Title

Quantifying uncertainty in precipitation climatology, twenty-first century change, and teleconnections in global climate models

Author

Langenbrunner, Baird

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Quantifying uncertainty in precipitation climatology, twenty-first century change, and teleconnections in global climate models

A dissertation submitted in partial satisfaction
of the requirements for the degree Doctor of Philosophy
in Atmospheric and Oceanic Sciences

by

Baird Grant Langenbrunner

2015

ABSTRACT OF THE DISSERTATION

Quantifying uncertainty in precipitation climatology, twenty-first century change, and teleconnections in global climate models

by

Baird Grant Langenbrunner

Doctor of Philosophy in Atmospheric and Oceanic Sciences

University of California, Los Angeles, 2015

Professor J. David Neelin, Chair

The ability of global climate models (GCMs) to simulate climatological precipitation and other features of the hydrological cycle accurately is acceptable by some metrics, especially at large scales. Regionally, however, there can be substantial discrepancy in a multi-model ensemble, both in the annual or seasonal historical precipitation climatology as well as in end-of-century changes. Characterizing this intermodel spread and identifying leading uncertainty patterns and underlying physical pathways is important in constraining climatological biases and projections of future change. This dissertation looks at three aspects of precipitation uncertainty in ensembles.

First, El Niño-Southern Oscillation (ENSO) teleconnections are analyzed in an atmosphere-only ensemble to gauge the ability of atmospheric components of GCMs to reproduce ENSO precipitation teleconnections. This serves as a test for how well models simulate the atmospheric response to sea surface temperature forcing in the immediate ENSO vicinity, as well as how accurately they reproduce the large-scale tropical-to-midlatitude dynamics leading to teleconnected precipitation. While individual models have difficulty in simulating the exact spatial pattern of teleconnections, they demonstrate skill in regional amplitude measures and sign agreement of the precipitation teleconnections at the grid point level, which lends value to the use of such measures in global warming projections.

Next, objective spatial analysis techniques are applied to a fully-coupled GCM ensemble in order to visualize patterns of uncertainty in end-of-century precipitation changes and in the historical

climatology. Global patterns are considered first, with the tropics exerting a clear dominance in intermodel spread, mainly within zones of deep convection or along convective margins. Regional domains are considered second, with a focus on the wintertime midlatitude Pacific storm track. A key region of end-of-century precipitation change uncertainty is identified at the terminus of the storm track, and large-scale circulation processes related to model differences in upper-level jet increases are found to play a role. These results help pinpoint a source of intermodel spread in projected precipitation changes along the North American west coast, especially for the Southern California region.

Last, an existing perturbed physics ensemble is examined in order to understand the parameter sensitivity of climatological precipitation and other fields. This ensemble consists of integrations in which four parameters in the deep convection scheme were systematically varied. Models of parameter dependence are constructed for precipitation, and this process—termed metamodeling—is a computationally cheap alternative to brute-force sampling of parameter space in the GCM. A quadratic metamodel performs generally well but fails to capture sensitive regions of high nonlinearity for certain parameter ranges. A second metamodel is constructed by combining an approach from the engineering literature with the spatial uncertainty patterns used above, and it proves adept at capturing sensitive regions where its quadratic counterpart fails. Finally, when more than one field is optimized simultaneously, it is often the case that a set of parameter values that optimizes one field can degrade performance in another. Concepts from multiobjective optimization are used to quantify these tradeoffs.

The dissertation of Baird Grant Langenbrunner is approved.

Alexander Dean Hall

Glen Michael MacDonald

Carlos R. Mechoso

J. David Neelin, Committee Chair

University of California, Los Angeles

2015

DEDICATION

to my parents, Mary and John,

and to my brother, Adam

Table of contents

List of figures

ACKNOWLEDGEMENTS

Chapter 1 contains versions of figures from work published in Sheffield et al. (2013a), Sheffield et al. (2013b), Maloney et al. (2014), and Neelin et al. (2013). These publications were supported in part by the NOAA Climate Program Office Modeling, Analysis, Predictions and Projections (MAPP) Program as part of the CMIP5 Task Force.

Chapter 2 is a version of Langenbrunner and Neelin (2013), which was supported in part by NSF grant AGS1102838 and by the NOAA Climate Program Office Modeling, Analysis, Predictions, and Projections (MAPP) Program under grant NA11OAR4310099, as part of the CMIP5 Task Force.

Chapter 3 is a version of Langenbrunner et al. (2015), and I would especially like to thank coauthors Benjamin R. Lintner and Bruce T. Anderson for their help in shaping this study. BL and JDN were supported in part by NSF grants AGS-1102838 and AGS-1540518 as well as by NOAA grants NA11OAR4310099 and NA14OAR4310274. BRL was supported in part by NSF grant AGS-1312865, and BTA was supported in part by DOE grant DE-SC0004975.

Chapter 4 is a draft of a manuscript in progress to be coauthored with J. David Neelin. For this work, BL and JDN were supported by NSF grants AGS-1102838 and AGS-1540518 and NOAA grants NA11OAR4310099 and NA14OAR4310274. I would also like to thank Diana N. Bernstein for providing the raw model data used in this chapter.

The first three chapters contain versions of work published in *Journal of Climate*, © Copyright 2013–2015 American Meteorological Society (AMS). I would like to acknowledge a Chancellor's Fellowship and a Dissertation Year Fellowship from the UCLA Graduate Division, both of which helped fund the work in this dissertation. I would also like to thank Helen and Lance Bosart for their kindness and generosity to me and other students in the AOS department. I am grateful to Joyce E. Meyerson for her wizardry at plotting and data analysis. Finally, I thank J. David Neelin for his patience and guidance as a coauthor and advisor.

Baird Grant Langenbrunner

Los Angeles, California

December, 2015

Baird Grant Langenbrunner

EDUCATION

University of California, Los Angeles, CA

M.S. Atmospheric and Oceanic Sciences — June 2013

Brown University, Providence, RI

Sc.B. Geophysics — May 2009

FELLOWSHIPS AND AWARDS

Dissertation Year Fellowship (UCLA Graduate Division) — Fall 2014–Fall 2015

Morris Neiberger Award – excellence in graduate teaching (UCLA AOS) — Fall 2014

AOS Fellowship – contribution to departmental outreach (UCLA AOS) — Fall 2013

Brian Bosart Award – for service to students and the department (UCLA AOS) — Fall 2012

Chancellor's Prize – awarded on basis of academic merit (UCLA Graduate Division) — 2010–2013

NSF Research Experience for Undergraduates (REU) Fellowship (CSU Fort Collins) Summer 2008

Undergraduate Teaching and Research Award (Brown Dept. of Geological Sciences) Summer 2008

CRC Press Award for Achievement in Undergraduate Chemistry (Brown Dept. of Chemistry) 2007

PUBLICATIONS

Langenbrunner, B., J. D. Neelin, B. R. Lintner, and B. T. Anderson, 2015: Patterns of precipitation change and climatological uncertainty among CMIP5 models, with a focus on the midlatitude Pacific storm track. *Journal of Climate*, **28**, 7858–7872.

Anderson, B. T., B. R. Lintner, B. Langenbrunner, J. D. Neelin, E. Hawkins, and J. Syktus, 2015: Sensitivity of terrestrial precipitation trends to the structural evolution of sea surface temperatures. *Geophysical Research Letters*, **42 (4)**, 1190–1196.

Berg, N., A. Hall, F. Sun, S. Capps, D. Walton, B. Langenbrunner, and J. D. Neelin, 2014: Twenty-first-century precipitation changes over the Los Angeles region. *Journal of Climate*, **28 (2)**, 401–421.

Maloney, E. D., S. J. Camargo, E. Chang, B. Colle, R. Fu, K. L. Geil, Q. Hu, X. Jiang, N. Johnson, K. B. Karnauskas, J. Kinter, B. Kirtman, S. Kumar, B. Langenbrunner, K. Lombardo, L. N. Long, A. Mariotti, J. E. Meyerson, K. C. Mo, J. D. Neelin, Z. Pan, R. Seager, Y. Serra,

A. Seth, J. Sheffield, J. Stroeve, J. Thibeault, S.-P. Xie, C. Wang, B. Wyman, and M. Zhao, 2014: North american climate in CMIP5 experiments: Part III: Assessment of Twenty-first Century projections. *Journal of Climate*, **27 (6)**, 2230–2270.

Sheffield, J., S. J. Camargo, R. Fu, Q. Hu, X. Jiang, N. Johnson, K. B. Karnauskas, S. T. Kim, J. Kinter, S. Kumar, B. Langenbrunner, E. Maloney, A. Mariotti, J. E. Meyerson, J. D. Neelin, S. Nigam, Z. Pan, A. Ruiz-Barradas, R. Seager, Y. L. Serra, D.-Z. Sun, C. Wang, S.-P. Xie, J.-Y. Yu, T. Zhang, and M. Zhao, 2013b: North american climate in CMIP5 experiments. Part II: Evaluation of historical simulations of intraseasonal to decadal variability. *Journal of Climate*, **26 (23)**, 9247–9290.

Sheffield, J., A. P. Barrett, B. Colle, D. Nelun Fernando, R. Fu, K. L. Geil, Q. Hu, J. Kinter, S. Kumar, B. Langenbrunner, K. Lombardo, L. N. Long, E. Maloney, A. Mariotti, J. E. Meyerson, K. C. Mo, J. David Neelin, S. Nigam, Z. Pan, T. Ren, A. Ruiz-Barradas, Y. L. Serra, A. Seth, J. M. Thibeault, J. C. Stroeve, Z. Yang, and L. Yin, 2013a: North american climate in CMIP5 experiments. Part I: Evaluation of historical simulations of continental and regional climatology. *Journal of Climate*, **26 (23)**, 9209–9245.

Neelin, J. D., B. Langenbrunner, J. E. Meyerson, A. Hall, and N. Berg, 2013: California winter precipitation change under global warming in the Coupled Model Intercomparison Project phase 5 ensemble. *Journal of Climate*, **26 (17)**, 6238–6256.

Langenbrunner, B., and J. D. Neelin, 2013: Analyzing ENSO teleconnections in CMIP models as a measure of model fidelity in simulating precipitation. *Journal of Climate*, **26 (13)**, 4431–4446.

Chapter 1

Introduction

1.1 Background and overview

Global climate models (GCMs) show widespread uncertainty in their ability to model regional precipitation (P) characteristics. GCM ensembles often agree on large-scale, average P features, though this doesn't always translate to consensus at the local or model grid point level. This regional uncertainty lurks both in seasonal climatologies and in end-of-century global warming P projections, and it is useful to seek observational constraints on this uncertainty in the current climate of GCMs and reduce barriers to prediction of future hydrological cycle change.

Intermodel disagreement over P features is at its core a very complex issue. GCM "ensembles of opportunity" are groups of models built independently by different modeling groups, so spread within these ensembles is a result of model structure, resolution, and numerical schemes, as well as differences in internal variability in the coupled ocean-atmosphere system. Another source of intermodel spread is differences in how GCMs simulate the dynamic and thermodynamic processes that interact to produce rain and snow. These interactions require models to bridge large-scale dynamics with sub-grid scale physics and parameterizations, which themselves can differ widely among models.

This dissertation discusses published and to-be-published work on intermodel uncertainty, with a major focus on historical P climatologies and end-of-century changes in model ensembles. The majority of the model simulation data is from the historical and global warming simulations available through the Coupled Model Intercomparison Project phase 5 (CMIP5) as well as the previous phase 3 (CMIP3).

Chapter 2 presents results from work done in Langenbrunner and Neelin (2013), where El Niño-Southern Oscillation (ENSO) P teleconnections were analyzed within the CMIP3 and CMIP5 ensemble archives. These teleconnections are calculated from atmosphere-only runs of the models—coupled GCMs driven by identical observed sea surface temperatures (SSTs)—and compared to observations. In using atmosphere-only simulations, one is able to probe the accuracy of the atmosphere alone in responding to SST anomalies and accurately producing the large-scale dynamics that lead to remote P anomalies.

Chapter 3 discusses work done in Langenbrunner et al. (2015), in which objective matrix decomposition techniques are employed to visualize patterns of model uncertainty in P climatology and end-of-century change in the CMIP5 model ensemble. Further analysis is done to relate model spread in P characteristics to spread in the larger-scale circulation and SSTs, and the contribution of internal variability in the historical and pre-industrical control simulations is quantified.

Chapter 4 discusses current work in preparation to be published. Distinct from the above work, this final chapter uses an available perturbed physics ensemble to explore model uncertainty related to sub-grid scale physics parameterizations. The ensemble is described in Bernstein and Neelin (2015, submitted) and is composed of branch runs from the National Center for Atmospheric Research (NCAR) Community Earth System Model version 1.1 (CESM1.1), in which parameters related to deep convection and moist processes are systematically changed to see the range of climates that will result. These simulations are used to visualize uncertainty patterns akin to those of Chapter 3, and these uncertainty patterns are then employed to create models for the parameter sensitivity of P and other climatological fields.

The remainder of the current chapter discusses published work to which I have contributed analysis, as well as some of the major conclusions of those papers.

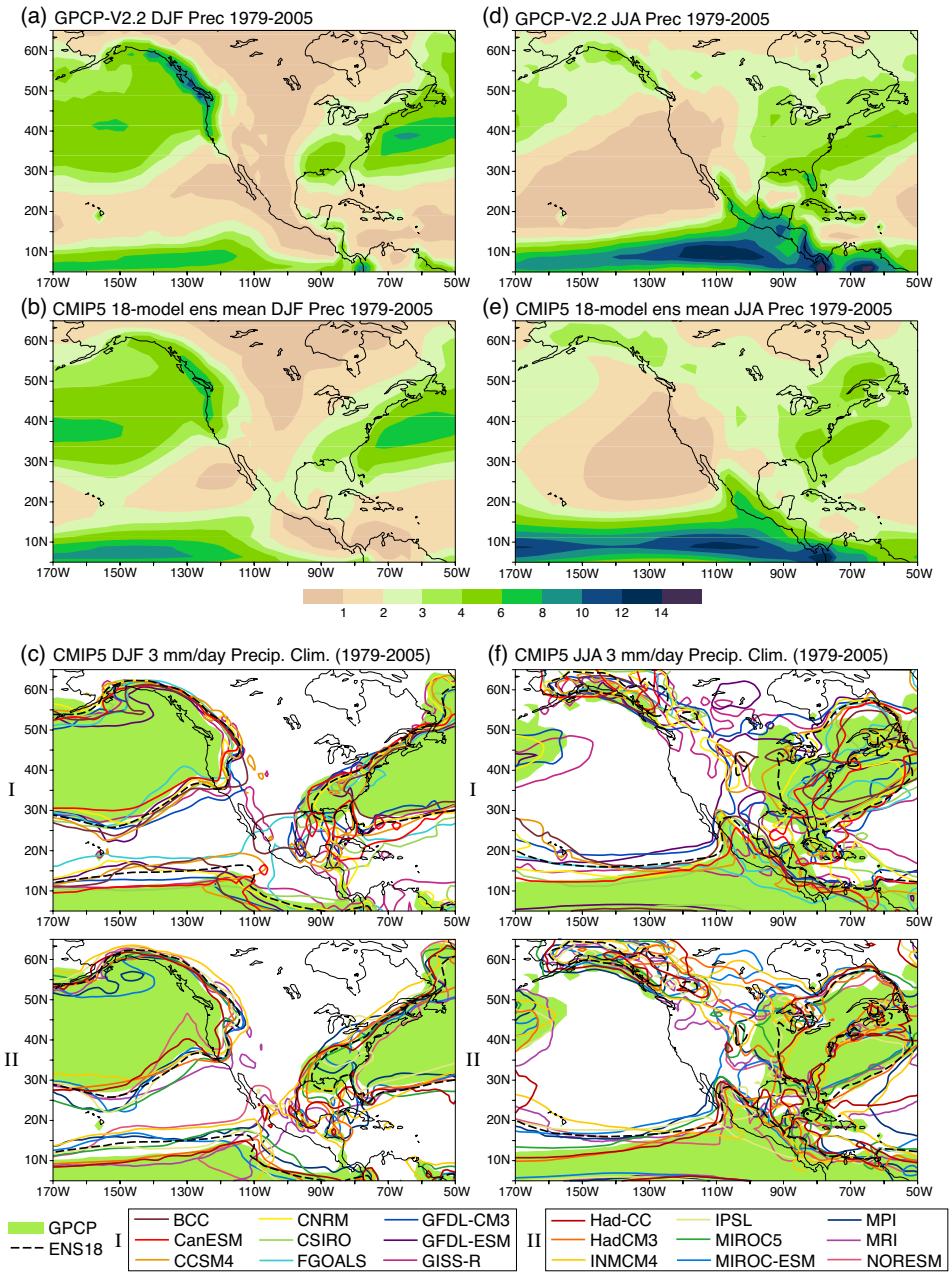

Figure 1.1: Precipitation climatology (mm day^{-1}) for (left) DJF and (right) JJA during 1979–2005. (a) Observational estimate of observed precipitation for DJF. (b) Multi-model ensemble mean over the 18 models for DJF; for models with multiple runs, all runs are averaged before calculating the mean. (c) Comparison of individual models to observations using the 3 mm day^{-1} contour as an index of major P features: half the models are shown in each of keys I and II with the legend giving the color coding for the models. Shading shows the regions where the observations exceed 3 mm day^{-1}. (d)–(f) as in (a)–(c), but for JJA. Observations are taken from the global precipitation climatology project (GPCP) version 2.2 (Adler et al., 2003), and model data were regridded to the GPCP resolution via bilinear interpolation prior to analysis. Figure appears in Sheffield et al. (2013a), their Fig. 1.

1.2 Other work in characterizing the simulation of climatology and climate change projections in multi-model ensembles

1.2.1 The North American climate in CMIP5 simulations

This was a three-part study that concentrated on three aspects of North American climate in the CMIP5 ensemble. The three studies are described briefly, along with some key figures from the published articles.

Part I: Sheffield et al. (2013a)

Part I of this study evaluates the historical simulations, focusing on continental and regional climatologies. I contributed to a section evaluating seasonal P climatologies, comparing an 18-model CMIP5 ensemble to observed P climatologies over North America during the 1979–2005 time period. Figure 1.1 shows some results from this work. Comparing observed P climatologies to the ensemble mean, one can see the models do fairly well at reproducing large-scale P features, and the biases over North America in the ensemble mean are calculated to be approximately +12% for DJF and –1% for JJA. However, regional scale disagreement is manifest in individual model biases, which range widely and typically exceed that of the ensemble mean (see Sheffield et al., 2013a, their Table 3). Generally, there is overestimation of P in more humid and cooler regions and underestimation in drier regions over North America. One notable aspect that the models reproduce is the DJF Pacific storm track, specifically the angle of the storm track as it approaches the North American west coast, though specific placement of the 3 mm day^{-1} contour does vary notably in space. The model grids are too coarse to resolve the complex topography along the west coast, so the exact magnitudes are slightly too large inland and not intense enough at the coast. Finally, the performance of CMIP5 models in representing climatological P was shown not to have improved significantly between CMIP3 and CMIP5.

Part II: Sheffield et al. (2013b)

Part II of this study focuses on the seasonal-to-decadal variability of the CMIP5 models. A portion of this analysis discusses the model accuracy of interannual variability, including ENSO

teleconnections in the coupled models. The results show a mixture of performance skill, and ENSO teleconnections are an aspect of climate variability that models show a range of accuracy in reproducing. Furthermore, no one model stands out as better than others when all types of variability are considered together, though certain models fare better on specific features. Based on these conclusions, the discussion points out that no model is particularly unskillful, lending credence to the utility of a multi-model ensemble. Additionally, this study notes that an overall ranking of the models would not be possible and found—as in Part I—that no significant improvement is seen in CMIP5 when compared to CMIP3 in these variability measures.

Part III: Maloney et al. (2014)

Part III of this study analyzes twenty-first century climate projections using the Representative Concentration Pathway 8.5 (RCP 8.5, Taylor et al., 2012) experiment. This paper builds on Parts I and II, considering first the continent-wide changes over North America and then shifting focus to selected regional climate features. My contribution was an analysis of end-of-century P changes. Figure 1.2 shows the DJF and JJA end-of-century P anomalies for the ensemble mean (left column). Anomalies are calculated for the 2070–2100 climatology in each model relative to a 1960–1990 base period. Figure 1.2 (right column) shows model agreement on positive and negative P changes for the 17-model ensemble. During DJF, Figs. 1.2a,c both show P increases to the north and decreases to the south, with Southern California lying at a latitude between where there is much less model consensus on predicted change. During JJA, Figs. 1.2b,d show robust drying extending from Mexico and Central America into the Atlantic Ocean, as well as robust P increases across Canada and Alaska, though between these regions there is less agreement on predicted change.

1.2.2 California winter P change in the CMIP3 and CMIP5 ensembles

Another study to which I have contributed concerns twenty-first century California wintertime P changes (Neelin et al., 2013). As noted above, Southern California lies in a tenuous region between P increases to the north and decreases to the south. This study focuses first on large-scale features of P change occurring throughout the eastern Pacific Ocean basin and North American west coast and then discusses the implications this has for California in particular. From Figs. 1.1

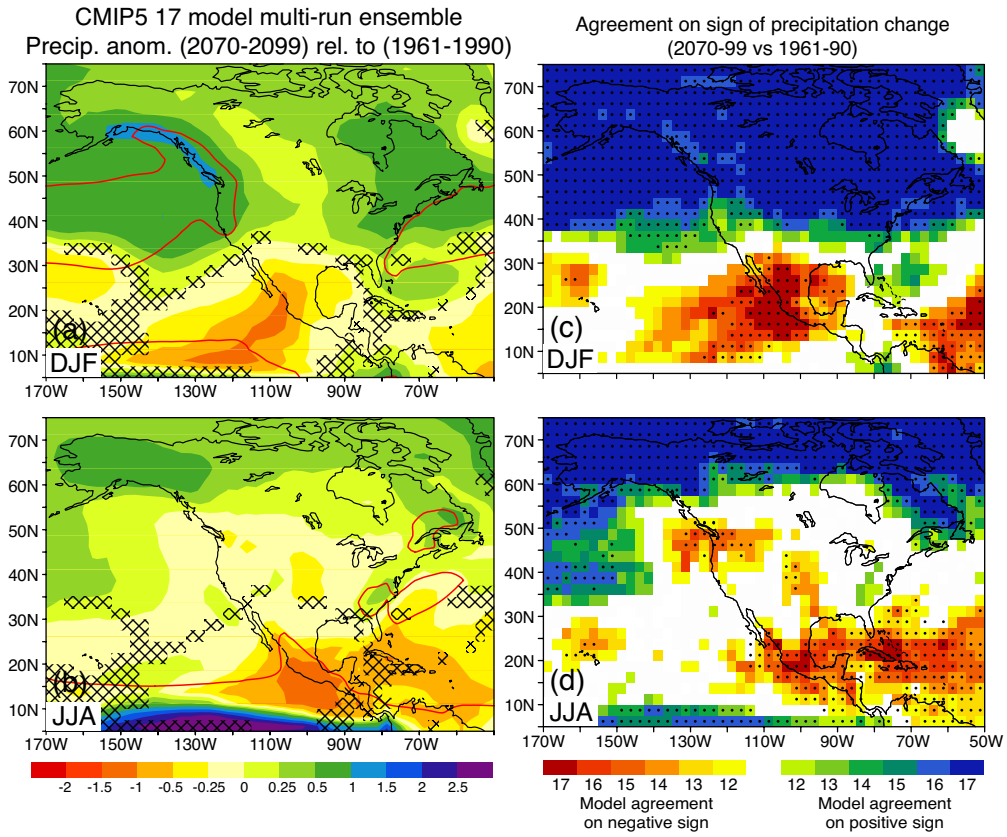

Figure 1.2: (a,b) CMIP5 17-member ensemble mean P change (mm day^{-1}) for RCP8.5 during 2070–2100 relative to 1960–1990 for (top) DJF and (bottom) JJA. The red line is the 4 mm day^{-1} contour of the historical climatology. Grid points are cross-hatched where the ensemble mean does not pass a two-tailed t test for differences of the mean with respect to interannual variability at the 95% level. (c,d) Plots of model agreement on sign of end-of-century precipitation change for the RCP8.5 scenario. Red colors indicate the number of models (out of 17) that agree on a negative P change; blue colors indicate the number of models that agree on a positive P change. The color shaded areas (12 or more models agreeing on sign) pass a binomial test rejecting the hypothesis of 50% probability of either sign at the 95% level; areas not passing at this level are left unshaded. Stippled areas use a version of the Neelin et al. (2006) criterion to show grid points where more than half (9+) of the models both pass a two-tailed t test at the 95% confidence level and agree on sign. Figures appear in Maloney et al. (2014), their Figs. 1 and 2.

and 1.2, one can see robust subtropical P reductions and midlatitude increases, often referred to as the rich-get-richer or wet-get-wetter mechanism (see the introduction to Chapter 3 for more detail). Though this will set the stage for P changes at large scales, dynamical feedbacks become important at the regional scale, and changes to the circulation affecting the climatological storm track has implications for the end-of-century P changes occurring along the California coast. In Neelin et al. (2013), a region encompassing the southeastern terminus of the storm track is discussed as a "jet extension" region, and it is evaluated critically in this study as an area with significant intermodel

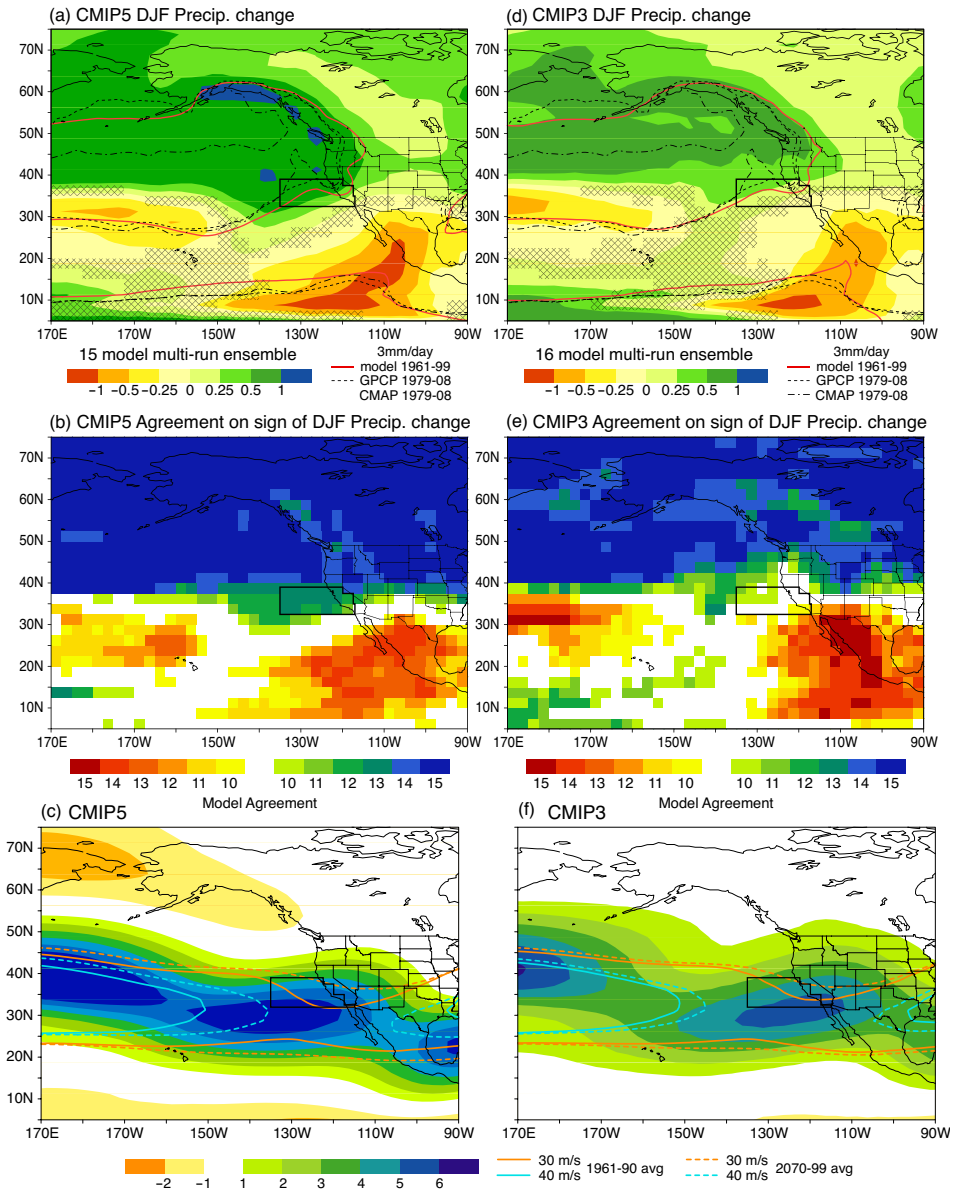

Figure 1.3: DJF end-of-century P changes for CMIP5. (a,d) Ensemble mean precipitation change in units of mm day^{-1} for CMIP5 and CMIP3. Red line shows the 3 mm day^{-1} contour from the ensemble mean base period climatology; dashed and dash-dot lines show the same contour from the GPCP and CMAP data sets during 1979–2008, respectively. Regions that do not pass a significance test for the ensemble mean relative to internal variability at the 95% level are cross-hatched. (b,e) Agreement on the sign of precipitation change among the model ensemble for CMIP5 and CMIP3. Blue and green colors indicate a higher number of models (out of 15) agreeing on positive precipitation change; red colors indicate agreement on a negative precipitation change. Grid points with 10 models agreeing on sign pass a binomial test at the 94% level; grid points not passing at this level are left blank. (c,f) End-of-century changes to 200 hPa zonal winds for CMIP5 and CMIP3. Boxes indicate the region used for averaging in later figures. Figures appear in Neelin et al. (2013), their Figs. 1, 2, and 6.

disagreement on sign of P change, with implications for Southern California wintertime P changes.

Figure 1.3 provides a comprehensive comparison between the CMIP3 and CMIP5 ensembles. Comparing Figs. 1.3a and 1.3d, one can see that the region of P increase to the north of Southern California is extended farther south in CMIP5 relative to CMIP3, which has implications for wintertime P throughout the state and, if CMIP5 is trustworthy, provides an optimistic outlook for future water in California. Figures 1.3b and 1.3e show relatively higher agreement in CMIP5 P increases here as well, adding a measure of robustness to the mean change plots. Figures 1.3c and 1.3f show the twenty-first century changes to the winds at the jet level. One can see that the jet is getting stronger at the latitude of its core and extends farther east as it approaches the North American west coast, implicating its role in steering incoming storms toward the California coast.

Figure 1.4 provides a summary of P changes in the model ensemble. Area-averaged P change is calculated for each model over a storm track termination outline shown in Fig. 1.3. CMIP3 and CMIP5 have comparable scatter for projected P change, though CMIP3 has more noticeable outliers, and the mean of CMIP5 is shifted toward a positive P change relative to CMIP3. These differences in mean are subject to t tests and are shown to be significant at or above the 90% level, implying that the CMIP5 model ensemble shows distinct changes from the CMIP3 ensemble and indicates an ensemble preference for a slight P increase in this region.

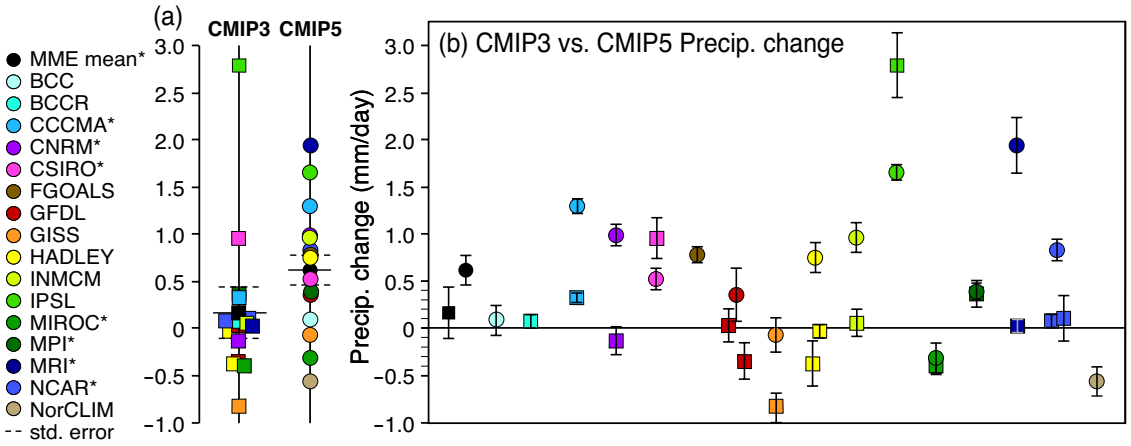

Figure 1.4: DJF P changes for an area average over the ocean just off the coast of California and extending into the California coastline (area shown in Fig. 1.4) for CMIP3 models (squares) and CMIP5 models (circles). (a) Values for all models with the ensemble mean for CMIP3 and CMIP5 respectively. (b) The ensemble mean for each of CMIP3 and CMIP5 shown at the left with error bars corresponding to ±1 standard error among the ensemble values (0.20 and 0.17 mm day^{-1}, respectively). Values from individual models from the same modeling group are stacked vertically. Error bars denote a ±1 standard error estimate due to natural variability from a given model. Asterisks in the legend denote models with multiple runs included in the average. Figure appears in Neelin et al. (2013), their Fig. 5.

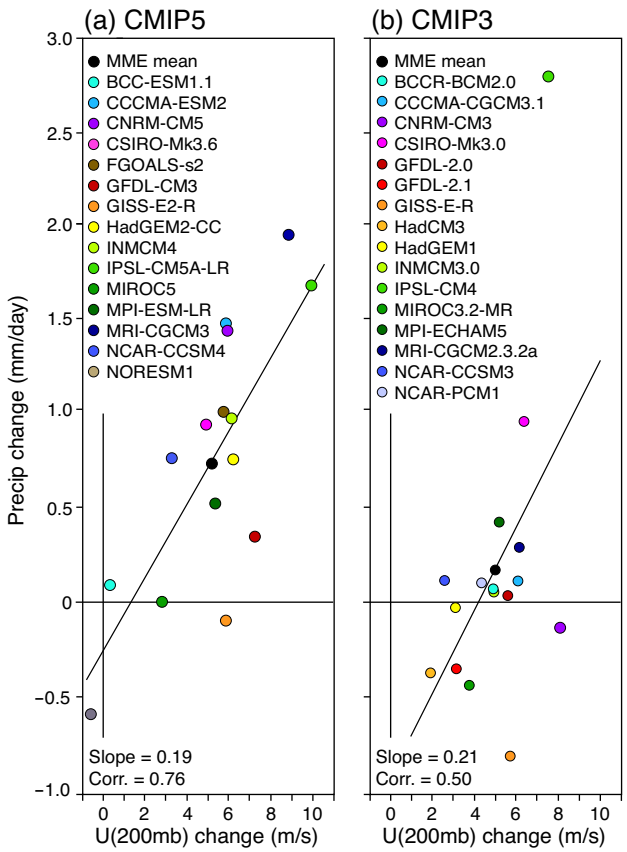

Figure 1.5: DJF P change averaged over the region in Fig. 1.3 plotted against DJF 200 hPa zonal wind change averaged over the same region, for (a) CMIP5 and (b) CMIP3 models. The correlation coefficients ($r = 0.76$ and 0.50) and slopes (0.19 and 0.21 mm day^{-1} [m s^{-1}]$^{-1}$, respectively) are associated with the linear regression line shown. Figure appears in Neelin et al. (2013), their Fig. 7.

As a final point, the relationship between twenty-first century change in jet-level winds and P in the averaging region is shown in Fig. 1.5. A majority of models in both ensembles indicate a positive change in both quantities, though the CMIP5 ensemble shows a higher correlation ($r = 0.76$). The CMIP3 ensemble has a lower correlation ($r = 0.50$) but a similar slope, so that a significant jet contribution to P change in this region is apparent in both ensembles, though it takes a less prominent role in CMIP3.

This connection between model spread in P changes along the California coast and its relation to shifts in the jet-level winds is a nontrivial source of P change disagreement at the latitude of Southern California at the end of the storm track, and this is the subject of the latter part of Chapter 3. But first, Chapter 2 discusses the ability of the CMIP3 and CMIP5 atmosphere-only models to simulate ENSO P teleconnections in the DJF season.

Chapter 2

Analyzing ENSO teleconnections in CMIP models as a measure of model fidelity in simulating precipitation

Abstract

The accurate representation of precipitation (P) is a recurring issue in climate models. El Niño-Southern Oscillation (ENSO) P teleconnections provide a test bed for comparison of modeled to observed P. The simulation quality for the atmospheric component of models in the Coupled Model Intercomparison Project phase 5 (CMIP5) is assessed here, using the ensemble of runs driven by observed sea surface temperatures (SSTs). Simulated seasonal P teleconnection patterns are compared to observations during 1979–2005 and to the ensemble of CMIP phase 3 (CMIP3). Within regions of strong observed teleconnections (equatorial South America, the western equatorial Pacific, and a southern section of North America), there is little improvement in the CMIP5 ensemble relative to CMIP3 in amplitude and correlation metrics of P. Teleconnection patterns within each region exhibit substantial departures from observations, with correlation coefficients typically less than 0.5. However, the atmospheric models do considerably better in other measures. First, the amplitude of the P response (root-mean-square error over each region) is well estimated by the mean of the amplitudes from the individual models. This is in contrast with the amplitude of the multi-model ensemble mean, which is systematically smaller (by about 30–40%) in the selected teleconnection regions. Second, high intermodel agreement on teleconnection sign provides a good predictor for high model agreement with observed teleconnections. The ability of the model ensemble to yield amplitude and sign measures that agree with the observed signal for ENSO P teleconnections lends supporting evidence for the use of corresponding measures in global warming

projections.

2.1 Introduction

El Niño-Southern Oscillation (ENSO) is a leading mode of interannual climate variability originating in the tropical Pacific. ENSO teleconnections are a reflection of the strong coupling between the tropical ocean and global atmosphere, and SST anomalies in the equatorial Pacific can have substantial remote effects on climate (Horel and Wallace, 1981; Ropelewski and Halpert, 1987; Trenberth et al., 1998; Wallace et al., 1998; Dai and Wigley, 2000).

In recent decades, measurable progress has been made in simulating ENSO dynamics and associated teleconnections within ocean-atmosphere coupled general circulation models (GCMs) (Neelin et al., 1992; Delecluse et al., 1998; Davey et al., 2002; Latif et al., 2001; DeWeaver and Nigam, 2004; AchutaRao and Sperber, 2006; Randall et al., 2007). A number of studies use the fully coupled GCMs to assess twentieth-century ENSO variability and teleconnections against observations (Doherty and Hulme, 2002; Capotondi et al., 2006; Joseph and Nigam, 2006; Cai et al., 2009). Others examine the evolution of ENSO and these teleconnections under climate change (Doherty and Hulme, 2002; Van Oldenborgh et al., 2005; Merryfield, 2006; Meehl et al., 2007; Coelho and Goddard, 2009). Problems persist in the ability of the models to accurately represent the tropical Pacific mean state, annual cycle, and ENSO's natural variability (Guilyardi et al., 2009b; Cai et al., 2012). Additional uncertainties remain in the role of the atmospheric components of coupled GCMs in setting the dynamics of ENSO and its teleconnections (Guilyardi et al., 2004, 2009a; Lloyd et al., 2009; Sun et al., 2009; Weare, 2013), as well as how ENSO will behave under climate change (Collins et al., 2010).

