipitation distributions for end-users in spite of the myriad challenges and questions that exist.

Materials and Methods

Conceptual Framework

The challenges associated with the landscape of developing downscaled projections motivates the construction of a rational conceptual framework to which research activities can be mapped and ultimately contribute. Figure 13 presents such a framework through a process-chain that yields DoD facility-level projections that balance the information contained in historical observations with the information provided by a climate model: the relative change in the downscaled, bias-corrected model between future and historical simulations. This process-chain describes a rational framework for ingesting climate model simulations and observations, comparing the two, developing bias corrections and then ultimately producing projections.

At a high level, this framework seeks to manage and balance the information contained in historical observations, historical climate model runs, and future climate model projections. As shown in the upper-right of Figure 13, the framework relies on the long-term data records that are obtained from discrete weather stations. These data provide surface air temperature and precipitation information. There are numerous instances of issues with these data, ranging from outages to biases, but they do provide long-term data records, and since they are based on observations, are necessarily borne of all of the relevant physical mechanisms that lead to the observed temperature and precipitation patterns. Statistical distributions can be constructed from these data to estimate long return-value events (e.g., the 50-year return value in daily temperature or precipitation). Because these observational data are discrete, however, they must be gridded in order to be compared against model simulations. The latter is generally quite coarse in spatial resolution, so downscaling techniques, which balance the coarse information and constraints provided by the model with the spatial resolution needed for measurement model intercomparison, is needed. Even with the downscaling, the historical biases must be corrected for the results to have utility for end users. Additionally, the development of future projections must address the unlikelihood that historical simulations will exhibit biases but future, uncorrected projections will not.

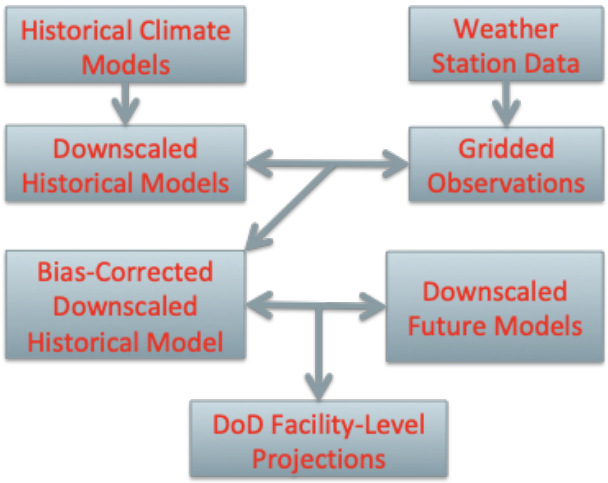


Figure 13: Flow-chart indicating how CMIP model projections are translated into model facility-level risk. Adapted from Feldman et al, JWRPM (Submitted 2019).

The specific implementation of this generic framework, and the issues that must be addressed as part of that implementation, are the focus of this investigation. The details of each step and how the different steps interact matter and must be evaluated critically to ensure, most especially, that biases are not introduced in the course of developing these projections.

Observational Datasets and Gridding

In point of fact, a seemingly straightforward part of the process chain in Figure 13 involves the development of observational products from discrete weather station data records. Figure 14 shows the distribution of these stations across the CONUS.

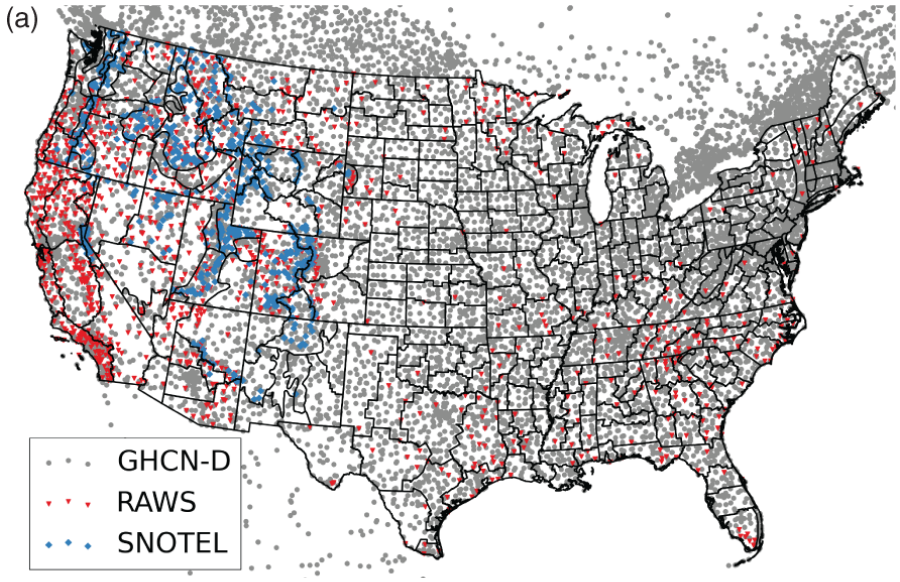


Figure 14: Station networks include the daily Global Historical Climatology Network (GHCN-D), Remote Automatic Weather Stations (RAWS) network, and the Snowpack Telemetry (SNOTEL) network. There are 14,087 stations in total. Each station has ≥5 years of raw data for each month. Boundaries represent US climate divisions. From Oyler et al, 2016.

A number of different gridded products have been developed for these purposes that incorporate spatial statistics to develop gridding, with the Livneh et al, (2013) dataset being widely-utilized for hydrometeorological analyses. Nevertheless, the use of historical precipitation observations to characterize model biases during their historical simulations is surprisingly non-trivial. The principal reason for the need for caution is that precipitation is fractal (Lovejoy and Mandelbrot, 1985). Consequently, the act of gridding existing observationally-based precipitation datasets reduces the magnitudes of extremes: the gridding process effectively spreads localized extremes over multiple grid cells. The spatial correlation patterns of extreme precipitation events are anisotropic and vary seasonally across the CONUS, as shown in Figure 15. As a result, existing precipitation datasets, again derived from observations, exhibit disagreement about which observed events are actually extreme (Timmermans et al., 2018). The work of Risser et al. (2019a) shows that reversing the analysis process — analyzing the statistics of extreme precipitation prior to gridding, and then gridding the parameters of the analysis — can increase

20-year daily precipitation return values over CONUS by over 30 mm compared to existing datasets, as shown in the scatterplots of Figure 16.

The spatial details of the location of these biases are also revealing and are shown in Figure 17: widely-used gridded precipitation datasets, such as from Livneh et al, (2013), are biased low throughout much of the CONUS, with the West Coast and the Southeast exhibiting the greatest low biases, while over the Rocky Mountains, the 20-year daily precipitation return values are biased high by over 30 mm. This implies that the procedures for gridding extreme precipitation depend on location and source of precipitation. The contrast in the sign of the extreme precipitation bias between high-altitude complex terrain and the rest of the CONUS suggests that further caution is needed to develop robust precipitation products against which models can be compared, and that the meaning of model precipitation within a grid box must be specified, especially in complex terrain.

This result also points to the need to ensure that bias-correction is, itself, unbiased, since projections of future precipitation are generally derived from utilizing the relative change in precipitation (Hempel et al, 2013).

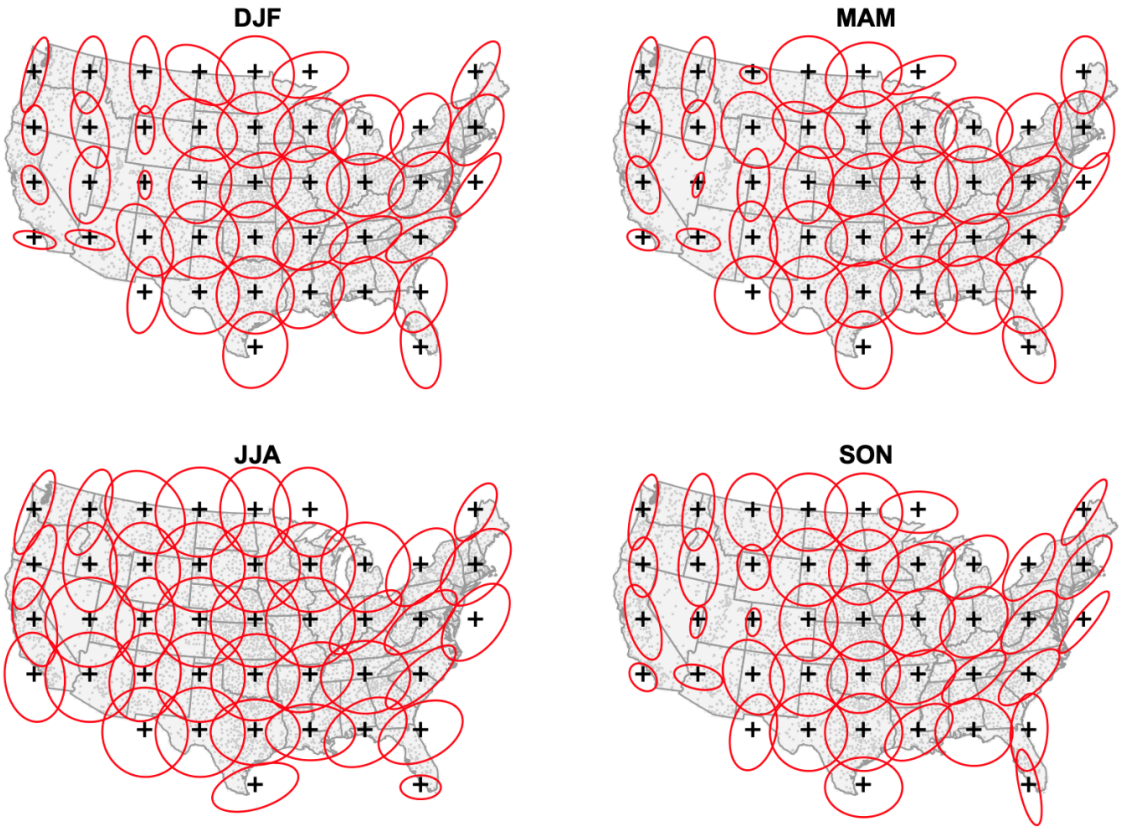


Figure 15: Directional spatial length-scale for the location intercept $\mu_0(\boldsymbol{s})$ of a Type-I GEV distribution for each season, estimated empirically from the data. The ellipses are a heuristic representation of the magnitude and direction of spatial dependence in $\mu_0(\boldsymbol{s})$. From Risser et al (2019a).

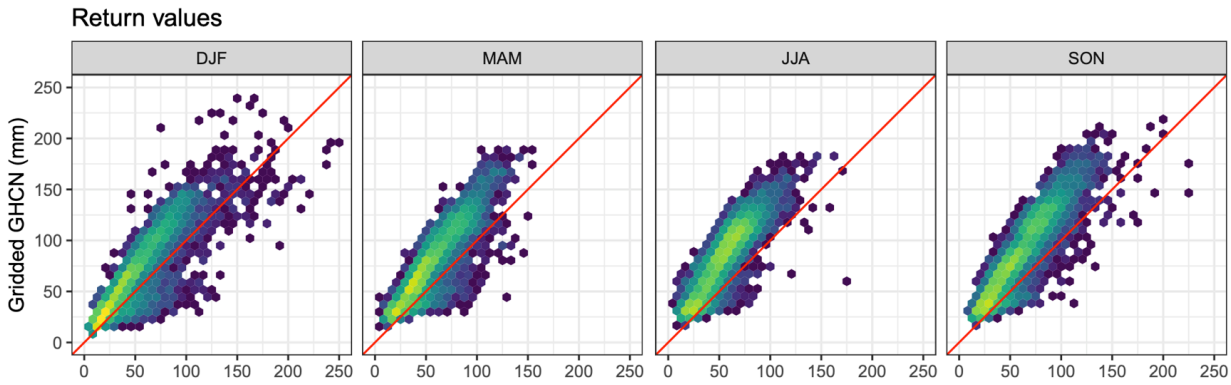


Figure 16: Seasonal density scatter plots of return values, with quantities from the Livneh data product on the x-axis and corresponding quantities from GHCN station data analysis on the y-axis. The red line indicates the 45° line. From Risser et al (2019a).

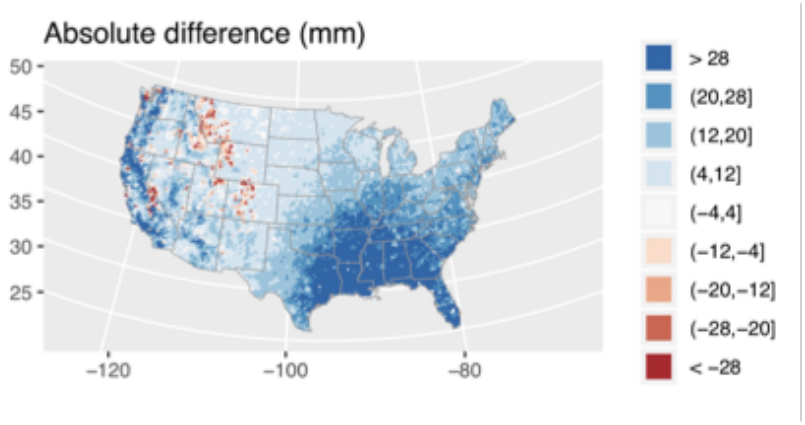


Figure 17: Map of the difference between 20-year daily precipitation return value when extremes are preserved in the gridding process. Red indicates bilinear gridding overestimates extreme precipitation, while blue indicates underestimation. From Risser et al, (2019a).

Needless to say, the other steps in the process-chain described in Figure 13 need to be carefully evaluated. For the first part of this investigation, we took a series of steps to understand the challenges associated with the rest of that process-chain. We started this effort by developing projections based on a straightforward statistical framework using regression principles. The purpose for that effort is to determine what, if any, numerical challenges arise from the development of such a framework for the construction of projections of means and extremes of temperature and precipitation, as well as determining how correlations change. These challenges would preclude the development of useful downscaled projections for end-users, but the lack of such challenges should not be taken as an indication that a regression-based approach is warranted. Rather, the regression-based approach can reveal the general challenges and opportunities with developing a single projection from a multi-model ensemble.

Statistical Analysis Approach

A diagram showing the steps required to project changes in the distributions of temperature and precipitation at a given location is shown in Figure 18. There are several steps shown in this figure: first, probability distribution functions (PDFs) and cumulative distribution functions (CDFs) are constructed from a set of downscaled climate model simulations over the historical record. By comparing these projections against PDFs and CDFs from gridded observational datasets, we construct a set of multi-linear regression (MLR) coefficients. We then develop bias-correction estimates for the historical model runs for each quantile for each GCM. Finally, we use a form of quantile mapping (QM) called quantile delta mapping (QDM) (Cannon et al, 2015), which scales the value of a given quantile of observed temperature by the difference or observed precipitation by the ratio of the future modeled to historical modeled quantile.

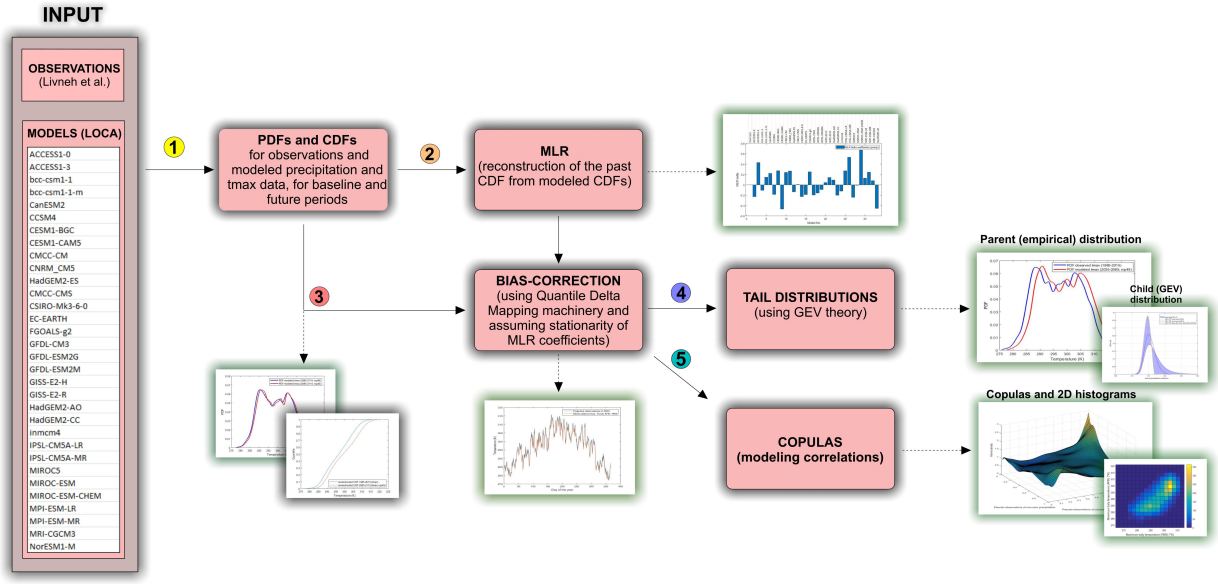


Figure 18: Flow diagram of a regression-based analysis. (1) Input data (observational and model data) are first converted into empirical Probability Distribution Functions (PDFs) and Cumulative Distribution Functions (CDFs). (2) Multi-linear regression (MLR) coefficients are constructed to reconstruct observational CDFs using model CDFs. (3) Quantile Delta Mapping (Cannon et al, 2015) QDM is applied to develop future CDFs. (4) We apply Generalized Extreme Value (GEV) theory and examine change in the likelihood of extreme events. Finally, (5) copula theory is applied to assess whether climate change increases risk of simultaneous or collocated extreme events. From Tadić and Biraud (2019, Submitted).

As part of this framework, we used generalized extreme value (GEV) theory (see Coles, 2001, for detailed information) to model the smallest or largest value among a large set of independent, identically distributed, random values representing measurements or observations. The generalized extreme value distribution combines three simpler distributions (Gumbel, Fréchet and Weibull) into a single form, allowing a continuous range of possible shapes that includes all three of the simpler distributions. Extreme Value Theorem (Rudin, 1976) states that a GEV distribution is the only possible limit distribution of properly normalized maxima of a sequence of independent and identically distributed random variables. In practice, modeling a GEV distribution of extreme events assumes taking block maxima (or minima) and fitting the GEV distribution to data.

In this study, a natural choice of the block size is one year to avoid aliasing first-order seasonal cycle signals into the analysis, so we use annual extremes ("block maxima") of T_{max} and precipitation over 30-year-wide windows for the reference baseline period (1985-2015), and for future periods centered around years 2020, 2050 and 2100 for both emission scenarios, and then fit GEV distributions to the resulting subset of data. Then we estimate probabilities for resulting extreme values that exceed chosen thresholds.

Copulas are functions that capture the dependence structure of joint distribution functions and provide a way to model dependencies among variables (Charpentier et al., 2007; Schmidt, 2007). The basic requirement for modeling copulas is having uniform marginal probability distributions of the variables whose dependency is modeled. This requires prior transformation of the original data into new entities that we will refer to as pseudo-observations throughout this paper. Modeling copulas is partially based on Sklar's theorem (Sklar, 1959), which states that if the marginal distributions for all variables whose modeled dependence structures are continuous random variables, then their dependence structures (copulas) are unique. Notice that T_{max} and (non-zero) precipitation represent continuous random variables, and thus fulfill the requirements for modeling using copulas.

Copulas, like common PDFs and CDFs, can be modeled using parametric or non-parametric approaches. The empirical copula can be thought of as an estimate of the underlying dependency structure of the data, unconstrained by any *a priori* models of how the data interact with each other. In this study we choose a non-parametric approach and model dependency structures using empirical copulas, due to potentially complex dependency structures. As shown below, this assumption is warranted.

We first create pseudo-observations using a rank-based approach. Then we estimate copula density directly from data using kernel methods. Specifically, we use beta kernels (Chen, 1999, Brown and Chen, 1999). Both the beta kernel and the copula have a bounded support on the interval [0,1], which eliminates boundary bias. In addition, the beta kernel changes shape with respect to the input, even if the kernel density estimation bandwidth remains the same, and thus can be viewed as an adaptive kernel density estimation approach. Finally, Charpentier et al., 2007 and Chen, 1999 showed that the variance of the estimator decreases with increase of the sample size.

We estimate the empirical copula densities using 0.03 kernel bandwidth and estimate density at 100 equal-spaced points on the [0,1] interval for both modeled variables. After copula density is estimated, we sample copulas using a Monte-Carlo approach, generate numerous pseudo-observational pairs, and then subject them to further statistical analysis.

Intercomparison of Downscaling Solutions

In addition to this statistical framework for developing projections based on historical model performance, we also engaged with colleagues at UCLA who have developed a series of downscaling intercomparisons, to provide deeper insight into when statistical methods are and are not appropriate.

This downscaling research builds on two prior studies that compared statistical and dynamical downscaling products over California. First, (Pierce et al. 2012) used three Regional Climate Models (RCMs) and two statistical techniques (bias correction with constructed analogues (BCCA) and bias correction with spatial disaggregation (BCSD)) to downscale temperature and precipitation from 16 GCMs for a total of 45 runs. Not all dynamical and statistical methods were applied to each GCM. Downscaled temperature projections were similar between statistical and dynamical techniques, with all runs resolving the pattern of reduced warming along the California coastline. Larger differences between downscaling techniques were seen in the precipitation changes, particularly on a seasonal basis. For instance, regional-scale features like the North American monsoon can be amplified in dynamical downscaled runs. Statistical methods, on the other hand, largely inherited the coarse-scale representation of the monsoonal features in their downscaled runs. Second, Walton et al. (2017) compared end-of-century snowpack-related temperature changes over California's Sierra Nevada between a dynamically downscaled run and two statistical downscaling techniques (BCSD and BCCA, as in (Pierce et al. 2012). It was found that BCSD and BCCA largely do not capture the snow albedo feedback, and consequently underestimate warming amplification in regions of the Sierra Nevada that experience significant snowpack loss. The regional dynamical model simulated a sharper warming gradient between coastal and inland regions compared to statistical techniques, particularly during the summertime.

