Observations and Simulations of the Western United States' Hydroclimate

by

Kristen Guirguis

Department of Civil and Environmental Engineering
Duke University

Date:_____

Approved:

_____
Dr. Roni Avissar, Supervisor

_____
Dr. Ana P. Barros

_____
Dr. Amilcare Porporato

_____
Dr. Sandra Yuter

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Civil and Environmental Engineering in the Graduate School
of Duke University

2009

ABSTRACT

Observations and Simulations of the Western United States' Hydroclimate

by

Kristen Guirguis

Department of Civil and Environmental Engineering
Duke University

Date:_____

Approved:

_____
Dr. Roni Avissar, Supervisor

_____
Dr. Ana P. Barros

_____
Dr. Amilcare Porporato

_____
Dr. Sandra Yuter

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Philosophy in the Department of
Civil and Environmental Engineering in the Graduate School
of Duke University

2009

Abstract

While very important from an economical and societal point of view, estimating precipitation in the western United States remains an unsolved and challenging problem. This is due to difficulties in observing and modeling precipitation in complex terrain. This research examines this issue by (i) providing a systematic evaluation of precipitation observations to quantify data uncertainty and (ii) investigating the ability of the Ocean-Land-Atmosphere Model (OLAM) to simulate the winter hydroclimate in this region. This state-of-the-art, non-hydrostatic model has the capability of simulating simultaneously all scales of motions at various resolutions.

This research intercompares nine precipitation datasets commonly used in hydrometeorological research in two ways. First, using principal component analysis, a precipitation climatology is conducted for the western U.S. from which five unique precipitation climates are identified. From this analysis, data uncertainty is shown to be primarily due to differences in (i) precipitation over the Rocky Mountains, (ii) the eastward wet-to-dry precipitation gradient during the cold season, (iii) the North American Monsoon signal, and (iv) precipitation in the desert southwest during spring and summer. The second intercomparison uses these five precipitation regions to provide location-specific assessments of uncertainty, which is shown to be dependent on season, location.

Long-range weather forecasts on the order of a season are important for water-scarce regions such as the western U.S. The modeling component of this research looks at the ability of the OLAM to simulate the hydroclimate in the western U.S. during the winter of 1999. Six global simulations are run, each with a different spatial resolution

over the western U.S. (360 km down to 11 km). For this study, OLAM is configured as for a long-range seasonal hindcast but with observed sea surface temperatures. OLAM precipitation compares well against observations, and is generally within the range of data uncertainty. Observed and simulated synoptic meteorological conditions are examined during the wettest and driest events. OLAM is shown to reproduce the appropriate anomaly fields, which is encouraging since it demonstrates the capability of a global climate model, driven only by SSTs and initial conditions, to represent meteorological features associated with daily precipitation variability.

Contents

List of Tables

List of Figures

Acknowledgements

I thank my advisor, Roni Avissar for all his support, guidance, and encouragement throughout this long doctoral process. I especially thank him for his understanding of my need to work away from the office so often, as I feel few advisors would have been as accommodating. I also greatly appreciate the support and advice I received from the members of my graduate committee: Drs Ana Barros, Amilcare Porporato, and Sandra Yuter.

I am thankful to Robert Walko for his work on OLAM and for those hours spent with me on issues related to my research specifically. I am enormously thankful to Dr. Martin Otte who provided an endless amount of help to me throughout my time at Duke. He is an invaluable part of Dr. Avissar's group and I benefited greatly from working with him. I thank Dr. David Medvigy for his work on OLAM and for the assistance he provided to me. I would also like to thank current and past students of the Avissar research group for the encouragement, assistance, and advice I received over the years. I wish them all the best in the future.

This research was funded by the National Oceanic and Atmospheric Administration (NOAA) under Grants NA04OAR4310078 and NA050AR4310014. Additional support was provided by the National Aeronautics and Space Administration (NASA) Headquarters under the Earth System Science Fellowship Grant NNG04GQ60H.

CHAPTER 1

Introduction

1.1 Background

The arid to semi-arid western United States is increasingly feeling the strain of population growth on its limited and variable water supply. This situation is unlikely to be resolved in the coming decades given current projections of population growth. The region's surface water resources derived from snowmelt are already allocated annually among users with little if any left unused, and ground water use in the region far exceeds aquifer recharge rates. Interannual climate variability has the potential to significantly alter the amount of water stored as snow in the Western Cordillera, which ultimately affects river flow and consequently impacts the region's surface water resources. The frequency and occurrence of extreme precipitation events vary markedly from year to year; and reliable forecasts of the spatial and temporal features of such storms remain elusive. Need notwithstanding, prediction skill in the western U.S. diminishes very quickly at spatial scales of interest for flash flood forecasting and at temporal scales of interest for water management. Therefore, improved hydrometeorological predictability on all scales is urgently needed.

The western U.S. hydroclimate is strongly influenced by interannual climate variability such as the El Niño-Southern Oscillation (ENSO). Generally, warm (cold)

ENSO events are associated with wet (dry) conditions in the southwest U.S., while the opposite relationship is observed for the Pacific Northwest (Dracup and Kahya 1994, Cayan 1996, Cayan et al. 1999). Climate patterns developing over longer time scales modulate ENSO and its associated predictability (Mantua et al. 1997, Trenberth 1990). Accordingly, an understanding of slowly-varying climate patterns and their interactions provide predictability to the hydrometeorological system (Gershunov and Cayan 2003). However, better predictions of the western U.S. hydroclimate are also dependent on an improved understanding and numerical representation of the interactions between climate variability and land-surface processes such as orography, soil, and vegetation.

There is substantial evidence to suggest that anomalies in soil and vegetation fields play a strong role in modulating precipitation variability in the western United States. For example, sensible and latent heat flux associated with evaporation from wet soil can affect the stability of the boundary layer, thereby enhancing convection and the occurrence of precipitation (Barros and Hwu, 2002). Studies have shown a positive feedback to exist between antecedent soil moisture conditions and summer rainfall (e.g. D'Odorico and Porporato, 2004), as well as between winter snow cover and warm season precipitation in the southwestern U.S. (Lo and Clark, 2002; Gutzler, 2000; Gutzler and Preston, 1997). Since soils and vegetation exhibit a long memory due to their capacity to store/extract moisture accumulated during previous seasons, their improved representation in numerical models could lend predictability to the hydrometeorological system and bring marked improvements in both accuracy and lead time.

One of the major obstacles to predictability in the West has been our incomplete understanding of the orographic impact of the Western Cordillera on precipitation in the region. The hydroclimate in the western U.S. demonstrates a strong dependence on thermal and dynamical, local and remote orographic processes (e.g. Smith, 1997). Mechanical forcing by orography regulates the transport of marine moisture inland and is instrumental in driving mesoscale reversals in circulation patterns that mark the transition of seasons (Smith et al., 1997; Broccoli and Manabe, 1992). Thermal effects include the role of orography in establishing the sensible heat distribution of land and subsequently the land-sea temperature gradient that drives warm season precipitation in the southwest.

Less well understood but equally important are the effects of land surface heterogeneities. Mountainous terrain exhibits steep gradients in topography, vegetation, soil moisture, and snow cover such that land surface albedo and roughness vary considerably over close spatial distances and can impact the atmospheric state at the surface level. Such land-surface heterogeneities can affect the structure of the boundary layer and consequently the vertical transport of heat and moisture on scales of a few kilometers or less, while also affecting larger scales through the occurrence and maintenance of organized mesoscale circulations (Giorgi and Avissar, 1997). The occurrence supercooled water pockets, which can enhance precipitation through the "seeder-feeder" mechanism (Barros and Lettenmaier, 1994), have been shown to occur on scales on the order of 1 km (Sassen et al., 1990). Therefore, orography has the ability to transfer large-scale weather phenomena down to a local response and vice versa. A key issue for modeling precipitation in orographic regions is that a range of spatial scales must be simultaneously considered, which ultimately requires cost-benefit consideration

of model complexity, spatial resolution, and computational resources (Barros and Lettenmaier 1994, Barros 1995).

1.2 Estimating Precipitation in the Western U.S.

Estimating precipitation in the western U.S. remains an unsolved and challenging problem. This is partly due to weakness in model parameterizations in their ability to approximate complex land-atmosphere dynamics in complex terrain. Another factor contributing to lack of forecast skill in the region is uncertainty in observations. Model evaluation and diagnostics require the use of systematic and high quality observations, which are difficult to obtain in mountainous regions. If a high level of observational data uncertainty exists, then it is possible that the choice of dataset for model evaluation could affect conclusions regarding model skill. The goal of this dissertation is to investigate the hydrometeorlogy/hydroclimatology of the western United States using observations and model simulations. For the observational data analysis, the goal is to provide a systematic evaluation of observation-based precipitation datasets to quantify data uncertainty in the region. The modeling portion of the study assesses the ability of a newly developed weather and climate model, the Ocean-Land-Atmosphere Model (OLAM), to simulate the hydroclimatology of the western U.S.

1.3 Observational Data Analysis

There are several sources of precipitation data available in the western United States including estimates from rain gauges, ground radar, satellite, and reanalysis. These data sources have their own strengths and weaknesses and it is difficult to know which

Table 1.1: Main characteristics of the datasets included in this study.

Dataset	Spatial resolution	Temporal domain	Data source
$GPCC_a$	2.5° x 2.5°	From 1986	Rain gauge
$GPCP_b$	2.5° x 2.5°	From 1979	Rain gauge, satellite
$CMAP_c$	2.5° x 2.5°	From 1979	Rain gauge, satellite
$NCEP2_d$	210 x 210 km^2	From 1979	Reanalysis
$GMFD_e$	1° x 1°	1948–2000	Reanalysis, rain gauge, satellite
$USMex_f$	1° x 1°	From 1948	Rain gauge
$NARR_g$	32 x 32 km^2	From 1979	Reanalysis
VIC_h	1/8° x 1/8°	1950–2000	Rain gauge with orographic adjustment
$PRISM_i$	4 x 4 km^2	From 1890	Rain gauge with orographic adjustment

[a]Global Precipitation Climatology Center monitoring product
[b]Global Precipitation Climatology Project Combined Precipitation Dataset, version 2
[c]Climate Prediction Center Merged Analysis of Precipitation
[d]National Centers for Environmental Prediction–Department of Energy Reanalysis 2
[e]Global Meteorological Forcing Dataset for land surface modeling
[f]Climate Prediction Center retrospective United States and Mexico daily precipitation analysis
[g]North American Regional Reanalysis
[h]Variable Infiltration Capacity Retrospective Land Surface Dataset
[i]Parameter-elevation Regressions on Independent Slopes Model

would be most appropriate for evaluating a numerical model. Each data product contains error that is space and time variant, and it is difficult to know a priori which data product is most reliable and best suited for model evaluation, and the choice of dataset for model evaluation could affect conclusions regarding model skill.

Data reliability depends on issues such as latitude, topography, and seasonality. The sampling error associated with rain gauge data, for example, can become large over high elevations where gauge coverage is sparse, and rain gauge data are additionally known to suffer low bias due to wind and evaporation (e.g. Legates and Willmott 1990). Ground Doppler radar estimates are available at a relatively high spatial/temporal resolution over much of the central and eastern U.S., but mountainous terrain interferes with low-level radar scans making Doppler precipitation estimates in the western U.S.

problematic (Serafin and Wilson 2000, Maddox et al. 2002). Error in radar estimates also occurs from brightband effects and algorithm misinterpretation of ground clutter (e.g. Klazura et al. 1999). Satellite data are available over mountains but precipitation estimates become less reliable poleward of approximately 40° where geostationary satellite measurements are unavailable and precipitation estimates rely on polar-orbiting satellites, which have a poor temporal sampling rate. Additionally, IR-based precipitation estimates from geostationary satellite platforms are based on cloud-top brightness temperatures, which have only a limited empirical relationship with surface precipitation (Arkin and Meisner 1987, Arkin and Xie 1994). Merged precipitation products use information from various sources including rain gauges and satellites as a means of rectifying latitudinal variations in sampling error and to correct for bias present in the individual data sources (Huffman et al. 1997, Xie and Arkin 1997, Adler et al. 2003). However, sampling error remains an issue over mountainous regions as these data products use rain gauge data as a major data component. Global and regional reanalysis data are also available for the western U.S. These precipitation estimates are heavily influenced by model parameterizations (Kalnay et al. 1996, Kistler et al. 2001, Kanamitsu et al. 2002, Mesinger et al. 2006), which do not perform well for all locations and for all weather regimes.

This analysis of observations is conducted as two parts. The first uses principal component-based regionalization methods to isolate the unique precipitation climates in the western United States, and the second part focuses on long-term averages, interannual variability, statistical moments, distributions, time series, and persistence. In each case, comparative analyses are conducted using the bulk of the observation-based precipitation

data products that are available for moderately long-term (15 yr) climatology studies of the western United States. Specifically, these are the (i) Global Precipitation Climatology Center (GPCC) monitoring product, (ii) Global Precipitation Climatology Project (GPCP) version 2, (iii) Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP), (iv) NCEP-DOE Reanalysis-2, (v) North American Regional Reanalysis (NARR), (vi) CPC Retrospective United States and Mexico Daily Precipitation Analysis (USMex), (vii) Global Meteorological Forcing Dataset for Land Surface Modeling (GMFD), (viii) Variable Infiltration Capacity (VIC) Retrospective Land Surface Dataset, and (ix) Parameter-elevation Regressions on Independent Slopes Model (PRISM). These nine datasets and their spatial/temporal characteristics are shown in Table 1.1. Chapter 3 provides a detailed description of each dataset.

1.4 Model Simulations

Long-range weather forecasts on the order of a season are important for resource and response planning, particularly for water-scarce regions such as the western U.S. Very little skill of instantaneous atmospheric fields exists for forecasts beyond 10-14 days. However, numerical climate models can be used with some skill for predicting seasonal anomalies (Barnston 1999, Livezey 1990, Goddard 2001). The skill is from surface-atmosphere relationships where slowly evolving fields such as ocean temperatures and land surface conditions strongly affect the atmospheric state. Surface conditions are provided initially by observations and can be updated during the integration period using coupled ocean and land surface models. These types of long-range forecasts have used both General Circulation Models (GCMs) and Regional

Climate Models (RCMs). GCMs have the advantage of their global domain, which does not require input at the lateral boundary. RCMs are forced at their lateral boundaries by GCM forecasts, which produce a one-way interaction and does not allow for any internal effects to propagate out to a global scale and back. The advantage of using RCMs is that they are able to represent the land surface and land-atmosphere interactions at a higher spatial resolution, which has been shown to improve seasonal forecast skill (Giorgi et al. 1994, Hong and Leetma 1999, Leung et al. 1999, Cocke and Larrow 2000, Fenessy and Shukla 2000, Hui et al. 2002, Hui and Shum 2005; Ramos da Silva and Avissar 2006, Ramos da Silva et al. 2008, Medvigy et al. 2008a).

This research investigates the ability of a newly developed weather and climate model to simulate the hydroclimatology of the western U.S. and to provide new insights into this complex issue. The Ocean-Land-Atmosphere Model (Walko and Avissar 2008a, 2008b) is a global, non-hydrostatic model with a refinement option for resolving a region (or regions) of interest at a higher spatial grid size within a global domain. Thus all interactions between gridcells having different spatial scales is two-way interactive, and there is no need for lateral boundary forcing. This solves some issues that are potentially limiting to regional climate prediction. Namely, (i) in a nested GCM-RCM configuration, different parameterizations used in the different models can result in nonmeteorological gradients near the lateral boundary and (ii) when an RCM is forced at its lateral boundary by a GCM, there is limited interaction between wavelengths of different scales because the higher resolution RCM solution cannot feed back to the large scales (Kumar et al. 1997). The issue of scale interaction is particularly important for representing and understanding meteorological teleconnections whereby local processes

in one part of the globe affect weather and climate in another (Werth and Avissar 2002, Avissar et al. 2002, Avissar and Werth 2005).

A series of simulations are run over the winter season of 1999 (1 December-28 February) with successive increases in the horizontal grid size over the western U.S. The winter of 1999 occurred during the cold phase of the El Nino Southern Oscillation (ENSO) with wet conditions observed in the Pacific Northwest including flooding events (NOAA, 1999). For these simulations, OLAM is configured with a global grid having a horizontal characteristic length scale (CLS) of 360 km, which is the square root of the gridcell area. For tests of sensitivity to horizontal grid size, additional grid refinement levels are configured over the western U.S. Five additional simulations are run, which have a CLS over the western U.S. of 180, 90, 45, 23, and 11 km. The model is initialized on 21 November and evaluated over 1 December – 28 February. OLAM is not nudged with observations during the simulations, so the evaluation is based on statistical comparisons with observations without comparing specifically timed events. The unique precipitation climates identified during the observational data analysis are used to localize the analysis of simulation results. The results of the observational data analysis are also used to select verification datasets used for the modeling study.

1.5 Dissertation Overview

The outline for the dissertation is as follows. In Chapter 2 a description of OLAM is provided. This includes a description of the dynamic core, grid structure, and subgrid-scale parameterizations. In Chapters 3 and 4 studies of observational data uncertainty are presented where nine state-of-the-art precipitation datasets are compared in order to

quantify data uncertainty. This information is used to localize assessments of simulation skill and to select data sources for use in the evaluation of OLAM. Chapter 5 is a numerical investigation of OLAM's ability to simulate the hydroclimatology of the western U.S. during the 1999 winter season. A summary and general conclusions are provided in Chapter 6.

CHAPTER 2

The Ocean-Land-Atmosphere Model (OLAM)

2.1 Introduction

The Ocean-Land-Atmosphere Model is a new generation of Earth System models, which has capabilities including an unstructured grid with a refinement option for resolving geographic regions of interest within a global domain. It was developed to overcome some important shortcomings of both General Circulation Models and Regional Climate Models. For example, in a global model, any effects of land surface processes are allowed to propagate out to a global scale and back, whereas a limited area model is forced at its boundary by observations or GCM output. Thus the global model allows for a more complete understanding of the totality of land-surface effects including any teleconnections that may occur (Werth and Avissar 2002, Avissar and Werth 2005). The advantage of RCMs is that they are able to represent the land surface and land-atmosphere interactions at a higher spatial resolution, which has been shown to improve forecast skill (Giorgi et al. 1994, Hong and Leetma 1999, Leung et al. 1999, Cocke and Larrow 2000, Fenessy and Shukla 2000, Hui et al. 2002, Hui and Shum 2005; Ramos da Silva and Avissar 2006, Ramos da Silva et al. 2008, Medvigy et al. 2008a). OLAM offers the advantage of high spatial resolution locally within a global simulation.

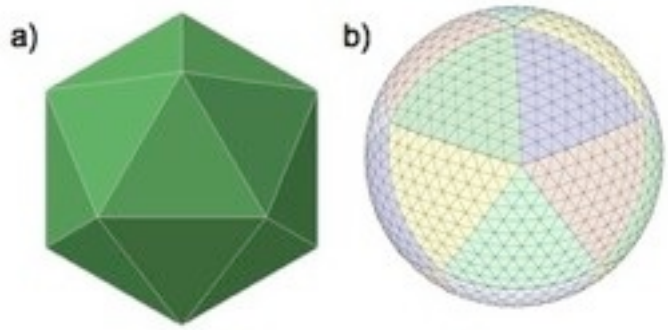

Figure 2.1: (a) The icosahedron and (b) OLAM grid triangulation.

The OLAM shares many parameterizations of the Regional Atmospheric Modeling System (RAMS) developed by Pielke et al. (1992) including those for microphysics, sub-grid cumulus convection, radiative transfer, and land and vegetation water and energy balances. However, its dynamic core is entirely new. The development of OLAM uses the approach of adapting a regional model to a global domain, so its parameterizations have been well tested in regional models RAMS and MM5 for applications on high-resolution grids. OLAM's new dynamical core solves the full compressible Navier-Stokes equations (i.e. non-hydrostatic with no Boussinesq approximation), which are spatially discretized using the finite volume method. OLAM's unstructured grid is seamless in the sense that all communication between gridcells having different resolutions is done by conservation of advective and diffusive transport, with no need for nesting algorithms (Walko and Avissar 2008b). Topography is represented using a volume-fraction method (Adcroft et al., 1997) in which model levels are strictly horizontal rather than terrain following and therefore intersect topography, which provides for superior surface flux representation in regions of elevated terrain.

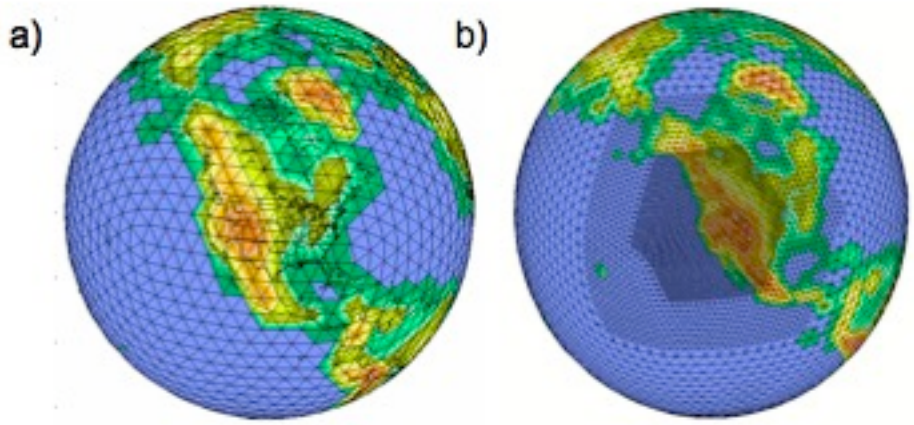

Figure 2.2: OLAM grid configured with (a) a uniform global horizontal resolution, and (b) with increased horizontal resolution over the western United States. The field plotted is topography.

2.2 Grid Structure

The grid structure implemented in OLAM to discretize atmospheric processes is a triangulation method based on the icosahedron, where the icosahedron (Fig. 2.1a) is a polyhedron with 20 equilateral triangle faces, 30 edges, and 12 vertices. In OLAM, each of the 20 original triangular faces is further divided into quasi-equilateral triangles (Fig. 2.1b), where the number of divisions is dependent on the desired spatial resolution. The underlying rule of OLAM's triangulation is that each triangle must share edges with exactly three other triangles. The use of the triangular grid aims to solve some important issues associated with the more commonly encountered coordinate systems, namely spherical and polar stereographic coordinates. The spherical coordinate system, which discretizes the atmosphere as a series of regularly spaced rectangles, is relatively easy to implement and makes intuitive sense, however it suffers from some important problems. These are (i) the grid cells do not take the same shape over the globe due to their closer spacing at the poles relative to the equator and such closely spaced grid cells can cause

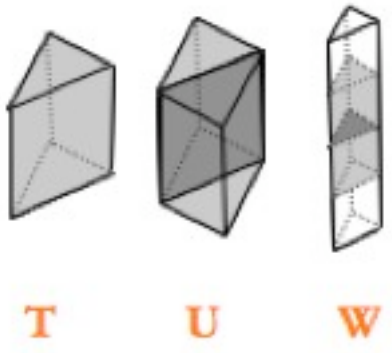

Figure 2.3: Illustration of OLAM control volumes.

numerical instability, which must be remedied with either a very small timestep, which might be computationally unfeasible, or by the application of filtering methods in order to adequately represent gravity waves at longer timesteps and (ii) the poles exhibit properties of a numerical singularity where, as air flows across a pole point, velocity abruptly changes direction and can cause instability. The polar stereographic coordinate system, such as that implemented in RAMS, discretizes the atmosphere by projecting the spherical shape of the simulation domain onto to a plane, which is divided into a specified number of subvolumes. This type of configuration would be inappropriate for a global model such as OLAM, as the projected plane would become prohibitively large. OLAM's unstructured grid gives the model "telescoping" capabilities that allow for high resolution over a region of interest without the need for nesting algorithms (Fig. 2.2). Therefore, direct communication occurs between grid cells regardless of spatial scale.

OLAM solves the non-hydrostatic, full compressible, finite volume analog of the Navier-Stokes equations and exactly conserves mass, momentum, and internal energy. The Arakawa-C grid stagger is implemented in OLAM due to its superior coupling of mass and momentum fields (Walko and Avissar, 2008). Scalar variables are prognosed

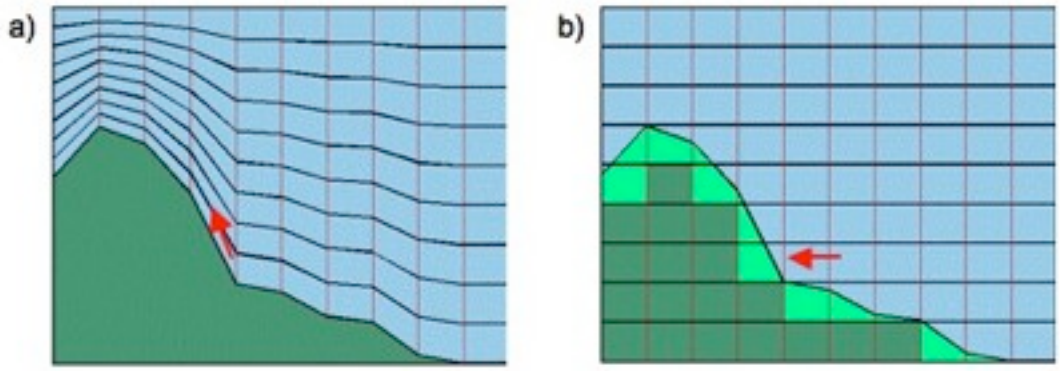

Figure 2.4: (a) Terrain-following and (b) volume-fraction coordinate system.

at "T-points" and represent a volume average centered within the volume triangle, or prism, while the momentum flux is calculated at "U-points" or "W-points" occurring at each prism face and which is strictly horizontal or vertical (Fig. 2.3).

2.3 Topography

Topography is represented in OLAM by the volume-fraction or shaved grid cell method (Adcroft et al., 1997; Marshall et al., 1997). In contrast to the more familiar terrain-following coordinate system (Gal-Chen and Somerville, 1975), model levels are strictly horizontal and therefore intersect topography. The degree of blockage by topography for a given grid cell is accounted for by reducing its face area, which appears explicitly in the finite volume equations (Section 2.4). This method resolves some important shortcomings of the terrain-following coordinate system. Namely, these are the coupling of horizontal flux with vertical gradients and the forced narrowing of atmospheric columns over hills and mountains (Fig. 2.4). This method has been shown superior at representing horizontal gradients of atmospheric fields near the land surface over topographically complex regions (e.g. Avissar and Mahrer 1998).

2.4 Governing Equations

The governing equations in OLAM are the Navier-Stokes equations, which are derived from the fundamental laws of physics and include the conservation of momentum, mass, heat, moisture and scalar fields, as well as the equation of state. These are represented, respectively, as follows

$$\frac{\partial V_i}{\partial t} = -\nabla \bullet (v_i \vec{V}) - (\nabla p)_i - \rho g_i - \left(2\rho\vec{\Omega} \times \vec{v}\right)_i + F_i \tag{2.1}$$

$$\frac{\partial \rho}{\partial t} = -\nabla \bullet \vec{V} \tag{2.2}$$

$$\frac{\partial (\rho\Theta)}{\partial t} = -\nabla \bullet (\Theta\overrightarrow{V}) + H \tag{2.3}$$

$$\frac{\partial (\rho q_n)}{\partial t} = -\nabla \cdot (q\vec{V}) + Q_n \tag{2.4}$$

$$\frac{\partial (\rho\chi_m)}{\partial t} = -\nabla \cdot (\chi\vec{V}) + S_m \tag{2.5}$$

$$p = \left[(\rho_d R_d + \rho_v R_v)\theta\right]^{\frac{C_p}{C_v}} \left(\frac{1}{p_0}\right)^{\frac{R_d}{C_v}}. \tag{2.6}$$

Here ρ is air density, v is velocity, V≡ ρv is momentum, p is pressure, θ is potential temperature, Θ is ice-liquid potential temperature, g and Ω respectively represent the Earth's gravity and angular momentum, q represents moisture (with n=1,2 or 3 for the three phase states of water), χ represents scalar fields, ρ_d and R_d are the density and the gas constant for dry air, ρ_v and R_v are the density and gas constant for water vapor, C_p and C_v are the specific heats of dry air at constant pressure and volume, and $\vec{k}$ is the unit vector in the vertical direction. Sources and sinks of momentum, heat,

moisture and scalar fields are represented by the forcing terms F, H, Q, and S, respectively.

Equations (2.1)–(2.5) are discretized by representing the atmosphere as a finite 3-dimensional grid and then solving these discretized equations for each control volume (cf. Fig. 2.3) at each timestep. Specifically, Eqns. (2.1)–(2.5) are solved by integrating over all control volumes Ψ, which are bounded by surface areas σ. The application of the Gauss divergence theorem, which transforms the control volume integrals into surface integrals, then allows changes in a volume quantity to be represented as a surface flux, as given by

$$\int_{\Psi} \nabla \cdot \vec{\Phi} d\Psi = \oint_{\sigma} \vec{\Phi} \cdot d\vec{\sigma}, \tag{2.7}$$

where $d\sigma$ is the surface element and $d\vec{\sigma} = \hat{n} d\sigma$, and where $\hat{n}$ is the unit vector whose direction is outward normal from the surface element $d\sigma$. The governing equations can then be expressed in integral form as

$$\frac{\partial}{\partial t}\int V_t d\Psi = -\oint\left(v_t \vec{V}\right)\cdot d\vec{\sigma} - \int \frac{\partial p}{\partial x_i} d\Psi - \left(2\rho\vec{\Omega}\times\vec{v}\right)_i d\Psi + \int \rho g_t d\Psi + \int F_i d\Psi \tag{2.8}$$

$$\frac{\partial}{\partial t}\int \rho d\Psi = -\oint\left(\vec{V}\right)\cdot d\vec{\sigma} \tag{2.9}$$

$$\frac{\partial}{\partial t}\int (\rho\Theta) d\Psi = -\oint\left(\Theta\vec{V}\right)\cdot d\vec{\sigma} + \int H d\Psi \tag{2.10}$$

$$\frac{\partial}{\partial t}\int \rho q_n d\Psi = -\oint\left(q_n\vec{V}\right)\cdot d\vec{\sigma} + \int Q_n d\Psi \tag{2.11}$$

$$\frac{\partial}{\partial t}\int \chi_m d\Psi = -\oint\left(\chi_m\vec{V}\right)\cdot d\vec{\sigma} + \int S_m d\Psi. \tag{2.12}$$

Each volume integral in (2.8)-(2.12) is discretized in OLAM by its representation as the grid cell average, as given by

$$\int \phi d\Psi = \bar{\phi}\Psi, \tag{2.13}$$

where ϕ is any dependent variable and $\bar{\phi}$ is its mean value over the subvolume. Each surface integral is discretized by its representation as the sum over each face area σ bounding the control volume Ψ. The surface integrals are additionally partitioned into a resolved and subgrid scale (SGS), such that (2.8) becomes

$$\oint\left(\phi\vec{V}\right)\cdot d\vec{\sigma} = \sum\left[\int\left(\phi_j\vec{V}_j\right)\cdot d\vec{\sigma}_j\right] = \sum\left[\left(\bar{\phi}\bar{V}_j + SGS\left\{\phi_j V_j\right\}\right)\sigma_j\right], \tag{2.14}$$

and comparable equations are used for (2.9)-(2.12). Here j represents the subvolume face being operated on. The fundamental equations in their discretized form and illustrating the resolved and subgrid components are given by

$$\frac{\partial\bar{V}_i}{\partial t}\Psi = -\sum_j\left[\left(\bar{v}_{ij}\bar{V}_j + SGS\left\{v_{ij},V_j\right\}\right)\sigma_j\right] - \frac{\partial\bar{p}}{\partial x_i}\sigma_i\Delta x_i - \left(2\bar{\rho}\bar{\vec{\Omega}}\times\bar{\vec{v}}\right)_i\Psi + \bar{\rho}g_i\Psi + \bar{F}\Psi \tag{2.15}$$

$$\frac{\partial\bar{\rho}}{\partial t}\Psi = -\sum_j\left[\bar{V}_j\sigma_j\right] \tag{2.16}$$

$$\frac{\partial\overline{\rho\Theta}}{\partial t}\Psi = -\sum_j\left[\left(\bar{\Theta}_j\bar{V}_j + SGS\left\{\Theta_j,V_j\right\}\right)\sigma_j\right] + \bar{H}\Psi \tag{2.17}$$

$$\frac{\partial\overline{\rho_q}}{\partial t}\Psi = -\sum_j\left[\left(\bar{q}_j\bar{V}_j + SGS\left\{q_j,V_j\right\}\right)\sigma_j\right] + \bar{Q}\Psi \tag{2.18}$$

$$\frac{\partial\overline{\rho\chi}}{\partial t}\Psi = -\sum_j\left[\left(\bar{\chi}_j\bar{V}_j + SGS\left\{\chi_j,V_j\right\}\right)\sigma_j\right] + \bar{S}\Psi. \tag{2.19}$$

The advective fluxes in (2.15)-(2.18) ($\overline{b\bar{V}_j}$, where b=v_{ij}, Θ_j,q_j, or χ_j) are calculated at each face of the control volume with the exception of vertical momentum, which must be interpolated to the upper and lower faces.

2.5 Physical Parameterizations

The purpose of physical parameterizations is to approximate the combined effects of processes occurring on the subgrid-scale in order to prognose their impact on the resolved scale. These include a bulk parameterization for moist microphysics, a model for land surface processes, and parameterizations for radiative transfer, and cumulus convection.

a. Bulk Microphysics

The OLAM represents subgrid-scale microphysics using the bulk microphysical parameterization scheme of Walko et al. (1995). The scheme distinguishes between 7 hydrometeor classes: cloud water, rain, pristine ice, snow, aggregates, graupel, and hail, which differ according to density, shape, size, liquid versus ice content, fall speeds, and radiative properties. Ice and snow are allowed to assume 5 different habits (shapes), and the user can control the complexity of the scheme by specifying which hydrometeor types can occur. Hydrometeor distributions are assumed to have the form of the gamma distribution

$$n(D) = \frac{N_t}{\Gamma(\nu)}\left(-\frac{D}{D_n}\right)^{\nu-1}\frac{1}{D_n}\exp\left(-\frac{D}{D_n}\right), \tag{2.20}$$

where N_t is the number concentration, ν is the shape parameter, Γ is the complete gamma function of ν, D is the diameter of the hydrometer, and D_n is the characteristic diameter of the distribution. Hydrometeor number density and mass mixing ratio are respectively represented by

$$n(D) = N_t f_{gam}(D) \tag{2.21}$$

and

$$r = \frac{N_t}{\rho_a} \alpha_m D_n^{\beta m} \frac{\Gamma(v + \beta_m)}{\Gamma(v)}, \tag{2.22}$$

where α and β are empirical parameters and N_t is the total concentration in each category. The physical processes represented by the bulk microphysics scheme include cloud droplet and ice nucleation, liquid and ice diffusional growth, evaporation and sublimation, freezing and melting, shedding, sedimentation, collisions between hydrometeors, and secondary ice production.

b. Cumulus Convection

In general, a cumulus convection parameterization scheme is a numerical algorithm that uses large-scale variables such as temperature, moisture, and momentum to determine if and how cumulus convection occurs at a given time step. There are many different parameterization schemes available that vary by their complexity, their assumptions, and the conditions under which they perform optimally. The simulations presented in this work use the Grell-Dévényi scheme (Grell et al., 1993; Grell and Dévényi, 2002), which is described here.

The Grell-Dévényi (GD) scheme makes use of the variety of dynamic closure assumptions employed in the different formulations of cumulus parameterizations used in numerical models. These closures are used to obtain an approximation for cloud base mass flux (m_b). The first dynamic closure assumption is based on Arakawa and Schubert (AS, 1974) and uses a stability equilibrium requirement, as measured by the cloud work function, between the large scale forcing and the cloud response, where the cloud work function is an integrated measure of the difference between moist static energy in a cloud

and that of the environment. The second dynamic closure is also an AS type similar to the first except the climatological value of the cloud work function is incorporated in its calculations. The third dynamic closure assumption is based on Kain and Fritsch (1992), which is a stability adjustment scheme in which stability is simply removed from the atmospheric column following convection. The fourth dynamic closure is based on moisture convergence following Krishnamurti et al. (1983). Finally, the fifth dynamic closure assumption is similar to Brown (1979) and Frank and Cohen (1987), which assumes the cloud base mass flux is equal to the environmental mass flux at a lower atmospheric level. Sensitivity of the closure assumptions to parameters within the model is considered by either perturbing the cloud base mass flux about its solution or by perturbing sensitive parameters prior to calculating the cloud base mass flux. The dynamic closures result in an ensemble of 16 solutions for cloud base mass flux.

In addition to dynamic control closures, an ensemble approach is also employed for static control closure assumptions, where static control is terminology for the cloud model used to approximate cloud properties. These include approximations for precipitation efficiency and the maximum depth of the capping inversion, where a cap is a thermal inversion that can suppress convection. Uncertainty in these parameters can affect results of simulated precipitation. Using the GD scheme, these variables are allowed to vary within a bounded range, producing a 3-variable ensemble of solutions for the precipitation efficiency and maximum cap height. The cap height maximum is perturbed around the empirical parameter of 100 mb. The precipitation efficiency ($1-\beta$) is approximated in OLAM based on vertical wind shear according to Fritsch and Chappel (1980) using the empirical function

$$(1-\beta) = 1.591 - 0.639\frac{\Delta V}{\Delta z} + 0.0953\left(\frac{\Delta V}{\Delta z}\right)^2 - 0.00496\left(\frac{\Delta V}{\Delta z}\right)^3,$$

and (2.23)

$$(1-\beta) = 0.9 \quad for \quad \frac{\Delta V}{\Delta z} < 1.35.$$

The precipitation efficiency is then perturbed about this value. Here V is wind velocity, z is height, and β is the fraction of resolvable vertical vapor flux that increases the moisture of the column (Kuo 1974).

