Downscaling with Constructed Analogues: Daily Precipitation and Temperature Fields over the United States

Hugo Hidalgo
University of Costa Rica

Michael D. Dettinger
University of California, San Diego

Daniel R. Cayan
University of California, San Diego

Arnold Schwarzenegger
Governor

DOWNSCALING WITH CONSTRUCTED ANALOGUES: DAILY PRECIPITATION AND TEMPERATURE FIELDS OVER THE UNITED STATES

PIER FINAL PROJECT REPORT

Prepared For:
California Energy Commission
Public Interest Energy Research Program

Prepared By:
Hugo G. Hidalgo, Michael D. Dettinger, and Daniel R. Cayan

January 2008
CEC-500-2007-123

California Climate Change Center
Report Series Number 2007-027

Prepared By:
Hugo G. Hidalgo[1], Michael D. Dettinger[2,1], and Daniel R. Cayan[1,2]

[1]Scripps Institution of Oceanography, University of California, San Diego
[2]United States Geological Survey

Commission Contract No. 500-02-004
Commission Work Authorization No: UC MR-025

Prepared For:
Public Interest Energy Research (PIER) Program
California Energy Commission

Guido Franco
Contract Manager

Guido Franco
Project Manager

Linda Spiegel
Acting Program Area Lead
Energy-Related Environmental Research

Kelly Birkinshaw
Acting Office Manager
Energy System Research

Martha Krebs
Deputy Director
ENERGY RESEARCH & DEVELOPMENT DIVISION

Melissa Jones
Executive Director

Acknowledgments

The authors thank Alexander Gershunov from the Scripps Institution of Oceanography (SIO) for providing the comprehensive daily meteorological Climate Research Division dataset for the United States. Mary Tyree provided support with data management. This work is funded by grants from the California Energy Commission through the California Climate Change Center at Scripps and from the United States Department of Energy, and by the U.S. Geological Survey's Priority Ecosystems Science Initiative.

Please cite this report as follows:

Hidalgo, H. G., M. D. Dettinger, and D. R. Cayan. 2008. *Downscaling with Constructed Analogues: Daily Precipitation and Temperature Fields Over the United States.* California Energy Commission, PIER Energy-Related Environmental Research. CEC-500-2007-123.

Preface

The Public Interest Energy Research (PIER) Program supports public interest energy research and development that will help improve the quality of life in California by bringing environmentally safe, affordable, and reliable energy services and products to the marketplace.

The PIER Program, managed by the California Energy Commission (Energy Commission), conducts public interest research, development, and demonstration (RD&D) projects to benefit California's electricity and natural gas ratepayers. The PIER Program strives to conduct the most promising public interest energy research by partnering with RD&D entities, including individuals, businesses, utilities, and public or private research institutions.

PIER funding efforts are focused on the following RD&D program areas:

- Buildings End-Use Energy Efficiency
- Energy-Related Environmental Research
- Energy Systems Integration
- Environmentally Preferred Advanced Generation
- Industrial/Agricultural/Water End-Use Energy Efficiency
- Renewable Energy Technologies
- Transportation

In 2003, the California Energy Commission's Public Interest Energy Research (PIER) Program established the **California Climate Change Center** to document climate change research relevant to the states. This center is a virtual organization with core research activities at Scripps Institution of Oceanography and the University of California, Berkeley, complemented by efforts at other research institutions. Priority research areas defined in PIER's five-year Climate Change Research Plan are: monitoring, analysis, and modeling of climate; analysis of options to reduce greenhouse gas emissions; assessment of physical impacts and of adaptation strategies; and analysis of the economic consequences of both climate change impacts and the efforts designed to reduce emissions.

The California Climate Change Center Report Series details ongoing center-sponsored research. As interim project results, the information contained in these reports may change; authors should be contacted for the most recent project results. By providing ready access to this timely research, the center seeks to inform the public and expand dissemination of climate change information, thereby leveraging collaborative efforts and increasing the benefits of this research to California's citizens, environment, and economy.

Downscaling with Constructed Analogues: Daily Precipitation and Temperature Fields Over the United States is the final report for the Continuing Climatic Data Collection, Analyses, and Modeling project (contract number 500-02-004, work authorization UC MR-025) conducted by Scripps Institution of Oceanography and the United States Geological Survey (USGS).

For more information on the PIER Program, please visit the Energy Commission's website www.energy.ca.gov/pier/ or contract the Energy Commission at (916) 654-5164.

Table of Contents

Preface iii

Abstract ix

Executive Summary 1

1.0 Introduction 5

2.0 Data 9

3.0 Methods 13

3.1. Preliminary Demonstration of the Method 13

3.2. Selection of the Variables to Use as Candidate Predictors 14

3.3. Diagnosis: Constructing a Coarse-resolution Analogue 18

3.4. Prognosis: Downscaling a Weather Pattern 21

4.0 Results 23

4.1. Selection of Most-suitable Predictors 23

4.2. Daily Skill 23

4.3. Monthly Skill 31

5.0 Domain Size and Skill over a Smaller Region: California Example 37

6.0 Conclusions and Discussion 39

7.0 References 43

8.0 Glossary 47

List of Figures

Figure 1. The storm of January 27, 1983, over California as recorded in precipitation (mm) and temperature (°C) estimates from the NCEP/NCAR reanalysis (top panel) and VIC meteorological data from Maruer et al. (2002) (bottom panel). The three locations in the bottom-left figure are pertinent to Figure 10. 13

Figure 2. Downscaled precipitation (mm) and average air temperature (°C) versions of the January 27, 1983, California storm using constructed analogues, applied to coarse VIC data (top), station data (middle), and medium-range forecast (bottom). 15

Figure 3. Schematic of the method of constructed analogues for downscaling reanlysis fields from 2.5° x 2.5° grids to 1/8° x 1/8° 18

Figure 4. Change in the median correlation skill of downscaled precipitation and temperature estimates by increasing the number of suitable predicators (n) 19

Figure 5. Median (dots) and interquartile range (error bars) of the distances between each pattern to be downscaled and its predictor weather patterns (top panel), and percentage variance explained by the first principal component of the most-suitable subset of predictors at each time of year (bottom panel). The spread of the quantiles represent the results of 25 downscaled years. Negative distances (positive) mean that given a weather pattern at a certain time of the year, more analoguous patterns are more frequently found earlier (later) during the year. 24

Figure 6. Cross-validated statistics comparing (square roots of the) downscaled daily precipitation rates to observed values (see text). The statistics shown are based on daily values from 25 years of winter (December through February), summer (June through August), and annual (all days of the year) seasons. 25

Figure 7. Same as Figure 6, but for daily average air temperatures 26

Figure 8. Difference in the means (top panel) and standard deviations (bottom panel) between downscaled minus observed fields of daily precipitation for January, July, and all days (annual) 26

Figure 9. Same as Figure 8, but for daily average air temperatures 27

Figure 10. Scatter plots of downscaled versus observed daily (square root of) precipitation for three U.S. locations indicated in lower left panel of Figure 1 (left panel) and percentile distributions of the same time series (right panel); observations from Maurer et al. (2002). 28

Figure 11. Average extra number of wet days per year in downscaled precipitation fields (compared to observations) and “zero-equivalent” precipitation. The zero-equivalent precipitation is the threshold below which the downscaled values would need to be set to

zero in order to remove the (positive) biases in the number of wet days. A wet day is defined as a day when P>0.1 mm. 29

Figure 12. Same as Figure 10, but for daily average air temperatures 30

Figure 13. Same as Figure 6, but for monthly precipitation averages 32

Figure 14. Same as Figure 7, but for monthly mean air temperatures 33

Figure 15. Same as Figure 10, but for monthly averages 34

Figure 16. Same as Figure 11, but for monthly mean air temperatures 35

Figure 17. Effect of changing the domain of the predictor fields on day-to-day correlations between observed and downscaled precipitation and temperature anomalies in California. The average correlation and root mean square errors are shown........ 38

List of Tables

Table 1. Median of correlation and RMSE between downscaled and observed daily P and Tavg using the reanalysis and VIC coarse data sets as predictors for the conterminous United States. The first (second) correlation represents the daily values with and (without) the seasonal cycle. The statistics shown are the median values from 52,135 gridpoints that cover the conterminous United States at 1/8 degree resolution. 16

Table 2. Median Brier skill scores comparing the constructed analogue downscaling method to three different reference approaches: (a) Using a randomly selected weather pattern from a pool of 2275 patterns (see text) as downscaling estimation, (b) Using only the climatological mean values at each grid cell, and (c) using the closest naturally occurring analogue to the pattern to be downscaled as downscaling estimation. The statistics shown are the median BSS values from 52,135 gridpoints that cover the conterminous US at 1/8 degree resolution. 20

Table 3. Lag-1 day autocorrelation coefficients of the anomalies of observed and downscaled precipitation and average air temperature. The coefficients shown are the median values from all 52,135 gridpoints at 1/8 degree resolution covering the conterminous United States. 31

Table 4. Same as Table 1, but for two cases 31

Abstract

Daily precipitation and average temperature patterns for the contiguous United States were downscaled from a 2.5 x 2.5 degree (coarse) resolution grid to a 1/8 x 1/8 degree (fine) resolution grid using a constructed-analogues method. Choice of predictors, and the selection of subsets of most-suitable historical dates to be included in the constructed analogues proved to be important determinants of the method's skill, especially for precipitation. The downscaling method skillfully reproduces daily variations of precipitation and average temperature anomalies, as well as seasonal cycles, across the contiguous United States. The method tends to overestimate the number of wet days, producing a very light "drizzle" on many of the effectively dry days. There are also biases in the monthly climatologies of precipitation and average temperature in some regions, which tend to average out at annual timescales. Averaging daily downscaled patterns into monthly means yielded even more skillful results, capturing about 55 percent of the variations of monthly precipitation anomalies and about 80 percent of the variations of average temperature monthly anomalies across the contiguous United States. The choice of the domain of the predictor also influences the skill. For example, in California, the most skillful precipitation downscaling was obtained when the precipitation predictors covered the state, whereas average temperature downscaling was most skillful when average temperature predictors included continent-wide patterns. Overall, the method showed encouraging results for downscaling daily precipitation and average temperature continental-wide patterns in North America—in particular, those of the western United States.

Keywords: Climate change, general circulation model, GCM, climate model, statistical downscaling

Executive Summary

Introduction

The general circulation models used for climate simulations typically have horizontal spatial resolutions of a few hundred kilometers. However, to properly evaluate regional and local effects of climate variations (for example, in projecting hydrologic changes in a watershed), researchers require a much higher level of detail—on the order of 10 kilometers—because many effects are sensitive to the nuances of the local climate. *Downscaling* is the process of transferring the climate information from a coarse-scale climate model to the fine scale required by models that address effects on climate. Although downscaling can be achieved using a regional climate model, it is computationally expensive and currently is not practical for processing multidecade and/or multimodel simulations from general circulation models. A viable alternative is to use statistical downscaling, which has the advantage of requiring considerably less computational resources. Underpinning this approach is the assumption that analogues from the present climate can be identified and applied, statistically, to determine the day-to-day patterns that will characterize the future climate.

Purpose

This report presents the development of a new method for statistically downscaling daily precipitation and temperature from general circulation models using the "constructed analogues" method. The method is based on the premise that an analogue for a given coarse-scale daily weather (target) pattern (for example, from a general circulation model simulation) can be constructed by combining the weather patterns for several days (predictors) from a library of previously observed patterns. In this application the analogue pattern is constructed at coarse scale, but a similar construction can be made using a companion library of high-resolution patterns using the same days as the coarse-scale predictors. Thus, a fine resolution downscaled estimate is created for the given pattern of that particular day.

The purpose of the constructed analogues downscaling method is to produce daily precipitation and temperature maps, at fine spatial resolution, from a general circulation model. These estimates can be used to study impacts of climate changes and climate variability, for example, as input to a distributed hydrological model focusing on California watersheds.

Project Objectives

The research team sought to develop a new method to statistically downscale daily patterns from a climate general circulation model. By applying this "constructed analogues" method to historical weather patterns, the team evaluated its performance in downscaling precipitation and temperature over the conterminous United States, and, in particular, the western United States.

Project Outcomes

- The constructed analogues method performed quite well in downscaling daily precipitation and temperature in North America, in particular in the western United States.

- The constructed analogues method has very high skill in downscaling daily and seasonal variations in temperature and reasonably high skill in downscaling daily and seasonal variations in precipitation.

- The skill in downscaling precipitation is highest in coastal states and during winter.

- The constructed analogues method tends to slightly overestimate the number of wet days in parts of the arid and semi-arid regions.

- For monthly timescales the results for both precipitation and temperature are even better than for daily timescales.

- Biases in the monthly climatologies of precipitation and temperature tend to average out at annual timescale

Conclusions

- The constructed analogues method is quite simple to apply. When used to downscale precipitation and temperature over the conterminous United States, it performed quite well in capturing 50 percent of daily high-resolution precipitation variance and 67 percent of daily temperature variance as an average across all seasons and across the contiguous United States. These percentages increase to 55 percent and 80 percent, respectively, when the downscaled daily estimates are accumulated into monthly means.

- The downscaled linearly constructed weather patterns are spatially similar to observed weather patterns, and the temporal autocorrelation is only modestly (10 percent) underestimated.

- The biases in the monthly climatologies are small for both precipitation and temperature, but the downscaling method tends to overestimate the number of wet days.

- The variables used as predictors and the selection of the most suitable patterns in the diagnostic part of the method are very important determinants of the method's skill, especially for precipitation.

- The geographic domain selected for the predictors affects the skill of downscaled precipitation and temperature patterns over the California region. Daily temperature patterns are best captured from large-scale predictor fields. In contrast, daily precipitation patterns were best captured using predictors over smaller domains, closer to the typical size of storm systems, or roughly the state scale.

- The constructed-analogues method presented here can be used to downscale output from a variety of different climate model applications, including downscaling of reanalyzed historical climate variations, diagnostic experiments (such as climate experiments with and without tropical sea-surface temperature anomalies), medium-range to season weather and climate forecasts, and long-term climate-change projections. Additionally, it could be used to patch during the period covered by various reanalyses, generally beginning about 1948 through present.

Recommendations

In a climate-change application, it will be important for the downscaling method to capture the high-resolution effects of largely unprecedented trends at general circulation model scales. Projected precipitation and temperature trends will derive from changes in the frequencies and amplitudes of various daily weather patterns, some of which have already been witnessed in historical archives but which will come more or less frequently or strongly under the changed climates, and some of which have not been witnessed before. Capturing the high-resolution consequences of the former, already-witnessed patterns is a particular strength of the method presented here; capturing the high-resolution consequences of the latter (truly new) patterns will depend on the flexibility offered by the construction of analogues. For example, to the extent that temperature trends take the form of extremely persistent zonally banded warming patterns, the downscaling process will have to either extract these "truly new" patterns from historical analogues or construct them wholesale from collections of other historical temperature anomalies.

