UNIVERSITY OF CALIFORNIA

Los Angeles

Linking global and local scales of climate impacts, adaptation, and intervention

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Environment and Sustainability

by

William Krantz

2024

ABSTRACT OF THE DISSERTATION

Linking global and local scales of climate impacts, adaptation, and intervention

by

William Krantz

Doctor of Philosophy in Environment and Sustainability

University of California, Los Angeles, 2024

Professor Alexander Dean Hall, Co-Chair

Professor Edward Parson, Co-Chair

The work in this dissertation brings a focus on climate variability and global teleconnections to inform adaptation planning and explore the risks of emerging climate intervention technologies. In the first half of the dissertation, I demonstrate a modeling framework for evaluating the global risks of a regional climate intervention, demonstrating how physical pathways for potential teleconnections can be identified using global climate models. I pair this analysis with an evaluation of the regulatory landscape for research on climate cooling techniques, tracing out a complex landscape of physical and geopolitical risks that will govern the development of climate intervention technologies. In the second half of the dissertation, I

evaluate how features of global climate change drive regional impacts and use this understanding to evaluate global climate models for adaptation planning in California, showing how incorporating measurements of large-scale atmospheric circulation along with measurements of local climate model performance can lead to more robust and decision-relevant climate model evaluation.

The thesis of William Krantz is approved.

Naomi Goldenson

William Clarence Boyd

J David Neelin

Alexander Dean Hall, Committee Co-Chair

Edward Parson, Committee Co-Chair

University of California, Los Angeles

2024

Table of Contents

List of Figures

performance. Metrics evaluating similar processes are grouped together by label color, and metrics not used for total error score are in gray. Models are ordered from worst (top) to best (bottom) based on the total euclidean distance error score.

Figure 4-2. The total process-based error score for each model, with lower values representing better model performance. Error bars indicate the estimated standard deviation from an ensemble of model runs due to CESM2-derived internal climate variability.

Figure 4-3. The local climate error score (x-axis) compared to the process-based climate error score (y-axis) for models in CMIP6. Error bars represent estimated uncertainty due to internal variability among simulations from the same model. Better performance is indicated by a lower error score on both axes (lower left corner).

Figure 4-4. The scatterplot as shown in Figure 4-3, but with a color scale showing model performance on the subset of process-based extreme precipitation metrics (a) and local 10-year variability (b).

Figure 4-5. Time series of monthly precipitation and detrended temperature anomaly in California for seven selected CMIP6 simulations in the SSP3-7.0 scenario. Anomalies are defined relative to a historical period of 1950-2014. Highlighted sections indicate each of four types of extreme events.

Figure 4-6. Changes in mean (left) and interannual variability (right, measured as the standard deviation of annual mean values) of precipitation and surface air temperature within California. Values are calculated as the change between a 30-year climatology at the end of the 20^{th} century (1971-2000) and end of the 21^{st} century (2071-2100), normalized by the number of doublings of CO_2 over the same time period. Normalization allows simulations from different emissions scenarios to share the same axes. Grey points show all simulations in CMIP6 with available data.

Figure 4-S1. The average change in rank across all evaluated models (CMIP5 and CMIP6) when each metric is removed in turn. The redundancy removal and calculation of the total error score is performed independently after each metric is removed.

Figure 4-S2. As in Figure 4-1 but including CMIP5 models.

Figure 4-S3 As in Figure 4-2 but including CMIP5 models

Figure 4-S4. Full matrix of local metric scores, as in Pierce (2022)

Figure 4-S5 R^2 score for regression model predicting each local metric from the full set of process-based metrics (above) or each process-based from the full set of local metrics (below).

Figure 4-S6 As in Figure 4-3 but including CMIP5 models

Figure 4-S7 Violin plots of the aggregated local and process metric scores across all CMIP5 and CMIP6 models evaluated.

Figure A1-1. (a) Seasonal precipitation change in the WRF simulations compared

to the original CMIP6 simulations, from Rahimi 2023, (b) The watershed definitions used in this study.

Figure A1-2. Time-series of Standardized Precipitation Evapotranspiration Index (SPEI) for in four river basins for WRF downscaled simulations. Shading indicates when three or four of the basins are simultaneously in drought condition, defined as an SPEI below -1. SPEI is normalized to the period of 1980-2015 for all models.

Figure A1-3. (a) The change in average months per year in drought conditions (SPEI < -1) between the historical period (1980-2010) and the end of the century (2070-2100), (b) The

List of Tables

Acknowledgements

For chapter two we acknowledge the contributions of co-authors J David Neelin and Fiaz Ahmed for scientific guidance, modeling support, manuscript drafting, and review. In addition, this work was supported in part by National Science Foundation Grant AGS-2414576, AGS-1936810 and AGS-2225956 (FA). We would like to acknowledge high-performance computing support from the Derecho system (doi:10.5065/qx9a-pg09) provided by the NSF National Center for Atmospheric Research (NCAR), sponsored by the National Science Foundation. The authors would like to thank Ted Parson for his insights guiding the original design of this study. For chapter three we acknowledge co-author Ted Parson for mentorship, project guidance, and extensive review and editing of the manuscript. For chapter four we acknowledge co-authors Naomi Goldenson for providing mentorship, scientific guidance, support developing the data processing and statistical methods, and contributions to the manuscript. We acknowledge co-authors David W Pierce, Isla Simpson, Jesse Norris, Yen-Heng Lin, Di Chen for contributing metric calculations and data for the analysis, as well as contributions to the manuscript. We also acknowledge Daniel Cayan, Stefan Rahimi, Alexander Hall for scientific guidance and review of the manuscript.

Vita

2024-	Eagle Rock Analytics
2014-2018	Instrumentation Engineer – Planet Labs
2012-2014	Nuclear Policy Working Group – University of California Berkeley
2012	Bachelor of Arts in Physics – University of California Berkeley
2009-2012	Experimental Cosmology Group – University of California Berkeley

Publications and Presentations

Rahimi, S., Huang, L., Norris, J., Hall, A., Goldenson, N., Krantz, W., Bass, B., Thackeray, C., Lin, H., & Chen, D. (2023a). An overview of the Western United States dynamically downscaled dataset (WUS-D3). Geoscientific Model Development Discussions, 2023, 1–56.

Krantz, W., Bass, B., Dennis, E. J., Rahimi, S., & Hall, A. D. (2023). Understanding and Constraining Projections of Concurrent western US Drought at the Watershed-Scale. 2023(1245), GC41K-1245.

Goldenson, N., Leung, L. R., Mearns, L. O., Pierce, D. W., Reed, K. A., Simpson, I. R., Ullrich, P., Krantz, W., Hall, A., & Jones, A. (2023). Use-inspired, process-oriented GCM Selection: Prioritizing models for regional dynamical downscaling. Bulletin of the American Meteorological Society, 104(9), E1619–E1629.

Dennis, E., Goldenson, N., Krantz, W., Rahimi-Esfarjani, S. R., Hall, A., & Nasseri, I. (2023). Application of Downscaled Extreme Precipitation to Flood Control Agency Operations: A Framework for Stakeholder-driven Climate Science. Authorea Preprints.

Rahimi, S., Krantz, W., Lin, Y., Bass, B., Goldenson, N., Hall, A., Lebo, Z. J., & Norris, J. (2022). Evaluation of a reanalysis-driven configuration of WRF4 over the western United States from 1980 to 2020. *Journal of Geophysical Research: Atmospheres*, *127*(4), e2021JD035699.

Sykora-Bodie, S. T., Jones, J. L., Hastings, Z., Lombardi, E., Barnett, M., Davis, O. N., Ferrari, O. M., Garcia Polanco, V., Hofner, A. N., & Hunter, B. (2022). Graduate student perspectives on transforming academia. Conservation Science and Practice, 4(1), e556.

Krantz, W., Neelin, J. D., & Ahmed, F. (2022). Limiting global impacts of regional climate intervention: A case study in the Great Barrier Reef. 2022, GC21B-02.

Thomas, N., Dennis, E., Krantz, W., Doherty, O. M., Jones, A. D., & Rahimi, S. (2022). Informing Decision-making about Extreme Events with a Dynamically-Downscaled GCM Ensemble. Fall Meeting 2022.

Goldenson, N. L., Dennis, E., Krantz, W., Doherty, O. M., Jones, A. D., & Rahimi, S. (2022). Informing Decision-making about Extreme Events with a Dynamically-Downscaled GCM Ensemble. 2022, A55C-01.

Goldenson, N., Mearns, L., Pierce, D., Reed, K., Ullrich, P., Hall, A., Leung, L. R., Jones, A., Krantz, W., & Rahimi, S. (2021). Prioritizing GCMs for Regional Dynamical Downscaling. 2021, A53E-01.

Krantz, W., Goldenson, N., Pierce, D., Simpson, I., Lin, Y.-H., Chen, D., Norris, J., Rahimi, S., & Hall, A. (2021). A case study for decision-relevant GCM selection in the western US. 2021, SY24B-01.

Krantz, W., Pierce, D., Goldenson, N., & Cayan, D. (2021). Memorandum on evaluating global climate models for studying regional climate change in California. Tech. Rep.).

Rahimi, S., Goldenson, N., Hall, A., Lebo, Z., Pierce, D., Cayan, D., & Krantz, W. (2021). Modification of downscaled climate trends depending on bias correction before dynamical downscaling increases uncertainty in regional projections. 2021, A15H-1763.

Krantz, W., Rahimi, S., Bass, B. J., Goldenson, N. L., & Hall, A. D. (2020). Development and validation of a dynamically downscaled historical dataset over the Western United States. 2020, A093-0006.

Rahimi, S., Krantz, W., Bass, B. J., Goldenson, N. L., Hall, A. D., & Lebo, Z. J. (2020). Multi-resolution dynamical downscaling of CMIP6 GCMs across the Western United States using WRF4: Quantifying climate change uncertainty down to the convective scale. 2020, A093-0018.

Buck, H. J., Martin, L. J., Geden, O., Kareiva, P., Koslov, L., Krantz, W., Kravitz, B., Noël, J., Parson, E. A., & Preston, C. J. (2020). Evaluating the efficacy and equity of environmental stopgap measures. Nature Sustainability, 3(7), 499–504.

Chapter 1 - Introduction

Overview

Over the past few decades, the impacts of global climate change have emerged through heat waves, prolonged droughts, wildfires, intensifying storm systems, and severe disruptions to terrestrial and marine ecosystems. Amid a broad consensus that these threats will continue to increase in frequency and severity over the coming decades (IPCC, 2022), local policymakers and planners urgently need to invest in adaptation measures that will moderate the harm to human wellbeing and the natural systems that we depend on. Adaptation measures may take the form of infrastructure to better manage threatened water supplies, economic incentives to shift agricultural practices, social programs to directly support impacted communities, or more dramatic attempts to intervene in the regional climate. Protecting communities and ecosystems from these complex and compounding threats requires coordinated mobilization across scientific, political, and social arenas. Paramount to enabling effective adaptation is understanding the projected changes to regional climate, quantifying the uncertainty around these projections, and translating these climate projections into estimates of the real human impacts. The projects in this dissertation explore these questions at the boundary of physical science and societal decision-making around climate adaptation. The results will guide policymakers and planners to make informed decisions to offset the most harm using the best tools from climate science.

There are two major domains of work in this dissertation. The first domain examines regional climate interventions as a climate adaptation tool to protect vulnerable ecosystems or populations. The two projects in this domain explore the physical and geopolitical consequences of regional climate interventions and establish a framework for planning research and deployment of regional interventions in a responsible manner.

The second domain focuses on climate impacts across the western United States and how they are represented in state-of-the-art climate models. The work in this domain closely examines physical processes that influence regional climate hazards and evaluate how they are represented by global climate models. These evaluations are used to produce better understanding of what specific regional impacts are forecast and how much irreducible uncertainty remains.

These two domains of research are unified by their focus on how large-scale modes of climate variability and teleconnections link global and regional scales of climate change. In the work focused on western US climate impacts, I examine the ways that large-scale climate modes interact to produce regional climate hazards. In the climate intervention research, I examine regional-to-global linkages in the other direction, demonstrating how regional changes propagate globally through teleconnected processes. By spanning both science and policy questions across regional and global scales, this dissertation illustrates the broad and interdisciplinary perspective that is needed to face the climate crisis.

Regional climate interventions

The second chapter of this dissertation presents work evaluating the risks of a regional climate intervention triggering unintended global consequences in the climate through teleconnections. The original motivation for this study emerged from discussions regarding early tests of marine cloud brightening (MCB) over the Great Barrier Reef (Tollefson, 2021). These conversations highlighted a gap in our understanding of whether international trans-boundary harms were a legitimate concern for these regional-scale climate interventions. Although existing studies of global climate intervention through marine cloud brightening (MCB) showed that cooling large patches of the ocean could influence global circulation patterns and produce significant remote impacts to temperature and precipitation (Jones and Haywood, 2012), no one had clearly demonstrated what scale of intervention was small enough that non-local influences would not be a meaningful concern. By conducting a modeling study to explore this question for the specific case of the Great Barrier Reef, we identified the unique physical pathways by which a regional cooling perturbation could influence global circulation patterns, while also showing how an operational intervention can be designed to meet the local ecological goals without triggering any significant non-local impacts.

Through the process of designing this modeling study, we developed a general framework for effectively using a global climate model to evaluate and avoid the risks of remote impacts from regional climate interventions. Recognizing that the specific physical pathways that could cause unintended remote effects are unique to each geographic location, and that teleconnection processes are represented differently in every climate model, our hope is that this framework will aid and encourage further studies to evaluate and cross-validate this type of risk assessment for a variety of proposed regional climate interventions.

The conclusion from this modeling study that small-scale ocean cooling efforts pose little risk of generating trans-boundary harms has meaningful implications for the international governance of marine cloud brightening and similar technologies, which is examined in the next chapter of this dissertation. In chapter three I present a policy analysis reviewing the regulatory frameworks that would govern marine cloud brightening research in the United States. This analysis examines how existing environmental regulations and research oversight mechanisms might apply to MCB field trials, drawing on the history of US regulations on weather modification and recent developments in MCB research plans.

Regional climate impacts across the western US

In chapter four I turn to the second domain of research, focusing on how to translate simulations from global climate models into actionable climate data for regional planners and local governments. Although modern climate models continue to make measurable improvements in their ability to simulate global climate dynamics (Cannon, 2020; Pierce et al., 2022; Simpson et al., 2020), they still contain biases in important regional processes that present a significant challenge for anyone seeking to make well-informed climate adaptation decisions (Abdelmoaty et al., 2021; Kim et al., 2020; Priestley et al., 2022). These limitations necessitate a careful evaluation and curation process when selecting climate data for regional planning and impact assessment.

The work presented in this chapter arose out of the need to produce a new dataset to support California's fifth climate assessment, and the recognition that previous model evaluations under-emphasized the role of large-scale climate variability and hemispheric circulation patterns in driving climate impacts in California (Goldenson et al., 2023). The manuscript presents a study that was both a practical undertaking to select the best set of models for studying climate change in the western US, and an in-depth investigation of how to effectively incorporate metrics evaluating key climate processes into a model selection process. The results indicate that the process-based metrics that we introduced provide a complementary evaluation to existing methods, resulting in a more well-informed selection of models while also highlighting blind-spots that should be improved in future evaluations. In this study I also develop a unique process for balancing the curation of a model ensemble to meet simultaneous goals that are relevant to adaptation planning: providing representation of uncertainty from model differences, internal variability, and the presence of impactful extreme events.

Based on the results of the manuscript presented here, an ensemble of climate simulations has subsequently been downscaled over the western US, creating a unique and valuable dataset being used for science and adaptation planning across California (Rahimi et al., 2024a). With the completion of the downscaled data, there is a new set of opportunities to explore how the relationship between large-scale climate and regional impacts is carried into the higher resolution simulations. Two additional pieces of work that have relevance for interpreting the downscaled climate data and for informing the next generation of climate model evaluation were completed during this dissertation. These projects have not yet been incorporated into a manuscript and are included in Appendix 1 and 2.

The project in the first Appendix examines the large-scale drivers of drought in key watersheds across the western US, investigating how well global climate models capture the teleconnections between Pacific sea-surface temperature and drought conditions. By combining drought measurements from the downscaled climate models and large-scale patterns of variability from global climate models, I was able to investigate these linkages in more granular detail than most studies relying only on GCM data. Beyond identifying the systematic shortcomings in model representations of key teleconnections, the results from this study also indicate that the overall strength of Pacific teleconnections can partially explain the tendency of a particular model to produce widespread concurrent droughts that span multiple watersheds. These results help guide interpretation of the risk of severe concurrent drought events in the current downscaled model ensemble and will help inform future generations of climate model evaluation.

The second appendix also turns an eye towards the next generation of climate model evaluations, by retrospectively examining the relationship between our climate model evaluation and the resulting biases in the eventual downscaled model ensemble. During the downscaling process, a strong tendency for the regional climate model to amplify precipitation was identified, necessitating an intermediate bias correction step (Rahimi et al., 2024b). In response, I conducted a retrospective evaluation to determine if our model evaluation process could be predictive of the GCMs that produced the largest precipitation biases when downscaled either with or without a bias correction step. The results identified a subset of the metrics that are helpful in predicting downscaled precipitation bias, indicating a potential set of priorities for future model evaluations seeking to reduce bias.

Future prospects and synthesis

Although the two domains of research in this dissertation may seem distinct from each other, the opportunity to work on both has built my appreciation for the important crosslinks between them. As regional climate interventions receive more research, funding, and public attention, the lines between climate adaptation and intervention become less clear. In the same way that cloud seeding for precipitation enhancement is used in several US states attempting to combat drought risks, other types of larger scale climate interventions may soon be in the adaptation planning arsenal of state agencies or even private climate tech companies. In this scenario there is a strong need to understand what regional climate hazards can effectively offset and what the side effects may be at a variety of scales.

The future of modeling larger scale climate intervention scenarios also has a strong need for the same type of regionally informed climate model evaluation that is demonstrated in this dissertation. In our study of global teleconnections from regional climate interventions, we acknowledge a major limitation of only considering a single GCM, which was chosen due to compute access rather than a rigorous evaluation. While an ensemble of GCMs would provide a more robust result, this approach is impractical for most studies of individual proposed interventions. Over the next several years as we are likely to see a wider range of proposals for regional climate interventions targeting unique geographies, each with their own potential pathways to trigger teleconnections or influence global circulation. The ideas from our evaluation could be developed into systems for knowing how to select the best GCMs for investigating the risks of each intervention.

Finally, the future of both domains will be significantly influenced by emerging developments in climate modeling methodology. Variable resolution models and machine learning approaches offer new opportunities for connecting regional and global scales, but also create new challenges for model evaluation and validation. Particularly for AI-based methods that do not explicitly simulate physical processes, robust evaluation frameworks that incorporate process-based understanding of atmospheric dynamics will become increasingly vital. The model evaluation framework developed in this dissertation advances existing methods by emphasizing the importance of accurately representing underlying physical processes. As climate models continue to evolve, maintaining this focus on physical processes while adapting evaluation methods to new modeling approaches will be essential for both understanding regional climate impacts and assessing potential interventions.

Bibliography for Chapter 1

Abdelmoaty, H. M., Papalexiou, S. M., Rajulapati, C. R., and AghaKouchak, A.: Biases Beyond the Mean in CMIP6 Extreme Precipitation: A Global Investigation, Earths Future, 9, e2021EF002196, https://doi.org/10.1029/2021EF002196, 2021.

Cannon, A. J.: Reductions in daily continental-scale atmospheric circulation biases between generations of global climate models: CMIP5 to CMIP6, Environ. Res. Lett., 15, 064006, https://doi.org/10.1088/1748-9326/ab7e4f, 2020.

Goldenson, N., Leung, L. R., Mearns, L. O., Pierce, D. W., Reed, K. A., Simpson, I. R., Ullrich, P., Krantz, W., Hall, A., Jones, A., and Rahimi, S.: Use-Inspired, Process-Oriented GCM Selection: Prioritizing Models for Regional Dynamical Downscaling, Bull. Am. Meteorol. Soc., 104, E1619–E1629, https://doi.org/10.1175/BAMS-D-23-0100.1, 2023.

IPCC: Summary for Policymakers, in: Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, NY, USA, 3–33, 2022.

Jones, A. and Haywood, J. M.: Sea-spray geoengineering in the HadGEM2-ES earth-system model: radiative impact and climate response, Atmospheric Chem. Phys., 12, 10887–10898, https://doi.org/10.5194/acp-12-10887-2012, 2012.

Kim, Y.-H., Min, S.-K., Zhang, X., Sillmann, J., and Sandstad, M.: Evaluation of the CMIP6 multi-model ensemble for climate extreme indices, Weather Clim. Extrem., 29, 100269, https://doi.org/10.1016/j.wace.2020.100269, 2020.

Pierce, D. W., Cayan, D. R., Goodrich, J., Das, T., and Munévar, A.: Evaluating Global Climate Models for Hydrological Studies of the Upper Colorado River Basin, JAWRA J. Am. Water Resour. Assoc., 58, 709–734, https://doi.org/10.1111/1752-1688.12974, 2022.

Priestley, M. D. K., Ackerley, D., Catto, J. L., and Hodges, K. I.: Drivers of biases in the CMIP6 extratropical storm tracks. Part 1: Northern Hemisphere, J. Clim., 1, 1–37, https://doi.org/10.1175/JCLI-D-20-0976.1, 2022.

Rahimi, S., Huang, L., Norris, J., Hall, A., Goldenson, N., Krantz, W., Bass, B., Thackeray, C., Lin, H., Chen, D., Dennis, E., Collins, E., Lebo, Z. J., Slinskey, E., Graves, S., Biyani, S., Wang, B., Cropper, S., and the UCLA Center for Climate Science Team: An overview of the Western United States Dynamically Downscaled Dataset (WUS-D3), Geosci. Model Dev., 17, 2265–2286, https://doi.org/10.5194/gmd-17-2265-2024, 2024a.

Rahimi, S., Huang, L., Norris, J., Hall, A., Goldenson, N., Risser, M., Feldman, D. R., Lebo, Z. J., Dennis, E., and Thackeray, C.: Understanding the Cascade: Removing GCM Biases Improves Dynamically Downscaled Climate Projections, Geophys. Res. Lett., 51, e2023GL106264, https://doi.org/10.1029/2023GL106264, 2024b.

Simpson, I. R., Bacmeister, J., Neale, R. B., Hannay, C., Gettelman, A., Garcia, R. R., Lauritzen, P. H., Marsh, D. R., Mills, M. J., Medeiros, B., and Richter, J. H.: An Evaluation of the Large-Scale Atmospheric Circulation and Its Variability in CESM2 and Other CMIP Models, J. Geophys. Res. Atmospheres, 125, e2020JD032835, https://doi.org/10.1029/2020JD032835, 2020.

Tollefson, J.: Can artificially altered clouds save the Great Barrier Reef?, Nature, 596, 476–478, https://doi.org/10.1038/d41586-021-02290-3, 2021.

Chapter 2: A framework for minimizing remote effects of regional climate interventions: Cooling the Great Barrier Reef without teleconnections

Abstract

Climate interventions like Marine Cloud Brightening (MCB) have gained attention for their potential to protect vulnerable marine ecosystems from the worst impacts of climate change. Early modeling studies raised concerns about potential harmful global side effects stemming from regional# interventions. Here we propose a modeling framework to evaluate these risks based on using maximal deployment scenarios in a global climate model to identify potential pathways of concern, combined with more realistic large intervention levels. We demonstrate this framework by modeling a cooling intervention over the Great Barrier Reef using the Community Earth System Model (CESM2). We identify potential impacts on tropical convection that could produce remote impacts, and show that limiting intervention duration to deployment in the key season largely eliminates these risks. Overall we illustrate that the local ecological goals can be achieved at a level of cooling well below what poses a risk of significant remote effects.

2.1 Introduction

In recent years, the widespread and disruptive impacts of global climate change have driven increased interest in climate intervention techniques. A variety of methods have been proposed for partially offsetting the warming effects of climate change, including stratospheric aerosol injection, marine cloud brightening (MCB), and surface albedo enhancement. Each of these methods shares the goal of reflecting a fraction of incoming solar radiation to cool the surface of the earth. While stratospheric aerosol injection is aimed primarily at a global scale, marine cloud brightening or surface albedo enhancement can also be used at a regional scale to counter the impacts on vulnerable natural systems (Latham et al., 2014). Targeted cooling interventions have been proposed for protecting tropical reefs, preserving arctic sea ice, and reducing tropical storm intensity (Field et al., 2018; Gertler et al., 2020; Latham et al., 2014; Tollefson, 2021). Recent research on protecting the Great Barrier Reef in Australia has highlighted the promise of MCB for shielding reefs from damaging marine heat waves through regional modeling and early field testing (Bay et al., 2019; Harrison et al., 2019; Hernandez-Jaramillo et al., 2023; Readfearn, 2020; MacMartin et al., 2023).

Early model studies of marine cloud brightening as a global cooling strategy found a strongly heterogeneous response of surface temperature and precipitation (Jones et al., 2009; Jones & Haywood, 2012; Latham et al., 2008; Rasch et al., 2009). These studies typically simulate increased cloud albedo over large patches of ocean where marine stratocumulus clouds are the most persistent, and found that each patch of cooling produced a unique global pattern of impacts on precipitation and temperature through teleconnections (Jones & Haywood, 2012). These strong heterogeneous impacts raise concerns about the viability of cloud brightening as a global cooling intervention, and raise the question of whether regional cooling could be deployed without triggering harmful impacts outside the intended area of intervention. The ability of regional cooling to trigger remote teleconnections has also been studied as a potential tool, with two studies exploring the possibility of intentionally perturbing teleconnections to reduce drought in the Sahel (Ricke et al., 2021) or in the western United States (Wan et al., 2024). MacMartin et al. used a GCM to assess the potential for transboundary effects of regional cooling over the Great Barrier Reef and the Gulf of Mexico, but only using a single scenario with a relatively low magnitude of cooling.

Collectively these studies highlight the importance of understanding when a regional cooling intervention carries a credible risk of adverse effects outside the targeted region of intervention. Research frameworks for MCB have repeatedly highlighted the importance of characterizing the large-scale circulation response to cooling interventions, and have proposed an "off-ramp" to halt further work on the technology if the risks of global disruptions are deemed too high (Diamond et al., 2022; Feingold et al., 2024).

In this study we demonstrate a modeling approach for studying the risk of teleconnections and remote climate impacts from regional climate interventions. This approach seeks to address a gap between studies of global MCB deployment that have shown widespread teleconnected effects and regionally focused simulations not well suited to detecting teleconnections. Drawing from our findings studying the case of cooling the Great Barrier Reef, we propose a general framework for identifying potential pathways for remote impacts and minimizing the risks of remote impacts.

The primary goals of our approach are to:

- Identify potential pathways of concern by simulating a high intensity intervention.

- Demonstrate how modifications to deployment scenarios can avoid triggering potential teleconnection pathways.
- Illustrate the margin of safety between the level of intervention required to achieve local goals and the level that poses a risk of harmful side effects.

In section 2.2 we outline the steps of our general framework, in section 2.3 we show the results of a modeling study implementing this framework over the Great Barrier Reef, and in section 2.4 we discuss the conclusions and takeaways from this study.

2.2 Framework for identifying potential pathways of concern from regional interventions

Although global climate models (GCMs) are an essential tool identifying potential remote impacts or teleconnections from a regional intervention, the limited spatial resolution and physical parameterizations of these models present challenges for designing appropriate scenarios. Additionally, finding teleconnection pathways and distinguishing their effects from natural variability presents a significant statistical challenge. The framework we describe here seeks to address these difficulties, and was developed through efforts to study the potential for remote impacts from cooling the Great Barrier Reef. Although we focus on MCB, this framework is applicable more generally to the process of using a GCM to identify and minimize potential risks of remote impacts from regional climate interventions.

1. Design an appropriate prototype of intervention

Examining the potential for unintended side effects of a regional intervention through teleconnections or shifts to large-scale circulation requires the use of a global climate model

(GCM). In practice an intervention like MCB will depend on both physical interactions and operational decisions at a sub-grid scale, necessitating a simplified representation for a GCM simulation. For a study focused on potential teleconnections, producing the desired magnitude and seasonal timing of surface cooling are more important than explicitly modeling the physical processes that produce a cooling effect. However, realistic choices for the magnitude of regional cooling should be informed by more detailed regional or process-level simulations when available.

2. Define regional criteria for successful intervention

The goal of this modeling framework is to determine whether a regional intervention can achieve the intended local benefits without posing significant risks to surrounding areas. Thus, it is paramount to first define the local goals of the intervention in a meaningful and measurable way. Whether the goal is the protection of an ecosystem, the preservation of natural resources, or shielding humans from extreme conditions, the criteria for a successful intervention should be grounded in scientific studies of the relevant impact. Because global climate models typically lack the detail to explicitly model the relevant ecological or meteorological processes, additional references are needed to translate the desired outcome (e.g. avoiding coral bleaching) into an appropriate variable and threshold in the global climate model (e.g. reducing peak annual SST by 0.5 C).

3. Test upper limits of physical plausible deployment to identify potential pathways of concern

Given the intense scrutiny on climate interventions at any scale, we advocate a modeling approach designed to identify any potential pathways for teleconnections stemming from the region of intervention. Simulating a maximal deployment scenario at the limits of what is physically plausible produces a stronger signal of climate response relative to natural variability. A better understanding of any potential pathways helps set priorities for monitoring of early deployments, and can inform operational changes that can minimize remote impacts.

Considering a maximal deployment scenario provides a safety factor against the many layers of uncertainty surrounding both the effectiveness of the intervention, the representation of teleconnections in climate models, and operational decisions about deployment. As research progresses, simulations in other climate models and field testing will further reduce these uncertainties and feed back into model scenario design.

Simulating such exaggerated scenarios comes with a risk of amplifying the perceived risk of an intervention. Accompanying these results with simulations of a more realistic magnitude and clear communication about the results is important to avoid the risks of the results being misconstrued.

4. Evaluate potential pathways of concern in the context of modeled natural variability

To contextualize the magnitude of any remote impacts, natural variability of the control run should be compared to the observational record when possible. Many studies of climate interventions use pre-industrial conditions or simplified forcing scenarios which have the potential to reduce natural variability relative to real-world observations and thus amplify the apparent significance of observer anomalies. While this is helpful for our goals of identifying physical pathways of impacts, it needs to be acknowledged when characterizing their significance under real-world conditions. Additionally, thorough statistical methodology is needed to adjust for both the tendency of climate fields to be temporal and spatial auto-correlated.

5. Adjust intervention to minimize activation of teleconnection pathways

With an understanding of potential teleconnection pathways from the intervention region, the scenario of regional deployment can be modified to minimize activation of these mechanisms. The intensity, spatial extent, and timing of the intervention can be modified to characterize the extent that remote impacts can be minimized while still achieving the local goals established in step 2. To understand the risk of a particular intervention, it is helpful to understand how much overlap there is between the desired local impact and the perturbations that trigger teleconnection pathways.

6. Consider off-ramp if regional goals can not be met without significant remote impacts

If simulations demonstrate that the local benefits can not be achieved without significant remote impacts, this should trigger a potential "off-ramp" of further research into the intervention, as described in Diamond et al. (2022). Given the significant differences between GCMs in representing global circulation and teleconnections, we recommend replicating the findings across multiple studies before proceeding with confidence with a positive or negative finding.

2.3 Simulations of regional cooling over the great barrier reef

We apply the framework described above to the case of a cooling intervention over the Great Barrier Reef with the goal of protecting the sensitive ecosystem from coral bleaching. This proposed intervention has received the most research to date and illustrates a case where the local goals of the intervention are well understood, while the potential for unintended remote effects is still under-explored.