The dynamic and static control closures are allowed to interact with each other, thereby producing a total of 16*3*3=144 ensemble members for the variable m_b. In the GD scheme, an average of the ensemble solutions is provided as feedback to the model. Convective precipitation is approximated as a function of cloud base mass flux, precipitation efficiency, and cloud condensate.

c. Radiative Transfer

OLAM uses the longwave and shortwave radiation schemes of Harrington (1997). These schemes use a two-stream model with three solar (0.245-0.7 μm, 0.7-1.53 μm, 1.53-4.64 μm) and five infrared bands (4.64-8.33 μm, 8.33-9.01 μm, 9.01-X μm, X-20 μm, 20-104 μm). The schemes are coupled to the bulk microphysical parameterization (Section 2.5a). The two-stream model requires three cloud optical parameters for the various hydrometeors: the extinction coefficient (Q_{ext}), single-scatter albedo (ω), and asymmetry parameter (g), where the difficulty in computing these optical properties is that they vary according to hydrometeor size and wavelength. The integral necessary to be solved for a given optical property (Slingo and Schrecker, 1982) is given by

$$\overline{\sigma}_{opt} = \int_{\Delta\lambda}\left[\int_{0}^{\infty} A(D)\sigma_{opt}(D,m,\lambda)n(D)dD\right]E_{\lambda}d\lambda \Big/ \int_{\Delta\lambda} E_{\lambda}d\lambda \qquad (2.24)$$

$$E_{\lambda} = S_{\lambda} \text{ (Solar)} \qquad E_{\lambda} = B(\lambda, T_s) \text{ (Infrared)}$$

where A(D) is the cross-sectional area of the hydrometeor, m is the complex index of refraction, E_λ is the solar/infrared energy density, T_s=273 K, σ_{opt} is an optical property such as Q_{ext}, ω, or g. The quantity n(D) is the hydrometeor size distribution given in (2.20).

Approximations exist for Q_{ext}, ω, and g, which can be used for substitution into (2.24). The Harrington scheme uses Anomalous Diffraction Theory (van de Hulst 1957) to approximate Q_{ext} and ω and uses the relationship between optical properties and effective radius (Hu and Stamnes 1993) to approximate g. However, even using the simplified, approximated forms of these optical properties, Eqn. (2.24) is still too costly to compute during a simulation. The parameterization of Harrington approximates these optical properties with a computationally efficient curve fit for each wavelength and hydrometeor class over a range of D_n and fixed shape ν. The curve fit has an exponential sum fit of the form

$$\omega_p = \omega_0 + \omega_1 e^{\omega_2 D_n} + \omega_3 e^{\omega_4 D_n}$$

$$g_p = g_0 + g_1 e^{g_2 D_n} + g_3 e^{g_4 D_n} \qquad (2.25)$$

$$Q_{ext} = b_0 + b_1 e^{b_2 D_n}$$

where ω_i (i=1,4) are the fit coefficients for the single-scatter coefficient, g_i (i=1,4) are the fit coefficients for the asymmetry parameter, and b_i (i=1,2) are the fit coefficients for the extinction coefficient.

d. Land Ecosystem-Atmosphere Feedback Model (LEAF)

OLAM's land surface scheme known as "LEAF" (Walko et al. 2000) numerically accounts for energy and moisture exchange between the different components of the ground surface including multiple soil and snow layers and vegetation, as well as the exchange between the surface and atmosphere. The hydrology model (TOPMODEL, Beven and Kirkby 1979, Beven 1982, 1984, Hornberger et al. 1985, Sivapalan et al. 1987) inclusive in LEAF represents lateral and downslope water transport in saturated soil, and has been updated to include frozen soils and runoff. LEAF represents the small-scale variations of the land surface such as vegetation and soil type, terrain slope, soil moisture, and bodies of water. Multiple surface types can exist under a single atmospheric column. Each land surface "patch" lying beneath an atmospheric column has multiple components including soil and snow layers, vegetation, and canopy air. Figure 2.5 illustrates the flux transport within a LEAF patch.

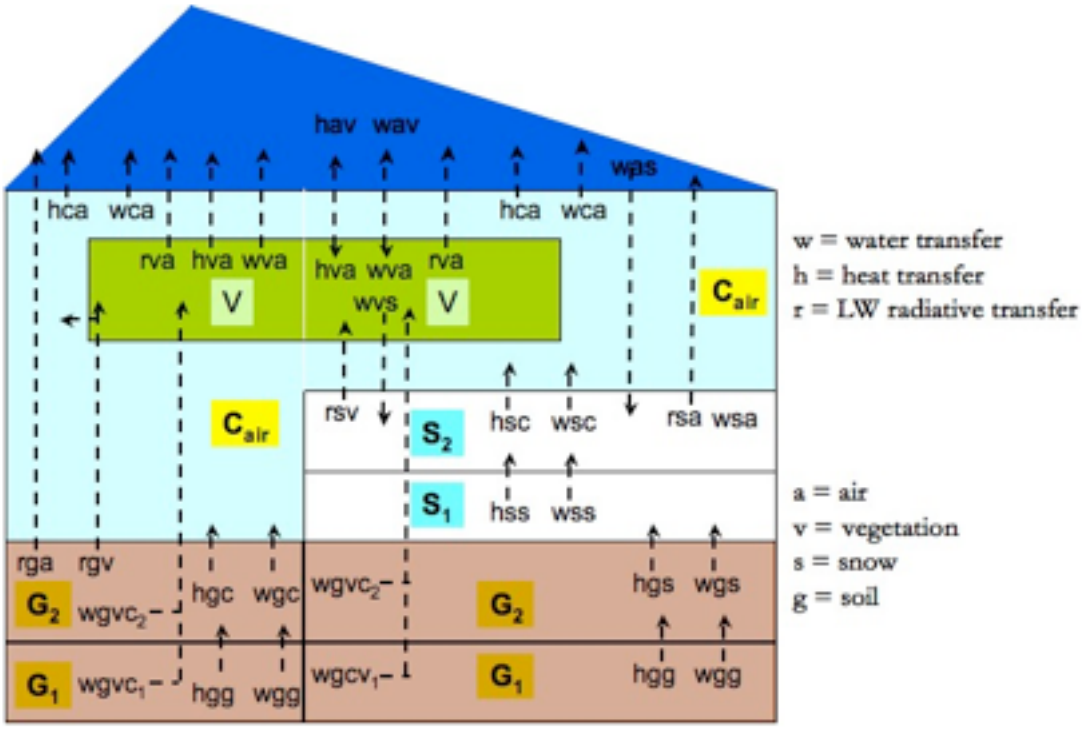

Figure 2.5: Flux transport in LEAF.

CHAPTER 3

A Precipitation Climatology and Dataset Intercomparison for the Western United States

3.1 Introduction

The western United States is a complex landscape consisting of coastal zones, mountains, basins, and plateaus. Its precipitation climatology is diverse as a result of interactions between orography and atmospheric dynamical processes occurring on multiple temporal and spatial scales. Competing climate controls and local differences in terrain and proximity to coastal moisture sources are such that unique precipitation climates are found within relatively short spatial distances (Mock 1996). This can be challenging for climate modeling because localities in close proximity may respond differently not only to real climate features but also to model errors. For example, model weaknesses associated with convection parameterizations will likely cause more error for the parts of the domain experiencing convective thunderstorm activity, whereas the failure to adequately resolve topography will be a greater problem for regions exhibiting complex terrain. As such, the systematic analysis of simulation results for each unique precipitation climate in the western United States is likely to lend insight into model strengths and weaknesses as well as into the study of local weather or climate characteristics or land-atmosphere feedback.

A useful approach for regionalizing a spatial domain into subdomains based on local precipitation climatology is principal component analysis (PCA). Many studies have used PCA and related eigen techniques to identify patterns of precipitation covariability and isolated regions that are spatially cohesive with respect to their precipitation regimes (e.g., Beaudoin and Rousselle 1982; Richman and Lamb 1985; Ehrendorfer 1987; Richman and Lamb 1987; White et al. 1991; Carter and Elsner 1997; Comrie and Glenn 1998; Hawkins et al. 2002; Dinpashoh et al. 2004; Gutzler 2004; Diem 2006; Diem and Brown 2006; Gochis et al. 2006). The regions identified by PCA have been used for many purposes such as identifying persistent patterns of thunderstorm activity (Easterling 1990), developing a statistical rainfall forecasting model (Carter and Elsner 1997), linking synoptic-scale circulation patterns to precipitation anomalies (Diem 2006), developing an index of North American monsoon (NAM) variability (Gutzler 2004), identifying snow–NAM teleconnections (Ellis and Hawkins 2001; Hawkins et al. 2002), elucidating atmosphere–rainfall relationships (Diem and Brown 2006), and many others.

However, past PCA-based regionalization studies have relied on rain gauge networks, which limit the utility of the method because many parts of the world are covered sparsely by, or are void of, these networks. Additionally, for the western United States with its complex terrain, rain gauge coverage tends to be clustered in the low elevations, whereas coverage in orographic zones is sparse. Given that numerous datasets are available for the western United States, including satellite data and reanalysis, it is not clear a priori whether rain gauge data would yield the most representative results in a regionalization study. Additionally, the ability to extend a study domain beyond land

areas using global products would be useful for numerical modeling studies because model domains often include an oceanic region. However, all precipitation products contain errors and many are available at a coarse spatial resolution on the order of 2.5° x 2.5°, and it is not clear whether this coarse resolution would affect the regionalization results. Therefore, this study investigates the utility of alternate data products for regionalization studies by conducting a comparative regionalization of the precipitation climate of the western United States using several observation-based precipitation datasets.

First, we conduct a series of PCAs using nine precipitation datasets rescaled to a common 2.5° x 2.5° grid and quantitatively compare their loading patterns. Next, we regionalize the domain based on the relative strength of gridcell loadings for each dataset, and then we compare the results for the different datasets and physically interpret any differences in the initial precipitation data. Finally, we test the stability of the method to differences in spatial/temporal data features. Section 3.2 describes the datasets and section 3.3 discusses rescaling. Section 3.4 describes the methodology of the PCA and regionalization. Section 3.5 contains a discussion of the results. Conclusions are given in section 3.6. This chapter is summarized in Guirguis and Avissar (2008a)

3.2 Data

The nine datasets included in this study and their spatial/temporal characteristics are shown in Table 1.1. A brief description is provided below.

a. GPCC monitoring product

The GPCC monitoring product provides monthly precipitation data from 1986 for global land areas based on rain gauge data from approximately 7500 stations (Fuchs et al. 2007). Source data are surface synoptic weather reports (SYNOP) and monthly climate bulletins (CLIMAT), which are quality controlled by automated and manual processes (Rudolf et al. 1994). The SYNOP provide information on precipitation totals for time intervals ranging from 1 to 24 h. Incomplete monthly time series are not uncommon and can result in local monthly precipitation estimates that are based on less than optimal temporal information (Rudolf et al. 1994; Rudolf and Schneider 2005). Following the quality-control (QC) processes, a precipitation value is assigned to a location based on source reliability information. The data are interpolated to a 0.5° x 0.5° grid and spatially averaged over 2.5° x 2.5° (Rudolf et al. 1994; Rudolf and Schneider 2005).

b. GPCP Combined Precipitation Dataset, version 2

The GPCP (Adler et al. 2003) is a global, merged satellite-gauge data product that is available monthly from 1979 at a 2.5° x 2.5° spatial resolution. The combination method uses a multistep process in which lesser-biased data sources are used to calibrate higher-biased sources to reduce the overall bias of the end product. For latitudes spanning 40° north–south, precipitation estimates from the IR-based Geostationary Operational Environmental Satellite (GOES) precipitation index (GPI; Arkin and Meisner 1987) are compared with Special Sensor Microwave Imager (SSM/I) data when both are simultaneously available. The GPI–SSM/I bias is calculated and used to calibrate the GPI precipitation estimates everywhere. Poleward of 40°, the satellite component is a combination of SSM/I and Television and Infrared Observation Satellite (TIROS)

Operational Vertical Sounder (TOVS) data. The multisatellite data are adjusted to agree with large-scale (5 x 5 grid cells each at 2.5°) averages of rain gauge data to reduce satellite bias. The corrected satellite data are then merged with rain gauge data using inverse error weighting to incorporate gauge-measured local variability (Huffman et al. 1997). Beginning in 1986, the gauge component of the GPCP is the GPCC monitoring product (section 3.2a), which is corrected for systematic rain gauge error using Legates and Willmott (1990). In the pre-SSM/I period (before July 1987), satellite estimates are from the outgoing longwave radiation precipitation index (Xie and Arkin 1998).

c. CPC Merged Analysis of Precipitation

The CMAP (Xie and Arkin 1996) is a monthly product that is available from 1979 at a 2.5° x 2.5° resolution. It combines precipitation estimates from satellite IR, SSM/I, and Microwave Sounding Unit (MSU) with rain gauge data in two steps. The first step is to reduce random error, and the second step is to reduce bias. Over land, the first step uses inverse (random) error variance weighting to combine the satellite products, which estimates the random error from large-scale averages of rain gauge data. The second step merges the combined satellite data with rain gauge data to reduce overall bias. Using the blending algorithm of Reynolds (1988), the combined product is assumed equal to rain gauge data for those grid areas having adequate coverage, whereas other values are calculated by solving a form of the Poisson equation. The result is that the satellite product determines the precipitation pattern, and the rain gauge data constrains the amplitude. From 1986, the rain gauge component is the GPCC monitoring product (section 3.2a).

d. CPC retrospective U.S. and Mexico daily precipitation analysis

The USMex is a gridded rain gauge product available at a 1° x 1° resolution from 1948. The U.S. data are based on the CPC Unified Raingauge Data (URD; Higgins et al. 2000), which are composed of (i) the daily CPC Cooperative (COOP) dataset (7000 rain gauges from 1992), (ii) the National Climatic Data Center (NCDC) COOP dataset (8000 rain gauges from 1948), and (iii) the hourly precipitation dataset as described in Higgins et al. (1996; 2500 rain gauges from 1948). Accounting for overlap between networks, the URD typically represents 13 000–15 000 daily rain gauge sites; however, prior to 1992 the number of rain gauges approaches 8000. The Mexican data are based on approximately 200 rain gauges prior to 1990 and 600 rain gauges thereafter. The data are quality controlled for duplicate and extreme values and bias corrected for spurious zero measurements against collocated ground radar data. The gridding of the data uses a modified Cressman scheme (Cressman 1959). Although there is likely some overlap between this USMex rain gauge data product and the GPCC monitoring product described above, the GPCC data contains only a fraction of the information contained in the USMex dataset (i.e., 7500 rain gauges globally for the GPCC data versus 13 000–15 000 in the United States for the USMex data). These datasets also differ in their interpolation methods, and interpolation alone has been shown to affect both correlation and bias with respect to actual point measurements (Chen et al. 2002).

e. Precipitation-elevation Regressions on Independent Slopes Model

PRISM is an analytical model that uses point source precipitation observations together with a digital elevation model (DEM) to provide gridded precipitation data

products for the United States (Daly et al. 1994). The development of PRISM was motivated by the need for precipitation data over elevated terrain where rain gauge data are sparse. It attempts to resolve some important problems associated with other orographic interpolation methods, which rely heavily on neighboring observations to approximate data-void localities even though neighboring sites may exhibit important orographic differences. The PRISM methodology is i) to use a DEM to estimate the elevation of each available gauge station, and ii) to group stations according to their orographic commonality (or "facet") with respect to orographic features such as elevation, face orientation, slope, and proximity to coasts. Precipitation at a given DEM cell is estimated using a regression of precipitation versus elevation, where the stations sharing the same facet as the grid cell of interest are used to determine the parameters of the regression. The primary point source estimates used as input to PRISM are from the NCDC Historical Climate Network (1200 stations), CPC COOP (8000 stations), and Natural Resources Conservation Service snowpack telemetry (SNOTEL) data (730 sites). PRISM data are available from 1890 as monthly or annual averages with a spatial resolution of 32 x 32 km^2.

f. NCEP–DOE Reanalysis 2

NCEP2 (Kanamitsu et al. 2002) is a global data product created from a spectral data assimilation system that merges observations from many sources. These include rawinsondes for upper air variables, satellites for vertical temperature soundings and cloud drift winds, aircraft for wind and temperature, ocean reports of surface variables, synoptic weather reports over land, and pentad CMAP data as a precipitation correction

to improve soil moisture fields (Kalnay et al. 1996; Kanamitsu et al. 2002). Following extensive QC, the data are assimilated using a T62 (210 km) global spectral model, which has 28 vertical levels and includes parameterizations of the major physical processes such as convection, large-scale precipitation, boundary layer physics, and so on. The assimilation process generates additional variables that are derived from observed fields. These derived variables (class C) include clouds, surface fluxes, and precipitation. Their reliability is heavily dependent on the performance of the model and its parameterizations (Kistler et al. 2001). The NCEP2 data are available in 6-h increments from 1979.

g. North American Regional Reanalysis

NARR (Mesinger et al. 2006) is an atmospheric and hydrology dataset covering North America. NARR uses the regional Eta Model (forced at its lateral boundaries by NCEP2) to assimilate observations in a similar manner used for global reanalysis products. In addition to a higher spatial (32 x 32 km^2 with 45 levels) and temporal (3 hourly) resolution as compared to global reanalysis, NARR uses additional and improved input data and contains a better representation of land surface hydrology and land–atmosphere interactions. A particular improvement is the assimilation of observed precipitation. NARR uses precipitation observations to correct atmospheric moisture and energy fields, which subsequently leads to improvements in the model-derived precipitation fields. NARR data are available from 1979.

h. VIC Retrospective Land Surface Dataset

The VIC land surface dataset is based on 50-yr (1950–2000) simulations with the

VIC land surface hydrology model. The simulations were designed to provide a high-resolution (3 hourly and 1/8° x 1/8°) dataset of land surface states and fluxes for the continental United States and for parts of Canada and Mexico (Maurer et al. 2002). The model includes a soil–vegetation–atmosphere transfer scheme and is forced by soil and land use data and observations of precipitation, air temperature, wind, humidity, and radiation. Model-derived variables include snow water and multiple-layer soil moisture tendencies, surface temperature, and latent and sensible heat fluxes. The VIC dataset also provides the model-forcing data including precipitation. The precipitation data are based on daily totals from the CPC COOP stations, with an approximate density of one gauge per 700 km^2 (Maurer et al. 2002), which is gridded to a 1/8° x 1/8° resolution (Shepard 1984; Widmann and Bretherton 2000). Variation due to orography is incorporated in the data using an adjustment factor for each month and grid cell, which is calculated as the ratio of monthly mean precipitation from PRISM (section 3.2e) to that of the COOP data.

i. Global Meteorological Forcing Dataset for Land Surface Modeling

GMFD (Sheffield et al. 2006) is a high-resolution (3 hourly and 1° x 1°) dataset spanning 1948–2000, developed by Princeton University's Land Surface Hydrology Research Group to provide near-surface meteorology variables on space–time scales necessary for forcing land surface and hydrology models. The forcing data includes precipitation fields, which are developed using NCEP global reanalysis as the primary input and uses other data to correct known reanalysis errors and to disaggregate the corrected precipitation fields to the desired 3-hourly, 1° resolution. The precipitation observations are the Climatic Research Unit (CRU) monthly climate variables (which

include rain day statistics), the GPCP daily precipitation product, and the Tropical Rainfall Measuring Mission (TRMM) 3-hourly precipitation. The corrections to the NCEP data are for (i) rain day frequency statistics (Sheffield et al. 2004), which involves resampling NCEP daily precipitation to match the rain day statistics of CRU, GPCP, and TRMM observations; and (ii) rain gauge undercatch, which uses monthly adjustment ratios from Adam and Lettenmaier (2003). Spatial downscaling uses probabilistic relationships between precipitation intensity observed on a daily reanalysis grid, with observed gridcell fractional coverage as determined from higher-resolution precipitation observations. Temporal downscaling uses probability density functions derived from TRMM.

3.3 Data rescaling

The precipitation datasets used in this study have spatial resolutions ranging from 1/8° to 2.5°. To compare these data directly, the data were rescaled prior to the analysis to a common 2.5° x 2.5° grid, extending from the Pacific coast to slightly east of the Rocky Mountains and from northern Mexico to slightly north of the U.S.–Canadian border (27.5°–50.0°N and 130.0°–102.5°W). For datasets having a horizontal resolution smaller than 2.5°, simple box averaging was employed to rescale the data. For those datasets having an original horizontal grid size of 1° or smaller, upscaling was done iteratively to prevent the propagation of undefined values that occur over the Pacific Ocean for rain gauge data as well as to prevent the spatial propagation of small-scale, intense precipitation events to the coarser scale. Bilinear interpolation was then employed to adjust the 2.5° data such that all datasets have collocated gridcell centers. Because

PCA has been shown to be sensitive to domain shape (e.g., Richman and Lamb 1985; Richman 1986), we opted to modify the PRISM data at its southern boundary to account for its lack of data over northern Mexico by filling five data-void grid cells centered at 28.75 °N with values of their nearest northern neighbor. The temporal resolutions of the nine datasets vary from 3 hourly to monthly. However, monthly averages of temporally high-resolution data are often available. Monthly products were used for all datasets except the USMex, for which we constructed monthly mean values from daily data. The period of study is January 1986–July 2000, which is the time of maximum overlap between datasets. Units of measure have been converted to mm month $^{-1}$ for all datasets.

3.4 Methods

S-mode, rotated PCA operating on the correlation matrix is used to disaggregate the domain of the western United States into regions that are unique and spatially cohesive with respect to their precipitation climate. Prior to the analysis, monthly precipitation values are weighted by $(\cos\varphi)^{1/2}$, where φ is latitude, to correct for latitudinal differences in grid spacing (e.g., Wilks 2006). The correlation matrix is used as the dispersion matrix because there are large differences in precipitation variance across the domain and using the correlation matrix (rather than the covariance matrix) prevents the high-variance grid cells from disproportionately influencing the results (Jolliffe 2002; Wilks 2006). The correlation coefficient can be sensitive to data nonnormality, and monthly precipitation data are skewed. However, normality is not required for PCA to be valid (Wilks 2006). In preliminary analyses, we tested the sensitivity of the methodology to transformations of the input data (which brought the

data closer to normal) and found no notable differences in the resulting principal component (PC) loading patterns.

The appropriate number of PCs to rotate is selected using the scree test (Cattell 1966), the log eigenvalue (LEV) diagram, and the North et al. (1982) eigenvalue separation test. The scree test and LEV diagram use a plot of the eigenvalue spectrum (eigenvalues in de creasing order) and the log eigenvalue spectrum, respectively, and the appropriate truncation point is observed as a discontinuity in slope. For the scree plot, this appears as a transition point between the steeply sloped part of the spectrum to the left and the gently sloped portion to the right (Wilks 2006), which marks the point where the cumulative variance explained with each added PC is approaching zero. For the LEV diagram, the log eigenvalues that form an approximately straight line indicate the exponential decay of PCs dominated by uncorrelated noise (e.g., Wilks 2006), so these and subsequent PCs would not be retained. The eigenvalue separation test is applied to avoid separating closely spaced PCs. Neighboring PCs whose eigenvalues are close in size are associated with large sampling errors whereby different samples may produce different linear combinations of the closely spaced PCs. As such, the patterns produced by closely spaced PCs ("effectively degenerate multiplets") are not independent but rather random mixtures of the true population PCs (North et al. 1982), and these should not be separated in a rotated PCA (North et al. 1982; Jolliffe 2002). Neighboring multiplets are identified as those having overlapping standard error bars, where standard error is estimated by $\delta=\lambda(2/N)^{1/2}$, and where is the eigenvalue and N is the sample size (North et al. 1982). Because monthly precipitation data are autocorrelated, the standard error approximation is modified in this study to use an effective sample size.

Specifically, we use $N^*=N(1-r^2)/(1+r^2)$, which is appropriate for variances (Bretherton et al. 1999), where r is the domain-averaged autocorrelation coefficient. The autocorrelation ranges from 0.34 to 0.39 for the nine datasets.

The PCs are rotated using the direct oblimin oblique rotation method. Before rotating, each PC is weighted by the square root of their corresponding eigenvalue, which makes interpretation more intuitively meaningful because the weighted PC elements are then a measure of the correlation between the original data and the PC time series (e.g., Richman and Lamb 1987; Wilks 2006). The rotated PCA is used because rotation has been shown to improve the interpretability of the solution and can eliminate some problems caused by the orthogonality constraint associated with unrotated PCs such as Buell patterns, which are artificial and misleading patterns resulting from domain shape rather than any physical relationship between variables (Buell 1979) and larger sampling errors as discussed above (Richman 1986; Richman and Lamb 1987; Jolliffe 2002; Wilks 2006). Additionally, rotated PCs (RPCs) show more similarity to the original data than unrotated solutions (Richman 1986; Wilks 2006) and, for regionalization studies in particular, rotated PC solutions have been shown to be more stable to changes in domain size and shape (Richman and Lamb 1985; Richman 1986; White et al. 1991). The oblique rotation rather than an orthogonal rotation is chosen because oblique methods have been shown to be more stable with respect to changes in spatial and temporal domain and also superior in achieving simple structure (White et al. 1991).

The rotated PCs are used to regionalize the domain of the western United States according to the maximum loading method (Comrie and Glenn 1998). This approach assigns each grid cell to the RPC onto which it loads most strongly, and the grid cells

Table 3.1: Interpretation of congruence coefficient representing the degree of likeness between rotated loading vectors

≥0.98	Excellent	
≥0.92	Good	<0.98
≥0.82	Borderline	<0.92
≥0.68	Poor	<0.82
	Terrible	<0.68

assigned to the same RPC are then grouped together into distinct "regions," which results in the regionalization of the domain.

The degree of similarity between rotated PC solutions obtained from different datasets is assessed by quantitatively comparing their respective loading matrices using the congruence coefficient (Harman 1976):

$$g = \frac{\sum_{j=1}^{n} \left(b_{ja} b_{jb} \right)}{\left[\left(\sum_{j=1}^{n} b_{ja}^2 \right) \left(\sum_{j=1}^{n} b_{jb}^2 \right) \right]^{1/2}} \tag{3.1}$$

where b_a is a loading from rotated loading vector A from one solution, b_b is a loading from rotated loading vector B from another solution, and n is the number of variables in each eigenvector, which for S-mode PCA corresponds to spatial positions. A value of +1 for the congruence coefficient represents perfect agreement between solutions, and a value of -1 represents perfect inverse agreement and a value of 0 indicates no agreement. The congruence coefficient is preferred to the correlation coefficient for measuring pattern similarity because it preserves the mean (whereas the correlation coefficient measures deviations from the mean), and the mean is an important feature of a PC loading vector (Richman 1986). Following Richman (1986) and the references therein, the guidelines shown in Table 3.1 are adopted as an indicator of the degree of likeness

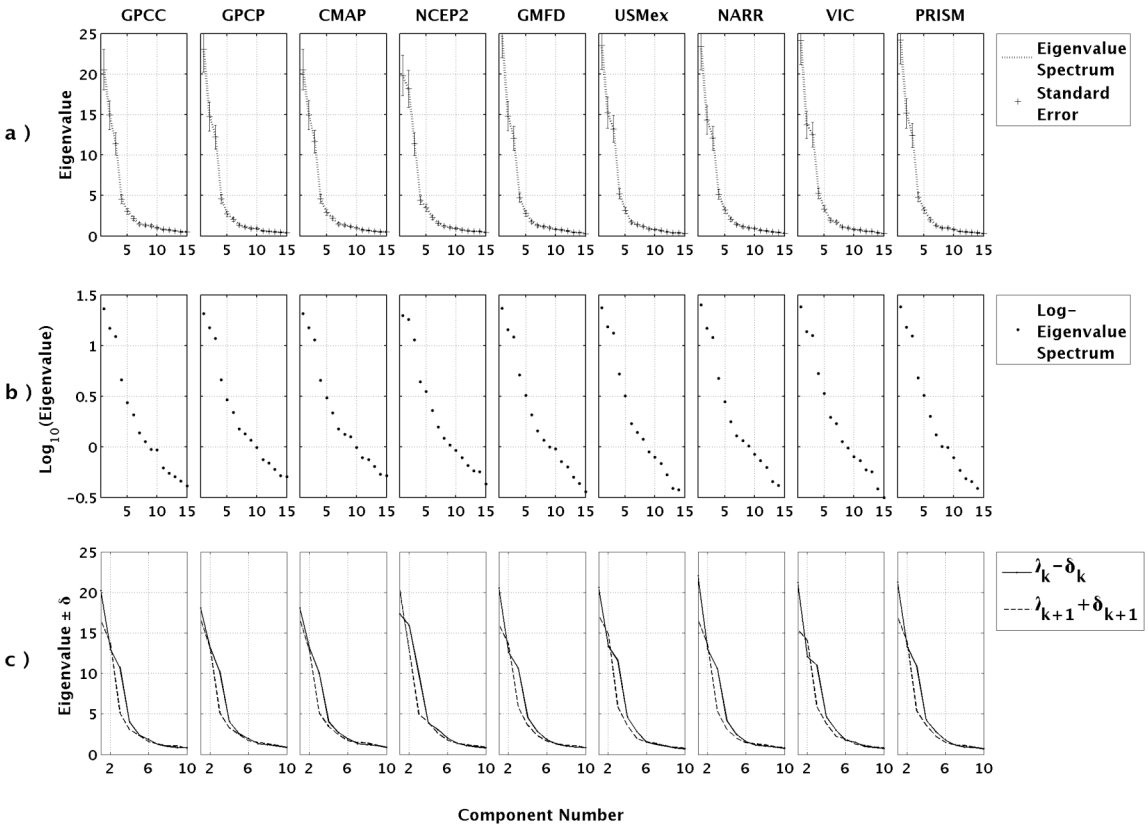

Figure 3.1: (a) Scree plot, (b) LEVdiagram (b), and (c) eigenvalue separation test for the nine precipitation datasets.

between solutions, which are based on the Monte Carlo studies of the distribution of the congruence coefficient by Korth and Tucker (1975). From this table, the congruence coefficient is biased toward higher values as compared to the correlation coefficient and, according to Richman (1986), any coefficient smaller than 0.7 represents a match that is not any better than would be expected by "randomly spinning" the PC axes prior to matching a set of PC loading vectors.

Two additional comparative RPC analyses are performed following the methodology just described to test the sensitivity of the method to differences in spatial resolution and temporal domain. The USMex and GMFD data are used for these

sensitivity tests because they have long records and a desirable original 1° x 1° spatial resolution. The 1° scale is small enough to provide a meaningful comparison with the 2.5° data, while not being so small as to become computationally difficult to process. The RPC solutions obtained using 1° versus 2.5° data are compared using USMex and GMFD data. Next, using 1° data from USMex and GMFD, the affect of temporal sample size is tested by comparing results obtained using data from January 1986 to July 2000 versus the longer time series of January 1950–December 2000. In each case, the loading patterns are compared using the congruence coefficient as the indicator of similarity.

3.5 Results

a. Principal component analysis and regionalization

Figure 3.1 gives the scree graphs, LEV diagrams, and eigenvalue separation tests for each dataset. Based on our interpretation of Figs. 3.1a–c, five PCs are retained and rotated for each dataset. The slope discontinuity observed on the scree and LEV graphs occurs at the fifth or sixth eigenvalue; however, the choice of retaining five versus six PCs is subjective, and the level of ambiguity varies by dataset. . The eigenvalue separation test shown in Figure 3.1c is a plot of λ_k-δ_k and λ_{k-1}+δ_{k-1}, where λ_k is the k^{th} eigenvalue, λ_{k-1} is its neighboring eigenvalue, and δ_k and δ_{k-1} are their corresponding standard error estimates. The intersection of these two lines is an indication that neighboring eigenvalues are not adequately separated. In Fig. 3.1c, effective degeneracy occurs at the sixth PC for nearly all datasets and, therefore, five PCs are retained. The first five unrotated PCs collectively account for between 76% and 85% (depending on dataset employed) of the variance contained in the original data (Table 3.2).

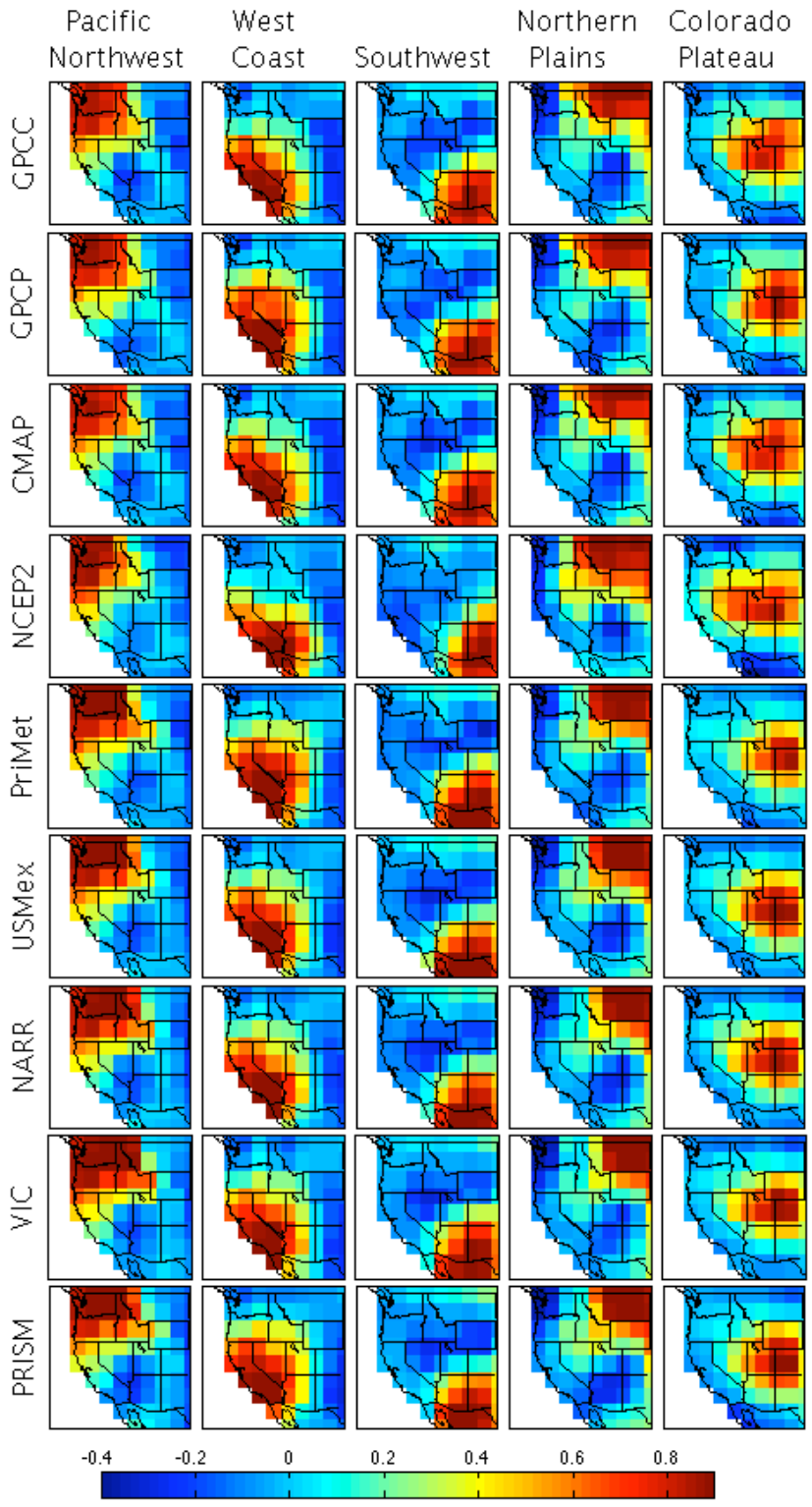

Figure 3.2: Rotated loading patterns corresponding to the five PCs for each dataset.

Table 3.2: Percent variance explained by the first five principal components.

Dataset	PC number				
	1	2	3	4	5
GPCC	28.95	20.99	16.03	6.4	4.30
GPCP	32.42	20.80	17.2	6.44	3.85
CMAP	28.94	21.05	16.45	6.45	4.10
NCEP2	27.92	25.60	16.02	6.20	4.93
GMFD	35.32	20.88	16.99	6.66	3.93
USMex	33.16	21.45	18.62	7.36	4.47
NARR	32.94	20.19	17.04	7.25	4.51
VIC	34.02	19.37	17.62	7.42	4.74
PRISM	34.08	21.35	17.48	6.75	4.52

The loading patterns corresponding to the five RPCs for each dataset are shown in Fig. 3.2 and are named and ordered according to their geographic centers: Pacific Northwest, West Coast, Southwest, Northern Plains, and Colorado Plateau. The rotated loading patterns given in Fig. 3.2 show the important influence of topography on the precipitation climate of the western United States. All datasets indicate a spatial pattern of variation centered over the Pacific Northwest, which is bounded to the east/southeast by the Rocky Mountains. There is also a California-centered pattern bounded to the east by the Rocky Mountains. The Southwest loading pattern is centered over northern Mexico and extends into Arizona and New Mexico, and is bounded to the north by the Colorado Plateau. The Northern Plains loading pattern is centered over the northern Great Plains at the state of Montana, and the Rocky Mountains bind the south/southwestward extent of the precipitation variability pattern. The fifth RPC for all datasets is centered over the high-elevation areas of the central Rocky Mountains near the Colorado Plateau.