As the previous studies indicate, dynamical and statistical downscaling techniques can produce different solutions in key areas, such as mountainous watersheds important to the western U.S. water supply. Two principal reasons could explain these differences. First, the statistical and dynamical models could be operating relative to different base climatologies. A statistical model is typically trained on historical observations, while an RCM evolves its own climatology that is often different from observations due to errors or biases, potentially in either the model or observations. We refer to this source of divergence between statistical and dynamical model projections as a "training data" difference. Second, the statistical model may differ from a dynamical model in simulating relationships between coarse and fine spatial scales and how those relationships change in the future. An example of the different assumptions in the dynamical versus statistical models is the interplay of local warming and snowpack (i.e., simulating the snow albedo feedback with physical equations versus empirical relationships). Differences of this second type are independent of the training data and are referred to as a "model relationship" difference.

The typical evaluation of statistical versus dynamical downscaling is to compare RCM output (which has its own climatology) to a statistical model trained on gridded observations. This approach conflates the two sources of dynamical/statistical model divergence into the final comparison. Thus, it is unclear from many previous studies whether dynamical and statistical outputs differ more because the training dataset used for the statistical model does not match the RCM climatology or because the downscaling techniques themselves produce different

responses to a changing climate. This study seeks to distinguish between these sources of divergence between RCM and statistically downscaled results by using RCM output from an historical simulation as a training dataset to a statistical model (similar to the "perfect model" approach, e.g. Dixon et al. 2016). Specifically, we compare three downscaled datasets over the Sierra Nevada to examine differences based on training data and differences based on model relationships. The datasets are (1) dynamically downscaled results from the Weather Research and Forecast (WRF) model, a state-of-the-art RCM; (2) statistically downscaled results using Localized Constructed Analogs (LOCA), a state-of-the-art statistical model, trained on historical WRF output, which is termed "LOCA-WRF"; and (3) statistically downscaled results using LOCA, this time trained on the (Livneh et al. 2013, 2015) station-based gridded observational dataset, which is termed "LOCA-Livneh". By comparing downscaled projected future results from WRF and LOCA-WRF, we remove the effect of differences between the training dataset and WRF climatology, and isolate disagreements due to different downscaling procedures between WRF and LOCA. On the other hand, comparing LOCA-WRF and LOCA-Livneh isolates differences due to the choice of the training dataset's climatology (WRF versus Livneh) in the statistical model.

For the work discussed here that colleagues at UCLA performed, three downscaled climate datasets are compared in this study and each of them use the North American Regional Reanalysis (NARR, Mesinger et al. 2006) as their coarse-scale forcing dataset. NARR (https://www.esrl.noaa.gov/psd/data/gridded/data.narr.html) has a spatial resolution of 32 km and contains 3-hourly outputs of roughly 40 land-surface and atmospheric variables simulated for observed conditions from January 1979 to near current day. Precipitation in NARR, including over the Western US, was shown to be significantly improved compared to other reanalyses such as NCEP-DOE and ERA-40 (Bukovsky and Karoly 2009).

Dynamical Downscaling: WRF

A "pseudo-global warming" approach (Schär et al. 1996; Rasmussen et al. 2011; Sato et al. 2007) is used to dynamically downscale a future climate over California. First, a 1991-2000 historical simulation of observed conditions is performed by forcing WRF with boundary and initial conditions from NARR. Then, a 2091-2100 future climate is simulated by perturbing 1991-2000 NARR boundary conditions with large-scale end-of-century changes (2081-2100 minus 1981-2000) from the CNRM-CM5 model obtained from the Coupled Model Intercomparison Project Phase 5 (CMIP5) archive under the Representative Concentration Pathway 8.5 (RCP8.5). Specifically, monthly climatological perturbations to temperature, specific humidity, winds, and geopotential height throughout the atmospheric column are applied to the NARR boundary conditions that force WRF for the future run. In this downscaling framework, natural variability between 2091-2100 and 1991-2000 is identical, but synoptic events entering WRF's outermost domain freely evolve in this perturbed, warmer climate.

For the dynamical downscaling of the historical NARR and future NARR + delta-CNRM-CM5 runs we use WRF version 3.5 (Skamarock et al. 2008). Two nested domains of 27 and 9 km resolution are used to resolve the fine-scale topography and local climate processes over

California for the historical and for the future simulations. Spectral nudging of temperature, winds, and geopotential heights is applied to the 27 km domain to help limit model drift away from the forcing data. The WRF configuration that best simulated precipitation and snowpack over this region included the Noah land surface model with multi-parameterizations (Niu et al. 2011), the Thompson microphysics scheme (Thompson et al. 2008), the Dudhia shortwave radiation scheme (Dudhia 1989), the Rapid Radiative Transfer Model (RRTM) longwave radiation scheme (Mlawer et al. 1997), the Mellor-Yamada and Nakanishi-Niino (MYNN) level-2.5 surface and boundary layer scheme (Nakanishi and Niino 2004), and the original Kain-Fritsch cumulus convection scheme (Kain and Michael Fritsch 1990). Additional details on the model configuration can be found in (Schwartz et al. 2017; Walton et al. 2017).

Statistical Downscaling: LOCA-WRF and LOCA-Livneh

Localized Constructed Analogs (LOCA; Pierce et al. 2014) is a statistical downscaling approach designed to improve spatial and temporal characteristics of downscaled fields compared to previous constructed analog-based techniques. Previous analog-based techniques have their roots in weather forecasting efforts: The basic idea was to first find a historical day (the "analog day") that best matches current conditions, then take the forecast as the time period immediately following the analog days. One problem with this approach is the difficulty of finding a suitable analog day, given the immense degrees of freedom in the atmosphere. As originally described by van den Dool (1994), constructed analogs ameliorate this problem by forming the weighted average of multiple close-matching historical days to "construct" a best-fit analog day. The forecast is then the same-weighted average of days subsequent to all the selected historical analog days.

Hidalgo et al. (2008) and Maurer and Hidalgo (2007) adapted this constructed analog weather forecasting approach to spatially downscale a GCM. Daily temperature or precipitation patterns from historical observations (the "training data") are a key component of a statistical downscaling process and are emphasized in their work. The training data are first coarsened to the GCM's grid and used to bias-correct the GCM data via quantile mapping. After bias-correction, the N (typically 30) GCM-grid observed days that best match the GCM day being downscaled are identified. Optimal weights are computed such that the weighted average of the N GCM-grid observed days minimizes the RMSE difference between the GCM day being downscaled and the weighted average of the observed days. Finally, the same weights are used to form the weighted average of the N observed days on the original fine-scale observational grid, producing the downscaled result. This approach has three key drawbacks (Pierce et al. 2014): 1) extreme days are muted due to the use of the weighted average of N source days, a process that tends to damp extremes; 2) related to the first drawback, the weighted average of N source days' over the domain smoothens the spatial structure, so for example, resulting precipitation fields produce unrealistic drizzle; and 3) the solution is domain-sensitive with larger errors as the domain size increases, since the GCM and coarsened observations are matched over the entire domain.

LOCA (Pierce et al. 2014) improves on these earlier constructed analog downscaling methods using a multi-scale matching approach. First, the N best-fit coarsened observed days are chosen only in the region that is positively correlated with the fine-resolution grid cell being downscaled

to, rather than over the entire domain. This makes the approach insensitive to domain extent for large domains. Then, the single one of the N days that best matches conditions in a 1-degree by 1-degree box around the grid cell being downscaled to is selected as the analog. This approach means that the final downscaled result is consistent with the coarse GCM field on both the synoptic and local (1-degree box) scales. Also, using a single day rather than averaging multiple days to produce the final result means that extreme events are not suppressed, spatial structure is not artificially smoothed and extraneous drizzle is not generated. LOCA also uses quantile mapping bias correction over the historical period. But future changes are handled by applying the GCM's predicted changes factors by quantile, additively for temperature and multiplicatively for precipitation. This preserves the GCM predicted future change in the downscaled output, unlike quantile mapping applied to future projections (Pierce et al. 2014). For example, if the GCM projections indicate that median (50th percentile) precipitation increases by 5%, but extreme (95th percentile) precipitation increases by 15%, then the future bias-corrected median precipitation is 5% larger than the historical median precipitation, while the future bias corrected 95th percentile precipitation is 15% larger than the historical 95th percentile value.

LOCA requires a coarse-grid representation of the variable being downscaled. These data are produced by GCMs, but it would not be appropriate to use them here directly, given the need to compare LOCA to the pseudo-global warming approach described above. The pseudo-global warming approach does not produce a coarse-resolution representation of the precipitation or temperature field to be downscaled by LOCA either, since the future WRF simulation is driven by adding CNRM-CM5's monthly atmospheric perturbations to the NARR historical time sequence at the domain's boundaries. To produce coarse-resolution data in a way that is consistent with the LOCA approach and also to produce LOCA output that can be sensibly compared with the pseudo-global-warming output, we applied the CNRM-CM5 projected climate change factors, evaluated by quantile, to the NARR historical values analyzed onto each grid point on the coarse grid. The climate change factors were additive for temperature and multiplicative for precipitation. This process mimics the bias correction procedure in LOCA as described above (Pierce et al., 2015). It also produces a coarse resolution precipitation field that includes the CNRM-CM5-predicted future climate changes to be downscaled subsequently with LOCA. These data are referred to as "bias corrected NARR" in the following sections.

Results and Discussion

Focus on TAFB and FBR (Tadić and Biraud, JAMC, Submitted)

From the purely statistical regression-based framework as described above, the primary result and achievement is that we demonstrated the numerical stability and scope of the statistical tools that are needed to develop projections of temperature and precipitation from a multi-model ensemble at the DoD-facility level. As described above, the ability to do this represents a necessary, but not sufficient condition for the development of robust projections of change at the facility level.

Using the approach outlined in Figure 18, we first focused on the development of projections of daily precipitation and surface air temperature around two selected DoD facilities: Travis Air Force Base (TAFB) near Fairfield, California, and Fort Bragg (FBR) in North Carolina. Below

is a summary of average and extreme changes for these two locations. Summaries of the changes in average precipitation and maximum daily surface temperature for 3 separate 30-year periods centered around 2020, 2050, and 2100 for the two facilities are shown in Tables 4 and 5. The changes of mean temperature are not necessarily monotonic with time and with emissions scenario, though they are large, and the changes in mean precipitation are moderate.

	TAFB			
	RCP45		RCP85	
Year (centered around)	Precipitation (mm/day) (delta)	T_{max} (K) (delta)	Precipitation (mm/day) (delta)	T_{max} (K) (delta)
2020	+0.1	+1.1	+0.2	+1.1
2050	-0.0	+2.0	+0.0	+1.9
2100	-0.1	+2.2	+0.0	+2.7

Table 4: Changes in average daily precipitation and maximum daily surface temperature values (future-past) for TAFB for three future periods centered around years 2020, 2050 and 2100, derived from bias-corrected future projections using MLR and past observations, for two emission scenarios (RCP4.5 and RCP8.5).

	FBR			
	RCP45		RCP85	
Year (centered around)	Precipitation (mm/day) (delta)	T_{max} (K) (delta)	Precipitation (mm/day) (delta)	T_{max} (K) (delta)
2020	-0.5	+0.9	-0.3	+0.1
2050	-0.4	+1.2	-0.3	+1.5
2100	-0.5	+1.6	-0.4	+2.2

Table 5: Changes in average daily precipitation and maximum daily surface temperature values (future-past) for FBR for three future periods centered around years 2020, 2050 and 2100, derived from bias-corrected future projections using MLR and past observations, for two emission scenarios (RCP4.5 and RCP8.5).

Additionally, we can use this approach to develop estimates of changes in extreme values of precipitation and daily maximum temperature in the 21st Century and associated confidence intervals using GEV parameter estimates. These changes are a function of emissions scenario, which becomes a large driver of the divergence in estimates by the end of the 21st Century (Hawkins and Sutton, 2009).

	TAFB				FBR			
	Precipitation (mm/day) (95% CI)		T_{max} (K) (95% CI)		Precipitation (mm/day) (95% CI)		T_{max} (K) (95% CI)	
	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5	RCP4.5	RCP8.5
2005-2035	1.09 (0-16.5)	2.77 (0-20)	4.35 (0-21.4)	4.63 (0-21.5)	0.30 (0-10)	0.47 (0-8)	7.58 (0-35.5)	1.86 (0-26)
2035–2065	0.20 (0-11)	1.87 (0-17)	11.35 (0-31.1)	11.63 (0-31.5)	0.23 (0-9)	0.33 (0-8)	11.83 (0-38)	17.65 (0-41.5)
2085–2115	0.06 (0-10)	2.97 (0-19)	13.60 (0-33.4)	22.45 (0.7-42.5)	0.21 (0-9)	0.25 (0-9)	19.98 (0-45)	31.04 (0-52.5)

Table 6: The ratio of probability of the extreme event in the future that corresponded to $p<0.01$ in the past, and its probability for three future periods (centered around years 2020, 2050 and 2100), for both AFB and FBR locations under two emission scenarios, based on GEV distribution. The uncertainty is reported as 95% Confidence Intervals (CI).

			RCP45		RCP85	
	Year (centered around)	Exceedance Probability	Precipitation (mm/day) (95% CI)	T_{max} (K) (95% CI)	Precipitation (mm/day) (95% CI)	T_{max} (K) (95% CI)
TAFB	2020	P (0.05)	73.8 (45.7-137.8)	318.2 (316.1-321.8)	80.2 (47.7-159.5)	318.2 (316.2-321.6)
		P (0.02)	83.0 (47.5-185.5)	318.9 (316.3-324.1)	92.7 (50.0-227.8)	319.0 (316.4-323.6)
		P (0.01)	89.1 (48.4-229.2)	319.4 (316.4-325.9)	101.7 (51.2-295.5)	319.5 (316.6-325.2)
	2050	P (0.05)	67.3 (44.3-109.1)	319.1 (317.0-322.6)	76.5 (46.2-138.0)	319.1 (317.0-322.5)
		P (0.02)	74.7 (46.5-135.1)	319.8 (317.3-324.9)	87.6 (49.9-183.3)	319.9 (317.3-324.5)
		P (0.01)	79.5 (47.6-155.9)	320.3 (317.4-326.6)	95.2 (50.4-223.6)	320.3 (317.4-325.9)
	2100	P (0.05)	66.0 (43.4-103.2)	319.3 (317.2-322.7)	80.45 (46.4-155.8)	319.9 (317.8-323.2)
		P (0.02)	72.8 (45.7-	320.1	94.3 (49.5-	320.6 (318.1-

			123.5)	(317.5-324.9)	219.4)	325.0)
		P (0.01)	77.0 (46.9-138.5)	320.6 (317.7-326.6)	104.3 (51.1-280.1)	321.0 (318.2-326.4)

Table 7: Exceedance probabilities and confidence intervals at TAFB as a function of emissions scenario for different 30-year intervals centered around 2020, 2050, and 2100 for daily maximum surface air temperature and precipitation.

	Year (centered around)	Exceedance Probability	Precipitation (mm/day) (95% CI)	T_{max} (K) (95% CI)	Precipitation (mm/day) (95% CI)	T_{max} (K) (95% CI)
FBR	2020	P (0.05)	95.8 (65.9-140.6)	314.5 (312.4-318.5)	91.3 (64.1-136.3)	313.9 (311.7-317.8)
		P (0.02)	105.6 (69.9-164.4)	314.9 (312.4-320.0)	103.2 (68.4-168.8)	314.3 (311.8-319.5)
		P (0.01)	111.6 (72.1-181.4)	315.0 (312.5-321.1)	111.4 (71.0-195.6)	314.5 (311.8-320.6)
	2050	P (0.05)	93.5 (63.4-134.7)	314.9 (312.7-318.7)	92.5 (65.1-134.8)	315.3 (313.1-319.1)
		P (0.02)	103.4 (69.6-157.5)	315.3 (312.8-320.2)	103.1 (69.4-161.4)	315.8 (313.3-320.8)
		P (0.01)	109.6 (72.0-173.8)	315.5 (312.9-321.3)	110.1 (71.8-181.8)	316.1 (313.3-322.0)
	2100	P (0.05)	94.6 (66.4-135.5)	315.3 (313.1-319.4)	93.0 (65.3-134.2)	315.9 (313.7-319.8)
		P (0.02)	104.1 (70.5-157.1)	315.7 (313.2-321.1)	103.2 (69.6-158.0)	316.4 (313.9-321.4)
		P (0.01)	109.9 (72.8-172.2)	315.9 (313.2-322.2)	109.6 (72.0-175.4)	316.6 (313.9-322.5)

Table 8: Exceedance probabilities and confidence intervals at FBR as a function of emissions scenario for different 30-year intervals centered around 2020, 2050, and 2100 for daily maximum surface air temperature and precipitation.

In Table 6, we present what is often refered to as a risk-ratio, which is the ratio of the probability that a specific extreme event (in terms of temperature or precipitation absolute value) will occur in future periods as compared to its probability of occurring under present conditions. This table indicates that the extremes of temperature and precipitation are becoming more extreme and provides confidence intervals on these risk-ratios which, themselves are quite broad. Meanwhile, Tables 7 and 8 show exceedence probabilities and their associated ranges to provide concrete, numbers for planning purposes.

We can develop detailed information about the shifts statistical distributions of surface air temperature and precipitation. Figure 20 shows the shift in surface air temperature from the period centered around 2000 to the period centered around 2050 for TAFB for the RCP4.5 and RCP8.5 emissions scenarios.

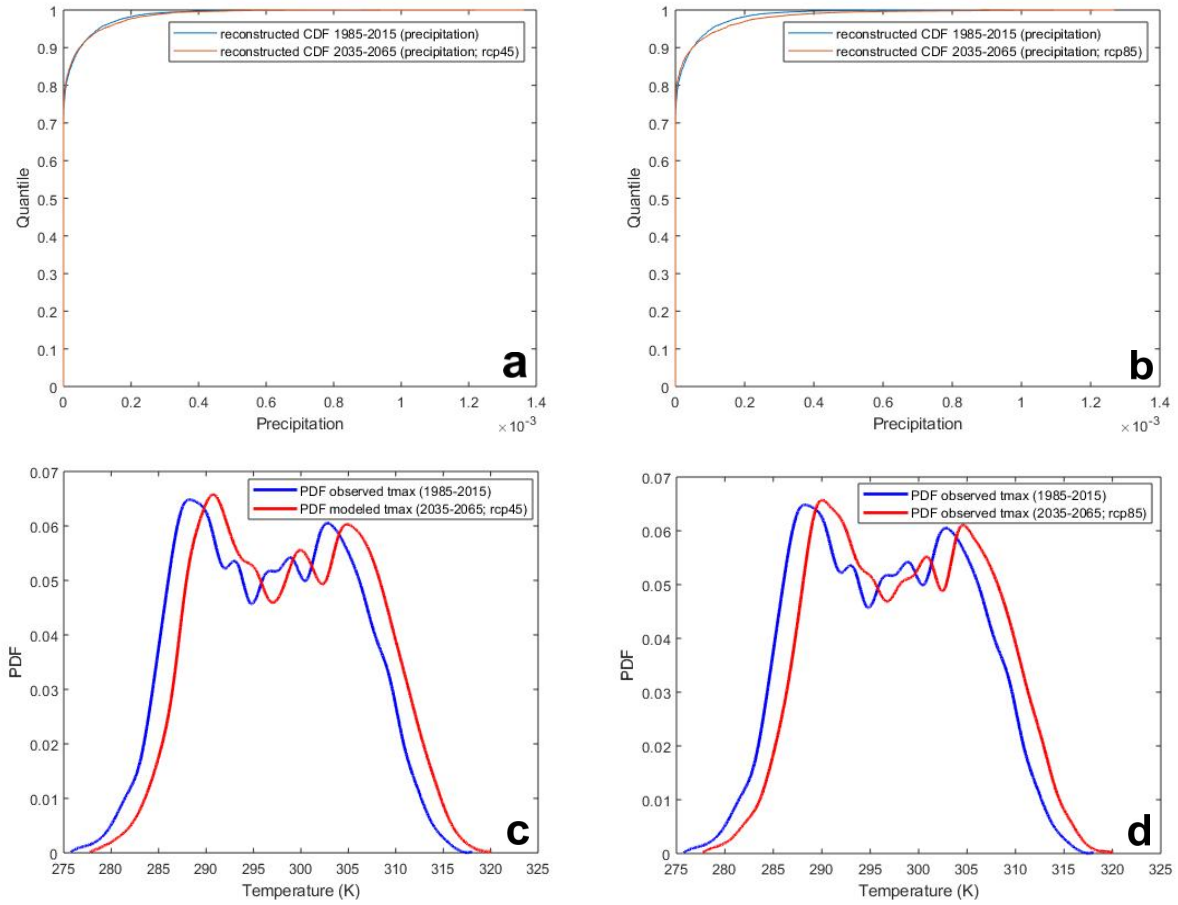


Figure 19: Sample CDFs for precipitation (a-b) and PDFs for T_{max} (c-d) for the TAFB location, comparing the 2035–2065 period to the baseline period. The changes in central tendency and distribution are apparent for T_{max}, and less apparent for precipitation. From Tadić and Biraud (2019, Submitted)

The development of this framework also indicates the numerical robustness of a regression-based approach. Figure 20 shows the variability in MLR coefficients for the CMIP5 models that informed the MLR development, but, in spite of this, Figure 21 indicates that the actual historical

CDF derived from observations can be reconstructed from the regression coefficients with negligible error. This is a necessary, but not sufficient condition, for skill in an MLR projection: it must not incur error simply from the regression itself.

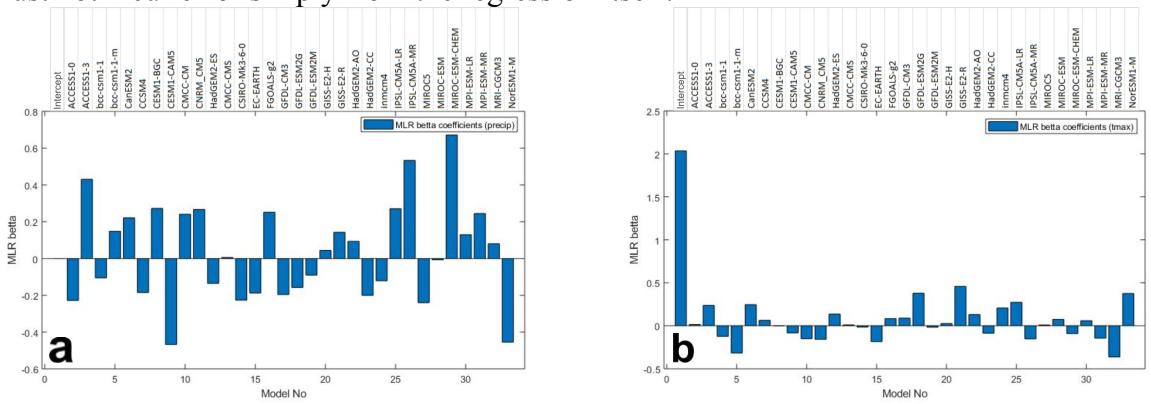


Figure 20: MLR β coefficients (and intercepts) (a-b) obtained by reconstructing observational precipitation and T_{max} CDF using 32 model outputs downscaled by LOCA, for the 1985–2011 period. From Tadić and Biraud (2019, Submitted).