2.3.1 Model Setup

The simulations for this study were conducted using the Community Earth System Model (CESM2, Danabasoglu et al. 2020; Computational and Information Systems Laboratory, 2023)) running in a “slab ocean” configuration. In this mode a simplified ocean simulates the mixed layer of the ocean and its interactions with the atmosphere without simulating full ocean circulation. The exchange of heat between the mixed layer and the deeper ocean is approximated with a prescribed monthly Q-flux that is derived from a fully coupled run of the model. All simulations are carried out under preindustrial conditions, with a 50-year spin up period to ensure the model reaches a stable equilibrium before perturbations are applied.

Cooling interventions are implemented using an additional Q-flux term to specify a forcing at the surface of the ocean. The advantage of this approach is that it allows direct control of the location and amount of forcing applied, and can be used to represent cooling that is achieved through cloud brightening, surface albedo enhancement, deep ocean pumping, or any combination of intervention technologies. The limitations of this approach are that it does not include any dynamic adjustments or feedbacks that are specific to an albedo modification or cloud intervention. The simplified ocean also neglects the advection of cooled surface waters due to ocean circulation. This modeling configuration is not aimed at evaluating the efficacy of particular intervention technology at producing cooling, but whether a given level of regional cooling would produce significant impacts outside the region of intervention, in line with step 1 of our framework. Despite these limitations, the SST variability in the Coral Sea region in the slab-ocean simulation is quite similar to both a fully-coupled pre-industrial run of CESM2 and with ERA5 reanalysis (Figure S1).

For each level of forcing, three separate simulations are branched from the control run at five year intervals and run for 13 years. The first year of each branch is discarded as the system is adjusting, producing 36 years of data for each scenario. Using several shorter runs rather than a single long run allows us additional statistics to examine impacts that emerge in the first decade of deployment, when increased scrutiny will be placed on both the effectiveness and any unintended side effects of an intervention. Spacing the branches by five years allows more sampling across conditions of natural variability.

2.3.2 Intervention Scenarios

To represent an intervention designed to protect the Great Barrier Reef from damaging marine heat waves and elevated peak summer water temperatures, we prescribe a negative forcing in the full Coral Sea Marine Protected Region (Figure 1a). Consistent with the goals of our framework, this represents a much larger intervention area than what would be required to cool the reef itself, and corresponds roughly with the largest scenario simulated in regional modeling studies (Bay et al., 2019). The forcing is tapered at the edges of the intervention domain to avoid overly steep temperature gradients in the model. In Figure 2-1 and all subsequent maps, stippling is used to indicate regions with significant difference from the control experiment using a field significance calculation, correcting for the false discovery rate with a threshold of $\alpha_{FDR} = 0.1$ (Wilks, 2016). This correction is important for accurate measurements of significance in spatially autocorrelated data like climate variables (as per step 4 described above). .

Modeling studies that have explicitly simulated cloud-albedo-change from increased cloud condensation nuclei have produced local forcings of a wide range of 10-40 W m^{-2} (Ahlm et al., 2017; Stjern et al., 2018; Wan et al., 2024) Choosing the higher end of this range, we prescribe cooling at 40 W m^{-2} during three months (DJF) leading up to the hottest part of the year. To test the sensitivity of global circulation to temperature perturbation in this region and indicate risks that might occur from cooling in other seasons, we also simulate year-round coolings at 20, 40, and 60 W m^{-2}. The year-round forcing at 60 W m^{-2} nears the upper limit of physical plausibility, designed to draw out weak teleconnection signals that may not otherwise be visible, as described in step 3 of our framework.

In evaluating the regional temperature response to this forcings, we compare to a typical threshold characterizing coral bleaching events is a water temperature one degree celsius higher than the maximum monthly temperature, and ecological studies of interventions to protect coral have identified cooling between 0.5 and 1 degree as threshold for reducing harm (Baird et al., 2020; Harrison et al., 2019; Logan et al., 2012). In line with step 2 of our framework, we consider this a rough target for a successful regional intervention.

The seasonal forcing at 40 W m^{-2} reduces the peak SST in the intervention region by 0.5 °C, but matches the temperature of the control run during the winter months (Figure 1c). The year-round cooling reduces the annual peak temperature by 0.4, 0.8, and 1.3 °C at 20, 40, and 60 W m^{-2} respectively, with a similar level of cooling throughout the year (Figure 1b). Although the year-round intervention produces more cooling during the hottest months (JFM) for the same forcing intensity, we consider the seasonal deployment to meet the essential goal of lowering maximum temperature by at least 0.5 °C. The comparison between the year-round and seasonal

40 W m^{-2} scenario allows us to examine how simple operational changes can dramatically change the risks of remote effects without sacrificing the core goal of the intervention.

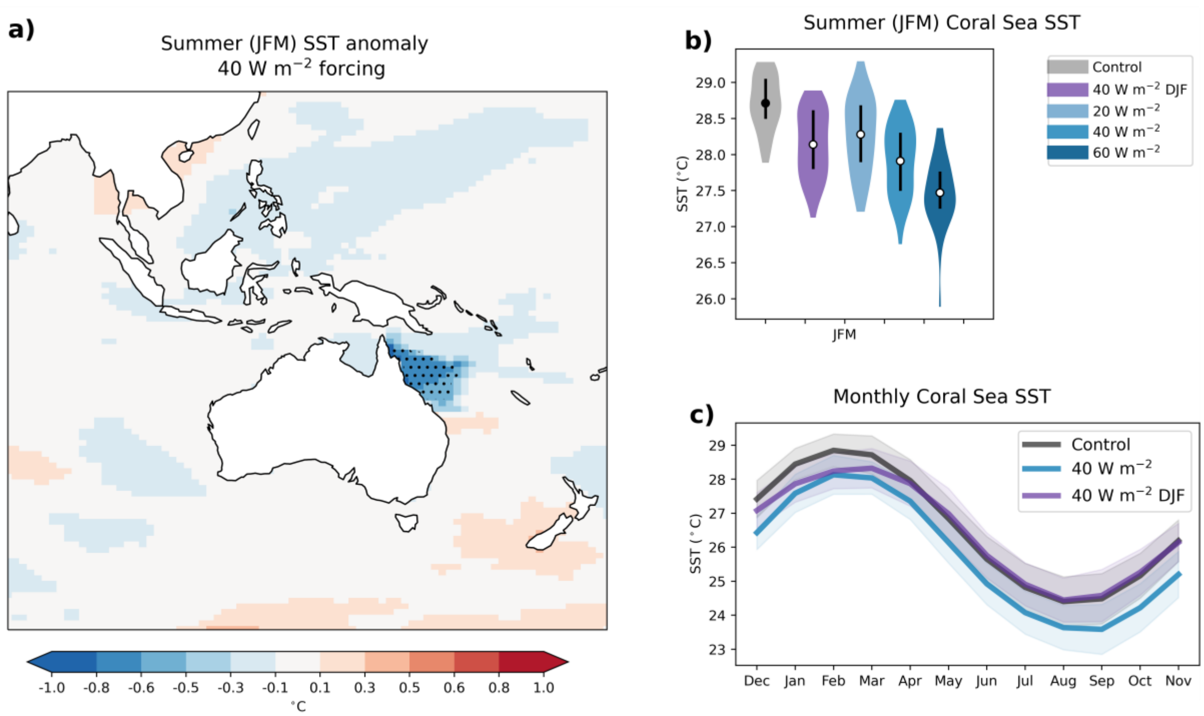

Figure 2-1. a) SST anomalies in the three hottest months (JFM) for the year-round 40 W m^{-2} cooling scenario relative to the control scenario, with the intervention domain outlined in magenta. Regions with significant anomalies relative to the control are stippled. b) Annual average SST within the region of intervention for each scenario. c) The seasonal cycle of SST within the region of intervention for the year-round and seasonal intervention.

2.3.3 Global impacts of regional intervention

Overall, we find very few impacts outside the intervention domain that appear significant relative to interannual variability, even for the maximal deployment scenarios simulated. The 20 and 40 W m^{-2} intervention scenarios that represent more realistic levels of forcing do not produce any statistically significant impacts outside of the intervention region for precipitation, sea surface temperature, or surface air temperature (Figures S2-S4). The 60 W m^{-2} simulation produces some significantly reduced precipitation in the areas directly adjacent to the intervention domain (Figure S4), and some remote impacts on sea surface temperature (Figure S2).

To thoroughly investigate any potential risks and understand the physical mechanisms driving them, we closely examine the strongest signals globally that could plausibly be triggered by the cooling intervention. Although the remote signals observed here are weak overall, there are large uncertainties around the strength of teleconnection processes in GCMs. Identifying the potential pathways for teleconnections stemming from a particular intervention can help inform future modeling studies as well as the operational design and monitoring needs of an eventual deployment.

2.4 Potential pathways of concern

2.4.1 Suppression of nearby deep tropical convection

The Coral Sea region exhibits deep tropical convection during the summer months, which is suppressed by the cooling intervention. The vertical velocity over the intervention region shows a significant reduction in upward velocity that extends from near the surface to the upper troposphere (Figure 2a), with associated anomalies in geopotential height (Fig. Sx). This type of deep baroclinic anomaly (Gill, 1980; Sardeshmukh & Hoskins, 1988) has the potential to propagate globally through Rossby and Kelvin waves, and represents a plausible pathway of concern that is worth flagging for careful examination in this and any future studies.

Although the strongest vertical velocity anomaly occurs during the same season as the maximum SST we are trying to reduce, the more targeted seasonal intervention shows a greatly reduced anomaly that is below the threshold of significance (Figure 2b). While some of this reduction can be attributed to the slightly lower peak cooling in the seasonal deployment, the strongest deep convection signal spans the months of November to January, where overall cooling is much lower in the seasonal DJF intervention (Figure S5). This illustrates a simple but effective modification of the deployment scenario that appears to significantly reduce the risk of a potential teleconnection pathway. This example also illustrates the importance of carefully understanding the temporal and spatial overlap between the intended effects of the intervention and the triggers of potential teleconnection pathways. With more detailed analysis and adjustments to the timing and location of cooling, the impact on deep convection and any associated risk of teleconnection could likely be reduced even further.

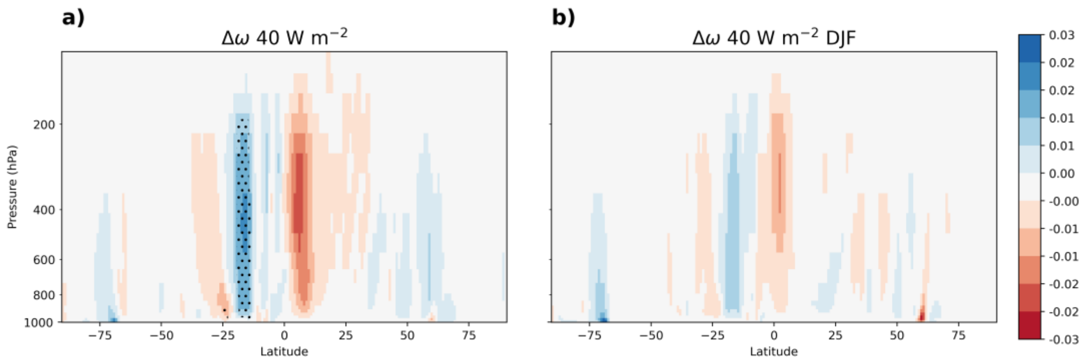

Figure 2-2. Vertical pressure-velocity anomaly (positive indicates anomalous descent) in the annual (a) and seasonal (b) 40 W m^{-2} intervention scenarios relative to the control simulations during the summer months (JFM), zonally averaged around the intervention domain (145°W - 151°W). Stippling indicates a significant change relative to the control.

2.4.2 Precipitation in Southeast Asia

Stemming from the suppression of convection, the immediate area of intervention shows a significant reduction in summer precipitation for all of the simulations. In the year-round intervention scenarios, there is a separate off-season drying effect that is apparent over much of Southeast Asia during the winter months (Figures 3a, 3b). There is also a slight increase in precipitation seen over the same region during JFM, but this is not present in all of the scenarios and does not pass a test of significance. While no individual grid cells in this area pass the field significance test, the seasonal precipitation over land aggregated within the domain indicated in Figure 2-3a and 2-3b experiences a significant shift in the 40 and 60 W m^{-2} intervention scenarios (indicated by a white point in Figure 2-3c). Notably, the seasonal intervention scenario avoids producing the off-season reduction in precipitation, as shown in the violin plots of Figure 2-3c. The probability density functions (PDFs) of seasonal precipitation in the violin plots help to indicate the magnitude of the interannual variability compared to the signal (note that many details of the differences in PDF would be due to sampling). Further analysis would be required to determine the specific physical mechanism linking the cooling over the Coral Sea to Southeast Asian precipitation, but given that the effect can almost entirely be avoided by focusing the forcing on the target season, the risk from this pathway appears low.

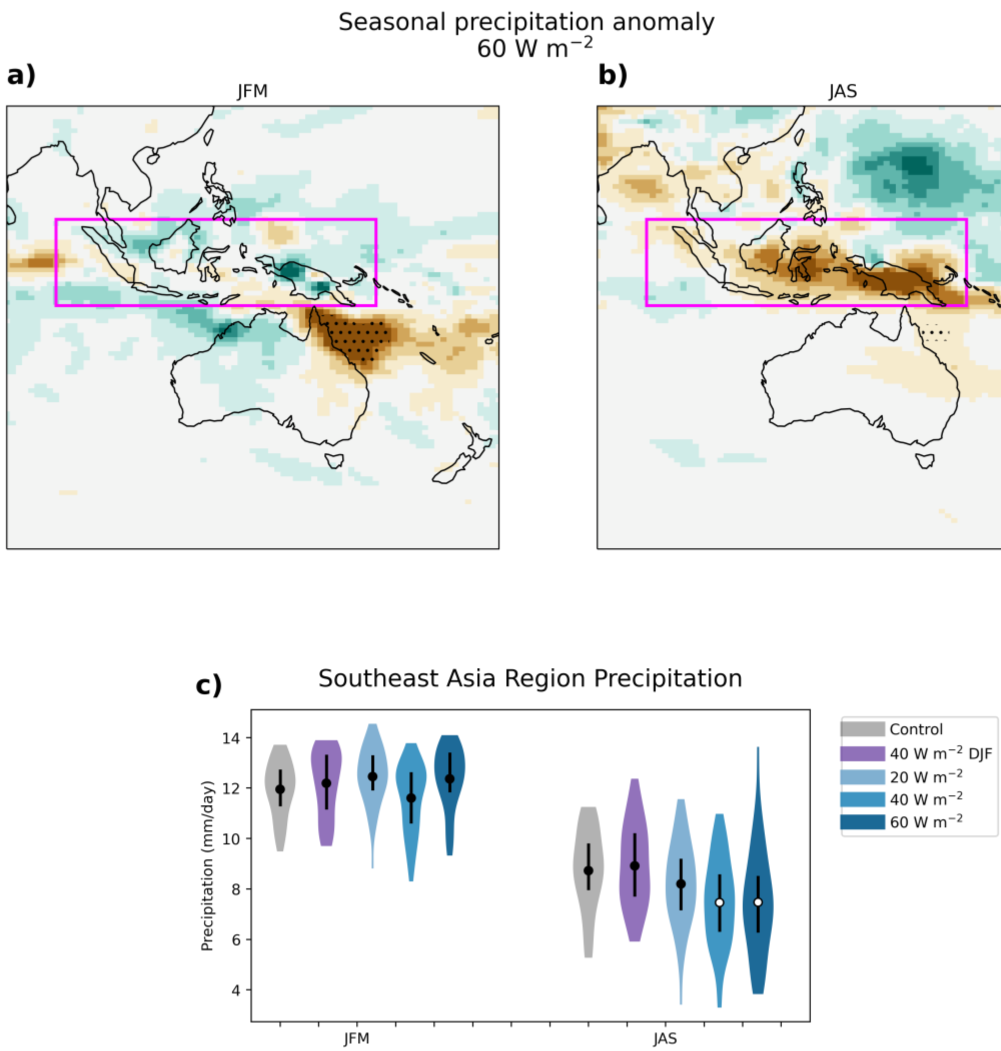

Figure 2-3. Seasonal precipitation anomaly in the year-round 60 W m^{-2} forcing scenario for (a) JFM and (b) JAS. (c) Violin plots of the land-masked seasonal precipitation within the highlighted domain for the control run and all four intervention scenarios. For each violin plot, the mean is plotted along with error-bars indicating the 25th to 75th percentile range. For distributions where the mean is shifted significantly relative to control simulation, the mean is plotted in white.

2.4.3 Remote SST anomalies

In the most extreme intervention scenario in this study (60 W m^{-2} year-round forcing), some remote anomalies in sea surface temperature appear significant relative to the control scenario (Figure 4a). Given the convection anomalies discussed above, it initially appears plausible that these are the result of global teleconnections stemming from the intervention. The two strongest remote signals are a warming in the southern ocean and a cooling over the northern Pacific. As before, the anomaly over the strongest area of impact is spatially averaged and shown in the violin plots in Figure 4b. Unlike precipitation, SST signals can be persistent year-to-year, which reduces the effective number of independent measurements present in our relatively short simulations. This increases the chance of a spurious fluctuation due to natural variability appearing significant. To account for this, as emphasized in step 4 of our framework, the significance calculation for the violin plots in Figure 4b and 4c incorporates an effective degrees of freedom adjustment based on the temporal autocorrelation within each region Wilks (2011). The Southern Ocean region and northern Pacific region show a 1-year lag autocorrelation of 0.39 and 0.57 respectively. After this adjustment, the 40 and 60 W m^{-2} interventions pass our significance test in the Southern Ocean, while the 20 and 40 W m^{-2} interventions pass in the northern Pacific. The absence of the cooling signal in either of the 40 W m^{-2} simulations for the northern Pacific casts some additional doubt on whether this signal is a real causal impact or a spurious signal of natural variability that happened to align in several of these simulations. Given the inconsistency of these signals, along with the fact that no clear causal chain is present in global geopotential height or streamfunction (Figures S6, S7), they represent areas flagged for attention in future studies, but are not identified here as a clear impact.

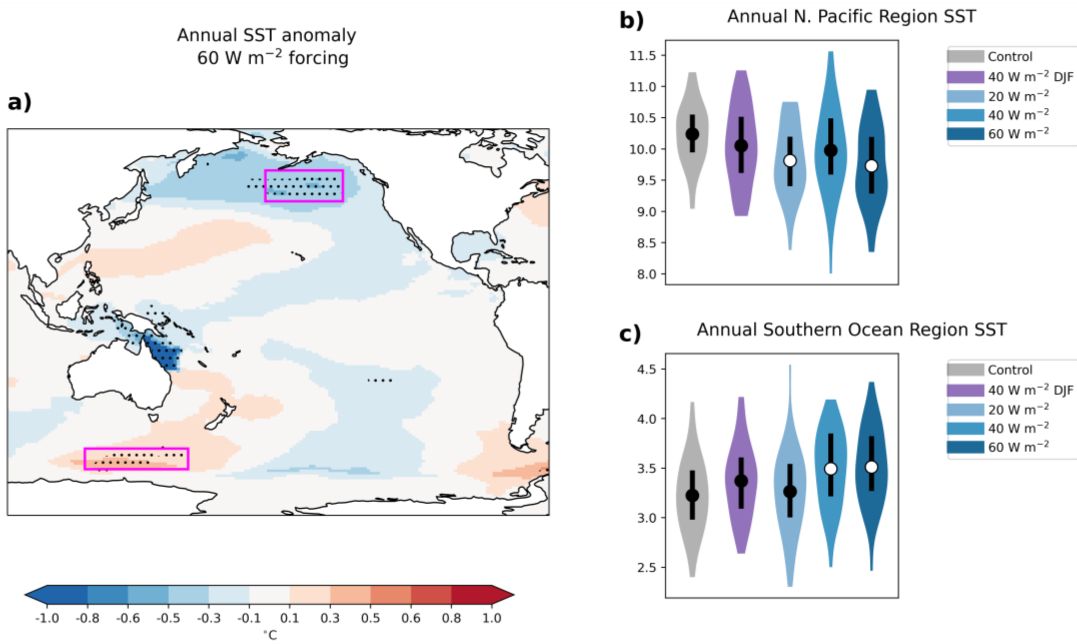

Figure 2-4. Global SST anomaly in the 60 W m^{-2} year-round forcing relative to control (a) and violin plots of the annual anomaly aggregated over the highlighted domains for all four intervention scenarios. As in Figures 1 and 2, areas of significant impact are stippled, and distributions with a significant shift in the mean are indicated by a white point.

2.4 Discussion and Conclusions

This study presents a modeling approach for studying the risks of teleconnections and unintended remote impacts from regional cooling interventions. In the case of the Great Barrier Reef, we find that an intervention designed to reduce heat extremes in the Coral Sea by 0.5-1 °C carries a low risk of triggering remote impacts through global teleconnections.

However, we define a potential pathway of concern as a teleconnection process that could be relevant for further evaluation. For Great Barrier Reef cooling we identify at least one such pathway: strong cooling in the region produces a suppression of deep convection which in turn produces a baroclinic anomaly with the potential to induce remote impacts. Evidence of this potential is illustrated by reduced seasonal precipitation over Southeast Asia and remote SST anomalies over the Southern Ocean and the northern Pacific Ocean.

For all these anomalies, limiting the forcing scenario to focus on cooling only the warmest months significantly reduced their intensity, illustrating step 5 of our framework. Given the high operational cost and complexity of deploying over such a large area, it is likely that an operational deployment will seek to cool over the smallest area and time necessary to achieve an ecological benefit. Given that our simulations only showed remote impacts above the threshold of significance for very large intervention areas and forcing at the upper end of what is feasible, we conclude these potential pathways are of very low practical risk. In the context of step 6 of our framework, we do not find risks that merit an off-ramp of further research, but encourage future studies to consider the potential teleconnection pathway we have identified.

This finding highlights the importance of characterizing potential pathways of concern by how much they overlap in time and space with the goal of the regional intervention. In this case the strongest potential pathways were triggered outside the summer season when the cooling intervention is most important, allowing them to be mitigated without reducing the effectiveness of the intervention. The potential pathways of concern will be unique to any climate intervention, and it may not always be possible to reduce them while maintaining the efficacy of the intervention. For any other cooling intervention in a tropical region, we highlight the potential of suppressed deep convection as a pathway for teleconnections and emphasize the importance of including similar analysis in future studies.

The potential remote impacts that we identify in this study only emerge as significant relative to natural variability with several decades of statistics, even with a spatial extent and forcing level much larger than necessary to meet the ecological goals. However, given the scrutiny that any regional climate intervention is likely to receive, there is a potential risk of falsely attributing regional anomalies from natural variability in the early years following the intervention. Our simulations produce many large regional anomalies in individual years, and the statistics enabled by several independent simulations are required to show these are due to natural variability rather than a forced response. Any real-world intervention will not have the benefits of these statistics, but using the modeling approach demonstrated here to understand the physical mechanisms and pathways that do pose a plausible risk can help guide monitoring and governance of interventions.

The framework we describe here to identify potential pathways of concern from regional interventions relies on a strategy of simulating larger-than-plausible intervention scenarios. These exaggerated perturbations to the climate system allow relatively weak signals to appear more visible, and aid in understanding the physical mechanisms underlying any potential teleconnections. This also serves as a margin of safety to account for modeling uncertainties when estimating thresholds for avoiding significant remote impacts. We note the risk that these maximal scenarios could be misinterpreted as impacts that are likely to occur—and emphasize that these are bounding cases. Given the uncertainties associated with modeling teleconnections, replicating findings across several GCMs is important to raise confidence in the results. Our hope in outlining this modeling framework along with our findings is to encourage other modeling studies to similarly characterize the risks from teleconnections in the case of the Great Barrier Reef and for any other proposed regional climate interventions.

Acknowledgments

This work was supported in part by National Science Foundation Grant AGS-2414576, AGS-1936810 and AGS-2225956 (FA). We would like to acknowledge high-performance computing support from the Derecho system (doi:10.5065/qx9a-pg09) provided by the NSF National Center for Atmospheric Research (NCAR), sponsored by the National Science Foundation. The authors would like to thank Ted Parson for his insights guiding the original design of this study.

Additional Figures for Chapter 2

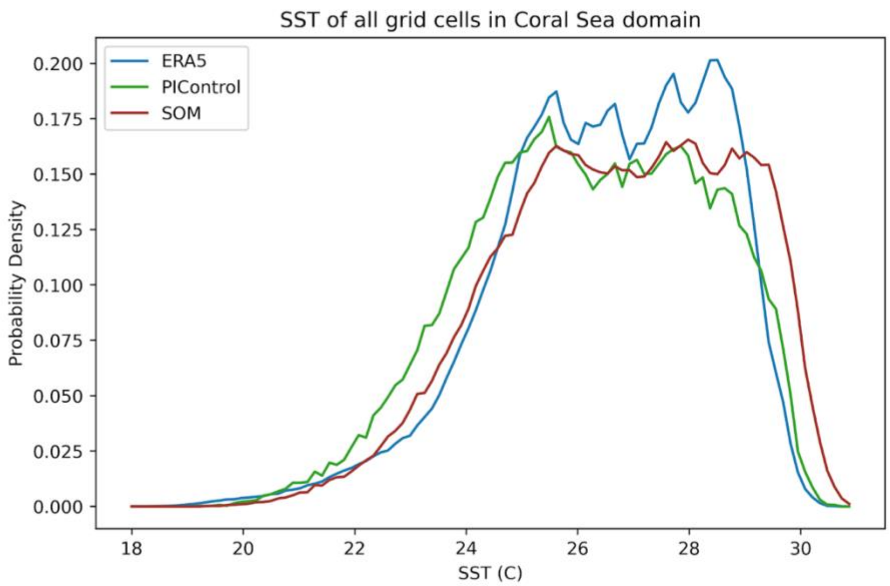

Figure 2-S1 - Probability density function of daily mean SST in all grid cells in the Coral Sea domain for 50 years of data from ERA5 reanalysis, CESM2 pre-industrial control, and CESM2 Slab Ocean Model (SOM) control run.

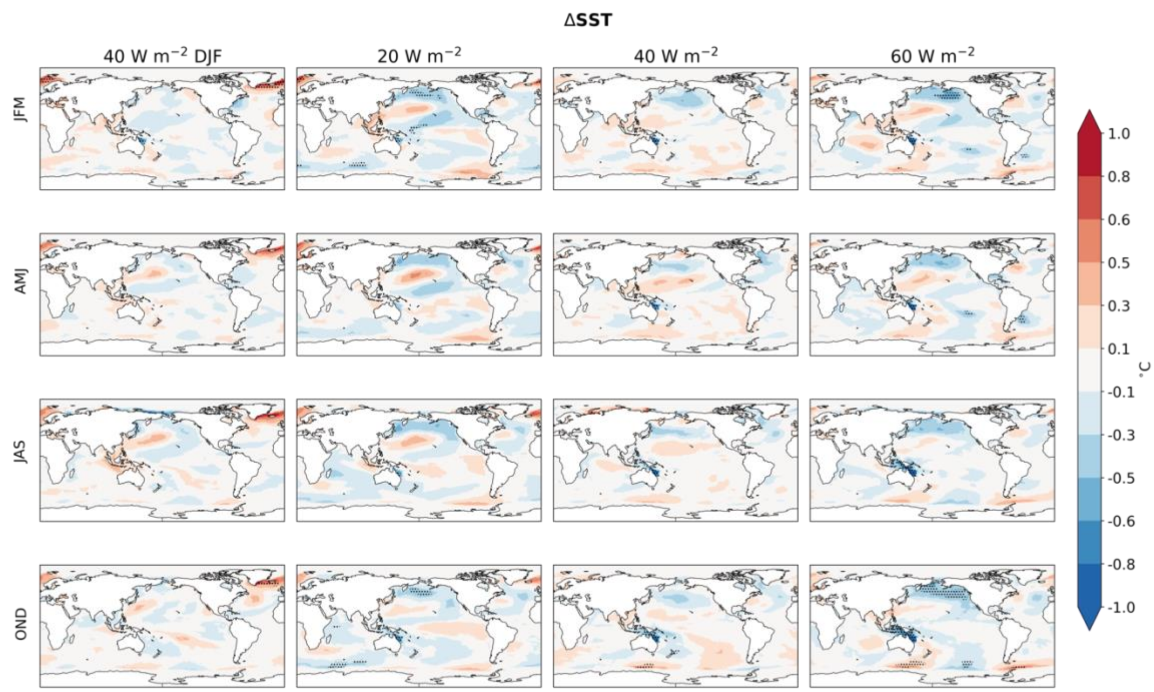

Figure 2-S2 - Seasonal sea surface temperature anomaly relative to control for seasonal 40 W m^{-2} and annual 20, 40, and 60 W m^{-2} intervention.

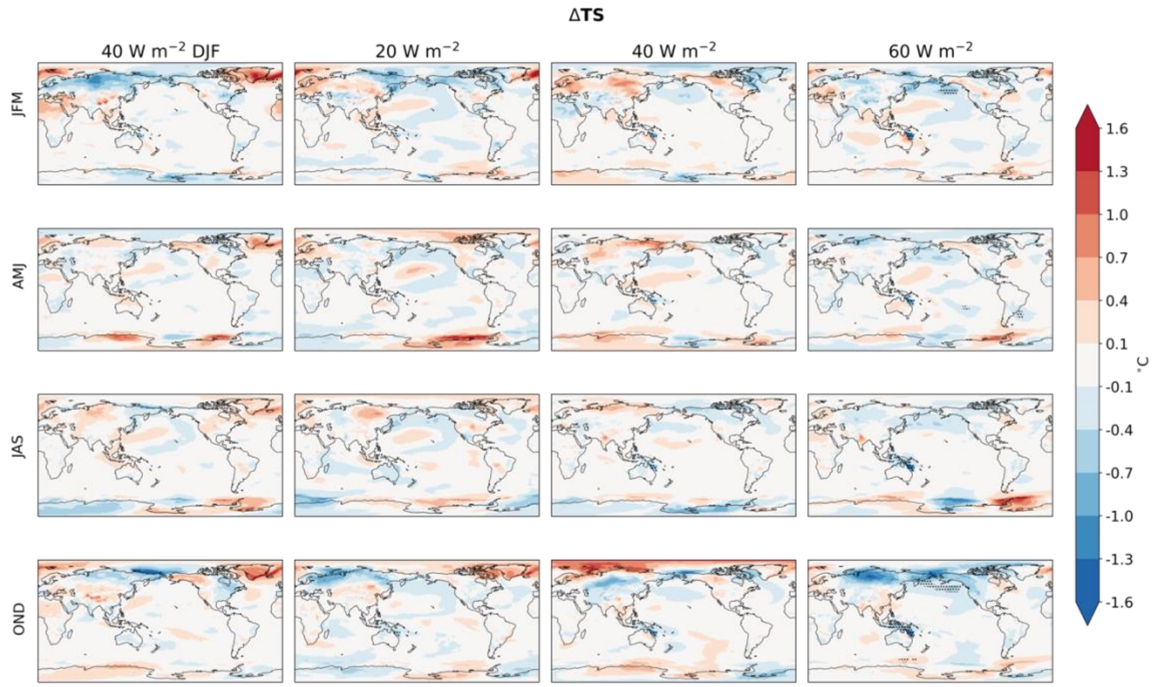

Figure 2-S3 - Seasonal surface air temperature anomaly, normalized by standard deviation per grid-cell, relative to control for seasonal 40 W m^{-2} and annual 20, 40, and 60 W m^{-2} intervention

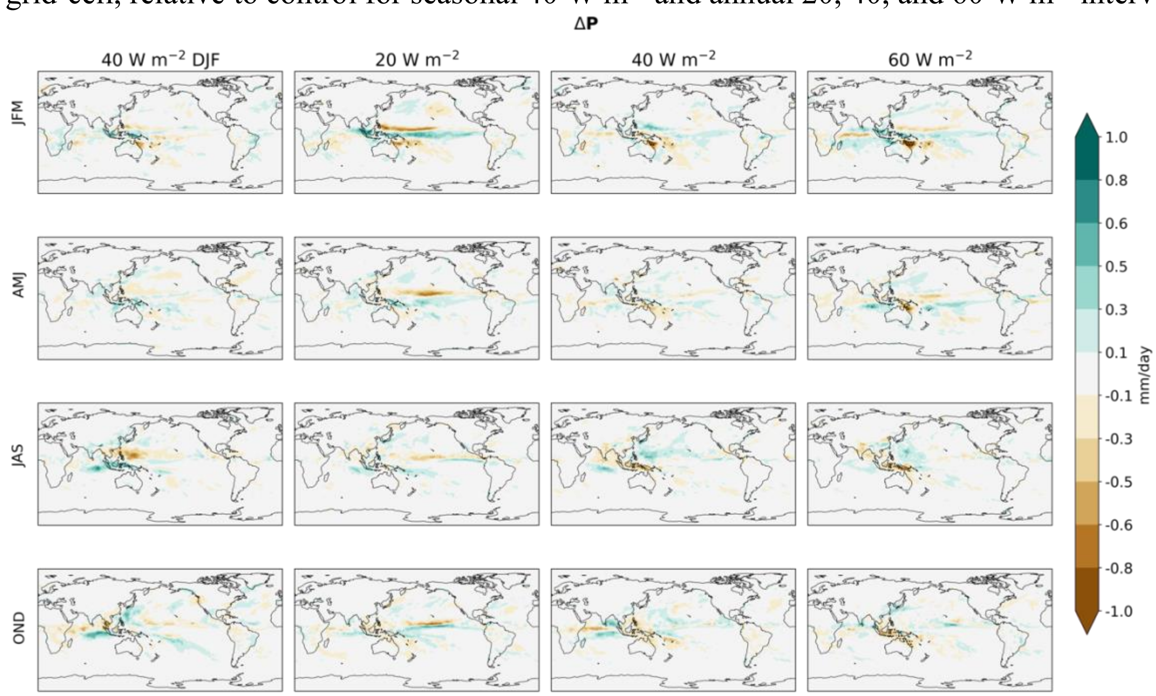

Figure 2-S4 - Seasonal precipitation anomaly relative to control for seasonal 40 W m^{-2} and annual 20, 40, and 60 W m^{-2} intervention.