The results of the principal component–based regionalization using the maximum

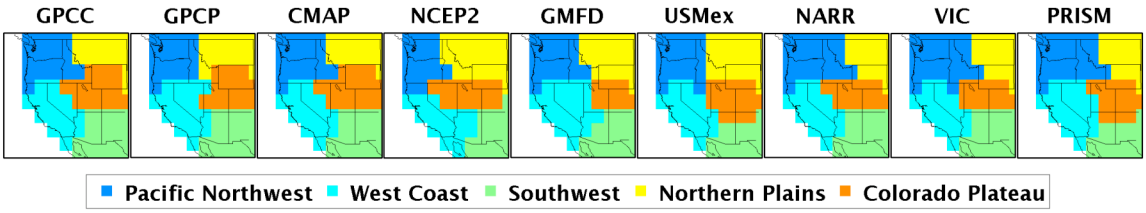

Figure 3.3: The five precipitation regions obtained from each dataset using the maximum loading method.

loading method are shown in Fig. 3.3. Figure 3.4 gives the monthly climatology for each of the five regions shown in Fig. 3.3 according to the nine datasets. The Pacific Northwest region extends from the Pacific coast to the Rocky Mountains, and its southern boundary is near the northern limit of the Sacramento Valley and the southern limit of the Cascades. This region has a cold-season precipitation regime with a maximum in November-January (Fig. 3.4) that is associated with cyclonic storms emanating from the northern Pacific and dry summers. The West Coast region extends from the Oregon California border southward to include the entire state of California and extends eastward to include all or part (depending on the dataset) of Nevada. The eastern boundary of the West Coast region is in the vicinity of where the Great Basin meets the Rocky Mountains. This region also has a winter precipitation regime similar to the Pacific Northwest but with its maximum occurring January–March. This seasonal shift in the cold-season maximum is attributed to a more southern concentration of cyclonic storm activity and the southward progression of the jet stream (Trewartha 1981; Mock 1996). The Southwest region is bound to the west at approximately the Arizona–California border in the vicinity of the Mojave Desert, which marks the transition between the Sonora Desert and the higher elevations of the Great Basin. Its northern boundary is near

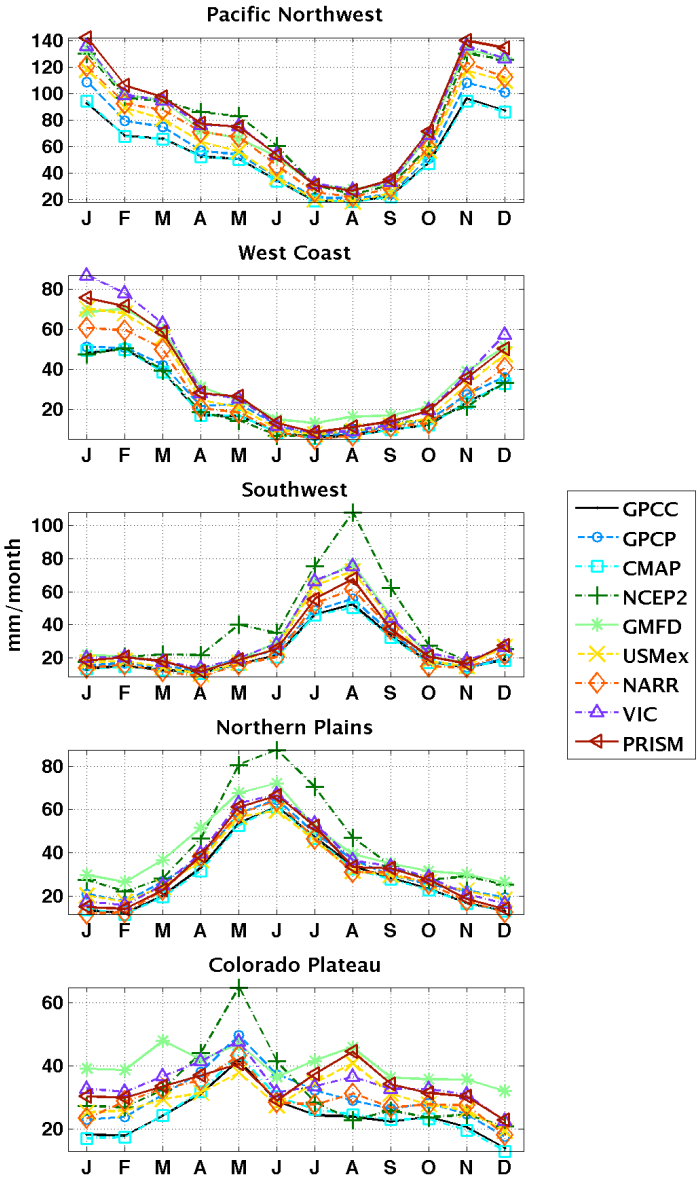

Figure 3.4: Monthly climatology for the five regions.

the Mogollon Rim, which marks the southern edge of the Colorado Plateau in Arizona. This Southwest region experiences a summer precipitation climate, with a peak from July-September marking the arrival and duration of the North American monsoon. The Northern Plains region is bound to the west by central or eastern Idaho (depending on the dataset) near where the Rocky Mountains intersect the northern Great Plains. This

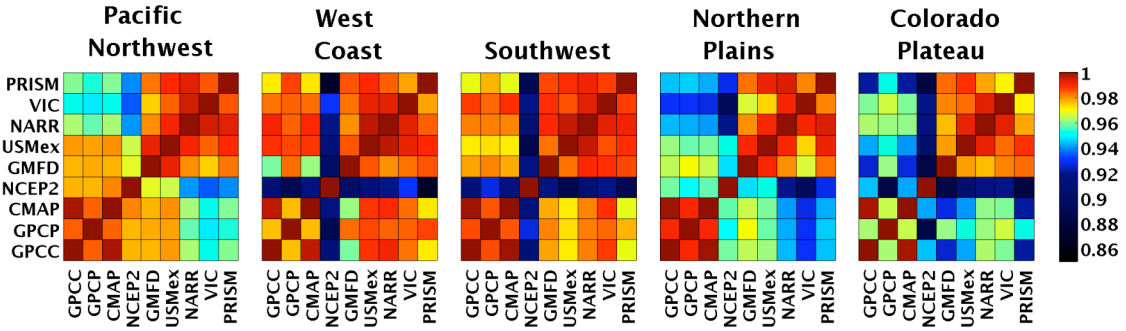

Figure 3.5: Congruence coefficient for each data pair.

region includes most or all of the state of Montana; however, the horizontal extent varies among datasets with respect to the proportion of Idaho and Wyoming contained. The Northern Plains region experiences an early summer (May–July) precipitation regime arising from the summer land–sea temperature gradient, which allows for advection of Gulf of Mexico moisture into the deep continental interior (Trewartha 1981). The Colorado Plateau region spans the high-elevation areas of Colorado and Utah (and Wyoming according to the CPCC, GPCP, and CMAP data). This region exhibits a bimodal precipitation regime that includes an early summer precipitation maximum similar to the precipitation climate of the Northern Plains and another maximum in late summer during the North American monsoon. For the Colorado Plateau, the positioning of the region itself (Fig. 3.3) explains some of the differences observed among the datasets with respect to the monthly climatology. For example, the summer peak centered in August is not observed according to the GPCP, CMAP, or GPCC data. This is probably because the Colorado Plateau region extends farther north according to these three datasets, thus the region is out of the range of primary North American monsoon influence.

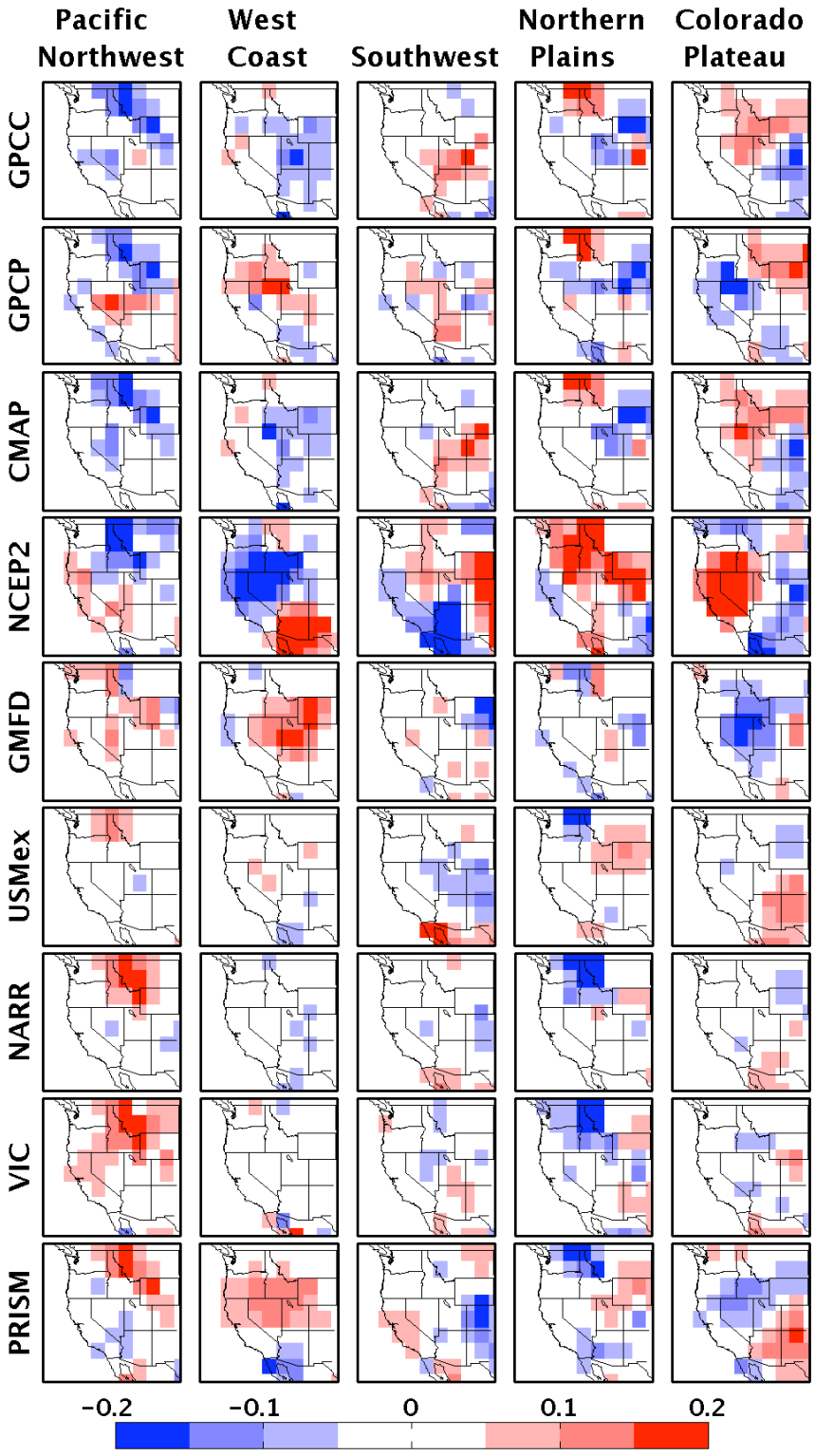

Figure 3.6: Difference from ensemble of rotated loading patterns, where the ensemble was created from the average of the nine datasets for each RPC.

b. Dataset Intercomparison

Figure 3.5 gives the congruence coefficients for all data pairs for each of the five rotated loading patterns shown in Fig. 3.2. For the Pacific Northwest loading pattern, there is a high level of similarity among data with congruence coefficients above 0.93 (the good-to-excellent range) for all data pairs. For the West Coast and Southwest, the similarity among data is very high in general, with congruence coefficients at or above 0.97 for most data pairs. The exception is NCEP2, which generally exhibits only "borderline" agreement with the other data (g=0.86–0.93). For the Northern Plains loading pattern, congruence is generally in the good-to-excellent range (g=0.92) except for the comparisons between NCEP2 with NARR and VIC, which show only borderline agreement (g=0.89–0.92). For the Colorado Plateau, there is good-to-excellent congruence (g=0.92) among datasets with the exception of NCEP2, which exhibits borderline congruence (g=0.88–0.92) with all but two datasets (GPCC and CMAP, g=0.94). The clustering of high-congruence coefficients in the southwest and northeast quadrants of each panel in Fig. 3.5 highlights two data groups that exhibit very similar loading patterns. These are the GPCP, CMAP, and GPCC datasets (Group 1) and the NARR, USMex, GMFD, VIC, and PRISM datasets (Group 2). These datasets show a higher level of congruence within groups (g=0.97–1.0, g=0.98–1.0, g=0.98–1.0, g=0.97–1.0, g=0.96–1.0 for the Pacific Northwest, West Coast, Southwest, Northern Plains, and Colorado Plateau, respectively) and lesser congruence between groups (g=0.95–0.98, g=0.96–0.99, g=0.97–0.99, g=0.93–0.97, g=0.92–0.97 for the Pacific Northwest, West Coast, Southwest, Northern Plains, and Colorado Plateau, respectively).

Figure 3.6 shows the difference between each loading pattern and an ensemble

mean, where the ensemble mean was created from the average of the nine datasets for each RPC as

$$E(i,r) = \frac{1}{N}\sum_{n=1}^{N} x_n(i,r) \tag{3.2}$$

where $E(i,r)$ is the ensemble mean for the rth RPC at position i, and $x_n(i,r)$ is the nth ensemble member for the rth RPC at position i, and N is the number of ensemble members. To relate the differences observed in Fig. 3.6 with the raw precipitation data, we consider the time-varying amplitudes of the five PCs. Specifically, we isolate the months in the January 1986–July 2000 record in which the PC amplitudes are strongest, which we take as those above the 80th percentile. The mean precipitation for those months corresponding to the largest amplitudes is shown in Fig. 3.7 for each dataset.

In Fig. 3.6 for the Pacific Northwest, loading pattern differences between datasets occur primarily along the Rocky Mountains from Idaho to Colorado. This is probably related to the differences in original spatial resolution among datasets where the higher resolution data (NARR, USMex, GMFD, VIC, and PRISM) are better able to represent the orographic enhancement of precipitation that occurs along the western Rocky Mountains than the coarser data. Some of this high-resolution information is carried over to the coarser scale when the data are regridded to the 2.5° spatial resolution (Chapter 4). This enhanced precipitation along the Rockies for the Group B datasets is also seen in the raw precipitation data shown in Fig. 3.7. In Fig. 3.6 for the West Coast, the NCEP2 data are shown to differ from the ensemble over most of the domain, which is a result of a shift to the southeast of the West Coast loading pattern for the NCEP2 data as compared to the others (Fig. 3.2). Considering Fig. 3.7, this shift appears to be related to the overestimation of precipitation in the vicinity of Arizona, which causes grid cells in

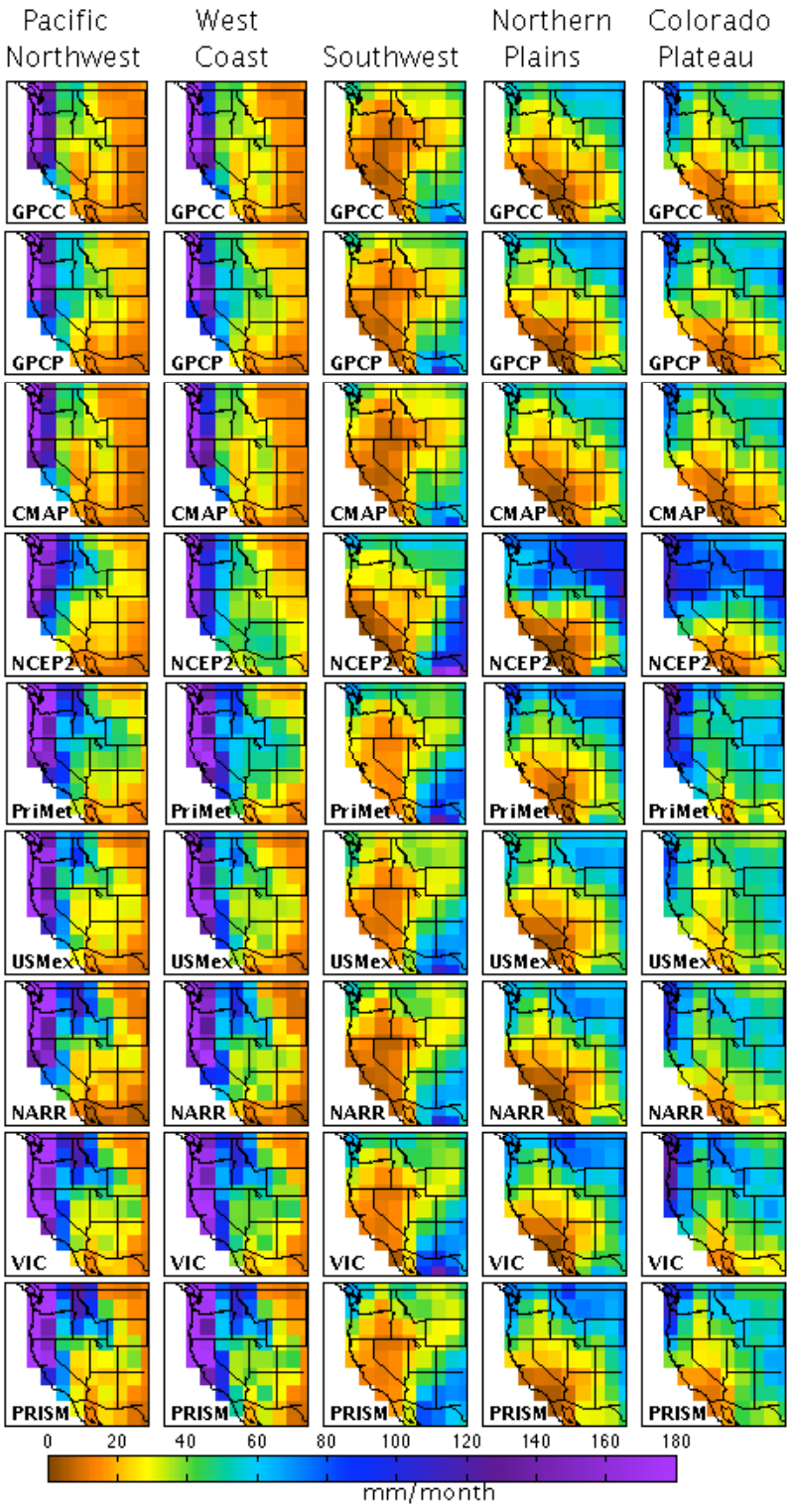

Figure 3.7: Mean precipitation for those months in which the PC amplitudes exceed the 80th percentile

Dataset	PC number				
	1	2	3	4	5
GPCC	28.95	20.99	16.03	6.4	4.30
GPCP	32.42	20.80	17.2	6.44	3.85
CMAP	28.94	21.05	16.45	6.45	4.10
NCEP2	27.92	25.60	16.02	6.20	4.93
GMFD	35.32	20.88	16.99	6.66	3.93
USMex	33.16	21.45	18.62	7.36	4.47
NARR	32.94	20.19	17.04	7.25	4.51
VIC	34.02	19.37	17.62	7.42	4.74
PRISM	34.08	21.35	17.48	6.75	4.52

Table 3.3: Congruence coefficients for (a) precipitation data at 1° vs 2.5° and (b) precipitation data for January 1950-December 2000 vs January 1986-July 2000.

the area to load more strongly onto the West Coast RPC than the other datasets. Also for the West Coast loading pattern and considering the other datasets, differences are observed in the intermountain region in Fig. 3.6, which in Fig. 3.7 appears to be related to the degree of severity of the eastward wet-to-dry precipitation gradient. For example, the GPCC and CMAP show a much stronger precipitation gradient with drier inland conditions, and GMFD shows a weaker gradient with wetter intermountain conditions. For the Southwest, the loading pattern differences appear to result from differences among data in their representation of the North American monsoon. Specifically, in Figs. 3.6 and 3.7, the data differ with respect to the strength of the monsoon signal as well as the spatial extent of monsoon influence. In Fig. 3.6 for the northern plains, the data differences are observed primarily over the Rocky Mountains, which in Fig. 3.7, are related to the tendency of the Group 1 datasets to represent drier conditions over Idaho, eastern Washington, Wyoming, Utah, and Colorado as compared to the Group 2 datasets. Also in Fig. 3.6 for the Northern Plains, the NCEP2 data demonstrates much stronger loadings onto the northern plains RPC as compared to the other datasets, which (in Fig.

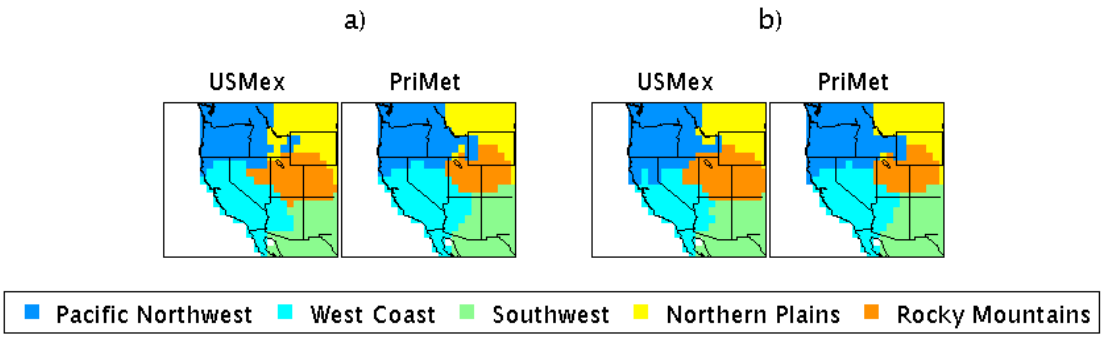

Figure 3.8: The five precipitation regions obtained from 1° USMex and GMFD data over (a) the January 1986-July 2000 time domain and (b) the longer period of January 1950-December 2000.

3.7) appears to be a result of the comparatively wet conditions observed over the northern half of the domain for the NCEP2 data. In Fig. 3.6 for the Colorado Plateau, differences between datasets are observed over much of the domain. These differences are primarily a result of (i) the drier conditions observed over the Southwest region for the GPCC, GPCP, and CMAP data as compared to the other datasets, and (ii) the wetter conditions observed for the intermountain region and along the Rocky Mountains according to PRISM, VIC, NARR, USMex, and GMFD. Also for the Colorado Plateau loading pattern, the NCEP2 data differs greatly from the other datasets, which in Fig. 3.2 is observed as an elongation to the west of the NCEP2 Colorado Plateau RPC relative to those of the other datasets. In Fig. 3.7, this appears to be caused by overestimated precipitation along the northern part of the domain according to NCEP2.

c. Space–Time Sensitivity Studies

Table 3.3 gives the results of two comparative RPC analyses using data from USMex and GMFD, which were designed to test the sensitivity of the methodology to differences in spatial resolution and temporal domain. The spatial resolution sensitivity

study uses Eq. (3.1), where b_A is the RPC from the dataset at its original 1° resolution, and b_B is the loading pattern from the same dataset rescaled to 2.5°. Because the calculation represents a point-to-point comparison, linear interpolation is applied to the coarse-resolution loading vector during the scale sensitivity analysis. The analysis of sensitivity to temporal sample size uses Eq. (3.1), where b_A is the loading vector obtained from 1° data over a long record (January 1950–December 200), and b_B is the loading vector obtained from 1° data from same dataset over a shorter record (January 1986–July 2000).

In Table 3.3, the methodology is shown to be stable to changes in spatial resolution and temporal domain. The rotated loading patterns obtained using a shorter versus longer time domain are shown to be very similar with congruence coefficients exceeding 0.98, representing excellent agreement, for all five RPCs and for both datasets. For the spatial resolution sensitivity analyses, the loading patterns obtained using 1° versus 2.5° data are shown to exhibit good agreement with congruence coefficients in the range of 0.95–0.97. Figures 3.8a and 3.8b give the results of the regionalization for the 1° USMex and GMFD data over the shorter and longer records, respectively. A visual comparison of Fig. 3.8a with Fig. 3.3 shows the regionalization results to be very similar for the USMex and GMFD data, whether 1° or 2.5° data are used. The regionalization results for the short-term versus the long-term data are shown to be very similar (cf. Figures 3.8a versus 3.8b). The minor differences observed are generally a result of the migration of one or two grid cells into a neighboring region. Because grid cells bordering two neighboring regions are often modestly loading onto both corresponding PCs, it is not unexpected to observe this shift from one experiment to the next.

3.6 Summary and Conclusions

This study has identified the precipitation climate regions of the western United States while testing the utility of the methodology, which has been traditionally applied to rain gauge data, to alternate data products. Specifically, we employed rotated PCA with observational precipitation data to identity the dominant patterns of covariability in the western United States and to isolate spatially cohesive regions that experience similar temporal variability. This analysis was conducted on nine precipitation datasets including two reanalysis datasets, four rain gauge datasets (two of which use orographic adjustment), two satellite–rain gauge combined datasets, and one multiple-sourced product.

The dominant covariability patterns corresponding to the first five RPCs are centered over Washington, central California, northern Mexico at the border with Arizona and New Mexico, Montana, and Wyoming/ Colorado. Using the congruence coefficient as a measure of loading pattern similarity, we compared the patterns obtained from each of the nine precipitation datasets and found the methodology to be reasonably stable to alternate data sources. That is, the congruence coefficient indicated good-to-excellent similarity between most datasets, with the exception of NCEP2, which frequently demonstrated only borderline agreement with the other datasets. Also from this analysis, loading pattern differences were shown to be related to differences among datasets primarily in their representation of (i) the precipitation over the Rocky Mountains, (ii) the eastward wet-to-dry precipitation gradient that occurs during the cold season, (iii) the magnitude and spatial extent of the North American monsoon signal, and (iv) the precipitation in the desert southwest during spring and summer. The

overestimation of precipitation by NCEP2 was shown to be responsible for the general lack of similarity between NCEP2 and the other datasets.

To regionalize the domain of the western United States into subregions representing unique precipitation climates, we applied the maximum loading principle, which groups those grid cells exhibiting a similar temporal variability. This process produced five distinct regions whose centers and borders were found to be physically reasonable and which highlight the relationship between the precipitation climatology and the local topography.

Sensitivity experiments were conducted in which we compared the RPC solutions obtained from data having different spatial resolutions and temporal domains but alike otherwise. These results showed the methodology to be stable to spatial/temporal data differences. However, it is worth mentioning that the spatial/temporal characteristics of precipitation data can affect the outcome of the regionalization when the eigenvalue separation test is used for deciding the number of PCs to rotate because standard error calculations are dependent on the sample size (measured in the time domain) and autocorrelation structure (which can be affected by spatial resolution) of the underlying data.

The results of this study have applications for the installation and planning of ground-based instrumentation and/or for climate research. Furthermore, we see interesting uses for numerical modeling experiments. The boundaries obtained from PCA-based regionalization can be used for selecting the extent of a model domain or for the positioning of focused high-resolution grids within a coarser-scale domain. The methodology can also be used to make decisions about grid resolution, where the goal is

to ensure that there are an adequate number of grid cells available to represent each distinct region. Additionally, for model evaluation purposes, this methodology provides an analytical way to disaggregate the domain into smaller subdomains to focus on model performance for specific climate or weather features and/or to study local land–atmosphere feedback. Previous precipitation regionalization studies have relied on rain gauge data. For numerical modeling studies, however, satellite data are potentially more useful because climate modeling experiments often include an oceanic region. Additionally, for a domain composed of complex topography as the western United States, it is not clear a priori that rain gauge data are the best choice for a regionalization study. The findings that the methodology shows stability to the choice of precipitation dataset therefore extends the utility of the method to include domains where there is uncertainty in observations.

A detailed intercomparison of the nine precipitation datasets is described in Chapter 4. There we provide a multiple dataset analysis of precipitation variability and persistence as well as season- and location-specific assessments of observational precipitation data uncertainty for the western United States and for its five subregions.

CHAPTER 4

An Analysis of Precipitation Variability, Persistence, and Observational Data Uncertainty in the Western United States

4.1 Introduction

Precipitation data intercomparison studies aim to provide a measure of the current state of observational data quality and uncertainty by comparing different datasets or precipitation estimation algorithms. Xie and Arkin (1995) compared IR and microwave satellite precipitation estimates with rain gauge data and found general agreement for warm seasons over the tropical Pacific but poorer correspondence over land areas during cold seasons. Similar findings were reported by Ebert et al. (1996), who compared three precipitation estimation algorithms against rain gauge data and found good agreement in the tropical western Pacific and over Japan in summer but poor agreement over Europe in winter. Costa and Foley (1998) compared six precipitation datasets for the Amazon Basin and found general agreement among gauge-based datasets of long-term average climatology but noted important differences with interannual variability and significant bias in comparisons with reanalysis. Janowiak et al. (1998) compared a merged product (Huffman et al. 1997) with reanalysis and found strong large-scale similarity but noted poor agreement for some regional features. Gruber et al. (2000) compared two merged satellite–gauge products (Huffman et al. 1997; Xie and

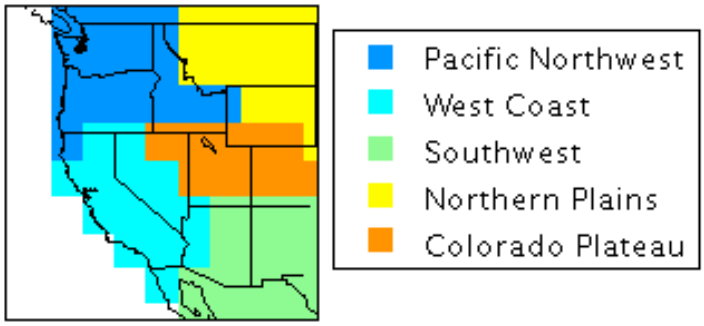

Figure 4.1: Composite regionalization for the western United States.

Arkin 1997) and found strong spatial and temporal correlation but also noted significant differences, which they attribute to differences in the use of atoll rain gauge data and aerodynamic gauge corrections. Gottschalck et al. (2005) compared several daily and subdaily datasets for the continental United States and found that, when used to force a land surface hydrology model, observational data differences produced large differences in some prognostic hydrology fields.

This chapter contributes to these efforts by providing an intercomparison for the western United States using nine state-of-the-art precipitation datasets. The research presented here differs from other intercomparison studies; this research includes the bulk of observational precipitation datasets available for moderately long-term studies in the western United States. Additionally, location-specific assessments of uncertainty are provided using the results of Chapter 3 to focus on specific regions within the domain of the western United States that have been found to have distinct precipitation climates. Data and methods are described in Section 4.2 and Section 4.3 discusses the effects of rescaling. The results are given in Sections 4.4 and 4.5. Section 4.4 focuses on general climate features including seasonality, interannual variability, and persistence. Section

4.5 provides a quantitative assessment of observational data uncertainty. A summary and conclusions are presented in section 4.6. This chapter is summarized in Guirguis and Avissar (2008b).

4.2 Data and Methods

The datasets used in this study are described in Section 3.2 and Table 1.1. These data are rescaled as described in Section 3.3. The time domain for this analysis is January 1986–July 2000, which is the period of maximum overlap between datasets. The results are stratified by season throughout. Winter is December–February (DJF), spring is March–May (MAM), summer is June–August (JJA), and fall is September–November (SON). Results are also stratified by location. Specifically, by using the results of Chapter 3, which identifies five unique precipitation climates in the western United States using principal component analysis. There, regionalization results are presented for the nine precipitation datasets. In this chapter, a composite of those regionalization results is used to subdivide the domain for spatial averaging and discussion. The five unique precipitation regions are referred to as the Pacific Northwest, West Coast, Southwest, Northern Plains, and Colorado Plateau Regions. They are mapped in Fig. 4.1.

a. Notation

Precipitation anomaly fields are calculated by subtracting the long-term climatic monthly mean from each monthly observation as

$$x'_m(i,k) = x_m(i,k) - \bar{x}_m(i), \tag{4.1}$$

where the prime represents an anomaly field, $x_m(i,k)$ represents monthly precipitation at position i and time k occurring in month m and $\bar{x}_m(i)$ is the long-term climatology for month m at position i. Brackets denote spatial averages as

$$\langle x \rangle = \frac{1}{p} \sum_{i=1}^{p} x, \tag{4.2}$$

where p is the number of 2.5-degree grid cells in the region being averaged over, which ranges from 10 for the Colorado Plateau region to 18 for the Southwest region. Overbars represent temporal averages as

$$\bar{x} = \frac{1}{t} \sum_{k=1}^{t} x, \tag{4.3}$$

where t is the temporal sample size. The standard deviation is represented by S_x and σ_x for spatial and temporal spread, respectively. The correlation coefficient for data pairs x and y is represented by r_{xy} and ρ_{xy} for correlation in the space and time domain, respectively.

It is convenient for some analyses to compare each dataset against a common reference dataset to identify data differences. For this, we use an ensemble dataset as the reference dataset, which is taken as the mean over all datasets as

$$E(i,k) = \frac{1}{N} \sum_{n=1}^{N} x_n(i,k), \tag{4.4}$$

where $E(i,k)$ is the ensemble at position i and time k, $x_n(i,k)$ is the n^{th} ensemble member at position i and time k, and N=9 is the number of ensemble members.

The PRISM dataset does not include data for northern Mexico. Therefore, for the Southwest region, any analyses involving PRISM data use only those grid cells located in the U.S.

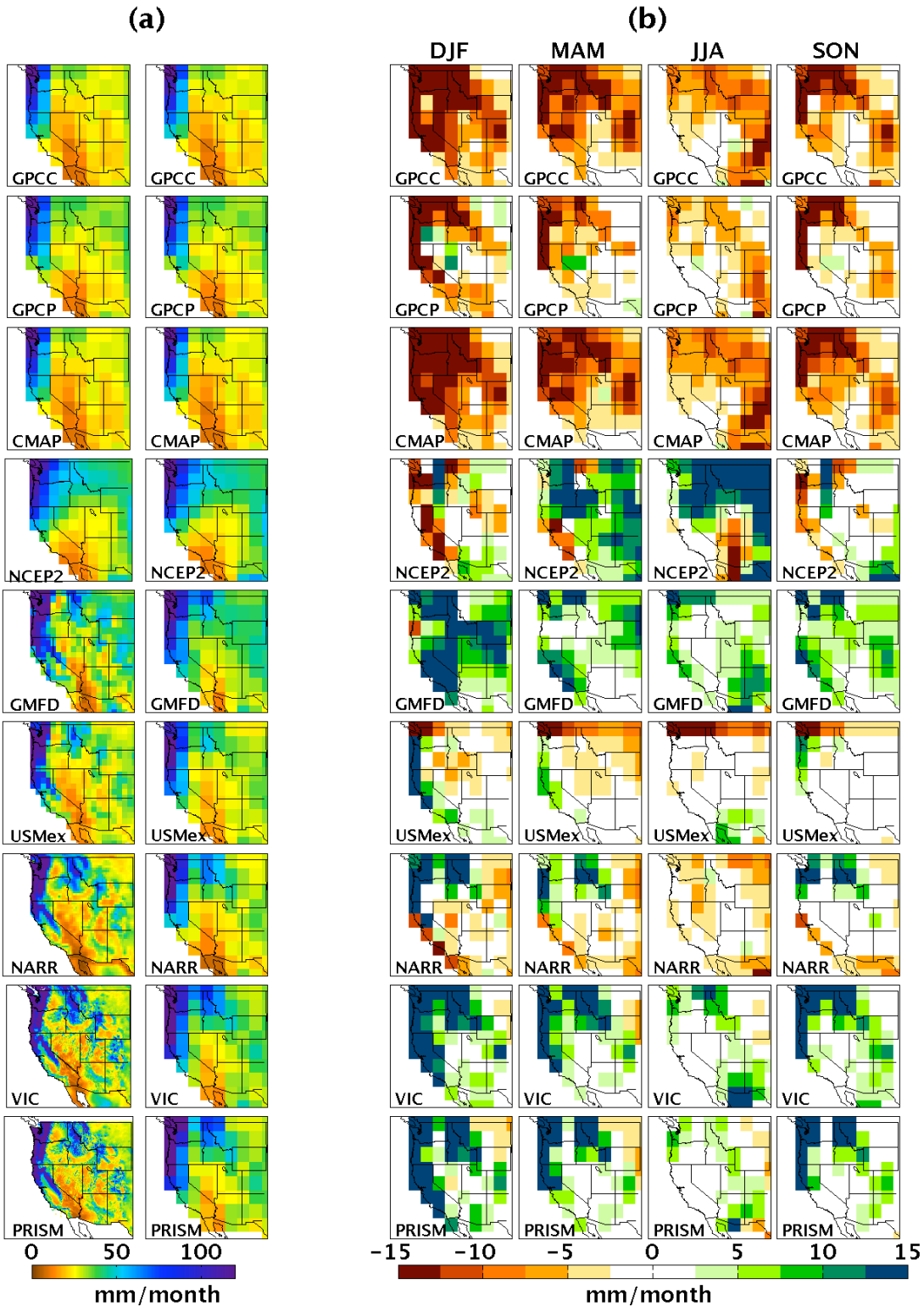

Figure 4.2: (a) Spatial distribution of long-term mean precipitation for the datasets at (left) their original resolution and (right) rescaled to 2.5° x 2.5°. (b) Bias as measured against the ensemble reference dataset for winter, spring, summer, and fall.