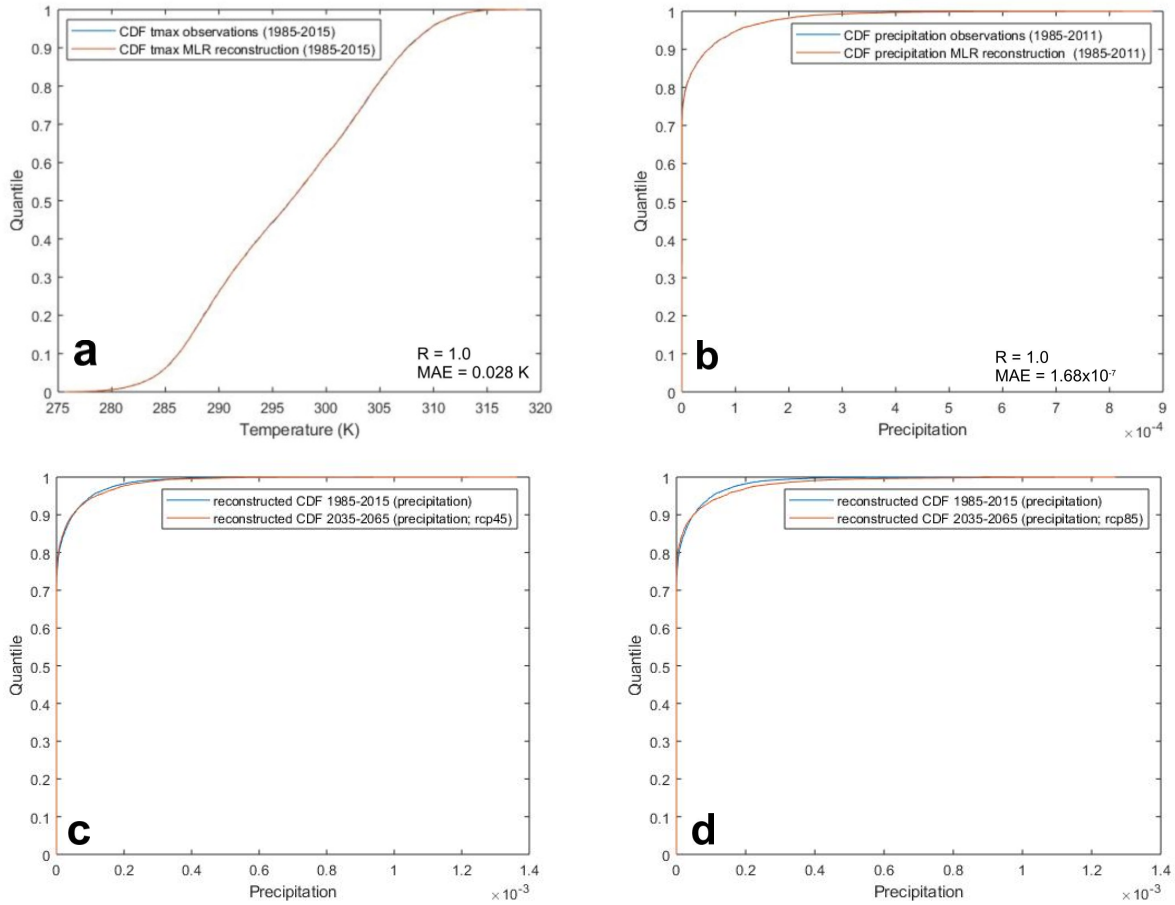


Figure 21: (a-b) CDF of the reconstructed observational precipitation and T_{max} CDF using 32 model outputs downscaled by LOCA, for the 1985–2011 period. (a-b) Show the CDF of reconstructed T_{max} from 32 model the excellent fit between observational and reconstructed CDFs for precipitation (c) and T_{max} (d). From Tadić and Biraud (2019, Submitted).

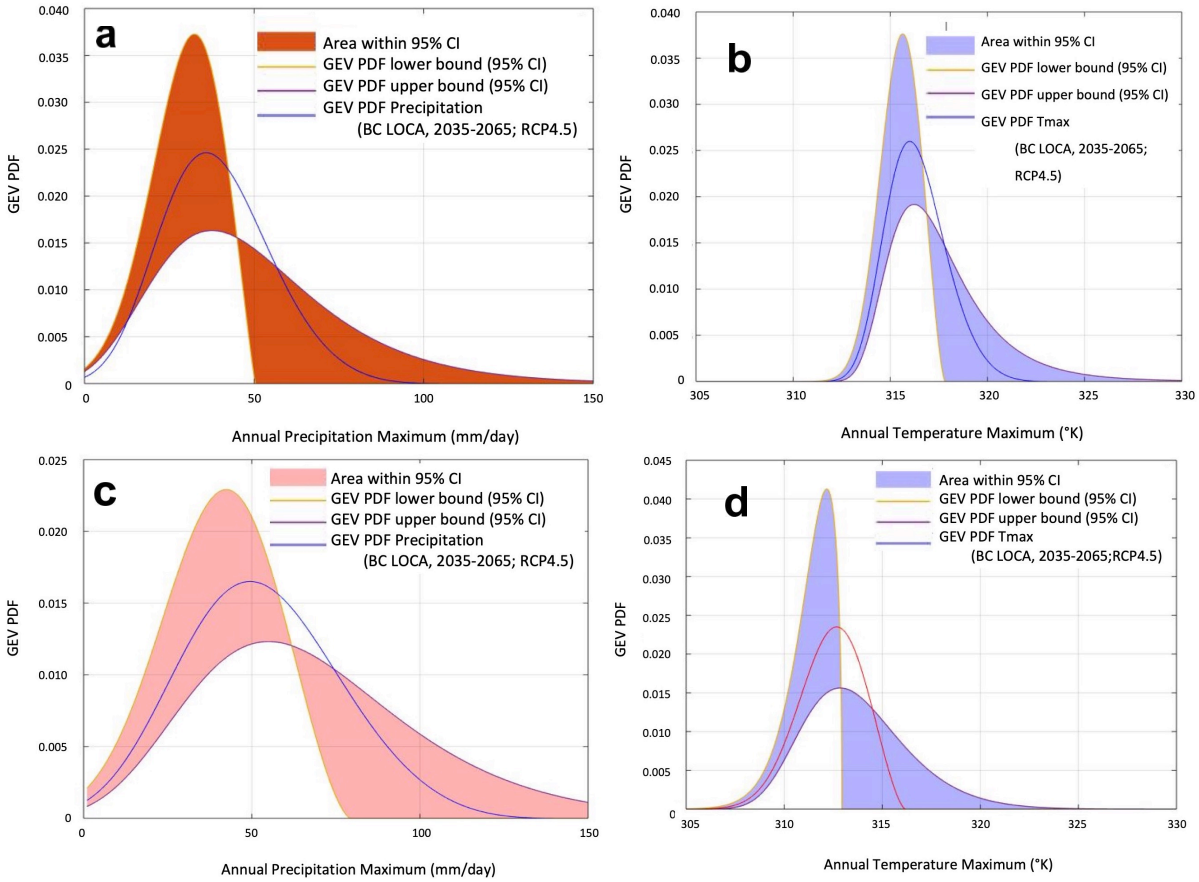


Figure 22: (a) GEV PDF for annual precipitation maximum values for 2035–2065, TAFB, under RCP4.5 emission scenario, (b) GEV PDF for annual T_{max} maximum values for 2035–2065, TAFB, under RCP4.5 emission scenario, (c) GEV PDF for annual precipitation maximum values for 2035–2065, FBR, under RCP4.5 emission scenario, and (d) GEV PDF for annual T_{max} maximum values for 2035–2065, FBR, under RCP4.5 emission scenario. From Tadić and Biraud (2019, Submitted).

Given the numerical stability of the regression-based approach, we can use GEV theory to develop highly-detailed projections of the distributions of temperature and precipitation, which are shown in Figure 22. These graphs form the basis of the confidence intervals for TAFB and FBR shown in Table 6.

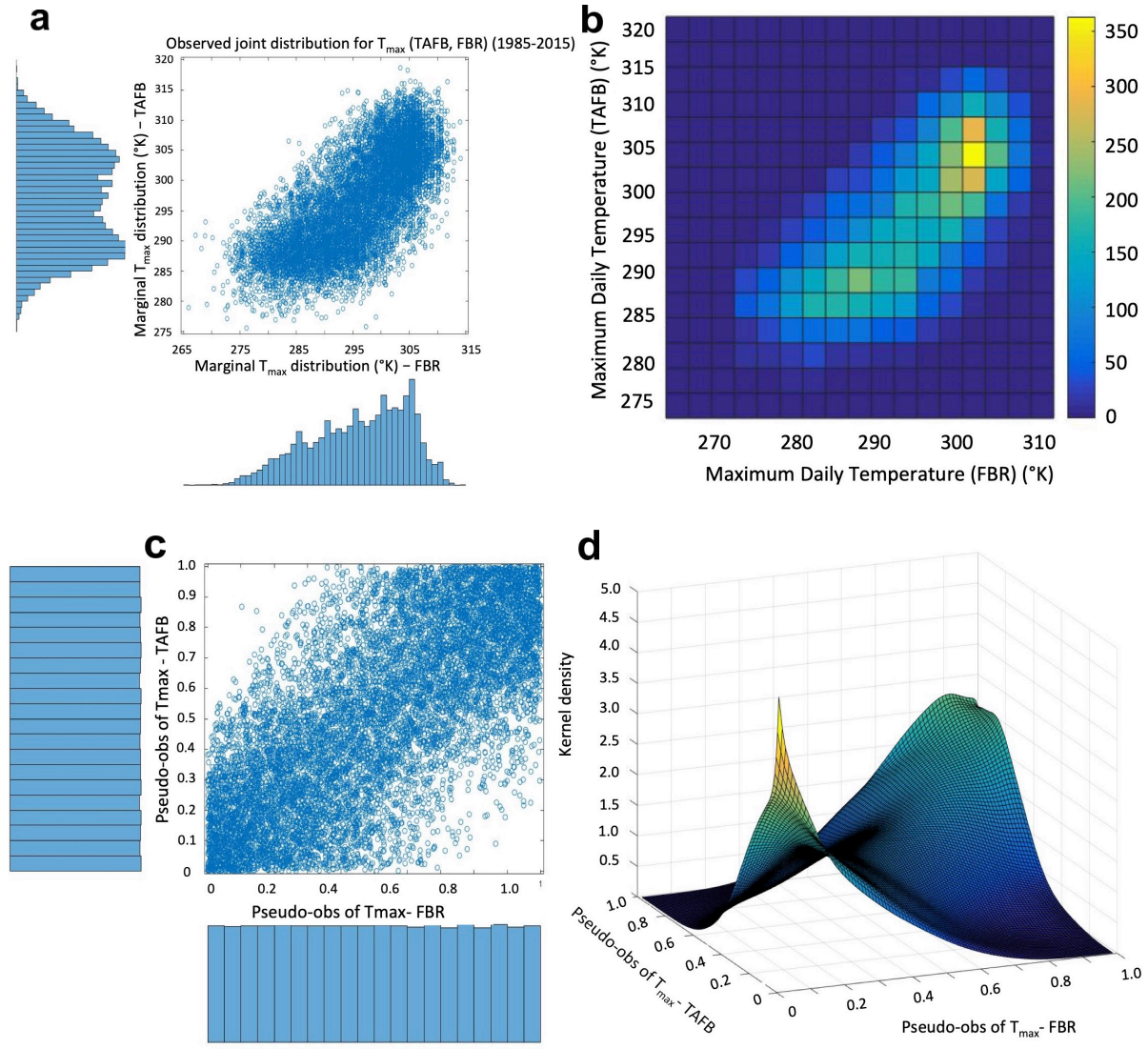


Figure 23: (a) Marginal and joint distribution of observations of T_{max} at TAFB and FBR during 1985–2011 period. (b) 2D histogram shown in tile format. (c) Marginal and joint distributions of T_{max} pseudo-observations, prepared as input to copula-modeling. (d) Fitted empirical copula. From Tadić and Biraud (2019, Submitted).

Furthermore, the use of copulas, as described above, presents a formal framework for understanding correlated shifts in means and extremes of surface air temperature and precipitation. They can even be used for different variables, such as correlations of extreme temperature and precipitation. While the TAFB and FBR facilities are geographically distant from each other and generally not meteorologically related, we show the results of the copula analysis in Figure 23-26. These figures not only show the lack of correlation in extremes between the two sites, as expected, but also show the level of detail and analysis that can be applied to understand how facility-level risks may be correlated in the future, especially for DoD facilities that are more geographically proximal.

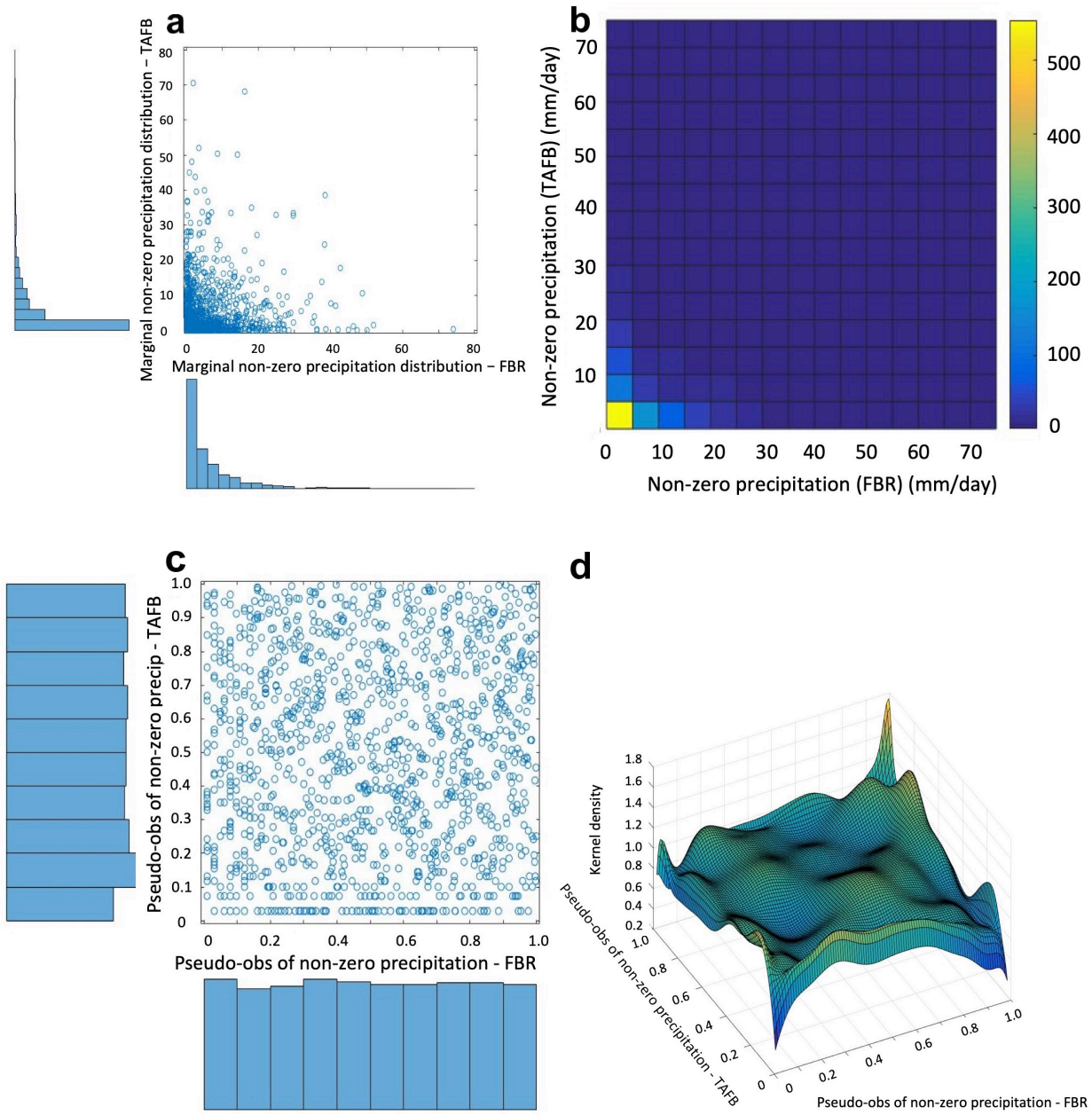


Figure 24: (a) Marginal and joint distribution of precipitation observations at TAFB and FBR during 1985–2011 period. (b) 2D histogram shown in tile format. (c) Marginal and joint distributions of precipitation pseudo-observations, prepared as input to copula-modeling. (d) Fitted empirical copula. From Tadić and Biraud (2019, Submitted).

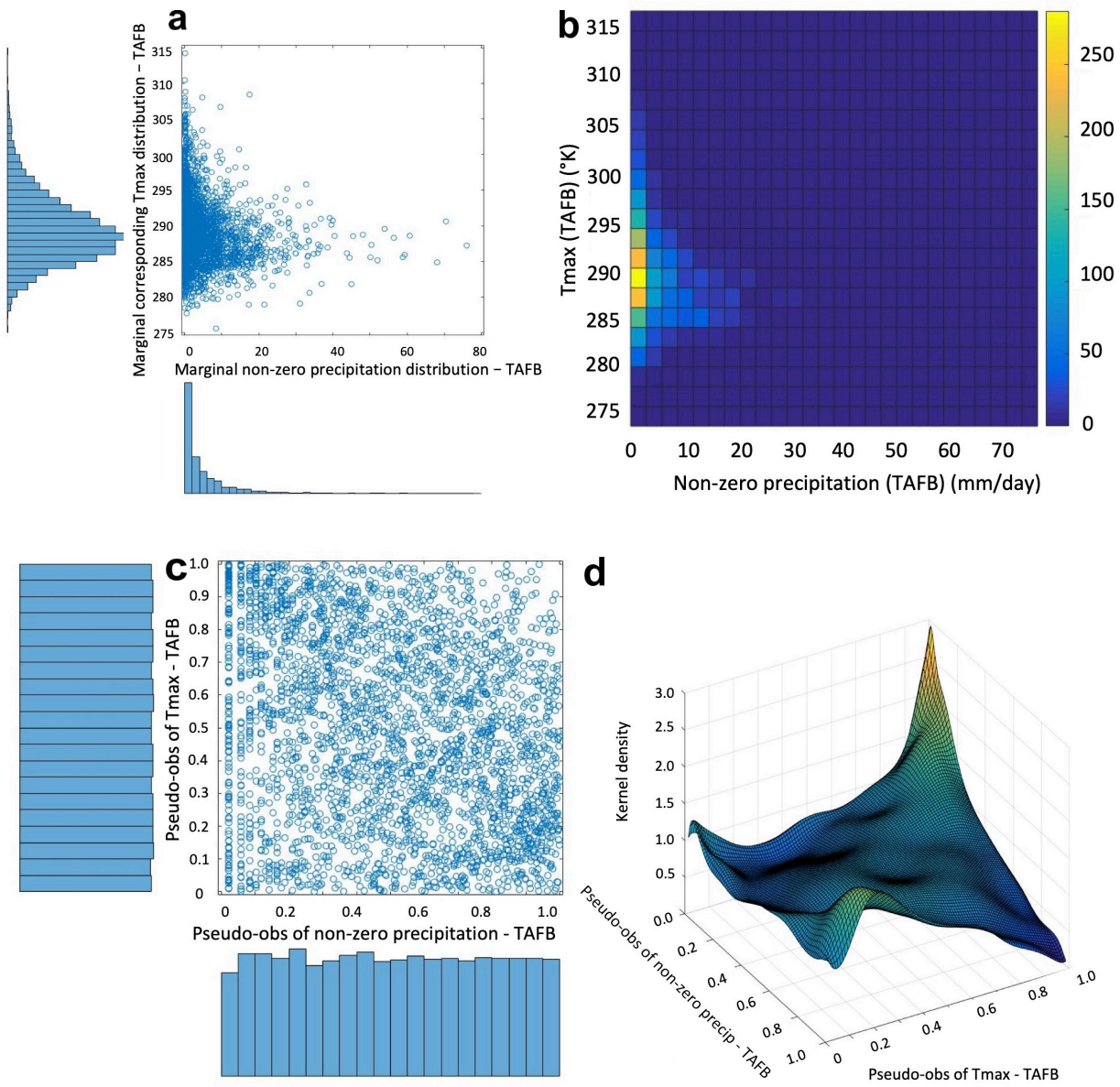


Figure 25: (a) Marginal and joint distribution of observations of non-zero precipitation and corresponding T_{max} at TAFB, during 1985–2011 period. (b) 2D histogram shown in tile format. (c) Marginal and joint distributions of non-zero precipitation and T_{max} pseudo-observations, prepared as input to copula modeling. (d) Fitted empirical copula. From Tadić and Biraud (2019, Submitted).