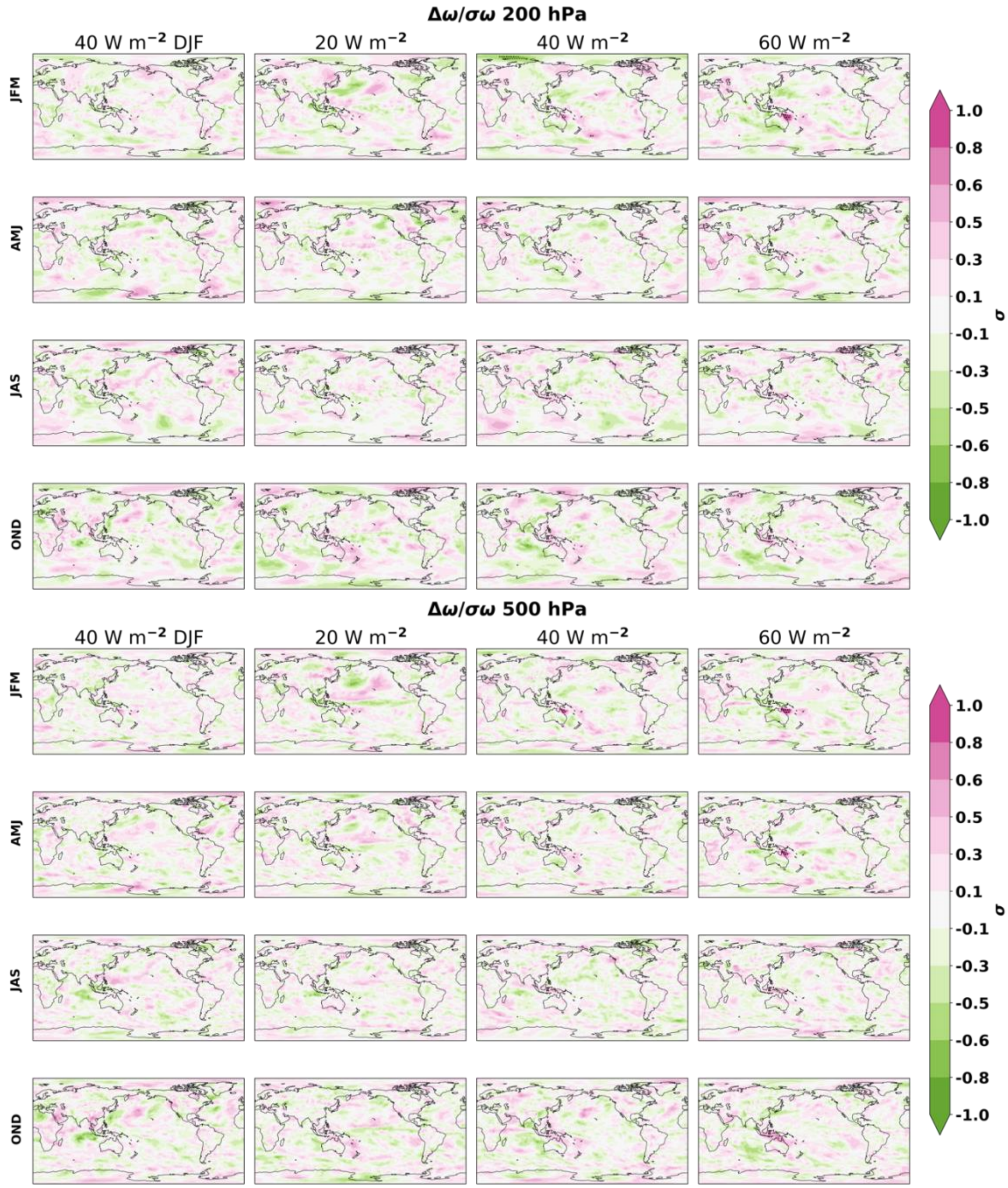

Figure 2-S5 - Seasonal vertical velocity anomaly, normalized by standard deviation per grid-cell, relative to control for seasonal 40 W m-2 and annual 20, 40, and 60 W m-2 intervention.

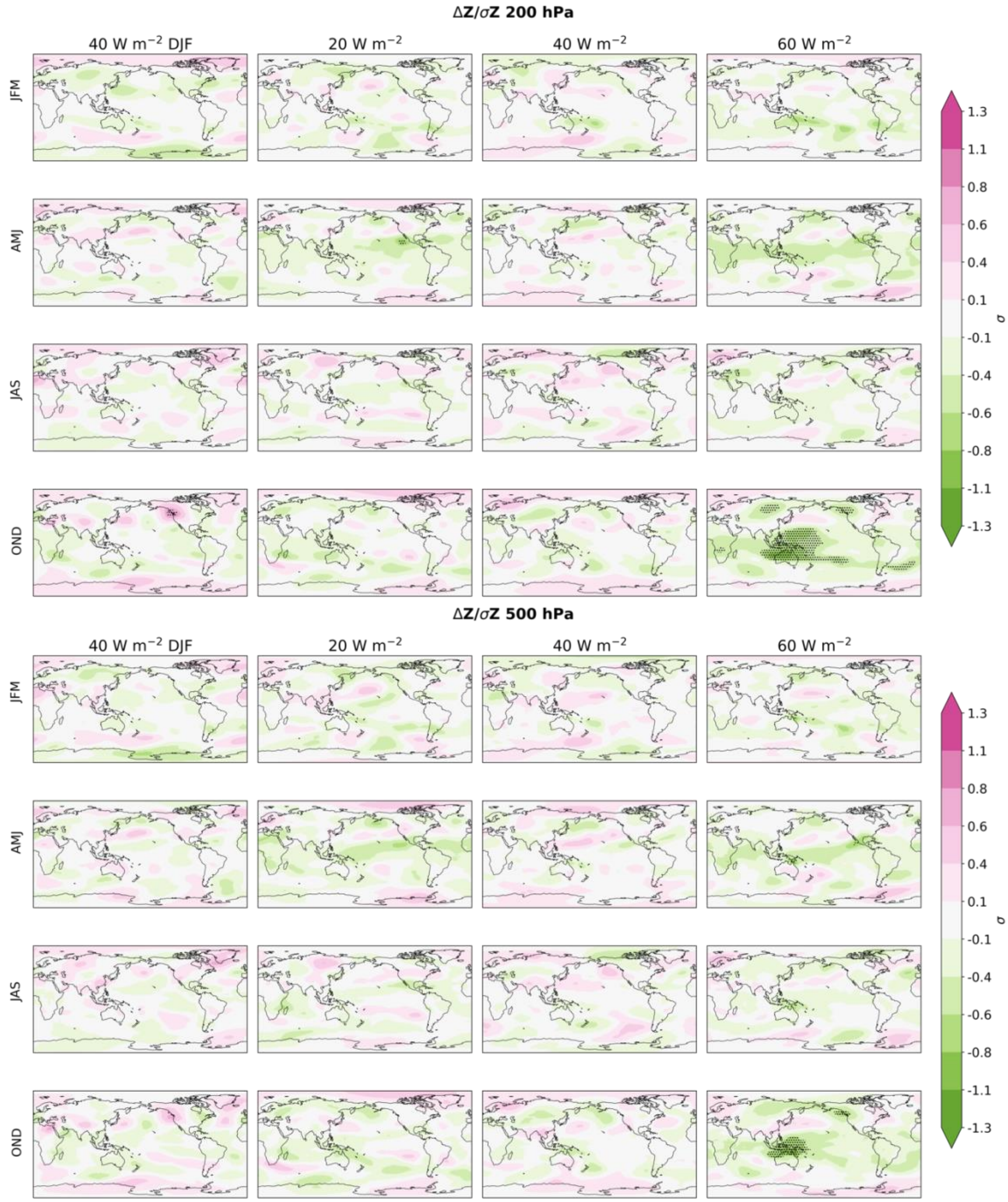

Figure 2-S6 - Seasonal geopotential height anomaly, normalized by standard deviation per grid-cell, relative to control for seasonal 40 W m^{-2} and annual 20, 40, and 60 W m^{-2} intervention.

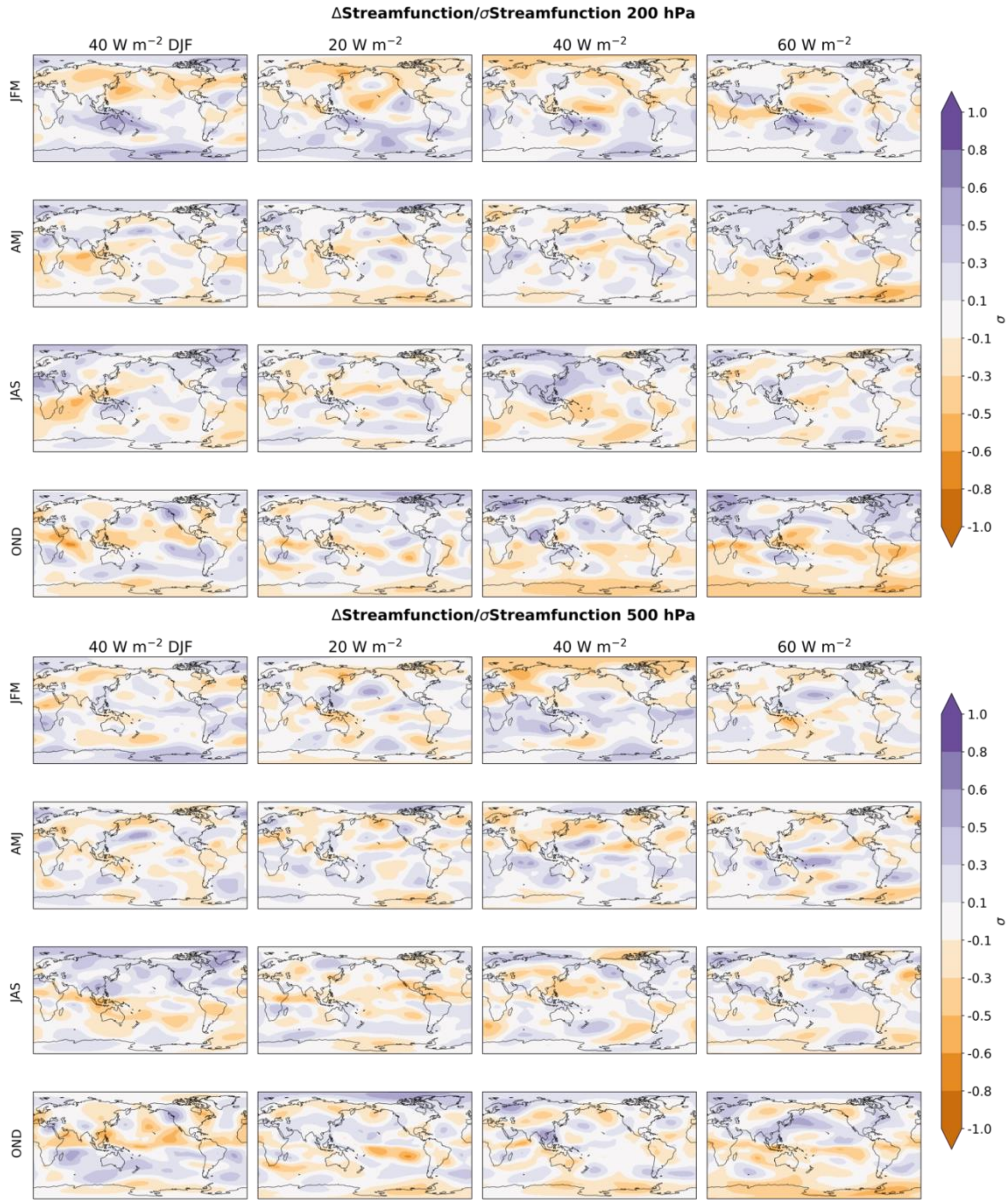

Figure 2-S7 - Seasonal streamfunction anomaly, normalized by standard deviation per grid-cell, relative to control for seasonal 40 W m-2 and annual 20, 40, and 60 W m-2 intervention.

Bibliography for Chapter 2

Ahlm, L., Jones, A., Stjern, C. W., Muri, H., Kravitz, B., & Kristjansson, J. E. (2017). Marine cloud brightening—As effective without clouds. *Atmospheric Chemistry and Physics*, *17*(21), 13071–13087. https://doi.org/10.5194/acp-17-13071-2017

Baird, M. E., Green, R., Lowe, R., Mongin, M., & Bougeot, E. (2020). Optimising cool-water injections to reduce thermal stress on coral reefs of the Great Barrier Reef. *PLOS ONE*, *15*(10), e0239978. https://doi.org/10.1371/journal.pone.0239978

Bay, L. K., Rocker, M., Boström-Einarsson, L., Babcock, R., Buerger, P., Cleves, P., Harrison, D., Negri, A., Quigley, K., Randall, C. J., van Oppen, M. J. H., & Webster, N. (2019). *Reef Restoration and Adaptation Program: Intervention Technical Summary. A report provided to the Australian Government by the Reef Restoration and Adaptation Program*. Australian Institute of Marine Science (AIMS). https://gbrrestoration.org/resources/reports/

Diamond, M. S., Gettelman, A., Lebsock, M. D., McComiskey, A., Russell, L. M., Wood, R., & Feingold, G. (2022). To assess marine cloud brightening's technical feasibility, we need to know what to study—And when to stop. *Proceedings of the National Academy of Sciences*, *119*(4), e2118379119. https://doi.org/10.1073/pnas.2118379119

Feingold, G., Ghate, V. P., Russell, L. M., Blossey, P., Cantrell, W., Christensen, M. W., Diamond, M. S., Gettelman, A., Glassmeier, F., Gryspeerdt, E., Haywood, J., Hoffmann, F., Kaul, C. M., Lebsock, M., McComiskey, A. C., McCoy, D. T., Ming, Y., Mülmenstädt, J., Possner, A., … Zheng, X. (2024). Physical science research needed to evaluate the viability and risks of marine cloud brightening. *Science Advances*, *10*(12), eadi8594. https://doi.org/10.1126/sciadv.adi8594

Field, L., Ivanova, D., Bhattacharyya, S., Mlaker, V., Sholtz, A., Decca, R., Manzara, A., Johnson, D., Christodoulou, E., Walter, P., & Katuri, K. (2018). Increasing Arctic Sea Ice Albedo Using Localized Reversible Geoengineering. *Earth's Future*, *6*(6), 882–901. https://doi.org/10.1029/2018EF000820

Gertler, C. G., O'Gorman, P. A., Kravitz, B., Moore, J. C., Phipps, S. J., & Watanabe, S. (2020). Weakening of the Extratropical Storm Tracks in Solar Geoengineering Scenarios. *Geophysical Research Letters*, *47*(11), e2020GL087348. https://doi.org/10.1029/2020GL087348

Gill, A. E. (1980). Some simple solutions for heat-induced tropical circulation. *Quarterly Journal of the Royal Meteorological Society*, *106*(449), 447–462. https://doi.org/10.1002/qj.49710644905

Harrison, D. D., Baird, Mark Edward, Harrison, L., Utembe, S., Schofield, R., Escobar Correa, R., Mongin, M., & Rizwi, F. (2019). *Environmental modeling of large-scale solar* *radiation management* [A report provided to the Australian Government from the Reef Restoration and Adaptation Program.]. Australian Institute of Marine Science.

Hernandez-Jaramillo, D. C., Harrison, L., Kelaher, B., Ristovski, Z., & Harrison, D. P. (2023). Evaporative Cooling Does Not Prevent Vertical Dispersion of Effervescent Seawater Aerosol for Brightening Clouds. *Environmental Science & Technology*, *57*(49), 20559–20570. https://doi.org/10.1021/acs.est.3c04793

Jones, A., Haywood, J., & Boucher, O. (2009). Climate impacts of geoengineering marine stratocumulus clouds. *Journal of Geophysical Research: Atmospheres*, *114*(D10). https://doi.org/10.1029/2008JD011450

Jones, A., & Haywood, J. M. (2012). Sea-spray geoengineering in the HadGEM2-ES earth-system model: Radiative impact and climate response. *Atmospheric Chemistry and Physics*, *12*(22), 10887–10898. https://doi.org/10.5194/acp-12-10887-2012

Krantz, Will; Neelin, David; Ahmed, Fiaz (Forthcoming 2024). Data from: A framework for minimizing remote effects of regional climate interventions: Cooling the Great Barrier Reef without teleconnections [Dataset]. Dryad. https://doi.org/10.5061/dryad.zpc866tj4

Latham, J., Gadian, A., Fournier, J., Parkes, B., Wadhams, P., & Chen, J. (2014). Marine cloud brightening: Regional applications. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *372*(2031), 20140053. https://doi.org/10.1098/rsta.2014.0053

Latham, J., Rasch, P., Chen, C.-C., Kettles, L., Gadian, A., Gettelman, A., Morrison, H., Bower, K., & Choularton, T. (2008). Global temperature stabilization via controlled albedo enhancement of low-level maritime clouds. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *366*(1882), 3969–3987. https://doi.org/10.1098/rsta.2008.0137

Logan, C. A., Dunne, J. P., Eakin, C. M., & Donner, S. D. (2012). *A framework for comparing coral bleaching thresholds*.

MacMartin, D. G., Kravitz, B., & Goddard, P. B. (2023). Transboundary effects from idealized regional geoengineering. *Environmental Research Communications*, *5*(9), 091004. https://doi.org/10.1088/2515-7620/acf441

Rasch, P. J., Latham, J., & Chen, C.-C. (Jack). (2009). Geoengineering by cloud seeding: Influence on sea ice and climate system. *Environmental Research Letters*, *4*(4), 045112. https://doi.org/10.1088/1748-9326/4/4/045112

Readfearn, G. (2020, April 16). Scientists trial cloud brightening equipment to shade and cool Great Barrier Reef. *The Guardian*.

https://www.theguardian.com/environment/2020/apr/17/scientists-trial-cloud-brightening-equipment-to-shade-and-cool-great-barrier-reef

Ricke, K., Ivanova, D., McKie, T., & Rugenstein, M. (2021). Reversing Sahelian Droughts. *Geophysical Research Letters*, *48*(14), e2021GL093129. https://doi.org/10.1029/2021GL093129

Sardeshmukh, P. D., & Hoskins, B. J. (1988). The Generation of Global Rotational Flow by Steady Idealized Tropical Divergence. *Journal of the Atmospheric Sciences*, *45*(7), 1228–1251. https://doi.org/10.1175/1520-0469(1988)045<1228:TGOGRF>2.0.CO;2

Stjern, C. W., Muri, H., Ahlm, L., Boucher, O., Cole, J. N. S., Ji, D., Jones, A., Haywood, J., Kravitz, B., Lenton, A., Moore, J. C., Niemeier, U., Phipps, S. J., Schmidt, H., Watanabe, S., & Kristjánsson, J. E. (2018). Response to marine cloud brightening in a multi-model ensemble. *Atmospheric Chemistry and Physics*, *18*(2), 621–634. https://doi.org/10.5194/acp-18-621-2018

Tollefson, J. (2021). Can artificially altered clouds save the Great Barrier Reef? *Nature*, *596*(7873), 476–478. https://doi.org/10.1038/d41586-021-02290-3

Wan, J. S., Chen, C.-C. J., Tilmes, S., Luongo, M. T., Richter, J. H., & Ricke, K. (2024). Diminished efficacy of regional marine cloud brightening in a warmer world. *Nature Climate Change*, 1–7. https://doi.org/10.1038/s41558-024-02046-7

Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Academic Press.

Wilks, D. S. (2016). "The Stippling Shows Statistically Significant Grid Points": How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It. *Bulletin of the American Meteorological Society*, *97*(12), 2263–2273. https://doi.org/10.1175/BAMS-D-15-00267.1

Chapter 3 - Regulatory frameworks for marine cloud brightening research in the United States

3.1. Introduction

The growing severity and visibility of impacts from anthropogenic climate change have driven the international consensus that urgent action is required to prevent further irreparable damage to global ecosystems and human society [1]. Despite renewed commitments to targets for reduction in emissions by mid-century, the possibility of reaching a world with at least 1.5 °C remains more likely than not even under a very low emission scenario[2]. In this context governments, scientists, and private investors have shown an increased interest in researching climate interventions to potentially offset the most harmful impacts of climate change or avoid global tipping points. The physical science underlying most proposed climate intervention technologies remains highly uncertain, and extensive research along with gradual scaling up of field tests will be required to determine the feasibility and risks of these methods. Developing research programs for better understanding the potential use of climate intervention technologies

---

[1] CLIMATE CHANGE 2022 – IMPACTS, ADAPTATION AND VULNERABILITY: WORKING GROUP II CONTRIBUTION TO THE SIXTH ASSESSMENT REPORT OF THE INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE, (H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama ed., 2022).

[2] CLIMATE CHANGE 2021 – THE PHYSICAL SCIENCE BASIS: WORKING GROUP I CONTRIBUTION TO THE SIXTH ASSESSMENT REPORT OF THE INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE, (Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou ed., 1 ed. 2023).

is urgent, but also faces significant hurdles navigating regulations and political controversy from the local to international scale.

Solar Radiation Management (SRM) describes a set of drastic interventions to reflect sunlight away from earth and cool the planet. This can be done at a variety of levels: through solar shades in space, particles suspended in the high in the atmosphere in the stratosphere, through low-lying clouds in the troposphere just above the earth's surface, or on the surface itself, like brightening the ocean or the roofs of buildings. There are efficiency, cost, and practicality trade-offs to all of these efforts. Stratospheric intervention is considered the most efficient from a cost perspective, and could produce the most temperature change with the lowest cost.[3] It produces a global cooling, with limited control over regional effects due to the nature of circulation in the stratosphere.[4] On the other hand, surface deployments provide the opportunity for very regional cooling, like offsetting the urban heat island effect that exacerbates deadly heat waves, or to cooling the ocean around critical regions of ice.[5] These surface actions are also the easiest to start with the tools and technologies we have today, but the least efficient and least scalable. Between these two is marine cloud brightening. It occupies a unique space as being somewhat easier to begin at a small scale compared to stratospheric intervention, but with more potential for scaling to global influence than surface interventions. Unlike other methods, there is potential for marine

---

[3] *Id.*

[4] Walker Lee et al., *Expanding the design space of stratospheric aerosol geoengineering to include precipitation-based objectives and explore trade-offs*, 11 EARTH SYSTEM DYNAMICS 1051–1072 (2020).

[5] Peter J. Irvine, Andy Ridgwell & Daniel J. Lunt, *Climatic effects of surface albedo geoengineering*, 116 JOURNAL OF GEOPHYSICAL RESEARCH: ATMOSPHERES (2011), https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JD016281 (last visited Dec 20, 2019).

cloud brightening to be used either as a targeted regional intervention to cool crucial resources like the water around sensitive reefs or ice, or as a scaled up to bring down the average temperature globally.[6]

Although stratospheric intervention has been given the most attention and funding so far, recent developments indicate that MCB has rapidly gained attention and additional funding sources in recent years. The inherent global nature of SAI has led to an enormous field of potential objectors to research and testing. Thus far, no field experiments have yet taken place, due to a combination of technical complexity, regulatory hurdles, and public opposition. MCB is perceived as being farther behind in theoretical development and has received less explicit attention in research and regulation. Despite this, the usefulness of MCB as tool for targeted regional cooling has allowed it to move more quickly towards potential adoption. Two research groups have already moved towards small outdoor tests of cloud brightening technology with varying degrees of success. The Australian government has approved and funded a plan to research MCB as a means to cool the water around the Great Barrier Reef to control coral bleaching,[7] and a small scale outdoor test has already taken place.[8] The University of Washington

---

[6]John Latham et al., *Marine cloud brightening: regional applications*, 372 PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A: MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES 20140053 (2014).

[7]Jan McDonald et al., *Governing geoengineering research for the Great Barrier Reef*, 19 CLIMATE POLICY 801–811 (2019).

[8]Putting the Great Barrier Reef marine cloud brightening experiment into context, , C2G (2020), https://www.c2g2.net/putting-the-great-barrier-reef-marine-cloud-brightening-experiment-into-context/ (last visited May 30, 2021).

also recently conducted a brief field experiment in Alameda California, which was quickly halted by the local city council.[9]

The experimental program in Australia and recent controversy in Alameda have illustrated that more work is needed to understand how current international and national regulatory regimes apply to marine cloud brightening. This paper examines the legal frameworks that could apply to MCB research across international, national and state scales. The Australian research program and current weather modification operations in the United States provide reference points to consider how an MCB research program based in the US may navigate the complex regulatory space. Despite the lack of frameworks specifically designed for MCB, existing environmental laws and regulatory processes offer potential pathways for research to proceed, though significant uncertainty remains about how these frameworks will be applied in practice.

3.2. Research pathways for marine cloud brightening

3.2.1 Physical basis of Marine Cloud Brightening

Marine cloud brightening relies on the physical phenomenon that clouds containing smaller water droplets reflect more light. Introducing aerosol particles into the atmosphere where conditions are favorable to cloud formation results in more water droplets forming and

---

[9] Alameda City Council Votes to Stop Cloud Brightening Test - The New York Times, https://www.nytimes.com/2024/06/05/us/alameda-cloud-brightening-climate-change.html (last visited Nov 20, 2024); To Slow Global Warming, Scientists Test Solar Geoengineering - The New York Times, https://www.nytimes.com/2024/04/02/climate/global-warming-clouds-solar-geoengineering.html (last visited Nov 20, 2024).

subsequently brighter clouds.[10] The process does not rely on introducing any additional water into the atmosphere, but simply redistributing it into particles the optimal size for reflecting light. The phenomenon is commonly observed as “ship tracks”, where a row of bright clouds will form along the path of large shipping vessels due to the aerosols output from their exhaust. Scientists have proposed producing the same effect by spraying sea water into the air, where salt particles should have the same brightening effect on the clouds without concerns of pollution. A large enough fleet of vessels pumping and spraying sea water could produce a significant local or even global cooling effect.[11]

Although the idea of cooling through cloud brightening was first proposed John Latham in 1990,[12] many fundamental scientific uncertainties technical hurdles remain. The cloud-aerosol interactions at the core of the brightening effect remain uncertain and difficult to model. The design and testing of a sprayer nozzle capable of producing particles of the correct size and spreading them into the atmosphere consistently is also a significant challenge. Finally, verifying and quantifying an actual brightening and cooling effect from a ship spraying aerosols is difficult given the relatively small effect relative to the natural variability of marine clouds.

The research necessary to fully characterize marine cloud brightening as a potential climate intervention ranges from fundamental research on cloud formation to highly specialized engineering tests and remote sensing campaigns. For the purpose of better understanding how

---

[10] John Latham et al., *Marine cloud brightening*, 370 PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A: MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES 4217–4262 (2012).

[11] *Id.*

[12] John Latham, *Control of global warming?*, 347 NATURE 339–340 (1990).

this body of research will interact with regulatory frameworks, this paper draws on a research plan proposed by the Marine Cloud Brightening Project at the university of Washington as a reference for how research might proceed.[13]

3.2.2 Phases of research

The first set of experiments outside of a laboratory that might garner any sort of regulatory attention would be outdoor tests of a spray system to study the physical characteristics of a sea salt aerosol plume. These tests would be done on land, away from the ocean, intentionally removed from the cloud interactions that introduce much greater complexity. On-site measurement equipment would be sufficient to study the physical characteristics of the aerosol plume and the effectiveness of the sprayer technology.

The second stage of experimentation would involve placing a sprayer on a fixed platform in a coastal location to begin testing interactions between the salt spray and marine atmospheric conditions. Locations along California's coastline have been proposed as possible sites for this testing due to their favorable atmospheric conditions. Through a combination of downwind measurement platforms and research aircraft, these experiments could answer many questions about the physical and chemical interactions between clouds and aerosols. Results from these experiments could greatly improve models of MCB, allowing better predictions of its effectiveness as well as indications of any unexpected side effects.

---

[13] Wood, *supra* note 11.

The final stage of research would involve mounting the spray system to a ship and testing it in open waters. For our purposes in understanding the regulatory implications, it is useful to separately consider experiments that would take place near shore within state or domestic waters and those that could take place further out to sea in international waters. Additional observations by other ships, aircraft, or satellites would likely be used in this phase to measure the larger scale effect on cloud brightness.

3.3. Potential harms from small-scale testing of marine cloud brightening

The scope of regulation that could apply to MCB research depends on the range of credible risks of harm to human health or to the environment that could result from research activities. Here we are careful to differentiate between potential risks from large-scale, sustained deployment (either at a regional or global scale) and risks from research activities and small-scale tests. For our purposes, we define small-scale testing to be anything between an outdoor sprayer test aimed at characterizing aerosol size distribution and a multi-platform test dispersing aerosols for weeks or months with the goal of producing a measurable change to local albedo. This range of activities is in line with what researchers have identified as necessary to thoroughly characterize the viability of MCB as a climate intervention technique [14]. Although the upper end of these research activities would be a significant undertaking, it is still orders of magnitude smaller than what

[14] Robert Wood & Thomas P. Ackerman, *Defining Success and Limits of Field Experiments to Test Geoengineering by Marine Cloud Brightening*, 121 CLIM. CHANGE 459 (2013); Graham Feingold et al., *Physical Science Research Needed to Evaluate the Viability and Risks of Marine Cloud Brightening*, 10 SCI. ADV. eadi8594 (2024).

would be necessary to produce a measurable global cooling impact[15]. There may be some overlap in the scale of the largest "small-scale" test (e.g. a proposal of 100x100 km test area[16]) and the smallest useful "operational" deployment of MCB for regional cooling (e.g. targeting particularly vulnerable coral reefs during marine heat waves). However, it is useful to include the largest possible activity that is justifiable as research in our analysis to ensure a thorough review of possible of regulatory jurisdiction.

What follows is a review of risks and possible harms from MCB research activities that may intersect with existing laws or regulations.