The P response to interannual climate variations like ENSO also continues to be a challenge for coupled GCMs (Dai, 2006). In the tropics, equatorial wave dynamics spread tropospheric temperature anomalies, which induce feedbacks within convection zones in surrounding regions (e.g., Chiang and Sobel, 2002; Su et al., 2003). At midlatitudes, wind anomalies generated by Rossby wave trains interact with storm tracks to create P anomalies (Held et al., 1989; Chen and van den Dool, 1997; Straus and Shukla, 1997). These moist teleconnection processes share physical mechanisms with feedbacks active in climate change (e.g., Neelin et al., 2003). Examination of

ENSO P teleconnections can therefore contribute to assessing the accuracy of models for these pathways, but note that this is distinct from the discussion in the literature that the tropical Pacific may experience "El Niño-like" climate change.

One difficulty with assessing teleconnections from coupled models is that errors in the ENSO dynamics (e.g., in amplitude or distribution of the main SST anomaly in the equatorial Pacific) degrade the quality of the simulation at the source region before the teleconnection mechanisms even begin (Joseph and Nigam, 2006; Coelho and Goddard, 2009). To isolate the atmospheric portion of the teleconnection pathway, it is useful to employ atmospheric component simulations forced by observed SSTs, referred to as Atmospheric Model Intercomparison Project (AMIP or atmosphere-only) runs (Gates et al., 1999). In coupled model runs, errors in position or amplitude of the main equatorial ENSO SST signal can have a substantial impact on the teleconnections (Cai et al., 2009), and it is quite challenging for the models to accurately simulate regional signals in P, even when observed SSTs are specified.

A few studies use atmosphere-only runs to examine ENSO teleconnections. Risbey et al. (2011) do so for teleconnections over Australia, noting errors in the modeled amplitude and pattern coherence. Spencer and Slingo (2003) find that issues in the sensitivity of P to tropical Pacific SSTs lead to errors in the Aleutian low despite otherwise accurate tropical ENSO teleconnections. Cash et al. (2005) compare two uncoupled atmospheric GCMs forced with identically prescribed SSTs, finding noticeable variations between the two models in the response of extratropical 500 hPa height and regional P. They force these models with climatological SST fields and SSTs representative of a response to a CO2 doubling experiment and find that P difference patterns between the two models are similar for either case, implying that the differences between the atmospheric components of GCMs are "relatively insensitive" to the prescribed SST fields.

Because challenges persist in correctly simulating a P teleconnection response (e.g., Rowell, 2013), analysis of the CMIP phase 5 (CMIP5) atmosphere-only ensemble can provide a way to gauge the fidelity of the current generation of models in simulating large-scale atmospheric processes leading to rainfall. In particular, we evaluate December-January-February (DJF) ENSO P teleconnections during 1979–2005 in the CMIP5 atmosphere-only models, and we compare these to observations and to the earlier CMIP phase 3 (CMIP3) atmosphere-only ensemble.

In standard evaluation measures of teleconnection patterns and amplitude, substantial differences exist among models and when compared to the observations. In light of such differences, we turn to other measures in which the multi-model ensemble (MME) may contain useful information. These include amplitude measures, a comparison of individual models to the ensemble mean, and measures of sign agreement. In these alternative measures, the CMIP5 model ensemble does unexpectedly well compared to observations. The performance on sign agreement measures is decent enough to motivate questions regarding the optimal way to apply significance tests within multi-model ensembles. We provide some explanation in the discussion section, noting that even though a full answer may not yet exist, such alternative measures are relevant to the evaluation of P change in global warming.

2.2 Datasets and analysis

2.2.1 Data

To produce ENSO P teleconnection patterns, we use modeled and observed monthly mean SST and P data during the DJF months for the years 1979–2005. For SST observations, we use the Extended Reconstructed Sea Surface Temperature (ERSST) version 3 data set (Xue et al., 2003; Smith et al., 2008). For monthly P observations, we employ the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) archive (Xie and Arkin, 1997). For modeled teleconnections, we have downloaded monthly atmosphere-only P and surface temperature (TS) data from the CMIP3 and CMIP5 archives, as detailed in Table 2.1 (for more information on atmosphere-only runs, see Gates et al., 1999, and references therein). All modeled P data are regridded to a $2.5^\circ \times 2.5^\circ$ grid via bilinear interpolation prior to calculating teleconnection patterns. This is the native grid of the CMAP P data set and facilitates direct comparison of modeled teleconnections to the observations.

2.2.2 Analysis

Linear regression and Spearman's rank correlation are used to calculate DJF P teleconnections for the selected time period. Linear regression is widely used for assessing the relationship

Table 2.1: CMIP5 and CMIP3 modeling centers and models used, and the number of atmosphere-only runs available at the time of analysis.

CMIP5 atmosphere-only	**runs**	**CMIP3 atmosphere-only**	**runs**	**Modeling center or group**
BCC-CSM1.1	3	—	—	Beijing Climate Center, China Meteorological Administration (BCC), Beijing, China
CanAM4	4	—	—	Canadian Centre for Climate Modeling and Analysis (CCCMA), Québec, Canada
CCSM4	1	CCSM3	1	National Center for Atmospheric Research (NCAR), Boulder, Colorado, USA
CNRM-CM5	1	CNRM-CM3	1	Centre National de Recherches Météorologiques (CNRM), Toulouse, France
CSIRO-Mk3.6.0	1	—	—	Commonwealth Scientific and Industrial Research Organization (CSIRO) in collaboration with Queensland Climate Change Centre of Excellence, Victoria, Australia
FGOALS-s2	3	FGOALS-g1.0	3	State Key Laboratory of Numerical Modelling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences (CESS), Beijing, China
GFDL-HiRAM-C180	3	GFDL-CM2.1	1	NOAA Geophysical Fluid Dynamics Laboratory (NOAA GFDL), Princeton, New Jersey, USA
GISS-E2-R	5	GISS-E-R	4	NASA Goddard Institute for Space Studies (NASA GISS), New York, New York, USA
HadGEM2-A	5	UKMO-HadGEM1	1	UK Met Office (UKMO) Hadley Centre, Exeter, United Kingdom
INMCM4.0	1	INMCM3.0	1	Institute for Numerical Mathematics (INM), Moscow, Russia
IPSL-CM5A-LR	5	IPSL-CM4	5	L'Institut Pierre-Simon Laplace (IPSL), Paris, France
MIROC5	2	MIROC3.2(hires)	1	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Tokyo, Japan
MPI-ESM-LR	3	MPI-ECHAM5	3	Max Planck Institute (MPI) for Meteorology, Hamburg, Germany
MRI-CGCM3	3	MRI-CGCM2.3.2	1	Meteorological Research Institute (MRI), Tokyo, Japan
NorESM1-M	3	—	—	Norwegian Climate Centre (NCC), Bergen, Norway

between global P and tropical Pacific SSTs, where P at a grid point is regressed against a spatially averaged SST time series. Here we use the Niño-3.4 index, defined from 5°S–5°N and 190°–240°E; for information on El Niño indices, see Trenberth (1997). One caveat is that linear regression assumes the P data follow a Gaussian distribution, whereas in reality they are zero-bounded and exhibit non-Gaussian behavior. Spearman's rank correlation—in which the rank of the data is used to compute the correlation coefficient (Wilks, 2011)—does not make such assumptions, and it is used here to provide a check on the sensitivity of teleconnection patterns to the statistical methods employed. For examples of other studies that use rank correlation, see Whitaker and Weickmann (2001) or Münnich and Neelin (2005).

Appropriate t tests are used in both the linear and rank methods to resolve grid points that meet or pass certain confidence levels (von Storch and Zwiers, 1999). The majority of this paper will focus on a t test applied to teleconnections resolved via linear regression. This test is based on

calculating a two-tailed p value where the null hypothesis is a linear regression slope of zero. Note that our use of the Niño-3.4 index yields standard teleconnection patterns, which provide a good basis for comparison of models to observations. We recognize, however, that there is interesting work addressing the next level of distinction among different "flavors" of ENSO and the remote impacts of SST anomalies that have a central (rather than eastern) Pacific signature (Ashok et al., 2007; Kao and Yu, 2009; Trenberth and Smith, 2009).

2.3 Evaluating modeled patterns and amplitudes of P teleconnections

2.3.1 Teleconnection patterns resolved via linear regression and rank correlation

Figures 2.1 and 2.2 show observed and modeled P teleconnections for the DJF season as estimated by linear regression and Spearman's rank correlation, respectively. We show both methods to check that teleconnected rainfall patterns are robust against the statistical assumptions going into the calculation (ENSO composites, not shown, yield similar results). Spearman's rank correlation is insensitive to extreme values and so can bring regions with different variance onto a common footing. This statistical method also offers a significance test that does not assume Gaussian statistics. Linear regression, by contrast, is easier to interpret in terms of a change of the physical variables, which in this case is P rate per degree change of SST in the Niño-3.4 region. Beyond this, comparing modeled to observed teleconnections raises some interesting questions about the restrictions of the statistical significance tests. The most pertinent question to arise is how best to use the collective information offered by a multi-model ensemble. Substantial intermodel variations also occur, and they are discussed in Subsections 2.3.2–2.3.4. Other aspects of the restrictive nature of these significance tests will be discussed in Section 2.4.

Figures 2.1b and 2.2b show teleconnection patterns obtained from the model ensemble. Note that there are several ways to obtain a regression representative of all data contained in the 15-model ensemble. The option we choose provides a straightforward test of statistical significance. Specifically, we perform the regression over all 15 models simultaneously; a straightforward way to interpret (and program) this is as a concatenated time series of the 15 available models, and so

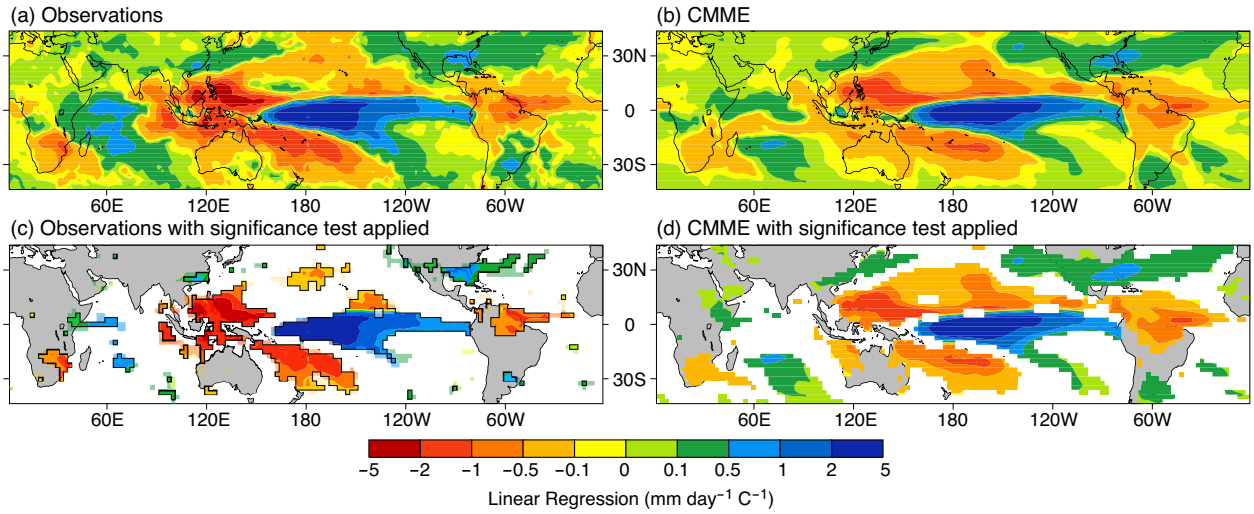

Figure 2.1: DJF P teleconnections for the years 1979–2005, as diagnosed through a linear regression analysis of P against the Niño-3.4 index (mm day^{-1} °C^{-1}). (a) Observed teleconnections. (b) Concatenated multi-model ensemble (CMME) teleconnections for the CMIP5 atmosphere-only 15-model ensemble. (c) As in (a), but with a two-tailed t test applied to the regression values and shaded at 95% confidence (black outline) and 90% confidence (lighter shading). (d) As in (b), but shaded only where a t test yields grid points significant at or above the 95% level.

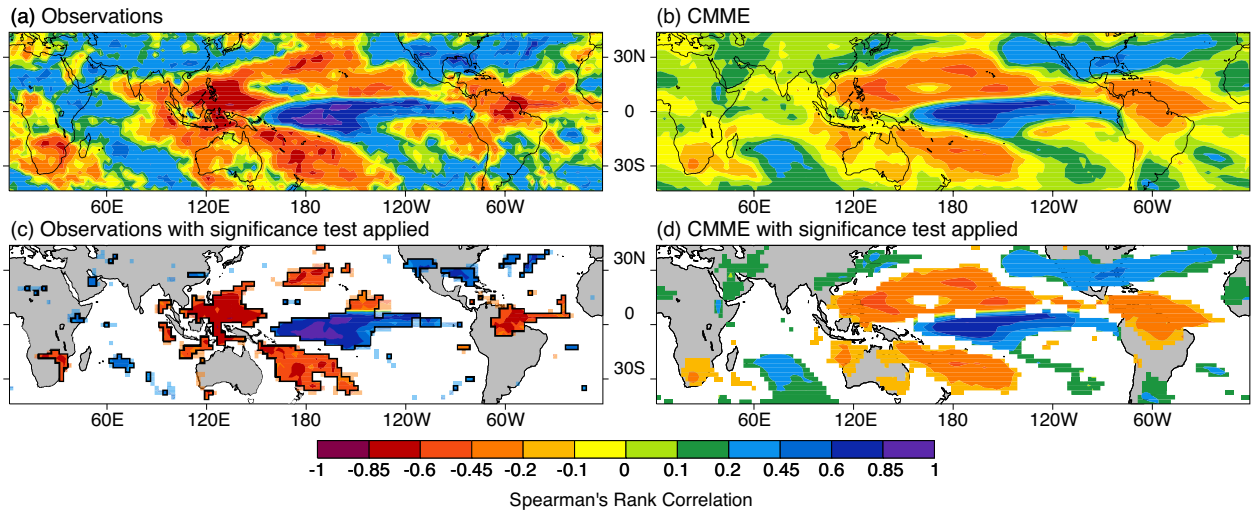

Figure 2.2: As in Fig. 2.1, but for Spearman's rank correlation analysis between grid point P and the Niño-3.4 index. Note here that the color bar is unitless and corresponds to the Spearman's rank correlation coefficient, with a minimum of +1.0 and a maximum of −1.0. (a),(b) The teleconnection patterns from the rank correlation applied to the observations and the CMME, respectively. (c) As in (a), but shaded only where grid points pass the 95% confidence level (black outline) and the 90% confidence level (lighter shading) of a statistical significance test for the rank correlation analysis. (d) The CMME teleconnections shaded for grid points that pass at the 95% significance level in the rank correlation analysis.

we will refer to this as the concatenated multi-model ensemble (CMME), when it is necessary to distinguish it.

The more classical approach of obtaining a single map of teleconnections for a 15-model ensemble is to calculate the teleconnections for each member individually and average the 15 patterns together afterward, discussed previously as the ensemble mean. While this is more widely used, obtaining a test of statistical significance becomes complicated, as one cannot easily take an average of significance tests across 15 models. Thus in Figs. 2.1 and 1.2, the variant shown is the first one, although the ensemble mean (not shown) and CMME patterns are nearly identical, with a global correlation coefficient of $r > 0.999$. The high correlation between these two methods is to be expected if the variance in each model is similar and stably estimated. In the remainder of this paper, we will focus on the ensemble patterns seen in both Figs. 2.1b and 1.1d, and we will refer to them using ensemble mean and CMME interchangeably.

In Fig. 2.1, we show CMME linear regression DJF teleconnection patterns (Figs. 2.1b,d) alongside observations (Figs. 2.1a,c). The ensemble pattern in Fig. 2.1b reproduces a number of observed features. A broad region of reduced P over equatorial South America, stretching out through the Atlantic intertropical convergence zone (ITCZ), is qualitatively simulated, although the region of the most intense anomalies is displaced from the observations. The region of increased P starting off the coast of California and extending through Mexico, the Gulf States, and beyond Florida into the Atlantic storm track is also qualitatively reflected in the CMME regression. In the western Pacific, and surrounding the main ENSO region to the north and south, there is a broad "horseshoe" pattern of reduced P, which the CMME captures reasonably well in terms of the low-amplitude parts, although the location of the most intense anomalies is off.

Figures 2.1c,d show the same data as Figs. 2.1a,b, but with a two-tailed t test applied to the regression at each grid point. One can see in Fig. 2.1d that the CMME regression passes a 95% confidence level criterion over fairly broad areas in each major teleconnection region, thanks to the large amount of information available in the 15-model ensemble. Each of the areas discussed above passes this significance test, as do some smaller regions, such as southeastern Africa. Figure 2.1c displays observed teleconnections masked to show only grid points that pass the 90% and 95% confidence levels, indicating a relatively limited area over which the grid point-based regressions

meet these confidence criteria. Specifically, linear regressions in Fig. 2.1 produce statistically significant teleconnections at 36.8% of grid points across the globe in the CMME. The average of the individual 15 models is 17.6% of grid points, while that of the observations is 16.1%. Thus the local significance tests for individual models, not shown, are qualitatively similar to the extent of the observations in Fig. 2.1c.

Given that the CMME yields a statistically significant prediction for the sign of the signal over the main teleconnection regions, a one-tailed t test (on the side predicted by the CMME) could be used on the observations (and note that the 90% confidence level of a two-tailed test would correspond to the 95% confidence level of a one-tailed test). However, when loosening the confidence restriction from 95% to 90% for observed teleconnections, we only see a small increase in the extent of regions that pass the significance test. In comparing Figs. 2.1c and 2.1d, one can see that the CMME is significant at 95% confidence over a broader area than the observations.

Figure 2.2 displays the same information as in Fig. 2.1, but for Spearman's rank correlation applied to the CMME and observations. The teleconnection patterns that result using either the linear or rank method are similar overall, implying that ENSO P teleconnections are robust despite assumptions made about the distribution of rainfall events a priori. Differences may be noted between the two methods in particular regions, such as the rank correlation deemphasizing the narrow band along the equator in South America in the CMME (Fig. 2.2b) relative to the linear regression (Fig. 2.1b), although not in the observations (Fig. 2.2a). The region passing significance criteria at the 95% level under the rank correlation of the observations (Fig. 2.2c) is comparable to that produced for the linear regression of the observations (Fig. 2.1c), and likewise for the CMME. In the remainder of this chapter, we focus on linear regression teleconnection patterns on account of the simpler interpretation of the amplitudes.

2.3.2 Regional model disagreement

Another point that can be made with Figs. 2.1 and 2.2 is the large-scale agreement between teleconnected P patterns in the CMME and in the observations. For reasons discussed in Section 2.5, this agreement is apparent over broader regions where the CMME passes the t test at 95% confidence, not just the narrower regions where observations pass this test. However, regional

disagreement between observations and the CMME pattern is also seen, especially in regions where the observations have intense P.

In addition, the CMME exhibits a general "smoothing" of teleconnection patterns. These overly smoothed teleconnection patterns in the CMME can be understood when examining individual model patterns. Figure 2.3 shows teleconnections for one run of each model in CMIP5, displayed for the equatorial Americas; substantial regional variability is easily seen. Qualitatively

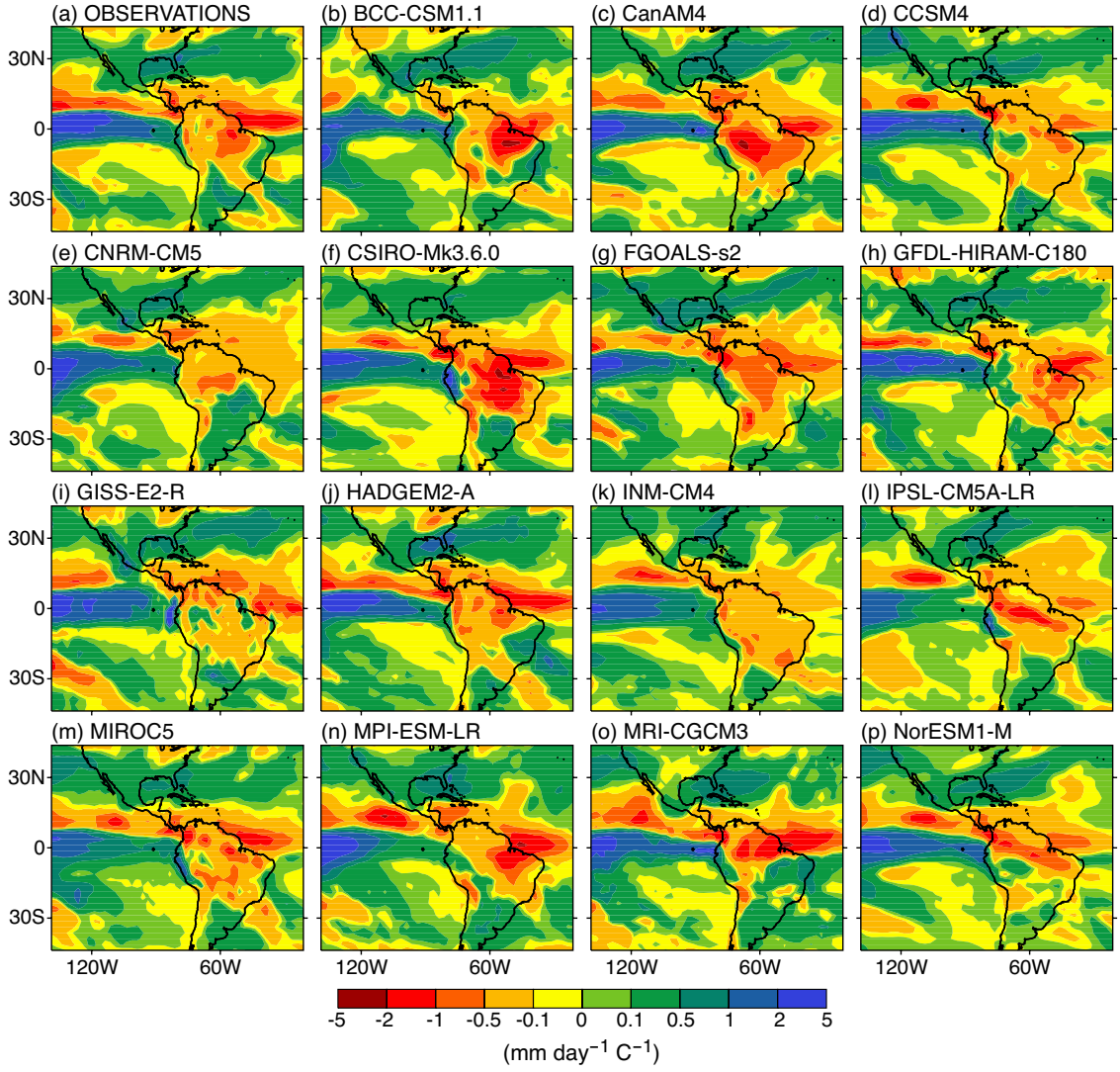

Figure 2.3: DJF P teleconnections shown for (a) the observations and (b)–(p) 1 run from each of 15 available CMIP5 atmosphere-only models (listed alphabetically by model acronym; see Table 2.1 for expansions of model names). Teleconnections here are resolved via the linear regression analysis as in Fig. 1, with an identical color bar that has units of mm day^{-1} °C^{-1}. Patterns are plotted for the equatorial Americas to highlight regional, intermodel disagreement among the ensemble members.

similar figures highlighting regional disagreement have been produced in other studies that use coupled GCMs to examine ENSO teleconnections and P characteristics (e.g., Dai, 2006, his Fig. 9). Difficulties in simulating these teleconnections in coupled GCMs persist in the atmosphere-only models shown here: variations in the location of the strongest P anomaly in Fig. 2.3 are common from model to model, even though these are the areas that most easily pass significance criteria on an individual model basis. Over the region where the CMME regression passes a t test at the 95% level, however, one can see that the overall teleconnection pattern is plausible at large scales in each of the models. Thus, Fig. 2.3 provides a visual sense of the tradeoffs to be quantified: disagreement among models at regional scales, excessive smoothing relative to observations in the CMME, and yet some possibility that there is useful information about the teleconnection patterns in the 15-model ensemble, if it can be suitably extracted.

2.3.3 Taylor diagram analysis of modeled teleconnections

The regional variation among atmosphere-only models leads to a distinction between their ability (1) to reproduce patterns of teleconnections and (2) to represent the amplitudes of these patterns. To examine individual model fidelity in simulating patterns and amplitude of rainfall teleconnections, we look at four regions that show a robust ENSO response; each region displays a continuous teleconnection signal significant at the 95% confidence level in observations (see Fig. 2.1c).

These four regions include (a) the equatorial Pacific (the "cold tongue" region; positive DJF ENSO signal), (b) the horseshoe-shaped region in the western Pacific (negative signal), (c) equatorial South America (negative signal), and (d) a southern section of North America (positive signal). The equatorial Pacific region is shown for reference, since this is the source region and is directly forced by the largest ENSO-related SST anomalies. We consider the other three regions the "teleconnection regions," since to accurately simulate teleconnected rainfall in each domain, the models must capture the pathways leading to remote P change. The Taylor diagrams (Taylor, 2001) in Fig. 2.4 show the correlations between the observations and each model plotted against the root-mean-square error of each model's pattern (i.e., the standard deviation σ_{mod}) normalized by observations (σ_{obs}); we refer to this measure as the teleconnection amplitude. For models with

multiple runs, correlations and amplitudes are calculated for each run first and then averaged among them; each individual model is given equal weight in the ensemble mean. Note we use the ensemble mean here, and not the CMME, although Taylor diagrams using the latter (not shown) are nearly identical. Additionally, some of the individual models have small negative correlations with observations in certain regions. These models are used in calculating the ensemble mean, although for simplicity the domain of the Taylor diagrams is not extended to display these points.

Figure 2.4 allows easy comparison between CMIP3 and CMIP5 atmosphere-only runs. There is little (if any) improvement from CMIP3 to CMIP5 in reproducing teleconnected rainfall patterns in these regions. Additionally, models exhibit generally low correlations with observations (ranging from less than 0.2 to a few instances exceeding 0.7, with an average correlation coefficient of about $r = 0.40$). In every region, one can also see that the ensemble mean is typically more accurate than the majority of individual models in reproducing patterns. However, the ensemble mean amplitude is substantially lower than that of the individual ensemble members, and it underestimates the observations in every region outside of the central equatorial Pacific. As a final point, we note that Taylor diagrams of the corresponding rank correlation method (not shown) indicate consistent results.

2.3.4 Teleconnection amplitude in major impact regions

The varied agreement in amplitude measures from Fig. 2.4 suggests that it may be more reasonable to use amplitude information from individual ensemble members, rather than using that of the ensemble mean. To get a better sense of how teleconnection amplitude of individual models might be affected by internal variability within the models themselves, we take advantage of atmosphere-only models with multiple realizations, and we assess the internal variability among these runs for each model. We then compare this to the amplitude range of the 15-model ensemble. Figure 1.5 displays the radial axis from the Taylor diagrams discussed previously, but where multiple runs from each model are available, we plot them individually (43 total runs for 15 models in CMIP5; 26 total runs for 13 models in CMIP3; see Table 2.1).

The vertical extent of the black lines in Fig. 2.5, representing ± 1 standard deviation of the amplitudes for the runs of a given model, is a measure of internal variability for that model.

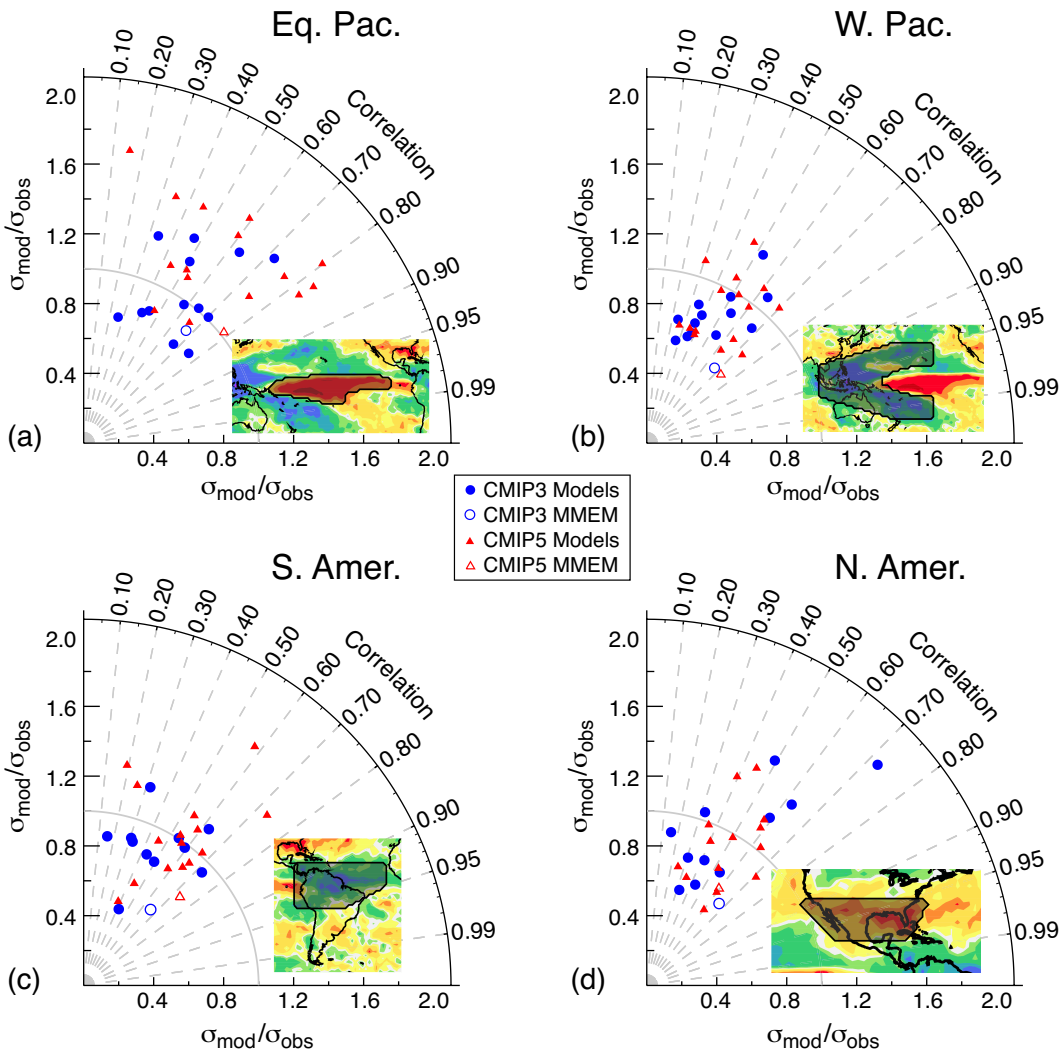

Figure 2.4: Taylor diagrams for the standardized amplitude and correlation of P teleconnections in four selected regions, as indicated in the inset of each panel: (a) the equatorial Pacific (central ENSO) region, (b) the "horseshoe" region in the western equatorial Pacific, (c) an equatorial section of South America, and (d) a southern section of North America. On the Taylor diagrams, angular axes show correlations between modeled and observed teleconnections; radial axes show standard deviation (root-mean-square error) of the teleconnection signals in each area, normalized against that of the observations. Shaded red triangles (15 total) and blue circles (11 total) denote each of the CMIP5 and CMIP3 atmosphere-only models, respectively. The unshaded red triangle is the CMIP5 ensemble mean; the unshaded blue circle is the CMIP3 ensemble mean. Note that some models have negative correlations with the observed teleconnections in a few regions; while we include them in calculating the ensemble mean, we do not plot them individually in the diagrams.

The vertical extent of each green bar is ±1 standard deviation of the ensemble mean amplitude, and it serves as a measure of intermodel variability. Notable points from this diagram include the following: (1) The ensemble mean systematically underestimates the spread and central tendency of intermodel variability, with a low bias of about 20–40% outside of the immediate ENSO region.

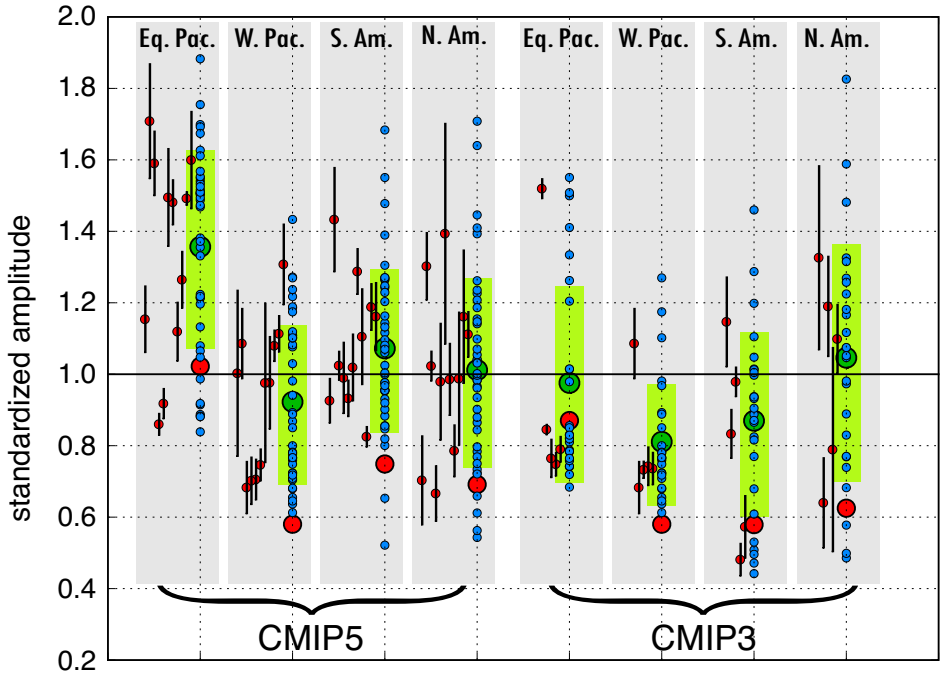

Figure 2.5: Standardized amplitude of P teleconnections in each of the four regions identified in Fig. 2.4. The calculation for this amplitude is discussed in the caption of Fig. 4 and in the text. CMIP5 models (15 models, 43 runs) are shown on the left; CMIP3 models (13 models, 26 runs) on the right; see Table 2.1 for models used. Each blue dot represents a separate model run, and where multiple runs are available for a given model, a blue dot is plotted for each. Black bars represent the spread among the multiple runs for one model (when available), centered at that model's average amplitude among the multiple runs ($\pm 1\sigma$ of the amplitude measure). The green dots and green bars denote the average teleconnection amplitude and its spread ($\pm 1\sigma$) for the entire ensemble in each region. The red dot is the ensemble mean including all available models and runs, weighted so that each separate model contributes equally.

(2) The regional disagreement among models owes itself partly to internal model variability, but intermodel variability contributes to the majority of the regional disagreement seen in Fig. 2.3. (3) Individual models are overestimating the amplitude in the immediate ENSO region for CMIP5, even though their spread is more symmetric about the observations in remote regions. (4) When comparing CMIP5 to CMIP3, CMIP5 shows no consistent improvement or change due to model development. Although the ensemble mean may fall closer to observed amplitudes in some regions for CMIP5, this comes at the expense of a tendency for individual models to overestimate P teleconnections in the central ENSO region.

Figure 2.5 suggests that serious errors can result from considering only information available in the ensemble mean. While its patterns correlate better with observations than most individual models, the ensemble mean teleconnection amplitude is routinely too low in the remote regions

considered. It is therefore useful to consider measures of teleconnection amplitude and spread from individual models, in addition to the ensemble mean, in situations where regional disagreement can dampen the ensemble mean amplitudes due to averaging varied model signals.

2.4 Sign agreement plots in ENSO teleconnections, and an argument for agreement plots of P change in global warming scenarios

Agreement plots for the sign of P change under global warming scenarios are commonly used in multi-model studies (e.g., Randall et al., 2007; Meehl et al., 2007), often as complementary information to the ensemble mean. Agreement-on-sign tests can be viewed as relatively weak statements regarding the P change at individual grid points for the model ensemble, and it has been argued that sign agreement should be used in conjunction with requirements on individual models that grid points pass statistical significance tests for change in mean P (e.g., Neelin et al. (2006), hereafter N06; Tebaldi et al. (2011), hereafter T11).

Here we examine agreement-on-sign measures based on the ENSO P regression patterns for each model. Because we can assess these against observations, we can use this to examine the procedure as a means of inferring its usefulness. If a procedure that identifies high model agreement at a grid point also correctly predicts the sign of the observations at that grid point, it can help build confidence in using corresponding procedures for the global warming case.