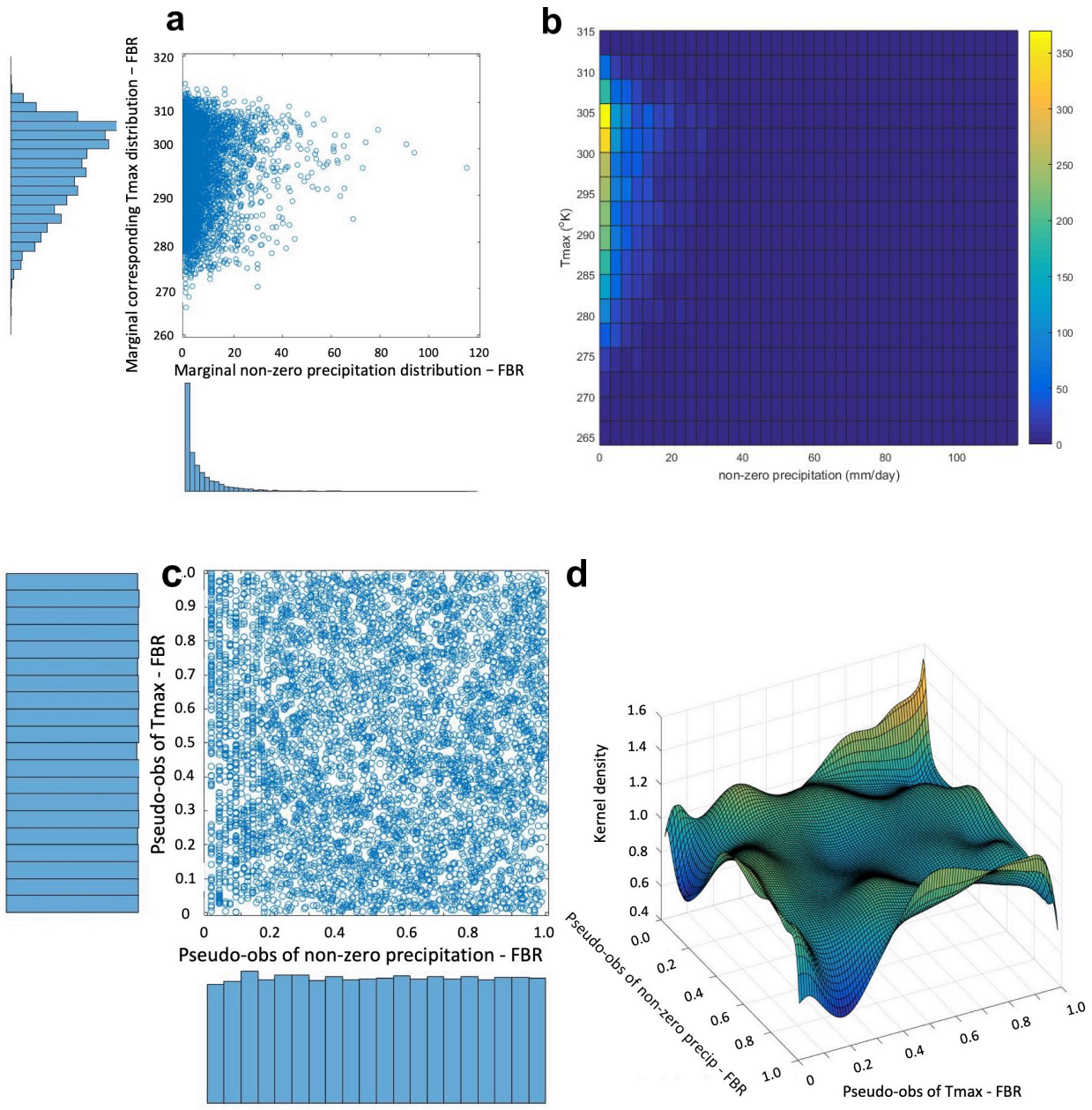


Figure 26: (a) Marginal and joint distribution of observations of non-zero precipitation and corresponding T_{max} at FBR, during 1985–2011 period. (b) 2D histogram shown in tile format. (c) Marginal and joint distributions of non-zero precipitation and T_{max} pseudo-observations, prepared as input to copula-modeling. (d) Fitted empirical copula. From Tadić and Biraud (2019, Submitted).

In addition to the highly-detailed results shown over TAFB and FBR, we can use the regression-based framework to develop distributions across the CONUS. Figure 27 shows a summary of those results by indicating the change in extreme surface air temperature and precipitation over the CONUS, by showing the change in the 20-year daily return value of

maximum daily surface air temperature and precipitation as derived from fitting parameters to a Type-I GEV distribution. This figure indicates that extreme temperatures are likely to increase across much of the CONUS and will be exacerbated with higher emissions. Extreme precipitation will also worsen in the 21st Century with some locations, especially in the Southeastern US, facing more extreme precipitation events.

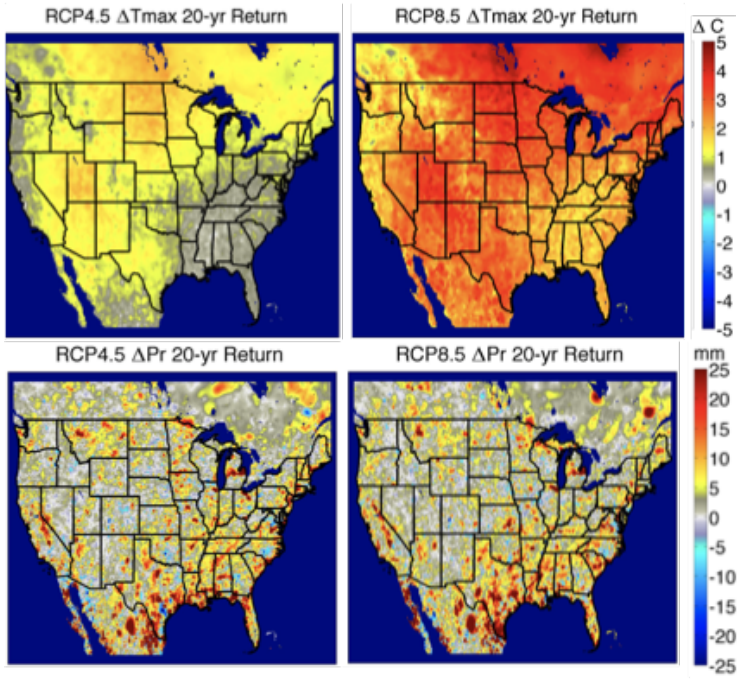


Figure 27: Change in 20-year daily return value of maximum surface air temperature and precipitation based on the MLR approach across the CONUS for two emissions scenarios from 1985-2015 to 2070-2100 .

Model Projections for California (Feldman et al, JWRPM, Submitted)

We can also consider focusing on a specific region and down-selecting models and downscaling them to provide practical datasets for planning in the face of climate change. While an ensemble of climate model simulations may be preferred, the range in emissions scenarios (*van Vuuren et al.,* 2011) and climate model responses (*Taylor et al.*, 2012; *Andrews et al.*, 2012) provide far too many simulations for the purposes of water planning exercises: it is simply infeasible to develop water plans for more than a handful of datasets. This challenge was recently recognized, and preliminary guidelines were developed to use a limited set of tailored scenarios to represent the spectrum of climate scenarios optimally (*Ntegeka et al*, 2014).

At the local level, practical limitations in resources and technical capacity often restrict the ability of resource management agencies to evaluate multiple projections of future conditions. To complicate matters further, climate may be just one of several future

uncertainties that an agency may need to address (e.g., population growth, land use patterns) and exploration of future conditions across multiple dimensions of uncertainty can quickly exceed an agency's capacity or resources. In order to simplify these analyses, it is often useful to provide a more limited set of future climate conditions upon which to bound and anchor the analysis.

We focus here on California because it presents a particular challenge in that the state spans a vast range of hydroclimates, from temperate forests in the north to arid deserts in the southeast, and a large area with a Mediterranean-type climate in between. Resource planners must consider how precipitation, and to a lesser extent, temperature, will change, especially where water resources are the result of a dozen or fewer synoptic systems.

In response to these water resource planning needs and following the approach indicated in *Ntegeka et al*, (2014), we develop here a process for producing the temporal distribution of precipitation in California at hydroclimatically-relevant spatial scales for less extreme, wetter conditions and more extreme, drier conditions, and discuss how the results are used for planning purposes across the State. As part of the development of this range, we highlight some inherent numerical challenges, indicates the resulting range in water resources that these projections imply, and discuss a path towards reducing the outstanding sources of error from downscaling and bias correction.

Different GCMs perform best for different metrics, and when multiple metrics were considered, no individual model emerged as the "best" model for California. Recognizing the need for multiple GCMs, as well as the requirement for a smaller set of simulations, *CCTAG*, (2015) performed a model evaluation effort as part of the 4th California Climate Assessment (Thorne et al, 2018) which aimed to identify a smaller set of GCMs by removing or "culling" the models that did not perform as well for a set of different evaluation metrics. This approach downselected the number of California-relevant climate models to 10: ACCESS-1.0, CCSM4, CESM1-BGC, CMCC-CMS, CNRM-CM5, CanESM2, GFDL-CM3, HadGEM2-CC, HadGEM2-ES and MIROC5. The output from these models from both the RCP4.5 and RCP8.5 scenarios has been used to construct a multi-model ensemble, but specific scenarios that provide a range is also needed.

Two cases were developed for use in these programs (CCTAG, 2015; Pierce et al, 2018). The first, known as Wetter Moderate-Warming (WMW) describes a future for California that exhibits moderate warming while also generally receiving more precipitation. The CNRM-CM5 model, run with the RCP4.5 emissions scenario, was chosen as the GCM/RCP projection pair because this projection generally exhibits the combination of least warming and largest increase in precipitation among the 20 GCM/RCP projection pairs in the ensemble chosen by *CCTAG* (2015). This model exhibits a mid-range value of ECS (3.25 K/2xCO_2) (Andrews et al, 2012), and while it generally exhibits regional biases, the model does not exhibit precipitation biases in California relative to Global Precipitation Climatology Project (GPCP) observations (Voldoire et al, 2013).

The second case, known as Drier Extreme Warming (DEW) exhibits extreme warming while also generally receiving less precipitation. The HadGEM2-ES model [*Jones et al,*

2011a], run with the RCP8.5 scenario, was chosen as the GCM/RCP projection pair because this projection generally exhibits the combination of most warming and largest decrease in precipitation among the 20 GCM/RCP projection pairs in the ensemble chosen by *CCTAG* (2015). This model exhibits the high value of diagnosed ECS (4.69 K/2xCO_2) (Andrews et al, 2012) and has superior performance with respect to precipitation relative to historical observations over North America (Sheffield et al, 2013). Together, these cases help establish a nominal range of conditions that state and local planning agencies across California can use (CCTAG, 2015).

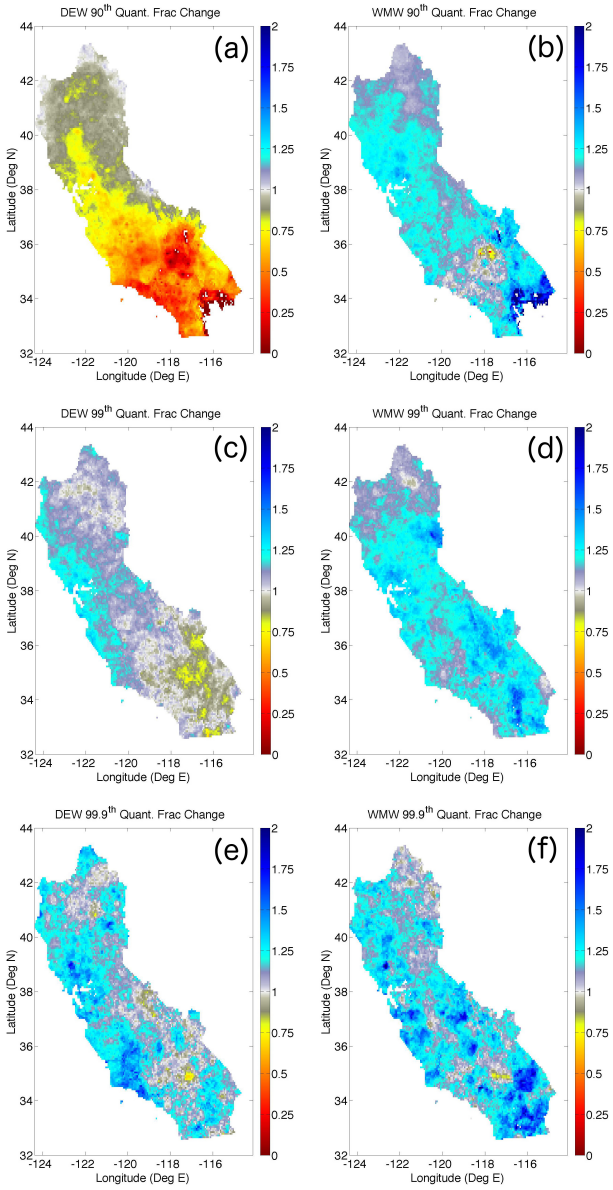


Figure 28: (a) Fractional change in precipitation for the DEW case for the 90th quantile between 1980-2010 and 2056-2085. (b) Same as (a) but for the WMW case. (c-d) Same as (a-b), respectively, but for the 99th quantile. (e-f) Same as (a-b), respectively but for the 99.9th quantile. From Feldman et al, JWRPM, Submitted.

Figure 28 shows the relative change in annual upper-quantile daily precipitation for DEW and WMW cases. At the 90th quantile, the DEW and WMW cases lead to decreased and increased precipitation, respectively, while for the 99th and 99.9th quantiles, both cases show significant increases even while the DEW case indicates more extreme coastal precipitation while the WMW case indicates increases in extreme precipitation in southern California and the Sierra Nevada.

There is an apparent need to use climate model fields to develop precipitation projections in order to estimate water-resource availability in the future, but, at a high level, the use of raw model output of regional and local precipitation projections associated with climate change is quite challenging. Many models are unable to reproduce both low and high-order moments of observed historical distribution of precipitation at regional spatial scales (Christensen et al, 2008; Mearns et al, 2012; Sillmann et al, 2013).

Figure 29 specifically indicates the need for bias-correction. We compare an observationally-derived data product against present-day runs from the DEW and WMW models. The panels of this figure indicate that both models are biased low in northwestern California for the 90th, 99th, and 99.9th quantiles, and the models are biased low across the Sierra for the 99th, and 99.9th quantiles. In particular, these panels show that in Northern California, biases in downscaled model output can exceed 30 percent of the total precipitation value. For planning purposes, these biases must be addressed, and while a univariate bias-correction does not address the source(s) of model bias or how they propagate into future projections, it does recognize that model bias exists and so adjustments to the projections must be made in order to account for this bias practically.

This motivates a hybridized approach wherein statistical information derived from present-day observations is merged with climate-model projections. Historical observations can be used to bias-correct historical climate-model projections, and, while the bias-correction approach is controversial (Ehret et al, 2012; Maraun 2012; Maraun 2013), it has been shown to be tenable (Gudmundsson et al, 2012; Teutschbein and Seibert, 2012; Teutschbein and Seibert, 2013).

Since the spatiotemporal distribution of precipitation across any given area often exhibits zero absolute values, bias correction techniques that are used for interval values (e.g., temperature) will necessarily lead to negative, and therefore unphysical, values. To address this issue, we can utilize a ratio approach to bias correction, which obviates this apparent problem. By using a ratio to scale bias-corrected precipitation, we avoid the creation of an artificially continuous spatiotemporal precipitation field when it should be discontinuous. Additionally, a ratio technique may be able to preserve the relative change contained within a model or the multi-model ensemble between future and present-day precipitation.

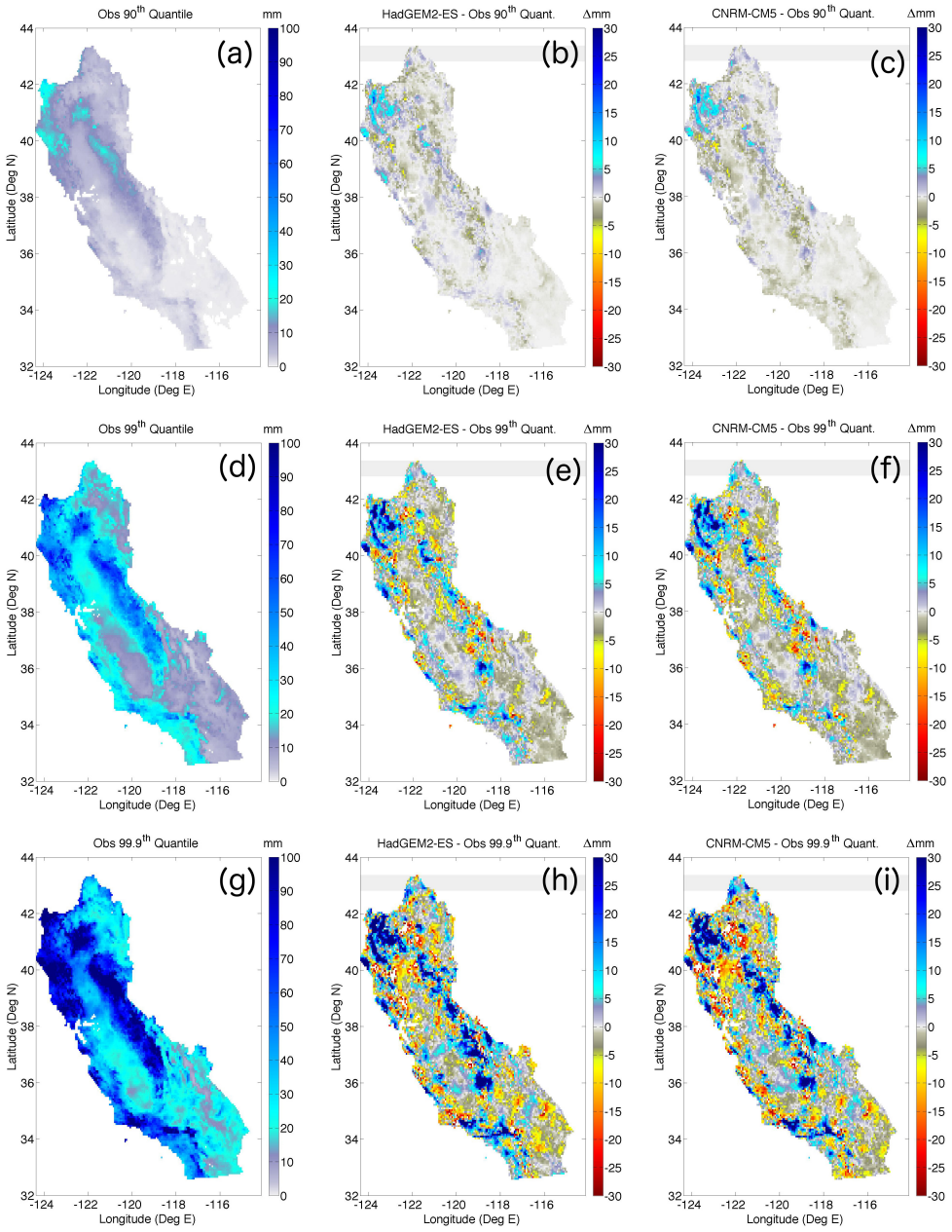


Figure 29: (a) Map of the 90th quantile of daily total precipitation from the observationally-derived dataset of Livneh et al, [2013]. (b) Difference between the HadGEM2-ES model run over the historical period (DEW) and the Livneh et al, [2013] dataset. (c) Same as (b) but showing the difference between the CNRM-CM5 model (WMW) and the Livneh et al, [2013] dataset. (d-f) Same as (a-c) respectively but for the 99th quantile. (g-i) Same as (a-c) respectively but for the 99.9th quantile. From Feldman et al, JWRPM, Submitted.

However, if appropriate caution is not exercised in this process chain, the bias-corrected, downscaled projections will be problematic, especially for extreme precipitation. It has been recognized that traditional quantile mapping (QM), wherein the effects of climate

change are realized by scaling the historical precipitation cumulative distribution function by the ratio of the mean future modeled distribution to the mean historical modeled distribution, can lead to artificial corruption where the mapped precipitation quantities do not reflect the parent model's relative change in precipitation across quantiles (Hagemann et al, 2011; Themeßl et al, 2012; Maraun, 2013; Maurer and Pierce, 2014; Cannon et al, 2015). An example of the artificial corruption that can arise from traditional quantile mapping is shown in Table 9. This shows that QM, by not preserving the relative change across quantiles, produces mean seasonal precipitation changes that are substantially overestimated, including by hundreds or even thousands of percent, from the parent model.

Basin	Run	QM Pr Change %	QM Temp Change °C	GCM Pr Change %	GCM Temp Change °C	Precip Delta (QM-GCM)	Temp Delta (QM-GCM)	Precip % change (QM-GCM)/|GCM|	Temp % change (QM-GCM)/GCM
Central California Coastal	DEW/HadGEM2-ES RCP8.5(2056-2085)	5.0	7.9	-1.4	7.6	6.35	0.3	454%	4%
San Francisco Bay	DEW/HadGEM2-ES RCP8.5(2056-2085)	4.6	7.6	0.2	7.3	4.42	0.3	2210%	4%
Lower Colorado	WMW/CNRM-CM5 RCP4.5(2056-2085)	15.0	3.4	7.6	3.5	7.4	-0.1	97%	-3%
Southern Mojave	WMW/CNRM-CM5 RCP4.5(2056-2085)	24.4	3.5	12	3.5	12.4	0	103%	0%

Table 9: Quantification of the bias in several California regions relative to the parent model when using traditional quantile mapping methods. From Feldman et al, JWRPM, Submitted.

A recently-developed approach called Quantile Delta Mapping (QDM) is designed, in principle, to preserve the relative model change in bias-corrected precipitation across all quantiles (Cannon et al, 2015). QDM also seeks to avoid the systematic overestimation of precipitation that arises when a univariate scaling is applied to perform quantile mapping, since that paper found that QM techniques led to large overestimates of precipitation relative to the parent model near Hudson Bay, Victoria Island, and Ellesmere Island in Canada.

QDM represents a transfer function, and it can be used to map model-projected precipitation at a given time to a bias-corrected, relative-change preserving value. The approach that it takes, therefore, is first to determine the quantile to which the original model-projected value corresponds. In the next step, the relative change is determined from the ratio of modeled projected precipitation for that quantile value in the model-projected cumulative distribution function (CDF) to the modeled historical precipitation for that quantile value in the modeled historical CDF. Finally, the historical precipitation from the same quantile value but from the observed historical CDF is scaled by the model ratio. The functional form of this procedure operates on an individual grid box is as follows:

$$\hat{x}_{m,p}(t) = \hat{x}_{o:m,h:p}(t)\,\frac{x_{m,p}(t)}{F_{m,h}^{-1}[\tau_{m,p}(t)]} \qquad (1)$$

where $\hat{x}_{m,p}(t)$ represents the bias-corrected future projection of precipitation at a given time, $\tau_{m,p}(t)$ represents the modeled quantile value of the model projected value from the model-projected CDF, $F_{m,h}^{-1}$ represents the inverse CDF function of the modeled historical CDF, $x_{m,p}(t)$ represents the original model-projected precipitation value, and $\hat{x}_{o:m,h:p}(t)$ represents the observed precipitation value in the historical record corresponding to the historical observed CDF at quantile $\tau_{m,p}(t)$.