3.3.1 Influence on global circulation patterns

The most discussed potential side-effect from marine cloud brightening in academic literature is disruption to global circulation patterns, and the subsequent impacts to regional climate conditions. The high profile of this concern stems from climate simulations of global-scale deployment of MCB that produce significant and highly heterogeneous patterns of temperature and precipitation change.[17] Subsequent studies of the priorities for MCB research have repeatedly highlighted the need to further characterize the risk of harmful regional impacts that

---

[15] Robert Wood, *Assessing the Potential Efficacy of Marine Cloud Brightening for Cooling Earth Using a Simple Heuristic Model*, 21 ATMOSPHERIC CHEM. PHYS. 14507 (2021).
[16] Wood and Ackerman, *supra* note 14.
[17] A. Jones & J. M. Haywood, *Sea-Spray Geoengineering in the HadGEM2-ES Earth-System Model: Radiative Impact and Climate Response*, 12 ATMOSPHERIC CHEM. PHYS. 10887 (2012); Ben Kravitz et al., *Sea Spray Geoengineering Experiments in the Geoengineering Model Intercomparison Project (GeoMIP): Experimental Design and Preliminary Results*, 118 J. GEOPHYS. RES. ATMOSPHERES 11,175 (2013).

could result from cooling large patches of the ocean from MCB.[18] However, the scale of MCB deployment necessary to produce global shifts in circulation patterns is orders of magnitudes larger than any proposed field testing campaign.[19] Recent modeling studies have specifically studied whether smaller scale regional MCB cooling interventions targeting the Great Barrier Reef or the Gulf of Mexico could result in unintended non-local climate effects through teleconnections, and found no significant results outside the area of intervention, even for years of sustained cooling over a large region.[20]

Collectively these results indicate that global circulation impacts and trans-boundary harms from teleconnections are a risk associate with large-scale deployment of MCB, but they have not produced any evidence to support the idea that field tests, even at the largest scale, pose such risks. Given the temporal and spatial scale of a cooling perturbation required to test for a significant global impact in the presence of natural climate variability, these risks will be evaluated by modeling studies rather than field experiments.

3.3.2 Disruptive change to local weather patterns

While remote impacts to climate conditions are unlikely to result from testing MCB, smaller-scale atmospheric adjustments and influence on meteorology within the experimental area are

---

[18] Feingold et al., *supra* note 14; Michael S. Diamond et al., *To Assess Marine Cloud Brightening's Technical Feasibility, We Need to Know What to Study—and When to Stop*, 119 PROC. NATL. ACAD. SCI. e2118379119 (2022).
[19] Jessica S. Wan et al., *Diminished Efficacy of Regional Marine Cloud Brightening in a Warmer World*, NAT. CLIM. CHANGE 1 (2024).
[20] Douglas G. MacMartin, Ben Kravitz & Paul B. Goddard, *Transboundary Effects from Idealized Regional Geoengineering*, 5 ENVIRON. RES. COMMUN. 091004 (2023).

more likely. Tests like the ones that have been conducted in Australia and Alameda involving only brief periods of salt water spray are unlikely to have any meaningful impact on local weather. However, experiments intended to produce a measurable change in cloud albedo would, by definition, be altering the amount of sunlight reaching the surface over some area nearby the test site. The spatial extent and magnitude of this effect would depend on the size of the experiment, and research on precipitation enhancement has demonstrated that attributing any local weather changes to cloud seeding activity is difficult even with thorough instrumentation.[21]

Ocean-based experiments conducted far from populated areas would face fewer regulatory hurdles related to local weather impacts, and may provide better conditions for measuring the effectiveness of cloud brightening. However, early research to characterize spray systems and verify model predictions of aerosol-cloud interactions will likely need to occur closer to shore where more extensive measurement equipment can be deployed. The regulatory frameworks governing these coastal experiments will need to carefully weigh the scientific value of the research against potential impacts to local communities.

3.3.3 Particulate matter and air pollutants

One potential environmental hazard from MCB testing that has not been thoroughly explored by existing literature is impacts on local air quality. There are two areas of concern associated with generating aerosolized salt particles: health impacts from the aerosols themselves as particulate

[21] Andrea I. Flossmann et al., *Review of Advances in Precipitation Enhancement Research*, 100 BULL. AM. METEOROL. SOC. 1465 (2019); David Reynolds, *A Review of the Sierra Cooperative Pilot Project*, BULL. AM. METEROROLOGICAL SOC. (1986); LAKE OROVILLE RUNOFF ENHANCEMENT PROJECT - FINAL REPORT, (1995).

matter, and the generation of reactive halogens through chemical interactions of the salt aerosols with the boundary layer atmosphere.

Airborne aerosol particles are a widely recognized air pollutant associated with significant human health impacts[22]. Particulate matter is typically classified and regulated by the diameter of the particles, with "course particles" (diameter up to 10 μm) and "fine particles" (diameter up to 2.5 μm) receiving the most research and regulation.[23] Newer research has begun to recognize the health impacts of even smaller "ultrafine" particles with diameters below 0.1 μm, finding respiratory and cardiovascular impacts associated with even short-term exposure.[24] Although the precise physiological mechanisms of harm are not yet fully understood, typically the damaging potential of particulate matter is associated with its size and concentration in the air rather than the specific chemical composition of the aerosols.[25]

For marine cloud brightening, aerosols with a diameter of less than 0.5 μm are suitable for producing a brightening effect, with the ideal size range for mass efficiency being between 0.03 and 0.06 μm.[26] These aerosol particles would be classified as ultrafine particulate matter.

---

[22] Ki-Hyun Kim, Ehsanul Kabir & Shamin Kabir, *A Review on the Human Health Impact of Airborne Particulate Matter*, 74 ENVIRON. INT. 136 (2015).
[23] Morton Lippmann, *Particulate Matter (PM) Air Pollution and Health: Regulatory and Policy Implications*, 5 AIR QUAL. ATMOSPHERE HEALTH 237 (2012).
[24] Dean E. Schraufnagel, *The Health Effects of Ultrafine Particles*, 52 EXP. MOL. MED. 311 (2020); Liqun Liu et al., *Size-Fractioned Particulate Air Pollution and Cardiovascular Emergency Room Visits in Beijing, China*, 121 ENVIRON. RES. 52 (2013); George D. Leikauf, Sang-Heon Kim & An-Soo Jang, *Mechanisms of Ultrafine Particle-Induced Respiratory Health Effects*, 52 EXP. MOL. MED. 329 (2020); Robert B. Devlin et al., *Controlled Exposure of Humans with Metabolic Syndrome to Concentrated Ultrafine Ambient Particulate Matter Causes Cardiovascular Effects*, 140 TOXICOL. SCI. OFF. J. SOC. TOXICOL. 61 (2014).
[25] Leikauf, Kim, and Jang, *supra* note 12; Schraufnagel, *supra* note 12.
[26] Wood, *supra* note 4; Jack Foster et al., *Continuing Results for Effervescent Aerosol Salt Water Spray Nozzles Intended for Marine Cloud Brightening*, 11 INT. J. GEOSCI. 563 (2020).

Engineering a spray nozzle capable of producing the ideal size distribution for cloud brightening remains a significant challenge, and in practice the spray produced contains a mixture of ultrafine and fine particulate matter.[27] Given that further developing and testing spray systems is a significant focus of MCB research and much is still unknown about the growth rates and evolution of the aerosols in the atmosphere, it is reasonable to assume that tests could be producing particulate matter at a range of sizes that would classify as ultrafine, fine, and possible course particulate matter.

A second potential concern is the production of reactive halogen species through chemical interactions in the atmosphere. Naturally occurring sea salt aerosols play a complex role in atmospheric chemistry, and are the major source of tropospheric reactive chlorine and bromine.[28] These reactive halogens play a significant role in the oxidation reactions of ozone, methane, and nitrogen oxides, and a local change in sea salt aerosols could set off a complex chain of chemical reactions with implications for both air quality and radiative forcing.[29] Although one study estimated that the net impact of increased sea-salt aerosols from MCB would be the reduction in tropospheric ozone (an air pollutant at the surface) along with a slight increase in the lifetime of methane, the exact effects are expected to vary significantly with local conditions.[30] The

---

[27] Jack Foster et al., *Continuing Results for Effervescent Aerosol Salt Water Spray Nozzles Intended for Marine Cloud Brightening*, 11 INT. J. GEOSCI. 563 (2020); Diana C. Hernandez-Jaramillo et al., *Evaporative Cooling Does Not Prevent Vertical Dispersion of Effervescent Seawater Aerosol for Brightening Clouds*, 57 ENVIRON. SCI. TECHNOL. 20559 (2023).

[28] Rainer Vogt, Paul J. Crutzen & Rolf Sander, *A Mechanism for Halogen Release from Sea-Salt Aerosol in the Remote Marine Boundary Layer*, 383 NATURE 327 (1996).

[29] Hannah M. Horowitz et al., *Effects of Sea Salt Aerosol Emissions for Marine Cloud Brightening on Atmospheric Chemistry: Implications for Radiative Forcing*, 47 GEOPHYS. RES. LETT. e2019GL085838 (2020).

[30] *Id.*

atmospheric chemistry impacts of sea-salt spraying are not clearly negative for human health or the environment, but the reactive potential of salt aerosols in the troposphere has been underrecognized in existing writings on MCB. The potential of salt spray to influence concentrations of greenhouse gasses and air pollutants significantly increases the cross-section of existing environmental regulations that could potentially apply to research or eventual deployment.

In the case of the experiments in Alameda, potential health impacts on local residents played a large role in the discussion and ultimate decision to halt the experiments. A technical memorandum prepared by an engineering firm at the request of Alameda City Council found no apparent health risk to the surrounding community from the testing planned at the USS Hornet.[31] The conclusion of the report was based on the chemical composition of the artificial sea water being used, and the relatively short total duration of the spray tests. Despite the findings in the report, the city council members expressed explicit concern about the health impacts of fine particulate matter, and identified the need for a more medically rigorous assessment.[32]

3.4. Governance mechanisms applying to MCB research

---

[31] ANDREW ROMOLO, *Technical Memorandum*, https://alameda.legistar.com/View.ashx?M=F&ID=12975036&GUID=58E1E01B-25F2-4C21-831E-97F3C208A284 (last visited Nov 17, 2024).
[32] Alameda City Council Hearing - 6/4/2024, (2024).

3.4.1 International

UN Convention on Biological Diversity

The UN Convention on Biological Diversity (CBD) has the goal of protecting biological diversity and ensuring sustainable use of biological resources.[33] The key obligations of member parties include to "develop national strategies, plans or programs for the conservation and sustainable use of biological diversity", and to actively mitigate and manage the impacts of actions that "have or are likely to have significant adverse impacts on the conservation or sustainable use of biological diversity".[34] Although the US is not a party to the CBD, it remains an influential treaty that frames much of the dialog on the international regulation of MCB.

The CBD is notable for passing one of the only international resolutions specifically mentioning geoengineering. After passing a decision in 2008 aimed specifically at discouraging ocean fertilization activities,[35] a more general set of guidelines on geoengineering were passed in 2010.[36] These guidelines recommend that until sufficient global governance is implemented, "no climate-related geo-engineering activities that may affect biodiversity take place, until there is in place an adequate scientific basis on which to justify such activities and appropriate consideration of the associated risks for the environment and biodiversity and associated social, economic and cultural impacts".[37] The definition of "geo-engineering activities" agreed upon by

[33] CONVENTION ON BIOLOGICAL DIVERSITY, (1992).
[34] *Id.* at Art. 6(a), 7(c), 8(l).
[35] REPORT OF THE CONFERENCE OF THE PARTIES TO THE CONVENTION ON BIOLOGICAL DIVERSITY ON THE WORK OF ITS NINTH MEETING, DECISION IX/116, (2008).
[36] REPORT OF THE CONFERENCE OF THE PARTIES TO THE CONVENTION ON BIOLOGICAL DIVERSITY ON THE WORK OF ITS ELEVENTH MEETING, DECISION XI/20, (2012).
[37] *Id.* at Art. 8.

the COP includes any “deliberate intervention in the planetary environment of a nature and scale intended to counteract anthropogenic climate change and its impacts”.[38] This definition encompasses operational deployment of MCB with the goal of reducing any particular climate impact, but suggests that research activities with neither the scale nor the intent to counteract climate impacts would not be considered geo-engineering. More importantly, the CDB has a specific exemption to the restrictions on geo-engineering for “small scale scientific research studies that could be conducted in a controlled setting” provided they are justified by a need to collect specific scientific data and undergo an assessment of their potential impacts on the environment.[39] Although the requirement for a “controlled setting” is not well defined, these exemptions seem to align well with the scope of all the MCB research activities discussed above, and the vetting processes from national and state level regulations would satisfy the CBD’s requirements for environmental review.

Given that the United States is not a party to the CBD, and considering the explicit exemptions for research activities, the Convention has little practical bearing on MCB research programs conducted within U.S. jurisdiction. However, the CBD remains an influential framework that would need to be carefully considered in planning any future international MCB operations requiring cooperation between multiple states. The Convention's treatment of

[38] REPORT OF THE CONFERENCE OF THE PARTIES TO THE CONVENTION ON BIOLOGICAL DIVERSITY ON THE WORK OF ITS ELEVENTH MEETING, DECISION XI/20, ART. 5(B), (2012), https://www.cbd.int/decision/cop/default.shtml?id=13181.
[39] *Id.* at Art. 8.

geoengineering activities provides important context for how similar activities might be regulated under other international agreements.

London Convention and London Protocol

The London convention and London Protocol are multilateral treaties that aim to prevent marine pollution from dumping.[40] The parties to these treaties have also passed amendments explicitly governing specific geoengineering activities, which currently only includes ocean iron fertilization. In contrast to the decisions on geoengineering under the CBD, the LCLP is legally binding. Although the amendments have not been fully ratified, they are the only legally binding international instrument governing geoengineering.[41]

The general terms of the London Convention require that parties regulate dumping of particular prohibited substances into the ocean within their jurisdiction.[42] The London Protocol strengthens those restrictions to only allow the dumping of a small set of allowed substances.[43] The US has ratified the London Convention but not the London Protocol, and so it is technically only bound by the less restrictive of these two measures. The pertinent questions for whether the restrictions of the London Convention and Protocol apply to ocean-based MCB research are then if spraying aerosols above the ocean could be considered dumping, and whether the material sprayed is prohibited. The London Convention's definition of dumping contains an exception for

---

[40] Convention on the Prevention of Marine Pollution by Dumping of Wastes and Other Matter, (1972).

[41] JESSE L. REYNOLDS, THE GOVERNANCE OF SOLAR GEOENGINEERING: MANAGING CLIMATE CHANGE IN THE ANTHROPOCENE (1 ed. 2019), https://www.cambridge.org/core/product/identifier/9781316676790/type/book (last visited Apr 6, 2021).

[42] "London Convention", *supra* note 30.

[43] Protocol to the Convention on the Prevention of Marine Pollution by Dumping of Wastes and Other Matters, (1996).

matter placed into the ocean "for a purpose other than mere disposal thereof",[44] and so deposition of particles sprayed from a ship for the purpose of MCB would likely not be considered dumping. Furthermore, because MCB ships would only be dispersing salt directly from sea water, the material being dispersed is not prohibited and would be difficult to construe as dumping of waste.

The fact that MCB vessels would only be emitting sea water also serves as crucial factor in determining the applicability of the amendments that govern "marine geoengineering". The definition of marine geoengineering is "a deliberate intervention in the marine environment to manipulate natural processes, including to counteract anthropogenic climate change and/or its impacts, and that has the potential to result in deleterious effects, especially where those effects may be widespread, long lasting or severe".[45] This is a broader definition than the one given by the CBD and arguably encompasses MCB research that occurs at sea. However, the substance of the amendment only restricts "the placement of matter into the sea", which has been interpreted so far not to include spraying sea water.[46]

For activities that do meet the definition and involve placement of matter into the sea, parties of the convention must vote to explicitly add these activities to an annex to restrict their use.[47]

---

[44]"London Convention", *supra* note 30 at Art. I.

[45]Resolution LP.4(8) on the Amendment to the London Protocol to Regulate the Placement of Matter for Ocean Fertilization and Other Marine Geoengineering Activities, (2013).

[46]Harald Ginzky & Robyn Frost, *Marine Geo-Engineering: Legally Binding Regulation under the London Protocol*, 2014 CCLR 82–96 (2014); KERRYN BRENT, WIL BURNS & JEFFREY MCGEE, *Governance of Marine Geoengineering* (2019), https://www.cigionline.org/publications/governance-marine-geoengineering/ (last visited May 25, 2021).

[47]"London Convention", *supra* note 30 at Art. IV.

Presently, ocean iron fertilization is the only activity listed in the annex and restricted by the London Protocol amendments. Although more actions could be added to the annex, MCB research or even full scale operation do not fit into easily into the definitions offered. The only remaining restriction applicable to marine MCB testing under the London Convention would be the obligation to protect against conventional "wastes generated on the course of operation of vessels".[48]

UN Convention on the Law of the Sea

The UN Convention on the Law of the Sea (UNCLOS) provides a foundational regime for governance of marine activity, and obligates member states to "protect and preserve the marine environment"[49] by taking all measures to "prevent, reduce and control pollution of the marine environment".[50] Despite not ratifying it, the US acknowledges most of the provisions of the convention as customary international law.

The definition of pollution under UNCLOS explicitly includes pollution "from or through the atmosphere"[51] and extends to pollutants emitted on land that end up in the marine environment.[52] The definition even encompass "matter or energy" introduced into the marine environment that "results or is likely to result in …deleterious effects". The inclusion of added energy as a potential source of harm expands this definition to potentially encompass MCB ships, even if they are only pumping and re-circulating sea water. Despite this expansive definition, the

---

[48]*Id.* at (LC XII).
[49]United Nations Convention on the Law of the Sea, Art. 192 (1982).
[50]*Id.* at Art. 194.1.
[51]*Id.* at Art. 212.1.
[52]Reynolds, *supra* note 31.

UNCLOS provides little that would hinder MCB research on land or at sea. Provided that basic requirements for monitoring for potential harmful effects are met, UNCLOS affirms the right of each state to conduct scientific research within their domestic waters and on the high seas given "due regard for the interests of other states,"[53] particularly when that research is designed to protect the marine environment in alignment with the goals of the convention. Similar to the UNCBD, UNCLOS also contains basic requirements for reporting any expected transboundary effects. [54]

Other agreements

Depending on the results of early MCB testing and further modeling of impacts on atmospheric chemistry, other international agreements such as the convention on long-range transboundary air pollution (CLRTAP) or the Montreal Protocol could apply. With the current models predicting a lack of harmful pollutants produced by salt spray and the containment of any impact on ozone to the troposphere, it is unlikely that either of these treaties would be relevant. Other agreements such as The International Convention for the Prevention of Pollution from Ships (MARPOL) or the Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and Their Disposal ("Basel Convention"), govern releases of dangerous material from ships into the ocean, but would not apply to MCB given that it would not be transporting or introducing any new material into the ocean. As illustrated throughout this

[53]UNCLOS, *supra* note 39 at Art 87.
[54]*Id.* at Art. 206.

section, the fact that sea salt is a suitable aerosol for cloud brightening significantly reduces the regulatory complications of MCB.

3.4.2 US legal framework

For research on MCB conducted within the US the National Environmental Policy Act is the primary regulatory statute that could influence a research program. In some cases, provisions from the Clean Air Act, the Clean Water Act, and the Marine Protection, Research, and Sanctuaries Act may also apply

National Environmental Policy Act

The National Environmental Policy Act (NEPA) requires federal agencies to consider and report the potential environmental impacts of their actions. While it does not mandate that any particular impacts be avoided, it does require agencies to list the potential impacts of alternative actions and submit their findings for a period of public and agency review. A full review in the form of an *Environmental Impact Statement(EIS)* is required for any federal action with a risk of "significantly affecting the quality of the human environment."[55] Smaller actions may complete a less thorough *Environmental Assessment(EA)* to determine if they meet the threshold of significance to proceed with a full EIS.[56] A number of actions are also categorically excluded from the need to complete an EA or and EIS, based on a blanket determination that they pose no significant environmental impacts.[57]

[55] 42 U.S.C., § 4321 et seq., 4332(2)(C).
[56] 42 U.S.C., *supra* note 45.
[57] *Id.* at § 1501.4.

The applicability of NEPA to MCB research depends on the scale of the program and which federal agency is responsible for it. Short term research projects that are designed to characterize uncertain cloud-aerosol interactions and monitor for potential environmental effects would likely fall below the threshold required to trigger NEPA review. By the very nature of the research it would be difficult to state that it poses a significant threat before carrying out the research itself. Particularly because these initial studies are less likely to be part of a federally funded research program or require a federal permit, they may need no engagement with NEPA at all. If the point is reached the MCB technology has matured and research ships plan to deploy aerosols with the intention of producing a measurable (if small) changes in cloud brightness and surface temperature, NEPA review could become more likely.

Certain federal agencies also carry more potential for categorical exemption than others. Research projects led or funded by the NSF generally qualify for a categorical exemption, however there is an explicit requirement to carry out at least an Environmental Assessment for "weather modification, or other techniques that may alter a local environment".[58] The Department of Energy on the other hand has a broader categorical exemption for "small-scale research and development projects;… and small-scale pilot projects".[59]

Although most individual experiments in an MCB research program would not merit NEPA review, there is also the possibility that the federal government would undertake a

---

[58] 45 C.F.R. § 640.3(b)(3), (4), .
[59] 10 C.F.R. Pt. 1021, Subpt. D, App. B, B3.6, .

programmatic review of a coordinated geoengineering research program.[60] This approach has been advocated as a way to bring legitimacy to solar geoengineering research and address concerns about its risks in a public and transparent manner.[61]

Weather Modification and Reporting Act

The weather modification and reporting act (WMRA) act was passed in 1972 as research into weather modification and cloud seeding were increasing across the US. It creates a procedural requirement for anyone who carries out weather modification to report their activities to NOAA.[62] The definition of weather modification includes "Modifying the solar radiation exchange of the earth or clouds, through the release of gases, dusts, liquids, or aerosols into the atmosphere" and "Seeding or dispersing of any substance into clouds or fog, to alter drop size distribution",[63] which clearly covers MCB activities even at a small scale. Any MCB research activity would be required so submit fairly simple reports covering the nature of the project and the materials being used.

Clean Air Act

A few provisions of the Clean Air Act (CAA) are relevant to MCB research, though none are likely to significantly restrict activities. The CAA regulates the ambient levels of six criteria pollutants, including particulate matter, nitrogen oxides, and surface ozone.[64] The salt aerosols

---

[60] CLIMATE ENGINEERING AND THE LAW: REGULATION AND LIABILITY FOR SOLAR RADIATION MANAGEMENT AND CARBON DIOXIDE REMOVAL, (Michael B. Gerrard & Tracy Hester eds., 2018), https://www.cambridge.org/core/books/climate-engineering-and-the-law/CC93F3228AE1FBB028ED18C6DE6E19B8

[61] Charles Corbett, *"Extraordinary" and "Highly Controversial": Federal Research of Solar Geoengineering Under NEPA*, NW. U. L. REV. COLLOQUY (2021), https://scholarlycommons.law.northwestern.edu/nulr_online/308.

[62] 15 CFR 908.3, .

[63] *Id.*

[64] 42 U.S.C., §§ 7401–7431, 7501–7515.

released in any of the tests described in this paper would be considered particulate matter, meaning even an inland test of an aerosol plume could fall under some CAA regulation. Chemical reactions between salt aerosols and other atmospheric species would likely decrease the concentrations of nitrogen oxides and ozone locally.[65] The CAA does not directly regulate small individual sources, and the administration of restrictions on particulate emissions would be at the discretion of the air quality district wherever a test were taking place. Unless it were a very long duration test the additional aerosol load from an MCB research plume would likely be on an insignificant scale for regional air quality under the CAA. The CAA also regulates emissions from mobile sources that could apply to ships carrying out MCB research, but regulations only apply to emissions from the ship's engine and not other sources like a sprayer.[66]

Other provisions

The Clean Water Act puts limits on discharges into US waterways, but deposition of aerosols sprayed into the atmosphere would not qualify as a point source that is regulated under the act.[67] The Marine Protection, Research, and Sanctuaries Act ("MPRSA") regulates dumping of material into U.S. waters and implements responsibilities of the London Convention. Consistent with the London Convention, it would not apply to ships that are transporting or dispersing material for purposes other than dumping.

---

[65] Horowitz et al., *supra* note 18.

[66] TRACY HESTER, *Remaking the World to Save it: Applying U.S. Environmental Laws to Climate Engineering Projects* (2011), https://papers.ssrn.com/abstract=1755203 (last visited May 31, 2021).

[67] Tracy D. Hester, *Remaking the World to Save It: Applying U.S. Environmental Laws to Climate Engineering Projects*, 38 ECOLOGY L.Q. 851–902 (2011).

3.4.3 State laws and regulations

Within California, local agencies are responsible for administering obligations under the Clean Air Act and other state regulations through a review and permitting process for any new research activities. By far the most significant regulatory consideration will be the applicability of CEQA.

The California Environmental Quality Act

The California Environmental Quality Act (CEQA) is the regulatory tool most likely to apply to outdoor research of cloud-aerosol interactions and MCB testing. It applies to any project undertaken or funded by public agencies in California, as well as any projects requiring governmental participation, financing, or approval.[68] The core requirement of CEQA is to evaluate the potential environmental impacts of a project, and to mitigate those impacts whenever possible. This happens primarily through the production of an Environmental Impact Report (EIR), that is reviewed by the public and the permitting agency before a project is allowed to commence.[69] The expansive scope of CEQA leaves little doubt that an MCB research program in California would merit some degree of review under the CEQA process, but local agencies would have significant discretion over how much review is required. By design, CEQA relies on public scrutiny to ensure environmental impacts are thoroughly considered. For a potentially controversial project like MCB, navigating CEQA may be a balancing act of gaining trust through transparency and avoiding obstructionist delays.

---

[68]California Environmental Quality Act Statute and Guidelines, § 15002, subd. (a)(1)-(4)) (2021).
[69]*Id.* at § 21061.

Scope of CEQA applicability

The definition of "projects" within CEQA is broad enough that even short term research endeavors or temporary installations are captured, so long as they require some amount of funding or discretionary approval from a state body.[70] Before continuing with an EIR however, projects may be deemed exempt, or they may undergo a brief "initial study" to determine if a full EIR is required.[71] Even the earliest outdoor research of aerosol plumes would fit the definition of a project and must interact with CEQA at this level, even if they are later determined to not require a full review.

Exemptions

CEQA contains 33 classes of categorical exemptions that apply to types of projects that rarely require environmental review. A collection of statutory exemptions carves out further specific actions that are excused from the need to conduct EIRs. The only categorical exemption that may be applicable to MCB related research would be class 6, which covers "basic data collection, research, experimental management, and resource

evaluation activities which do not result in a serious or major disturbance to an environmental Resource".[72] This exemption covers activities like air and water quality monitoring, boreholes for ground water testing, and other sample collection or measurement activities. It is unclear whether it could extend to research projects that involve emitting matter into the environment as

[70] *Id.* at § 21065.
[71] *Id.* at § 15063.
[72] *Id.* at § 15306.

in a test aerosol plume, although is has been used to cover outdoor testing of pesticides and small controlled burns.[73]

Initial studies

Without a clear exemption, any outdoor testing project would require the lead agency to conduct an initial assessment to survey possibly significant environmental impacts and how easily they could be mitigated. CEQA guidelines provide an extensive list of potential impacts that must be considered, including "emissions adversely effecting a substantial number of people", "obstruction of the applicable air quality plan", "adverse effect on any riparian habitat or other sensitive natural community", and any "impacts that are individually limited but cumulatively considerable".[74] The exact process for conducting the initial assessment and what level of analysis is required will depend on the agency conducting the study and the permitting body responsible for approving it. Many small projects receive either a "negative declaration" or a "mitigated negative declaration" after an initial study, indicating they do not have significant environmental impacts that merit a full EIR.

Necessity of full EIR

The decision of whether to conduct a full EIR for a research project is at the discretion of lead agency. If the decision not to conduct an EIR is challenged, the court will order the completion of a full EIR if there is a *fair argument* that the project may have a significant impact

---

[73] DEPARTMENT OF FISH AND WILDLIFE, *Ocean Ranch Unit Pilot Project* (2020), https://ceqanet.opr.ca.gov/2020030433/2.
[74] CEQA Guidelines, *supra* note 58 at Appendix G.

on the environment. This low standard for the completion of an EIR puts pressure on agencies to proactively complete an EIR if there is any chance the initial study will be the subject of significant scrutiny. Particularly in the case of a project with the potential for controversy, there are trade offs to be considered

For an MCB research plan that has several phases, ranging from experiments with very potential for environmental impact to projects that certainly justify and EIR, agencies would need to make a strategic decision about when to undertake the more extensive review. Completing an EIR an early stage could require significant time and resources up front, but expedite subsequent review processes by generating a record of successful reviews. Due to the decentralized administration of CEQA, it is likely that separate studies and/or EIRs would be required for each phase of research, especially if each project is taking place in a different location and will be lead by a different local agency.

Joint NEPA/CEQA review

In the case that a NEPA review is taking place for an MCB research program, it is encouraged to complete a joint assessment under CEQA and NEPA. The state provides extensive guidelines on coordinating joint reviews to avoid duplicated work.[75] Since the threshold for significance is higher in NEPA, it is almost always the case that a project undergoing NEPA review, even at just the level of an Environmental Assessment, will need a CEQA EIR. In this case the state agency must ensure that the additional consideration of mitigation measures above

---

[75]NEPA and CEQA: Integrating Federal and State Environmental Reviews, (2014).

and beyond NEPA requirements are included.[76] A more likely scenario is that CEQA review is required but the necessity of NEPA is more ambiguous. In this case, following the guidelines for coordination with federal agencies while completing a CEQA review can greatly ease integrating those studies into a NEPA review if one is eventually conducted.[77]

3.5. Lessons from cloud seeding in California

To better understand the way CEQA may apply to MCB research in California, we can learn from the example of the ongoing precipitation enhancement and cloud seeding program. This program shares many physical and operational similarities with marine cloud brightening, and has been weaving its way through the California regulatory environment for decades.

Precipitation enhancement through cloud seeding involves burning a flare either from the ground or from an aircraft that releases silver iodide particles into the atmosphere. Adding these particles to a convective storm system increases the formation of ice crystals within the clouds, leading to more rain or snow droplets and increased precipitation within the storm.[78] It is extremely difficult to determine the large scale effect of cloud seeding on precipitation,[79] but nonetheless it has been a popular method to enhance water resources across the western United States.

---

[76] *Id.*

[77] *Id.*

[78] Don A. Griffith et al., *The Santa Barbara Cloud Seeding Project in Coastal Southern California, Summary of Results and Their Implications*, 37 JOURNALWEAMOD 21–27 (2005).

[79] Sarah A. Tessendorf et al., *An Assessment of Winter Orographic Precipitation and Cloud-Seeding Potential in Wyoming*, 59 JOURNAL OF APPLIED METEOROLOGY AND CLIMATOLOGY 1217–1238 (2020).

Although MCB and cloud seeding have somewhat opposite effects on clouds (MCB seeks to spread water out across more smaller droplets, while cloud seeding tries to encourage larger droplets that will rain out), the processes are similar and they raise parallel sets of concerns about their impact on the physical environment. In both cases, regulators and the public might be concerned about impacts to local air quality from the release of aerosols, downwind deposition of chemicals into water or soil, or the broader ecological and climatological effects of the cloud manipulation. Cloud seeding operations in California thus provides an illustrative reference for how a controversial climate technology has undergone public and regulatory scrutiny.

3.5.1 Regulatory treatment in California

Cloud seeding operations regularly take place in around a dozen locations along the California coast and the Sierra Nevada range.[80] Local agencies within each county manage contracting and permitting private operators of the cloud seeding projects. The state requires that local agencies file a notice of intent for new operations, renewed every five years, submit reports to the California Department of Water Resources, and report to NOAA to satisfy the requirements of the Weather Modification Reporting Act.[81] Finally, the state requires that all agencies sponsoring cloud seeding operations comply with CEQA [82].

