Drought Reconstructions for the Continental United States*

EDWARD R. COOK

Tree-Ring Laboratory, Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

DAVID M. MEKO

Laboratory of Tree-Ring Research, The University of Arizona, Tucson, Arizona

DAVID W. STAHLE AND MALCOLM K. CLEAVELAND

Tree-Ring Laboratory, Department of Geography, University of Arkansas, Fayetteville, Arkansas

(Manuscript received 8 December 1997, in final form 25 June 1998)

ABSTRACT

The development of a 2° lat × 3° long grid of summer drought reconstructions for the continental United States estimated from a dense network of annual tree-ring chronologies is described. The drought metric used is the Palmer Drought Severity Index (PDSI). The number of grid points is 154 and the reconstructions cover the common period 1700–1978. In producing this grid, an automated gridpoint regression method called ''point-by-point regression'' was developed and tested. In so doing, a near-optimal global solution was found for its implementation. The reconstructions have been thoroughly tested for validity using PDSI data not used in regression modeling. In general, most of the gridpoint estimates of drought pass the verification tests used. In addition, the spatial features of drought in the United States have been faithfully recorded in the reconstructions even though the method of reconstruction is not explicitly spatial in its design.

The drought reconstructions show that the 1930s ''Dust Bowl'' drought was the most severe such event to strike the United States since 1700. Other more local droughts are also revealed in the regional patterns of drought obtained by rotated principal component analysis. These reconstructions are located on a NOAA Web site at the World Data Center-A in Boulder, Colorado, and can be freely downloaded from there.

1. Introduction

Drought occurrence remains a serious concern in the United States (U.S.). In 1996, the most severe drought of the past 20 years struck the Southwest. Such events place huge demands on rural and urban water resources and quality, and place huge burdens on agricultural and energy production. The 1996 drought was preceded by a severe drought in the late 1980s in California (Roos 1994; Haston and Michaelsen 1997), in 1986 in the Southeast (Bergman et al. 1986; Cook et al. 1988), in 1976–77 in the West (Namias 1978; Matthai 1979), in the 1960s in the Northeast (Namias 1966; Cook and Jacoby 1977), in the 1950s in Texas (Namias 1955; Stahle and Cleaveland 1988), and in the 1930s in the northern Great Plains (Warrick 1980; Stockton and Meko 1983). Clearly, drought is a common occurrence in the U.S. and can occur anywhere. Understanding the causes of drought, especially the severe multiyear events, is necessary if reliable methods of forecasting are to be developed. A major difficulty in using the available meteorological records to model drought in the U.S. is the limited time span covered by such records. There are often too few realizations of proposed forcing mechanisms of drought in the short instrumental records to test any of them in a statistically rigorous way. We hope to alleviate this problem through the use of centuries-long, annual tree-ring chronologies.

In this paper we describe the development of a gridded network of drought reconstructions covering the continental U.S. that is derived from a large collection of climatically sensitive tree-ring chronologies. The reconstructions cover the period 1700–1978 and are based on the Palmer Drought Severity Index (PDSI; Palmer 1965), a well-known and widely used measure of drought and wetness. Previous efforts in reconstructing drought from tree rings in the U.S. have been highly successful (e.g., Blasing and Duvick 1984; Cook and

* Lamont-Doherty Earth Observatory Contribution Number 5896.

Corresponding author address: Dr. Edward R. Cook, Lamont-Doherty Earth Observatory, Palisades, NY 10964.
E-mail: drdendro@ldeo.columbia.edu

Jacoby 1977; Cook et al. 1988; Cook et al. 1992; Graumlich 1993; Haston and Michaelsen 1994, 1997; Meko 1992; Mitchell et al. 1979; Stahle and Cleaveland 1988; Stahle et al. 1985, 1988; Stockton and Meko 1975, 1983; Woodhouse and Meko 1997). However, most of these efforts have only involved reconstructing local or regional drought histories. Earlier efforts by Fritts (1976, 1991) to reconstruct seasonal temperature and precipitation across the U.S. met with limited success because the tree-ring data used were restricted to western North America. Since that pioneering work, the coverage of tree-ring chronologies across the U.S. has increased enormously. Consequently, an effort to expand these drought reconstructions in a homogeneous way to the entire continental U.S. was initiated by Meko et al. (1993) and developed further by Cook et al. (1996).

