THE CONTINUUM OF DROUGHT IN WESTERN NORTH AMERICA

by

Toby Rollin Ault

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF GEOSCIENCES

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2011

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Toby Rollin Ault entitled "The Continuum of Drought in Western North America" and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Date:

Julia E. Cole

Date:

Jonathan T. Overpeck

Date:

Joellen Russell

Date:

Julio L. Betancourt

Date:

Bette Otto-Bliesner

Date:

Michael N. Evans

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.
I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Date:

Dissertation Director: Julia E. Cole

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

SIGNED: Toby Rollin Ault

ACKNOWLEDGEMENTS

I am grateful for the investment of time and energy that my advisor, Julia Cole, and my committee, Jonathan Overpeck, Julio Betancourt, Joellen Russell, Michael Evans and Bette Otto-Bliesner, have put into my research. I received significant lab assistance from Heidi Barnett, Ali Kimbrough, and Sarah Truebe. I also thank Cody Routson, Sheridan Stone, Dave Anderson and Mihai Ducea for logistical and technical assistance throughout the project, and I thank Dave Meko, Connie Woodhouse, Jon Pelletier, Jay Quade, Malcolm Hughes and Clara Deser for their feedback and helpful comments. Several former and current graduate students have assisted me in my research including Scott St. George, Greg T. Pederson, Suz Tolwinski, Nick McKay, Alison Macalady, Diane Thompson, Jeremy Weiss, Jen Wagner, Tim Shanahan and Kevin Anchukaitis. Support for my research has come from NSF, GSA, NOAA, IoE, and the University of Arizona Department of Geosciences.

DEDICATION

To Jenny.

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

LIST OF FIGURES

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF TABLES

ABSTRACT

The continuum of western North American hydroclimate during the last millennium is analyzed here using instrumental records, proxy data, and global climate model (GCM) simulations. We find that variance at long timescales (low frequencies) is generally more substantial than variance at short timescales (high frequencies). We find that local sources of autocorrelation (e.g., soil moisture storage) likely explain the tendency for variance to increase from monthly to interannual timescales, but that variance at longer timescales requires remote climate sources of variability. Our analysis of global climate model data indicates that at least one fully coupled GCM can reproduce the characteristics of the continuum on short (interannual) and long (multicentury) timescales, but that proxy spectra and GCM spectra disagree about the amount of variance present on intermediate (decadal to centennial) timescales. Since instrumental records, as well as multiple independent types of paleoclimate records, provide evidence that variance increases with timescale at these frequencies, and because numerical experiments indicate that local autocorrelation is not a likely source of variance at these timescales, we argue that climate model simulations underestimate the full range of low-frequency drought variability. Moreover, the models may also underestimate the risk of future megadroughts, which we attempt to quantify using a new method that combines frequency information from observational data with projections of 21st century hydroclimate. Our results indicate that the risk of a severe, decadal-scale drought during the coming century is at least 1-in-10 for most of the US Southwest, and may be as high as 1-in-3. These findings should be incorporated into adaptation and mitigation strategies to cope with regional climate variability and climate change.

INTRODUCTION

0.1 Motivation

During the 20th century, North America experienced two periods of prolonged aridity: the 1930s "dust bowl" and 1950s Southwest drought. Archives of hydroclimate fluctuations during the last few millennia indicate that droughts like these occur naturally one or two times per century, and that much more prolonged periods of aridity (megadroughts) are also part of the natural range of variability (e.g., Woodhouse and Overpeck, 1998). In the past, megadroughts have had devastating consequences for people and societies worldwide. The first step in preventing similar impacts in the future is to assess how likely these events are to occur. To do so requires us to understand the spatial and temporal characteristics of megadroughts, and ideally to elucidate their mechanistic origins. Since megadroughts occur infrequently and on long timescales, 20th century records alone paint an incomplete picture of the full range of low-frequency hydroclimate behavior (e.g., Cook et al., 2004, 2010). To explore a more comprehensive range of megadrought behavior, we must therefore turn to paleoclimate archives of hydroclimate fluctuations in the past, and multi-century climate model simulations. Both of these types of data provide a more robust view of the frequency, magnitude, and duration of megadroughts.

To compare paleoclimate indicators of drought and hydroclimate variables in climate model simulations, we rely heavily on spectral analysis, which allows us to ignore the time-evolving characteristics of drought and focus instead on the relative importance of different frequency components across a continuum of timescales. This research is particularly motivated by the observation that drought variance is generally greater at long timescales than at short timescales – a phenomenon referred to as "redness." Although there are many plausible sources of redness, we focus on two extreme cases: redness caused by non-climate sources of local "autocorrelation," and redness caused by the climate system. In other words, we are interested in determin-

ing if redness in hydroclimate arises because drought variables behave like weakly-damped, stochastically-forced processes (Delworth and Manabe, 1989; Hasselmann, 1976), or because the climate continuum itself is red (Huybers and Curry, 2006). It is important to distinguish between these two possibilities for several reasons. First, redness caused by local autocorrelation could give rise to megadroughts, but at a much more infrequent rate (e.g., Pelletier and Turcotte, 1997). Hence understanding the continuum of hydroclimate variability could also improve our risk assessments of drought in the future. Second, if the sources of redness in drought variables are local, then drought variability may not be very predictable on timescales longer than a few months or seasons. On the other hand, if there are climate sources of redness that can identified, they may offer the most potential for making decadal "forecasts" of climate, which in turn would be highly valuable tools for policymakers and water resource managers during the coming years (e.g., Meehl et al., 2009).

0.2 Overview of Appendices

Presented here are four studies that characterize several crucial aspects of drought and megadrought variability in western North America. They represent a unified subset of more general research into the decadal-scale variability in western North America that has included: (1) analysis of tropical Pacific coral archives of sea surface temperature variability (Ault et al., 2009); (2) estimates of the importance of decadal to multidecadal precipitation variability (Ault and St George, 2010; St. George and Ault, 2011); and (3) diagnosis of the synoptic, interannual, and long-term controls on the timing of spring in western North America (Ault et al., 2011). To varying degrees, these four chapters examine the phenomena of megadrought in the context of a null hypothesis that assumes these events occur as a consequence of high-frequency fluctuations that have been integrated through time (e.g., Hasselmann, 1976)

In the first Appendix (A), we test whether or not the amplitude of low-frequency drought variance in observations, paleoclimate data, and climate model simula-

tions is significantly different from what would be expected from a weakly damped, stochastically-forced process (Hasselmann, 1976; Delworth and Manabe, 1988). We develop this null hypothesis using a Monte Carlo approach to generate realizations of an order-1 autoregressive (AR(1)) process, whose autocorrelation term operates at the monthly timescale to reflect local sources of "memory" (such as soil moisture, ecological processes and snow accumulation). We then employ spectral analysis to compare the variance of these Monte Carlo realizations with the variance of instrumental, paleoclimate, and global climate model (GCM) data. We use spectral analysis because it allows us to overcome the reality that the time evolution of hydroclimate fluctuations would have evolved differently in climate models than in nature, even if the two types of data exhibited a fundamentally similar range of low-frequency hydroclimatic variability.

The results in Appendix A indicate that a simple (AR(1)) null hypothesis cannot explain the full range of low-frequency behavior in instrumental data, lake and cave records of hydroclimate, and long, moisture-sensitive tree-ring chronologies. On the other hand, the spectral densities of hydroclimatic variability computed from various GCM experiments, as well as the densities computed from many tree-ring based reconstructions of hydroclimate, are consistent with the expectations developed from the null hypothesis. In other words, high-frequency variance, autocorrelation, and sufficient time could produce a range of megadrought variability that is consistent with the range that has been reconstructed from tree-rings and is simulated by long, unforced climate model simulations. We argue that this view is incomplete, however, because tree-ring data are typically "pre-processed" using methods that remove long-term non-climate (and climate) variability, and because the long control simulations lack the time-evolving boundary conditions that have operated during the last few millennia. When long tree-ring records, other proxies, and transient climate model simulations are considered, hydroclimate variance emerges along a continuum of drought that can be characterized by a "red" power spectrum. Yet there is still a discrepancy between the paleoclimate and climate model data on decadal (10yr) to multicentury timescales, with the former supporting substantially more

low-frequency variance than the latter. This discrepancy, in turn, has implications for our view of megadrought risk, much in the same way that long-term variability in stock markets influences our view of economic risk (Mandelbrot and Hudson, 2004), a point we explore further in Appendix B.

The second Appendix (B) asks "how likely are megadroughts in the coming century?" To answer this question, we broaden our view beyond North America to consider global megadrought risk as it is represented by 18 GCMs used in the IPCC 4th Assessment (Solomon et al., 2007). We summarize the relationship between timescale and variance using power-law coefficients (m) that relate frequency (f) to spectral density ($S(f)$) such that: $f^{-m} \propto S(f)$. Since higher values of this parameter are associated with greater megadrought risk, we refer to the parameter as the megadrought risk parameter, which we calculate from observations, paleoclimate data, and climate models. We also calculate this parameter from a new reconstruction of Colorado River flow (presented for the first time in this study), which preserves variance across the widest possible range of timescales.

We find that the values of the megadrought risk parameter (m) differ greatly between observational data (instrumental and paleoclimate) and GCMs. Observational data support positive values of m, while values of m calculated from GCMs are near zero or negative, particularly in the tropical pacific where interannual variability is driven by highly-energetic El Nino/La Nina variability (Guilyardi et al., 2009).

In Appendix B, we also develop a definition of "megadrought" that characterizes the well-known 1150's interval of aridity that occurred in reconstructed Colorado River streamflow (Meko et al., 2007). Using this definition, we map out the risk of such an event during the coming century, and find that in most of the arid regions the risk of a megadrought is at about 1-in-10, and higher than 1-in-2 in the most extreme cases. Since we argue that the GCMs underestimate the megadrought risk parameter (m), we conduct a second megadrought-risk analysis whereby the precipitation projections are rescaled to exhibit a value of m that is consistent with observational data. When we perform this analysis in the Southwest, the level of

risk of a megadrought in the next century rises from 1-in-10 to 1-in-4.

In Appendix C we turn our attention to the Sonoran Desert, a region already experiencing serious water resource challenges (e.g., MacDonald, 2010). Importantly, many of these water resources come from underground aquifers, and hence we have developed a new speleothem proxy for regional hydroclimate over the last 1500 years. To aid in our interpretation of this record, and to compare it with precisely dated tree-ring chronologies from the Southwest, we develop a Monte Carlo age modeling scheme. This scheme allows us to develop families of age models and explicitly quantify the effects of our assumptions on the distribution of these age model families. It also allows us to treat correlations with annually-precise tree-ring records in a probabilistic context. These modeling experiments, and experiments from an earlier study (Truebe et al., 2010), suggest that the most robust (and interpretable) timescales recorded by the record are multidecadal and longer. Findings also suggest that the cave record is winter sensitive, and that it has recorded several megadroughts over the past millennium and a half. It also depicts the interval from 1000AD to 1500AD as a period of enhanced variability.

In Appendix D we return to the null hypothesis for drought considered in Appendix A, but we modify it to account for spatial variability as well. That is, we are interested in determining if low-frequency drought variance in large-scale spatial patterns might arise from high-frequency fluctuations with spatial structure and autocorrelation. To develop this null hypothesis further, we begin by examining the time evolution of the worst droughts of the 20th century. We then identify robust patterns of large-scale drought variability that are present in a variety of drought indices, and across a wide range of timescales. We explore the seasonal characteristics of these patterns, and examine their association with hemispheric climate variability. All of these analyses show that atmospheric variability at the synoptic timescale can synchronize regional hydroclimate anomalies, and they further show that autocorrelation can cause these anomalies to persist.

To rigorously test the null hypothesis that autocorrelation and spatial correlation explain persistent, large-scale megadroughts, we run 1000 millennial-scale statisti-

cal simulations of North American climate. Simulations are run at the monthly timestep, and they preserve the spatial correlations of precipitation and temperature, but do not contain any climate sources of persistence outside of the annual cycle. We then calculate drought from the statistically-simulated climate fields, and use multi-variate spectral methods (Mann and Park, 1999) to calculate spatial patterns of variability in these climates across a wide range of frequencies. We also estimate large-scale patterns of variability at different timescales in reconstructed drought (Cook et al., 2004), drought in a long GCM control run (Collins et al., 2006), and a transient simulation (Kaufman et al., 2009).

Even with the more rigorous null hypothesis used in Appendix D, we find that the spatial patterns of drought in nature are more energetic than stochasticity, autocorrelation, and spatial correlation would predict. We also find that climate model simulations of drought are more energetic than the null would predict, and that both climate model simulations and reconstructions of drought are consistent with each other. As in Appendix A, however, we stress that this finding relies on what is likely a conservative estimate of low-frequency hydroclimate variance because the authors of the reconstructed drought dataset explicitly remove variance at long timescales (see supplementary material of Cook et al., 2004). Nonetheless, the similarity between global climate model drought patterns and paleoclimate patterns of drought is remarkable, and it implies that the mechanisms of drought may be similar across a continuum of timescales.

0.3 Implications

In general, low-frequency hydroclimate variability in North America is more energetic than high-frequency variability. On short timescales (interannual to decadal) and long timescales (multi-century to millennial), paleoclimate proxies and global climate model simulations provide a consistent view of hydroclimate continuum in western North America (see schematic drawing in Figure 1). On short timescales, redness arises from local autocorrelation, either because of soil-moisture storage or

other sources of "memory" at monthly (and shorter) time intervals. On the very long timescales, redness arises from the dynamic nature of the boundary conditions themselves. There is considerable disagreement, however, on decadal to centennial timescales between the "proxy view" of hydroclimate and the GCM view. Proxies indicate that the spectrum is red, but GCMs do not simulate this redness. Since redness of proxy records cannot be explained by interannual variability and autocorrelation alone, it implies that the background state of the whole hydroclimate system is always varying because of the influence of large-scale climate sources of variability. Since the GCMs do not simulate the observed redness on decadal to centennial timescales, it is possible that there are important climate phenomena that operate at these timescales that the GCMs are not simulating.

There are several known, large-scale limitations to the dynamics of GCMs used in the IPCC 4th assessment, all of which could be linked to North American hydroclimate. First, there has been an observed widening of the Hadley cell that GCMs do not simulate (Seidel et al., 2008); second, the winds are faster and winter precipitation is greater in GCMs than in nature (McAfee et al., 2011); and thirdly, the CMIP3 generation of GCMs simulate highly-energetic 2-4 year fluctuations in the tropical Pacific (Guilyardi et al., 2009), whereas observational records of tropical Pacific variability indicate that it too may exhibit red spectrum on long timescales (Urban et al., 2000; Cobb et al., 2003; Conroy et al., 2008; Ault et al., 2009). Of these three candidates, the third source has the most potential to explain the differences between the observed and modeled North American hydroclimate continuum because tropical Pacific variability is closely linked to western North American hydroclimate (Ropelewski and Halpert, 1987; Seager et al., 2005a; Herweijer and Seager, 2008; Findell and Delworth, 2010; Schubert et al., 2009).

If redness in western North American drought does indeed reflect redness in tropical Pacific sea surface temperature variability, this finding would call into question the existing paradigm of El Niño/Southern Oscillation (ENSO) variability being a fundamentally interannual phenomena (Neelin et al., 1998). Nonetheless, this view is not outside the scope of several existing, alternative paradigms. For in-

stance, Clement and Cane (1999) showed that ENSO-like variability emerges in a very long (150,000 year) control run of the Zebiak-Cane model Zebiak and Cane (1987), and Clement et al. (2011) has shown that red spectra arise in the thermodynamic mixed layer, "slab ocean," experiments of several IPCC models. On decadal timescales Meehl and Hu (2006) has shown that slow-moving Rossby waves modulate thermocline depth in the western Pacific, which triggers a variety of feedbacks in the region leading to sustained, low-frequency variability in a long coupled control run. In multi-millennial, fully coupled simulations with fixed (pre-industrial) boundary conditions, Wittenberg (2009) and Delworth (2010) have demonstrated that multicentury-scale variability arises in the variance of the interannual ENSO signal. These findings encompass a myriad of modeling paradigms that give rise to ultra-low ENSO-like variability. Isolating the physical mechanisms of variability across a wide range of timescale in the tropical Pacific, therefore, may afford the most promise for understanding the physical drivers of the continuum of drought variability in western North America.

Even if the ultimate origins of megadroughts are currently unknown and unpredictable, the results of these studies strongly support the view that megadrought risk in the coming century is substantial. Any long-term water resource management strategies would benefit from developing plans to cope with droughts that are more severe and prolonged than anything witnessed during the last century. At the same time, future paleoclimate and climate modeling studies should work to elucidate the physical processes that contribute variance to the continuum, which will ultimately help refine our view of future drought and megadrought risk.

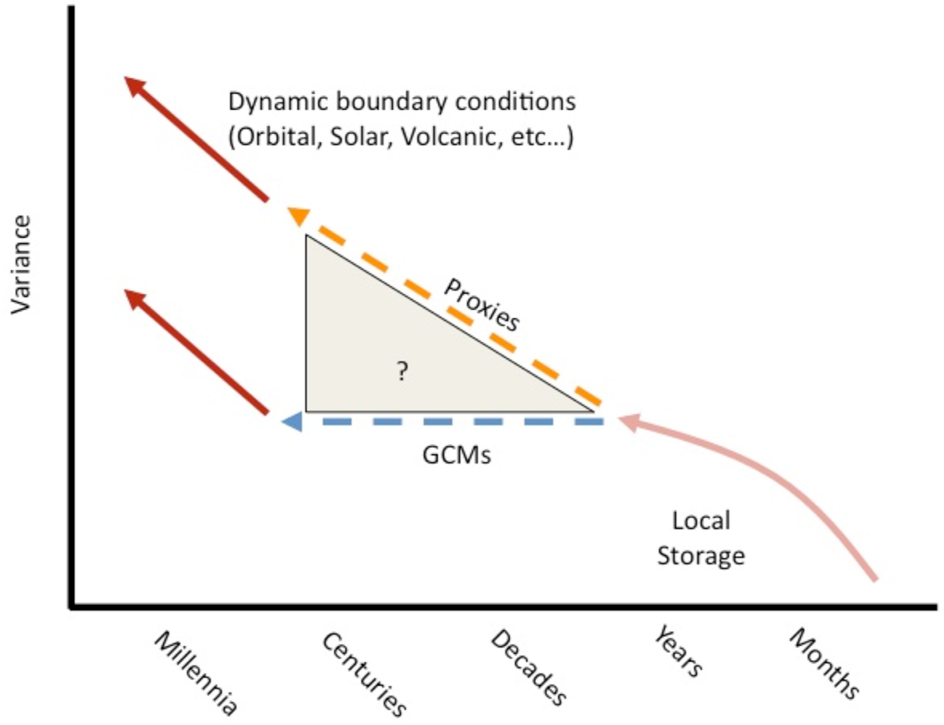

Figure 1: Schematic representation of the power spectrum of western North American hydro climate. At high frequencies, "redness" arises from local autocorrelation. At intermediate (decadal to centennial) frequencies, proxy drought records and global climate models disagree about the shape of the spectrum: proxies (orange line) indicate that variance increases at lower frequencies, climate models (blue line) exhibit a relatively flat spectrum. On long timescales (centennial and longer) the proxies and the models are back in agreement (red lines), with both providing evidence that dynamic boundary conditions contribute redness to the lowest resolvable frequencies. The gray area represents the region where disagreement between proxies and models leads to uncertainty in the "true" shape of the power spectrum.

APPENDIX A

AMERICAN DROUGHT: RED, WHITE OR BLUE?

T.R. Ault[1], J.E. Cole[1], G.T. Pederson[2], J.T, Overpeck[1,3], S. St. George[4], B. Otto-Bliesner[5], C. Deser[5], and C. Woodhouse[6]

1. Dept. of Geosciences, Univ. of Arizona, Tucson, AZ
2. Northern Rocky Mountain Science Center, USGS, Bozeman, MT
3. Institute of the Environment, Univ. of Arizona, Tucson, AZ
4. Department of Geography, University of Minnesota, MN
5. National Center for Atmospheric Research, Boulder, CO
6. Department of Geography, Univ. of Arizona, Tucson, AZ

A.1 Abstract

The distribution of natural climatic variance across the frequency spectrum has substantial importance for anticipating how future climate change will evolve. Here we analyze the spectrum of drought-related climate variance in a large array of instrumental, proxy, and climate model data for western North America (WNA). We estimate the linear relationship between log-transformed frequency and log-transformed variance (e.g., the power-law coefficient, β) in data and model output. We then use these values of β to test the significance of low-frequency variability against the null hypothesis that it arises from local sources of autocorrelation (e.g., soil moisture). We find evidence that in several types of hydroclimatic time series, a simple autoregressive (AR(1)) model cannot explain the full range of low-frequency variability. We show that observed values of β vary spatially and by timescale: at interannual (2 to 10 year) periods the precipitation spectrum resembles white noise, whereas the spectra of soil moisture and PDSI is much redder. At longer periods (greater than

10 year) values of β agree more closely among observed variables. Global climate models simulate realistic power-law scaling only when forced with realistic transient boundary conditions or observed global SSTs, but not when run with fixed radiative forcing conditions or 20th-century forcings. This finding suggests that the risk of severe droughts, which are part of the natural variability, may be underestimated by projections of future climate change.

A.2 Introduction

An important insight from many paleoclimate studies has been that intervals of aridity unlike anything seen in the 20^{th} occur naturally in western North America (Woodhouse and Overpeck, 1998; Stahle et al., 2007; Cook et al., 1999; Meko et al., 2007). Woodhouse and Overpeck (1998) showed that droughts like those of the 1950's and 1930's occurred once or twice per century and that far more severe and more persistent dry intervals occurred as recently as the 16^{th} century. Tree-ring based reconstructions of flow on the Colorado River over the last 1200 years (Meko et al., 2007) show that the mid-12th century experienced a 15% reduction in the long-term mean flow. In addition, Woodhouse, Gray, and Meko (2006) showed that the early part of the 20th century-the period upon which Colorado River allocations were made-was one of the wettest in the last 500 years. Reconstructed Palmer Drought Severity Index (PDSI) (Cook et al., 2004) and a collection of other proxy records (Graham et al., 2007) provides further evidence for a prolonged period of widespread drought between ca. 800 to 1200 years ago, as well as multiple decadal scale intervals matching or exceeding the range of 20^{th} century PDSI variability (Cook et al., 2004).

Megadroughts provide an important benchmark of natural variability for assessing climate model simulations. That is, if climate models can accurately assess the risk of future hydroclimate regimes, they should also reproduce a realistic range of natural drought variability when properly forced and integrated for many centuries. If they do not simulate a realistic natural range of drought variability (either when

forced with realistic boundary conditions or allowed to run for many centuries), then they may underestimate the risk of megadroughts in the future. Since many long control, 20th century, and paleoclimate simulations have been performed on the Community Climate System Model III (CCSM3) and Community System Model 1.4 (CSM1.4) (both based at the National Center for Atmospheric Research), they provide a test bed for this hypothesis.

Several factors complicate proxy-model drought comparisons. First, the time evolution of droughts in long model simulations will differ from the actual paleodrought history, even if both types of data exhibit similar underlying behavior. Second, mean model precipitation is greater than observed precipitation throughout much of western North America (WNA). This means that modeled "droughts" can be wetter than observed average conditions. Third, proxy records of hydroclimate have been filtered by the climate sensor itself (trees, lakes, peat bogs, etc...), which can distort the underlying climate signal. In contrast, there is no error in the measurement of model variables.

We attempt to minimize the limitations above by focusing on the power spectra of drought indices calculated from model output and observed drought indicators (e.g., instrumental records and indices, tree-ring reconstructions and chronologies, and paleoclimate records of hydroclimate variability). Specifically, we estimate the power-law scaling parameter (β) that relates hydroclimate variance to timescale (e.g., Huybers and Curry, 2006; Pelletier and Turcotte, 1997). This approach allows us to ignore the time evolution of drought variability and data-model differences in the mean and variance of drought. If the model simulates a realistic spectrum, then it should also simulate a realistic relationship between timescales of variability, and hence capture the relative risk and magnitude of megadroughts. Since a power-law distribution is not necessarily the *best* fit to the power spectra of all records considered here (e.g. Clauset et al., 2009), we refer to the values of β we estimate as "spectral slopes," which give us insight into the underlying relationship between time scale and variance across disparate types of data. We use values of β to compare different types of data and also to determine if the spectrum of hydroclimatic

variability in western North America is redder than would be expected by stochastic variability alone.

Power-laws relate variance to timescale such that $S(f) \propto f^{-\beta}$ where f is frequency and $S(f)$ is the power spectrum. Values of β near zero describe "white" spectra, which have an approximately uniform distribution of variance across frequencies and no significant autocorrelation. "Red" spectra, and "redness" in the power spectrum, are associated with time series that exhibit more variance at longer timescales. If time series are red and stationary, they are usually characterized by β values between 0.5 and 1.0 (e.g., Pelletier and Turcotte, 1997). Non-stationary processes are described by β values greater than 1 and, by definition, never stabilize around a long-term mean. Negative values of β, on the other hand, describe "blue" spectra. Such spectra can arise in systems with negative autocorrelation, because from one time step to the next the system is pushed away from its previous value.

Many studies have shown that power spectra of temperature fluctuations in the ocean and atmosphere are characterized by a power-law distribution, with β ranging from approximately zero to 0.5 on land, and near 1.0 for most of the ocean (Huybers and Curry, 2006; Fraedrich and Blender, 2003). Hydroclimate variables, which reflect the terrestrial P-E balance, have also been shown to exhibit similar scaling behavior, with β estimated near 0.5 for globally averaged spectra of river discharge, tree-ring widths, and precipitation records (Pelletier and Turcotte, 1997). Indeed, the earliest studies into the relationship between timescale and variance (Hurst, 1951) found the scaling coefficient H to be near 0.7 for Nile River discharge and for tree-ring widths in the western US (H is related to β such that $\beta = -(2 * H - 1)$, Heneghan and McDarby (2000)). These findings, and similar behavior in other complex systems, led Mandelbrot and Wallis (1968) to propose a family of statistical models of precipitation. Termed "fractional Gaussian noises" these time series exhibit more variance on long timescales, realistic scaling coefficients, and instability or non-stationarity of the long-term mean depending on the value of β specified. However, as noted by Klemes (1974) these models are limited by the fact that they do not represent the underlying physical processes responsible

for power-law scaling behavior. Pelletier and Turcotte (1997) therefore proposed that a simple 2-layer advection-diffusion model of global moisture transport in the atmosphere to explain the observed spectrum. While their model elucidates the underlying characteristics of ocean-atmosphere interactions that may cause power-law scaling, it does not provide insight into the regions that are most susceptible to this type of behavior, nor does it highlight the specific dynamical mechanisms that cause hydroclimate to vary on long timescales. More recently, Vyushin and Kushner (2009) and Vyushin et al. (2009) have used global climate model (GCM) and reanalysis data to link power-law behavior in the troposphere to tropical SST variability and similar behavior in the stratosphere to volcanic activity.

To some extent, low-frequency variability in many drought indicators must arise because of the way in which the phenomenon is defined, and how it is recorded. For example, one of the most widely used drought indicators, the Palmer Drought Severity Index (PDSI), has a built-in autocorrelation term of 0.897 at the monthly time step (Alley, 1984). Although this number is specified *a priori* in PDSI, it agrees well with autocorrelation in simulated soil moisture during the 20^{th} century (Figure A.1). Likewise, biological, geochemical, and physical archives of past droughts integrate hydroclimate variability through time: tree-ring widths during a given year may be influenced by previous years; lakes may store information from one year to the next. Importantly, autocorrelation tends to tends attenuate high-frequency fluctuations while enhancing low-frequency variance. Here we explore this tendency by testing the possibility that local autocorrelation related to surface processes accounts for the low-frequency energy in drought indicators.

Understanding hydroclimate scaling behavior in North America is important because it helps characterize the risk of persistent megadroughts. Without any scaling, drought indices could not persist below (or above) their long-term mean for any consequential amount of time. Clearly, this is not how 20th century droughts or paleodroughts tend to behave. Hence some frequency scaling is required to explain the tendency for arid anomalies to persist. The higher the value of β describing a particular time series, the longer it can deviate from its long-term mean. For

instance, Pelletier and Turcotte (1997) showed that the risk of a megadrought on the Colorado is five times higher during any given century if the river's underlying spectrum is best characterized by a power-law distribution rather than an AR(1) process.

A.3 Data & Methods

A.3.1 Data

We estimated the parameter β and its significance from a variety of sources (Table A.4):

1. Gridded precipitation and temperature from the University of East Anglia's TS2.1 datasets for WNA (20°N to 55°N; 125°W to 100°W). The TS2.1 product is derived from interpolating monthly station data to a 0.5° x 0.5° grid and is continuous from January 1901 through December 2002 (Mitchell and Jones, 2005).

2. Gridded Palmer Z-Index (a measure of P-E before autocorrelation has been introduced) and PDSI (Palmer, 1965; Alley, 1984) calculated from CRU precipitation and temperature in WNA.

3. Simulated soil moisture of the 20$^{\text{th}}$ Century using the Variable Infiltration Capacity (VIC) model (Liang et al., 1994; Andreadis et al., 2005). The forcing for this data was gridded (0.5° x 0.5° resolution) observations from the US HCN. We report β values here for the second layer in the VIC model, which represents depths of 0.2 to 2.4m. The spatial domain of data available in WNA from this product only covers US soil.

4. Gridded (2.5° x 2.5°) reconstructed PDSI from the North American Drought Atlas (Cook et al., 2004) spanning 1000AD to 2000AD. Again, only gridpoints in WNA were considered.

5. Twenty-five precipitation and 8 streamflow reconstructions from tree-ring studies at various sites across WNA (Tables A.1 and A.2). These data have been processed as described in original references; all data and references are available on the NOAA Paleoclimatology web page (http://www.ncdc.noaa.gov/paleo/recons.html). All available reconstructions of WNA hydroclimate were used.

6. Ten multicentury moisture-sensitive tree-ring chronologies from a recently developed 5-needle pine database (Kipfmueller and Salzer, 2010). Chronologies were selected if they (a) showed a significant ($p - value < .01$) relationship with local precipitation; (b) had at least 900 years of data from 1000AD to 2000AD; and (c) were at least 250m below treeline, which we required to ensure stable sensitivities through time (e.g., Salzer et al., 2009). We use long chronologies (with high sample depth) to minimize the effects of the "segment length curse," which removes variance from chronologies at long timescales (i.e., Cook et al., 1995). Finally, we normalized each chronology to exhibit unit variance over the last 200 years for comparison with modern records. This simple rescaling ensures that variance outside the range of the last 200 years is represented by spectral energies that are correspondingly greater, but it does not alter the shape of the power spectrum.

7. Two high-resolution lake records and two high-resolution cave records from WNA. Records were included if they were high-resolution (decadal), spanned at least 1000 years, and contained data from after 1800 A.3. As with the tree ring chronologies, we normalized each time series so that it would exhibit unit variance over the last 200 years.

8. Model data from several simulations using NCAR's CCSM3 (Collins et al., 2006). We used 8 simulations of the 20th century (Solomon et al., 2007), as well as 5 long control runs and a paleoclimate simulation (Table A.6). The 20th century simulations were fully coupled and forced with evolving greenhouse gasses (GHG) and aerosols from 1870 to 2000. The control simulations were

also fully coupled and run with forcings fixed at pre-industrial (four runs) or 1990 (one run) values. We analyzed climate model output from one very long (2300 year) transient "Mid-Holocene" CCSM3 simulation as well (details of the model are described in the supplementary material of Kaufman et al., 2009).

9. We also examined one long (1000 year) simulation using NCAR's Community Climate System Model version 1.4 (CSM1.4) (Ammann et al., 2007). This model was forced with evolving reconstructed solar variability (Lean, 2000) from 1000 AD to 2000AD and reconstructed volcanic forcing (Ammann et al., 2007).

10. Finally, we computed values of β for output from two types of sea surface temperature (SST) forcing experiments run on the Community Atmospheric Model III (CAM3). The first were "Tropical Ocean/Global Atmosphere" (TOGA) simulations, in which gridded instrumental SST values were used to force the atmosphere from $20°S$ to $20°N$, while SST values poleward of these latitudes were fixed at their climatological averages. In the "Global Ocean/Global Atmosphere" (GOGA) simulations, time-evolving SSTs were used throughout the world. Both the TOGA and GOGA simulations consist of 5-member ensembles and were performed with SST values from 1950-2000 with no additional forcing. We averaged variables in time from the 5-member ensembles to isolate the SST-forced component of variability.

We estimated power spectra from instrumental and simulated 20^{th} century annual drought indicators using the multi-taper method (Thomson, 1982). Since paleoclimate data (excluding tree-rings) tend to be unevenly sampled in time, we used the Lomb-Scargle method (Lomb, 1975; Scargle, 1982) to estimate their spectra to minimize frequency bias (Schulz and Stattegger, 1997). We normalized all paleoclimate data to exhibit unit variance prior to spectral analysis. As in Huybers and Curry (2006), we estimated the slope (β) of individual power spectra by performing a least-squares regression of log-transformed spectral density

against log-transformed frequency. To avoid biasing the regression towards the higher frequencies (Huybers and Curry, 2006), we averaged spectral density estimates to evenly spaced log-frequency bins prior to calculating β. We ensured that the values of β were not unduly influenced by long-term trends, which would artificially inflate variance at the lowest frequencies, by performing the regression over the frequency domain f_q to $2/N$, where f_q is the highest resolvable frequency (the Nyquist Frequency) and N is the length of the record in years. This step effectively sets the upper limit of our limit of our analyses to half the dataset length (e.g., to 50yr periods for instrumental data).

A.3.2 A Null Hypothesis for Drought

Since the power spectrum of white noise would be characterized by a spectral slope close to zero, we test the values of β against the null hypothesis that they are not significantly different from zero. To do so, however, requires the null hypothesis to be applied differently depending on the variable being examined and the frequency domain being considered. That is, atmosphere-only variables (e.g., precipitation and the Palmer Z-index, a metric of P-E) do not have a built-in autocorrelation term, hence we evaluate the significance of β calculated from these variables as the $p - value$ of the linear regression of log-transformed spectral density ($S(f)$) against log-transformed frequency (f).

In contrast to the atmosphere-only variables, drought indicators such as PDSI, soil moisture, tree-ring records, and other paleoclimate records are subject to autocorrelation that arises locally and does not reflect low-frequency behavior in the climate system. Here, we develop a null hypothesis for the power spectrum of these drought indicators by using the autocorrelation parameter that is built in to the PDSI model (Alley, 1984). Specifically, we are interested in determining if the PDSI's monthly autocorrelation term is sufficient to explain the power spectrum of simulated soil moisture, instrumental and reconstructed PDSI, reconstructed streamflow and precipitation, normalized paleoclimate records of drought, and PDSI calculated from climate models. We begin by examining the expected power spec-

trum of an autoregressive process with lag-1 autocorrelation ρ (e.g., Bartlett, 1978):

$$S(f) = S_o \frac{1-\rho^2}{1-2\rho\cos\pi\frac{f}{f_N}+\rho^2} \tag{A.1}$$

Where f is frequency, f_N is the Nyquist frequency (in this case frequencies corresponding to 2 month wavelengths), and S_o is mean spectral density, which is related to the white noise variance (σ^2) by:

$$S_o = \frac{\sigma^2}{1-\rho^2}. \tag{A.2}$$

Figure A.3a shows an example of this analytical expectation. In the case where the input of the white noise variance (σ^2) is unknown, but the time series have been normalized to unit variance *a posteriori*, the value of S_o will be equal to unity (Figure A.3b). However, here we will be considering time series that have been annualized, but that are assumed to exhibit monthly autocorrelation close to 0.897 (i.e., the built-in autocorrelation in the PDSI model). In this case, the spectral density on interannual timescales will be given by:

$$S(f_{yr}) = \left(\frac{S_o}{< S(f) >}\right) \frac{1-\rho^2}{1-2\rho\cos\pi\frac{f_{yr}}{f_N}+\rho^2}, \tag{A.3}$$

Where only frequencies corresponding to wavelengths of 2 years or longer ($0 < f_{yr} \leq 0.5$) are considered, S_o is given by equation A.2, and $< S(f) >$ is the spectral average of $S(f)$ in equation A.1. This spectrum is shown in Figure A.3. Note that autocorrelation, even at the monthly timescale, reddens the power spectrum out to decadal (10 year) frequencies. We therefore consider the slope of the power spectrum on interannual timescales (β_I) separately from decadal-and-longer timescales (β_D).

We use a Monte Carlo procedure to establish confidence limits for the power spectra of drought indicators and for the overall value of β, the interannual spectral slope β_I and the decadal spectral slope β_D. First, we generated 1000 realizations of a random, autoregressive (AR(1)) process with monthly autocorrelation set to 0.897. Each realization was run for 12000 months, averaged to produce synthetic annual time series, and normalized to exhibit unit variance. We then computed the power

spectrum of each of these realizations to generate upper and lower 95% confidence limits of the variance in the power spectrum (Figure A.4a) as well as the spectral slope (β) on interannual (Figure A.4b) and decadal (Figure A.4b) timescales.

A.4 Results

Results from our Monte Carlo analysis give us upper and lower confidence limits for the values of β calculated from an annually averaged, unit variance time series with monthly autocorrelation equal to 0.897. Overall, the upper and lower 90% confidence limits of β are 0.11 and 0.42, respectively. On interannual timescales, β_I these limits are 1.06 and 1.56, and on decadal β_D timescales the limits are -0.27 and 0.31 (i.e., close to zero). We deem values outside of these limits significant. These expectations are consistent with the shape of the analytical spectra in A.3 and A.4, which show how autocorrelation reddens the spectrum at 1/2 year to 1/10 year frequencies.

Figure A.5 shows frequency-averaged spectral densities for western North America from Precipitation, Z-index, PDSI and soil moisture. As in Pelletier and Turcotte (1997), we find that the spectrum of precipitation does not exhibit much redness ($\beta_I = 0.03$) on interannual timescales. We also observe that the spectra of Z-index and precipitation are virtually identical. PDSI, in contrast, exhibits a much steeper (albeit non-significant) slope at interannual timescales ($\beta_I = 0.74$), as would be expected from the analytical AR(1) red-noise spectrum (Figure A.5c, dotted line). The change in slope is not as dramatic for soil moisture ($\beta_I = 0.27$, $\beta_D = 0.2$). On longer timescales, the slope of precipitation, Z-index, PDSI, and soil moisture ($\beta_D = 0.27, 0.26, 0.33$, and 0.2, respectively) are more similar. The higher value of the PDSI slope (0.33) reflects the slight attenuation of high frequencies from autocorrelation.

Figure A.6 shows the value of β mapped out across WNA for precipitation, Z-index, PDSI, and soil moisture. Values of β calculated from precipitation and Z-index are significantly different from zero ($p - value < 0.1$) throughout much of

the region (Figure A.6, top row, first two columns). Values of β calculated from PDSI and Z-index are also significantly different from the range of values expected from our AR(1) null hypothesis (Figure A.6, top row, third and fourth columns). At interannual timescales, significant values of β_I are restricted to small regions of the Pacific Northwest and northern Mexico in precipitation and Z-index (Figure A.6, second row, first two columns). The values of β_I calculated from PDSI and soil moisture are not significant throughout most of WNA. In contrast, at long timescales there are large regions where the value of β_D is significantly different from zero in precipitation and Z-index and significantly different from the AR(1) null hypothesis in PDSI and Z-index (Figure A.6, third row). Importantly, the maps for precipitation, Z-index and PDSI are nearly identical. Positive values ($\beta_D > 0.5$) occur in the Southwest, the Rocky Mountains, and the northern US Great Plains. The pattern of β_D is similar for soil moisture (Figure A.6).

The frequency-averaged power spectrum for tree-ring reconstructed PDSI is shown in Figure A.7a. The spectral slope of the average reconstructed WNA PDSI power spectrum is not significantly different (at the 90% confidence limit) from the AR(1) expectation at interannual ($\beta = 0.70$) or decadal ($\beta_D = -0.08$) timescales. Overall the value is close to zero ($\beta = 0.09$), which is also non-significant (at the 90% confidence level). Also shown in Figure A.7 are maps of β, β_I and β_D calculated from WNA. They indicate that the slope is significant at interannual timescales, but not overall and not at decadal timescales (Figure A.7b-d).

Figure A.8 shows spectral averages from reconstructed, normalized precipitation and streamflow, as well as other paleodrought indicators from across WNA. At high frequencies, average spectra from normalized reconstructions and from the long chronologies are more energetic than the AR(1) expectation. At lower frequencies, the reconstructions are well within the AR(1) expectation, but the mean spectra from the longest 5-needle pine chronologies are not. Spectral variance in other paleodrought records (Figure A.8, symbols) is within the AR(1) expectation on decadal to 100 year timescales. At longer timescales, however, there is more variance in these records than the null hypothesis alone can account explain. Interestingly,

the variances and slopes of the paleodrought records are in good agreement with the variance and slope records of the 5-needle pine chronologies.

We now turn our attention to three types of coupled modeling experiments: long control runs, 20$^{\text{th}}$ century simulations used in the IPCC's 4$^{\text{th}}$ Assessment, and paleoclimate simulations of the late Holocene. First, we examine the mean power spectra of PDSI in the unforced, coupled model simulations (Figure A.9a). The shape of these spectra closely resembles the analytical AR(1) expectation, and their spectral slopes are not significant overall, or at interannual or decadal timescales (Table A.6). In the "20th century" simulations, the shape of the spectra is again consistent with the analytical AR(1) expectation, and most of the overall, decadal and interannual slopes A.6. Two exceptions occur (β from 20th Century "run-a" as well as β and β_D from 20th Century "run-f"), but further inspection of these spectra showed the significant values were spurious and non-significant at a slightly higher (95%) confidence limit. Finally, we examine PDSI spectra calculated from long, transient, paleoclimate simulations of CCSM3 and CSM1.4. Both of these spectra exhibit redness at interannual timescales that is quite similar to the AR(1) analytical spectrum, and neither one produces significant values of β, β_I or β_D. On long timescales, the spectrum of PDSI calculated from the CSM1.4 simulation is quite similar to the AR(1) spectrum. The mid-Holocene transient simulation run on CCSM3 also exhibits some redness in PDSI on multi-centennial (and longer) timescales that is clearly not explained by the AR(1) spectrum. Although the decadal spectral slope (β_D) is not significant, slopes calculated at even longer (e.g., centennial to millennial timescales) are significant.

A.5 Discussion

A.5.1 Megadroughts: a consequence of time and memory?

Our results suggest that β varies with hydro-climate indicator, geography, and frequency domain. These results are consistent with the idea that scaling arises from a variety of sources (Klemes, 1974) and is not uniform through space

(Kantelhardt et al., 2006). For example, precipitation exhibits almost no persistence on interannual timescales, yet the autocorrelation of the PDSI reddens the spectrum considerably at these frequencies in observations (Figure A.5) and in models. While the built-in PDSI autocorrelation coefficient was derived from local observations of soil moisture at a small spatial scale (Palmer, 1965; Alley, 1984), it is within the range of simulated soil moisture autocorrelation in the west (Figure A.1). If this estimate is indeed realistic for WNA, then we cannot rule out the possibility that interannual redness in PDSI, soil moisture, and tree-ring records can arise solely from local persistence, and does not exclusively reflect remote sources of climate variability.

Analysis of reconstructed PDSI, precipitation, and streamflow reveals that most, if not all, of the low-frequency variance of these records can be explained by monthly autocorrelation related to local storage. This result is clear in Table B.1, Figure A.7 and Figure A.8: values of β and β_D are not significant and the variance in these spectra is within the AR(1) expectation. This finding implies that we cannot rule out the null hypothesis as a possible explanation for the range of low-frequency variance in these tree-ring based hydroclimate reconstructions. In other words, stochastic variability in the atmosphere, realistic autocorrelation, and sufficient time may produce a range of drought variability that is consistent with tree ring reconstructions of North American hydroclimate during the last millennium. If this is the case, then long control runs and forced simulations of the 20th Century century exhibit low-frequency variability that is in fundamentally good agreement with tree-ring based reconstructions and the monthly AR(1) null hypothesis. We stress that this finding refers to the underlying continuum of hydroclimate variability and not to specific narrow-band features that may arise in nature and in coupled models as a consequence of deterministic sources of variability (e.g., Meehl and Hu, 2006). We also stress that there are a number of practical and statistical limitations to the inferences about low-frequency variability that one can make from tree-ring reconstructions Cook et al. (1995); Moberg et al. (2005). Most notably, the "segment-length curse" and removal of non-climate, low-frequency variance would be expected to

diminish variance at the longest timescales.

Despite the good agreement between the AR(1) spectrum and spectra calculated from reconstructed PDSI, precipitation, and streamflow, there is evidence that more low-frequency variance exists in WNA hydroclimate than the AR(1) expectation alone can explain. This evidence comes from the spectrum of the longest, moisture-sensitive chronologies in the 5-needle pine database, and from the spectra of several high-resolution paleodrought indicators (Table B.1, Figure A.8, and supp. mat.). Moreover, the null hypothesis cannot account for low frequency scaling of observed PDSI and soil moisture at timescales longer than 10 years. Since the spectrum at these frequencies is relatively free of local autocorrelation effects, the spectral slope reflects the behavior of the underlying climate system. Finally, the null hypothesis cannot account for the low-frequency scaling of precipitation because precipitation fluctuations are not autocorrelated at timescales longer than a few days (Kantelhardt et al., 2006). Trends in precipitation may influence the apparent spectral slope, but we removed frequencies lower than $1/(2N)$ from our calculation of β_D to avoid this bias. We also computed decadal spectral slopes (β_D) for precipitation from different 20th century time windows, detrended time series, Historical Climate Network station data, and wet-day frequency. Each of these analyses supported a mean value of β_D that was positive and significant for western North American precipitation. Since observations and proxies all highlight the importance low-frequency redness, we argue that this behavior is a robust feature of precipitation in western North America. That is, we reject the AR(1) null hypothesis as an explanation for low-frequency hydroclimate variance on decadal-and-longer timescales.

A.5.2 Sources of redness in WNA hydroclimate

The spectra of PDSI variability calculated from CCSM3's long control runs and 20th century simulations are consistent with the AR(1) expectation, but they do not capture the low-frequency variance inferred from instrumental and proxy records at longer timescales. However, when transient boundary conditions are used to

force the model, substantially more variance appears at the low frequency end of the spectrum than the AR(1) expectation can explain. Since the proxy records analyzed here record variance on these ultra-low frequencies, it is likely that the spectral slope on centennial to millennial timescales reflects a forced response to dynamical boundary conditions.

At intermediate (decadal to centennial) frequencies, however, there is generally more redness in nature than there is in the long Mid-Holocene simulation (i.e., there is a "flat" region of the spectrum in Figure A.10). The 20th Century and control simulations are also not in agreement with observations at these frequencies. One possible explanation for this discrepancy is that the model oceans exhibit inadequate low-frequency variability, and hence fail to produce realistic long-term droughts on land. Alternatively, internal variability in the GCM atmosphere may distort low-frequency information from the oceans before it reaches land. To discern between these two possibilities, we examine global maps of β calculated from precipitation in two idealized SST-forcing experiments (TOGA and GOGA). We focus here on precipitation to isolate the atmospheric component, which allows us to ignore the effects of local autocorrelation.

Global SST-forced precipitation patterns of β are shown for the GOGA and TOGA experiments in Figure A.11. In the TOGA simulations (Figure A.11a), precipitation fluctuations in the tropical Pacific and along the west coast of North America exhibit sufficient variance at high frequencies to produce blue ($\beta < 0$) spectra. However, when forced with the entire global ocean (Figure A.12b), the negative values of β disappear from western North America and from most of the tropical Pacific. The results in Figure A.11 help elucidate the near-zero values of β calculated from control runs and 20th century simulations: the ENSO region of CCSM3 is highly energetic on a biennial time scale (Collins et al., 2006; Guilyardi et al., 2009) and is strongly teleconnected to western North American hydroclimate (e.g., Ropelewski and Halpert, 1987; Seager et al., 2005a,b; Herweijer and Seager, 2008). These slopes are estimated only through periods of 25yr, but suggest that the tropical Pacific is a source of interannual variability strong enough to minimize apparent

power-law behavior over these time scales in teleconnected regions.

This interpretation is further supported by Figure A.12, which shows the value of β calculated from observed (Kaplan et al., 1998) and simulated (using pre-industrial boundary conditions) SST. Model values of β are generally lower than those calculated from instrumental data (zonal averages in panels (b) and (d) of Figure A.12, and this difference is most dramatic in the tropical Pacific. In fact, in the western tropical Pacific ($\backsim$ 180°W) instrumental values of β are positive ($\beta = 0.5$) while CCSM3 values are negative ($\beta = -0.5$). Yet this discrepancy cannot be the sole cause of the spectral differences in modeled versus observed precipitation because when CAM3 was forced with only tropical Pacific observations in the TOGA experiments, the spectra along the western coast of North America remained blue (Figure A.11. It is therefore possible that the differences in spectral slopes arise because the atmospheric model itself responds too weakly to mid-latitude, low-frequency variability in the oceans, which has been implicated as a possible source of low-frequency variability in North American hydroclimate (e.g., McCabe et al., 2004; Kushnir et al., 2010; Schubert et al., 2009). It is also possible that low-frequency variability in the tropical Pacific, which is not well resolved in instrumental records (Cole et al., 1993; Urban et al., 2000; Cobb et al., 2003; Conroy et al., 2008; Ault et al., 2009), is responsible for the redness in WNA hydroclimate at decadal timescales and longer. Finally, modeled slopes may be incompletely estimated because our analysis of these 50 year simulations does not extend beyond periods of 25 years.

A.5.3 Implications for Drought Risk

The differences between modeled hydroclimate and observational hydroclimate spectra are not very noticeable when considering only interannual to decadal timescales: the redness imposed by the PDSI in the model mimics redness in the observed system. However, at longer timescales, the discrepancies between modeled hydroclimate variability and proxy-inferred hydroclimate become much more important. To illustrate this point, we generated two different sets of 1000 year long Monte

Carlo "red-noise" time series with variance equal to the Southwest PDSI average. The first set was generated by calculating AR(1) time series with monthly autocorrelation of 0.897, then analyzing them. The second set was generated from 1000 year realizations of random, annually resolved white-noise time series that were then filtered to exhibit spectral slopes that were at the lower end of the instrumental and proxy range ($\beta = 0.4$ to 0.6). We generated these red noise realizations using the same technique employed by Pelletier and Turcotte (1997) and described more thoroughly in Pelletier (2008). We examine drought regimes across different timescales in these two sets of red noise time series by tallying the number of years in each realization when 3/5 (60%) of the antecedent years were below a given threshold. We then calculated the percentage of time each MC realization spent at a given duration-magnitude combination, and averaged these percentages across all realizations. The resulting average distributions are contoured in Figure A.13.

At interannual to decadal timescales, the differences in drought-risk distributions between AR(1) time series and those scaled by realistic spectral slopes are not very noticeable. However, at longer timescales, the two types of red noise data present very different views of drought-risk. For instance, according to the AR(1) average distribution, there is no risk of a 30 year or a 50 drought regime when 60% of the years experience PDSI values at or below -1. In contrast, the MC realizations that have been rescaled by a realistic spectral slope tend to spend about 15% of the time in such a regime at the 30 year timescale, and they spend about 10% of their time in this regime at the 50 year timescale. We can also interpret these findings in terms of other hydroclimate variables, such as streamflow on the Colorado River, by simply rescaling the Monte-Carlo realizations to exhibit the same variance as the variable of interest. The -1 threshold (1 standard deviation) would correspond to 4.37 MAF, which, for reference, is roughly the state of California's entire allocation. Clearly, any decadal drought regimes at this magnitude would pose unprecedented challenges to water resource managers. Yet, our results suggest that this type of behavior is realistic for the region.

Since forcing experiments with dynamical boundary conditions and long control

runs do not necessarily simulate positive spectral slopes at low-frequencies, the risk of events that occur on long timescales is difficult to assess. However, our results indicate that their spectra tend to resemble the AR(1) spectrum, which according to Figure A.13, likely underestimate the risk of low-frequency megadroughts.