4.3 Effects of Rescaling

Figure 4.2a shows the spatial distribution of long-term mean precipitation for each dataset at its original spatial resolution and rescaled to a common 2.5° x 2.5° grid. Considering first the original data in Fig. 4.2a (left column), all datasets, regardless of spatial scale, represent the general features of the precipitation climate. The Pacific Northwest and northern California are shown to receive much more precipitation than the rest of the domain, and southern California, western Arizona, and northern Mexico are comparatively dry. All datasets represent the enhancement of precipitation along the windward side of the coastal ranges and Cascades in northern California, Oregon, and Washington where Pacific airstreams encounter orographic uplift. The higher-resolution data having spatial scales of 1° or smaller (NARR, USMex, GMFD, VIC, and PRISM) additionally represent orographic precipitation associated with the Sierra Nevada and coastal ranges in central and southern California. These datasets also represent the zone of precipitation enhancement that occurs on the windward side of the Rocky Mountains, which is observed as a roughly linear zone of high precipitation, running from the northwest part of the domain over Idaho to the southeastern part of the domain over Colorado. The very high-resolution data having spatial scales smaller than 1° (NARR, VIC, and PRISM) are further able to show the rain shadow effect that occurs over the San Joaquin Valley, which lies between the coastal ranges and the Sierra Nevada in central California. The effect of rescaling the high- and very high-resolution data to a 2.5° x 2.5° grid (Fig. 4.2a, right column) is to lose much detail of precipitation spatial variability. However, some important details are transferred to the larger grid. The NARR, USMex, GMFD, VIC, and PRISM still show the enhancement of precipitation over the Rocky

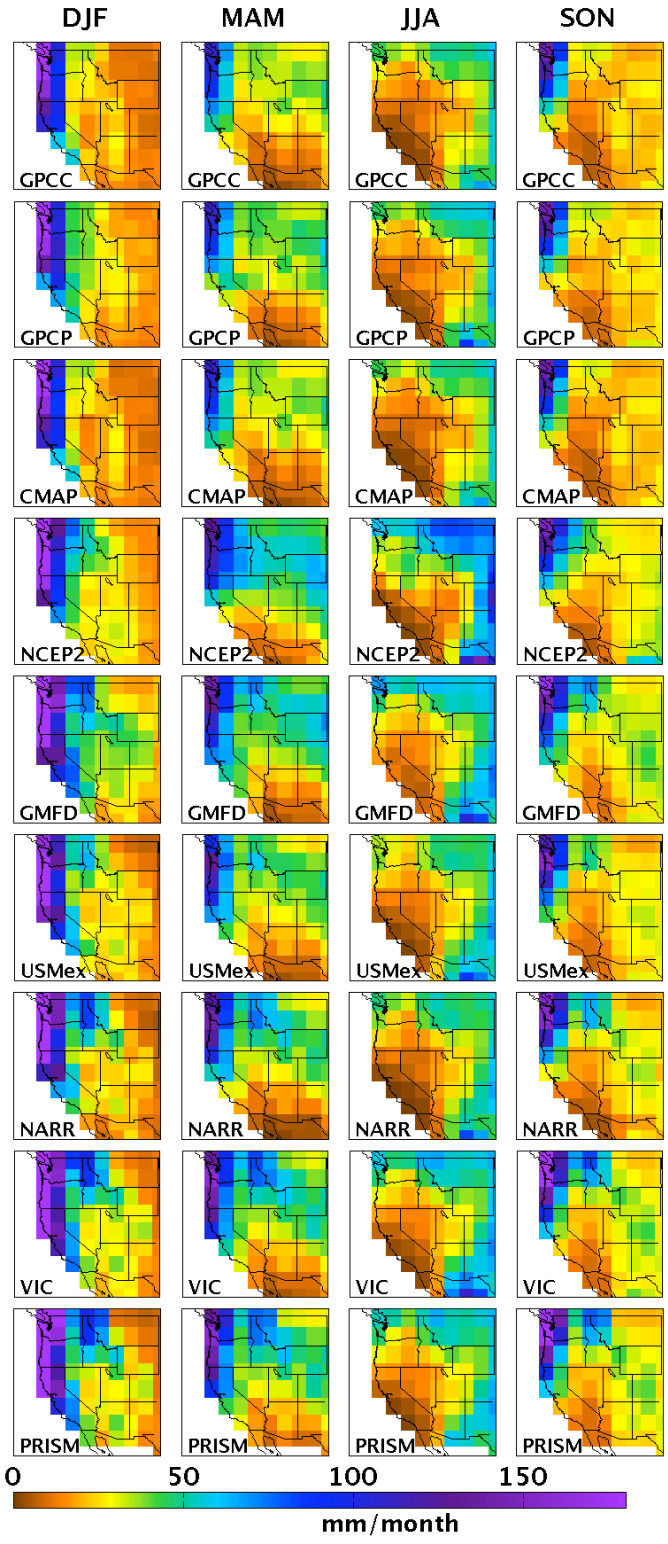

Figure 4.3: Long-term seasonal precipitation for (left to right) winter, spring, summer, and fall.

Mountains, which is not as clearly represented by other datasets. This orographic signal is muted according to the GMFD as a result of wetter conditions in the northern plains.

4.4 Data Intercomparison of Climatology

a. Seasonality

Long- term seasonal mean precipitation is given in Fig. 4.3. Here, all datasets capture the general seasonal features of the western United States. During winter, all data show the eastward wet–dry precipitation gradient associated with orographic uplift of Pacific air-streams along the western coast and the subsequent rain shadow effect to the east. During spring, wetter conditions are observed over the Colorado Plateau and Northern Plains regions, which are associated with the advection of Gulf of Mexico moisture into the deep continental interior, whereas drier conditions are seen along the northern coastal regions. During summer, the West Coast and Pacific Northwest regions are shown to be at the peak of their dry season, whereas precipitation increases occur in the Southwest region, Northern Plains, and eastern Colorado, which are associated with the North American monsoon (NAM) and Great Plains low-level jet. During fall, the domain transitions from the warm- to cold-season precipitation regime and, therefore, fall exhibits many of the same characteristics as winter but to a lesser degree.

Figure 4.2b gives the seasonal bias for each dataset n calculated against the ensemble reference dataset E. Specifically, Fig. 4.2b shows

$$\beta_n(i) = \frac{1}{t}\sum_{k=1}^{t}\left[x_n(i,k) - E(i,k)\right]. \tag{4.5}$$

The overall mean precipitation $\langle \bar{x}_n \rangle$ is shown in Table 4.1, and the mean bias $\langle \beta_n \rangle$ is given in Table 4.2. In Figs. 4.2b and 4.3 and Tables 4.1 and 4.2, the GPCP, CMAP, and

Table 4.1: Long-term mean precipitation (mm month^{-1}).

		GPCC	GPCP	CMAP	NCEP2	GMFD	USMex	NARR	VIC	PRISM
Domain	DJF	34.1	39.2	34.0	43.7	54.6	45.5	46.3	57.3	58.8
	MAM	31.2	35.3	31.0	46.3	44.4	37.1	38.8	45.3	46.4
	JJA	28.9	31.7	28.4	45.8	40.2	34.1	32.2	38.9	35.5
	SON	27.3	29.7	26.8	34.9	38.7	32.9	33.8	40.6	40.6
Pacific Northwest	DJF	78.5	87.0	78.4	99.2	115.3	101.0	115.5	134.1	134.8
	MAM	54.9	58.6	54.6	79.3	75.9	65.9	77.4	87.0	84.9
	JJA	23.9	26.3	23.8	42.9	37.2	26.7	28.7	36.6	34.4
	SON	53.1	56.0	52.1	65.5	75.2	64.3	73.4	86.5	84.8
West Coast	DJF	43.9	48.1	44.2	50.5	70.8	62.2	53.8	74.4	71.4
	MAM	24.0	28.4	24.0	33.1	39.4	33.7	29.5	38.7	37.9
	JJA	6.9	8.3	6.7	11.3	12.6	8.7	6.7	9.5	9.7
	SON	15.0	17.4	14.9	18.4	25.5	19.8	17.0	22.9	22.8
Southwest	DJF	15.6	17.0	15.4	23.0	23.6	21.0	17.0	22.1	22.2
	MAM	13.3	15.5	12.9	23.9	16.7	15.5	12.0	16.9	19.8
	JJA	39.7	42.6	38.6	56.7	55.1	51.0	44.4	56.2	48.3
	SON	21.5	22.5	21.1	32.2	27.3	25.8	22.0	28.4	27.4
Northern	DJF	12.9	18.3	12.7	19.7	25.8	16.6	18.3	22.9	22.8
	MAM	36.3	42.4	36.0	51.8	52.2	38.7	42.5	46.4	45.0
	JJA	45.3	51.2	45.1	73.8	55.9	48.1	47.9	53.7	52.0
	SON	23.5	27.1	23.2	28.9	32.0	26.0	28.1	31.6	31.5
Colorado	DJF	17.0	24.3	16.5	22.2	37.4	23.6	23.5	29.6	28.9
	MAM	30.7	35.7	30.9	47.7	43.0	35.2	37.2	41.4	39.9
	JJA	24.8	25.9	24.2	39.7	34.7	30.7	29.0	31.9	32.0
	SON	21.4	23.9	20.9	25.8	32.8	26.8	27.3	31.0	30.3

GPCC data show drier conditions relative to other datasets for all seasons and regions. Conversely, the GMFD, VIC, and PRISM generally show wetter conditions. Spatially, dataset differences are greatest along the western coast and along the path of the Rocky Mountains. This may be related to the original spatial scale of the datasets, owing to the ability of the higher resolution data to represent orographic precipitation enhancement over the Sierra Nevada in California and over the western flanks of the Rocky Mountains (cf. Fig. 4.2a). However, sampling error associated with poor gauge coverage by the

Table 4.2: Same as Table 4.1 but for the mean bias as measured against the reference dataset (mm month-1).

		GPCC	GPCP	CMA	NCEP	GMF	USMe	NAR	VIC	PRIS
Domain	DJF	-11.3	-6.2	-11.4	-1.7	9.2	0.1	0.8	11.9	11.7
	MAM	-8.0	-3.9	-8.2	7.1	5.2	-2.1	-0.3	6.1	5.3
	JJA	-6.5	-3.6	-6.9	10.4	4.9	-1.3	-3.2	3.6	1.3
	SON	-6.4	-3.9	-6.8	1.3	5.1	-0.7	0.2	7.0	6.3
Pacific Northwest	DJF	-24.3	-15.8	-24.4	-3.6	12.5	-1.8	12.8	31.3	32.0
	MAM	-14.8	-11.1	-15.1	9.6	6.2	-3.8	7.6	17.3	15.2
	JJA	-7.3	-5.0	-7.4	11.6	6.0	-4.6	-2.5	5.3	3.2
	SON	-13.5	-10.6	-14.5	-1.1	8.5	-2.3	6.7	19.9	18.2
West Coast	DJF	-13.4	-9.1	-13.1	-6.8	13.5	4.9	-3.5	17.1	14.1
	MAM	-8.1	-3.8	-8.1	0.9	7.2	1.5	-2.6	6.6	5.7
	JJA	-2.4	-1.0	-2.6	2.1	3.3	-0.6	-2.6	0.2	0.5
	SON	-4.6	-2.2	-4.6	-1.2	5.9	0.2	-2.6	3.3	3.2
Southwest	DJF	-4.1	-2.7	-4.3	3.3	3.9	1.3	-2.7	2.4	2.0
	MAM	-2.6	-0.4	-3.0	8.0	0.8	-0.4	-3.9	1.0	1.0
	JJA	-8.5	-5.5	-9.5	8.5	7.0	2.9	-3.7	8.1	2.4
	SON	-3.7	-2.7	-4.0	7.1	2.2	0.6	-3.2	3.2	1.7
Northern Plains	DJF	-6.3	-0.9	-6.5	0.5	6.6	-2.5	-0.9	3.7	3.6
	MAM	-7.8	-1.7	-8.1	7.6	8.0	-5.4	-1.6	2.3	0.9
	JJA	-7.5	-1.6	-7.7	21.0	3.1	-4.7	-4.9	0.9	-0.8
	SON	-4.5	-0.9	-4.8	0.9	3.9	-2.0	0	3.6	3.5
Colorado Plateau	DJF	-8.6	-1.4	-9.1	-3.5	11.7	-2.1	-2.1	4.0	3.2
	MAM	-7.4	-2.4	-7.2	9.5	4.9	-3.0	-1.0	3.3	1.8
	JJA	-5.8	-4.7	-6.5	9.1	4.0	0	-1.6	1.2	1.4
	SON	-5.5	-3.0	-6.0	-1.1	5.9	0	0.4	4.1	3.5

GPCC dataset (also used in the CMAP and GPCP) may also be a factor. The orographic adjustment employed in the development of the VIC and PRISM would contribute to the higher precipitation estimates by these datasets over these parts. Also notable in Figs. 4.2b and 4.3 are differences between the GPCP and CMAP over eastern Oregon, Nevada, and the northern plains, especially in winter, where larger precipitation amounts are observed for the GPCP compared to the CMAP. This is likely a result of the systematic gauge corrections used in the development of the GPCP data, whereas the CMAP gauge

data are uncorrected, and most systematic errors occur as undercatch (Fuchs et al. 2007). The GMFD data and especially the NCEP2 data show wetter conditions over the Northern Plains in spring and summer relative to the other data, which may be related to convection parameterizations used in the derivation of NCEP2 precipitation fields. The GMFD data show wetter conditions over the Colorado Plateau region and the state of Nevada in winter than is observed by any other dataset. During summer, the GMFD, VIC, USMex, and especially NCEP2 data show a stronger monsoon signal in the Southwest region, whereas the NAM signal is much more subtle in the CMAP and GPCC data (Fig. 4.3). For the NCEP2, the spatial extent of the monsoon signal is restricted westward, as observed by comparatively dry conditions over the western part of northern Mexico and extending northward into Arizona and Utah (Fig. 4.2b). Another important difference observed in Figs. 4.2 and 4.3 is that, whereas (Fig. 4.3) the NARR data generally show a similar spatial pattern of precipitation as other high-resolution data, much drier conditions are seen along the California coast during winter and to a lesser extent in spring and fall (Fig. 4.2b). This is also observed for the NCEP2, suggesting that model parameterizations used in the data assimilation process may be an issue here for both reanalysis products.

To estimate overall bias among datasets and its spatial distribution, we consider the spread among ensemble members (datasets) and the corresponding coefficient of variation (as in Yang and Arritt 2002). The spread is represented by the ensemble standard deviation as

$$S_E(i,k) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x_n(i,k) - E(i,k)\right)^2}, \tag{4.6}$$

and the corresponding coefficient of variation is given by

$$CV(i,k) = \sigma_E(i,k) / E(i,k). \quad (4.7)$$

Figure 4.4 gives temporal averages of the ensemble mean, spread, and coefficient of variation. The ensemble spread (Fig. 4.4b) is shown to be large for the Pacific Northwest during spring, fall, and particularly winter; the Rocky Mountains in winter, spring and summer; coastal California during winter; and the Great Plains and Southwest in summer. In Fig. 4.4c, the coefficient of variation is large for dry regions where the precipitation frequency distributions are heavily skewed toward zero, and the mean precipitation [denominator in (4.7)] is small. The spread among datasets for California reaches up to 55% of the mean during winter, 90% during spring and fall, and up to 137% for some coastal areas during summer. This implies that data uncertainty for the West Coast region can be larger than the mean precipitation received there. This is similarly observed for the Southwest region but to a lesser degree. Here, the spread can reach 93% of the mean during spring and 75% during winter, summer, and fall. The spread for the Pacific Northwest, whereas generally larger in magnitude (Fig. 4.4b), represents a smaller ratio of the mean (Fig. 4.4c), as here the coefficient of variation ranges from 15%–70% depending on season and location. Smaller spreads are observed for the Northern Plains and Colorado Plateau regions. For the Northern Plains, the ensemble standard deviation ranges from 5 to 19 mm month^{-1}, which corresponds to a coefficient of variation of 22%–58%. The spread for the Colorado Plateau region ranges from 5 to 21 mm month^{-1}, which represents 21%–51% of the mean.

Empirical probability densities of precipitation values are shown in Fig. 4.5, stratified by season and region. Each data vector represented in Fig. 4.5 is of length t∗p, where t is the temporal sample size (t=42–45 depending on season), and p is the number

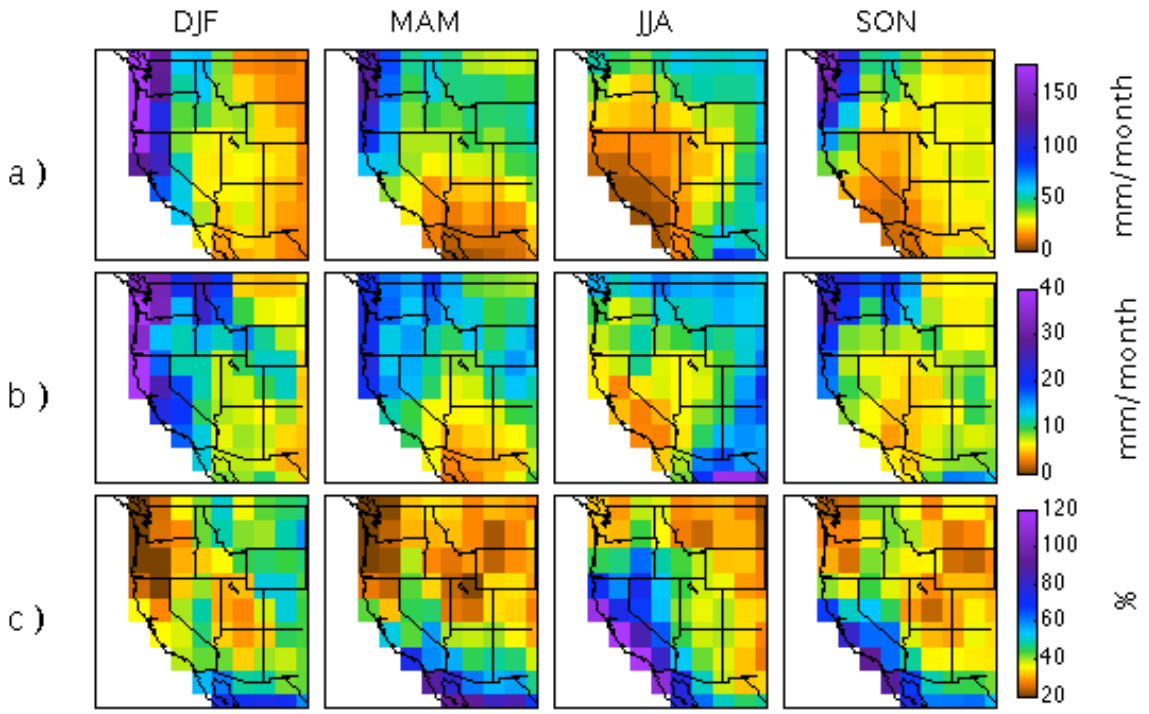

Figure 4.4: (a) Ensemble mean, (b) standard deviation, and (c) coefficient of variation.

of spatial grid cells (p=10–18 depending on region). We apply the two-sample Kolmogorov–Smirnov (K–S) test to compare the distributions of precipitation between each pair of datasets. The null hypothesis for the K–S test is that data vectors x and y are drawn from the same continuous distribution, against the alternative that they are from different distributions. The test statistic is $\max\left|F_x(j) - F_y(j)\right|$, where F_x and F_y are the empirical cumulative density functions (CDFs) from sample vectors x and y. The null hypothesis is rejected if the test statistic exceeds a critical value.

Figure 4.6 gives the results of the K–S test. The shaded parts indicate that differences between distributions are significant at the 95% level. Here, the null hypothesis that x and y are drawn from the same population is rejected more often than not, implying a significant degree of data uncertainty. Specifically, the null is rejected

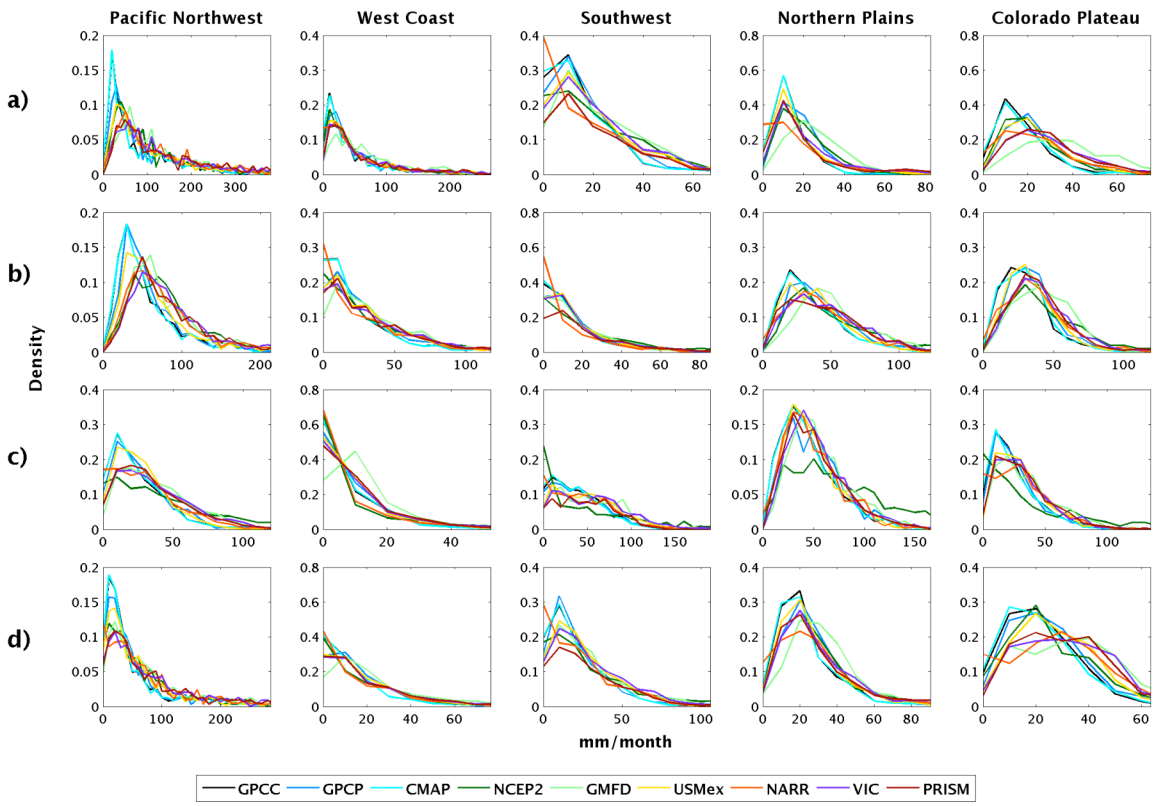

Figure 4.5: Empirical cumulative density functions for (a) winter, (b) spring, (c) summer, and (d) fall.

579 times out of $5 * 4 \sum_{n=1}^{N-1} n = 720$ comparisons, where 5 is the number of regions, 4 is the number of seasons, and n is the number of datasets. Figure 4.7 gives the results of the K–S test for anomaly fields in which the climatology has been removed, as in (4.1). Here, the null hypothesis is rejected less frequently, implying that seasonal anomaly distributions are more similar. However, uncertainty remains an issue, as significant differences are observed for 39% of the dataset comparisons (283 out of 720). In Fig. 4.7, the greatest similarity is observed during summer and fall for the Southwest, Northern Plains, and Colorado Plateau regions and in spring for the Northern Plains.

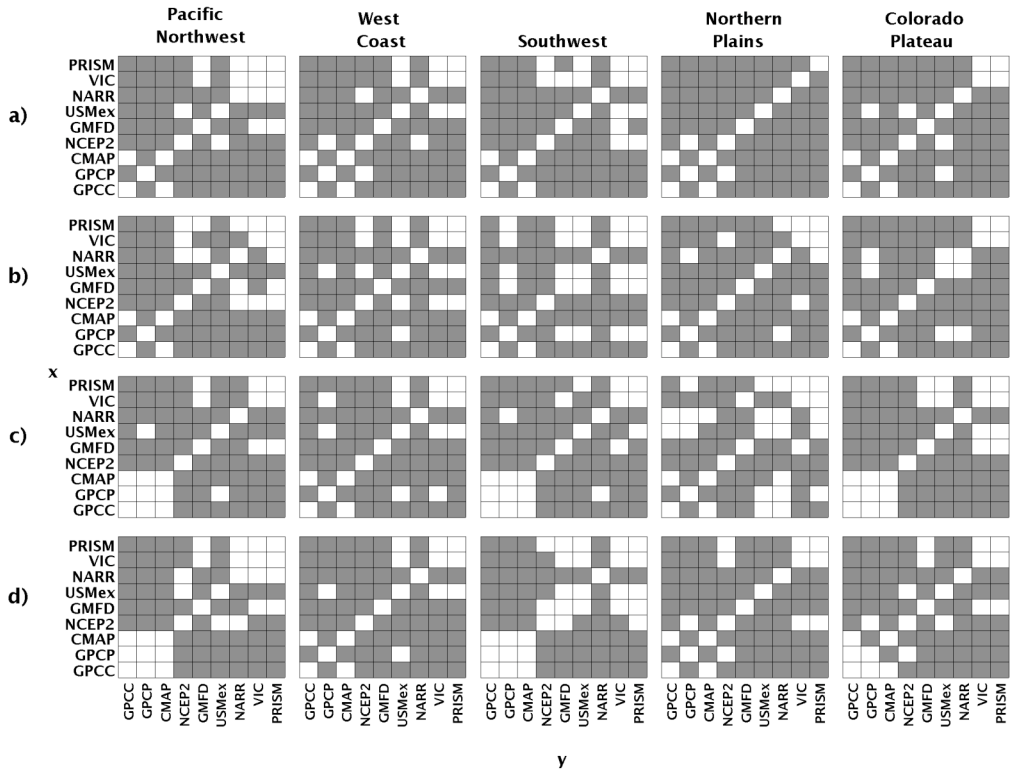

Figure 4.6: Results of the two-sample K–S test for (a) winter, (b) spring, (c) summer, and (d) fall. Shading indicates that the difference the distributions of datasets x and y is significant at the 95% level.

b. Interannual Variability

Figure 4.8a shows spatially averaged annual anomaly fields in which December–November is used as the averaging period to maintain seasonal congruence (e.g., calendar year 1990 represents the period December 1989–November 1990). Also shown in Fig. 4.8a is the phase and strength of El Niño–Southern Oscillation (ENSO) according to the monthly multivariate ENSO index. There is strong agreement between data sources with respect to interannual peaks and troughs. All data show a Pacific Northwest dry period between 1986 and 1994, after which time the region entered a wet phase that extended through 1999. Extremes occurring in the Pacific Northwest are the negative anomalies in 1987 and 1992 coinciding with warm phases of ENSO and the positive anomalies during

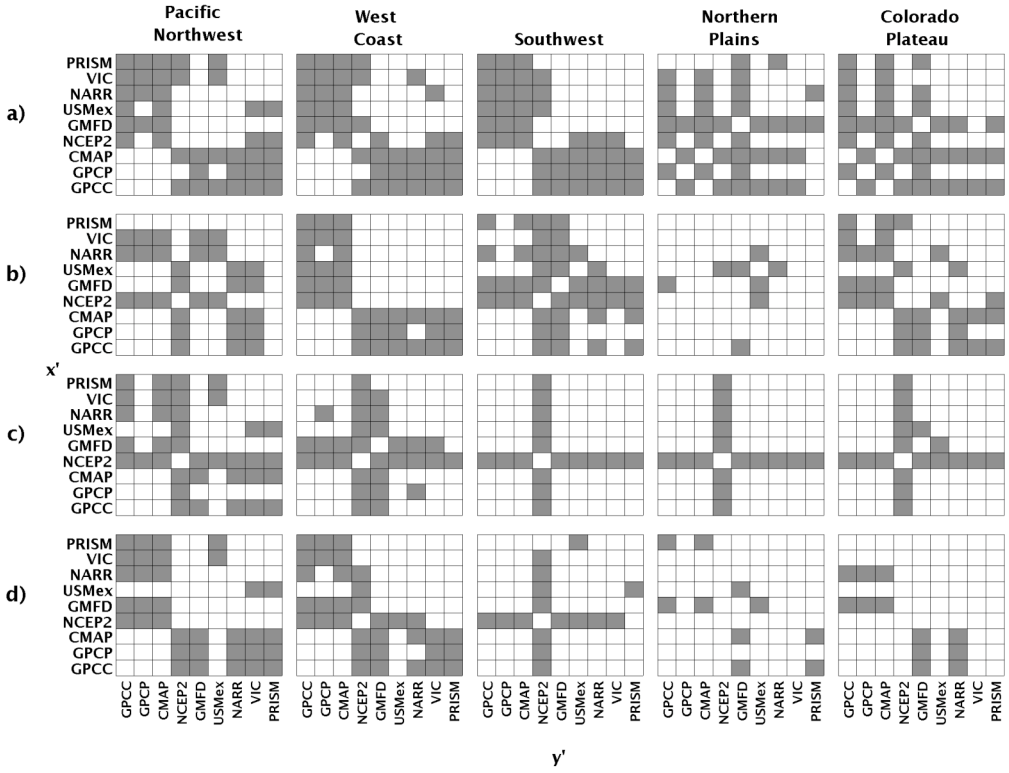

Figure 4.7: As in Fig. 4.6 but for precipitation anomaly fields x′ and y′.

1995–97, which occurred during a cold ENSO period. The West Coast region shows a similar pattern of dry conditions during the first part of the record and wetter conditions later. Strong positive anomalies are observed for the West Coast region in 1993, 1995, and 1998, all of which coincide with the warm phase of ENSO. The Southwest region shows an opposite anomaly pattern as compared to the Pacific Northwest and West Coast, with wet conditions dominating prior to 1994 and drier conditions observed thereafter. Precipitation extremes for the Southwest include the relatively dry 1989 and the wet 1992, which occur during cold and warm phases of ENSO, respectively. The Northern Plains and Colorado Plateau regions demonstrate some evidence of a linear positive trend, with regularly spaced cycles of peaks and troughs that do not deviate

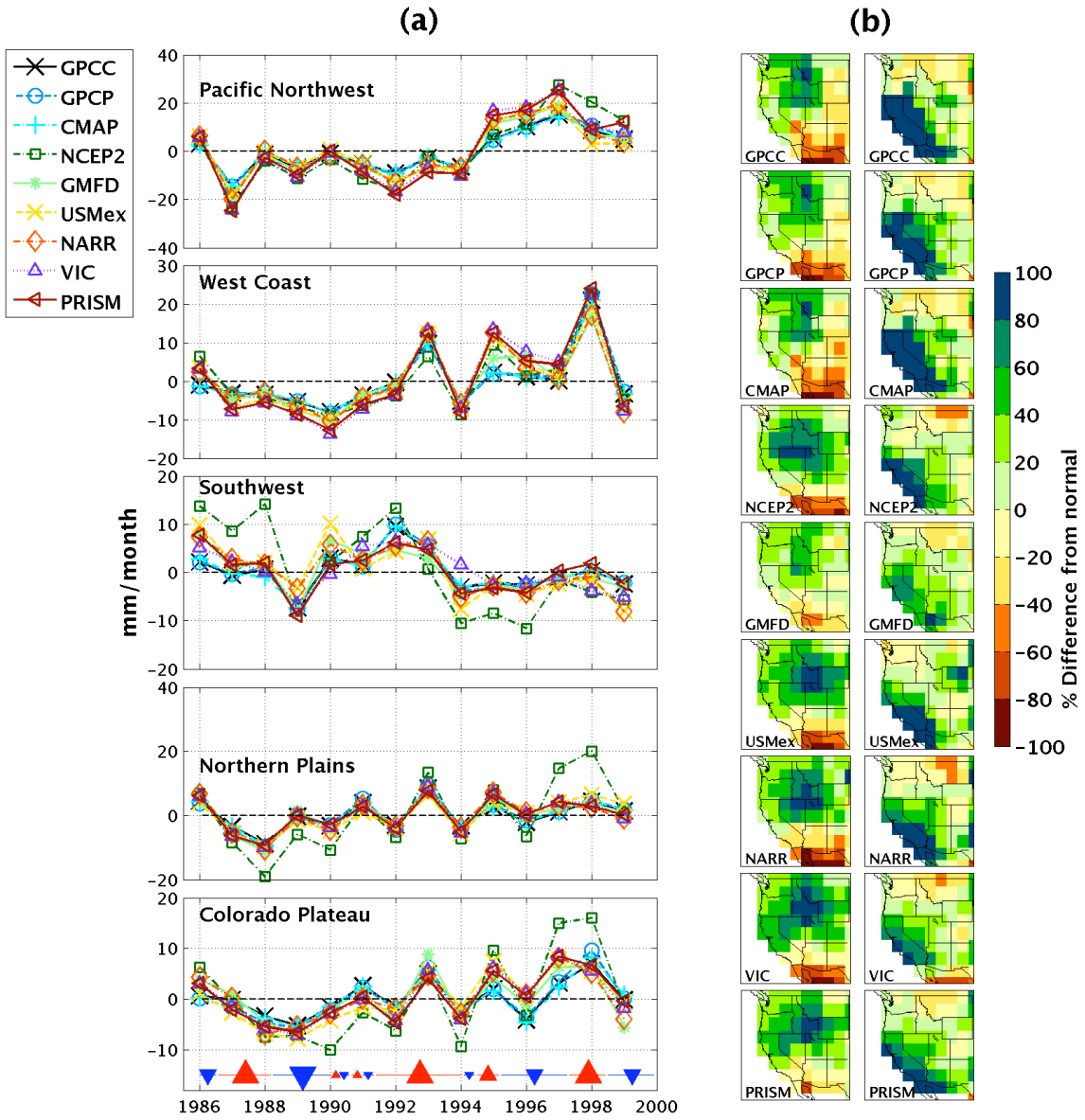

Figure 4.8: (a) Annual precipitation anomaly fields and the phase and strength of ENSO according to the monthly multivariate ENSO index. The warm (cold) phases of ENSO are represented by red and upward (blue and downward) triangular markers. Markers are centered at the midpoint of a warm or cold period and their size is weighted by the magnitude of the ENSO anomaly. (b) Spatial distribution of precipitation anomalies for the winters of 1996/97 and 1997/98. The anomalies were calculated by subtracting the long-term DJF mean.

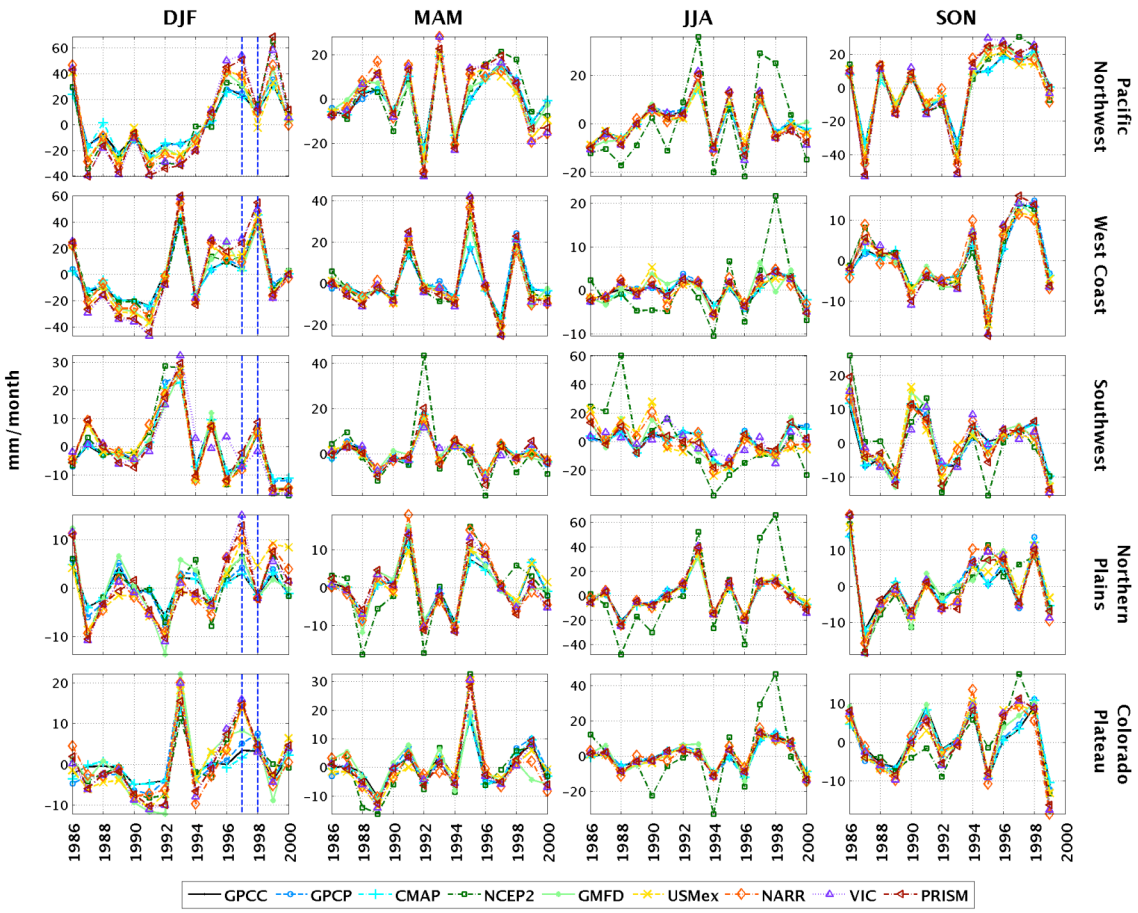

Figure 4.9: Spatially averaged seasonal precipitation anomaly fields. The vertical dashed lines in the DJF time series correspond to the anomaly fields shown in Fig. 4.8b.

dramatically from normal. The observed peaks and troughs appear to correspond fairly consistently with warm and cold phases of ENSO, respectively.

The data series for seasonal anomalies are shown in Fig. 4.9. Strong similarity is observed among data for all regions and seasons with respect to interannual highs and lows. The NCEP2 data are shown to exaggerate some anomalies relative to the other datasets, especially in summer. The difference between data in their estimation of $\langle x' \rangle$ in Fig. 4.9 (measured on the y axis as the difference between two time series) ranges from 2.3 to 35.8 mm month^{-1} for the Pacific Northwest, 1.0–25.7 mm month^{-1} for the West

Coast, 2.8–57.6 mm month^{-1} for the Southwest, 1.8–58.0 mm month^{-1} for the Northern Plains, and 0.9–39.1 mm month^{-1} for the Colorado Plateau region. For the Southwest, Northern Plains, and Colorado Plateau regions, the maximum data difference drops to 26.4, 10.5, 14.5 mm month^{-1}, respectively, if the NCEP2 data are not considered. Figure 4.10 shows the seasonal time series in which the climatology has not been removed. Here, the difference between data in their estimation of $\langle x \rangle$ ranges from 10.8 to 92.3 mm month^{-1} for the Pacific Northwest, 2.7–55.6 mm month^{-1} for the West Coast, 4.0–73.4 mm month^{-1} for the Southwest, 6.5–83.7 mm month^{-1} for the Northern Plains and 6.0–50.1 mm month^{-1} for the Colorado Plateau region. If the NCEP2 data are not considered, the upper limit is reduced to 34.1, 21.5, and 29.2 mm month^{-1} for the Southwest, Northern Plains and Colorado Plateau regions, respectively.