In practice, the level of review carried out to comply with CEQA varies substantially from county to county. One of CEQA's defining features is the lack of a single agency to administer it, and so this level of heterogeneity is not uncommon. Most counties file regular

---

[80] PRECIPITATION ENHANCEMENT - A RESOURCE MANAGEMENT STRATEGY OF THE CALIFORNIA WATER PLAN, (2016).
[81] *Id.*
[82] *Id.*

Mitigated Negative Declarations for their cloud seeding operations, indicating that they conducted an Initial Assessment and determined that any potentially significant risks to the environment were appropriately managed.[83] At least some locations operate cloud seeding projects with only a notice of intent, with the county agency relying on findings from previous studies to justify not completing an EIR or an initial study.[84] Within the negative declarations that have been filed in recent years, some contain detailed modeling and scientific assessment of potential risks to justify the decision,[85] while others simply complete the minimum checklist of potential impacts provided by CEQA documentation.[86]

In the mitigated negative declarations that have been published recently, the environmental impacts that were deemed not significant "due to mitigation actions" included disturbance to sensitive wildlife habitats, public safety hazards from burning flares and the release of Silver Iodide, and alteration to local hydrology and flood risk.[87]

It could be surprising that actions like cloud seeding, designed to have a significant effect on the local environment, passes through the CEQA process without an EIR or even an initial assessment in some cases. The statewide Water Plan from the California Department of Water Resources cites findings from prior federal and state level studies carried out as early as 1977 to

---

[83]The Governor's Office of Planning and Research, *CEQAnet*, https://ceqanet.opr.ca.gov.

[84]Charlie Unkefer, *PG&E weather modification plan raises concerns*, MOUNT SHASTA HERALD - MOUNT SHASTA, CA , https://www.mtshastanews.com/article/20081105/NEWS/311059945

[85]Los Angeles County, *County of Los Angeles Weather Modification Project*, https://ceqanet.opr.ca.gov/2009071101 (last visited May 31, 2021).

[86]Livermore Valley Joint Unified School District, *Santa Barbara County and Twitchell Reservoir Cloud Seeding*, https://ceqanet.opr.ca.gov/2000101010 (last visited May 31, 2021).

[87]County, *supra* note 75; District, *supra* note 76.

indicate that environmental impacts are not a major concern.[88] These findings have also been publicly acknowledged by PG&E officials as justification for not carrying out CEQA review of modern cloud seeding projects.[89]

The fact that cloud seeding is treated as a routine undertaking that garners little special attention from CEQA could lead to the conclusion that pilot MCB programs would be treated similarly. However, a closer look reveals a history of state and federal environmental review that proceeded the current status quo. In 1986 the Department of the Interior and the state of California undertook the Sierra Cooperative Pilot Project to investigate cloud seeding to increase winter precipitation in the Sierra Nevada mountains. This project underwent an environmental assessment under NEPA which resulted in a "finding of no significant impact", and was not reviewed by CEQA.[90] In 1991, a full joint EIR/EIS was completed for the Oroville runoff enhancement project, a three year study and pilot project of winter cloud seeding undertaken by the California Department of Water resources and the Bureau of Reclamation. This project involved extensive monitoring of air quality, aerosol plume characteristics, and changes in meteorological conditions that contributed extensively to the knowledge of cloud seeding practices. Independent of federal research programs, California agencies completed full EIRs on pilot cloud seeding projects on several more occasions throughout the late 80s and early 90s. Together with several more recent studies outside of the CEQA process on the potential for

---

[88]PRECIPITATION ENHANCEMENT - A RESOURCE MANAGEMENT STRATEGY OF THE CALIFORNIA WATER PLAN, *supra* note 70.
[89]Unkefer, *supra* note 74.
[90]David Reynolds, *A Review of the Sierra Cooperative Pilot Project*, BULLETIN AMERICAN METEROROLOGICAL SOCIETY (1986).

contamination of water or soils,[91] these studies have formed a basis of knowledge and prior risk assessment that is perceived to alleviate the need for continued EIRs for cloud seeding. Although the application of CEQA over the years has changed significantly, the early environmental impact reviews of cloud seeding may offer a better comparison for new MCB research than the way current cloud seeding projects are regulated in California.

3.6. Conclusions

A review of the regulatory frameworks that could apply to MCB research reveals an interesting mismatch between the most commonly discussed international governance regimes and the regulations that are likely to meaningfully impact near-term research programs. Although decisions under the CBD and amendments to the London Convention and Protocol have created the only legally binding international instruments specifically addressing geoengineering, their applicability to MCB research is limited by explicit exemptions for scientific research activities. Additionally, the requirements of UNCLOS and other agreements that aim to prevent transboundary environmental harms are unlikely to restrict small-scale research given the limited spatial and temporal extent of proposed experiments. These international frameworks may play a crucial role in governing eventual deployment of MCB, particularly at scales that could produce measurable changes to regional climate, but they provide little practical constraint on the research activities necessary to characterize the feasibility and risks of the technology.

---

[91] PRECIPITATION ENHANCEMENT - A RESOURCE MANAGEMENT STRATEGY OF THE CALIFORNIA WATER PLAN, *supra* note 70.

Within the United States, federal and state environmental regulations provide a more immediate set of requirements for MCB research programs to navigate. However, the applicability of these regulations depends significantly on the specifics of how research activities are funded and where they take place. Projects occurring in state waters or requiring state permits would trigger state environmental review under laws like CEQA, but research activities taking place further offshore might avoid these requirements entirely. The discretion given to agencies in determining the scope of environmental review creates further uncertainty in how these regulations would apply in practice.

Given the limited role of international agreements and the variable applicability of federal and state environmental regulations, local governments and agencies are positioned to have an outsized influence on early MCB research activities. As illustrated by the recent experience in Alameda, California, city councils and other local bodies can effectively block research activities regardless of federal or state level approvals. Even in locations where research is allowed to proceed, local air quality districts would have significant discretion over restrictions on particulate emissions from spray testing. This situation is likely to persist until MCB research matures to the point where it can move to ocean-based platforms outside the jurisdiction of local and state governments. The path taken by cloud seeding in California, where extensive environmental review of early research projects eventually led to more routine treatment by local regulators, may provide a model for how MCB research could gradually establish legitimacy within existing regulatory frameworks.

Bibliography for Chapter 3

Alameda City Council Votes to Stop Cloud Brightening Test—The New York Times. (n.d.). Retrieved November 20, 2024, from https://www.nytimes.com/2024/06/05/us/alameda-cloud-brightening-climate-change.html

Devlin, R. B., Smith, C. B., Schmitt, M. T., Rappold, A. G., Hinderliter, A., Graff, D., & Carraway, M. S. (2014). Controlled exposure of humans with metabolic syndrome to concentrated ultrafine ambient particulate matter causes cardiovascular effects. *Toxicological Sciences: An Official Journal of the Society of Toxicology*, *140*(1), 61–72. https://doi.org/10.1093/toxsci/kfu063

Diamond, M. S., Gettelman, A., Lebsock, M. D., McComiskey, A., Russell, L. M., Wood, R., & Feingold, G. (2022). To assess marine cloud brightening's technical feasibility, we need to know what to study—And when to stop. *Proceedings of the National Academy of Sciences*, *119*(4), e2118379119. https://doi.org/10.1073/pnas.2118379119

Feingold, G., Ghate, V. P., Russell, L. M., Blossey, P., Cantrell, W., Christensen, M. W., Diamond, M. S., Gettelman, A., Glassmeier, F., Gryspeerdt, E., Haywood, J., Hoffmann, F., Kaul, C. M., Lebsock, M., McComiskey, A. C., McCoy, D. T., Ming, Y., Mülmenstädt, J., Possner, A., … Zheng, X. (2024). Physical science research needed to evaluate the viability and risks of marine cloud brightening. *Science Advances*, *10*(12), eadi8594. https://doi.org/10.1126/sciadv.adi8594

Flossmann, A. I., Manton, M., Abshaev, A., Bruintjes, R., Murakami, M., Prabhakaran, T., & Yao, Z. (2019). Review of Advances in Precipitation Enhancement Research. *Bulletin of the American Meteorological Society*, *100*(8), 1465–1480. https://doi.org/10.1175/BAMS-D-18-0160.1

Foster, J., Cooper, G., Galbrath, L., Jain, S., Ormond, R., & Neukermans, A. (2020). Continuing Results for Effervescent Aerosol Salt Water Spray Nozzles Intended for Marine Cloud Brightening. *International Journal of Geosciences*, *11*(9), Article 9. https://doi.org/10.4236/ijg.2020.119030

Hernandez-Jaramillo, D. C., Harrison, L., Kelaher, B., Ristovski, Z., & Harrison, D. P. (2023). Evaporative Cooling Does Not Prevent Vertical Dispersion of Effervescent Seawater Aerosol for Brightening Clouds. *Environmental Science & Technology*, *57*(49), 20559–20570. https://doi.org/10.1021/acs.est.3c04793

H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (Ed.). (2022). *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://doi.org/10.1017/9781009325844

Horowitz, H. M., Holmes, C., Wright, A., Sherwen, T., Wang, X., Evans, M., Huang, J., Jaeglé, L., Chen, Q., Zhai, S., & Alexander, B. (2020). Effects of Sea Salt Aerosol Emissions for Marine Cloud Brightening on Atmospheric Chemistry: Implications for Radiative Forcing. *Geophysical Research Letters*, *47*(4), e2019GL085838. https://doi.org/10.1029/2019GL085838

Jones, A., & Haywood, J. M. (2012). Sea-spray geoengineering in the HadGEM2-ES earth-system model: Radiative impact and climate response. *Atmospheric Chemistry and Physics*, *12*(22), 10887–10898. https://doi.org/10.5194/acp-12-10887-2012

Kim, K.-H., Kabir, E., & Kabir, S. (2015). A review on the human health impact of airborne particulate matter. *Environment International*, *74*, 136–143. https://doi.org/10.1016/j.envint.2014.10.005

Kravitz, B., Forster, P. M., Jones, A., Robock, A., Alterskjær, K., Boucher, O., Jenkins, A. K. L., Korhonen, H., Kristjánsson, J. E., Muri, H., Niemeier, U., Partanen, A.-I., Rasch, P. J., Wang, H., & Watanabe, S. (2013). Sea spray geoengineering experiments in the geoengineering model intercomparison project (GeoMIP): Experimental design and preliminary results. *Journal of Geophysical Research: Atmospheres*, *118*(19), 11,175-11,186. https://doi.org/10.1002/jgrd.50856

Lake Oroville runoff enhancement project—Final Report. (1995). U.S. Department of the Interior.

Leikauf, G. D., Kim, S.-H., & Jang, A.-S. (2020). Mechanisms of ultrafine particle-induced respiratory health effects. *Experimental & Molecular Medicine*, *52*(3), 329–337. https://doi.org/10.1038/s12276-020-0394-0

Lippmann, M. (2012). Particulate matter (PM) air pollution and health: Regulatory and policy implications. *Air Quality, Atmosphere & Health*, *5*(2), 237–241. https://doi.org/10.1007/s11869-011-0136-5

Liu, L., Breitner, S., Schneider, A., Cyrys, J., Brüske, I., Franck, U., Schlink, U., Marian Leitte, A., Herbarth, O., Wiedensohler, A., Wehner, B., Pan, X., Wichmann, H.-E., & Peters, A. (2013). Size-fractioned particulate air pollution and cardiovascular emergency room visits in Beijing, China. *Environmental Research*, *121*, 52–63. https://doi.org/10.1016/j.envres.2012.10.009

MacMartin, D. G., Kravitz, B., & Goddard, P. B. (2023). Transboundary effects from idealized regional geoengineering. *Environmental Research Communications*, *5*(9), 091004. https://doi.org/10.1088/2515-7620/acf441

Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (Ed.). (2023). *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157896

Report of the Conference of the Parties to the Convention on Biological Diversity on the Work of its Eleventh Meeting, Decision XI/20, Art. 5(b). (2012). Secretariat of the Convention on Biological Diversity. https://www.cbd.int/decision/cop/default.shtml?id=13181

Reynolds, D. (1986). A Review of the Sierra Cooperative Pilot Project. *Bulletin American Meterorological Society*.

Romolo, A. (n.d.). *Technical Memorandum*. Terraphase Engineering. Retrieved November 17, 2024, from https://alameda.legistar.com/View.ashx?M=F&ID=12975036&GUID=58E1E01B-25F2-4C21-831E-97F3C208A284

Schraufnagel, D. E. (2020). The health effects of ultrafine particles. *Experimental & Molecular Medicine*, *52*(3), 311–317. https://doi.org/10.1038/s12276-020-0403-3

To Slow Global Warming, Scientists Test Solar Geoengineering—The New York Times. (n.d.). Retrieved November 20, 2024, from https://www.nytimes.com/2024/04/02/climate/global-warming-clouds-solar-geoengineering.html

Vogt, R., Crutzen, P. J., & Sander, R. (1996). A mechanism for halogen release from sea-salt aerosol in the remote marine boundary layer. *Nature*, *383*(6598), 327–330. https://doi.org/10.1038/383327a0

Wan, J. S., Chen, C.-C. J., Tilmes, S., Luongo, M. T., Richter, J. H., & Ricke, K. (2024). Diminished efficacy of regional marine cloud brightening in a warmer world. *Nature Climate Change*, 1–7. https://doi.org/10.1038/s41558-024-02046-7

Wood, R. (2021). Assessing the potential efficacy of marine cloud brightening for cooling Earth using a simple heuristic model. *Atmospheric Chemistry and Physics*, *21*(19), 14507–14533. https://doi.org/10.5194/acp-21-14507-2021

Wood, R., & Ackerman, T. P. (2013). Defining success and limits of field experiments to test geoengineering by marine cloud brightening. *Climatic Change*, *121*(3), 459–472. https://doi.org/10.1007/s10584-013-0932-z

Chapter 4 - Decision relevant GCM evaluation for studying climate change in California

Abstract

Latest-generation global climate models (GCMs) developed for the sixth phase of the Coupled Model Intercomparison Project (CMIP6) provide state-of-the-art simulations of how Earth's systems are expected to change through the end of the century with continued emissions of greenhouse gasses. However, the low spatial resolution and regional biases in GCMs present decision makers with challenges in developing local-scale future climate assessments, thus necessitating the need for sub-selecting model simulations to regionally downscale. Here, we present a comprehensive evaluation of CMIP6 GCMs for downscaling that combines two complementary approaches: an assessment of large-scale processes important to the western United States (e.g., Northern Hemispheric circulation, jet stream climatology, ENSO) and an independent ranking of GCMs' abilities to accurately simulate regional surface air temperature and precipitation. By comparing these two evaluation methods, we identify GCMs that perform well in both frameworks, indicating robust representation of both regional climate and the physical processes driving it. Importantly, we find that the process-based evaluation does poorly in evaluating long-term variability, while the evaluation of local historical conditions has a blind-spot for the drivers of extreme precipitation events. Finally, we demonstrate how this dual evaluation framework can guide the selection of an ensemble of climate simulations that balances uncertainty in future change signals, inter-model differences, and extreme events. The result is an ensemble of downscaled simulations specifically designed to aid stakeholder

adaptation planning, and a flexible model evaluation process that can be expanded for future studies.

4.1 Introduction

As the impacts of climate change become more prominent around the world, scientists and decision makers need reliable predictions of the regional hazards threatening societies, economies, and ecosystems in the coming century. The latest generation of global climate models (GCMs) developed for the sixth phase of the Coupled Model Intercomparison Project (CMIP6) provide state-of-the-art simulations of how Earth's systems are expected to continue changing through the end of the century with continued emissions of greenhouse gasses, production of anthropogenic aerosols, and changes to land use. However, leveraging this wealth of data for practical adaptation planning brings several challenges. For one, the data is of too low spatial resolution to capture many of the physical processes that impact conditions on the scale of a city or a watershed, especially in regions of complex topography. Further, although the models in CMIP6 have shown measurable improvements in their representation of key climate processes (Cannon, 2020; Simpson et al., 2020; Pierce et al. 2022), they can still contain biases in important regional processes (Abdelmoaty et al., 2021; Kim et al., 2020; Priestley et al., 2022; Pierce et al. 2022).

Downscaling addresses the first of these difficulties. In this work we focus on dynamical downscaling: running a high-resolution regional weather or climate model with boundary conditions forced by a GCM produces climate projections where crucial details like extreme heat events, snowpack patterns, runoff, and orographic precipitation are far better resolved (Giorgi et al., 1994; Berg & Hall, 2017; Racherla et al., 2012; Rahimi et al., 2024). However, a regional climate model can only produce reliable simulations if the GCM used as an input is accurate. Because a regional climate model will propagate whatever large-scale conditions and biases are present in the GCM into the region of interest, it is even more crucial to address the second difficulty: selecting GCMs that are skilled at simulating the larger scale weather and climate conditions in the specific region of study.

A wide range of literature demonstrates procedures for selecting GCM simulations for downscaling (Jury et al., 2015; McSweeney et al., 2012, 2015; Pierce et al., 2009; Virgilio et al., 2022). Most of these studies follow a similar structure of selecting relevant performance metrics for the region of interest, devising a method to weight and combine metrics into a single score per model, and then selecting some number of the top models to downscale. Studies selecting models for statistical downscaling are primarily concerned with the accuracy of the model within the region of study, and emphasize performance metrics comparing local climate conditions to observations (Pierce et al., 2009, 2022; Xue & Ullrich, 2021). In addition, the importance of key global-scale dynamical processes in influencing climate in remote regions has long been recognized; for example, the evaluation of the El Nino/Southern Oscillation, Pacific Decadal Oscillation, and teleconnected sea level and temperature responses to Pacific Ocean variability have all been used as performance metrics for statistical downscaling in the western United States (Pierce et al. 2009, 2022). However these non-local metrics gain additional emphasis when selecting GCMs to supply boundary conditions for dynamic downscaling, where accurate simulation of the large-scale meteorological processes driving the flow of energy and moisture into the domain is essential (Goldenson et al., 2023; Karmalkar et al., 2019; McSweeney et al., 2015; Meng-Zhuo Zhang et al., 2022). These process-based evaluation methods encompass hemispheric circulation patterns as well as teleconnections and modes of global variability that are important to the region under study.

There are many approaches to selecting an ensemble from the best performing GCMs depending on the goals of the study and the resources available for downscaling. A simple approach would be to try to ensure that the statistics of the subset of GCMs to be downscaled match the larger set of best performing GCMs to the extent possible to avoid introducing a bias (Karmalkar et al., 2019). On the other hand, some studies have tried to use ensemble selection or weighting to reduce future uncertainty, with the hypothesis that models with more accurate historical conditions will produce more trustworthy future change projections (Brunner et al., 2020; Lorenz et al., 2018). Many studies emphasize the value of sampling inter-model variability and avoiding models that are too structurally similar to each other (Goldenson e t al., 2023; Karmalkar et al., 2019; Meng-Zhuo Zhang et al., 2022; Sanderson et al., 2017), as has been done in California previously (Lynn et al., 2015). Whether an ensemble should avoid edge case models or intentionally select for worst-case scenarios depends on the specific goals of the study. Ultimately the choice of ensemble has elements that are inherently subjective, and when providing data to guide stakeholder decision-making, the justification for the choices must be clearly communicated to allow correct interpretation of the ensemble's behavior.

This paper outlines a process for selecting an ensemble of GCMs for studying regional climate change through dynamical downscaling across the western US, with a focus on California. We demonstrate this process for the specific case of producing stakeholder-relevant climate projections to inform adaptation planning in California. Following the guidelines in Goldenson et al. (2023), we develop a set of process-based metrics targeting large-scale drivers of climate for California. We focus in particular on conditions associated with extreme heat, wildfire risk, prolonged drought, and extreme precipitation. To further ensure the robustness of our evaluation and the flexibility of the resulting ensemble, we combine the process-based ranking with a set of local metrics designed to inform statistical downscaling. Finally we demonstrate selecting an ensemble of simulations for downscaling that balances goals of sampling inter-model differences, impactful extreme events, and the range of future change signals.

Sections 4.2 and 4.3 describe the specific data and metrics used in this evaluation respectively, Section 4 shows the process of combining them to produce a model ranking. In Section 4.5, we combine this ranking with a ranking based on historical local climate conditions. Finally, in section 4.6, we discuss selecting a representative set of model simulations for downscaling.

4.2 Data

The model data for this study was taken from the Earth System Grid Federation (ESGF) archive for CMIP5 and CMIP6 GCMs (O'Neill et al., 2016; Taylor et al., 2012). The focus of this study is identifying CMIP6 models for downscaling, but the previous generation of GCMs is included to compare performance across model generations. The performance of the GCMs is evaluated against the European Center for Medium Range Weather Forecasting (ECMWF) Reanalysis, version 5 (ERA5; Hersbach et al., 2020). Both the model and the reanalysis data are interpolated to a common grid (either a 1˚ or 2˚ grid depending on the metric, as described below). Except where otherwise specified in the description of metrics below, the performance of models is evaluated and compared to ERA5 over the historical reference period of 1979-2014.

4.3 Process-based metrics for regional climate

4.3.1 Standardizing climate metrics

To ensure that metrics for a wide range of physical properties can be combined into an overall skill score, it is important that each metric be unitless, computed consistently, and span a similar numerical range. This could be accomplished by normalizing or re-scaling each metric, but such normalization would exaggerate small differences in metrics where all models perform similarly well. Instead, we standardize each metric using the Normalized Mean Square Error (NMSE). The NMSE was originally proposed by Williamson (1995), and has been used by Simpson et al. (2020) to evaluate the performance of Community Earth System Model 2 (CESM2) against the rest of the CMIP6 models. For the climatological average of a model field X_m, the NMSE is given by:

$$NMSE(X_m) = \frac{\overline{(X_m - X_o)^2}}{\overline{(X_o')^2}}$$

Where X_o is the climatological average of the observed data field (from ERA5 in this case), the overbar represents an area weighted spatial average, and the prime represents the deviation from the spatial average. Computing the NMSE from climatologically averaged fields measures the mean bias and spatial differences relative to the spatial variability in that field. This allows for the direct comparison of errors in climate variables with very different spatial and temporal characteristics, though it is limited in its ability to capture differences in temporal variability.

4.3.2 Process-based climate metrics

The GCMs in this study were evaluated by their accuracy in simulating the large-scale circulation patterns and physical processes that strongly influence climatological conditions and extreme weather events in California. In particular, we target Northern Hemispheric circulation features such as Pacific storm tracks and regional jet dynamics, as well as patterns of atmospheric variability such as ENSO. Because the regional climate models used for downscaling are primarily driven by GCM conditions at the lateral and lower boundaries of the domain, particular attention is paid to winds and moisture transport over the Pacific where climate information is passed into the downscaling model. A full overview of the metrics used in this evaluation is given below.

4.3.2.1. Northern hemisphere circulation

- Seasonal 300-hPa eddy stream function ($\psi*$)
- Seasonal 850-hPa zonal winds
- Seasonal 10-day high-pass-filtered eddy meridional wind variance ($v'v'$)

One of the highest priorities for this model evaluation is capturing the large-scale circulation patterns across the entire northern hemisphere. Just over half of the metrics used in the evaluation measure seasonal patterns in physical fields related to northern hemisphere circulation: the 300-hPa eddy stream function ($\psi*$), 850-hPa zonal winds, and 10-day high-pass-filtered eddy meridional wind variance ($v'v'$). These metrics were previously used to evaluate the performance of CESM2 relative to other CMIP models (Simpson et al., 2020), and are expanded here to evaluate a slightly larger set of CMIP6 models.

4.3.2.2. Blocking

- Summer and winter blocking
- Summer and winter mean-fixed blocking

A northern hemisphere blocking index captures the frequency of days where a significant circulation anomaly blocks the westerly mid-latitude winds. These blocking events can produce heat waves in the summer and extreme cold events in the winter. The blocking metric used is originally described in Masato et al. (2013) and here we use the specific implementation calculated for the CMIP6 ensemble in Simpson et al. (2020). Blocking is calculated between 25°N and 75°N for the historical period, with separate metrics calculated for the winter and summer months. The metric originally calculated for a subset of the CMIP6 ensemble by Simpson et al. (2020) has been calculated for additional CMIP6 models for this evaluation. A supplementary blocking metric developed by Scaife et al. (2010) and also computed in Simpson et al. (2020) normalizes for biases in the mean flow to remove the contribution of these biases to the blocking calculation in order to represent the portion of the error that is due to errors in synoptic variability alone. This metric is included as “mean fixed” blocking.

4.3.2.3. Wind shear

Seasonal wind shear

Early dynamical downscaling experiments as part of this project identified a poor representation of off-shore wind shear which correlated with biases in dynamically downscaled simulations across the California region. In this work a wind-shear metric was added to identify those GCMs that realistically capture the off-shore vertical structure of the winds at the boundaries. The monthly difference is taken between the zonal winds at two levels in the atmosphere (250 hPa minus 850 hPa) over the northern Pacific Ocean (between 150-230˚E and 20-66˚N). An NMSE is calculated from this two-dimensional grid of wind shear across the domain on a seasonal timescale by comparing to ERA5 reanalysis, with all data re-gridded to a common 2˚ grid through bilinear interpolation.

4.3.3.3. California extreme precipitation

- Integrated Water Vapor (IVW) during extreme precipitation events
- Sea Level Pressure (SLP) during extreme precipitation events
- 250 hPa zonal wind (u250) during extreme precipitation events
- California Precipitation Mode
- Self Organizing Map (SOM) nodes characterizing synoptic patterns of integrated vapor transport during heavy precipitation events

Three sets of metrics focused on capturing processes related to extreme precipitation in California are calculated as described below.

Metrics are used to evaluate the GCM's representation of average vertically integrated column water vapor (IWV), sea level pressure (SLP), and 250 hPa zonal wind (u250) on days of extreme precipitation in California. Following the method used in Norris et al. (2021), extreme wet days are identified as those with average California precipitation above the 95th percentile of all wet season days (November-April). The NMSE is computed comparing each GCM to ERA5 over this same reference period within a domain bounded by 20˚N-60˚N and 150˚W-100˚W. For total column water vapor (CWV), only ocean points are considered due to the difficulty of vertically integrating over land with data on limited pressure levels.

The California precipitation mode (CPM) is a distinct mode of atmospheric pressure over the North Pacific that strongly influences extreme precipitation and dry days in California. The mode is identified as the 3rd empirical orthogonal function (EOF) of 500 hPa geopotential height (Z500) anomalies (20-75N, 90-170W), with positive anomalies in this mode strongly associated with extreme (>99th percentile) precipitation days and negative anomalies strongly associated with dry (<10th percentile) days (Chen et al., 2021). An NMSE is computed comparing the pattern of the 3rd Z500 EOF of each GCM to ERA5 over a reference period of 1982-2014, which captures how well the GCMs re-produce the location and strength of this mode.

Southern California faces unique patterns of change to precipitation variability and extreme precipitation that may not be well captured by metrics that focus on the entire state (Swain et al., 2018). A set of metrics examining the meteorological circulations that cause heavy precipitation in Los Angeles County help ensure these patterns are included in the evaluation. Data from 553 heavy precipitation days in winter (October to April) from 1950 to 2019 are clustered using a self-organizing map (SOM). The days were selected when any 24-hour rain gauge measurement in Los Angeles Country was greater than its 2-year return value. The SOM was applied to the combination of standardized 250-hPa stream function (ST250) anomaly and integrated water vapor transport (IVT) anomaly over the domain of 195W-110W and 20N-50N, based on the National Center for Environmental Prediction (NCEP) 20th century reanalysis (Compo et al., 2011). Four modes from the SOM with the highest cross-model variance that are associated with heavy precipitation days are used to compute metrics. An NMSE is calculated for each mode by comparing the two-dimensional map of the average value of IVT or ST250 across all the days associated with the mode between the GCM and ERA5.

4.3.2.4. Santa Ana winds

- Sea-level Pressure (SLP) during Santa Ana events

Characterized by a strong horizontal gradient in sea-level pressure (SLP), the Santa Ana winds in southern California are among the strongest drivers of fire risk in the region. Given the importance of understanding wildfire risk for the stakeholder planning in California, particularly within the electricity sector, we include a metric for the large-scale synoptics associated with Santa Ana Winds. The method described in Abatzoglou et al. (2013) is used to identify Santa Ana events between 1979 and 2004 in both the GCMs and ERA5 reanalysis, and an NMSE is calculated from the mean SLP across the western US and eastern Pacific (130W-100W, 30N-50N) on days with strong Santa Ana events. While the empirical relationship used for this metric (from Abatzoglou 2013) was developed relative to Los Angeles. We suspect that models that capture this well would behave similarly well in terms of the large-scale boundary conditions controlling the development of comparable winds in nearby Santa Barbara.

4.3.2.5. El Niño Southern Oscillation

- First Empirical Orthogonal Function (EOF) of ENSO
- ENSO Skew

Large-scale modes of natural climate variability can influence temperature and precipitation in California via teleconnections, for example the ENSO. We evaluate each model's representation of ENSO through two metrics: one for the overall spatial pattern of the ENSO SST anomaly, and another for the non-linear behavior of ENSO progression. We identify the sea surface temperature (SST) pattern associated with ENSO variability by calculating the first empirical orthogonal function (EOF), of the sea surface temperatures over the Pacific Ocean basin, north of 60˚S. The pattern is compared via the NMSE with that derived via the same procedure from the NOAA Extended Reconstructed SST V3b dataset over the period 1854-2005. The second metric evaluates the representation of non-linear dynamics in ENSO through the skewness of SST anomalies, as shown in An et al. (2020). This is a feature that is known to be challenging for GCMs (An et al., 2005a; T. Zhang & Sun, 2014) and is likely to identify significant differences between models. Skewness in the distribution of sea surface temperature anomaly is calculated for each grid cell in the equatorial pacific (Between 10N-10S, 120E-90W). An NMSE is computed from comparison to the same observational dataset as above.

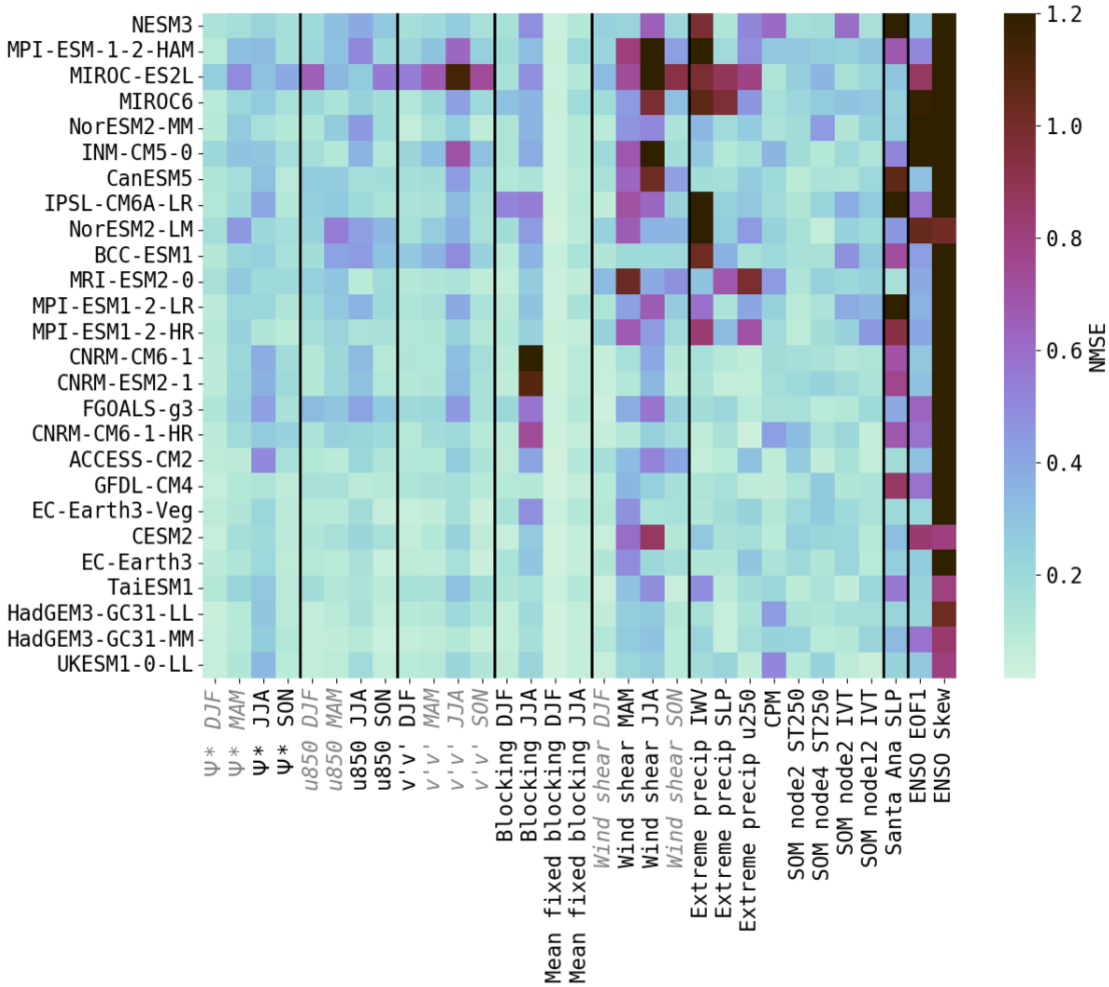

Figure 4-1. Normalized mean square error (NMSE) values across all process-based climate metrics (x-axis) for each GCM (y-axis). Lower NMSE scores indicate better model

performance. Metrics evaluating similar processes are grouped together by label color, and metrics not used for total error score are in gray. Models are ordered from worst (top) to best (bottom) based on the total euclidean distance error score.