Future research must evaluate the ability of the constructed analogues method to accommodate these new "deformations" of the weather patterns associated with natural climate variability by externally derived trend patterns associated with anthropogenic climate change. Depending on the strength and distribution of new patterns (which in turn may depend on the model and trace gases emission scenario analyzed), it may become increasingly more difficult to find good analogues as the climate changes proceed into the future, because the predictors are based on a period in which the effects of anthropogenic climate change are small compared to natural variability. Fortunately, it is possible to monitor the skill with which large-scale analogues can be constructed, and that monitoring provides a tool for diagnosing when the historical record is failing to provide adequate predictors. In this regard, the preliminary analyses by the authors are encouraging, because they show, for the National Center for Atmospheric Research's Parallel Climate Model and a single emission scenario, that large-scale analogues of projected climate-change trends could be constructed from historical patterns without degradation throughout the twenty-first century (results to be presented separately). Previous studies using other models and emission scenarios obtained good analogues of climate change weather patterns using historically observed patterns. Thus, the downscaling method presented here offers many potential uses, relying where possible on the particular strengths of the large-scale models while providing flexibility for capturing unexpected changes in the large-scale climate projections and weather predictions.

Benefits to California

Because general circulation models are calculated at a very coarse scale—approximately 200 kilometers horizontally—it is necessary to downscale the general circulation model projections to much finer scale, commensurate with the structure of the California landscape. The constructed analogues method developed for this project is an important tool in helping to determine the effects of climate change on California's hydrological budget, and on other sectors. These assessments are key to learning how the state can adapt to and possibly mitigate climate change impacts. Many human activities and ecosystems health depend on the hydrology, such as agricultural production, power generation, and wildfire potential. In particular, studies of future changes in the water supply from local and remote sources are crucial to developing strategies to cope with climate change and for mid-term and long-term water planning.

1.0 Introduction

Downscaling is the process of using simulations by climate models at coarse spatial resolutions (on the order of a few hundred kilometers) to produce estimates, forecasts, or projections of climate variations at higher resolutions (in some cases, down to a few kilometers). Downscaling is generally required when the impacts of various climate variations and changes cannot be estimated from coarse resolution predictions or projections because the impacts are especially sensitive to nuances of local climate (e.g., those associated with differences in topography and exposure) or when surface characteristics that typically vary in the order of 10 kilometers (6 miles) or less (like topographic boundaries) dictate the form and extent of regional impacts (Timbal et al. 2003).

Two broad approaches to downscaling have been used: *dynamical* and *statistical.* Comparisons of the skills of the two approaches can be found elsewhere (Murphy 1999; Hay and Clark 2003; Hanssen-Bauer et al. 2003; Wood et al. 2004). Dynamical downscaling uses higher-resolution, limited-area, physically based regional climate models (RCMs) nested within the outputs of coarser models to increase the spatial resolution of climate simulations, predictions, and projections (e.g., Roads et al. 2003; Antic et al. 2004; Weaver 2004). Regional climate models represent the same physical processes (and occasionally more physical processes) as do the coarser-resolution global climate models, except at significantly higher spatial resolutions, and they cover smaller overall areas. Their higher resolutions allow inclusion of more detailed representations of topography and surface contrasts, as well as more spatially detailed depictions of gradients within the atmosphere (Zangl 2004; Leung and Wigmosta 1999). The main disadvantages of RCMs are that their implementation usually requires computational efforts as large or larger than the original coarse resolution model simulations, and that they remain somewhat experimental in climate applications (that is, the extent to which they add skill to predictions of climate variations and changes remains a subject of ongoing assessments and probably varies considerably from place to place and situation to situation in ways that are not yet well understood (Wood et al. 2004)). Moreover, statistical post-processing procedures are needed in many cases to remove systematic biases from RCM outputs (Wood et al. 2004).

Although future improvements in RCMs may remove many of these biases, for certain applications, the skill of RCMs is currently (at best) comparable to the results obtained through statistical methods (Murphy 1999; Hay and Clark 2003; Wood et al. 2004). Statistical downscaling methods apply statistical relations between historical climate records at coarse resolutions and high resolutions to interpolate from coarse model outputs to higher resolutions. Statistical downscaling has the advantages of requiring much less computational effort and is several orders of magnitude less time-consuming, and of (usually) transparently eliminating most biases as part of the derivation of the underlying statistical relations. One drawback of statistical downscaling methods is that they represent extreme simplifications of the physical relations at work and thus generally can only be expected to capture the high-resolution, dependent-variable outcomes of variations in the particular limited sets of coarse resolution, independent-variable inputs used, and can only be expected (a priori) to perform well by the particular statistical measures optimized in the training period. Furthermore, the statistical relations can only replicate the historical forms of the relations between coarse climate fields and high-resolution variations and thus cannot be expected to capture changed forms of those relations, should they change in the future.

This report presents a deterministic, linear approach for statistically downscaling from coarse-resolution (~250 kilometers (km), or ~155 miles), continental regional-wide patterns of daily precipitation (P) and average temperature (Tavg) to their high-resolution (~12 km, or ~7.5 miles) equivalents by using a variation on the constructed analogue method of van den Dool (1994, hereafter VDD) and van den Dool et al. (2003). VDD's method is a statistical approach to climate prediction based on the notion that, if one could find an exact analogue (in the historical record) to the weather field today, weather in the future should replicate the weather following the time of that exact analogue; the more exact the correspondence between today's weather patterns (across all climatic variables), the longer the prediction by analogy would be expect to work. Lorenz (1969) showed that naturally occurring analogues at hemispheric scales of acceptable quality for such a use are highly unlikely to be found, given the relatively short historical records of observations and the high number of degrees of freedom (d.o.f.) of atmospheric circulations. In fact, VDD estimated that it would take around 10^{30} years to find two identical (within current observational error) Northern Hemisphere 500 hectopascal (hPa) geopotential height patterns. However, rather than waiting for an acceptable analogue to emerge, VDD developed an approach whereby analogues were constructed from linear combinations of past atmospheric patterns. By linear regressions with the current weather or climate pattern as the dependent variable and selected historical patterns as independent variables, very high quality analogues can be constructed. VDD then assumed that the same linear combination (using the same regression coefficients) of the future evolutions of each of the historical patterns that contributed to the constructed analogue should tend to describe the evolution of weather or climate into the future, for a time (van den Dool 2003). The assumption and method provides forecasts that are surprisingly accurate, indicating that the contributions from the evolving historical patterns do not begin to interact nonlinearly or to randomize overwhelmingly for long enough times to allow useful predictions.

A related method of analogues has been used for downscaling output from climate models (Zorita et al. 1995; Martin et al. 1996; Zorita and von Storch 1999; Timbal and McAvaney 2001; Fernández and Sáenz 2003; Timbal et al. 2003; Timbal 2004; Wetterhall et al. 2005; Diez et al. 2005; Gangopadhyay et al. 2005). The single-analogue downscaling method used in these studies was based on finding the closest analogue for each particular weather pattern simulated by the climate model, from a library of historically observed coarse-resolution patterns. Once the date of the best analogues for a weather pattern from the climate model has been identified, the high-resolution field corresponding to the coarse-resolution analogue date is the downscaled version of the climate model pattern. By contrast, in the method of constructed analogues pursued here (and in Fernández and Sáenz 2003), an improvement over the single-analogue method is obtained by producing a downscaled estimation based on the linear combination of multiple patterns, much like those constructed analogues developed by VDD for prediction purposes. Procedurally, a collection of historically observed coarse-resolution climate patterns is linearly regressed to form a best-fit constructed analogue of a particular coarse-resolution climate-model output. The constructed analogues method develops a downscaled, higher resolution climate pattern associated with the climate-model output from the (same) linear combination of historical high-resolution patterns as was fitted to form the coarse-resolution analogue. Thus, the regression coefficients that form the best-fit combination of coarse resolution daily maps (at, say 250 km resolution, to reproduce a given climate-model daily pattern are applied to the high-resolution (say, 12 km resolution) maps from the same (historical) days. Conceptually, rather than assuming that the same linear combination of the future evolutions from the contributing patterns yields a useful forecast (as in VDD), in its downscaling form, the constructed analogues method assumes that the (same) linear combination of high-resolution patterns can provide a faithful depiction of the high-resolution climate pattern associated with the model output of interest. Thus the high-resolution climate variations are obtained from this form of the constructed analogues method. With this form of the constructed analogues the temporal evolution of those patterns are not projected, but rather is determined by the more completely represented simulation of the large-scale climate model. That is, to downscale a weather pattern for the following day, the coarse-scale pattern for the next day is obtained from the climate model output and a new linear regression is performed to construct a new (high-resolution) analogue.

Several versions of the method of analogues have been used for downscaling (Zorita et al. 1995; Martin et al. 1996; Zorita and von Storch 1999; Timbal and McAvaney 2001; Timbal et al. 2003; Timbal 2004; Salathe 2003; Fernández and Sáenz 2003; Gangopadhyay et al. 2005). In general, these previous studies have demonstrated more skill for downscaling Tavg than for P, in particular when using atmospheric circulations as predictors. Like some of these previous studies, the approach presented here downscales to continental scales, and, in particular, downscales patterns of P and Tavg across the contiguous United States. However, this downscaled approach differs from all previous efforts in the way analogues are selected and constructed, in the type of predictors used, and in that the analogues method presented here is generally simpler than other approaches; for example, no reduction of the d.o.f. using Principal Component Analysis (PCA; e.g., Gangopadhyay and Clark 2005)—or canonical correlations (Fernández and Sáenz 2003)—is required. It was also common in these previous studies to use atmospheric circulations as the coarse scale (predictor) field used for downscaling the surface variables (precipitation or temperature). This approach is designated as the "indirect analogue downscaling method," and its success depends critically on identifying an atmospheric circulation field that is very closely related to the surface variable of interest to obtain adequate skill. That most of these downscaling efforts produced better temperature than precipitation fields is an indication of the difficulty of identifying any single atmospheric circulation descriptor that encapsulates the processes controlling daily precipitation amounts and patterns. In contrast, this study used the coarse scale versions of the surface fields themselves as predictors (designated as a direct analogue downscaling method), in order to include amalgams of essentially all the atmospheric circulations and fields simulated by the climate models in the analogues, as will be discussed in the next section. As will be seen in the following sections, this approach resulted in a significant improvement in the skill, in particular for precipitation.

The data and method used are described in sections 2 and 3. Results and various measures of downscaling skill are described in Section 4. Section 5 presents an evaluation of the skill of the method for California's climate, with special attention to the effect that the choice of the domain of the coarse-resolution analogues has on the downscaling skill. Conclusions are drawn and possible applications discussed in Section 6.

2.0 Data

The downscaling experiments presented here can be accomplished with historical climate data from individual stations or on grids; for geographic completeness, this study focuses mostly on downscaling onto high-resolution gridded fields. To accomplish this, gridded daily fields corresponding to the actual weather on historical days were required, both to serve as high-resolution analogues and to measure the success of the downscaling procedure. Daily P, maximum air temperature (Tmax), and minimum air temperature (Tmin) fields from 1950 to 1999, gridded over the entire contiguous Unite States at 1/8 x 1/8 degree resolution, were obtained from the Surface Water Modeling Group at the University of Washington.[1] These historical data were developed as described by Maurer et al. (2002), and were used herein to characterize the daily local variations of weather and climate. The P grids are based on the daily observations at Cooperative Observer (COOP) stations, from the National Oceanic and Atmospheric Administration (NOAA), which were interpolated onto a 1/8 degree grid, and then rescaled to match the long-term averages of the gridded data in the Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994). The Tmax and Tmin grids also are based on daily observations at NOAA's COOP stations, which were interpolated onto the same 1/8 degree grid using the environmental lapse rate to account for topography (Maurer et al. 2002; Sheffield et al. 2004). The 1/8 degree grid of the Maurer et al. (2002) meteorological data is the same grid as used for gridded vegetation and soil properties data in the Land Data Assimilation Systems (LDAS);[2] and the same grid as used by the Variable Infiltration Capacity (VIC) model (Liang et al. 1994) of macroscale, land-surface hydrologic fluxes and energy balances at the continental scale. The P and Tavg fields are primary forcing inputs for the VIC model. The present analysis uses these P and Tavg fields as a source of temporally and spatially high-resolution climate analogues for the downscaling experiments.

In this report, Tmax and Tmin were averaged to produce daily Tavg, but the procedure delineated here can be used for temperature extremes separately. The P data were transformed by taking the square root of the daily totals with the objective of removing some of the skewness of daily P data (e.g., Wilson 1997; Dettinger et al. 2004). All P and Tavg fields were mean-centered by subtracting long-term daily mean values at each grid cell, calculated from a calibration subset of the data (to be described later).

A variety of different large-scale, coarse-resolution fields could serve as potential contributors to the constructed analogues, although large-scale P and Tavg actually perform very well (to be discussed later) and make the procedure implemented here particularly simple. Daily, gridded analyses of several atmospheric parameters were obtained from the National Center of Environmental Prediction (NCEP) and the National Center of Atmospheric Research (NCAR) Reanalysis (Kalnay et al. 1996, and updates thereto), hereafter called "Reanalysis." The Reanalysis data at 2.5 x 3.5 degree resolution were used to test different predictors for

---

[1] See www.hydro.washington.edu/Lettenmaier/gridded_data/index_maurer.html.

[2] See http://ldas.gsfc.nasa.gov/.

downscaling P and Tavg to the VIC 1/8 degree grid, as will be discussed in the Methods section. Parameters from the Reanalysis that were considered included P, Tavg, 500 hPa and 700 hPa geopotential height fields (Z_{500} and Z_{700} respectively), and sea level pressures (SLP). The Reanalysis P was transformed by taking the square root of the daily values and all reanalysis data were mean-centered.

Although gridded P and Tavg data are the focus here, comparisons to the original station data increases confidence in the results. A dataset of daily P and maximum and minimum temperatures from first-order weather stations covering the contiguous United States was obtained from Scripp's Climate Research Division (CRD) digital archives (A. Gershunov, personal communication 2005). The CRD data is a subset of a larger set of stations from the National Climatic Data Center (NCDC),[3] selected as having minimal missing data and a relatively uniform spatial distribution. In this dataset, the stations (reporting data at least 90% of the time) were selected in densely sampled areas to have one station representing a radius of around 70 km (43 miles). In the mountains, the most complete record from within each 70 km radius was selected, but also the highest elevation station was included. In poorly sampled areas, serial completeness standards were relaxed to include stations with as few as 80% of the daily data present. The coverage of the dataset declines in northern Canada and parts of Mexico (especially Sonora), as does the data continuity (serial completeness). For this study, researchers discarded stations from the CRD dataset. The discarded data consisted of stations having more than 7% of missing values from 1950 to 1999.

All skill measurements presented in the study were computed on validated datasets, with daily data from the even-numbered years from 1950 to 1999 used as a library or pool of potential analogues, (calibration) and data for the odd-numbered years used to (cross) validate the downscaling results. Both the calibration and validation datasets were centered by removing the daily means of the calibration dataset. Therefore, differences in the means of the calibration and verification datasets will be fitted through the downscaling technique. This emphasis on climate anomalies is not a necessary part of the procedure for the demonstrations presented herein, but makes future transfer of the methods to other applications (using multiple climate models, or applying to nonstationary climates) easier.

In addition to long-term (1950–1999) downscaling experiments that will form much of the Results section, a brief application of the method to a historical medium-range weather forecast (MRF) is also presented as an example of a possible application. P and Tavg estimates from an MRF initiated with conditions on January 26, 1983 (0000 Coordinated Universal Time [UTC]) and valid for the following 36 hours (essentially describing the forecast of weather for January 27,1983) were obtained from NOAA's Climate Diagnostic Center.[4] The MRF data represent the ensemble mean of 15 runs of NCEP's MRF model (Kanamitsu 1989; Kanamitsu et al. 1991;

---

[3] See www.ncdc.noaa.gov/oa/ncdc.html.