The winters of 1996/97 and 1997/98 are known for being particularly eventful for the western United States, and we consider them in more detail. The 1996/97 winter season brought heavy and extensive flooding to the Pacific Northwest, Nevada, and California. The Sierra Nevada region was especially affected, and December precipitation in parts of Idaho was recorded at more than 300% of normal (Lott et al. 1997). The 1997/98 winter season saw record-breaking warm and wet conditions across the United States. The western United States was particularly hard hit during February, which brought 4 weeks of near-continuous storm activity to California; and parts of the Southwest and Northern Plains regions were also affected (Ross et al. 1998). In Fig. 4.9, the 1996/97 winter season is wetter for the Pacific Northwest, Northern Plains, and Colorado Plateau, and the 1997/98 winter season is wetter for the West Coast and Southwest. Also in Fig. 4.9, the spread among data are relatively large for the 1996/97

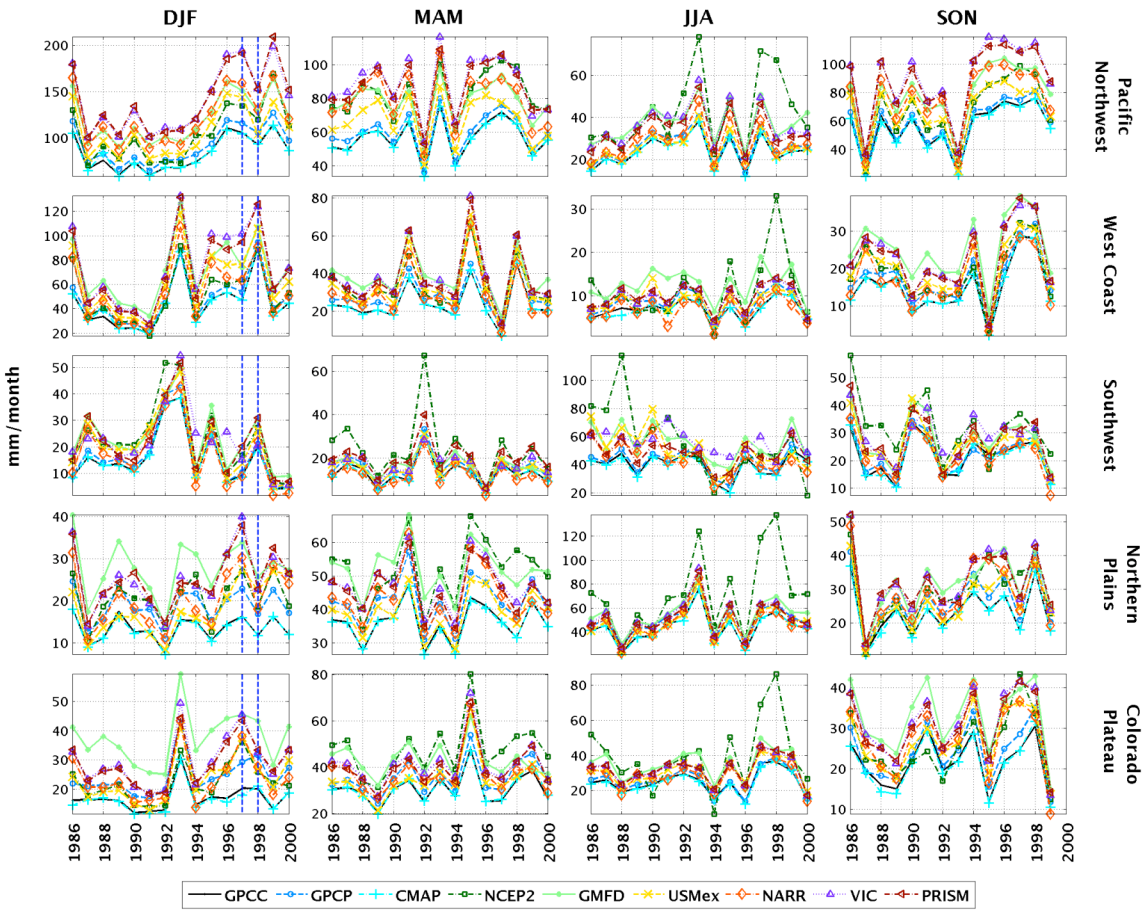

Figure 4.10: As in Fig. 4.9 but without removing the climatology.

winter season (6.1–33.0 mm month^{-1}), whereas a smaller spread is observed for the 1997/98 winter season (2.4–18.0 mm month^{-1}).

The spatial distributions of the 1996/97 and 1997/98 winter anomalies according to each dataset are shown in Fig. 4.8b. All datasets show the 1996/97 winter season as having widespread wet conditions over the Pacific Northwest, western Montana, Nevada, and California. However, there are some important discrepancies between data sources. For example, eastern Montana, Wyoming, and Colorado are shown to be mildly dry according to the GPCP, CMAP, and GPCC data, whereas the other datasets suggest wet conditions.

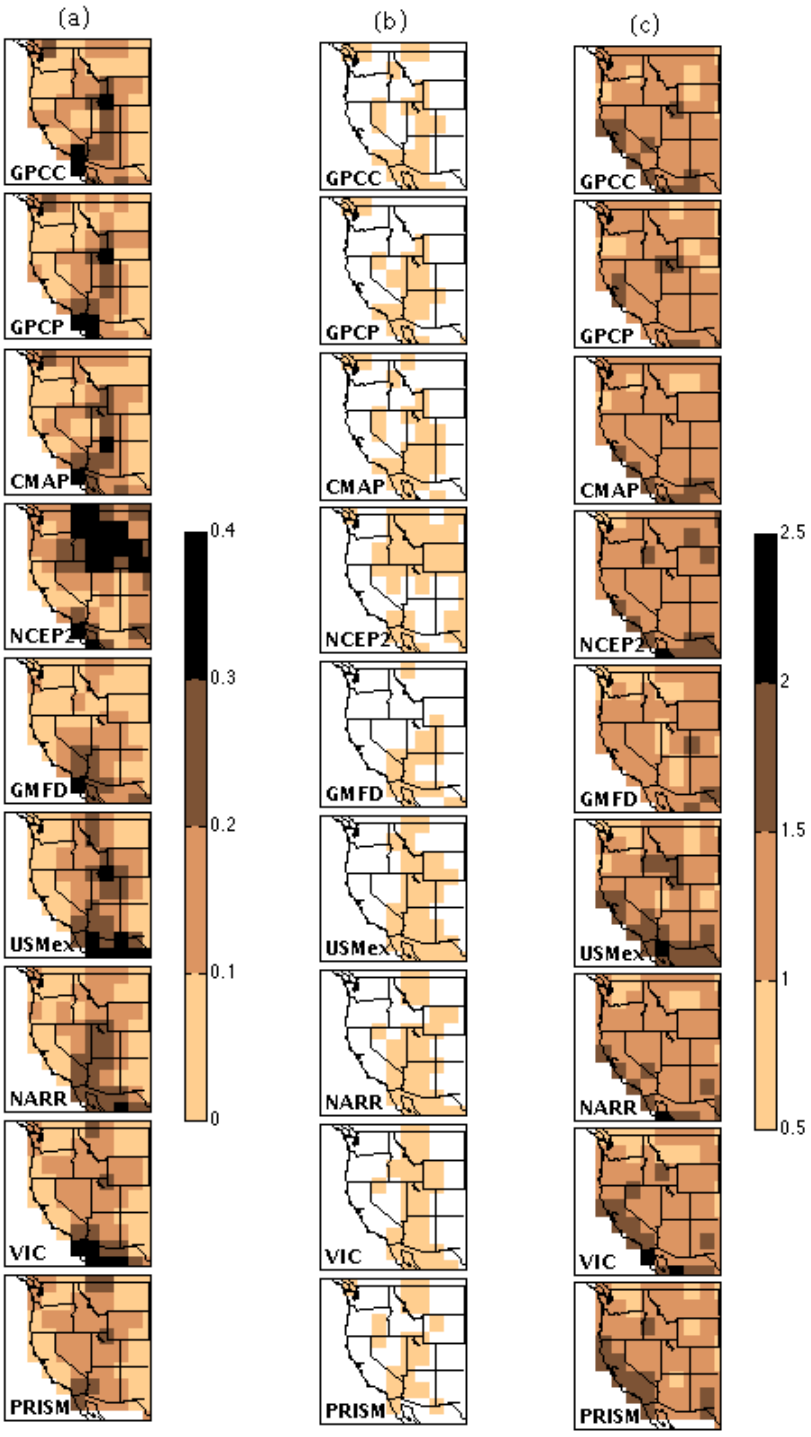

Figure 4.11: (a) One-month lag autocorrelation coefficient, (b) correlations significant at the 95% level shown as shaded, and (c) characteristic persistence time scale in months.

There is also disagreement over the severity of the anomaly over northern Nevada and Idaho during 1996/97. The NARR, USMex, and VIC data suggest dramatically wet conditions over Idaho as compared to the other data, and the GPCC, GPCP, and CMAP data show drier conditions over northern Nevada. For 1997/98, all the datasets show anomalously wet conditions over California. However, the magnitude of the anomaly is highly varied. The CMAP and GPCC data show extremely widespread and wet conditions covering the entire state of California and western Nevada, whereas the PRISM, VIC, USMex, and NARR data show intensely wet conditions only over the southern and coastal parts of California.

b. Persistence

For our analysis of persistence, we calculate the one-month lag autocorrelation coefficient and estimate the characteristic time scale of persistence. The autocorrelation is calculated as

$$\alpha(\tau) = \sum_{k=1}^{t-\tau} (x_k - \bar{x})(x_{k+\tau} - \bar{x}) / \sigma_x^{\,2}, \tag{4.8}$$

where τ is the lag in months, x is precipitation, t is the number of months in the time series, and σ_x is the standard deviation of x. The data are detrended prior to the analysis to remove any linear tendency in the data series.

The characteristic timescale of persistence is estimated using a count of consecutive, like-signed monthly anomalies as in Liu and Avissar (1999). A run of length l represents the time period after an anomaly is observed in which $x' = x - \bar{x}$

maintains the same sign. Beginning with the first anomaly $x'_{k=k1}$, we count the number of consecutive months in which x' maintains the same sign and this value is denoted l_1. The next run (which is opposite in sign) begins with $x'_{k=k2}$ and the length of the second run is denoted l_2. This process repeats until the end of the time series, however, the first and last runs are omitted to avoid truncation at the end points. An anomaly that maintains its sign for only one month demonstrates no persistence and, therefore, l=0. The characteristic timescale of persistence is estimated as the average length-scale according to

$$L = \frac{1}{\eta}\sum_{k=1}^{n} l_k , \tag{4.9}$$

where η is the number of persistent runs in the time series. We also consider the probability distribution of l_k for each dataset and region using the Kolmogorov-Smirnov test to determine the degree of similarity between datasets in their representation of persistence.

The one-month lag autocorrelation coefficients are shown in Fig. 4.11a. The coefficients range from 0.08 to 0.41 overall and are shown to vary geographically and by dataset. Grid cells with statistically significant correlations are shown in Fig. 4.11b. The correlations are statistically significant at the 95% level (critical value of 0.148 for 173 degrees of freedom) only in the intermountain region, northern Mexico, and southern California. The VIC, USMex, and NARR data give larger coefficients over northern Mexico relative to the other data. The NCEP2 data show larger coefficients for grid cells spanning Idaho, western Montana, and Wyoming, which may be related to moisture recycling efficiency in the spectral model. Figure 4.11c shows the estimated characteristic persistence time scale. The time scale L ranges from 0.68 to 2.2 months

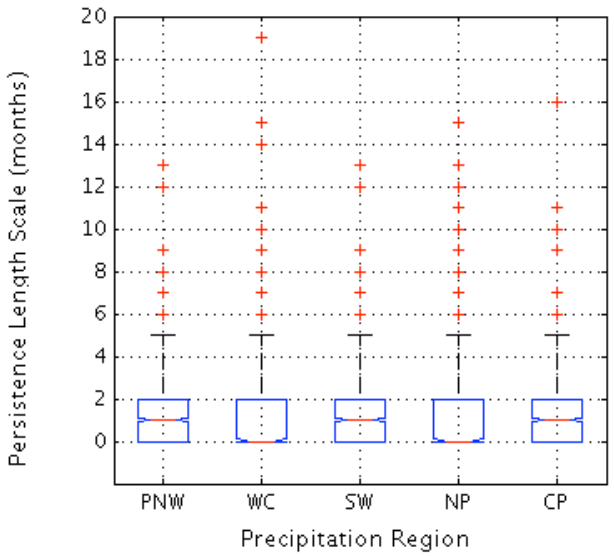

Figure 4.12: The distribution of persistence length scales for each precipitation region. There is overlap of outliers such that each plus sign (+) represents more than one length scale occurrence.

overall, with most of the domain showing persistence on the order of 1–1.5 months. Longer time scales of 1.5–2 months occur for parts of California and northern Mexico. The USMex and NCEP2 data show longer time scales for northern Mexico, and the PRISM, VIC, and USMex data demonstrate longer time scales over a larger portion of California.

Figure 4.12 shows the distribution of persistence length scales for each precipitation region. Here, the ensemble reference dataset is used as the precipitation data, from which we calculate $l_1, l_2, \ldots, l_\eta$ for each grid cell, and these persistence length scale time series are then aggregated over each precipitation region to give the distributions shown in Fig. 4.12. The median length scale for the Pacific Northwest, Southwest, and Colorado Plateau regions is one month, and the median length scale is zero for the West Coast and Northern Plains. The spread as represented by the interquartile range (IQR) is two months, in which the 25th percentile is zero, and the 75th

percentile is two months for all regions. Outliers are shown as those larger than 1.5 IQR, or l_k>5 months. The majority of outliers correspond to negative anomalies (not shown), suggesting that droughts are more likely to persist for long periods than for periods of precipitation excess. The proportion of positive-to-negative anomalies exceeding five months is (displayed as positive:negative) 12:10, 12:26, 23:53, 6:21,and 6:17 for the Pacific Northwest, West Coast, Southwest, Northern Plains, and Colorado Plateau regions, respectively.

The degree of similarity between datasets in their representation of persistence is determined by compiling length scale distributions as described above for each of the nine datasets. Then we apply the two-sample K–S test to compare the length scale distributions among datasets. The null hypothesis that two samples comprised of l_k from two different precipitation datasets are drawn from the same distribution is rejected only 4 times out of $5 * \sum_{n=1}^{N-1} n = 180$ comparisons, where 5 is the number of regions and N is the number of datasets. Statistically significant differences occur for comparisons of the GPCP data against the USMex, NARR, and VIC data for the West Coast region, and between the PRISM and NCEP2 data for the Northern Plains region. For the West Coast, there are slight differences between the cumulative densities of l_k for the GPCC, GPCP, and CMAP data as compared to the USMex, NARR, VIC, and PRISM data. Specifically, the GPCC, GPCP, and CMAP data show higher densities for l_k in the range of 0–2 months and lower densities for l_k in the range of 2–4 months relative to the USMex, NARR, VIC, and PRISM data; this difference is large enough to achieve statistical significance for comparisons between the GPCP data with the USMex, NARR, and VIC data. For the Northern Plains, the NCEP2 data show higher densities for l_k in the range of

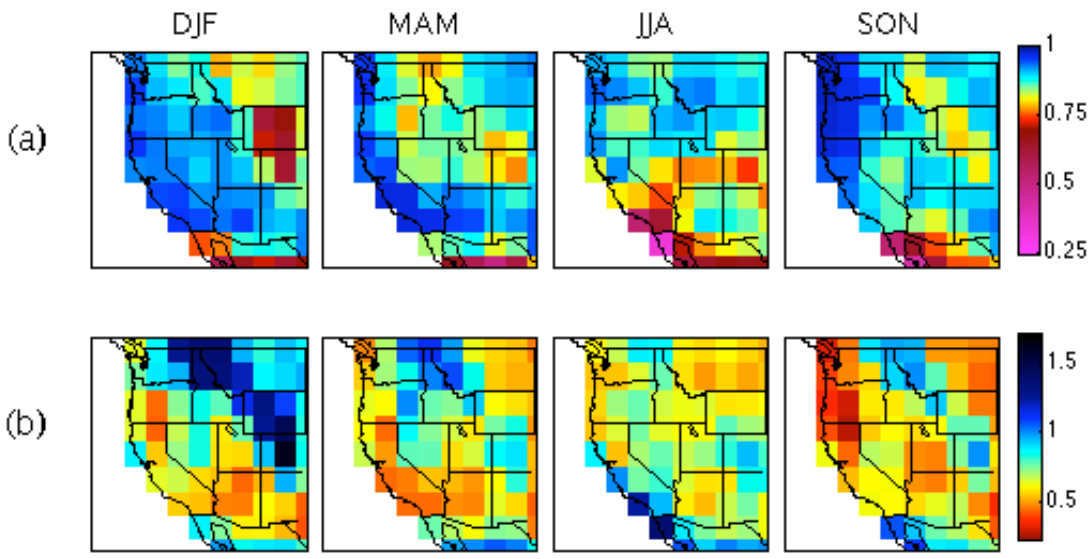

Figure 4.13: (a) Average correlation coefficient and (b) NRMSE between datasets.

5–10 months and lower densities for l_k in the range of 0–2 months; this difference is large enough to reach significance for the comparison between the NCEP2 and PRISM data.

4.5. Observational Data Uncertainty

a. Spatial Distribution of Uncertainty

Figure 4.13 is used to assess the spatial distribution of observational data uncertainty. For each grid cell (i), the correlation coefficient and normalized root-mean-square error (NRMSE) are calculated, in which the NRMSE normalization factor is the standard deviation. These quantities are calculated for each pair of datasets according to

$$\rho_{xy}(i) = \sum_{k=1}^{t} (x(i,k) - \bar{x}(i))(y(i,k) - \bar{y}(i)) / \sigma_{x(i)}\sigma_{y(i)}, \tag{4.10}$$

and

$$NRMSE_{xy}(i) = \frac{1}{\sigma(i)}\left[\frac{1}{t}\sum_{k=1}^{t}\left[x(i,k) - y(i,k)\right]^2\right]^{1/2}, \tag{4.11}$$

Table 4.3: Range of values by season and precipitation region for the correlation coefficient and NRMSE presented in Figure 13.

	Season	ρ	NRMSE
Domain	DJF	0.59-0.96	0.41-1.62
	MAM	0.50-0.97	0.43-1.13
	JJA	0.34-0.94	0.50-1.37
	SON	0.43-0.98	0.28-1.11
Pacific Northwest	DJF	0.85-0.96	0.43-1.37
	MAM	0.78-0.97	0.47-1.13
	JJA	0.83-0.94	0.55-0.88
	SON	0.83-0.98	0.30-0.97
West Coast	DJF	0.75-0.96	0.42-0.87
	MAM	0.83-0.97	0.44-0.86
	JJA	0.34-0.93	0.55-1.37
	SON	0.53-0.97	0.28-1.07
Southwest	DJF	0.59-0.95	0.41-0.97
	MAM	0.50-0.95	0.43-1.06
	JJA	0.61-0.89	0.54-1.07
	SON	0.43-0.94	0.37-1.11
Northern Plains	DJF	0.63-0.89	0.81-1.36
	MAM	0.82-0.93	0.46-1.07
	JJA	0.83-0.93	0.50-0.71
	SON	0.81-0.92	0.44-1.00
Colorado Plateau	DJF	0.63-0.93	0.56-1.62
	MAM	0.77-0.88	0.53-0.90
	JJA	0.74-0.90	0.62-0.95
	SON	0.80-0.91	0.47-0.99

respectively, where σ in (4.11) is the average of σ_x and σ_y. The quantities (4.10) and (4.11) are calculated for each dataset pair, and the average of these quantities over the 36 dataset combinations is taken as an estimate of data uncertainty for each grid cell. Specifically, in Figure 4.13 we show

$$\rho(i) = \frac{1}{N}\sum_{n=1}^{N}\rho_n(i), \tag{4.12}$$

and

$$NRMSE(i) = \frac{1}{N}\sum_{n=1}^{N} NRMSE_n(i) \tag{4.13}$$

where N=36 is the number of dataset combinations.

The correlation coefficient provides a measure of the phase association (ρ), or the phase error (1- ρ) between datasets. In Fig. 4.13a, the average dataset correlation is in the range 0.34–0.98 for the domain as a whole over all seasons. The correlation is lowest over the Rocky Mountains in winter, spring, and fall; over northern Mexico for all seasons; and over southern California and the intermountain region in summer. Table 4.3 gives the range of values of ρ(i) for each season and region. The correlation across seasons is in the range of 0.78–0.98, 0.34–0.97, 0.43–0.95, 0.63–0.93, and 0.63–0.93 for the Pacific Northwest, West Coast, Southwest, Northern Plains, and Colorado Plateau regions, respectively. The lowest correlations occur in spring for the Pacific Northwest, in summer for the West Coast, in fall for the Southwest, and in winter for the Northern Plains and Colorado Plateau regions.

The NRMSE provides a measure of amplitude differences between datasets. In Fig. 4.13b, the average NRMSE is in the range of 0.28–1.62 for the domain as a whole over all seasons. The NRMSE is generally high over the Rocky Mountains, with the largest values observed in these parts in winter and the smallest values observed in summer. Large amplitude differences are also observed in coastal California in summer and in northern Mexico during all seasons. Values of NRMSE(i) across seasons are in the range of 0.30–1.37, 0.28–1.37, 0.37–1.11, 0.44–1.36, and 0.47–1.62 for the Pacific Northwest, West Coast, Southwest, Northern Plains, and Colorado Plateau regions, respectively. In Table 4.3, the largest errors are observed in winter for the Pacific Northwest, Northern Plains, and Colorado Plateau regions, whereas the largest errors are

observed in summer for the West Coast and Southwest regions.

b. Decomposition of the Mean Squared Error

To give an overall indication of the degree of agreement among datasets, we use the mean square error (MSE) and apply the decomposition of Murphy (1988). This methodology is commonly used in quantitative forecast verification to compare forecasted and observed fields to assess forecast skill. Here, we use the method to compare observations from one dataset against observations of another dataset. The MSE operating on anomaly fields is represented as

$$MSE_{x'y'} = \frac{1}{m}\sum_{i=1}^{m}\left(x_i' - y_i'\right)^2, \tag{4.15}$$

where x' and y′ are calculated as in (4.1) for two sets of observations (datasets) and m is the number of paired observations for a given 2-dimensional field (map). The MSE can be decomposed according to Murphy (1988) as

$$MSE_{x'y'} = \left(\langle x'\rangle - \langle y'\rangle\right)^2 + S_{x'}^2 + S_{y'}^2 - 2S_{x'}S_{y'}r_{x'y'}, \tag{4.16}$$

where the brackets represent spatial averages, $S_{x'}$ and $S_{y'}$ represent the spatial standard deviations of x' and y′,and $r_{xy'}$ represents the spatial anomaly correlation coefficient. Equation (4.16) can be manipulated algebraically by dividing by $S_{y'}^2$ and completing the square to give the normalized MSE, represented as

$$NMSE_{x'y'} = \underbrace{\left(\frac{\langle x'\rangle - \langle y'\rangle}{S_{y'}}\right)^2}_{A^2} + \underbrace{\left(r_{x'y'} - \frac{S_{x'}}{S_{y'}}\right)^2}_{B^2} + \underbrace{\left(1 - r_{x'y'}^2\right)}_{C^2}, \tag{4.17}$$

where A^2 is a non-dimensional measure of the unconditional bias, B^2 is a measure of the conditional bias, C^2 is the phase error, and NMSE is the normalized MSE, where

Table 4.4: Median value of the NMSE and terms of its decomposition taken over all data pairs.

		NMSE	A^2	B^2	C^2
Domain	DJF	0.42	0.05	0.08	0.27
	MAM	0.42	0.05	0.07	0.29
	JJA	0.70	0.06	0.15	0.45
	SON	0.41	0.04	0.08	0.29
Pacific Northwest	DJF	0.71	0.22	0.20	0.27
	MAM	0.73	0.18	0.16	0.32
	JJA	0.85	0.21	0.18	0.41
	SON	0.66	0.18	0.16	0.29
West Coast	DJF	0.85	0.25	0.26	0.34
	MAM	1.09	0.31	0.34	0.39
	JJA	1.22	0.22	0.42	0.49
	SON	0.87	0.22	0.22	0.38
Southwest	DJF	1.19	0.38	0.33	0.50
	MAM	0.85	0.17	0.24	0.41
	JJA	1.53	0.43	0.43	0.60
	SON	1.04	0.22	0.34	0.51
Northern Plains	DJF	1.46	0.48	0.46	0.53
	MAM	0.86	0.20	0.24	0.40
	JJA	0.71	0.14	0.16	0.41
	SON	0.95	0.26	0.30	0.43
Colorado Plateau	DJF	1.45	0.65	0.27	0.46
	MAM	1.43	0.62	0.32	0.49
	JJA	1.23	0.35	0.37	0.48
	SON	1.03	0.43	0.20	0.40

$NMSE = MSE / S_y^2$. For analyses of fields (maps), the terms in (4.17) can be physically interpreted as representing the contribution to NMSE due to phase error (C^2), amplitude differences (B^2), and error due to map mean differences (A^2) (based on Livezey et al. 1995).

Figure 4.14 provides the NMSE (averaged spatially and temporally) and the terms

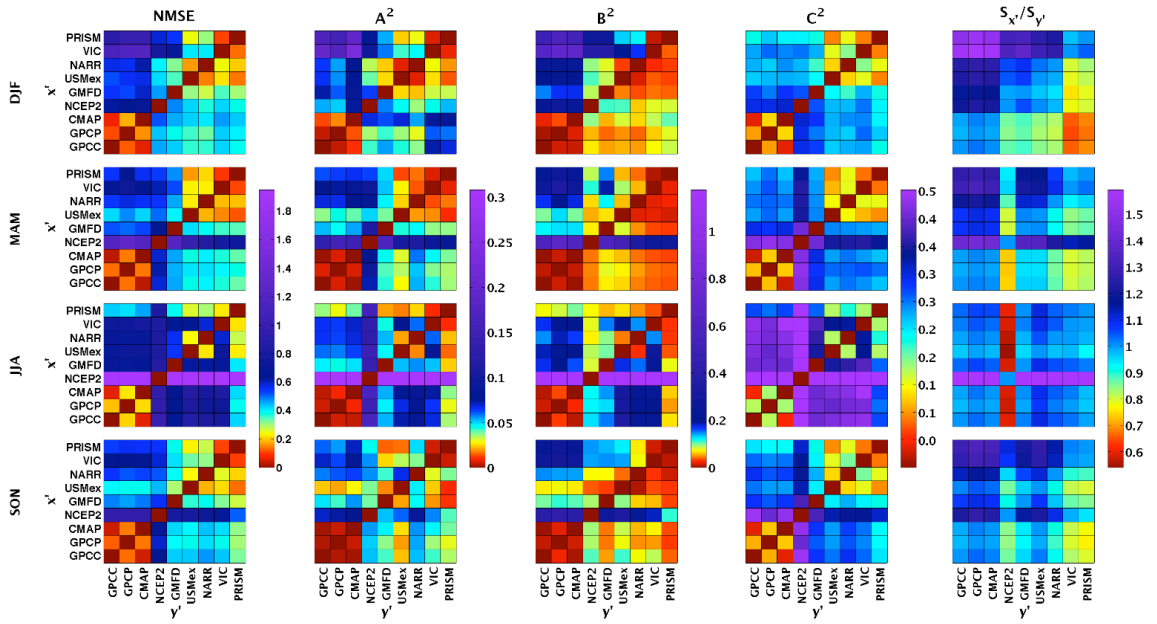

Figure 4.14: NMSE, the terms of its decomposition, and the ratio of standard deviations for the domain of the western United States. The datasets listed on the y axis (x axis) represent x′ (y′) in Eq. (4.16).

of its decomposition for all data combinations, and Table 4.4 gives the median value taken over all data pairs. Also shown in Fig. 14 is the term $S_{x'}/S_{y'}$ to describe how standard deviation differences might affect the NMSE and terms A^2 and B^2.

An examination of Fig. 14 and Table 4.4 shows that greater similarity between datasets is observed in winter, spring, and fall as compared to summer, with median NMSE values of 0.42, 0.42, 0.41, and 0.70, respectively. This summer reduction in similarity is a result of a reduction in phase association and an increase in amplitude differences. The clustering of low NMSE values in the southwest and northeast quadrants of Fig. 4.14 highlights two groups of "similar" data: Group 1 are GPCC, GPCP, and CMAP data, and Group 2 are USMex, NARR, VIC, and PRISM. These datasets generally show strong phase association and low bias within groups, and notably lesser phase association and higher bias between groups. The GPCC, GPCP, and CMAP data

Table 4.5: Seasonal averages of the spatial standard deviation Sy′ for each dataset.

		GPCC	GPCP	CMAP	NCEP	GMFD	USMex	NAR	VIC	PRISM
Domain	DJF	23.3	22.90	23.2	27.9	26.4	28.8	29.0	36.7	36.1
	MAM	17.9	17.6	17.8	24.9	18.9	19.4	21.2	23.2	23.1
	JJA	16.8	16.4	16.6	30.8	16.9	15.8	16.3	16.9	16.2
	SON	18.7	18.3	18.4	22.3	19.9	19.1	21.1	24.0	24.4
Pacific Northwest	DJF	29.7	28.6	29.3	31.8	27.4	33.5	34.5	43.4	41.9
	MAM	19.3	18.1	19.5	24.1	17.1	20.4	22.1	25.4	24.3
	JJA	10.2	10.4	10.2	19.3	10.4	8.7	11.2	11.7	11.1
	SON	22.3	21.4	22.2	23.1	21.0	22.3	24.4	29.2	29.0
West Coast	DJF	22.8	20.5	23.1	25.2	24.3	28.3	25.8	39.6	36.4
	MAM	12.8	12.6	12.7	16.4	13.7	15.0	15.4	19.3	18.2
	JJA	5.5	5.6	5.2	13.0	6.5	5.4	6.3	5.7	5.6
	SON	11.2	10.3	11.2	13.3	12.1	11.5	11.9	14.8	15.0
Southwest	DJF	8.9	7.5	8.9	11.9	9.8	10.0	10.4	11.9	11.2
	MAM	9.2	8.2	9.0	15.8	8.9	8.7	9.5	10.2	10.4
	JJA	18.2	15.6	18.1	30.6	16.8	16.0	16.3	16.9	14.9
	SON	12.6	11.0	12.2	15.5	11.2	10.1	11.3	11.8	12.2
Northern Plains	DJF	5.3	6.1	5.3	6.1	7.9	6.7	9.1	8.7	8.7
	MAM	12.5	12.9	12.5	15.1	13.9	10.5	13.2	12.6	12.6
	JJA	14.4	15.4	14.3	22.1	12.7	12.0	12.7	13.6	14.1
	SON	9.2	9.4	9.2	10.0	9.8	9.0	11.5	11.5	12.3
Colorado Plateau	DJF	6.3	7.6	6.1	7.3	9.9	6.8	9.1	8.2	8.0
	MAM	9.4	9.2	9.1	14.7	9.8	8.8	11.4	10.1	9.7
	JJA	9.2	8.2	8.8	17.4	9.4	8.0	9.7	8.8	9.1
	SON	7.6	7.5	7.5	9.8	9.1	7.4	9.5	8.6	8.5

are shown to be very similar, with NMSE less than 0.22 for all seasons. The conditional and unconditional bias between the GPCC, GPCP, and CMAP data is small (A^2, B^2<0.03) and phase association is strong ($1-C^2 = r_{x'y'}^2 = 0.82-0.97$). Strong similarity is also observed between the VIC and PRISM data, with NMSE less than 0.13 in winter, spring, and fall, and NMSE=0.22 in summer. The VIC and PRISM data demonstrate very little conditional or unconditional bias (A^2, B^2<0.04) and strong phase association ($r_{x'y'}^2 = 0.82-0.92$). The USMex and NARR data show relatively strong similarity

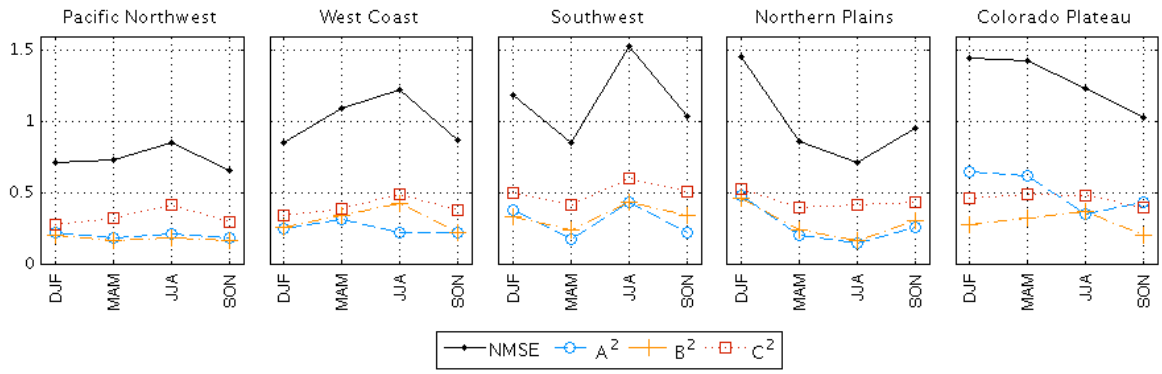

Figure 4.15: Median NMSE and decomposition terms for each season.

between each other and with the VIC and PRISM data during winter, spring and fall, with NMSE ranging from 0.12 to 0.41, squared anomaly correlation coefficients ranging from 0.80 to 0.91, and conditional (unconditional) bias ranging from 0.01 to 0.21 (0.01–0.07). In summer, comparisons of the USMex and NARR data against the VIC data show a reduced similarity (NMSE=0.58–0.73), which is primarily a result of an increase in phase error (C^2=0.38). The PRISM and VIC data demonstrate a minimum similarity with Group 1 during winter (NMSE 0.38–1.20) as a result of increased conditional (B^2=0.06–0.69) and unconditional (A^2=0.06–0.23) bias. The NCEP2 data are shown to be generally dissimilar to the other data during all seasons except winter, and it is especially dissimilar in summer. In Fig. 4.14, the NCEP2 data are shown to exhibit lower phase association and higher conditional and unconditional bias than is observed for comparisons between other datasets. The GMFD data generally show a greater similarity with Group 2 than with Group 1 (mean NMSE 0.47 and 0.58, respectively).

Table 4.5 provides seasonal averages of the spatial standard deviation for each dataset, which affects the NMSE through the term $S_{x'}/S_{y'}$. Standard deviation differences between datasets are shown in Fig. 4.14 to contribute to the bias terms A^2 and

B^2 and to the NMSE in winter, spring, and fall. During summer, the term $S_{x'}/S_{y'}$ is near one for all data pairs except the NCEP2, which has a much larger spread than the other data. Aside from the NCEP2, the standard deviation differences are greatest between Group 1 and Group 2 datasets, where for Group 1 S_x=16.4–23.3 and for Group 2 S_x=15.8–36.7 mm day^{-1}. The standard deviations of the GMFD and NCEP2 data generally occur in the midrange of these two groups except during summer, when the NCEP2 spread is large, as discussed above.

Figure 4.15 and Table 4.4 give the median NMSE and decomposition terms for each season and region. The NMSE is largest in the Southwest region in summer, and in the Northern Plains and Colorado Plateau regions in winter (median NMSE 1.53, 1.46, and 1.45, respectively). The summer reduction in dataset similarity in the Southwest is a result of increases of all error components. For the Northern Plains in winter, the increase in NMSE is a result of increases in all error terms, but conditional and unconditional bias rise more sharply in winter than phase error. For the Colorado Plateau in winter, the higher NMSE is a result of an increase in unconditional bias. For the Pacific Northwest region, the largest NMSE is observed in summer, which is primarily a result of a reduction in phase association because the bias terms change very little across seasons. For the West Coast region, the NMSE peaks in summer, which is a result of increases in conditional bias and phase error.

4.6. Summary and Conclusions

A comparison of precipitation between nine data products revealed two groups of similar datasets. The GPCP, CMAP, and GPCC (Group 1) were shown to behave

similarly as were the USMex, NARR, VIC, and PRISM (Group 2). In general, Group 1 depicts drier conditions relative to other data during all seasons and for most of the domain. In previous studies, the satellite components of the GPCP and CMAP data have been shown to produce biased estimates of precipitation in the midlatitudes (e.g., Adler et al. 2001), which could help explain these findings. However, over land the GPCP and CMAP are strongly influenced by gauge observations. This study has shown a strong similarity between the GPCP, CMAP and the GPCC rain gauge datasets, and any dry bias is similarly expressed by all three datasets. Therefore, it is more likely that sampling issues related to sparse gauge coverage over the western United States by the GPCC rain gauge dataset (which is also used in the GPCP and CMAP) explains the drier conditions observed by Group 1 relative to the other datasets.

The largest differences within Group 1 occur as higher precipitation reported by the GPCP. This is likely a result of gauge corrections used in the development of the GPCP data, which are not used in the CMAP and GPCC. This is consistent with other studies involving the GPCP and CMAP over land. For example, Gruber et al. (2000) noted these differences over northern hemispheric land areas including central North America, which they also attribute to the GPCP gauge corrections. The largest data differences within Group 1 were observed in the Pacific Northwest and Colorado Plateau regions in winter and in the northern plains during all seasons (Table 4.1 and Fig. 4.2b).

Group 2 data generally show wetter conditions relative to other datasets, particularly in the Pacific Northwest in winter, spring, and fall and in the West Coast in winter (Table 4.1 and Fig. 4.2b). Within this group, the VIC and PRISM exhibit the wettest conditions among the group as a result of the orographic adjustment employed in

their development. These two datasets were found to be very similar, which is not unexpected because the rain gauge data used in the VIC is adjusted using data from the PRISM. The greatest differences among Group 2 datasets in long-term mean precipitation occur in the Pacific Northwest in winter, spring, and fall and in the West Coast in winter (Table 4.1). However, when error components are considered, the largest differences are observed in summer for comparisons of the USMex and NARR against the VIC as result of a summertime increase in phase errors. With respect to data distributions (as represented by empirical CDFs), comparisons of Group 2 datasets were shown to be significantly different for most regions and seasons except for comparisons between the VIC and PRISM. However, removing the climatology generally eliminates any distribution differences between Group 2 datasets (Figs. 4.6 and 4.7).

The NCEP2 precipitation was generally observed in the midrange of the two groups during winter and fall; however, during spring and summer, the NCEP2 depicts wetter conditions than Group 2 (Table 4.1 and Fig. 4.2b). Previous studies have shown the NCEP2 to overestimate precipitation in the central United States during spring and summer (Higgins et al. 1996) and in the central-western United States during the North American monsoon (Janowiak et al. 1998). This has been attributed to inadequacies in the treatment of the Great Plains low-level jet by the spectral model, which results in the transport of excess moisture into the central United States (Higgins et al. 1996; Mo and Higgins 1996). This is consistent with our results; the NCEP2 excesses are observed to occur predominantly in the Northern Plains and Colorado Plateau regions in spring and summer and in the Southwest region where the NAM dominates during summer and fall (the NAM season is typically late summer and into September). Our results also show

higher precipitation values for the NCEP2 relative to other data for the Pacific Northwest in spring and summer. In general, the pattern of precipitation bias associated with the NCEP2 over Washington and Oregon (Fig. 4.2b) is consistent with Widmann and Bretherton (2000), who compared the NCEP reanalysis to gauge-based observations adjusted with PRISM climatology. Their study showed that the NCEP data (relative to gauge observations) underestimated precipitation west of the Cascades, while over-estimating it to the east as a result of coarse representation of topography by the spectral model.