A.6 Conclusions

Our analysis uses a wide range of observed and simulated datasets to examine the distribution of variance in the hydroclimatology of western North America. By comparing spectral slopes against a clearly defined null hypothesis of variance distribution, we are able to identify where hydroclimatic variance appears to follow a power-law distribution, in which variability at low frequencies is significantly stronger than that at high. One of our most compelling findings is that nearly all observational datasets indicate substantial power-law behavior at decadal and longer periods, with the exception of tree-ring reconstructions (in which such variance has been minimized by preprocessing). Such behavior is only rarely apparent in any of the climate simulations we analyze here (nor have we observed it in other coupled models; Ault et al. 2011 and unpublished analyses). If such a discrepancy continues in the latest, AR5 generation of coupled models, it may lead to underestimation of future megadrought risk in the next IPCC AR5 analysis.

Our analysis supports the idea that variability in western North America hydroclimate scales with a power-law distribution and a spectral slope between about 02-0.8. We have examined whether paleoclimate data, which define the larger end of this range, might overestimate this slope due to internal, proxy specific processes. Where such processes add autocorrelation (e.g. groundwater storage terms for cave records Truebe et al., 2010), they would simply create an AR(1) spectrum that should flatten at multidecadal periods; this is not what we observe here. The low-frequency slopes defined by instrumental data are also likely conservative, as our detrending process removes variance with a period longer than half the record (for instrumental data, roughly 50 years).

Several different processes likely explain the significant, positive spectral slopes of North American hydroclimate. These processes vary depending on frequency interval and geography. At interannual timescales, precipitation and the Palmer Z-index exhibit spectral slopes that are near zero ($\beta = 0.03$) and non-significant, whereas soil moisture and PDSI are considerably redder ($\beta = 0.52$ and $\beta = 0.42$, respectively). At decadal and lower frequencies, the spectrum of precipitation converges with the PDSI and soil moisture spectrum, implying that at low frequencies, terrestrial hydroclimate variability and precipitation input are fundamentally similar. Moisture sensitive tree-ring chronologies, lake and cave records all support significant, positive values of β. The TOGA and GOGA simulations we describe point to the tropical Pacific as a source of exceptionally strong interannual variance in the model. Coupled simulations indicate that transient forcings can impart substantial low-frequency variance (presumably as in the real world).

Future modeling experiments may help clarify certain aspects of our analysis. First, we were limited to SST forcing experiments run for only 50 years, which gives us only a modest understanding of the low frequency response of precipitation scaling to SST forcing. Longer, continuous SST forcing experiments may help determine if the scaling we observed in the GOGA experiments extends to lower frequencies still. Extending model simulations using coral-based reconstructions of SST (e.g., Cane, 2005) may help determine if tropical variability imparts low frequency scaling when sufficiently long windows of time are considered. Analysis of AR5 coupled models is planned to assess whether model improvements, particularly in the tropical Pacific, have yielded more realistic representations of low-frequency hydroclimatic variability. Meanwhile, assessments of future drought risk may need to incorporate paleoclimatic insights if they are to encompass the full range of variability apparent in hydroclimatic observations.

A.7 Tables and Figures

Table A.1: Estimates of the parameter β from a variety of precipitation reconstructions in WNA. Bold numbers are significant against the AR(1) null hypothesis at th 90% confidence limits.

Region	Reference	$N/2$	β	β_I	β_D
Durango, CO	Cleaveland et al. (2003)	304	0.15	0.08	0.28
Chihuahua, Mexico	Diaz et al. (2002)	173	-0.15	0.23	-0.19
D7, OR	Garfin and Hughes (1996)	138	0.25	0.27	**0.45**
D8, OR	Garfin and Hughes (1996)	138	0.28	0.47	**0.4**
Bighorn, WY	Gray et al. (2004a)	370	-0.32	0.08	-0.8
Uinta, UT	Gray et al. (2004b)	388	0.42	0.83	0.21
Malpais, NM	Grissino-Mayer (1995)	1065	**0.53**	0.1	**0.55**
D6, NV	Hughes and Gramlich (1996)	1498	0.12	0.05	0.22
D1, AZ	Ni et al. (2002)	495	0.35	0.23	**0.37**
D2, AZ	Ni et al. (2002)	495	0.21	0.69	0
D3, AZ	Ni et al. (2002)	495	0.42	0.57	0.3
D4, AZ	Ni et al. (2002)	495	**0.47**	0.79	**0.31**
D5, AZ	Ni et al. (2002)	495	0.29	0.46	0.16
D6, AZ	Ni et al. (2002)	495	0.3	0.45	0.21
D7, AZ	Ni et al. (2002)	495	**0.6**	0.53	**0.52**
D1, NM	Ni et al. (2002)	495	0.32	0.59	0.11
D2, NM	Ni et al. (2002)	495	**0.52**	0.62	**0.48**
D3, NM	Ni et al. (2002)	495	**0.49**	0.4	**0.39**
D4, NM	Ni et al. (2002)	495	**0.53**	0.59	**0.48**
D5, NM	Ni et al. (2002)	495	**0.62**	0.49	**0.48**
D6, NM	Ni et al. (2002)	495	**0.56**	0.42	**0.41**
D7, NM	Ni et al. (2002)	495	**0.53**	0.41	**0.41**
D8,NM	Ni et al. (2002)	495	**0.54**	0.48	**0.46**
Saltillo, Mexico	Pohl et al. (2003)	110	0.07	0.13	**0.47**
Colorado Plateau	Salzer and Kipfmueller (2005)	709	-0.06	0.55	-0.51

Table A.2: Estimates of β, β_I and β_D from several reconstructed streamflow in several basins in WNA.

Region	Reference	$N/2$	β	β_I	β_D
Yellowstone R., MT	Graumlich et al. (2003)	147	-0.76	0.36	-1.76
Clear Creek, CO	Woodhouse (2000)	152	-0.15	0.23	-0.19
Salinas, CA	Griffin (2005)	298	0.1	0.27	0.05
Boulder Creek, CO	Woodhouse (2001)	143	0.28	0.47	**0.4**
Feather R., CA	Meko (2001)	539	-0.55	0.08	-1.09
Upper Colorado R.	Meko et al. (2007)	622	0.37	0.83	0.16
Lees Ferry, AZ	Woodhouse et al. (2006)	255	**0.52**	0.1	**0.56**
Sacramento R., CA	Meko et al. (2001)	555	0.05	0.05	0.11

Table A.3: Estimates of β_D from four high-resolution, moisture-sensitive paleoclimate records from WNA. Since the shortest resolvable frequencies in these records are decadal, β_D and β would be the same. For clarity, we report the values in terms of β_D

Variable	Reference	β_D
Moon Lake Salinity	Laird et al. (1996)	**0.84**
Pyramid Lake $\delta^{18}O$	Benson et al. (2002)	**1.16**
Carlsbad Cave Band Thickness	Polyak and Asmerom (2001)	**0.63**
Hidden Cave Band Thickness	Rasmussen et al. (2006)	**0.82**

Table A.4: Summary of spectral slopes calculated from mean spectra for western North America (20°N to 55°N; 125°W to 100°W). Slopes were calculated as the linear least squares regression between frequency and spectral density. Estimates of β were made from 2 year to 50 year timescales (i.e., the shortest resolvable frequencies to frequencies corresponding to one half of the lowest resolvable wavelength), as well as from 2 to 10 year timescales (β_I) and 10 to 50 year timescales (β_D). In the case of precipitation and the Z-index, significance of the slope was determined by the p-value of the regression line. Significance of the slope of PDSI and soil moisture was determined using a Monte Carlo method. Bold indicates values that were significant at the 90% confidence limit or higher.

Variable	Source	β	β_I	β_D
Precipitation	CRU TS2.1 (Mitchell and Jones, 2005)	**0.2**	0.03	**0.27**
Z-Index	CRU TS2.1 (Mitchell and Jones, 2005)	**0.19**	0.03	**0.22**
PDSI	CRU TS2.1 (Mitchell and Jones, 2005)	**0.52**	0.74	**0.33**
Soil Moist.	VIC (Andreadis et al., 2005)	0.42	0.27	0.2

Table A.5: Same as Table A.4, but for paleoclimate records. Since the highest resolvable frequencies from the lake and cave records used here correspond to 10 year wavelengths, estimates of β_D and β would be the same. For clarity, we report these values in terms of β_D.

Variable	Reference	β	β_I	β_D
N. American Drought Atlas	Cook et al. (2004)	0.09	0.7	-0.08
Precip. Reconstructions	See Table A.1	0.1	0.24	0.01
Streamflow Reconstructions	See Table A.2	-0.02	0.19	-0.14
5-Needle Pine Chronologies	Kipfmueller and Salzer (2010)	**0.59**	0.3	**0.57**
WNA Lakes and Caves	See See Table A.3	–	–	**0.86**

Table A.6: Same as Tables A.4 and B.1, but for PDSI calculated from climate model data. Values that are significant at the 90% confidence limit are shown in bold. The column marked N_yrs indicates the number of years included in the run.

Model	Resolution	Experiment	N_{yrs}	β	β_I	β_D
CCSM3	T42	1k pre-industrial control	1000	0.2	0.85	0.09
CCSM3	T31	Pre-industrial control	100	0.15	0.77	-0.24
CCSM3	T42	Pre-industrial control	200	0.29	0.94	0.06
CCSM3	T85	Pre-industrial control	200	0.4	0.82	0.24
CCSM3	T85	1990 control	600	0.19	0.84	0.06
CCSM3	T85	20th Century (run-a)	130	**0.44**	0.82	0.11
CCSM3	T85	20th Century (run-b)	130	0.34	0.72	-0.12
CCSM3	T85	20th Century (run-c)	130	0.29	0.7	-0.26
CCSM3	T85	20th Century (run-d)	130	0.37	0.7	-0.08
CCSM3	T85	20th Century (run-e)	130	0.22	1.08	-0.3
CCSM3	T85	20th Century (run-f)	130	**0.43**	0.66	**0.37**
CCSM3	T85	20th Century (run-g)	130	0.39	0.72	0.21
CCSM3	T85	20th Century (run-h)	130	0.22	1.09	-0.44
CCSM3	T31	Mid-Holocene Transient	2300	0.36	1.02	0.3
CSM1.4	T31	Last Millennium	1150	0.17	0.76	0.03

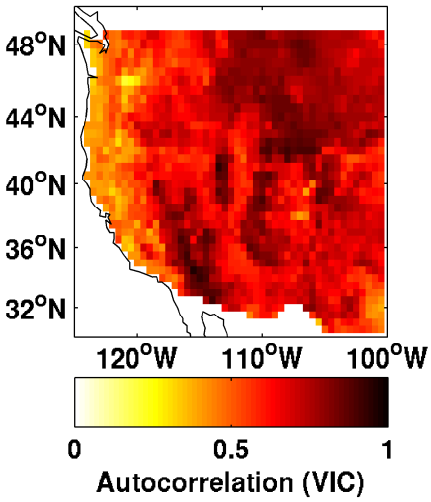

Figure A.1: Monthly autocorrelation parameter estimated from a VIC soil moisture model simulation of the 20th century using instrumental data (Andreadis et al., 2005).

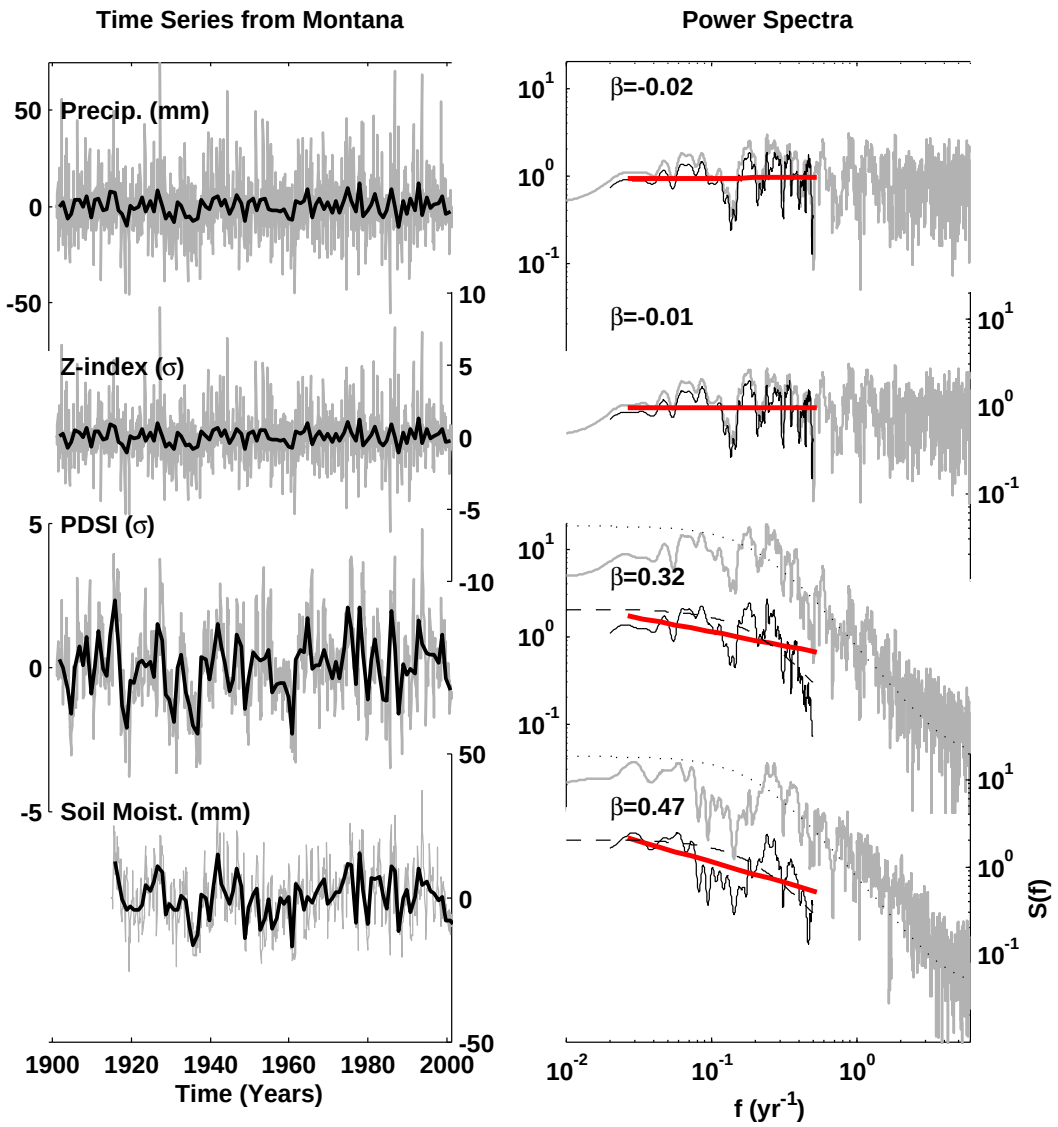

Figure A.2: Time series (left) and spectra (right) for hydro-climate indicators averaged for "Montana" (114°W to 105°W; 45°N to 50°N) as an example of how β is calculated from annually averaged precipitation, Z-index, PDSI and Soil Moisture. Grey lines show the original monthly time series and spectra, black lines show the time series and spectra of the annual average. Time series were normalized to unit standard deviation before computing the spectrum. Annual spectra are offset from monthly spectra because they exhibit less variance. For reference, the analytical AR(1) spectrum is shown on the bottom two panels on the right-hand side (thin, dotted line). Red lines on the right panel show the overall spectral slope (β) calculated from 1/2 year to 1/50 year frequencies.

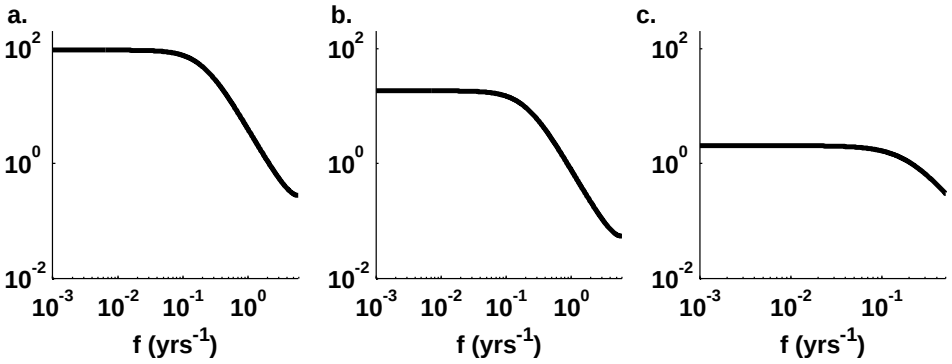

Figure A.3: Examples of the analytical AR(1) spectrum to illustrate the effects of different normalizations. (**a**) An AR(1) process with unit white-noise variance and autocorrelation of 0.897 at the monthly time step (Equation A.1). (**b**) An AR(1) process that has been normalized to exhibit unit variance overall (Equation A.2. **(c)** An AR(1) process with monthly autocorrelation of 0.897 that has been averaged to annual resolution and normalized to unit variance (Equation A.3.

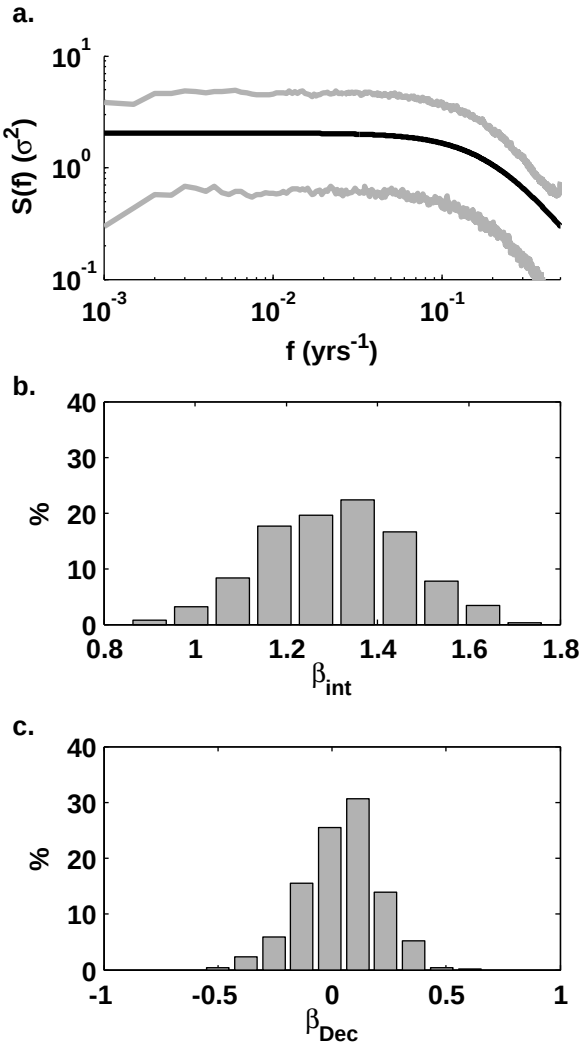

Figure A.4: Results of Monte Carlo simulations of an AR(1) process to establish confidence limits for the spectral slope overall (β) and at interannual (β_I) and decadal (β_D) timescales. (**a**) The upper and lower 90% confidence limits from all Monte-Carlo spectra (gray) shown with the analytical expectation (black line). (**b**) The distribution of β_I values at interannual timescales (2-10 years). (**c**) The distribution of β_D on > 10 year timescales.

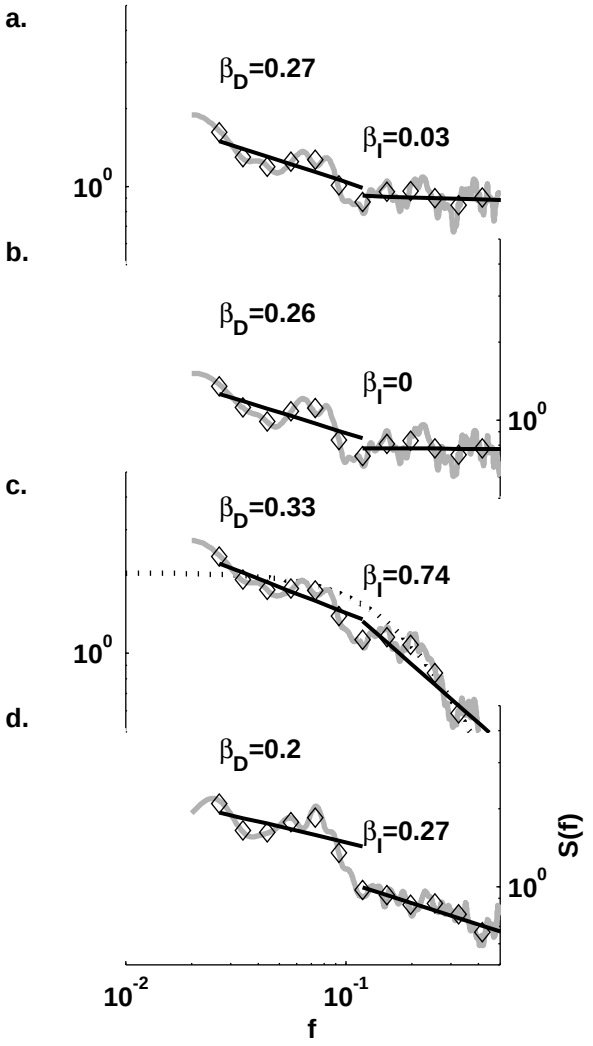

Figure A.5: Mean North American spectra (gray lines) and corresponding values of β computed over interannual and decadal frequencies (black lines) for: (**a**) CRU TS2.1 precipitation; (**b**) the Palmer Z-index (a measure of the P-E balance) calculated from CRU TS2.1 precipitation and temperature; (**c**) PDSI also calculated from CRU TS2.1 data; and (**d**) soil moisture simulated by the VIC model (Andreadis et al., 2005). Diamonds indicate the mean value of the spectrum at the points used to perform regression calculations, and they are evenly spaced in log-transformed units frequency. The dotted line on panel (**c**) shows the theoretical AR(1) spectrum for red noise with autocorrelation equal to the built-in monthly autocorrelation of the PDSI.

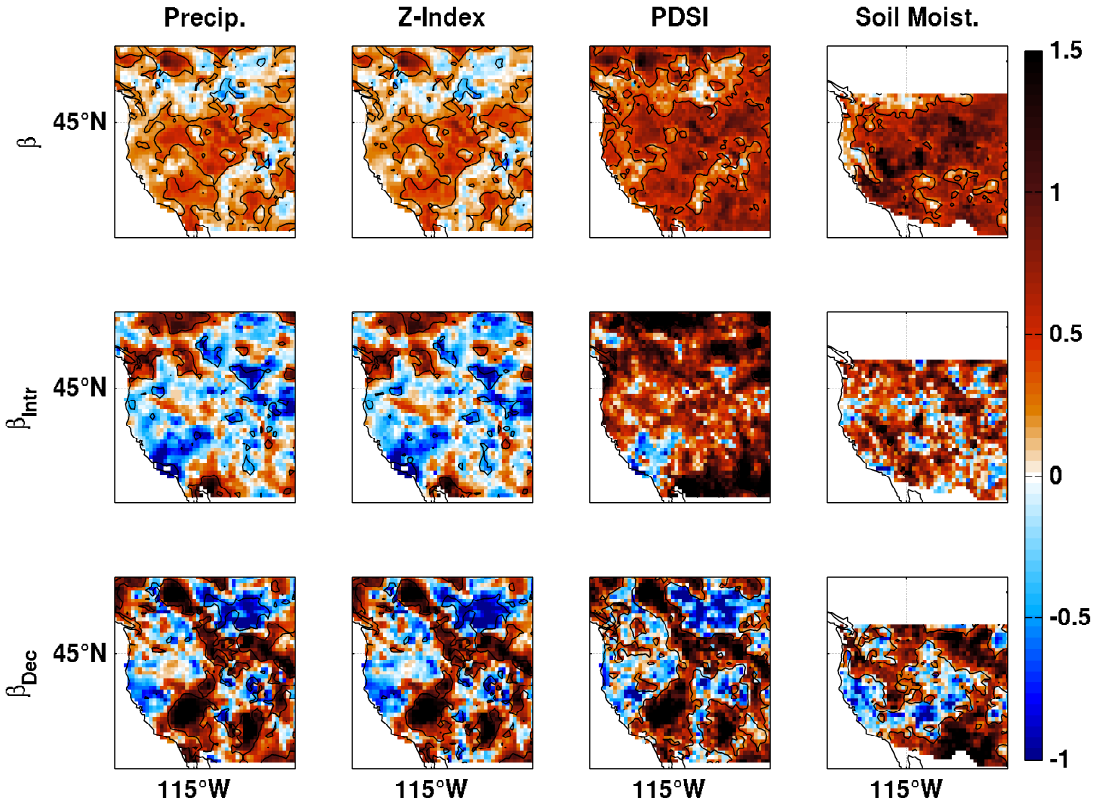

Figure A.6: Scaling exponents calculated for various instrumentally-based datasets from western North America (125°W to 100°W; 20°N to 55°N) at 2 to 50 year, interannual (2 to 10 years), and decadal (10 to 50 years) timescales.

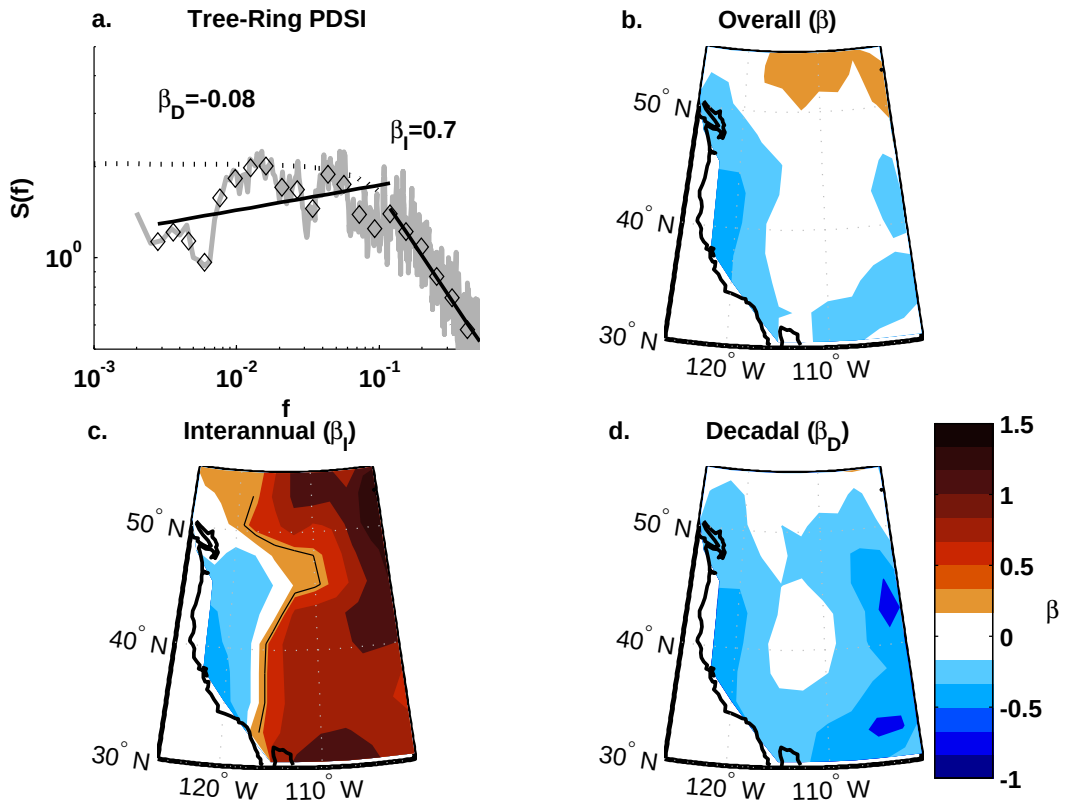

Figure A.7: (**a**) Mean spectral density (i.e., the WNA average at each frequency) calculated from tree-ring reconstructed PDSI (Cook et al., 2004). As in Figure A.5, the gray line shows the mean power spectrum, the straight black lines show the spectral slopes at interannual and decadal frequencies (β_I and β_D), and the black diamonds show the spectral bins that were used to calculate those slopes. The thin dotted line shows the analytical AR(1) spectrum. (**b**) The overall value of β (there are no significant values at the 90% confidence limit). (**c**) Estimates of β_I. Significant spectral slopes are outlined in black (values to the west of this line are significant at the 90% confidence limit). (**d**) the decadal spectral slope β_D, which is not significant anywhere in WNA.

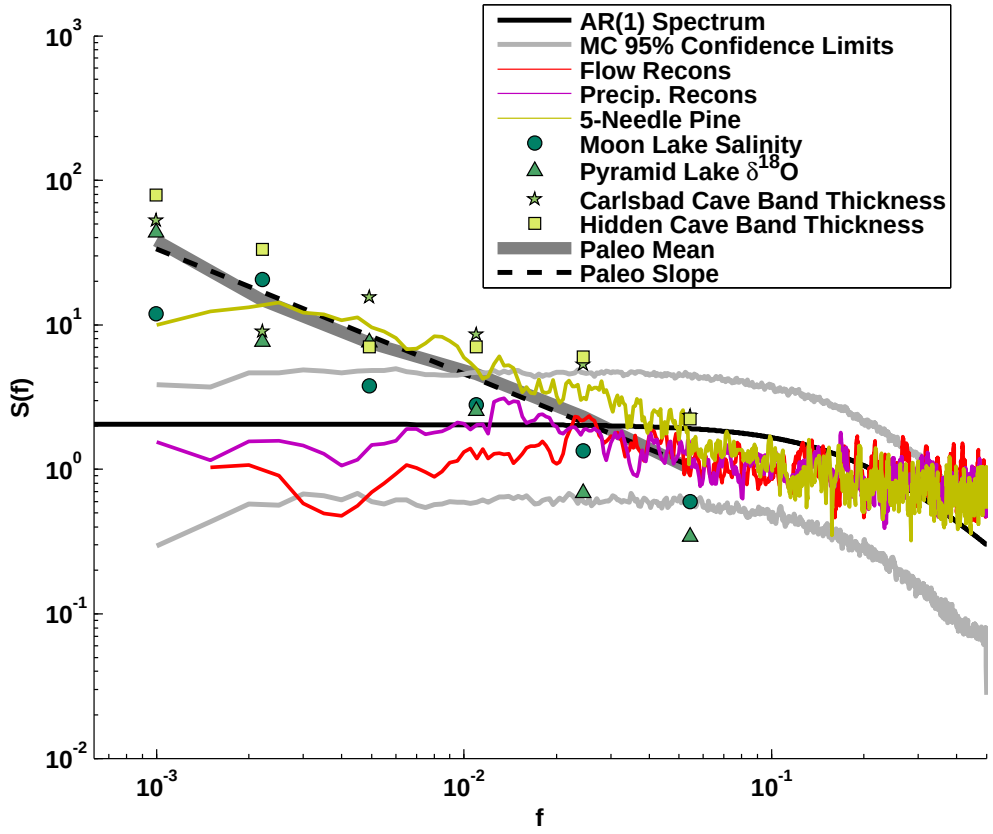

Figure A.8: Analytical AR(1) spectrum (black) and upper and lower 90% Monte-Carlo confidence limits shown with spectral averages from a variety of paleoclimate archives (see supp. mat. for information about the individual records that contribute to spectral averages shown here). The red line shows the spectral average of reconstructed streamflow from 8 sites in WNA, and the purple line shows the spectral average of 25 precipitation reconstructions. Mean spectral density from 10 moisture sensitive, long, 5-needle pine chronologies are shown in gold. Symbols show the spectra of four paleoclimate time series that have been averaged into evenly spaced log-transformed bins. These records include: a diatom-inferred salinity record from Moon Lake, ND (Laird et al., 1998) (circles); an oxygen isotope record from Pyramid Lake, CA (Benson et al., 2003) (triangles); a speleothem band-thickness record from Carlsbad Caverns, NM (Polyak and Asmerom, 2001) (stars); and a cave δ^{18}O record from Hidden Cave, NM (Rasmussen et al., 2006) (squares). The thick gray line is the average spectral density from the four paleoclimate records, and the dashed black line shows the regression line for that average. The regression line (dotted) for the Colorado River power spectrum is also shown

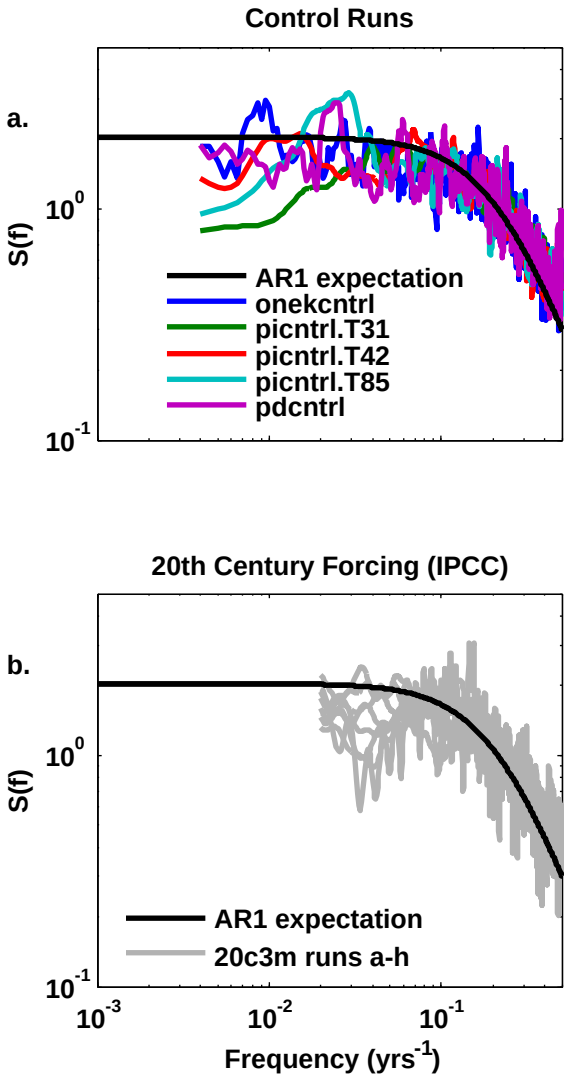

Figure A.9: North American spectral means for PDSI computed from five long control runs (**a**) and eight 20th century simulations (**b**). The smooth black line shows the theoretical mean spectrum of PDSI.

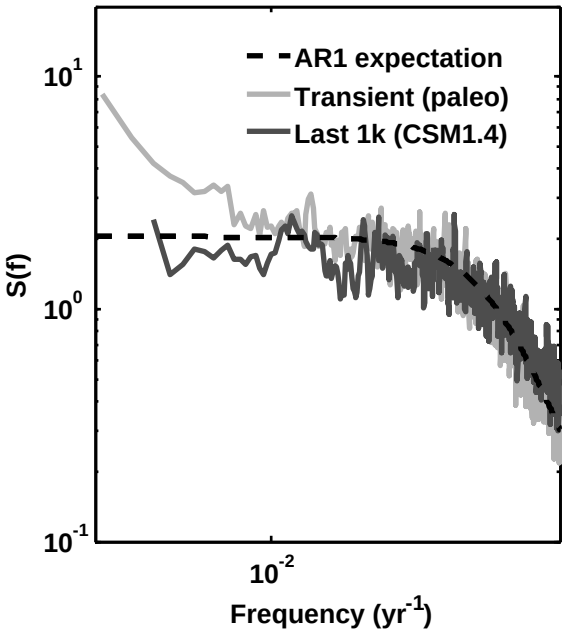

Figure A.10: Mean PDSI spectra for North America from two paleoclimate simulations: (gray line) a 4000BC to 2700BC simulation with transient solar and stochastic volcanic forcing using CCSM3, and (black line) a 1000AD to 2000AD simulation using reconstructed solar variability and stochastic volcanic forcing from a simulation on CSM1.4.

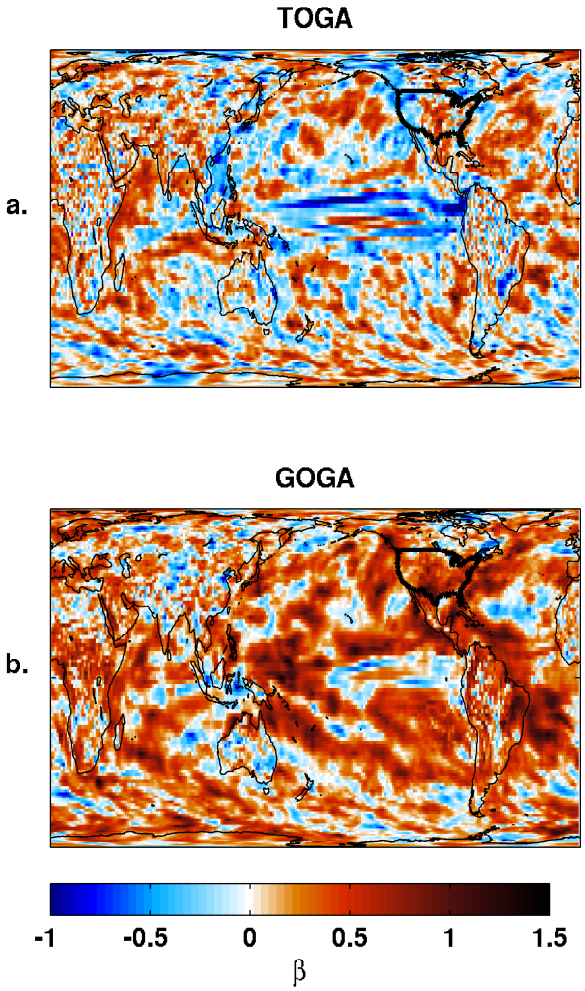

Figure A.11: Global maps of β values computed from two 5-member ensemble averages from CAM3 SST-forcing experiments. (**a**) shows the values of β computed between 1/2 and 1/25 year frequencies from the TOGA experiments and the bottom panel shows the values of β computed from ensemble averages of the GOGA forcing experiments.

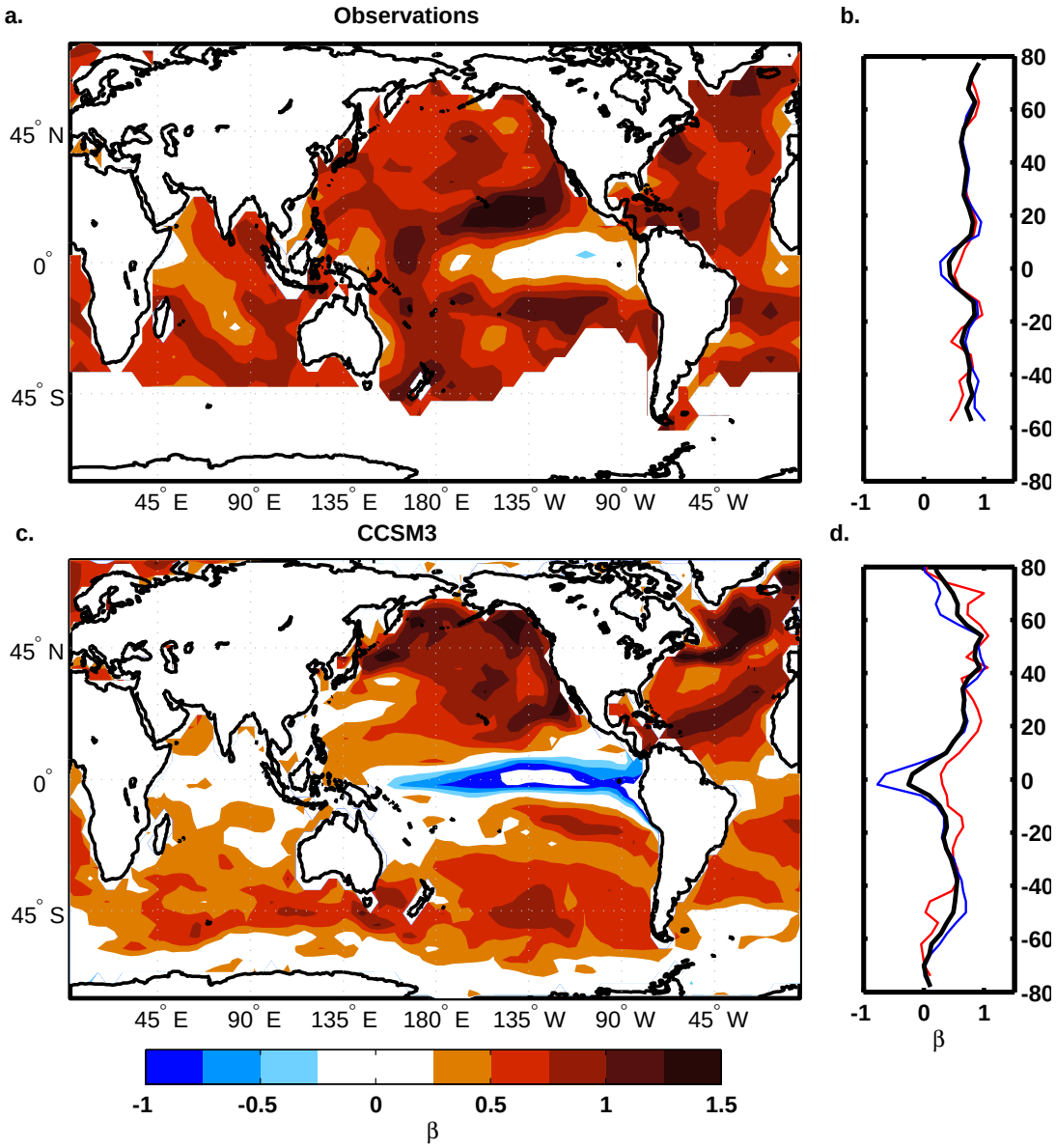

Figure A.12: (**a**) Spectral slopes computed from instrumental SSTs from 1870-2000 (Kaplan et al., 1998). (**b**) The zonal means for the global oceans (black), Pacific Ocean (blue), and Atlantic Ocean (red). Choice of data product does not substantially impact this result. (**c**) β for a long CCSM3 control run. (**d**) Zonal mean values of β from the long CCSM3 control run for all oceans (black), the Pacific (blue) and the Atlantic (red).

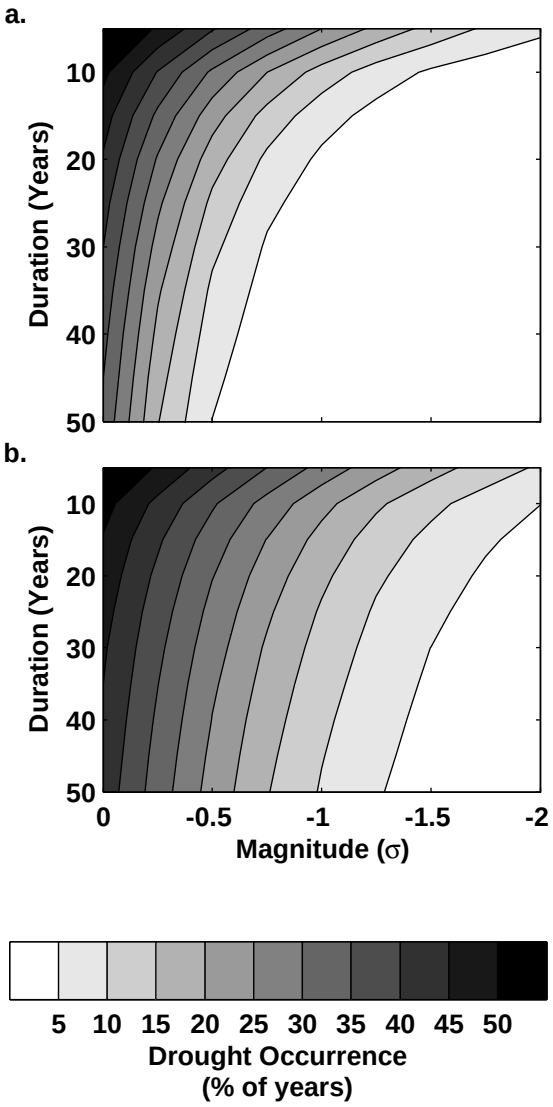

Figure A.13: Duration versus magnitude diagram of Monte Carlo (MC) PDSI simulations of droughts in the Southwest. In this case, a drought is defined as a year in which 3/5 of the antecedent years exhibited PDSI values below a given threshold. The thresholds are shown on the x-axis and the length of the time window considered is shown on the y-axis. The shading indicates the percentage of time (averaged across MC realizations) spent at different duration-magnitude combinations. **(a)** shows the average distribution for AR(1) time series; **(b)** shows the average distribution for time series that have been rescaled to exhibit a realistic spectral slope.

APPENDIX B

HOW LIKELY ARE MEGADROUGHTS IN THE FUTURE?

T.R. Ault[1], G.T. Pederson[2], J.E. Cole[1], J.T, Overpeck[1,3], D. Meko [4]

1. Dept. of Geosciences, Univ. of Arizona, Tucson, AZ
2. Northern Rocky Mountain Science Center, USGS, Bozeman, MT
3. Institute of the Environment, Univ. of Arizona, Tucson, AZ
4. Laboratory of Tree-Ring Research, Tucson, AZ

B.1 Abstract

Projected changes in global rainfall patterns have the potential to dramatically alter water supplies and ecosystems in semiarid regions during the coming century. Instrumental and paleoclimate data indicate that low-frequency hydroclimate fluctuations tend to be more energetic than high-frequency fluctuations. This tendency can be quantified using a "megadrought-risk coefficient" (equivalent to a power-law coefficient) that relates spectral density to frequency. Here we show that global climate models used in the 4th IPCC assessment do not generally reproduce the megadrought-risk coefficients observed in nature, even when run for many centuries or forced by 20th century boundary conditions. Our findings suggest that these models underestimate the risk of future decadal and longer megadroughts. To include quantitative information from paleoclimate records in future megadrought-risk assessments, we propose a method for rescaling projected precipitation changes using local estimates of the megadrought-risk coefficient. Where observational data are reliable, this method can provide a more complete view of megadrought risk. In the Southwest, for instance, IPCC projections suggest the risk of a megadrought in the coming century is about one in ten; our analysis suggests that the risk is at least

one in six for most of the Southwest, and as high as one in three in certain regions. Our results also indicate that observed southwest megadroughts in Colorado River streamflow were approximately 30% more severe than previous estimates, which by itself has major implications for drought risk management in the region.

B.2 Introduction

Many studies have shown that the 20th century does not depict the full range of drought variability Woodhouse and Overpeck (1998); Stahle et al. (2007); Cook et al. (2004); Meko et al. (2007). Paleoclimate records, for instance, provide evidence of dry intervals unlike anything seen during the last century Woodhouse and Overpeck (1998). These multidecadal epochs of aridity (megadroughts) occur naturally and throughout the world Woodhouse and Overpeck (1998); Shanahan et al. (2009); Buckley et al. (2010); Haug et al. (2003); deMenocal (2001). If they were to occur again, the ecological and societal consequences would be severe. In a megadrought during the 1150s, for example, the 25-year average of reconstructed Colorado River streamflow dropped by 15% (2.25 MAF) of the 20th century mean for a Meko et al. (2007). In modern terms, this would be roughly equivalent to losing the entire allocation for the state of Arizona, on average, year after year.

Assessing the risk of megadroughts in the future is challenging because the 20th century record is too short to reliably estimate the rate at which these events occur. For example, the instrumental record of Colorado River streamflow alone is insufficient to determine if decadal-scale droughts would be expected to strike every 100 years or every 500 years Pelletier and Turcotte (1997). This limitation has been partly addressed by using reconstructions of Colorado River streamflow to estimate the underlying distribution of variance across timescales in the reconstruction, and using that distribution to generate Monte Carlo realizations of flow Meko et al. (2010). This approach indicates that events like the 1150s drought would be expected to occur every 400 to 600 years, which is consistent with the

existing reconstructed record Meko et al. (2010, 2007). This view of risk is incomplete, however, because 21st century climate will likely be dramatically influenced by rising temperatures from anthropogenic greenhouse gases Solomon et al. (2007), implying that the risk of future megadroughts will depend on the natural rate at which these events occur as well as the forced component of change. In the US Southwest, for instance, precipitation is projected to decrease as a consequence of forced changes in the average position of winter storm tracks Seager et al. (2007b). Any assessment of future megadrought risk, therefore, should account for both the natural variability inferred from the past few centuries and the changes in rainfall patterns that are projected to occur in the coming century. To that end, we present here a megadrought-risk analysis that uses instrumental and paleoclimate records of hydroclimate to make statistical adjustments to global climate model (GCM) projections of precipitation during the 21st century.

We compute power spectra from drought proxies, instrumental records, and GCMs, and we use power laws to summarize the information in these spectra across a wide range of time scales (here we use the term "time scale" to refer to a group of similar frequencies). This approach allows us to ignore the time evolution of drought events, which differs between models and observations, and focus on the importance of low-frequency (multidecadal to multicentury) variance with respect to high-frequency (interannual) variance. Although a power-law distribution might not necessarily be the *best* fit to the power spectra of all instrumental, model, and paleodrought records (e.g., Clauset et al. (2009)), power laws give us critical insight into the underlying relationship between time scale and variance across disparate types of data.

Power-law coefficients have been widely used to characterize the power spectra of temperature fluctuations in the ocean and atmosphere on monthly to multimillennial time scales Wunsch (2003); Huybers and Curry (2006); Pelletier (1998). Such coefficients relate frequency f to spectral density $S(f)$ such that $S(f) \propto f^{-m}$. Power spectra with higher values of m correspond to time series that exhibit more variance at lower frequencies. For terrestrial temperature records, the parameter m

ranges from approximately zero to 0.5, whereas in the oceans it tends to be somewhat higher (0.5 to 1.5) Huybers and Curry (2006); Fraedrich and Blender (2003). In contrast to temperature, the power spectrum of precipitation tends to resemble that of "white noise" ($m \approx 0$) on interannual time scales Rajagopalan and Lall (1998); Pelletier and Turcotte (1997). On longer time scales, and in other types of hydroclimate variables, variance tends to increase at lower frequencies Pelletier and Turcotte (1997); Kantelhardt et al. (2006); Koscielny-Bunde et al. (2006), with power-law coefficients estimated near 0.5 for globally averaged spectra of river discharge, instrumental precipitation records, and tree-ring widths from moisture-sensitive trees Pelletier and Turcotte (1997).

B.3 Data and Methods

We use paleoclimate data, instrumental records, and climate model simulations to estimated power spectra and power-law coefficients (m) for various hydroclimate indicators. Since these coefficients relate the amount of variance at long timescales to the variance at short timescales, they also indicate how often persistent phenomena (such as megadroughts) are expected to occur Pelletier and Turcotte (1997). We therefore refer to this coefficient (m) as the "megadrought-risk coefficient:" higher values of m will be associated with more frequent intervals of severe, persistent aridity. Figure B.6 illustrates this concept in more detail using Monte Carlo realizations of Colorado River streamflow that have been scaled by different megadrought-risk coefficients. For values of m near zero, megadrought conditions are uncommon; for higher values of m, these conditions become considerably more prevalent (Figure B.6).

The shortness of the instrumental record makes it difficult to estimate the megadrought-risk coefficient (m) on multidecadal, centennial, and longer time scales. We therefore examine the power spectra of paleoclimate records from the central and southwestern US and West Africa. We use a new reconstruction of flow on the Colorado River and four precipitation-sensitive paleoclimate records from the World

Data Center for Paleoclimatology (Supplementary Table B.1). The new Colorado River flow reconstruction is described in detail in the Supplementary Materials section. It was developed using methods aimed at preserving low-frequency variance, and multiple different methodological choices were tested to ensure that our results were stable. We also analyzed other paleoclimate records from the western US (125°W to 100°W, 30°N to 48°N) if they met all of the following criterion: (1) they were at least 1000 years long; (2) they contained data from the last 500 years; (3) they exhibited at least decadal (10 year) resolution on average; and (4) they were primarily drought indicators. Lastly, we calculated m for two paleoclimate records from Lake Bosumtwi, Ghana, which was the only record that fit our selection criterion for western sub-Saharan Africa. Since paleoclimate data tend to be unevenly sampled in time, we used the Lomb-Scargle method Lomb (1975); Scargle (1982) to estimate their spectra. We normalized all paleoclimate data to exhibit unit variance prior to spectral analysis.