Figure 2.6a shows the traditional agreement-on-sign plot for ENSO teleconnections in the CMIP5 atmosphere-only ensemble. At each grid point, we count the number of models that agree on a positive or negative DJF teleconnection signal for the linear regression over Niño-3.4, so that the plot shows the integer value of models that agree on a wet or dry response during ENSO. The sign of the regression slope at each grid point is equivalent to the sign of the expected DJF P response during an El Niño event. Areas with 12 or more models agreeing on sign are shaded based on a binomial test. Specifically, if we consider the null hypothesis that the value of an ENSO P signal for a given point is equally likely to be positive or negative (i.e., drawn from a binomial distribution with a probability of $p > 0.5$), then when 12 or more models agree on sign, the null hypothesis for this 50–50 probability can be rejected at a confidence level greater than 98% (for 15 models, a sign agreement of 12 or more corresponds to a confidence level of about 98.6%, and 11

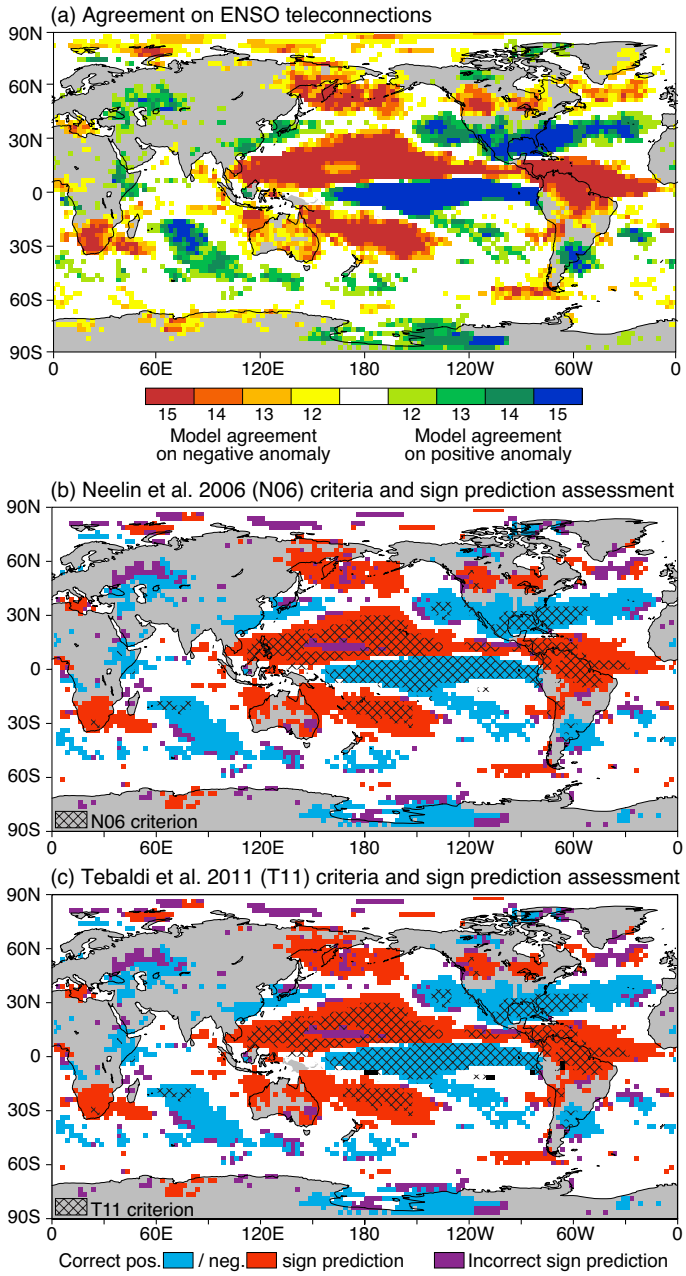

Figure 2.6: (a) Agreement on a positive teleconnection signal (linear regression) within the 15-model ensemble. Blue (red) colors represent high agreement on a positive (negative) P response during ENSO events. Note that in an ensemble of 15 models, an agreement count of 12 implies that 80% of models agree on the sign of the P teleconnection at that grid point, which is the area passing a binomial test at 98% confidence level (discussed in text). (b) Neelin et al. (2006), or N06, significance criteria (cross-hatching) overlaid on the sign prediction of the 15-model ensemble (colored shading). (c) Tebaldi et al. (2011), or T11, significance criteria (cross-hatching) overlaid on the sign prediction of the ensemble, as in (b). Details of the N06 and T11 cross-hatching criteria and sign prediction shading are outlined in the text. The cross-hatching is shown as an overlay in (b) and (c) to highlight the restrictive nature of the N06 and T11 criteria relative to the more extensive coverage over which the 15-model ensemble passes the binomial test at the 98% level and exhibits an accurate prediction of the observed teleconnection signals.

or more corresponds to 95.8%; both yield fairly similar patterns, so we use the more conservative 12).

The grid points with high sign agreement that pass the binomial test at the 98% level in Fig. 2.6a cover a region similar to the areas passing the two-tailed t test applied to the CMME (Fig. 2.1d) at the 95% level. However, the areas of high sign agreement cover a much larger region than those passing the t test at the 95% level for individual model realizations, which are similar to the areas passing the t test at this level for observations (see Fig. 2.1c and the discussion in Section 2.3.1).

This last point suggests two comparisons. First, we can contrast regions of high sign agreement identified by the binomial test with examples of criteria that have been considered in the global warming literature that combine t tests on individual models with sign agreement criteria from the ensemble. Second, in this ENSO teleconnection test bed, we can evaluate the model ensemble's sign prediction against observations. These results are displayed in Figs. 2.6b and 2.6c. These panels display hatching according to the N06 or T11 criteria, respectively, overlaid on a plot that assesses the prediction of the model ensemble for the sign of the teleconnection signal; details of these criteria are outlined below.

To produce the cross-hatching in Fig. 2.6b, we follow the N06 procedure: 1) at each grid point, count the number of models in the ensemble that have a slope significantly different from zero at the 95% confidence interval, and 2) cross-hatch grid points where greater than 50% of models are significant and also agree on the sign of the P teleconnection. The N06 criteria impose a requirement that at least half of models both be significant and agree on sign.

To produce the cross-hatching in Fig. 2.6c, we follow the T11 procedure: 1) at each grid point, count the number of models with a teleconnection significant at the 95% confidence interval (as in N06); 2) for grid points where more than 50% of models show a significant rainfall response, cross-hatch if 80% or more of significant models agree on the sign of the response; and 3) if fewer than 50% of models agree on the sign, shade the grid point black.

The underlying color shading in Figs. 2.6b and 2.6c is identical and evaluates the sign prediction of the atmosphere-only CMME for the teleconnection signal, produced in the following way: (1) Take the regions of high sign agreement passing the binomial test at the 98% significance

level in Fig. 6a as a prediction of the sign of the observed teleconnection pattern and compare that to the observations at the same grid point. (2) If the observations and the model prediction agree on sign, shade blue for a positive or red for a negative ENSO P signal, representing a correct prediction by the intermodel agreement plot (Fig. 2.6a). (3) If the observations and Fig. 2.6a disagree on the sign, shade the grid point purple to indicate an erroneous prediction. (4) If the agreement on sign does not pass the binomial test criterion of Fig. 2.6a, no prediction is made and the grid point is left unshaded.

When examining Figs. 2.6b and 2.6c, the most important point is that the model ensemble prediction of sign does very well when assessed against observations. In major regions for which model agreement passes the binomial test at 98% confidence, almost the whole area yields the correct sign. The scattered, incorrect grid points tend to be either isolated or at the edges of correct regions, such that a scientific assessment of likely areas of increase or decrease based on the predicted areas (color shading in Figs. 2.6a and 2.6b) would be highly accurate. Potential physical mechanisms for the success of the sign prediction are discussed in the next section.

Another obvious point in Figs. 2.6b and 2.6c is the similarity between the N06 and T11 approaches. In practice, the T11 test employed here is equivalent to the N06 test defined at a 40% threshold ($80\% \times 50\% = 40\%$). The one difference is that T11 further specify those grid points where more than 50% of models are significant but fewer than 80% agree on sign, which they classify as "no prediction." This last T11 criterion may be useful in evaluating P change under global warming, where at a given grid point, statistical significance of the P change for individual models does not necessarily mean they will agree on sign. In comparing the N06 and T11 procedures to the regions over which the models correctly predict sign of the observations, it is immediately apparent that the N06 and T11 tests are highly conservative. Although they do remove the modest fraction of points for which the sign would have been incorrectly predicted based on high agreement (passing the binomial test at the 98% level), they do so at the cost of excluding substantial regions that are correctly predicted. This is evident in Figs. 2.6b and 2.6c, where the hatched areas are restricted in extent relative to the broader shaded regions.

To show the sign agreement of the model ensemble with observations in more detail, we display in Fig. 2.7a the number of individual ensemble members that agree on sign with observations

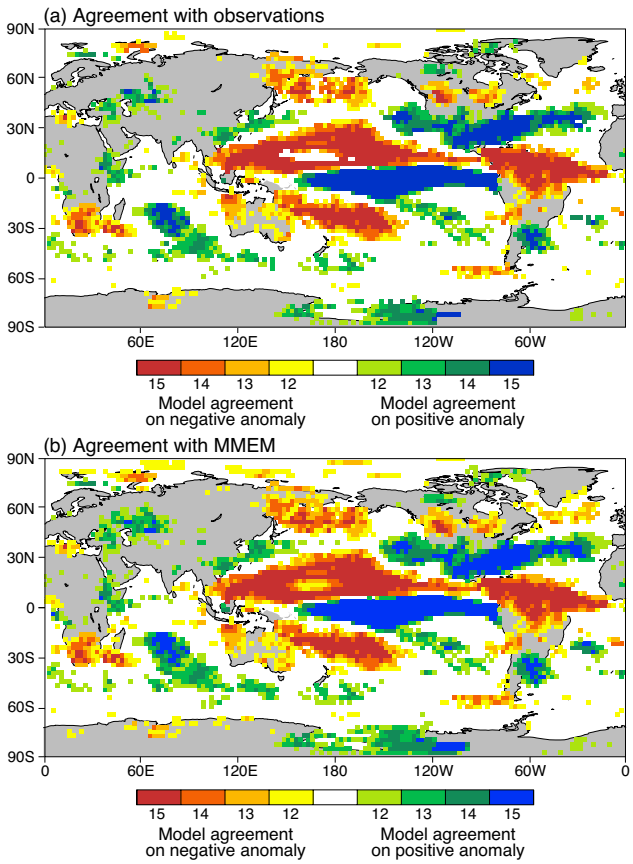

Figure 2.7: (a) Sign agreement of P teleconnections between each of 15 CMIP5 atmosphere-only models and the observations. (b) Sign agreement of P teleconnections between the CMIP5 atmosphere-only models and the ensemble mean, calculated using one run from each model. For (b), each model is individually removed from the ensemble mean before determining its sign agreement. Both (a) and (b) use Niño-3.4 teleconnection patterns diagnosed via linear regression. Red areas denote models that agree with the observations or ensemble mean on a negative P signal during ENSO events; blue areas imply agreement on a positive P signal.

for ENSO teleconnections. The same criterion for displaying high model agreement (12 or more models) is used as in Fig. 2.6a. Within this region, it may be seen that there are large portions in which the number of models agreeing on sign with observations is even higher, including substantial areas where 100% of models agree with the sign of the observations. To obtain a counterpart of this plot from the model ensemble, Fig. 2.7b shows the number of models agreeing with the sign of the ensemble mean. Note that in producing this, we exclude each model's contribution to the ensemble mean when determining agreement, so as to avoid inflating the count. The similarities between Figs. 2.7a and 2.7b indicate that high sign agreement with the ensemble mean can serve as a predictor for sign agreement with the observations.

2.5 Discussion

As discussed in the previous section, Figs. 2.6 and 2.7 suggest that there are substantial regions where models from the CMIP5 atmosphere-only ensemble are providing useful information on the sign of rainfall teleconnections, despite individual models and the observations failing to meet t test criteria at the 95% level in parts of these regions. We argue below that this is a combined consequence of the larger size of the model ensemble relative to individual runs, the nature of the quantity being tested (the sign), and the models' skill in predicting the observed sign.

Before addressing this, we consider the possibility that the broader region of skill at sign prediction in the ensemble (relative to individual model runs) could simply be an issue with applicability of the t test due to the inherent non-Gaussianity of the rainfall distribution, even at seasonal time scales. This was addressed in Fig. 2.2 by repeating the teleconnection calculations using Spearman's rank correlation, which makes no assumptions of Gaussianity for the grid point rainfall distributions, and an accompanying statistical significance test. This yields results similar to those of the linear regression t test.

We now consider an explanation based on the fact that the sign agreement both uses information from the full model ensemble and tests a different hypothesis than difference from zero. Because the collective 15-model ensemble contains a much larger set of realizations of internal variability, it is natural that regions of smaller signal should pass a given significance criteria in measures that use all 15 models. This is evident in comparing Fig. 2.6a to Fig. 2.1d, where areas of high sign agreement (passing the binomial test at the 98% level) tend to coincide with areas that pass a t test on the CMME at 95% confidence. In both cases the broad regions of statistical significance come from using all 15 models.

Taking this into account, we consider the question of why the models agree so well with the observations on the sign of the teleconnection patterns, despite doing poorly at detailed distribution. There are two aspects to this question: one statistical, and the other physical. The statistical aspect is that where the models exhibit sign agreement of 80%, the best estimate of the parameter P in the binomial distribution is 0.8. While it is beyond the scope of the paper to establish Bayesian posterior probability density functions or other measures of margin of error on the inferred P, the point needed to interpret the results here is straightforward: if the models are sufficiently good

representations of observations such that the observed signal can be considered to be drawn from a binomial distribution with a similar value of P at each point, then one would expect the high level of agreement seen. Thus, the 15-model ensemble shows success at predicting the sign of the observations in broader regions than those where teleconnection signals pass t tests applied to individual models or observations. If we consider the fact that these broader regions are those that pass the 98% confidence level of the binomial test, this success of the ensemble at sign prediction is completely consistent with expectations and with the statement that the models are doing well at simulating the observed sign.

The ability of models to provide information beyond what a particular significance test may suggest is not a new concept in modeled P studies. Risbey et al. (2011) resolve significant teleconnections in an atmosphere-only model using a 30-yr record and a two-tailed t test. The authors note that the number of grid points passing a 95% significance criterion is much fewer than the same method applied to a century of historical data. As a result, they loosen their restriction to an 80% confidence interval, noting that the associated teleconnection patterns are similar for records of either length. Power et al. (2011) evaluate projected P changes from the coupled CMIP3 model ensemble, and they demonstrate using the binomial distribution that model consensus on the sign of end-of-century rainfall anomalies is itself a strong argument for confidence in ensemble agreement patterns.

That the ensemble does, in fact, get broad areas of small-amplitude change correct in our teleconnection analysis adds to the discussion in the literature that projected change is worth assessing even in regions that do not meet t test criteria applied to individual runs (Tebaldi et al., 2011; Power et al., 2011) if these regions do meet significance tests applied to the ensemble. This is particularly relevant in global warming studies, where a modest regional P anomaly in a ensemble mean implies substantial changes in regional P budgets.

An important physical question that arises from the present teleconnection results is this: Why does the 15-model ensemble perform better at predicting the sign of the observed signal (including in broad areas of modest P amplitude response) and at yielding the amplitude of the observed response than the individual models do at reproducing detailed patterns of observed teleconnections? The unimpressive correlations (Fig. 2.4) are affected by poor individual model skill

in positioning high-amplitude signals. We suggest that this may be associated with the multiple physical processes operating in ENSO teleconnections. Specifically, there are atmospheric processes at work that will have smaller intermodel uncertainty and smaller internal variability but are widespread lly. An example of these processes is an increase in tropospheric temperature driving changes in radiative fluxes, as well as driving an increase in water vapor and a corresponding increase in the threshold for convection (the thermodynamic process sometimes referred to as the "rich-get-richer" mechanism; Chou and Neelin (2004); Held and Soden (2006); Trenberth (2011)). At the same time, feedbacks associated with dynamical changes in moisture convergence can produce large excursions from expected values of P, both in intermodel and temporal variability. The models contain reasonable approximations to each of these processes, but the location of strong P changes can be highly sensitive to factors such as model convective parameterizations, including the threshold for convective onset (Kanamitsu et al., 2002; Neelin et al., 2010).

2.6 Summary and conclusions

Atmosphere-only runs from the CMIP3 and CMIP5 ensembles provide one standard by which we can judge the ability of the atmospheric components of GCMs to reproduce dynamical feedback processes that lead to remote seasonal P anomalies. We focus on standard teleconnection patterns associated with the ENSO Niño-3.4 index. Comparisons among the ensemble of models and with the observations are made using P teleconnection patterns for the DJF for the years 1979–2005. The patterns and amplitudes of these teleconnections are analyzed in several regions with robust ENSO feedbacks, including the eastern tropical Pacific, the horseshoe region in the western tropical Pacific, a southern section of North America, and equatorial South America.

Teleconnection patterns are examined using three methods: linear regression, Spearman's rank correlation, and compositing techniques (not shown), all with similar results. The rank correlation method provides an alternative significance test, which is useful in narrowing some of the questions that arise for regions of low-amplitude signal. Teleconnection patterns defined with linear regression are useful for questions that involve the amplitude of the signal; as such, we focus on results from the linear regression.

How well the models perform at reproducing the observed teleconnection patterns (ampli-

tudes and patterns) depends strongly on the quantity for which they are assessed. In standard measures of correlation, taken over the regions outlined above, the CMIP3 and CMIP5 atmosphere-only models exhibit strong regional disagreement with one another and with observations. Comparing patterns visually, this is associated with regions of strong P change varying substantially from model to model and with respect to observations, yielding low correlations between modeled and observed teleconnection patterns (average correlation coefficients on the order of $r = 0.40$ in the defined regions).

The ensemble mean performs marginally better than most individual models in correlation measures, largely because the regions of strongest and varying change have been smoothed. However, the ensemble mean systematically underestimates amplitude measures of the regional P response by 30%–40%, typically falling more than one standard deviation below the central tendency of the 15-model ensemble. This underestimation is again associated with regional disagreement among ensemble members, a well-documented artifact in P studies of GCM ensembles (e.g., N06, Räisänen, 2007; Knutti et al., 2010; Neelin et al., 2010; Schaller et al., 2011). The average of individual CMIP5 atmosphere-only amplitudes, by contrast, is an accurate predictor for the observations in all regions but the central ENSO region, where models overestimate the P response. Sizeable internal variability of P teleconnections is also shown to exist within each model, although it does not dominate the intermodel spread.

One thing underlined by the low correlations in individual models is that even in atmosphere-only experiments, where only the atmospheric components of GCMs are being compared, simulation of ENSO teleconnections is fairly challenging for the models. While coupled models will have additional feedbacks, the atmosphere-only experiments provide a first line of assessment. Furthermore, because we can compare atmosphere-only simulations to observations, we can assess how the model simulations fare under other metrics commonly used in assessment of ensemble patterns and intermodel agreement.

Sign agreement measures for a P response in model ensembles are often used for assessing global warming P changes. Examining sign agreement for the teleconnection patterns, the model ensemble has broad regions with high consensus on sign, passing a binomial test (to reject the null hypothesis of 50–50 probability of either sign) at the 98% level. These regions are more spatially

extensive than the regions for which individual models (or observations) would pass a two-tailed t test at the 95% (or 90%) level. Furthermore, the regions passing the binomial test correspond well to the set of points passing a t test (at the 95% level) applied to the 15-model ensemble. Thus the larger region with high agreement on sign, relative to regions passing criteria (e.g., N06 or T11) that make use of t tests on individual models, is primarily the result of the sign agreement test making use of the 15-model ensemble.

For these teleconnection patterns, the sign prediction can be tested against observations. The models exhibit high sign agreement with observations over similarly broad regions, implying that high sign agreement within the model ensemble (grid points passing the binomial test at the 98% level) is a good predictor for sign agreement with observations. One can infer from this that the model ensemble is producing useful information regarding the teleconnected P signal in regions that do not pass a t test at the 95% level for individual models, provided they pass a significance test that makes use of information from the full ensemble.

The evaluation of the model simulations for ENSO teleconnections may be used, with due caution, to draw inferences for assessment of P in global warming projections. Many of the physical processes leading to rainfall teleconnections are analogous to the global warming case. In particular, widespread tropospheric warming initiates tropical dynamics that cause similar global P change in both teleconnections and global warming. In both cases, one can trace localized P anomalies with high amplitude and sizeable intermodel spread back to tropical regions of strong convergence feedbacks and regions where large-scale wave dynamics interact with midlatitude storm tracks.

The unimpressive skill of models at capturing the precise regional distribution of large-amplitude rainfall teleconnections compared to observations is consistent with poor intermodel agreement on a precise pattern of P change in global warming. However, the skill of individual models at reproducing the observed teleconnection signal amplitude (assessed from the mean of the individual model amplitudes, not the ensemble mean) suggests that corresponding measures for global warming P change may be trustworthy. Furthermore, sign agreement plots for the atmosphere-only ensemble prove skillful at predicting the sign of observed teleconnections. While agreement plots for end-of-century P change obviously have different patterns than the signals considered here, the fact that sign agreement plots are skillful at predicting spatially extensive

ENSO remote P impacts—which are challenging simulation targets that share physical pathways with global warming P signals—provides a supporting argument in favor of using sign agreement plots in global warming studies to make predictions of change from an ensemble of models.

Chapter 3

Patterns of precipitation change and climatological uncertainty among CMIP5 models, with a focus on the midlatitude Pacific storm track

Abstract

Projections of modeled precipitation (P) change in global warming scenarios demonstrate marked intermodel disagreement at regional scales. Empirical orthogonal functions (EOFs) and maximum covariance analysis (MCA) are used to diagnose patterns of disagreement in the simulated climatology and end-of-century P changes in the Coupled Model Intercomparison Project phase 5 (CMIP5) archive. The term "principal uncertainty pattern" (PUP) is used for any robust mode calculated when applying these techniques to a multi-model ensemble. For selected domains in the tropics, leading PUPs highlight features at the margins of convection zones and in the Pacific cold tongue. The midlatitude Pacific storm track is emphasized given its relevance to wintertime P projections over western North America. The first storm track PUP identifies a sensitive region of disagreement in P increases over the eastern midlatitude Pacific where the storm track terminates, related to uncertainty in an eastward extension of the climatological jet. The second PUP portrays uncertainty in a zonally asymmetric meridional shift of storm track P, related to uncertainty in the extent of a poleward jet shift in the western Pacific. Both modes appear to arise primarily from intermodel differences in the response to radiative forcing, distinct from sampling of internal variability. The leading storm track PUPs for P and zonal wind change exhibit similarities to the leading uncertainty patterns for the historical climatology, indicating important and parallel sensitivities in the eastern Pacific storm track terminus region. However, expansion coefficients for climatological uncertainties tend to be weakly correlated with those for end-of-century change.

3.1 Introduction

Accurate prediction of end-of-century P change as a result of global warming is critical to the assessment of future changes in the hydrological cycle, especially at regional scales relevant to water resource management and decision-making. Global climate models (GCMs) run as part of the Coupled Model Intercomparison Project phases 3 and 5 (CMIP3 and CMIP5; Meehl et al., 2007; Taylor et al., 2012) exhibit robust agreement over large-scale changes related to thermodynamic arguments that regions of climatological moisture convergence will get wetter (the tropics and mid-to-high latitudes), while regions of climatological moisture divergence will get drier (the subtropics). This concept was first noted by Manabe and Stouffer (1980) and Manabe and Wetherald (1980) and is now referred to as the wet-get-wetter or rich-get-richer mechanism (Chou and Neelin, 2004; Held and Soden, 2006; Trenberth, 2011). It has been documented extensively in the CMIP archives, and the patterns it describes can be seen in individual models as well as in the multi-model ensemble (MME) mean of projected P changes (see Figs. 1a,b). Recent studies have noted that these changes are detectable in ocean observations (Durack et al., 2012), but their validity can break down at the regional or grid point level (Chadwick et al., 2012; Roderick et al., 2014), especially for changes over land (Greve et al., 2014). Alongside these thermodynamic changes are notable dynamic changes in the atmospheric circulation that ultimately affect P projections, typically framed in terms of a poleward expansion of subtropical dry zones or the descending branch of the Hadley cell, as well as a poleward shift, upward expansion, and slight widening of the climatological storm track (Yin, 2005; Lu et al., 2007; Seager and Vecchi, 2010; Scheff and Frierson, 2012b,a; Chang et al., 2012, see Figs. 1a,b).

Outside of a large-scale or mean meridional framework, however, intermodel agreement on P changes can diminish greatly. The standard deviation of grid point (local) P changes across models is one way of quantifying this spread (see Figs. 3.1c,d), and complex P changes occurring in the tropics or along convective margins can lead to large disagreement in tropical P projections (Neelin et al., 2003; Chou and Neelin, 2004; Chou et al., 2009), especially over land (e.g., Yin et al., 2013), even when models agree on bulk measures of tropical change, such as distributions of P intensity (Lintner et al., 2012). Substantial regional uncertainty also exists in areas with large internal variability in the P climatology, leading to small signal-to-noise ratios in future projections

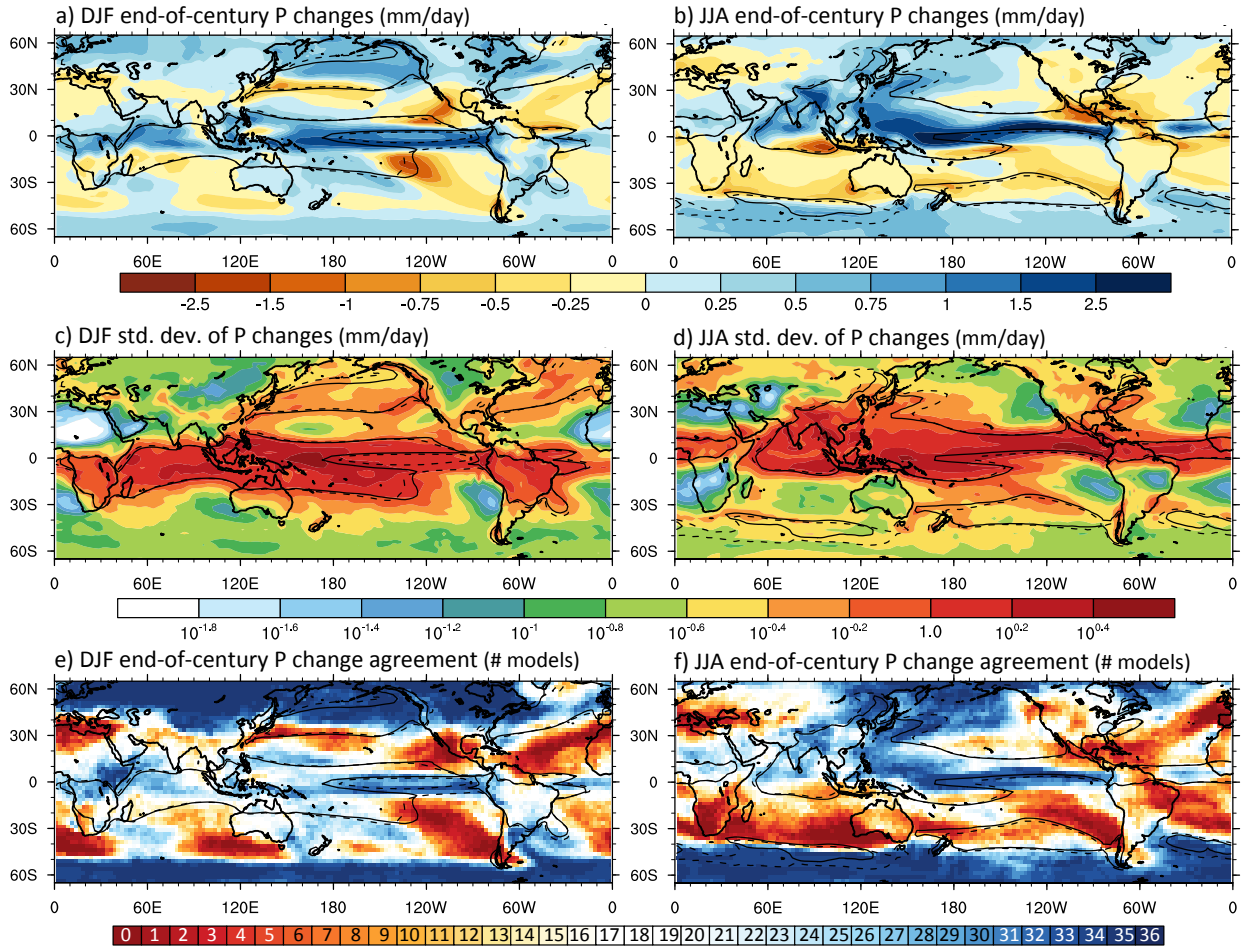

Figure 3.1: (a,b) 36-member ensemble mean end-of-century P changes for the DJF and JJA seasons, respectively. Anomalies are calculated for the 2070–2100 climatology relative to a 1960–1990 base period. (c,d) Standard deviation of P anomalies across the 36-model ensemble. (e,f) Agreement on positive P anomalies for the same models and seasons. In all maps, solid (dashed) lines represent the 1960–1990 (2070–2100) 4 mm day^{-1} P climatology.

and to noticeable uncertainty and conflicting model changes (Giorgi and Francisco, 2000; Allen and Ingram, 2002; Neelin et al., 2006; Tebaldi and Knutti, 2007; Hawkins and Sutton, 2011; Tebaldi et al., 2011; Deser et al., 2012; Mahlstein et al., 2012; Knutti and Sedlácek, 2013; Roderick et al., 2014). These regions are often located between P increases at mid-to-high latitudes and decreases in the subtropics, where thermodynamic arguments become less dominant and projections are susceptible to disagreement in dynamic feedbacks at regional scales (Chou et al., 2009; Seager and Vecchi, 2010; Shepherd, 2014).

In addition to uncertainty in P projections, models exhibit a range of abilities in simulating historical P climatology and internal variability in CMIP3 and CMIP5 (Deser et al., 2012; Flato

et al., 2013). This is especially true for the tropics, where considerable ocean-atmosphere feedbacks, as well as differences in model parameterization of deep convection, can cause discrepancies in the P climatology (e.g., Mechoso et al., 1995; Zhang, 2001; Lin, 2007; Brown et al., 2013) and projected changes (Collins et al., 2011; Brown et al., 2012) among models. GCM biases also exist outside of the tropics, especially at regional scales where local dynamics come into play (e.g., Kumar et al., 2013; Langford et al., 2014; Mehran et al., 2014). Finally, only modest (if any) improvement has been found between CMIP3 and CMIP5 in these biases or in constraining regional model disagreement in end-of-century changes (Knutti and Sedlácek, 2013; Sheffield et al., 2013a,b; Maloney et al., 2014; Hirota and Takayabu, 2013).

One region that is particularly sensitive to P change uncertainty is the North American west coast in the latitude range 30–50°N. This area lies at the eastern terminus of the Pacific wintertime storm track and is situated between robust P decreases at lower latitudes and increases at higher latitudes (Neelin et al., 2013; Berg et al., 2014; Seager et al., 2014). In this region, P exhibits substantial variability on interannual to interdecadal timescales—especially in the North American southwest and in southern California (Dettinger et al., 1998)—so that in future projections, signal-to-noise ratios tend to be small and dynamical changes complex (Seager and Vecchi, 2010; Seager et al., 2014). Studies have also noted GCM uncertainty in simulating the climatological Pacific storm track feeding into this region (Delcambre et al., 2013a), as well as widespread disagreement in end-of-century storm track changes (Yin, 2005; Bengtsson et al., 2006; Ulbrich et al., 2008; Ihara and Kushnir, 2009; Chang et al., 2012; Chang, 2013; Delcambre et al., 2013b; Grise and Polvani, 2014; Simpson et al., 2014).

In this chapter, we visualize intermodel uncertainty patterns for P projections and climatologies in the CMIP5 ensemble, and we seek coupled or associated patterns of uncertainty in circulation and temperature fields. We do this by applying empirical mode decomposition techniques—commonly used in the space-time domain—to the "space-model index" domain. In the latter half of our analysis, we emphasize the wintertime midlatitude Pacific storm track region, and in doing so we illustrate the benefits of cross-checking with complementary methods and show that using several in tandem can strengthen the conclusions drawn for each.

Table 3.1: Models and affiliated modeling groups used in the analysis. When additional ensemble members are used for certain models, this is marked in the N_E column. When pre-industrial control runs are used to estimate internal model variability, the number of distinct 30-year intervals are reported in the N_{30} column. Asterisks mark the models for which direct atmosphere-only and coupled comparisons were made (25 total).

Coupled models	$\mathbf{N_E}$	$\mathbf{N_{30}}$	**Atmosphere-only models**	**Modeling center or group**
ACCESS1-0		16	ACCESS1-0*	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia (CSIRO-BOM)
ACCESS1-3		16	ACCESS1-3*	
bcc-csm1-1-m		13	bcc-csm1-1-m*	Beijing Climate Center, China Meteorological Administration, Beijing, China (BCC)
bcc-csm1-1		16	bcc-csm1-1*	
BNU-ESM		18	BNU-ESM*	College of Global Change and Earth System Science, Beijing Normal University, Beijing, China (GCESS)
CanESM2	4	33	CanAM4*	Canadian Centre for Climate Modelling and Analysis, Québec, Canada (CCCMA)
CCSM4	5	35	CCSM4*	National Center for Atmospheric Research, Boulder, Colorado, USA (NCAR)
CESM1-BGC		16	—	
CESM1-CAM5		10	CESM1-CAM5*	
CMCC-CESM		9	—	Centro Euro-Mediterraneo per I Cambiamenti Climatici, Lecce, Italy
CMCC-CM		10	CMCC-CM*	
CMCC-CMS		16	—	
CNRM-CM5	4	28	CNRM-CM5*	Centre National de Recherches Météorologiques, Toulouse, France
CSIRO-Mk3-6-0	9	16	CSIRO-Mk3-6-0*	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence, Victoria, Australia
EC-EARTH	5	14	EC-EARTH*	EC-EARTH consortium
FGOALS-g2		23	FGOALS-g2*	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
GFDL-CM3		16	GFDL-CM3*	NOAA Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA
GFDL-ESM2G		16	—	
GFDL-ESM2M		16	—	
—			GFDL-HIRAM-C180	
—			GFDL-HIRAM-C360	
GISS-E2-H		7	—	NASA Goddard Institute for Space Studies, New York, New York, USA
GISS-E2-R		18	GISS-E2-R*	
—			HadGEM2-A	Met Office Hadley Centre, United Kingdom (additional HadGEM2-ES realizations contributed by Instituto Nacional de Pesquisas Espaciais)
HadGEM2-AO		23	—	
HadGEM2-CC		7	—	
HadGEM2-ES		9	—	
inmcm4			inmcm4*	Institute for Numerical Mathematics, Moscow, Russia
IPSL-CM5A-LR		33	IPSL-CM5A-LR*	Institut Pierre Simon Laplace, Paris, France
IPSL-CM5A-MR		9	IPSL-CM5A-MR*	
IPSL-CM5B-LR		9	IPSL-CM5B-LR*	
MIROC5	2		MIROC5*	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies, Tokyo, Japan
MIROC-ESM-CHEM		8	—	
MIROC-ESM		17	—	
MPI-ESM-LR		33	MPI-ESM-LR*	Max Planck Institute for Meteorology, Hamburg, Germany
MPI-ESM-MR		33	MPI-ESM-MR*	
—			MRI-AGCM3-2H	Meteorological Research Institute, Tokyo, Japan
—			MRI-AGCM3-2S	
MRI-CGCM3		16	MRI-CGCM3*	
NorESM1-ME		8	NorESM1-ME*	Norwegian Climate Centre, Bergen, Norway
NorESM1-M		16	NorESM1-M*	

3.2 Data, methods, and terminology

3.2.1 CMIP5 data

The primary ensemble in our analysis consists of 36 fully-coupled GCMs; historical forcing and Representative Concentration Pathway 8.5 (RCP8.5) simulations were used for each model. An additional ensemble of 30 atmosphere-only runs was also used. Table 3.1 lists model information for both ensembles. In some cases, additional ensemble members for a given coupled model were used in order to estimate internal variability; for these models, the total number of additional realizations is included in the second column of Table 3.1.

We downloaded the following fields, all at monthly resolution: precipitation (P), zonal and meridional winds at 200 hPa (U200 and V200) and 850 hPa (U850 and V850), surface air temperature (TAS), and skin temperature (TS). We treat the models in each ensemble as independent samples, though we acknowledge that some of them share structural cores and may have common biases (Jun et al., 2008b,a; Knutti et al., 2010, 2013). Our goal is to understand intermodel uncertainty relative to the ensemble mean, not to the "true climate" or observations.

Seasonal end-of-century changes for the RCP8.5 scenario were calculated for each model by differencing the climatology at the end of the 21st Century (2070–2100) and over a 20th Century base period (1960–1990). The historical climatologies of the atmosphere-only models cover 1979–2009. 30-year averages were chosen to minimize interannual-to-decadal model variability; all fields were regridded via bilinear interpolation to a common $2.5^\circ \times 2.5^\circ$ grid prior to analysis.

3.2.2 Methods

The methods are described in more detail in the Chapter 3 supplemental information, though a brief summary is provided here, taking P as the example variable. We use EOF analysis to visualize patterns of model uncertainty in end-of-century P changes. These EOFs are calculated across the model dimension—as opposed to the conventional time dimension—and the resulting modes show patterns of intermodel disagreement over end-of-century P changes relative to the ensemble mean. Grid point correlation maps between expansion coefficients of P change uncertainty patterns and temperature or winds allow one to find associated uncertainties in other variables. We

also use maximum covariance analysis (MCA; sometimes referred to as singular value decomposition, or SVD). MCA is performed on the covariance matrix between two variables (e.g., P changes and U200 changes), producing pairs of uncertainty patterns that represent coupled ensemble disagreement.

Our methods complement other studies that have used similar methods to examine uncertainty patterns in GCMs, including intermodel disagreement in tropical P change and its connection to SST uncertainties (Li and Xie, 2012; Ma and Xie, 2012; Li and Xie, 2013), GCM skill in modeling historical midlatitude jets, uncertainty in jet changes, and relationships with tropical SST uncertainties (Delcambre et al., 2013a,b), and relationships between large-scale SST uncertainties and global warming trends in land surface P (Anderson et al., 2015). Similar methods have also been used in the weather forecasting community for medium-range ensemble forecasts (e.g., Harr et al., 2008; Keller et al., 2011; Zheng et al., 2013; Chang and Zheng, 2013).

3.2.3 Principcal Uncertainty Pattern (PUP) terminology

For brevity, we refer to all modes that arise from the matrix decomposition techniques described above as principal uncertainty patterns (PUPs), and we use this term interchangeably with "patterns" or "modes" from the EOF and MCA analyses when discussing results. This is done to emphasize that these techniques all seek to capture robust patterns of uncertainty common among models in an ensemble, even though the methodology used to calculate each may differ.

3.3 End-of-century P changes in the CMIP5 ensemble

To lay the basis for our discussion, Figs. 3.1a,b show the ensemble mean global end-of-century P change in the RCP8.5 scenario, for the December–January–February (DJF) and June–July–August (JJA) seasons. Seasonal 4 mm day^{-1} contours for the base (solid) and future (dashed) periods are shown for the ensemble mean to establish approximate geographical boundaries for regions of deep convection and storm track precipitation.

Localized standard deviation plots (Figs. 3.1c,d) give a sense of how individual model P changes spread about the ensemble mean. In both seasons, the largest spread is over the oceans and

within tropical zones of deep convection, implicating intermodel uncertainty in changes occurring to moisture fluxes and deep convective processes in these regions. In DJF (Fig. 3.1c), the largest intermodel spread is over the western Pacific warm pool and South Pacific Convergence Zone (SPCZ). Secondary regions include the cold tongue region off the coast of Peru and Ecuador, the Atlantic ITCZ, and the Indian Ocean. In JJA (Fig. 3.1d), the western and eastern Pacific are still major areas of spread, as are the tropical Indian and Atlantic Oceans and the Asian monsoon region. Note these hotspots of disagreement do not always coincide with areas of largest absolute P change (Figs. 3.1a,b) or P change as a percent of the historical climatology (Figs. S3.1a,b).

We also show agreement plots on the sign of P changes (Figs. 3.1e,f; plotted for agreement on positive change). Relative to the standard deviation maps, the white regions in agreement maps reach wider into the extratropics and midlatitudes, where intermodel spread is a result of changes to both regional dynamics and local thermodynamics. Our concern here is P change disagreement at regional scales (i.e., at the scale of climate features such as convection zones or storm tracks), though an equivalent global analysis is shown in the Chapter 3 supplemental information for absolute P changes (Figs. S3.2 and S3.3) and P change as a percent of the base period climatology (Figs. S3.4 and S3.5). These patterns show tropical dominance in intermodel disagreement, even with normalization by the climatology, and so we turn to a regional PUP analysis.