[4] See www.cdc.noaa.gov/reforecast/.

Caplan et al. 1997). Details on the characteristics and skill of the model can be found in Hamill et al. (2004).

3.0 Methods

3.1. Preliminary Demonstration of the Method

Before describing details of the method, a quick demonstration of what the method accomplishes in a particular application will help to illustrate the problem to be solved (re-capturing high-resolution features of a daily weather pattern) and thus motivate the effort.

Coarse-scale depictions of weather and climate—whether from analyses, data assimilations, or model simulations—only capture a fraction of the rich spatial details of P and Tavg variations depicted in the higher resolution patterns. In Figure 1, the P and Tavg patterns on the day of a particular storm in California (January 27, 1983) are shown. In the upper panel of this figure, the coarse-resolution Reanalysis version (2.5 degrees) of the day's weather is shown. In the lower panel, the corresponding weather patterns at the VIC resolution (1/8 degree) are shown. The VIC resolution data are able to resolve the influence of the mountain ranges in the western United States and the patchiness of precipitation in the southeast.

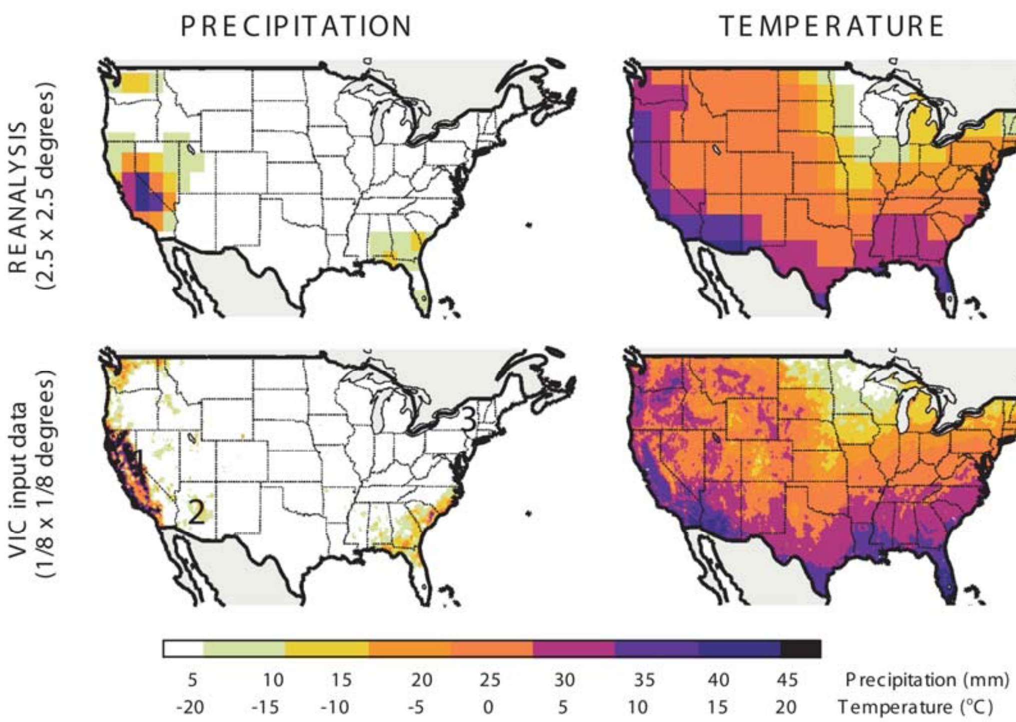

Figure 1. The storm of January 27, 1983, over California as recorded in precipitation (mm) and temperature (°C) estimates from the NCEP/NCAR reanalysis (top panel) and VIC meteorological data from Maruer et al. (2002) (bottom panel). The three locations in the bottom-left figure are pertinent to Figure 10.

In Figure 2, high-resolution versions of the weather on January 27, 1983, in each instance derived by downscaling (with the method described here) of the top panels of Figure 1, are shown. In the top row of Figure 2, the downscaled gridded P and Tavg patterns for the same storm are shown, downscaled from the coarse weather patterns following the procedures that will be enumerated here. As can be seen, the P and Tavg patterns for this particular day were well replicated by the constructed-analogue method. In the case of P, in addition to the large precipitation system over California, smaller storms in the southeast and Pacific Northwest are captured with significant skill. The downscaled Tavg pattern strongly resembles the observed pattern, although temperatures are overestimated in the southern regions (on average by 0.46°C, or 0.83°F) and underestimated in the northern regions (on average by 0.54°C, or 0.97°F), in particular in the upper Midwest that were significantly underestimated by around 4°C (7.2°F) (Figures 1 and 2). In the middle row of Figure 2 the results of downscaling of the coarse weather patterns to station data are shown. Not surprisingly, the downscaled station patterns are less spatially coherent than the gridded data, but in general there is substantial in reproducing the defining P and Tavg features of the storm. In particular, the main and secondary storms are represented in the P patterns from the station data, although slightly overestimated values were obtained over Utah and Colorado (on the order of 5 millimeters [mm] per day) and the southeastern secondary storm is shifted inland and has less intensity in the downscaled estimate than in the observed pattern. The bottom row of Figure 2 is an example of the application of the analogues method for downscaling MRFs (in this case for the 36 hours forecast from January 26, 1983 (0000 UTC) initial conditions). Even though the patterns shown in figures 2e and 2f include the forecast uncertainties of the MRF model, the downscaled MRF captures most of the nuances of local weather caused by elevation and exposure with reasonable accuracy.

3.2. Selection of the Variables to Use as Candidate Predictors

The procedure for using constructed analogues for prediction or downscaling can be divided in two parts: diagnosis and prognosis (following VDD). Given a particular coarse-resolution weather pattern to be downscaled, the diagnosis step consists of selecting a subset of weather patterns from a large library of historical patterns at coarse resolution (e.g., 2.5 degree resolution) and then determining the linear combination of those patterns that best match the given (target) pattern. The subset is formed so as to contain a limited number of patterns that are most suitable for constructing a faithful analogue of the target pattern. In the downscaling application of constructed analogues, the "prognosis" step is the derivation of the high-resolution pattern by applying the linear fit developed from the subset of most suitable, coarse-resolution predictors, where "predictors" in this downscaling application are the historical coarse-resolution patterns used in the diagnosis step. The regression coefficients derived for each coarse-resolution pattern in the diagnosis step are applied directly to the corresponding fine-resolution weather patterns for the same days (Figure 3).

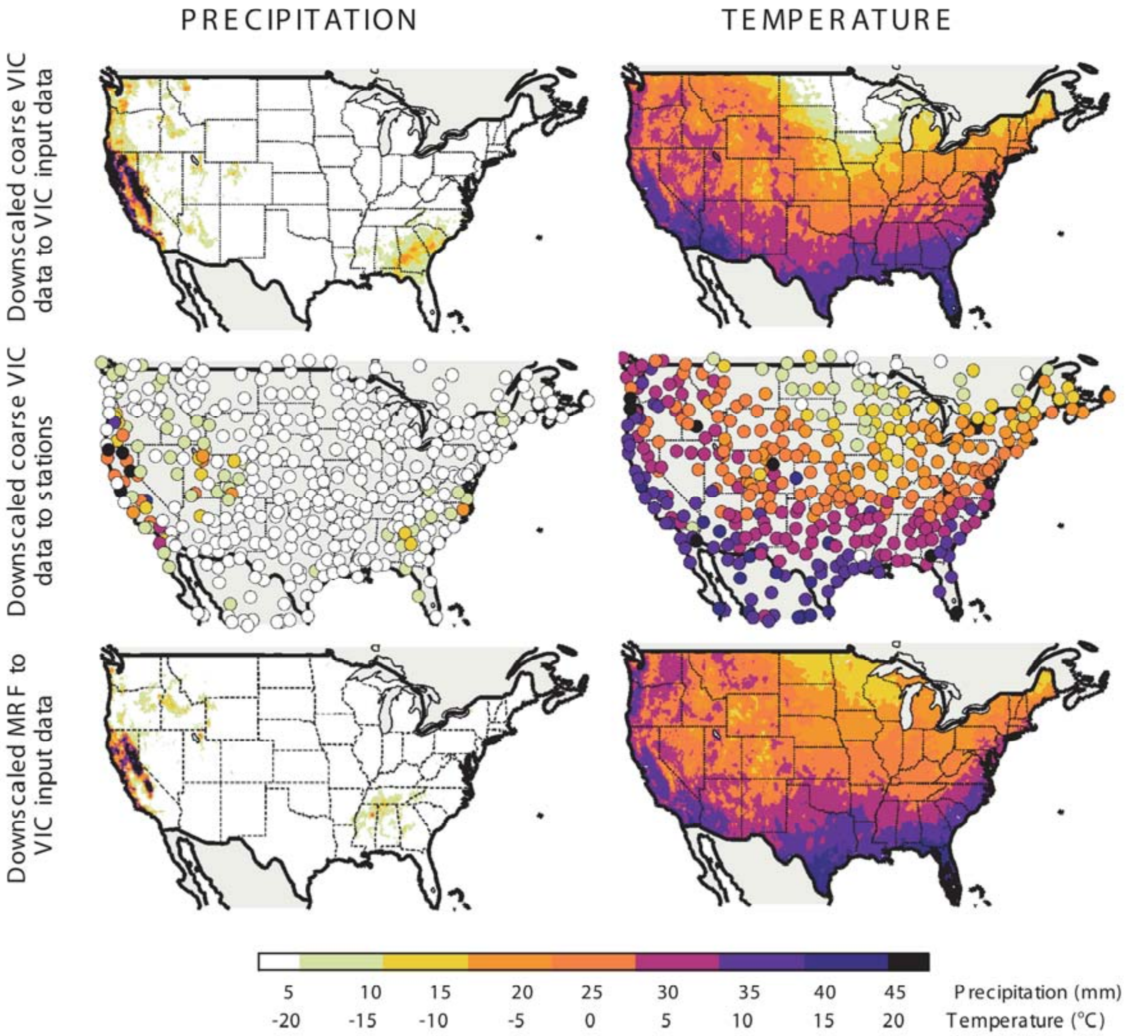

Figure 2. Downscaled precipitation (mm) and average air temperature (°C) versions of the January 27, 1983, California storm using constructed analogues, applied to coarse VIC data (top), station data (middle), and medium-range forecast (bottom).

Several meteorological parameters (obtained from the Reanalysis) covering different domains were tested for their suitability as predictors for downscaling VIC-scale P and Tavg (not shown). The Reanalysis parameters tested included Z_{700}, Z_{500}, Tavg, P, SLP, and the divergence of the Z_{500}. In general, the method showed skill in downscaling coarse-scale Tavg to the 1/8 degree grid using Reanalysis circulations (e.g., Z700 and Z500, not shown) and Tavg data (Table 1). However, the procedure for downscaling P showed skill in reproducing the continental patterns only when P was used as the predictor (Table 1).

Table 1. Median of correlation and RMSE between downscaled and observed daily P and Tavg using the reanalysis and VIC coarse data sets as predictors for the conterminous United States. The first (second) correlation represents the daily values with and (without) the seasonal cycle. The statistics shown are the median values from 52,135 gridpoints that cover the conterminous United States at 1/8 degree resolution.

	Precipitation$^{1/2}$	**Temperature**
(a) Reanalysis data as the coarse scale field		
R	0.52 (0.46)	0.96 (0.78)
RMSE	1.02 (mm day)$^{1/2}$	2.82 (°C)
(b) VIC data aggregated to 2.5 x 2.5 degrees as the coarse scale field		
R	0.72 (0.70)	0.97 (0.87)
RMSE	0.72 (mm day)$^{1/2}$	2.16 (°C)

The "$^{1/2}$" in the table indicates that the researchers took the square root of the precipitation before conducting the analysis.

Interestingly, large-scale models and predictions of P are notoriously inaccurate (e.g., Palmer et al. 2005), and Reanalysis precipitation fields are likewise understood to be fairly poor, if for no other reason than that the Reanalysis process did not assimilate or use any actual observations of precipitation (Kalnay et al. 1996). Thus, this study's authors suspect that the usefulness of the P fields in the downscaling procedure derives not from their adequacy as representations of the actual precipitation patterns on a given day but rather because they are low-dimensional representations of all the large-scale atmospheric influences that create potentials for precipitation on that day. One could use a Z500 field (or divergence field) as a P predictor, but would in the process generally fail to capture the influences of, say, humidities and temperatures on P, or of circulations at other height levels. Of course, one might choose to also include some of these other predictors, but parts of the atmospheric state that are associated with P would still be neglected.

Adding more and more predictors would accomplish two things: (1) introduce large degrees of freedom and collinearity into the prediction process, and (2) if all went well, effectively converge on a set of relations approximating those that are also the aim of builders of physically based climate models. It is better to accept that the Reanalysis P field is a compact representation of our best understanding of all the various atmospheric processes and variables that converge to yield the large-scale potential for precipitation. As such the Reanalysis P fields are viewed as summaries of the many (large-scale) atmospheric precipitation forcings. In support of this philosophy, note that Widmann et al. (2003) found that numerically simulated P from the Reanalysis is a better predictor than atmospheric circulations, Tavg, and humidity for downscaling monthly P in Washington and Oregon. A discussion on the suitability of predictors for downscaling P can be found in Wilby and Wigley (2000).