We also estimated power spectra and megadrought-risk coefficients (m) from instrumental and simulated 20th century annual precipitation records using the multitaper method Thomson (1982). Instrumental precipitation data originate from the University of East Anglia's Climate Research Unit's (CRU) TS 2.1 database Mitchell and Jones (2005), which spans 101 years from 1901-2002. Precipitation time series at each grid point were averaged to annual (Jan-Dec) resolution. We focus on precipitation because of its obvious link to hydroclimate, but temperature may also play an important role in exacerbating drought. To assess this possibility, we estimated m from the Palmer Z-Index (a measure of P-E) as well. We also estimated m for precipitation data from 18 GCMs used in the IPCC 4th Assessment Solomon et al. (2007) that were annualized in the same way as the instrumental CRU TS2.1 data. The simulations considered here were the forced "Climates of the 20th Century" experiments (77 runs), the longest (>500 year) unforced "Pre-industrial" control runs (6 runs), and projections of the 21st century from the SRESA1B (57 runs) and SRESA2 (37 runs) Solomon et al. (2007) scenarios.

As in Huybers and Curry (2006), we estimated the value of m from individual

power spectra by performing a least-squares regression of log-transformed spectral density against log-transformed frequency. To avoid biasing the regression towards the higher frequencies Huybers and Curry (2006), we averaged spectral density estimates to evenly spaced log-frequency bins prior to calculating m. We ensured that the values of m were not unduly influenced by long-term trends, which would artificially inflate variance at the lowest frequencies, by performing the regression over the frequency domain f_q to $2/N$, where f_q is the highest resolvable frequency (the Nyquist Frequency) and N is the length of the record in years (Supplementary Table S1). Since the power spectrum of white noise would be expected to be characterized by a value of m near zero, we calculate the p-value of the regression slope to determine the probability that the value of m is significantly different from zero. Finally, we used an alternative method (detrended fluctuation analysis, DFA, Peng et al. (1994)) to test the robustness of the spatial patterns of m shown here (see supplementary material).

Global megadrought risk during the 21st century was estimated from unadjusted and rescaled versions of the SRESA1B and SRESA2 climate model simulations. The rescaling was done using an approach similar to the one described by Pelletier and Turcotte (1997) and explained thoroughly in Pelletier (2008). In the earlier work, the authors first used a random number generator to produce white noise time series ($x(t)$, where t is time). They then calculated the Fourier transform ($x(k)$, where k are the standard Fourier frequencies) of these time series, and filtered them to conform to a predefined value of m. These filtered Fourier transforms ($\hat{x}(k)$) where then used to calculate filtered time series ($\hat{x}(t)$) by taking the real part of the inverse Fourier transform of $\hat{x}(k)$. Finally, the mean and variance of the rescaled time series were adjusted to desired values. We applied essentially the same algorithm, but used precipitation projections instead of white noise time series. We also forced the mean and variance of the rescaled precipitation projections to be equal to the original time series. The values of m that were used to rescale precipitation globally were derived from instrumental data (CRU), and (for the US southwest only) proxy data. Figure B.7 illustrates the application of this method to a single

time series from one model.

Here, we define a megadrought as a period of 10 years when the running 25-year projected precipitation mean drops below 85% of the 20th century mean of a given model at a given grid point. We chose this definition because it would characterize a well-known megadrought that occurred on the Colorado River during the 1150s Meko et al. (2007). We calculated the risk of such an event by querying each SRESA1B projection from each model (57 runs total) to determine whether such an event occurred. Finally, we mapped out the percentage of all runs that experienced a megadrought for unadjusted and rescaled versions of projected precipitation data. We subtracted the two maps from one another to show where rescaling might matter most for future megadrought risk.

B.4 Results and Discussion

The time series of our new Colorado River reconstruction is shown in Figure B.1, and discussed further in the supplementary material. During the 20th Century, it is in very good agreement with naturalized streamflow from the Colorado river (Figure B.1a) and the reconstruction in another recent study Meko et al. (2007). Over the last millennium, however, it differs from the earlier reconstruction in several important regards. First, the long-term mean of the 25-year moving average of our reconstruction is 12.43 MAF, whereas the long-term mean of the earlier reconstruction was 14.66 MAF. This offset represents primarily lower mean streamflow values prior to the 20th century (see supplementary material), and the presence of megadroughts in our reconstruction that are even more severe than those previously identified Meko et al. (2007). Second, the variance of the 25-year moving average is somewhat greater in our reconstruction (1.8 MAF^2) as opposed to the variance in the earlier reconstruction (0.58 MAF^2) Meko et al. (2007). This difference can be clearly seen in the distributions of the 25-year moving averages (Figure B.1c), and most notably in the power spectra of the reconstructions shown in Figure B.1d.

The megadrought-risk coefficient (m) is positive ($m = 0.42$) for the power

spectrum of our new 1200-year Colorado River flow reconstruction (Figure B.2a). This indicates that long-term fluctuations in streamflow exhibit more variance than shorter-term fluctuations. Also shown in Figure B.2a are the individual and composite spectra from four independent paleodrought records in the US Southwest. These are lower resolution (decadal) records that span a longer time interval than the Colorado River reconstruction, and they too support a positive value of m ($m = 0.84$ for the mean spectrum of all four records).

In Figure B.2b we show the power spectrum of West African annual precipitation (averaged over 5°N to 15°N and 17.5°W to 20°E) along with two drought records from Lake Bosumtwi, Ghana. They span the last 2,500 years and support a value of m close to 1.04, which is in good agreement with precipitation for the region ($m = 0.94$).

Global estimates of the megadrought-risk parameter (m) are mapped in Figure B.3 for 20th century instrumental precipitation, and results show that most regions of the world experience more low-frequency hydroclimate variability than high-frequency variability. The highest values of the megadrought-risk coefficient (m) occur in Australia, Brazil, West Africa, and the western and northern parts of North America. Accordingly, low-frequency (decadal) variance in these regions is substantially more energetic than variance at higher frequencies (e.g., interannual). We found that the values of m calculated from the Palmer Z-Index (a measure of P-E) were virtually identical to those calculated from precipitation alone (Supplementary Figure B.8). Estimates of a related parameter using DFA Peng et al. (1994) also support the values of m we have shown in Figure B.3.

The GCMs do not appear to simulate the geographic pattern, nor the magnitude, of the megadrought-risk coefficient (m) in precipitation (Figure B.10). This is true for most of the individual models (Figure B.11) and also the longest (>500 year) available pre-industrial control experiments from the IPCC 4th assessment (Figures B.12 and B.13). Instead, values of m are close to zero for most parts of the world, and negative (indicating very energetic high-frequency variance) in the equatorial tropical Pacific. This likely reflects the effect of interannual "El Niño" and "La

Niña" variability that is exceptionally energetic in GCM simulations Guilyardi et al. (2009).

Figure B.4 shows the results of our global megadrought-risk analysis from raw (Figure B.4a) and rescaled (Figure B.4b) SRESA1B projections. In many regions, our rescaling does not increase the likelihood of megadroughts (Figure B.4c). This is partly because in higher latitudes (e.g., poleward of 45-50°N in Europe and North America) and near the equator, the projections suggest it will get wetter Solomon et al. (2007), making megadroughts less likely. It is also partly because, in many regions, the coefficient during the 20th century is not very different from zero, so rescaling precipitation has little impact. However, for several regions where drought is already common (e.g. Australia, the US Southwest, and parts of West Africa), megadroughts are considerably more likely under the rescaled analysis than under the unadjusted projections. Our results also imply that megadrought risk is underestimated across parts of South Asia, South Africa, and the Amazon region of South America.

The megadrought-risk maps in Figure B.4 are sensitive to the choice of the megadrought-risk coefficient, and in the US Southwest estimates of this value differ between 20th century instrumental data ($m = 0.25$) and proxies ($m = 0.42$ to 1.0). We therefore construct another set of megadrought-risk maps for the US Southwest using unadjusted SRESA1B precipitation (Figure B.5a), rescaled precipitation using 20th century estimates of m (Figure B.5b), and rescaled precipitation using the low (conservative) end of our proxy estimates of m ($m = 0.4$) (Figure B.5c). According to the raw SRESA1B precipitation estimates, the risk of a megadrought is less than 10% for most of the domain. If only instrumental estimates of m are used, our results imply that there is a 10% to 20% chance of an 1150s-style megadrought during the coming century (Figure B.5b). However, if we rescale SRESA1B precipitation using proxy estimates of m, we see that there is at least a one-in-six (15%) chance of such an event in most of the Southwest, and a one-in-four (25%) or one-in-three (30%) chance in some areas. These estimates are conservative: we used the low end of the proxy range of m, and the SRESA1B scenario assumes some emission cuts in

the coming century. Maps of megadrought risk in the SRESA2 scenario show even higher likelihoods (i.e., up to 30% to 50%; Figures B.14 and B.15).

B.5 Conclusions

We have focused our analysis of megadrought risk to a relatively small part of the globe (the US Southwest) because there are abundant high-resolution and high-quality hydroclimate proxies from this area, which give us confidence in the values of m we have estimated for this region. Here, cold season precipitation has been, and likely will be, the primary driver of megadrought variability. However, in the southern part of this region (as in many other areas of the world), monsoonal variability contributes considerable amounts precipitation during the warm season, and some GCMs may not simulate the intensity, nor certain dynamics, of summertime precipitation Liang et al. (2008); Cavazos et al. (2008). If this is the case, then the GCMs may provide an incomplete view of the full range of future hydroclimatic changes that could occur in monsoon-sensitive regions, and hence our megadrought-risk assessments may also be constrained by these limitations. Future research could therefore apply the tools we have developed here to a subset of GCMs that simulate desirable characteristics of monsoonal variability.

Future work should also identify the likely physical mechanisms that contribute variance to the power spectrum of precipitation at different timescales. One possibility is that high-latitude, low-frequency forcing paces hydroclimate in nature Kushnir et al. (2010); Schubert et al. (2009), but that in models the higher frequency, tropical Pacific fluctuations dominate the power spectrum. Another possibility is that low-frequency variability in the tropical Pacific, which is not well resolved in instrumental records Cole et al. (1993); Urban et al. (2000); Cobb et al. (2003); Conroy et al. (2008); Ault et al. (2009) or GCMs Guilyardi et al. (2009), is responsible for the low-frequency behavior in hydroclimate at decadal timescales and longer in some regions. It is also possible that the recent observed "widening of the tropics," which models do not completely simulate Seidel et al. (2008), is linked

to natural components of variability that influence low-frequency precipitation variance.

B.6 Implications

Droughts in the past have had dramatic human and financial costs. In the US alone, for instance, the Federal Crop Insurance Corporation spent $1.7 billion in losses per year from 1980-2005 on average Stephenson (2007). In the future, such losses might be curtailed if the full range of "natural" and "forced" hydroclimatic variability can be included in megadrought-risk mitigation strategies. Here, we have described a method for combining insights from observational data and projections from climate models to estimate the risk of persistent intervals of aridity in the coming century. In some regions (such as northern Africa), this risk is extremely high (>50%) whether unadjusted or rescaled projections are considered. In Australia, the US Southwest, and Central America, the risk is about 15% to 30% if unadjusted projections are considered, but higher (20% to 50%) if observational estimates of the megadrought-risk coefficient (m) are used to rescale precipitation. In the US Southwest, where high-quality proxy records of hydroclimate are abundant, risk is about 10% for most of the region if unadjusted projections are used, but as high as 30% if proxy-based megadrought-risk coefficients are used.

B.7 Acknowledgments

We thank Steve Gray, Connie Woodhouse, and the US Bureau of Reclamation for assistance, and we appreciate the helpful reviews we received from Joellen Russel, Paul Cross and Scott St. George. This project was supported in part by an NSF fellowship to T. Ault, NOAA CCDD (NA07OAR4310054) and NSF P2C2 (0903093) support for J. Cole, and NOAA Climate Program Office support for CLIMAS (J. Overpeck). Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

B.8 Figures

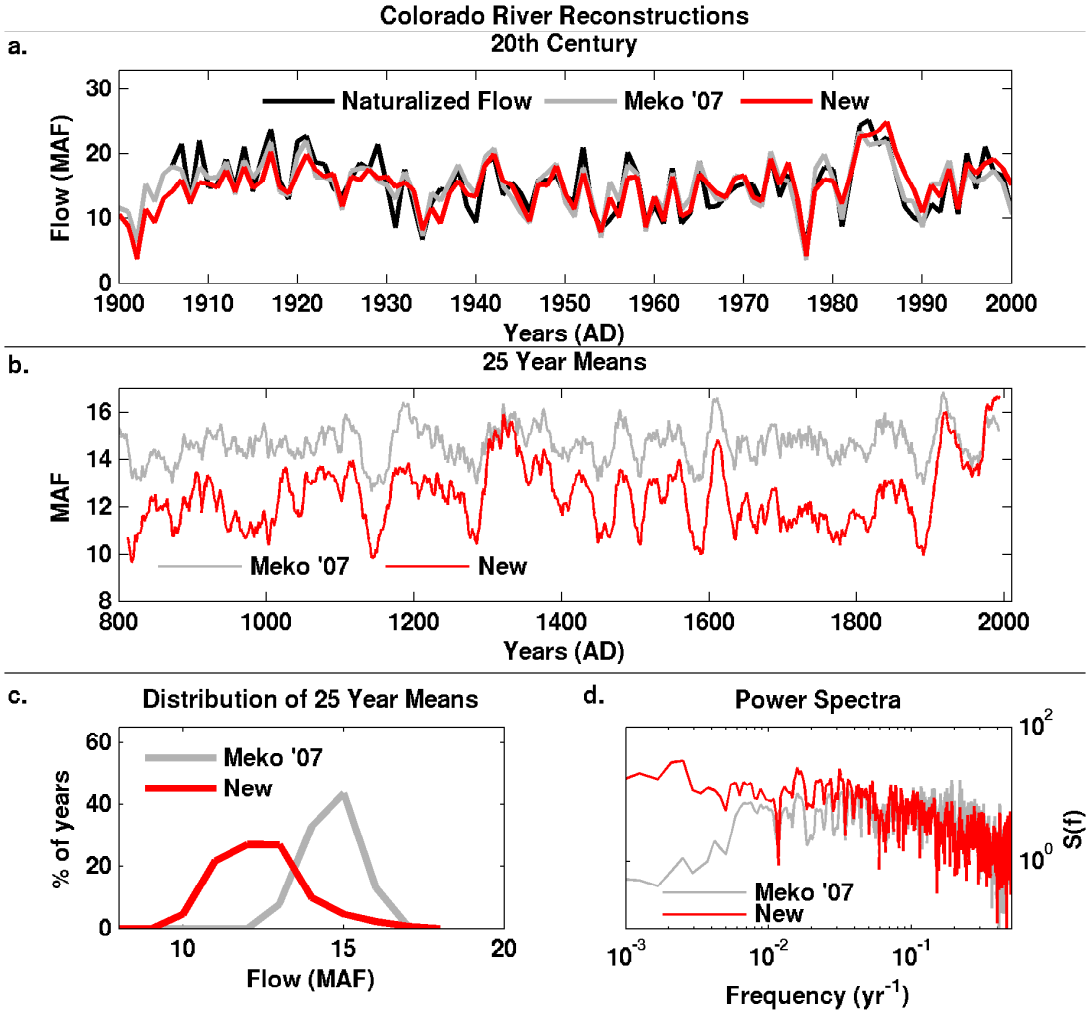

Figure B.1: **a.** Naturalized flow on the Colorado River at Lee's ferry (black line) shown with a recent reconstruction Meko et al. (2007) (gray) as well as our new reconstruction (red) during the 20th century. **b.** 25-year running mean of our new reconstruction (red) shown with the 25-year mean of the reconstruction in Meko et al. (2007) (gray). **c.** Distribution of 25-year running means in the earlier reconstruction Meko et al. (2007) and our reconstruction to illustrate that the mean of our reconstruction is lower (12.43 MAF compared with 14.66 MAF in Meko et al. (2007)) and the decadal variance is higher (1.8 MAF^2 as opposed to 0.58 MAF^2). This finding is further illustrated in **d.**, which shows the power spectra of the two reconstructions. Clearly, more low-frequency variance is present in our reconstruction, which can be seen by the more energetic power spectrum (red line in **d.**; see supplementary material for further discussion of this finding).

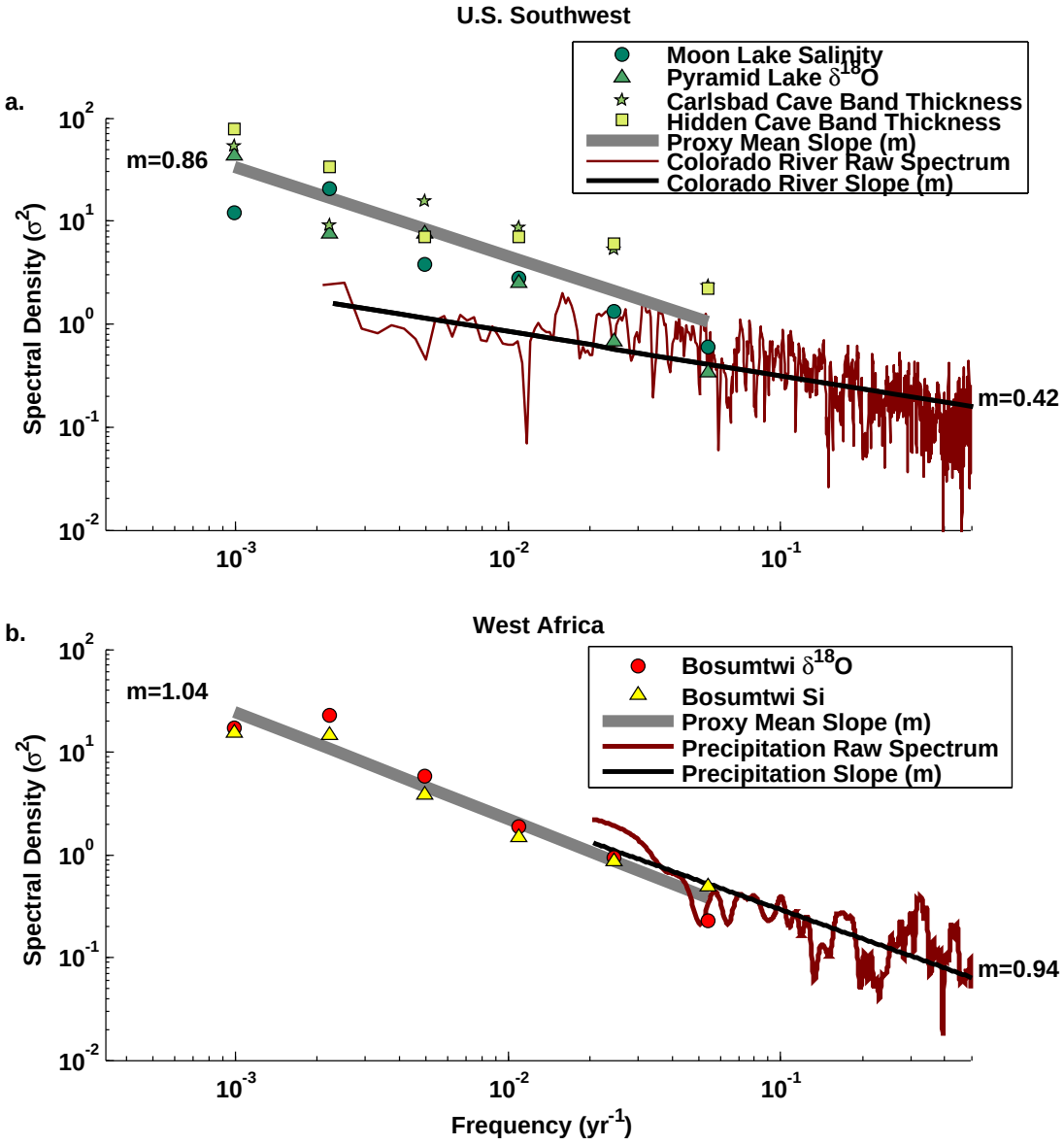

Figure B.2: *See caption on following page*

Figure B.2: **a.** Spectra and megadrought-risk coefficients (m) from reconstructed Colorado River Flow (red line) and four paleodrought indicators (symbols). These records include: a diatom-inferred salinity record from Moon Lake ND Laird et al. (1998) (circles); an oxygen isotope record from Pyramid Lake, CA Benson et al. (2003) (triangles); a speleothem band-thickness record from Carlsbad Caverns, NM Polyak and Asmerom (2001) (stars); and a cave δ^{18}O record from Hidden Cave, NM Rasmussen et al. (2006) (squares). The least-squares regression line is shown in gray and was obtained by regressing the log-transformed average spectral density of all four proxies against log-transformed frequency. The megadrought-risk coefficient (e.g., the slope of the least squares regression line) is 0.86 for the proxy average. The least squares regression line for the Colorado River power spectrum shown in black ($m = 0.42$). **b.** The West African precipitation spectrum (red) and spectra of paleodrought proxies from Lake Bosumtwi (symbols) Shanahan et al. (2009). The Lake Bosumtwi δ^{18}O record indicates evaporation during dry epochs, and the Lake Bosumtwi Si record is a proxy for terrigenous flux into the lake, which tends to increase during lake low stands. Again, the regression line for the proxy spectral average is shown in gray ($m - 1.04$), and the regression line for precipitation is shown in black ($m = 0.94$). Spectral densities of all records were averaged into evenly spaced log-frequency bins and used to calculate m.

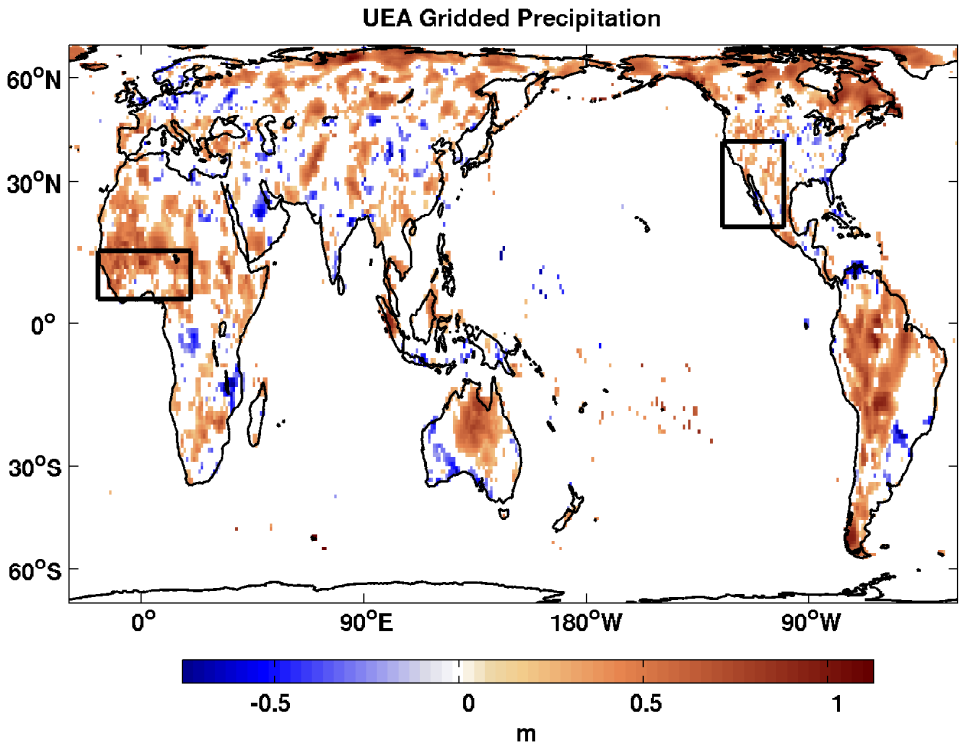

Figure B.3: Map of the megadrought-risk coefficient (m) calculated from annually averaged (Jan-Dec) precipitation. The parameter m was estimated by regressing log-transformed precipitation power spectra against log-transformed frequency at each grid point. To avoid introducing bias from long term trends, the value of m was calculated between periods of 2 years and 50 years. Regions where the value is significantly different from zero ($p < 0.01$) are shaded. Warm colors indicate regions where long-term moisture anomalies are more energetic than short-term fluctuations (blue colors indicate the opposite)

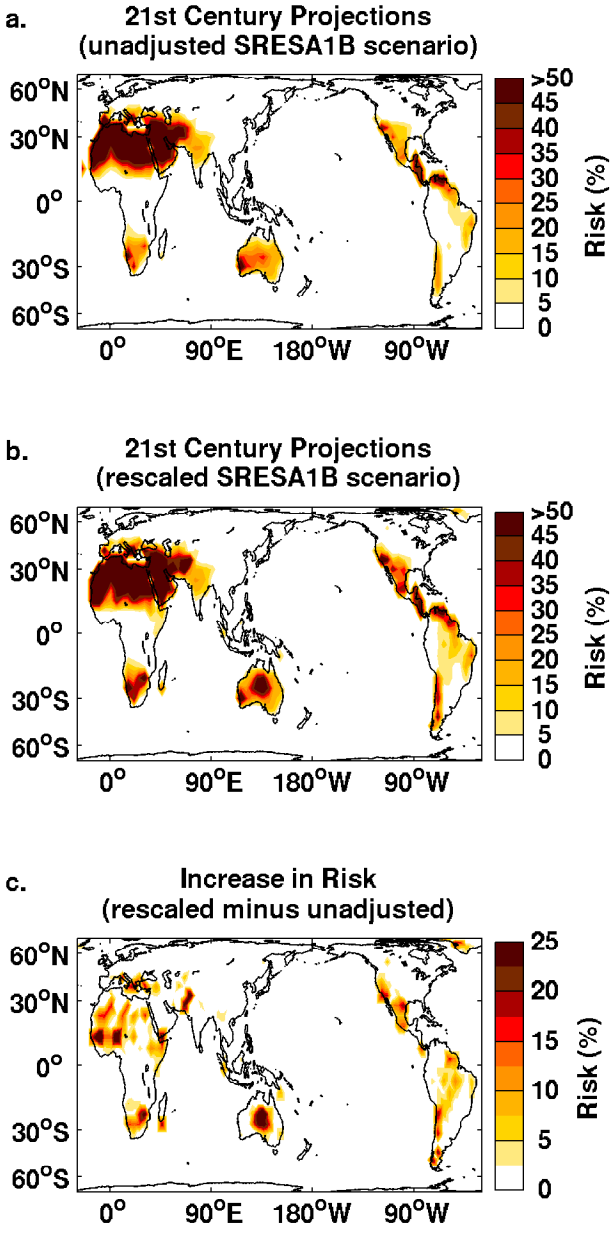

Figure B.4: Risk of a 21st century megadrought in SRESA1B IPCC AR4 simulations (shown as percentage of all runs). **a.** Risk in unadjusted SRESA1B annual precipitation projections. **b.** Risk in rescaled precipitation projections. **c.** Difference in percentages obtained by subtracting **a** from **b**.

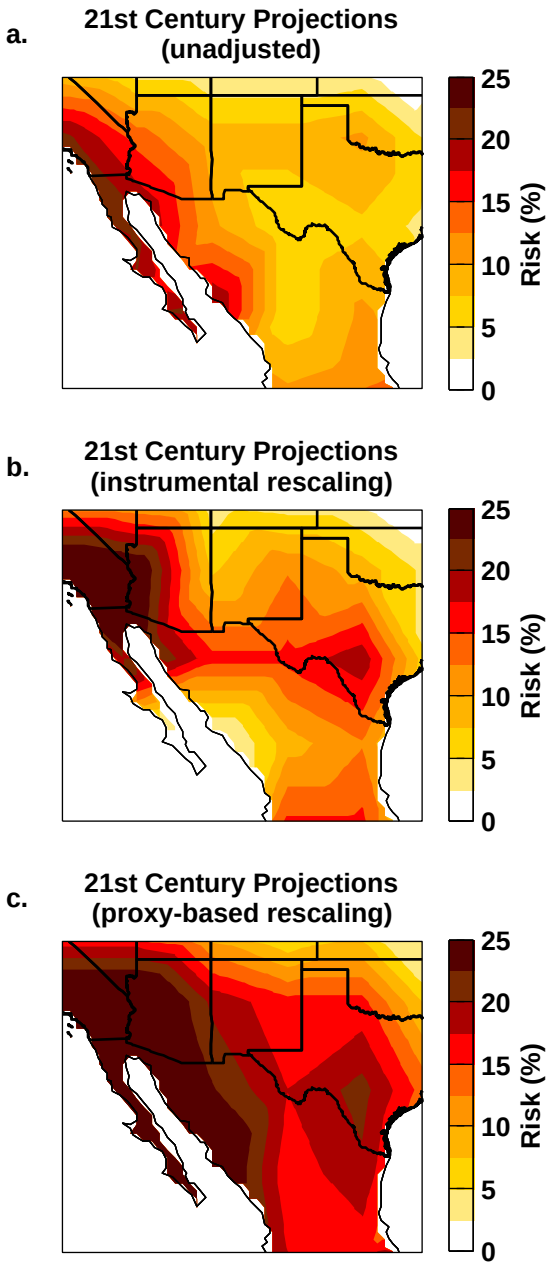

Figure B.5: Risk of a 21st century megadrought in the US Southwest (as in Figure B.4, shown as a percentage of runs). **a.** Risk in unadjusted SRESA1B annual precipitation projections; **b.** risk in precipitation projections rescaled using instrumental values of m; **c.** risk in precipitation projections rescaled using our proxy-based estimate of m ($m = 0.4$).

B.9 Supplementary Material

B.9.1 Monte Carlo Realizations of Streamflow

To illustrate usefulness of the power-law coefficient (m) as a drought risk indicator, we performed Monte Carlo simulations of Colorado River streamflow. In each Monte Carlo simulation, a random (white noise) time series was generated with the same mean and variance as 20th century observations of Colorado River flow, then rescaled by a predefined power law using methods described in the text. Next, we tallied the total number of years where the 25-year running mean dropped below the 20th century mean, and averaged this count by the total number of centuries (twelve) in our simulation. This provides us with an average of the number of years that experienced megadrought conditions for each of the 1000 Monte Carlo realizations at each value of m (we chose values between zero and 1.6 in increments of 0.2). Note that this average number of years differs slightly from the definition of "a megadrought" used in the text: this choice was made to include a more complete view of the average number of *years that experienced megadrought conditions* instead of the *total number of decadal megadroughts* per century.

B.9.2 Colorado River Reconstruction

Methodological Details

We produced a new Colorado River reconstruction using methods that preserved variance across the widest possible range of time scales given the available tree-ring data. We began with raw ring-width measurements from 9 sites in the upper Colorado River basin (Table B.2 and Figure B.16); these were all used in an earlier study (1), referred to as M07 from here on. Raw data (ring-width files) are all available from the International Tree-Ring Data Bank (http://www.ncdc.noaa.gov/paleo/treering.html). First, we selected only segments that were at least 400 years long, then detrended these segments using splines whose frequency response was 95% at 400 year wavelengths. This left us with 283 samples,

which were averaged together to produce a "master chronology." We then used least-squares regression to derive a linear model to predict 20th century Colorado River flow at Lee's Ferry from the master chronology. The model was then validated using a "leave-one-out" technique: for the 99 years of overlap, 98 years were used to calibrate the model, and 1 observation was used to validate it. This process was repeated 99 times and the root-mean-squared error and R^2 values were calculated from the validation procedure. Since regression-based reconstructions can deflate variance, we also reconstructed flow by rescaling the mean and variance of the master chronology to match 20th century flow. We obtained such similar reconstruction and validation statistics that we only report the regression-based results here (Table B.3). However, the mean and variance rescaling approach suggested that the low-flow intervals in our reconstruction may have been even more severe, hence our results may be a slightly conservative estimate of the worst droughts on the Colorado River.

The methodology of our reconstruction differs from M07 in three regards. First, the earlier study selected segments that were 250 years long and longer, whereas we used segments that were at least 400 years long. Naturally, this left us with fewer segments (283 total). Second, M07 detrended segments using splines whose frequency response was 95% at twice the segment length. This would mean that the splines used to detrend the shortest series would have a 95% frequency response at 500 years, but that the frequency response would be different for segments of different lengths. In our reconstruction the frequency response is nearly the same for each segment because we used splines with a 95% frequency response at 400 year wavelengths, and used only segments that were 400 years long or longer. This removed almost all variance at wavelengths longer than 400 years from all segments. In effect, we used conservative detrending as a high-pass filter to maintain more control over the frequency response. Thirdly, M07 used the first principal component time series of site chronologies in nested models, but we constructed a single master chronology and used it as a predictor of Colorado River flow. Despite the differences our methodologies and those employed by M07, the resulting reconstructions are

remarkably similar (Figure B.17 and B.18).

We compared our reconstruction with M07 by constructing two additional master chronologies and repeating the regression analysis (summarized in Table B.3). We first returned to the raw ring-width measurements and selected segments that were 250 years or longer as in M07. We detrended these segments using splines with a 95% frequency response at 100 years. Next, we detrended the network of 283 samples used in our reconstruction using the more flexible splines (95% response at 100 years). The results are reported in rows b and c of Table S3.

The 25-year means of these additional reconstructions are shown in Figure B.18. During the 20th century, all three of our reconstructions are in good agreement with M07 as well as naturalized 20th century flow. Validation statistics for our reconstruction are also quite good (Table S3). In earlier centuries, the dry intervals appear drier in our reconstruction from conservatively detrended segments (Figure B.3, top left). In contrast, the 25-year means of our other two reconstructions (which removed more low-frequency variance) are nearly identical to the 25-year mean of the M07 reconstruction. Clearly, the reconstruction is more sensitive to the detrending choices we made than to the network of trees we used, which helps us rule out the possibility that differences in data account for differences in the reconstructions. Instead, we argue that our conservative detrending methods account for nearly all of the differences between our reconstruction and M07.

To further examine the impact of our different detrending methods, we modeled hypothetical power law distributed spectra whose low-frequency variance had been removed using the two different criteria used here (95% at 100 years and 95% at 400 years). The stiffness, or flexibility, of a spline is determined in part by the spline parameter, p, which is usually specified to yield a desired frequency response (2). Since the definition of the spline parameter in (2) is somewhat different than the definition used in MATLAB©(the software we use), we apply the equations provided by J. Dupouey (INRA, Forest Ecology and Ecophysiology Unit, Chanmpenoux, France) to determine the expected frequency response of a given spline with parameter p:

$$u(f) = \frac{1}{\left[1 + 12\left(\frac{1-p}{p}\right)\frac{(\cos 2\pi f - 1)^2}{(\cos 2\pi f + 2)}\right]} \tag{B.1}$$

Where f is frequency. Note that $u(f)$ is normalized so that it attains a maximum value of 1.0 (i.e., 100% of the variance would be removed below some frequency interval, f_0). The residual energy in the power spectrum, therefore, could be approximated by: $S'(f) = S(f) - u(f)$, where $S(f)$ is spectral density.

Since the spline parameter p varies slightly from sample to sample, the expected frequency response will also vary. We therefore use equation (1) to model a set of hypothetical spectra (gray lines in Figure B.19 and B.20). These spectra are assumed to have a power-law distribution ($S(f) \propto f^{-m}$) with $m = 0.5$ and have reduced variance at low frequencies according to each of the values of p used to detrend the individual ring-width measurements in our reconstruction. This is only an approximation, however, because in reality ring-width measurements are detrended by dividing by a spline curve, not by subtracting that curve. Nonetheless, the hypothetical spectra give us a good sense for the point in the power spectrum where detrending the individual ring-widths is likely to impact the final master chronology. When conservative detrending is used, this point occurs at multicentury time scales (Figure B.19); when more aggressive detrending is used, it occurs at multidecadal time scales (Figure B.20).

Key Features of the New Reconstruction

A distinctive difference in our new reconstruction, when compared to M07, is an early 1900s shift, centered on about 1905 (Figure B.18). This makes the entire 20th century look wet with respect to the previous 1000 years. We examine the source of the early 20th century shift in our reconstruction by averaging the ring-with indices that have at least 90 years of data during the 20th century (Figure B.21a), which rules out the possibility that the shift is driven by changes in the network. Next, we divided all ring-width indices into pre- and post-1900's intervals, then took the mean of each index during those two intervals. The distribution of these pre- and

post-1900's means are shown in the box plots in Figure B.21b. Importantly, this results shows that the shift doesn't seem to be driven by outliers. Finally, we rebuilt the chronologies used in M07 after filtering out the segments that were shorter than 400 years or had less than 90 years of data during the 20th century (Figure B.22). This step confirmed that the shift is present in most of the individual chronologies. Taken together, these three analyses suggest that the shift is not an artifact of the data or methods, but instead reflects a fairly ubiquitous increase in ring width that occurred near the turn of the 20th century.

References Cited in Supplementary Material

1. D. M. Meko, et al., Geophysical Research Letters 34, L10705 (2007).

2. E. Cook, K. Peters, Tree-Ring Bulletin 41, 45 (1981).

B.9.3 Supplementary Figures

Table B.1: Summary of paleoclimate data from the southwestern United States and southern Great Plains used in this study. The parameter m was calculated between the shortest resolvable periods (twice the sampling resolution) and periods half the length of the record; the resolution of each record is shown in the fourth column ("**Resolution**"), and the length of the record divided by two is shown in the fifth column ("**N/2**").

Type of record	**Reference**	**Location**	**Resolution**	**N/2**
Diatom inferred Salinity	Laird *et al.*, 1998	Moon Lake (46.6°N, 98.16°W)	5yr	1167
Lake $\delta^{18}O$	Benson *et al.*, 2002	Pyramid lake (40.066°N, 119.56°W)	4-8yr	2236
Speleothem Band Thickness	Polyak *et al.*, 2001	Carlsbad Caverns (32°N, 104°W)	1yr	2230
Speleothem $\delta^{18}O$	Rasmussen *et al.*, 2006	Hidden Cave (32°N, 104°W)	1yr	2265

Table B.2: Basic information about tree-ring sites used in our Colorado River reconstruction. The first column is the site name. The second is the original file name of the ".rwl" (ring-width) file that contain the segments used to build our master chronology. The "species" column is the species code (PSME for *Pseudotsuga menziesii*, PIED for *Pinus edulis*). Latitude, longitude, and elevation (in meters above sea level) are shown in columns 4-6. The number of ring-width indices used from the original chronology are shown in the last column (N).

Site	File	Species	Lat	Lon	Elev	N
Harmon Canyon	har2	PSME	36.7°N	110.5°W	2012	43
Beef Basin	bfb3	PSME	36.7°N	110.5°W	2012	23
Wells Draw	wed2	PIED	40°N	110.2°W	2027	13
Eagle	egl11	PSME	39.4°N	105.2°W	2103	20
Pump House	pum86	PSME	40°N	106.5°W	2194	22
Wild Rose	wil44	PSME	39°N	108.2°W	2636	26
Trail Gulch	trg55	PSME	39.7°N	107°W	2210	14
Lands End	lan34	PSME	39°N	108.2°W	2987	26
Green Mtn Res	gmr20	PSME	39.9°N	106.2°W	2514	26

Table B.3: Reconstruction statistics for three different pre-processing methods: **a.** conservative detrending and long segments; **b.** flexible detrending and short segments; and **c.** flexible detrending and long segments. A spline designed to remove 95% of the variance at periods 400 years and longer was used for "conservative detrending," whereas a spline designed to remove 95% of the variance at 100 years and longer was used for "flexible detrending." Segments were included in the "long segment" network if they were at least 400 years long, and included in the "short segment" network if they were 250 years long or longer. The network is therefore exactly the same in **a.** and **c.**

	Model	N	R^2	R^2_{adj}	RE	$RMSE_v$	r	PRESS	D-W	Years
a.	PFIT5	99	66.1	65.7	64.85	2.56	0.813	659.24	1.57	800-2007
b.	PFIT3	99	74.9	74.6	73.77	2.20	0.865	491.84	1.96	800-2007
c.	PFIT3	99	73.6	73.3	72.46	2.26	0.858	516.50	1.99	800-2007

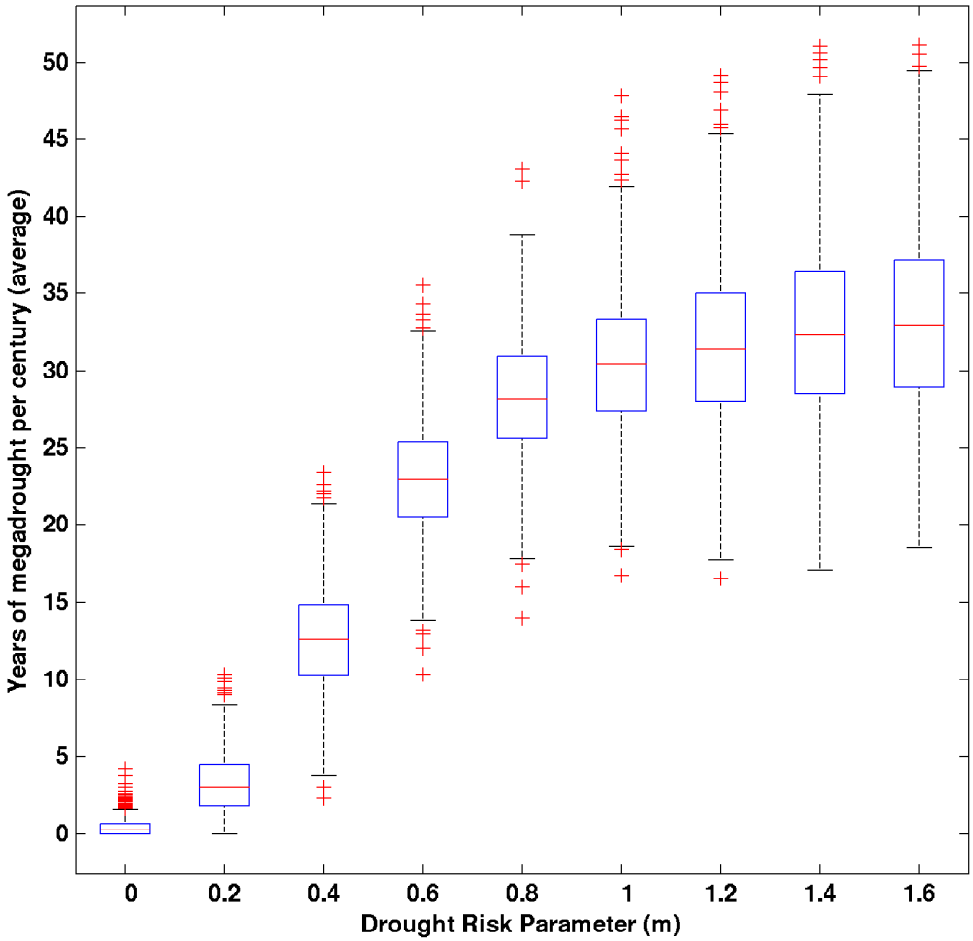

Figure B.6: An illustration of how the "drought-risk coefficient" (m) relates to the frequency of years that experience megadrought conditions. Here 1000 Monte Carlo simulations of flow on the Colorado river were conducted with predefined megadrought-risk coefficients (shown on the x-axis), and the average number of megadrought years per century are shown on the y-axis for each of the 1000 realizations for each of the given values of m.

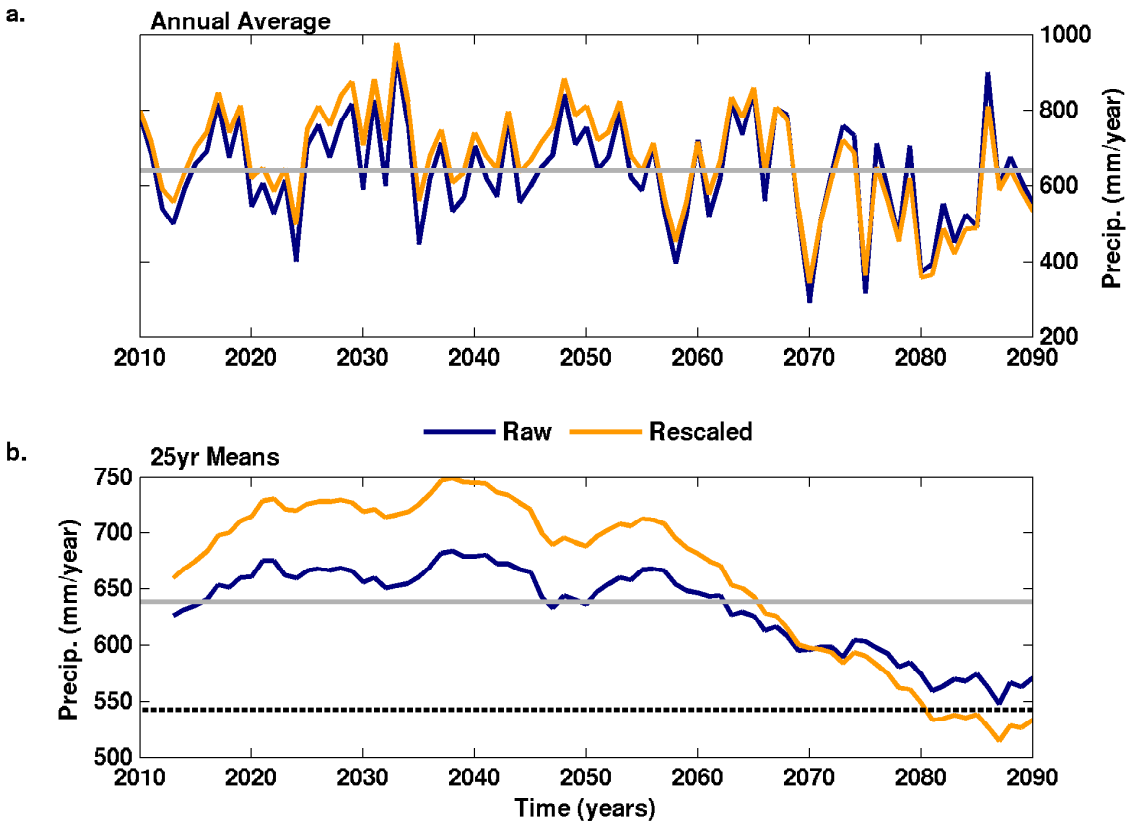

Figure B.7: (**a**) Raw (blue) and rescaled (orange) projections of annual precipitation amounts from the GFDL CM2.0 model show to illustrate the effects of rescaling on megadrought occurrence. (**b**) 25-year means of the precipitation amounts shown in (**a**), along with the 20th century mean (gray) and the megadrought threshold we use to identify events. In the raw projection, no megadrought is detected in the 25-year mean, in the rescaled projection, a megadrought occurs starting in 2080.

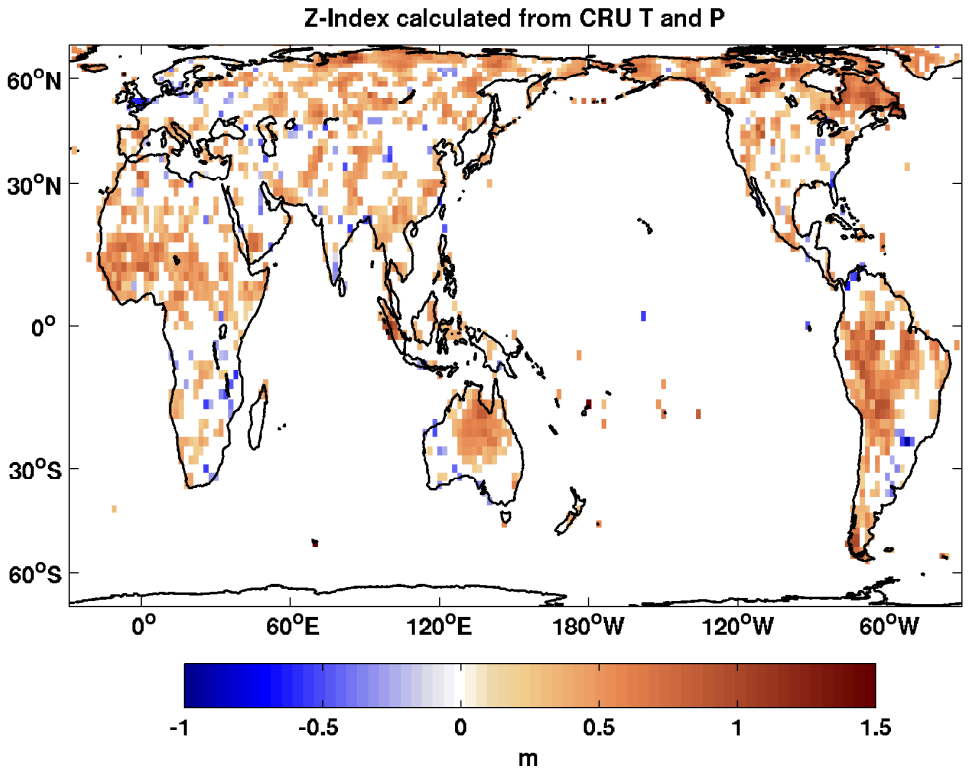

Figure B.8: Same as Figure 1a, but for the Palmer Z-Index calculated from instrumental CRU TS2.1 precipitation and temperature.

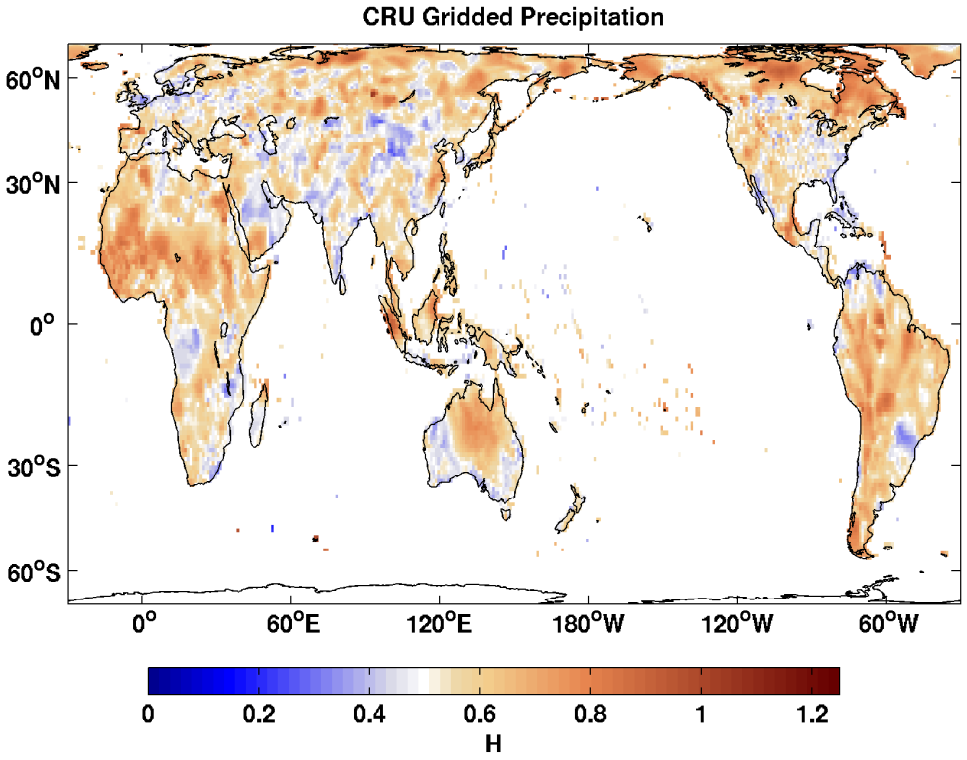

Figure B.9: Hurst scaling exponents (H) obtained from Detrended Fluctuation Analysis (*5*) performed on global annual precipitation. As in Figure 1, warm colors indicate regions with more energy at longer time scales. The Hurst scaling exponent is related to m by $m = -(2 * H - 1)$ (*8*); the values of H in, say, Brazil (0.8 to 1.2) are consistent with the values of m (0.5 to 1) in the same region.