3.4 Regional PUPs of P change disagreement

3.4.1 Criteria for selecting regional PUP domains

We select regional domains using the following criteria: (1) if the regional PUPs are to have any significance, there should be sizeable disagreement in the region of interest; (2) objective pattern-seeking techniques are sensitive to the domain size and characteristics, and edge effects can arise as domain boundaries change; (3) seasonality comes into play when disagreement patterns are the objective, since zones of heavy P and storm track domains will shift seasonally, and it is important not to segment climatological features.

3.4.2 Regional PUPs using EOF analysis

For DJF, we produce P change PUPs for the following regions: the Indian Ocean, the tropical Pacific, the N. Pacific storm track region, and equatorial S. America. For JJA, the regions include: the Asian monsoon region, the central and western tropical Pacific, the S. Pacific storm track region, Central America (encompassing the eastern tropical Pacific), and the equatorial Atlantic region. This regional analysis reveals second-order and structural characteristics that are not revealed in simpler analyses like those in Figs. 3.1c–f.

The first two EOF PUPs for these domains are shown in Figs. 3.2 (DJF) and 3.3 (JJA). In order to plot PUPs on a common color bar, each mode is normalized by its standard deviation. Figs. 3.2c and 3.3c show the percent variance that successive modes account for in the EOF analyses (cut off after 15 modes). For DJF, two regions of uncertainty include the southern edge of the Indian Ocean convection zone and the tropical Pacific at the western edge of the cold tongue. For JJA, one can see uncertainty along the southern edge of the ITCZ and along the northern edge of the cold tongue region in the tropical Pacific.

Table 3.2 lists details about normalization, latitude and longitude range of the domains, and the percent variances accounted for by the displayed modes. Leading regional PUPs tend to explain 20%–30% of intermodel uncertainty, and features commonly occur on edges of convection zones, within storm tracks, and over the equatorial cold tongue. We also include the correlation coefficient between expansion coefficients for the regional and global domains, "r(reg, glob)," noting that in some cases PUPs can change order (or be spread across several modes). These correlations tend to be higher for the tropical domains in Figs. 3.2 and 3.3, reinforcing the dominance of the tropics in global-scale uncertainty, although there are also notable correlations at higher latitudes.

We focus on the DJF midlatitude Pacific storm track region for the remainder of this study. We chose the Pacific storm track region after noting strong separation and clear signals in the first two PUPs (see Figs. 3.2a,b for the modes in this domain, as well as Fig. 3.2c to get a sense of mode separation).

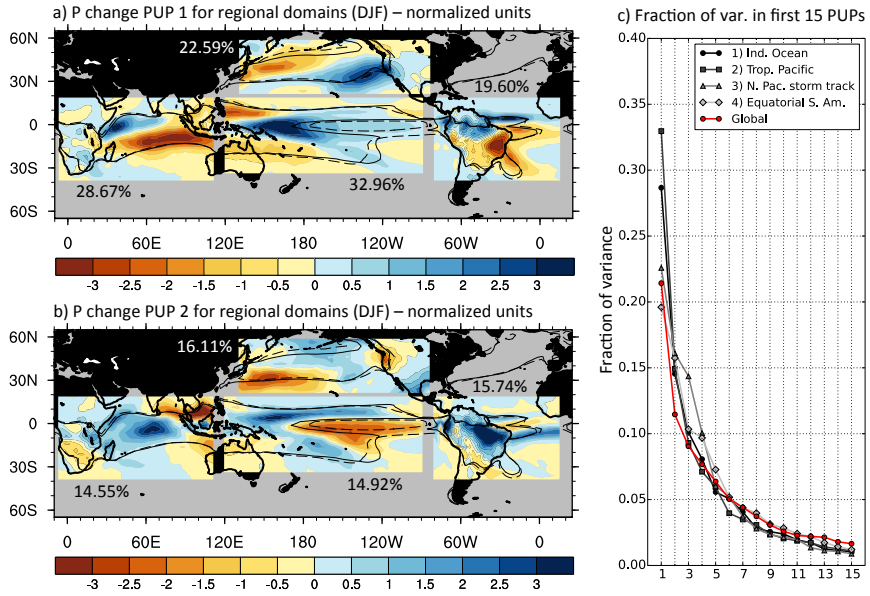

Figure 3.2: (a,b) First and second PUPs for the regions shown in DJF. From left to right, these regions are: The Indian Ocean region; the tropical Pacific region; the North Pacific storm track region; the equatorial South America and Atlantic region. For each mode and region, the variance of intermodel disagreement that the PUP accounts for is written as a percent. (c) The fraction of variance accounted for in each region for the first 15 modes. Tropical regions show a much stronger response when P units are used, so to plot patterns on a common color bar, each region has been normalized by the standard deviation of its EOF values; a unit of ± 1 therefore represents a "standard" departure from zero in that region. More detail on these regions and units can be found in Table 3.2.

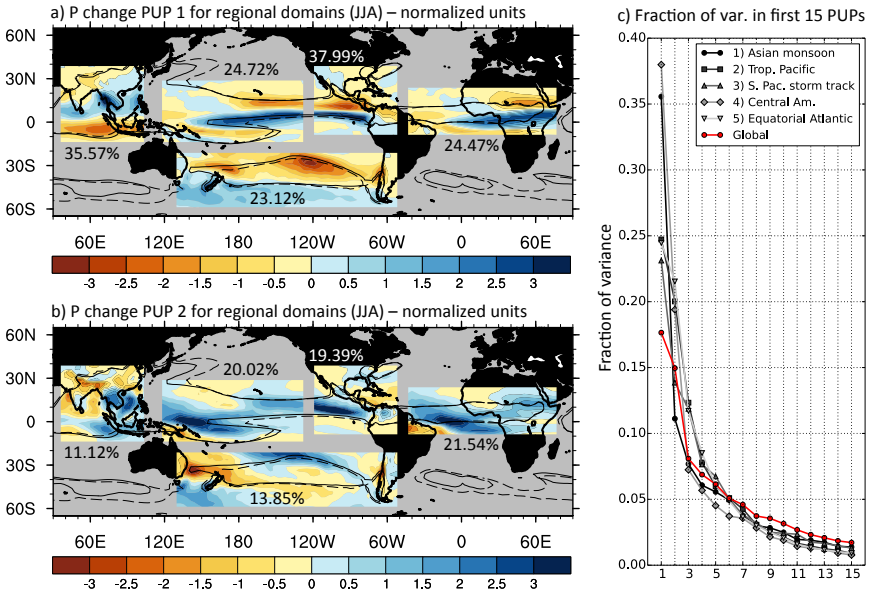

Figure 3.3: As in Fig. 3.2, but for the JJA season. (a,b) First and second PUPs for each of the regions shown in JJA. From left to right, these regions are: The Asian monsoon region; the tropical Pacific region; the South Pacific storm track region; the Central America region; and the equatorial Atlantic. The variance fraction accounted for by that mode is included next to each region as in Fig. 3.2. (c) The fraction of variance accounted for in each region for the first 15 modes. The units and color bar for PUP patterns are discussed in Table 3.2.

Table 3.2: Details for the regional PUP analysis shown in Figs. 3.2 and 3.3. Regional information is shown in column 1. For the first two modes in each region, we list the percent variance accounted for in that mode (column 3), the correlation coefficient between the expansion coefficients for that mode and the expansion coefficients for the corresponding global modes 1 and 2 (columns 4 and 5), and the standard deviation of the EOF in that domain (column 6, used for normalizing patterns in Figs. 3.2 and 3.3). Because the sign of any EOF mode is arbitrary, we show the absolute value of the correlation coefficient. Correlations are bold where they pass a two-tailed test at the 95% confidence level.

Region (DJF)	**reg mode**	**% var.**	$\|r$**(reg, glob1)**$\|$	$\|r$**(reg, glob2)**$\|$	σ**(mm day**$^{-1}$**)**
Indian Ocean	1	28.67%	**0.57**	**0.71**	0.44
40°S–20°N, 10°E–120°E	2	14.55%	**0.53**	0.31	0.32
Tropical Pacific	1	32.96%	**0.98**	0.16	0.70
35°S–20°N, 125°E–265°E	2	14.92%	0.03	**0.53**	0.47
N. Pacific storm track	1	22.59%	0.03	0.12	0.21
20°S–60°N, 135°E–270°E	2	16.11%	**0.45**	**0.61**	0.18
Equatorial South America	1	19.60%	0.14	**0.80**	0.32
40°S–20°N, 270°E–365°E	2	15.74%	0.11	**0.35**	0.27
Region (JJA)	**reg mode**	**% var.**	$\|r$**(reg, glob1)**$\|$	$\|r$**(reg, glob2)**$\|$	σ**(mm day**$^{-1}$**)**
Asian monsoon	1	35.57%	**0.94**	0.10	0.76
15°S–40°N, 65°E–125°E	2	11.12%	0.04	**0.62**	0.42
Tropical Pacific	1	24.72%	0.31	**0.78**	0.67
15°S–30°N, 135°E–235°E	2	20.02%	**0.83**	0.33	0.54
S. Pacific storm track	1	23.12%	0.06	0.06	0.18
60°S–20°N, 145°E–300°E	2	13.85%	**0.58**	0.17	0.14
Central America	1	37.99%	0.08	**0.91**	0.71
10°S–40°N, 240°E–300°E	2	19.39%	**0.75**	0.03	0.50
Equatorial Atlantic	1	24.47%	0.25	**0.80**	0.73
10°S–25°N, 305°E–400°E	2	21.54%	**0.86**	0.30	0.57

3.5 Uncertainty in P change in the midlatitude Pacific storm track domain

3.5.1 Storm track P change PUPs and associated uncertainties in circulation changes

In the midlatitude Pacific, the wintertime storm track is a preferred region of cyclone activity, generated by baroclinic instability and acting to transport momentum, heat, and moisture poleward. We use the 4 mm day^{-1} contour as an approximate storm track outline in the base and future periods. Cyclones that produce precipitation in this region are steered by the upper-level jet across the Pacific and trail into the North American west coast, and so we inspect uncertainties in P change in the context of the broader upper- and lower-level circulation. Delcambre et al.

(2013a,b) have extensively documented uncertainties in the jet for this region, and we will discuss model disagreement in a similar way by highlighting the two leading modes:

1. The "jet extension" mode describes disagreement in the eastward extension of the jet and resulting P changes at the eastern terminus of the storm track, where the 4 mm day^{-1} contour hooks slightly northeast at ∼35°N, 220°E.

2. The "meridional shift" mode describes uncertainty in the extent of a meridional shift or poleward displacement of the storm track P, with increases on the poleward edge and decreases on the equatorward edge of the 4 mm day^{-1} contour (see Figs. 3.1a,c).

Figure 3.4 shows the first (a,c) and second (b,d) PUPs of the EOF analysis on P changes in the storm track region (replotted from Figs. 3.2a,b but with units of mm day^{-1}). Vectors representing coupled disagreement in upper- and lower-level wind changes are overlaid, the components of which are regressions of end-of-century U-wind and V-wind changes onto the expansion coefficients for each mode.

We identify the first PUP (Figs. 3.4a,c) as a jet extension mode, with an east-to-west dipole showing P increases on the southeast flank of the storm track as well as decreases over the western Pacific. The wind vectors form a cyclonic circulation pattern slightly to the northwest of this center at both upper (Fig. 3.4a) and lower (Fig. 3.4c) levels, implying that to a first approximation this pattern represents an equivalent barotropic mode of disagreement.

We identify the second PUP (Figs. 3.4b,d) as a meridional shift mode, showing drying along the equatorward edge of the storm track in the western half of the domain and implying disagreement among models in the location and magnitude of poleward P shifts in this region. Upper-level wind regressions (Fig. 3.4b) show associated anticyclonic uncertainties in circulation along the equatorward edge of the storm track.

We use the terminology "extension" and "shift" to give names to the uncertainty patterns, though the changes occurring to the midlatitude jet and resulting P in the midlatitude Pacific are asymmetric and complex. One robust feature found in the CMIP5 archive is a poleward jet shift in the western Pacific but equatorward shift in the east (Simpson et al., 2014; Park and An, 2014). Simpson et al. (2014) attribute this equatorward shift in the eastern Pacific to a barotropic

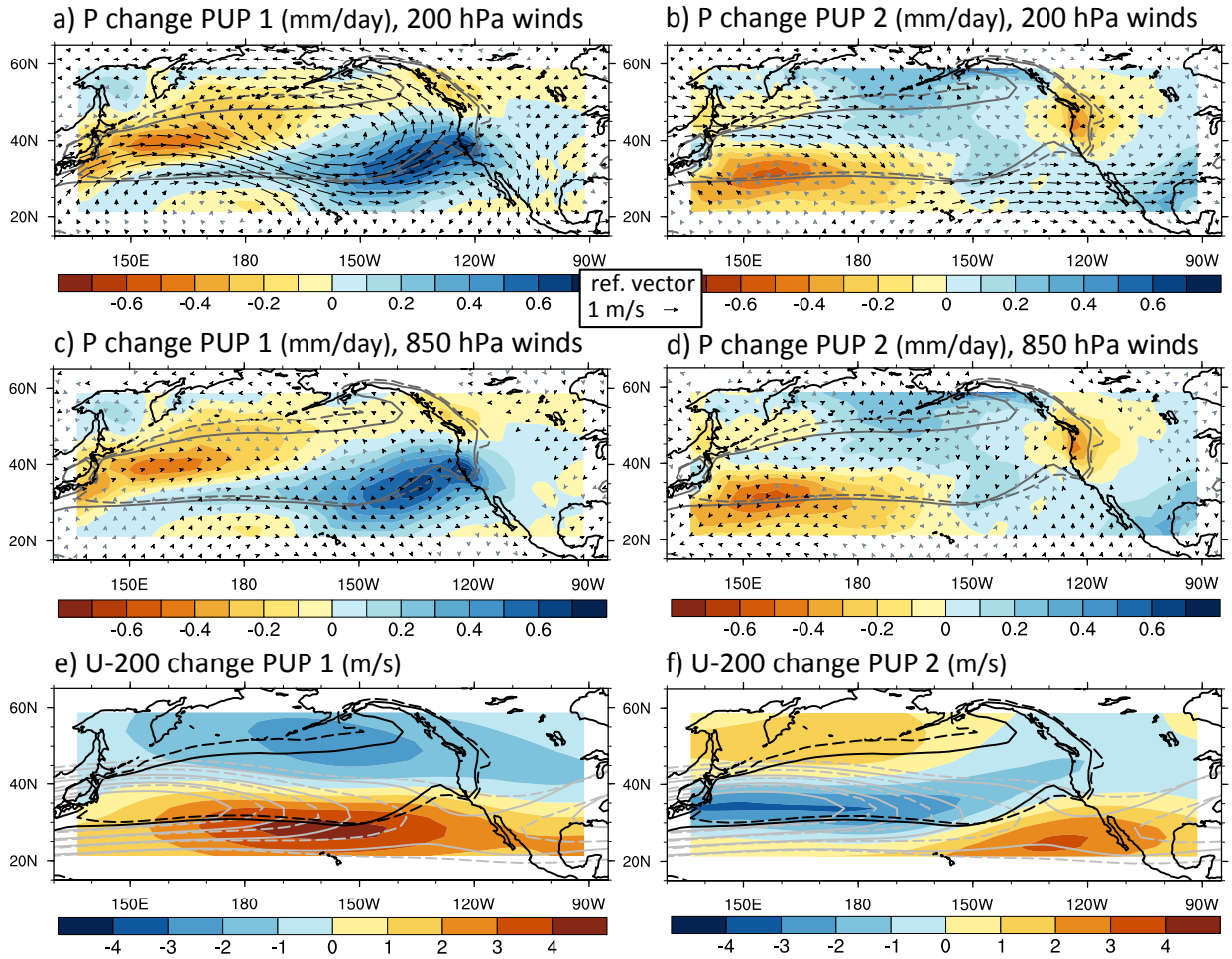

Figure 3.4: (a-d) First and second P change PUPs for the midlatitude Pacific storm track region during DJF. (a,b) P change PUPs 1 and 2 (shading) with 200 hPa U- and V-winds change regressions. (c,d) as in (a,b) but for winds at 850 hPa. Black solid and dotted contours are as in Fig. 3.1. Wind vectors are drawn in proportion to the 1 m s^{-1} reference vector. Vectors are plotted black where at least *one* component passes a two-tailed t test at the 90% confidence level, and gray otherwise. (e,f) First and second EOF U200 change PUPs during DJF. Gray solid contours show isotachs for the base period ensemble mean, and dashed contours show isotachs for the end-of-century ensemble mean, starting at 30 m s^{-1} and plotted at intervals of 10 m s^{-1}. The first P change PUP accounts for 22.59% of intermodel variance in P change projections, and the second P change PUP an additional 16.11%. The first U200 change PUP accounts for 41.80% of intermodel variance, and the second mode an additional 23.13%

stationary wave anomaly off the west coast of North America. Neelin et al. (2013) investigated the CMIP5 ensemble mean P changes occurring in this region, noting a slight but statistically significant increase in the region where the climatological storm track approaches the California coast, between ∼33°N and 42°N. Seager et al. (2014) attributed this increase primarily to enhanced mean flow moisture convergence. The circulation regressions in Figs. 3.4a,b may reflect model uncertainty in

changes to the mean circulation (and associated stationary wave patterns), which would feed into uncertainty in P change in the storm track termination region.

The first and second U200 change PUPs are displayed in Figs. 3.4e,f. Note that the first mode (Fig. 3.4e) reinforces the jet extension interpretation: models exhibit disagreement in U200 increases in the jet core, providing a source of uncertainty in the extent to which storms are steered onto the North American west coast. The second mode (Fig. 3.4f) seems to suggest intermodel spread over the extent of a poleward shift of the jet in the western Pacific. However, it is actually the third mode for U200 changes (not shown) that is more strongly related to the meridional shift P change PUP (see supplementary information for details). To better see coupled uncertainties between jet and P change uncertainty, we show MCA results in Section 2.5.3.

3.5.2 Internal variability versus intermodel uncertainty

To get a sense of how individual models contribute to the uncertainty patterns in Fig. 3.4, expansion coefficients for P change PUPs are shown in Figs. 3.5a,b. Red dots represent expansion coefficients for each of the 36 models in the ensemble and characterize total ensemble spread—i.e., a combination of internal variability, climatological uncertainty, and intermodel differences in the response to radiative forcing. The ensemble mean is represented by the zero line, and the box and whisker plots to the right show the overall spread of models about the mean (see caption for details). For models with additional ensemble runs available, black dots show the projection of those P changes onto the PUPs in Figs. 3.4a–d. The relative spread of the black and red dots indicates that internal model variability is small compared to overall intermodel uncertainty.

We also estimate internal variability using pre-industrial control runs, which vary in length from 240–1050 years in the models. We calculate the DJF P climatology for non-overlapping 30-year periods and produce N_{30} climatologies for each model (see Table 3.1 for details). We center these climatologies by that model's long-term mean at each grid point and then take all possible pairwise combinations without repeat. The difference between each pair is taken, and the result is projected onto the P change PUPs in Figs. 3.4a–d. The resulting values represent expansion coefficients for internal variability. The error bars in Figs. 3.5a,b show ± 1 standard deviation of these values for each model, centered at zero on the vertical axis, representing the spread can that

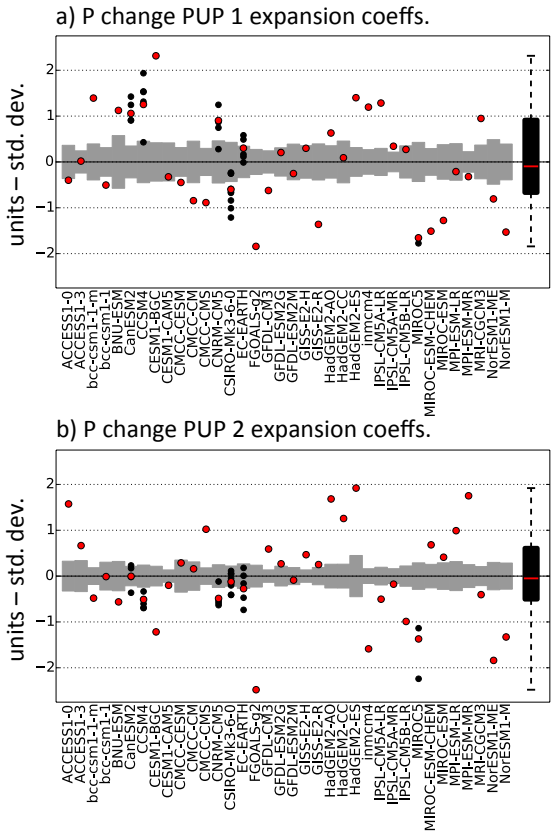

Figure 3.5: (a,b) Expansion coefficients for the first and second P change PUPs, respectively. Red dots represent each of the 36 models in the ensemble. Black dots show expansion coefficients for additional ensemble members (see text and Table 3.1 for details). Error bars show internal variability estimation using pre-industrial control runs in the models (see text for details).

can be attributed to internal P variability alone in the storm track region. This tests whether the intermodel spread seen by the red dots in Figs. 3.5a,b could arise solely from internal variability. The magnitude of these error bars is a modest fraction of the ensemble spread, providing evidence that to leading order, intermodel differences in the response to radiative forcing are likely driving much of the spread captured by these uncertainty patterns, and internal variability does not contribute substantially. We explore the role of interannual SST variability and in particular ENSO forcing in the next section.

3.5.3 Extension and shift modes from MCA

MCA helps reveal coupled patterns in P and U200 change uncertainties that are seen qualitatively in Fig. 3.4. In Fig. 3.6, we show MCA results for coupled P and U200 change PUPs, along

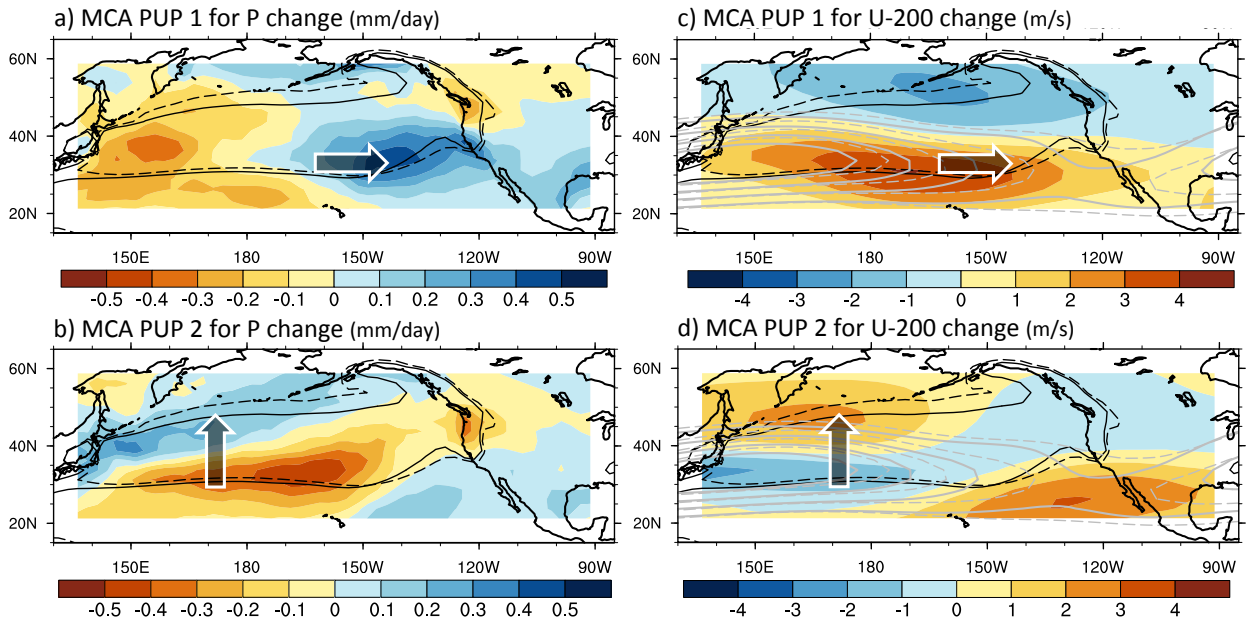

Figure 3.6: First and second MCA PUPs showing end-of-century coupled disagreement between DJF P and U-200 changes. (a,b) First and second left singular vectors for P in units of mm day^{-1}. (c,d) First and second right singular vectors for U-200 change in units of m s^{-1}. Note the color bar is reversed in (a,b) vs. (c,d) in order for blue to coincide with positive P changes, and red to coincide with positive U-200 changes. Schematic arrows are included to depict the jet extension and shift interpretations. The covariance fraction is 31.83% for the first PUP and 23.47% for the second. Black and gray contours are as in Figs. 3.4e,f.

with diagrammatic arrows to illustrate the extension and shift modes. Figures 3.6a,c show the first coupled mode patterns; an arrow in Fig. 3.6a shows location of strong P change disagreement in the storm track termination region, and an identical arrow in Fig. 3.6c shows collocated model uncertainty in U200 increases in the historical jet core. Figures 3.6b,d show the second coupled mode patterns; the arrow in Fig. 3.6b depicts intermodel disagreement in the poleward displacement of the storm track in the western half of the domain, and the arrow in Fig. 3.6d shows associated disagreement where the U200 jet is displaced from its historical maximum. MCA here becomes a useful tool for isolating coupled patterns of uncertainty in P and U200 change uncertainties that are not obvious in the separate EOF analyses.

Lastly, we have tested the patterns in this storm track region and note they are not sensitive to domain size, in that changing the domain boundaries by a few degrees latitude or longitude causes the fraction accounted for by individual modes to change only within a few percent, and the appearance and qualitative details of the PUPs themselves are stable.

3.6 Multivariate PUPs for P and surface temperature

Figures 3.7a,b show global correlation maps between end-of-century TAS changes and the expansion coefficients of P change PUPs in the storm track region. TAS anomalies were pattern-scaled prior to any calculations, in order to minimize the influence of model differences in warming; this was done by dividing each model by its global annual end-of-century TAS change between the 1960–1990 and 2070–2100 periods. When this pattern scaling is applied to P changes prior to calculating EOFs (not shown), results are similar.

The correlation map for TAS and the first PUP (Fig. 3.7a) identifies a strong connection between the jet extension and a localized meridional TAS gradient in the mid-to-high latitude Pacific. For the second mode, the TAS correlations (Fig. 3.7b) show a connection between the Pacific cold tongue region and storm track P change disagreement.

The tropical correlations in Fig. 3.7b may suggest that uncertainty in storm track P changes is tied to uncertainty in tropical SSTs via tropical-to-midlatitude teleconnections. Such telecon-

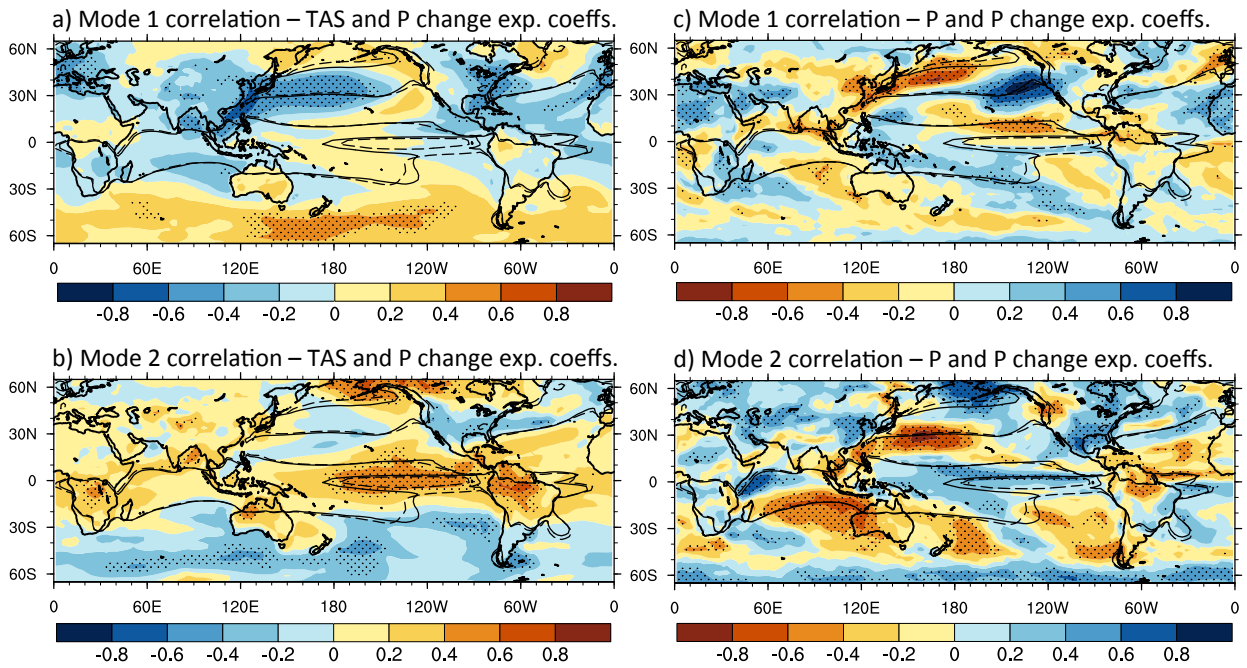

Figure 3.7: Correlation maps between expansion coefficients from P change PUPs over the storm track region (as in Figs. 3.4a-d) and either (a,b) end-of-century model TAS changes across the model ensemble at each grid point, or (c,d) P changes. For (a,b), TAS anomalies were pattern scaled by the annual end-of-century TAS change (global mean for each model) prior to analysis. Stippling shows where correlations pass a two-tailed t test for statistical significance at the 95% confidence level.

nections are sometimes referred to as "ENSO-like," though we take caution in using this term and define ENSO forcing as a specific SST pattern arising from ocean dynamics in the cold tongue region, as well as deep tropospheric latent heating anomalies in the tropical Pacific that initiate teleconnections (see Zhang et al., 1997). If the uncertainties in SST were truly ENSO-like, one may expect a region of high correlation in tropical P changes in the central-to-eastern Pacific that resemble typical ENSO P anomalies. Figures 3.7c,d show correlation maps between grid point P changes and the expansion coefficients in Fig. 3.5. For both modes, the tropical Pacific correlation pattern for P is quite different from a canonical ENSO P anomaly.

To better quantify the role that intermodel differences in ENSO play in these PUPs, we produce DJF ENSO composite maps for SST and P in each model's ∼150-year historical run, project these onto the maps in Fig. 3.7, and then calculate the correlation between these projection values and the corresponding TAS or P change PUP expansion coefficients. This projection method tests whether model differences in the strength and location of historical ENSO SST and P anomalies are associated with uncertainties in end-of-century P changes in the Pacific storm track. Projections are done for ocean-only grid points between 30°N and 30°S, and ENSO events are defined when the Niño 3.4 index (Trenberth, 1997) is more than one standard deviation above (positive) or below (negative) the long-term DJF mean. Composites were calculated by extracting linearly detrended P and SST fields for ENSO events, averaging positive and negative events separately, and subtracting the two averages (positive-minus-negative). Correlation values are $r = 0.18$ and $r = 0.18$ for the first two correlation maps of TAS (Figs. 3.7a,b), respectively. Correlation values are $r = 0.15$ and $r = 0.54$ for the correlations maps of P change PUPs (Figs. 3.7c,d). These correlation values imply that uncertainties in ENSO composites are able to explain at most about 28% of intermodel uncertainty associated with the two leading PUPs. Correlation values are comparable when SST is substituted for TAS in Figs. 3.7a,b, and ENSO composites calculated for the RCP8.5 runs give similar results.

We have also defined ENSO-like SST forcing using the first EOF of internal SST variability for each model, which bypasses any requirement that models have correctly placed SST anomalies. These patterns were calculated for detrended DJF SST data between 30°N and 30°S in the historical period. We proceed as above by projecting each model's tropical SST pattern onto Figs. 3.7a,b

and correlate the resulting values with model expansion coefficients. Correlations remain low at $r = 0.20$ and $r = 0.03$, and we conclude that P change PUPs in the Pacific storm track region are not a manifestation of ENSO variability among models.

That said, common physical mechanisms will be active in the historical and end-of-century time periods, and so the uncertainty patterns and modes of internal variability may still resemble another. The extension and shift modes in the storm track region are similar to the those described by Lau (1988), which are related to a "pulsing" and "wobbling" of the Pacific midlatitude jet and have implications in storm development and P variability along the North American west coast (e.g., Wettstein and Wallace, 2010; Athanasiadis et al., 2010). These jet-driven storm track fluctuations interact with other modes of circulation variability in the Northern Hemisphere (e.g., Wallace and Gutzler, 1981). For instance, the western Pacific (WP) dipole pattern seen in the second mode of U200 change in MCA-based PUPs (Fig. 3.6d) is similar to the WP pattern investigated by Linkin and Nigam (2008), who identify this as a circulation associated with the North Pacific Oscillation (NPO) and discuss its link to Pacific storm track fluctuations. Cayan (1992) found that NPO and WP variability at midlatitudes can have broad-scale effects on sensible and latent heat fluxes in the Pacific ocean, affecting tropical SSTs. Complementary studies have discussed the effect of the NPO on sea level pressure anomalies in the midlatitude Pacific, which induce broad SST anomalies that persist into the following summer, affecting trade winds and tropical Pacific climate variability (Vimont et al., 2003; Anderson, 2004; Chiang and Vimont, 2004). Caballero and Anderson (2009) show that the WP pattern can also influence the strength of the descending branch of the Hadley cell via stationary wave forcing, which will also affect tropical trade winds and tropical Pacific climate.

It is therefore not out of the question that physical mechanisms underlying P change PUPs in the midlatitude Pacific storm track have qualitative analogues to internal climate variability. However, we contend that the processes giving rise to P change PUPs in the storm track region are to first order distinct from internal climate variability.

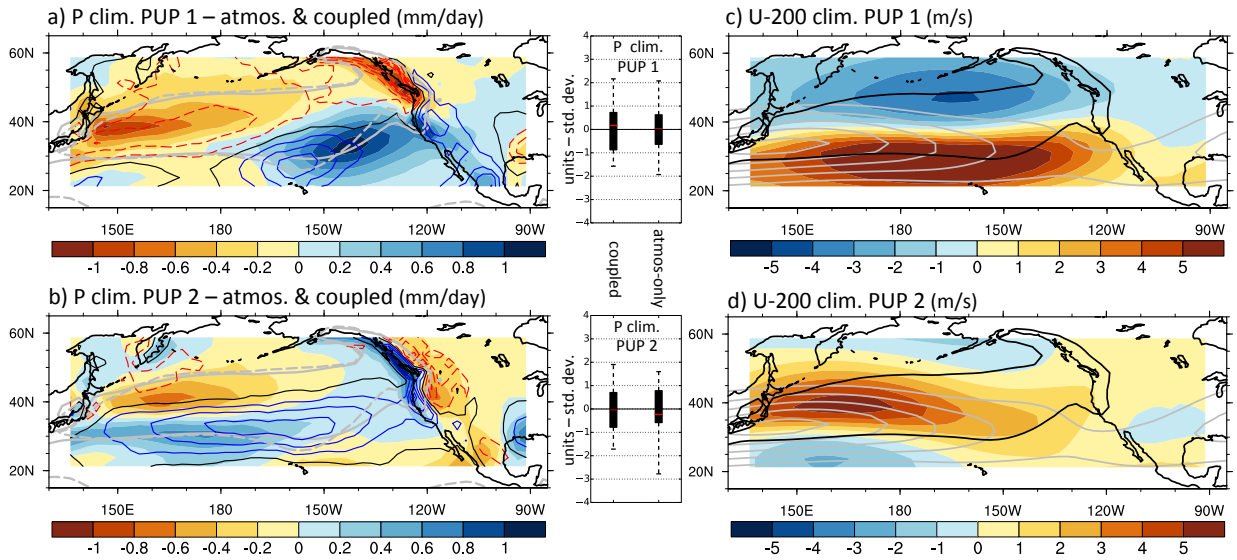

Figure 3.8: (a,b) First and second PUPs for the historical P climatology, in both coupled and atmosphere-only runs. Shaded contours show PUPs for fully-coupled historical runs (36-model ensemble, for the period 1960–1990), accounting for 26.16% and 16.48% of the variance, respectively. Blue and red contours are overlaid to show PUPs for atmosphere-only runs (30-model ensemble, for the period 1979–2009), accounting for 25.75% and 18.88% of the variance, respectively; lines are spaced at an interval of 0.2 mm day^{-1} with the zero contour drawn in black. Box and whisker plots to the right of each mode show the median (red line), the first and third quartiles (box limits), and the total range (whiskers) of expansion coefficients for the coupled and atmosphere-only runs; units are in standard deviation (i.e., the "typical departure" from the ensemble mean). (c,d) First and second EOF PUPs for the base period U200 climatology (atmosphere-only U200 PUPs not shown). The first two modes capture 52.55% and 18.16% of the variance. The color bar is in units of m s^{-1}. End-of-century P and U-200 contours are included to show the storm track outline and the climatlogical jets in the base (solid) and future (dashed) period.

3.7 Patterns of climatological uncertainty and parallels to P change PUPs

We also ask whether uncertainties in the base period P climatology are related to those for P changes. Figures 3.8a,b show PUPs calculated across model P climatology fields during the base period. Shading is for the coupled ensemble (36 members), and blue/red contours depict PUPs for the atmosphere-only ensemble (30 members). These maps represent intermodel uncertainties relative to the MME, and we refer to them as climatology PUPs.

Comparing atmosphere-only (contours) and coupled (shading) PUPs, the disagreement patterns are similar. While there are geographical shifts in emphasis, the first mode (Fig. 3.8a) depicts an approximate jet extension, along with intermodel spread in the exact latitude that wintertime P falls along the North American west coast. The second mode (Fig. 3.8b) appears to represent a combination of climatological disagreement over orographic P or land-sea contrasts (where the

signal is strongest along the coast), as well as a meridional shift over the western Pacific. In general, the enhanced intermodel spread along the North American coast in Figs. 3.7a,b is not seen in Fig. 3.4. This implies that similar orographic or land-sea contrast uncertainties are common to both the historical and end-of-century periods and hence are not present in differences between the two. As we found previously, this continues to suggest that end-of-century PUPs are depicting intermodel uncertainty in structural *changes* to climatological patterns, not amplification of those patterns.

To understand how important the coupled ocean is in producing these climatological uncertainties, we calculate PUPs for the 25 coupled models whose atmospheric component is the same in the atmosphere-only runs (see Table 3.1 for details). We then project the atmosphere-only climatologies in the storm track region (centered by the ensemble mean) onto the first two coupled climatological PUPs and correlate these projection values with the coupled expansion coefficients. This allows one to see how much of the overall uncertainty in coupled runs is present in the atmosphere-only runs. Correlation values are $r = 0.65$ and $r = 0.91$ for the first two modes, respectively, implying that uncertainty within the atmospheric component of the models alone is responsible for a significant amount of the disagreement seen in coupled climatological PUPs.