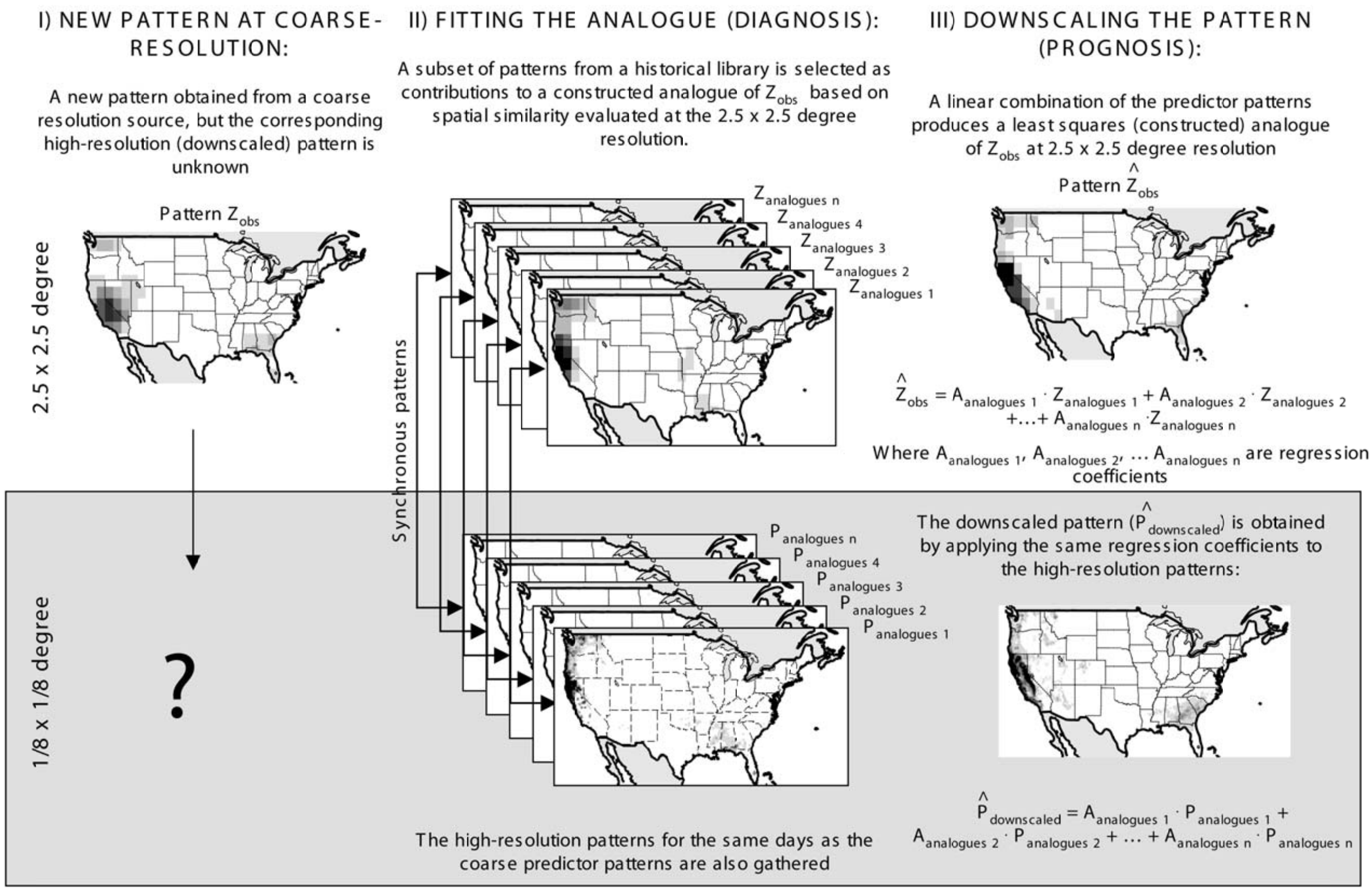

Figure 3. Schematic of the method of constructed analogues for downscaling reanlysis fields from 2.5° x 2.5° grids to 1/8° x 1/8°

For the purposes of the demonstrations presented here, to eliminate the influence of some of the historical inaccuracies of the Reanalysis fields, the Maurer et al. (2002) meteorological dataset at 1/8 degree resolution (hereafter referred to as the *VIC dataset*) was aggregated into averages at 2.5 degree resolution (hereafter referred to as the *coarse VIC dataset*) to be used as an alternative more directly observationally based (coarse resolution) predictor (than the modeled Reanalysis P values). Among all predictors tested, the coarse VIC dataset yielded the highest skill in reproducing daily P in the high-resolution validation dataset—even better than the Reanalysis P (Table 1). The increase in the skill from using the VIC coarse dataset suggests that inconsistencies between the Reanalysis P estimates and the observations previously noted by other authors (see Wilby and Wigley 2002; Charles et al. 2004) were limiting the downscaling process. Since the historical coarse VIC dataset is available and (apparently) eliminates some of the Reanalysis biases and errors, it is reasonable to use it for the demonstrations here, as well as in other downscaling efforts not involving the Reanalysis model. That is, why mix Reanalysis errors and biases with those from other models when we can reasonably estimate large-scale predictors from more direct observational fields? Preliminary studies by the authors downscaling control and climate-change runs from the NCAR Parallel Climate Model (to be presented separately) using the coarse VIC fields to construct analogues yield results for the western United States that encourage use of these predictors. In the rest of this report, the P from the coarse VIC calibration dataset will be used as a P predictor to downscale the independent validation dataset. For consistency, a coarse Tavg VIC dataset will also be used (Table 1). Thus, the results presented hereafter represent the best skill that can be obtained from downscaling observed coarse-resolution P and Tavg continental patterns from the Maurer et al. (2002) dataset by the method of constructed analogues.

3.3. Diagnosis: Constructing a Coarse-resolution Analogue

Several schemes were tested for selecting the subset of predictors from the large library of available weather patterns. In principle, all previous weather patterns could be included in the regressions that form the analogue, but this approach results in large inefficiencies, because there is a point where the correlation downscaling skill does not increase significantly by adding more predictors, and the variance downscaling skill is reduced somewhat compared to more parsimonious approaches (Fernández and Sáenz 2003). Alternatively, a very few patterns known to be very similar to the target pattern could be used to form the analogue; however, this approach reduced the d.o.f. of the process and much reduced the correlation downscaling skill (Figure 4). An intermediate approach using a moderate number of historical patterns to construct the analogues proved to yield better results. Therefore, for each weather pattern to be downscaled, a subset of the 30 patterns most spatially similar to the target pattern was selected for use in the constructed analogue for that day. The 30 patterns were selected from a pool of 2275 seasonally confined patterns potential patterns (see below). The number of historical analogue patterns to incorporate into the constructed analogues (n=30, in this report) was selected to keep numeric computational requirements low (the size of the matrix to be inverted depends on n), while maintaining correlation skills that are high for both precipitation and temperature (Figure 4). For n=30, precipitation correlation skill reached a plateau and therefore the inclusion of more patterns would not be justified, while, by that number of analogues, the temperature correlation skill was already very high. Spatial similarity was measured in terms of the spatial root mean square error (RMSE) between the weather pattern to be downscaled and each of the 2275 potential patterns from the library of historical weather patterns. The pool of 2275 potential predictors was composed by the weather patterns from the calibration dataset that were within ±45 days of the day of year of the target pattern (i.e., seasonally confined) for all the calibration years available in the historical record (25).

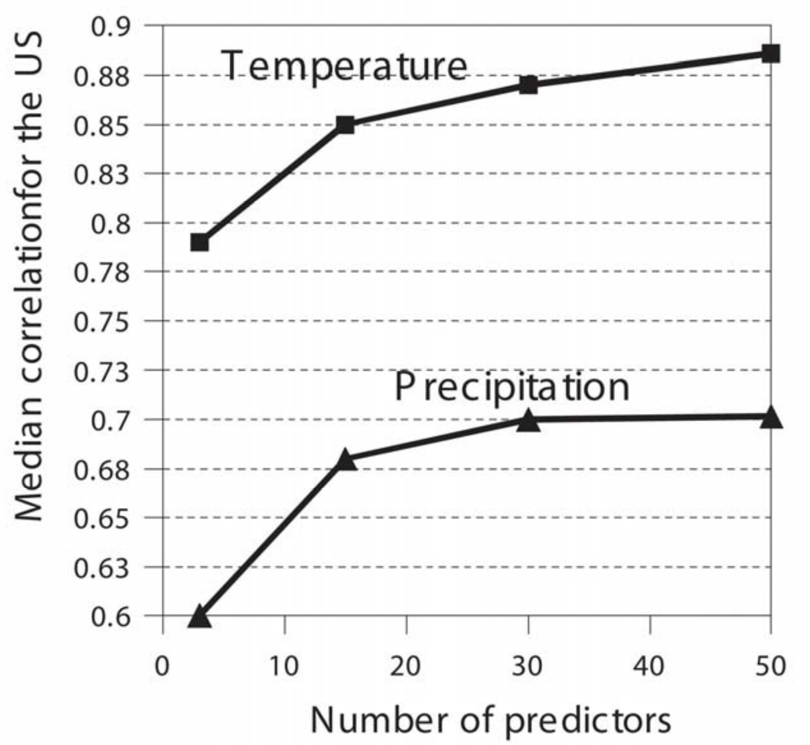

Figure 4. Change in the median correlation skill of downscaled precipitation and temperature estimates by increasing the number of suitable predicators (n)

The relative skill of the constructed analogues (in reproducing observed fine-scale P and Tavg variations at daily time scales), compared tcompared to other alternative reference methods can be measured with the Brier skill score (BSS, Brier 1950). Defining MSE and MSEref as the mean square errors from the constructed analogue downscaling method with the preferred 30-member best-suited subset and the reference methods, the BSS can be formulated as:

$$BSS = \left(1 - \frac{MSE}{MSEref}\right) x100 \qquad (1)$$

The lower (or more negative) the BSS value, the more poorly the preferred method performed relative to the reference method, while BSS=100 would be associated with performance by the preferred method that is much better than the reference method. In Table 2, three reference methods are compared to our preferred method: (1) picking one of the 2275 potential predictors from the library at random for use in the downscaling method, (2) using the mean average daily patterns as downscaling estimate (climatology), and (3) using the single most spatially similar pattern (according to RMSE) from the pool of potential predictors—that is, using the closest natural occurring analogue (Table 2). The constructed analogues method showed significant higher skill at reproducing the historical high-resolution P and Tavg patterns over all the reference methods (all the BSS scores are positive). As expected, the preferred constructed analogues showed the highest skill in comparison to the reference method that consisted in a random selection of patterns. In the case of P, the BSS values of the preferred method using climatology as reference are comparable to the BSS values using the closest naturally occurring single analogue as reference. This suggests that for each day the closest single P analogue is really not very similar to the target pattern, and also suggests that climatology is just as good an analogue (see VDD). In contrast, the preferred method outperforms more significantly the climatology method than the naturally occurring analogue reference method; indicating that that the Tavg naturally occurring analogues are better than the P's analogues.

Table 2. Median Brier skill scores comparing the constructed analogue downscaling method to three different reference approaches: (a) Using a randomly selected weather pattern from a pool of 2275 patterns (see text) as downscaling estimation, (b) Using only the climatological mean values at each grid cell, and (c) using the closest naturally occurring analogue to the pattern to be downscaled as downscaling estimation. The statistics shown are the median BSS values from 52,135 gridpoints that cover the conterminous US at 1/8 degree resolution.

	Precipitation$^{1/2}$	**Temperature**
(a) Randomly selected pattern	72	88
(b) Climatological mean values at each cell	42	74
(c) Single most spatially similar pattern	42	57

Once the pool of predictor patterns has been selected for a given coarse-resolution Tavg or P pattern for a certain day and year (Z_{obs}), an analogue of that pattern ($\hat{Z}_{obs}$) can be constructed as a linear combination of the (preferred 30-member most-suitable subset of) predictor patterns, according to:

$$Z_{obs} \approx \hat{Z}_{obs} = Z_{analogues} A_{analogues} \qquad (2)$$

where $Z_{analogues}$ is a matrix of the column vectors comprising the most-suitable subset of coarse-resolution patterns identified above specifically for Z_{obs}, and $A_{analogues}$ is a column vector of fitted least-squares estimates of the regression coefficients that are the linear proportions of the contributions of each column of $Z_{analogues}$ to the constructed analogue. The dimensions of the Z_{obs} matrix are p_{coarse} x 1, where pcoarse is the number of considered gridpoints contained in each coarse-resolution weather pattern; that is, Z_{obs} is a column vector. The dimensions of $Z_{analogues}$ are

p_{coarse}, x n, where n is the number of patterns in the most suitable predictors subset (i.e., 30), and the dimension of $A_{analogues}$ is $n \times 1$.

Assuming $Z_{analogues}$ has full rank (n), and using the definition of the pseudo-inverse

(Moore-Penrose inverse), $A_{analogues}$ is obtained from Equation 2 by:

$$A_{analogues} = \left[\left(Z'_{analogues} Z_{analogues} \right)^{-1} Z'_{analogues} \right] Z_{obs} \tag{3}$$

where the ' superscript denotes the transpose of the matrix.

3.4. Prognosis: Downscaling a Weather Pattern

To downscale the Z_{obs} pattern, the coefficients $A_{analogues}$ from Equation 3 are applied to the high-resolution weather patterns corresponding to the same days as the coarse-resolution predictors $Z_{analogues}$, according to:

$$\hat{P}_{downscaled} = P_{analogues} A_{analogues} \tag{4}$$

From Equation 3:

$$\hat{P}_{downscaled} = P_{analogues} \left[\left(Z'_{analogues} Z_{analogues} \right)^{-1} Z'_{analogues} \right] Z_{obs} \tag{5}$$

where $\hat{P}_{downscaled}$ is a constructed high-resolution analogue (e.g., a P pattern on the VIC 1/8 degree grid) and $P_{analogues}$ is the set of high-resolution historical patterns corresponding to the same days as the $Z_{analogues}$. The dimension of the $\hat{P}_{downscaled}$ vector is p_{VIC} x 1, and the dimension of the $P_{analogues}$ matrix is p_{VIC} x 1, where p_{VIC} is the number of gridpoints in the high-resolution weather patterns. Note that the matrix, $Z'_{analogues}Z_{analogues}$, inverted with each application of the procedure is only of dimension $n \times n$, and therefore the numerical computational resources needed to downscale the weather patterns are determined by the number of the patterns included in the most-suitable subset, and can be quite small.

4.0 Results

4.1. Selection of Most-suitable Predictors

Although predictor patterns in the most-suitable subsets are selected from different (calibration) years than the target patterns, for Tavg the most-suitable patterns turn out to be drawn mostly from the same times of year as the target patterns. For P, interquartile range (IQR) of distance of the predictors to the target patterns is comparable to the IQR obtained by selecting 30 daily patterns at random from 91-day windows, suggesting that suitable predictors are not particularly drawn close to the time of the year of the target pattern, and instead can be found at other times within the seasonal window. During certain times of the year the predictors selected are not centered on the target pattern. This approach could potentially affect the estimation of the monthly climatologies, especially if the patterns are more frequently selected from distant days forward or backward from the pattern to be downscaled.

Although no reduction of the d.o.f. of the predictors through PCA is needed in the method presented here, the percentage of variance explained by the first principal component (VPC1) of the most-suitable subsets of predictors patterns provides insights into how similar the predictors are to each other as a function of the day of year (Figure 5). High VPC1 is associated with systems with lower spatial d.o.f. Higher VPC1 does not necessarily result in better fit, but a system with fewer d.o.f. presumably could be modeled using fewer independent predictors. For both P and Tavg, the subsets tend to have more similar and spatially smooth patterns (as indicated by large VPC1) during the wintertime (Figure 5). The seasonal cycle of VPC1 for P is notably more pronounced than for Tavg, with a large reduction in summer, suggesting that the summertime P analogue patterns tend to be much more spatially patchy and dissimilar from each other than are the P wintertime predictor patterns.

4.2. Daily Skill

Various measures of the skill with which downscaled P and Tavg for all days of the odd-numbered years from 1950 to 1999 reproduce the observed patterns on those same days are shown in Figures 5 to 7. As in previous studies, the downscaled patterns more accurately reproduce the observed patterns of Tavg than P.In this case, downscaling skill is measured by the temporal correlation coefficients at each grid cell, e.g., top two panels, Figure 6. Notice, however, that P correlations across large parts of the United States are greater than 0.75 ($p<0.0001$) overall as well as in winter, indicating impressive skill at capturing day-today precipitation variations on a 12 km grid from 2.5-degree gridded inputs. The highest overall skill scores in the day-to-day downscaled P patterns were obtained along the west coast of the United States, especially in winter, but downscaled wintertime P fields in the eastern United States yielded the most widespread high skills. Wintertime P in the western U.S. coast region typically contains many of the most intense storms in the country and, since the P data were not standardized locally, the resulting linear fits from Equation (4) tend to emphasize fitting the variability in this region, so that the region enjoys the highest wintertime skill.