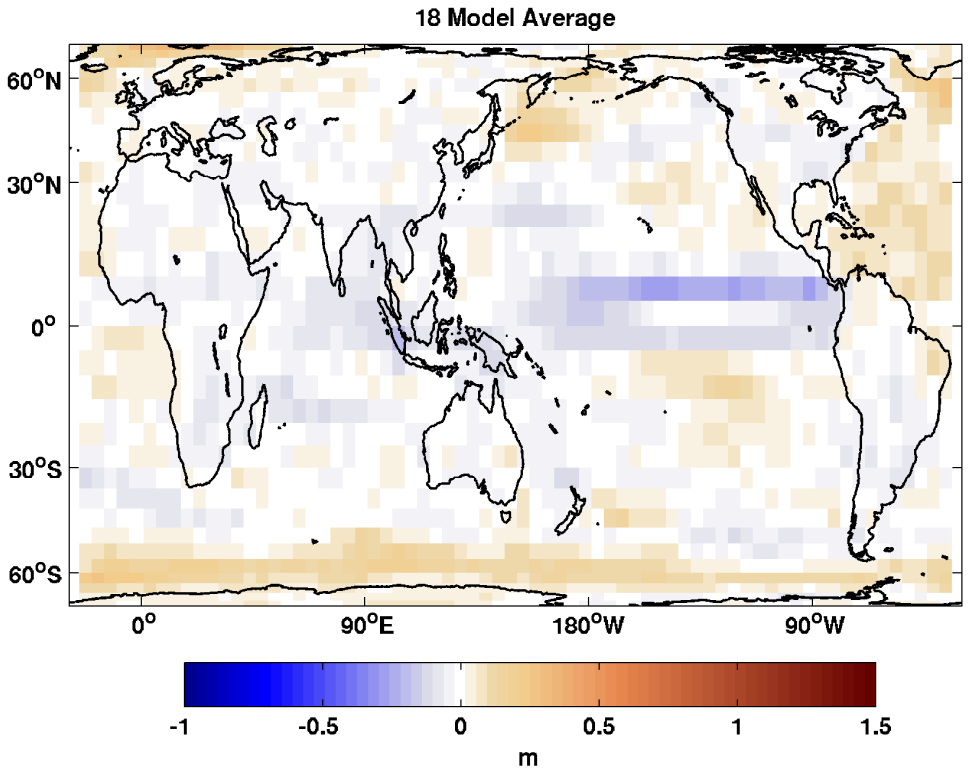

Figure B.10: Map of m calculated for precipitation fields from 18 IPCC AR4 simulations of the 20th century and averaged across models. When multiple ensemble members were available for a single model, the ensemble-average map of m was calculated first and used in the multimodel average shown here.

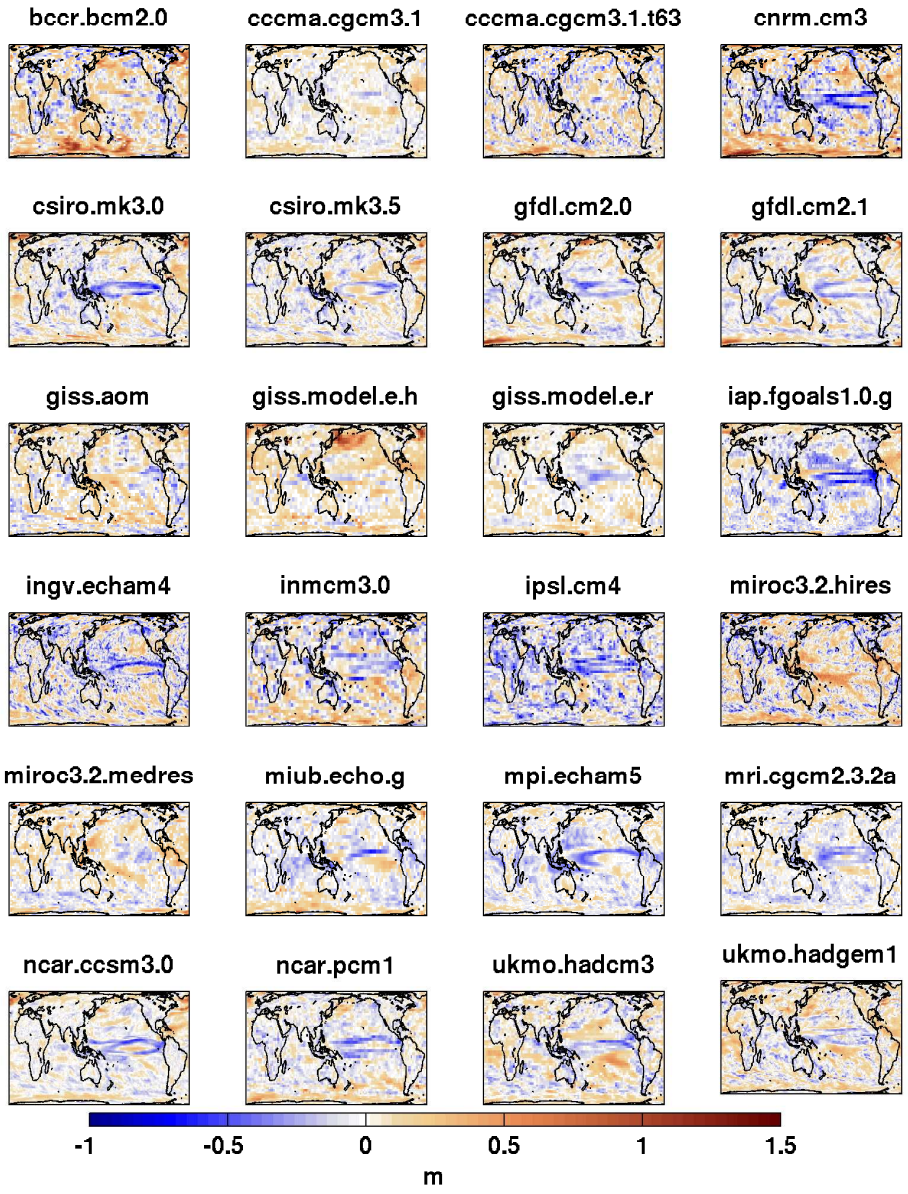

Figure B.11: Scaling coefficient (m) mapped out for each of the 18 individual model "20th Century" simulations used to make the composite map in Figure 3. In the cases where multiple ensembles were available for a single model, m was calculated for each member individually, then averaged (the ensemble average is shown here). All data were obtained from the "Earth System Grid" FTP site (ftp-esg.ucllnl.org).

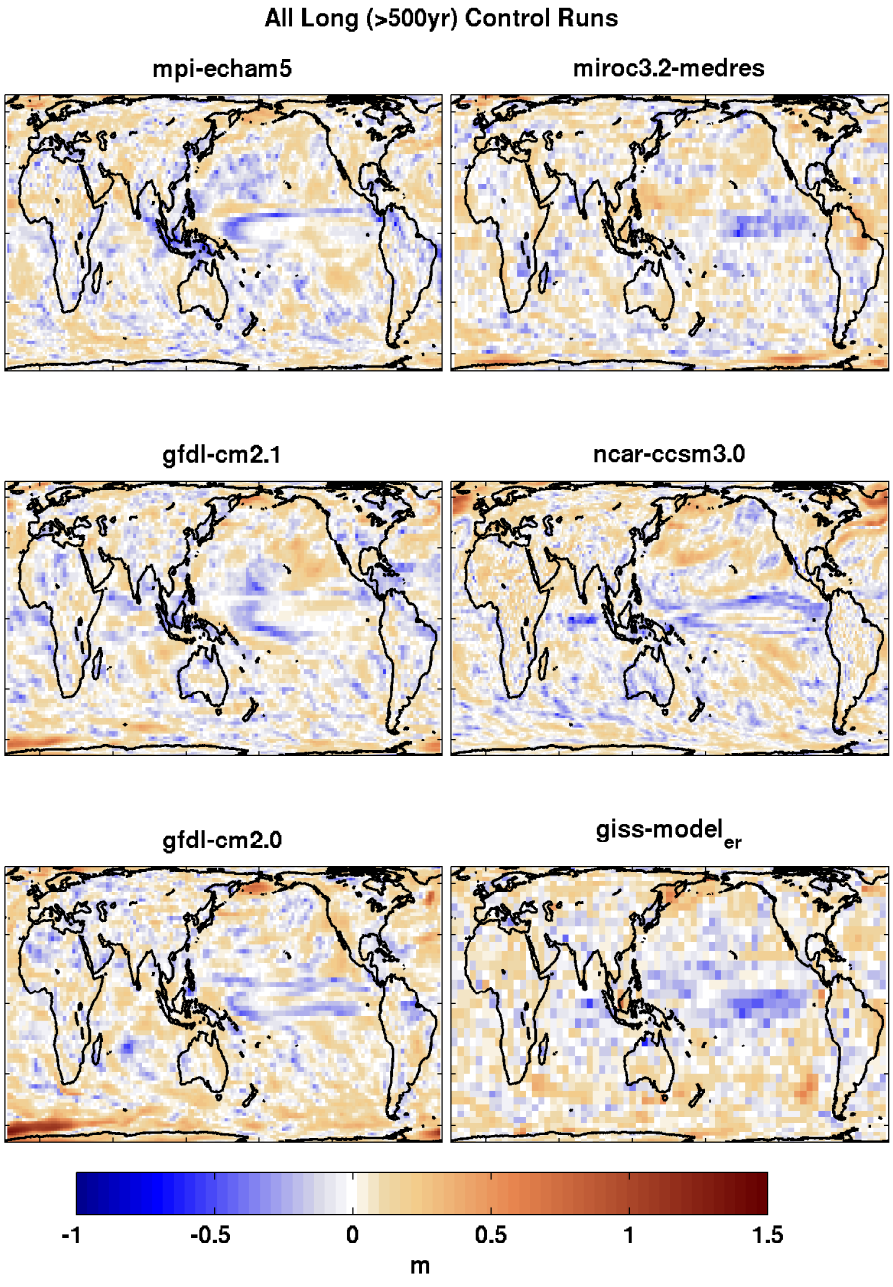

Figure B.12: Maps of m for each of the individual pre-industrial control simulations used in Figure B.11.

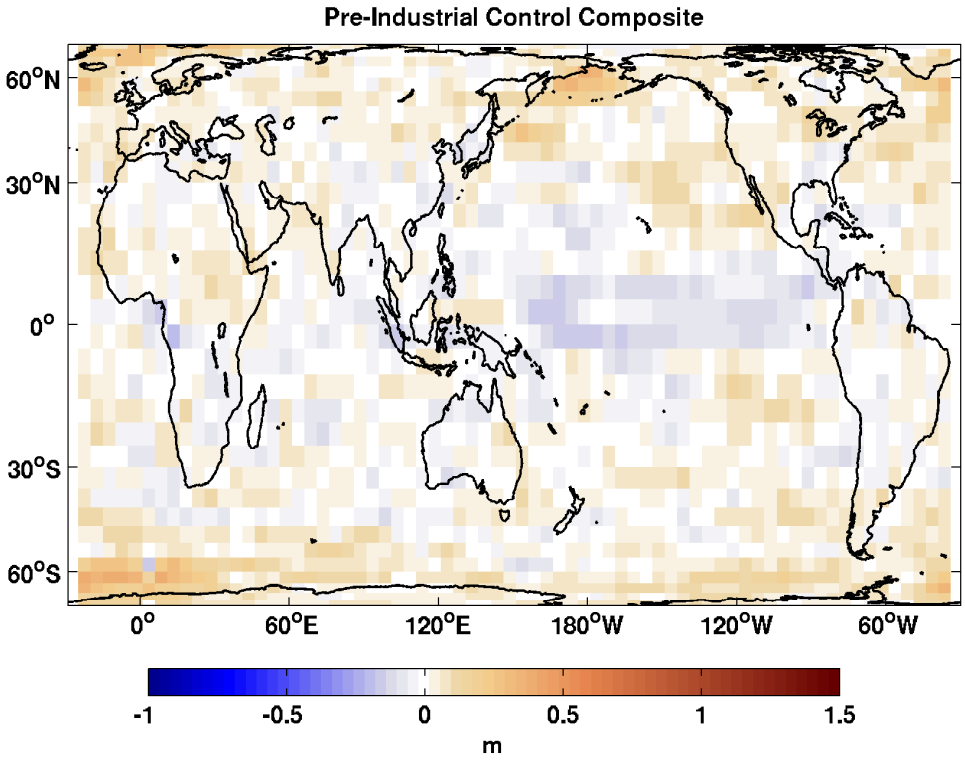

Figure B.13: Mean value of m averaged across 6 long (>500 year) pre-industrial control simulations.

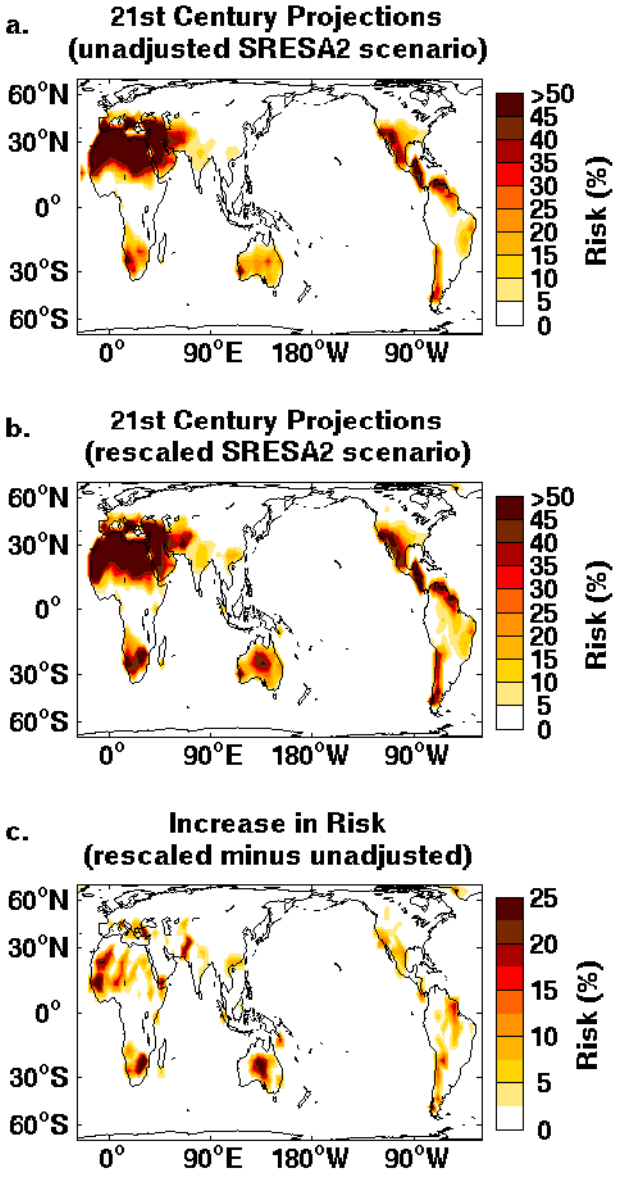

Figure B.14: Same as Figure 3 in the text, but using the SRESA2 scenario: **a.** Risk in unadjusted SRESA2 annual precipitation projections. **b.** Risk in rescaled precipitation projections. **c.** Difference between **a** and **b**.

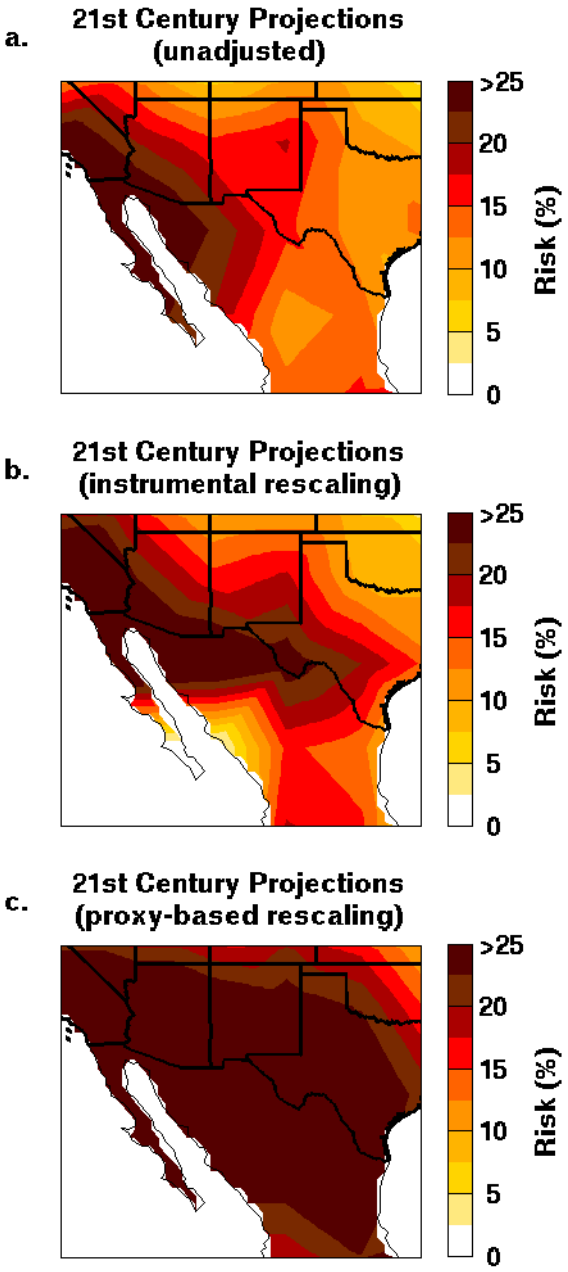

Figure B.15: Risk of a 21th century megadrought in the US Southwest (same as Figure 4 in the text, but for the SRESA2 scenario). **a.** Risk in unadjusted SRESA2 annual precipitation projections. **b.** Risk in precipitation projections rescaled from instrumental values of m. And, **c.** risk in precipitation projections rescaled using our proxy-based estimate of m ($m = 0.4$).

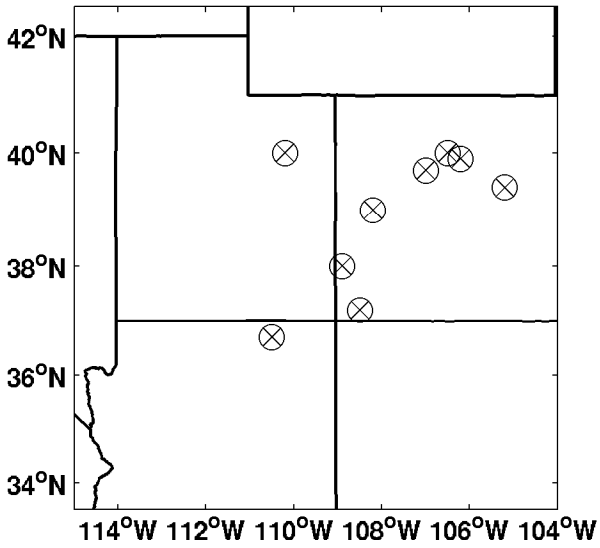

Figure B.16: Map showing the location of each site where tree-ring samples were taken and used in our Colorado River reconstruction. Data are all available through the NOAA Paleoclimatology pages (www.ncdc.noaa.gov/paleo/treering.html)

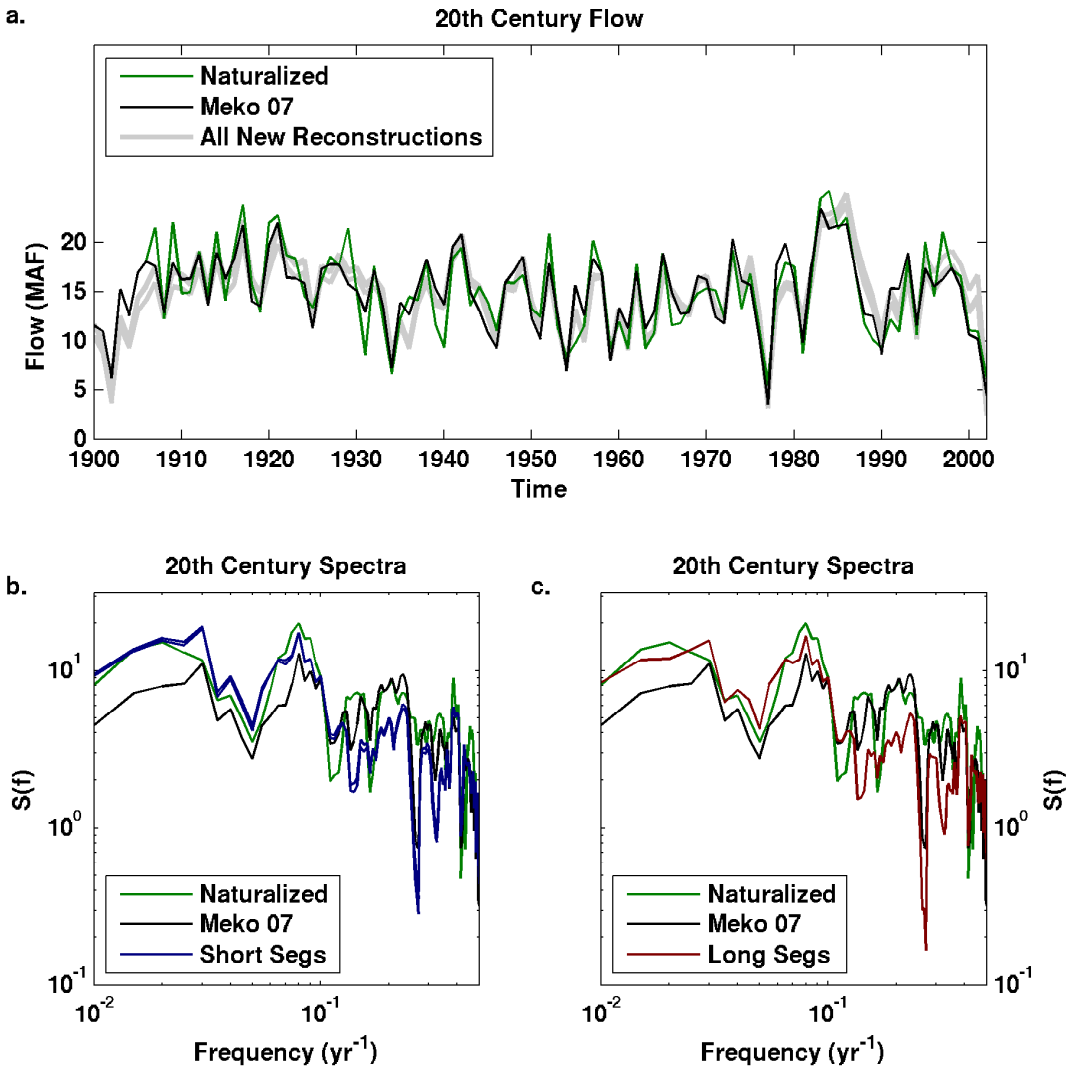

Figure B.17: Naturalized (e.g., observed) and reconstructed flow during the 20th century. **a.** Time series of the Colorado River at Lee's Ferry (green), the M07 reconstruction (black), and the new suite of reconstructions (gray). During the 20th century, they are all in very good agreement. **b.** Power spectra of naturalized and reconstructed flow using short (>250 year) segments for the time period when the naturalized and reconstructed records overlap (the 20th century). **c.** Power spectra of naturalized and reconstructed flow using long (>400 year) segments for the 20th century only.

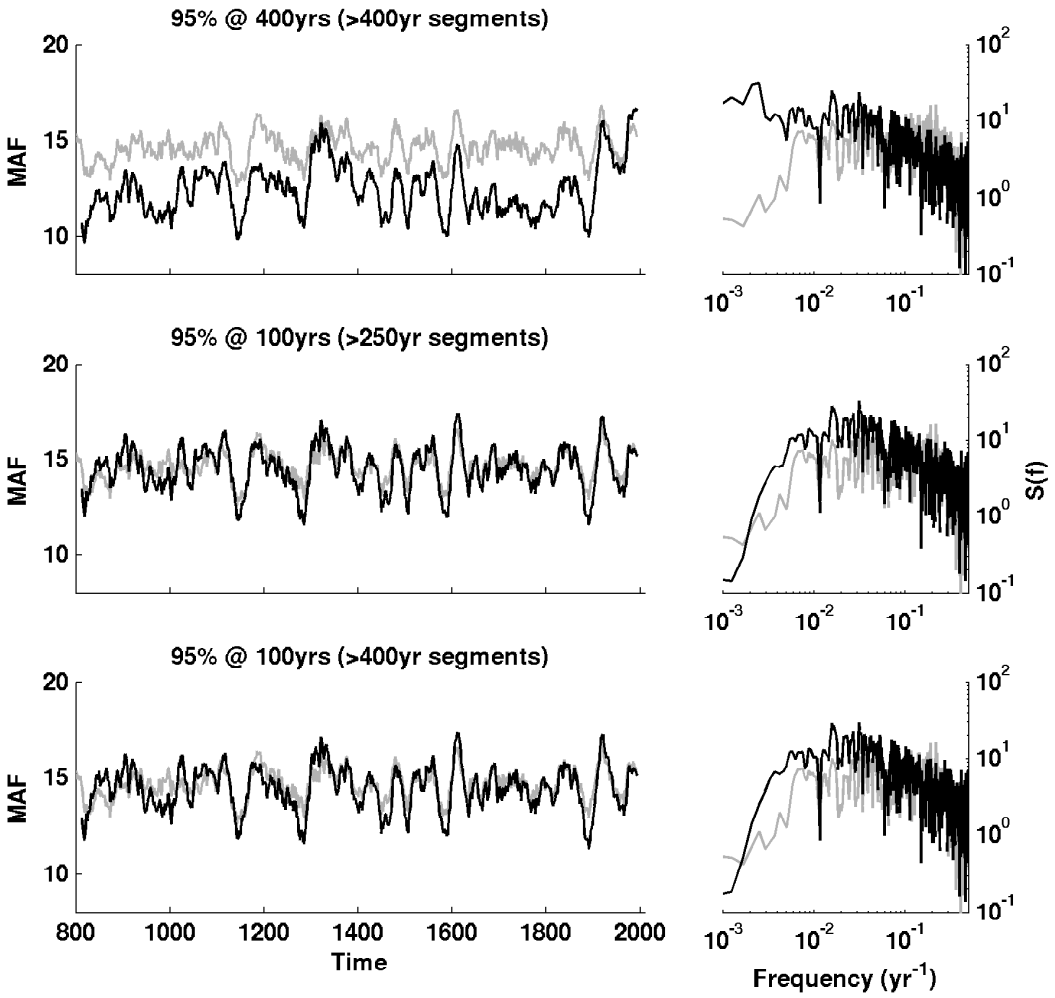

Figure B.18: **Top:** (left) 25 year mean of reconstructed Colorado River flow from M07 (gray) and from our new reconstruction (black); (right) power spectra of the M07 reconstruction (gray) and our new reconstruction (black). In our reconstruction segments were only included if they were at least 400 years long and were deterended with a conservative spline designed to remove 95% of the variance at periods 400 years and longer. **Middle:** Same as top, but for our reconstruction (black) using short segments (>250 year) and flexible deterending. **Bottom:** Same as Top, but for flexible detrending applied to the network of long segments. The network of samples used in the reconstructions on the top and bottom, therefore, were exactly the same (see Table 1).

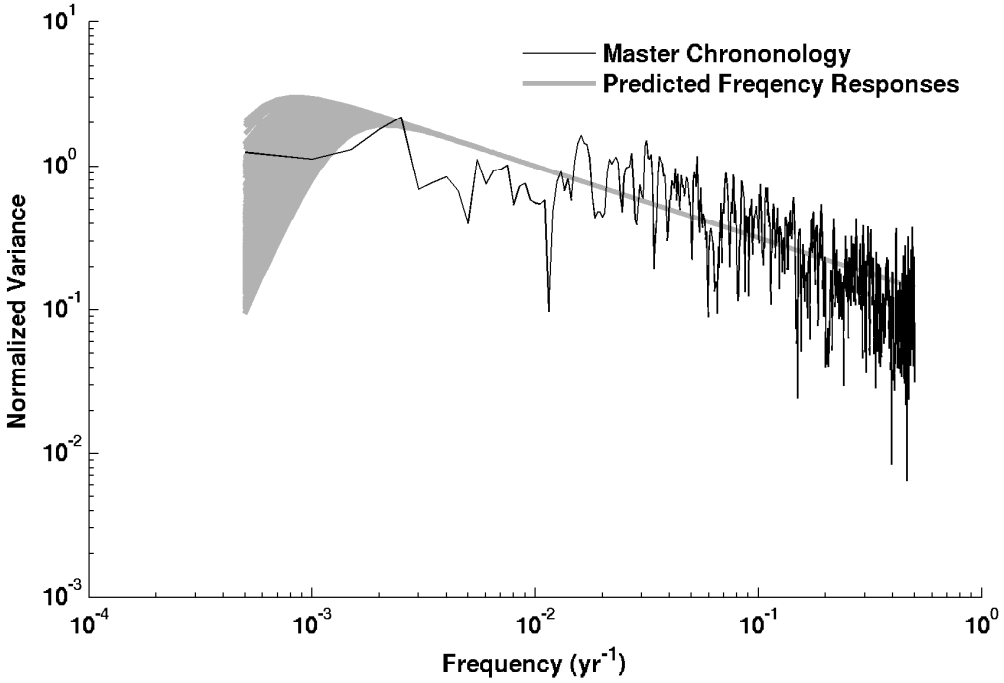

Figure B.19: Power spectrum (black line) of the raw master chronology used in our reconstruction. This master chronology is built from long segments (>400 years) that were detrended with a conservative spline designed to remove 95% of the variance removed at periods 400 years and longer. The gray lines show the expected frequency response from detrending segments who are assumed to follow a power law ($1/f^{-m}$) distribution with m set to 0.5. This analysis shows that our detrending methods are expected to preserve variance up to the millennial time scale, and that the master chronology generally exhibits more variance at lower frequencies.

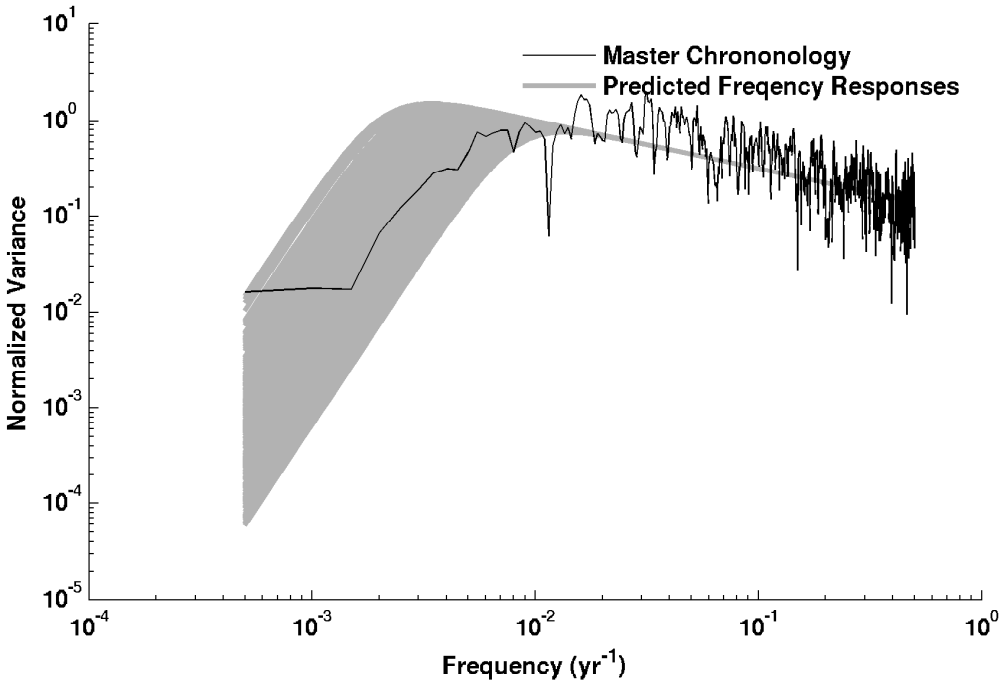

Figure B.20: Same as Figure 4, but for the power spectrum (black) the master chronology used in the reconstruction summarized in row "b" of Table S1. This chronology was built from short segments (>250 years) that were detrended with a spline designed to remove 95% of the variance at time scales longer than 100 year. The figure demonstrates that the use of shorter segments and flexible detrending is expected to damp variance at periods approximately 100 years and longer, which in turn would bias the calculation of the parameter m.

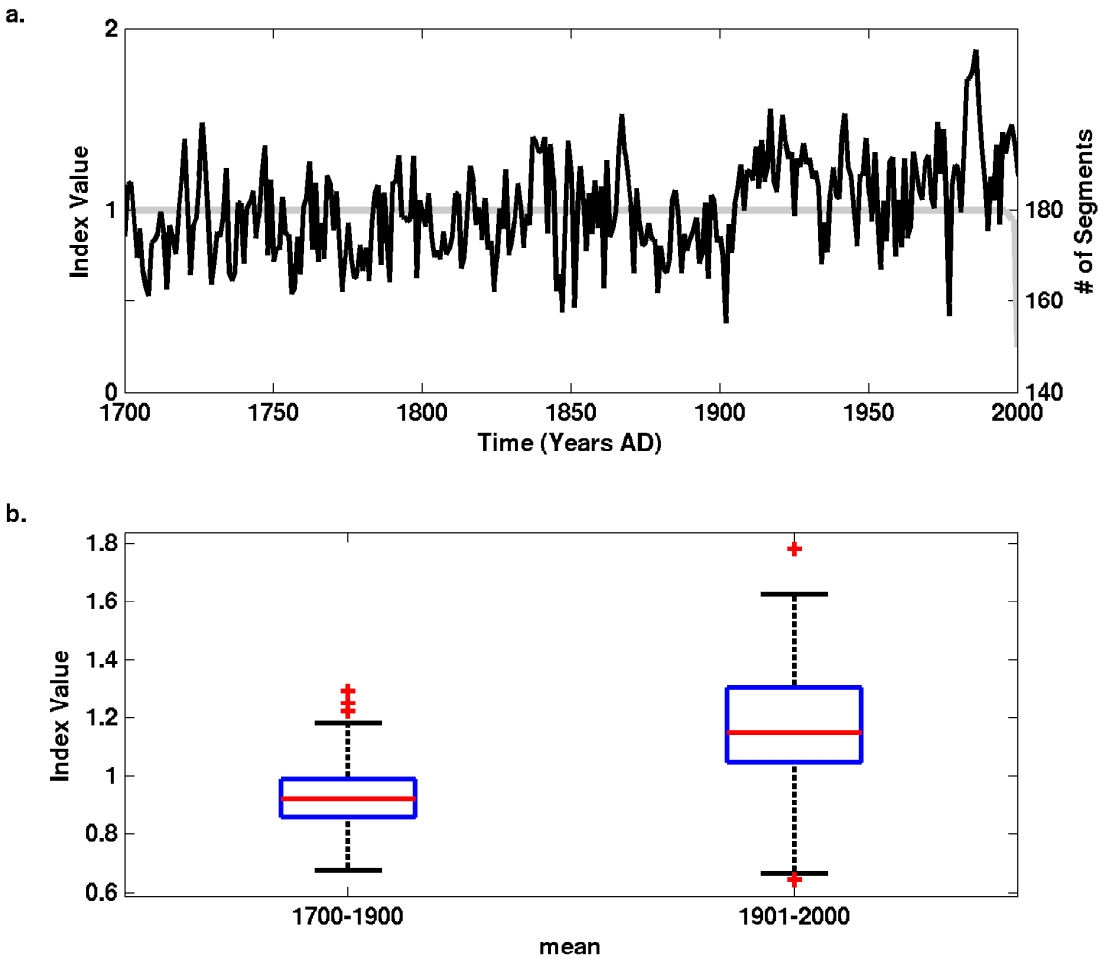

Figure B.21: **a.** Mean index value (black line) of 400 year-long segments with at least 90 years of data during the 20th century. The number of segments used to make the black line is shown in gray. **b.** Box plots of the population of means of each ring-width index before and after 1900. The red line on the box plot indicates the median, and the box encompasses the upper and lower quartiles. The whiskers show the 95% limits of the distribution of the data, and the "+" symbols indicate outliers.

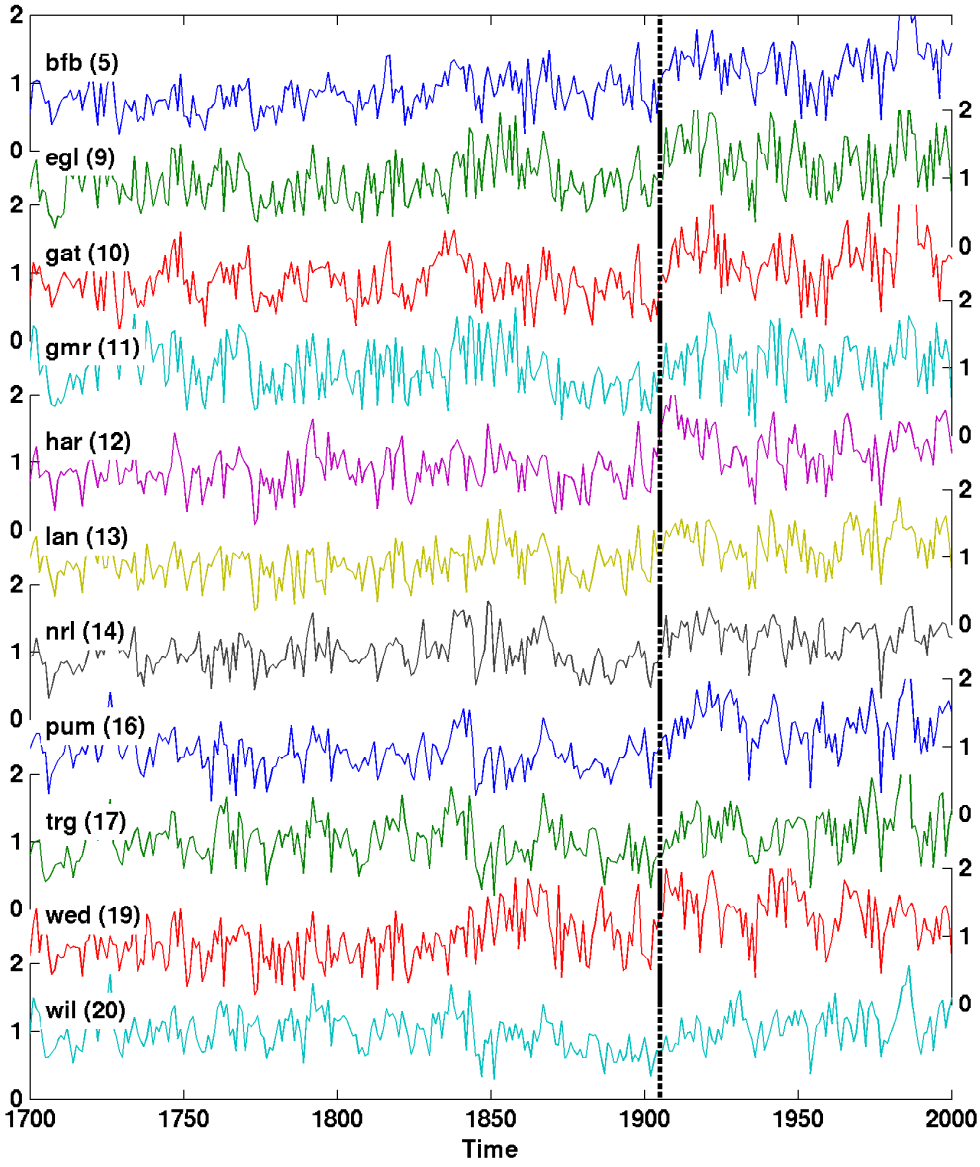

Figure B.22: Time series of new chronologies that were re-built from only segments with at least 400 years of data and 90 years of coverage during the 20th century. The name of the chronology is printed in bold with the number of segments used in parenthesis. The dotted line marks the 1905 shift present in our Colorado River reconstruction. These findings further substantiate the result that the early 20th century change is real and driven by climate.

APPENDIX C

MONTE CARLO AGE MODELS

Novel tools for analyzing paleoclimate data and their application to a new speleothem record of Southwest hydroclimate during the last 1500 years.

Toby R. Ault[1], Ali Kimbrough[1], Julia E. Cole[1], and Sarah S. Truebe[1]

1. Dept. of Geosciences, Univ. of Arizona

C.1 Abstract

Understanding the full range of past hydroclimatic variability requires multiproxy reconstructions to address the full spectrum of variance. As a complement to regional tree-ring records, we have developed a new millennium-length δ^{18}O record from a cave in Southern Arizona. Comparing this record with absolutely dated dendroclimatic reconstructions requires that we quantify dating uncertainties. Here we develop a new Monte Carlo approach to model time uncertainty of radiometrically dated records, whereby "families" of age models are generated from a given set of assumptions, then applied to the δ^{18}O series. Results indicate that traditional approaches of using a single linear or spline-based age model may be misleading because these models are, in effect, single realizations of time depth curves drawn from a much richer probability distribution. As an alternative, we propose a method for generating stochastic models of age uncertainty. We also develop a method for correlating entire families of age models with single time series, and apply this method to correlate our new δ^{18}O record with annual tree-ring proxies of hydroclimate in the region. We find that variance at long timescales (multidecadal and longer) is much less sensitive to our age-modeling choices than variance at shorter timescales. The

new speleothem record confirms that several well-known megadroughts (and pluvials) extended well into the Sonoran desert, and that there has been a long-term trend towards wetter conditions during the last 1500 years.

C.2 Introduction

The southwestern U.S. has been one of the fastest-growing regions of the country during the last few decades. To cope with the increasing demand on regional water resources, managers require realistic estimates of future water availability. This entails integrating the natural range of hydroclimatic variability with projections of the future. Since the instrumental record alone is not long enough to reliably estimate the natural range (e.g., Cook et al., 2004), researchers have turned to a variety of high-resolution proxies during recent years to develop regional climate histories of change during the last few centuries. Tree-ring records have been widely used in this capacity to reconstruct a wide range of hydroclimate variables, most notably Colorado River flow during the last 1200 years (Meko et al., 2007) and the Palmer Drought Severity Index (PDSI) (Cook et al., 2004). In a now-classic paper reviewing a large number of hydroclimate proxies from the west, Woodhouse and Overpeck (1998) showed that droughts such as the 1930's Dust Bowl and 1950's drought occurred once or twice per century since 1600AD, and that multidecadal "megadroughts" occurred once every 500 years during the last few millennia (Woodhouse and Overpeck, 1998).

Despite the pressing need for reliable records of paleodrought in the region, there are few continuous, high-resolution records of regional hydroclimate that span the last millennium. This is partly because tree-ring reconstructions are limited to the higher-elevations where trees can grow, and partly because the length of tree-ring records is limited by the lifespan of trees. Moreover, dendroclimatic reconstructions are absolutely dated and linked to hydroclimatic variability with exceptional statistical rigor, but they likely miss ultra-low frequency (century-scale and longer) variability due to the "segment-length curse" described by Cook et al. (1995). A

complete picture of the full range of drought variability requires that paleorecords sensitive to low-frequency variability be developed to complement the dendroclimatic and instrumental perspectives. However, there are very few lakes in the Southwest that could yield sediment-based histories of the last few millennia. Some insights have emerged from packrat middens (Holmgren et al., 2003) and cave deposits (Asmerom et al., 2007; Polyak and Asmerom, 2001; Rasmussen et al., 2006), but linking these to precisely-dated annual tree-ring reconstructions is a notable challenge (Betancourt et al., 2002; Tan et al., 2006). Nonetheless, a recent analysis of instrumental and paleoclimatic records from the western U.S. argues that the natural range of low-frequency variability is considerable and needs to be constrained for future adaptation and mitigation strategies (Ault et al.).

In arid regions, changes in speleothem oxygen isotope ratios reflect changes in the hydrologic budget at the surface. Radiometric dating methods provide precise constraints on the progression of isotopic changes. This should make speleothems ideal proxies for recording the timing, magnitude, duration and severity of paleodroughts during the past. Ideally, we would want to compare and calibrate the isotopic variability in cave records with instrumental and tree-ring records of hydroclimate to build confidence in our interpretations of climate variability inferred from speleothems. This poses a challenge because even well-dated, high-resolution cave records of climate lack the age precision of instrumental and dendroclimatic records. We require a method to quantify and constrain the impact of age model uncertainty on such comparisons.

Here, we approach the challenge of comparing cave-based and tree-ring based climate reconstructions, using a new interannual $\delta^{18}O$ record from an Arizona cave spanning the past 1400 years, and a new set of methods for age model development. Our approach explicitly recognizes that single age-depth curve fits do not fully represent the potential range of age-depth models that are possible for a given suite of radiometric dates (and their associated uncertainties). We offer two alternative methods: a set of statistical spline-based fits of varying stiffness to age-depth points (as in Blaauw, 2010), and a statistical approach that simulates annual speleothem

accumulation as a stochastic process constrained by explicit assumptions and by the radiometric dates. We represent our δ^{18}O record as a distribution of plausible time series given our age model uncertainty. This process allows us to compare our family of reconstructions with dendroclimatically based records and more rigorously assess whether, for example, particular extreme intervals are truly coeval. We conclude that the cave record does track multidecadal wet and dry anomalies, a finding that holds promise for extending this cave record back in time.

C.3 Methods

In 2002 we removed a 19.5cm-long stalagmite formation from "Fort Huachuca Cave" (FHC) ($32°N$,$111°W$). The formation was then cored with a 48mm-diameter drill bit, and the center portion of that core was sliced into a 5mm thick slab. The top 146mm of the center portion of that slab was divided along the apparent axis of growth, so that we could use one side (side "L") for our radiometric (U/Th) dating and the other (side "R") for our isotopic analysis (Figure C.1). Consecutive 500μm samples of the carbonate material were drilled from side "R" using a micromill. Isotopic analysis of these carbonate samples was conducted on a Micromass Optima dual inlet stable isotope mass spectrometer at University of Arizona (with <0.08 ‰ analytical precision for δ^{18}O). Eight pieces of material from various depths (Table C.1) were sent to Oxford University for U/Th dating on a Nu-Plasma MC-ICP-MS.

In 2007, we re-molded the cored formation and restored it to its original location inside the cave. Since 2007, we have visited the cave periodically to check the surface of the formation for any new calcite growth and collect water samples for isotopic analysis.

C.3.1 Monte Carlo Age Models

We develop a Monte Carlo age modeling approach that allows us to use the uncertainty in our U/Th to produce "families" of age models (Table C.1). First, we assume that the U/Th dates we obtained were normally distributed around the

mean date with the standard deviations given by the 1σ error bars of the U/Th dates (Table C.1, right column). Next, we develop spline-based Monte Carlo families of age models by drawing random numbers from each dates' probability distribution function (PDF), then fitting a spline through those randomly selected dates. This approach is quite similar to the one employed by Blaauw (2010). All age models are reported in "Years AD" for comparison with tree-ring records.

The "stiffness" (flexibility) of a spline is specified *a priori* using a single parameter, p (usually called the "spline parameter"). We therefore generate three different families of spline-based age models using different values of this parameter. To generate the first family, we use $p = 0$, which is equivalent to fitting a linear model to each Monte Carlo draw of U/Th dates . We generate a second family of spline-based age models using wigglier splines ($p = 10^{-5}$) that are sufficiently flexible to pass through the distribution of each U/Th date, which keeps growth rates from rapidly changing but sometimes prevents age models from being able to fit through the mean U/Th date of each depth. The final family of spline-based age models is generated using a more flexible spline parameter ($p = 10^{-4}$), which we selected to ensure that the splines would be flexible enough to fit through the midpoint of all dates. Increasing p further would lead to age reversals between dates, which we deemed unacceptable.

Although widely used, spline-based age models may be of limited utility because they do not necessarily reflect the actual physical processes that govern formation growth. They also do not allow for the possibility that growth rates may vary irregularly between points where dated material is available. An alternative approach (e.g., Ramsey (2008)) is to model accumulation of the material in a proxy archive as a stochastic variable that is constrained by the available radiometric dates. To develop such stochastic models of growth for Fort Huachuca Cave, we begin by treating time as a random variable as (X_d) whose distribution we want to model as a function of depth. At the points where we have U/Th dates, we expect the values of X to be normally distributed with means and variances specified by the dates we have: $\{X_k \backsim N(\mu_k, \sigma_k^2);\ k = 1, 2, \ldots, H\}$ where H is the total number of dated

pieces of material (in this case, the samples reported in Table C.1.

We now consider the distribution of possible ages at the first sample (X_i) located at a depth (d_i) directly above a dated piece of material (X_k). We assume that the distribution of possible ages corresponding to this piece of material (X_i) will depend on the age of the previous piece of (dated) material (X_k) and the addition of some known quantity of time. That is $X_i = X_k + Y$, where Y is a random variable whose distribution and parameters we will discuss below. We also assume that the point d_l directly below a piece of dated material will have a distribution that depends on the date at X_k *minus* some quantity of years: $X_l = X_k - Y$.

Under these assumptions, an age model can be thought of as a random walk between two dates (X_a and X_c) at two known depths (d_a and d_c), where a and c are the depths of dated pieces of material, and between those two points $d_a < d_j < d_c$ and therefore $X_a < X_j < X_c$. The greatest age uncertainty would be expected to occur at the midpoint between those dates. That is, if b is the midpoint (in depth) between two samples, then:

$$X_j = \begin{cases} X_{j-1} + Y, & \text{if } a < j < b; \\ X_{j+1} - Y, & \text{if } b \leq j < c. \end{cases} \tag{C.1}$$

To implement this framework numerically between each piece of dated material, we generate two suites of "random walkers." The first suite is comprised of "forward random walkers," which are generated from the first part of equation C.1. The second suite, the "backward random walkers," are generated from the second part of the equation. Finally, to adjoin the forward walkers with the backward walkers, we use a "Frankenstein" approach whereby the individual members of each suite are sutured to a nearby neighbor. This is achieved at the depth d_{b-1} by drawing a value from Y, adding it to the forward walker's value at X_{b-1} and determining if there is a backward walker whose value at X_b is less than or equal to the value drawn from Y and added to X_{b-1}. If no suitable backward walker is available, another backwards random walk is generated and the routine is repeated. Note that by adjoining the forward random walker to the backward random walker in

increments that are randomly selected from the distribution of Y, the distribution of ages between each depth remains the same between the two endpoints.

We need to select an appropriate distribution (and parameters) for Y. Ideally, we would be able to use some prior information about the processes that cause carbonate to precipitate to inform our choice. In reality, we do not completely understand the controls on carbonate precipitation or how these processes may vary on decadal to millennial timescales. Nonetheless, we can make several general requirements using intuition and the limited number of observations we have from ongoing monitoring efforts in Fort Huachuca Cave and other caves in southern Arizona. First, we note that carbonate does not precipitate during some years, but accumulates rapidly during others. Second, we do not expect there to be any age reversals because this would entail older carbonate material being contained in younger material; this unlikely and would be visually apparent. Finally, we note that the average growth rate, no matter how variable, must be equal to the linear slope between two depths.

The above stipulations on our distribution lead us to select a gamma distribution ($Y \backsim \Gamma(k, \theta)$), from which we will draw values to generate random walkers. We choose the gamma distribution because it is continuous and, depending on the "shape" (k) and "scale" (θ) parameters, will allow the distribution of years to be skewed towards zero most of the time, but occasionally larger. Very similar results could be (and were) obtained using a log-normal distribution or a Poisson distribution. Ultimately the specific type of distribution is less important than the parameters selected for it.

Since we want the mean of our Gamma distribution ($E(Y)$) to be equal to the linear age-depth slope (μ) between two pieces of dated material, we select k and θ by introducing another parameter m, such that: $\theta = \mu/m$ and $k = 1/m$. This preserves the mean slope because $E(\Gamma(k, \theta)) = \frac{\theta}{k}$ for the gamma distribution. The value of μ we estimate as the linear slope between two points, while the parameter m, which governs the variance of the age models, must be estimated more subjectively. We constrained values of with a "leave-one-out" approach in which we modeled ages between two dates and determined what values of m would allow the age-depth

curve to pass through a third, intermediate-age, date C.3. Clearly, if the value of m is too low, the models will be too linear, growth rates too stable, and the models would not generally be variable enough to encompass the range of dates we have. On the other hand, if m is too large, the distributions of years will be extremely skewed, which will lead to very blocky growth rates (long hiatuses followed by rapid growth). We picked values of $m = 1, 3$ and 5 to encompass a range of realistic growth rates and variances for this site, and used these values to produce three distinct gamma-based families of age models.