The GMFD was shown to be more similar to Group 2 than Group 1. However, it shows a high level of phase error and bias when compared to both groups. The GMFD shows the wettest conditions of long-term mean precipitation relative to other datasets for all seasons and regions except the Pacific Northwest during winter, spring and fall. Although the GMFD uses the NCEP reanalysis in its development, this study shows the GMFD to be dissimilar to the NCEP2.

The results of this analysis underscore the high level of uncertainty in precipitation observations for the western United States. Additionally, the uncertainty demonstrates important space/time dependencies, which makes it difficult to quantify data error outright. Uncertainty in observations should be considered when using precipitation data for numerical model evaluation. It is possible that the choice of the observational dataset selected for a particular study could affect conclusions regarding model skill.

One limitation of this analysis is that it is not possible to separate the effects of the rescaling from the effects of other data differences on the results. This is recognized as an

important issue, especially because the most similar datasets (GPCP, CMAP, and GPCC and PRISM and VIC) share common data sources as well as a common spatial scale. The similarities and differences between datasets are likely a result of some combination of scale and precipitation estimation methodology. However, the contribution of each to the uncertainty is difficult to quantify individually. Another limitation is our inability to draw any conclusions about superiority among datasets. This will likely remain a problem because the ability to quantify error requires that we make comparisons against the truth, and no dataset can be assumed to represent the truth free from error.

CHAPTER 5

Hydroclimatology of the western United States: A winter 1999 case study using the Ocean-Land-Atmosphere Model (OLAM)

5.1 Introduction

Hydrometeorological forecasting in the western U.S. is challenging, largely due to complex terrain, which makes issues of model spatial resolution extremely important. Failure to adequately resolve topography produces an incomplete representation of the interaction between orography and transient weather systems. Precipitation in the western U.S. is largely governed by the interaction between large-scale circulation and associated storm tracks with local topography (Cayan and Roads 1984, Soong and Kim 1996, Cayan et al. 1993). Topographic features such as elevation and slope orientation have a considerable effect on the occurrence and preferred location of precipitation (Daly et al. 1994). Circulation features such as prevailing wind speed and direction (Reynolds and Dennis 1986, Leung et al. 2003), and the strength of barrier jets are also important in driving precipitation in the region (Braun et al. 1999, Parish 1982, Smith et al., 2009).

Long-range forecasts of seasonal anomalies on the order of a season can be used with some skill (Barnston 1999, Livezey 1990, Goddard 2001) due to surface-atmosphere relationships where slowly evolving fields such as SSTs and land surface conditions strongly affect the atmospheric state. Such forecasts have used both General Circulation

Models and nested model configurations in which a Regional Climate Model is forced at its lateral boundaries by GCM output. The high spatial resolution afforded by RCMs has been shown to improve forecast skill. However, there are some limitations with nested GCM-RCM configurations. Most importantly is that the interaction between different scales is strictly one-way so there is no feedback to the large scale by the RCM.

This research investigates the ability of the Ocean-Land-Atmosphere Model to simulate the hydroclimatology of the western U.S. and to provide new insights into this issue. The OLAM is a global, non-hydrostatic model with a refinement option for resolving regions of interest at a higher spatial grid size within a global domain. Therefore, interactions between gridcells having different spatial scales is two-way interactive, with no need for lateral boundary forcing. Chapter 2 provides a description of OLAM including its grid structure, dynamic core, and subgrid parameterizations.

This Chapter investigates OLAM's seasonal predictability potential using the winter of 1999 as a case study by integrating over a season without external atmospheric updates. In such a long-range hindcast it is not expected that simulated and observed storms or atmospheric anomalies would evolve identically in time. This study focuses instead on whether heavy and light precipitation events occur under similar atmospheric conditions according to both OLAM and observations. Updates to the surface are from observations for SSTs and, for the land surface a coupled land surface model (LEAF) is used. Since observed SSTs are incorporated during the simulation this does not represent a true long-range seasonal hindcast, which would require a coupled ocean model. In the future, OLAM will be updated to include a coupled ocean model for use in these types of studies.

For this study, a series of simulations are run over the winter season of 1999 (1 December-28 February) with successive increases in the horizontal grid size over the western U.S. The winter of 1999 occurred during the cold phase of the ENSO with wet conditions observed in the Pacific Northwest including flooding events. Temperatures were much warmer than usual over the continental U.S. and a persistent trough was positioned over the Pacific Northwest (NOAA, 1999).

The OLAM configuration is discussed in Section 5.2. The verification data are discussed in Section 5.3. Section 5.4 compares OLAM simulated precipitation against observations and discusses the effect of increased horizontal grid size on model skill. Section 5.5 explores features of the synoptic-scale meteorology during heavy and light precipitation events according to observational data and OLAM simulations. A summary and conclusions are given in Section 5.6. This chapter is summarized in Guirguis et al. (2009).

5.2 Model Grid Configuration

High-resolution land-surface datasets for topography and land cover are used to initialize OLAM at its lower boundary. The USGS' 1-km digital elevation database is used for topography, and its 1-km Global Land Cover Characteristics database is used for land cover. The land cover data are derived from satellite Advanced Very High Resolution Radiometer (AVHRR) data. Ocean forcing is provided to OLAM by the National Climate Data Center's weekly 1°x1° SST dataset (Reynolds 1988). Initial soil temperature and moisture fields, snow cover, and sea ice are provided by NCEP-DOE Reanalysis 2 (Kanamitsu et al 2002).

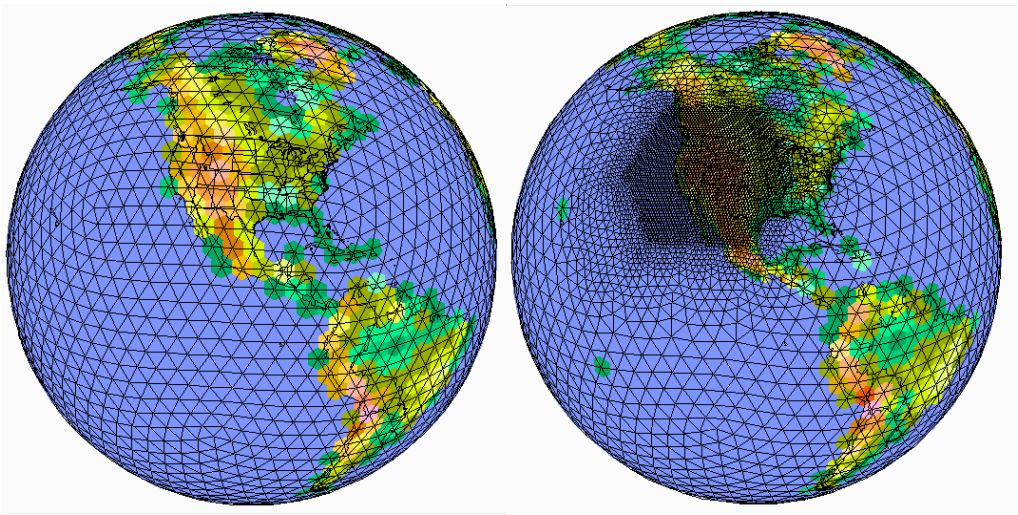

Figure 5.1: OLAM grid configuration corresponding to simulations OLAM0, and OLAM3. The field plotted is topography.

For these simulations, OLAM is configured with a global grid having a horizontal characteristic length scale of 360 km. For tests of sensitivity to horizontal grid size, additional grid refinement levels are configured over the western U.S. The grid refinement uses a 2:1 refinement ratio, where the transitions to each level occur over 2-5 grid cells. Five additional simulations are run, which have a CLS over the western U.S. of 180, 90, 45, 23, and 11 km. Vertically there are 53 layers, the first being 200 meters and with a stretching ratio of 1.05 applied to higher levels. The model is initialized on 21 November and evaluated over 1 December – 28 February. OLAM is not nudged with observations during the simulations, so the evaluation is based on statistical comparisons with observations without comparing specifically timed events. In this paper, $OLAM_0$ refers to the simulation having a 360 km global grid, and $OLAM_1$, $OLAM_2$, $OLAM_3$, $OLAM_4$, and $OLAM_5$ refers to those simulations having higher resolution over the western U.S. with a CLS of 180, 90, 45, 23, and 11 km respectively. The OLAM grid configuration for this study is shown in Fig. 5.1. For analysis and display, OLAM output is transformed to a rectangular grid.

5.3 Verification Data

Four precipitation datasets available at a temporal resolution of daily or sub-daily are used to evaluate OLAM's performance. These include one merged satellite-rain gauge product, one gridded rain gauge dataset, one rain gauge dataset with statistical orographic adjustment, and regional reanalysis. They are, respectively, (i) the Global Daily Merged Precipitation Analysis of the GPCP (GPCP1dd, Huffman et al. 2001), (ii) CPC Retrospective United States and Mexico Daily Precipitation Analysis (USMex, Higgins et al. 2000), (iii) VIC Retrospective Land Surface Dataset (Maurer et al. 2002), and (iv) North American Regional Reanalysis (NARR, Mesinger et al. 2006). The USMex, NARR, and VIC precipitation datasets are described in Sections 3.2d , 3.2g, and 3.2h, respectively.

The GPCP1dd is a global, merged satellite-rain gauge data product available daily from 1997 at a 1° x 1° spatial resolution. It is developed similarly as the GPCP monthly product (Section 3.2b). However, the GPCP1dd uses the threshold-matched precipitation index (TMPI) for its IR-based precipitation estimates whereas the monthly product uses the GOES Precipitation Index. The primary difference between the two estimation methods is that the GPI assumes a constant rain rate in the empirical relationship between cloud brightness temperatures and precipitation, whereas rain rate is allowed to vary when calculating the TMPI. Poleward of 40º, the data components are the same in both the daily and monthly products. The GPCP1dd is scaled to match the GPCP monthly product in terms of monthly accumulation.

The choice to use all four datasets for the evaluation of OLAM is due to the high level of uncertainty in observational precipitation data (Chapters 3 and 4). Merged

satellite-rain gauge products were shown in Chapter 4 to be nearly indistinguishable from the sparse rain gauge dataset used in their development and demonstrated a dry bias as compared to the other data. NARR and USMex represented the precipitation climatology in the midrange of the other data with a few datasets being generally drier and a few being generally wetter. VIC generally represented wetter conditions than the other datasets. USMex, being a rain gauge dataset, would suffer the problem of sampling error especially over mountains and aerodynamic issues such as low bias due to wind and evaporation (e.g. Legates and Willmott 1990). However, gauge coverage of the USMex dataset greatly surpasses that of the rain gauge dataset employed in the development of the GPCP data (monthly and 1-degree daily). NARR was shown in Chapters 3 and 4 to compare well with observations, but precipitation fields are dependent on the performance of subgrid-scale parameterizations, which may not be optimal over complex terrain. Chapter 4 showed NARR to share some features with the global NCEP Reanalysis such as a drier representation of coastal California during winter, suggesting that parameterizations used in the data assimilation process may be an issue here for both reanalysis products. The use of these four datasets should represent the range of observations. To assess OLAM's performance locally within the domain of the western U.S., specific regions that were found in Chapter 3 to have unique precipitation climates are considered. These unique precipitation regions are shown in Chapter 4, Fig. 4.1, and are referred to as the Pacific Northwest (PNW), West Coast (WC), Southwest (SW), Northern Plains (NP), and Colorado Plateau (CP) regions.

This study uses additional NARR variables to investigate the synoptic meteorology during heavy and light precipitation events. Specifically, 850-mb

geopotential height, temperature, wind, and specific humidity are considered. These variables are held to be more reliable than NARR precipitation because they are influenced directly by observations and are less dependent on model parameterizations (Kalnay et al. 1996). Geopotential height, temperature, and wind fields are "Class A" variables, which are the most reliable, and humidity is a "Class B" variable, which is influenced more by the model than Class A but still contains observational data input.

5.4 Results

a. Precipitation

The seasonal precipitation for each OLAM simulation and observational dataset is shown in Fig. 5.2. Considering first the observations, NARR, USMex, and VIC show the enhancement of precipitation over the Rocky Mountains, while this feature is absent in the GPCP1dd data. This inability to represent Rocky Mountain precipitation was also observed in Chapters 3 and 4 for the monthly GPCP data product, likely due to the influence of the sparse rain gauge data. The simulation results show OLAM to be capable of representing the general precipitation pattern, including precipitation over the Rockies. $OLAM_0$ underestimates precipitation everywhere except in the Southwest region over northern Mexico, Arizona, and New Mexico, which are wetter than observed. Also in $OLAM_0$ there are no distinguishing spatial features due to the coarse spatial scale. $OLAM_1$ shows much improvement over $OLAM_0$, however it fails to capture the spatial details over the Rockies. $OLAM_2$ shows improvements in the spatial pattern, however with too much precipitation occurring over the Idaho Rocky Mountains and northern California. The spatial details of precipitation are further defined by $OLAM_3$-$OLAM_5$.

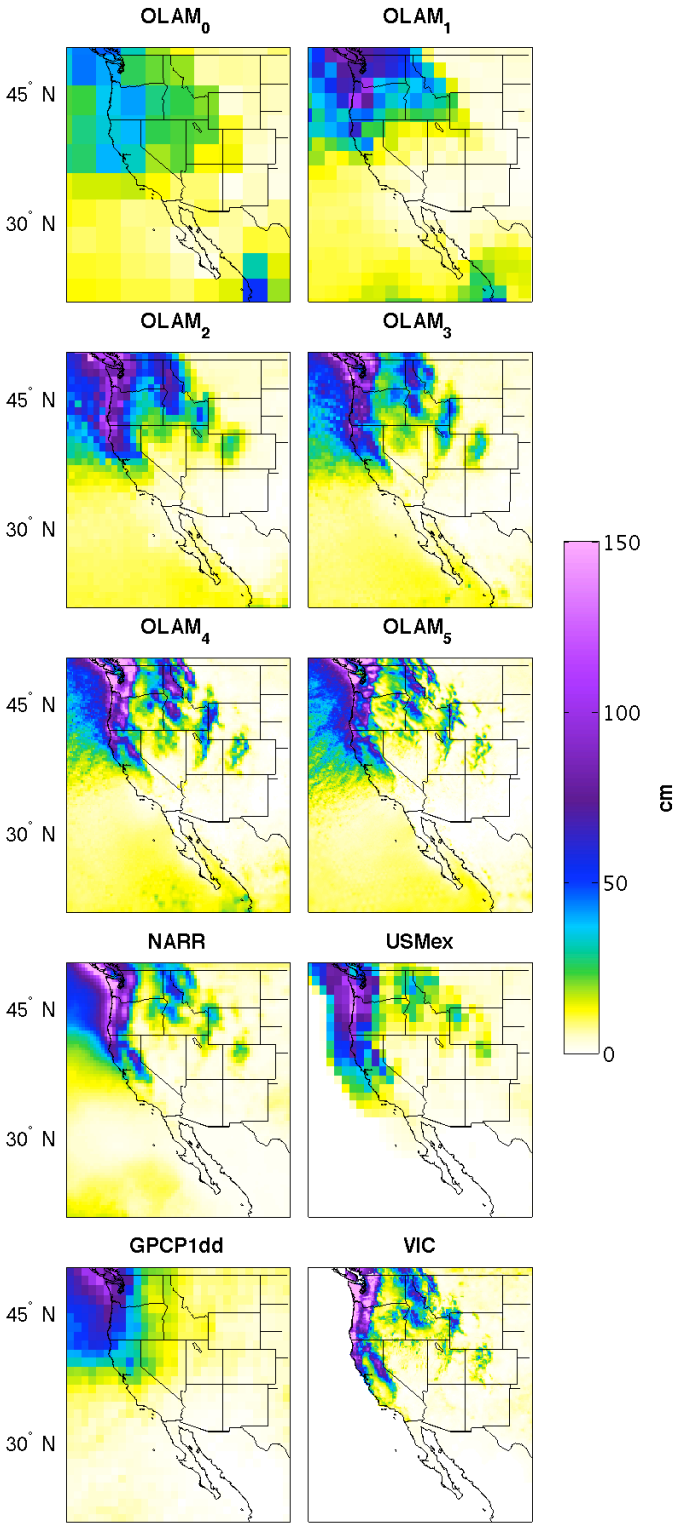

Figure 5.2: Seasonal precipitation totals for each OLAM simulation and observations.

$OLAM_3$ compares well in magnitude with VIC and is within the range of observations over the coast of the Pacific Northwest and northern California. $OLAM_4$ overestimates precipitation over the Rockies and eastern Washington and is similar to VIC in the Pacific Northwest. $OLAM_5$ is similar in magnitude and pattern to VIC and is within the range of observations over the coastal Pacific Northwest and California.

Zonal cross sections along three latitudes are shown in Fig. 5.3 to illustrate the precipitation-topography relationship. The sections are over 47.6, 44.4, and 37.1° N, which cross near Seattle, Eugene, and San Francisco, respectively. Fig. 5.4 shows the spatial representation of topography according to the six OLAM simulations over the western U.S. In both Figures, considering the cross section at 47.6° N, $OLAM_0$ and $OLAM_1$ do not distinguish between the Cascades to the west and the Rockies to the east but rather represent elevation as steadily increasing eastward. This distinction begins to be represented by $OLAM_2$ and for $OLAM_3$–$OLAM_5$ the ranges are well separated. Also for $OLAM_3$ and the higher resolution simulations, the Rocky Mountain elevation is higher and more detailed. Accordingly, in Fig. 5.3 $OLAM_0$ and $OLAM_1$ show a single and gradual peak in precipitation windward of the terrain, $OLAM_2$ shows a peak corresponding to uplift along the Cascades and another peak near −117° W as storms cross the Rockies, and $OLAM_3$ shows a double peak windward of the Cascades and a double Rocky Mountain peak due to resolved elevation details. The higher resolution $OLAM_4$ and $OLAM_5$ simulations show a similar spatial pattern to $OLAM_3$ but with a larger magnitude, most notably to the west of the Cascades. Considering the observations, the GPCP1dd data differs from the other observations to the east of the Cascades, where it fails to represent Rocky Mountain enhancement of precipitation.

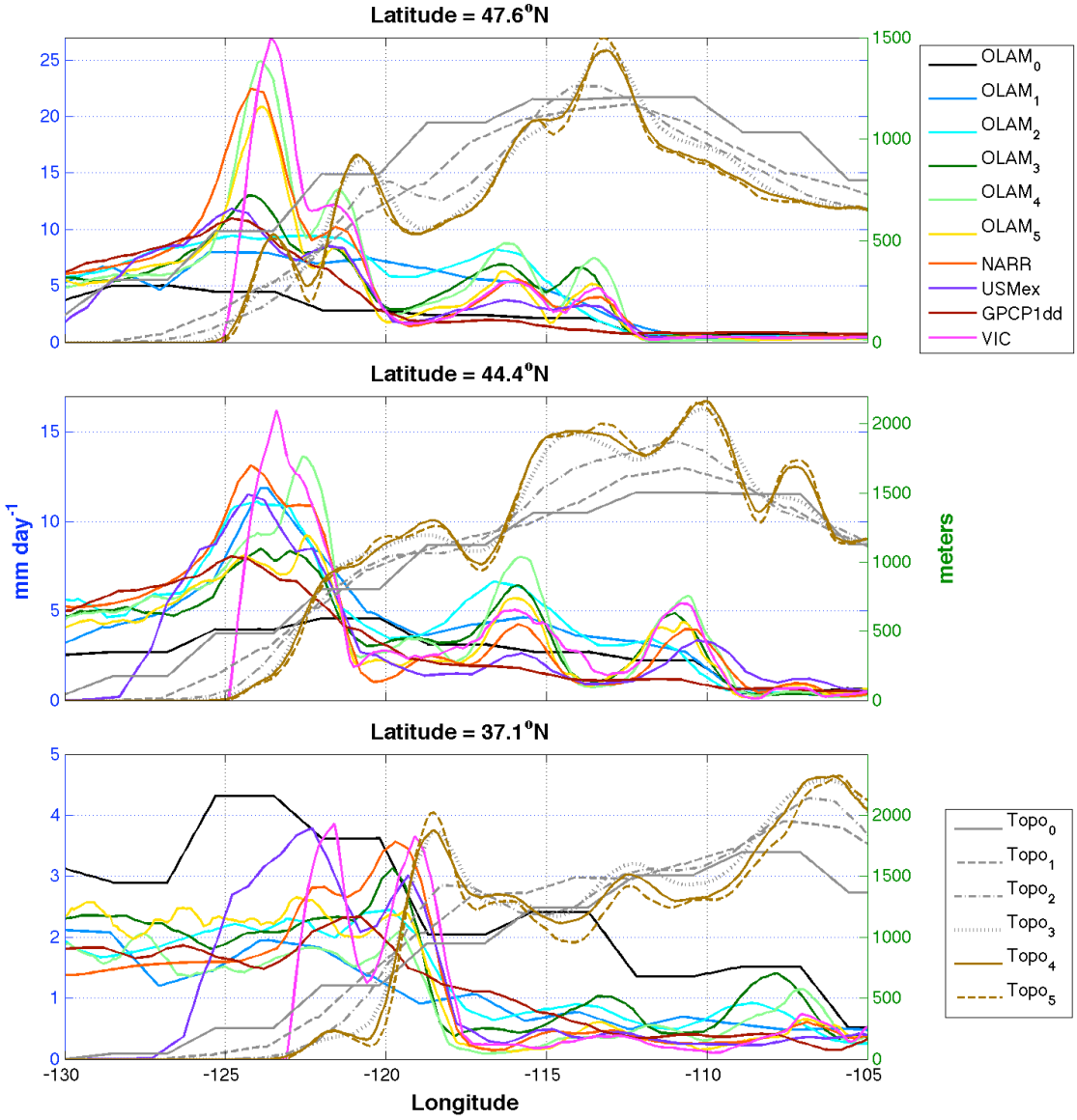

Figure 5.3: Zonal cross-sections of seasonal mean precipitation (OLAM and observations) and topography (OLAM) across three latitudes. Data is shown with a 7-point running average filter for clarity. $Topo_0$-Top_5 refers to topography fields from $OLAM_0$-$OLAM_5$

NARR and VIC are similar across the entire cross section in magnitude and pattern, but with slightly more precipitation west of the Cascades according to VIC. USMex is similar to GPCP1dd except with a better representation of the Rocky Mountain orographic signal. $OLAM_3$ is similar to USMex in terms of magnitude and spatial pattern, except $OLAM_3$ depicts a stronger orographic signal from the Rockies. $OLAM_4$

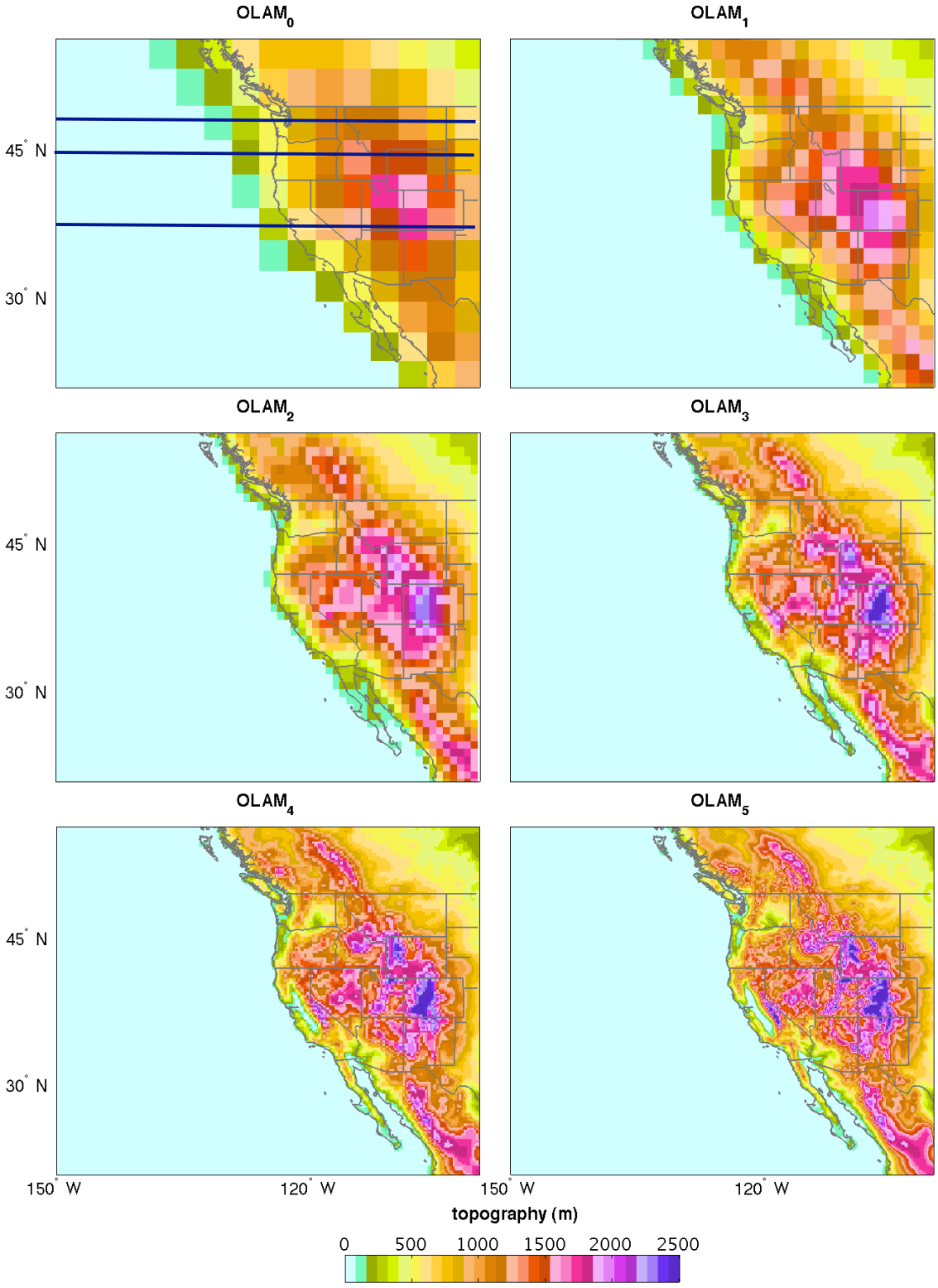

Figure 5.4: Topography as represented by the six OLAM simulations. Horizontal lines displayed for $OLAM_0$ correspond to the cross sections in Fig. 5.3.

and $OLAM_5$ compare well against the highest resolution data, NARR and VIC, except $OLAM_4$ is wetter over the Rockies.

From Fig. 5.3, the cross section along 44.4° N shows that $OLAM_2$ begins to resolve the Snake River Valley in southern Idaho (–117° W). This feature is well represented only starting in $OLAM_3$. $OLAM_3$-$OLAM_5$ also show details of the Rocky Mountains including the valley between the Sawtooth Range in Idaho (~ –15° W) and the Absoroka Range in Wyoming (~ –110° W). $OLAM_4$ and $OLAM_5$ additionally resolve the Bighorn Mountains (~ –107° W). $OLAM_0$ shows only a weak orographic signal windward of the Cascades, which is much better represented by $OLAM_1$. $OLAM_1$ also shows a very minor Rocky Mountain signal as seen in the mild peak near –116° W. $OLAM_2$ compares well with the higher resolution simulations except for a reduced precipitation peak near –110° W where details of the Rockies are not resolved by $OLAM_2$. $OLAM_3$–$OLAM_5$ are similar in their spatial pattern but differ in magnitude. West of the Cascades, the $OLAM_2$-$OLAM_5$ are generally within the range of observations, however, $OLAM_4$ and $OLAM_5$ show the precipitation peak shifted to the east. East of the Cascades, the Rocky Mountain impact on precipitation is generally stronger according to $OLAM_3$-$OLAM_5$ for the first precipitation peak (~ –116° W). For the second precipitation peak (~ –111° W) $OLAM_3$-$OLAM_5$ are within the range of observations. East of the Cascades $OLAM_0$ shows no orographic signal and $OLAM_1$ shows only a weak one.

Figure 5.3 for the cross section along 37.1° N shows that $OLAM_2$ begins to resolve the Sierra Nevada (~ –118° W) as distinct from the high elevations of Nevada to the east, and the Sierras are well defined by $OLAM_3$–$OLAM_5$. $OLAM_4$ and $OLAM_5$ additionally

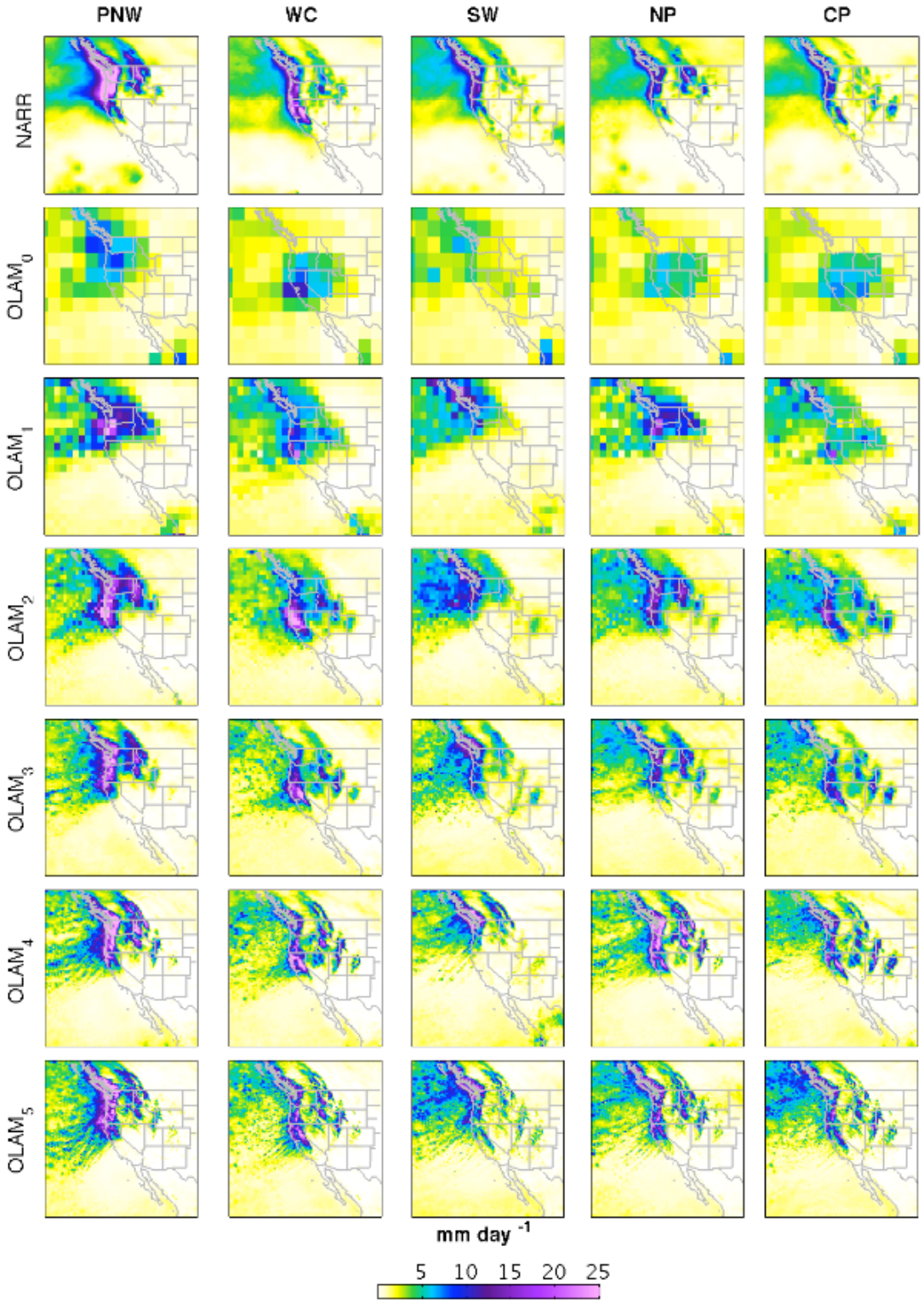

Figure 5.5 Daily mean precipitation for those days when precipitation exceeded the 85th percentile in the specified region.

resolve the valley (~ –121° W) between the Coast Ranges and the Sierras in California. These details are also seen in Figure 5.4 where elevation differences over the Sierras are notable between simulations, as is the representation of the Sacramento and San Joaquin Valleys. From Fig. 5.3, there is more pattern variation in the observations for the coastal and near coastal parts than was observed for the northern latitudes. East of the Sierra Nevada, the precipitation peaks are not as well defined as to the north, probably because the ranges in this intermountain region tend to be smaller and a majority of moisture has been lost during ascent over the Sierras. $OLAM_0$ shows a general high bias at this latitude, in contrast to the northern latitudes where $OLAM_0$ precipitation is low. West of the Sierras, $OLAM_5$ compares well against NARR. Also west of the Sierras $OLAM_2$, $OLAM_4$, and $OLAM_5$ show a similar magnitude as GPCP1dd, except OLAM precipitation demonstrates a double peak making it more similar in pattern to the other observational datasets. OLAM generally exhibits drier conditions to the west of the Sierras with the exception of $OLAM_5$ precipitation, which is in the range of observations.

Fig. 5.5 shows the daily mean precipitation for the wettest days in each region. This is taken as those events where precipitation exceeded the 85^{th} percentile (P85) in each of the regions shown in Fig. 4.1. According to NARR, during the wettest events in the PNW heavy precipitation is observed over western Washington and Oregon and the Idaho Rocky Mountains. When conditions shift to those bringing wet conditions to the WC, precipitation in the Pacific Northwest is reduced in magnitude and spatial extent along the coast and is less intense over the Idaho Rockies. Heavy rainfall in the SW brings reduced precipitation along the coast and more precipitation over Arizona and New Mexico, although the amount of rainfall received here is comparatively small.

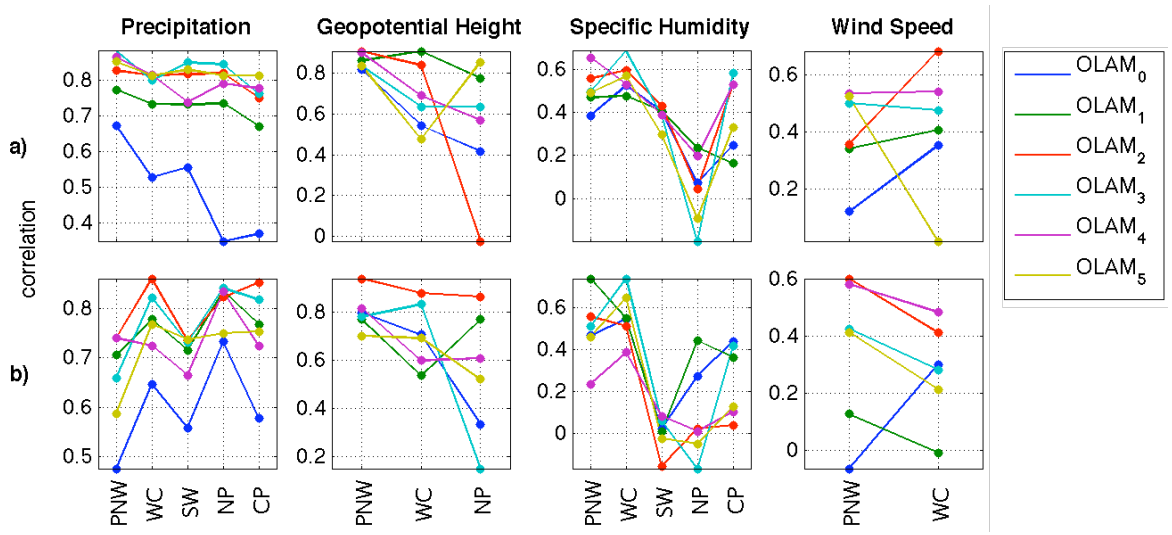

Figure 5.6: Spatial correlation measured against NARR for precipitation fields in Fig. 5.5 and the anomaly fields shown in Figs. 5.9-5.12 for (a) wet events and (b) dry events.