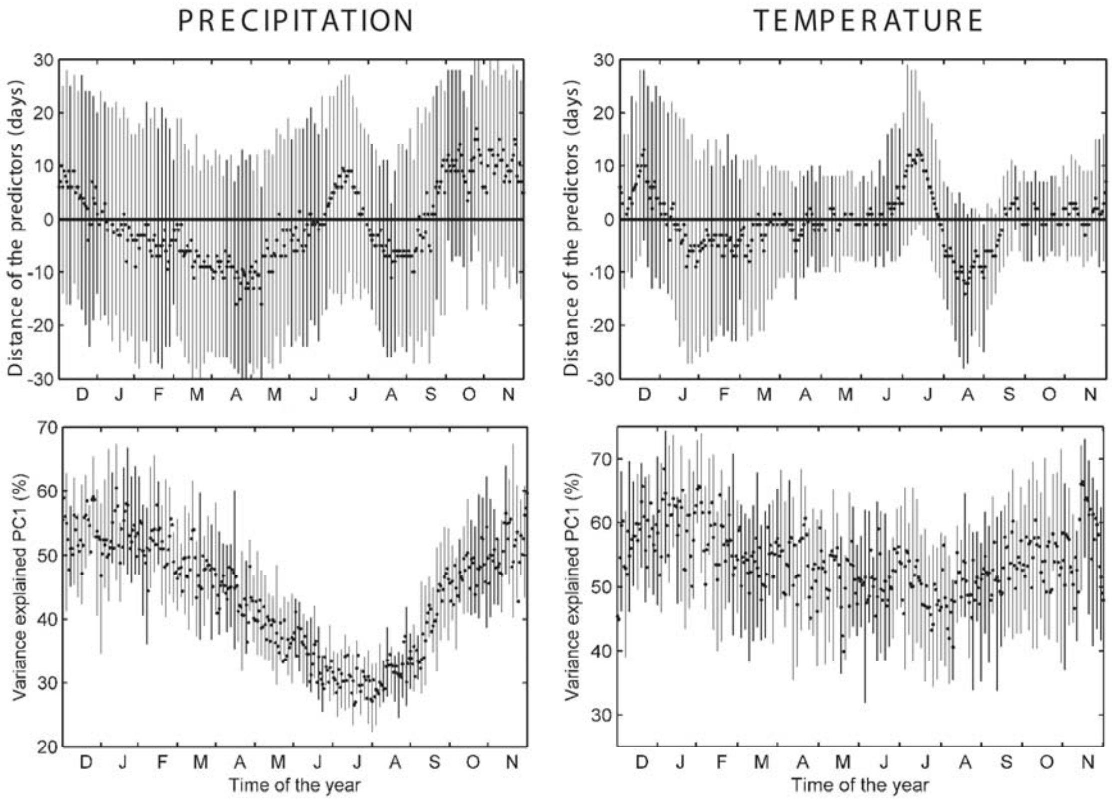

Figure 5. Median (dots) and interquartile range (error bars) of the distances between each pattern to be downscaled and its predictor weather patterns (top panel), and percentage variance explained by the first principal component of the most-suitable subset of predictors at each time of year (bottom panel). The spread of the quantiles represent the results of 25 downscaled years. Negative distances (positive) mean that given a weather pattern at a certain time of the year, more analoguous patterns are more frequently found earlier (later) during the year.

The correlations between downscaled and observed daily Tavg values are high throughout most of the country and most of the year (top two panels, Figure 7). Daily Tavg grid-wide RMSE errors in the downscaling are generally less than about 2°C (3.6°F), and are lowest during the summer (Figure 5). Thus the RMSE values for Tavg are relatively small compared to the seasonal normals throughout the year; in contrast, the RMSE of P during the summer (nearly 1 mm per day, Figure 5) is a considerable fraction of the seasonal values in many of the arid and semi-arid regions of the contiguous United States (Figures 6 and 7). Nonetheless, when the seasonal cycles of P and Tavg are removed (middle panels, Figure 6 and 7), the day-to-day correlations are still high. The primary weakness shown by the downscaled Tavg values are in the interior southwest during summertime.

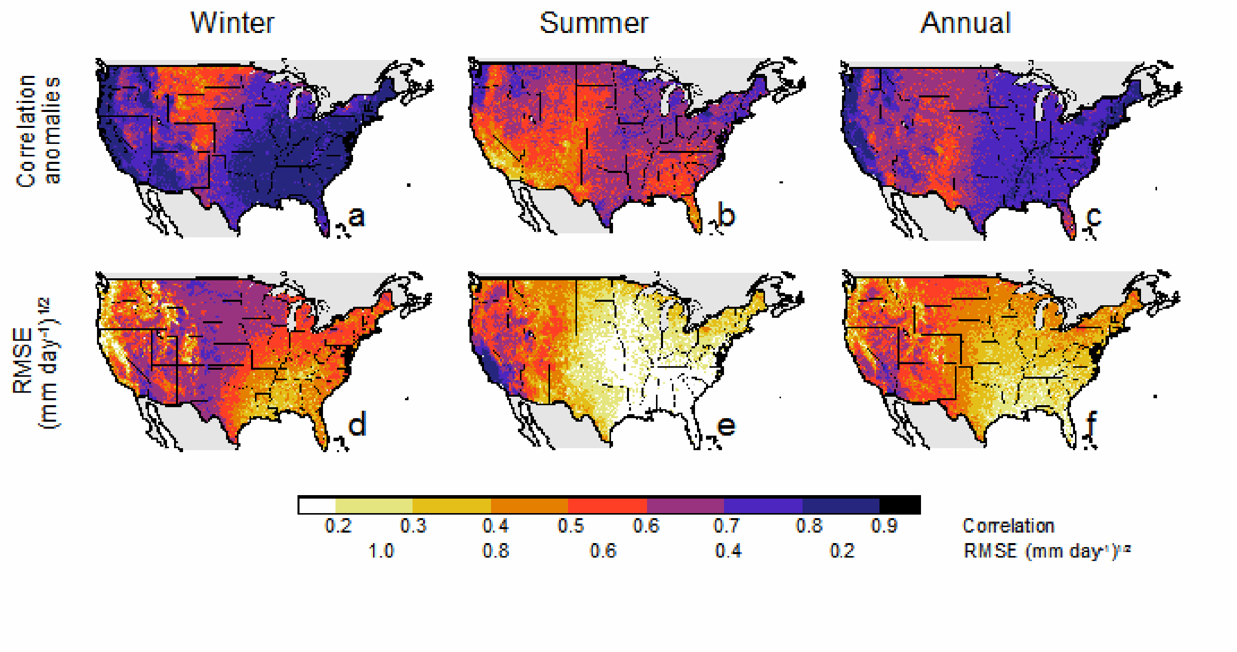

Figure 6. Cross-validated statistics comparing (square roots of the) downscaled daily precipitation rates to observed values (see text). The statistics shown are based on daily values from 25 years of winter (December through February), summer (June through August), and annual (all days of the year) seasons.

Downscaled P values tend to be overestimated in winter in the Pacific Northwest region and the northeastern region (Figure 8). During summer, downscaled P is overestimated in the northeastern region and underestimated in the northwest interior regions and eastern Texas (Figures 8). Biases in downscaled P values at the annual-mean scales are generally low, with slight underestimation in the interior regions. The P standard deviation is in general underestimated, especially in July (with the exception of California). When all the days of the year are considered, the western United States showed the smallest underestimates of P standard deviations.

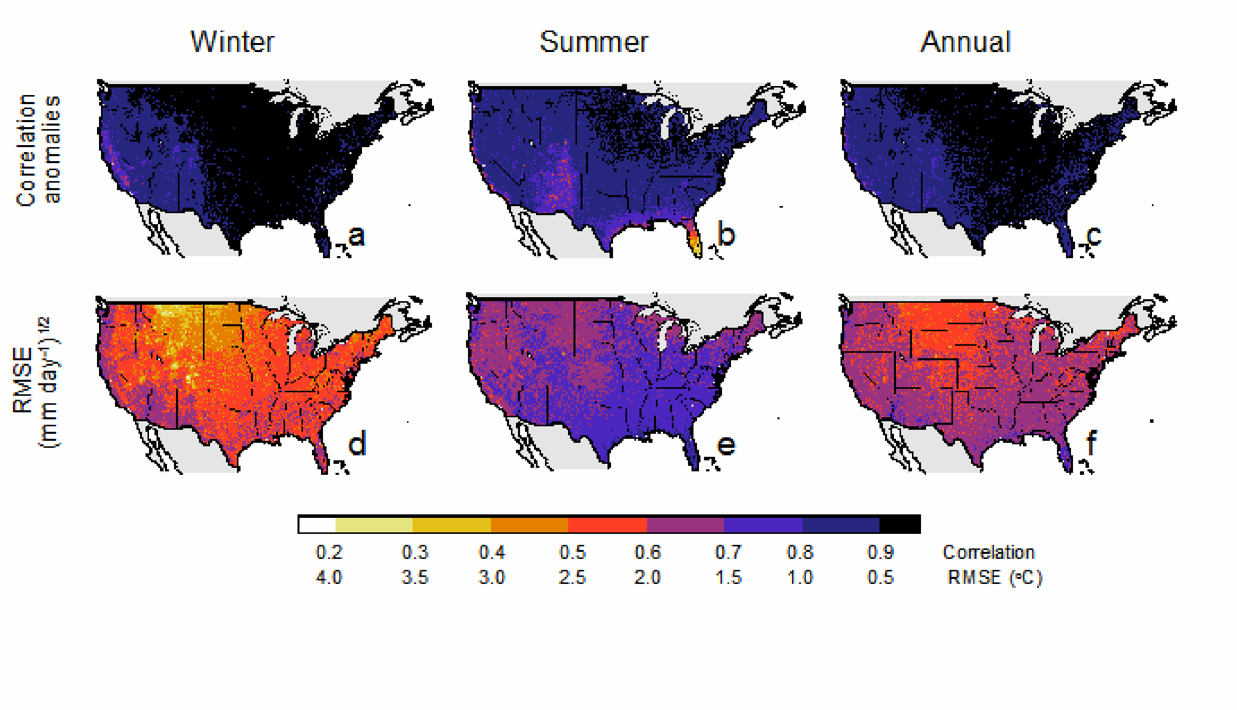

Figure 7. Same as Figure 6, but for daily average air temperatures

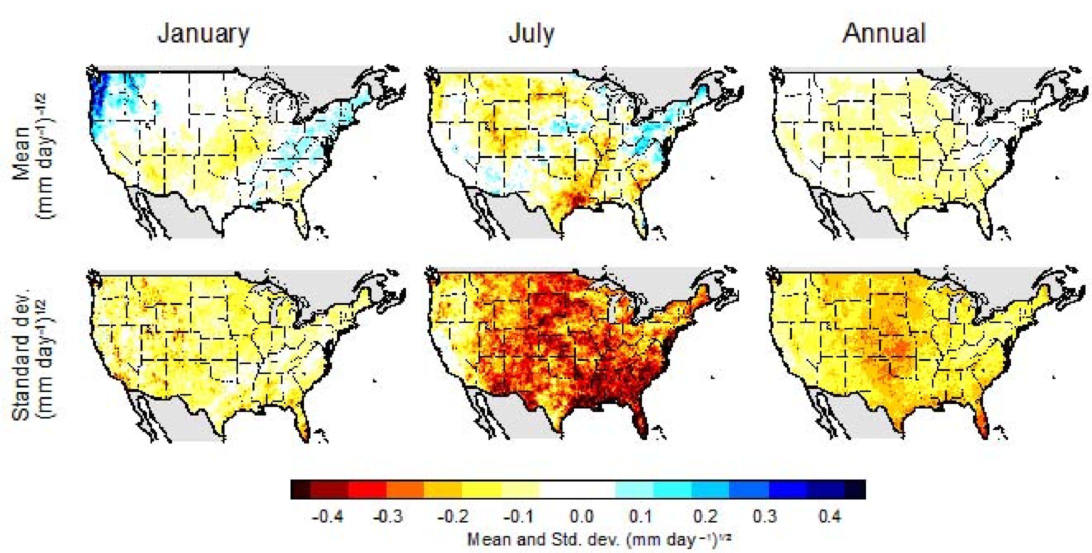

Figure 8. Difference in the means (top panel) and standard deviations (bottom panel) between downscaled minus observed fields of daily precipitation for January, July, and all days (annual)

Wintertime Tavgs are underestimated in the north-central states (Figure 9) and overestimated in the arid regions of the Western United States and eastern interior regions. Summertime Tavg overestimates are located in the West Coast states, and underestimates are found in the northeastern region. Accumulated to annual-means, however, the biases are virtually eliminated. Standard deviations of downscaled daily Tavgs are very similar to observed levels, with exception of modest underestimation in the interior west in winter and the southwest in summer (Figure 9).

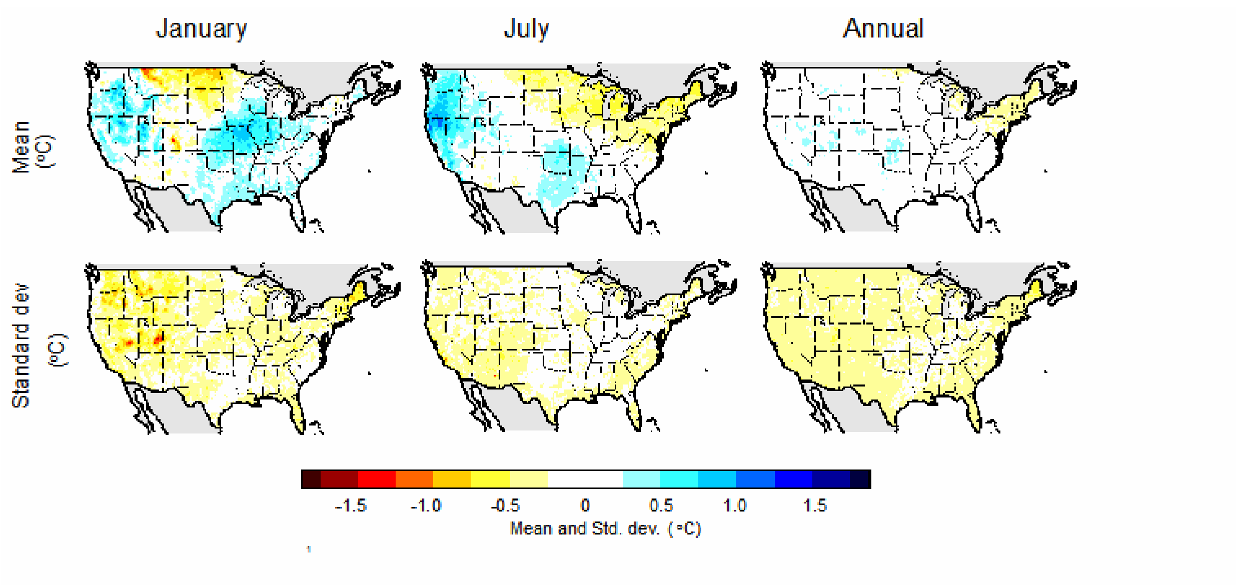

Figure 9. Same as Figure 8, but for daily average air temperatures

Scatter plots and percentile distributions of the downscaled versus observed P values for three example locations in the United States provide insights about the origins of the skill (Figure 10). Days with low P values are reproduced with lower skill than the high P days, and the numbers of wet days are overestimated in all three locations. In fact, a map of the overestimations of wet days shows that the downscaling procedure tends to "drizzle" in many regions of the United States (Figure 11a). However, in Figure 11b, a "zero-precipitation equivalent" is defined as the precipitation threshold below which the downscaled P values need to be set to zero in order to remove the (positive) bias in the number of wet days (as in Dettinger et al. 2004). (The zero-precipitation equivalent is also seen in the statistical distributions of Figure 10 as the downscaled P value that correspond to the highest percentile of all the observed dry days).   As can be seen, although the number of wet days in arid and semi-arid regions tends to be much too large, the zero-equivalent precipitation in those regions is quite low (0.3 mm per day and lower), indicating that the drizzle is quite light. Regions with higher zero-equivalent values tend to have lower biases in the number of wet days. In the case of Tavg, the daily skills and distributions of the downscaled estimations are very close to the observed data (Figure 12).