C.3.2 Correlation Analysis

To compare our record with tree-ring records, we calculate the linear (Pearson's) correlation coefficient between all members of each Monte Carlo family (both the spline-based families and the gamma-based families) and 25-year averages of recently developed local tree-ring chronologies. These chronologies are 355 year indices of summer and winter precipitation that have been developed from stands of trees within a few kilometers of Fort Huachuca cave, and they closely track 20th century variability (D. Griffin, personal communication). We also correlate the different age model families with 25-year running averages of the reconstructed Palmer Drought Severity Index (Cook et al., 2004) for the 5x5 degree grid point corresponding to our site. This reconstruction spans 1AD to 2003AD (2003 years total). We estimate the significance (p-value) of each individual correlation coefficient by approximating the degrees of freedom as $N' = N/w - 2$, where N is the number of years when the tree-ring data and the cave record overlap, and w is length of the smoothing window (25 years, in this case). This is equivalent to the number of non-overlapping 25-year windows. For the summer and winter precipitation chronologies, N' is 14; for the longer PDSI reconstruction, N' is estimated as 60. Since we are using an *a priori* significance criterion for our correlations, we would expect approximately 50 spurious significant correlations to occur at the 95% confidence limits.

C.3.3 Forward modeling and spectral analysis

To evaluate the continuum of hydroclimatic variability and the potential for periodic low-frequency forcing mechanisms, we calculated the spectra of each member of each age model family using the Lomb-Scargle method(Lomb, 1975; Scargle, 1982). We compare these spectra with results from an earlier study that simulated dripwater δ^{18}O using a two-layer model of "soil" (from the "leaky bucket" model proposed by Huang et al. (1996)) and "rock" (simulated as a well mixed reservoir). Herein we will refer to this model of isotopes in caves as the "isocave" model. The study found that even without low-frequency climate input, the combination of stochastic precipitation input and storage created variability at interannual and decadal time scales. The ratio of rock-to-soil (rock:soil) governs the shape of the modeled power spectrum and determines the point at which the cave record was expected to be relatively free from local "storage" effects. Here we compare the spectral densities of our new cave δ^{18}O time series with the median spectral densities of the earlier studies' 1:1, 2:1, 3:1, and 5:1 rock:soil experiments (black lines in Figure 2 of Truebe et al. (2010)). No adjustments were made to the parameters of the isocave model, nor the input data used to drive the model, because the earlier results are applicable to the climates and caves of Southern Arizona (i.e., it is not a site-specific model).

Spectral peaks associated with the Fort Huachuca record are deemed "significant" if they meet both of the following criterion. First, they should rise above the level of variance expected to occur as a consequence of cave processes predicted by the isocave model. Second, they should be relatively insensitive to the different age modeling choices we make. To reconstruct the temporal evolution of the timescales of variability that satisfy these criteria, we use singular spectrum analysis (SSA) (Ghil et al., 2002). SSA relies on estimating temporal patterns of variability from the lag-covariance matrix of a time series, and requires a user-defined "lag-window" (w), which we set to 100 years. However, we found that the leading reconstructed components (RCs) of variability were insensitive to this choice.

C.4 Results

The mean of our δ^{18}O record is -8.0 ‰(VPDP). We performed a "Hendy Test" (Hendy, 1971) and found that $\delta^{18}O$ and $\delta^{13}C$ did not covary along growth horizons. We further show that $\delta^{18}O$ and $\delta^{13}C$ do not generally covary through time (Figure C.2). Taken together, these two pieces of evidence suggest that the calcite formed under equilibrium conditions (Mickler et al., 2006). However, instrumental measurements of cave dripwater δ^{18}O from various excursions into the cave from 2007-2010 have yielded a value of -10.0 ‰VSMOW across several sites throughout the cave, and with almost no variability within or between years. Since the average annual temperature of the cave is 17°C, calcite precipitating in equilibrium would be expected to have a δ^{18}O value of -10.7 ‰VPDB (using the fractionation factors provided by Kim and O'Neill (1997)), hence there is an offset between the long-term mean value of cave calcite δ^{18}O and the predicted value from modern measurements. This offset is nearly the same as that reported in Wagner (2006) and in several other nearby caves currently being monitored (S. Truebe, personal communication).

C.4.1 Age modeling

Based on the full range of age model families, our new δ^{18}O record spans approximately the past 1500 years. The linear, spline-based ($p = 0$) family of Monte Carlo age models for the Fort Huachuca Cave record is shown in Figure C.4a, and the corresponding family of time series (one for each age model) is shown in Figure C.4b. The average rate of carbonate accumulation for the duration of the record was approximately 113μm/yr, which we calculated as the slope of the least squares regression line through our U/Th dates and their corresponding depths. However, this rate varies slightly, and during the most recent interval (from 1400AD onward) it was approximately 62μm/yr, whereas in the earlier portion of the record is was closer to 138μm/yr. In terms of the 500μm samples we collected, this means there are 4.4 yr/sample overall, 8.0 yr/sample during the most recent period, and 3.6 yr/sample during the earlier period. Clearly, a single "line of best fit" does not

capture the full range of growth rate variability in our record.

We turn next to the other two spline-based families of models (Figures C.5 and C.6). Both of these are able to accommodate the variability in U/Th dates, and this variability is reflected in the greater spread in the time series of δ^{18}O. The distributions of both of these families are in very good agreement with those developed using the "classical age modeling" software described by Blaauw (2010). However, the bends between points (indicative of changes in the growth rate) occur according to the spline parameter. For instance, we have U/Th dates at 142mm and 122.75mm that are dated at 625 ($\pm$11.5 yr) AD and 888 ($\pm$10.0 yr) AD, respectively. The spline-based families indicate that growth rates would have been fastest nearest these dates, and slowed down for some time in between them. However, this necessarily reflects the characteristics of the spline, and not any information pertaining to the characteristics of the record itself.

Finally, we examine the three families of stochastic (gamma-based) age models (Figures C.7, C.8, and C.9). As expected, the distribution of age models is greatest between points and for the larger values of m (panel **a.** of Figures C.7, C.8, and C.9). This translates into large uncertainties in the timing of individual events in the time series of δ^{18}O (panel **b.** of Figures C.7, C.8, and C.9)

C.4.2 Correlation Analysis

Figure C.10 shows the distribution of correlation coefficients between 25-year averages of several tree-ring indices of regional hydroclimate and all members of all age model families for the Fort Huachuca Cave record. The number of significant correlations is reported in Table C.2. Since we are using the 95% confidence limits to assess significance, we would expect 50 out of the 1000 correlations to be significant just by chance. For summer and winter chronologies (which are 355 years in length), the number of significant correlations is inconsistent depending on whether the spline-based families are considered or the gamma-based age models are considered. In the former case, winter correlations have a greater-than-expected number of significant correlations when the spline parameter (p) is zero or when $p = 10^{-5}$. However, we

observe a greater-than-expected number of significant correlations between summer chronologies and δ^{18}O when the more flexible spline parameters are used, or when the gamma-based age models are used. In all cases except for the spline-based linear model, however, the correlations with PDSI rise well above expectation (Table C.2, last column).

C.4.3 Power Spectra

The spectra calculated from each member of each family are shown in Figure C.11 (gray lines). For the linear model, there is almost no difference between the spectra individual models. As the splines are allowed to be more flexible, there tends to be greater disagreement between models at higher frequencies (left column, middle and bottom panel). This effect is even more pronounced for the gamma-based models. In all cases, however, the agreement between models is highest at the lowest frequencies. These frequencies are also more energetic than expected from our forward modeling experiments (Figure C.11), and hence we deem them significant.

Since the lowest frequencies characteristics of the Fort Huachuca Cave record are relatively insensitive to our choice of age model, we show the temporal evolution of these frequency components using the mean age model from the flexible spline-based family. Using just one age model makes it easier to depict the low-frequency behavior through time, and we are only considering timescales that are insensitive to age-model uncertainty, which means that the temporal evolution of these low-frequency components does not vary too much between individual family members or across families. The first three reconstructed components (RCs) from SSA exhibit variability at 1750, 140, and 90 year timescales. We sum these components together to produce a filtered δ^{18}O time series in Figure C.12. The filtered times series highlights a long-term trend towards more-negative values, and enhanced variability from 1000AD through 1500AD.

C.5 Discussion

We have developed a millennial-scale record of isotopic variability from a mid-elevation cave in the Sonoran Desert of Southern Arizona. We applied the standard test for disequilibrium effects in the record, and did not find any evidence that such effects might be present anywhere in the core. Nonetheless, there is an offset between the predicted value of δ^{18}O (-10.7 ‰) that we calculated from modern dripwater δ^{18}O and the long-term mean value of -8.0 ‰. If calcite is indeed precipitating in equilibrium, then the offset may reflect a difference between the temperature of the water at the formation's surface (where the calcite forms) and the mean air temperature of the cave. Alternatively, the offset could simply reflect a difference in the long-term mean from modern dripwater $\delta^{18}O$, or there could be disequilibrium effects that we are unable to detect using standard methods.

In a typical study of high-resolution proxies with age model uncertainty, it is common to develop a single age model based on a limited number of dated samples, then apply this model to the remaining samples and make interpretations. Our results suggest that this approach may be misleading for two reasons. First, spurious correlations may arise if only one model is used out of the many thousands of possible models given the uncertainty in the dated pieces of material. Second, in our case (and presumably others) we lack sufficient information to select one age modeling approach over another. Although we can probably rule out the spline-based "linear" model as a reasonable method (since it fits the U/Th dates so poorly), we cannot tell if a flexible spline is better than a stiff spline. On the one hand, if we do not completely trust our U/Th dates and we expect growth rates to be approximately linear, then a somewhat stiff spline might be a good choice (e.g., in Figure C.5). On the other hand, if we have high confidence in our radiometric dates, we would choose a more flexible spline, although this would necessarily introduce variability in growth rates between dates that was not mediated by any physical influences, but instead by the statistical characteristics of the spline itself (as in Figure C.6).

As an alternative to the spline-based age models, we have also presented families

of age models that are generated by assuming that time uncertainty at each depth increment is a function of a random value drawn from a pre-defined probability distribution function and added to (or subtracted from) the date of the adjacent depth. In this case, the variance of the individual models is greatest between U/Th dates, and it is highly sensitive to our choice of parameters for the distribution. We therefore have shown the results from several different parameter choices.

Uncertainties in time translate to uncertainties in the spectral estimates of individual models, which can clearly be seen in Figure C.11. When the age models are all constrained by a liner best fit, this effect is small. When more flexible splines are used, this effect gives rise to spectral density estimates that differ by 1-2 orders of magnitude, and it is even more dramatic when the stochastic (gamma-based) models are used. However, the variance estimates at longer timescales (> 100 years) are remarkably consistent within each of the six families, and even across families. This finding implies that not only the timing of low-frequency events, but also the variance of low-frequency events, may be more robust than events occurring at higher frequencies.

Despite the uncertainty we have depicted, there are several findings that appear to be insensitive to the different age modeling choices we have described. First, regardless of the choice of the parameter m, the distribution of correlations with tree-ring indices of summer rainfall, winter rainfall, and the PDSI are all remarkably similar (Figure C.10). In contrast, this is not the case for the spline-based age models. Second, the limited number of degrees of freedom (14) leave some ambiguity in the significance of the correlations with chronologies from adjacent mountains (summer or winter). The number of significant correlations between reconstructed PDSI, however, exceeds the number of significant correlations expected by chance for all families (except for the linear/spline-based family). These significant correlations are all negative and cluster near a value of -0.3, which would suggest that drier conditions inferred from tree-ring widths are associated with more enriched δ^{18}O values. This finding is consistent with previous studies of the relationship between winter precipitation δ^{18}O and precipitation amount (e.g., Wright et al. (2001);

Wagner et al. (2010)).

To show the most robust features of our data more clearly, and compare them with the tree-ring reconstructions of PDSI over the last millennium, we generate singe δ^{18}O averages from each family (excluding the poor-performing linear spline-based family). These time series are generated by first averaging the δ^{18}O values of each member in each family to evenly-spaced 25-year bins, then averaging these evenly-spaced 25-year averages across all members of a given family. The results are shown in Figure C.13 (gray lines), and are extremely similar to the SSA-filtered results shown in Figure C.12. From about 500AD to 1400AD there is a general trend towards more negative δ^{18}O values, and trend in reconstructed PDSI towards more positive values. Both of these trends suggest gradually increasing moisture. After about 1400 AD, the speleothem δ^{18}O and the PDSI reconstruction flatten out on average, but exhibit greater multidecadal variance than earlier. This variance includes multidecadal dry intervals (megadroughts) but also periods of increased moisture (pluvials). The megadroughts occurred around 630AD, 1150AD, and 1450AD – all well-known periods of persistent aridity. The wettest interval appears to have occurred during the 14th century, which again is consistent with previous studies of decadal-scale pluvials (e.g. Meko et al., 2007).

C.6 Conclusion

We have presented a new, high-resolution record of δ^{18}O from Fort Huachuca Cave in Southern Arizona. Even with the precise U/Th dates available for this record, age model uncertainty makes it extremely difficult to understand interannual- to decadal- scale climate variability in the Forth Huachuca Cave record (and probably others). This is evident from both the power spectra in Figure C.11 and also the smoothness of the time-averaged δ^{18}O in Figure C.13. This finding compliments an earlier study that suggested *in situ* cave processes may generate substantial non-climatic, decadal-scale variability, and hence cave records from this region might be most meaningful climate proxies on longer timescales (Truebe et al., 2010). De-

spite these limitations, it is likely that the cave record has archived important low-frequency information from the climate system over the last millennium.

Despite these limitations, it is likely that the cave record has archived important low-frequency information from the climate system over the last millennium. Correlations with PDSI, and earlier studies of 20th century variability, indicate that wetter conditions are associated with lower $\delta^{18}O$ values, while aridity is associated with more positive values. At least three megadroughts inferred from tree-rings are present the new cave record (630AD, 1150AD, and 1450AD), and isotopic variance appears to have increased from 1000AD to 1500AD. Overall, there is a long-term trend towards more negative values, which would indicate that the region has become wetter over the last millennium, a result consistent with tree-ring reconstructions.

To develop more refined age models in the future, research could focus on constraining the dynamics of carbonate growth rates so that a physical model of speleothem growth could be used instead of the stochastic models we have shown. Further work could also apply a Bayesian hierarchical modeling scheme to constrain the distribution of ages as a function of depth and the statistical or process model used to represent formation growth, as in Ramsey (2008). Finally, other, longer records from this same cave (or other similar caves) might provide valuable insights into the magnitude and behavior of century-scale, forced and unforced, climate variability of the region.

C.7 Figures

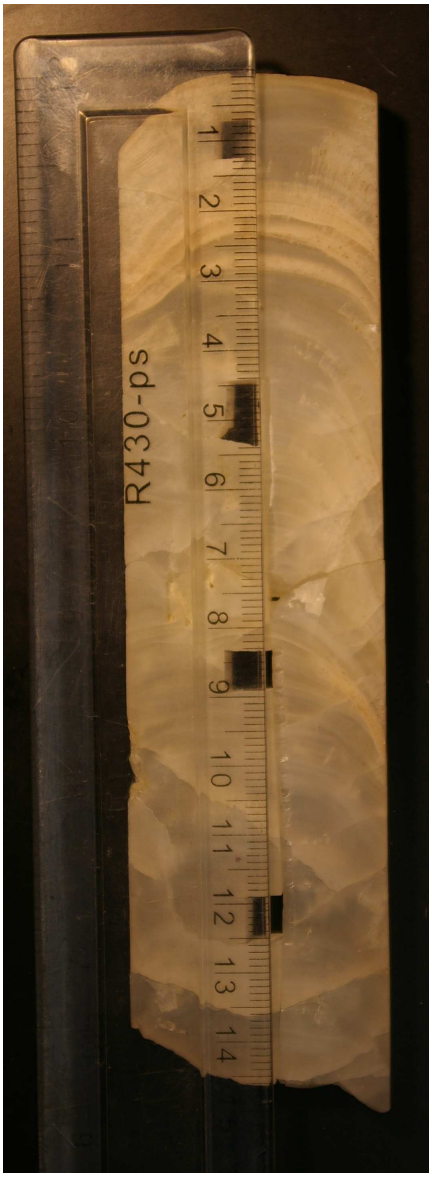

Figure C.1: Picture of the stalagmite slab from Fort Huachuca Cave used for analysis here. The portion on the right was used for isotopic analysis; the portion on the left was used for U/Th dating.

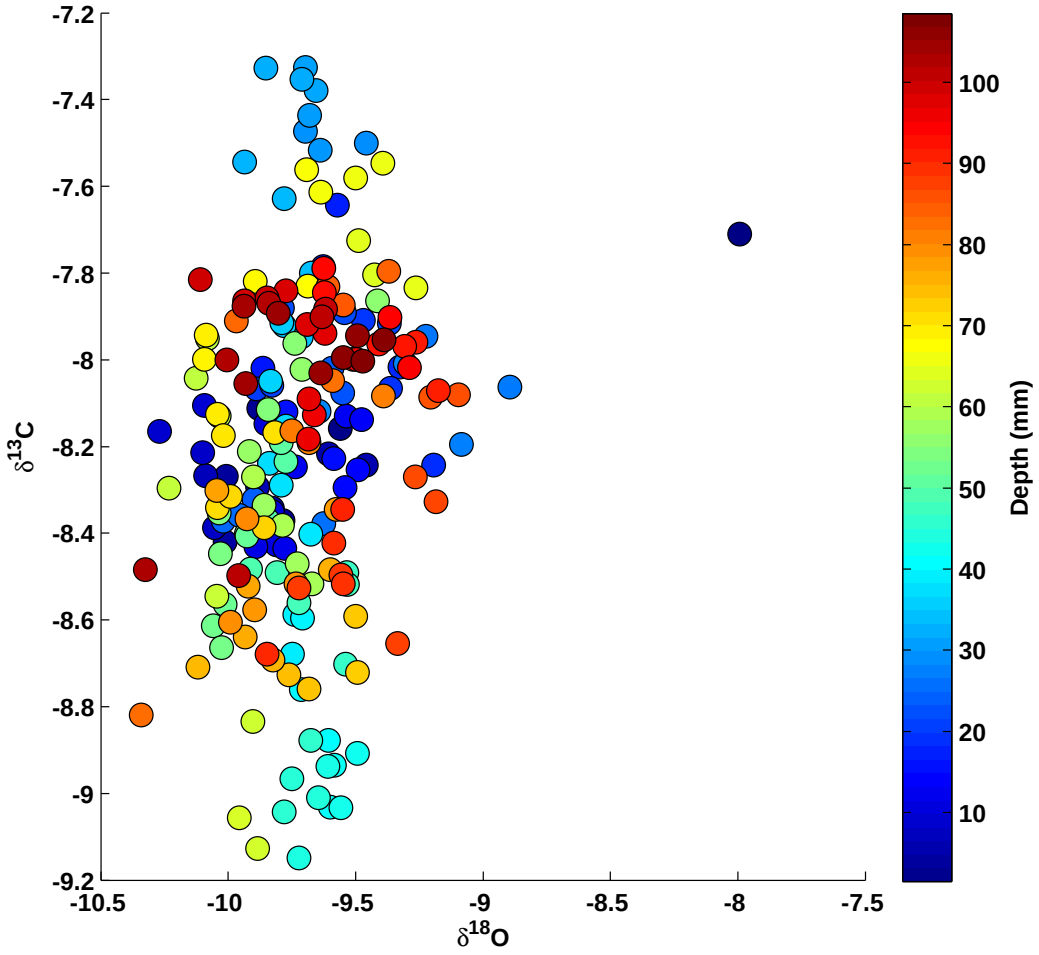

Figure C.2: Scatter plot of $\delta^{13}C$ against $\delta^{18}O$. Coloring indicates the depth of the sample and is included to illustrate that there are not any portions of the core where the two variables appear to covary.

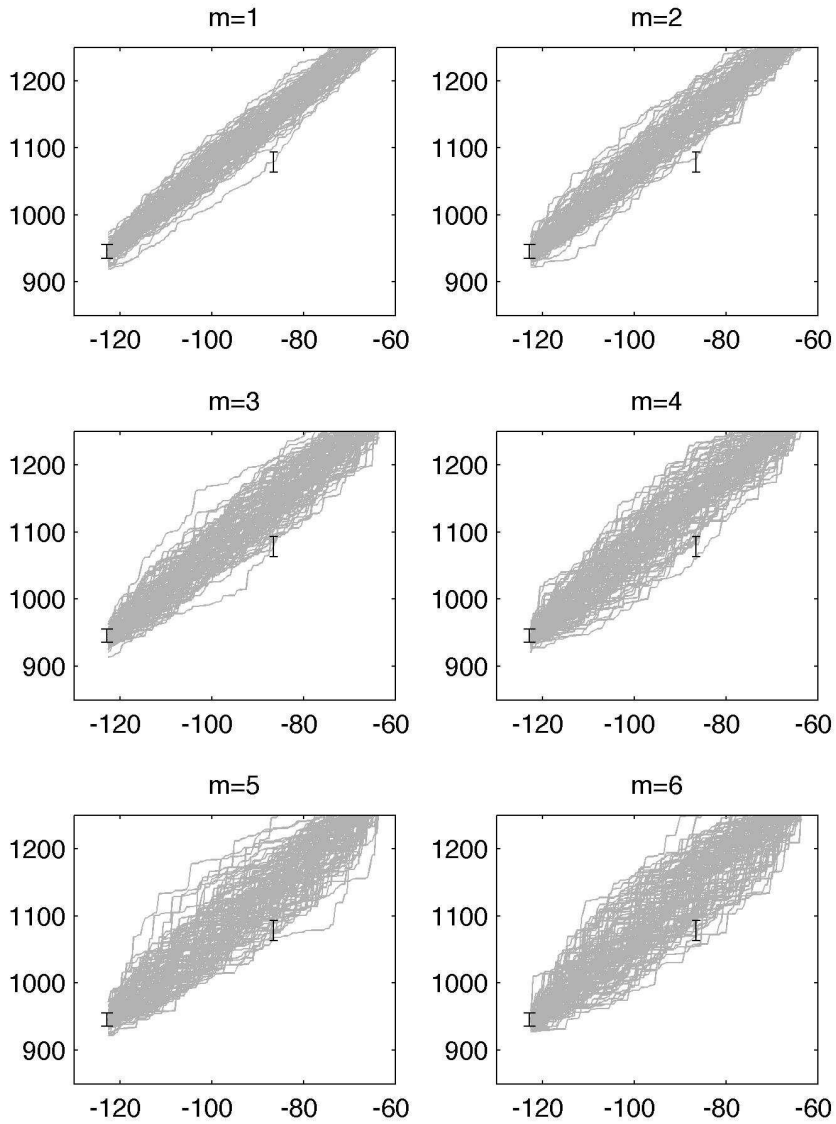

Figure C.3: Illustration of how the value of m is selected for the gamma distribution used in the second set of MC models. Here, three U/Th dates from the middle of the core are shown, and different values of m are used to generate a Monte Carlo family of age models, then compared with the third (middle) date. If small values of m are used, none of the models would be adequately wiggly to contain the third date. However, for larger values, this is not the case. We therefore suggest that a value of m greater than 3 might be most appropriate for the particular cave we are working in.

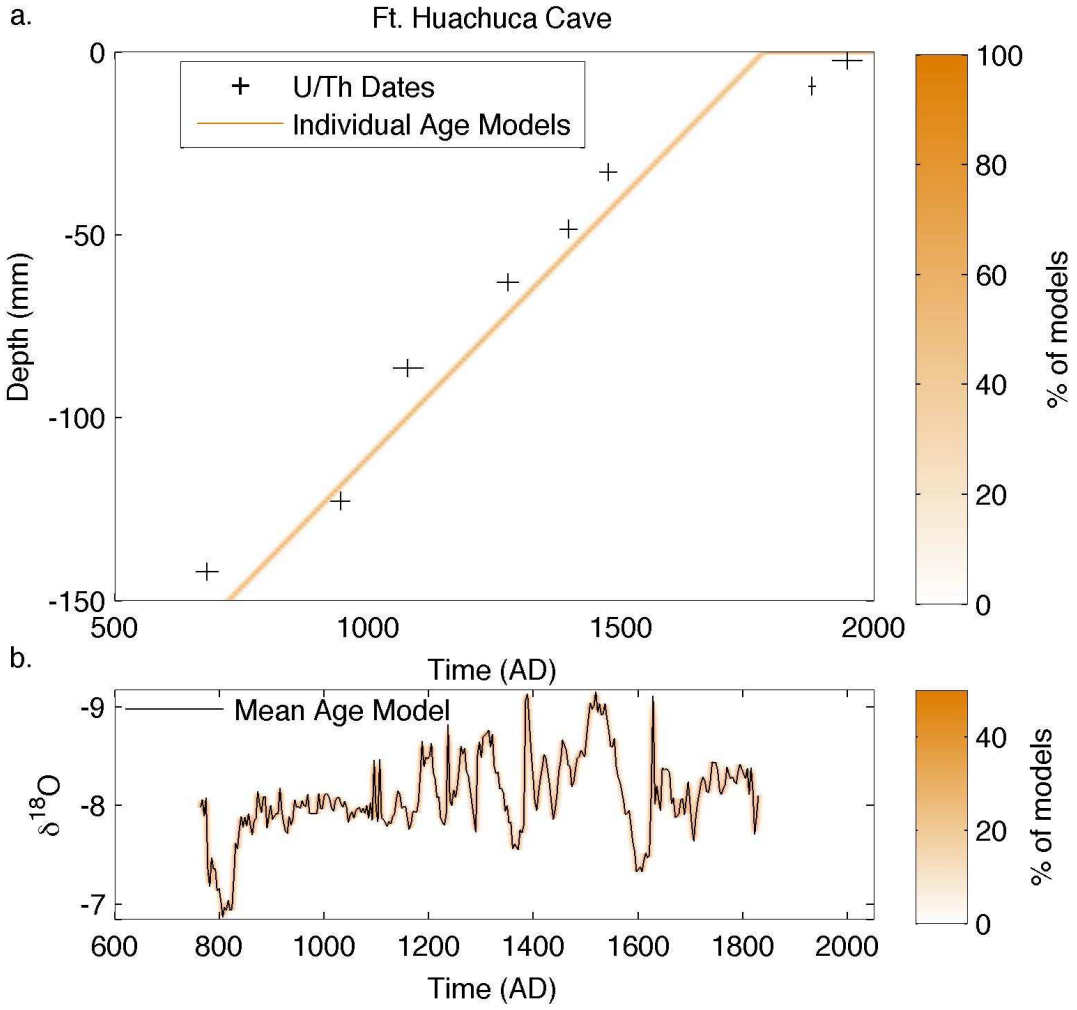

Figure C.4: (a) U/Th dates (squares) and our linear, spline-based ($p = 0$) Monte Carlo family of age models for the Fort Huachuca Cave δ^{18}O record. Depth (from surface) of the samples is shown on the y-axis, age (in years AD) is shown along the x-axis. The shading indicates the number of overlapping age models that run through given point (as indicated by the color scale on the right): areas that are lighter and more transparent have few overlapping age models. (b) The δ^{18}O time series for each of our 1000 Monte Carlo age models (one time series per model). Again, shading is indicated by the transparency of the line as shown by the color scale on the right. the thin black line indicates the time series that would be obtained by averaging together all age models and using this as a single age model for the δ^{18}O measurements.

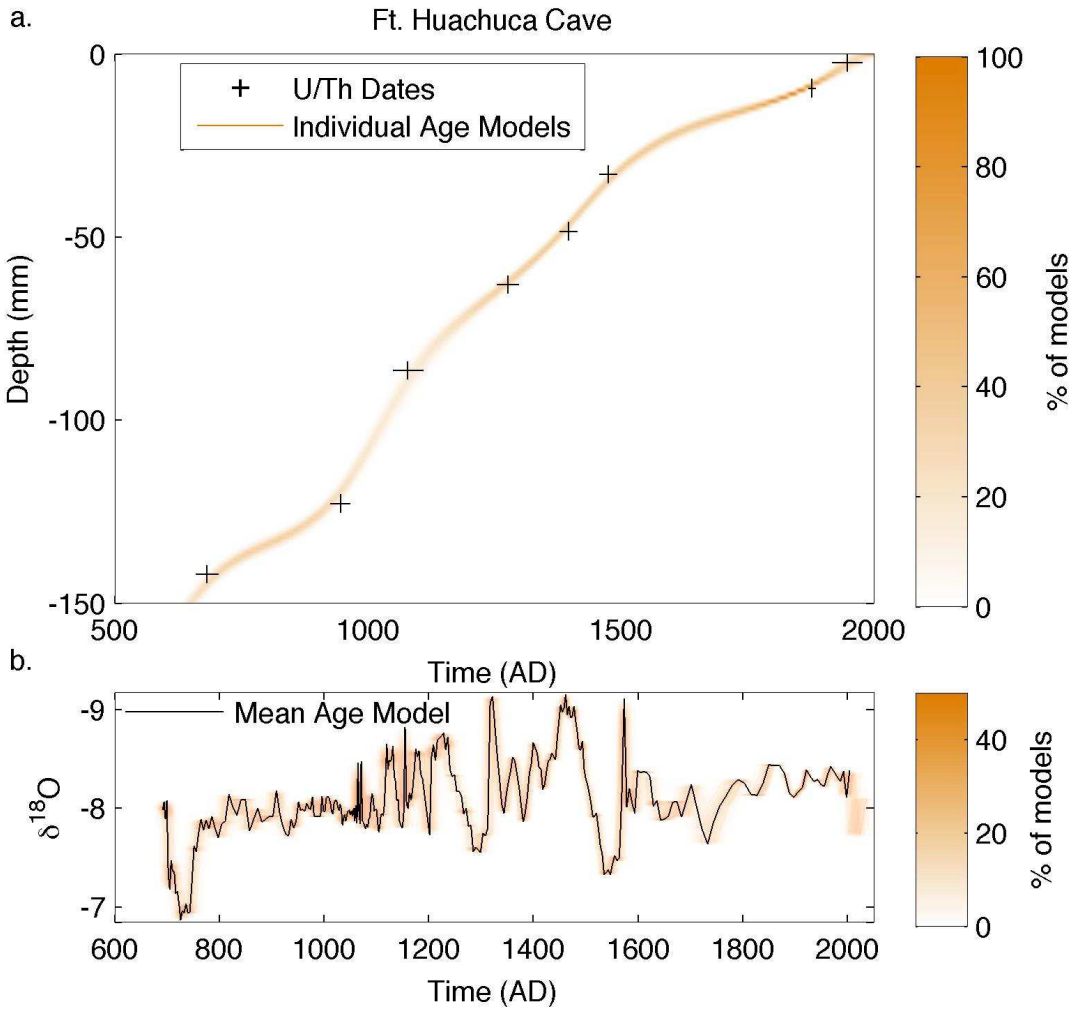

Figure C.5: Same as figure C.4, but for a more flexible spline parameter ($p = 0.00001$).

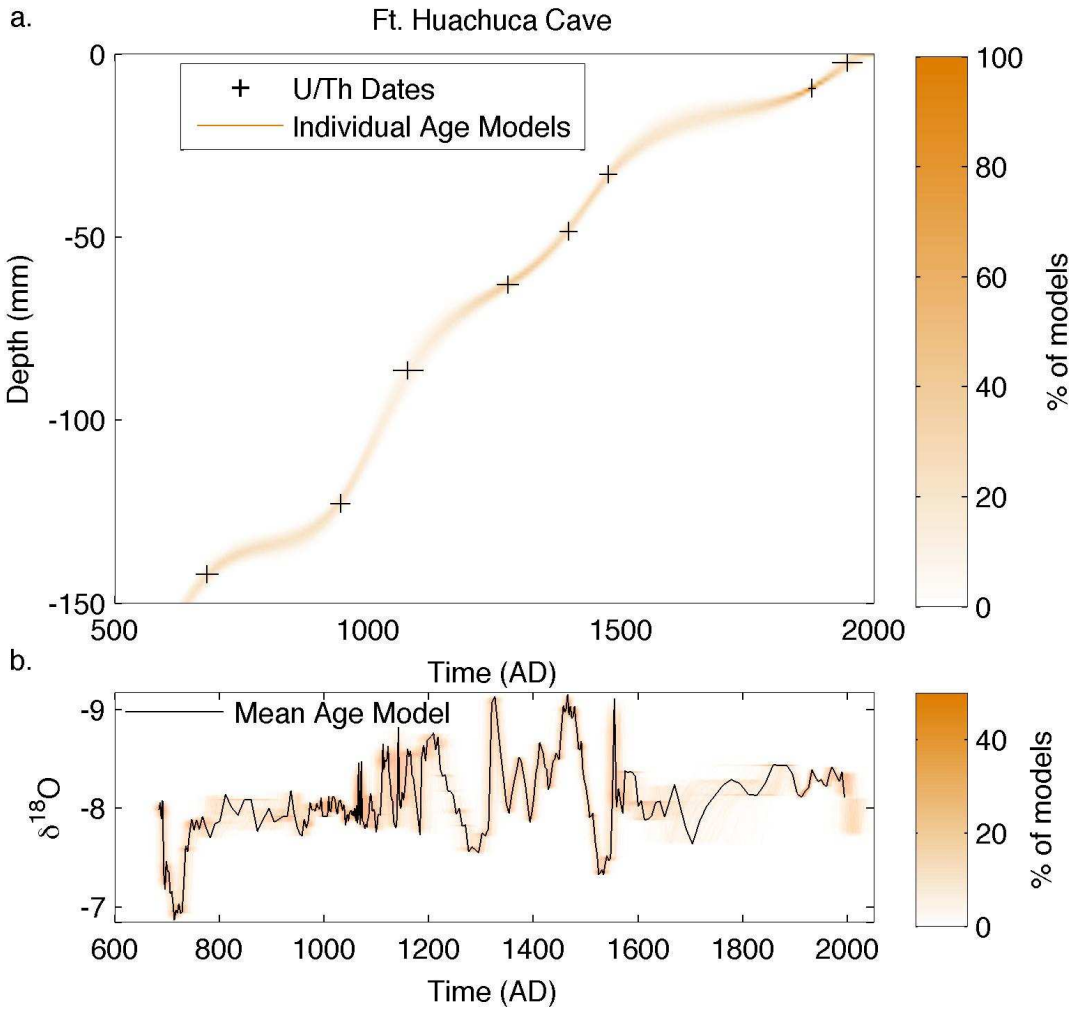

Figure C.6: Same as Figures C.4 and C.5 , but for a spline parameter that is sufficiently flexible as to be able to pass through the midpoint of each individual U/Th date ($p = 0.0001$.

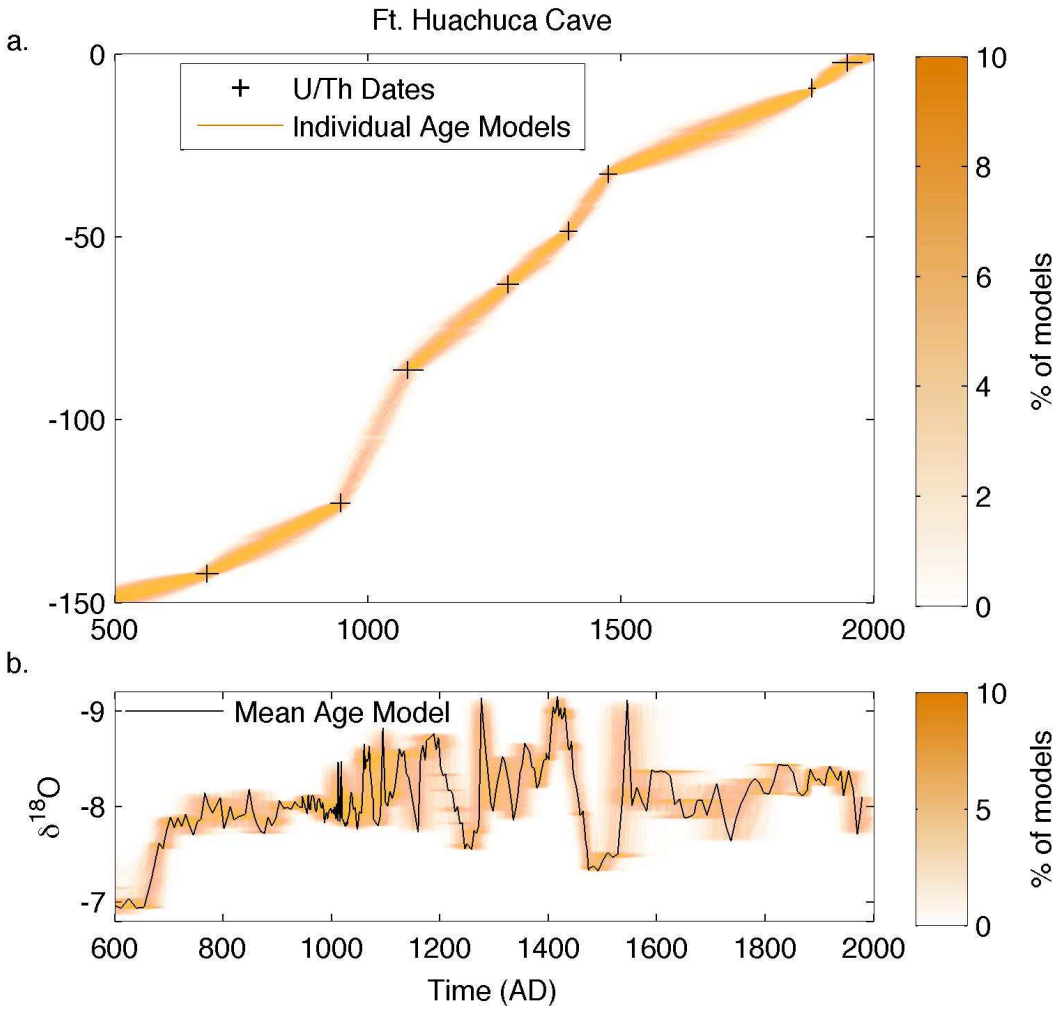

Figure C.7: Same as Figure C.4, but now a stochastic model has been used to model the distribution of time values at each depth. Here the distribution has been scaled to exhibit relatively little variance between each depth step, which is controlled by the parameter m (in this case $m = 1$).

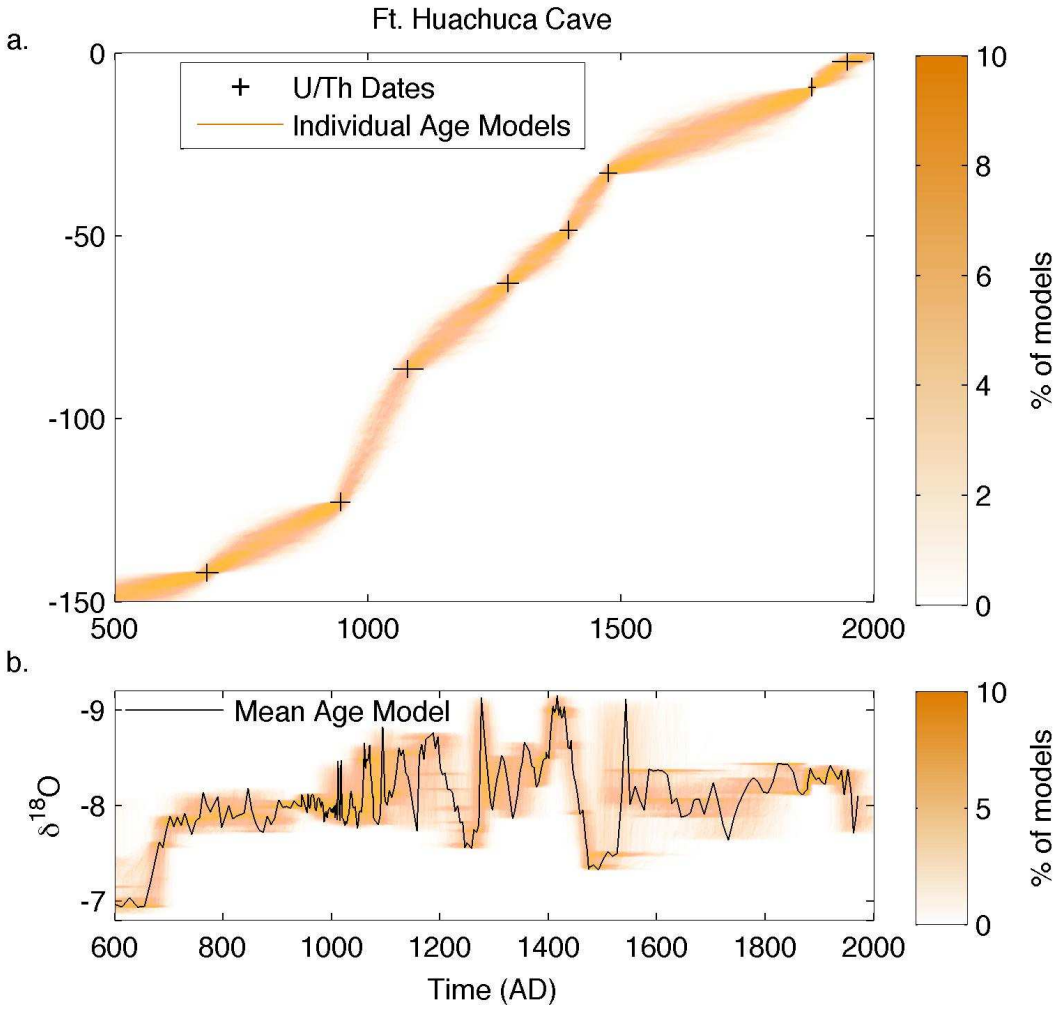

Figure C.8: Same as Figure C.7, but for $m = 3$.

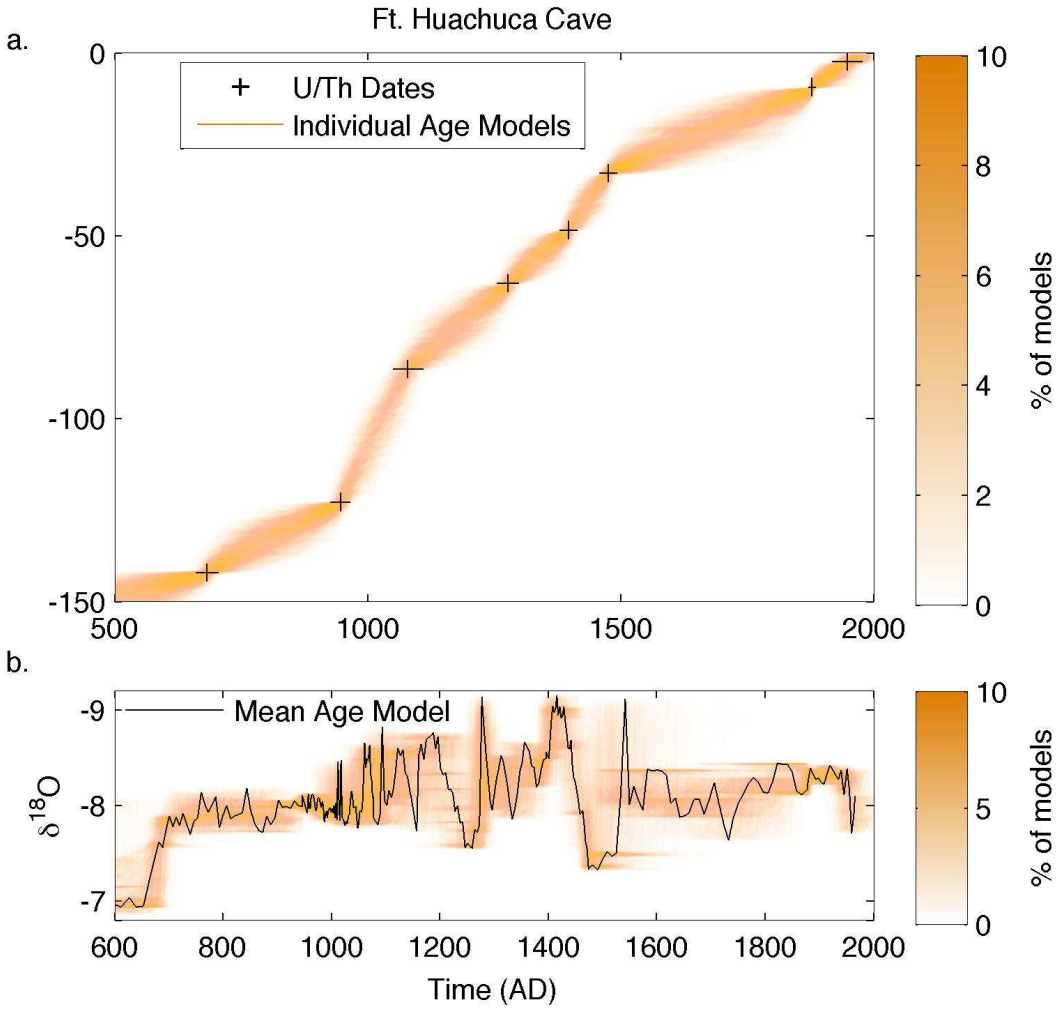

Figure C.9: Same as Figure C.7, but for $m = 5$.

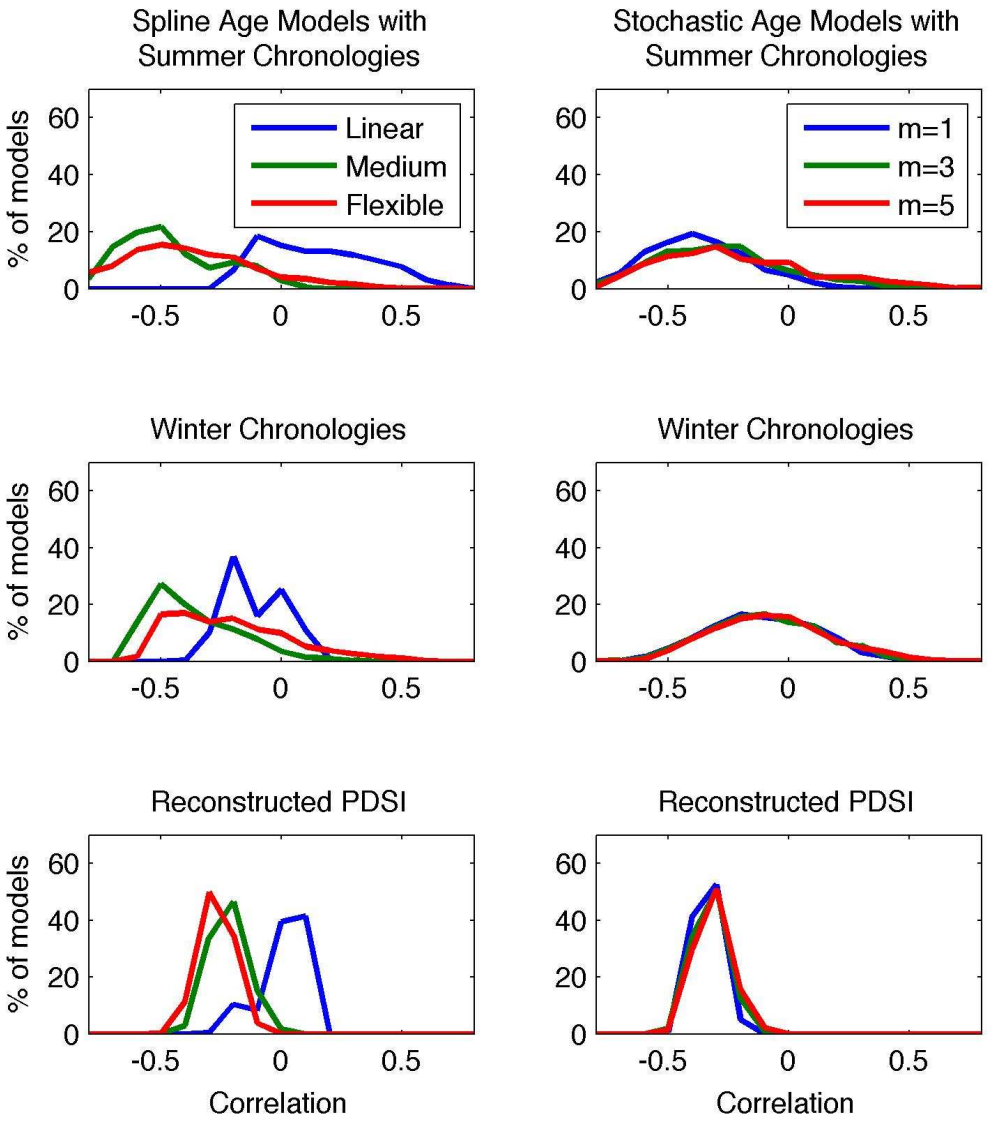

Figure C.10: Correlation coefficients obtained from correlating each δ^{18}O time series from each member of each family of age models with the 25yr averages of tree-ring data. (Left) distribution of correlation coefficients between tree-ring chronologies and each of the families of spline-based age models. (Right) Distributions of correlation coefficients between tree-ring data and each of the stochastic age models. The tree ring data used for comparison were: (top row) summer-sensitive chronologies from the Huachuca Mountains in Southern Arizona; (middle row) winter-sensitive chronologies from the Huachuca Mountains; and (bottom) reconstructed PDSI from southern Arizona. 25-year means of all chronologies were used to calculate correlations.

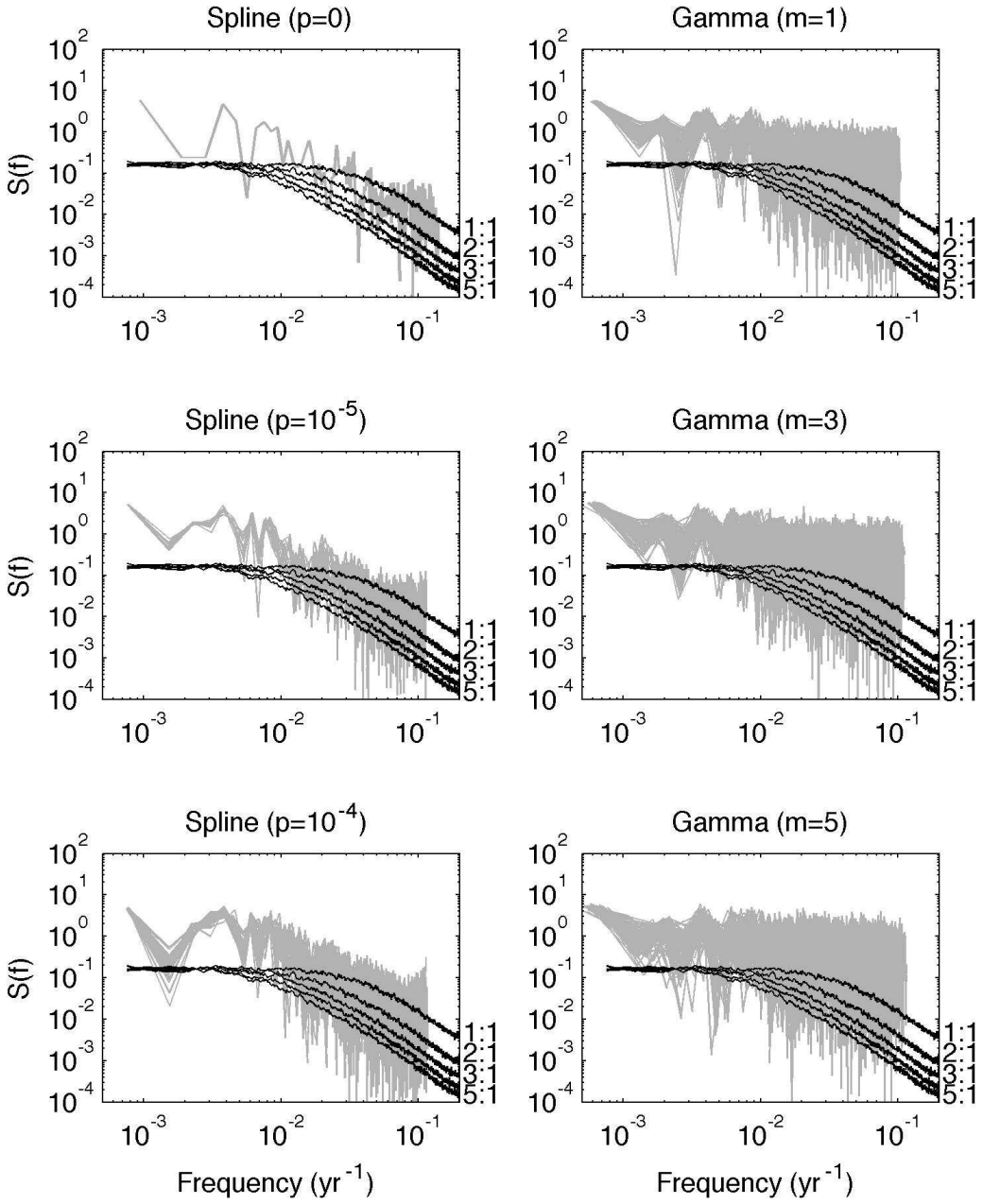

Figure C.11: Spectra of δ^{18}O variability from each member of each family of age models. Panels on the left side show results from the spline-based age models, and panels on the right show results from the gamma-based models. The thin black lines show expected median spectral densities from caves in the region based on a forward model of cave isotopic variability developed by Truebe et al. (2010): differences in these expectations arise from differences in the assumptions about the ratio of rock:soil reservoir sizes (which are indicated by the numbers to the right of the lines).