We created a centennial-length data record of daily precipitation, minimum, and maximum temperature that approximates the distribution of these variables under climate change conditions. In order to be relevant for hydrological modeling and planning purposes, we produce these variables at 6-km spatial resolution across the state of California, and, to a limited extent, over the areas of neighboring states that drain into rivers in California. There are a total or 11,367 grid boxes that are analyzed and a total of 35,429 days spanning a historical record ranging from January 1, 1915 to December 31, 2011, or a future record covering a period of comparable length.

We compare the results of traditional QM vs. a QDM approach in Figures 30-32, first to explore whether the relative change between historical and future precipitation in the model is captured by the mapping approach at different quantiles. Figures 4 and 5 indicate that QM is not capturing the relative change in modeled precipitation at the 90th, 99th, and 99.9th quantiles. QDM is able to capture the relative change in modeled precipitation at the 90th and 99th quantiles, but is problematic for the 99.9th quantile, due in large part to the numerical resolution of the quantiles (see QDM Numerical Considerations section for details).

The resulting 5-, 20-, and 50-year return values, derived from a parameter fit of the mapped precipitation on a Type-I Generalized Extreme Value Distribution, with QM and QDM approaches for the DEW case are shown in Figure 6. This figure indicates modest differences between QDM and QM for the DEW case and approaches for 5-year return, but these differences grow more pronounced for the 20-year and 50-year return maps. Similar behavior in return value maps, not shown, occur for the WMW case. In general, we also find that QDM produces a greater contrast in return values at coastal locations and over complex terrain. Finally, these results indicate that, while the DEW case is generally drier than the WMW case, even the drier future has very extreme wet events: the DEW case exhibits higher 50-year return values.

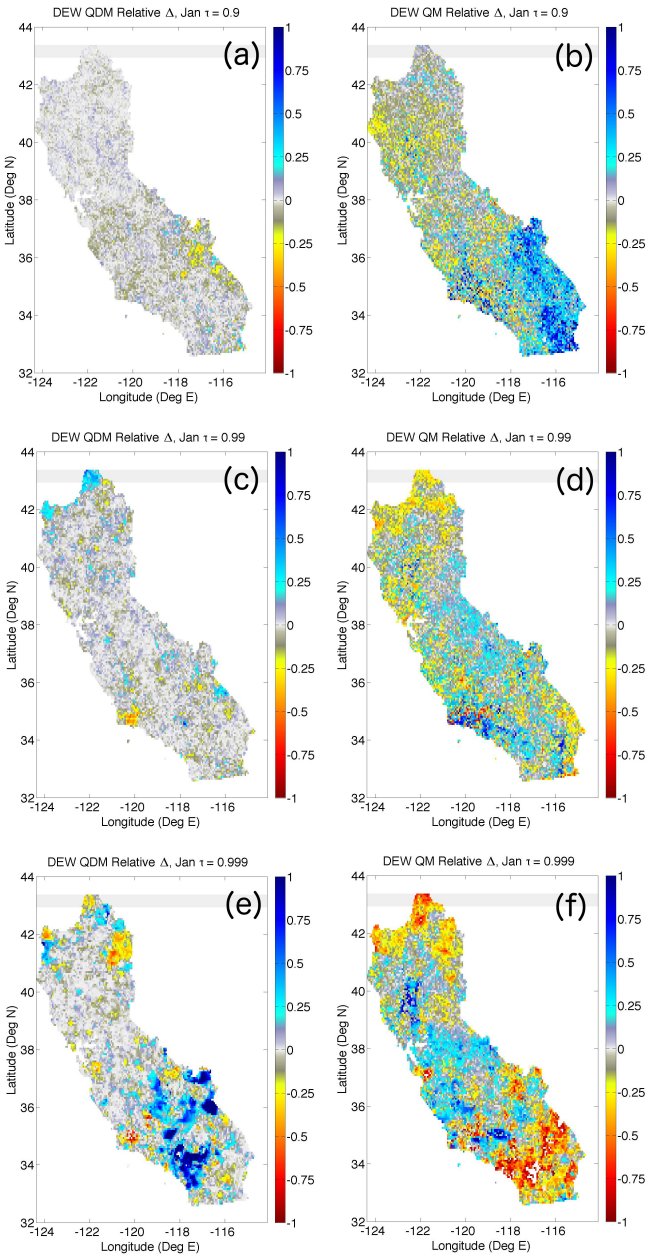


Figure 30: (a) Map of the difference in the relative change in precipitation in the 90th quantile of precipitation of QDM-derived precipitation for 2056-2085 and the observational record from 1981-2010 and the relative change in precipitation in the LOCA-downscaled HadGEM2-ES model for the RCP8.5 scenario from 2056-2085 (DEW) and the historical scenario from 1981-2010. (b) Same as (a) but for QM-derived precipitation instead of QDM-derived precipitation. (c-d) Same as (a-b), respectively but for the 99th quantile. (e-f) Same as (a-b). respectively but for the 99.9th quantile. From Feldman et al, JWRPM, Submitted.

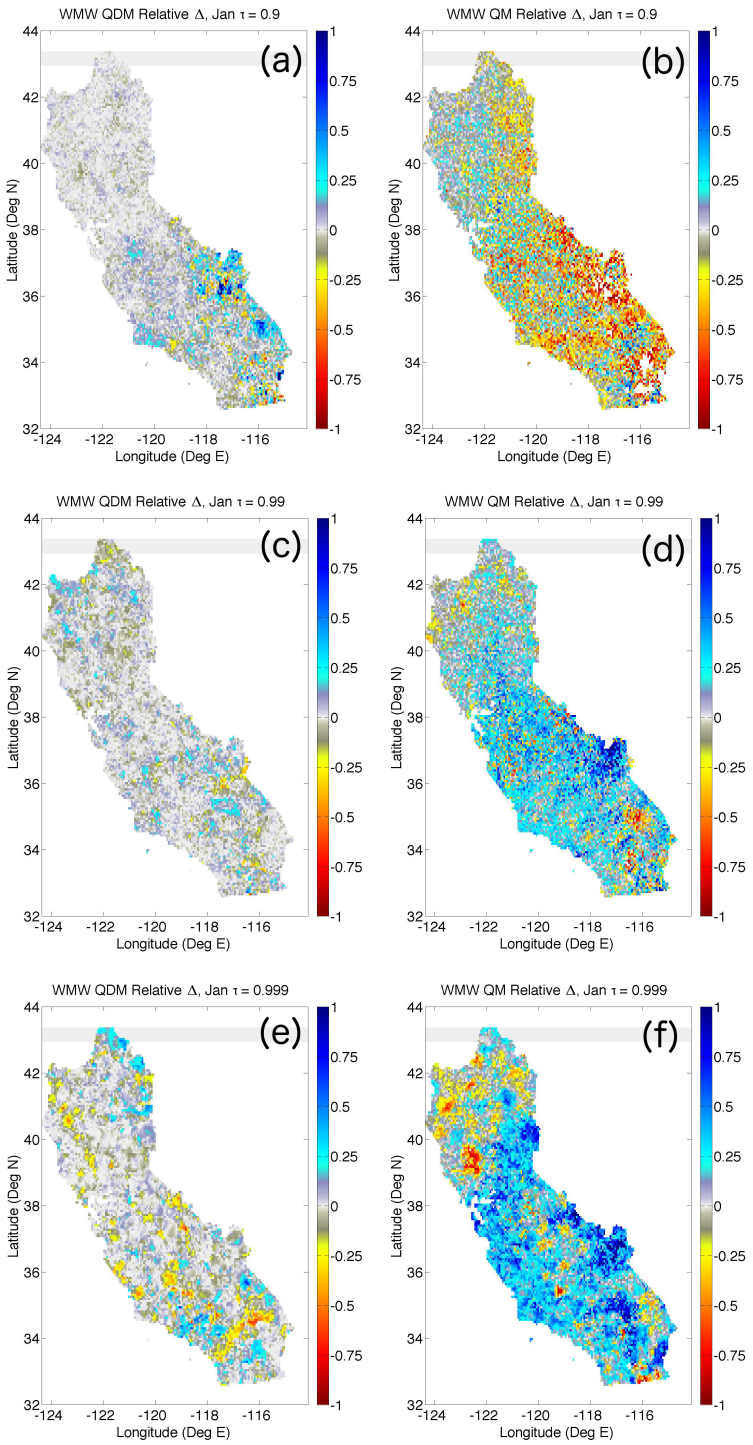


Figure 31:) Map of the difference in the relative change in precipitation in the 90th quantile of precipitation of QDM-derived precipitation for 2056-2085 and the observational record from 1981-2010 and the relative change in precipitation in the LOCA-downscaled CNRM-CM5 model for the RCP4.5 scenario from 2056-2085 (WMW) and the historical scenario from 1981-2010. (b) Same as (a) but for QM-derived precipitation instead of QDM-derived precipitation. (c-d) Same as (a-b), respectively but for the 99th quantile. (e-f) Same as (a-b). respectively but for the 99.9th quantile. From Feldman et al, JWRPM, Submitted.

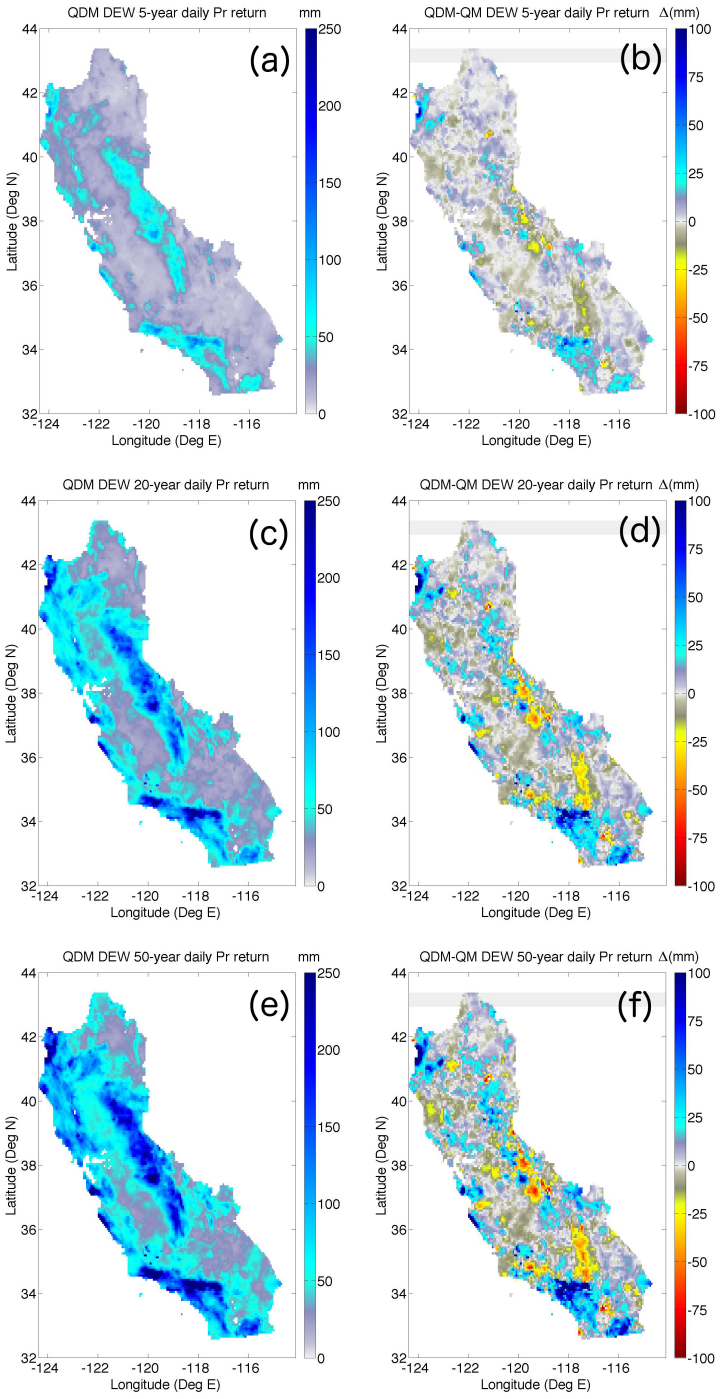


Figure 32: (a) Daily precipitation 5-year return value for the DEW case derived from QDM estimated from a parameter fit to a Type 1 Generalized Extreme Value Distribution. (b) Difference between (a) and the QM case. (c-d) Same as (a-b), respectively, but for the daily precipitation 20-year return value. (e-f) Same as (a-b), respectively, but for the daily precipitation 50-year return value. From Feldman et al, JWRPM, Submitted.

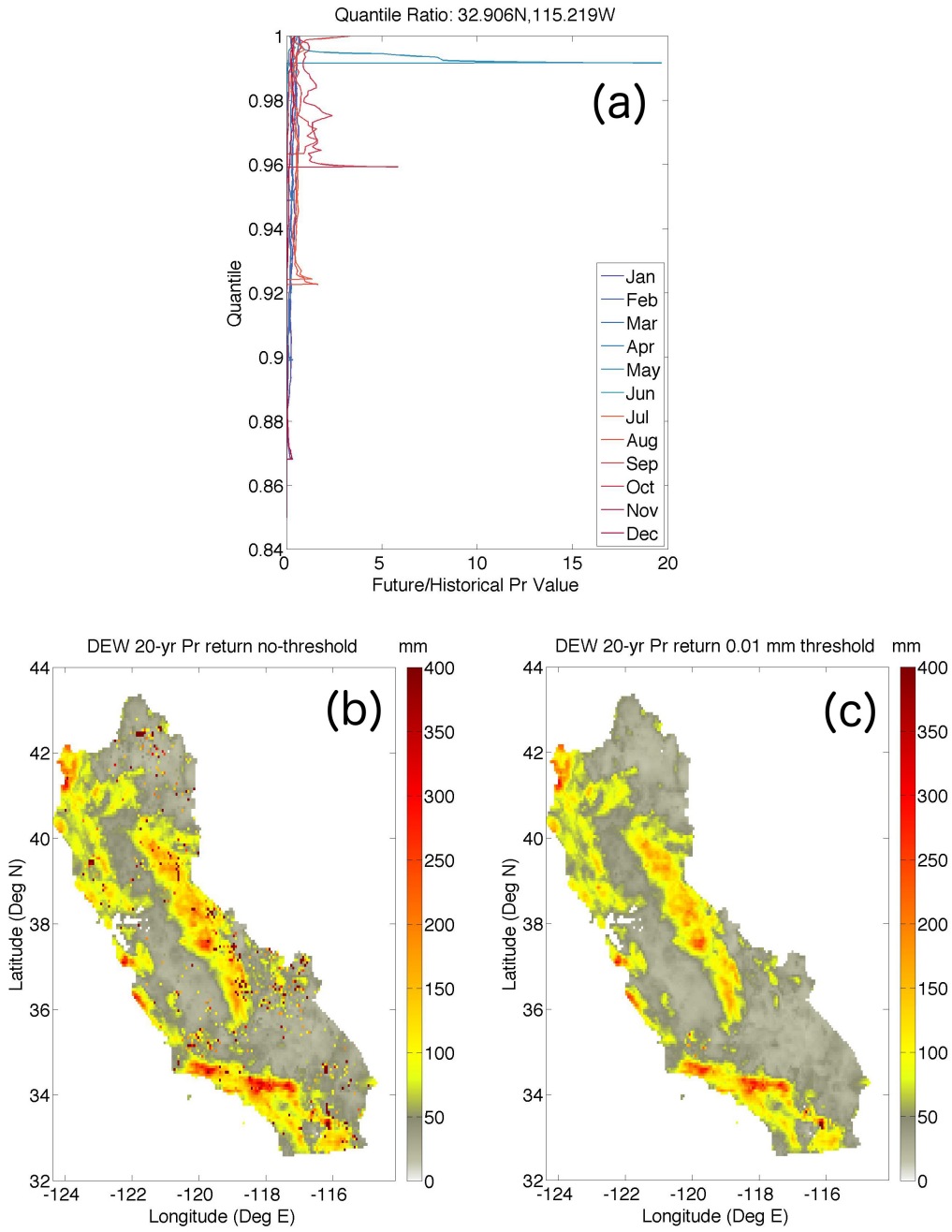


Figure 33: (a) Ratio of monthly-resolved precipitation quantiles between modeled future and modeled historical precipitation for sample grid box that exhibits an anomalously high 20-year return value. (b) Map of the 20-year daily precipitation return value for DEW conditions using QDM with no threshold for calculating the ratio of modeled future to modeled historical precipitation. (c) Same as (b) but where all historically-modeled precipitation values below 0.01 mm are mapped to exactly zero future precipitation. From Feldman et al, JWRPM, Submitted.

There are several numerical challenges that must be addressed in the practical implementation of QDM, however. First, the use of ratio-scaling to obtain mapped precipitation can lead to a numerical instability where the denominator scales differently from the numerator. QDM requires the construction of empirical CDFs from observationally-gridded products (Livneh et al, 2013) and downscaled climate model data (Pierce et al, 2014). There are thresholds associated with each of these products below which precipitation is set to zero. By simply calculating a ratio of the two CDFs, the mapped precipitation quantities can be severely distorted, but the distortion arises due to numerical instabilities in the small, but non-zero values of precipitation. Figure 33a shows that, especially for Mediterranean climates where the empirical CDF of precipitation has potentially very small values of precipitation at the lower quantiles, this numerical instability arises, and can lead to distortions in the mapped CDF. Figure 33b shows with the very high 20-year return values of precipitation that are interspersed on the map of California. This numerical instability can be managed simply by setting a threshold for when the QDM ratio for precipitation mapping should be set to zero. Figure 33c shows a map of the 20-year return values where the mapping ratio is set to zero when the historical modeled precipitation is less than 0.01 mm/day, wherein the high 20-year return values have been eliminated. This threshold is far below what is detectable from current instrumentation and has a negligible impact on hydrology, and so, as a practical matter, is recommended.

The second numerical challenge arises where precipitation values in the downscaled historical model simulations exceed the maximum observed precipitation in the historical record. Under these conditions, QDM will scale the maximum observed precipitation value in the historical record by the ratio of future to historical precipitation for the maximum quantile. This is what we do here, but we should note that could be problematic. There is no clear scaling limit (i.e., Clausius-Clapeyron) to extreme precipitation [*Lenderink et al*, 2017], and this remains an open issue that could create distortions in projections of extreme precipitation. At the very least, the occurrence of such cases can be determined, and Figure 8a indicates a map of the number of occurrences based on a historical simulation from DEW from 1981-2010, showing that in there can be as many as 100 cases of off-scale precipitation mapping in grid cells across the state and most grid cells having at least 20 case of off-scale precipitation.

The third numerical challenge faced in the creation of a QDM mapping approach pertains to a Mediterranean climate, where long-tailed precipitation distributions lead to large differences in precipitation values at the highest quantiles. As a practical matter, the implementation of a QDM requires a numerical discretization of quantiles, but the choice of this discretization can matter. QDM for precipitation mapping across California is extremely computationally and memory-intensive: an implementation wherein there are 1000 evenly-spaced quantiles requires 64 GB of memory to be run without disk-swapping. Therefore, there is a trade-off between discretization error and memory in terms of resolving the structure of the upper-precipitation quantiles. Figures 34(b-d) show the difference between mapped precipitation between 1000 and 10000 evenly-spaced quantiles, indicating that at the highest quantile, that an error of at least 5% can be incurred from discretization.

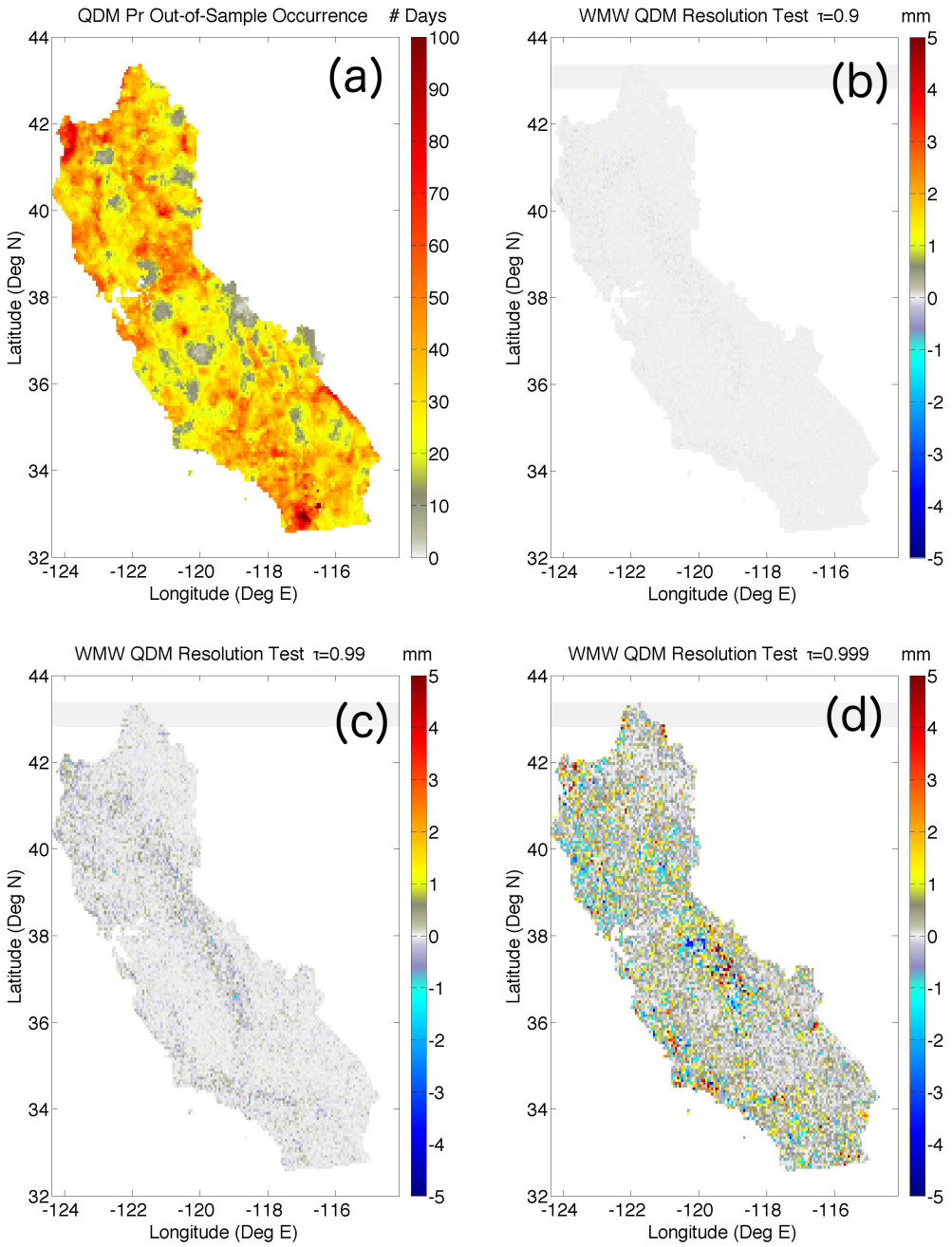


Figure 34: (a) Number of days during the 1981-2010 historical period where the downscaled CNRM-CM5 model exceeds the maximum daily precipitation value of the corresponding grid-box from the Livneh et al, [2013] dataset. (b) Difference in QDM-derived precipitation for WMW between 10000 and 1000 evenly-spaced quantiles for the 90th quantile. (c) Same as (b) but for the 99th quantile. (d) Same as (b) but for the 99.9th quantile. From Feldman et al, JWRPM, Submitted.