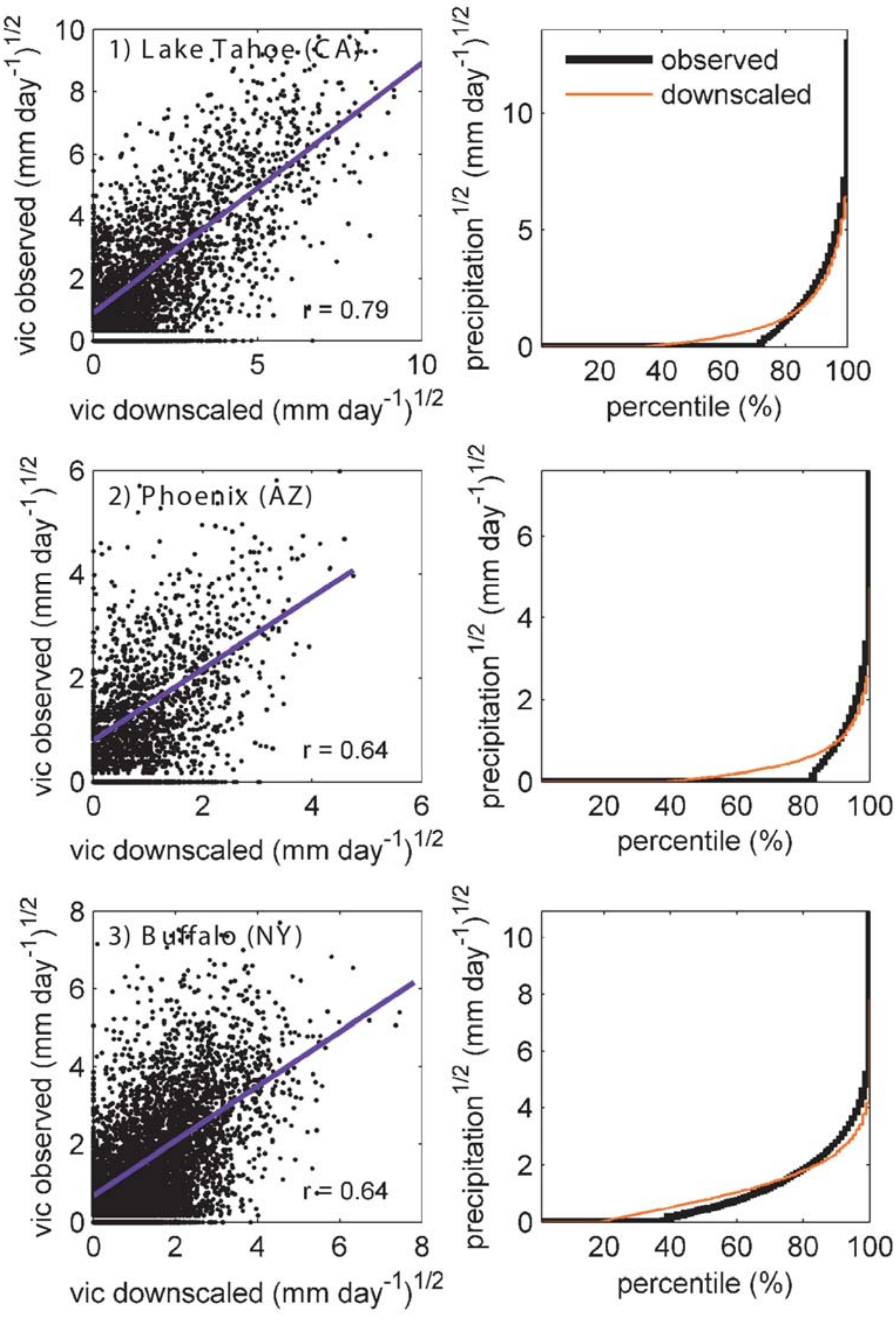

Figure 10. Scatter plots of downscaled versus observed daily (square root of) precipitation for three U.S. locations indicated in lower left panel of Figure 1 (left panel) and percentile distributions of the same time series (right panel); observations from Maurer et al. (2002).

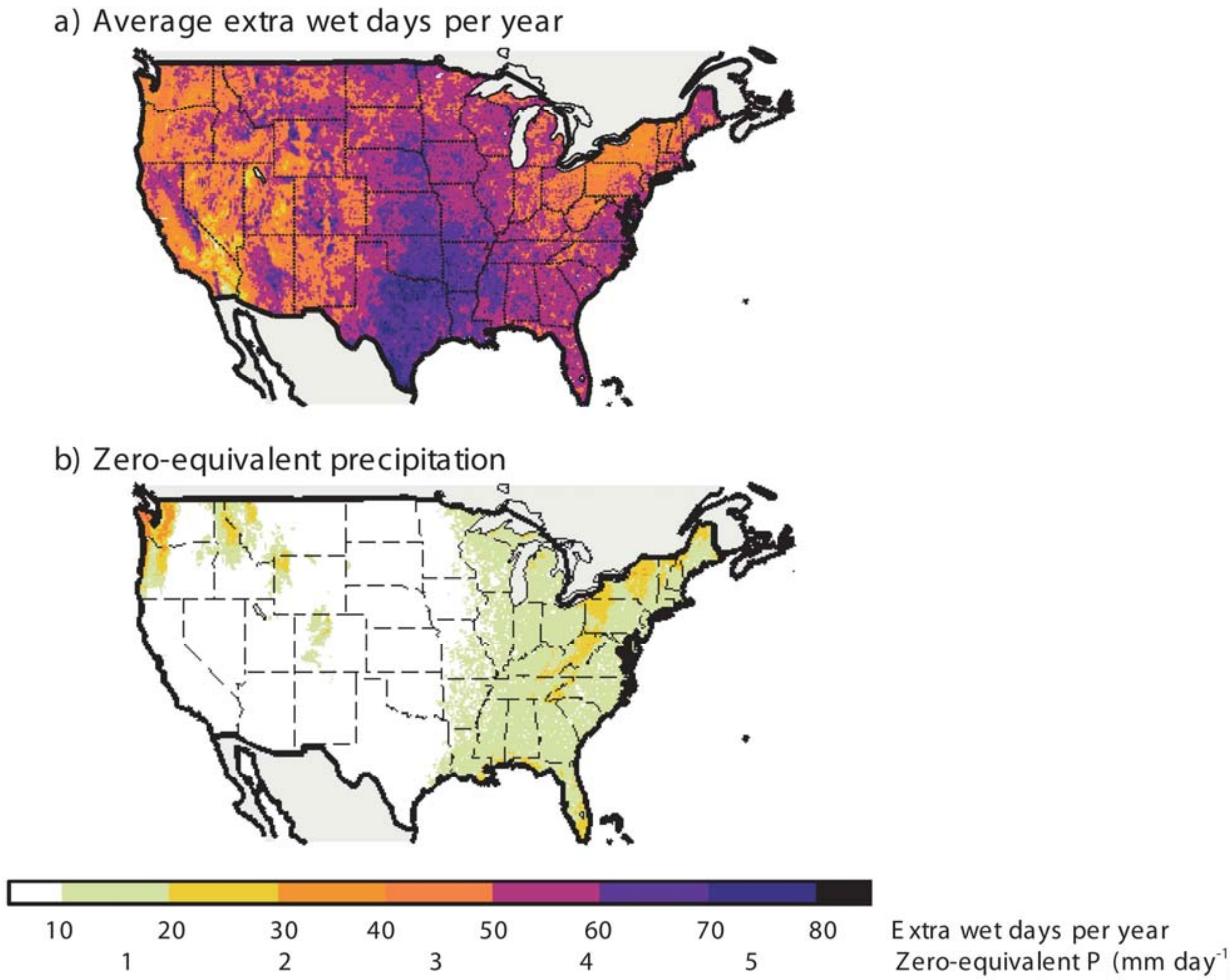

Figure 11. Average extra number of wet days per year in downscaled precipitation fields (compared to observations) and "zero-equivalent" precipitation. The zero-equivalent precipitation is the threshold below which the downscaled values would need to be set to zero in order to remove the (positive) biases in the number of wet days. A wet day is defined as a day when P>0.1 mm.

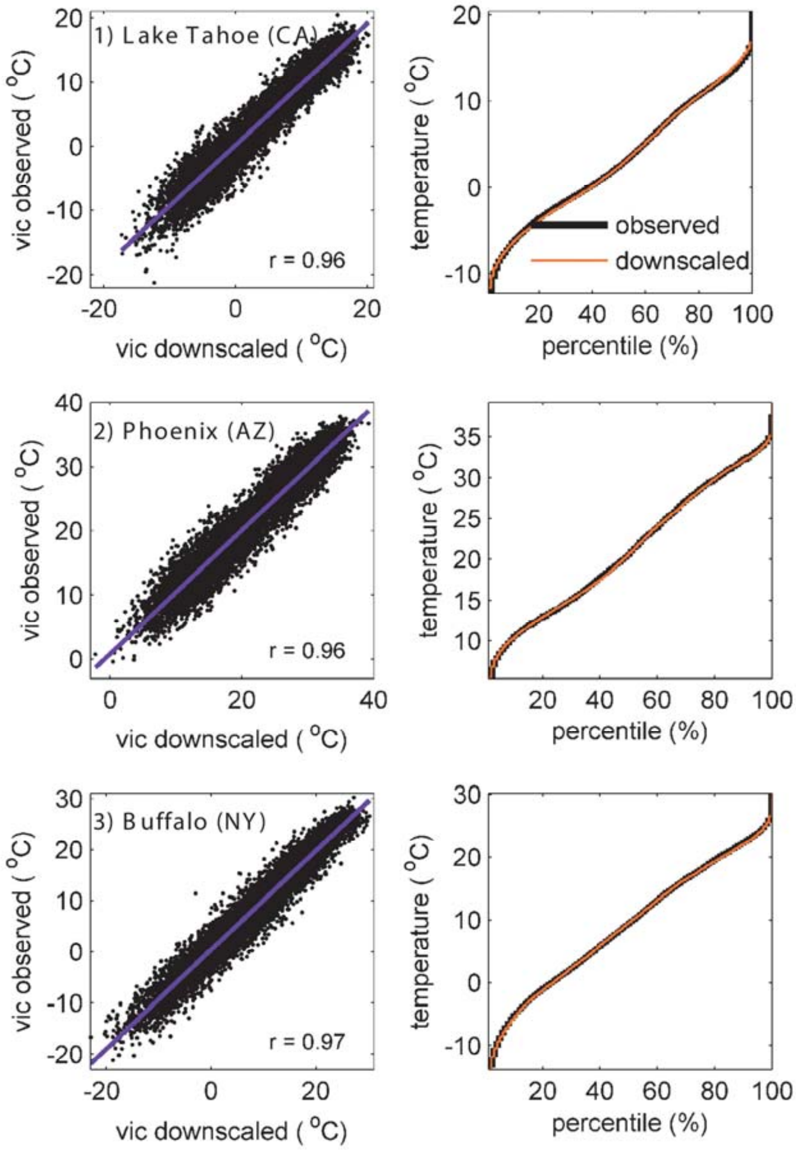

Figure 12. Same as Figure 10, but for daily average air temperatures

Implemented on a day-by-day basis, the downscaled weather patterns acquire any persistence or temporal structure from the temporal sequences of coarse-resolution fields that were downscaled. To see whether this transfer of temporal structure, and in particular day-to-day persistence, is adequate, the lag-one-day autocorrelation coefficients of the downscaled P and Tavg daily anomalies were compared to observed lag-correlations. Persistences in both the observed P and Tavg fields were approximately recaptured in the downscaled fields, although modestly underestimated in all cases (Table 3). The median one-day autocorrelation coefficients for observed and downscaled P are 0.57 and 0.51 and for observed and downscaled Tavg are 0.75 and 0.70 (Table 3). The spatial distribution of the observed lag-correlation coefficients is reproduced well in the downscaled fields (not shown).

Table 3. Lag-1 day autocorrelation coefficients of the anomalies of observed and downscaled precipitation and average air temperature. The coefficients shown are the median values from all 52,135 gridpoints at 1/8 degree resolution covering the conterminous United States.

	Observed	**Downscaled**
$Precipitation^{1/2}$	0.57	0.51
Temperature	0.75	0.70

Finally, to check for possible spatial smoothing of the weather patterns by the downscaling method, the daily magnitudes of P and Tavg gradients were computed at each gridpoint. In general, results showed insignificant differences between downscaled and observed gradient magnitudes, suggesting that the downscaled patterns are not significantly smoother (or sharper) than the observed weather patterns (not shown).

4.3. Monthly Skill

The application of the constructed analogues method to monthly-mean fields (instead of daily fields) yielded on average significant correlation skill for precipitation (Table 4), although the correlations are less than 0.5 for the internal continental regions, and the highest correlations were found for the western coastal states (not shown). For Tavg, the monthly-mean analogues resulted in high correlations that are comparable to the results obtained by averaging the daily downscaling estimates into monthly averages (Table 4). Over the entire year and for the whole United States, the median correlation skill between monthly downscaled P and observations is 0.61, while for Tavg is 0.88 (Table 4). For reference, similar correlation skill estimates for daily downscaling estimates are 0.72 for P and 0.97 for Tavg (Table 1).

Table 4. Same as Table 1, but for two cases

	$Precipitation^{1/2}$	**Temperature**
(a) averaging the daily downscaling results into monthly averages		
R	0.72 (0.86)	0.93 (0.99)
RMSE	0.20 (mm day)1/2	0.78 (°C)
(b) downscaling using monthly mean patterns		
R	0.61 (0.72)	0.88 (0.99)
RMSE	0.34 (mm day)1/2	0.92 (°C)

The monthly averaging of the daily downscaled results shows high skill for both P and Tavg for large regions of the contiguous United States (Figures 13 to 16). One reason why the daily P downscaling captures the observed variations at monthly scales significantly more skillfully than at the daily scale is that the largest relative errors occurred for low daily P values, while the variability of monthly mean tends to be dominated by larger P values. Another reason for the much better P fits of the monthly averages is that, despite the square root transformation used, the daily data are still skewed due to the many dry days present in the daily P time series. When the monthly averages are computed, the P data becomes more regular and less skewed with far fewer zero values. Overall, the aggregated monthly means of the daily Tavg and P downscaled estimations showed excellent agreement with the observations in large regions of the contiguous United States.