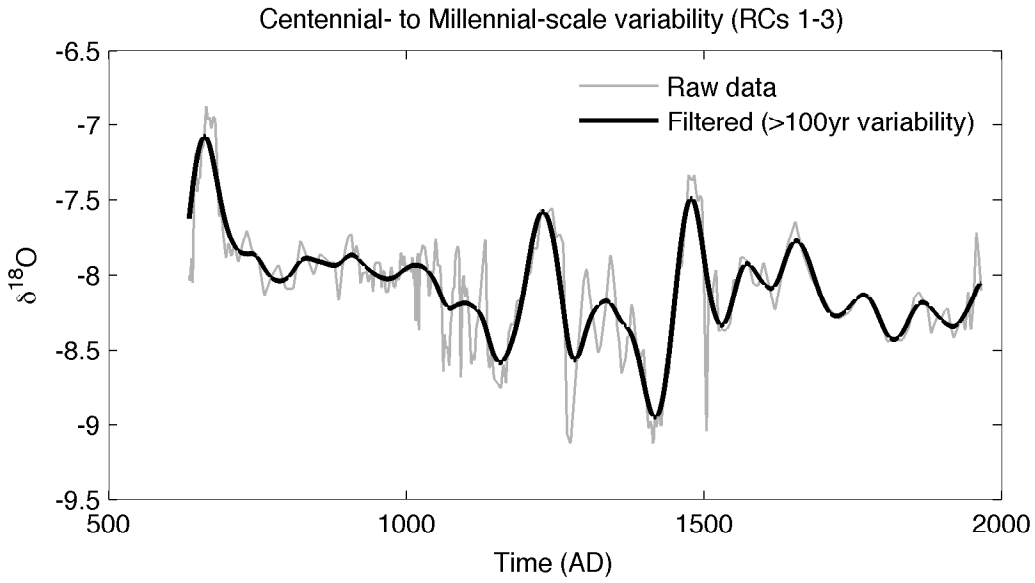

Figure C.12: Raw time series (gray) and low-frequency filtered time series (black) of the Fort Huachuca cave record. In this case, the average of all flexible spline-based models is used as a single age-depth curve. The low-frequency time series consists of the sum of three SSA-based reconstructed components with approximately 1750yr, 140yr, and 90yr timescales.

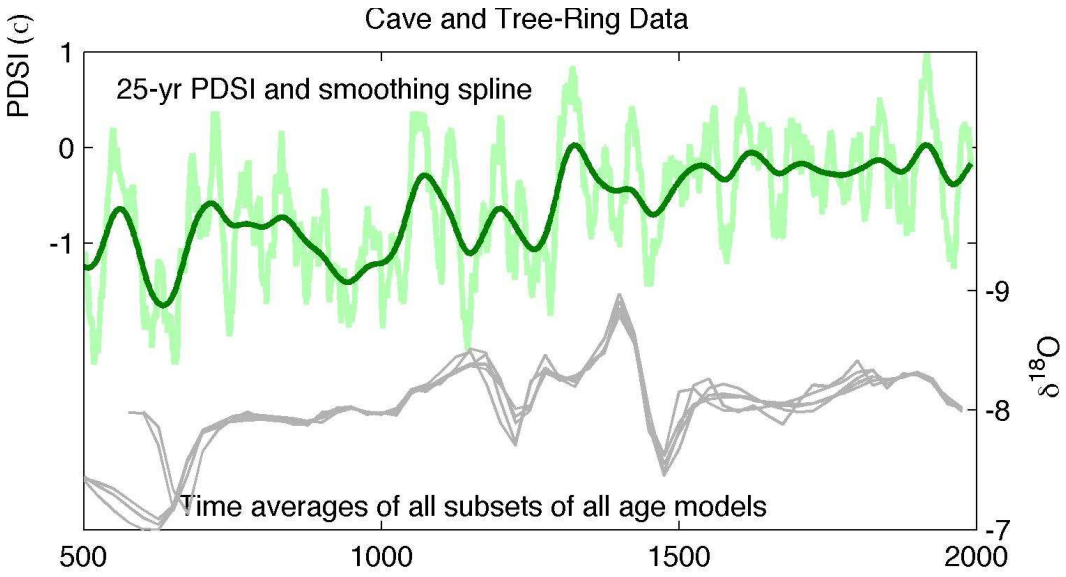

Figure C.13: (Top) Plot of the 25-year mean of reconstructed PDSI (light green), shown with a smoothing spline to highlight the decadal-to-century scale variability. (Bottom) Time-averages of all δ^{18}O time series of all 6 Monte Carlo families of age models considered here (gray lines).

Table C.1: Depths (left column), ages (second column) and standard deviations (third columns) of carbonate samples taken from 7 pieces of the Fort Huachuca Cave core.

Depth (mm)	Date (AD)	$\pm\sigma$
-2.5	1885.73	15.11
-9.5	1819.55	4
-33	1414	8.78
-48.5	1338.46	9
-63	1216.33	11.21
-86.5	1021.5	15
-122.75	888.43	10
-142	625	11.5

Table C.2: Number of significant (at the 95% confidence level) correlations between each tree ring index of hydroclimate (rows) and each family of 1000 age models (columns)

	Summer	Winter	PDSI
Spline ($p = 0$)	47	0	1
Spline ($p = 0.00001$)	410	158	315
Spline ($p = 0.0001$)	296	26	568
Gamma ($m = 1$)	221	24	926
Gamma ($m = 3$)	172	21	851
Gamma ($m = 5$)	169	16	788

APPENDIX D

NORTH AMERICAN DROUGHT PATTERNS

Toby R. Ault[1] and Julia E. Cole[1]
1. Dept. of Geosciences, Univ. of Arizona

D.1 Abstract

Continental-scale megadroughts have occurred in the past, and since they comprise part of the natural variability they will likely occur again. Understanding their causes requires an understanding of their spatial distribution, as well as a distinction between forced low-frequency variability and low-frequency patterns of drought variability that may arise simply from spatial correlation and autocorrelation. Here we examine how different indices of drought lead to different interpretations of their temporal and spatial evolution. We also identify patterns of drought variability that are robust across indices and a wide range of timescales. We test the spectral energy in large-scale spatial patterns of drought against the null hypothesis that it arises as a consequence of high-frequency variability with spatial structure and autocorrelation using 1000 millennial-scale statistical simulations of North American precipitation and temperature. These statistical simulations preserve the spatial correlations of temperature and precipitation, but do not contain low-frequency persistence on timescales longer the annual cycle. We compare the MTM-SVD spectra of these stochastic climates against MTM-SVD spectra estimated from reconstructed (from tree-rings) drought and drought in climate model simulations. Both reconstructed drought and modeled drought are more energetic at low frequencies that would be expected from our null hypothesis. This finding implies that a continuum of dynamically interesting mechanisms may be responsible for low-frequency droughts in western North America, but that at higher-frequency timescales it is not always

possible to disentangle remote and local sources of persistence, even if the droughts have large spatial scales.

D.2 Introduction

Decadal-scale droughts that encompass most of western North America (WNA) have been observed in a wide range of proxy records (Stine, 1990; Laird et al., 1996; Woodhouse and Overpeck, 1998; Graham et al., 2007; Cook et al., 2004; Meko et al., 2007). The prevailing view is that these large-scale, low-frequency "megadroughts" are sufficiently unusual, with respect to the 20th century, that they must arise from exotic sources of variability in the ocean or atmosphere, or that the must be forced by changes in the Earth's boundary conditions (Graham et al., 2007; Herweijer et al., 2007; Seager et al., 2007a). On the other hand, Meehl and Hu (2006) has shown that megadroughts in WNA may arise from internal variability alone, and Hunt (2010) has pointed out that sufficient time and static boundary conditions may be sufficient to generate megadroughts that are comparable to those that have occurred in the past.

Previous studies of both climate models and observational data have failed to put the observations of low-frequency variability in the context of an appropriate null hypothesis. That is, to some extent large-scale, low-frequency drought behavior is expected to occur simply because drought records and drought proxies are "auto-correlated," meaning that drought indicators carry information from one year to the next year. Autocorrelation, in turn, damps out higher-frequency fluctuations while emphasizing lower-frequency ones. Similarly, the interannual-scale and synoptic-scale mechanisms responsible for delivering moisture to WNA exhibit a high degree of spatial correlation, which reflects the size of winter storms, the mean position of the subtropical jet stream, and planetary waves (e.g., Barry and Chorley, 1998). It is therefore plausible that high-frequency variability in the patterns of moisture delivery, and autocorrelation from local sources, would together produce large-scale megadroughts if given enough time.

As an analogy, consider the Pacific Decadal Oscillation (PDO) Mantua et al. (1997): it has a well-defined spatial pattern and it exhibits a high degree of low-frequency variability. However, the amount of low-frequency variance in the PDO, as well as its spatial pattern, can be explained by high-frequency El Niño/Southern Oscillation (ENSO) teleconnections and heat storage in the ocean (Newman et al., 2003). Here, we test whether a similar paradigm might be applicable to the low-frequency energy and the large-scale patterns of western North American (WNA) drought variability.

Section D.3 reviews several widely-used indices of drought, and Section D.4.1 uses one of these indices to examine how four of the worst 20th century droughts evolved at the monthly timescale. Section D.4 then illustrates that the large-scale patterns of drought are relatively insensitive to our choice of index. We argue that the results in D.4.1 and D.4 support the view that interannual variability with large-scale spatial structures of covariability in the atmosphere may account for the droughts we have seen in the 20th century. We therefore test this hypothesis more rigorously in section D.5 by running 1000 statistical simulations of the last millennium that preserve the spatial correlations and autocorrelations of 20th century climate, but do not have longer-term climate variability embedded within them. These simulations serve as a benchmark against which we test the amount of variance in large-scale, low-frequency drought patterns in millennial tree-ring based drought reconstructions of drought and climate model simulations.

D.3 Data and Methods

We compute the standardized precipitation index (SPI) (McKee et al., 1993; Guttman, 1999), the Palmer "Z-Index" and the Palmer Drought Severity Index (PDSI) (Alley, 1984; Palmer, 1965) from gridded 20th century observations. The SPI was developed specifically to describe droughts on multiple timescales to address the variable responses of different components of the terrestrial system. That is, the top few centimeters of soil may respond relatively quickly to moisture anomalies, whereas

streams and river levels may take a bit more time, and large lakes, reservoirs, and aquifers may take longer still. Since the SPI may be computed over any timescale of interest, it can be easily adapted to the specific system being studied. Computation of the SPI is relatively straightforward. First the climatological distribution of precipitation is computed then transformed to fit a normal distribution. Precipitation values for any given month are similarly transformed and compared against their expected values from the (transformed) climatological distribution. SPI time series are therefore normally distributed with zero-mean and 95% of their values fall between -2 and 2. SPI values of -1, -2, -3 and lower may be interpreted as corresponding to moderate, serious, severe, and extreme drought conditions, respectively. SPI may be computed for longer intervals by summing together monthly scores over 3 months, 6 months, 12 months, or even several years. To calculate SPI, we used code written by N. Guttman (http://www1.ncdc.noaa.gov/pub/data/csd/palmer/spinew.f). In this algorithm, a Pearson's Type III distribution is fitted to the precipitation values, which is then transformed to a normal distribution and used to compute SPI values. Guttman (1998) found that such a distribution was a robust fit for most precipitation records. We computed SPI for 1, 12, 24, 48 and 120 month (1mo, 1yr, 2yr, 4yr, and 10yr) timescales using the "TS2.1" gridded precipitation dataset produced by the Climate Research Unit (CRU) at the University of East Anglia (Mitchell and Jones, 2005).

Calculation (and interpretation) of the PDSI is less transparent than SPI, and its strengths and weaknesses have been extensively explored (Alley, 1984; Guttman, 1999; Redmond, 2002; Wells et al., 2004; Dai et al., 2004). At its core it is a simplified, two-layer model of soil moisture that is exhibits approximately unit variance and mean of zero. In calculating PDSI, first the precipitation minus evaporation (and runoff) balance is calculated to produce a monthly index of anomalies called the Z-index. Autocorrelation is then added to the Z-index to mimic the persistence inherent to soil moisture at monthly timescales (Alley, 1984), and the resulting index is then re-normalized. We use Matlab code provided by D. Meko (pers. comm.) to calculate both the Z-index and PDSI. Again, we used CRU precipitation on a 1.5

x 1.5 degree grid, as well as CRU temperature on the same grid.

We also calculate SPI and PDSI from two long CCSM3 simulations (Collins et al., 2006). The first is a 1000 "1990" control run (with forcings described in Solomon et al., 2007, Chapter 8), and the second is a 2300 year forced, transient simulation of mid-Holocene (Kaufman et al., 2009). In the latter simulation, the Earth's boundary conditions evolved slowly by adjusting the orbital, solar, and greenhouse-gas forcing parameters from 6000 to 3600 BP. Vegetation adjustments to climate were resolved dynamically and volcanic variability was simulated stochastically (see supporting on line material of Kaufman et al., 2009).

Reconstructed PDSI fields spanning the last millennium are from the "North American Drought Atlas" (Cook et al., 2004). This product provides gridded PDSI time series for the entire North American domain going back to about 1400 AD, and back to 2 BC for western North America (24N to 52N, 95W to 125W). However, before about 1000 AD the number of trees used in the reconstruction is low (see supplementary material of Cook et al. (2004)). We therefore focus our analysis on the last millennium because it provides a reasonable compromise between extensive temporal coverage and reasonable spatial coverage. During the last millennium the reconstructions are quite reliable and have been shown to exhibit pronounced, continental-scale intervals of megadroughts (Cook et al., 2004; Herweijer et al., 2007).

We use principal components analysis (PCA) to identify large-scale patterns of SPI variability in instrumental, proxy and model data. PCA reduces large noisy datasets into a small set of orthogonal (in space) eigenvectors (empirical orthogonal functions called EOFs) and corresponding principal component (PC) time series. We use a Monte Carlo (rule N) approach to test the significance of the leading eigenmodes in instrumental and model data. We performed PCA on all the SPI fields, on the instrumental Z-index and PDSI fields, and on the tree-ring PDSI fields. We rotated the leading EOFs corresponding to the leading PCs to satisfy the varimax criterion described by Richman (1986). To interpret the large-scale controls on the leading PCs we identify, we correlate these time series with 250mb heights, zonal wind, and meridional wind from NCEP reanalysis data (Kalnay et al., 1996),

and with sea surface temperature (SST) data (Kaplan et al., 1998)

D.4 Patterns of Drought

D.4.1 20th century case studies (monthly resolution)

We use SPI to examine the temporal evolution of several of the worst droughts during the 20th century. These years can be readily identified by a "drought area index" (DAI), which we compute here as the percentage of North America experiencing 12mo-SPI values below -1 Figure D.1. Using this definition of DAI, the worst years of drought were 1931, 1934, 1936, 1956-1957, 1976-1977, and 1988. Different thresholds for calculating the DAI, and even different indices, also highlight these years (e.g., Cook et al., 2004; Schubert et al., 2004a; Seager et al., 2005b). However, the DAI time series calculated from 1mo SPI (gray line in Figure D.1) shows that there were several other short-lived (monthly) events of similar magnitude to the time windows we choose to examine here.

In the time-evolving maps of SPI (Figures D.2,D.3), we can observe how dry conditions worsened through the spring and summer of 1934 across most of the continent. Comparing the 12mo SPI maps in Figure D.2 with the 1mo SPI maps in Figure D.3, we can infer that much of the drought consisted of moderate (-1) SPI values at the monthly timescale, which are intensified considerably by the extreme SPI values (<-3) during May of 1934. We also note that the high degree of spatial structure apparent in the 12mo SPI is not present in the shorter, 1mo SPI field.

The spatial pattern of the 1956-1957 drought is quite different from the 1934 drought and is shown in Figure D.4. As in the earlier interval, most of the continent was already experiencing moderate drought (SPI values <-1), but conditions deteriorated in the late summer and fall of 1956 for the entire US Southwest. By January 1957, virtually the entire domain is in moderate (-1), severe (-2), or extreme (-3) drought. Once again, the 1mo SPI maps depict a different, noisier evolution of the drought: extreme aridity occurred in the southwest during March, June, and September of 1956, and remained present (although less severe) through November

and December (Figure D.5).

In the SPI depiction of the 1976-1977 drought, the event appears to have started as early as June of 1976 near the Great Lakes region, then spread through the following fall, winter, and spring until most of the domain is encompassed by SPI values below -1 (Figure D.6. By the summer of 1977, conditions had improved slightly in the great lakes region, but remained dry along the west coast of North America. From the 1mo SPI maps in Figure D.7, we can see that the much of the drought's intensity was related to extreme conditions in November and December in 1976, as well as February and April of 1977.

Finally, in 1988 the 12mo SPI data depict extreme mid-continent aridity during spring and summer of 1988 in the middle of the continent that did not completely abate until even December of 1989 (Figure D.8). This event appears far less remarkable, however, in the 1mo SPI maps, which suggest that April, May and June played the most important roles in initiation the drought, and conditions in October worked to sustain the drought.

D.4.2 PCA Results

To reduce the number of drought patterns down to a more manageable number (rather than considering every month of every year), we now turn our attention to the results from PCA performed on a variety of different drought indicators (e.g., Karl and Koscielny, 1982). These results depict at least two large-scale drought patterns that together explain 20% to 50% of the total variance in the 5 different drought datasets we examined (Figure D.10). These patterns explained more variance the 12mo SPI, PDSI, and reconstructed PDSI datasets than they explained in the 1mo SPI and Z-index data (but still comprised the top 2 PCs). Patterns are depicted here as correlations (as opposed to raw EOFs) between PC time series and the original raw data so that the units are comparable across datasets. The first pattern we show (which is either the first or second PC for each index; see figure for details) is a "domain-wide" feature with maximum strength in the center of the continent. The second is a north-south dipole pattern with "centers of action" of

opposite sign in the US Southwest and Northwest. We also identified an east-west pattern in some of the indices (as in Woodhouse et al., 2009), but found that this feature was less stable across indices and timescales. Other patterns (not shown) explained less variance and were also less robust with respect to our choice of index and our decisions about rotation. In general, however, the consisted of regionally-synchronous patterns, as opposed to the continental-scale ones shown here.

The temporal evolution of the domain-wide pattern is shown in Figure D.11 for the different datasets we analyzed, and the evolution of the second pattern is shown in Figure D.12. All PC time series are in good agreement with each other for both patterns. They show major intervals of drought associated with the domain-wide pattern in the 1930's, as well as the late 1970s and 1980s. The north-south pattern depicts Southwestern droughts during the 1910s, 1920s and most especially during the 1950s.

Figure D.13 shows the results of performing PCA on SPI datasets that have been constructed to emphasize variability at timescales from 1 month to 10 years. Again, two primary patterns emerge (a domain-wide pattern and an north-south pattern). From this analysis, it appears that the domain-wide pattern exhibits greater strength further west at longer timescales, whereas the north-south pattern is more timescale-independent. Again, the PCs corresponding to the domain-wide pattern in Figure D.14 are quite similar across timescales, although they necessarily become smoother as the length of the SPI timescale increases. At the longest timescale, it is evident that the domain-wide pattern characterizes the 1930s drought, and the north-south pattern characterizes the 1950s drought.

Spectra of the domain-wide and north-south patterns are shown in Figure D.16. Importantly, the spectra of both PC time series derived from the 1mo SPI and the Z-index (e.g., the indices that have no built-in "memory") are the whitest, meaning that they exhibit a uniform distribution of variance across timescales. In contrast, the 12mo SPI and PDSI are much redder, particularly on monthly to interannual timescales, while the tree-ring reconstructions of PDSI fall somewhere in between. This idea is consistent with the argument that soil moisture behaves like a weakly

damped, stochastically forced process (Delworth and Manabe, 1989).

Our PCA results indicate that the most robust patterns of large-scale drought variability are not sensitive to the choice of index. They do, however, highlight a limitation of smoothed indices, such as the SPI on timescales longer than 1 month, PDSI and reconstructed PDSI. This limitation can be seen quite clearly in Figure D.16 because the 1mo SPI and the Z-index are computed at monthly timescales, and hence best reflect the timescales of variability inherent to the atmospheric controls on drought. In contrast, the 12mo SPI and PDSI are smoother in time and redder in frequency, and hence their low-frequency components are emphasized. However, we can plainly see from Figures D.11, D.12 D.14, D.15, and D.16 that this tendency is not an intrinsic feature of the climate inputs to the drought indices, a point also made by Ault and St George (2010).

D.4.3 Seasonal characteristics

The domain-wide and north-south spatial structures shown in Figure D.10 may vary in their importance to North American hydroclimate on seasonal timescales. To test this idea, we map out the month-by-month expression of these patterns in Figures D.17 and D.18. These maps were generated by taking the 1mo PC amplitudes of each pattern at each month and correlating it with the original 1mo SPI data at that month for the entire length of the dataset. By focusing on the 1mo SPI we can ensure that the patterns are not driven by antecedent conditions, but rather the state of the atmosphere during the months of interest.

The domain-wide pattern is apparent throughout the year (Figure D.17), although it reaches its greatest strength and spatial extent during winter and fades somewhat as summer approaches. It then re-emerges in its importance during fall. Notably, during spring months (Mar, Apr, May) and in the beginning of summer (Jun) the pattern exhibits a di-pole pattern with centers of action in northern Mexico and just southwest of the US Great Lakes.

Like the domain-wide pattern, the north-south pattern is most pronounced during boreal winter, while in the fall its spatial extent is reduced to the US Southwest

and northern Mexico.

D.4.4 Teleconnections

Correlation fields between the domain-wide pattern and NCEP 250mb height fields are shown in Figure D.19. The winter-time correlations depict a wave-train emanating out of the western mid-latitude Pacific ocean that is strongly reminiscent of the Pacific North American (PNA) pattern (Wallace and Gutzler, 1981). High 250mb heights above northwestern North America correlate with drier PC values, and low heights correlate with negative PC values. The pattern is similar year-round, but in summer spatial extent of significant correlations is much less than in winter and fall.

As with the domain-wide pattern, the extent of the significant correlations between the north-south PC and 250mb heights is greatest in winter (D.20). These correlation patterns also vary considerably with season. In winter, high 250mb heights correspond along the US west coast correspond to aridity in the southwest and wetter conditions in the northwest. Highs along the US west coast are correlated with a wet Southwest and a dry Southwest.

Monthly correlations of SST with the domain-wide pattern are shown in Figure D.21. During winter (DJF), these correlation fields emphasize the relationship between drought anomalies and the North Pacific. During the rest of the year, the oceans do not seem to be significantly correlated with anomalies in the PC time series of the domain-wide pattern.

Correlations between the north-south pattern and tropical Pacific SST are positive during fall (OND) and late winter as well as spring (FMAM) (Figure D.22). Interestingly, the correlations during January are not as widely significant. Correlations are strongest with tropical Pacific SSTs during November and December. During summer, significant correlations are not generally present.

D.4.5 Discussion and Interpretation

Thus far, we have examined the temporal evolution of some of the worst drought episodes in North America during the 20th century, diagnosed the primary patterns of drought variability and their behavior through time, and elucidated certain seasonal characteristics of the large-scale atmospheric and oceanic correlations. To varying degrees, all three of these analyses emphasize that a non-unique variety of remote mechanisms can generate drought patterns of drought that are highly similar. For instance, the 1930s Dust Bowl droughts and the 1988 summer drought share many similar features including their seasonality and spatial expression, yet there is no firm consensus on their ultimate causes. The earlier drought has been explained recently in the context of anomalous SST forcing (Schubert et al., 2004b), dust forcing (Cook et al., 2008), and internal atmospheric (and stochastic) variability (Hoerling et al., 2009). The 1988 drought has been variably linked to anomalous SST forcing in the equatorial pacific (Trenberth et al., 1988; Trenberth and Branstator, 1992; Trenberth and Guillemot, 1996), atmospheric variability (Lyon and Dole, 1995; Chen and Newman, 1998), and most recently, variability in the strength of the Caribbean Low-Level Jet (Cook and Vizy, 2010).

From the perspective of our analysis, the idea that different causal mechanisms have similar impacts is not surprising: the domain-wide pattern exhibits variability year-round, and hence must be driven by fundamentally different dynamical processes depending on the season. Since ENSO exhibits a dipole-like pattern in western US precipitation during the winter (Rasmusson and Arkin, 1993; Dai et al., 2004; Woodhouse et al., 2009), variability in the tropical Pacific likely influences the structure of the di-pole pattern shown here. Nonetheless, this pattern is present even during summer months when it does not appear to be correlated to ENSO, hence other processes may also generate this pattern during summer. Earlier studies have also argued that the Northern Annular Mode (NAM) may exert a similar influence to ENSO on North American storm tracks and hence precipitation patterns (Quadrelli and Wallace, 2004; McAfee and Russell, 2008).

Our results also raise the possibility that much of the low-frequency character of drought arises from the definition of the phenomena, and not, as is usually assumed, by the inherent low-frequency behavior of the climate system. This clearly seems to be the case in our analysis of the worst droughts of the 20th Century (Figures D.2 through D.9, our PCA of different drought indices (Figures D.11 and D.12) and SPI at different timescales (Figures D.11 and D.12), and especially in the power spectra of the leading PC time series of the drought patterns we have described D.16. Essentially, our results raise the possibility that the spectrum of North American drought variability is consistent with a weakly-damped, stochastic forcing that exhibits some large-scale spatial covariability (Delworth and Manabe, 1989; Hasselmann, 1976). In the next section (D.5), we test this hypothesis using statistical simulations of North American climate (with 20th century statistics) for 1000 years. We calculate drought indices from these climates and compare them with long tree-ring based reconstructions of drought and with long climate model simulations.

D.5 Do megadroughts require a low-frequency mechanism?

Here we explore the possibility that low-frequency behavior in drought arises simply from high-frequency fluctuations with large-scale spatial structure and autocorrelation. To test this possibility, we turn our attention to reconstructed PDSI (Cook et al., 2004), PDSI calculated from GCM output (see section D.3), and PDSI calculated from 1000 year long bootstrap simulations of the 20th century. To generate bootstrap realizations of climate, which we term "null climates," we first pool all of the monthly temperature and precipitation maps from the CRU TS2.1 dataset (Mitchell and Jones, 2005) into 12 sets of 102 maps (one map for each month of each year). Simulations are run by starting in January and selecting a year at random from the 102 the CRU TS2.1 January precipitation and temperature fields. A small amount of noise is added to these fields (with noise variance equal to the variance of the raw data), then the simulations move on to the next month and repeat the

procedure until 1000 years of monthly data have been assembled. We then calculate PDSI from these simulations, and average PDSI values together to produce annual, gridded PDSI fields with 1000 years of data to compare with the annually-resolved PDSI reconstructions. Our bootstrap re-sampling of temperature and precipitation preserves three important characteristics of 20th century climate: (1) the means and standard deviations of temperature and precipitation at each grid point; (2) the spatial correlation structures of both precipitation and temperature; and (3) the covariance between temperature and precipitation. However, no low-frequency information is present at longer timescales. Essentially, we are simulating 1000 years of 20th century climate with monthly deviations from the annual cycle that are uncorrelated between time steps before PDSI is calculated.

We use two metrics that allow us to compare the range of PDSI variability in our null climates, reconstructed PDSI, and PDSI in climate models. First, we examine the distribution of the drought area index (DAI), which is a single time series calculated from each PDSI field. It indicates the percentage of western North American grid points experiencing PDSI values below -1 (Cook et al., 2004). Here we simply tally the total number of years that are spent at various DAI values, and express this tally as a percentage of all years, allowing us to compare time series of different lengths. Second, we examine the raw (unnormalized) MTM-SVD spectrum (Mann and Park, 1999). We use these metrics because they allow us to ignore the time-evolution of climate, which is expected to differ between the reconstructions, our stochastic simulations, and the long CCSM3 runs. The MTM-SVD method is particularly useful in this regard because it measures the amount of spectral variance associated with a (complex-valued) spatial pattern, hence both the spectra and the patterns of variability can be compared between the different types of data considered here.

D.5.1 Results

Figure D.23 shows the distribution of the DAI from 1000 null climates, tree-ring reconstructions, and two climate model simulations. The null climates are distributed

around a DAI value near 10%, meaning that about 10% of western North America would be expected to experience drought conditions at or below -1 most of the time. DAI values above 40% are not common in the null climates. In contrast, DAI values calculated from tree-ring reconstructions of PDSI exhibit a much wider distribution. This is also true for the CCSM3 control runs and the long transient CCSM3 simulation.

Our MTM-SVD results clearly show that there is more variance in PDSI reconstructed from tree-rings, and simulated by climate models, than would be expected from our null hypothesis (Figure D.24). However, on interannual timescales, the shape of the null climate spectra, the tree-ring spectra, and the climate model spectra are all very similar: variance tends to increase from 2yr to 10yr timescales. Since the null climates do not have any sources of low-frequency variability, outside of the autocorrelation imposed by calculating PDSI, it follows that the PDSI's built-in autocorrelation may explain some of the tendency for variance to increase from 2yr to 10yr timescales in both the reconstructions and in the climate model.

The CCSM3 control run and the tree-ring reconstructed MTM-SVD spectra agree closely across most frequencies, and they even exhibit similar peaks at the centennial timescale. At longer timescales they disagree, with the PDSI spectrum tending to support more variance than the CCSM3 control spectrum. The CCSM3 transient spectrum, on the other hand, exhibits similarly enhanced multi-centennial to millennial variance (reflecting the dynamical boundary conditions at those lower frequencies). The CCSM3 transient spectrum also generally comparable to the tree-ring PDSI spectrum on interannual to centennial timescales.

As a final comparison between the climate model simulations and the reconstructions of PDSI, we examine the spatial pattern of drought associated with a particular time scale of variability across biennial to centennial periods. Figures D.25 through D.27 show the variance weights we used were just the MTM-SVD spectra amplitudes corresponding to each frequency for each map. In the reconstructed PDSI field, the centennial band is characterized by a domain-wide pattern that evolves roughly synchronously throughout the region. On multidecadal timescales, the pat-

tern is more complicated: it begins with weak loadings in the west, evolves into a N-S dipole, then switches sign. We found that the interannual patterns were different at 4-7 year timescales than at 2-4 year timescales, and hence we show them as separate components here. Like the centennial timescale, the 4-7 year timescale is characterized by a domain-wide, in-phase pattern. At the higher frequency end of the interannual band (2-4 years), there is a N-S dipole that decays into a weak domain-wide pattern, then switches sign before decaying again.

In the CCSM3 control run (D.26, the patterns of variability at centennial, multidecadal, and interannual timescales are very similar to the patterns of reconstructed PDSI variability. There is a domain-wide, roughly-synchronous pattern at the centennial timescale. There are also N-S dipole patterns at the multidecadal and interannual (2-4 year) timescales. Again, the lower-frequency (4-7 year) interannual pattern is more regionally-expansive than its 2-4 year counterpart. The forced transient simulation, on the other hand, exhibits just one pattern that is roughly the same across timescales (Figure D.27).

D.5.2 Discussion

If high-frequency mechanisms and autocorrelation were a viable explanation for megadrought variance, we would have seen tree-ring spectra and climate model spectra that fell within the confidence limits determined by our statistical simulations of North American climate. Instead, we note that the MTM-SVD spectra of both tree-ring and model drought rise quit well above the null expectation on long timescales. This finding is important because it implies that dynamically interesting climate mechanisms are needed to account for the full range of decadal to centennial-scale drought behavior in the west, a point that has often been assumed but that we have now shown more conclusively. Moreover, we have shown that the MTM-SVD spectrum of reconstructed PDSI is in reasonably good agreement with at least one climate model (CCSM3). This agreement is remarkable when the raw MTM-SVD spectra of the CCSM3 control run and the reconstruction are compared, and also when comparing the transient simulation with PDSI. Variance in the forced simu-

lation is even higher in the CCSM3 transient simulation than in the reconstructed PDSI spectrum, indicating that the evolving boundary conditions have a marked effect on the power spectrum of drought on centennial to millennial timescales.

It should be noted, however, that the reconstructions of PDSI likely reflect a conservative estimate of low-frequency variability in drought because tree-ring data must be "pre-processed" to remove low-frequency, non-climate variability before they can be used in any reconstruction. In the case of the PDSI drought atlas, the details of this pre-whitening are given in the supplementary material of Cook et al. (2004). Briefly, they entailed (1) removal of growth trends from tree-ring chronologies; (2) pre-whitening of predictor variables; and (3) statistical rescaling of the autocorrelation term to better-agree with instrumental data. Since steps (1) and (2) would be expected to remove some low-frequency climate variability (as well as non-climate variability), we emphasize that the good agreement between the proxies and the models might be somewhat artificial. Nonetheless, the range of drought that has been discussed using the drought atlas is apparently consistent with an unforced, fully coupled 1000 year climate model simulation.

The patterns of drought in both the reconstruction and in the CCSM3 control run are very similar across timescales. Moreover, the patterns of PDSI variability are robust: we tested several other methods including "standard" principle components analysis and multichannel singular spectrum analysis Ghil et al. (2002), and found that these spatial structures and their corresponding timescales were generally consistent across methodologies. Since we have limited our analysis to the western half of North America (because of PDSI sample depth issues), the patterns shown in this section are only indirectly comparable to the patterns discussed in the previous section (section D.4). However, additional analysis suggested that the west-wide pattern shown in Figure D.25 corresponds to the domain-wide pattern in Section D.4, and the north-south pattern in D.25 corresponds to the north-south pattern in the earlier section. These findings confirm that the domain-wide and north-south patterns exhibit variance across a broad continuum of timescales.

Despite their differing time evolutions, we have shown that tree-ring reconstruc-

tions of PDSI and climate models are in agreement with each other in terms of the magnitude and patterns of low-frequency drought in WNA. This finding suggests that the spatial patterns of drought may reflect a deterministic response to certain aspects of the climate system that differ on interannual to centennial timescales. Understanding the dynamics of these patterns could, therefore, present the greatest potential for predicting climate decadal timescales.

D.5.3 Conclusion

The findings presented here allow us to confidently reject the null hypothesis that megadroughts arise simply from autocorrelation, spatial structure in the atmosphere, and sufficient time. Instead, more interesting mechanisms likely push substantial portions of western North America into severe droughts and keep those regions locked into those arid conditions for long periods of time. In contrast to the existing paradigm, we do not find evidence that these mechanisms need to be remote sources of climate forcing such as volcanic or solar activity, because a similar range of PDSI variability is seen in an unforced global climate model simulation. Dynamic boundary conditions do, nonetheless, enhance variance at the longest timescales.

Future work could expand our analysis of climate model data to include a larger spatial domain, which in turn could help elucidate the dynamics of decadal to centennial-scale variability in the models. Other models could also be diagnosed to determine how robust our results are across models.

D.6 Figures

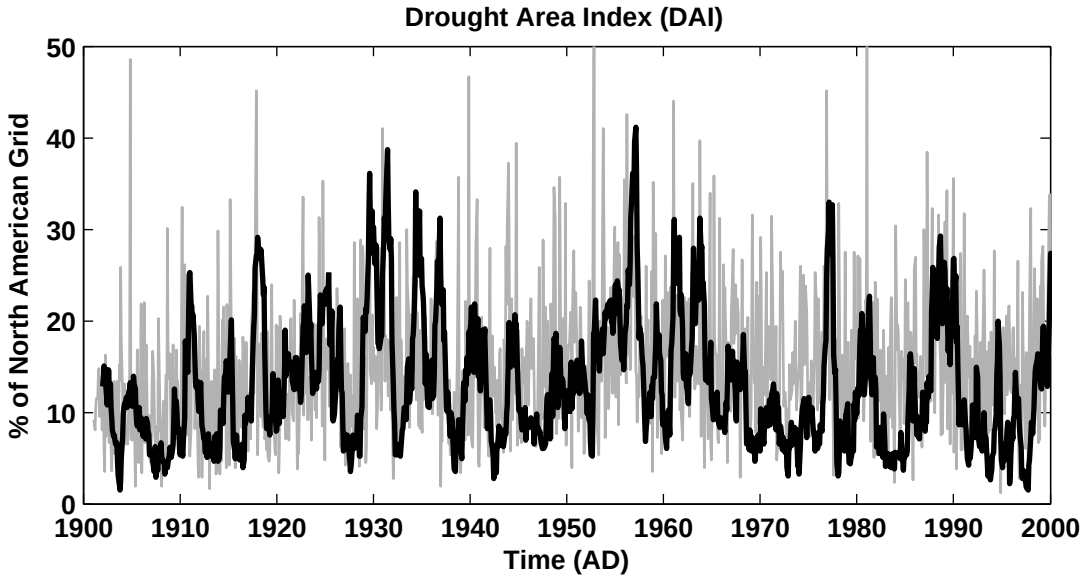

Figure D.1: Drought area index for 1mo (gray) and 12mo (black) 20th century SPI values, calculated as the percentage of the grid (y-axis) experiencing SPI values below -1 through time.

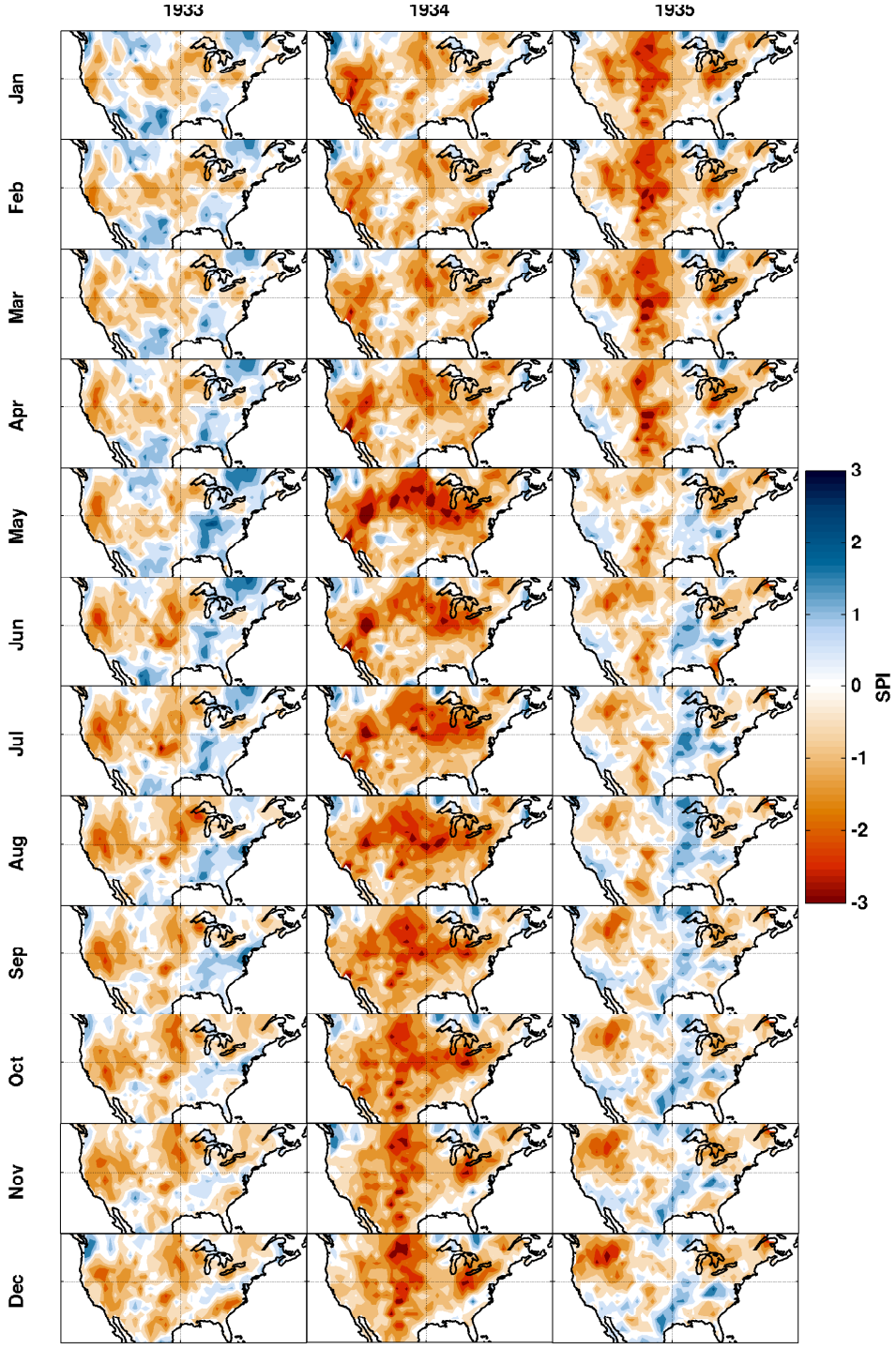

Figure D.2: Monthly values of the 12mo SPI calculated from CRU TS2.1 data. Coloring indicates the magnitude of SPI anomalies, which are normally distributed with unit variance and a mean of zero. Hence, a -3 value is at the 95% edge of the distribution and would be considered very rare (and hence very extreme) for that region. Here the evolution of one of the 1930s era "dust bowl" droughts centered on 1934.

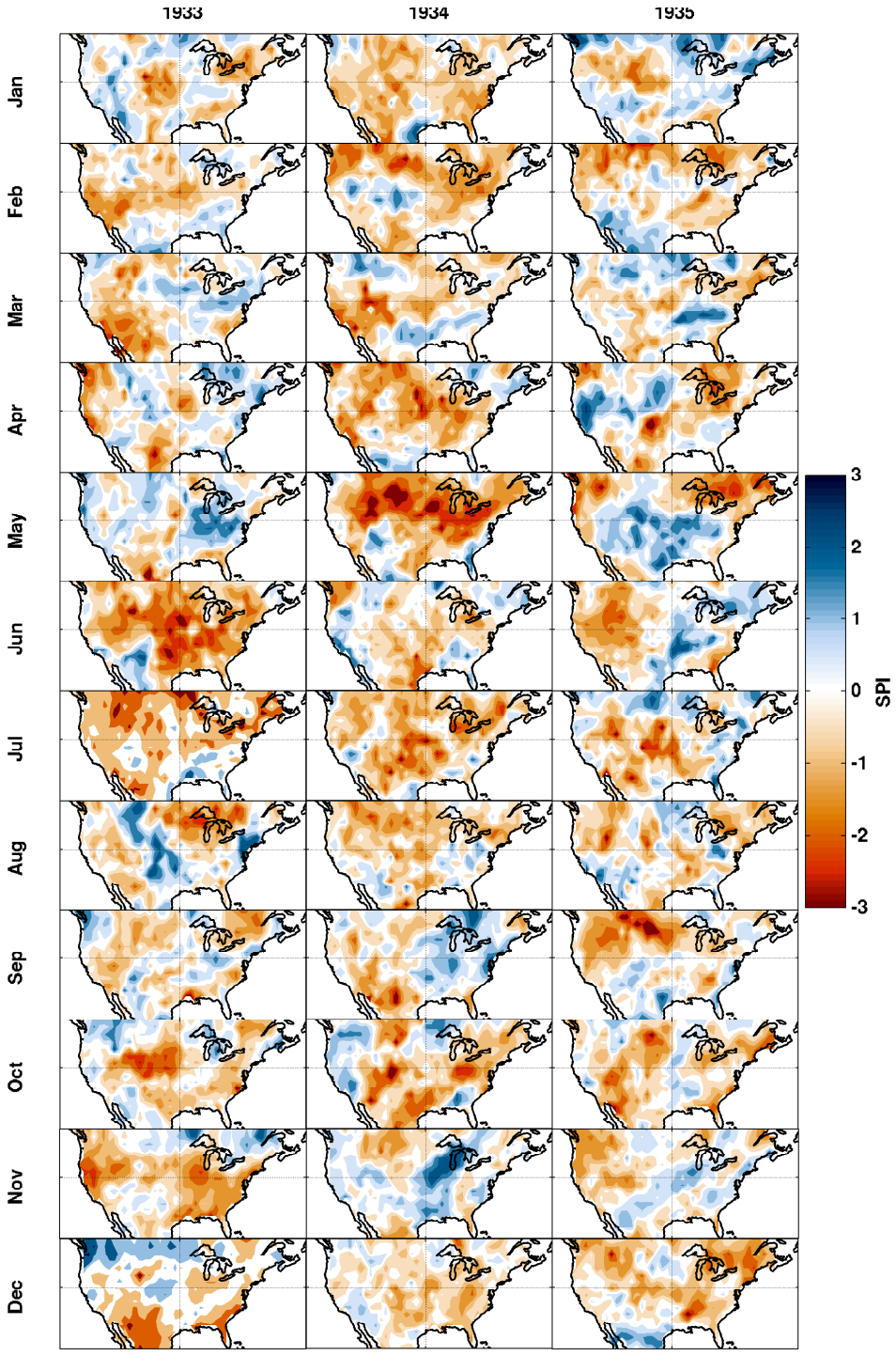

Figure D.3: Same as Figure D.2, but for the 1mo SPI.

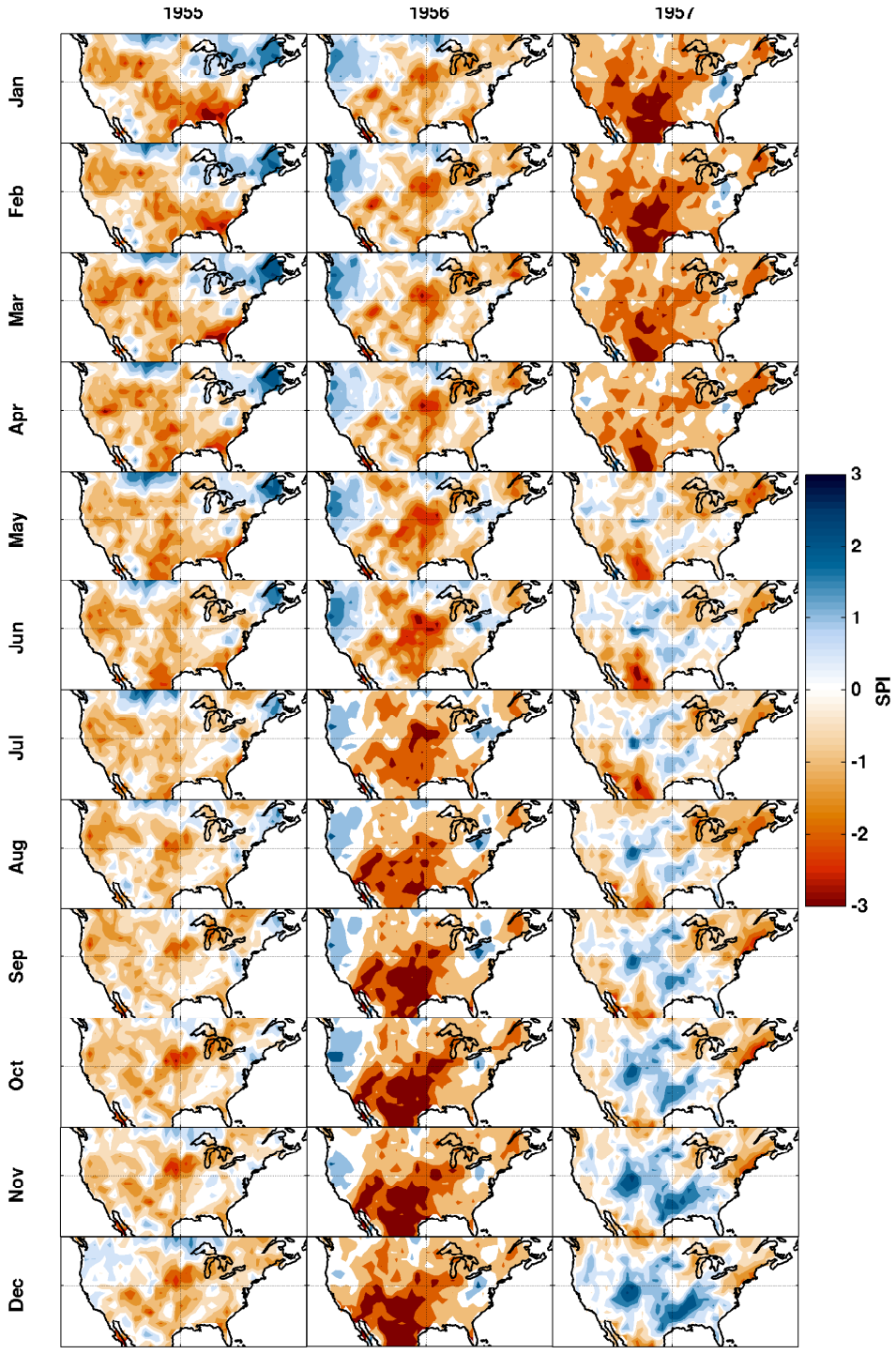

Figure D.4: Same as Figure D.2, but for 12mo SPI centered on the 1956 drought.