The QDM precipitation fields across the State of California for the DEW and WMW cases are used as inputs to a calibrated version of the Variable Infiltration Capacity (VIC) model. The

VIC model has been used to simulate regional hydrology for historical and future conditions for California as well as many major basins in the United States.

A version of the VIC Model has been configured at 1/16th degrees (approximately 6 km, or 3.75 miles) spatial resolution throughout California. The data from Livneh et al. (2013) were used as a preliminary dataset in the VIC Model setup. Parameterization within the model is performed primarily through adjustments to parameters describing the rates of infiltration and base flow as a function of soil properties, as well as the soil layer depths. When simulating in water balance mode, as done for this California application, the model is driven by daily inputs of precipitation, maximum and minimum temperature, and wind speed. The model internally calculates additional meteorological forcings, such as short- and long-wave radiation, relative humidity, vapor pressure, and vapor pressure deficits.

Five elevation bands are included for each 1/16th degree grid cell in the VIC Model to capture the precipitation and snow variability over the grid cell. In addition, the model includes a sub-daily (1-hour) computation to resolve transients in the snow model. The soil column is represented by three soil zones extending downward from the land surface to capture the vertical distribution of soil moisture. The land cover is represented by multiple vegetation types.

Rainfall, snow, infiltration, evapotranspiration, runoff, soil moisture, and base flow are computed over each grid cell on a daily basis for the entire period of simulation. The VIC Model routing tool processes the individual cell runoff and base flow terms, and routes the flow to develop streamflow at various locations in the watershed.

A streamflow routing network in the VIC Model at 1/16th degree was developed using ArcMap's Flow Direction and Flow Accumulation tools. The Flow Direction tool first assigns the flow direction for each VIC Model grid cell to its steepest downslope neighbor. Prior to processing the VIC Model grid through this tool, a stream network shapefile was burned into the digital elevation model (DEM) to enhance the performance of the flow direction tool by increasing the slope toward the closest stream. The VIC Model also requires that flow from each grid cell be directed out of the cell and into another one, and is unable to process sinks. Sinks in the DEM were filled to accommodate this. The Flow Accumulation tool then creates a raster dataset of accumulated flow to each cell by accumulating the number of all upstream cells that flow into each downslope cell.

Once the VIC Model grid is processed through these two tools, watershed delineations were determined based on downstream U.S. Geological Survey (USGS) gage locations and were compared to USGS watershed boundaries. Due to the topographic complexity of the high-elevation regions and the coarseness of the VIC Model grid, adjustments were made to the model watershed delineations to more accurately align with USGS watershed boundary delineation.

The out-of-the-box VIC Model had previously undergone only limited calibration for monthly streamflow for selected major river basins over the conterminous U.S. (Livneh et al., 2013). For this work here, further VIC Model calibration was performed for the 12 upper watershed locations in the Sacramento and San Joaquin River basins. The VIC Model was recalibrated for water years 1970-2003.

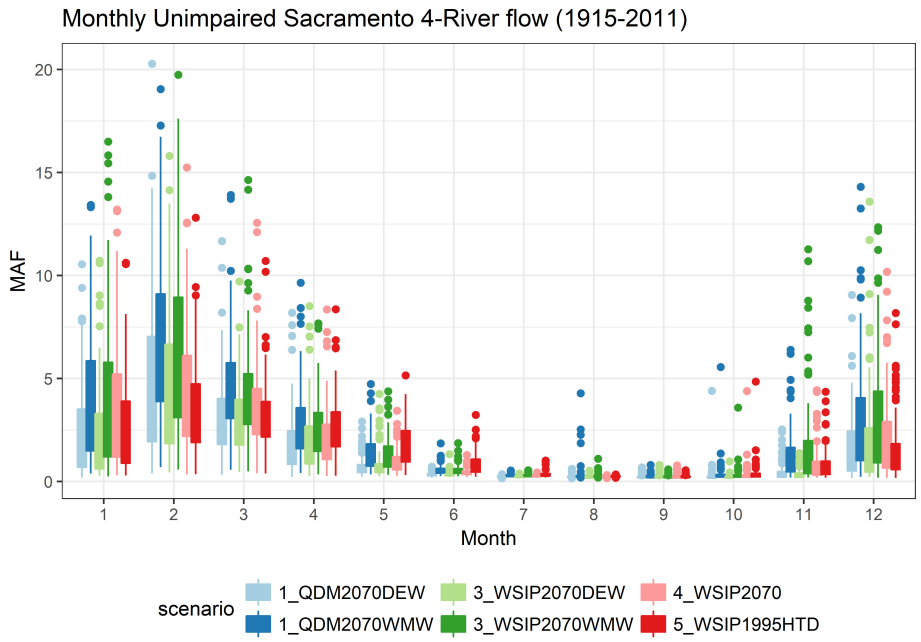


Figure 35: Results from the VIC model calibrated from historical streamflow data for California. Box-whisker plot of monthly unimpaired flows (in million acre-feet (MAF)) across four rivers in the Sacramento River basin. Light and dark blue correspond VIC output forced by QDM-derived temperature and precipitation, while light and dark red correspond to VIC output forced by QM temperature and precipitation. Red corresponds to the multi-model average.

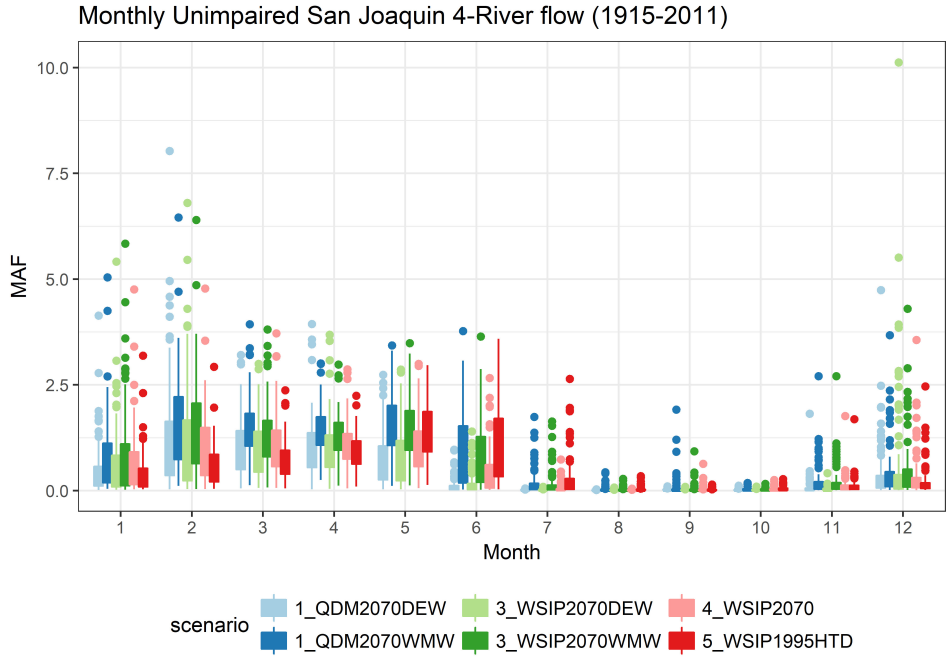


Figure 36: Results from the VIC model calibrated from historical streamflow data for California. Box-whisker plot of monthly unimpaired flows (in million acre-feet) across four rivers in the San Joaquin River basin. Light and dark blue correspond VIC output forced by QDM-derived temperature and precipitation, while light and dark red correspond to VIC output forced by QM temperature and precipitation. Red corresponds to the multi-model average.

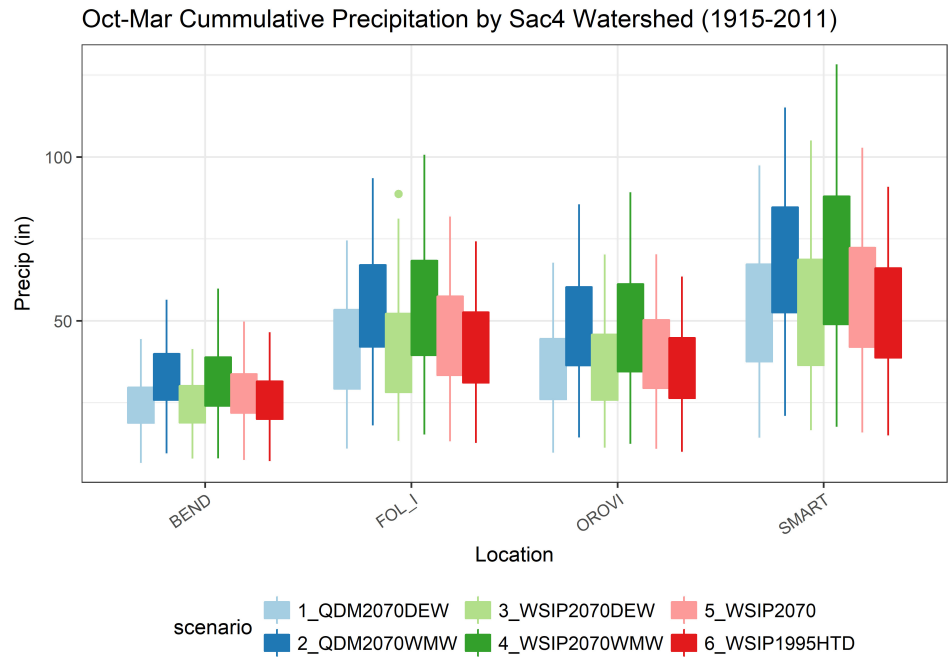


Figure 37: Rainy-season precipitation forcing (inches) from the VIC model calibrated from historical streamflow data for California across four rivers in the Sacramento River basin. From Feldman et al, JWRPM, Submitted.

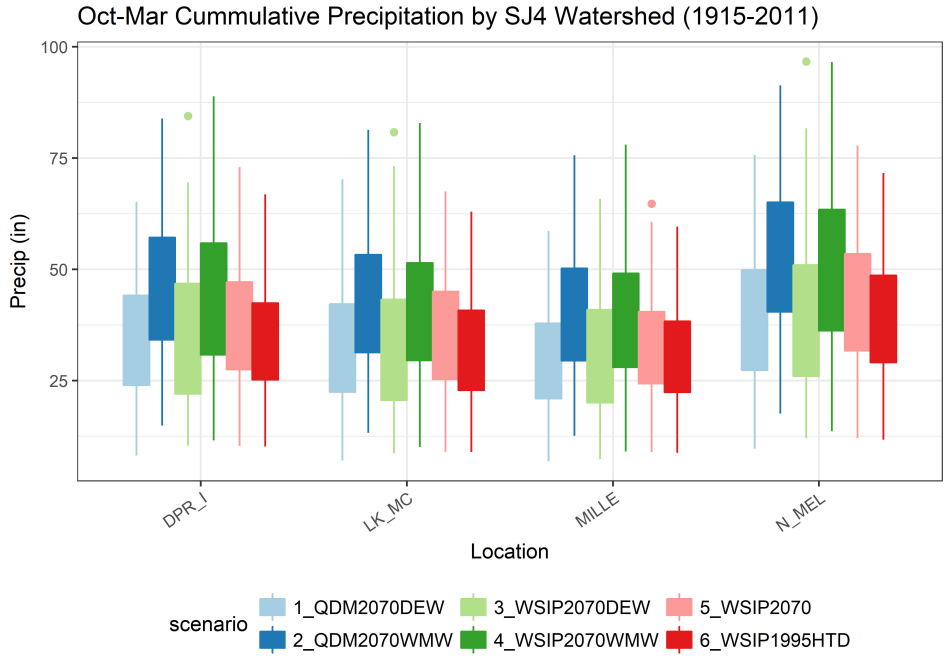


Figure 38: Rainy-season precipitation forcing (inches) from the VIC model calibrated from historical streamflow data for California across four rivers in the San Joaquin River basin. From Feldman et al, JWRPM, Submitted.

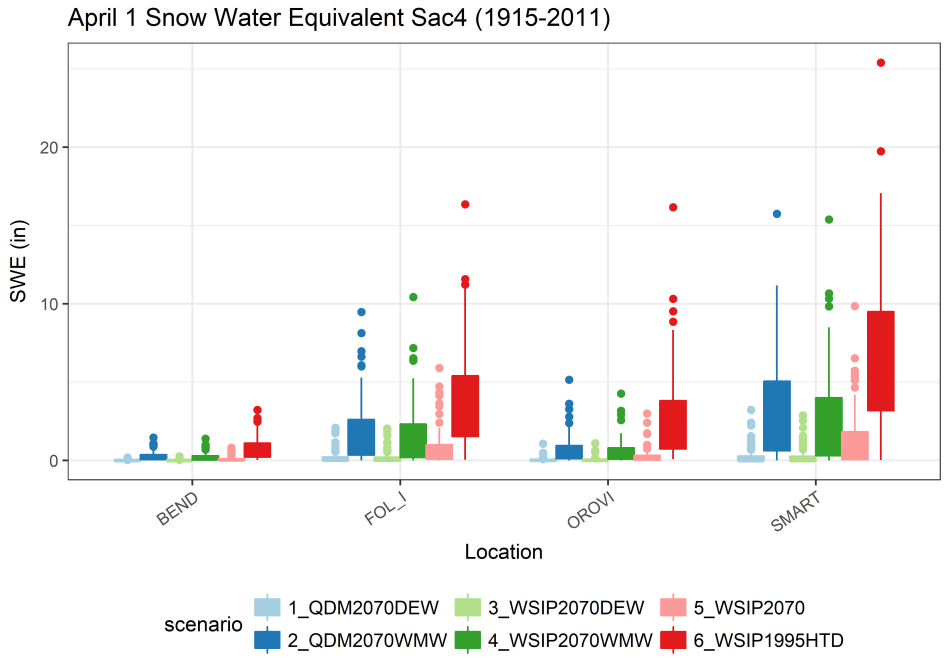


Figure 39: April 1 snow-water equivalent (SWE) from the VIC model calibrated from historical streamflow data for California across four rivers in the Sacramento River basin. From Feldman et al, JWRPM, Submitted.

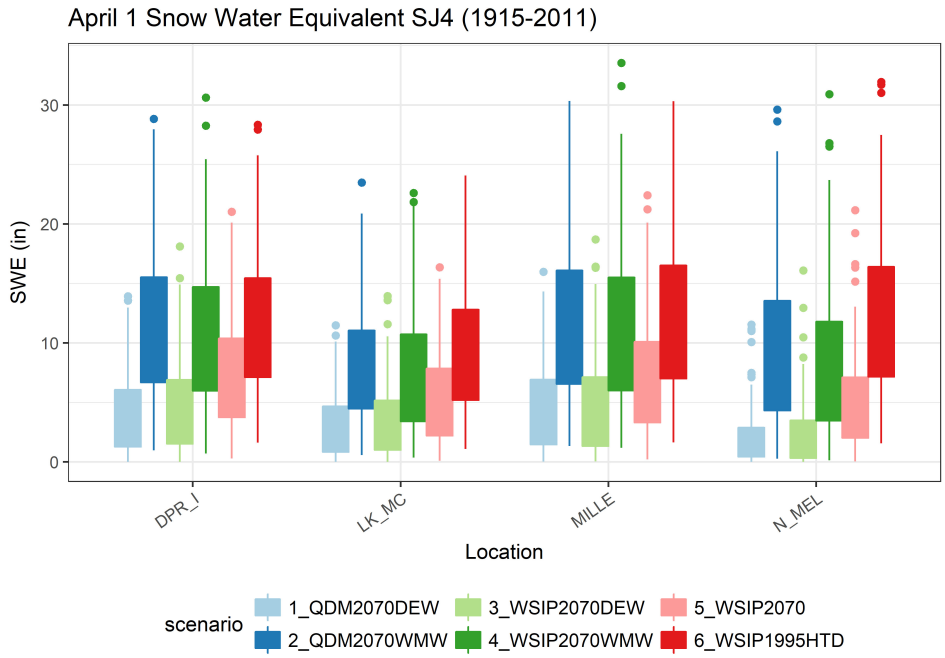


Figure 40: April 1 snow-water equivalent (SWE) from the VIC model calibrated from historical streamflow data for California across four rivers in the San Joaquin River basin. From Feldman et al, JWRPM, Submitted.

Daily VIC Model simulations were performed from 1915 to 2011. The daily runoff and base flow simulated from each grid cell was routed to various river flow locations. For the simulations performed here, streamflow was routed throughout the Sacramento and San Joaquin River basins. It is important to note that VIC Model routed flows are considered naturalized in that they do not include effects of diversions, imports, storage, or other human management of the water resource.

Even though the VIC Model is calibrated, the model bias still needs to be removed from the model outputs. These biases result from several factors, including spatial and temporal errors in gridded climate forcings, complex groundwater interactions, and other complexities normally inherent to VIC hydrologic model parameter calibration. In order to bias-correct VIC, a quantile mapping procedure was developed that used simulated and historical streamflows at each flow location to develop and rescale the CDF of streatmflow so that it is consistent with the annual scale.

The results of the calibrated VIC model forced by QDM-derived precipitation for the multi-model ensemble, the DEW, and WMW cases are shown in Figures 35-40. These plots indicate approximately a 50% range in water resource availability between the DEW and WMW cases over the Sacramento and San Joaquin River basins. While the middle panels show that the precipitation received by the WMW and DEW cases provides a bound for the multi-model ensemble, the lower panels indicate that that the multi-model ensemble predicts far more snow-water equivalent than either the DEW and WMW cases. This finding is largely a function of differences in elevation-dependent warming between these cases and the multi-model ensemble.

A Deeper Exploration into Nonstationarity (Walton et al, JAMC, In Review)

In spite of these detailed analyses for a set of DoD facilities, across the CONUS, and over California, our process chain still has not explored whether parent models and downscaling techniques are competent for handling no-analogue futures: can they actual capture the relevant nonstationarity. Through an intercomparison of downscaling techniques, we can develop clues as to which physics that may or may not break stationarity are relevant.

One of the key features differentiating dynamically and statistically downscaled temperature patterns in California is the regional imprint of snow albedo feedback changes in dynamically downscaled temperature patterns (Walton et al. 2017). However, we do not understand yet whether the deficiency of snow albedo feedback effects in statistically downscaled patterns is due to the absence of the feedback in gridded observational products used to train the statistical model (Walton et al. 2018) or whether it is due to differences in the dynamical versus statistical methods themselves. Thus, our first objective in this section is to compare April maximum temperatures across the downscaled datasets. April is used because this is when the greatest area of the Sierra Nevada consistently experiences snow albedo feedback (Walton et al., 2017). Our second objective is to evaluate how annual mean precipitation is projected to change according to the dynamical and

statistical datasets. These comparisons are guided by the following questions: How much of the disagreement between dynamical and statistical output can be explained by different precipitation magnitudes in the historical high-resolution training products (WRF vs. Livneh)? Do statistical models trained on RCM output agree more closely with future dynamical output than statistical models trained on observational products?

That being said, the findings shown in Figure 36 indicate that the growth in errors in downscaling solutions between the historical record and the end of the 21st Century suggests that either parent models or downscaling solutions are not able to capture important nonstationary effects.

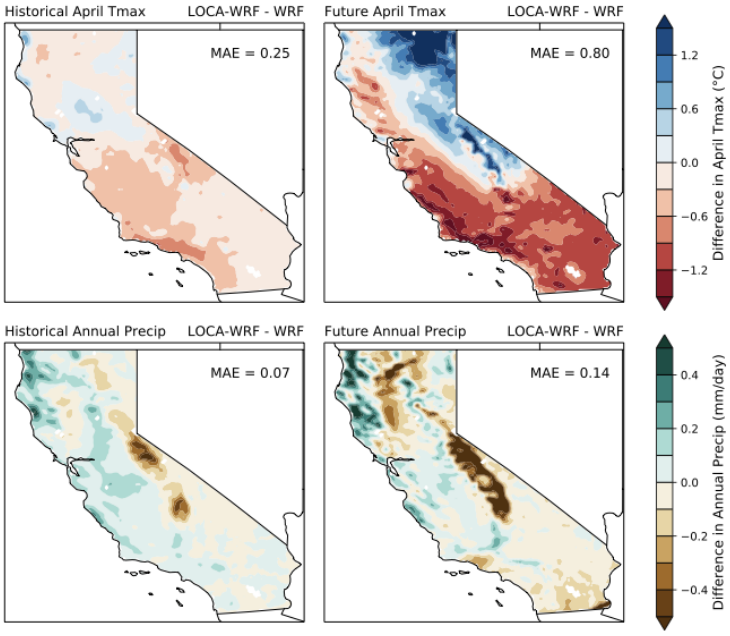


Figure 41: Differences between LOCA-WRF and WRF projections for historical and future April maximum temperature (panels a and b, respectively) and for historical and future annual mean precipitation (panels c and d, respectively). Unit for (a) and (b) is degrees C. Unit for (c) and (d) is millimeters per day. The mean absolute error (MAE) is reported in the upper right margin of each panel. From Walton et al, JAMC, In Review.