Northern Mexico remains relatively dry during SW storms. The NP wet periods show widespread precipitation over Montana while wet events in the CP region occur as spatially isolated clusters of heavy precipitation over the Central Rockies. These features are generally well represented by $OLAM_1$-$OLAM_5$, where the correlation with NARR is above 0.67 (Fig. 5.6). The highest resolution simulations show stronger correlations, which are near 0.8, excepting $OLAM_4$ in the NP. The greatest increase in correlation occurs when spatial grid size is increased from 360 to 180 km ($OLAM_0$ to $OLAM_1$), and secondarily from 180 to 90 km ($OLAM_1$ to $OLAM_2$). Bias (Fig. 5.7) is negative for P85 events in the PNW and the CP, during which OLAM shows drier and wetter conditions than NARR, respectively. Fig. 5.6b (5.7b) gives the correlation (bias) for the driest days, taken as days when precipitation fell below the 15th percentile (P15). Here there is generally a lower level of skill than is observed for the wettest events. However, OLAM still compares well against NARR with correlations above 0.7 for $OLAM_1$-$OLAM_5$, with a few exceptions. Bias is largest for PNW P15 events and is positive for P15 events in

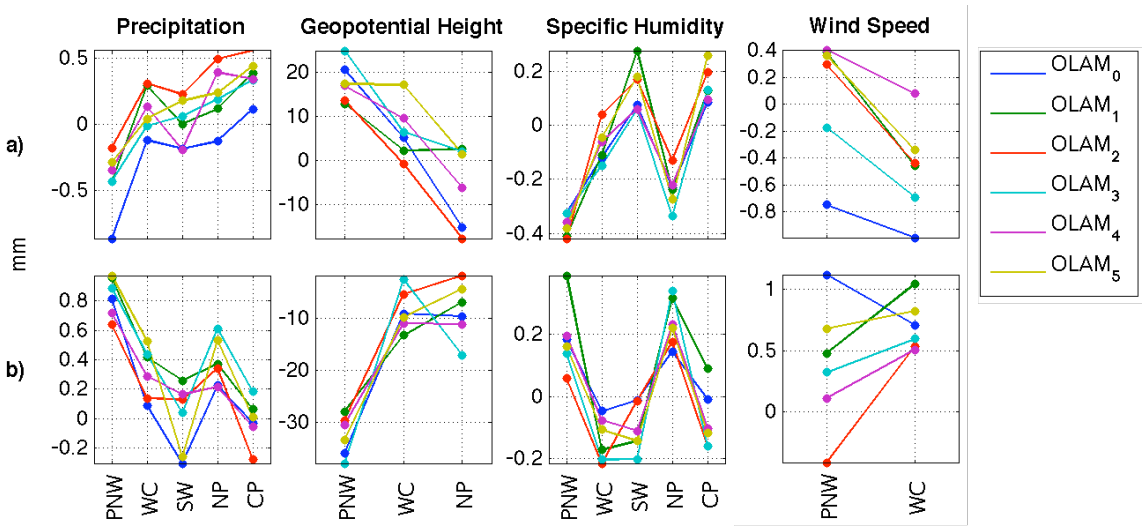

Figure 5.7: As in Fig. 5.6 but for bias.

general. The simulations show a reduction in skill for dry events as compared to wet ones, particularly for extremes in the PNW and WC.

b. Synoptic Meteorology

This section explores features of the synoptic-scale meteorology during the 1999 winter season according to NARR data and OLAM simulations. The OLAM data are from a seasonal integration of a numerical model with no external data input during the simulation except for observed SSTs. The NARR data are output from a numerical model that is run in assimilation mode, where observations are input to the model and then a 3-hour forecast is generated. The data are a collection of these 3-hour forecasts. Thus, there is no predictive capability represented in the NARR output. Fig. 5.8 shows spatial averages over each region of NARR precipitation and 850-mb atmospheric fields. For the Pacific Northwest, strong precipitation events are associated with local low-pressure as represented by geopotential height (Fig. 5.8b), higher temperatures (5.8c),

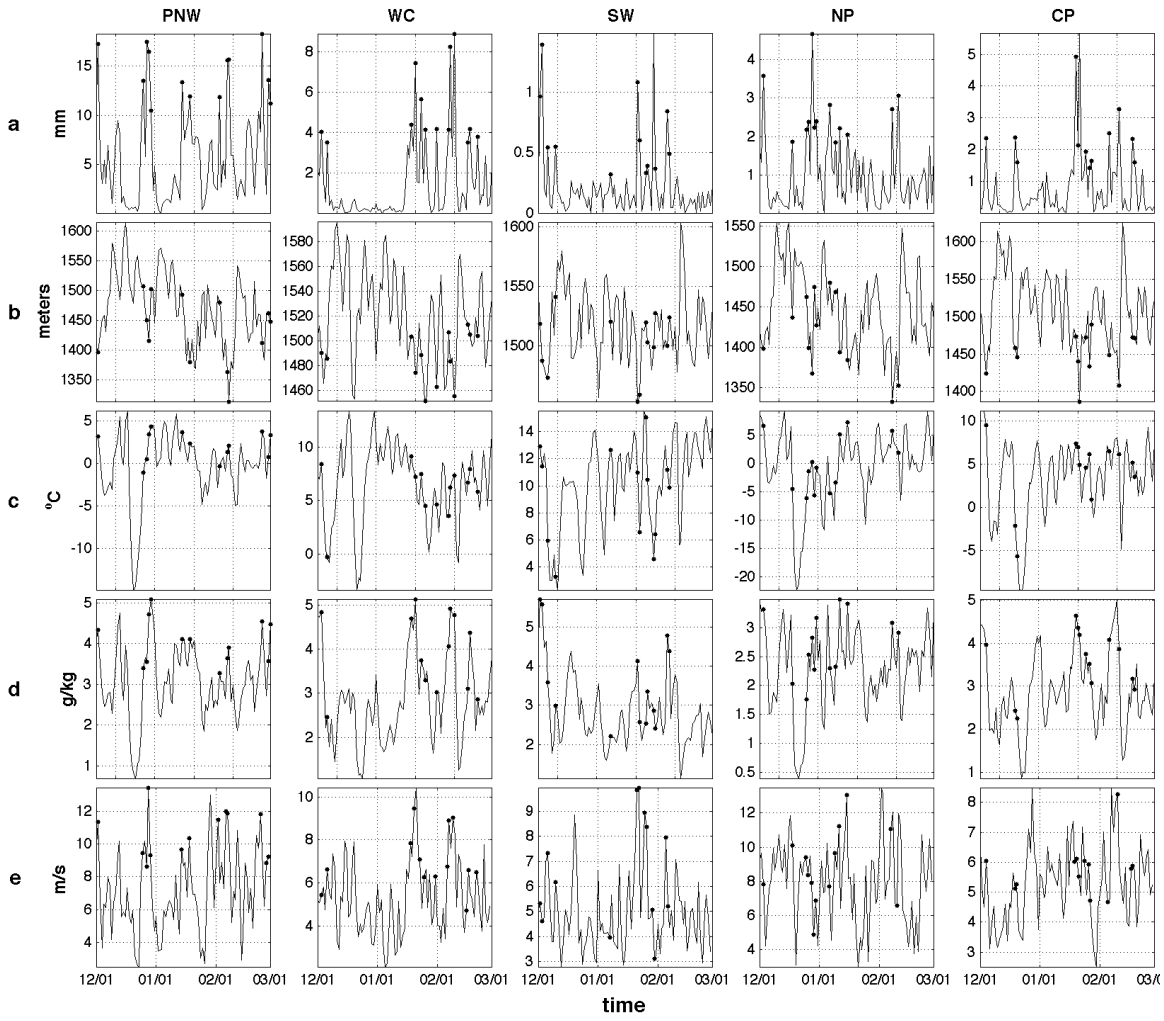

Figure 5.8: Spatial averages over each region of NARR (a) precipitation, (b) 850-mb geopotential height, (c) 850-mb temperature, (d) 850-mb specific humidity, and (e) 850-mb wind speed. Markers (•) indicate those days where precipitation exceeds the 85th percentile.

high humidity (5.8d), and strong winds (5.8e). This is similarly true for the West Coast where these conditions are observed during the second half of the season, although the temperature-precipitation relationship is less apparent. Table 5.1 gives the correlation between the precipitation time series (5.8a) with the other atmospheric variables (5.8b-e) for these and the other three regions. Precipitation is associated with low-pressure systems in all regions, and this relationship is strongest for the PNW, WC, and NP (ρ = –0.56-0.57) and weakest for the SW and CP (ρ = –0.28). Other studies have

Table 5.1: Correlation with precipitation

	PNW	WC	SW	NP	CP
Geopotential Height	-0.57	-0.56	-0.28	-0.56	-0.28
Temperature	0.33	-0.03	-0.11	-0.03	-0.11
Specific Humidity	0.63	0.72	0.48	0.72	0.48
Wind Speed	0.76	0.59	0.33	0.10	0.08

similarly shown height anomalies to be inversely correlated with anomalies of precipitation (e.g. Stidd 1954, Cayan and Roads 1984, Klein and Bloom 1987, McCabe 1994, 1995). Warmer temperatures are weakly-to-moderately correlated with precipitation in the Pacific Northwest ($\rho = 0.33$) but no linear temperature-precipitation relationship is seen in the other regions. Leung et al. (2003b) found greater precipitation frequency at higher temperatures in the Pacific Northwest (which was more evident according to simulations than observations) while for other locations the temperature-precipitation relationship was less apparent. Specific humidity is positively correlated with precipitation in all regions ($\rho = 0.48$-0.72). Wind speed is strongly associated with precipitation in the Pacific Northwest ($\rho = 0.76$) and to a lesser extent the West Coast ($\rho = 0.59$), with a weak correlation observed in the Southwest ($\rho = 0.33$), and no relationship observed for the other regions. This precipitation-wind relationship has been studied previously. For example, Cayan and Roads (1984) found a positive correlation between precipitation and zonal and meridional wind in California and the Pacific Northwest.

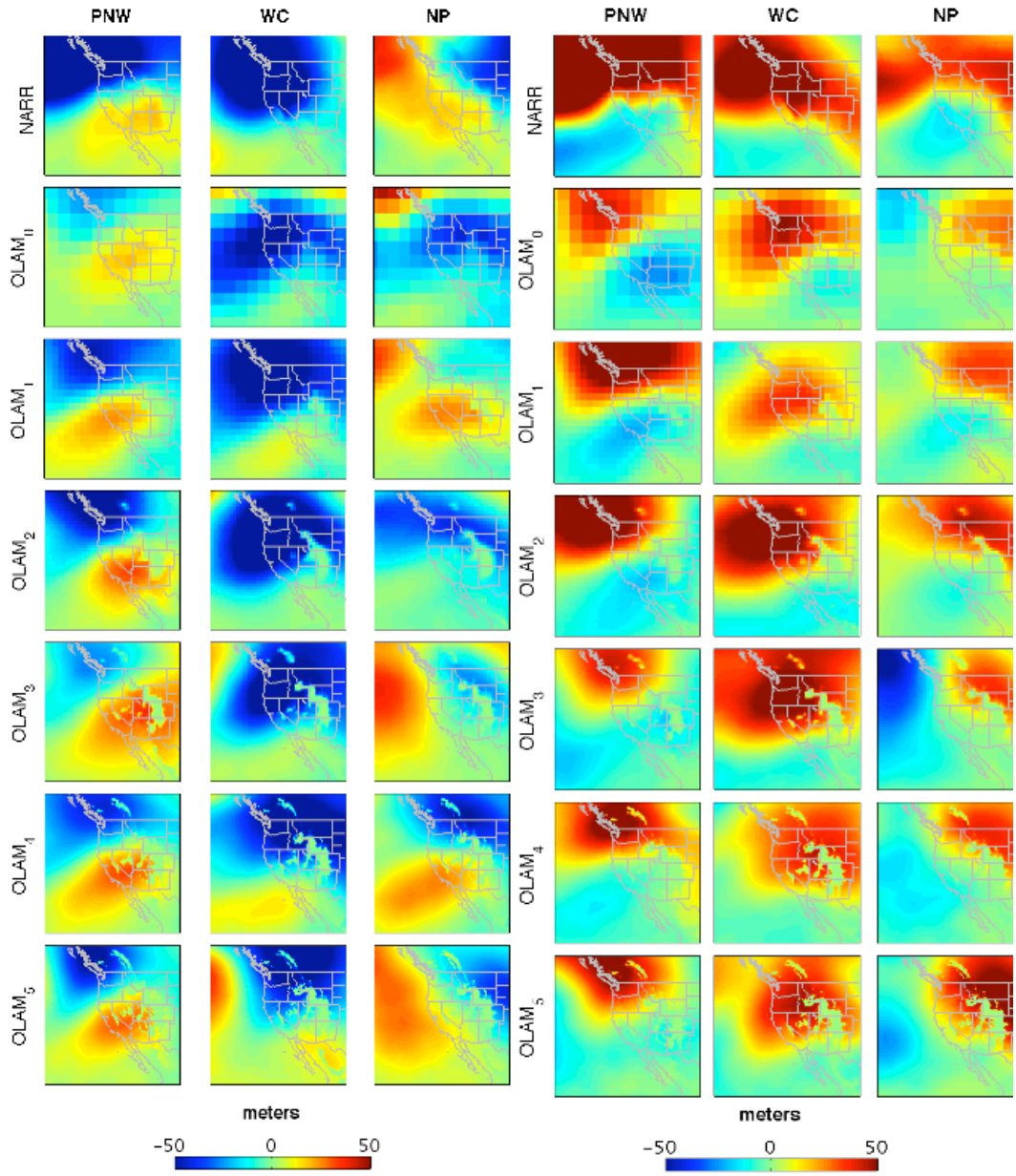

Figure 5.9: 850-mb geopotential height anomalies for (left) P85 and (right) P15 events in the specified region. The regions shown are those where the linear association between precipitation and geopotential height is moderate or better (ρ>0.4, Tab. 5.1)

Figs. 5.9-5.11 illustrate the synoptic conditions for the wettest days in a given region for NARR and the six OLAM simulations to show how the synoptic meteorology differed on the wettest days as compared to the overall mean. In these plots, those days where precipitation exceeds the 85th percentile for the specified region are determined,

and then the difference between the mean over these days and the seasonal mean are plotted for 850-mb geopotential height, wind speed, and specific humidity. Specifically, the field plotted is

$$\frac{1}{N}\sum_{k=1}^{N} X_k - \frac{1}{M}\sum_{k=1}^{M} X_k , \tag{5.1}$$

where X is the 850-mb atmospheric variable, N is the number of days when precipitation exceeds the 85th percentile and M is the number of days in the season. Figures 5.9, 5.11, and 5.12 show the anomaly plots for the driest days when precipitation fell below the 15th percentile in the specified region. Only those regions where the linear association between precipitation and the atmospheric variable of interest is moderate or better according to NARR (ρ>0.4,Tab. 5.1) are shown.

Fig. 5.9 shows the geopotential height anomalies for NARR and the OLAM simulations for P85 events. Considering first the NARR fields, when the PNW is wettest a dipole pattern is observed with low pressure over the Pacific Northwest and Northern Plains and a high-pressure anomaly to the south. This low-pressure anomaly migrates south to extend over northern California when the WC is wettest, during which time the high-pressure anomaly to the south is less pronounced. Storms in the NP occur when low pressure is centered locally and with high-pressure anomalies to the west. OLAM reproduces the PNW dipole pattern well, with correlation coefficients above 0.82 for all simulations (Fig. 5.6a). The strength of the high- and low-pressure anomalies shows a tendency to increase with increasing spatial resolution. Bias (Fig. 5.7a) is positive due to the stronger positive anomaly represented by OLAM. Bias is largest for $OLAM_0$ due to the comparatively weak low-pressure anomaly and for $OLAM_3$ where the high-pressure center is stronger and wider spread. During the wettest days in the WC region, the

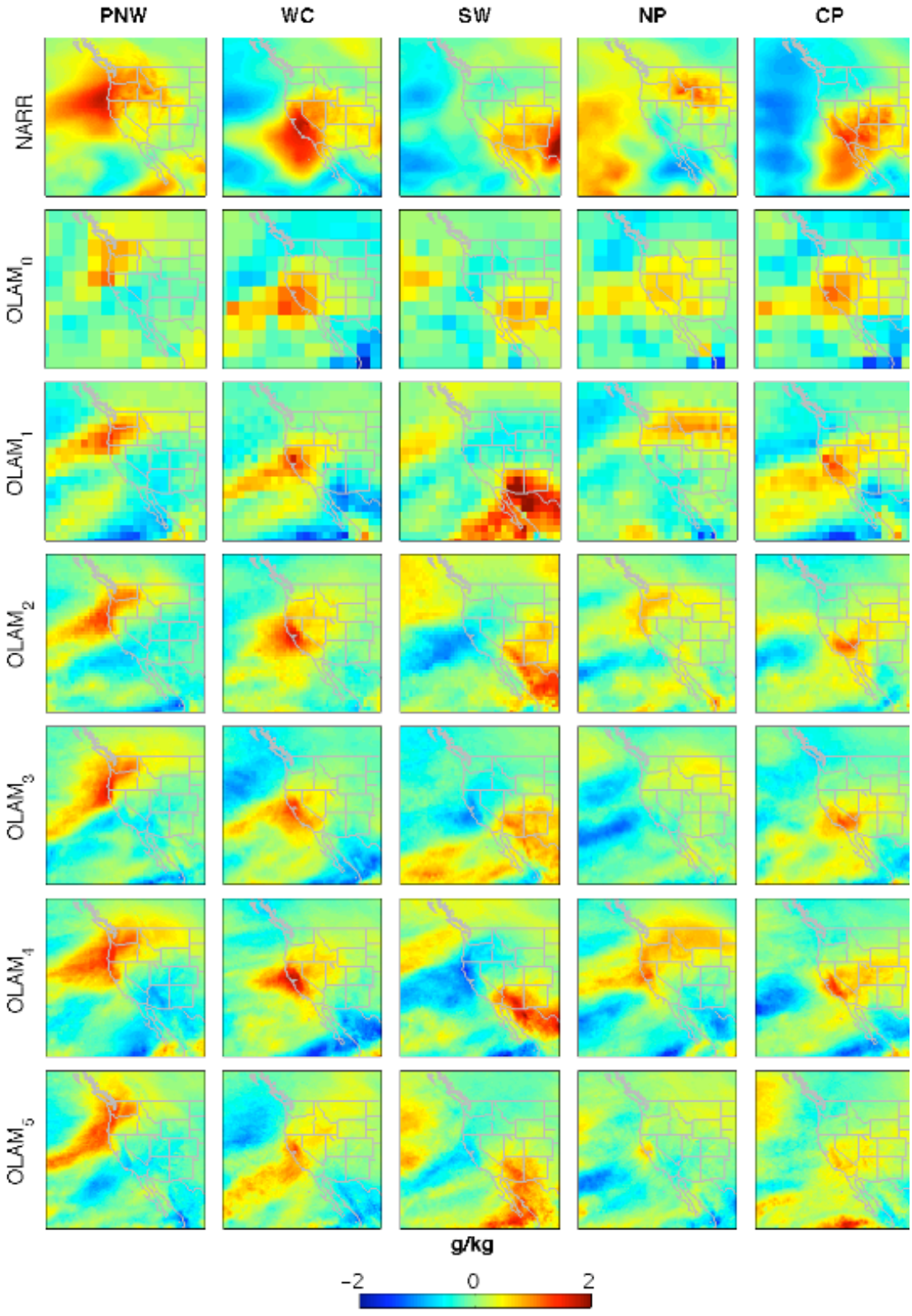

Figure 5.10: As in Fig. 5.9 but showing specific humidity anomalies during P85 events.

southward migration of the low-pressure anomaly to reach northern California is observed in the OLAM simulations as in NARR, and (like NARR) the high-pressure anomaly to the south is less pronounced than for wet events in the PNW. The correlation with NARR is less than for the PNW (0.48-0.91) due to the extension of the low-pressure anomalies eastward in the OLAM simulations. $OLAM_5$ shows the lowest correlation with NARR due to the high-pressure anomaly in the Pacific. For the NP, all OLAM simulations show a local high-pressure anomaly during P85 events, although the strength and extent of the anomaly varies. During NP storms, a high-pressure anomaly over the Pacific is observed for the OLAM simulations and NARR. An important exception is $OLAM_2$, which does not represent this feature. This anomaly is shifted south according to $OLAM_4$ and two high-pressure anomalies are seen for $OLAM_1$. The bias (Fig. 5.7a) is negative for $OLAM_0$, $OLAM_2$, and $OLAM_4$, which show larger low-pressure anomalies and smaller or weaker high-pressure ones. During dry precipitation events, also shown in Fig. 5.9, the geopotential height pattern is generally the inverse of what is seen for P85 events. The general features are captured by OLAM. Bias (Fig. 5.7b) is negative in the PNW and WC for dry days because of smaller high-pressure anomalies according to OLAM. During P15 events in the NP, $OLAM_3$-$OLAM_5$ show a low-pressure anomaly over the Pacific Ocean, which is not represented by NARR.

Fig. 5.10 shows the specific humidity anomalies corresponding to the wettest days in each region. NARR shows that heavy precipitation is associated with high specific humidity locally. When it is wet in the PNW, specific humidity is high over this region and this is similarly true for the other regions. This is also observed by OLAM. For the PNW and WC, OLAM shows the appropriate increase in local specific humidity,

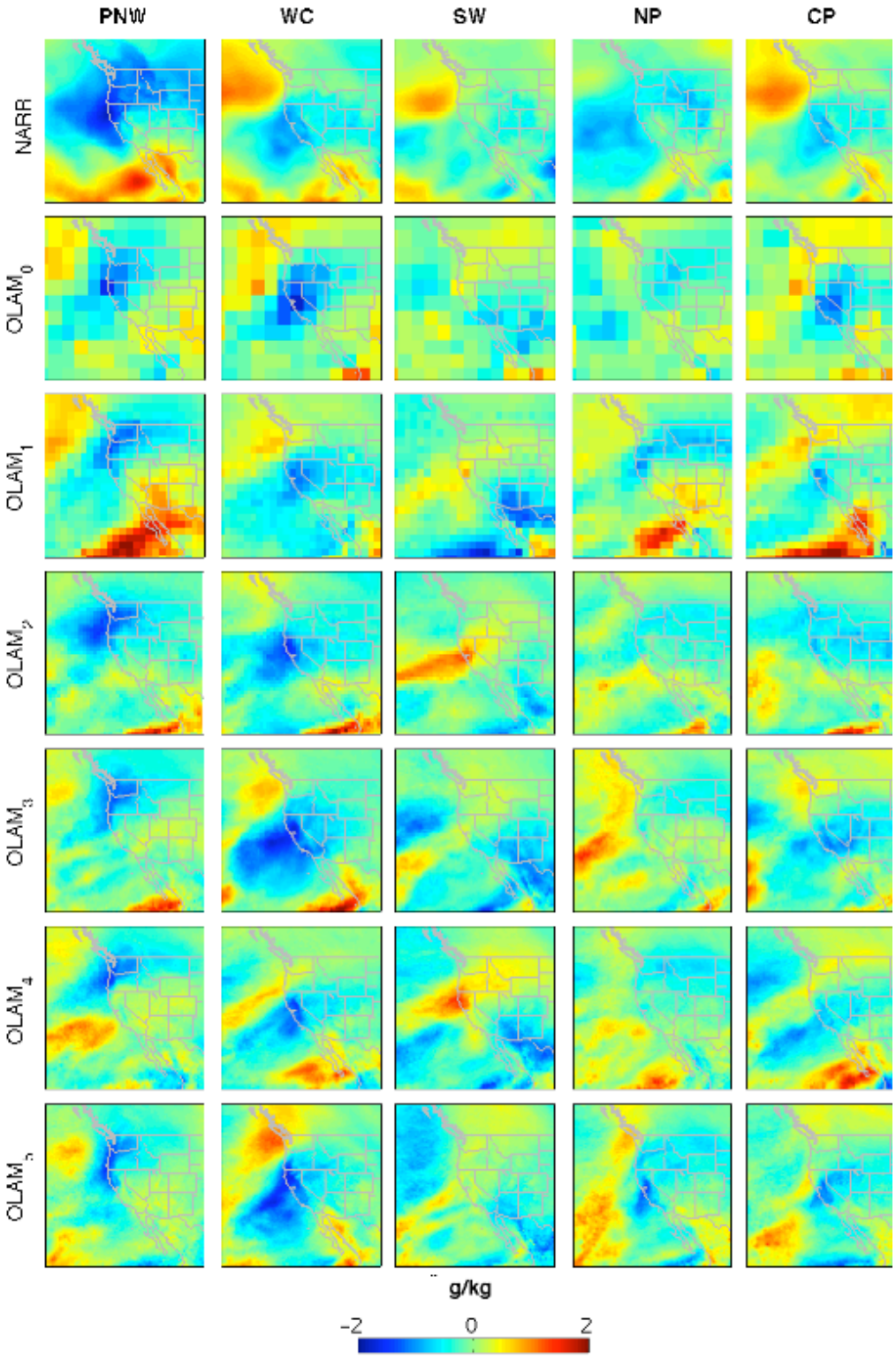

Figure 5.11: As in Fig. 5.10 but during P15 events.

however, the spatial extent of the moist anomaly is smaller than for NARR. The NARR-OLAM correlation (Fig. 5.6a) for the PNW (WC) is in the range of 0.39-0.56 (0.48-0.68) and bias (Fig. 5.7a) in both regions is generally negative indicating a weaker high-moisture anomaly. For the SW, the anomaly over Arizona, New Mexico and northern Mexico is similar in OLAM and NARR except for $OLAM_1$ and $OLAM_2$, which show a weaker and stronger anomaly, respectively. Differences between OLAM and NARR are also apparent over the Pacific during wet events in the SW. The correlation between anomaly fields during strong SW storms SW is 0.30-0.41. In the NP region, OLAM shows the local anomaly similarly as NARR but with important differences. The anomaly is weak according to $OLAM_3$ and $OLAM_5$, the anomaly is shifted to the west according to $OLAM_1$ and $OLAM_2$, and differences are also observed over the Pacific. The correlation between NARR and OLAM is weak during heavy NP storms ($\rho = -0.20$-0.24). In the CP, OLAM shows a positive anomaly centered over California and spreading into the CP region. The strength of the anomaly is much weaker locally over the CP region than is observed by NARR. The low moisture seen over the Pacific according to NARR is not similarly observed by OLAM. The correlation (Fig. 5.6a) between NARR and OLAM during CP storms is 0.16-0.59 and bias (Fig. 5.7a) is positive, largely due to differences over the Pacific. During dry events (Fig. 5.11), negative moisture anomalies are observed locally according to NARR and OLAM. The correlation (Fig. 5.6b) is similar as for the wet events with the greatest association observed for the PNW and WC and the poorest association seen for the PNW moisture fields. Bias (Fig. 5.7b) is positive in the PNW and WC and negative for the WC and NP due to weaker and stronger low-moisture anomalies, respectively, as compared to NARR.

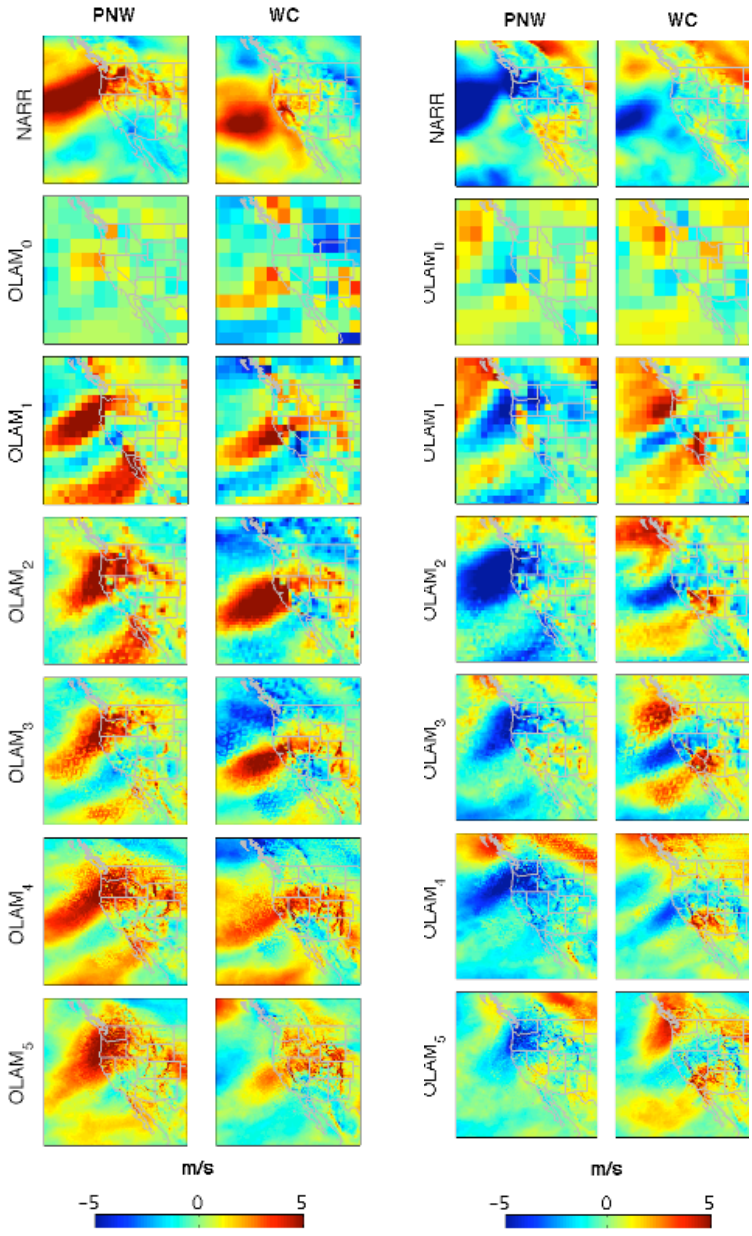

Figure 5.12: As in Fig. 5.9 but for 850-mb wind speed.

Fig. 5.12 gives the wind speed anomalies for the wettest storms over the PNW and WC. According to NARR, heavy precipitation in these regions is associated with strong winds locally. During PNW P85 events, strong winds are observed over the Pacific Northwest and Northern Plains. During wet WC events, strong winds are

observed over California and Nevada. This is also observed for the OLAM simulations. However, the positive wind speed anomalies are very subtle according to $OLAM_0$. The OLAM simulations differ from NARR during PNW storms in that strong winds are also observed to the south in the vicinity of Baja California according to the OLAM simulations. The correlation (Fig. 5.6a) with NARR is generally moderate ($0.34 \leq \rho \leq 0.53$), excepting $OLAM_0$ during strong PNW storms and $OLAM_5$ during strong WC storms. During dry events, also shown in Fig. 5.12, weak winds occur locally according to both NARR and OLAM. However the negative anomalies represented by OLAM during WC dry events are much weaker for $OLAM_1$, $OLAM_2$, and $OLAM_5$. Further, these negative anomalies are shifted inland according to the OLAM simulations. The correlation (Fig. 5.7b) between NARR and $OLAM_2$-$OLAM_5$ for PNW dry events is moderate ($0.41 \leq \rho \leq 0.60$), and this association is reduced for WC p15 events ($0.21 \leq \rho \leq 0.48$). For WC dry events, while OLAM generally shows weak winds locally, strong winds are seen over much of the Pacific and coastal regions, which is not similarly observed by NARR.

Precipitation in the western U.S. is associated with flow direction (Daly 1994, Leung et al. 2003a, 2003c). In the Pacific Northwest and California strong precipitation is associated with a southwesterly flow direction, though not independently from moisture flux, which is also higher during southwesterly flow (e.g. Leung et al. 2003a, 2003c). Fig. 5.13 shows a plot of precipitation versus wind direction for each region. According to NARR, in the PNW and WC the strongest precipitation occurs when wind flows southwesterly. Strong precipitation in the NP and CP show a preference for a southwesterly to westerly flow. The wind direction-precipitation relationship is less

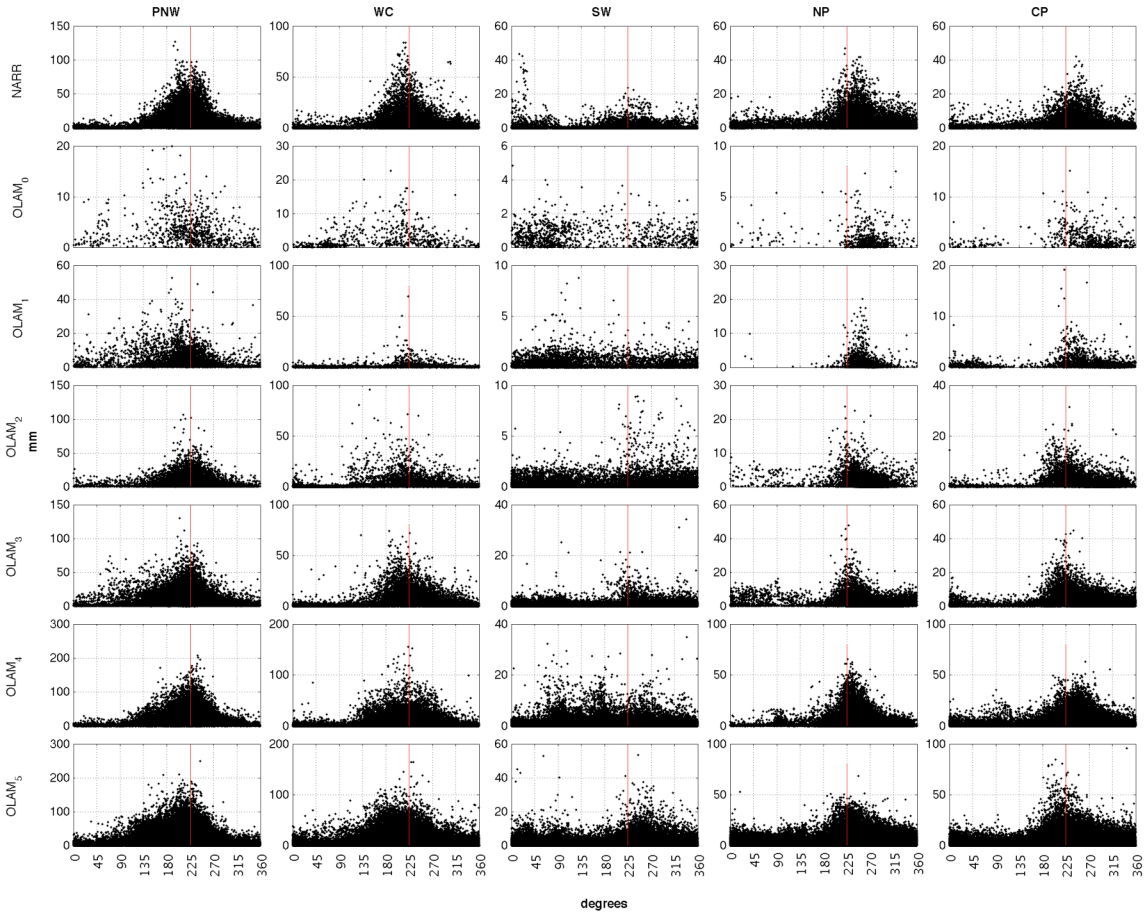

Figure 5.13: Wind direction versus precipitation for each region. The vertical line marks the southwesterly wind direction (225°).

apparent in the SW, although there appears to be some preference for storms to occur during southwesterly or northerly (near 0°) wind conditions. In the PNW, OLAM represents the wind direction-precipitation relationship well in general especially for $OLAM_2$-$OLAM_5$. However there are differences in the range of precipitation magnitudes, which is related to the model spatial resolution. For $OLAM_0$ the data are sparser and the wind direction-precipitation relationship more poorly defined, and with $OLAM_1$ there is more heavy events during southerly and southeasterly flow. In the WC, the grouping of high precipitation events during southwesterly wind is observed for $OLAM_3$-$OLAM_5$. $OLAM_0$ shows more strong events occurring with southerly and

southeasterly winds than is observed by NARR. A greater preference for southerly storms is also seen by $OLAM_4$ and $OLAM_5$. For the SW, the OLAM data differ from NARR, possibly because there is a weaker precipitation-wind direction relationship in this region. No directional preference is seen for $OLAM_0$, $OLAM_1$ or $OLAM_4$, and a northwesterly peak is seen for $OLAM_2$. $OLAM_3$ and $OLAM_5$ show the greatest similarity with NARR in the SW. In the NP, OLAM generally reproduces the southwesterly to westerly peak observed by NARR, however, with a notable shift towards strictly southerly flow for $OLAM_3$-$OLAM_5$. Generally, OLAM reproduces the wind direction-precipitation relationship in the CP, with the strongest precipitation events clustered around southerly or southeasterly flow. As in the NP region, in the CP the OLAM curves show a shift towards strict southwesterly flow as compared with NARR.

5.5. Summary and Conclusions

The utility of numerical models for long-range hydroclimatological forecasts has been investigated by a number of researchers. Most recently, the use of regional models nested within GCMs has shown improvements in seasonal hindcasts over those conducted by coarser resolution GCMs alone. This study has looked at the ability of the Ocean-Land-Atmosphere Model, which has grid refinement capabilities, to simulate the hydroclimate of the western U.S. during the winter of 1999. Six simulations were run, each with a different spatial resolution over the western U.S. OLAM was configured with a global grid having a horizontal characteristic length scale of 360 km, and for tests of sensitivity to horizontal grid size, additional grid refinement levels were configured over the western U.S. to have a CLS of 180, 90, 45, 23, and 11 km.

An examination of simulated versus observed precipitation showed the greatest improvements for spatial grid size increases between 360 and 45 km. This is likely due to the fact that at 45 km, some important features of topography are resolved such as the distinction between the Cascades and the Rockies, and the separation between the Sierra Nevada and the high desert elevations to the east. Further improvement is seen in the representation of topography by increasing the spatial scale beyond 45 km. $OLAM_3$-$OLAM_5$ represented the magnitude and spatial variability of precipitation well, which was generally found to be within the range of observational data uncertainty. $OLAM_3$ compared well against NARR, which has a comparable spatial resolution (23 x 23 km^2 and 32 x 32 km^2, respectively). However, the Rocky Mountain orographic signal is stronger according to $OLAM_3$. $OLAM_5$ (11 x 11 km^2) compared well against VIC (14 x 14 km). However, $OLAM_5$ produced drier conditions west of the Cascades and Sierra Nevada Mountains. $OLAM_2$ (0.8° x 0.8°) performed well against the 1° x 1° observational dataset, excepting the tendency for OLAM to produce wetter conditions east of the Cascades. The GPCP1dd does not represent the Rocky Mountain orographic signal so $OLAM_2$ performed poorly against this dataset for inland parts of the domain. Similarly, the other observational datasets were in disagreement with the GPCP1dd in these parts.

Observed and simulated synoptic meteorological conditions at 850-mb were examined during the heaviest precipitation events (precipitation>85^{th} percentile) and for the driest days (precipitation<15^{th} percentile). Generally, OLAM was able to reproduce anomalies in atmospheric fields that were observed during wet and dry periods. Simulated anomalies in geopotential height were better represented over the Pacific

Northwest and West Coast than for the Northern Plains. For specific humidity, OLAM correctly simulates moist anomalies over the region when heavy precipitation occurs, however the size of the anomaly is generally smaller and/or weaker than observed. Similarly, OLAM correctly represents negative humidity anomalies during dry periods, although the anomaly tends to be smaller according to OLAM as compared to NARR. OLAM better represents the humidity fields during precipitation extremes occurring in the Pacific Northwest and West Coast than it does for extremes occurring in other regions. For extremes in the Northern Plains simulated specific humidity anomalies compare poorly with observations and there are large pattern differences across simulations.

For wind speed, OLAM shows an association between strong winds locally and heavy precipitation in the Pacific Northwest and West Coast similar to that observed. However, for events in the WC, there are large pattern differences between the OLAM simulation and NARR. In general, a moderate correlation exists between the OLAM simulations and observational anomaly maps. The OLAM simulations show a similar relationship between wind direction and heavy precipitation to that observed. Generally heavy precipitation in the Pacific Northwest and the West Coast is associated with southwesterly flow and storms in the Northern Plains and Colorado Plateau show a preference for a southwesterly to westerly flow. Some OLAM simulations showed a shift to the strictly southwesterly in the wind direction-precipitation curves for the NP and CP. The wind direction-precipitation relationship is poorly defined in the Southwest according to observations and OLAM.