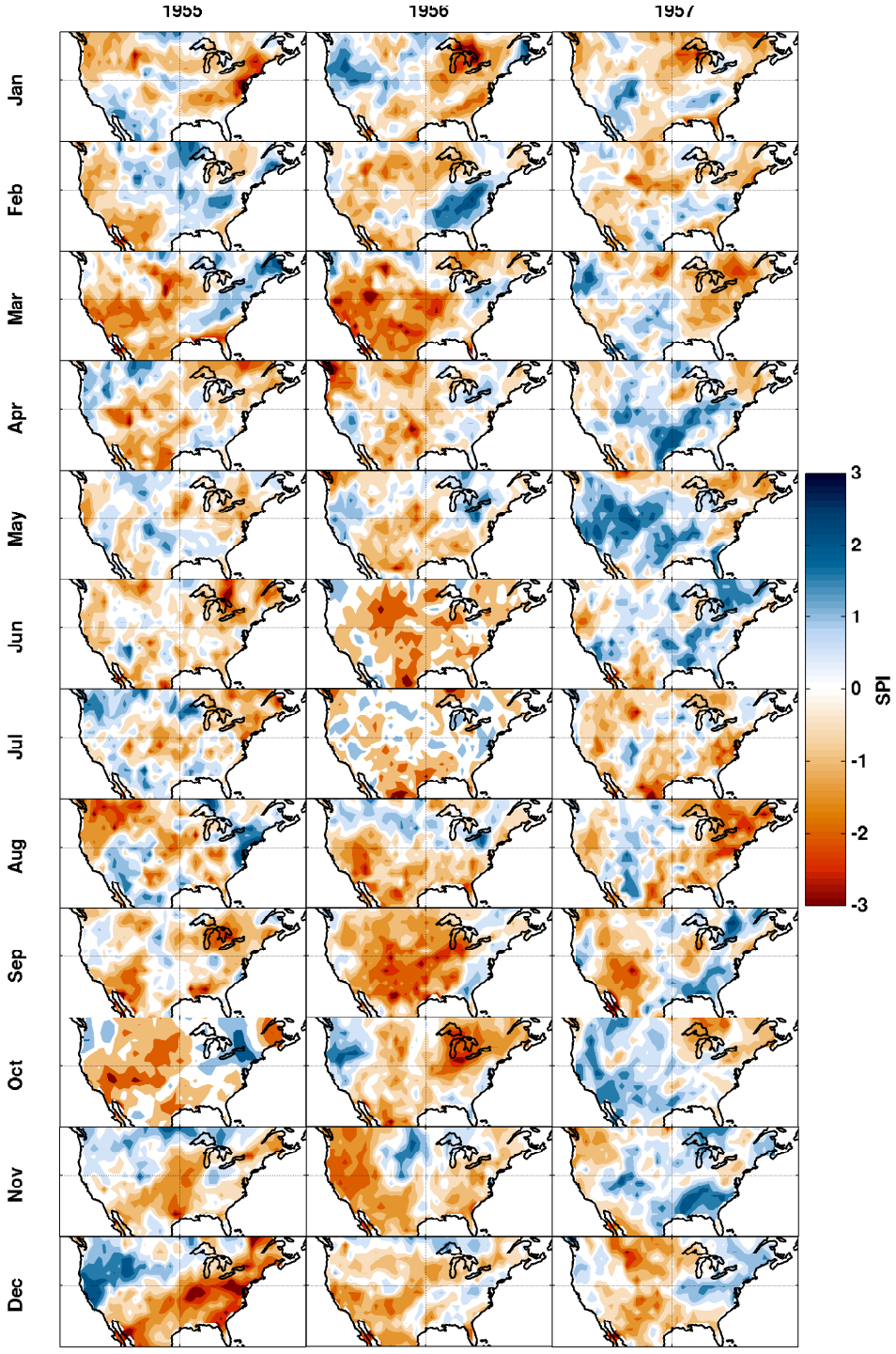

Figure D.5: Same as Figure D.4, but for 1mo SPI.

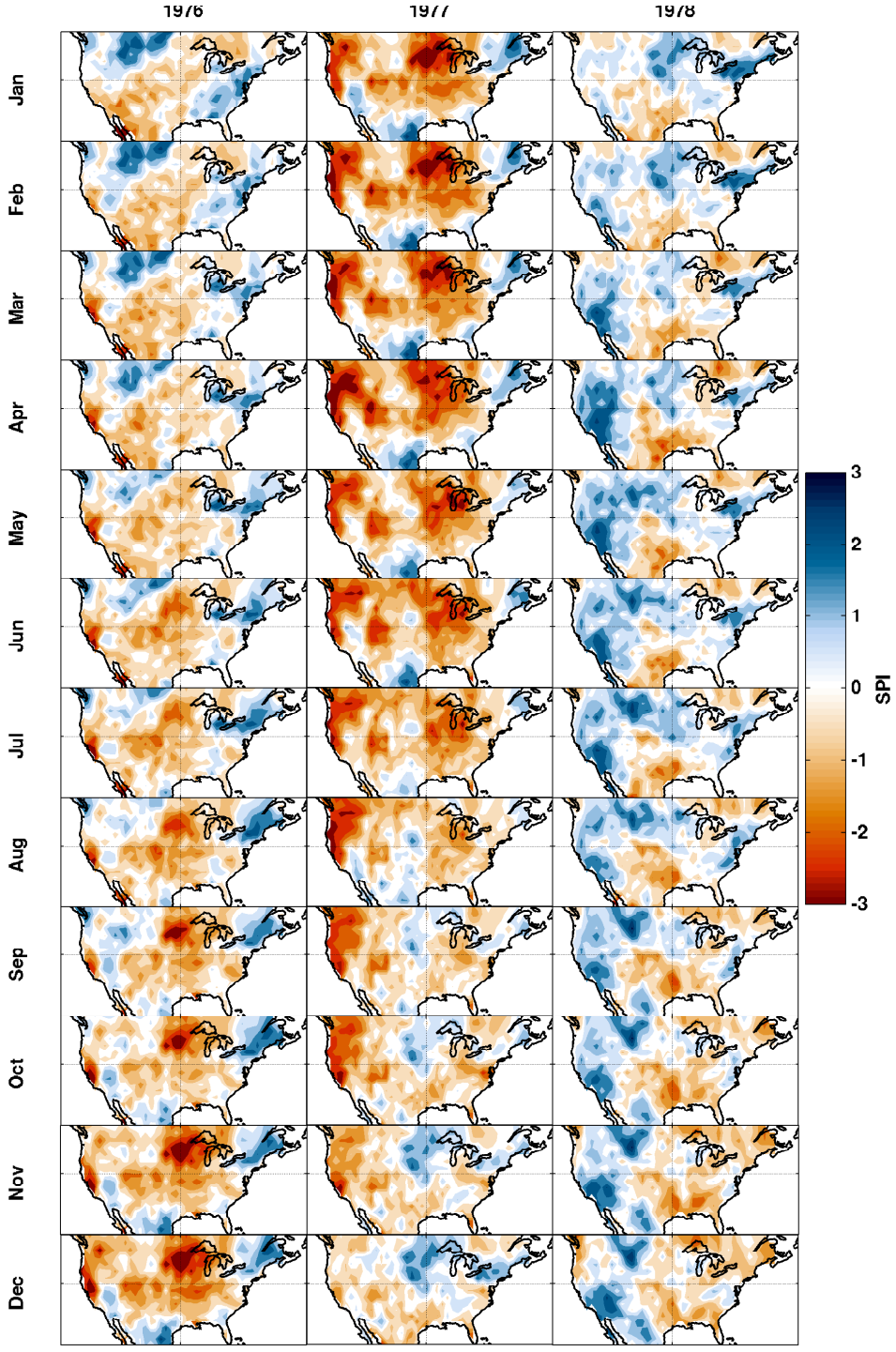

Figure D.6: Same as Figure D.2, but for 12mo SPI centered on the 1977 drought.

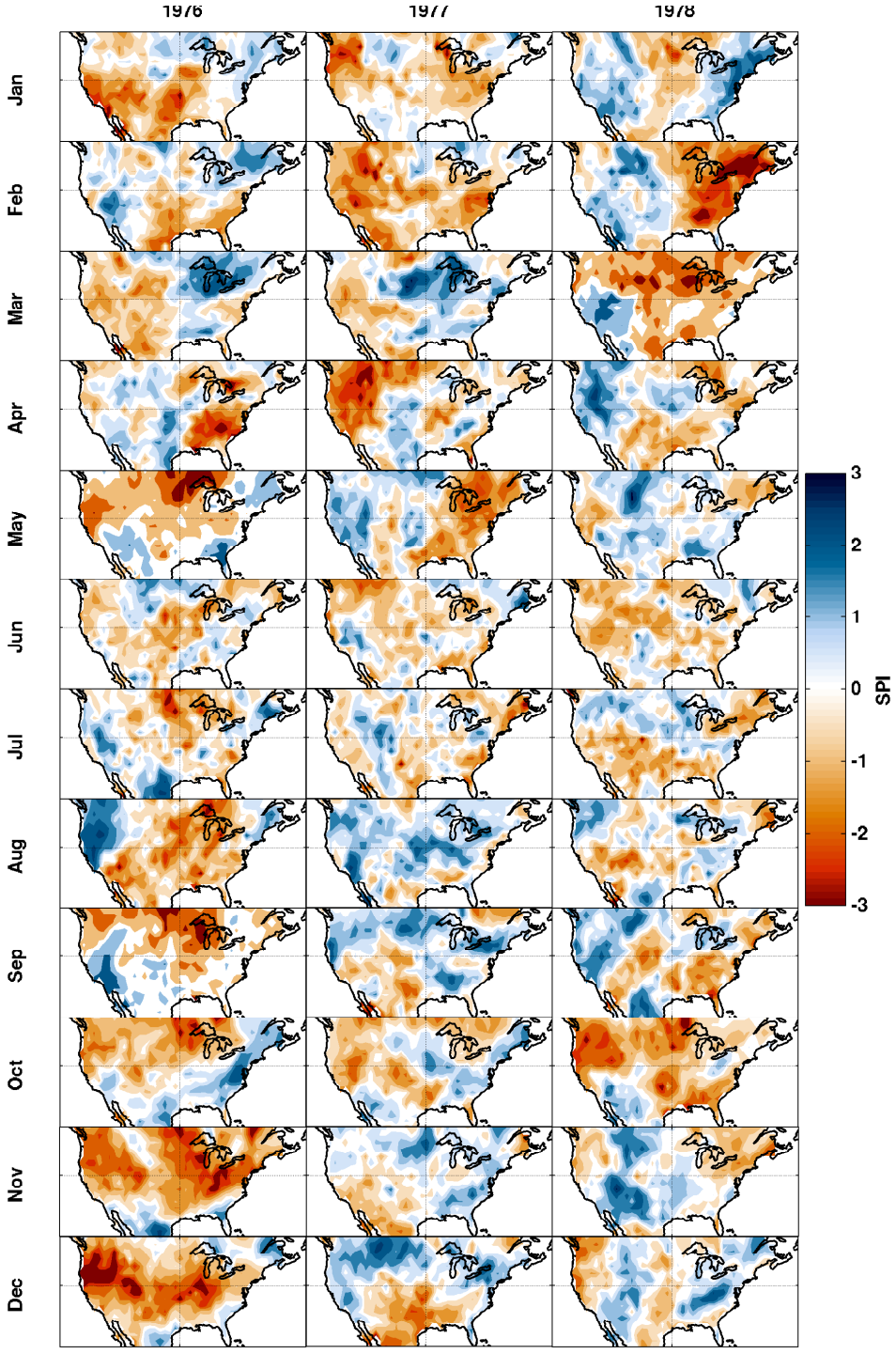

Figure D.7: Same as Figure D.6, but for 1mo SPI.

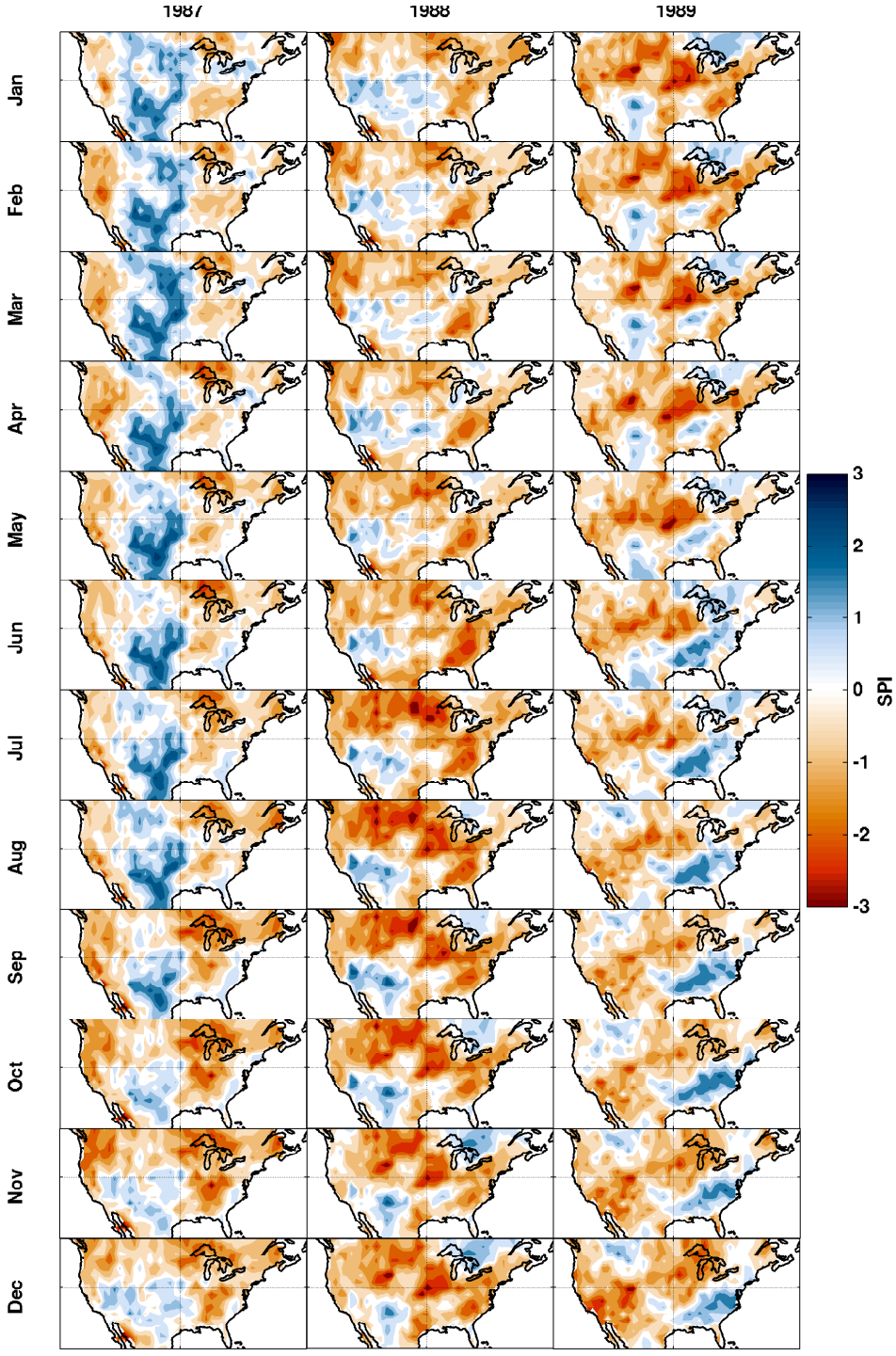

Figure D.8: Same as Figure D.2, but for 12mo SPI centered on 1988.

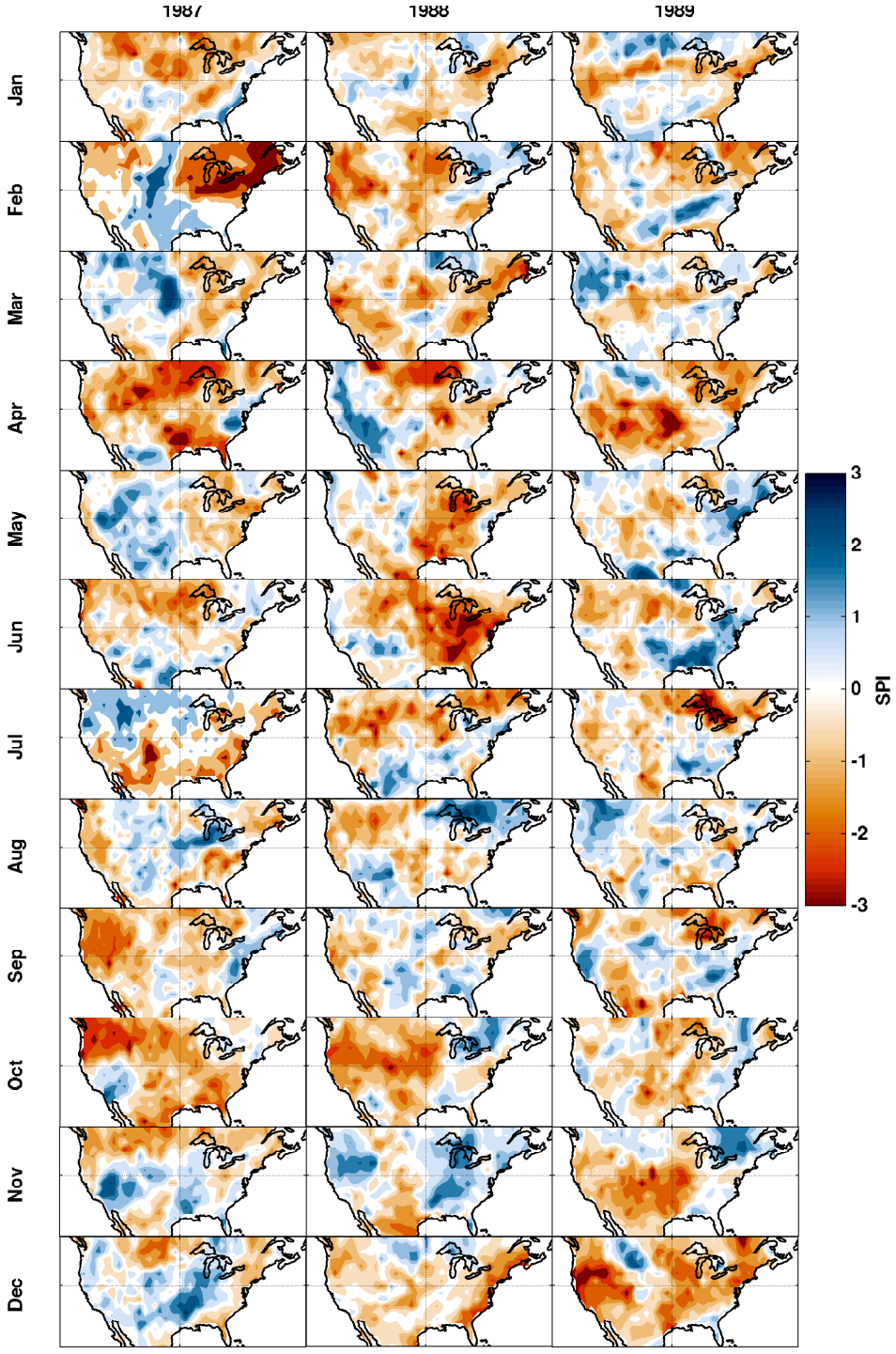

Figure D.9: Same as Figure D.8, but for 1mo SPI.

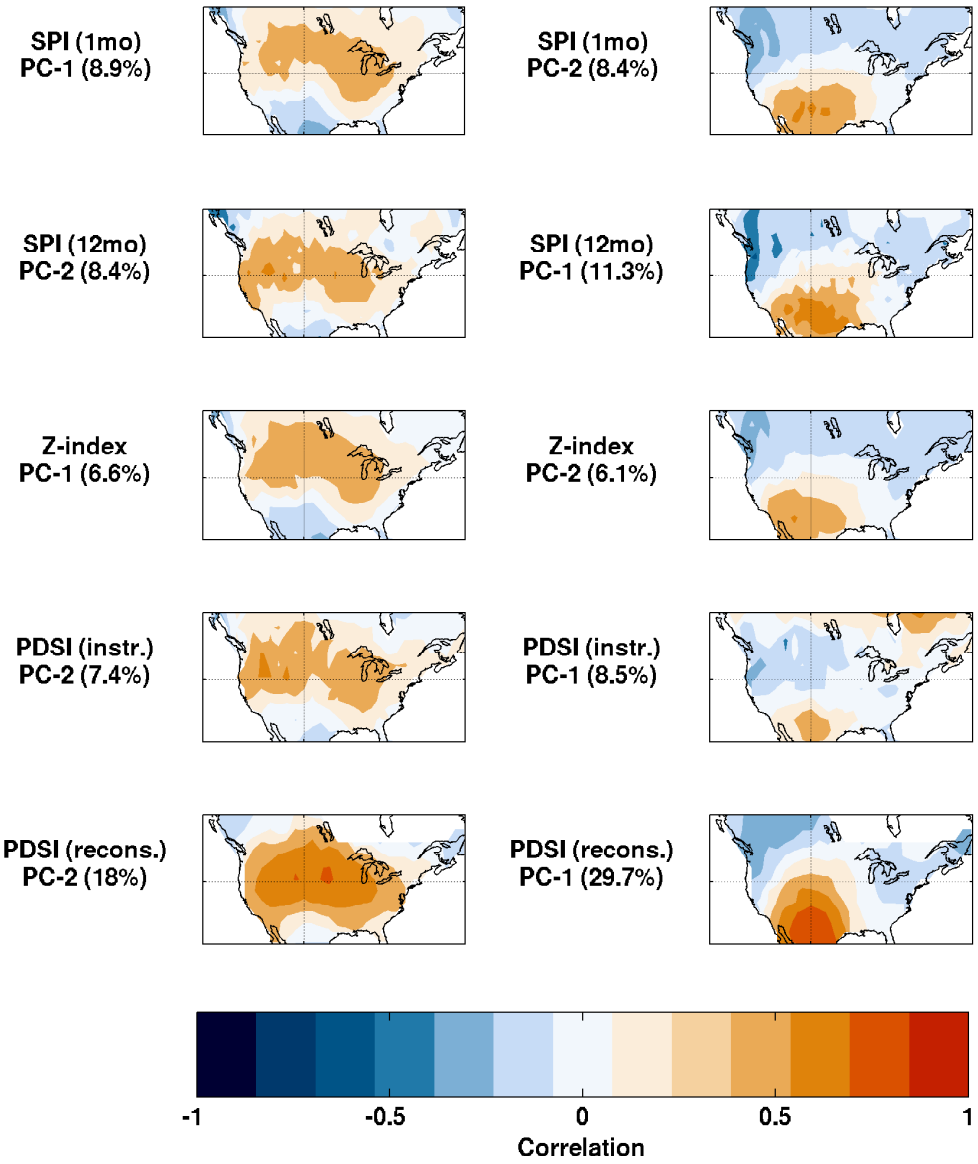

Figure D.10: Maps of correlation coefficients between the PC time series corresponding to the domain-wide pattern (left) an the north-south Pattern (right). The PC number and variance explained is indicated to the left of each map for each data product.

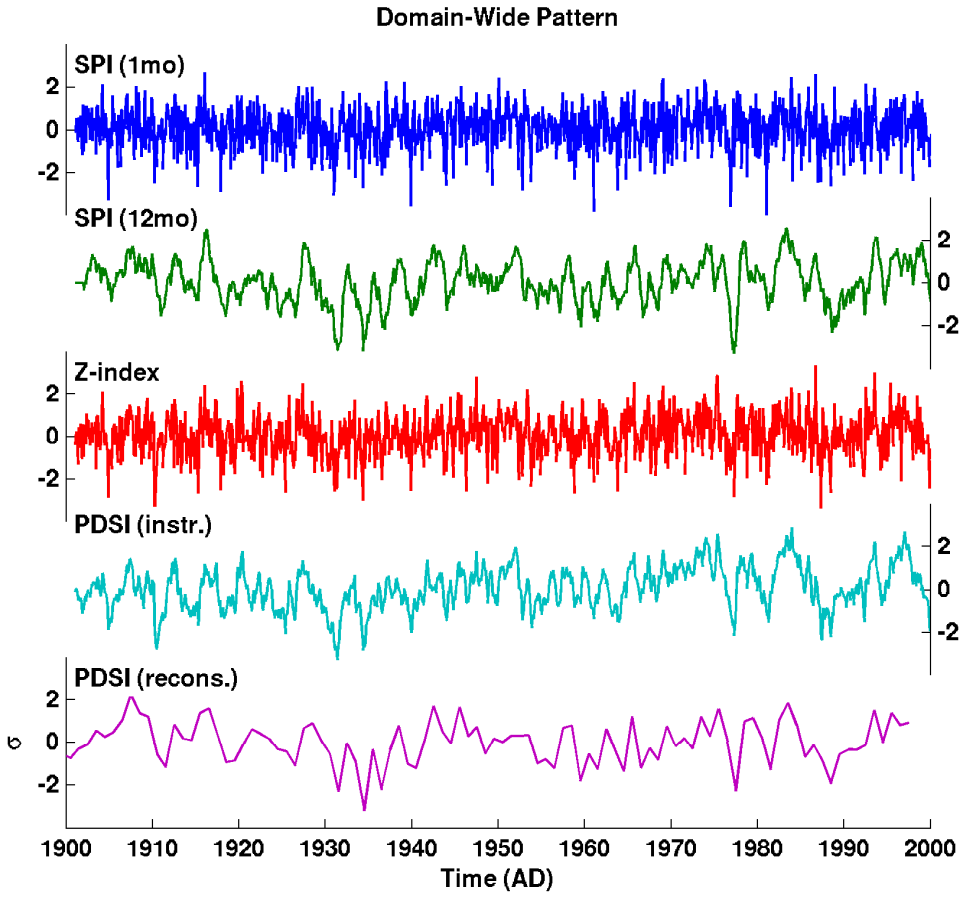

Figure D.11: PC time series associated with the domain wide pattern from each drought index considered here.

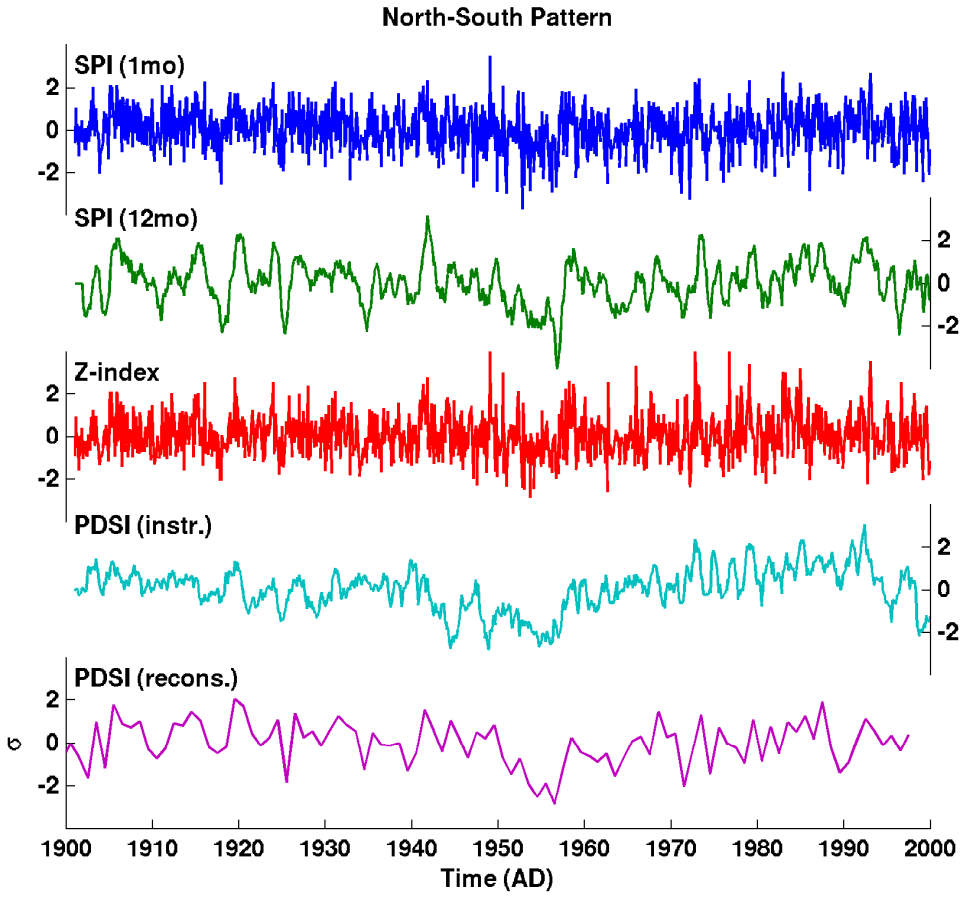

Figure D.12: Same as Figure D.11, but for the north-south pattern.

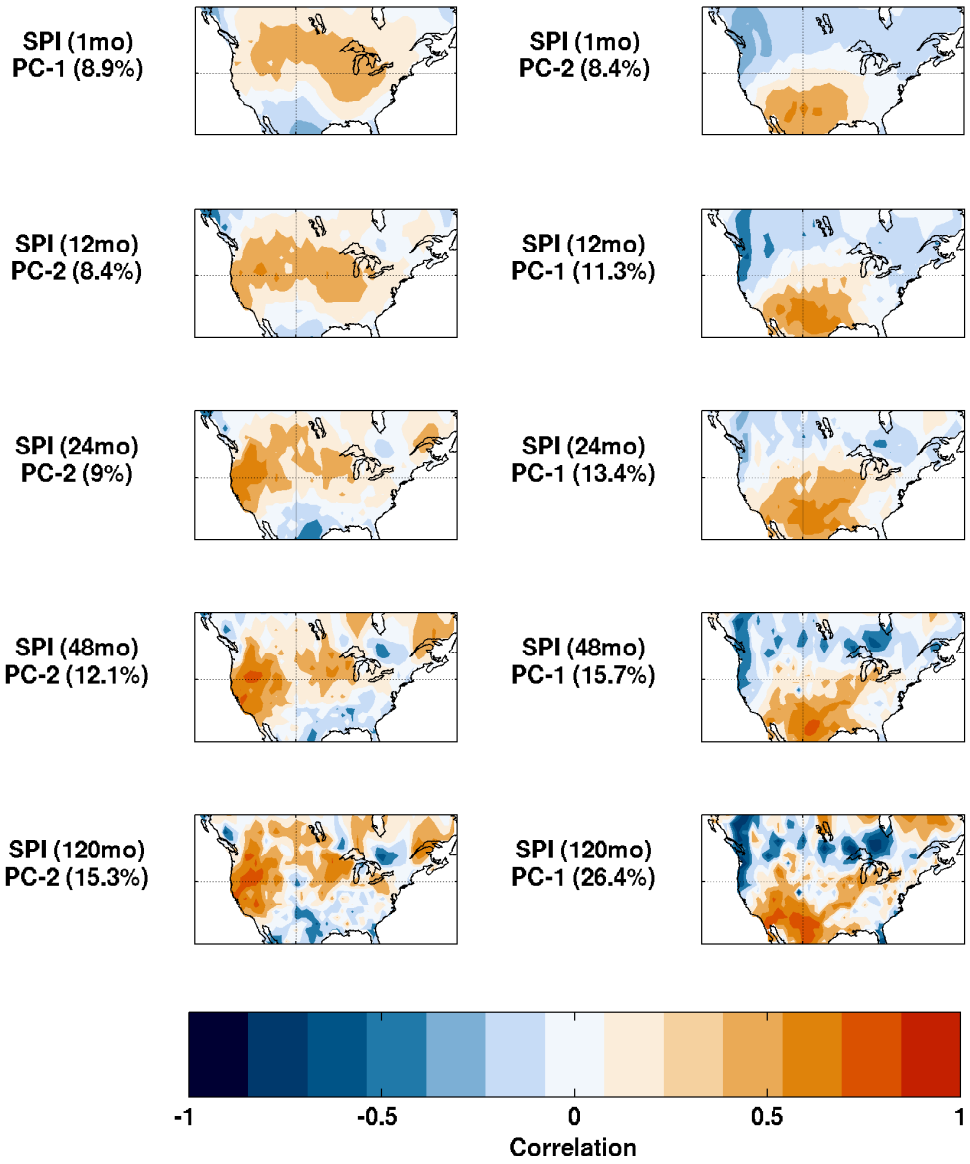

Figure D.13: Same as Figure D.10, but for the domain-wide (left) and north-south (right) patterns derived from SPI datasets calculated at different timescales. The SPI timescales are indicated on the left hand side of the maps along with the percentage of variance explained by the corresponding EOF (in parenthesis).

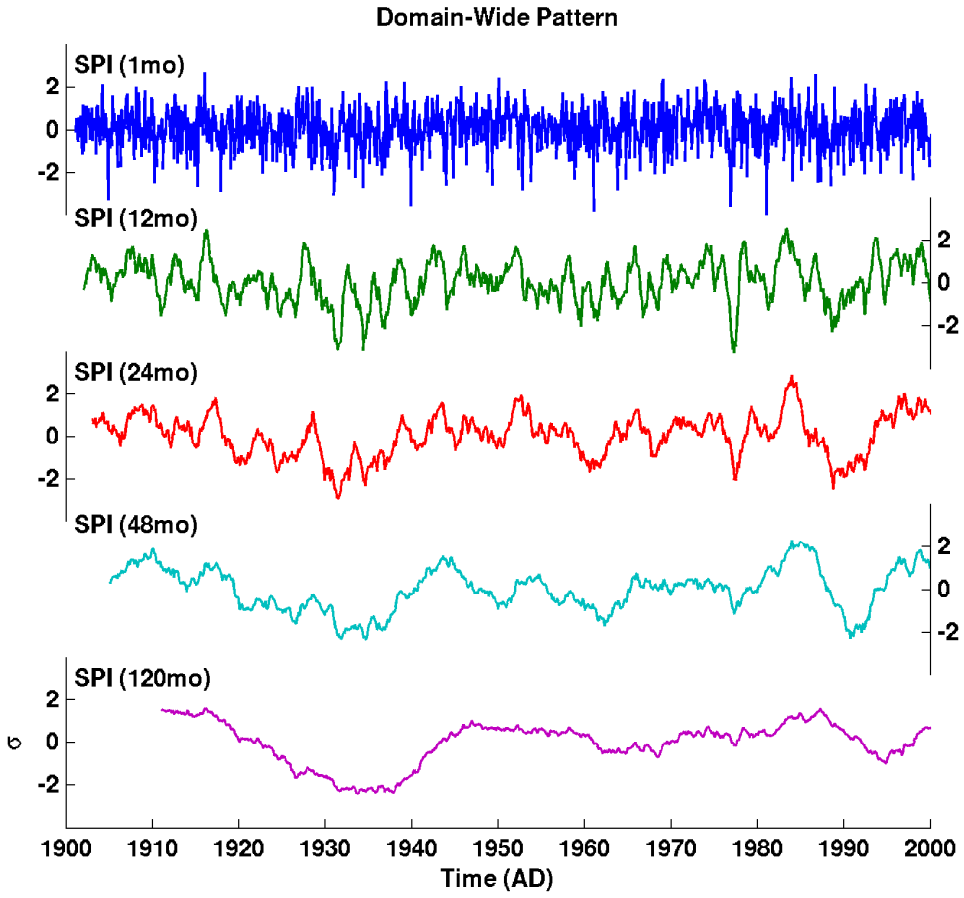

Figure D.14: Principal component time series of the domain-wide pattern from various SPI timescales (indicated on the top right of each panel).

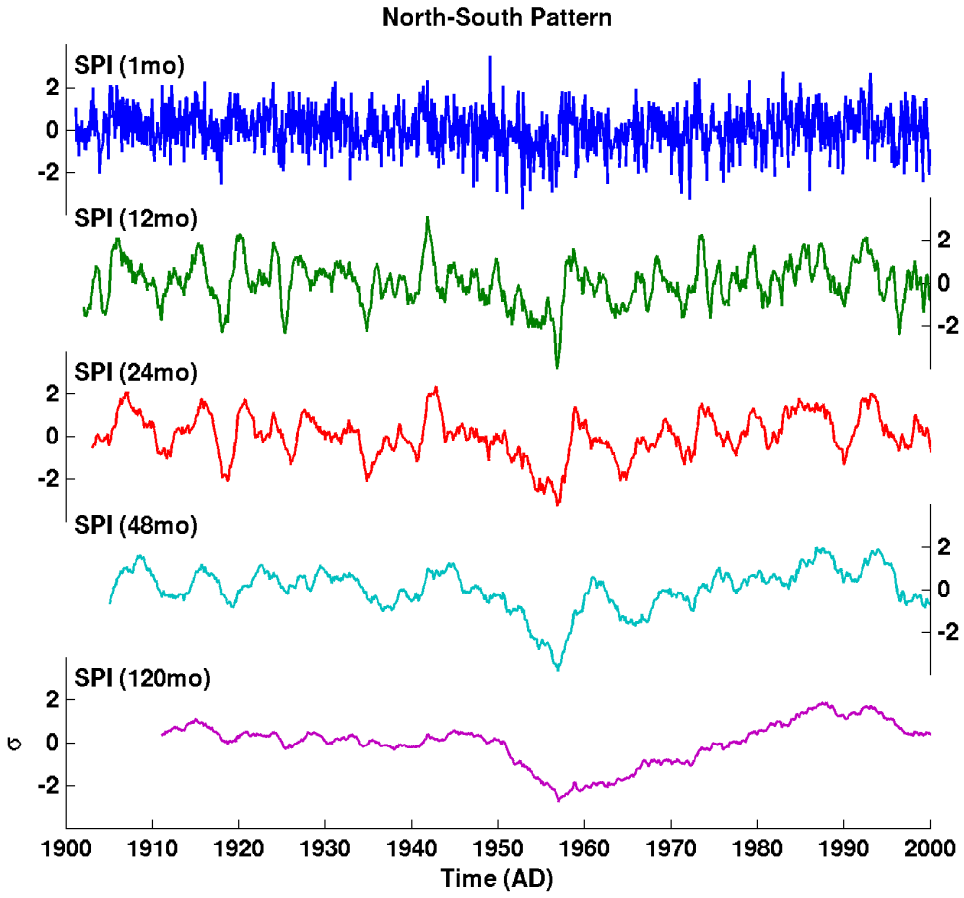

Figure D.15: Same as Figure D.14, but for the north-south Pattern.

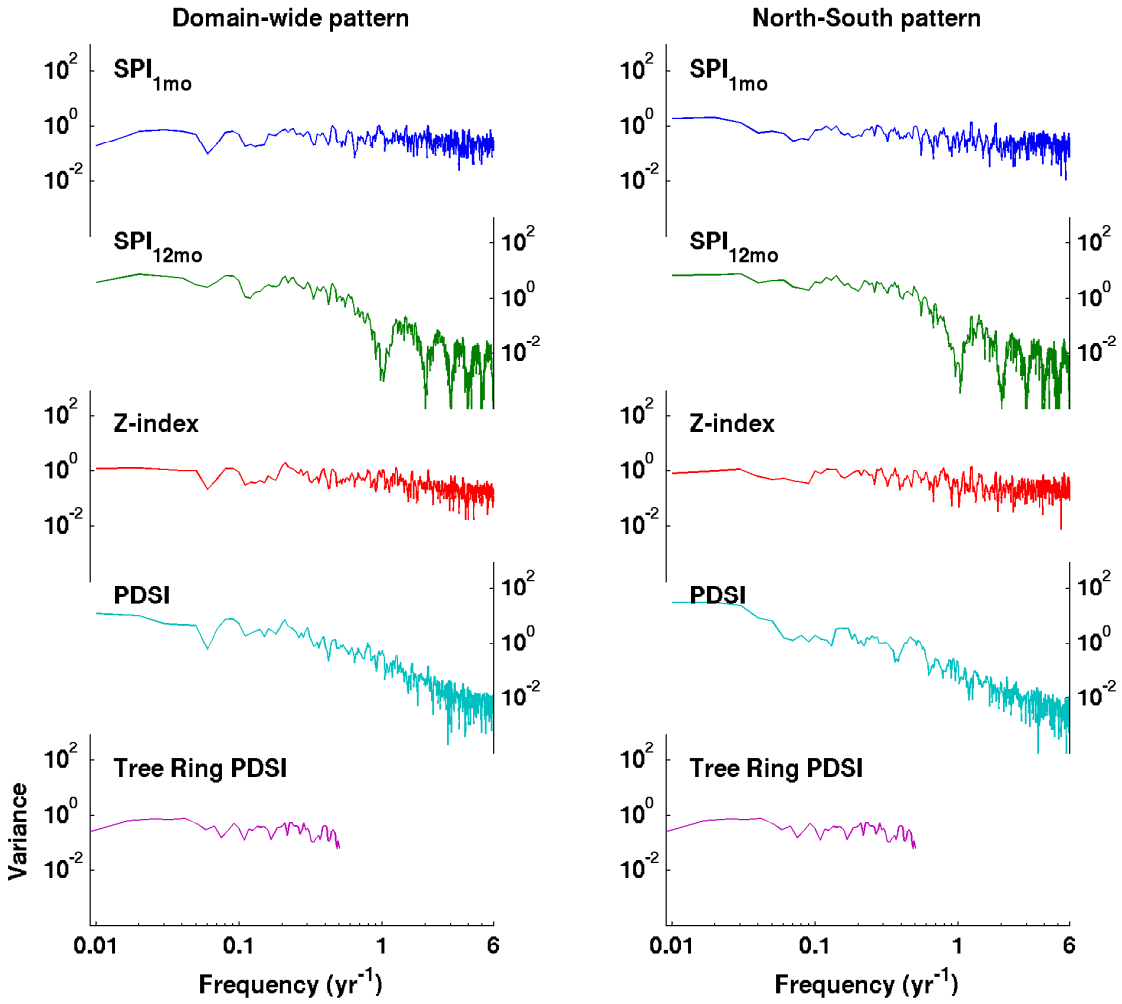

Figure D.16: Power spectra calculated from the five drought indices considered in Figure D.10 for the domain-wide pattern and the north-south pattern.

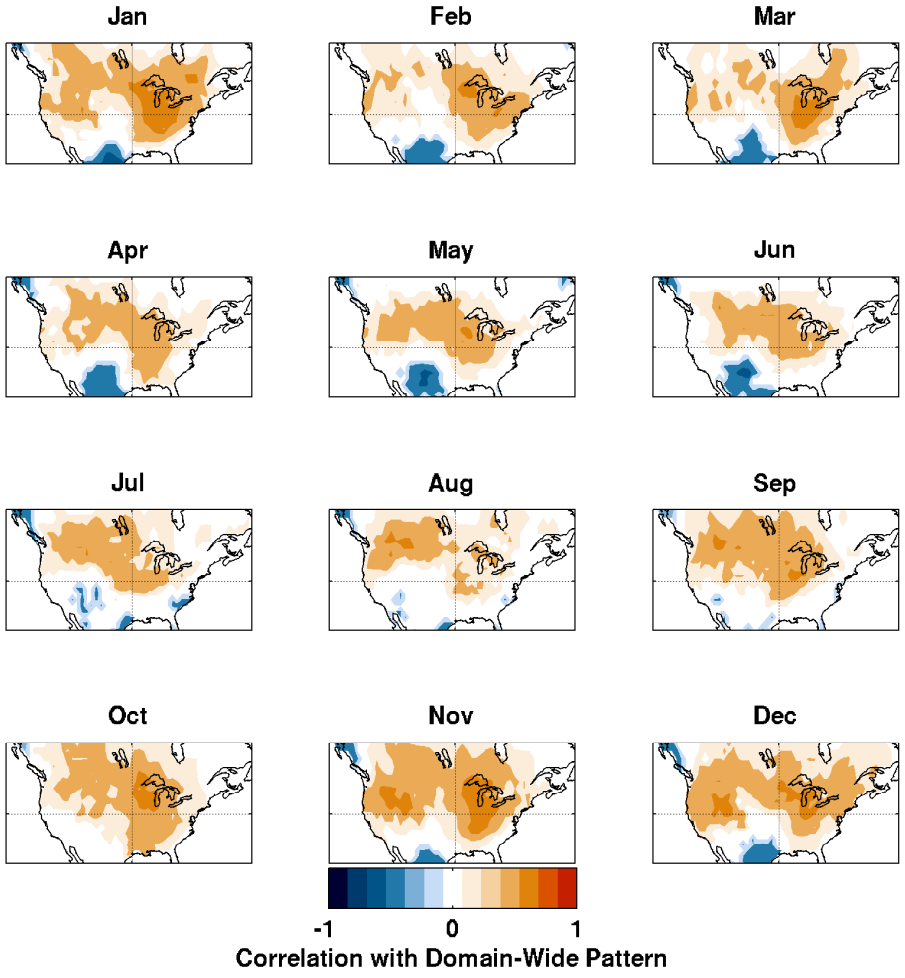

Figure D.17: Correlations between the 1mo PC of the domain-wide pattern and monthly raw 1mo SPI. Correlations are calculated from all years at each month indicated on the 12 panels above. All shaded correlations are significant at the 95% confidence limit.

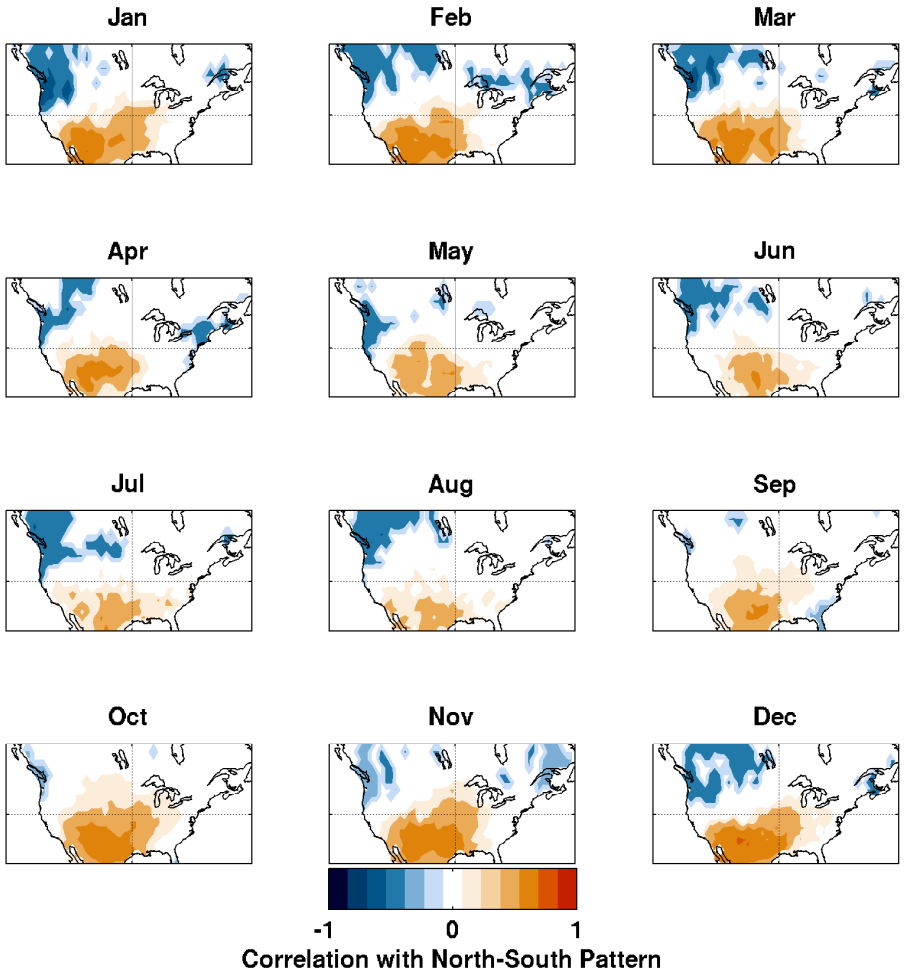

Figure D.18: Same as D.17, but for the north-south pattern.

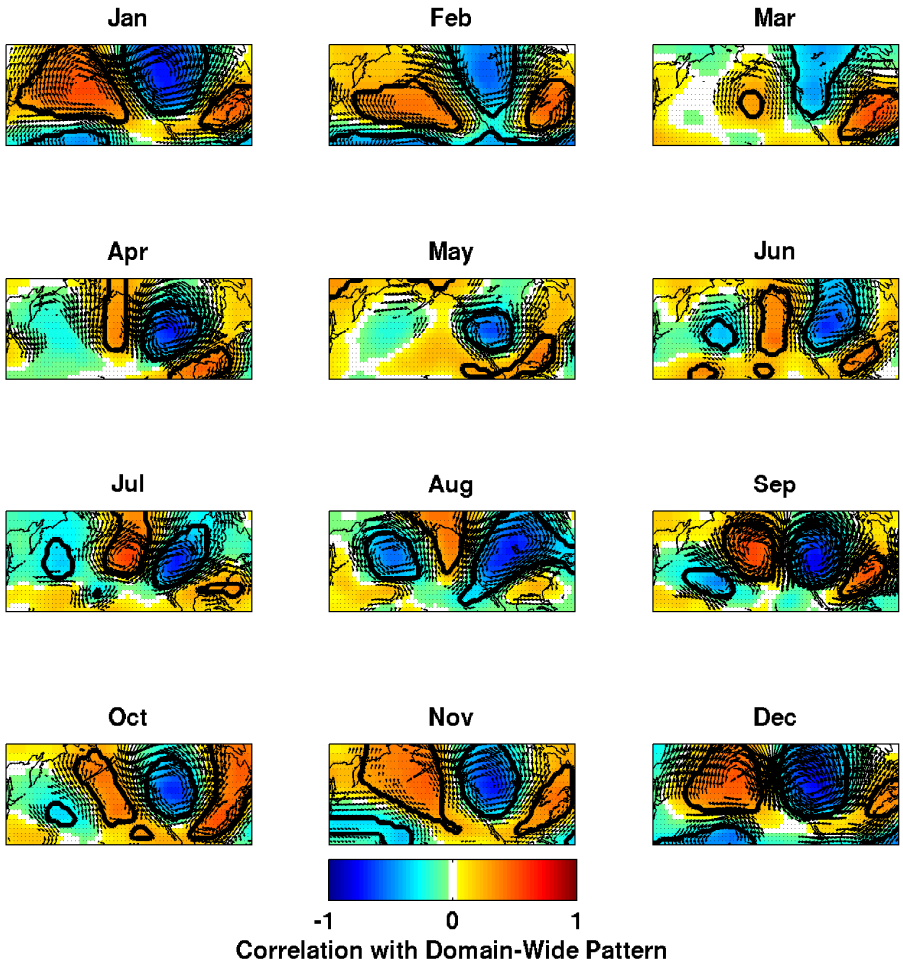

Figure D.19: Correlations between the values of the domain-wide pattern and monthly 250mb heights, meridional, and zonal winds from NCEP reanalysis data. As in Figure D.17, correlations are calculated between the values of the PC and the spatial fields at each month indicated in the 12 panels. Thick black lines indicate regions where correlations are significant at the 95% confidence limit.

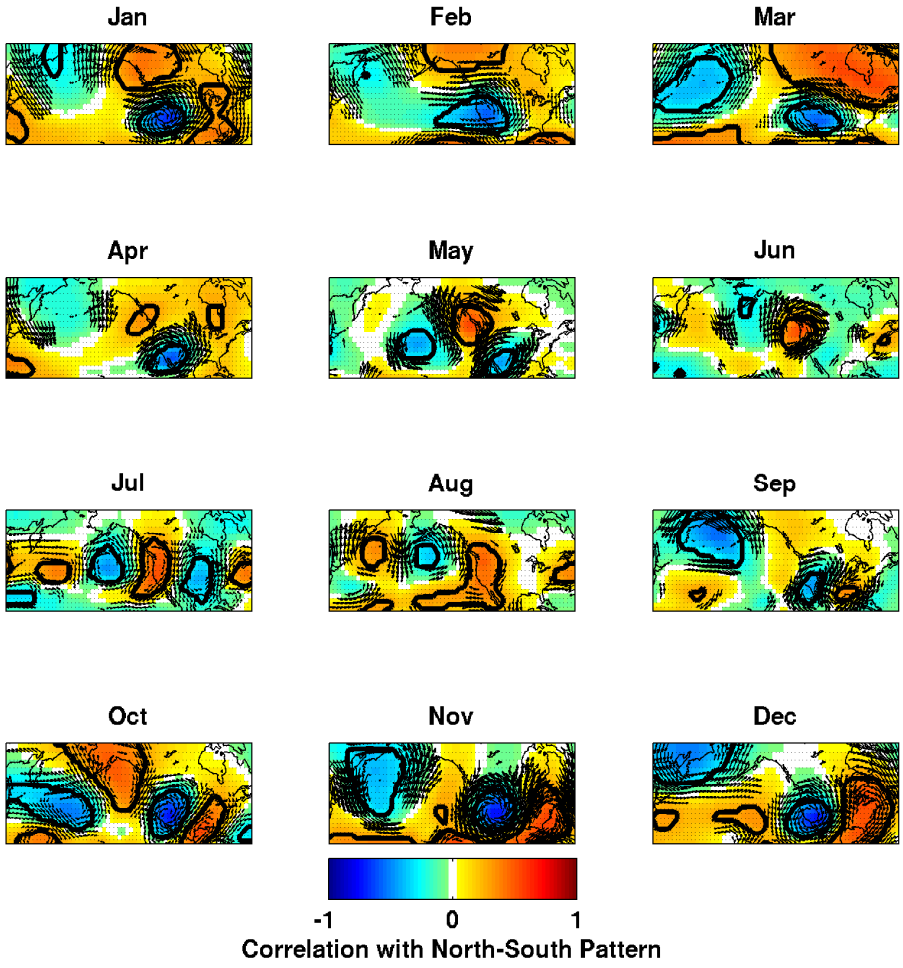

Figure D.20: Same as D.19, but for the north-south pattern.