To quantify whether climatological P uncertainties can predict P change uncertainties among models, we project centered historical climatologies for the 36-model ensemble onto the change PUPs in Figs. 3.4a–d and then correlate these projections with the corresponding expansion coefficients. The resulting correlations are $r = 0.12$ and $r = 0.47$ for the first and second modes, respectively (and cross-mode correlations hover between these values). This implies that within our ensemble, uncertainty patterns in the storm track P climatology do not serve as strong predictors for the uncertainty seen in the extension mode, though they do have some predictive capacity for the shift mode. When comparing climatological PUPs within the historical and RCP8.5 periods, the relationship is much stronger. Using the same technique by calculating climatology PUPs in the RCP8.5 runs and projecting historical runs onto them, the correlations are $r = 0.89$ and $r = 0.94$ for modes one and two. In other words, there is a strong relationship between intermodel uncertainty in the historical and RCP8.5 climatologies, but when considering the end-of-century *differences*, the relationship drops significantly. This also supports our earlier point that P change PUPs represent uncertainty in radiatively forced changes to individual models' climatologies.

Figures 3.8c,d show PUPs for uncertainties in the U200 climatology. We repeat the analysis above by projecting U200 climatology fields onto the U200 change PUPs from Fig. 3.4, and we then calculate the correlation between these projection values and the original U200 change PUP expansion coefficients. The correlation values are $r = 0.47$ and $r = 0.42$ for the first two modes, implying that U200 climatological uncertainties can have a modest predictive value for end-of-century change uncertainties. In other words, information about jet shift uncertainty in the historical climatology might provide some information about the jet shift response to radiative forcing, but this does not translate well to information about P change.

We have also explored relationships between U200 PUPs and global P signals for changes and the base period climatology, as well as internal variability in the historical period (see the Chapter 3 supplemental information for details). We find that there are qualitative similarities among all three cases, especially in the storm track region, though their global relationships appear to be different, and the correlations patterns are fairly weak. We conclude that storm track P change uncertainty patterns cannot be explained by any measure of model differences in internal variability, but are driven primarily by uncertainty in individual model responses to radiative forcing.

3.8 Summary and conclusions

In this chapter, we analyze uncertainties in P climatologies and end-of-century changes in the CMIP5 ensemble. We borrow objective matrix decomposition techniques from the time series domain to calculate modes of variability across models. This produces patterns of intermodel disagreement in P projections, and we refer to them generally as principal uncertainty patterns (PUPs) to emphasize that the relationships they show among fields are of interest independent of the specific method used to calculate them. PUPs are produced using EOF analysis as well as MCA and correlation/regression techniques, allowing one to identify leading modes of uncertainty in one field and possible relationships with other fields.

Focusing on a domain encompassing the wintertime midlatitude Pacific storm track, the first P change PUP represents intermodel uncertainty in the eastward extension of the steering jet and resulting P changes over the North American west coast; this is labeled the "jet extension" mode. The second PUP depicts intermodel disagreement in the extent of a meridional storm track

shift, but this signal is confined to the western Pacific and shows zonal asymmetry farther east; this is labeled the "meridional shift" mode. Regressions of changes in upper- and lower-level winds onto these PUPs reveal uncertainties in circulation that reinforce these interpretations, and MCA analysis provides a check on coupled uncertainty in the the zonal wind and P fields.

Measures of internal model variability using both additional ensemble members and pre-industrial control runs show that internal model variability is not a dominant source of intermodel uncertainty in P change PUPs in this region. This is found for both long term variability and interannual ENSO forcing.

Correlations of TAS changes with the extension and shift PUPs reveal intermodel uncertainties in larger-scale temperature fields associated with Pacific storm track P change PUPs. The first mode shows that uncertainties in changes to meridional temperature gradients are associated with uncertainties in P increases at the eastern terminus of the Pacific eddy-driven jet. The second mode shows positive correlations between tropics-wide TAS increases and a poleward storm track P shift in the western Pacific.

In the historical climatologies, P and U200 PUPs show patterns that are similar to the extension and shift modes. While models demonstrate a strong relationship between historical and future climatological uncertainties, these relationships are weak for climatological PUPs and end-of-century *change* PUPs in the storm track region. We conclude that the P change PUPs in the storm track likely arise from intermodel uncertainty in radiatively forced structural changes to individual model climatologies.

Finally, we reiterate the value of this approach as a set of tools for analyzing patterns of intermodel spread, and we note the applicability of the "PUP" acronym to any objective process seeking robust uncertainty patterns in an ensemble. Our results highlight the importance of informed domain selection and cross-checking interpretations with complementary procedures. These methods have been used in previous model studies and contribute to the growing body of work in addressing, characterizing, and constraining intermodel uncertainties within an ensemble.

Chapter 3 supplementary material

S3.1 End-of-century P changes as a percent of the climatology in CMIP5

Figures S3.1a,b show end-of-century P change calculated as a percent of the base period climatology. In general, the magnitude of percent increases in certain regions are substantially larger than percent decreases (as the latter should be bounded by -100%). Note also that these maps are for the ensemble mean, though several individual models give unphysical results, specifically with percent increases in the Sahel region and the central equatorial Pacific (and these models are excluded from the ensemble mean in Fig. S3.1). These unrealistic percent values make EOF analysis difficult, which is discussed later in this supplement.

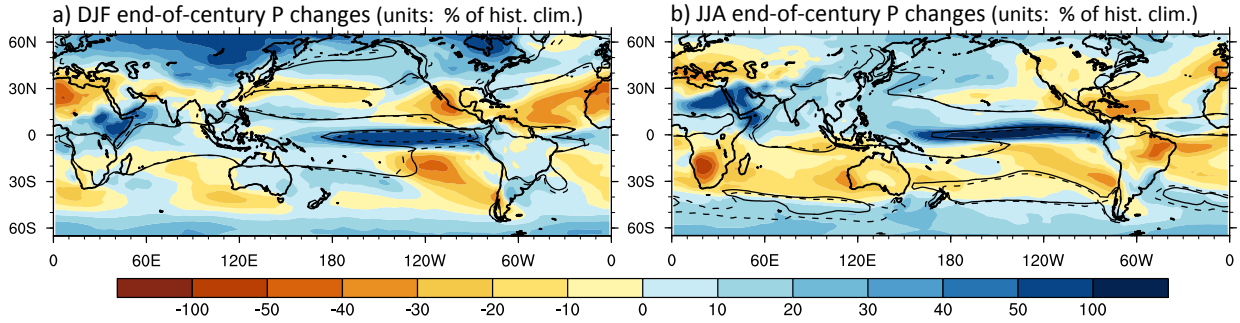

Figure S3.1: P change as a percent of the base period climatology at each model grid point, shown for (a) DJF and (b) JJA.

S3.2 Methods in more detail

In the following subsections, we go into partial detail of the mathematics behind our analyses and refer the reader to a few helpful publications that discuss the techniques more comprehensively: Horel (1981), Wallace and Gutzler (1981), Richman (1986), and Hannachi et al. (2006) for EOFs

or rotated EOFs; Bretherton et al. (1992) and Wallace et al. (1992) for coupled methods like MCA and correlation maps; von Storch and Zwiers (1999) and Wilks (2011) for textbook discussions of all of these methods, as well as alternatives.

S3.2.1 EOF analysis for single variables

To aid in our explanation, let $\mathbf{X}$ be a $N \times M$ matrix consisting of all data from the ensemble and with N rows and M columns. Here, M is the number of models in the ensemble and N is the total number of grid points in the chosen domain, while the matrix X can represent end-of-century changes for a field of interest at N grid points (e.g., end-of-century P changes in CMIP5). Furthermore, let $\mathbf{X'}$ be the ensemble matrix centered by the mean of each row of $\mathbf{X}$ (i.e., centered by the ensemble mean at each grid point). Finally, let $\mathbf{C} = (1/M)(\mathbf{X'X'^T})$ be the $N \times N$ covariance matrix of the centered ensemble matrix.

To visualize leading modes of P change disagreement within the 36-model ensemble, we begin with global and regional EOF analyses on end-of-century P change fields, which amounts to an EOF analysis on the covariance matrix $\mathbf{C}$. The decomposition yields a set of N eigenvectors, which are the modes of variability or disagreement across the ensemble of M models, as well as N corresponding eigenvalues, which give insight into the percent of overall variance in the ensemble that each mode accounts for. Note in this analysis that M is typically much less than N, and only the leading M eigenvalues and eigenvectors (out of N total) have any physical meaning. Positive and negative values in modes should not be interpreted strictly as P increases or decreases, but rather model uncertainty in the exact placement of these increases and decreases—as well as the boundaries between them—relative to the ensemble mean.

Once the EOFs are obtained, expansion coefficients (or principal components) are calculated by projecting that model's original centered P change map onto the EOF mode of interest. This produces 36 different scalar expansion coefficients for the end-of-century EOF calculations, one for each model. The magnitude of each expansion coefficient represents the contribution of that model to the disagreement pattern; the larger the magnitude, the more it contributes (either positively or negatively, relative to the ensemble mean) to intermodel uncertainty. We refer to these values as either expansion coefficients or model weights.

S3.2.2 MCA for coupled modes of intermodel disagreement

We explore coupled modes of uncertainty between P and U200 fields using maximum covariance analysis (MCA). MCA is a matrix decomposition performed on the cross-covariance matrix between two fields. As an example, let the matrix $\mathbf{P}$ be a $N \times M$ matrix of precipitation anomalies, for M models at N grid points. Let $\mathbf{U}$ be a $K \times M$ matrix consisting of zonal wind anomalies for M models at K grid points. Let $\mathbf{P'}$ and $\mathbf{U'}$ be the original matrices centered by their column means at each grid point and converted into standardized units. This standardization step prevents differences in units from affecting the results and is done by dividing the centered P data by the standard deviation of the flattened P anomalies matrix, and likewise for the centered U200 data and $\mathbf{U}$ matrix. Finally, let $\mathbf{S} = (1/M)(\mathbf{P'U'^T})$ be the $N \times K$ cross-covariance matrix between the two fields.

MCA calculates patterns explaining the maximum amount of covariance in a data set. MCA will factor the matrix $\mathbf{S}$ into three matrices, $\mathbf{S} = \mathbf{\Pi \Sigma Y}^T$. Columns in $\mathbf{\Pi}$ are referred to as left singular vectors, columns in $\mathbf{Y}^T$ are right singular vectors, and each pair has a singular value associated with it along the diagonal in the matrix $\mathbf{\Sigma}$. Each pair of singular vectors comprises a coupled mode of covariability between the two fields (i.e., coupled disagreement pattern between the two variables relative to the ensemble mean of each field).

For every pair of singular vectors, the singular value in $\mathbf{\Sigma}$ can be used to calculate the covariance fraction (CF) between the two fields in that mode. We choose to report this instead of the squared covariance fraction; see Cheng and Dunkerton (1995) for further discussion.

S3.3 Global PUPs of P change disagreement

Figures S3.2 and S3.3 show the first and second PUPs from a global EOF analysis on P changes, for the DJF and JJA seasons, respectively. To the right of each mode, we display the expansion coefficients for each ensemble, normalized to have unit variance so that a value of ± 1 on the vertical axis implies one standard deviation departure from the ensemble mean P change. In these results, global EOFs are computed over all latitudes, though the patterns are weak poleward of about 60°, and so we crop the maps slightly in the figures to show better detail.

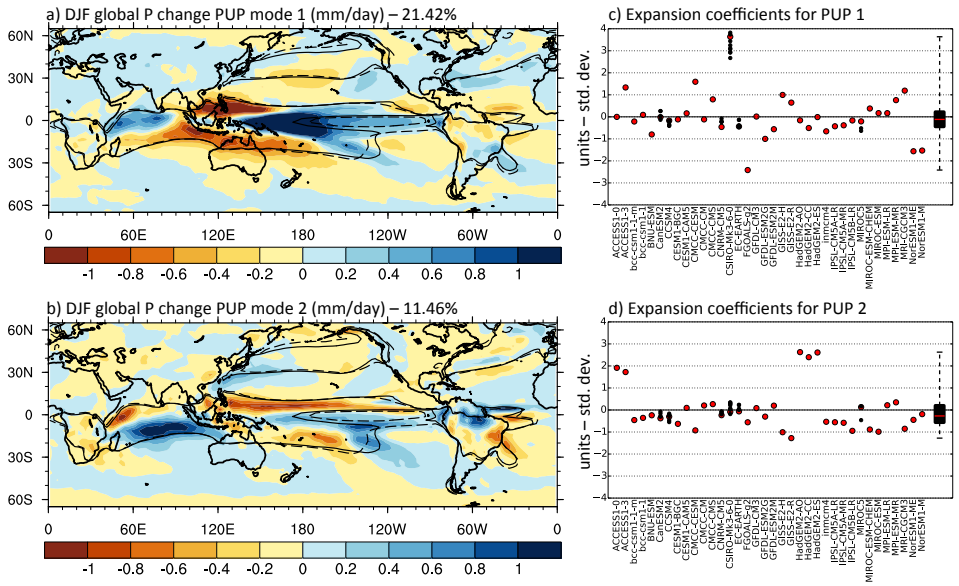

Figure S3.2: (a,b) First and second P change PUPs from a global EOF analysis on the 36-model ensemble P change maps, calculated for the DJF season in mm day^{-1}. Solid and dashed P contours are identical to those of Fig. 3.1. Percent variances accounted for by each PUP are included at the top of each figure. (c,d) Individual model expansion coefficients corresponding to the modes on the left, in units of standard deviation relative to the ensemble mean P change. Red and black dots and box and whisker plots are as described in the main text.

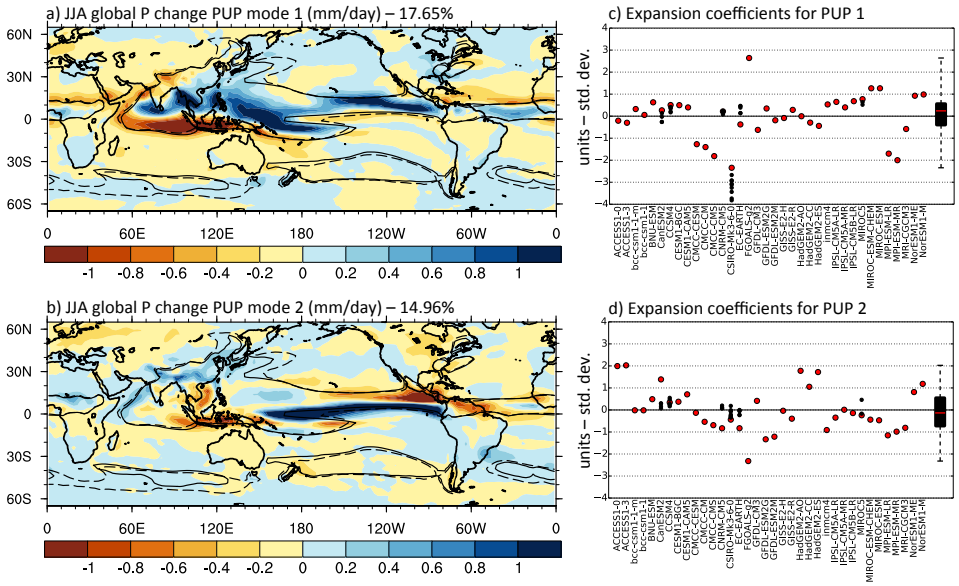

Figure S3.3: As in Fig. S3.2, but for the JJA season.

In both seasons, the first few modes are dominated by model disagreement in P change within the tropics. For DJF, the first PUP (Fig. S3.2a) has a strong uncertainty signal in the western Pacific warm pool; the second PUP (Fig. 3.2b) reflects disagreement primarily in the amount of P decreases over the eastern flank of the SPCZ. For JJA, the first PUP (Fig. S3.3a) appears to highlight model disagreement in the position of the eastern Pacific ITCZ, though the entire tropics have patches of uncertainty of similar magnitude. The second PUP (Fig. S3.3b) is even more tropically confined, again showing disagreement in displacement of the ITCZ and showing continuous signal along convective margins throughout the Pacific Ocean basin.

S3.3.1 Intermodel disagreement versus internal model variability

The expansion coefficients shown in Figs. S3.2c,d and S3.3c,d give a sense of how individual models contribute to the corresponding PUP. Each red point represents an individual run from the model on the horizontal axis and characterize total ensemble spread, a combination of internal variability and differences in the individual models' response to radiative forcing. In some cases, a few particular models are noticeable outliers and project strongly onto the global PUPs, but this is not consistent across seasons. There is, however, notable consistency among groups of models that are taken from the same modeling center. For example, the signs and magnitudes of the expansion coefficients show little variation among the three HadGEM2 models, and this is true in all seasons and PUPs. The robustness of these patterns against the removal of outlying models has been tested by eliminating CSIRO and is discussed in the following section.

The box and whisker plots on the right in Figs S3.2c,d and S3.3c,d show the overall spread of the red points. Note the red line within each box and whisker plot represents the median of the red points, and the horizontal zero line represents the ensemble mean. The sign of the median gives information about the model distribution about the ensemble mean for this given pattern, with a negative median indicating right skew (a longer tail in the positive direction), and a positive median indicating left skew.

Along with red points in Figs. S3.2c,d and S3.3c,d, the smaller black points give a sense of how internal model variability compares to intermodel spread. For six models, we have downloaded additional ensemble runs from the CMIP5 database to calculate end-of-century P changes for

each (see Table 3.2 for more detail), and these maps are projected onto the full-ensemble PUPs. Additional runs of a given model (black points) exhibit a variability range that is a small fraction of the intermodel spread (red points). More quantitatively, in both leading global modes and seasons, the variance of the black points for a given model is less than 5% of the entire ensemble variance, except for the CSIRO-Mk-3-6-0 model in the DJF season, which is about 13%. This is approximately true for the midlatitude Pacific storm track region as well (see the main text), though internal model variability is a slightly larger fraction.

S3.3.2 EOFs of P change as a percent of the base period climatology

Given that global PUPs calculated on absolute P change are dominated by the tropics, it is useful to consider normalizing model P changes prior to EOF analysis in a way that could emphasize higher latitudes while down-weighting the tropical influence. One possible approach—dividing the P change at each grid point by the base period climatology—becomes difficult in practice because nearly half of the models in the ensemble show unphysical percent P increases on the order of $10^3\%$ or more, particularly in the Sahel region and the central equatorial Pacific.

Masking these unrealistic grid points also proves unfruitful. Figures S3.4 and S3.5 show PUPs where grid points with P changes outside of the range $[-100\%, +250\%]$ were masked prior to EOF analysis (and show up as white grid points in the figures). While this approach does allow one to see more detail on signals of intermodel uncertainty outside of the tropics, the criteria for masking are somewhat arbitrary, and the information one gains about regional uncertainty patterns from Figs. S3.4 and S3.5 relative to Figs. S3.2 and S3.3 does not increase substantially. This is the reasoning behind the focus on absolute P change PUPs in Chapter 3 and the emphasis on regional domain analyses.

S3.3.3 Rotated EOFs

We have tested whether varimax EOF rotation (Kaiser, 1958) provides a clearer understanding of the P change PUPs in both the global and midlatitude Pacific storm track domains. Rotation of global modes does not clarify any disagreement patterns already evident in the local standard deviation plots and original EOFs. Rotation of the midlatitude Pacific storm track PUPs

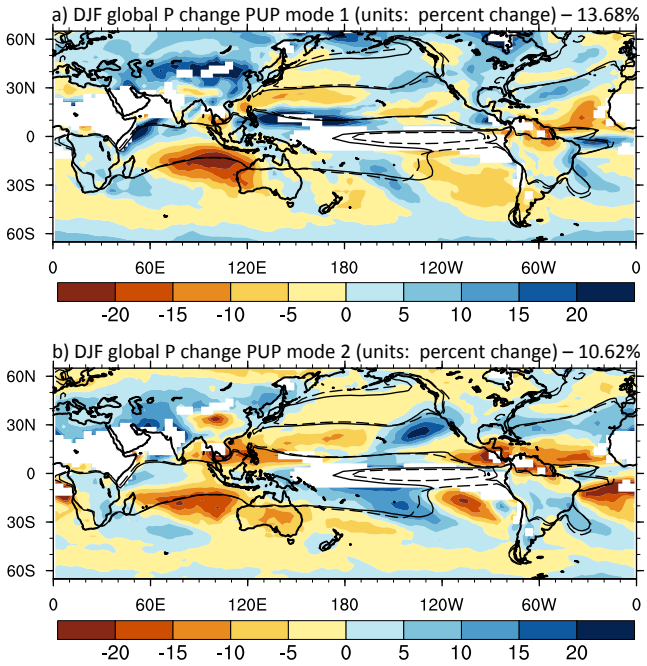

Figure S3.4: (a,b) First and second P change PUPs calculated over P changes as a fraction of the base period climatology for DJF. Areas with a P change outside of the range $[-100\%, +250\%]$ were masked prior to EOF analysis.

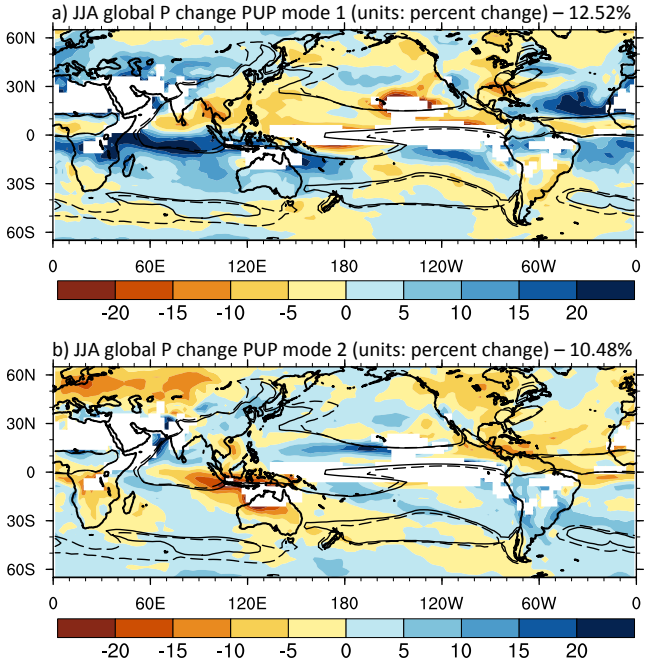

Figure S3.5: (a,b) Same as in Fig. S3.4 but for the JJA season.

redistributes the jet extension and shift modes across the first three rotated modes, though this does not change our original interpretation. EOF rotation therefore confirms that the P change PUPs in Fig. 3.6 are meaningful and stable, though it does not help gain more information about regional model disagreement.

S3.4 Testing the sensitivity of PUPs to removal of model outliers

We have calculated the robustness of global P change PUPs to removal of the CSIRO model (shown in Figs. S3.6 and S3.7). We compare the resulting partial-ensemble PUPs with the original PUPs (Figs. S3.2 and S3.3) and quantify the similarity by means of a correlation of leading modes (though this is a blunt tool, and a full analysis would require more advanced methods). Our results show that removing the CSIRO model does not affect the PUPs substantially.

For DJF, the first and second global PUPs are robust to removal of CSIRO, with correlations of $r = 0.90$ and $r = 0.9997$. For JJA, these correlations are $r = 0.87$ and $r = 0.88$. Furthermore, despite any small changes in the patterns, one can still draw the major conclusions from the partial-ensemble PUPs: that global uncertainty patterns are tropically dominated; that the largest type of uncertainty involves P increases in the deep tropics; that these uncertainties occur along the edges of tropical convection zones and within the ITCZ. Finally, when the same analysis is applied to the N. Pacific storm track region during DJF, correlation values are $r = 0.9995$ and $r = 0.9997$ for the first and second modes, respectively.

S3.5 Relationships between P and the U200 jet

As a final point, we investigate relationships between EOFs of U200 in the Pacific storm track region and global modeled P fields in the GCMs. These maps give insight into where model uncertainties in storm track U200 are associated with uncertainties in tropical and storm track P, and to what extent these resemble modes of internal P and U200 variability.

Figures S3.8a,b show end-of-century P changes correlated with the expansion coefficients of the first and third U200 change PUPs, respectively. Note that we have chosen to display the first and *third* PUP for U200 changes, as P change correlations with the second U200 change PUP appear

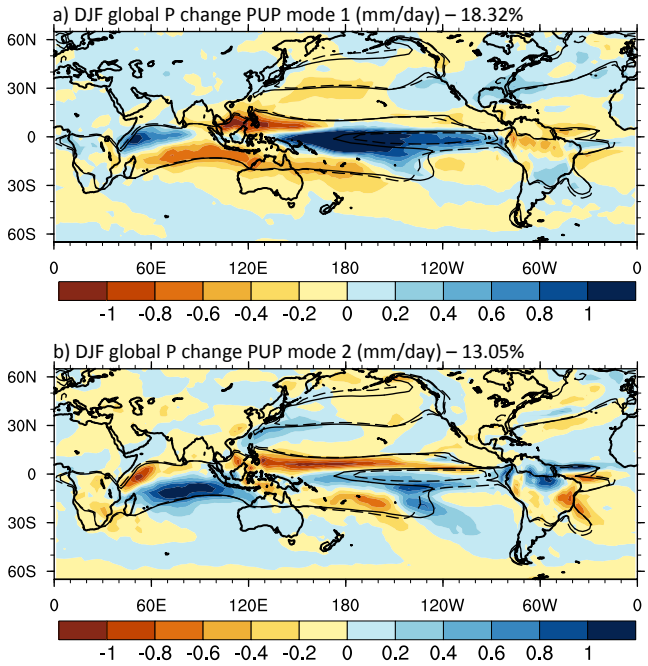

Figure S3.6: First and second DJF P change PUPs from a global analysis on a partial ensemble omitting CSIRO-Mk3-6-0.

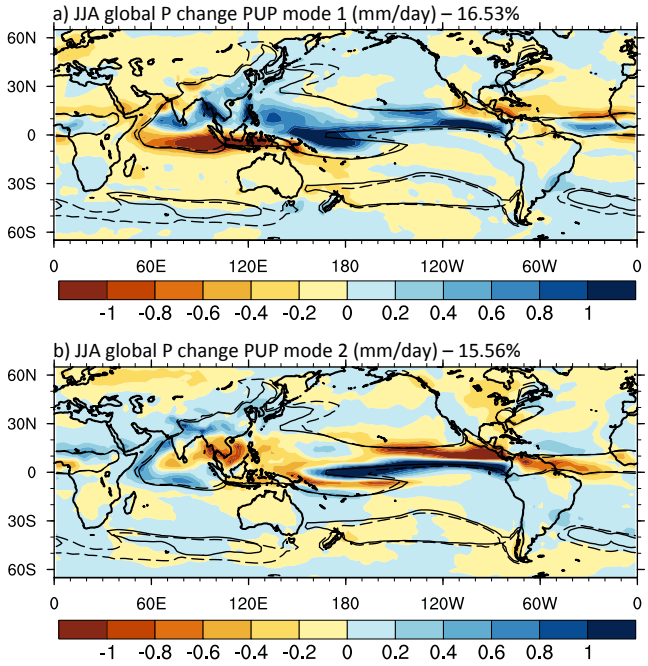

Figure S3.7: Same as in Fig. S3.4, but for JJA.

to represent another variant of the extension mode. For the first U200 mode, P change uncertainty in the midlatitude Pacific depicts a jet extension mode and is associated with uncertainty in the southern tip of the SPCZ (in the central Pacific, between the equator and 20°S). There is also a region of strong correlation of opposite sign to the north, which may reflect localized uncertainties in the descending branch of the Hadley circulation associated with jet uncertainties. Correlations with the third U200 change PUP exhibit a drying along the equatorward edge of the storm track, which may represent a poleward P shift pattern in the western-to-central midlatitude Pacific, with an associated tropical P change signal.

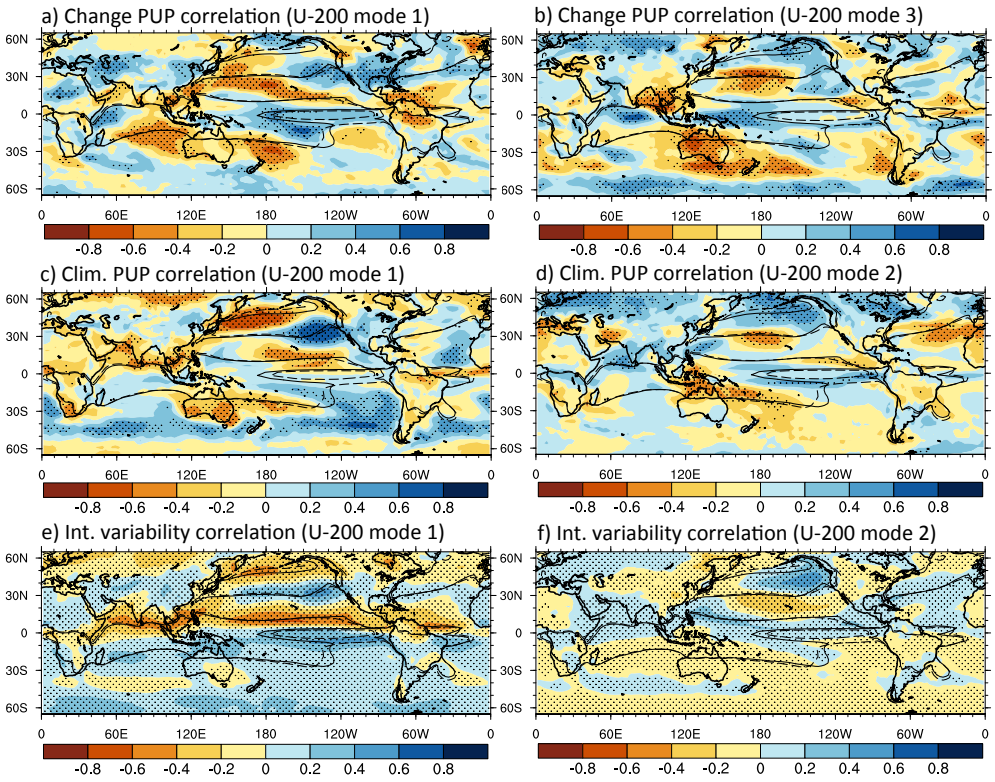

Figure S3.8: Correlations of P with U200, normalized to mm day^{-1} for a standard U200 deviation. (a,b) DJF P change values correlated with DJF U200 change PUPs one (as shown in Fig. 3.4e) and three (which accounts for 19.38% of the variance). (c,d) DJF model P climatologies correlated with expansion coefficients of U200 climatology PUPs one and two (as shown in Figs. 3.8c,d). (e,f) Monthly DJF P values correlated with internal U200 variability in the storm track region for EOFs one and two. Stippling in all plots shows where the correlation passes a t test at the 95% confidence level. Modes of internal U200 variability were calculated using linearly detrended DJF values across all 36 models at once (here in the time dimension), with each model's climatology removed prior to analysis. The two leading patterns of internal jet variability in the storm track region represent (1) a jet extension (or pulsing) mode, which accounts for 40.95% of the total variability, and (2) a meridional jet shift (or wobbling) mode, accounting for an additional 15.35% (these patterns are not shown here but can be seen in other studies (e.g., see Delcambre et al., 2013a, their Figs. 2a,b)).

Figures S3.8c,d show P climatologies correlated with the expansion coefficients of U200 climatology PUPs (as seen in Figs. 3.8c,d). In the first mode (Fig. S3.8c), the storm track region displays a similar jet extension pattern, though the zero contour is displaced and associated tropical P signals are much more localized along the southeastward sloping SPCZ axis. Again, a region of significant correlation of opposite sign appears north of the equator and may reflect localized interaction with the Hadley cell. The second mode (Fig. S3.8d) exhibits a meridional P shift in the midlatitude Pacific storm track region, though the tropics look quite different from that of Fig. S3.8b, with a strong signal south of the equator along the 4 mm day^{-1} contours, as well as a signal of strong opposite sign over the Pacific warm pool.

Figures S3.8e,f show monthly P values correlated with the first two modes of internal U200 variability in the midlatitude Pacific storm track region. Details of internal variability calculations are found in the caption, and the first two modes in U200 variability (not shown) are extension and shift modes, in that order. P correlations with the first mode (Fig. S3.8e) reveal a jet extension pattern in the midlatitude Pacific, and this is associated with tropical signals of opposite sign on either side of the equator. The second mode correlation map (Fig. S3.8f) shows a possible shift in storm track P, though the WP dipole is absent, and a more central-to-eastern Pacific shift appears in its place, with associated tropical variability evident as another cross-equatorial dipole, this time concentrated over the western tropical Pacific and maritime continent.

We conclude that while intermodel uncertainty patterns for end-of-century change in the Pacific storm track region may qualitatively resemble those for the historical climatologies or internal variability, the underlying processes appear to be distinct. Our results show that change PUPs are not primarily driven by differences in internal model variability on decadal+ timescales, though there do appear to be similarities between uncertainties in the P/U200 plots that are reminiscent of internal variability.

Chapter 4

Identifying leading sensitivity patterns in a perturbed physics ensemble, and using these to explore the parameter dependence of climatological precipitation

Abstract

Proper GCM simulation of precipitation (P) and the hydrological cycle requires that models correctly capture the interaction of large-scale dynamics with local, sub-grid scale processes, which themselves rely heavily on parameterizations. Perturbed physics ensembles address this aspect of sensitivity by isolating uncertainty from model physics, distinct from structural or numerical uncertainty. The dependence of a climate field on parameter values can be significantly nonlinear, and this is an important factor affecting model uncertainty, though its explicit form has not yet been thoroughly addressed in ensemble studies. In this chapter, an existing perturbed physics ensemble is used to investigate the parameter dependence of historical precipitation climatologies. Fitting models to a field's parameter dependence is a computationally cheap alternative that can help interpolate the GCM response to parameter value combinations not sampled in the integrations. This approach, termed metamodeling, is used here to reconstruct model fields at specified parameter combinations. A quadratic metamodel is employed first, which is created using second-order polynomial fits of a field at each grid point. This model reproduces global, seasonal root-mean-square error (rmse) relative to observations with decent skill, though it fails over certain highly sensitive parameter ranges. A trust region works well at the cost of excluding these ranges. An alternative metamodel is constructed which combines approaches used in the engineering literature with the principal uncertainty pattern (PUP) methodology from Chapter 3. For a given parameter, the first two or three PUP modes perform well at reconstructing the rmse measure, even in

highly sensitive ranges. This alternative metamodel is then used to explore questions of parameter optimization while considering multiple parameters and multiple objective functions associated with different climate fields. Concepts from multiobjective optimization are used to quantify the tradeoffs encountered when optimal parameter values for multiple fields are contradictory.

4.1 Introduction

As discussed in the previous chapters, global climate models (GCMs) exhibit notable uncertainty in simulating the current and future hydrological cycle. A large source of uncertainty not yet discussed is model sensitivity to the parameterization of convective processes leading to precipitation (P). Within a single GCM, grid point uncertainties due to choice of parameter value can be equal to (or greater in magnitude than) that arising in a multi-model ensemble from internal variability or climate change.

Perturbed physics ensembles are good tools for understanding this sensitivity, where the parameter values of a given GCM are sampled systematically within a plausible range. Recent work has shown that the parameter dependence of a given climate field can be highly nonlinear (Neelin et al., 2010; Bracco et al., 2013), and identifying the degree of this nonlinearity in parameter space is important in model tuning and parameter optimization.

The number of parameters to be sampled N in a perturbed physics ensemble can easily be on the order of tens, and sampling each at k different values requires k^N simulations (i.e., an order k^N problem). Given the inherent impracticality of such a brute-force sampling, optimization methods can reduce the order of the problem and alleviate computational burden. One relatively simple method is fitting the parameter dependence with a low-order polynomial regression and is termed "metamodeling" (Wang and Shan, 2006; Shan and Wang, 2010). A good starting point is a second-order polynomial regression, which we refer to as a quadratic metamodel in this chapter, and several studies have shown this form to be quite adept at modeling parameter dependence (Neelin et al., 2010; Bellprat et al., 2012; Bracco et al., 2013). A basic assumption in a low-order approach is that the parameter dependence is smooth across a feasible parameter range. Additionally, nonlinear two-way interactions (or cross-terms) are taken into account for different parameters (i.e., changing two parameters at once will not give the same result as changing them independently and adding

the outcome), though in this case three-way and higher-order interactions are neglected.

Bernstein and Neelin (2015, submitted) have created a branch-run perturbed physics ensemble for the fully coupled Community Earth System Model version 1.0.2 (CESM1). They analyze the magnitude and spatial distribution of the parameter sensitivity of P, as well as the degree of nonlinearity in the parameter dependence of P climatology and future climate change fields. The parameter dependence shows nonlinear behavior in certain ranges, most notably for low values of the entrainment fraction parameter (dmpdz), which is labeled a "dangerous range." Identifying such dangerous ranges in GCM simulations can help inform choices made by modelers as well as constrain a GCM's uncertainty in the present-day P climatology and end-of-century changes.

This chapter examines several features related to parameter dependence and sensitivity of historical P climatology and other fields. First, the principal uncertainty pattern (PUP) methodology is applied to the perturbed physics ensemble, and patterns of model uncertainty (or sensitivity) are shown for individual parameters. Second, the accuracy of a quadratic metamodel is explored for the parameter dependence of P and surface temperature (TS), with special consideration given to the dmpdz parameter which exhibits highly sensitive low-end nonlinearity. Third, these sensitivity PUPs are used in an alternative metamodel approach labelled PUP-cut-HDMR. This PUP-based metamodel does not require a priori assumptions about the specific type (linear, second-order, etc.) of parameter dependence and is a successful alternative where its quadratic counterpart fails. Lastly, P and TS are considered simultaneously, and optimization tradeoffs are illustrated by borrowing concepts from multiobjective optimization.

4.2 Data

4.2.1 Perturbed physics ensemble

Bernstein and Neelin (2015, submitted) have integrated CESM1 during the historical and representative concentration pathway 8.5 (RCP8.5) scenarios, akin to the experiments in the Climate Model Intercomparison Project phase 5 (CMIP5) ensemble (Taylor et al., 2012). This integration is performed using default parameter values until years 1976 and 2071, at which point branch integrations are carried out for an additional 30 years by varying the values of four parameters in

Table 4.1: The four parameters modified in the perturbed physics ensemble. First column lists the parameter shorthand as used in this chapter, with the full parameter name in the second column. The third column shows the parameter values used in the CESM1 integrations, with the control run values (the parameter default values) shown with an asterisk. Note in the the text that the first two dmpdz values are discussed as a dangerous range, so quadratic metamodel fits in this chapter exclude the model runs for the bracketed values of dmpdz (third column). Units and a short description are listed in the fourth and fifth columns. More information on this ensemble can be found in Bernstein (2014) and Bernstein and Neelin (2015, submitted). For more information on CESM1 or the deep convection scheme, see the Community Atmosphere Model 4.0 (CAM 4.0) documentation at Neale et al. (2010).

Parameter	**Name**	**Units**	**Values**	**Description**
dmpdz	deep convective entrainment parameter	$\times 10^{-3}$ m^{-1}	[0, 0.08,] 0.16, 0.25, 0.5, **1***, 1.5, 2	Turbulent entrainment of environmental air into deep convective plume
τ	deep convective timescale	minutes	30, **60***, 120, 180, 240	Time scale for consumption rate deep of Convective Available Potential Energy (CAPE) by cumulus convection; necessary for closure of deep convection scheme
α	downdraft fraction	fraction (out of 1.0)	0, **0.1***, 0.25, 0.5, 0.75	Fraction or proportionality factor that determines the mass flux of an ensemble downdraft, taking into account precipitation and evaporation
k_e	evaporation efficiency	$\times 10^{-6}$ kg $[\text{m}^{-2}\text{s}^{-1}]^{-1/2}$ s^{-1}	0.1, 0.5, **1***, 5, 10	Evaporation efficiency of precipitation

***** indicates standard or control value

the deep convection scheme. The first 10 years are discarded to allow for model equilibration. This produces a perturbed physics ensemble of 20 members in both the historical and RCP8.5 scenarios (including a control run, where default values are used for all parameters). The name, units, and a short description for the parameters that have been modified are listed in Table 4.1. The data used here are from the historical integrations of CESM1, and the December-January-February (DJF), June-July-August (JJA), and annual climatologies are calculated for each parameter value. Future work will incorporate the RCP8.5 integrations.