4.3.3 Process-based metric results

Figure 4-1 shows the NMSE values across all the metrics for each CMIP6 model evaluated (CMIP5 is shown in Figure 4-S1). The total number of models included is limited by the availability of high-frequency data in all the required variables at the time this analysis was performed. Out of 169 CMIP5 and CMIP6 models listed on ESGF 111 had enough data available to calculate at least some metrics, and only 27 had the necessary data available to calculate all 31 metrics used in this evaluation. To expand the number of models that could be evaluated, models missing values for 4 or fewer metrics are also included. The missing metric values are filled with the mean NMSE for that metric across the other models to avoid giving any relative advantage or disadvantage for the missing metrics.

Since our goal is to identify meaningful differences between models, we draw special attention to the metrics shown in Figure 4-1 that have a wide range of NMSE values across the GCMs. Notably, the first ENSO EOF, summer blocking, sea level pressure during Santa Ana events, and seasonal wind shear measurements in the spring and summer all show a wide range of NMSE values. Because of this higher variance, these metrics contribute more weight to differentiating model performance. On the other hand, with metrics like winter blocking and the winter eddy stream function, the model scores show little variation. These metrics still represent processes that are important for western US climate, but they do not capture significant differences for model performance. As noted above, capturing the skew of ENSO temperature anomalies is particularly challenging for GCMs, and is the only metric where all models perform especially poorly.

4.4 Ranking and uncertainty

4.4.1 Reducing redundancy within metrics

The individual metrics we have chosen have some notable redundancy across the physical processes that they cover. For example, moisture transport during extreme precipitation is measured through the IWV across all of California and as the IVT focused on southern California precipitation events. Further, circulation patterns at 850 hPa and 250 hPa are captured individually, but also in a wind shear metric that captures differences between them. This set of metrics was chosen for their thorough coverage of important physical processes rather than their complete independence from one another. As a result, some physical processes or qualities of the model may receive extra weight in our evaluation without a step to remove redundancy between metrics. Similarly, some metrics are calculated for all four seasons, which may or may not evaluate independent information about model performance.

Several other model evaluation efforts have implemented some form of redundancy removal step, often by using principal component analysis or principal feature analysis to reduce the dimensionality of the metrics used before computing a final score (Pierce et al., 2009, 2022; Xue & Ullrich, 2021). In testing, we found a similar effectiveness for removing redundancy and more interpretable results using an iterative trimming based on the variance inflation factor (VIF), as in Lorenz et al. (2018). The VIF quantifies the multicollinearity of any one metric using a least squares regression against the rest of the metrics. To avoid overfitting these regressions based on the large number of predictors, we use a forward stepwise multiple linear regression to compute the VIF. At each step, the metric with the highest VIF is removed and the process is repeated until all metrics have a VIF under a standard threshold of 5 (Craney & Surles, 2002). Through this process, we drop some of the seasonal metrics of meridional wind variance, 850 hPa zonal winds, wind shear, and 300 hPa eddy stream function. None of these metrics are dropped across all four seasons, so some representation of each physical process is retained in the final list in Table S1. Comparing the model ranking with and without this redundancy removal step shows only small changes to the total error score, and no re-ordering of model rank.

4.4.2 Model Ranking

4.4.2.1. Overall error score

An overall error score is computed for each model from the metrics remaining after the redundancy removal step. This overall score is computed as the Euclidean distance between the point representing perfect model skill and the point represented by the model's error score on each metric. Perfect model skill would be an NMSE value of zero for all metrics and any larger value represents decreased performance. Figure 4-2 shows the total error scores for each model, calculated using the subset of metrics retained after redundancy removal.

4.4.2.2. Sensitivity to included metrics

While the redundancy removal step above ensures that each retained metric is contributing independent information, we also perform a sensitivity analysis to determine if any individual metric carries too much influence on the final model ranking. Ideally a robust GCM evaluation will not change dramatically based on the inclusion or exclusion of any single metric. To test this, each metric is individually dropped in turn and the process of VIF truncation and overall ranking is repeated in each case. The resulting collection of model rankings shows minimal variation, with 14 of the 31 metrics producing no change to the overall ranking when dropped (Figure 4-S2). Only two metrics, ENSO EOF and ENSO Skew, produce any change to the top 10 models in the ranking when removed, each swapping 2 of the top 10. Although these metrics have stronger individual influence, we retain them because of ENSO's strong influence on regional temperature and precipitation variability.

4.4.2.3. Uncertainty from natural variability

Most of the models in CMIP6 have been run through the end of the 21st century several times for each future climate scenario. Each of these model realizations contains a different expression of natural climate variability. The range of total error scores resulting from this variability represents an irreducible uncertainty in the overall score of each GCM. Estimating this uncertainty is helpful for determining when differences in model performance are significant relative to natural variability. Given that real world observations are also the result of processes driven by natural variability, it is appropriate to only expect models to match historical observations within an estimated range of natural variability.

Because the metrics used require extensive daily data, most of the CMIP6 models do not have sufficient outputs available at this time to compute a full error score for more than one realization. To overcome this limitation, we compute a full error score for one model (CESM2) that has the necessary data available for ten model realizations. The standard deviation of this ten member ensemble is used to approximate the uncertainty in the rest of the CMIP6 models, which is shown as error bars in Figure 4-2. Although this can only serve as a rough approximation, the approach is supported by the results in Pierce et al. (2022), which shows similar magnitudes of uncertainty from natural variability across many CMIP6 models when evaluating their local climate performance (See Section 5).

4.4.3 Process-based model ranking results

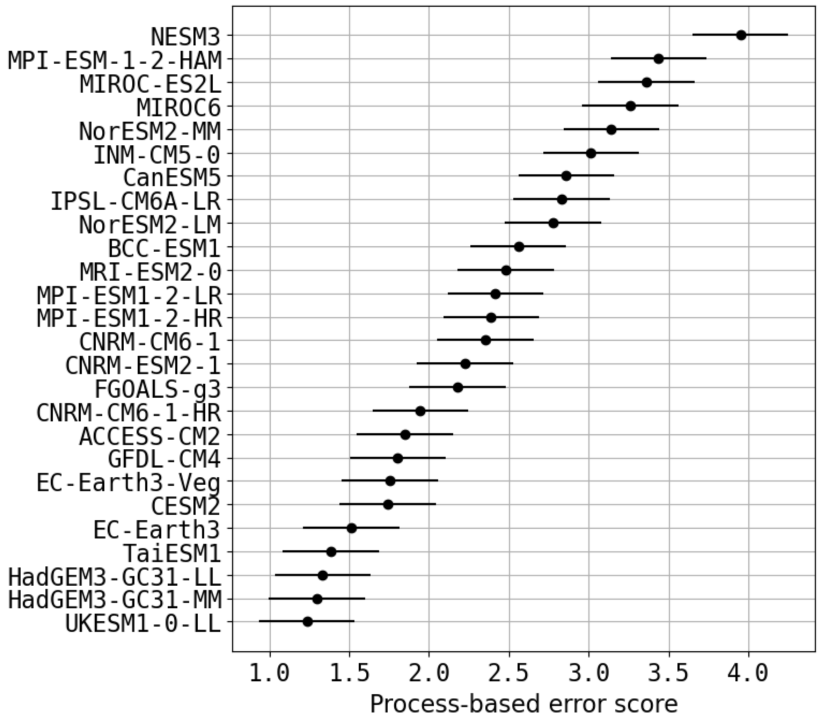

Figure 4-2. The total process-based error score for each model, with lower values representing better model performance. Error bars indicate the estimated standard deviation from an ensemble of model runs due to CESM2-derived internal climate variability.

4.5 Comparison to local climate metrics

To cross-check this model ranking and ensure we are selecting GCMs with the most accurate representation of regional climate, we combine the process-based ranking described above with an independent ranking using local climate metrics. Rather than large-scale climatological features, these metrics directly evaluate each model's representation of surface air temperature and precipitation within the region of interest. This independent ranking provides a way to check for blind spots in the process-based metric rankings. Models that perform well in the process-based evaluation but poorly in the local climate evaluation may indicate an impactful physical feature that is not well captured by the process-based metrics. Similarly, models that perform well in the local climate evaluation but poorly in the process-based metrics could indicate statistical coincidence, offsetting physical biases, or tuning to match historical values despite deficiencies in the underlying model physics. By considering these two complementary evaluation systems simultaneously, we have increased confidence that our model selections will robustly capture the climate system of the western US.

4.5.1 Overview of local metrics

The local climate metrics that we employ are based on the methodology developed by Pierce et al. (2009, 2022) to evaluate GCMs for regional climate studies across the western US. These metrics are computed by comparing each model to observational data on a common 1˚ latitude/longitude grid over a domain spanning 32 to 42 North and 125 to 114 West (a box covering California and Nevada). The dataset used for historical observations of precipitation and temperature is from Livneh (2015), over a time period of 1950-2005. Mean temperature and precipitation are measured as seasonal averages over the historical period, and interannual variability of both variables is measured on three timescales by first averaging the data into 1-, 5-, and 10-year blocks then taking the standard deviation. The phase and amplitude of the seasonal cycle of precipitation and temperature are measured using a best-fit sinusoid. Finally, the standard deviation of monthly temperature and precipitation for January and July are used to evaluate shorter timescale variability. All together, this creates a set of 40 measurements that capture the historical regional climate conditions in each model.

These measurements are evaluated against observations and used to form a set of skill scores following the method in Pierce et al. (2022). Before the model data is compared against observations, a simple bias correction is applied. Because GCMs are often bias corrected before downscaling (For example, see Rahimi et al. 2024, Risser et al. 2024), this allows the local metrics to evaluate the remaining errors that would be passed to the regional or statistical model. As described in Pierce et al. 2022, the simple bias correction subtracts off the time- and space-averaged error of the GCM with respect to the observations. Since the simple bias correction uses only a single number for all times and locations, GCMs can still sensibly be evaluated for their ability to reproduce spatial and temporal variability in comparison to the observations.

After bias correction, a skill score is calculated for each measurement by computing a z-score at each grid point and taking the root mean square across the domain:

$$SS(X_m) = 1 - \ RMS\left(\frac{X_m - X_o}{\sigma_o}\right)$$

Where X_m and X_o represent the climatological average or standard deviation of the model field and observational data field respectively, and σ_o represents the standard deviation over time of the observational data field. The full set of 40 local climate skill scores for each model is shown in Figure 4-S4. As with the process-based metrics, a redundancy removal step is taken before computing an overall model score using the Euclidean distance between the point representing perfect model skill and the point represented by the model's error score on each metric (Pierce et al., 2022)

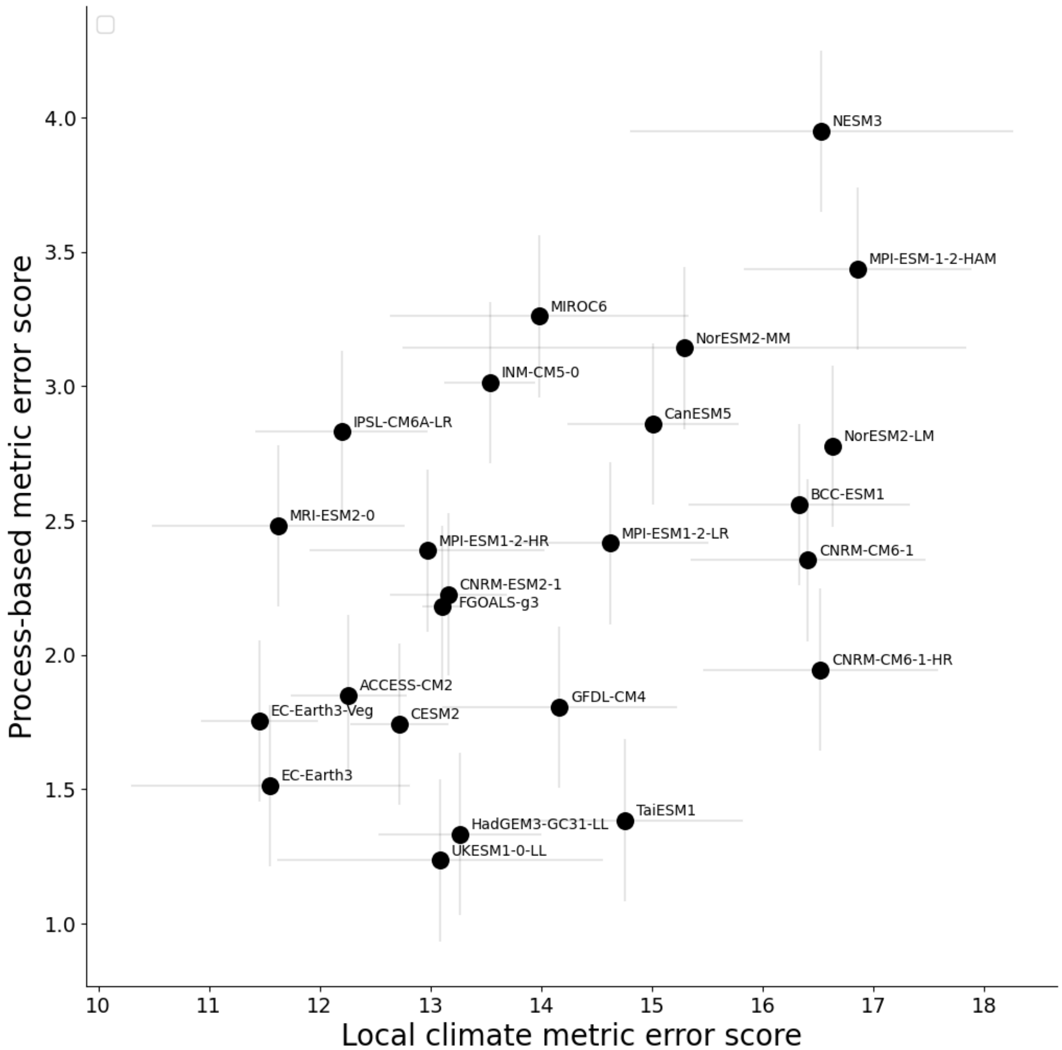

Figure 4-3. The local climate error score (x-axis) compared to the process-based climate error score (y-axis) for models in CMIP6. Error bars represent estimated uncertainty due to internal variability among simulations from the same model. Better performance is indicated by a lower error score on both axes (lower left corner).

4.5.2 Comparing process-based and local climate rankings

Both the process-based and local climate evaluations capture unique qualities of model performance that are important for studying regional climate change. Rather than standardizing the metrics from both evaluations and combining them into a single score, we preserve the individual methodologies and compare the final results of both rankings by presenting them on each axis of a scatter plot (Figure 4-3). This approach reveals an overall trend of agreement between the two rankings, while also revealing corner cases that lend insight into the differences between the two evaluation methods. Even though the scale and range of the two ranking schemes differs, the relative position of the models on the scatterplot provides a useful way of summarizing their performance without needing to choose an explicit weight for each evaluation.

The models that perform well in both evaluations (lower left corner of Figure 4-3, e.g. EC-Earth3, EC-Earth3-Veg, CESM2, ACCESS-CM2) constitute the strongest candidates for regional downscaling in California. These models demonstrate an accurate representation of California's historical climate, and do so with a skillful representation of the large-scale physical processes that are driving those conditions. In contrast, models in the upper left quadrant (e.g. IPSL-CM6A-LR, MRI-ESM2-0, INM-CM5-0) have a fairly accurate historical climate on average, but their poor relative performance in the process-based evaluation indicates they may be "right for the wrong reasons", or lack a robust representation of processes that drive important extreme events. On the other hand, models falling on the lower-right corner of the plot (e.g CNRM-CM6-1-HR, TaiESM1, GFDL-CM4), demonstrate that skillful representation of large-scale processes is not sufficient to produce accurate local climate conditions, even after accounting for biases in the mean state.

A natural question that arises when comparing these two evaluation schemes is whether the process-based metrics are actually incorporating new information, or if these two sets of metrics are effectively redundant. Under the hypothesis that the physical processes evaluated are solely responsible for driving the local climate conditions, it would be possible to predict the local climate metric values with some appropriate linear combination of process-based metrics.

If all the important influences of the climate process are measured in local conditions, then we could predict the process metrics from the local ones.

To test this, we perform a multiple linear regression on each metric to find how completely it can be modeled from metrics in the other set. Due to the small sample size (37 models across CMIP5 and CMIP6) relative to the number of predictors (22 process metrics or 40 local metrics), there is a risk of overfitting a model and exaggerating the predictive power. To counter this, we use a stepwise-forwards regression that adds predictors one at a time until none of the remaining metrics contribute a significant ($p<0.1$) improvement to the model fit.

The regression models using process-based metrics explain 44% of the variance of the local metrics on average, while the local metrics predict 51% of the variance of the process-based metrics on average (individual R^2 values shown in Figure 4-S5). We do not attribute much significance to the absolute value of these R^2 values, but the low relative values within each set of metrics reveal the “blind spots” of model behavior that the predicting set of metrics could have evaluated. Among the local metrics, most of the lowest R^2 values occur on the 5- and 10-year variability metrics, indicating that the process-based metrics that we selected may be systematically missing the drivers of long-term variability. Among the process-based metrics, metrics associated with extreme precipitation and winter blocking have relatively lower R^2 values, indicating that the local climate metrics used may be missing shorter timescale extremes.

Having identified these blind spots, we can test whether they help explain the behavior of the corner case models by computing a specialized skill score for each one. Figure 4-4 illustrates model performance within these subsets in the context of the comparison between the process and local evaluations. The models that rank highly based on the local climate evaluation but relatively lower in the process-based evaluation tend to be brought down by their representation of extreme precipitation conditions (Figure 4-4a). Conversely, the models that rank highly in the process-based evaluation and relatively lower in the local climate evaluation perform very poorly on measurements of decadal variability (Figure 4-4b). Overall, this indicates the value of combining the two rankings, and lends advice towards crafting a thorough set of metrics for future evaluations.

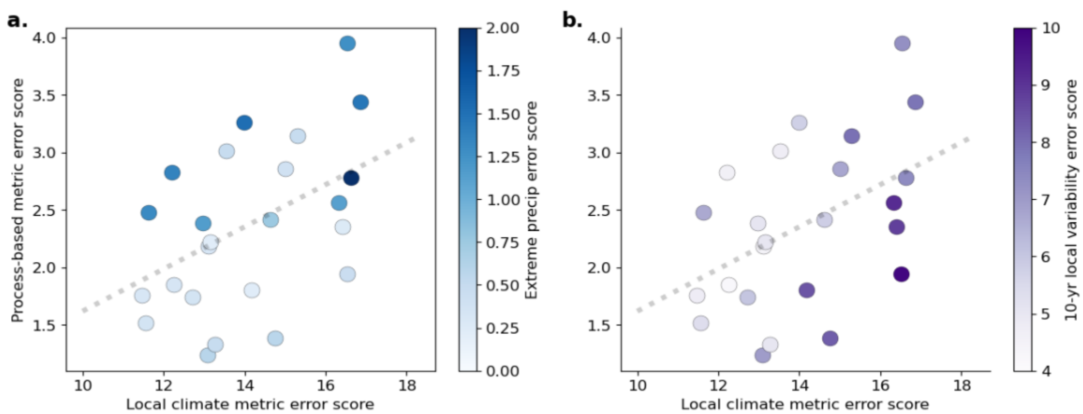

Figure 4-4. The scatterplot as shown in Figure 4-3, but with a color scale showing model performance on the subset of process-based extreme precipitation metrics (a) and local 10-year variability (b).

4.6 Selecting an ensemble for downscaling

The analysis discussed so far identifies the best GCMs for producing realistic climate projections and boundary conditions for downscaling over California. The following section covers the process of selecting an ensemble of individual realizations to downscale from the top performing models. At the time this analysis was performed, two GCMs (CESM2 and MPI-

ESM1-2-LR) had each already had one realization downscaled based on earlier iterations of the evaluation criteria. Given the computational resources available, we sought to select five additional simulations to expand the ensemble. (Further simulations were subsequently added using different criteria, but we illustrate here how our methods yielded five additional simulations.) Rather than strictly selecting the top-ranking models in order until the ensemble is full, we consider a slightly broader cluster of models and select among them to satisfy additional priorities for the ensemble as a whole, as described below. As the error bars representing estimated variation from natural variability on Figure 4-2 indicate, the performance of many of the models are statistically indistinguishable given the sampling uncertainty. For the remaining steps of ensemble selection, we consider the following models as having the best performance across the processed-based and local metrics, as indicated in Figure 4-3, with the number of realizations available indicated parenthetically:

ACCESS-CM2 (2)

CESM2 (9)

CNRM-ESM2-1 (1)

EC-Earth3 (1)

EC-Earth3-Veg (1)

FGOALS-g3 (2)

HadGEM3-GC31-LL (1)

MPI-ESM1-2-HR (10)

MPI-ESM1-2-LR (29)

MRI-ESM2-0 (1)

UKESM1-0-LL (3)

This list excludes GFDL-CM4 and TaiESM1 due to them lacking the complete set of atmospheric variables necessary for dynamical downscaling at the time this analysis was performed. Among this set of models and realizations that perform well on the metrics described above we do not further differentiate model skill, and choose an ensemble to achieve the remaining goals of model diversity, storyline candidates, and a representative future change signal.

4.6.1 Model diversity

Similar to other model ensemble selection studies, we emphasize model diversity—that is, sampling models with a range of simulation approaches. Overly similar models can be identified as having similar tendencies in model output (Brunner et al., 2020), or by shared code and physics schemes (shared model genealogy). Here we implement a simple version of the latter by preventing more than one variant of the same model from appearing in our ensemble. Specifically, we only allow one each of EC-Earth3/EC-Earth3-Veg, of HadGEM3-GC31-LL/UKESM1-0-LL, and of MPI-ESM1-2-LL/MPI-ESM1-2-LL, despite all six of these models performing relatively well in the evaluations. Although the atmospheric component of ACCESS-CM2 uses a version of the UK Met Office Unified Model shared by HadGEM3 and UKESM, we allow it to remain to avoid overly constraining the ensemble. The tight clustering of each of these model pairs across both the process-based and local evaluations in Figure 4-3 lends evidence to the hypothesis that structural similarities are driving similar model performance. To maximize the number of different models in the ensemble, we also only select one realization from each GCM for our ensemble, even if multiple realizations from the same GCM would satisfy the subsequent criteria.

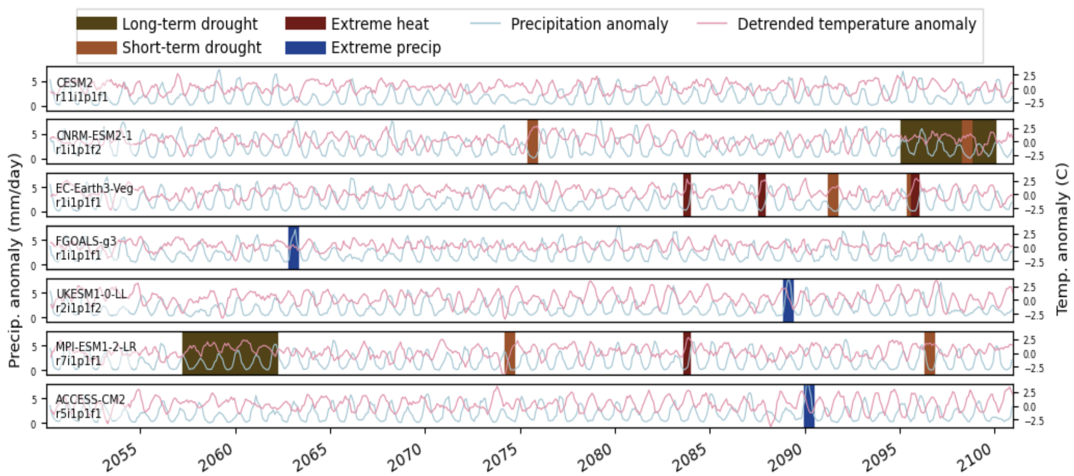

Figure 4-5. Time series of monthly precipitation and detrended temperature anomaly in California for seven selected CMIP6 simulations in the SSP3-7.0 scenario. Anomalies are defined relative to a historical period of 1950-2014. Highlighted sections indicate each of four types of extreme events.

4.6.2 Extreme event storylines

On a regional scale, the most disruptive effects of climate change will manifest in the form of extreme events. These events are expected to increase in intensity and frequency with continued global warming (Berg & Hall, 2015; Goss et al., 2020; Pendergrass et al., 2017; Swain et al., 2018). Long-term infrastructure investments in water systems and the energy grid rely on a clear understanding of extreme heat, precipitation, and drought events over the next few decades. Downscaled simulations provide the spatial and temporal resolution to study the local impact of these events, but a relatively small ensemble like ours lacks the statistical power to robustly characterize the changing frequency of rare events. One solution to this limitation is a "storyline" approach: choosing specific model realizations because they include representative extreme events that can serve as benchmarks to evaluate the resilience of adaptation plans (Shepherd et al., 2018). To support this approach, we design our ensemble to contain a variety of suitable storyline candidates while being careful not to over-represent the frequency of extreme events.

We consider four types of extreme events that have large societal impacts and high relevance for adaptation planning: extreme heat events, extreme precipitation events, severe short-term seasonal drought, and long-term multi-year drought. We define each of these events relative to the historical conditions for each model, compiling statistics from all available historical realizations. Extreme precipitation events are defined as months where the average precipitation across California exceeds the 99.99th percentile level of historical monthly averages. Short-term drought events are defined as years where the 6-month rolling average of precipitation drops below the 0.01st percentile of the historical rolling average. Long-term droughts are designated as years below the 0.15th percentile of the 5-year rolling precipitation average. Extreme heat events are defined as months where the average detrended monthly temperature exceeds the 99.99th percentile level of historical monthly averages. Future temperatures are detrended by subtracting a 10-year rolling mean, which is necessary to prevent models with the strongest mean warming signal from dominating measures of short-term heat events. These thresholds are intentionally set to define very rare events that only occur in 30-50% of all CMIP6 model realizations for the SSP3-7.0 scenario.

Counting the occurrences of these four extreme events across all the candidate model realizations listed above allows us to produce a set of all possible ensembles that proportionally represent the extremes in CMIP6 as a whole. We consider any ensemble where each extreme type is present in 2-4 of the seven ensemble members. Starting with the two members that had previously been downscaled, this yields 36 possible combinations of five additional simulations that would meet this condition. For the ensemble of seven simulations discussed below, time series highlighting the occurrences of extreme events is shown in Figure 4-5.

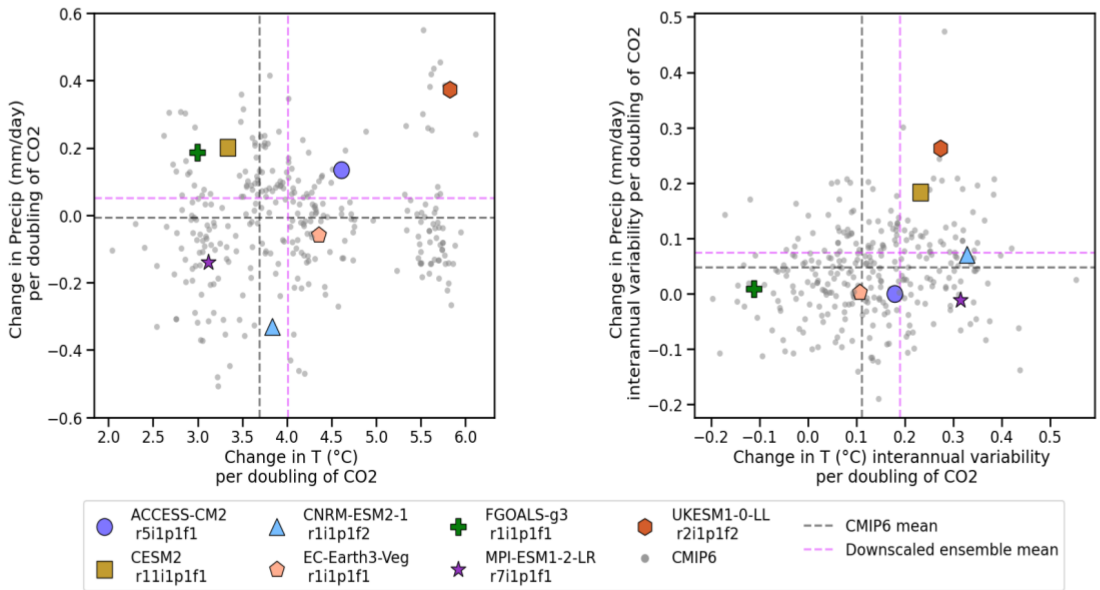

Figure 4-6. Changes in mean (left) and interannual variability (right, measured as the standard deviation of annual mean values) of precipitation and surface air temperature within California. Values are calculated as the change between a 30-year climatology at the end of the 20th century (1971-2000) and end of the 21st century (2071-2100), normalized by the number of doublings of CO_2 over the same time period. Normalization allows simulations from different emissions scenarios to share the same axes. Grey points show all simulations in CMIP6 with available data.

4.6.3 Representative Future Change

The final ensemble is selected from the possible candidates by choosing the set of simulations that most closely matches the future change signal of the full CMIP6 ensemble. We consider the change in 30-year climatologies at the end of the 20th and 21st centuries for the mean and inter-annual variability of temperature and precipitation across California. Figure 4-6 shows the spread of the selected ensemble relative to CMIP6 across these four variables. The axes in both subfigures are normalized by the number of projected CO_2 doublings between the two time periods to allow simulations from multiple SSP scenarios to be shown on the same plot. To select the ensemble that minimizes overall bias in future change signal relative to the CMIP6 ensemble mean, the total distance between each ensemble mean and the CMIP6 mean was calculated in four-dimensional space. To ensure roughly equal weighting of each of the four change measurements, they are each normalized by their standard deviation before calculating the distance. To prevent models with large ensembles from dominating the CMIP6 mean, the ensemble mean of each model is taken first, then averaged across all CMIP6 models. Matching only the mean change signal carries the risk of selecting an ensemble that is too tightly clustered in the center or consists only of extreme edge cases. In practice however, the candidate ensembles span the range in such a way that including a target standard deviation for the ensemble does not change the results appreciably.

The final ensemble shown in Figure 4-6 cannot perfectly match the mean of CMIP6, and exhibits a slight bias towards increased temperature and precipitation, as well as the associated (higher) variability. The inclusion of UKESM in the ensemble notably shifts the mean of the ensemble to an extent that selecting among the remaining realizations cannot counteract. For data practitioners concerned about overestimating a change signal due to an ensemble bias, using a global warming levels approach can mitigate the influence of hot models like UKESM, and is often recommended for working with CMIP6 data (James et al., 2017).