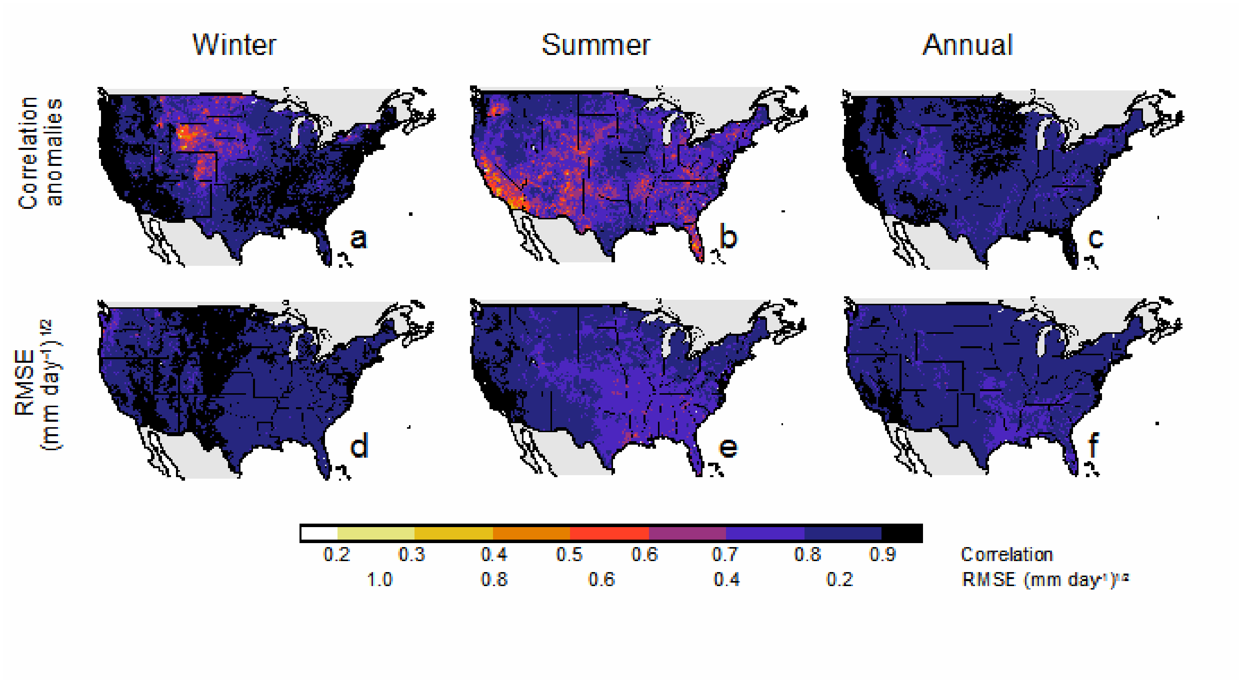

Figure 13. Same as Figure 6, but for monthly precipitation averages

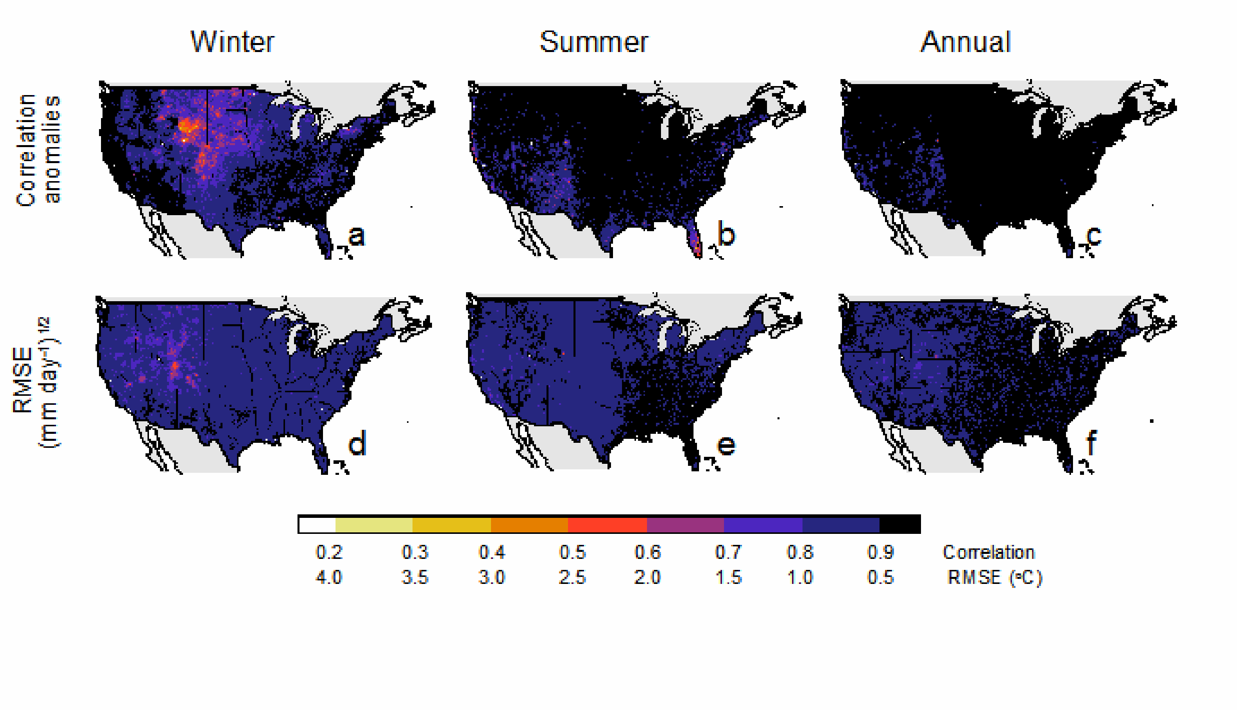

Figure 14. Same as Figure 7, but for monthly mean air temperatures

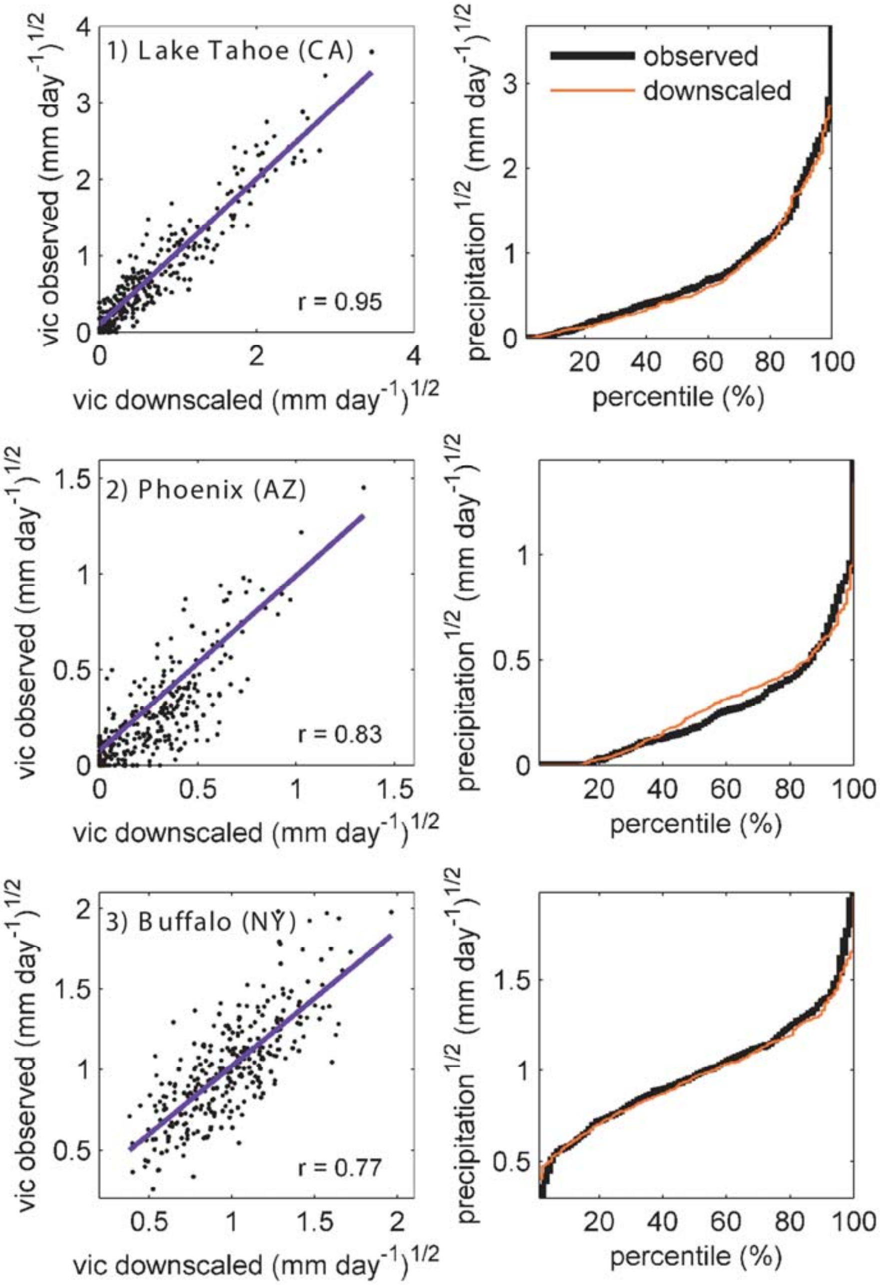

Figure 15. Same as Figure 10, but for monthly averages

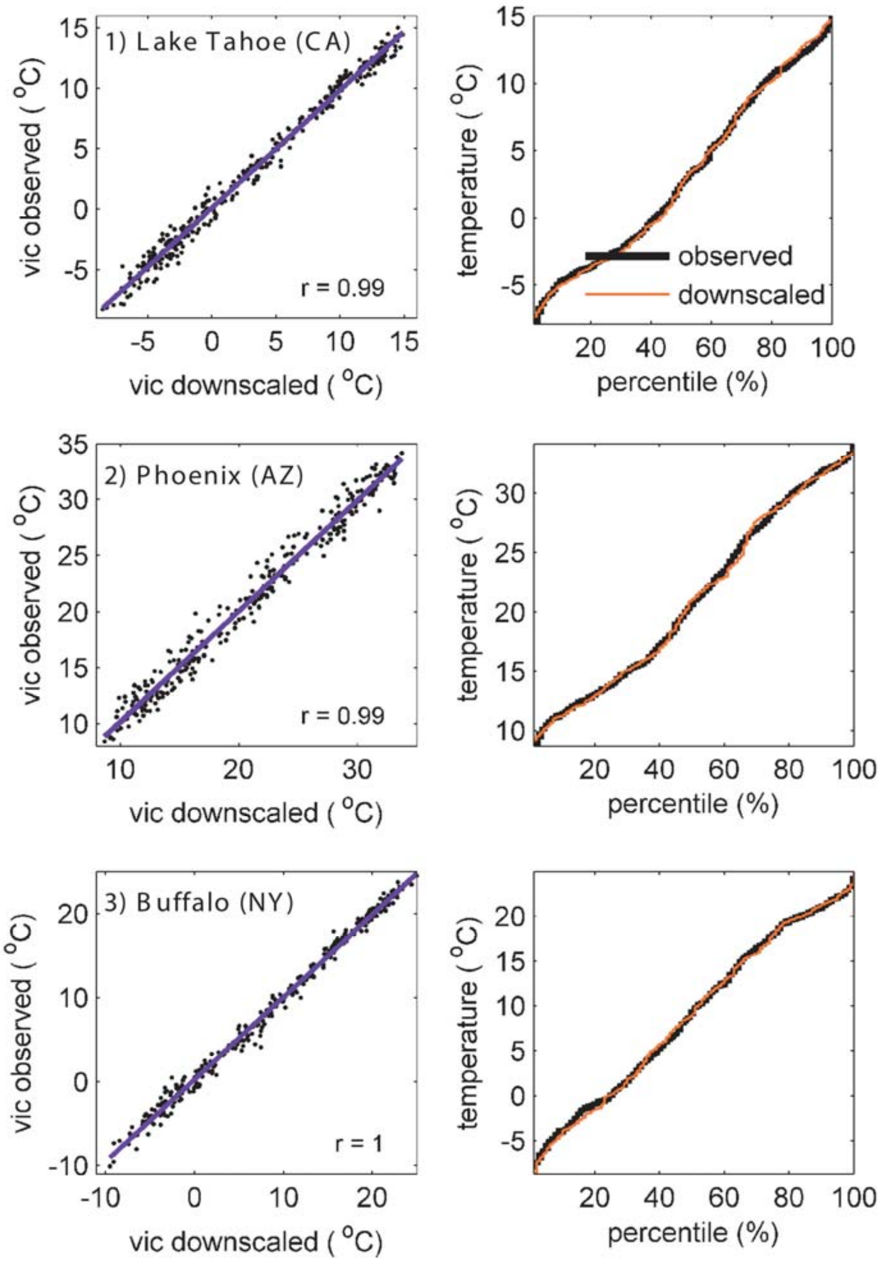

Figure 16. Same as Figure 11, but for monthly mean air temperatures

5.0 Domain Size and Skill over a Smaller Region: California Example

The downscaling method proves to be skillful at scales less than nationwide as well, and can be used to provide high-resolution climate fields for selected regions and even individual states. In producing regional estimates, though, decisions must be made regarding the domains of predictor patterns. That is, nothing in Equation 3 requires that the domains of $Z_{analogues}$ and $P_{analogues}$ be the same. Indeed, changing the domain of the predictors has a significant skill of the downscaled weather patterns over California (Figure 17). Reducing the domain of the predictors from the continental scale considered thus far, down to the 11 gridpoints that cover California at 2.5 degree resolution, proved to markedly increase the daily skill of the method in reproducing the P anomalies (top panels, Figure 17), but at the cost of reducing the skill in reproducing the Tavg anomalies (bottom panels). From this domain-dependence, it is evident that detailed P patterns over California have relatively small characteristic spatial scales (a more local nature) and relatively local atmospheric causes compared to Tavg. In contrast, the Tavg patterns derive from much larger, continental-scale atmospheric footprints, and therefore information from a larger region is needed in order to best downscale Tavg. These results suggest that selecting the predictors from a region close to the size of the western United States would result in high skill for both Tavg and P for many of the California gridpoints; however, because in the present implementation the P and Tavg predictors are chosen separately, the best downscaled results can be obtained by using state-scale predictors for P and continental-scale predictors for Tavg.