4.2.2 Observational and reanalysis data sets

Monthly P data are used from the global precipitation climatology project (GPCP) version 2.2, a merged satellite and surface rain gauge data set (Adler et al., 2003). Monthly skin temperature data are used from the National Center for Environmental Prediction (NCEP) reanalysis version 2. Both data wets were provided by NOAA/OAR/ESRL PSD in Boulder, CO, USA and can be downloaded at http://www.esrl.noaa.gov/psd/.

4.3 Methods

4.3.1 PUP calculations

PUPs are calculated across each parameter axis (i.e., varying each parameter as in Table 4.1 while holding all other parameters at their control values). This is done via empirical orthogonal function (EOF) analysis on the DJF, JJA, and annual climatologies separately, and the reader is referred to Chapter 3 for details of these calculations. The EOFs are calculated on fields centered at each grid point by the control run, so that EOF patterns represent departures from the CESM1 control integration. Note it will be common practice to take parameter values and fields relative to the control throughout this chapter.

4.3.2 Quadratic metamodel

The parameter dependence of a given climate field ϕ is emulated using a second-order polynomial regression. This methodology was adapted from the engineering literature (Wang and Shan, 2006; Shan and Wang, 2010) and applied to perturbed physics ensembles by Neelin et al. (2010). A simple but important distinction here is that the quadratic dependence is applied to each model grid point and season. This allows emulation of the model fields directly, as opposed to metamodels of objective functions *for* those fields. Bellprat et al. (2012) and Bracco et al. (2013) also follow this approach. The model field ϕ depends on a set of parameter values $\boldsymbol{\mu}$ such that $\phi = \phi(\boldsymbol{\mu})$. The combination of parameters $\boldsymbol{\mu}$ can be written as $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N) = (\text{dmpdz}, \tau, \alpha, k_e)$, where $N = 4$ as seen in Table 4.1. The metamodel is calculated over these parameter values centered

by the control or default value,

$$\mu_i' = \mu_i - \mu_{\text{cont}} \ .$$

Using this notation, the quadratic metamodel approximating a given field ϕ can be written for each time step or climatology and at each grid point as:

$$\tilde{\phi} \ = \ \phi_{\text{cont}} \ + \ \sum_{i=1}^{N} \mathbf{a}_i \mu_i' \ + \ \sum_{i=1}^{N} \mathbf{b}_{ii} {\mu_i'}^2 \ + \ \sum_{1 \le i < j}^{N} \sum \mathbf{b}_{ij} \mu_i' \mu_j' \ , \tag{4.1}$$

where $\tilde{\phi}$ is a function of the centered parameter values only:

$$\tilde{\phi} = \tilde{\phi}(\mu_1', \mu_2', \ldots, \mu_N')$$

Equation 4.1 allows for any number of parameters to be varied at once, which can be written here as $\tilde{\phi} = \tilde{\phi}(\text{dmpdz}', \tau', \alpha', {k_e}')$. The first term on the right-hand side is the control integration, which is a function of all parameters held at the control value, $\phi_{\text{cont}} = \phi(\text{dmpdz}_{\text{cont}}, \tau_{\text{cont}}, \alpha_{\text{cont}}, k_{e\,\text{cont}})$. The second, third, and fourth terms are each the linear, purely quadratic, and interaction terms. $\mathbf{a}$ is a vector containing the linear coefficients from a second-order polynomial regression at each grid point across N parameter fields. $\mathbf{b}$ is a square matrix containing the purely quadratic (ii) and interaction (off-diagonal, or ij) terms, with $\mathbf{b}_{ij} = \mathbf{b}_{ji}$.

In the metamodel analyses in this chapter, parameter interactions are neglected, so the final term in equation 4.1 is not calculated and $\mathbf{b}$ can be thought of as a symmetric matrix with quadratic coefficients on the diagonal and zeros elsewhere ($\mathbf{b}_{ij} = 0$ when $i \neq j$). The regression expands about the control run, which is why ϕ_{cont} must be added back in equation 4.1.

4.3.3 Cut-HDMR

An alternative way of framing the metamodeling approach above is that of high-dimensional model representation (HDMR), which is borrowed from the engineering literature and is part of the larger field of global sensitivity analysis (Rabitz and Alis, 1999; Rabitz et al., 1999). At its heart, HDMR deals with high-dimensional input-output problems in which a given system (physical, chemical, etc.) has a large number of input variables that can be independently altered and will

produce a complex response in the output. The GCM used in this chapter is one such system, and understanding the behavior or functional form of its parameter sensitivity is the overall goal.

The theoretical formulation of HDMR expresses a given physical system or model output as a function of all input variables $f(\boldsymbol{\mu}') = f(\mu_1', \mu_2', \ldots, \mu_N')$. If the output is a model field (e.g., ϕ as above), then this simply becomes $f(\boldsymbol{\mu}') = \phi$. HDMR is expressed as a finite expansion in terms of these input parameters:

$$\begin{aligned} f(\boldsymbol{\mu}') \;=\; f_0 \;+\; \sum_{i=1}^{N} f_i(\mu_i') \;+\; \sum_{1\le i<j\le N} f_{ij}(\mu_i', \mu_j') \\ +\; \sum_{1\le i<j<k\le N} f_{ijk}(\mu_i', \mu_j', \mu_k') \;+\; \ldots \;+\; f_{12\ldots N}(\mu_1', \mu_2', \mu_3', \ldots, \mu_N') \;. \end{aligned} \tag{4.2}$$

The first term on the right-hand side, f_0, is a constant representing the mean response to $f(\boldsymbol{\mu})$. The second term represents any variable μ_i acting independently, so this term can include first-order but also any higher-order terms (quadratic, cubic, etc.). The third term represents all possible two-way interactions between two different input variables (and is therefore a quadratic or second-order term). The fourth term represents any three-way interactions among the variables. Subsequent terms represent all possible interaction combinations, and the final term represents the residual N^{th}-order interaction among *all* variables at once.

This HDMR formulation has several benefits. First, it is a finite expansion and is always exact at the grid points over which evaluations are done. Second, it requires fewer assumptions to be made about the underlying functionality of the parameter dependence. Third, and most importantly, while the number of terms in the expansion can be daunting for large N, most systems tend to have weak interactions beyond quadratic (Li et al., 2001), and it is often sufficient to approximate this as

$$f(\boldsymbol{\mu}') \;\approx\; f_0 \;+\; \sum_{i=1}^{N} f_i(\mu_i') \;+\; \sum_{1\le i<j\le N} f_{ij}(\mu_i', \mu_j') \;, \tag{4.3}$$

with the third term on the right-hand side representing two-way interactions.

The precise set of choices going into the HDMR expansion can vary (for a discussion of common approaches, see Rabitz et al., 1999). One option for this problem is cut-HDMR (Rabitz

and Alis, 1999; Li et al., 2001), in which the expansion is calculated about a reference point, and each successive function is defined along *cut* lines, planes, volumes, etc. through that reference point. Note the quadratic metamodel in equation 4.1 is equivalent to the cut-HDMR metamodel in equation 4.3 when (1) the control run is set as the reference point such that $f_0 = f(\boldsymbol{\mu}_{\text{cont}}) = \phi_{\text{cont}}$, and (2) the second term in equation 4.3 is restricted to include only linear and quadratic terms.

While this chapter only considers the linear and purely quadratic terms, we do plan to incorporate integrations representing off-diagonal interaction points in future work. We justify the simplification used here by noting that Neelin et al. (2010) tested the impact of including the interaction terms on a quadratic metamodel and found the sensitivity is modest and in most cases small. Furthermore, as will be shown later, metrics of model error relative to observations and reanalyses demonstrate that the near-optimal parameter combinations often tend to occur in the vicinity of the control parameter values. Therefore the analysis here can be thought of as the first step in a process deciding how worthwhile it is to integrate the GCM at off-diagonal points.

The utility of these metamodels is their ability to estimate the response of a field to any parameter combination without explicitly sampling the entire parameter space with GCM integrations. Metamodels for the perturbed physics ensemble here require $\mathcal{O}(N)$ calculations to obtain coefficients for $\mathbf{a}$ and the diagonals along $\mathbf{b}$ in equation 4.1. Resolving the coefficients for the interaction terms (off-diagonals in $\mathbf{b}$) requires $\mathcal{O}(N^2)$ calculations, which becomes impractical for high-density evaluation. Furthermore, restricting calculations to the parameter axes is more directly relevant to a physical interpretation of parameter sensitivity.

4.3.4 PUP-cut-HDMR metamodel

One lingering issue associated with a quadratic metamodel is that making assumptions a priori about the functional form of the sensitivity limits the model's flexibility, and it may not be adequate to describe parameter sensitivity over a highly nonlinear range. Motivated by making the parameter sensitivity more intuitive and simplifying the reconstruction, we construct a new metamodel which blends PUPs from Chapter 3 and cut-HDMR into a modified "PUP-cut-HDMR" metamodel. In the results and discussion section, we give examples of visualizing the spatial sensitivity of fields and displaying the along-axis dependence of the climatological rmse relative to

observations and reanalyses. The remainder of this section presents a formal contruction of the PUP-cut-HDMR metamodel.

As a reminder, the subscript k is used to represent sampling density for one parameter μ_i. Suppose an EOF analysis is done across the parameter axis for a climate field $\vec{\phi}_i$ with shape ($k \times$ latitude $\times$ longitude). The vector notation $\vec{(\cdot)}$ is used to distinguish this field from that in equation 4.1 and represents the sampling dimension k along a parameter axis. Each of the four parameters in Table 4.1 can be expressed as

$$\vec{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ik})\ ^{1}\ .$$

EOF analysis of $\vec{\phi}_i$ results in m EOF modes along with m expansion coefficient (or principal component) vectors $\vec{c}_i$, each of which has length k and corresponds to an EOF mode:

$$\vec{c}_{im} = (c_{im1}, c_{im2}, \ldots, c_{imk})\ .$$

Recall that this EOF analysis is done on the field centered by the control run at each grid point, $\phi'_{ik} = \phi_{ik} - \phi_{\text{cont}}$ for all k in $\vec{\phi}_i$.

The goal here is to use a pruned set of leading EOF modes to reconstruct the original field for each parameter value μ_{ik}. If we assume that the important part of the sensitivity is contained within the first M EOFs (with $M \leq k$), then the k^{th} field for each parameter value μ_i can be approximated by reconstructing it using the leading M modes:

$$\phi_{ik\,\text{approx}} \;=\; \phi_{\text{cont}} \;+\; \sum_{m=1}^{M} c_{imk}\,\text{EOF}_{im}\ , \tag{4.4}$$

where the approximated field is a function of μ_{ik}: $\phi_{ik\,\text{approx}} = \phi_{\text{approx}}(\mu_{ik})$. In this analysis, each EOF mode EOF_{im} is normalized to carry units of the original field, and the vector $\vec{c}_{im}$ is normalized to have unit variance.

The PUP-cut-HDMR metamodel is first consructed along each parameter axis using equation 4.4 repeated for all k parameter values, and a new field $\vec{\phi}_{i\,\text{approx}}$ is built with dimension ($k \times$

[1]To be explicit about this notation, taking the parameter α as an example, $\vec{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ik})$ corresponds to $\vec{\alpha} = (0, 0.1, 0.25, 0.5, 0.75)$, as read from Table 4.1, and $k = 5$.

latitude × longitude). This represents an approximation to the perturbed physics ensemble across a single parameter μ_i using the first M EOFs. This analysis is repeated for all four parameters $\mu_i = \text{dmpdz}, \tau, \alpha, k_e$. Reconstructions using the first two and three EOFs will be shown in the results of this chapter.

Finally, from the approximated ensemble above, we can extend the metamodel *off* parameter axes in order to explore the full parameter space. This is done by borrowing concepts from cut-HDMR. We make the assumption that the contribution to parameter sensitivity for each variable acts independently and is additive, as discussed for equations 4.2 and 4.3. This is perhaps easiest to think about in terms of varying two parameters simultaneously. Suppose dmpdz and α are both varied off-axis, and the other two variables τ and k_e are held at their control values. The resulting field $\phi(\text{dmpdz}', \alpha', \tau_{\text{cont}}, k_{e\,\text{cont}})$ can be approximated by summing the control run and separate approximations from dmpdz and α:

$$\begin{aligned}\phi(\text{dmpdz}', \alpha', \tau_{\text{cont}}, k_{e\,\text{cont}}) \;\approx\; \phi_{\text{cont}} \;+\; & \phi(\text{dmpdz}', \alpha_{\text{cont}}, \tau_{\text{cont}}, k_{e\,\text{cont}}) \\ & +\; \phi(\text{dmpdz}_{\text{cont}}, \alpha', \tau_{\text{cont}}, k_{e\,\text{cont}})\ .\end{aligned}$$

To minimize extrapolation, the field is only constructed at combinations of the points already sampled in the perturbed physics ensemble (see Table 4.1). The resulting field will have dimension (latitude × longitude) and is labeled the PUP-cut-HDMR metamodel:

$$\begin{aligned}\tilde{\phi}(\mu'_{1k}, \mu'_{2k}, \ldots, \mu'_{Nk}) \;&=\; \phi_{\text{cont}} \;+\; \sum_{i=1}^{N} \left(\phi_{ik\,\text{approx}} - \phi_{\text{cont}} \right) \\ &=\; \phi_{\text{cont}} \;+\; \sum_{i=1}^{N} \left(\sum_{m=1}^{M} c_{imk}\, \text{EOF}_{im} \right)\ ,\end{aligned} \tag{4.5}$$

where equation 4.4 was substituted into the second line above. Note that since EOF calculations are done relative to the control field, ϕ_{cont} must be subtracted for each approximation and then added at the end. This control-centered approach has the benefit of being easily modified if the modeler chooses to add more integrations at other parameter values. As before, $N = 4$ represents the parameters in Table 4.1, and the value for k does not have to be uniform for the N parameters.

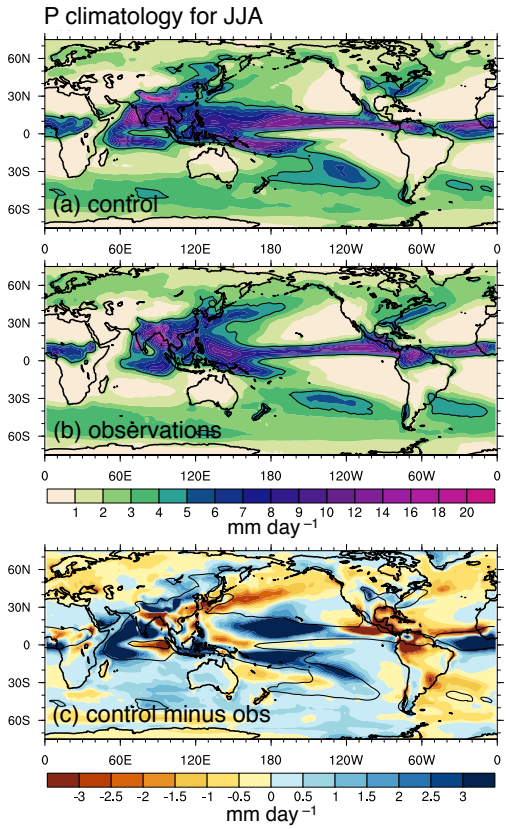

Figure 4.1: 1986–2005 P climatology during JJA for (a) the CESM1 control run, (b) observations, and (c) control-minus-observations. Black contours show the 4 mm day^{-1} contour from the CESM1 control run in (a) and (c), and for the observations in (b). Observed P data are from the GPCP version 2.2 data set.

4.4 Results and discussion

4.4.1 Sensitivity patterns for P

To simply the presentation of maps, JJA is used as the example season, though DJF and annual sensitivity measures will be discussed later. Figures 4.1a,b display the JJA P climatology for the control run and observations. Figure 4.1c shows the control-minus-observed climatology, giving a sense of where CESM1 is over- and underestimating P. Note the excessive P on either side of the equator in the central Pacific, as well as underestimated P in Latin and Central America, more broadly across the Pacific cold tongue, and at midlatitudes. The Asian monsoon region and the Indian Ocean are also notable in the magnitude and spatial complexity of P bias. For a related discussion of parameter sensitivities in the historical climatologies, see Bernstein (2014). Figure

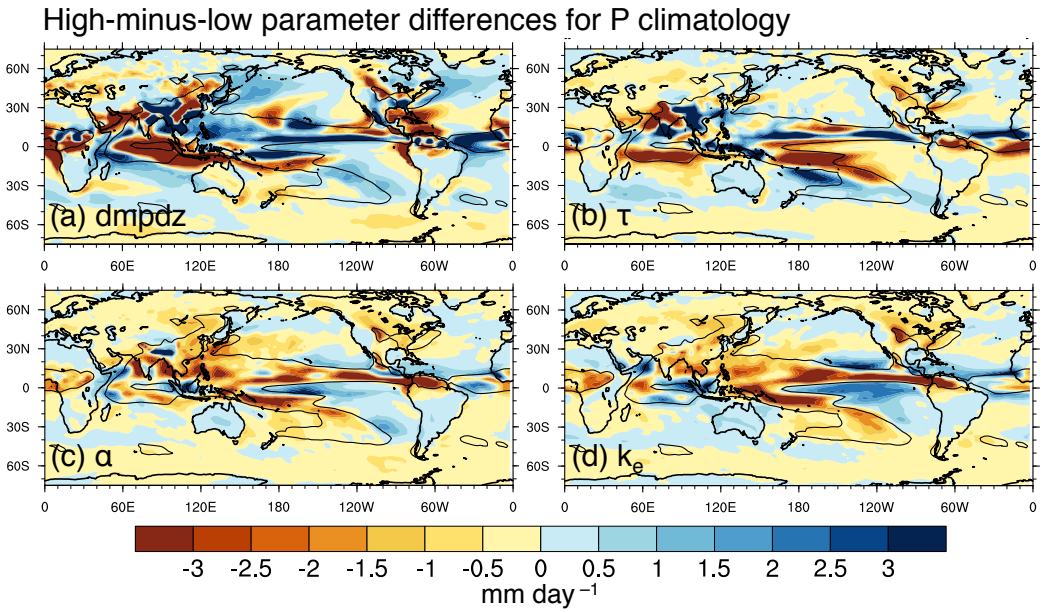

Figure 4.2: Differences in the 1986–2005 JJA P climatology for the highest-minus-lowest parameter values (as noted in Table 4.1). Note that the low value used for dmpdz is 0 (so that this analysis includes bracketed values in Table 4.1). This figure has been adapted from Bernstein (2014), her Fig. 3.2.1.2.

4.1c serves as a common example of GCM bias in the CMIP5 archive and highlights why parameter optimization—with the hope of constraining this bias—is a worthwhile activity.

Figure 4.2 displays climatological P sensitivity across individual parameters in the perturbed physics ensemble. These figures show the result of taking the P climatology for the lowest parameter value and subtracting it from that for the highest parameter value (i.e., high-minus-low parameter differences; see Table 4.1 for details). The areas that show the highest sensitivity are within zones of deep convection (as visualized by the 4 mm day^{-1} control run climatology), as well as on these margins and in the cold tongue region. The magnitude of the sensitivity is large (well over 3 mm day^{-1} in certain regions), especially over the tropics, where deep convection plays a central role in setting tropospheric temperature and moisture profiles.

P climatology PUPs across parameter axes

This perturbed physics ensemble lends itself well to a PUP analysis, and the first and second EOF modes for all four parameters are displayed in Figs. 4.3 and 4.4. There is striking similarity between the first mode for each parameter in Fig. 4.3 to its counterpart in Fig. 4.2.

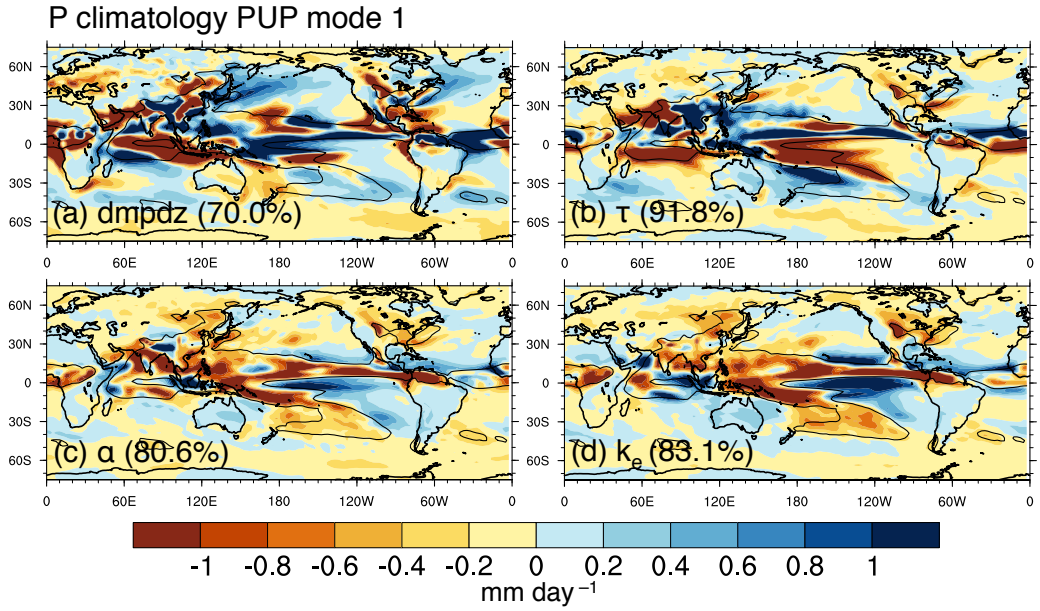

Figure 4.3: First mode from a PUP analysis applied to the 1986–2005 JJA P climatology across each parameter axis separately. Percent variances accounted for by each mode are included on the maps.

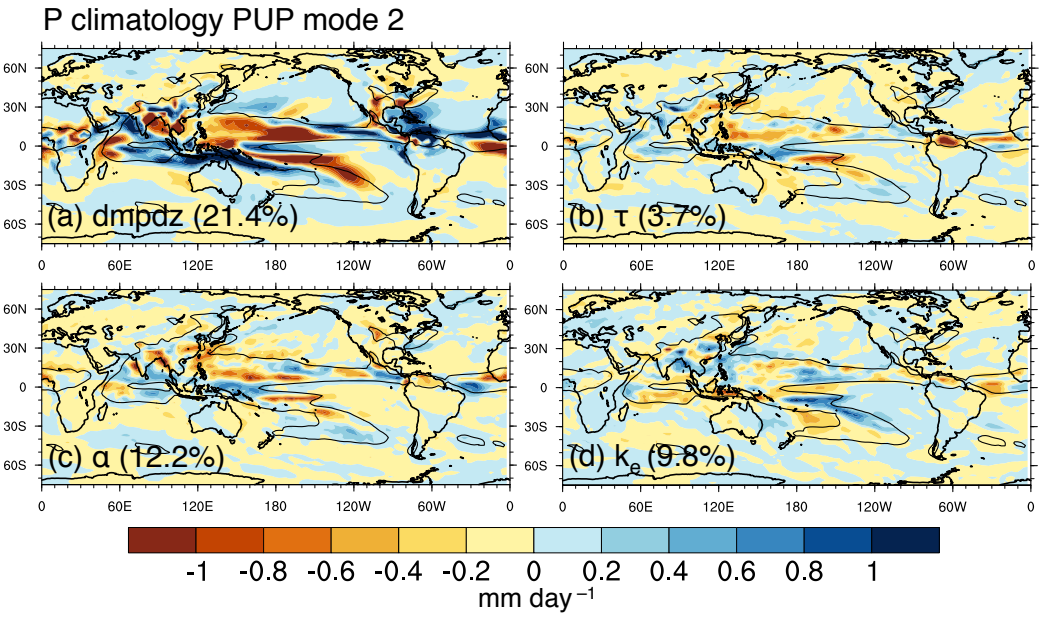

Figure 4.4: Second mode from a PUP analysis applied to the 1986–2005 JJA P climatology across each parameter axis separately. Percent variances accounted for by each mode are included on the maps.

For all parameters except dmpdz, the first PUP captures over 80% of the variance in the model sensitivity. This value is lower for dmpdz at 59.9%. The second EOF modes are shown in Fig. 4.4, and one can see that for dmpdz in particular, substantial P sensitivity is carried into the second mode, explaining an additional 29.7% of its variance. This particular behavior for dmpdz is associated with a strong nonlinearity that will be discussed below. The other three parameters, in contrast, have comparatively weak signals in the second mode, capturing less than 10%. The patterns for the third modes (not shown) are even smaller in magnitude and each capture around 5% or less of the variance.

Quadratic metamodel results

Transitioning from spatial maps of P sensitivity, we now focus on the global rmse of a given P climatology relative to observations and reanalyses. Bernstein and Neelin (2015, submitted) display the global rmse of P climatologies as a function of parameter value, and this information is repeated in Fig. 4.5 (solid color lines). Note that the magnitude of error in the P climatology can be decreased 10–20% simply by varying parameter values, and that in certain cases optimal parameter values exist at the endpoints of the sampled parameter range.

Note that the rmse in Fig. 4.5 is an example of an objective function that describes some aspect of model performance or error. In this chapter, we will use rmse and mean-square-error (mse) relative to reanalyses or observations to objectively assess model performance in simulating a given field. Multiobjective optimization entails the optimization of multiple objective functions at once, and this will be discussed later.

In Fig. 4.5, the rmse exhibits a fairly linear dependence for some parameters (e.g., for τ in Fig. 4.5b), though dmpdz and α show notable sensitivity and nonlinearity. The colored dashed lines show the rmse reconstructed using the quadratic metamodel. The second-order metamodel performs generally well and is able to reproduce the nonlinear nature of the rmse along the parameter axes, especially when the nonlinearity is low-order like that of α. Certain parameter ranges, however, do not lend themselves well to the quadratic metamodel. This can be seen in Fig. 4.5b, where the metamodel is unable to capture the precise characteristics of the rmse around $\tau = 175$, and also in Fig. 4.5a, which displays high sensitivity and nonlinearity in the low-end dmpdz (low

entrainment) range.

Note here in Fig. 4.5a, the quadratic metamodel has been trained over the entire dmpdz range. Looking more closely, the highest sensitivity where the metamodel has difficulty is in the first two parameter values (0 and 0.08). The metamodel could also be trained over a shorter range or "trust region" where the nonlinearity is more aptly described by a second-order metamodel. Figure 4.6a compares the performance of a quadratic metamodel for dmpdz when trained over the full parameter range (dashed line) or over the trust region excluding the first two values (dotted line). Relatively speaking, when the metamodel fit includes the highly sensitive values at low

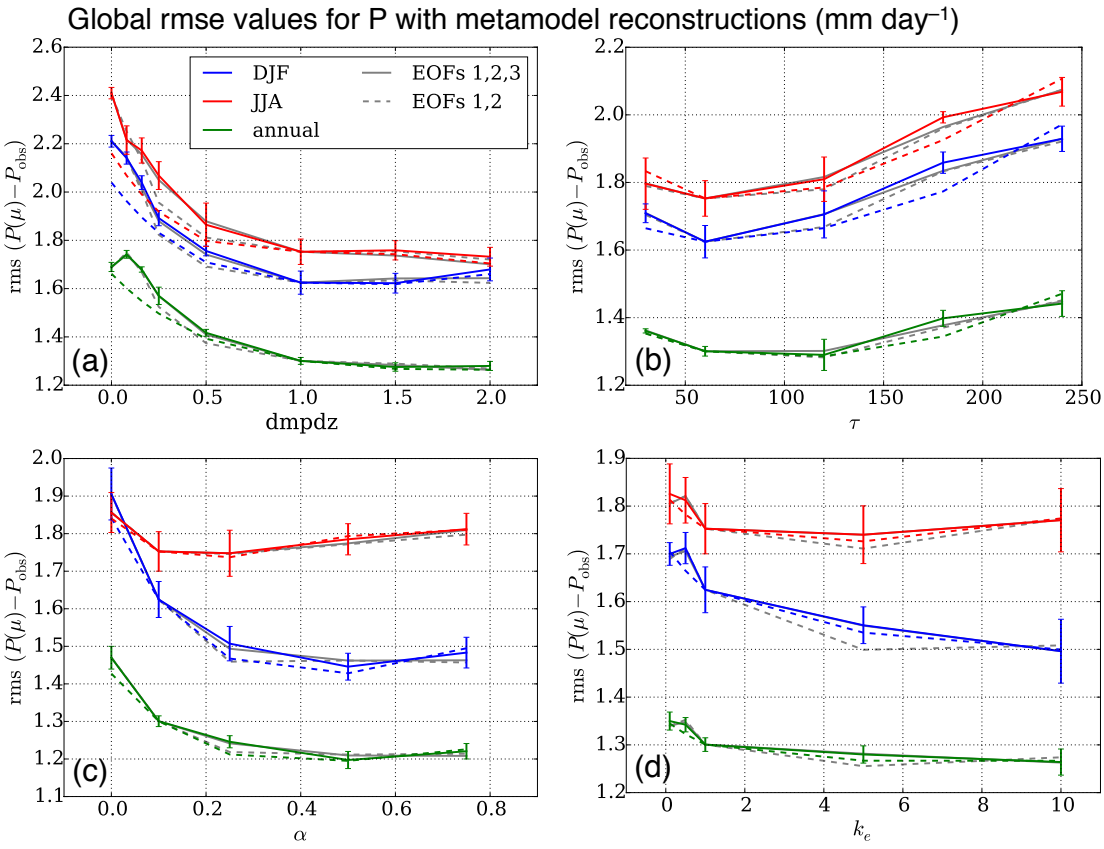

Figure 4.5: Solid colored lines show global rmse values for P climatologies (model-minus-observations) in mm day^{-1}, as a function of parameter value, for (a) dmpdz, (b) τ, (c) α, and (d) k_e (see Table 4.1 for more information). Dashed colored lines show the reconstruction of global rmse values by the quadratic metamodel, fit over all parameter values in each case. The solid and dashed gray lines show rmse reconstructions using the first two and first three EOF PUPs, respectively, and these are also done across all parameter values. Rmse values are calculated relative to P observations in the GPCP version 2.2 data set. Error bars are estimated by splitting the historical simulation into non-overlapping five-year chunks and taking the standard deviation of their rmse values at each parameter value.

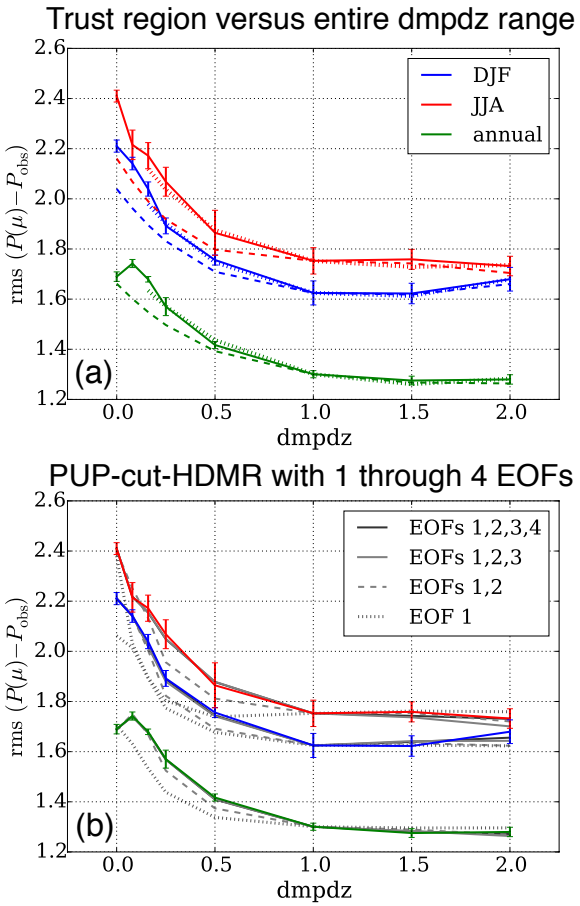

Figure 4.6: (a) and (b) as in Fig. 4.5a, with solid lines showing the full rmse across the dmpdz parameter range. (a) shows a comparison between a quadratic metamodel fit over the entire dmpdz range (colored dashed line) and a fit over the trust region, which excludes the first two dmpdz values (colored thick dotted line). (b) shows a PUP-cut-HDMR fit over the entire dmpdz parameter range, showing the rmse reconstruction using the leading mode (gray dotted line), the first two leading modes (gray dashed line), the first three leading modes (light gray solid line), and the first four leading modes (dark gray solid line).

entrainment, it is not able to capture the sensitivity in this region, and in such a case modifying the fit to focus on the trust region yields a more successful metamodel.

PUP-cut-HDMR metamodel results

The gray lines in Fig. 4.5 show rmse reconstructions using the PUP-cut-HDMR metamodel on each parameter axis. Gray dashed lines show the first two EOFs ($M = 2$) and gray solid lines show the first three ($M = 3$). These reconstructions are quite adept at capturing the degree of nonlinearity in the rmse curves, and in every case the first two modes perform better than the

quadratic metamodel, while the first three modes are nearly indistinguishable from the full model results.

Given the highly sensitive low-end range of dmpdz, it is useful to ask how it behaves for each EOF in the PUP-cut-HDMR analysis. Figure 4.6b shows reconstructions of the global rmse using anywhere from the first EOF to the first four EOFs. Using just the first mode (dotted line), the PUP-cut-HDMR method is very poor at reconstructing global rmse, giving insight into the failure of the quadratic metamodel over the full range in Fig. 4.6a. With the addition of the second mode in Fig. 4.6b (dashed line), however, the performance of the PUP-cut-HDMR metamodel is able to capture the strong sensitivity at low dmpdz values.

To get a sense of how nonlinear the contribution from each EOF is across the parameter axes, Fig. 4.7 shows the seasonal breakdown for the first three modes. These projections are normalized to carry P units of mm day^{-1}, with a magnitude that translates to the relative contribution of the individual EOFs, and with the zero line representing the control run. The behavior of individual modes in Fig. 4.7 helps to understand the nonlinearity and performance of the metamodel in Fig. 4.6. Focusing on the first modes (Fig. 4.7 solid lines) for all seasons and parameters, one can broadly see that they display varying degrees of nonlinearity. Sensitivity is highest for the dmpdz parameter (top row of Fig. 4.7). The first modes show significant nonlinearity of what appears to be fairly low order, which explains the similarity between the first mode in the PUP-cut-HDMR metamodel (dotted gray lines in Fig. 4.6b) and the quadratic metamodel fit over the entire dmpdz range (dashed colored lines in Fig. 4.6a). The second dmpdz modes in Fig. 4.7 highlight the particularly sensitive and nonlinear form of the low-end range. The improved performance of PUP-cut-HDMR relative to the quadratic metamodel in Fig. 4.6 is apparent here: PUP-cut-HDMR is able to recover both the low- and high-end nonlinearities in the parameter dependence with as little as two EOFs.

Figure 4.7 also gives insight into the contribution of each parameter value to the EOF modes shown in Figs. 4.3 and 4.4. For example, the leading EOF for τ (second row of Fig. 4.7) shows a fairly linear dependence across the range, implying variations in τ contribute to Fig. 4.3b in a linear fashion. By contrast, the leading EOF for dmpdz, associated with Fig. 4.3a, begins with a strong and high-gradient influence at low dmpdz values, increases monotonically through its range

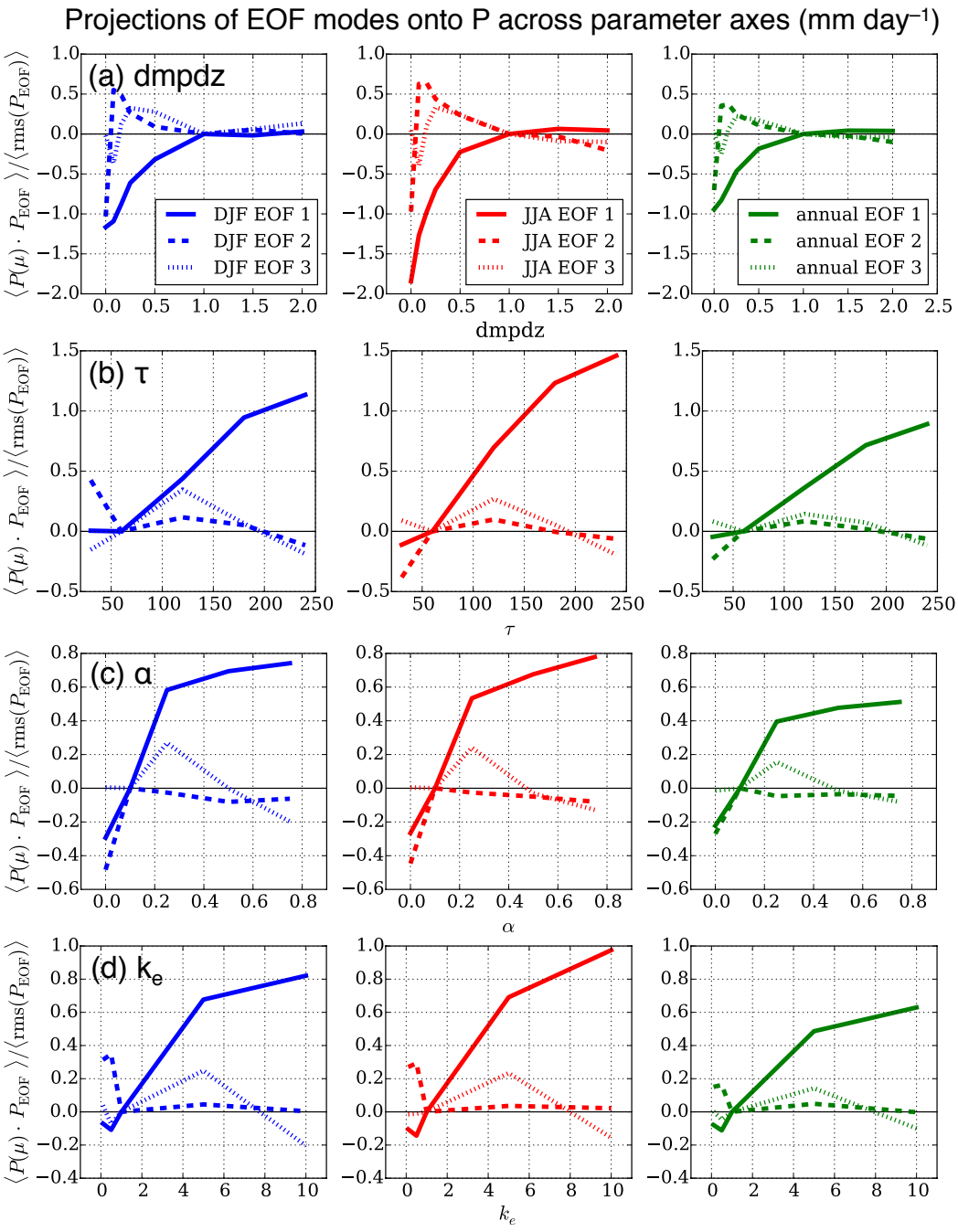

Figure 4.7: Projections of the first three EOFs for each variable onto the difference between global P at a given parameter value and the control run for that parameter. Specifically, this projection calculation is $\langle \mathrm{P}(\mu_i') \cdot \mathrm{P}_{\mathrm{EOF}} \rangle \,/\, \langle \mathrm{P}_{\mathrm{EOF}} \rangle$, with the term in angle brackets $\langle \cdot \rangle$ representing a spatial average. Units for vertical axes are in mm day^{-1}. Units on horizontal axes are as in Table 4.1.

with a low-order structure, and tapers off at the high end. To capture the additional and highly nonlinear structure in the low-end dmpdz range seen in Fig. 4.6, the second EOF is necessary, which is notably nonlinear at low dmpdz in Fig. 4.7. A modeler interested in which regions or climate features are most impacted by parameter variations could use Figs. 4.3 and 4.4 to build intuition about the spatial structure of the sensitivity and then use Fig. 4.7 to evaluate the form of the nonlinearity and the parameter value ranges that are most influential.