4.7 Discussion

This study illustrates the process of selecting an ensemble of climate simulations for studying regional climate change impacts in California. We have specifically focused on selecting models with historical skill in the northern hemisphere indicating they would provide accurate boundary conditions for dynamical downscaling. Following the framework laid out by Goldenson et al. (2023), we designed a set of process-based metrics that evaluate the most impactful large-scale meteorological patterns for our domain in California. Both the selection of the metrics for model evaluation and the process of ensemble selection have been executed with the goal of creating decision-relevant climate data for a diverse range of stakeholders. In practice, this means selecting metrics that evaluate key threats to a variety of sectors, ensuring that the ensemble contains relevant storylines of extreme events, and minimizing the bias of the ensemble as a whole.

Incorporating a complementary set of local metrics into the model evaluation illustrates both the added value and limitations of the process-based evaluation. Ultimately our results support using a combination of process-based metrics and measurements of historical local conditions for future model evaluations. Our finding that the process-based metrics were notably missing a robust characterization of decadal variability illustrates that good metric coverage is derived from both the careful choice of physical fields and the methodology used to quantify them. Our use of NMSE for process-based metrics is robust for evaluating spatial characteristics of model fields, but inherently limited in capturing temporal variability. Thus, part of the value derived from combining two sets of model evaluation metrics in our case is the diversity of numerical methodology. There may be a choice of error metric that is appropriate for measuring temporal and spatial variability across local and process-based model fields, but we found a surprising number of benefits from simply taking the intersection of best performance in two independent ranking schemes.

The ensemble selection process in this study focused on sampling future uncertainty with a diverse set of models. Constructing a seven-member ensemble allows us to sample a representative range of future conditions by selecting from models in the top half of our performance evaluation. A tradeoff was made for this ensemble to focus on only one emissions scenario (SSP3-7.0) in order to sample more individual models and simplify the interpretation of the ensemble mean. Although this comes at a cost of representing a wider range of possible climate futures, techniques like using global warming levels allow planners to consider climate impacts separately from timing dictated by a particular SSP. Our priorities for ensemble design are also dependent on the size of the ensemble and have evolved as plans for the study have grown in scope. When constrained to a very small ensemble of 2-3 realizations, the change signal relative to the CMIP6 mean becomes more important. The high warming signal of UKESM would overly skew the average of a small ensemble, making it a poor choice despite its performance in the evaluation. Curating interesting extreme storylines is a goal that could only be considered with the additional degrees of freedom in a 5-7 member ensemble. For any larger ensemble, there becomes a tradeoff between increased sampling of inter-model differences and the lower scoring performance of the remaining models. Dynamically downscaling lower-ranking models can have diminishing benefits for stakeholder-relevant decision making, but does have unique benefits for emergent constraint approaches and providing additional statistics for studying extreme events. It is difficult to formally define a threshold for when a model is too low-ranking to justify downscaling, but at a certain point resources are better put towards including additional variants or realizations of top-ranking models or sampling other emissions scenarios. The method described in this study was originally used to select an ensemble to support California's Fifth Climate Change Assessment (*Cal-Adapt*, 2023), but additional resources have since allowed additions to the ensemble that were chosen to achieve some of these secondary goals (Rahimi et al., 2024).

The analysis in the text focuses only on the CMIP6 models, but including CMIP5 models in the analysis provides additional insight into the differences between process-based and local climate evaluation. Figure 4-S6 and 4-S7 illustrate that overall CMIP6 outperforms CMIP5 on process-based metrics, echoing the findings in Simpson et al. (2020) and Pierce et al. (2022). The local metrics by comparison do not show a systematic improvement in the newer generation of models, with the caveat that the set of models included is not comprehensive for either generation. This could support the hypothesis that some models have accurate local metrics by statistical chance. It could also arise from the fact that the local metrics have been bias corrected, whereas the process-based metrics were not. If more accurate representation of physical processes in CMIP6 is producing lower mean-state bias in regional temperature and precipitation, that improvement would not be captured here. Further examining the relationship between process-based skill, mean-state bias in GCMs, and the bias in their downscaled counterparts is an area of future work that could provide useful insight for improving future model evaluations.

4.8 Summary and Conclusions

This study demonstrates the value of combining process-based and local climate metrics when evaluating GCMs for regional downscaling applications. Through a systematic evaluation of CMIP6 models focused on California, we found that neither process-based nor local climate metrics alone provided a complete picture of model performance. While process-based metrics effectively captured spatial patterns and physical mechanisms driving regional climate, they showed limitations in evaluating temporal variability, particularly at decadal scales. Conversely, local climate metrics proved adept at identifying models with accurate historical climate conditions but could potentially miss deficiencies in the representation of key atmospheric processes driving extreme events. The complementary nature of these approaches highlights the importance of a comprehensive evaluation framework that spans both spatial and temporal scales.

Our framework provides a practical methodology for selecting GCMs for regional downscaling while balancing multiple competing priorities. The process detailed here successfully identified an ensemble of models that maintain high performance scores while sampling both inter-model diversity and the range of projected future changes. The method proved particularly valuable for stakeholder-relevant applications, allowing for the intentional inclusion of extreme event storylines while maintaining ensemble-mean characteristics similar to the broader CMIP6 distribution. While this study focused on California, the approach is readily adaptable to other regions by adjusting the process-based metrics to capture regionally relevant physical mechanisms and atmospheric patterns.

Future work in model evaluation and selection must evolve to meet the growing demands of adaptation planning. In particular, better understanding the relationship between process-based skill and bias in downscaled results will help support practical decision making in the context of significant climate uncertainty. Additionally, the development of metrics that can more comprehensively evaluate temporal variability across multiple timescales would address a key limitation identified in this study. As the number of available climate simulations continues to grow, robust evaluation frameworks incorporating the physical drivers of unique regional climate impacts will become increasingly crucial for identifying the most suitable models for adaptation planning.

Supplementary Material For Chapter 4

Metric	VIF Initial	VIF final
v'v' MAM	57.5	Dropped
u850 DJF	46.2	Dropped
v'v' DJF	36.4	2.7
v'v' SON	29.2	Dropped
u850 SON	18.3	3.0
v'v' JJA	17.4	Dropped
Ψ∗ DJF	14.5	Dropped
Ψ∗ MAM	13.0	Dropped
wind shear DJF	11.3	Dropped
wind shear SON	11.3	Dropped
u850 MAM	10.8	Dropped
Ψ∗ SON	8.5	4.7
wind shear MAM	7.3	3.5
extreme precip IWV	6.5	2.5
extreme precip u250	5.9	3.2
Ψ∗ JJA	5.4	3.5

	u850 JJA	5.2	2.5
JJA	mean fixed blocking	4.5	2.8
	ENSO Skew	4.3	2.0
	SOM node12 IVT	4.1	2.6
	Santa Ana SLP	4.1	2.7
	blocking DJF	3.5	4.2
DJF	mean fixed blocking	3.2	2.9
	extreme precip SLP	2.9	2.6
	wind shear JJA	2.7	3.3
	blocking JJA	2.2	1.8
	SOM node2 IVT	2.0	1.7
	ENSO EOF1	1.8	1.6
	SOM node4 ST250	1.7	1.3
	SOM node2 ST250	1.5	1.4
	CPM	1.2	1.2

Table 4-S1. The variance inflation factor for each process-based metric before and after redundancy removal. Metrics were iteratively dropped until no metric had a VIF value over 5.

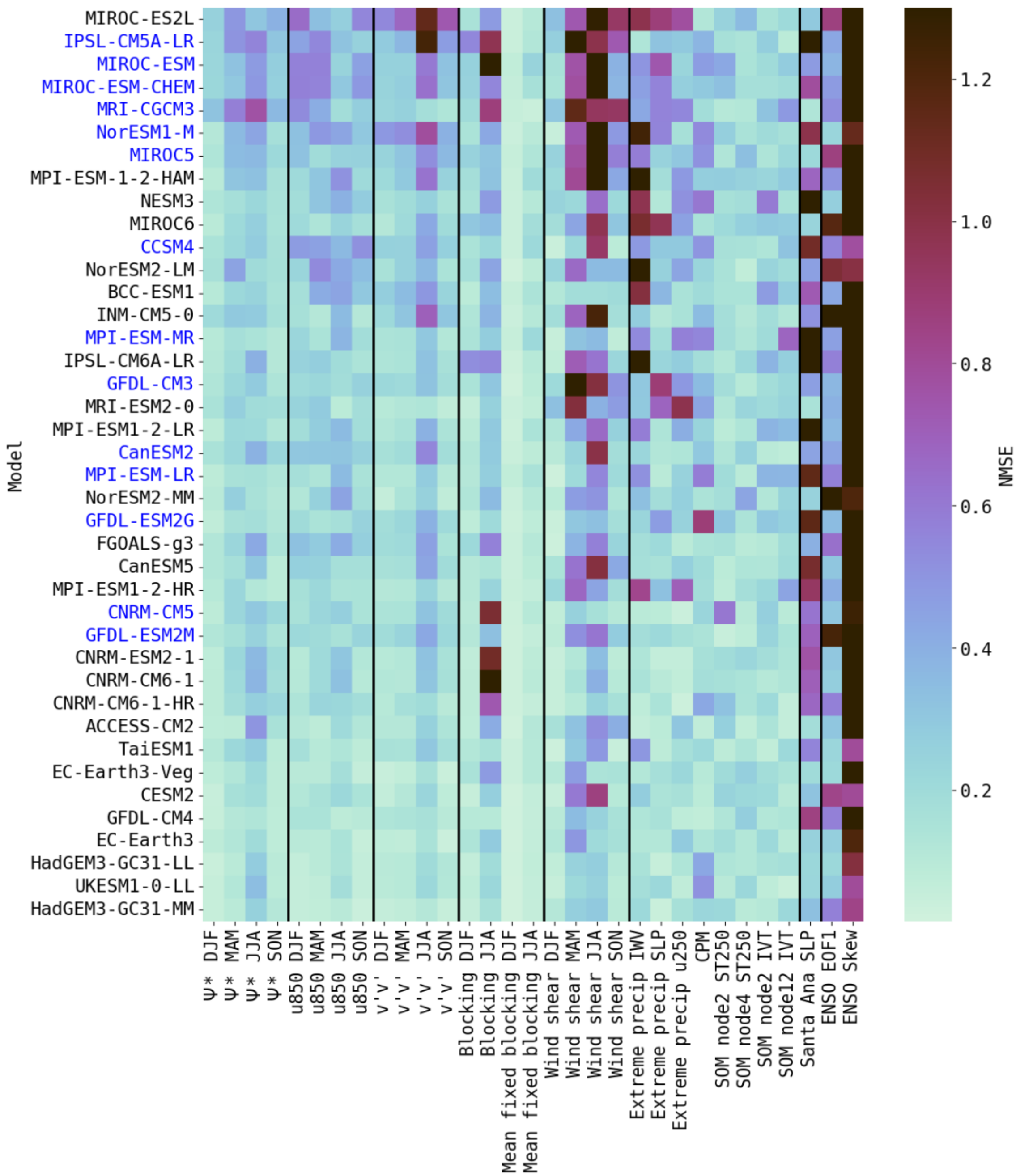

Figure 4-S2. As in Figure 4-1 but including CMIP5 models.

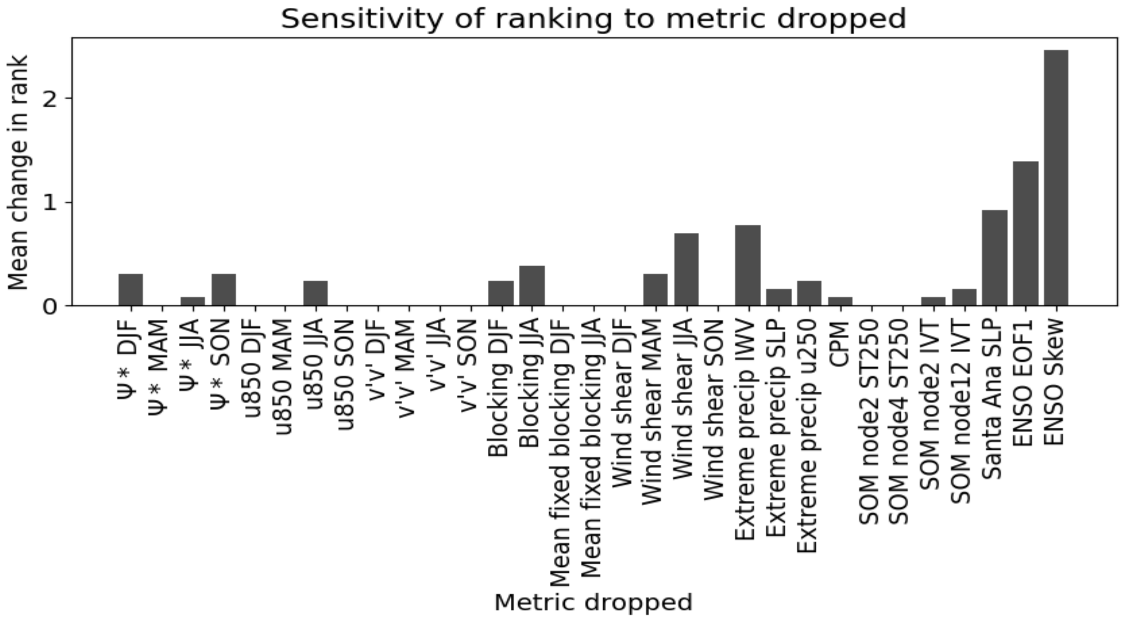

Figure 4-S1. The average change in rank across all evaluated models (CMIP5 and CMIP6) when each metric is removed in turn. The redundancy removal and calculation of the total error score is performed independently after each metric is removed.

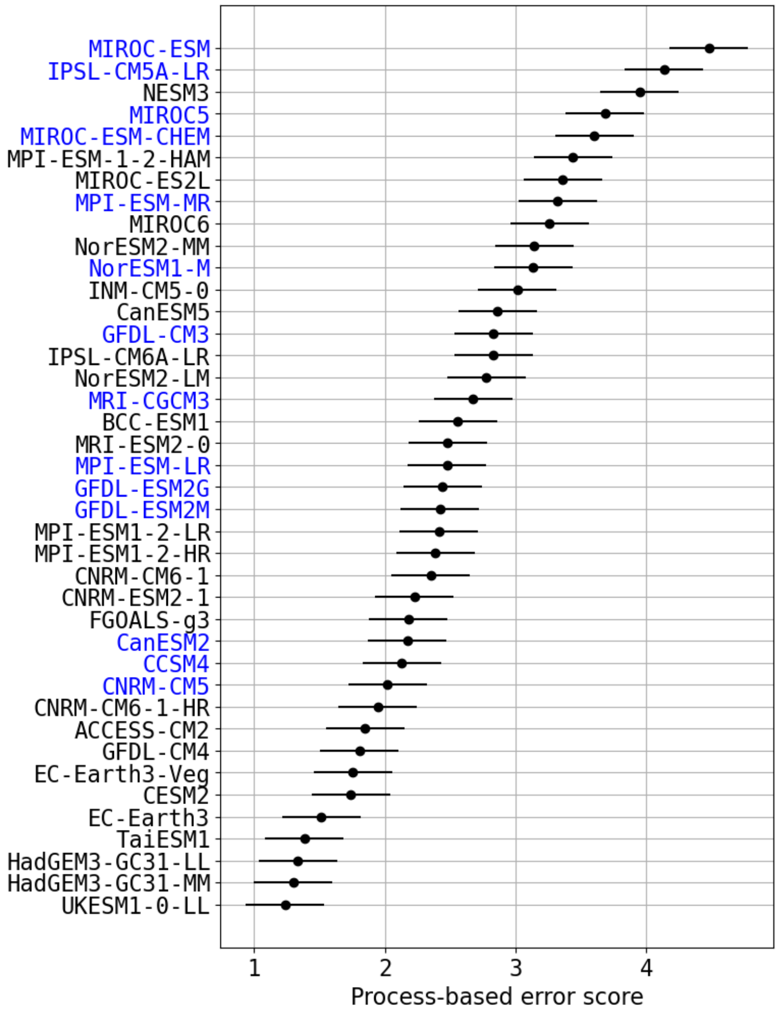

Figure 4-S3 As in Figure 4-2 but including CMIP5 models

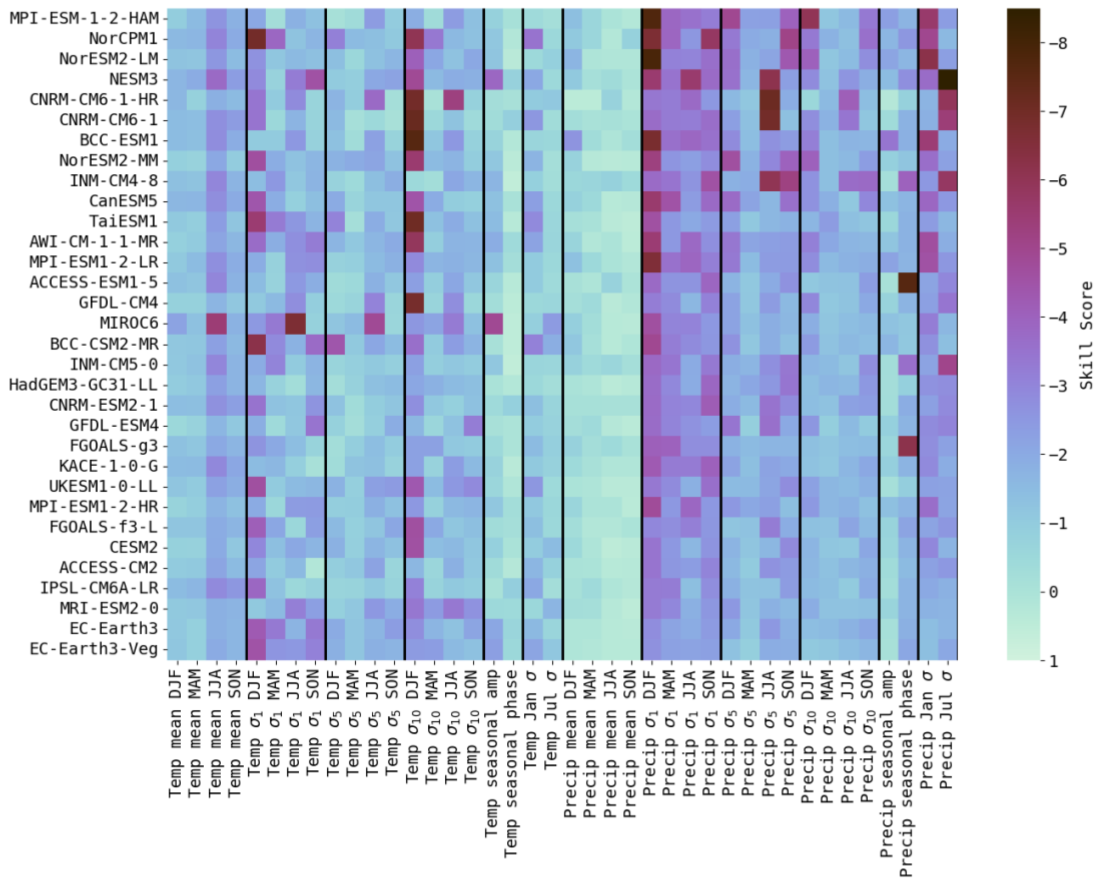

Figure 4-S4. Full matrix of local metric scores, as in Pierce (2022)

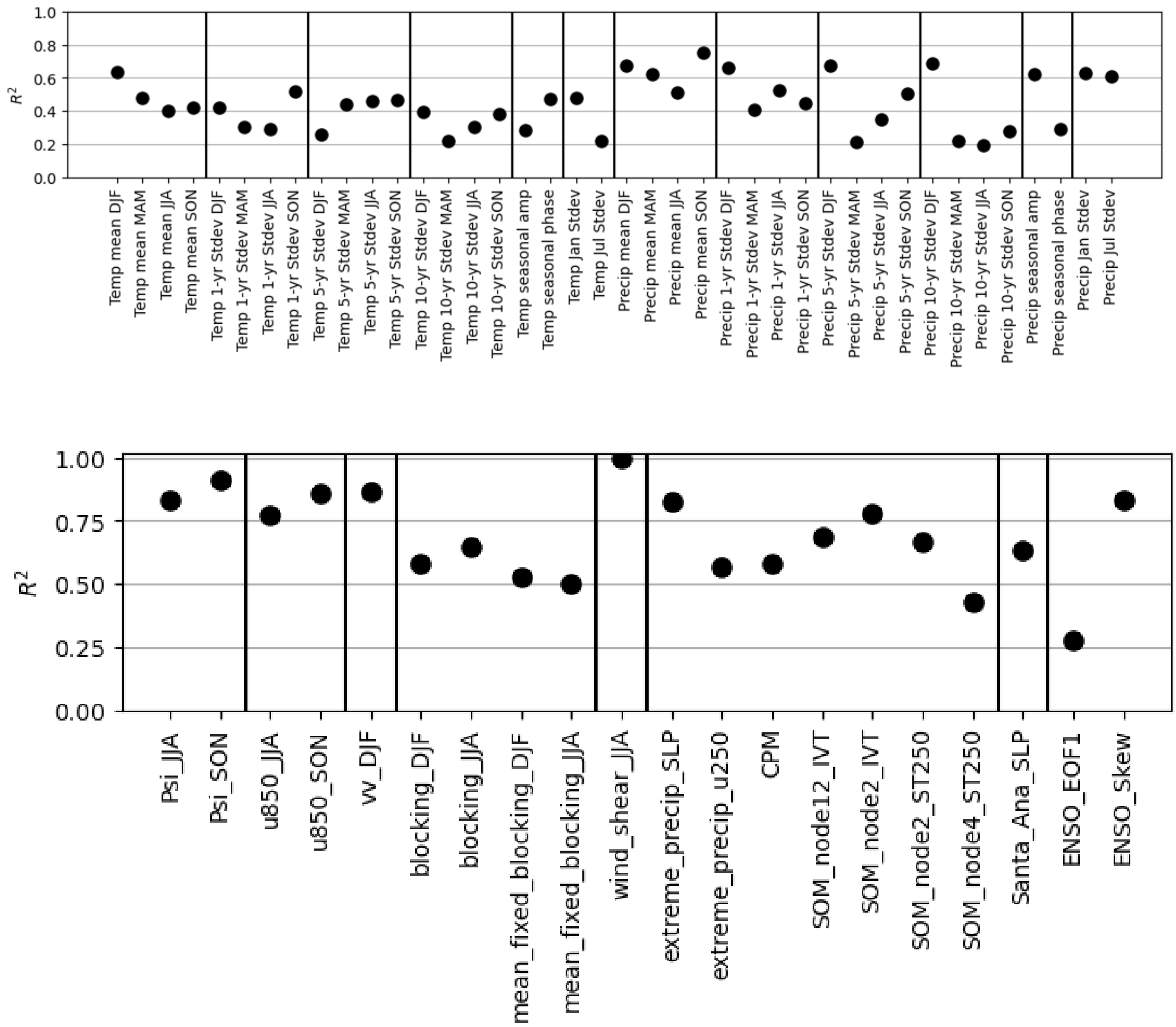

Figure 4-S5 R^2 score for regression model predicting each local metric from the full set of process-based metrics (above) or each process-based from the full set of local metrics (below).

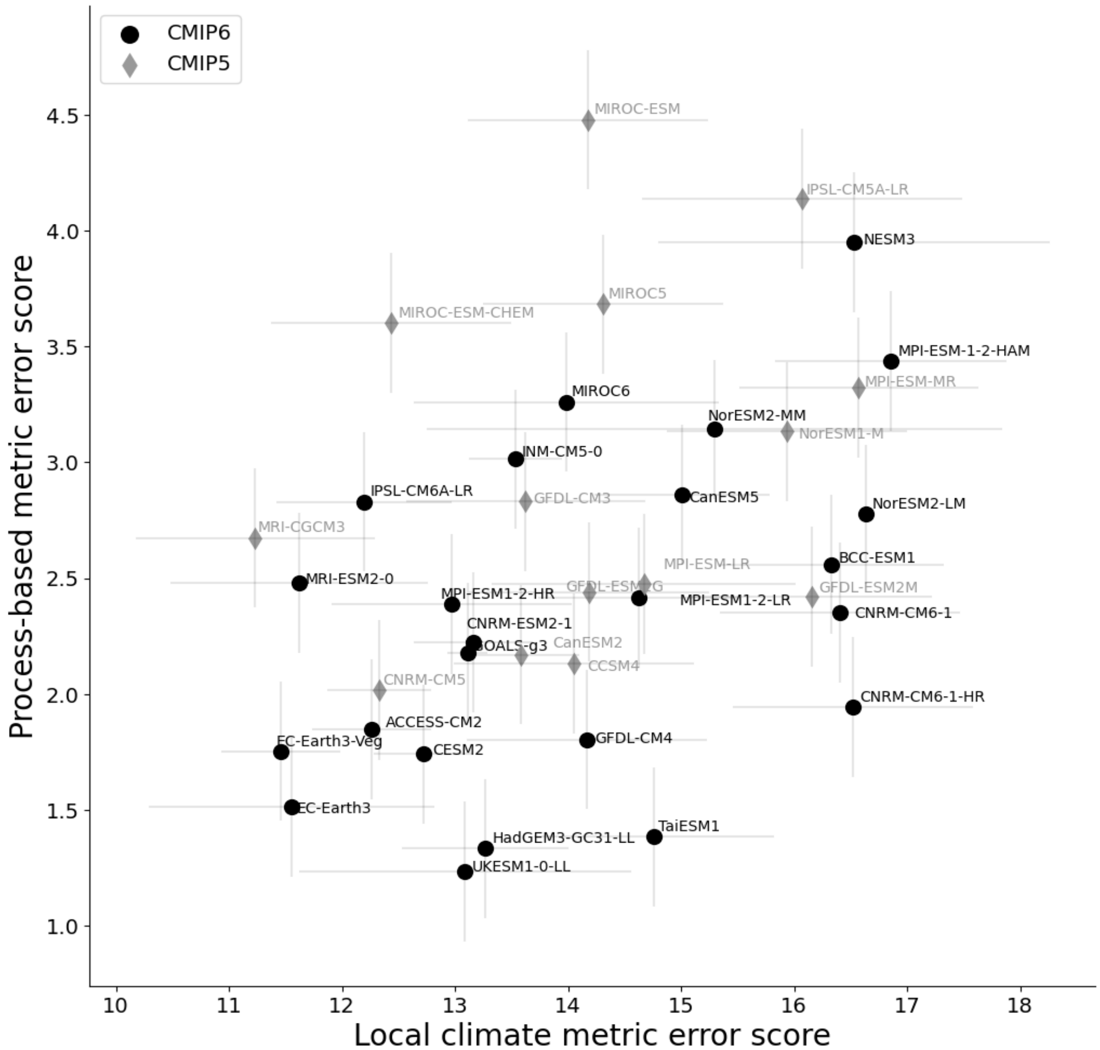

Figure 4-S6 As in Figure 4-3 but including CMIP5 models

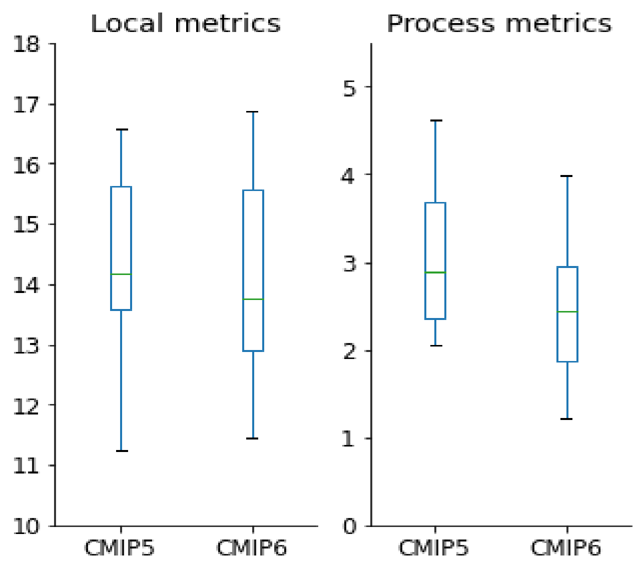

Figure 4-S7 Violin plots of the aggregated local and process metric scores across all CMIP5 and CMIP6 models evaluated.

Bibliography for Chapter 4

Abdelmoaty, H. M., Papalexiou, S. M., Rajulapati, C. R., & AghaKouchak, A. (2021). Biases Beyond the Mean in CMIP6 Extreme Precipitation: A Global Investigation. *Earth's Future*, *9*(10), e2021EF002196. https://doi.org/10.1029/2021EF002196

Berg, N., & Hall, A. (2017). Anthropogenic warming impacts on California snowpack during drought. *Geophysical Research Letters*, *44*(5), 2511–2518. https://doi.org/10.1002/2016GL072104

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics*, *11*(4), 995–1012. https://doi.org/10.5194/esd-11-995-2020

Goldenson, N., Leung, L. R., Mearns, L. O., Pierce, D. W., Reed, K. A., Simpson, I. R., Ullrich, P., Krantz, W., Hall, A., Jones, A., & Rahimi, S. (2023). Use-Inspired, Process-Oriented GCM Selection: Prioritizing Models for Regional Dynamical Downscaling. *Bulletin of the American Meteorological Society*, *104*(9), E1619–E1629. https://doi.org/10.1175/BAMS-D-23-0100.1

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, et al. "The ERA5 Global Reanalysis." Quarterly Journal of the Royal Meteorological Society 146, no. 730 (2020): 1999–2049. https://doi.org/10.1002/qj.3803.

James, Rachel, Richard Washington, Carl-Friedrich Schleussner, Joeri Rogelj, and Declan Conway. "Characterizing Half-a-degree Difference: A Review of Methods for Identifying Regional Climate Responses to Global Warming Targets." *WIREs Climate Change* 8, no. 2 (March 2017): e457. https://doi.org/10.1002/wcc.457.