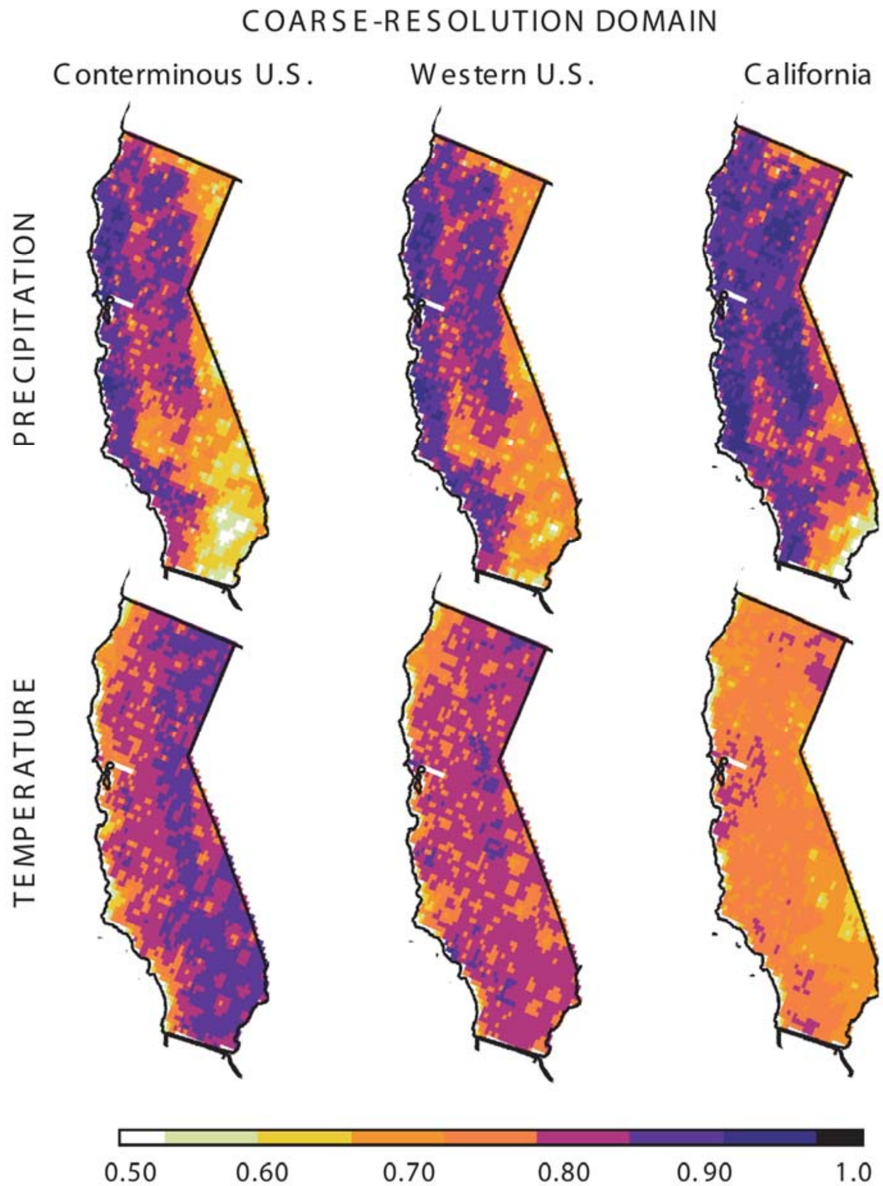

Figure 17. Effect of changing the domain of the predictor fields on day-to-day correlations between observed and downscaled precipitation and temperature anomalies in California. The average correlation and root mean square errors are shown.

6.0 Conclusions and Discussion

The downscaling method presented here proves to be quite simple and accurate, capturing an average of 50% of daily high-resolution P variance and an average of around 67% of Tavg variance, across all seasons and across the contiguous United States. The downscaled P variations capture as much as 62% of observed variance in the coastal regions during the winter. When the downscaled daily estimations are accumulated into monthly means, an average 55% of the variance of monthly P anomalies and more than 80% of the variance of Tavg monthly anomalies are captured. The downscaled linearly constructed weather patterns are spatially similar to observed patterns, and the temporal autocorrelation of the Tavg and P anomalies is only modestly (about 10%) underestimated. The biases in monthly climatologies are small for both P and Tavg, but the downscaling method tends to overestimate the number of wet days as it tends to produce drizzle during many observed dry days. This latter discrepancy can be fixed by various bias-correction procedures (e.g., Dettinger et al. 2004; Wood et al. 2004). However, no bias-correction was applied to the results shown in this study to allow a better evaluation of the true skill of the method. Moreover, preliminary analysis of hydrologic simulations of the downscaled data using the VIC model (Liang et al. 1994) for California have shown good agreement with similar simulations using observed meteorological data (to be presented separately).

The variables used as predictors and the selection of the most suitable patterns in the diagnostic part of the method are very important determinants of the method's skill, especially for P. The smaller-scale nature of daily P forcings and patterns, along with the many factors associated with P production, makes it more difficult to capture intermittent P patterns from smoothly varying atmospheric circulation patterns than it is to capture daily Tavgs. Because coarse resolution P fields (whether modeled, analyzed, or observation-based) are the variables that most closely match the spatial and temporal variability of the real-world P, and because they represent the best synthesis of the overall potential for P from the entire collection of large-scale atmospheric conditions, large-scale P fields are also the only predictor that yielded skillful results in reproducing the continental P patterns. In contrast, the daily Tavg patterns are more closely related to large-scale circulations changes, and therefore, although coarse Tavg fields were the most skillful predictors, significant (but somewhat less) skill could also be obtained by using atmospheric-circulation indicators, like Z_{500}, as the predictors.

The choice of predictor domains also affected the skill of downscaled P and Tavg patterns over the California region. Daily Tavg patterns were best captured from synoptic or larger scale predictor fields. In contrast, daily P patterns were captured better using P predictors over smaller domains, domains closer to the typical sizes of storm systems or roughly the state scale. When downscaling is implemented for evaluation of climate-change impacts on the state's resources, however, a more regional predictor domain will probably be a useful compromise, because (1) the loss of skill in using regional (western United States) rather than state-scale predictors is not large, and (2) California's resource systems—including water and power (Cayan et al. 2003), as well as air quality and ecosystems—are regionally linked and dependent.

Because the downscaled fields are derived from a deterministic combination of historical patterns, the results depend only on easily understood decisions regarding which and how many predictor patterns to include and how the most suitable of those patterns are selected. Consequently, the downscaled results should be very reproducible from group to group. Because no stochastic component is imposed externally upon the downscaled fields, subtle and potentially unexpected changes and differences in the intensities, frequencies, and even day-to-day timing of various weather types in model outputs will be directly reflected in the downscaled fields, only a single (reproducible) downscaled field is generated from each daily coarse-resolution weather field.

The constructed-analogues method presented here can be used to downscale output from a variety of different climate model applications, including downscaling of reanalyzed historical climate variations, diagnostic experiments (e.g., climate experiments with and without tropical sea-surface temperature anomalies), medium-range to season weather and climate forecasts, and long-term climate-change projections. Presumably it could be used to patch missing data, at least during the period covered by various reanalyses (1948–present).

In a climate-change application, it will be important for the downscaling method to capture the high-resolution effects of largely unprecedented trends at general circulation model (GCM) scales. Projected P and Tavg trends will derive from changes in the frequencies and amplitudes of various daily weather patterns, some of which have already been witnessed in historical archives (but which will come more or less frequently or strongly under the changed climates) and some of which have not been witnessed before. Capturing the high-resolution consequences of the former (already witnessed) patterns is a particular strength of the method presented here; capturing the high-resolution consequences of the latter (truly new) patterns will depend on the flexibility offered by the construction of analogues. For example, to the extent that temperature trends take the form of extremely persistent zonally banded warming patterns, the downscaling process will have to either extract these "new" patterns from historical analogues or construct them wholesale from collections of other historical temperature anomalies.

The ability of the method presented here to accommodate these new "deformations" of the weather patterns associated with natural climate variability by externally derived trend patterns associated with anthropogenic climate change must be evaluated. Depending on the strength and distribution of new patterns (which in turn may depend on the model and trace gases emission scenario analyzed), it may become increasingly more difficult to find good analogues as the climate changes proceed into the future, because the predictors are based on a period in which the effects of anthropogenic climate change are small compared to natural variability. Fortunately, it is possible to monitor the skill with which large-scale analogues can be constructed, and that monitoring provides a tool for diagnosing when the historical record is failing to provide adequate predictors. In this regard, the preliminary analyses by the authors are encouraging, because they show, for the NCAR Parallel Climate Model and a single emission scenario, that large-scale analogues of projected climate-change trends could be constructed from historical patterns without degradation throughout the twenty-first century (results to be presented separately). Previous studies using other models and emission scenarios obtained good analogues of climate change weather patterns using historically observed patterns (e.g., Zorita et al. 1995). Thus, the downscaling method presented here offers many potential uses, relying where possible on the particular strengths of the large-scale models while providing flexibility for capturing unexpected changes in the large-scale climate projections and weather predictions.

7.0 References

Antic, S., R. Laprise, B. Denis, and R. de Elia. 2004. "Testing the downscaling ability of a one-way nested regional climate model in regions of complex topography." *Climate Dynamics* 23: 473–493.

Brier, G. W. 1950. "Verification of forecasts expressed in terms of probability." *Mon. Wea. Rev.* 78: 1–3.

Caplan, P., J. Derber, W. Gemmill, S.-Y. Hong, H.-L. Pan, and D. Parrish. 1997. "Changes to the 1995 NCEP operational medium-range forecast model analysis-forecast system." *Wea. Forecasting* 12: 581–594.

Cayan, D. R., M. D. Dettinger, K. T. Redmond, G. J. McCabe, N. Knowles, and D. H. Peterson. 2003. The transboundary setting of California's water and hydropower systems - Linkages between the Sierra Nevada, Columbia River, and Colorado River hydroclimates, in Diaz, H. F., and B. Woodhouse, eds., Climate and Transboundary Issues: 25 p.

Charles S. P., B. C. Bates, I. N. Smith, and J. P. Hughes. 2004. "Statistical downscaling of daily precipitation from observed and modeled atmospheric fields." *Hydrological Processes* 18: 1373–1394.

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.

Dettinger, M. D., D. R. Cayan, M. Meyer, and A. E. Jeton. 2004. "Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American River basins, Sierra Nevada, California, 1900–2099." *Climatic Change* 62: 283–317.

Diez, E., C. Primo, J. A. Garcia-Moya, J. M. Gutierrez, and B. Orfila. 2005. "Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts." *Tellus series A* 57: 409–423.

Fernández, J., and J. Sáenz. 2003. "Improved field reconstruction with the analog method: Searching the CCA space." *Climate Research* 24: 199–213.

Gangopadhyay, S., M. Clark, and B. Rajagopalan. 2005. "Statistical downscaling using K-nearest neighbors." *Water Resources Research* 41: Art. No. W02024.

Hamill, T. M., J. S. Whitaker, and X. Wei. 2004. "Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts." *Monthly Weather Review* 132: 1434–1447.

Hanssen-Bauer, I., E. J. Forland, J. E. Haugen, and O. E. Tveito. 2003. "Temperature and precipitation scenarios for Norway: Comparison of results from dynamical and empirical downscaling." *Climate Research* 25: 15–27.

Hay, L. E., and M. P. Clark. 2003. “Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States.” *Journal of Hydrology* 282: 56–75.

Kalnay, E., M. Kanamitsu, R. Kistler, et al. 1996. “The NCEP/NCAR 40-Year Reanalysis Project.” *Bull. Amer. Meteor. Soc.* 77: 437–471.

Kanamitsu, M. 1989. “Description of the NMC global data assimilation and forecast system.” *Wea. Forecasting* 4: 334–342.

Kanamitsu, M., J. C. Alpert, K. A. Campana, et al. 1991. “Recent changes implemented into the global forecast system at NMC.” *Wea. Forecasting* 6: 425–435.

Leung, L. R., and M. S. Wigmosta. 1999. “Potential climate chance impacts on mountain watersheds in the Pacific Northwest.” *Journal of the American Water Resources Association* 35: 1463–1471.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges. 1994. “A Simple Hydrologically Based Model of Land Surface Water and Energy Fluxes for GSMs.” *J. Geophys. Res.* 99(D7): 14,415–14,428.

Lorenz, E. N. 1969. “Atmospheric Predictability as Revealed by Naturally Occurring Analogues.” *Journal of the Atmospheric Sciences* 26: 636–646.

Martin, E., B. Timbal, and E. Brun. 1996. “Downscaling of general circulation model outputs: Simulation of the snow climatology of the French Alps and sensitivity to climate change.” *Climate Dynamics* 13: 45–56.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen. 2002. “A long-term hydrologically based data set of land surface fluxes and states for the conterminous United States.” *J. Clim.* 15: 3237–3251.

Murphy J. 1999. “An evaluation of statistical and dynamical techniques for downscaling local climate.” *Journal of Climate* 12: 2256–2284.

Palmer, T. N., G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung, and M. Leutbecher. 2005. “Representing model uncertainty in weather and climate prediction.” *Ann. Rev. of Earth and Plan. Sci.* 33: 163–193.

Roads J., S. C. Chen, and M. Kanamitsu. 2003. “US regional climate simulations and seasonal forecasts.” *Journal of Geophysical Research – Atmospheres* 108 (D16): Art. No. 8606.

Salathe, E. P. 2003. “Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin.” *International Journal of Climatology* 23: 887–901.

Sheffield, J., G. Goteti, F. Wen, and E. F. Wood. 2004. “A simulated soil moisture based drought analysis for the USA.” *J. Geophys. Res.* 109: D24108, doe:10.1029/2004JD005182.

Timbal, B. 2004. “Southwest Australia past and future rainfall trends.” *Climate Research* 26: 233–249.

Timbal, B., A. Dufour, and B. McAvaney. 2003. “An estimate of future climate change for western France using a statistical downscaling technique.” *Climate Dynamics* 20: 807–823.

Timbal, B., and B. J. McAvaney. 2001. “An analogue-based method to downscale surface air temperature: Application for Australia.” *Climate Dynamics* 17: 947–963.

van den Dool, H. M. 1994. “Searching for analogues, how long must one wait?” *Tellus* Ser. A 46: 314–324.

van den Dool, H., J. Huang, and Y. Fan. 2003. “Performance and analysis of the constructed analogue method applied to US soil moisture over 1981–2001. *Journal of Geophysical Research* 108(D16).

Weaver, C. P. 2004. “Coupling between large-scale atmospheric processes and mesoscale land-atmosphere interactions in the US Southern Great Plains during summer. Part I: Case studies.” *Journal of Hydrometeorology* 5(6): 1223–1246.

Wetterhall, F., S. Halldin, and C. Y. Xu. 2005. “Statistical precipitation downscaling in central Sweden with the analogue method.” *Journal of Hydrology* 306: 174–190.

Widmann, M., C. S. Bretherton, and E. P. Salathe. 2003. “Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor.” *Journal of Climate* 16: 799–816.

Wilby, R. L., and T. M. L. Wigley. 2000. “Precipitation predictors for downscaling: Observed and general circulation model relationships.” *International Journal of Climatology* 20 (6): 641–661.

Wilby, R. L., and T. M. L Wigley. 2002. “Future changes in the distribution of daily precipitation totals across North America.” *Geophysical Research Letters* 29. Art. No. 1135.

Wilson, R. C. 1997. Daily rainfall along the U.S. Pacific coast appears to conform to a square-root normal probability distribution. Proc. 13 Annual Pacific Climate Workshop, 19–32.

Wood, A. W., L. R. Leung, V. Sridhar, et al. 2004. “Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs.” *Climatic Change* 62: 189–216.

Zangl, G. 2004. “The sensitivity of simulated orographic precipitation to model components other than cloud microphysics.” *Quarterly Journal of the Royal Meteorological Society* 130: 1857–1875.

Zorita, E., J. P. Hughes, D. P. Lettenmaier, and H. von Storch. 1995. “Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation.” *Journal of Climate* 8: 1023–1042.

Zorita, E., H. von Storch. 1999. "The analog method as a simple statistical downscaling technique: Comparison with more complicated methods." *Journal of Climate* 12: 2474–2489.

8.0 Glossary

BSS	Brier skill score
COOP	cooperative observer
CRD	Scripp's Climate Research Division
d.o.f.	degrees of freedom
GCM	general circulation model
hPa	hectopascal
IQR	interquartile range
Km	kilometer
LDAS	Land Data Assimilation Systems
MRF	medium-range weather forecast
MSE	Mean square error
MSEref	Mean square error, reference
NCAR	National Center of Atmospheric Research
NCDC	National Climatic Data Center
NCEP	National Center of Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
P	precipitation
PCA	Principal Component Analysis
PRISM	Parameter-elevation Regressions on Independent Slopes Model
RCM	regional climate model
RMSE	root mean square error
SLP	sea level pressures
Tavg	average temperature
Tmax	maximum temperature
Tmin	minimum temperature
UTC	Coordinated Universal Time

VDD	van den Dool
VIC	Variable Infiltration Capacity
VPC1	variance explained by the first principal component