A closer analysis of the downscaling solutions is shown in Figures 37 and 38. Figure 37 shows dramatic differences between statistical and dynamical downscaled projections of April warming by the end of the century (2091-2100 average minus 1991-2000 average). We first focus on the geographical patterns (panels a, b, d, e, and f). Fig. 37a and 37b display changes within the raw GCM and the bias corrected NARR. The bias corrected NARR incorporates the GCM-predicted changes via the bias correction process. Together, the raw GCM and bias corrected NARR comprise the large-scale future forcing datasets to the dynamical and statistical procedures. Also shown are the downscaled changes according to WRF (Fig. 37b) and the two statistical datasets, LOCA-WRF (Fig. 37e) and LOCA-Livneh (Fig. 37f). Though the statewide statistics (minimum, maximum, and average) for the three downscaled projections (Figs. 37b, e, and f) are similar, their spatial patterns are not. The

raw GCM and bias corrected NARR changes (Figs. 37a and 37d) show a coastal gradient, with the least amount of warming occurring along the coast that gradually increases moving inland. However, the inland pattern of warming is inconsistent with California's actual topography and associated regional dynamics. These datasets project the greatest amount of warming (around 5-6°C) in northeastern and central California along the border with Nevada. This location of enhanced warming reflects the GCM's overly strong snow albedo feedback, which is also inaccurately positioned too far north and east in the domain. The erroneous patterns in the GCM climate change signal are reproduced by LOCA, and only appear to be interpolated to a finer grid. This is true even when LOCA is trained on WRF (Fig. 37e).

WRF's downscaling of the GCM provides a notably different solution than LOCA (Fig. 37b). In this case, warming patterns over the Sierra may have a more physically defensible interpretation, particularly over the Sierra Nevada. The physical mechanisms involved can be seen more clearly by viewing the elevation gradients of warming (Fig. 3c). Within the low-lying coastal zone and Central Valley, where elevations span from sea level to roughly 1000m, WRF projects up to a degree more of warming than the LOCA projections. This may be tied to declining cloud cover in WRF, particularly in the coastal zone of southern California that is already observed over this region (Williams et al. 2015). This phenomenon is very unlikely to be resolved by the GCM. LOCA has no means of producing it on finer scales upon downscaling, ultimately producing a weaker coastal warming gradient in LOCA relative to WRF.

Across the Sierra Nevada, where elevations extend to 3500m, WRF projects the greatest amount of April warming, around 4°C, along a narrow elevational band near 2500 meters. Warming then gradually decreases to below 3°C for elevations above 2500 meters. This gradient at least partly reflects WRF's ability to simulate the snow albedo feedback in the Sierra Nevada. Historically, the freezing line in April tends to hover around 2500 meters. Snowpack along or just higher than this elevation would be most susceptible to losses in a future, warmer climate. The snowpack losses would in turn amplify warming locally at these elevations. At the same time, snowpack above the 2500 meter freezing line would likely remain frozen in WRF. Therefore these elevations would experience less warming than mid-elevations. These physical arguments are consistent with what WRF projects in Fig. 37c: maximum warming around 2500m with lesser warming at higher elevations. Recall that besides spectral nudging of select variables along the outermost domain, WRF only receives information from the GCM along the domain boundaries. This makes WRF largely unaware of the GCM's snow albedo feedback located in the interior of the WRF domain. WRF instead generates its own climate within its domain and adds its own snow albedo feedback, which in this case aligns with our physical understanding of the Sierra Nevada system and is more accurate than the feedback contained in the GCM.

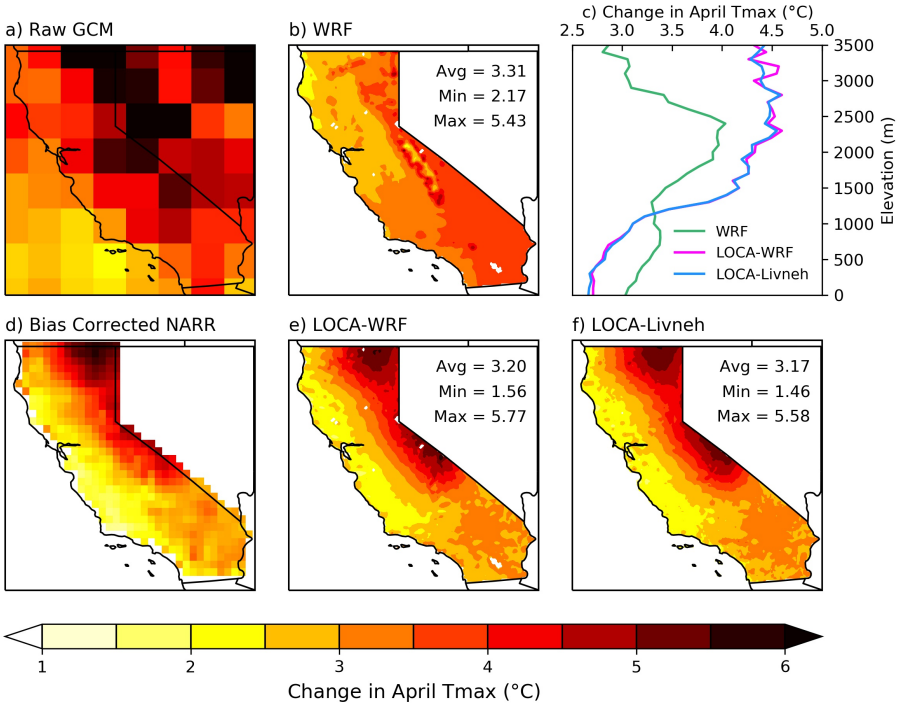


Figure 42: End-of-century (2091-2100 average minus 1991-2000 average) changes in April maximum temperature according to (a) the raw GCM (CNRM-CM5), (b) WRF, (d) bias corrected NARR, (e) LOCA-WRF, and (f) LOCA-Livneh. The statewide average, minimum, and maximum for each downscaled dataset are reported in the upper right margin of those panels. Unit for each panel is °C. Panel (c) compares change in April maximum temperature versus elevation for each grid cell within the WRF (green line), LOCA-WRF (magenta line), and LOCA-Livneh (blue line) projections. From Walton et al, JAMC, In Review.

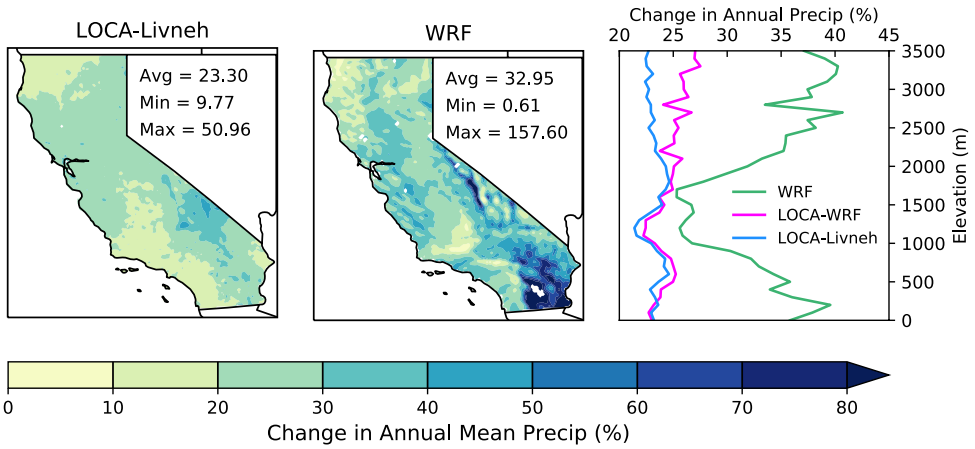


Figure 43: End-of-century (2091-2100 average minus 1991-2000 average) changes in annual mean precipitation according to (a) the raw GCM (CNRM-CM5), (b) WRF, (d) bias corrected NARR, (e) LOCA-WRF, and (f) LOCA-Livneh. The statewide average, minimum, and maximum for the downscaled products are reported in the upper right margin. Unit for each panel is percent. Panel (c) compares the change in annual mean precipitation versus elevation for each grid cell within the WRF (green line), LOCA-WRF (magenta line), and LOCA-Livneh (blue line) projections. From Walton et al, JAMC, In Review.

LOCA's elevation patterns of warming simply reflect the GCM warming signal draped on a high resolution topography, since unlike WRF, LOCA imparts little local snow albedo feedback. LOCA instead inherits the erroneous GCM signal, rendering its downscaled solution with inaccurate strength and position of the feedback across the Sierra Nevada (i.e. regions of max warming). These spatial inaccuracies are reflected in LOCA's elevational gradient of warming (Fig. 37c). LOCA projects amplified warming around, or just a bit below, 2500m (magenta and blue lines of Fig. 37c), but it fails to replicate the decline in warming for elevations above the freezing line. Again, this issue arises from the fact that LOCA only has higher warming at high elevations due to the smoothed and overly-strong snow albedo feedback inherited from the GCM (Walton et al. 2017); it adds no fine-scale information during its downscaling procedure. Overall, the results of Figure 37 indicate that LOCA - even when trained with historical WRF output - yields warming patterns that are physically inconsistent with the location and topography of the Sierra Nevada. Only the dynamical solution, which imparts its own snow albedo feedback, appropriately resolves the complex topographic-based warming signals over this domain.

Dynamical and statistical projections of annual-mean precipitation changes also exhibit striking differences (Figure 38), with some features that are similar to statistical/dynamical downscaling differences found in Pierce et al. (2012). The raw GCM and bias corrected NARR datasets both project modest wettening up to 20-30% for much of the state. Slightly larger changes appear to be concentrated in the northern half of the state, with the exception of a few strongly positive grid cells in the far southeastern corner of California in the raw GCM case. In either case, there are not any obvious fine-scale imprints in this pattern beyond the larger-scale pattern of modestly wetter future conditions, especially in the northern half of the state (discussed further by (Neelin et al. 2013)).

Both the dynamical and statistical projections show comparatively large relative increases over the dry parts of the domain, especially the Mojave Desert, the Coachella Valley and the Mojave and Imperial Valleys. The changes are much larger in WRF and are the result of small absolute changes occurring over a region with very low historical rainfall values, but may nonetheless be hydrologically significant. Since LOCA's downscaled solution is tied to the GCM's future change pattern, the results of Fig. 38 indicate that the GCM has a smaller relative increase in arid region precipitation than WRF does. Since larger precipitation increases in the extremes have a physical basis in basic thermodynamic arguments (Held and Soden 2006), the larger relative change seen in the arid regions may have a physical basis. In general, it is not clear whether WRF or LOCA's future precipitation signal is more realistic or whether either can be relied on as a quantitative projection. The WRF signals are handicapped by the pseudo-global-warming methodology, which only imposes the mean GCM changes, rather than the GCM changes during the synoptic conditions producing precipitation. LOCA, whether trained by WRF or Livneh, is limited by the library of analogs in the historical climate, which may not fully represent future conditions, and inherits biases from the GCM. Still, some of the basic qualitative features seen in Fig. 38, such as the lee-side enhancement of the precipitation increase in WRF, and the larger relative increase in the arid deserts seen with both techniques, seem physically plausible.

Conclusions and Implications for Future Research

The development of robust, defensible projections of temperature and precipitation patterns for the 21st Century at the spatial scales that are relevant and actionable for future infrastructure planning represents a far-reaching need for the DoD, and the RC18-1577 project made concrete progress towards addressing that need by defining a process chain that couples climate model projections to statistical, identifying the practical steps that are needed to develop these projections, and developing partnerships that can move forward with those practical steps.

This research has explored the landscape of developing projections of surface air temperature and precipitation at the DoD facility level from an ensemble of downscaled climate model projections. The research has shown that this very landscape is not easily navigated because the climate models have representation errors in the processes that can lead to nonstationarity and statistical downscaling solutions are constructed explicitly with a stationarity assumption that has not been tested.

Fortunately, there are tractable solutions to these challenges. Focused research efforts that develop a deeper understanding of the processes that "break" stationarity and evaluate model projections accordingly are needed. Additionally, it is critical that the researchers engaged in developing climate model projections bridge the communication divide between the modeling community and end-users so that they have the appropriate level of confidence in these projections. Ultimately, end-users need to understand the challenges and opportunities with using climate model projections for infrastructure and operations planning and so that they can utilize model projections judiciously. It is incumbent on those with knowledge and expertise in climate model projections to guide the utilization of these data for end-users.

Therefore, a targeted set of process studies, that follow the guidance of what was conducted in Walton et al, JAMC, In Review and discussed above, are needed to identify where parent models and statistical techniques are insufficient for developing temperature and precipitation projections.

One such prime example related to coastal processes associated Marine Boundary Layer (MBL). Figure 39 indicates how changes in these processes can "break" stationarity. Specifically, the marine boundary layer (MBL) plays a critical role in moderating surface air temperatures along the U.S. west coast during the summer season. The extent to which the MBL penetrates inland is largely driven by a simple interaction between the height of the marine boundary layer and the local topography.

Coastal mountains act as a barrier to further inland extent, and higher cloud [boundary layer] heights are required to penetrate further inland (Iacobellis and Cayan, 2013). The resolution of this process is substantially finer than a GCM, and either statistical or dynamical downscaling approaches could conceivably be employed. However, there are two nonlinearities in the interaction between the MBL and surface temperatures that might lead to substantial differences between the two approaches. The first nonlinearity could arise due to long-term increases in MBL heights (e.g., due to weakening subsidence (Vecchi et al, 2007) or increased urbanization (Williams et al., 2015)): if MBL heights reach the height of local topography, then it would

become common for the MBL to overcome the topographic barrier. This would represent a step-change beyond which the MBL, and associated cooling effect, can penetrate much further inland than it can in the present climate. The second nonlinearity could arise due to a long-term loss of MBL clouds (e.g., Clement et al., 2009), leading to increased surface insolation and temperatures (Iacobellis and Cayan, 2013).

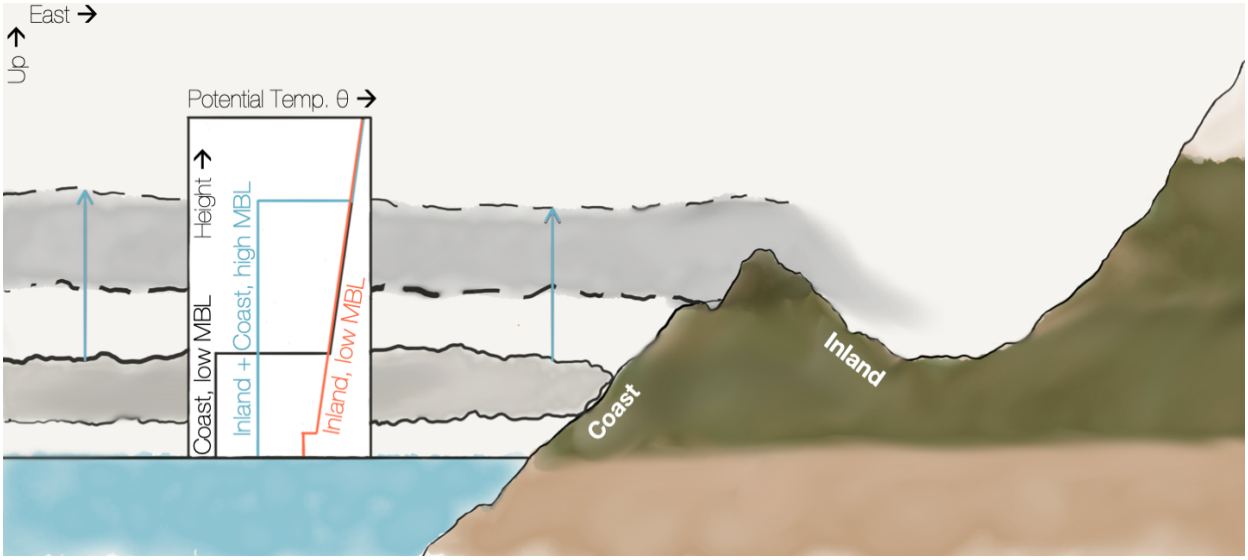


Figure 44: A diagram depicting the influence of MBL height on inland penetration of the MBL and MBL temperatures.

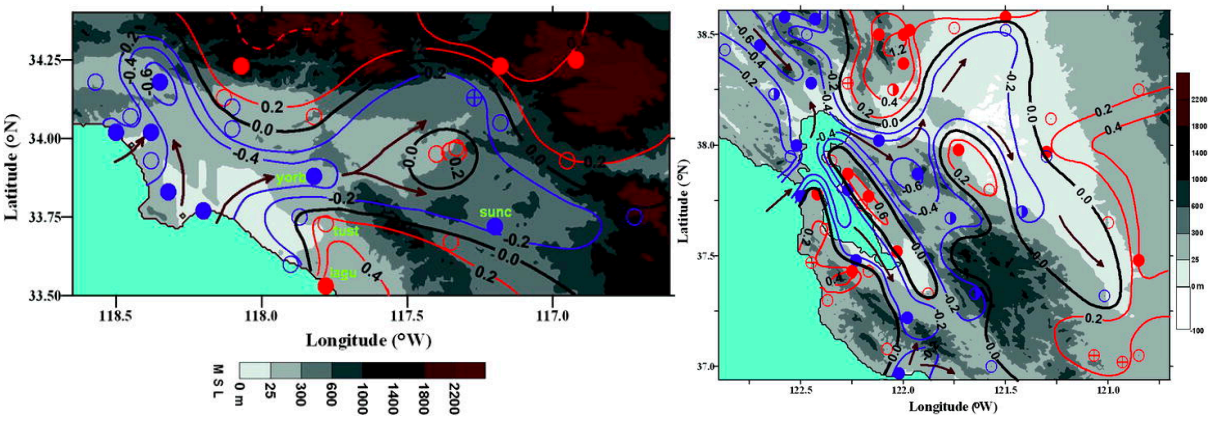


Figure 45: Historical trends over coastal California (left) from Southern California and (right) from Northern California over the period 1970-2005 with topography overlain. From Lebassi et al, 2009.

Based on this, two process-based metrics for coastal temperature change can be developed that represent a necessary condition for a model to exhibit skill in, and for users to have confidence pertaining to, that model's future projections. These metrics will both yield physical insight into the reasons that a given model needs to generally reproduce historically-observed trends in coastal and near-coastal temperatures (Lebassi et al. 2009; Potter, 2014) and highlight the (nonlinear) physical relationships between those two quantities, such that model performance

regarding these metrics is an indication that confidence can be placed in a model's future projections. The first metric will be focused on evaluating the relationship between trends in coastal and inland temperatures and trends in nearby ocean temperatures (e.g., via a regression analysis or conditional probability distribution analysis, such as in O'Brien et al., 2012). The second metric focuses on evaluating how changes in the frequency of marine air penetration events, based on the methodology of Wang and Ullrich (2018), relate to coastal/inland temperatures. Both metrics utilize commonly-available, large scale variables as both input (ocean surface temperature for the former, surface wind and temperature for the latter) and output (surface temperature for both metrics.

The primary benefit to the DoD of this seed project's research is to detail the scale of the work that is needed to address in a comprehensive fashion specific research tasks that go beyond a purely statistical approach to developing climate model projections from an ensemble of model results to provide insight into the models with which the DoD should have increased or diminished confidence. The challenges of using downscaled projections for infrastructure and operations planning should not be underestimated, because they are generally completely opaque to the end-user. These end-users can easily have either too much or too little confidence in the model projections. For the former case, the provision of a set of numbers, and even error bars, that can be fed into end-user models is a straightforward exercise, but the danger here is that those numbers can be misleading if they do not consider the processes that break stationarity.

For the latter case, the esoteric nature of climate models and downscaling can preclude their adoption even though it is very likely that long-term shifts in surface air temperature and precipitation distributions will occur. In this situation, planning exercises either use historical data, that does not include nonstationarity, or use generic information on increased variability as a placeholder for a deeper understanding of the risks associated with climate change, and this generic information may or may not be relevant to a specific location. In order to fulfill the needs expressed in RCSON-18-L2, a deeper understanding on how to use models and downscaling to capture relevant nonstationarity is needed. Confidence in downscaled projections is of paramount importance for infrastructure and planning purposes.

Literature Cited

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor (2012), Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models, *Geophys. Res. Lett.*, **39**, L09712, doi:10.1029/2012GL051607.

Bengtsson, L., K.I. Hodges, and E. Roeckner (2006) Storm Tracks and Climate Change. *J. Climate,* **19**, 3518–3543, doi:10.1175/JCLI3815.1.

Brown, B.M. and Chen, S.X., (1999) Beta-Bernstein smoothing for regression curves with compact supports. *Scandinavian Journal of Statistics*, **26**, 47-59, doi:10.1111/1467-9469.00136.

Bukovsky, M. S., & Karoly, D. J. (2007). A brief evaluation of precipitation from the North American Regional Reanalysis. *Journal of Hydrometeorology*, **8**(4), 837-846, doi: 10.1175/JHM595.1.

Cannon, A.J., and Coauthors, (2015) Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *J. Clim.*, **28**, 6938–6959, doi: 10.1175/JCLI-D-14-00754.1.

CCTAG: California Department of Water Resources, Climate Change Technical Advisory Group (2015): Perspectives and Guidance for Climate Change Analysis (2015), in: Lynn, E., Andrew Schwarz, Anderson, J., Correa, M. (Eds.), 140pp.

Charney, J. et al, National Research Council. (1979). *Carbon Dioxide and Climate: A Scientific Assessment*. Washington, DC: The National Academies Press. doi:10.17226/12181.

Charpentier, A., and Coauthors, 2007: The estimation of copulas: Theory and practice. In Rank, J., editor, Copulas: From Theory to Applications in Finance, pages 35–62. Risk Books.

Chen, S.X., 1999: A beta kernel estimation for density functions. Computational Statistics and Data analysis. 31. 131-135.

Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher (2008), On the need for bias correction of regional climate change projections of temperature and precipitation, *Geophys. Res. Lett.*, **35**, L20709, doi:10.1029/2008GL035694.

Cinquini, L., et al. (2014) The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data, *Future Generation Computer Systems*, 36, 400-417, doi: 10.1016/j.future.2013.07.002

Clement, Amy C., Robert Burgman, and Joel R. Norris (2009) Observational and Model Evidence for Positive Low-Level Cloud Feedback. *Science*, **325**(5939), 460–64. doi:10.1126/science.1171255.

Climate Change Planning Handbook Installation Adaptation and Resilience (2017) https://www.fedcenter.gov/_kd/go.cfm?destination=ShowItem&Item_ID=31041

Coles, S., 2001: An Introduction to Statistical Modeling of Extreme Values,. Springer-Verlag. ISBN 1-85233-459-2.