Jury, M. W., Prein, A. F., Truhetz, H., & Gobiet, A. (2015). Evaluation of CMIP5 Models in the Context of Dynamical Downscaling over Europe. *Journal of Climate*, *28*(14), 5575–5582. https://doi.org/10.1175/JCLI-D-14-00430.1

Karmalkar, A. V., Thibeault, J. M., Bryan, A. M., & Seth, A. (2019). Identifying credible and diverse GCMs for regional climate change studies—case study: Northeastern United States. *Climatic Change*, *154*(3–4), 367–386. https://doi.org/10.1007/s10584-019-02411-y

Kim, Y.-H., Min, S.-K., Zhang, X., Sillmann, J., & Sandstad, M. (2020). Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. *Weather and Climate Extremes*, *29*, 100269. https://doi.org/10.1016/j.wace.2020.100269

Lorenz, R., Herger, N., Sedláček, J., Eyring, V., Fischer, E. M., & Knutti, R. (2018). Prospects and Caveats of Weighting Climate Models for Summer Maximum Temperature

Projections Over North America. *Journal of Geophysical Research: Atmospheres*, *123*(9), 4509–4526. https://doi.org/10.1029/2017JD02799

McSweeney, C. F., Jones, R. G., & Booth, B. B. B. (2012). Selecting Ensemble Members to Provide Regional Climate Change Information. *Journal of Climate*, *25*(20), 7100–7121. https://doi.org/10.1175/JCLI-D-11-00526.1

McSweeney, C. F., Jones, R. G., Lee, R. W., & Rowell, D. P. (2015). Selecting CMIP5 GCMs for downscaling over multiple regions. *Climate Dynamics*, *44*(11–12), 3237–3260. https://doi.org/10.1007/s00382-014-2418-8

Meng-Zhuo Zhang, Zhongfeng Xu, Ying Han, & Weidong Guo. (2022). Evaluation of CMIP6 models toward dynamical downscaling over 14 CORDEX domains. *Climate Dynamics*. https://doi.org/10.1007/s00382-022-06355-5

Pierce, D. W., Barnett, T. P., Santer, B. D., & Gleckler, P. J. (2009). Selecting global climate models for regional climate change studies. *Proceedings of the National Academy of Sciences*, *106*(21), 8441–8446. https://doi.org/10.1073/pnas.0900094106

Pierce, D. W., Cayan, D. R., Goodrich, J., Das, T., & Munévar, A. (2022). Evaluating Global Climate Models for Hydrological Studies of the Upper Colorado River Basin. *JAWRA Journal of the American Water Resources Association*, *58*(5), 709–734. https://doi.org/10.1111/1752-1688.12974

Priestley, M. D. K., Ackerley, D., Catto, J. L., & Hodges, K. I. (2022). Drivers of biases in the CMIP6 extratropical storm tracks. Part 1: Northern Hemisphere. *Journal of Climate*, *1*(aop), 1–37. https://doi.org/10.1175/JCLI-D-20-0976.1

Racherla, P. N., Shindell, D. T., & Faluvegi, G. S. (2012). The added value to global model projections of climate change by dynamical downscaling: A case study over the continental U.S. using the GISS-ModelE2 and WRF models. *Journal of Geophysical Research: Atmospheres*, *117*(D20). https://doi.org/10.1029/2012JD018091

Rahimi, S., Krantz, W., Lin, Y.-H., Bass, B., Goldenson, N., Hall, A., Lebo, Z. J., & Norris, J. (2022). Evaluation of a Reanalysis-Driven Configuration of WRF4 Over the Western United States From 1980 to 2020. *Journal of Geophysical Research: Atmospheres*, *127*(4), e2021JD035699. https://doi.org/10.1029/2021JD035699

Rahimi, Stefan, Lei Huang, Jesse Norris, Alex Hall, Naomi Goldenson, Will Krantz, Benjamin Bass, et al. “An Overview of the Western United States Dynamically Downscaled Dataset (WUS-D3).” Geoscientific Model Development 17, no. 6 (March 20, 2024): 2265–86. https://doi.org/10.5194/gmd-17-2265-2024.

Sanderson, B. M., Wehner, M., & Knutti, R. (2017). Skill and independence weighting for multi-model assessments. *Geoscientific Model Development*, *10*(6), 2379–2395. https://doi.org/10.5194/gmd-10-2379-2017

Virgilio, G. D. , Ji, F., Tam, E., Nishant, N., Evans, J. P., Thomas, C., Riley, M. L., Beyer, K., Grose, M. R., Narsey, S., & Delage, F. (2022). Selecting CMIP6 GCMs for CORDEX Dynamical Downscaling: Model Performance, Independence, and Climate Change Signals. *Earth's Future*, *10*(4). https://doi.org/10.1029/2021EF002625

Xue, Z., & Ullrich, P. (2021). A Comprehensive Intermediate-Term Drought Evaluation System and Evaluation of Climate Data Products over the Conterminous United States. *Journal of Hydrometeorology*, *22*(9), 2311–2337. https://doi.org/10.1175/JHM-D-20-0314.1

Appendix 1: Drivers of drought across the western US

Introduction

Over the past two decades the western US has experienced the longest and deepest widespread drought in its recorded history (Williams et al., 2022), stressing ecosystems, agriculture, and human water supply. Regions like southern California currently rely on water imported from the northern reaches of the Sierra Nevadas, the Owens river valley, and the Colorado river basin. The geographic variety of water sources means southern California can rely more heavily on one source when another is dry. Macdonald et al. (2008) coined the term "perfect drought" to describe widespread droughts like our current one which threatens all regional water supplies simultaneously. The core goal of this project is to understand the climate processes that drive widespread drought across the western US and to produce better estimates of future drought risks. This understanding will in turn be crucial to water infrastructure planning and water management practices over the coming decades.

Drought risk across the western US is driven by a combination of natural variability in precipitation and a broad warming that increases evaporative demand (Stevenson et al., 2022). The deepest and most persistent droughts are directly influenced by widely-recognized modes of climate variability like the El Niño Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO) that steer the path of moisture transport across the Pacific (Allen & Anderson, 2018; Ault et al., 2018; Choi et al., 2015; Cook et al., 2018). Understanding how these modes and their interactions will change in a warming climate is fundamental to making credible predictions of future drought risk. Unfortunately, these large-scale modes are often poorly represented in climate models (Coburn & Pryor, 2021; Le et al., 2021). Moreover, the uncertainty about how they will change in the future is very large. In this project I evaluate climate models by mapping the linkages between climate modes and drought conditions for each major watershed in the western US.

Data and Methods

The data used for this study is fourteen CMIP6 simulations downscaled to 9km over the western United States using the Weather Research and Forecasting model (WRF) between 1980 and 2100 under the Shared Socioeconomic Pathway (SSP) 3-7.0 (Rahimi et al., 2023), along with ERA5 reanalysis downscaled using the same method for the historical period of 1980-2015 (Rahimi et al., 2022).

Four watersheds of interest are defined across the western US: the Colorado river basin, the Columbia river basin, the western Sierra Nevadas including the Sacramento river basin, and the eastern Sierra Nevadas (Figure A1-1). Drought is characterized within each basin by the Standardized Precipitation Evapotranspiration Index (SPEI) on a 2-year timescale. A threshold of -1 is used to define drought conditions in a particular watershed. For each model and basin, a response-guided precursor detection method (Vijverberg 2022) is used. SST patterns associated with drought are identified by correlating the 2-year SPEI value in the spring months (MAM) with a mean of the SST in the preceding 12 months. The similarity between any two SST precursor patterns is quantified as a spatial correlation after a gaussian smoothing kernel is applied to both maps. The overall strength of the link between the SST and watershed drought is defined as the mean absolute value of the correlation at points with a significance of $\alpha < 0.5$.

Results

All models in the study show an increase in drought conditions in the Columbia and Colorado basin by the end of the century (Figure A1-3). More uncertainty exists in the simulations of the Sierra Nevada and Eastern Sierra watersheds, with some models showing no change or a decrease in drought conditions (Figure A1-3a). All but one of the downscaled WRF simulations indicate an increase in the frequency and intensity of concurrent drought across all four watersheds by the end of the century (Figure A1-2). Much of this change is driven by the warming signal captured in the SPEI metric. There is no systematic change in how tightly correlated the drought time-series are, but the downscaled WRF simulations do consistently underestimate the correlation between the Columbia river basin's drought time-series and that of the other watersheds (Figure A1-3b).

The drought precursor patterns in general display a strong similarity between the Colorado and Eastern Sierras. About half the models additionally show a similar precursor pattern for the Sierra Nevadas (Figure A1-4). The ERA data stands out as a strong outlier compared to the WRF downscaled CMIP6 simulations.

The precursor patterns at the timescales examined here capture the unique sensitivities of regional drought to large-scale forcings from major modes of climate variability within each model. This methodology reveals strong diversity among the models in both the spatial pattern and strength of that linkage. The differences in SST forcing in the precursor patterns appear to have a meaningful influence on the resulting patterns of regional drought that can partially explain tendencies towards concurrent drought in some models (Figure A1-5a). Models with an overall stronger coupling between SST patterns and regional drought tend to have less of a tendency towards concurrent drought (Figure A1-5b).

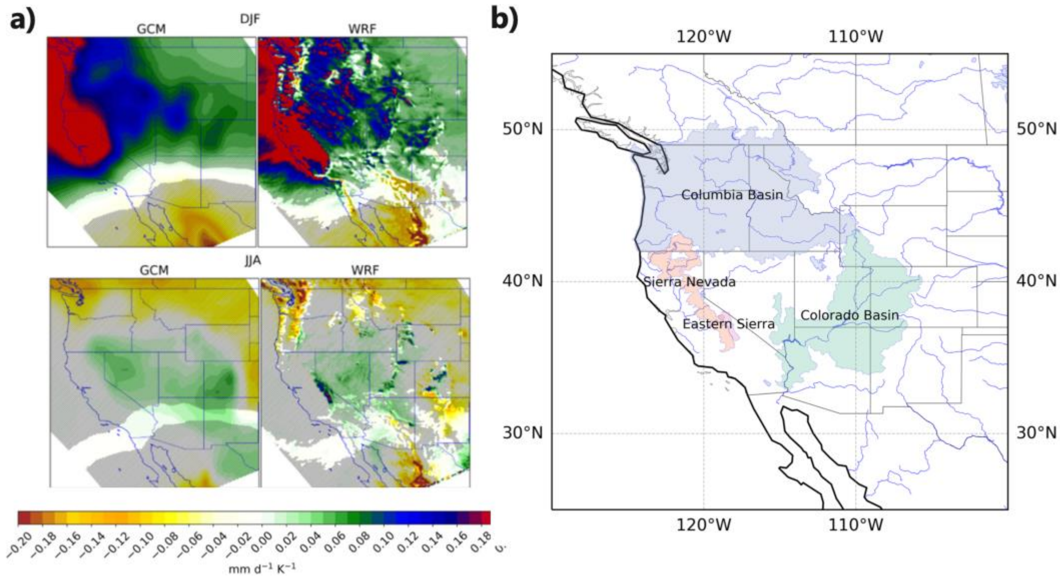

Figure A1-1. (a) Seasonal precipitation change in the WRF simulations compared to the original CMIP6 simulations, from Rahimi 2023, (b) The watershed definitions used in this study.

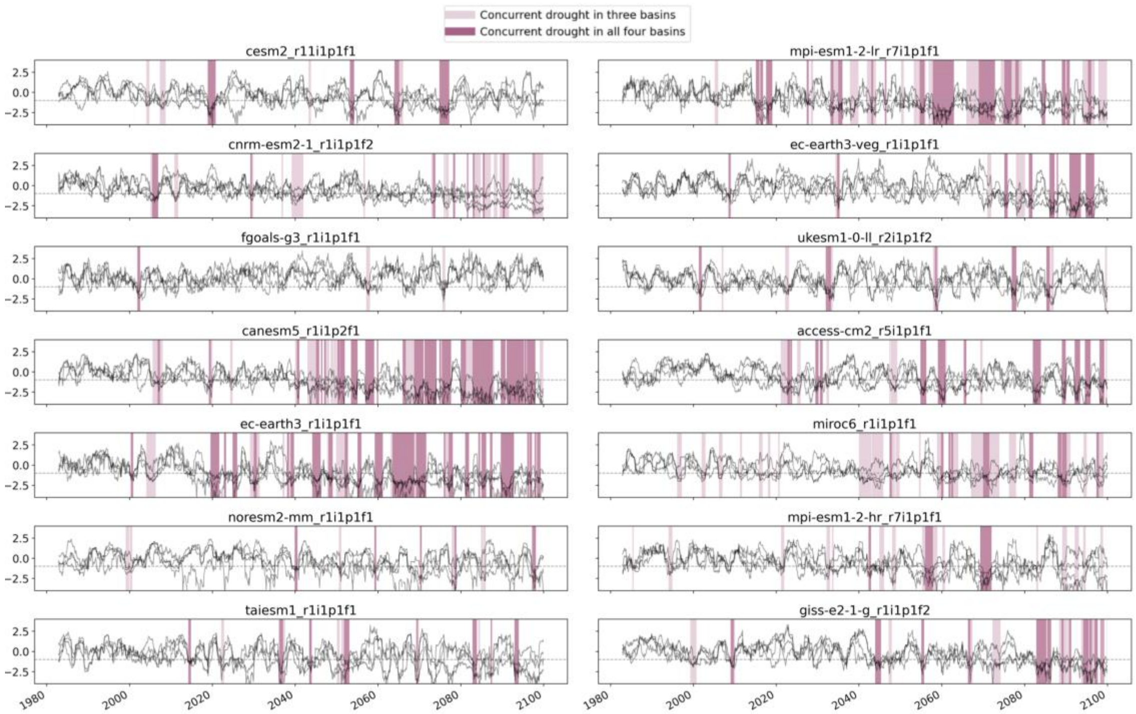

Figure A1-2. Time-series of Standardized Precipitation Evapotranspiration Index (SPEI) for in four river basins for WRF downscaled simulations. Shading indicates when three or four of the basins are simultaneously in drought condition, defined as an SPEI below -1. SPEI is normalized to the period of 1980-2015 for all models.

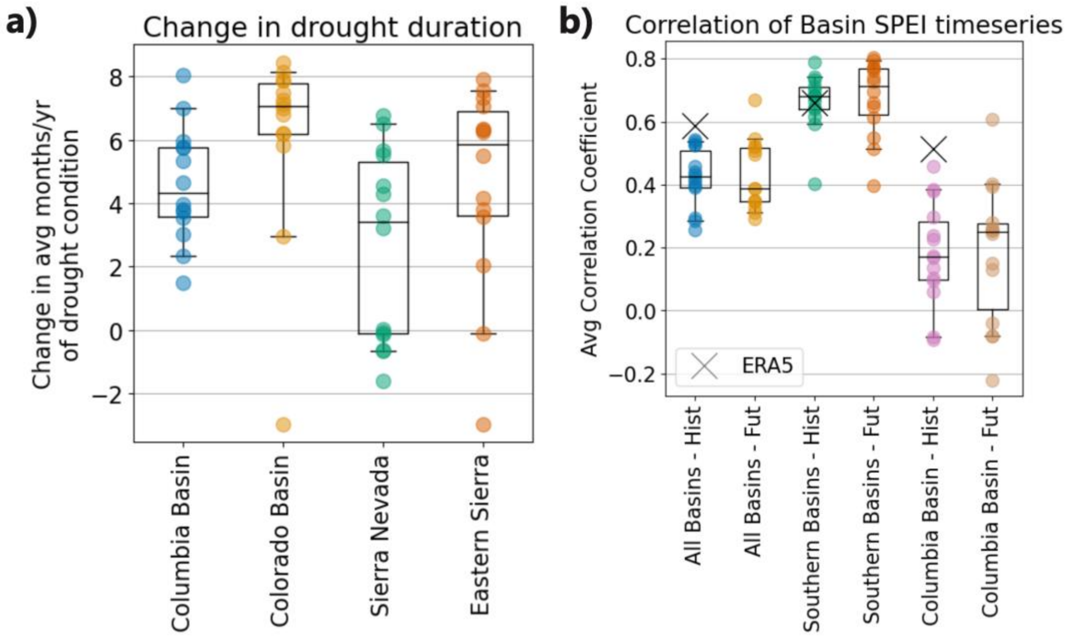

Figure A1-3. (a) The change in average months per year in drought conditions (SPEI < -1) between the historical period (1980-2010) and the end of the century (2070-2100), (b) The average of the correlation coefficient between the SPEI time-series of all four basins, the three southernmost basins (Colorado, Sierra Nevada, Eastern Sierra), and between the Columbia basin and each of the other three basins.

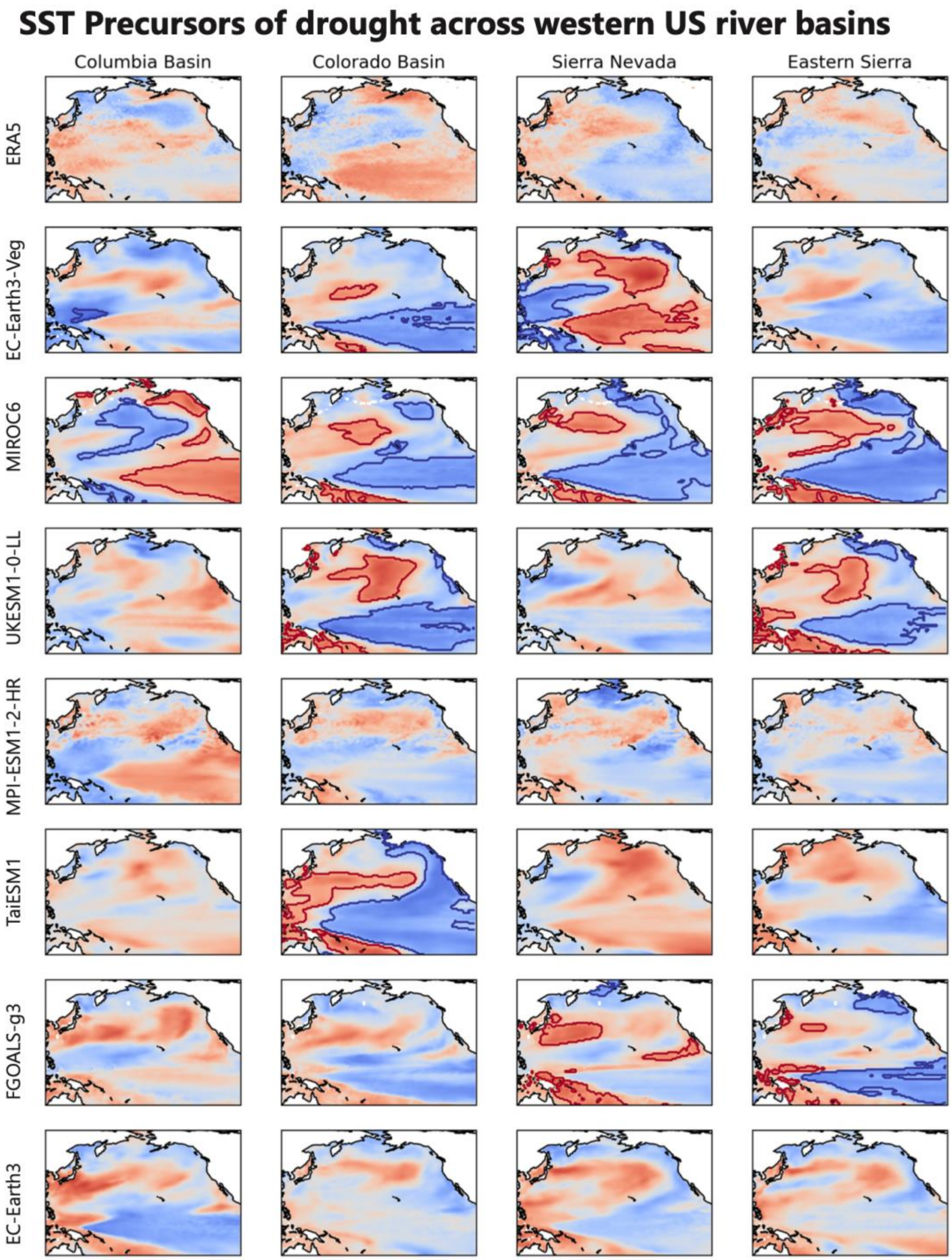

Figure A1-4. Time lagged correlation between CMIP6 model SST and basin SPEI from WRF downscaled model in the historical period (1980-2015). Clustered areas of significant correlation ($\alpha < 0.1$, corrected for field significance using false discovery rate) are outlined.

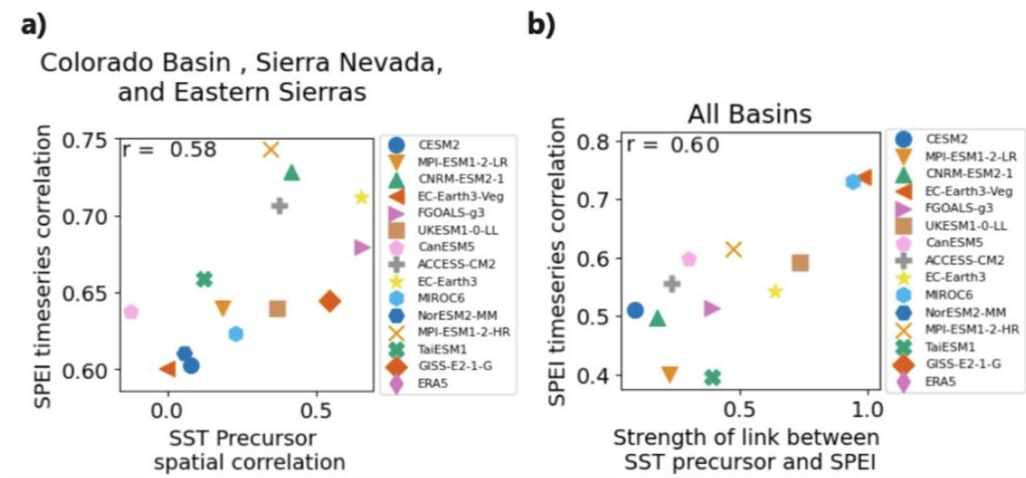

Figure A1-5. Correlation between the SST precursor spatial correlation and the SPEI timeseries correlation for the southern basins (a) and correlation between the normalized overall teleconnection strength and SPEI timeseries correlation for all basins (b).

Bibliography for Appendix 1

Allen, R. J., & Anderson, R. G. (2018). 21st century California drought risk linked to model fidelity of the El Niño teleconnection. *Npj Climate and Atmospheric Science*, *1*(1), Article 1. https://doi.org/10.1038/s41612-018-0032-x

Ault, T. R., George, S. S., Smerdon, J. E., Coats, S., Mankin, J. S., Carrillo, C. M., Cook, B. I., & Stevenson, S. (2018). A Robust Null Hypothesis for the Potential Causes of Megadrought in Western North America. *Journal of Climate*, *31*(1), 3–24. https://doi.org/10.1175/JCLI-D-17-0154.1

Choi, I., Li, B., Zhang, H., & Li, Y. (2015). Modelling space-time varying ENSO teleconnections to droughts in North America. *Stat*, *4*(1), 140–156. https://doi.org/10.1002/sta4.85

Coburn, J., & Pryor, S. C. (2021). Differential Credibility of Climate Modes in CMIP6. *Journal of Climate*, *34*(20), 8145–8164. https://doi.org/10.1175/JCLI-D-21-0359.1

Cook, B. I., Williams, A. P., Mankin, J. S., Seager, R., Smerdon, J. E., & Singh, D. (2018). Revisiting the Leading Drivers of Pacific Coastal Drought Variability in the Contiguous United States. *Journal of Climate*, *31*(1), 25–43. https://doi.org/10.1175/JCLI-D-17-0172.1

Le, P. V. V., Guilloteau, C., Mamalakis, A., & Foufoula-Georgiou, E. (2021). Underestimated MJO Variability in CMIP6 Models. *Geophysical Research Letters*, *48*(12), e2020GL092244. https://doi.org/10.1029/2020GL092244

Rahimi, S., Huang, L., Norris, J., Hall, A., Goldenson, N., Krantz, W., Bass, B., Thackeray, C., Lin, H., & Chen, D. (2023). An overview of the Western United States dynamically downscaled dataset (WUS-D3). *Geoscientific Model Development Discussions*, *2023*, 1–56.

Rahimi, S., Krantz, W., Lin, Y.-H., Bass, B., Goldenson, N., Hall, A., Lebo, Z. J., & Norris, J. (2022). Evaluation of a Reanalysis-Driven Configuration of WRF4 Over the Western United States From 1980 to 2020. *Journal of Geophysical Research: Atmospheres*, *127*(4), e2021JD035699. https://doi.org/10.1029/2021JD035699

Appendix 2: Model evaluation and bias for dynamical downscaling

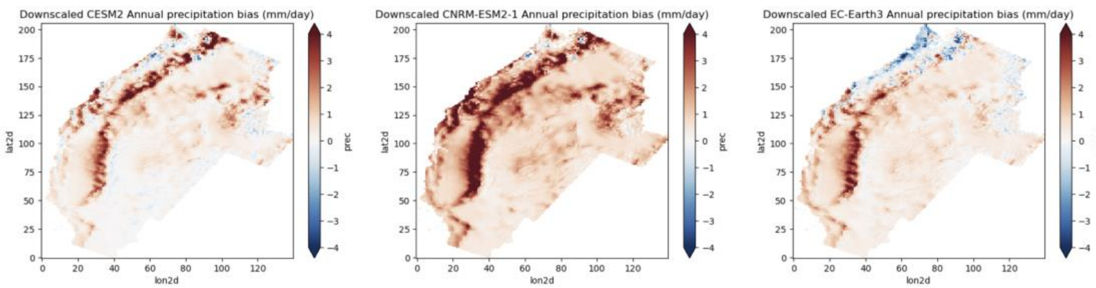

Figure A2-1. Precipitation bias in downscaled WRF simulations

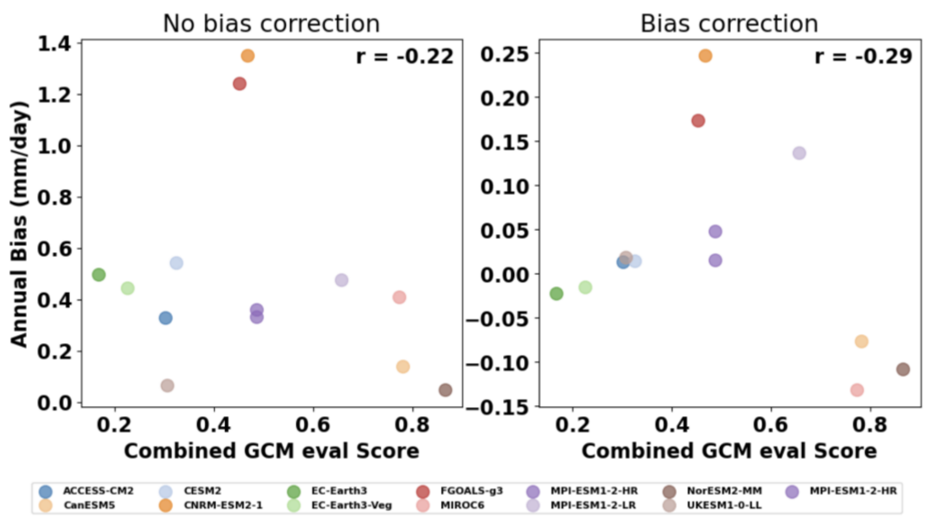

Figure A2-2. Combined GCM evaluation score and precipitation bias for bias corrected and non bias corrected WRF simulations

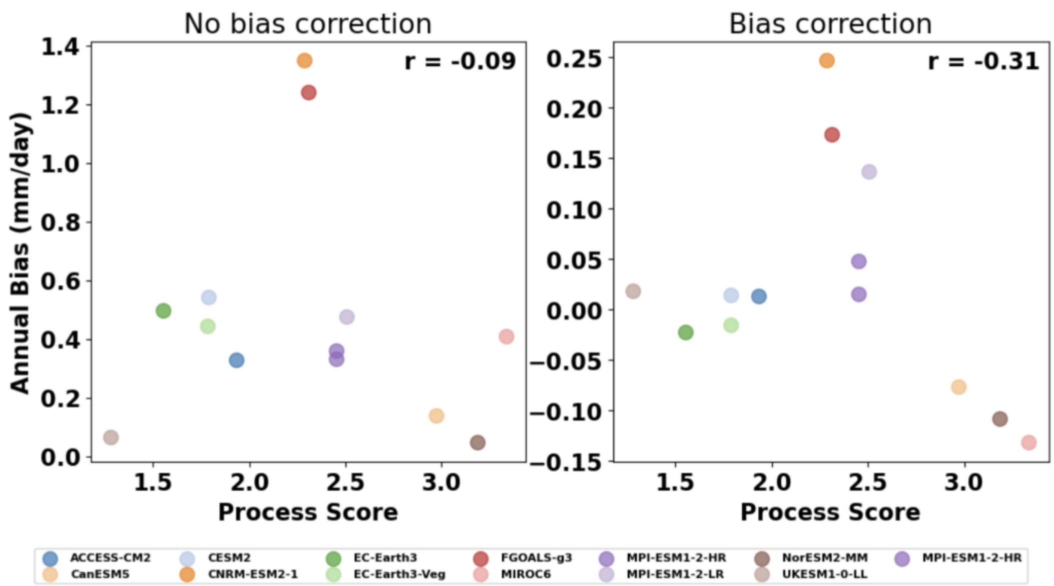

Figure A2-3. Process-based GCM evaluation score and precipitation bias for bias corrected and non bias corrected WRF simulations

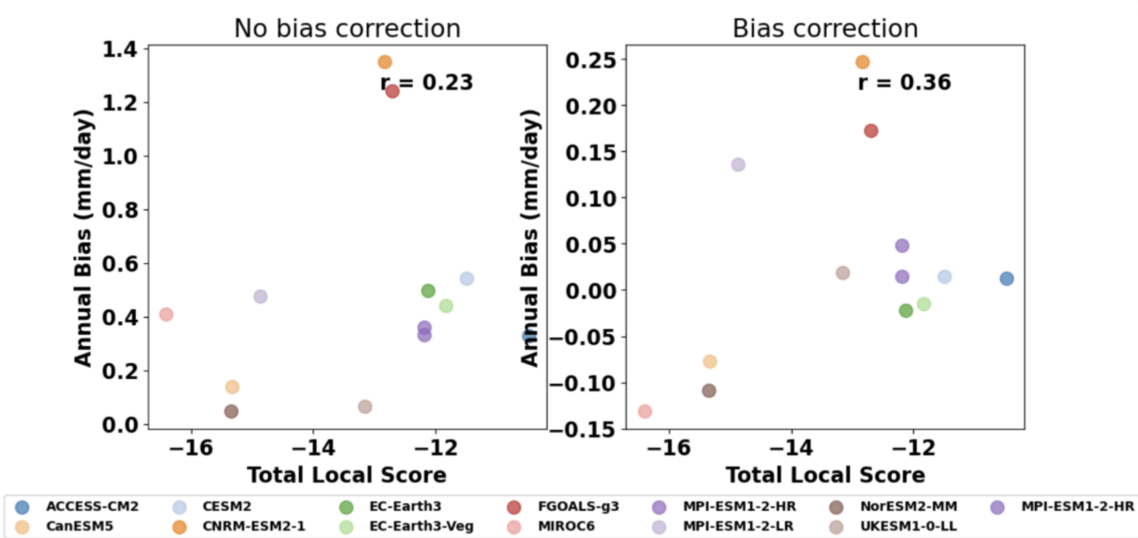

Figure A2-4. Local GCM evaluation score and precipitation bias for bias corrected and non bias corrected WRF simulations

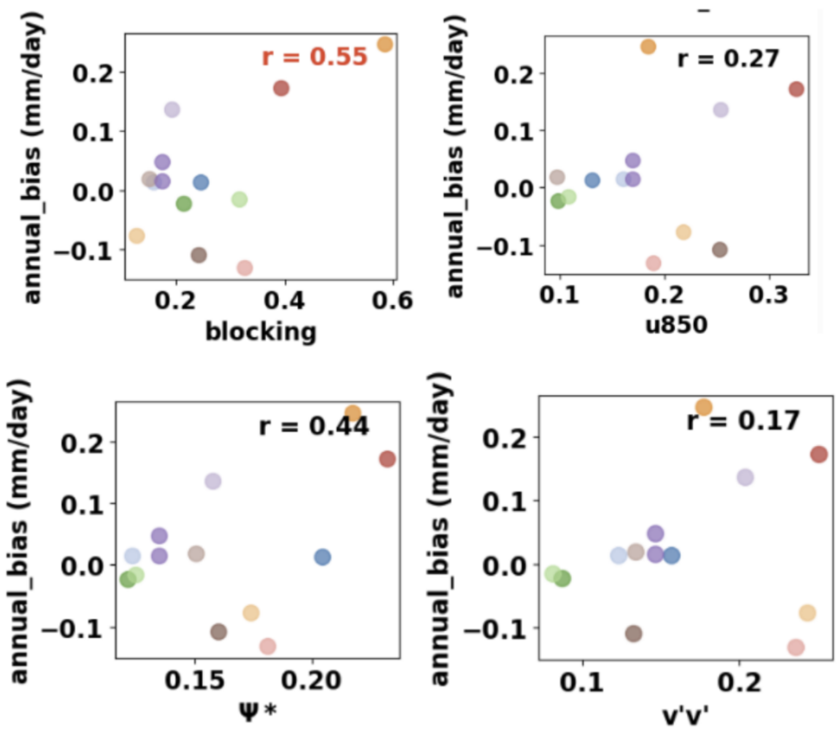

Figure A2-5. Individual process metrics with the strongest correlation to annual precipitation bias