4.4.2 Sensitivity patterns for TS

In the process of parameter tuning and optimization, the desired outcome is a set of optimal parameter values that minimize some objective function for the fields and variables of interest. Here, global rmse has been chosen as the objective function, and P has been the focus so far. Monthly TS climatologies are now examined, and a parallel set of figures is produced for this second field.

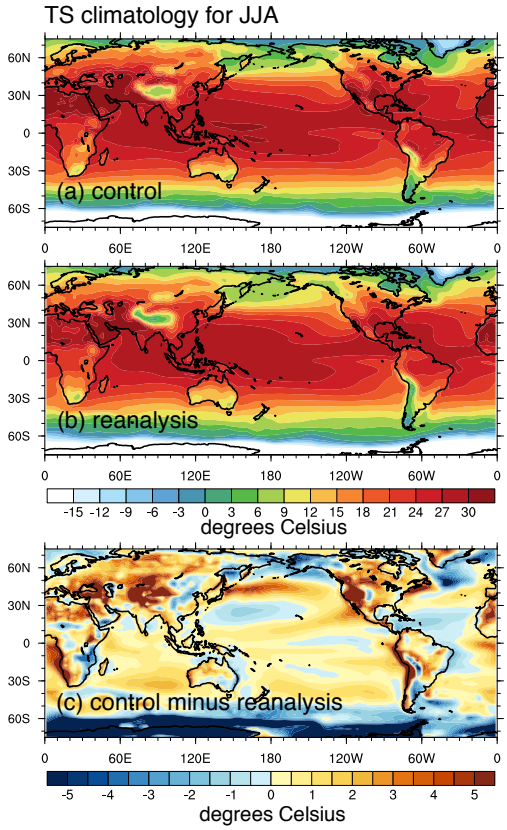

Figure 4.8: 1986–2005 TS climatology during JJA for (a) CESM1 control run, (b) reanalysis, and (c) control-minus-reanalysis. Skin temperature reanalysis data are from the NCEP-DOE Reanalysis version 2.

Figure 4.8 shows control, observed, and control-minus-observed TS climatology for JJA. The control shows a general warm bias of around one degree Celsius or less over the oceans and slightly higher warm biases over land. Notably large positive TS biases exist along the west coasts of North and South America as well as southern Africa and interior parts of Asia including the Tibetan Plateau region. These positive land biases are often collocated with adjacent negative biases and appear to coincide with large mountain ranges (Himalayas, Rockies, and Andes). Areas with significant cold biases include the north Pacific and high southern latitudes.

The high-minus-low parameter difference maps for TS during JJA are shown in Fig. 4.9. Besides the tropical Pacific increase in TS for dmpdz in Fig. 4.9a, increasing any of these parameter values tends to decrease TS in the JJA season, and land areas with complex topography again show the largest magnitudes of TS uncertainty. Comparing these difference maps to Fig. 4.2 for P, one can see that TS uncertainty is not collocated with areas of largest P uncertainty, which tend to occur in the tropics within regions of deep convection. Instead, the largest magnitude TS uncertainties occur mostly on land and without preference for latitude. Note that the parameters this ensemble has sampled are part of the deep convection scheme, which is not directly related to TS, highlighting the ease with which sub-grid scale parameterization uncertainty can propagate to other components of a coupled GCM.

Figures 4.10 and 4.11 show the first and second EOFs from the PUP analysis applied to TS JJA climatologies. As noted for P climatology PUPs, the spatial patterns for the first TS mode are very similar to the high-minus-low differences in Fig. 4.9. The first modes for τ, α, and k_e capture about 80–90% of the variance, and dmpdz falls in slightly lower at 72%. The second mode captures around 10% or less for all variables except dmpdz, which is higher at 18%. This behavior is similar to that for P climatology PUPs: a significant amount of the uncertainty is captured in the first mode, though the highly nonlinear low-end range for dmpdz causes a notable amount of sensitivity to spill over into the second mode.

Metamodel reconstruction of global rmse for TS is shown in Fig. 4.12. As in the P rmse analysis, parameter dependence for TS can be highly nonlinear in certain ranges, such as low-end dmpdz values. The quadratic metamodel in Fig. 4.12 (dashed colored lines) is adequate in some parameter ranges but fails to reconstruct global rmse error in some of the same places where it

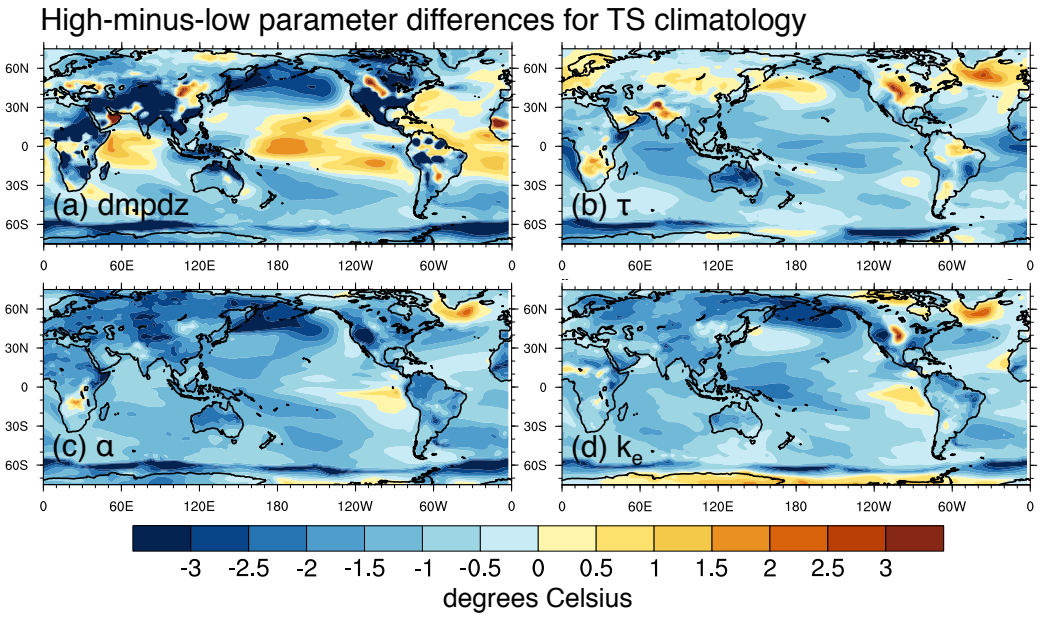

Figure 4.9: Differences in the 1986–2005 JJA TS climatology for the highest-minus-lowest parameter values (as noted in Table 4.1). Note that the low value used for dmpdz is 0 (so that this analysis includes bracketed values in Table 4.1).

fails for P. Also note that the minima of the rmse curves change between TS and P, leading to contradictions over which parameter values will optimize model performance for those fields. For example, if a modeler is interested in tuning dmpdz in order to minimize global rmse for P and TS simultaneously, the DJF season leads to a tradeoff where minimum rmse values for P are in the higher end of the range (dmpdz $\approx$ 1.0–1.5), but minimum rmse values for TS draw the modeler toward the lower end of the range (dmpdz $\approx$ 0.25–0.50). Therefore tradeoffs are encountered even when considering a single parameter.

The PUP-cut-HDMR metamodel reconstructed using two EOFs (gray dashed line) performs as well or better than the quadratic metamodel, confirming that the majority of quadratic parameter dependence is captured by the first PUP modes in Fig. 4.10. However, this metamodel still has issues in certain parameter ranges, including most of the intermediate τ and α values. The PUP-cut-HDMR metamodel that includes the first three EOFs (gray solid line) is notably better in these ranges, though it still shows some lack of skill for α around 0.25 and 0.5.

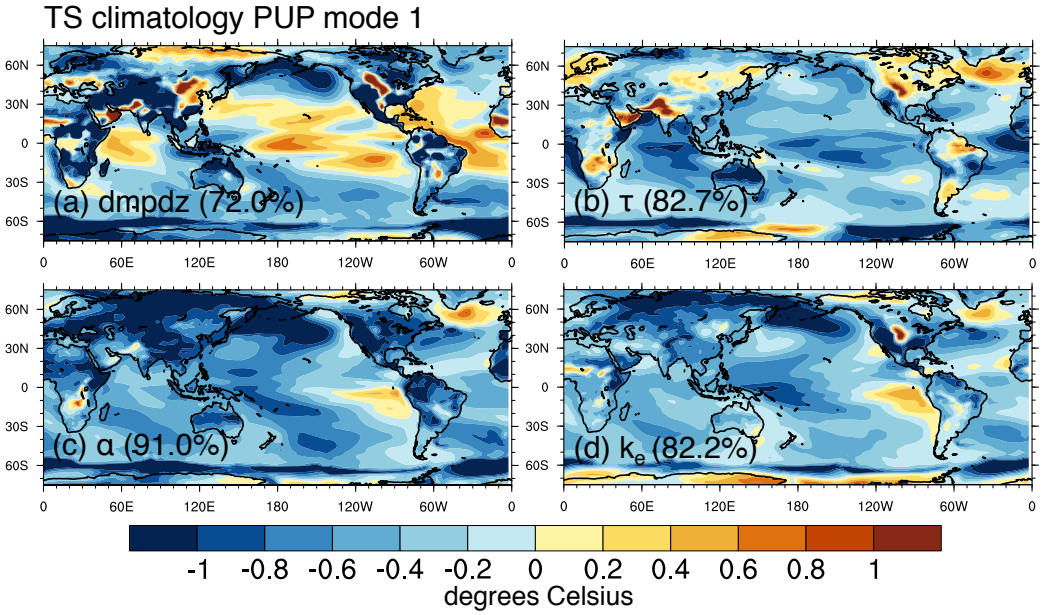

Figure 4.10: First mode from a PUP analysis applied to the 1986–2005 JJA TS climatology across each parameter axis separately. Percent variances accounted for by each mode are included on the maps.

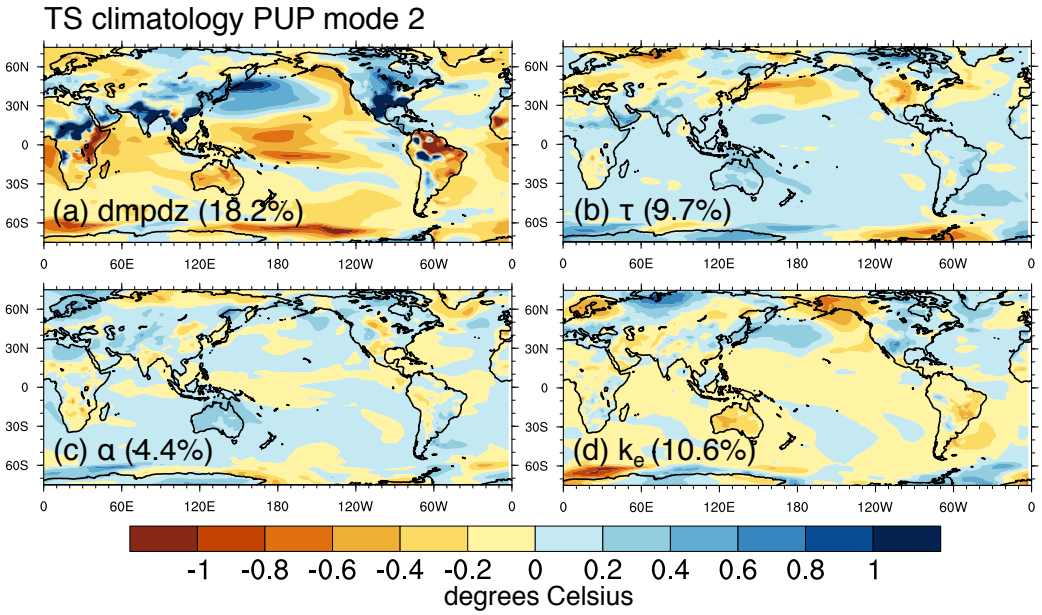

Figure 4.11: Second mode from a PUP analysis applied to the 1986–2005 JJA TS climatology across each parameter axis separately. Percent variances accounted for by each mode are included on the maps.

4.4.3 P and TS error surfaces as a function of two parameters

Comparing Figs. 4.5 and 4.12, it is obvious that two different climate fields will often have minima that occur at different parameter values. This makes the process of model tuning and optimization difficult, as changing a parameter value can improve model performance in one field but degrade it in another, and this is strongly dependent on region and season. Until now, we have looked at sensitivity taken one parameter axis at a time, but in a $N = 4$ parameter space sampled at density k, there are k^4 possible combinations for parameter values. The metamodels in this chapter can be used to approximate the model response to any of these parameter value

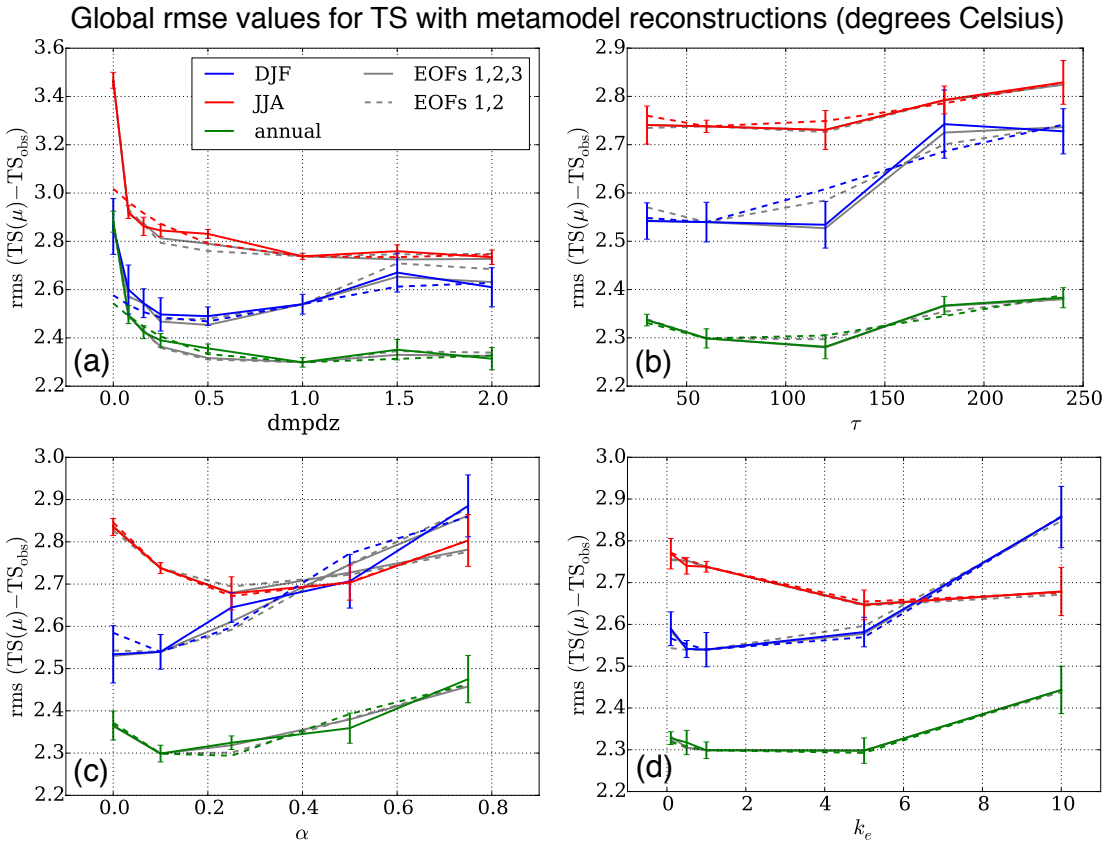

Figure 4.12: Solid colored lines show global rmse values for TS climatologies (model-minus-reanalysis) in degrees Celsius, as a function of parameter value, for (a) dmpdz, (b) τ, (c) α, and (d) k_e (see Table 4.1 for more information). Dashed colored lines show the reconstruction of global rmse values by the quadratic metamodel, fit over all parameter values in each case. The solid and dashed gray lines show rmse reconstructions using the first two and first three EOF PUPs, respectively, and these are also done across all parameter values. Rmse values are calculated relative to skin temperature reanalysis data from the NCEP-DOE Reanalysis version 2. Error bars are estimated by splitting the historical simulation into non-overlapping five-year chunks and taking the standard deviation of their rmse values at each parameter value.

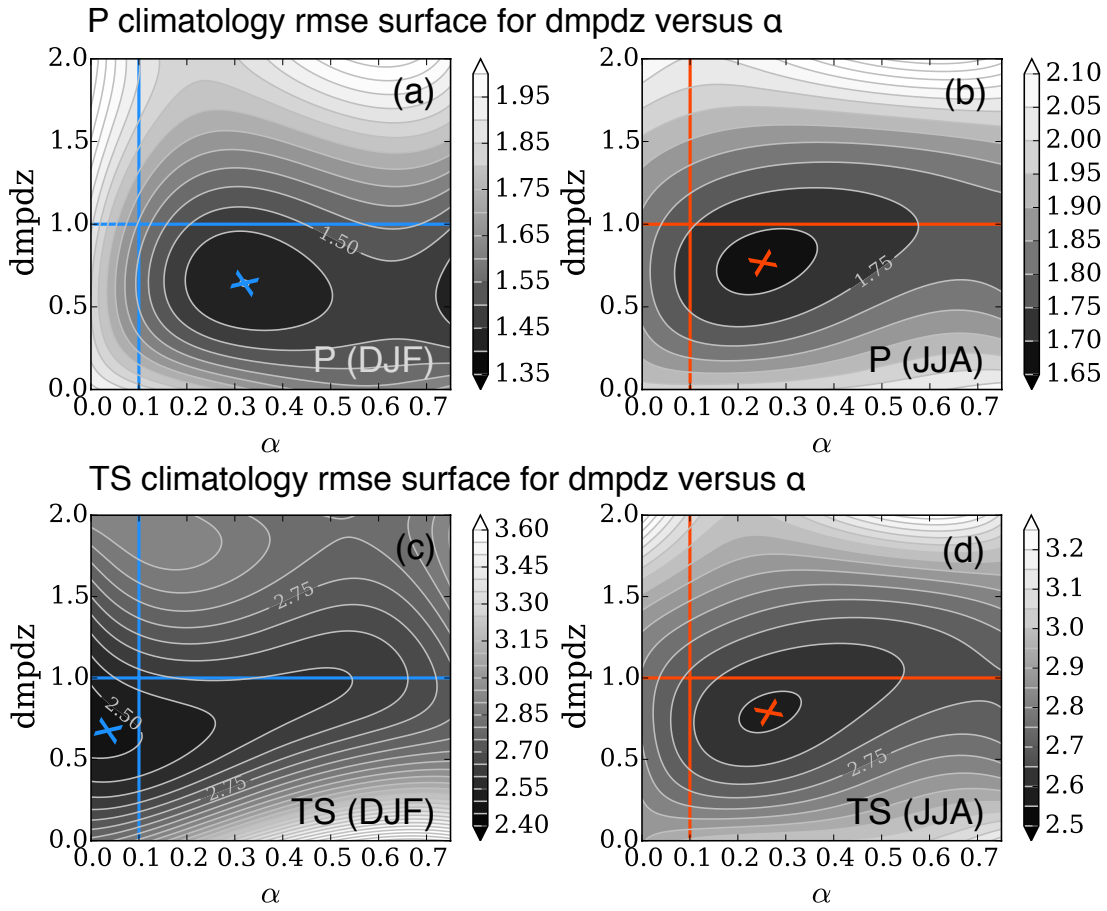

Figure 4.13: Rmse values as a function of dmpdz (vertical axis) and α (horizontal axis). Bivariate spline interpolation was used to extrapolate a surface extending from along-axis rmse values from the PUP-cut-HDMR metamodel (using the first three EOFs). This was accomplished using the Scipy 'SmoothBivariateSpline' class for surface interpolation. (a,b) DJF and JJA surfaces for P rmse values. (c,d) DJF and JJA surfaces for TS rmse values. P units (top row) are in mm day^{-1}, and TS units (bottom row) are in degrees Celsius. Blue and red lines show control axes for the α and dmpdz parameters, colored to match seasonal conventions in other figures. Contours vary in these plots but the light gray contour lines are spaced at a constant value of 0.05 mm day^{-1} or degrees Celsius. A colored **x** is placed by eye at the approximate locations of rmse minima.

combinations and, in the process, develop an understanding of the behavior of a field through its parameter space.

For the sake of illustration, we next vary two parameters at a time while holding the other two at control values. The two-dimensional rmse surface that results is in reality a four-dimensional rmse surface sliced at the control value by the two-dimensional plane of the parameters that are being varied. Note this is an estimate done without explicit interaction terms, as discussed earlier; while there are caveats to this approximation, we note there is already rich and nonlinear behavior

in the metamodel that is worth investigating.

Figure 4.13 shows two-dimensional rmse surfaces for global P and TS fields, calculated using PUP-cut-HDMR while varying the α and dmpdz parameters. To calculate these results, the PUP-cut-HDMR metamodel was applied to α and dmpdz separately, and model fields were reconstructed at all possible combinations of the sampled dmpdz and α values in Table 4.1 ($8 \times 4 = 40$ total combinations). Global rmse values were calculated, and the resulting surfaces were then smoothed using cubic bivariate spline interpolation to a 10×10 mesh. Blue and red lines represent the parameter axes for DJF and JJA, respectively; the control run is located at the intersection of these lines.

In both seasons and for both P and TS, moving from the control point to the lowest value on these rmse surfaces represents an improvement in model error by as much as 10%. Following along the parameter axes, the rmse values correspond to those seen for dmpdz and α in Figs. 4.5 and 4.12. Darker regions on these surfaces indicate lower rmse values, and for a modeler whose goal is to minimize some objective function or measure of model error, preferable parameter combinations lie in the darker regions.

Note that the details of the surface change both as a function of climate field (P versus TS) and season (DJF versus JJA). Such complex parameter sensitivity gives tradeoffs when attempting to optimize two fields at once. For example, the optimal α value for DJF P climatology (Fig. 4.13a) lies in the middle of the α range, but optimizing TS instead would draw α toward lower values. This type of contradiction is encountered frequently, and the tradeoffs involved in minimizing more than one objective function simultaneously are apparent in this two-dimensional illustration. Another noteworthy point is that the control run at the intersections of lines does not coincide with the lowest rmse point on these surfaces as approximated by each colored **x**. While only two parameters and two fields are considered here, this distance between the control and the minima on these surfaces highlight imperfectly constrained parameterizations in this GCM. Finally, high rmse gradients on these surfaces represent the most sensitive and nonlinear ranges. These are areas for the modeler to avoid whenever possible, bearing in mind tradeoffs from different objective functions.

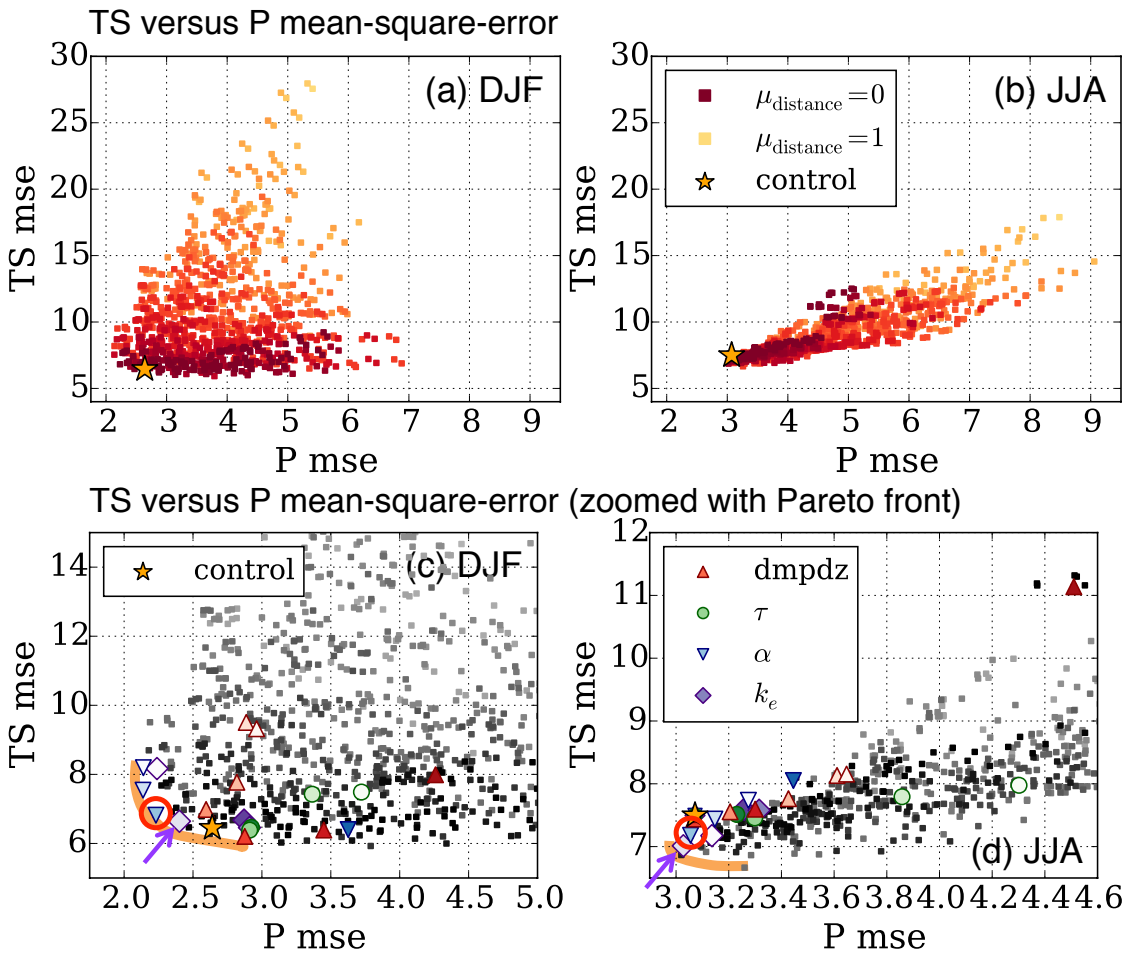

Figure 4.14: (a,b) DJF and JJA scatter plots of mse for TS versus P relative to their respective reanalysis and observational data sets, calculated for the global domain. Control points are indicated with stars. Yellow-orange-red squares represent the mse of each of 1000 parameter combinations (as calculated from Table 4.1) in the ensemble using PUP-cut-HDMR. Shading indicates the Euclidean distance (as a function of four parameters) to the control, normalized to have a maximum of 1 (dark red) and minimum of 0 (light yellow). (c,d) Same as top row, but zoomed closer, with colors now indicating along-axis points. The symbols in (c,d) represent distinct parameters according to the legend, with shading giving insight into its value. If a symbol has a darker shade than the legend, the parameter value is lower than its control value. If a symbol has a lighter tint than the legend, the parameter value is higher than its control value. Squares are plotted as gray to increase visibility for along-axis values. Vertical axis units are [degrees Celsius]^2, and horizontal axis units are mm^2 day^{-2}. Red circles are included in (c,d) for the value corresponding to an on-axis run with $\alpha = 0.25$. Purple arrows are included in (c,d) for the value corresponding to an on-axis run with $k_e = 5$ (see Table 4.1 for units).

4.4.4 Visualizing tradeoffs with Pareto fronts for global fields

The surfaces illustrated in Fig. 4.13 show the contradictions that often arise in the process of multiobjective optimization. Figure 4.13 allows one to visualize these tradeoffs when two parameters are varied at once, though it is more useful to visualize these tradeoffs when all $N = 4$ parameters

are varied. Figure 4.14 shows results from the PUP-cut-HDMR model calculated for all possible parameter combinations. Each point in Figs. 4.14a,b represents a unique parameter combination, $\boldsymbol{\mu} = (\text{dmpdz}, \tau, \alpha, k_e)$, from a set of $(8 \times 5 \times 5 \times 5) = 1000$ total points (see Table 4.1). These points can be viewed as projections of high-dimensional objective function vectors onto the P and TS mse plane. Colors represent the Euclidean distance in parameter space between a point's parameter values and the control values; the distances are normalized to have a maximum of one (light yellow) for the combination of all parameters at their highest endpoints, and zero (dark red) for all parameters at the control. The control integration itself is plotted as a star.

The edge of the cloud of points that lies closest to the origin in Figs. 4.14a,b represents a Pareto front. This is a boundary in the objective function space where improvement in one objective function comes at the cost of degradation in another, and points *on* this boundary together represent the "Pareto optimal" solutions. Here, the Pareto front is a line, but in reality it is a high-dimensional surface dictated by the number of objective functions being evaluated. In other words, the Pareto fronts here are visualized on pairs of objective function dimensions in what could be a much higher-dimensional space.

Figures 4.14c,d show a zoomed version of Figs. 4.14a,b, and schematic Pareto fronts have been drawn in light orange. Using these as a guide, the modeler can visualize when a given parameter value combination will improve the objective function (move the point closer to the Pareto front), or when there are solely tradeoffs to be considered (move the point along the Pareto front). Furthermore, one can estimate the percent improvement that would be obtained by updating from the control run to a new set of parameter values. In the case of these global mse measures, some positive news to the modeler is worth noting. First, the control integration sits near the edge of the Pareto front in DJF and fairly close in JJA, indicating that it is well tuned to observations at the global scale. Second, the points in Figs. 2.14a,b that are farthest from the Pareto front are also farthest from the control parameters (i.e., they tend to be more yellow than red). This latter observation implies that optimal parameter values for a model at the global scale will be within the vicinity of the control values, though specifics will differ based on objective functions.

To provide a practical example, we look more closely at Fig. 4.14d. We can start with the control run and ask what parameter changes might be made to improve the objective functions

both for P and for TS. In this case, updating k_e from 1 to 5 would accomplish this by placing the new point at the light purple diamond that sits on the Pareto front (as indicated by a purple arrow). This would lower both the P and TS error and shift the model closer to the Pareto front. Making this change, however, would cause a small increase in P mse during DJF by as much as 10%. This can can be seen in Fig. 4.14c, where the control run is located closer to the Pareto front than the purple diamond to its right. While the changes to model performance are by mse measures modest, the modeler could use this information together with other variables and objective functions—as well as empirical data for the parameters being changed—to make an informed decision on whether it is a sensible parameter update.

4.4.5 Visualizing tradeoffs and Pareto fronts for the tropical Pacific region

Until now, objective functions have been applied over a global domain, but modelers are also typically interested in optimizing regional fields. Figure 4.15 shows rmse surfaces for the domain 40°S–40°N and 120°E–200°E. This encompasses the tropical (and subtropical) Pacific Ocean and includes the intertropical convergence zone (ITCZ) and south Pacific convergence zone (SPCZ). This particular perturbed physics ensemble is therefore quite relevant given the importance of deep convection and moist processes in this region. Comparing Figs. 4.15a,b with those for the global analysis in 4.13, one can see that the tropical Pacific rmse is generally higher for P (given the sheer magnitude of P in the tropics) and lower for TS (since high southern latitudes showing large bias in Fig. 4.8c have been excluded from the domain). The shapes of the rmse surfaces and the locations of minima are similar across parameter and season, and the distance between each origin and the red or blue **x** gives a visual sense of how much the model can be improved in these two dimensions, typically between 5 and 10%.

Figure 4.16 shows the TS versus P mse scatter plot for all parameter combinations in the tropical Pacific. Note in Figs. 4.16a,b that the control run for the tropical Pacific domain lies slightly off the Pareto fronts in DJF and JJA for this region. This stands in contrast to the global analysis in Fig. 4.14, likely because the model has been historically tuned with global measures in mind. This also gives additional information to modelers who for many purposes want to reduce regional model error, especially with regard to P and the hydrological cycle. Approximate Pareto fronts are

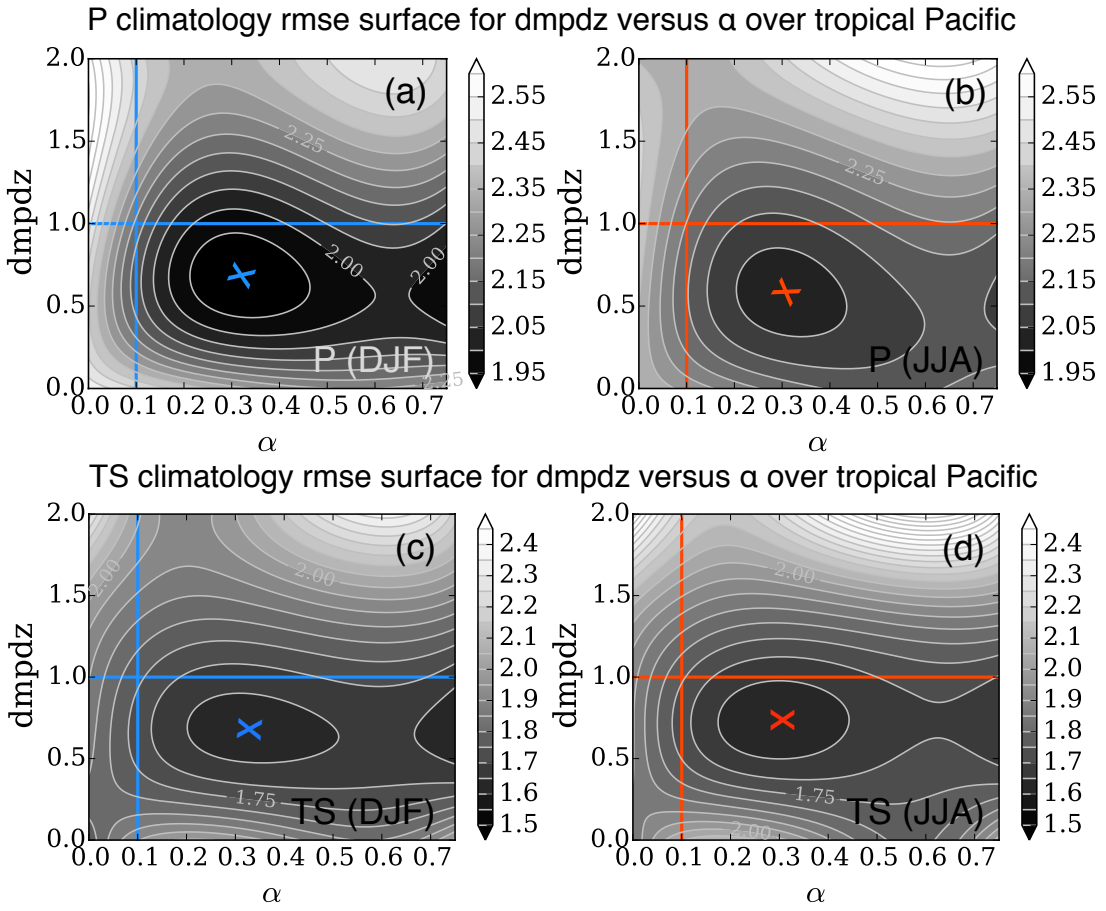

Figure 4.15: Rmse values as a function of dmpdz (vertical axis) and α (horizontal axis). Bivariate spline interpolation was used to extrapolate a surface extending from along-axis rmse values from the PUP-cut-HDMR metamodel (using the first three EOFs). This was accomplished using the Scipy 'SmoothBivariateSpline' class for surface interpolation. (a,b) DJF and JJA surfaces for P rmse values. (c,d) DJF and JJA surfaces for TS rmse values. P units (top row) are in mm day^{-1}, and TS units (bottom row) are in degrees Celsius. Blue and red lines show control axes for the α and dmpdz parameters, colored to match seasonal conventions in other figures. Contours vary in these plots but the light gray contour lines are spaced at a constant value of 0.05 mm day^{-1} or degrees Celsius. A colored **x** is placed by eye at the approximate locations of rmse minima.

drawn in Figs. 4.16c,d, and it is clear that several possible parameter combinations distinct from the control values or other sampled integrations can help can help improve model performance in this region.