Dixon, K.W., J. R. Lanzante, M. J. Nath, K. Hayhoe, A. Stoner, A. Radhakrishnan, V. Balaji, and C. F. Gaitán. (2016) Evaluating the Stationarity Assumption in Statistically Downscaled Climate Projections: Is Past Performance an Indicator of Future Results? *Climatic Change* 135 (3-4): 395–408, doi: 10.1007/s10584-016-1598-0.

Dudhia, Jimy. (1989) Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model. *Journal of the Atmospheric Sciences* 46 (20): 3077–3107, doi: 10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Easterling, D. R., J. R. Arnold, T. Knutson, K. E. Kunkel, A. N. LeGrande, L. R. Leung, R. S. Vose, D. E. Waliser, and M. F. Wehner, 2017: Precipitation Change in the United States. *Climate Science Special Report: Fourth National Climate Assessment, Volume I.* Wuebbles, D. J., D. W. Fahey, K. A. Hibbard, D. J. Dokken, B. C. Stewart, and T. K. Maycock, Eds., U.S. Global Change Research Program, Washington, DC, USA, 207–230. doi:10.7930/J0H993CC.

Ehret, U., E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert, (2012) HESS opinions: ‘‘Should we apply bias correction to global and regional climate model data?’’ *Hydrol. Earth Syst. Sci.*, **16**, 3391–3404, doi:10.5194/hess-16-3391-2012.

Eyring, V., et al., (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.*, **9**, 1937-1958, doi:10.5194/gmd-9-1937-2016.

Eyring, V., et al., (2019) Taking climate model evaluation to the next level, *Nature Climate Change*, **9**, 102-110, doi:10.1038/s41558-018-0355-y.

Feldman, D.R., J.M. Tadić, W. Arnold, A.Schwarz (2019) Establishing a Range of Extreme Precipitation Estimates in California for Planning in the Face of Climate Change, *Water Resources Research* (Submitted).

Forster, P.M., (2016) Inference of Climate Sensitivity from Analysis of Earth's Energy Budget, *Annual Review of Earth and Planetary Sciences*, **44**, 85-106, doi:10.1146/annurev-earth-060614-105156.

Giorgi, F., C. Jones, and G. R. Asrar, (2009) Addressing climate information needs at the regional level: the CORDEX framework. WMO Bull., 58, 175–183.

Gudmundsson, L., J. Bremnes, J. Haugen, and T. Engen-Skaugen, (2012) Technical note: Downscaling RCM precipitation to the station scale using statistical transformations—A

comparison of methods. *Hydrol. Earth Syst. Sci.*, **16,** 3383–3390, doi:10.5194/ hess-16-3383-2012.

Haarsma, R.J., et al., (2016) High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, *Geosci. Model Dev.*, **9**, 4185-4208, doi:10.5194/gmd-9-4185-2016.

Hagemann, S., C. Chen, J. O. Haerter, J. Heinke, D. Gerten, and C. Piani (2011) Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. *J. Hydrometeor.*, **12**, 556–578, doi:10.1175/2011JHM1336.1.

Hawkins, E., and R. Sutton (2009) The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, **90**, 1095–1107, doi:10.1175/2009BAMS2607.1.

Held, I.M. and Soden, B.J., (2006). Robust responses of the hydrological cycle to global warming. Journal of Climate, **19**(21), pp.5686-5699, doi: 10.1175/JCLI3990.1

Hempel, S., K. Frieler, L. Warszawski, J. Schewe, and F. Piontek (2013) A trend-preserving bias correction—The ISI-MIP approach. *Earth System Dyn.*, **4**, 219–236, doi:10.5194/esd-4-219-2013.

Hidalgo, H., M. D. Dettinger, and D. R. Cayan, (2008) Downscaling with constructed analogues: daily precipitation and temperature fields over the United States. California Energy Commission report CEC-500-2007-123. 48 pp.

Hourdin, F., T. Mauritsen, A. Gettelman, J. Golaz, V. Balaji, Q. Duan, D. Folini, D. Ji, D. Klocke, Y. Qian, F. Rauser, C. Rio, L. Tomassini, M. Watanabe, and D. Williamson, (2017) The Art and Science of Climate Model Tuning. *Bull. Amer. Meteor. Soc.*, **98**, 589–602, doi:10.1175/BAMS-D-15-00135.1.

Iacobellis, S. F., and D. R. Cayan (2013), The variability of California summertime marine stratus: Impacts on surface air temperatures, *J. Geophys. Res. Atmos.*, **118**, 9105–9122, doi:10.1002/jgrd.50652.

Jeevanjee, N. & D.M. Romps (2018) Mean precipitation change from a deepening troposphere, *Proc. Nat. Acad. Sci. USA*, **115**, 11465-11470, doi:10.1073/pnas.1720683115.

Jones, C., F. Giorgi, and G. Asrar, (2011) The Coordinated Regional Downscaling Experiment: CORDEX—An international downscaling link to CMIP5. CLIVAR Exchanges, Vol. 56, No. 16, International CLIVAR Project Office, Southampton, United Kingdom, 34–40.

Kain, J.S. and J.M. Fritsch, (1990) A One-Dimensional Entraining/Detraining Plume Model and Its Application in Convective Parameterization. *J. Atmos. Sci.,* **47**, 2784–2802, doi:10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2.

Knutti, R., G. Hegerl (2008), The equilibrium sensitivity of the Earth's temperature to radiation changes, *Nature Geoscience*, **1**, 735-743, doi:10.1038/ngeo337.

Knutti, R., M. A. A. Rugenstein, G.C. Hegerl, (2017), Beyond equilibrium climate sensitivity, *Nature Geoscience,* **10**, 727-736, doi:10.1038/ngeo3017.

Lebassi, Bereket, Jorge González, Drazen Fabris, Edwin Maurer, Norman Miller, Cristina Milesi, Paul Switzer, and Robert Bornstein. (2009) Observed 1970–2005 Cooling of Summer Daytime Temperatures in Coastal California. *Journal of Climate*, **22**(13), 3558–73, doi:10.1175/2008JCLI2111.1.

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

Lovejoy, S., Mandelbrot, B.B. (1985) Fractal properties of rain, and a fractal model, *Tellus A: Dynamic Meteorology and Oceanography*, 37(3), 209-232, doi: 10.3402/tellusa.v37i3.11668.

Lu, J., G. A. Vecchi, and T. Reichler (2007), Expansion of the Hadley cell under global warming, *Geophys. Res. Lett.*, 34, L06805, doi:10.1029/2006GL028443.

Maraun, D. (2012) Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophys. Res. Lett.*, **39**, L06706, doi:10.1029/2012GL051210.

Maraun, D., (2013) Bias correction, quantile mapping, and downscaling: Re-visiting the inflation issue. *J. Climate*, **26**, 2137–2143, doi:10.1175/ JCLI-D-12-00821.1.

Maurer, E. P., and H. G. Hidalgo. (2007) Utility of Daily vs. Monthly Large-Scale Climate Data: An Intercomparison of Two Statistical Downscaling Methods. *Hydrology and Earth System Sciences Discussions* **4** (5): 3413–40, doi: 10.5194/hess-12-551-2008.

Maurer, E. P., and D. W. Pierce, (2014) Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrol. Earth Syst. Sci.*, **18**, 915–925, doi:10.5194/hess-18-915-2014.

Mauritsen, T., et al. (2012). Tuning the climate of a global model, *Journal of Advances in Modeling Earth Systems*, **4**, doi:10.1029/2012MS000154.

Milly, P.C.D., J. Betancourt, M. Falkenmark, R.M. Hirsch, W.Z. Kundzewicz, D.P. Lettenmaier, and R.J. Stouffer. 2008. Stationarity Is Dead: Whither Water Management? *Science*. **319**, 573–574, doi:10.1126/science.1151915.

Mearns, L. O., et al (2012) The North American Regional Climate Change Assessment Program: Overview of phase I results. *Bull. Amer. Meteor. Soc.*, **93**, 1337–1362, doi:10.1175/ BAMS-D-11-00223.1.

Meehl, G.A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J.F. Mitchell, R.J. Stouffer, and K.E. Taylor, (2007) THE WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research. *Bull. Amer. Meteor. Soc.,* **88**, 1383–1394, doi:10.1175/BAMS-88-9-1383.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jović, et al. (2006) North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, **87** (3): 343–60, doi:10.1175/BAMS-87-3-343.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A. (1997), Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated‐k model for the longwave, *J. Geophys. Res.*,102(D14), 16663– 16682, doi:10.1029/97JD00237.

Moss, R.H., et al, (2017) Nonstationary Weather Patterns and Extreme Events: Informing Design and Planning for Long-Lived Infrastructure, ESTCP Project RC-201591, https://serdp-estcp.org/content/download/46665/436260/file/Nonstationarity%20RC%20Workshop%20Report.pdf

Nakanishi, M., and H. Niino. (2004) An Improved Mellor–Yamada Level-3 Model with Condensation Physics: Its Design and Verification. *Boundary-Layer Meteorology*, **112** (1): 1–31, doi:10.1023/B:BOUN.0000020164.04146.98.

Neelin, J. David, B. Langenbrunner, J. E. Meyerson, A. Hall, and N. Berg. (2013) California Winter Precipitation Change under Global Warming in the Coupled Model Intercomparison Project Phase 5 Ensemble. *Journal of Climate* **26** (17): 6238–56, doi: 10.1175/JCLI-D-12-00514.1

Niu, Guo-Yue, Zong-Liang Yang, Kenneth E. Mitchell, Fei Chen, Michael B. Ek, Michael Barlage, Anil Kumar, et al. (2011). The Community Noah Land Surface Model with Multiparameterization Options (Noah-MP): 1. Model Description and Evaluation with Local-Scale Measurements. *Journal of Geophysical Research* **116** (D12). doi:10.1029/2010jd015139.

Ntegeka, V., P. Baguis, E. Roulin, P. Willems (2014) Developing tailored climate change scenarios for hydrological impact assessments, *J. Hydrolog.*, **508**, 307-321, doi:10.1016/j.jhydrol.2013.11.001.

O'Brien, T. A., P. Y. Chuang, L. C. Sloan, I. C. Faloona, and D. L. Rossiter. (2012) Coupling a New Turbulence Parametrization to RegCM Adds Realistic Stratocumulus Clouds. *Geoscientific Model Development* **5**(4) 989–1008. doi:10.5194/gmd-5-989-2012.

O'Neill, B.C., Kriegler, E., Riahi, K. et al. (2014) A new scenario framework for climate change research: the concept of shared socioeconomic pathways, *Climatic Change*, **122**, 387, doi:10.1007/s10584-013-0905-2.

O'Neill, B.C., et al. (2017) The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century, *Global Environmental Change*, 42, 169-180, doi: 10.1016/j.gloenvcha.2015.01.004.

Oyler, J.W., S.Z. Dobrowski, Z.A. Holden, and S.W. Running, (2016) Remotely Sensed Land Skin Temperature as a Spatial Predictor of Air Temperature across the Conterminous United States. *J. Appl. Meteor. Climatol.,* **55**, 1441–1457, doi:10.1175/JAMC-D-15-0276.1.

Pendergrass, A.G. and D.L. Hartmann (2014) Changes in the Distribution of Rain Frequency and Intensity in Response to Global Warming. *J. Climate,* **27**, 8372–8383, doi:10.1175/JCLI-D-14-00183.1.

Pierce, D. W., T. Das, D. R. Cayan, E. P. Maurer, N. L. Miller, Y. Bao, M. Kanamitsu, et al. (2012) Probabilistic Estimates of Future Changes in California Temperature and Precipitation Using Statistical and Dynamical Downscaling. *Climate Dynamics*, **40** (3-4): 839–56, doi: 10.1007/s00382-012-1337-9.

Pierce, D. W., D. R. Cayan, and B. L. Thrasher, (2014) Statistical downscaling using Localized Constructed Analogs (LOCA). *J. Hydromet.*, **15**, 2558-2585, doi:10.1175/JHM-D-14-0082.1.

Pierce, D. W., D. R. Cayan, E. P. Maurer, J. T. Abatzoglou, and K. C. Hegewisch, (2015) Improved bias correction techniques for hydrological simulations of climate change. *J. Hydrometeorology*, 16, 2421-2442. doi:10.1175/JHM-D-14-0236.1.

Pierce, D. W., D. R. Cayan, J. F. Kalansky. (Scripps Institution of Oceanography). 2018. Climate, Drought, and Sea Level Rise Scenarios for the Fourth California Climate Assessment. California's Fourth Climate Change Assessment, California Energy Commission. Publication number: CCCA4-CEC-2018-006.

Potter, Christopher. (2014), Understanding Climate Change on the California Coast: Accounting for Extreme Daily Events among Long-Term Trends. *Climate*, **2**(1), 18–27. doi:10.3390/cli2010018.

Rasmussen, R., C. Liu, K. Ikeda, D. Gochis, D. Yates, F. Chen, M. Tewari, et al. (2011) High-Resolution Coupled Climate Runoff Simulations of Seasonal Snowfall over Colorado: A Process Study of Current and Warmer Climate. *Journal of Climate*, **24** (12): 3015–48, doi:10.1175/2010JCLI3985.1.

Risser, M. D. and Calder, C. A. (2015). Regression-based covariance functions for nonstationary spatial modeling. *Environmetrics*, **26**(4):284–297, doi:10.1002/env.2336.

Risser, M.D., C.J. Paciorek & D. A. Stone (2018) Spatially Dependent Multiple Testing Under Model Misspecification, With Application to Detection of Anthropogenic Influence on Extreme Climate Events, *Journal of the American Statistical Association*, doi: 10.1080/01621459.2018.1451335

Risser, M.D., Paciorek, C.J., O'Brien, T.A., Wehner, M.F., Collins, W.D., (2019a) A Probabilistic Gridded Product for Daily Precipitation Extremes Over the United States, *Climate Dynamics*, doi: 10.1007/s00382-019-04636-0.

Risser, M.D., Paciorek, C., O'Brien, T.A., Wehner, M.F., Collins, W.D. (2019b) Detected changes in precipitation extremes at their native scales derived from in situ measurements, *J. Climate*, (Accepted) https://arxiv.org/abs/1902.05977.

Ruane A C, Teichmann C, Arnell N W, Carter T R, Ebi K L, Frieler K, Goodess C M, Hewitson B, et al. (2016). The Vulnerability, Impacts, Adaptation and Climate Services Advisory Board (VIACS AB v1.0) contribution to CMIP6. *Geoscientific Model Development* **9** (9): 3493-3515. DOI:10.5194/gmd-9-3493-2016.

Rudin, W., 1976: Principles of Mathematical Analysis. New York: McGraw Hill. pp. 89–90. ISBN.

Rupp, D. E., J. T. Abatzoglou, K. C. Hegewisch, and P. W. Mote (2013), Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA, *J. Geophys. Res. Atmos.*, **118**, 10,884–10,906, doi:10.1002/jgrd.50843.

Sanderson, B. M., Knutti, R., and Caldwell, P. (2015) A representative democracy to reduce interdependency in a multimodel ensemble, *J. Climate*, **28**, 5171–5194, doi:10.1175/JCLI-D-14-00362.1.

Sanderson, B.M., Wehner, M., Knutti, R. (2017). Skill and independence weighting for multi-model assessments, *Geoscientific Model Development*, **10**, 2379-2395, doi:10.5194/gmd-10-2379-2017.

Sato, T., F. Kimura, and A. Kitoh. (2007) Projection of Global Warming onto Regional Precipitation over Mongolia Using a Regional Climate Model. *Journal of Hydrology*, **333** (1): 144–54, doi:10.1016/j.jhydrol.2006.07.023.

Schär, C., C. Frei, D. Lüthi, and H. C. Davies. (1996) Surrogate Climate- Change Scenarios for Regional Climate Models. *Geophysical Research Letters*, **23** (6): 669–72, doi: 10.1029/96GL00265.

Schmidt, G.A., D. Bader, L.J. Donner, G.S. Elsaesser, J.-C., Golaz, C. Hannay, A. Molod, R.B. Neale, S. Saha (2017) Practice and philosophy of climate model tuning across six US modeling centers, *Geosci. Mod. Dev.*, 10, 3207-3223, doi:10.5194/gmd-10-3207-2017.

Schmidt, T., 2007: Coping with copulas, in: Copulas – From Theory to Application in Finance, Risk Books, London, 3–34.

Schwartz, M., A. Hall, F. Sun, D. Walton, and N. Berg. (2017). Significant and Inevitable End-of-Twenty-First-Century Advances in Surface Runoff Timing in California's Sierra Nevada. Journal of Hydrometeorology 18 (12): 3181–97, doi: 10.1175/JHM-D-16-0257.1

Sheffield, J., A.P. Barrett, B. Colle, D. Nelun Fernando, R. Fu, K.L. Geil, Q. Hu, J. Kinter, S. Kumar, B. Langenbrunner, K. Lombardo, L.N. Long, E. Maloney, A. Mariotti, J.E. Meyerson, K.C. Mo, J. David Neelin, S. Nigam, Z. Pan, T. Ren, A. Ruiz-Barradas, Y.L. Serra, A. Seth, J.M. Thibeault, J.C. Stroeve, Z. Yang, and L. Yin, (2013) North American Climate in CMIP5 Experiments. Part I: Evaluation of Historical Simulations of Continental and Regional Climatology. *J. Climate,* **26**, 9209–9245, doi:10.1175/JCLI-D-12-00592.1.

Sillmann, J., V. Kharin, X. Zhang, F. Zwiers, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.*, **118**, 1716–1733, doi:10.1002/jgrd.50203.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 125 pp., doi:10.5065/D68S4MVH.

Sklar, A., 1959 : Fonctions de repartition a *n* dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris 8, 229–231.

Tadić, J.M. and S.C. Biraud (2019) Towards More Accurate Modeling of Future Weather: Challenges of Downscaling, Bias Correction, Correlation, Risk Assessment, and Non-Stationarity, *Journal of Applied Meteorology and Climatology*, (Submitted).

Taylor, K.E., R.J. Stouffer, and G.A. Meehl, (2012) An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteor. Soc.,* **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Teutschbein, C., and J. Seibert, (2012) Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *J. Hydrol.*, **456–457**, 12–29, doi:10.1016/j.jhydrol.2012.05.052.

Teutschbein, C. and Seibert, J. (2013) Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? *Hydrol. Earth Syst. Sci.*, **17**, 5061-5077, doi:10.5194/hess-17-5061-2013.

Themeßl, M. J., A. Gobiet, and G. Heinrich, (2012) Empirical- statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, **112**, 449–468, doi:10.1007/s10584-011-0224-4.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall. (2008) Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization. *Monthly Weather Review* **136** (12): 5095–5115, doi: 10.1175/2008MWR2387.1.

Thorne, James H., Joseph Wraithwall, Guido Franco. (2018). California's Changing

Climate 2018. *California's Fourth Climate Change Assessment*, California Natural Resources Agency.

Timmermans, B., Wehner, M., Cooley, D., O'Brien, T, Thibaud, E., & Krishnan, H. (2018). An evaluation of the consistency of extremes in gridded precipitation data sets. *Climate Dynamics*, 1-20, doi:10.1007/s00382-018-4537-0.

Ulbrich, U., J.G. Pinto, H. Kupfer, G.C. Leckebusch, T. Spangehl, and M. Reyers (2008) Changing Northern Hemisphere Storm Tracks in an Ensemble of IPCC Climate Change Simulations. *J. Climate,* **21**, 1669–1679, doi:10.1175/2007JCLI1992.1.

USGCRP, 2017: *Climate Science Special Report: Fourth National Climate Assessment, Volume I* [Wuebbles, D.J., D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C. Stewart, and T.K. Maycock (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, 470 pp, doi: 10.7930/J0J964J6.

Van den Dool, H. M. (1994) Searching for analogues, how long must we wait?. *Tellus A* **46.3** 314-324, doi:10.1034/j.1600-0870.1994.t01-2-00006.x

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G., Kram, T., Krey, V., Lamarque, J.-F., Masui, To., Meinshausen, M., Nakicenovic, N., Smith, S., Rose, S. K. (2011). The representative concentration pathways: an overview. *Climatic Change*, **109**, 5-3, doi:10.1007/s10584-011-0148-z.

Vecchi, Gabriel A., Brian J. Soden, Andrew T. Wittenberg, Isaac M. Held, Ants Leetmaa, and Matthew J. Harrison. (2006) Weakening of Tropical Pacific Atmospheric Circulation Due to Anthropogenic Forcing. *Nature*, **441**(7089), 73–76. doi:10.1038/nature04744.

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D. et al. (2013) The CNRM-CM5.1 global climate model: description and basic evaluation, *Clim Dyn,* **40**, 2091-2121, doi: 10.1007/s00382-011-1259-y.

Vose, R. S., D. R. Easterling, K. E. Kunkel, A. N. LeGrande, and M. F. Wehner, 2017: Temperature Changes in the United States. *Climate Science Special Report: Fourth National Climate Assessment, Volume I*. Wuebbles, D. J., D. W. Fahey, K. A. Hibbard, D. J. Dokken, B. C. Stewart, and T. K. Maycock, Eds., U.S. Global Change Research Program, Washington, DC, USA, 185–206. doi:10.7930/J0N29V45.

Walton, D. B., A. Hall, N. Berg, M. Schwartz, and F. Sun. (2017) Incorporating Snow Albedo Feedback into Downscaled Temperature and Snow Cover Projections for California's Sierra Nevada. *Journal of Climate*, **30** (4): 1417–38, doi: 10.1175/JCLI-D-16-0168.1.

Walton, D., and A. Hall. (2018). An assessment of high-resolution gridded temperature datasets over California. *Journal of Climate* **31**.10 , 3789-3810, doi: 10.1175/JCLI-D-17-0410.1

Wang, M., and P. Ullrich. (2018) Marine Air Penetration in California's Central Valley: Meteorological Drivers and the Impact of Climate Change. *Journal of Applied Meteorology and Climatology*, **57**(1), 137–54. doi:10.1175/JAMC-D-17-0089.1.

Williams, A. P., Schwartz, R. E., Iacobellis, S., Seager, R., Cook, B. I., Still, C. J., Husak, G. and Michaelsen, J.(2015), Urbanization causes increased cloud base height and decreased fog in coastal Southern California. *Geophys. Res. Lett.*, 42: 1527– 1536. doi: 10.1002/2015GL063266.

Yin, J. H. (2005), A consistent poleward shift of the storm tracks in simulations of 21st century climate, *Geophys. Res. Lett.*, **32**, L18701, doi:10.1029/2005GL023684.