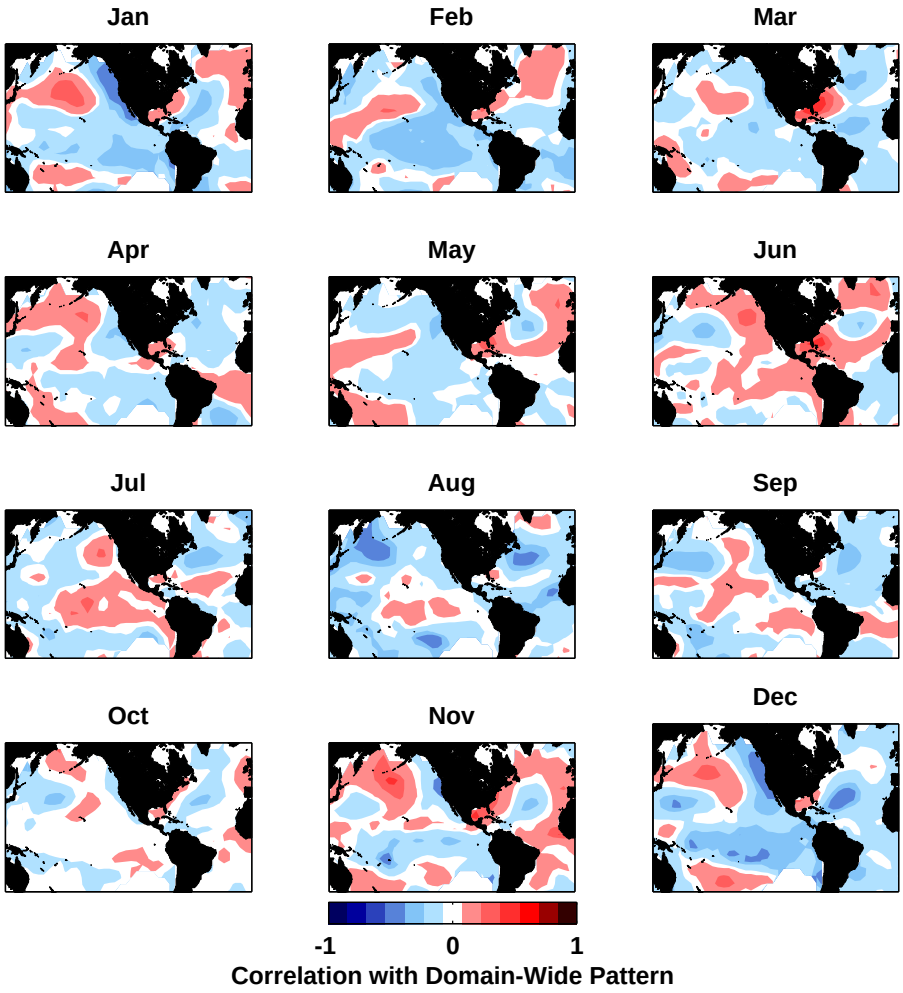

Figure D.21: Same as D.19, but for the correlations between the domain-wide pattern PC and SST (Kaplan et al., 1998).

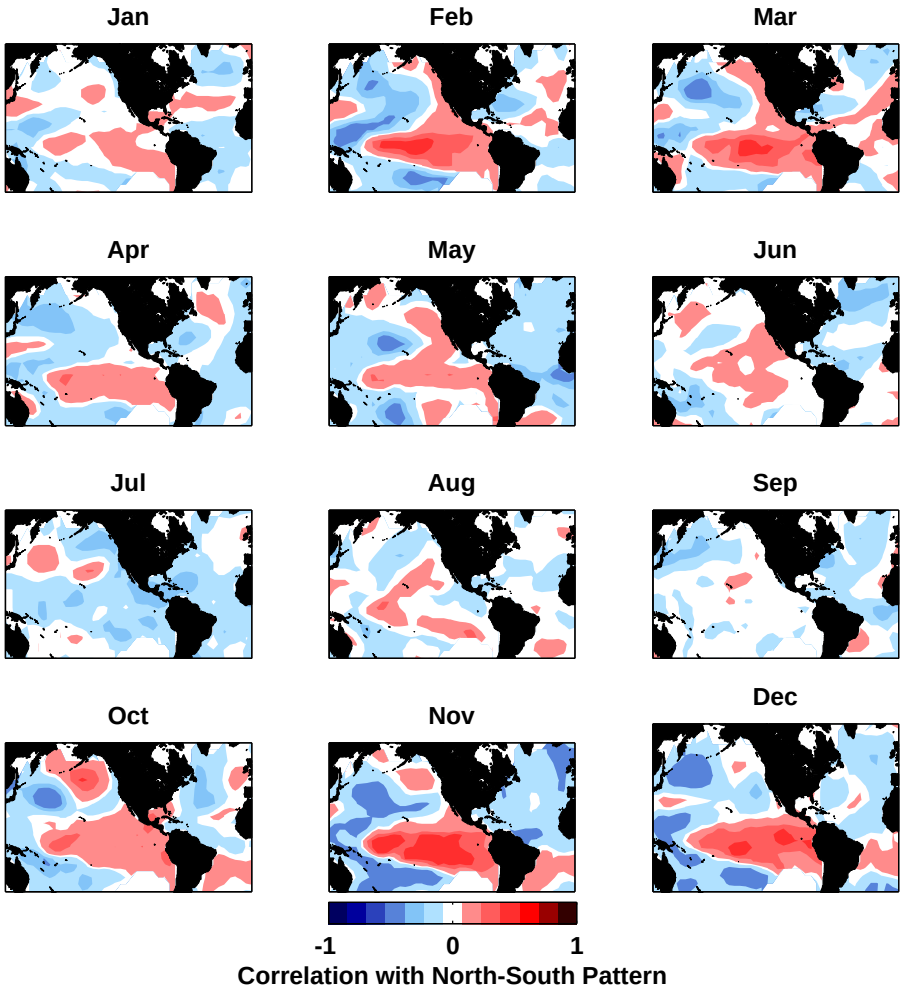

Figure D.22: Same as D.21, but for the correlations between the north-south pattern PC and SST (Kaplan et al., 1998).

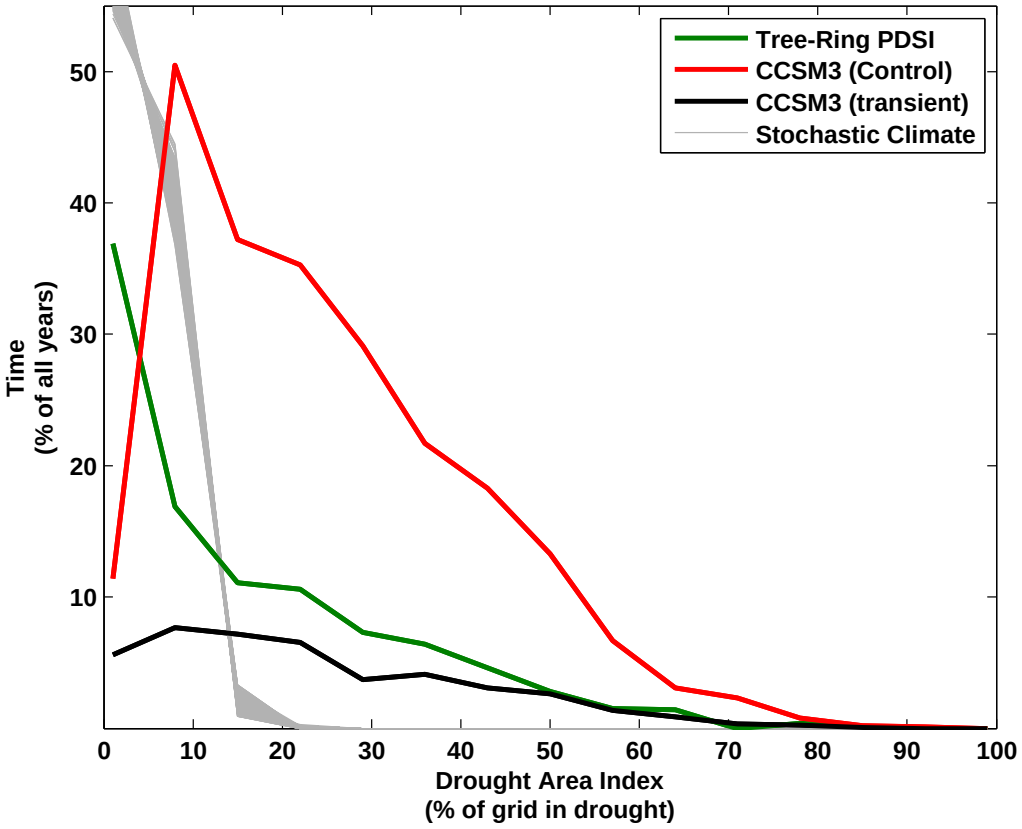

Figure D.23: Histograms of millennial-scale drought area index (DAI) values from from paleoclimate and climate model data, calculated for western North America (24N to 52N, 125W to 95W). Drought area indices were calculated (as in Cook et al. (2004)) by summing the number of grid points with PDSI values below -1, and expressing that count as a percentage of the total number of grid points through time. Histograms were then generated by tallying the total number of years spent at different DAI values (which are shown on the x-axis). To compare millennial-scale data of different lengths, we express these counts as percentages of all years (values on the y-axis). The gray lines show the results from our 100 "null climate" simulations'; the green line shows the distribution from reconstructed PDSI; the red line shows the DAI distribution from on 1,000 year long pre-industrial control run from CCSM3; the black line shows the DAI distribution from a 2,3000 year long mid-Holocene transient simulation.

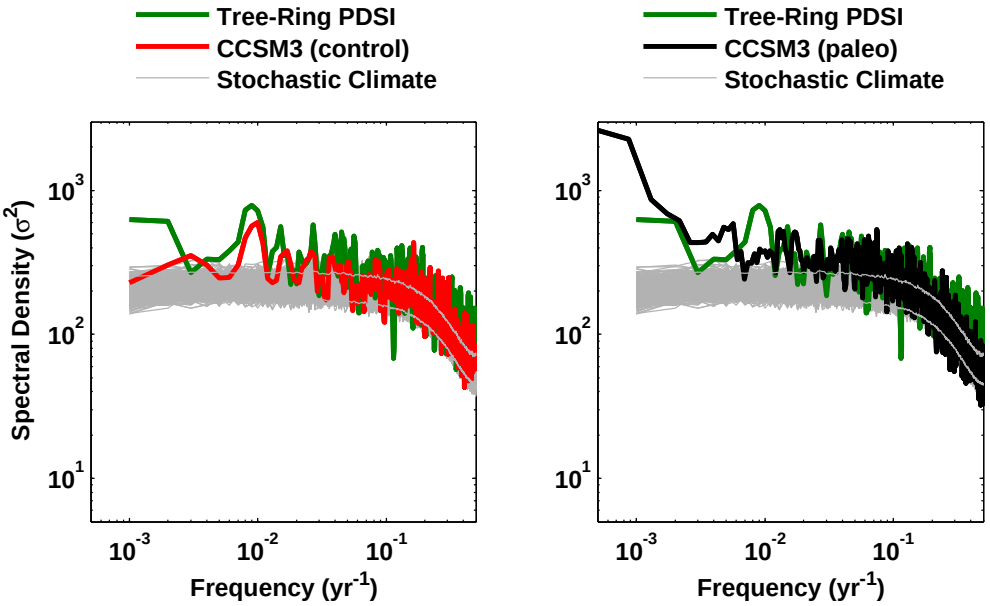

Figure D.24: Results from performing MTM-SVD (Mann and Park, 1999) on PDSI reconstructed from tree-ring data and calculated from climate model data as well as our null climate data. Shown here are the MTM-SVD spectra from each dataset; each point in the power spectrum depicts the amount of variance (the singular value from SVD at frequency) associated with a corresponding spatial pattern. Gray lines show spectra from MTM-SVD performed on PDSI from all 1000 null climate statistical simulations, and green lines show the spectra from MTM-SVD performed on gridded, reconstructed PDSI over the last millennium. The left panel also shows the MTM-SVD spectrum from a 1000 year CCSM3 control run (red), while the left panel shows the spectrum from a long transient (e.g., evolving boundary conditions) simulation on CCSM3 (black).

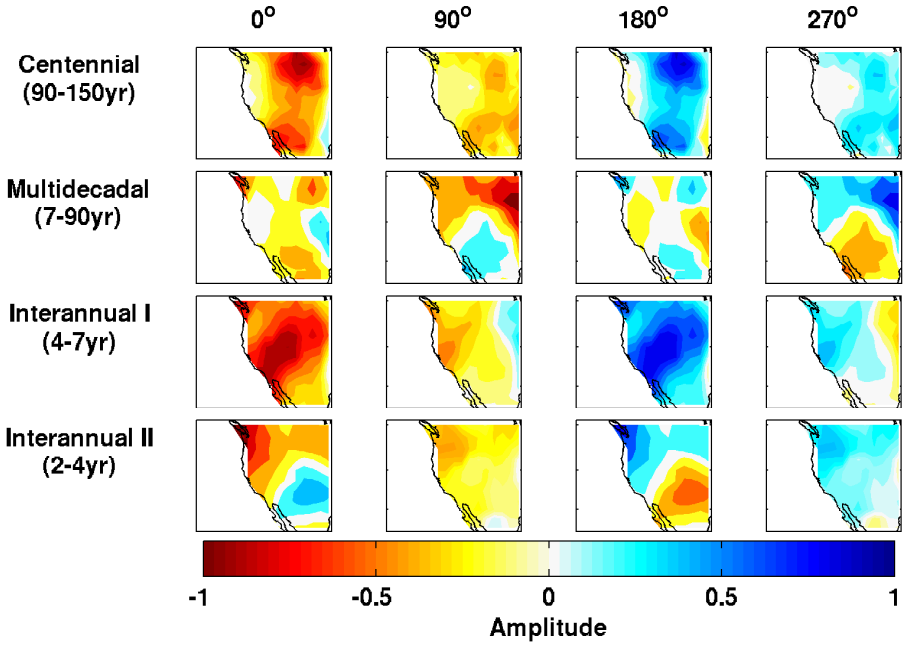

Figure D.25: Spatial patterns of drought variability (reconstructed PDSI from Cook et al., 2004) associated with different timescales of variability estimated from MTM-SVD. Here the maps are shown as an average (weighted by the variance in the power spectrum) of the MTM-SVD spectral EOFS over a range of frequencies. Since raw MTM-SVD spectral EOFs are complex numbers, which can be hard to visualize, we show their time evolution here in "phase-space" by calculating the amplitude of each spectrally-averaged pattern as it completes one cycle (passing from zero through 360°). The units are normalized so that the maximum value any point can have is 1.

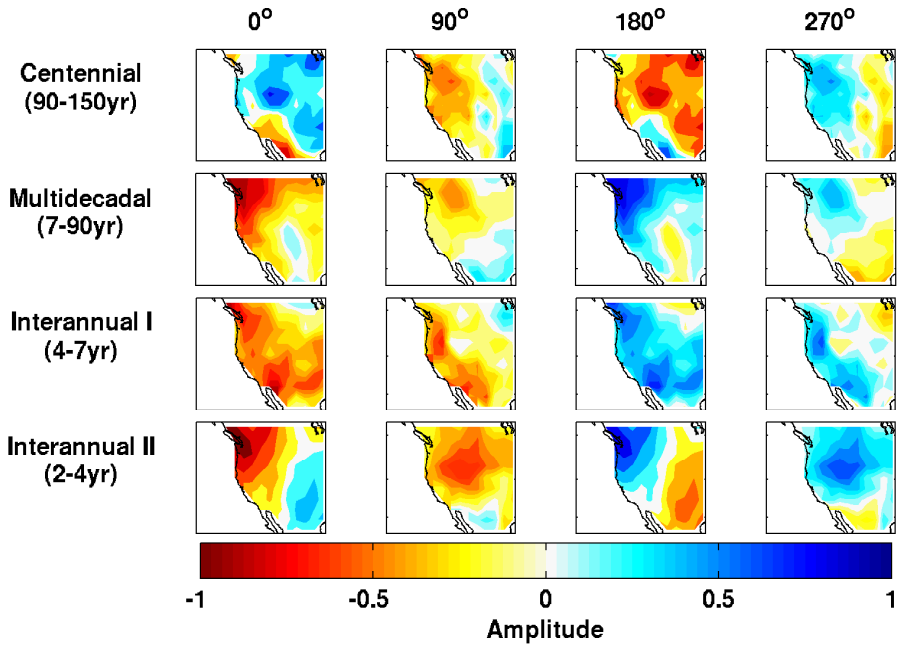

Figure D.26: Same as figure D.25, but for a long pre-industrial control run.

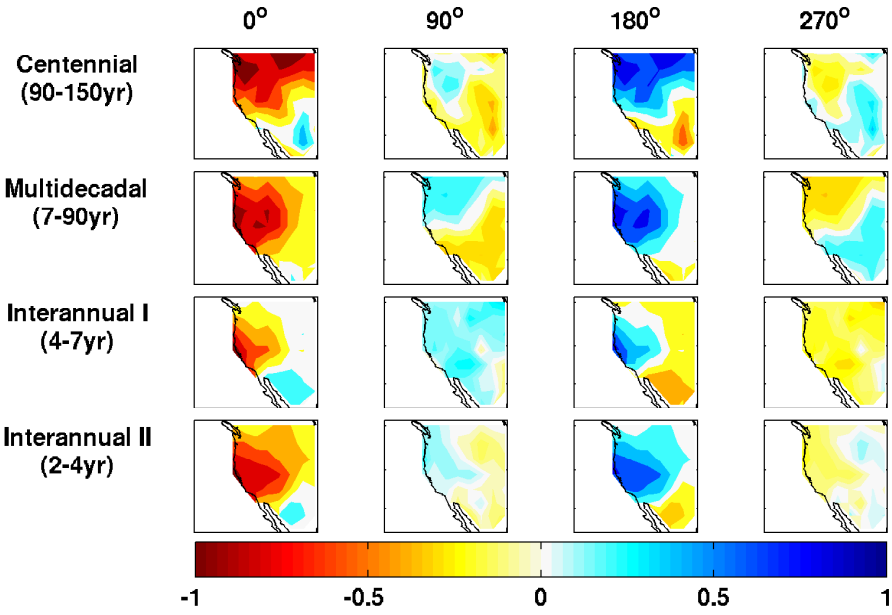

Figure D.27: Same as figure D.25, but for a long forced (mid-Holocene) transient simulation.

REFERENCES

Alley, W. M. (1984). The Palmer Drought Severity Index - Limitations and Assumptions. *Journal of Climate and Applied Meteorology*, **23**(7), pp. 1100–1109.

Ammann, C. M., F. Joos, D. S. Schimel, B. L. Otto-Bliesner, and R. A. Tomas (2007). Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(10), pp. 3713–3718.

Andreadis, K. M., E. A. Clark, A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier (2005). Twentieth-century drought in the conterminous United States. *Journal of Hydrometeorology*, **6**(6), pp. 985–1001.

Asmerom, Y., V. Polyak, S. Burns, and J. Rassmussen (2007). Solar forcing of Holocene climate: New insights from a speleothem record, southwestern United States. *Geology*, **35**(1), pp. 1–4.

Ault, T., C. J. E., J. Overpeck, G. T. Pederson, S. St. George, B. Otto-Bliesner, C. Woodhouse, and C. Deser (????). The continuum of drought variability in North America: a proxy-model comparison. *In preparation*.

Ault, T., A. Macalady, G. Pederson, and J. B. andM. Schwartz (2011). Northern Hemisphere modes of variability and the timing of spring in western North America. *Journal of Climate*.

Ault, T. R., J. E. Cole, M. N. Evans, H. Barnett, N. J. Abram, A. W. Tudhope, and B. K. Linsley (2009). Intensified decadal variability in tropical climate during the late 19th century. *Geophysical Research Letters*, **36**, p. L08602. doi:10.1029/2008GL036924.

Ault, T. R. and S. St George (2010). The Magnitude of Decadal and Multidecadal Variability in North American Precipitation. *Journal of Climate*, **23**(4), pp. 842–850. ISSN 0894-8755. doi:10.1175/2009JCLI3013.1.

Barry, R. G. and R. J. Chorley (1998). *Atmosphere, Weather, and Climate*. Routledge, New York, 8 edition.

Bartlett, M. S. (1978). *An introduction to stochastic processes, with special reference to methods and applications*. CUP Archive. ISBN 0521215854, 9780521215855.

Benson, L., M. Kashgarian, R. Rye, S. Lund, F. Paillet, J. Smoot, C. Kester, S. Mensing, D. Meko, and S. Lindstrom (2002). Holocene multidecadal and multicentennial droughts affecting Northern California and Nevada. *Quaternary Science Reviews*, **21**(4-6), pp. 659–682.

Benson, L., B. Linsley, J. Smoot, S. Mensing, S. Lund, S. Stine, and A. Sarna-Wojcicki (2003). Influence of the Pacific Decadal Oscillation on the climate of the Sierra Nevada, California and Nevada. *Quaternary Research*, **59**(2), pp. 151–159.

Betancourt, J. L., H. D. Grissino-Mayer, M. W. Salzer, and T. W. Swetnam (2002). A test of "annual resolution" in stalagmites using tree rings. *Quaternary Research*, **58**(2), pp. 197–199.

Blaauw, M. (2010). Methods and code for 'classical' age-modelling of radiocarbon sequences. *Quaternary Geochronology*, **5**(5), pp. 512–518.

Buckley, B. M., K. J. Anchukaitis, D. Penny, R. Fletcher, E. R. Cook, M. Sano, L. C. Nam, A. Wichienkeeo, T. T. Minh, and T. M. Hong (2010). Climate as a contributing factor in the demise of Angkor, Cambodia. *Proceedings of the National Academy of Sciences*, **107**(15), pp. 6748–6752. doi:10.1073/pnas.0910827107.

Cane, M. a. (2005). The evolution of El Nino, past and future. *Earth and Planetary Science Letters*, **230**(3-4), pp. 227–240.

Cavazos, T., C. Turrent, and D. P. Lettenmaier (2008). Extreme precipitation trends associated with tropical cyclones in the core of the North American monsoon. *Geophysical Research Letters*, **35**(21). Cavazos, Tereza Turrent, C. Lettenmaier, D. P.

Chen, P. and M. Newman (1998). Rossby wave propagation and the rapid development of upper-level anomalous anticyclones during the 1988 US drought. *Journal of Climate*, **11**(10), pp. 2491–2504.

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, **51**(4), pp. 661–703. ISSN 0036-1445. doi: 10.1137/070710111.

Cleaveland, M. K., D. W. Stahle, M. D. Therrell, J. Villanueva-Diaz, and B. T. Burns (2003). Tree-ring reconstructed winter precipitation and tropical teleconnections in Durango, Mexico. *Climatic Change*, **59**(3), pp. 369–388.

Clement, A., P. DiNizio, and C. Deser (2011). Rethinking the ocean's role in the Southern Oscillation. *Journal of Climate (in press)*.

Clement, A. C. and M. Cane (1999). A Role for the Tropical Pacific Coupled Ocean-Atmosphere System on Milankovitch and Millennial Timescales. Part I: A Modeling Study of Tropical Pacific Variability. *Geophysical Monograph*, **112**, pp. 363–371.

Cobb, K. M., C. D. Charles, H. Cheng, and R. L. Edwards (2003). El Nino/Southern Oscillation and tropical Pacific climate during the last millennium. *Nature*, **424**(6946), pp. 271–276.

Cole, J. E., R. G. Fairbanks, and G. T. Shen (1993). Recent Variability in the Southern Oscillation - Isotopic Results from a Tarawa Atoll Coral. *Science*, **260**(5115), pp. 1790–1793.

Collins, W. D., C. M. Bitz, M. L. Blackmon, G. B. Bonan, C. S. Bretherton, J. A. Carton, P. Chang, S. C. Doney, J. J. Hack, T. B. Henderson, J. T. Kiehl, W. G. Large, D. S. McKenna, B. D. Santer, and R. D. Smith (2006). The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, **19**(11), pp. 2122–2143.

Conroy, J. L., J. T. Overpeck, J. E. Cole, T. M. Shanahan, and M. Steinitz-Kannan (2008). Holocene changes in eastern tropical Pacific climate inferred from a Galapagos lake sediment record. *Quaternary Science Reviews*, **27**(11-12), pp. 1166–1180.

Cook, B. I., R. L. Miller, and R. Seager (2008). Dust and sea surface temperature forcing of the 1930s "Dust Bowl" drought. *Geophysical Research Letters*, **35**(8), pp. –.

Cook, E. R., K. J. Anchukaitis, B. M. Buckley, R. D. D'Arrigo, G. C. Jacoby, and W. E. Wright (2010). Asian Monsoon Failure and Megadrought During the Last Millennium. *Science*, **328**(5977), pp. 486–489.

Cook, E. R., K. R. Briffa, D. M. Meko, D. A. Graybill, and G. Funkhouser (1995). The Segment Length Curse in Long Tree-Ring Chronology Development for Paleoclimatic Studies. *Holocene*, **5**(2), pp. 229–237.

Cook, E. R., D. M. Meko, D. W. Stahle, and M. K. Cleaveland (1999). Drought reconstructions for the continental United States. *Journal of Climate*, **12**(4), pp. 1145–1162.

Cook, E. R., C. A. Woodhouse, C. M. Eakin, D. M. Meko, and D. W. Stahle (2004). Long-term aridity changes in the western United States. *Science*, **306**(5698), pp. 1015–1018.

Cook, K. H. and E. K. Vizy (2010). Hydrodynamics of the Caribbean Low-Level Jet and Its Relationship to Precipitation. *Journal of Climate*, **23**(6), pp. 1477–1494.

Dai, A. G., K. E. Trenberth, and T. T. Qian (2004). A global dataset of Palmer Drought Severity Index for 1870-2002: Relationship with soil moisture and effects of surface warming. *Journal of Hydrometeorology*, **5**(6), pp. 1117–1130.

Delworth, T. (2010). The Atlantic Meridional Overtuning Circulation and Climate - Variability, Predictability and Change. *Presentation at CLIVAR Workshop on Decadal Variability, Predictability, and Prediction*. Http://www.clivar.org/organization/wgomd/decadal/talks/Delworth.pdf.

Delworth, T. and S. Manabe (1988). The Influence of Potential Evaporation on the Variabilities of Simulated Soil Wetness and Climate. *Journal of Climate*.

Delworth, T. and S. Manabe (1989). The influence of soil wetness on near-surface atmospheric variability. *Journal of Climate*, **2**(12), pp. 1447–1462.

deMenocal, P. (2001). Cultural responses to climate change during the Late Holocene. *Science*, **292**(5517), pp. 667–673. ISSN 0036-8075.

Diaz, S. C., M. D. Therrell, D. W. Stahle, and M. K. Cleaveland (2002). Chihuahua (Mexico) winter-spring precipitation reconstructed from tree-rings, 1647-1992. *Climate Research*, **22**(3), pp. 237–244.

Findell, K. L. and T. L. Delworth (2010). Impact of Common Sea Surface Temperature Anomalies on Global Drought and Pluvial Frequency. *Journal of Climate*, **23**(3), pp. 485–503. ISSN 0894-8755. doi:10.1175/2009JCLI3153.1.

Fraedrich, K. and R. Blender (2003). Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Physical Review Letters*, **90**(10), pp. –.

Garfin, G. M. and M. K. Hughes (1996). Eastern Oregon Divisional Precipitation and Palmer Drought Severity Index from Tree-Rings. Technical report, Forest Service Intermountain Research Station.

Ghil, M., M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou (2002). Advanced spectral methods for climatic time series. *Reviews of Geophysics*, **40**(1), pp. –.

Graham, N. E., M. K. Hughes, C. M. Ammann, K. M. Cobb, M. P. Hoerling, D. J. Kennett, J. P. Kennett, B. Rein, L. Stott, P. E. Wigand, and T. Y. Xu (2007). Tropical Pacific - mid-latitude teleconnections in medieval times. *Climatic Change*, **83**(1-2), pp. 241–285.

Graumlich, L. J., M. F. J. Pisaric, L. A. Waggoner, J. S. Littell, and J. C. King (2003). Upper Yellowstone River flow and teleconnections with Pacific basin climate variability during the past three centuries. *Climatic Change*, **59**(1-2), pp. 245–262.

Gray, S. T., C. L. Fastie, S. T. Jackson, and J. L. Betancourt (2004a). Tree-ring-based reconstruction of precipitation in the Bighorn Basin, Wyoming, since 1260 AD. *Journal of Climate*, **17**(19), pp. 3855–3865.

Gray, S. T., S. T. Jackson, and J. L. Betancourt (2004b). Tree-ring based reconstructions of interannual to decadal scale precipitation variability for northeastern Utah since 1226 AD. *Journal of the American Water Resources Association*, **40**(4), pp. 947–960.

Griffin, R. (2005). Paleohydrology of the Salinas River, California, reconstructed from blue oak tree rings.

Grissino-Mayer, H. D. (1995). *The climate and fire history of El Malpais National Monument, New Mexico*. Ph.D. thesis, The University of Arizona, Tucson, AZ.

Guilyardi, E., A. Wittenberg, A. Fedorov, M. Collins, C. Wang, A. Capotondi, G. J. van Oldenborgh, and T. Stockdale (2009). Understanding El Nino in Ocean-Atmosphere General Circulation Models: Progress and Challenges. *Bulletin of the American Meteorological Society*, **90**(3), pp. 325+. ISSN 0003-0007. doi: 10.1175/2008BAMS2387.1.

Guttman, N. B. (1998). Comparing the Palmer Drought Index and the standardized precipitation index. *Journal of the American Water Resources Association*, **34**(1), pp. 113–121.

Guttman, N. B. (1999). Accepting the standardized precipitation index: A calculation algorithm. *Journal of the American Water Resources Association*, **35**(2), pp. 311–322.

Hasselmann, K. (1976). Stochastic Climate Models.1. Theory. *Tellus*, **28**(6), pp. 473–485.

Haug, G. H., D. Gunther, L. C. Peterson, D. M. Sigman, K. A. Hughen, and B. Aeschlimann (2003). Climate and the Collapse of Maya Civilization. *Science*, **299**(5613), pp. 1731–1735. doi:10.1126/science.1080444.

Hendy, C. H. (1971). Isotopic Geochemistry of Speleothems.1. Calculation of Effects of Different Modes of Formation on Isotopic Composition of Speleothems and Their Applicability as Palaeoclimatic Indicators. *Geochimica Et Cosmochimica Acta*, **35**(8), pp. 801–.

Heneghan, C. and G. McDarby (2000). Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. *Physical Review E*, **62**(5, Part A), pp. 6103–6110. ISSN 1063-651X.

Herweijer, C. and R. Seager (2008). The global footprint of persistent extra-tropical drought in the instrumental era. *International Journal of Climatology*, **28**(13), pp. 1761–1774.

Herweijer, C., R. Seager, E. R. Cook, and J. Emile-Geay (2007). North American droughts of the last millennium from a gridded network of tree-ring data. *Journal of Climate*, **20**(7), pp. 1353–1376. Herweijer, Celine Seager, Richard Cook, Edward R. Emile-Geay, Julien.

Hoerling, M., X. W. Quan, and J. Eischeid (2009). Distinct causes for two principal US droughts of the 20th century. *Geophysical Research Letters*, **36**. Hoerling, Martin Quan, Xiao-Wei Eischeid, Jon.

Holmgren, C. A., M. C. Penalba, K. A. Rylander, and J. L. Betancourt (2003). A 16,000 C-14 yr BP packrat midden series from the USA-Mexico Borderlands. *Quaternary Research*, **60**(3), pp. 319–329.

Huang, J., H. vandenDool, and K. Georgakakos (1996). Analysis of model-calculated soil moisture over the United States (1931-1993) and applications to long-range temperature forecasts. *Journal of Climate*, **9**(6), pp. 1350–1362. ISSN 0894-8755.

Hughes, M. K. and L. J. Gramlich (1996). Climatic variations and forcing mechanisms of the last 2000 years. *Multi-millenial dendroclimatic studies from the western United States, NATO ASI Series*, **141**, pp. 109–124.

Hunt, B. G. (2010). Global characteristics of pluvial and dry multi-year episodes, with emphasis on megadroughts. *International Journal of Climatology*, pp. n/a–n/a. ISSN 1097-0088. doi:10.1002/joc.2166.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, pp. 770–799. ISSN 0066-0604.

Huybers, P. and W. Curry (2006). Links between annual, Milankovitch and continuum temperature variability. *Nature*, **441**(7091), pp. 329–332.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**(3), pp. 437–471.

Kantelhardt, J., E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde, and S. Havlin (2006). Long-term persistence and multifractality of precipitation and river runoff records. *Journal of Geophysical Research-Atmospheres*, **111**(D1). ISSN 0148-0227. doi:10.1029/2005JD005881.

Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan (1998). Analyses of global sea surface temperature, 1856-1991. *Journal of Geophysical Research*, **103**, pp. 18,567–18,589.

Karl, T. R. and A. J. Koscielny (1982). Drought in the United States - 1895-1981. *Journal of Climatology*, **2**(4), pp. 313–329.

Kaufman, D. S., D. P. Schneider, N. P. McKay, C. M. Ammann, R. S. Bradley, K. R. Briffa, G. H. Miller, B. L. Otto-Bliesner, J. T. Overpeck, B. M. Vinther, and Arctic Lakes 2k Project Members (2009). Recent Warming Reverses Long-Term Arctic Cooling. *SCIENCE*, **325**(5945), pp. 1236–1239. ISSN 0036-8075. doi:10.1126/science.1173983.

Kim, S. T. and J. R. O'Neill (1997). Equilibrium and nonequilibrium oxygen isotope effects in synthetic carbonates. *Geochimica Et Cosmochimica Acta*, **61**(16), pp. 3461–3475.

Kipfmueller, K. F. and M. W. Salzer (2010). Linear trend and climate response of five-needle pines in the western United States related to treeline proximity. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **40**(1), pp. 134–142.

Klemes, V. (1974). Hurst Phenomenon - Puzzle. *Water Resources Research*, **10**(4), pp. 675–688.

Koscielny-Bunde, E., J. W. Kantelhardt, P. Braun, A. Bunde, and S. Havlin (2006). Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies. *Journal of Hydrology*, **322**(1-4), pp. 120–137. ISSN 0022-1694. doi:10.1016/j.jhydro1.2005.03.004.

Kushnir, Y., R. Seager, M. F. Ting, N. Naik, and J. Nakamura (2010). Mechanisms of Tropical Atlantic SST Influence on North American Precipitation Variability. *Journal of Climate*, **23**(21), pp. 5610–5628.

Laird, K. R., S. C. Fritz, and B. F. Cumming (1998). A diatom-based reconstruction of draught intensity, duration, and frequency from Moon Lake, North Dakota: a sub-decadal record of the last 2300 years. *Journal of Paleolimnology*, **19**(2), pp. 161–179.

Laird, K. R., S. C. Fritz, E. C. Grimm, and P. G. Mueller (1996). Century-scale paleoclimatic reconstruction from Moon Lake, a closed-basin lake in the northern Great Plains. *Limnology and Oceanography*, **41**(5), pp. 890–902.

Lean, J. (2000). Evolution of the sun's spectral irradiance since the Maunder Minimum. *Geophysical Research Letters*, **27**(16), pp. 2425–2428.

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 General-Circulation Models. *Journal of Geophysical Research-Atmospheres*, **99**(D7), pp. 14415–14428.

Liang, X. Z., J. H. Zhu, K. E. Kunkel, M. F. Ting, and J. X. L. Wang (2008). Do CGCMs simulate the North American monsoon precipitation seasonal-interannual variability? *Journal of Climate*, **21**(17), pp. 4424–4448. Liang, Xin-Zhong Zhu, Jinhong Kunkel, Kenneth E. Ting, Mingfang Wang, Julian X. L.

Lomb, N. R. (1975). Spectrographic Study of Beta-Centauri. *Monthly Notices of the Royal Astronomical Society*, **172**(3), pp. 639–647.

Lyon, B. and R. M. Dole (1995). A diagnostic comparison of the 1980 and 1988 US summer heat wave-droughts. *Journal of Climate*, **8**(6), pp. 1658–1675.

MacDonald, G. M. (2010). Water, climate change, and sustainability in the southwest. *Proceedings of the National Academy of Sciences*, **107**(50), pp. 21256–21262. doi:10.1073/pnas.0909651107.

Mandelbrot, B. and R. L. Hudson (2004). *The (mis)behavior of markets: a fractal view of risk, ruin, and reward*. Basic Books, New York.

Mandelbrot, B. and J. R. Wallis (1968). Noah Joseph and Operational Hydrology. *Water Resources Research*, **4**(5), pp. 909–918.

Mann, M. E. and J. Park (1999). Oscillatory spatiotemporal signal detection in climate studies: A multiple-taper spectral domain approach. *Advances in Geophysics, Vol 41*, **41**, pp. 1–131.

Mantua, N., S. Hare, Y. Zhang, J. Wallace, and R. Francis (1997). A Pacific interdecadal oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, **78**, pp. 1069–10079.

McAfee, S. A. and J. L. Russell (2008). Northern Annular Mode impact on spring climate in the western United States. *Geophysical Research Letters*, **35**(17). ISSN 0094-8276. doi:10.1029/2008GL034828.

McAfee, S. A., J. L. Russell, and P. J. Goodman (2011). Evaulating IPCC AR4 cool-season precipitation simulations and projections for impacts assessment over North America. *Climate Dynamics*.

McCabe, G. J., M. A. Palecki, and J. L. Betancourt (2004). Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(12), pp. 4136–4141.

McKee, T. B., N. J. Doeskin, and J. Kleist (1993). The Relationship of Drought Frequency and Duration to Time Scales. *Proc. 8th Conf. on Aplied Climatology*.

Meehl, G. A., L. Goddard, J. Murphy, R. J. Stouffer, G. Boer, G. Danabasoglu, K. Dixon, M. A. Giorgetta, A. M. Greene, E. Hawkins, G. Hegerl, D. Karoly, N. Keenlyside, M. Kimoto, B. Kirtman, A. Navarra, R. Pulwarty, D. Smith, D. Stammer, and T. Stockdale (2009). Decadal Prediction: Can It Be Skillful? *Bulletin of the American Meteorological Society*, **90**(10), pp. 1467+. ISSN 0003-0007. doi:10.1175/2009BAMS2778.1.

Meehl, G. A. and A. X. Hu (2006). Megadroughts in the Indian monsoon region and southwest North America and a mechanism for associated multidecadal Pacific sea surface temperature anomalies. *Journal of Climate*, **19**(9), pp. 1605–1623.

Meko, D. M. (2001). Reconstructed Sacramento River System Runoff From Tree Rings. Technical report, California Department of Water Resources.

Meko, D. M., M. D. Therrell, C. H. Baisan, and M. K. Hughes (2001). Sacramento River flow reconstructed to AD 869 from tree rings. *Journal of the American Water Resources Association*, **37**(4), pp. 1029–1039.

Meko, D. M., C. A. Woodhouse, C. A. Baisan, T. Knight, J. J. Lukas, M. K. Hughes, and M. W. Salzer (2007). Medieval drought in the upper Colorado River Basin. *Geophysical Research Letters*, **34**(10), p. L10705.

Meko, D. M., C. A. Woodhouse, and K. Morino (2010). Dendrochronology and links to streamflow. *Journal of Hydrology*. In press.

Mickler, P. J., L. A. Stern, and J. L. Banner (2006). Large kinetic isotope effects in modern speleothems. *Geological Society of America Bulletin*, **118**(1-2), pp. 65–81.

Mitchell, T. D. and P. D. Jones (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, **25**(6), pp. 693–712.

Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Karlen (2005). Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature*, **433**(7026), pp. 613–617.

Neelin, J. D., D. S. Battisti, A. C. Hirst, F. F. Jin, Y. Wakata, T. Yamagata, and S. E. Zebiak (1998). ENSO theory. *Journal of Geophysical Research-Oceans*, **103**(C7), pp. 14261–14290.

Newman, M., G. P. Compo, and M. A. Alexander (2003). ENSO-forced variability of the Pacific decadal oscillation. *Journal of Climate*, **16**(23), pp. 3853–3857.

Ni, F., T. Cavazos, M. Hughes, A. Comrie, and G. Funkhouser (2002). Cool-season precipitation in the southwestern USA since AD 1000: Comparison of linear and nonlinear techniques for reconstruction. *International Journal of Climatology*, **22**(13), pp. 1645–1662. ISSN 0899-8418. doi:10.1002/joc.804.

Palmer, W. C. (1965). Meteorological drought.

Pelletier, J. (1998). The power spectral density of atmospheric temperature from time scales of 10(-2) to 10(6) yr. *Earth and Planetary Science Letters*, **158**(3-4), pp. 157–164. ISSN 0012-821X.

Pelletier, J. and D. Turcotte (1997). Long-range persistence in climatological and hydrological time series: analysis, modeling and application to drought hazard assessment. *Journal of Hydrology*, **203**(1-4), pp. 198–208. ISSN 0022-1694.

Pelletier, J. D. (2008). *Quantitative Modeling of Earth System Processes*. Cambridge University Press, Cambridge, UK, 1 edition.

Peng, C., S. Buldyrev, S. Havlin, M. Simons, H. Stanley, and A. Goldberger (1994). Mosaic Organization Of DNA Nucleotides. *Physical Review E*, **49**(2), pp. 1685–1689. ISSN 1063-651X.

Pohl, K., M. Therrell, J. B. ad N. Ayotte, J. Hernandez, S. Castro, E. Oviedo, J. Elvir, M. E. D. Opland, J. Park, G. Pederson, S. Salazar, L. Selem, J. Diaz, and D. Stahle (2003). A cool season precipitation reconstruction for Saltillo, Mexico: 11th North American Dendroecological Fieldweek, Climatic Reconstruction Research Group, Saltillo, Mexico, August 2001. *Tree-Ring Research*, **59**(1), pp. 11–19.

Polyak, V. J. and Y. Asmerom (2001). Late Holocene climate and cultural changes in the southwestern United States. *Science*, **294**(5540), pp. 148–151.

Quadrelli, R. and J. M. Wallace (2004). A simplified linear framework for interpreting patterns of Northern Hemisphere wintertime climate variability. *Journal of Climate*, **17**(19), pp. 3728–3744.

Rajagopalan, B. and U. Lall (1998). Interannual variability in western US precipitation. *Journal of Hydrology*, **210**(1-4), pp. 51–67. ISSN 0022-1694.

Ramsey, C. B. (2008). Deposition models for chronological records. *Quaternary Science Reviews*, **27**(1-2), pp. 42–60.

Rasmussen, J., V. Polyak, and Y. Asmerom (2006). Evidence for Pacific-modulated precipitation variability during the late Holocene from the southwestern USA. *Geophysical Research Letters*, **33**(8). ISSN 0094-8276. doi:10.1029/2006GL025714.

Rasmusson, E. M. and P. A. Arkin (1993). A global view of large-scale precipitation variability. *Journal of Climate*, **6**(8), pp. 1495–1522.

Redmond, K. T. (2002). The depiction of drought - A commentary. *Bulletin of the American Meteorological Society*, **83**(8), pp. 1143–1147.

Richman, M. (1986). Rotation of Principal Components. *Journal of Climatology*.

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

Salzer, M. W., M. K. Hughes, A. G. Bunn, and K. F. Kipfmueller (2009). Recent unprecedented tree-ring growth in bristlecone pine at the highest elevations and possible causes. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(48), pp. 20348–20353. Salzer, Matthew W. Hughes, Malcolm K. Bunn, Andrew G. Kipfmueller, Kurt F.

Salzer, M. W. and K. F. Kipfmueller (2005). Reconstructed temperature and precipitation on a millennial timescale from tree-rings in the Southern Colorado Plateau, USA. *Climatic Change*, **70**(3), pp. 465–487.

Scargle, J. D. (1982). Studies in Astronomical Time-Series Analysis.2. Statistical Aspects of Spectral-Analysis of Unevenly Spaced Data. *Astrophysical Journal*, **263**(2), pp. 835–853.

Schubert, S., D. Gutzler, H. Wang, A. Dai, T. Delworth, C. Deser, K. Findell, R. Fu, W. Higgins, M. Hoerling, B. Kirtman, R. Koster, A. Kumar, D. Legler, D. Lettenmaier, B. Lyon, V. Magana, K. Mo, S. Nigam, P. Pegion, A. Phillips, R. Pulwarty, D. Rind, A. Ruiz-Barradas, J. Schemm, R. Seager, R. Stewart, M. Suarez, J. Syktus, M. Ting, C. Wang, S. Weaver, and N. Zeng (2009). A US CLIVAR Project to Assess and Compare the Responses of Global Climate Models to Drought-Related SST Forcing Patterns: Overview and Results. *Journal of Climate*, **22**(19), pp. 5251–5272. ISSN 0894-8755. doi:10.1175/2009JCLI3060.1.

Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister (2004a). Causes of long-term drought in the US Great Plains. *Journal of Climate*, **17**(3), pp. 485–503.

Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister (2004b). On the cause of the 1930s Dust Bowl. *Science*, **303**(5665), pp. 1855–1859.

Schulz, M. and K. Stattegger (1997). SPECTRUM: Spectral analysis of unevenly spaced paleoclimatic time series. *Computers & Geosciences*, **23**(9), pp. 929–945. ISSN 0098-3004.

Seager, R., N. Graham, C. Herweijer, A. L. Gordon, Y. Kushnir, and E. Cook (2007a). Blueprints for Medieval hydroclimate. *Quaternary Science Reviews*, **26**(19-21), pp. 2322–2336.

Seager, R., N. Harnik, W. A. Robinson, Y. Kushnir, M. Ting, H. P. Huang, and J. Velez (2005a). Mechanisms of ENSO-forcing of hemispherically symmetric precipitation variability. *Quarterly Journal of the Royal Meteorological Society*, **131**(608), pp. 1501–1527. Part B.

Seager, R., Y. Kushnir, C. Herweijer, N. Naik, and J. Velez (2005b). Modeling of tropical forcing of persistent droughts and pluvials over western North America: 1856-2000. *Journal of Climate*, **18**(19), pp. 4065–4088.

Seager, R., M. F. Ting, I. Held, Y. Kushnir, J. Lu, G. Vecchi, H. P. Huang, N. Harnik, A. Leetmaa, N. C. Lau, C. H. Li, J. Velez, and N. Naik (2007b). Model projections of an imminent transition to a more arid climate in southwestern North America. *Science*, **316**(5828), pp. 1181–1184.

Seidel, D. J., Q. Fu, W. J. Randel, and T. J. Reichler (2008). Widening of the tropical belt in a changing climate. *Nature Geoscience*, **1**(1), pp. 21–24. Seidel, Dian J. Fu, Qiang Randel, William J. Reichler, Thomas J.

Shanahan, T. M., J. T. Overpeck, K. J. Anchukaitis, J. W. Beck, J. E. Cole, D. L. Dettman, J. A. Peck, C. A. Scholz, and J. W. King (2009). Atlantic Forcing of Persistent Drought in West Africa. *Science*, **324**(5925), pp. 377–380. ISSN 0036-8075. doi:10.1126/science.1166352.

Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller, and Z. Chen (2007). *Climate Change 2007: The Physical Science Basis. Working Group I Contribution to the Fourth Assessment Report of the IPCC (Climate Change 2007)*. Cambridge University Press, United Kingdom and New York, NY, USA, 996 pp. ISBN 0521705967.

St. George, S. and T. R. Ault (2011). Is energetic decadal variability a stable feature of the central Pacific Coast's winter climate? *J. Geophys. Res.*

Stahle, D. W., F. K. Fye, E. R. Cook, and R. D. Griffin (2007). Tree-ring reconstructed megadroughts over North America since AD 1300. *Climatic Change*, **83**(1-2), pp. 133–149.

Stephenson, J. B. (2007). *Climate Change: financial risks to federal and private insurers in coming decades are potentially significant*. United States. Congress. Senate. Committee on Homeland Security and Governmental Affairs, U.S. Govt. Accountability Office.

Stine, S. (1990). Past Climate at Mono Lake. *Nature*, **345**(6274), pp. 391–391.

Tan, M., A. Baker, D. Genty, C. Smith, J. Esper, and B. G. Cai (2006). Applications of stalagmite laminae to paleoclimate reconstructions: Comparison with dendrochronology/climatology. *Quaternary Science Reviews*, **25**(17-18), pp. 2103–2117.

Thomson, D. J. (1982). Spectrum Estimation and Harmonic-Analysis. *Proceedings of the IEEE*, **70**(9), pp. 1055–1096.

Trenberth, K. E. and G. W. Branstator (1992). Issues in establishing causes of the 1988 drought over North-America. *Journal of Climate*, **5**(2), pp. 159–172.

Trenberth, K. E., G. W. Branstator, and P. A. Arkin (1988). Origins of the 1988 North-American drought. *Science*, **242**(4886), pp. 1640–1645.

Trenberth, K. E. and C. J. Guillemot (1996). Physical processes involved in the 1988 drought and 1993 floods in North America. *Journal of Climate*, **9**(6), pp. 1288–1298.

Truebe, S., T. Ault, and J. Cole (2010). A forward model of cave dripwater delta-18-O and application to speleothem records. *IOP Conference Series: Earth and Environmental Science*, **9**(012022). doi:doi:10.1088/1755-1315/9/1/012022.

Urban, F. E., J. E. Cole, and J. T. Overpeck (2000). Influence of mean climate change on climate variability from a 155-year tropical Pacific coral record. *Nature*, **407**(6807), pp. 989–993.

Vyushin, D. I. and P. J. Kushner (2009). Power-Law and Long-Memory Characteristics of the Atmospheric General Circulation. *Journal of Climate*, **22**(11), pp. 2890–2904.

Vyushin, D. I., P. J. Kushner, and J. Mayer (2009). On the origins of temporal power-law behavior in the global atmospheric circulation. *Geophysical Research Letters*, **36**.

Wagner, J. D. (2006). *Speleothem record of southern Arizona paleoclimate, 54 to 3.5ka*. Ph.D. thesis, University of Arizona.

Wagner, J. D. M., J. E. Cole, J. W. Beck, P. J. Patchett, G. M. Henderson, and H. R. Barnett (2010). Moisture variability in the southwestern United States linked to abrupt glacial climate change. *Nature Geoscience*, **3**(2), pp. 110–113.

Wallace, J. M. and D. S. Gutzler (1981). Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. *Monthly Weather Review*, **109**(4), pp. 784–812.

Wells, N., S. Goddard, and M. J. Hayes (2004). A self-calibrating Palmer Drought Severity Index. *Journal of Climate*, **17**(12), pp. 2335–2351.

Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, **36**, p. 5.

Woodhouse, C. (2000). Extending Hydrologic Records with Tree Rings. *Water Resources Impact*, **2**(4), pp. 25–27.

Woodhouse, C. A. (2001). A tree-ring reconstruction of streamflow for the Colorado Front Range. *Journal of the American Water Resources Association*, **37**(3), pp. 561–569.

Woodhouse, C. A., S. T. Gray, and D. M. Meko (2006). Updated streamflow reconstructions for the Upper Colorado River Basin. *Water Resources Research*, **42**(5).

Woodhouse, C. A. and J. T. Overpeck (1998). 2000 years of drought variability in the central United States. *Bulletin of the American Meteorological Society*, **79**(12), pp. 2693–2714.

Woodhouse, C. A., J. L. Russell, and E. R. Cook (2009). Two Modes of North American Drought from Instrumental and Paleoclimatic Data. *Journal of Climate*, **22**(16), pp. 4336–4347. Woodhouse, C. A. Russell, J. L. Cook, E. R.

Wright, W. E., A. Long, A. C. Comrie, S. W. Leavitt, T. Cavazos, and C. Eastoe (2001). Monsoonal moisture sources revealed using temperature, precipitation, and precipitation stable isotope timeseries. *Geophysical Research Letters*, **28**(5), pp. 787–790.

Wunsch, C. (2003). The spectral description of climate change including the 100 ky energy. *Climate Dynamics*, **20**(4), pp. 353–363. ISSN 0930-7575. doi:10.1007/s00382-002-0279-z.

Zebiak, S. and M. Cane (1987). A model El Nino/Southern Oscillation. *Monthly Weather Review*, **115**, pp. 2262–2278.