One major limitation of this work is the inclusion of only a single season when multiple seasons would be desirable for truly evaluating OLAM's predictive skill. However, a more thorough analysis requires computing capabilities that were not readily available. While this series of 90-day simulations is limited from a climate perspective, each simulation represents a 90-day realization experiment given that each day is different from the previous day. These findings are encouraging from a seasonal forecasting perspective, since it demonstrates the capability of a global climate model, driven only by SSTs and initial conditions, to represent the meteorological features associated with daily precipitation variability. SSTs and orographic forcing heavily influence weather in this part of the world and the level of skill may differ for other locations where local convective activity and soil-memory processes are more dominant. Additional simulations are planned for the future and any findings, supportive or contradictory to those presented here, shall be made available.

CHAPTER 6

General Summary, Conclusions and Recommendations for Future Work

Estimating precipitation in the western United States remains an unsolved and challenging problem due to deficiencies in observations and model simulations over complex terrain. A systematic evaluation of precipitation observations in the western U.S. has been presented in two parts. The first used principal component-based regionalization methods to isolate unique precipitation climates in the western U.S., and the second part focused on interannual variability, distributions, time series, and persistence. In each case, comparative analyses were conducted using nine state-of-the-art observational datasets commonly used in hydrometeorological research.

The results of the observational data analysis identified five unique precipitation climates within the western U.S., which have centers and boundaries that are physically reasonable and highlight the relationship between the precipitation climatology and local topography. These results were used to provide location-specific assessments of uncertainty and to evaluate model performance for the different regions individually. The regionalization methodology was shown to be stable to the choice of dataset and to changes in spatial and temporal resolution. Data differences do exist, however. By physically interpreting the regionalization and loading pattern differences in terms of raw precipitation data, the dominant data differences were identified as being due to

differences in the representation of (i) precipitation over the Rocky Mountains, (ii) the eastward wet-to-dry precipitation gradient that occurs during the cold season, (iii) the magnitude and spatial extent of the North American Monsoon signal, and (iv) precipitation in the desert southwest during spring and summer.

The dataset intercomparison study identified two groups of similar datasets. The GPCP, CMAP, and GPCC (Group 1) were shown to behave similarly as were USMex, NARR, VIC, and PRISM (Group 2). In general, Group 1 depicts drier conditions relative to the other. Sampling issues related to sparse rain gauge coverage over the western U.S. by the GPCC rain gauge dataset, which also used in the development of the GPCP and CMAP, likely explains the dry bias of Group 1.

Interannual variability is similar among data with respect to the timing of precipitation anomalies, but important differences exist in the spatial distribution of specific anomalous events. Data uncertainty is high with respect to distributions and low for studies of persistence. Spatially, the greatest uncertainty occurs along the path of the Rocky Mountains and in California. Overall, data similarity is lowest in summer.

For the modeling component of this research, OLAM was evaluated for its ability to simulate the hydroclimatology of the western U.S. during the winter of 1999. OLAM was configured as for a long-range seasonal hindcast but with observed SSTs. One simulation ($OLAM_0$) was run at 360-km globally, and five additional simulations were run with successively higher spatial resolutions over the western United States with a CLS of 180, 90, 45, 23, and 11 km ($OLAM_1$, $OLAM_2$, $OLAM_3$, $OLAM_4$, and $OLAM_5$, respectively).

An examination of simulated versus observed precipitation showed the greatest improvements for resolution increases between 360 and 45 km, likely because at 45 km some large scale features of topography are resolved such as the distinction between coastal and inland mountains. Further increases in spatial grid size resolves details of individual mountain ranges. $OLAM_3$-$OLAM_5$ represented the magnitude and spatial variability of precipitation well, which was generally found to be within the range of observational data uncertainty. $OLAM_3$ compared well against NARR, which has a comparable spatial resolution (23 x 23 km^2 and 32 x 32 km^2, respectively), however, with a stronger Rocky Mountain orographic signal according to $OLAM_3$. $OLAM_5$ (11 x 11 km^2) compared well against VIC (14 x 14 km^2), however, with drier conditions west of the Cascades and Sierra Nevada Mountains according to $OLAM_5$. $OLAM_2$ (0.8° x 0.8°) performed well against to the 1° x 1° observational dataset, excepting the tendency for OLAM to produce wetter conditions east of the Cascades.

Observed and simulated synoptic meteorological conditions were examined during the wettest and driest events. Specifically, we considered 850-mb geopotential height, specific humidity and wind speed when precipitation exceeded the 85th percentile or fell below the 15th in each of the unique precipitation regions identified in Chapter 3. While there were some magnitude and pattern differences, OLAM was shown to reproduce the appropriate anomaly fields, which is encouraging since it demonstrates the capability of a global climate model, driven only by SSTs and initial conditions, to represent meteorological features associated with daily precipitation variability.

One major limitation of this work is the evaluation over only a single season. To truly evaluate predictive skill, long-term simulations are needed to assess whether OLAM

adequately represents seasonal anomalies in the western U.S. such as those associated with ENSO. Medvigy et al. (2008a) showed OLAM capable of representing ENSO-related precipitation variability in the Amazon. However, the model's ability to predict seasonal extremes was shown heavily dependent on the spatial scale of the Andes. The representation of the Andes was additionally shown to affect precipitation in other parts of the globe. This type of investigation should be conducted for the western U.S. to improve understanding of teleconnections associated with the orography there, and to see if such a relationship exists between inter-seasonal predictive skill and model resolution. Such a study is not possible with a one-way nested GCM-RCM configuration since the effects of the high-resolution topography would not feed back to the large scale. In this regard, OLAM would be particularly well suited. Ultimately, global simulations that resolve at a higher spatial scale several regions that are known or suspected to have global impacts would enhance our understanding of teleconnection patterns, and may improve predictability or our understanding of predictive limits.

Additional work is needed to study model sensitivity to subgid-scale parameterizations, especially for cumulus convection. Model skill is known to be highly sensitive to cumulus parameterizations, and this sensitivity can vary geographically and seasonally. This study used the Grell-Dévényi convection scheme, which has several dependent parameters that can be modified (e.g. precipitation efficiency and the weighting of each closure assumption). Optimization techniques would be useful for finding the best parameters for the western U.S., and ultimately for other parts of the world, possibly varying them by season. The regionalization technique presented in Chapter 3 would be useful for such optimization studies, as it provides an objective

method to subdivide a domain. Similarly, regionally focused optimization techniques could be applied to the bulk microphysics scheme or any other OLAM parameterization. Medvigy et al. (2008b) conducted such an optimization study for OLAM's radiative transfer parameters. In that study a simple, somewhat subjective regionalization was employed.

There remains the issue of observational data uncertainty because observations are needed for model evaluation. For the western U.S. the selection of a small group of datasets was useful for understanding model performance, and this approach is recommended where available. However, most parts of the world are not so well covered by observations. Globally available options are the GPCP and CMAP merged satellite-gauge products, NCEP global reanalysis, and TRMM. However, TRMM is available over a shorter time domain (from 1998) and its coverage is limited to parts equatorward of 40°. When global datasets are needed, the recommendation based on this research is to use at least a pair of unlike datasets for model evaluation over land. The model would then be evaluated within a range of data uncertainty. The GPCP and CMAP demonstrate similar bias over land and contain many of the same data sources, therefore they are not considered 'unlike' datasets. NCEP2 should be used with knowledge of possible bias when representing convective precipitation events.

APPENDIX

List of Acronyms

AS	Arakawa and Schubert (1974)
AVHRR	Advanced Very High Resolution Radiomete
CDF	Cumulative Density Function
CLS	Characteristic Length Scale
CMAP	CPC Merged Analysis of Precipitation
CP	Colorado Plateau
CPC	Climate Prediction Center
CRU	Climate Research Unit
DEM	Digital Elevation Model
DOE	Department of Energy
ENSO	El-Niño Southern Oscillation
GCM	General Circulation Model
GD	Grell and Dévényi (2002)
GMFD	Global Meteorological Forcing Dataset
GOES	Geostationary Operational Environmental Satellites
GPCC	Global Precipitation Climatology Center
GPCP	Global Precipitation Climatology Project
GPCP1dd	Global Daily Merged Precipitation Analysis of the GPCP

GPI	GOES Precipitation Index
IQR	Interquartile Range
IR	Infrared
LEAF	Land Ecosystem-Atmosphere Feedback Model
LEV	Log-eigenvalue
MM5	Pennsylvania State University/NCAR Mesoscale Model
MSU	Microwave Sounding Unit
NAM	North American Monsoon
NARR	North American Regional Reanalysis
NASA	National Aeronautics and Space Administration
NOAA	National Atmospheric and Oceanic Administration
NCEP	National Centers for Environmental Prediction
NCEP2	NCEP–DOE Reanalysis 2
NRMSE	Normalized Root Mean Square Error
NP	Northern Plains
OLAM	Ocean-Land-Atmosphere Model
P85	precipitation exceeding the 85^{th} percentile
P15	precipitation below the 85^{th} percentile
PC	Principal Component
PCA	Principal Component Analysis
PNW	Pacific Northwest
PRISM	Precipitation-elevation Regressions on Independent Slopes Model
QC	Quality Control

RAMS	Regional Atmospheric Modeling System
RCM	Regional Climate Model
RPC	Rotated Principal Components
SGS	Subrid Scale
SSM/I	Special Sensor Microwave Imager
SST	Sea Surface Temperature
SW	Southwest
TIROS	Television and Infrared Observational Satellite
TMPI	Threshold-Matched Precipitation Index
TRMM	Tropical Rainfall Measuring Mission
TOVS	TIROS Operationsl Vertical Sounder
URD	Unified Raingauge Dataset
USMex	CPC retrospective United States and Mexico daily precipitation analysis
VIC	Variable Infiltration Capacity
WC	West Coast

REFERENCES

Adam, J.C., and D.P. Lettenmaier, 2003: Adjustment of global gridded precipitation for systematic bias. *J. Geophys. Res.,*108, 4257, doi:10.1029/2002JD002499.

Adcroft, A., C. Hill, and J. Marshall, 1997: Representation of Topography by Shaved Cells in a Height Coordinate Ocean Model. *Mon. Wea. Rev.*, **125**, 2293–2315.

Adler, R.F., A.J. Negri, P.R. Keehn, and I.M. Hakkarinen, 1993: Estimation of monthly rainfall over Japan and surrounding waters from a combination of low-orbit microwave and geosynchronous IR data. *J. Appl. Meteor.*, 32, 335–356.

Adler, R.F., C. Kidd, G. Petty, M. Morissey, and H.M. Goodman, 2001: Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP-3). *Bull. Amer. Meteor. Soc*., **82**, 1377–1396.

Adler, R.F., and Coauthors, 2003: The version-2 Global Precipitation Climatolog
y Project (GPCP) monthly precipitation analysis (1979–Present). *J. Hydrometeor*., **4**, 1147–1167.

Arakawa, A., and W.H. Schubert, 1974: Interaction of a Cumulus Cloud Ensemble with the Large-Scale Environment, Part I. *J. Atmos. Sci.*, **31**, 674–701.

Arkin, P.A., and B.N. Meisner, 1987: The relationship between large-scale convective rainfall and cold cloud over the Western Hemisphere during 1982–84. *Mon. Wea. Rev*., 115, 51–74.

Arkin, P.A., and P. Xie, 1994: The Global Precipitation Climatology Project: First algorithm intercomparison project. *Bull. Amer. Meteor. Soc*., 75, 401–419.

Avissar, R. and Y. Mahrer, 1988: Mapping frost-sensitive areas with a three-dimensional local scale numerical model. Part I: Physical and numerical aspects. *J. Appl. Meteor.,* **23**, 400-413.

Avissar, R., P.L. Silva Dias, M.A.F. Silva Dias, and C.A. Nobre, 2002. The Large-Scale Biosphere-Atmosphere Experiment in Amazonia (LBA): Insights and future research needs, *J. Geophys. Res.*, **107**, 8086, doi:10.1029/2002JD002704.

Avissar, R., and D. Werth, 2005: Global Hydroclimatological Teleconnections Resulting from Tropical Deforestation. *J. Hydrometeor.*, **6**, 134–145.

Barnston, A.G., Y. He, and D.A. Unger, 2000: A Forecast Product that Maximizes Utility

for State-of-the-Art Seasonal climate Prediction *Bull. Amer. Meteor. Soc.*, Vol. 81, No. 6, pp. 1271-1279.

Barros, A. P. and D. P. Lettenmaier, 1994: Dynamic Modeling of Orographically Induced Precipitation. *Rev. Geophys.*, **32,** 265–284.

Barros, A.P., 1995: Adaptive Multilevel Modeling of Land-Atmosphere Interactions. *J. Climate*, **8**, 2144–2160.

Barros, A. P. and W. Hwu, 2002: A Study of Land-Atmosphere Interactions During Summertime Rainfall Using a Mesoscale Model. *J. Geophys. Res.- Atmospheres*, **107**,10.1029/2000JD000254.

Beaudoin, P., and J. Rousselle, 1982: A study of space variations of precipitation by factor analysis. *J. Hydrol*., 59, 123–138.

Beven, K.J. and M.J. Kirby, 1979: A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.,* **24,** 43–69.

Beven, K.J., 1982: On subsurface stormflow: An analysis of response times. *Hydrol. Sci. J.,* **27,** 505–521.

Beven, K.J., 1984: Infiltration into a class of vertically non-uniform soils. *Hydrol. Sci. J.,* **29,** 424–434.

Braun, S.A., R. Rotunno, and J.B. Klemp, 1999: Effects of Coastal Orography on Landfalling Cold Fronts. Part I: Dry, Inviscid Dynamics. *J. Atmos. Sci.*, **56**, 517–533.

Bretherton, C. S., M. Widmann, V. P. Dymnikov, J. M. Wallace, and I. Bladé, 1999: The effective number of spatial degrees of freedom of a time-varying field. *J. Climate*, 12, 1990–2009.

Broccoli, A. J. and S. Manabe, 1992: The Effects of Orography on Midlatitude Northern-Hemisphere Dry Climates. *J. Climate*, **5,** 1181-1201.

Brown, J.M., 1979: Mesoscale unsaturated downdrafts driven by rainfall evaporation: A numerical study. *J. Atmos. Sci*. **36**, 313-338.

Buell, C.E., 1979: On the physical interpretation of empirical orthogonal functions. Preprints, Sixth Conf. on Probability and Statistics in the Atmospheric Sciences, Baniff, AB, Canada, Amer. Meteor. Soc., 112–117.

Carter, M.M., and J. B.Elsner, 1997: A statistical method for forecasting rainfall over Puerto Rico. *Wea. Forecasting*, 12, 515–525.

Cattell, R.B., 1966: The screen test for the number of factors. Multivariate *Behav. Res*., 1,

245–276.

Cayan D.R., and J.O. Roads, 1984: Local relationships between United States West Coast precipitation and monthly mean circulation parameters. *Mon. Wea. Rev.*, **112**, 1276–1282

Cayan D.R., and J.O. Roads, 1993: The influence of precipitation and temperature on seasonal dtreamflow in California, *Water Resour. Res*., **29**, 1127-1140, 1993.

Cayan D.R., L.G. Riddle, and E. Aguado, The influence of precipitation and temperature on seasonal dtreamflow in California, *Water Resour. Res*., **29**, 1127-1140, 1993.

Chen, M., P. Xie, J. E. Janowiak, and P.A. Arkin, 2002: Global land precipitation: A 50-yr monthly analysis based on gauge observations. *J. Hydrometeor*., 3, 249–266.

Cocke, S., and T.E. LaRow, 2000: Seasonal predictions using a regional spectral model embedded within a coupled ocean-atmosphere model. *Mon. Wea Rev.,* **128**, 689-708.

Colle, B.A., C.F. Mass, and K.J. Westrick, 2000: MM5 Precipitation Verification over the Pacific Northwest during the 1997–99 Cool Seasons. *Wea. Forecasting*, **15**, 730–744.

Comrie, A. C., and E. C. Glenn, 1998: Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability. *Climate Res*., 10, 201–215.

Costa, M.H., and J.A. Foley, 1998: A comparison of precipitation datasets for the Amazon basin. *Geophys. Res. Lett*., **25**, 155–158.

Cressman, G.P., 1959: An operational objective analysis system. *Mon. Wea. Rev*., 87, 367–374.

Daly, C., R.P. Neilson, and D.L. Phillips, 1994: A statistical–topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor*., **33**, 140–158.

Diem, J.E., 2006: Synoptic-scale controls of summer precipitation in the southeastern United States. *J. Climate*, 19, 613–621.

Diem, J.E. and D.P. Brown, 2006: Tropospheric moisture and monsoonal rainfall over the southwestern United States. *J. Geophys. Res*., 111, D16112, doi:10.1029/2005JD006836.

Dinpashoh, Y., A. Fakheri-Fard, M. Moghaddam, S. Jahanbakhsh, and M. Mirnia, 2004: Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *J. Hydrol*., 297, 109–123.

D'Odorico, P. and A. Porporato, 2004: Preferential states in soil moisture and climate dynamics. *Proc. Nat. Acad. Sci.*

Easterling, D.R., 1990: Persistent patterns of thunderstorm activity in the central United States. *J. Climate*, 3, 1380–1389.

Ebert, E.E., M.J. Manton, P.A. Arkin, R. J. Allam, G. E. Holpin, and A. Gruber, 1996: Results from the GPCP Algorithm Intercomparison Programme. *Bull. Amer. Meteor. Soc.*, **77**, 2875–2887.

Ehrendorfer, M., 1987: A regionalization of Austria's precipitation climate using principal component analysis. *Int. J. Climatol*., 7, 71–89.

Ellis, A.W., and T.W. Hawkins, 2001: An apparent atmospheric teleconnection between snow cover and the North American monsoon. *Geophys. Res. Lett., 28, 2653–2656.*

Fennessy, M.J., and J. Shukla, 2000: Seasonal prediction over North America with a regional model nested in a global model. *J. Climate*, **13**, 2605-2627.

Frank, W.M. and C. Cohen, 1987: Simulation of tropical convective systems. Part I: A cumulus parameterization. *J. Atmos. Sci*. 44, 3787-3799.

Fritsch, J.M., and C.F. Chappel., 1980: Numerical prediction of convectively driven mesoscale pressure systems. Part I: Convectve parameterzation. *J. Atmos. Sci*. **37**, 1722-1733.

Fuchs, T., U. Schneider, and B. Rudolf, 2007: Global precipitation analysis products of the GPCC. Global Precipitation Climatology Centre, Deutscher Wetterdienst, 12 pp.

Gal-Chen, T., and R.C.J. Somerville, 1975: Numerical solution of the navier-stokes equations with topography. *J. Comput. Phys*., **17**, 276-310, doi:10.1016/0021-9991(75)90054-6.

Gershunov, A. and D. Cayan, 2003: Heavy daily precipitation frequency over the contiguous United States: Sources of climatic variability and seasonal predictability. *J. Climate*, 16, 2752-2765

Giorgi, F., G.T. Bates, and S.J. Nieman, 1993: The Multiyear Surface Climatology of a Regional Atmospheric Model over the Western United States. *J. Climate*, **6**, 75–95.

Giorgi, F., C. Shields Brodeur, and G.T. Bates, 1994: Regional Climate Change Scenarios over the United States Produced with a Nested Regional Climate Model. *J. Climate*, **7**, 375–399.

Giorgi, F. and R. Avissar, 1997: Representation of heterogeneity effects in earth system modeling: Experience from land surface modeling. *Rev. Geophys.*, **35,** 413-438.

Gochis, D.J., L. Brito-Castillo, and W.J. Shuttleworth, 2006: Hydroclimatology of the

North American Monsoon region in northwest Mexico. *J. Hydrol*., 316, 53–70.

Goddard L., S.J. Mason, S.E. Zebiak, C.F. Ropelewski, R.E. Basher, and M.A. Cane, 2001: Current approaches to seasonal to interannual climate predictions. *Int. J. Climatol.*, **21**, 1111–1152

Gottschalck, J., J. Meng, M. Rodell, and P. Houser, 2005: Analysis of multiple precipitation products and preliminary assessment of their impact on global land data assimilation system land surface states. *J. Hydrometeor*., **6**, 573–598.

Grell, G.A., 1993: Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev*.,**121**, 764-787.

Grell, G.A. and Dévényi, D. 2005: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Gephys. Res. Lett*., **29**, 1693-1696.

Gruber, A., X. J. Su, M. Kanamitsu, and J. Schemm, 2000: The comparison of two merged rain gauge–satellite precipitation datasets. *Bull. Amer. Meteor. Soc*., **81**, 2631–2644.

Guirguis, K.J., and R. Avissar, 2008a: A precipitation climatology and dataset intercomparison for the western United States. *J. Hydrometeor*., **9**, 825–841.

Guirguis, K.J., and R. Avissar, 2008b: An analysis of precipitation variability, persistence, and observational data uncertainty in the western United States. *J. Hydrometeor*., 9, 843–865.

Guirguis, K., R. Avissar, R.L. Walko, M. Otte, and D. Medvigy, 2009: Hydroclimatology of the western United States: A winter 1999 case study using the Ocean-Land-Atmosphere Model (OLAM). In preparation.

Gutzler, D. S., 2004: An index of interannual precipitation vari-ability in the core of the North American monsoon region. J. Climate, 17, 4473–4480.

Gutzler, D. S., 2000: Covariability of spring snowpack and summer rainfall across the southwest United States. *J. Climate*, **13,** 4018-4027.

Gutzler, D. S. and J. W. Preston, 1997: Evidence for a relationship between spring snow cover in North America and summer rainfall in New Mexico. *Geophys. Res. Lett.,* **24,** 2207-2210.

Harman, H. H., 1976: Modern Factor Analysis. 3rd ed. The University of Chicago Press, 487 pp.

Hawkins, T.W., A.W. Ellis, J.A. Skindlov, and D. Reigle, 2002: Intra-annual analysis of

the North American snow cover–monsoon teleconnection: Seasonal forecasting utility. *J. Climate, 15, 1743–1753.*

Higgins, R. W., J.E. Janowiak, and Y.-P. Yao, 1996: A gridded hourly precipitation data base for the United States (1963–1993). NCEP/Climate Prediction Center Atlas 1, National Weather Service, NOAA, U.S. Department of Commerce, 47 pp.

Higgins, R.W., K.C. Mo, and S. D. Schubert, 1996: The moisture budget of the central United States in spring as evaluated in the NCEP/NCAR and the NASA/DAO reanalyses. *Mon. Wea. Rev*., **124**, 939–963.

Higgins, R.W., W. Shi, E. Yarosh, and R. Joyce, 2000: Improved United States precipitation quality control system and analysis. NCEP/Climate Prediction Center Atlas 7, National Weather Service, NOAA, U.S. Department of Commerce, 40 pp.

Higgins, R.W., A. Leetmaa, Y. Xue, and A. Barnston, 2000: Dominant factors influencing the seasonal predictability of U.S. precipitation and surface air temperature. *J. Climate*, **13**, 3994–4017.

Hong, S.Y., and A. Leetmaa, 1999: An Evaluation of the NCEP RSM for Regional Climate Modeling. *J. Climate*, **12**, 592–609.

Huffman, G.J., and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bull. Amer. Meteor. Soc*., **78**, 5–20.

Hui, T.W., Chang, W.L., and Shum, K.Y.: 2002, Long-range forecasting for Hong Kong with ensembles from a regional model - some preliminary results. Second APCN Working Group Meeting, Seoul, Republic of Korea,11-13 June 2002. Hong Kong Observatory Reprint No., 480.

Hui, T.W., and Shum, K.Y., 2005: Prediction of Seasonal Rainfall in Hong Kong Using ECPC's Regional Climate Model. The Sixth International RSM Workshop, Palisades, New York, USA, 11-15 July 2005. Hong Kong Observatory Reprint No. 602.

Janowiak, J.E., A. Gruber, C.R. Kondragunta, R.E. Livezey, and G.J. Huffman, 1998: A comparison of the NCEP–NCAR reanalysis precipitation and the GPCP rain gauge–satellite combined dataset with observational error considerations. *J. Climate***, 11**, 2960–2979.

Kain, J.S., and J.M. Fritsch, 1990: A One-Dimensional Entraining/Detraining Plume Model and Its Application in Convective Parameterization. *J. Atmos. Sci.*, **47**, 2784–2802.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc*., **77**, 437–471.

Kanamitsu, M., W. Ebisuzaki, J. Woolen, S.-K. Yang, J.J. Hnilo, M. Fiorino, and G.L. Potter, 2002: NCEP–DEO AMIP-II reanalysis (R-2*). Bull. Amer. Meteor. Soc*., **83**, 1631–1643.

Krishnamurti, T. N., S. Low-Nam, and R. Pasch, 1983: Cumulus parameterization and rainfall rates II. *Mon. Wea. Rev.,* **111,** 815–828..

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc*., **82**, 247–267.

Klazura, G.E., J.M. Thomale, D.S. Kelly, and P. Jendrowski, 1999: A comparison of NEXRAD WSR-88D radar estimates of rain accumulation with gauge measurements for high- and low-reflectivity horizontal gradient precipitation events. *J. Atmos. Oceanic Techno*l., 16, 1842–1850.

Klein, W.H., and H.J. Bloom, 1987: Specification of Monthly Precipitation over the United States from the Surrounding 700 mb Height Field. *Mon. Wea. Rev.*, **115**, 2118–2132.

Korth, B., and L.R. Tucker, 1975: The distribution of chance congruence coefficients from simulated data. *Psychometrika*, 40, 361–372.

Kuo, H.L., 1974: Further studies of the parameterization of the effect of cumulus convection on large scale flow. *J. Atmos. Sci*., **31**, 1232-1240.

Kumar, A., and M.P. Hoerling, 1995: Prospects and Limitations of Seasonal Atmospheric GCM Predictions. *Bull. Amer. Meteor. Soc.*, **76**, 335–345.

Legates, D.R., and C.J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol*., 10, 111–127.

Leung, L.R., A.F. Hamlet, D.P. Lettenmaier, and A. Kumar, 1999: Simulations of the ENSO Hydroclimate Signals in the Pacific Northwest Columbia River Basin. *Bull. Amer. Meteor. Soc.*, **80**, 2313–2329.

Leung, L.R.,and Y. Qian, 2003a: The Sensitivity of Precipitation and Snowpack Simulations to Model Resolution via Nesting in Regions of Complex Terrain. *J. Hydrometeor.*, **4**, 1025–1043.

Leung, L.R., Y. Quian, and X. Bian, 2003b: Hydroclimate of the Western United States Based on Observations and Regional Climate Simulation of 1981–2000. Part I: Seasonal Statistics. *J. Climate*, **16**, 1892–1911.

Leung, L.R., Y. Qian, X. Bian, and A. Hunt. 2003c. Hydroclimate of the western United States based on observations and regional climate simulation of 1981-2000. Part II: Mesoscale ENSO anomalies. *J. Climate* 16:1912-1928.

Liu, Y., and R. Avissar, 1999: A Study of Persistence in the Land–Atmosphere System Using a General Circulation Model and Observations. *J. Climate*, **12**, 2139–2153.

Livezey, R.E., J.D. Hoopingarner, and J. Huang, 1995: Verification of official monthly mean 700-HPA height forecasts: An update. *Wea. Forecasting*, **10**, 512–527.

Lo, F., and M.P. Clark, 2002: Relationships between Spring Snow Mass and Summer Precipitation in the Southwestern United States Associated with the North American Monsoon System. *J. Climate*, **15**, 1378–1385.

Lott, N., G. Ross, and M. Sittel, 1997: The winter of '96-'97 west coast flooding. NCDC Tech. Rep. 97-01, National Climate Data Center, 22 pp.

Maddox, R.A., J. Zhang, J.J. Gourley, and K. W. Howard, 2002: Weather radar coverage over the contiguous United States. *Wea. Forecasting*, 17, 927–934

Mantua, N.J., S.R. Hare, Y. Zhang, J.M. Wallace, and R.C. Francis, 1997: A Pacific decadal climate oscillation with impacts on salmon. *Bull. Amer. Met. Soc.*, Vol. 78, pp 1069-1079.

Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey, 1997: A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *J. Geophys. Res*., **102**, no. C3, 5753-5766.

Maurer, E., A. Wood, J. Adam, D. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Climate*, **15**, 3237–3251.

McCabe G.J. (1995): Relations between winter atmospheric circulation and annual streamflow in the western United States. *Clim. Res*. **5**, 139-148.

Medvigy, D., R.L. Walko, and R. Avissar (2008a), Modeling interannual variability of the Amazon hydroclimate, *Geophys. Res. Lett.*, *35*, L15817, doi:10.1029/2008GL034941.

Medvigy, D., Avissar, R., Walko, R.L., and Otte, M. (2008b), The Ocean-Land-Atmosphere Model (OLAM): evaluation, calibration, and resolution-dependence of the physical parameterizations. Submitted.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc*., **87**, 343–360.

Mo, K. C. and J. N. Paegle, 2000: Influence of sea surface temperature anomalies on the precipitation regimes over the southwest United States. *J. Climate*, **13,** 3588–3598.

Mo, K.C., and R.W. Higgins, 1996: Large-scale atmospheric moisture transport as

evaluated in the NCEP/NCAR and the NASA/DAO reanalyses. *J. Climate*, **9**, 1531–1545.

Mock, C.J., 1996: Climatic controls and spatial variations of precipitation in the western United States. *J. Climate*, 9, 1111–1125.

Murphy, A.H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev*., **116**, 2417–2424.

National Oceanic and Atmospheric Administration, 1999: December 1998-February 1999 The Climate in Historical Perspective, National Climatic Data Center, March 12, 1999 <http://lwf.ncdc.noaa.gov/oa/climate/research/1999/win/win99.htm>

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. Mon. Wea. Rev., 110, 699–706.
Parish, T., 1982: Barrier winds along the Sierra Nevada mountains. *J. Appl. Meteor.,* **12,** 925–930.

Pielke, R.A, and Coauthors, 1992: A comprehensive meteorological modeling system—RAMS. *Meteor. Atmos. Phys.,* **49,** 69–91.

Ralph, F.M., and Coauthors, 2005: Improving short-term (0–48 h) cool-season quantitative precipitation forecasting: Recommendations from a USWRP workshop. *Bull. Amer. Meteor. Soc*., **86**, 1619–1632.

Ramos da Silva, R., and R. Avissar, 2006: The Hydrometeorology of a Deforested Region of the Amazon Basin. *J. Hydrometeor.*, **7**, 1028–1042.

Ramos da Silva, R., D. Werth, and R. Avissar, 2008: Regional Impacts of Future Land-Cover Changes on the Amazon Basin Wet-Season Climate. *J. Climate*, **21**, 1153–1170.

Reynolds, R. W., 1988: A real-time global sea surface temperature analysis. *J. Climate*, 1, 75–87.

Richman, M.B., 1986: Rotation of principal components. *Int. J. Climatol*., 6, 293–335.

Richman, M.B., and P.J. Lamb, 1985: Climatic pattern analysis of three- and seven-day summer rainfall in the central United States: Some methodological considerations and a regionalization. *J. Climate Appl. Meteor*., 24, 1325–1343.

Richman, M.B., and P.J. Lamb, 1987: Pattern analysis of growing season precipitation in southern Canada. *Atmos.–Ocean*, 25, 137–158.

Ross, T., N. Lott, S. McCown, and D. Quinn, 1998: The El Nino winter of '97-'98. NCDC Tech. Rep. 98-02, National Climate Data Center, 28 pp.

Rudolf, B., and U. Schneider, 2005: Calculation of gridded precipitation data for the global land-surface using in-situ gauge observations. Proc. Second Workshop of the International Precipitation Working Group IPWG, Monterey, CA, EUMETSAT, 231–247.

Rudolf, B., H. Hauschild, W. Rueth, and U. Schneider, 1994: Terrestrial precipitation analysis: Operational method and required den-sity of point measurements. Global Precipitations and Climate Change, M. Desbois and F. Desalmond, Eds., NATO ASI Series 1, Vol. 26, Springer-Verlag, 173–186.

Serafin, R.J., and J.W. Wilson, 2000: Operational weather radar in the United States: Progress and opportunity. *Bull. Amer. Meteor. Soc*., 81, 501–518.

Sheffield, J., A.D. Ziegler, E.F. Wood, and Y. Chen, 2004: Correction of the high-latitude rain day anomaly in the NCEP–NCAR reanalysis for land surface hydrological modeling. *J. Climate*, **17**, 3814–3828.

Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088–3111.

Shepard, D.S., 1984: Computer mapping: The SYMAP interpolation algorithm. Spatial Statistics and Models, G. L. Gaile and C. J. Willmott, Eds., D. Reidel, 133–145.

Slingo, A. and H.M. Schrecker, 1982: On the shortwave radiative properties of stratiform clouds. *Quart. J. Roy. Met. Soc*. **108**, 407-426.

Smith, R., J. Paegle, T. Clark, W. Cotton, D. Durran, G. Forbes, J. Marwitz, C. Mass, J. McGinley, H.-L. Pan, and M. Ralph, 1997: Local and remote effects of mountains on weather: Research needs and opportunities. *Bull. Amer. Met. Soc.*, **78,** 877-892.

Smith, B. L., S.E. Yuter, P.J. Neiman and D.E. Kingsmill, 2009: Water vapor fluxes and orographic precipitation over Northern California associated with a land-falling atmospheric river. *Mon. Wea. Rev.*, submitted.

Soong, S., and J. Kim, 1996: Simulation of a heavy precipitation event in California, *Clim. Change*, **32**, 55-77.

Stidd C. K., 1954: The use of correlation fields in relating precipitation to circulation. *J. Meteor*, **11**, 202–213.

Susskind, 2001: Global precipitation at one-degree daily resolution from multi-satellite observations. *J. Hydrometeor.,* **2,** 36-50.

Trenberth, K.E., 1990: Recent Observed Interdecadal Climate Changes in the Northern Hemisphere. *Bull. Amer. Meteor. Soc.*, **71**, 988–993.

Trewartha, G.T., 1981: The Earth's Problem Climates. 2nd ed. University of Wisconsin Press, 371 pp.

Van de Hulst, H.C. 1957: Light scattering by small particles. New York, J.Wiley & Sons, Inc.

White, D.R., M. Richman, and B. Yarnal, 1991: Climate regionalization and rotation of principal components. *Int. J. Climatol*., 11, 1–25.

Widmann, M., and C.S. Bretherton, 2000: Validation of meso scale precipitation in the NCEP reanalysis using a new gridcell dataset for the northwestern United States. *J. Climate*, **13**, 1936–1950.

Wilks, D.S., 2006: Statistical Methods in the Atmospheric Sciences. Academic Press, 648 pp.

Walko, R.L., Cotton, W.R., Meyers, M.P., Harrington, J.Y., 1995: New RAMS cloud microphysics parameterization part I: the single-moment scheme. *Atm. Res*., **38**, 1-4.

Walko, R.L., L.E. Band, J. Baron, T.G.F. Kittel, R. Lammers, T.J. Lee, D. Ojima, R.A. Pielke, C. Taylor, C. Tague, C.J. Tremback, and P.L. Vidale, 2000: Coupled Atmosphere–Biophysics–Hydrology Models for Environmental Modeling. *J. Appl. Meteor.*, **39**, 931–944.

Walko, R.L., and R. Avissar, 2008a: The Ocean Land Atmosphere Model (OLAM). Part I: Shallow water tests., *Mon. Wea. Rev.*, **136**, 4033-4044.

Walko, R.L., and R. Avissar , 2008b: The Ocean Land Atmosphere Model (OLAM). Part II. Formulation and Tests of the Nonhydrostatic Dynamic Core., *Mon. Wea. Rev.*, **136**, 4045-4062.

Warner, T.T., R.A. Peterson, and R.E. Treadon, 1997: A Tutorial on Lateral Boundary Conditions as a Basic and Potentially Serious Limitation to Regional Numerical Weather Prediction. *Bull. Amer. Meteor. Soc.*, **78**, 2599–2617.

Werth, D., and R. Avissar, 2002. The local and global effects of Amazon deforestation, *J. Geophys. Res.*, **107**, 8087, doi:10.1029/2001JD000717.

Xie. P., and P.A. Arkin, 1995: An intercomparison of gauge observations and satellite estimates of monthly precipitation. *J. Appl. Meteor.*, **34**, 1143–1160.

Xie. P., and P.A. Arkin , 1996: Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J. Climate*, 9, 840–

858.

Xie. P., and P.A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc*., **78**, 2539–2558.

Xie. P., and P.A. Arkin, 1998: Global monthly precipitation estimates from satellite-observed outgoing longwave radiation. *J. Climate*, 11, 137–164..

Yang, Z., and R. W. Arritt, 2002: Tests of a perturbed physics ensemble approach for regional climate modeling. *J. Climate*, **15**, 2881–2896.

Yin, X., A. Gruber, and P. Arkin, 2004: Comparison of the GPCP and CMAP merged gauge–satellite monthly precipitation products for the period 1979–2001. *J. Hydrometeor.*, 5, 1207–1222.

Biography

Kristen was born on August 19, 1972 in Ann Arbor, Michigan. She received her B.S. degree in Earth Sciences from the University of California, San Diego in 1998. She continued her study at UCSD receiving her M.S. degree in 1999 from UCSD/Scripps Institution of Oceanography studying the radiative properties of clouds. This work contributed to the journal article "Radiative Transfer through Broken Clouds: Observations and Model Validation", which appeared in *Journal of Climate.* After graduating from UCSD Kristen worked for the U.S. Environmental Protection Agency in San Francisco before beginning her doctoral work at Duke. Her doctoral research focused on the hydrometeorology/hydroclimatology of the western United States using observations and model simulations. She published two journal articles in the *Journal of Hydrometeorology* entitled "Precipitation Climatology and Dataset Intercomparison of the Western United States" and "A Multiple Dataset Analysis of Precipitation Variability, Persistence, and Observational Data Uncertainty in the Western United States". A third article is expected from her simulation results using the Ocean-Land-Atmosphere Model (OLAM). While at Duke, Kristen was the recipient of the NASA Earth Science Fellowship from 2004 to 2007.