Focusing on Figs. 4.16c,d, it is useful to look for parameter changes that would move the model closer to the Pareto front in this regional domain. This is true for both DJF and JJA if we choose the light blue upside down triangle (circled in red), which corresponds to an increase in α from 0.1 to 0.25. Comparing this to the global domain, Figs. 4.14c,d also show that this α update

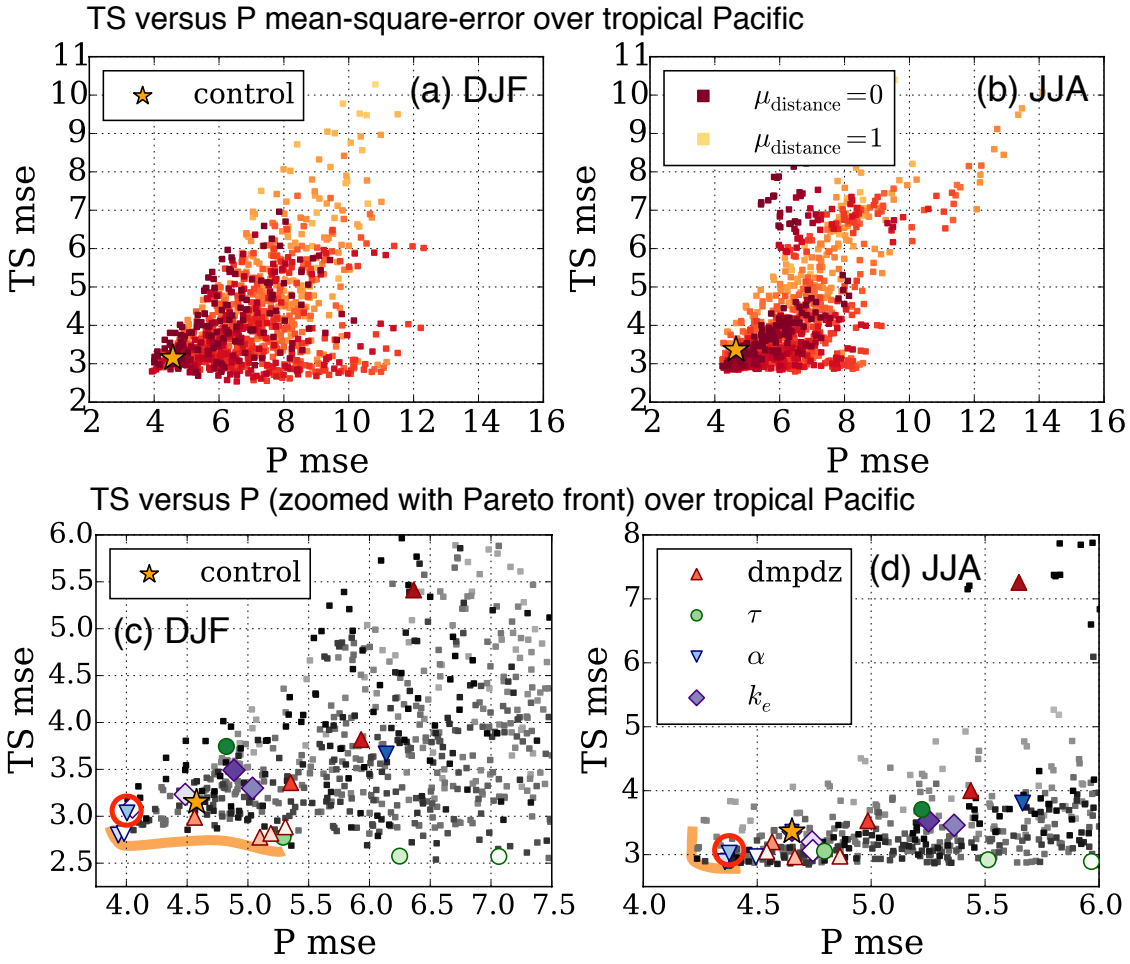

Figure 4.16: (a,b) DJF and JJA scatter plots of mse for TS versus P relative to their respective reanalysis and observational data sets, calculated for the tropical Pacific domain. Control points are indicated with stars. Yellow-orange-red squares represent the mse of each of 1000 parameter combinations (as calculated from Table 4.1) in the ensemble, calculated using PUP-cut-HDMR. Shading indicates the Euclidean distance (as a function of four parameters) to the control, normalized to have a maximum of 1 (dark red) and minimum of 0 (light yellow). (c,d) Same as top row, but zoomed closer, with colors now indicating along-axis points. The symbols in (c,d) represent distinct parameters according to the legend, with shading giving insight into its value. If a symbol has a darker shade than the legend, the parameter value is lower than its control value. If a symbol has a lighter tint than the legend, the parameter value is higher than its control value. Squares are plotted as gray to increase visibility for along-axis values. Vertical axis units are [degrees Celsius]^2, and horizontal axis units are mm^2 day^{-2}. Red circles are included in (c,d) for the value corresponding to an on-axis run with $\alpha = 0.25$.

causes the model to move closer to the Pareto front during JJA and along the Pareto front during DJF. This result, which requires only changing one parameter, could improve model performance in these objective error measures in the tropical Pacific domain by 10–15% for P and by 5–10% for TS, with a modest improvement globally as well. Again, a modeler would at this point be expected to bring in relevant information that helps inform whether an α update is sensible. For example,

making informed decisions about α based on knowledge of convective physics and incorporating observations when available will help determine whether the update is physically plausible.

As a final point, we focus on the Pareto front drawn in Fig. 4.16d. At the very corner of this front, one can see three separate parameter combinations from the metamodel (gray squares) that have a notable separation from other points in this scatter plot. From the figure itself (i.e., the gray shading), the parameter values that correspond to these points are not obvious, but it would be fairly straightforward to construct an interactive plot (or a very basic graphical user interface) that allows the user to click on these squares and obtain the parameter values for them. Armed with this knowledge, a modeler could then integrate the GCM using these parameter values and others nearby, and these additional points could be fed into the metamodel to check its predictive capacity and better constrain it in the vicinity of the Pareto front.

4.5 Summary and conclusions

An existing perturbed physics ensemble is used to explore several key aspects of parameter sensitivity for precipitation (P) and surface temperature (TS). Perturbed physics ensembles isolate uncertainties in sub-grid scale physics and parameterizations in a global climate model (GCM), and the ensemble used here consists of integrations in which four influential parameters in the deep convection scheme are systematically sampled.

For a GCM, varying N parameters each at a density k would require k^N total simulations to fully sample the parameter space. Such a calculation is unfeasible, and metamodeling is used as a cost-efficient alternative. A quadratic metamodel calculated over seasonal P and TS fields proves adept at capturing parameter sensitivity over most ranges of the sampled parameters, though it fails over highly sensitive regions, especially those associated with low entrainment fraction parameter values. This can in part be remedied by training the metamodel over a trust region that excludes these sensitive points.

An alternative metamodel is developed using leading modes from uncertainty pattern techniques developed in Chapter 3 in combination with a metamodeling approach from the engineering literature. We label this the PUP-cut-HDMR metamodel and show that it is able to capture the highly sensitive parameter dependence where the quadratic metamodel fails, serving as a skill-

ful alternative when parameter sensitivity is highly nonlinear. This metamodel is then used to extrapolate off parameter axes and sample the full parameter space in the model ensemble.

We demonstrate the utility of this metamodel by analyzing objective functions (model root-mean-square error and mean-square error relative to observations) over the global domain and the tropical Pacific. This is first illustrated as a function of two parameters at a time and gives quantitative information on how the model error varies in parameter space. We then evaluate combinations of parameters in objective function space, yielding information about what parameter combinations can improve multiple objective functions simultaneously. The concept of a Pareto front is used to visualize tradeoffs where updating parameterizations can yield improvement in one model field but degradation in another. We discuss particular cases where updating parameter values based on global and regional objective functions can result in measurable model improvement.

Exploring how a model and its fields behave in parameter space is an important step in model tuning and optimization. The contributions here provide quantitative and visualization tools for understanding changes in GCM fields as a function of multiple parameters. The type of analysis shown here for GCM performance in current climate could in future work be extended to include global warming simulations. We anticipate that constraining current GCMs by observations will not only be critical to improving their simulation of the current climate but may also help narrow the range of uncertainty seen in end-of-century projections.

Bibliography

AchutaRao, K., and K. R. Sperber, 2006: ENSO simulation in coupled ocean-atmosphere models: are the current models better? *Climate Dynamics*, **27 (1)**, 1–15.

Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P.-P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, and E. Nelkin, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, **4**, 1147–1167.

Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature*, **419 (6903)**, 224–232.

Anderson, B. T., 2004: Investigation of a large-scale mode of ocean–atmosphere variability and its relation to tropical Pacific sea surface temperature anomalies. *Journal of Climate*, **17 (20)**, 4089–4098.

Anderson, B. T., B. R. Lintner, B. Langenbrunner, J. D. Neelin, E. Hawkins, and J. Syktus, 2015: Sensitivity of terrestrial precipitation trends to the structural evolution of sea surface temperatures. *Geophysical Research Letters*, **42 (4)**, 1190–1196.

Ashok, K., S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, 2007: El niño modoki and its possible teleconnection. *Journal of Geophysical Research: Oceans*, **112 (C11)**.

Athanasiadis, P. J., J. M. Wallace, and J. J. Wettstein, 2010: Patterns of wintertime jet stream variability and their relation to the storm tracks*. *Journal of the Atmospheric Sciences*, **67 (5)**, 1361–1381.

Bellprat, O., S. Kotlarski, D. Lüthi, and C. Schär, 2012: Objective calibration of regional climate models. *Journal of Geophysical Research: Atmospheres*, **117 (D23)**.

Bengtsson, L., K. I. Hodges, and E. Roeckner, 2006: Storm tracks and climate change. *Journal of Climate*, **19 (15)**, 3518–3543.

Berg, N., A. Hall, F. Sun, S. Capps, D. Walton, B. Langenbrunner, and J. D. Neelin, 2014: Twenty-first-century precipitation changes over the Los Angeles region. *Journal of Climate*, **28 (2)**, 401–421.

Bernstein, D. N., 2014: Evaluation of regional sensitivities in climate modeling. Ph.D. thesis, Hebrew Univeresity of Jerusalem.

Bernstein, D. N., and J. D. Neelin, 2015: Identifying sensitive ranges in global warming precipitation change dependence on convective parameters, submitted.

Bracco, A., J. D. Neelin, H. Luo, J. C. McWilliams, and J. E. Meyerson, 2013: High dimensional decision dilemmas in climate models. *Geoscientific Model Development*, **6 (2)**, 2731–2767.

Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate*, **5 (6)**, 541–560.

Brown, J., A. Moise, and F. Delage, 2012: Changes in the south pacific convergence zone in ipcc ar4 future climate projections. *Climate Dynamics*, **39 (1-2)**, 1–19.

Brown, J., A. Sen Gupta, J. Brown, L. Muir, J. Risbey, P. Whetton, X. Zhang, A. Ganachaud, B. Murphy, and S. Wijffels, 2013: Implications of cmip3 model biases and uncertainties for climate projections in the western tropical pacific. *Climatic Change*, **119 (1)**, 147–161.

Caballero, R., and B. T. Anderson, 2009: Impact of midlatitude stationary waves on regional Hadley cells and ENSO. *Geophysical Research Letters*, **36 (17)**, L17 704.

Cai, W., A. Sullivan, and T. Cowan, 2009: Rainfall teleconnections with Indo-Pacific variability in the WCRP CMIP3 models. *Journal of Climate*, **22 (19)**, 5046–5071.

Cai, W., M. Lengaigne, S. Borlace, M. Collins, T. Cowan, M. J. McPhaden, A. Timmermann, S. Power, J. Brown, C. Menkes, A. Ngari, E. M. Vincent, and M. J. Widlansky, 2012: More extreme swings of the South Pacific Convergence Zone due to greenhouse warming. *Nature*, **488 (7411)**, 365–369.

Capotondi, A., A. Wittenberg, and S. Masina, 2006: Spatial and temporal structure of tropical Pacific interannual variability in 20th century coupled simulations. *Ocean Modelling*, **15 (3–4)**, 274–298.

Cash, B., E. Schneider, and L. Bengtsson, 2005: Origin of regional climate differences: role of boundary conditions and model formulation in two GCMs. *Climate Dynamics*, **25 (7-8)**, 709–723.

Cayan, D. R., 1992: Latent and sensible heat flux anomalies over the northern oceans: Driving the sea surface temperature. *Journal of Physical Oceanography*, **22 (8)**.

Chadwick, R., I. Boutle, and G. Martin, 2012: Spatial patterns of precipitation change in cmip5: Why the rich do not get richer in the tropics. *Journal of Climate*, **26 (11)**, 3803–3822.

Chang, E. K. M., 2013: Cmip5 projection of significant reduction in extratropical cyclone activity over north america. *Journal of Climate*, **26 (24)**, 9903–9922.

Chang, E. K. M., Y. Guo, and X. Xia, 2012: Cmip5 multimodel ensemble projection of storm track change under global warming. *Journal of Geophysical Research: Atmospheres*, **117 (D23)**, D23 118.

Chang, E. K. M., and M. Zheng, 2013: Medium-range ensemble sensitivity analysis of two extreme pacific extratropical cyclones. *Monthly Weather Review*, **141 (211-231)**.

Chen, W. Y., and H. M. van den Dool, 1997: Asymmetric impact of tropical sst anomalies on atmospheric internal variability over the north pacific. *Journal of the Atmospheric Sciences*, **54 (6)**, 725–740.

Cheng, X., and T. J. Dunkerton, 1995: Orthogonal rotation of spatial patterns derived from singular value decomposition analysis. *Journal of Climate*, **8**, 2631–2643.

Chiang, J. C. H., and A. H. Sobel, 2002: Tropical tropospheric temperature variations caused by ENSO and their influence on the remote tropical climate. *Journal of Climate*, **15 (18)**, 2616–2631.

Chiang, J. C. H., and D. J. Vimont, 2004: Analogous pacific and atlantic meridional modes of tropical atmosphere-ocean variability. *Journal of Climate*, **17 (21)**, 4143–4158.

Chou, C., and J. D. Neelin, 2004: Mechanisms of global warming impacts on regional tropical precipitation. *Journal of Climate*, **17 (13)**, 2688–2701.

Chou, C., J. D. Neelin, C.-A. Chen, and J.-Y. Tu, 2009: Evaluating the "rich-get-richer" mechanism in tropical precipitation change under global warming. *Journal of Climate*, **22 (8)**, 1982–2005.

Coelho, C. A. S., and L. Goddard, 2009: El Niño–induced tropical droughts in climate change projections. *Journal of Climate*, **22 (23)**, 6456–6476.

Collins, M., B. B. Booth, B. Bhaskaran, G. Harris, J. Murphy, D. H. Sexton, and M. Webb, 2011: Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. *Climate Dynamics*, **36 (9-10)**, 1737–1766.

Collins, M., S.-I. An, W. Cai, A. Ganachaud, E. Guilyardi, F.-F. Jin, M. Jochum, M. Lengaigne, S. Power, A. Timmermann, G. Vecchi, and A. Wittenberg, 2010: The impact of global warming on the tropical Pacific Ocean and El Niño. *Nature Geosciences*, **3 (6)**, 391–397.

Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, **19 (18)**, 4605–4630.

Dai, A., and T. M. L. Wigley, 2000: Global patterns of enso-induced precipitation. *Geophysical Research Letters*, **27 (9)**, 1283–1286.

Davey, M., M. Huddleston, K. Sperber, P. Braconnot, F. Bryan, D. Chen, R. Colman, C. Cooper, U. Cubasch, P. Delecluse, D. DeWitt, L. Fairhead, G. Flato, C. Gordon, T. Hogan, M. Ji, M. Kimoto, A. Kitoh, T. Knutson, M. Latif, H. Le Treut, T. Li, S. Manabe, C. Mechoso, G. Meehl, S. Power, E. Roeckner, L. Terray, A. Vintzileos, R. Voss, B. Wang, W. Washington, I. Yoshikawa, J. Yu, S. Yukimoto, and S. Zebiak, 2002: STOIC: a study of coupled model climatology and variability in tropical ocean regions. *Climate Dynamics*, **18 (5)**, 403–420.

Delcambre, S. C., D. J. Lorenz, D. J. Vimont, and J. E. Martin, 2013a: Diagnosing northern hemisphere jet portrayal in 17 CMIP3 global climate models: Twentieth-century intermodel variability. *Journal of Climate*, **26 (14)**, 4910–4929.

Delcambre, S. C., D. J. Lorenz, D. J. Vimont, and J. E. Martin, 2013b: Diagnosing northern hemisphere jet portrayal in 17 CMIP3 global climate models: Twenty-first-century projections. *Journal of Climate*, **26 (14)**, 4930–4946.

Delecluse, P., M. K. Davey, Y. Kitamura, S. G. H. Philander, M. Suarez, and L. Bengtsson, 1998: Coupled general circulation modeling of the tropical pacific. *Journal of Geophysical Research: Oceans*, **103 (C7)**, 14 357–14 373.

Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change projections: the role of internal variability. *Climate Dynamics*, **38 (3-4)**, 527–546.

Dettinger, M. D., D. R. Cayan, H. F. Diaz, and D. M. Meko, 1998: North–south precipitation patterns in western north america on interannual-to-decadal timescales. *Journal of Climate*, **11 (12)**, 3095–3111.

DeWeaver, E., and S. Nigam, 2004: On the forcing of ENSO teleconnections by anomalous heating and cooling. *Journal of Climate*, **17 (16)**, 3225–3235.

Doherty, R., and M. Hulme, 2002: The relationship between the soi and extended tropical precipitation in simulations of future climate change. *Geophysical Research Letters*, **29 (10)**, 113–1–113–4.

Durack, P. J., S. E. Wijffels, and R. J. Matear, 2012: Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. *Science*, **336 (6080)**, 455–458.

Flato, G., J. Marotzke, B. Abiodun, P. Braconnot, S. C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason, and M. Rummukainen, 2013: *Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chap. 9: Evaluation of climate models. Cambridge University Press.

Gates, W. L., J. S. Boyle, C. Covey, C. G. Dease, C. M. Doutriaux, R. S. Drach, M. Fiorino, P. J. Gleckler, J. J. Hnilo, S. M. Marlais, T. J. Phillips, G. L. Potter, B. D. Santer, K. R. Sperber, K. E. Taylor, and D. N. Williams, 1999: An overview of the results of the atmospheric model intercomparison project (amip i). *Bulletin of the American Meteorological Society*, **80 (1)**, 29–55.

Giorgi, F., and R. Francisco, 2000: Evaluating uncertainties in the prediction of regional climate change. *Geophysical Research Letters*, **27 (9)**, 1295–1298.

Greve, P., B. Orlowsky, B. Mueller, J. Sheffield, M. Reichstein, and S. I. Seneviratne, 2014: Global assessment of trends in wetting and drying over land. *Nature Geosci*, **7 (716-721)**.

Grise, K. M., and L. M. Polvani, 2014: The response of mid-latitude jets to increased co2: Distinguishing the roles of sea surface temperature and direct radiative forcing. *Geophysical Research Letters*, 2014GL061638.

Guilyardi, E., P. Braconnot, F.-F. Jin, S. T. Kim, M. Kolasinski, T. Li, and I. Musat, 2009a: Atmosphere feedbacks during ENSO in a coupled GCM with a modified atmospheric convection scheme. *Journal of Climate*, **22 (21)**, 5698–5718.

Guilyardi, E., A. Wittenberg, A. Fedorov, M. Collins, C. Wang, A. Capotondi, G. J. van Oldenborgh, and T. Stockdale, 2009b: Understanding el niño in ocean-atmosphere general circulation models: Progress and challenges. *Bulletin of the American Meteorological Society*, **90 (3)**, 325–340.

Guilyardi, E., S. Gualdi, J. Slingo, A. Navarra, P. Delecluse, J. Cole, G. Madec, M. Roberts, M. Latif, and L. Terray, 2004: Representing el niño in coupled ocean–atmosphere gcms: The dominant role of the atmospheric component. *Journal of Climate*, **17 (24)**, 4623–4629.

Hannachi, A., I. T. Jolliffe, D. B. Stephenson, and N. Trendafilov, 2006: In search of simple structures in climate: simplifying eofs. *International Journal of Climatology*, **26 (1)**, 7–28.

Harr, P. A., D. Anwender, and S. C. Jones, 2008: Predictability associated with the downstream impacts of the extratropical transition of tropical cyclones: Methodology and a case study of typhoon nabi (2005). *Monthly Weather Review*, **136 (9)**, 3205–3225.

Hawkins, E., and R. Sutton, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, **37 (1-2)**, 407–418.

Held, I. M., S. W. Lyons, and S. Nigam, 1989: Transients and the extratropical response to el niño. *Journal of the Atmospheric Sciences*, **46 (1)**, 163–174.

Held, I. M., and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *Journal of Climate*, **19 (21)**, 5686–5699.

Hirota, N., and Y. Takayabu, 2013: Reproducibility of precipitation distribution over the tropical oceans in cmip5 multi-climate models compared to cmip3. *Climate Dynamics*, **41 (11-12)**, 2909–2920.

Horel, J. D., 1981: A rotated principal component analysis of the interannual variability of the northern hemisphere 500 mb height field. *Monthly Weather Review*, **109 (10)**, 2080–2092.

Horel, J. D., and J. M. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the southern oscillation. *Monthly Weather Review*, **109 (4)**, 813–829.

Ihara, C., and Y. Kushnir, 2009: Change of mean midlatitude westerlies in 21st century climate simulations. *Geophysical Research Letters*, **36 (13)**, L13 701.

Joseph, R., and S. Nigam, 2006: ENSO evolution and teleconnections in IPCC's twentieth-century climate simulations: Realistic representation? *Journal of Climate*, **19 (17)**, 4360–4377.

Jun, M., R. Knutti, and D. W. Nychka, 2008a: Logal eigenvalue analysis of cmip3 climate model errors. *Tellus*, **60A (5)**, 992–1000.

Jun, M., R. Knutti, and D. W. Nychka, 2008b: Spatial analysis to quantify numerical model bias and dependence: How many climate models are there? *Journal of the American Statistical Association*, 934–947.

Kaiser, H., 1958: The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23 (3)**, 187–200.

Kanamitsu, M., A. Kumar, H.-M. H. Juang, J.-K. Schemm, W. Wang, F. Yang, S.-Y. Hong, P. Peng, W. Chen, S. Moorthi, and M. Ji, 2002: Ncep dynamical seasonal forecast system 2000. *Bulletin of the American Meteorological Society*, **83 (7)**, 1019–1037.

Kao, H.-Y., and J.-Y. Yu, 2009: Contrasting eastern-pacific and central-pacific types of enso. *Journal of Climate*, **22 (3)**, 615–632.

Keller, J. H., S. C. Jones, J. L. Evans, and P. A. Harr, 2011: Characteristics of the TIGGE multimodel ensemble prediction system in representing forecast variability associated with extratropical transition. *Geophysical Research Letters*, **38 (12)**, L12 802.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, 2010: Challenges in combining projections from multiple climate models. *Journal of Climate*, **23**, 2739–2758.

Knutti, R., D. Masson, and A. Gettelman, 2013: Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, **40**, 1194–1199.

Knutti, R., and J. Sedlácek, 2013: Robustness and uncertainties in the new cmip5 climate model projections. *Nature Climate Change*, **3 (4)**, 369–373.

Kumar, S., V. Merwade, J. L. Kinter, and D. Niyogi, 2013: Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. *Journal of Climate*, **26 (12)**, 4168–4185.

Langenbrunner, B., and J. D. Neelin, 2013: Analyzing ENSO teleconnections in CMIP models as a measure of model fidelity in simulating precipitation. *Journal of Climate*, **26 (13)**, 4431–4446.

Langenbrunner, B., J. D. Neelin, B. R. Lintner, and B. T. Anderson, 2015: Patterns of precipitation change and climatological uncertainty among CMIP5 models, with a focus on the midlatitude Pacific storm track. *Journal of Climate*, **28**, 7858–7872.

Langford, S., S. Stevenson, and D. Noone, 2014: Analysis of low-frequency precipitation variability in cmip5 historical simulations for southwestern north america. *Journal of Climate*, **27 (7)**, 2735–2756.

Latif, M., K. Sperber, J. Arblaster, P. Braconnot, D. Chen, A. Colman, U. Cubasch, C. Cooper, P. Delecluse, D. Dewitt, L. Fairhead, G. Flato, T. Hogan, M. Ji, M. Kimoto, A. Kitoh, T. Knutson, H. Le Treut, T. Li, S. Manabe, O. Marti, C. Mechoso, G. Meehl, S. Power, E. Roeckner, J. Sirven, L. Terray, A. Vintzileos, R. Voß, B. Wang, W. Washington, I. Yoshikawa, J. Yu, and S. Zebiak, 2001: Ensip: the el niño simulation intercomparison project. *Climate Dynamics*, **18 (3-4)**, 255–276.

Lau, N.-C., 1988: Variability of the observed midlatitude storm tracks in relation to low-frequency changes in the circulation pattern. *Journal of the Atmospheric Sciences*, **45 (19)**, 2718–2743.

Li, G., C. Rosenthal, and H. Rabitz, 2001: High dimensional model representations. *The Journal of Physical Chemistry A*, **105 (33)**, 7765–7777.

Li, G., and S.-P. Xie, 2012: Origins of tropical-wide SST biases in CMIP multi-model ensembles. *Geophysical Research Letters*, **39 (22)**, L22 703.

Li, G., and S.-P. Xie, 2013: Tropical biases in cmip5 multimodel ensemble: The excessive equatorial pacific cold tongue and double itcz problems. *Journal of Climate*, **27 (4)**, 1765–1780.

Lin, J.-L., 2007: The double-ITCZ problem in IPCC AR4 coupled GCMs: Ocean–atmosphere feedback analysis. *Journal of Climate*, **20 (18)**, 4497–4525.

Linkin, M. E., and S. Nigam, 2008: The north pacific oscillation–west pacific teleconnection pattern: Mature-phase structure and winter impacts. *Journal of Climate*, **21 (9)**, 1979–1997.

Lintner, B. R., M. Biasutti, N. Diffenbaugh, J.-E. Lee, M. Niznik, and K. Findell, 2012: Amplification of wet and dry month occurrence over tropical land regions in response to global warming. *Journal of Geophysical Research*, **117 (D11106)**.

Lloyd, J., E. Guilyardi, H. Weller, and J. Slingo, 2009: The role of atmosphere feedbacks during ENSO in the CMIP3 models. *Atmospheric Science Letters*, **10 (3)**, 170–176.

Lu, J., G. A. Vecchi, and T. Reichler, 2007: Expansion of the hadley cell under global warming. *Geophysical Research Letters*, **34 (6)**, L06 805.

Ma, J., and S.-P. Xie, 2012: Regional patterns of sea surface temperature change: A source of uncertainty in future projections of precipitation and atmospheric circulation. *Journal of Climate*, **26 (8)**, 2482–2501.

Mahlstein, I., R. W. Portmann, J. S. Daniel, S. Solomon, and R. Knutti, 2012: Perceptible changes in regional precipitation in a future climate. *Geophysical Research Letters*, **39 (5)**, L05 701.

Maloney, E. D., S. J. Camargo, E. Chang, B. Colle, R. Fu, K. L. Geil, Q. Hu, X. Jiang, N. Johnson, K. B. Karnauskas, J. Kinter, B. Kirtman, S. Kumar, B. Langenbrunner, K. Lombardo, L. N. Long, A. Mariotti, J. E. Meyerson, K. C. Mo, J. D. Neelin, Z. Pan, R. Seager, Y. Serra, A. Seth, J. Sheffield, J. Stroeve, J. Thibeault, S.-P. Xie, C. Wang, B. Wyman, and M. Zhao, 2014: North american climate in CMIP5 experiments: Part III: Assessment of Twenty-first Century projections. *Journal of Climate*, **27 (6)**, 2230–2270.

Manabe, S., and R. J. Stouffer, 1980: Sensitivity of a global climate model to an increase of co_2 concentration in the atmosphere. *Journal of Geophysical Research: Oceans*, **85 (C10)**, 5529–5554.

Manabe, S., and R. T. Wetherald, 1980: On the distribution of climate change resulting from an increase in co_2 content of the atmosphere. *Journal of the Atmospheric Sciences*, **37 (1)**, 99–118.

Mechoso, C. R., A. W. Robertson, N. Barth, M. K. Davey, P. Delecluse, P. R. Gent, S. Ineson, B. Kirtman, M. Latif, H. L. Treut, T. Nagai, J. D. Neelin, S. G. H. Philander, J. Polcher, P. S. Schopf, T. Stockdale, M. J. Suarez, L. Terray, O. Thual, and J. J. Tribbia, 1995: The seasonal cycle over the tropical pacific in coupled ocean–atmosphere general circulation models. *Monthly Weather Review*, **123 (9)**, 2825–2838.

Meehl, G. A., T. F. Stocker, W. D. Collins, P. Friedlingstein, A. T. Gaye, J. M. Gregory, A. Kitoh, R. Knutti, J. M. Murphy, A. Noda, S. C. B. Raper, I. G. Watterson, A. J. Weaver, and Z.-C. Zhao, 2007: *Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.*, chap. Global climate projections. Cambridge University Press.

Mehran, A., A. AghaKouchak, and T. J. Phillips, 2014: Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations. *Journal of Geophysical Research: Atmospheres*, **119 (4)**.

Merryfield, W. J., 2006: Changes to enso under co2 doubling in a multimodel ensemble. *Journal of Climate*, **19 (16)**, 4009–4027.

Münnich, M., and J. D. Neelin, 2005: Seasonal influence of ENSO on the Atlantic ITCZ and equatorial South America. *Geophysical Research Letters*, **32 (21)**.

Neale, R. B., J. H. Richter, A. J. Conley, S. Park, P. H. Lauritzen, A. Gettelman, D. L. Williamson, P. J. Rasch, S. J. Vavrus, M. A. Taylor, W. D. Collins, M. Zhang, and S.-J. Lin, 2010: Description of the NCAR Community Atmosphere Model (CAM 4.0). NCAR Technical Note TN-485, National Center for Atmospheric Research.

Neelin, J., M. Latif, M. Allaart, M. Cane, U. Cubasch, W. Gates, P. Gent, M. Ghil, C. Gordon, N. Lau, C. Mechoso, G. Meehl, J. Oberhuber, S. Philander, P. Schopf, K. Sperber, K. Sterl, T. Tokioka, J. Tribbia, and S. Zebiak, 1992: Tropical air-sea interaction in general circulation models. *Climate Dynamics*, **7 (2)**, 73–104.

Neelin, J. D., A. Bracco, H. Luo, J. C. McWilliams, and J. E. Meyerson, 2010: Considerations for parameter optimization and sensitivity in climate models. *Proceedings of the National Academy of Sciences*, **107 (50)**, 21 349–21 354.

Neelin, J. D., C. Chou, and H. Su, 2003: Tropical drought regions in global warming and el niño teleconnections. *Geophysical Research Letters*, **30 (24)**, 2275.

Neelin, J. D., B. Langenbrunner, J. E. Meyerson, A. Hall, and N. Berg, 2013: California winter precipitation change under global warming in the Coupled Model Intercomparison Project phase 5 ensemble. *Journal of Climate*, **26 (17)**, 6238–6256.

Neelin, J. D., M. Munnich, H. Su, J. E. Meyerson, and C. E. Holloway, 2006: Tropical drying trends in global warming models and observations. *Proceedings of the National Academy of Sciences of the United States of America*, **103 (16)**, 6110–6115.

Park, J.-H., and S.-I. An, 2014: Southward displacement of the upper atmosphere zonal jet in the eastern pacific due to global warming. *Geophysical Research Letters*, **41 (22)**, 7861–7867.

Power, S. B., F. Delage, R. Colman, and A. Moise, 2011: Consensus on twenty-first-century rainfall projections in climate models more widespread than previously thought. *Journal of Climate*, **25 (11)**, 3792–3809.

Rabitz, H., and O. F. Alis, 1999: General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry*, **25 (2–3)**, 197–233.

Rabitz, H., O. F. Alis, J. Shorter, and K. Shim, 1999: Efficient input–output model representations. *Computer Physics Communications*, **117 (1–2)**, 11–20.

Räisänen, J., 2007: How reliable are climate models? *Tellus A*, **59 (1)**, 2–29.

Randall, D. A., R. A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R. J. Stouffer, A. Sumi, and K. E. Taylor, 2007: *Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.*, chap. Climate models and their evaluation. Cambridge University Press.

Richman, M. B., 1986: Rotation of principal components. *Journal of Climatology*, **6 (3)**, 293–335.

Risbey, J. S., P. C. McIntosh, M. J. Pook, H. A. Rashid, and A. C. Hirst, 2011: Evaluation of rainfall drivers and teleconnections in an ACCESS AMIP run. *Australian Meteorological and Oceanographic Journal*, **61 (91-105)**.

Roderick, M. L., F. Sun, W. H. Lim, and G. D. Farquhar, 2014: A general framework for understanding the response of the water cycle to global warming over land and ocean. *Hydrol. Earth Syst. Sci.*, **18 (5)**, 1575–1589.

Ropelewski, C. F., and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly Weather Review*, **115 (8)**, 1606–1626.

Rowell, D. P., 2013: Simulating SST teleconnections to Africa: What is the state of the art? *Journal of Climate*, **26 (15)**, 5397–5418.

Schaller, N., I. Mahlstein, J. Cermak, and R. Knutti, 2011: Analyzing precipitation projections: A comparison of different approaches to climate model evaluation. *Journal of Geophysical Research: Atmospheres*, **116 (D10)**.

Scheff, J., and D. M. W. Frierson, 2012a: Robust future precipitation declines in cmip5 largely reflect the poleward expansion of model subtropical dry zones. *Geophysical Research Letters*, **39 (18)**.

Scheff, J., and D. M. W. Frierson, 2012b: Twenty-first-century multimodel subtropical precipitation declines are mostly midlatitude shifts. *Journal of Climate*, **25**, 4330–4347.

Seager, R., and G. A. Vecchi, 2010: Greenhouse warming and the 21st century hydroclimate of southwestern north america. *Proceedings of the National Academy of Sciences*, **107 (50)**, 21 277–21 282.

Seager, R., D. Neelin, I. Simpson, H. Liu, N. Henderson, T. Shaw, Y. Kushnir, M. Ting, and B. Cook, 2014: Dynamical and thermodynamical causes of large-scale changes in the hydrological cycle over north america in response to global warming. *Journal of Climate*.

Shan, S., and G. G. Wang, 2010: Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. *Structural Multidisciplinary Optimization*, **41 (2)**, 219–241.

Sheffield, J., A. P. Barrett, B. Colle, D. Nelun Fernando, R. Fu, K. L. Geil, Q. Hu, J. Kinter, S. Kumar, B. Langenbrunner, K. Lombardo, L. N. Long, E. Maloney, A. Mariotti, J. E. Meyerson, K. C. Mo, J. David Neelin, S. Nigam, Z. Pan, T. Ren, A. Ruiz-Barradas, Y. L. Serra, A. Seth, J. M. Thibeault, J. C. Stroeve, Z. Yang, and L. Yin, 2013a: North american climate in CMIP5 experiments. Part I: Evaluation of historical simulations of continental and regional climatology. *Journal of Climate*, **26 (23)**, 9209–9245.

Sheffield, J., S. J. Camargo, R. Fu, Q. Hu, X. Jiang, N. Johnson, K. B. Karnauskas, S. T. Kim, J. Kinter, S. Kumar, B. Langenbrunner, E. Maloney, A. Mariotti, J. E. Meyerson, J. D. Neelin, S. Nigam, Z. Pan, A. Ruiz-Barradas, R. Seager, Y. L. Serra, D.-Z. Sun, C. Wang, S.-P. Xie, J.-Y. Yu, T. Zhang, and M. Zhao, 2013b: North american climate in CMIP5 experiments. Part II: Evaluation of historical simulations of intraseasonal to decadal variability. *Journal of Climate*, **26 (23)**, 9247–9290.

Shepherd, T. G., 2014: Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geosci*, **7**, 703–708.

Simpson, I. R., T. A. Shaw, and R. Seager, 2014: A diagnosis of the seasonally and longitudinally varying midlatitude circulation response to global warming. *Journal of the Atmospheric Sciences*, **71 (7)**, 2489–2515.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to noaa's historical merged land–ocean surface temperature analysis (1880–2006). *Journal of Climate*, **21 (10)**, 2283–2296.

Spencer, H., and J. M. Slingo, 2003: The simulation of peak and delayed ENSO teleconnections. *Journal of Climate*, **16 (11)**, 1757–1774.

Straus, D. M., and J. Shukla, 1997: Variations of midlatitude transient dynamics associated with ENSO. *Journal of the Atmospheric Sciences*, **54 (7)**, 777–790.

Su, H., J. D. Neelin, and J. E. Meyerson, 2003: Sensitivity of tropical tropospheric temperature to sea surface temperature forcing*. *Journal of Climate*, **16 (9)**, 1283–1301.

Sun, D.-Z., Y. Yu, and T. Zhang, 2009: Tropical water vapor and cloud feedbacks in climate models: A further assessment using coupled simulations. *Journal of Climate*, **22 (5)**, 1287–1304.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, **106 (D7)**, 7183–7192.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of cmip5 and the experiment design. *Bulletin of the American Meteorological Society*, **93 (4)**, 485–498.

Tebaldi, C., J. M. Arblaster, and R. Knutti, 2011: Mapping model agreement on future climate projections. *Geophysical Research Letters*, **38 (23)**.

Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **365 (1857)**, 2053–2075.

Trenberth, K. E., 1997: The definition of El Niño. *Bulletin of the American Meteorological Society*, **78 (12)**, 2771–2777.

Trenberth, K. E., 2011: Changes in precipitation with climate change. *Climate Research*, **47 (1-2)**, 123–138.

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *Journal of Geophysical Research*, **103 (C7)**, 14 291–14 324.

Trenberth, K. E., and L. Smith, 2009: Variations in the three-dimensional structure of the atmospheric circulation with different flavors of El Niño. *Journal of Climate*, **22 (11)**, 2978–2991.

Ulbrich, U., J. G. Pinto, H. Kupfer, G. C. Leckebusch, T. Spangehl, and M. Reyers, 2008: Changing northern hemisphere storm tracks in an ensemble of ipcc climate change simulations. *Journal of Climate*, **21 (8)**, 1669–1679.

Van Oldenborgh, G. J., S. Y. Philip, and M. Collins, 2005: El Niõ in a changing climate: a multi-model study. *Ocean Science*, **1 (2)**, 81–95.

Vimont, D. J., J. M. Wallace, and D. S. Battisti, 2003: The seasonal footprinting mechanism in the Pacific: Implications for ENSO. *Journal of Climate*, **16 (16)**, 2668–2675.

von Storch, H., and F. W. Zwiers, 1999: *Statistical analysis in climate research*. Cambridge University Press.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly Weather Review*, **109 (4)**, 784–812.

Wallace, J. M., E. M. Rasmusson, T. P. Mitchell, V. E. Kousky, E. S. Sarachik, and H. von Storch, 1998: On the structure and evoluation of ENSO-related climate variability in the tropical Pacific: Lessons from TOGA. *Journal of Geophysical Research*, **103 (C7)**, 14 241–14 259.

Wallace, J. M., C. Smith, and C. S. Bretherton, 1992: Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *Journal of Climate*, **5 (6)**, 561–576.

Wang, G. G., and S. Shan, 2006: Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical Design*, **129 (4)**, 370–380.

Weare, B., 2013: El niño teleconnections in cmip5 models. *Climate Dynamics*, **41 (7-8)**, 2165–2177.

Wettstein, J. J., and J. M. Wallace, 2010: Observed patterns of month-to-month storm-track variability and their relationship to the background flow*. *Journal of the Atmospheric Sciences*, **67 (5)**, 1420–1437.

Whitaker, J. S., and K. M. Weickmann, 2001: Subseasonal variations of tropical convection and week-2 prediction of wintertime western north american rainfall. *Journal of Climate*, **14 (15)**, 3279–3288.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*, International geophysics series, Vol. 100. 3rd ed., Academic Press.

Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bulletin of the American Meteorological Society*, **78 (11)**, 2539–2558.

Xue, Y., T. M. Smith, and R. W. Reynolds, 2003: Interdecadal changes of 30-yr sst normals during 1871–2000. *Journal of Climate*, **16 (10)**, 1601–1612.

Yin, J. H., 2005: A consistent poleward shift of the storm tracks in simulations of 21st century climate. *Geophysical Research Letters*, **32 (18)**, L18 701.

Yin, L., R. Fu, E. Shevliakova, and R. Dickinson, 2013: How well can cmip5 simulate precipitation and its controlling processes over tropical south america? *Climate Dynamics*, **41 (11-12)**, 3127–3142.

Zhang, C., 2001: Double ITCZs. *Journal of Geophysical Research: Atmospheres*, **106 (D11)**, 11 785–11 792.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: Enso-like interdecadal variability: 1900–93. *Journal of Climateou*, **10 (5)**, 1004–1020.

Zheng, M., E. K. M. Chang, and B. Colle, 2013: Ensemble sensitivity tools for assessing extraopical cyclong intensity and track predictability. *Weather and Forecasting*, **1133-1156**.