We will describe in considerable detail the method used to develop the continent-wide grid of drought reconstructions and provide a number of results that document the generally high fidelity of the tree-ring estimates. This has been done in an effort to make the reconstructions as useful as possible to climatologists and modelers who might want to use these data for studying past temporal and spatial patterns of drought and their association with hypothesized forcing functions. To this end, the drought reconstructions have already been used with considerable success to reevaluate the putative connection between a bidecadal drought area rhythm in the western U.S. and solar/lunar tidal forcing (Mitchell et al. 1979; Cook et al. 1997). In addition, they are presently being used to study and better understand the teleconnection between drought/wetness and the El Niño–Southern Oscillation in the U.S. (Ropelewski and Halpert 1986; Cook et al. 1999; Cole and Cook 1998).

Much of what we have produced is, in a sense, an extension of the work done by Karl and Koscielny (1982; referred to hereafter as KK) in their study of drought in the U.S. The research of KK was based on a 60-point grid of instrumental PDSI data covering the period 1895–1981. In so doing, they described some important temporal and spatial features of drought in the U.S. As will be shown, our tree-ring reconstructions faithfully capture most of the properties of drought described by KK. In the process, the drought database for the U.S. has been extended back in time by a factor of 3 on a much denser 154-point grid, which allows for more detailed temporal and spatial analyses.

2. The PDSI grid

The PDSI grid used in this study is 2° lat × 3° long and is patterned after a study of runoff and drought across the United States by Langbein and Slack (1982). This grid is shown in Fig. 1. It is based on 1036 single-station monthly PDSI records estimated from the Historical Climatology Network (Karl et al. 1990) and modified according to Guttman (1991). However, the Death Valley, California, record was deleted prior to gridding because of anomalously high (e.g., >20) monthly PDSI values in 1958. All stations used begin no later than 1925 and end no earlier than 1989.

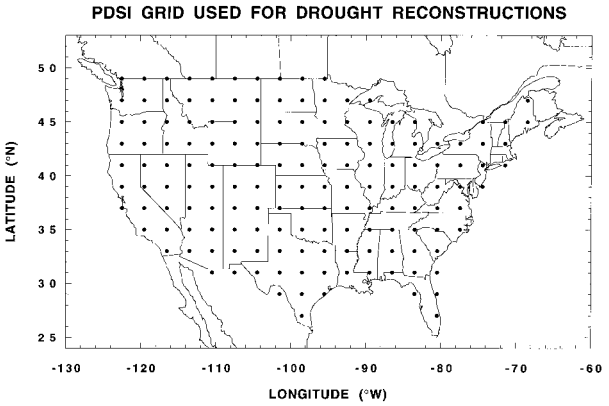

FIG. 1. Map of the continental U.S. showing the locations of the PDSI grid points used in this study. The grid spacing is 2° lat. × 3° long. and totals 155 points.

The choice of the gridding dimension was a trade-off between spatial resolution and the desire to reduce the size of the PDSI network. A 2° × 3° grid reduces the PDSI network to about 15% of its original size. Yet, the spatial definition of the grid should still be high enough to capture mesoscale patterns of wetness and dryness and to resolve regional drought patterns found by KK using a coarser 60-point PDSI grid.

Following Meko et al. (1993) and Cook et al. (1996), the single-station records were interpolated to the 155 grid points using inverse-distance weighting of the form

$$\mathrm{PDSI}_k = \left(\sum_{j=1}^{m} \frac{\mathrm{PDSI}_j}{d_j}\right) \Bigg/ \left(\sum_{j=1}^{m} \frac{1}{d_j}\right), \qquad (1)$$

where m is the number of stations within a given search radius of grid point k, PDSI_j is the jth PDSI station record, and d_j is the distance of station j from grid point k. The first 3 yr of data were deleted from each station record to eliminate starting-value transience in computing the monthly PDSIs (N. Guttman 1994, personal communication). Then, a 150-km search radius was used to locate stations local to each grid point. All stations found within that radius were used. If at least five stations were not found within 150 km, then the five closest stations were used. A minimum distance of 30 km was used in $1/d$ to avoid excessively weighting stations very close to the grid points. As stations dropped out prior to 1928, the weights of the remaining series were renormalized for interpolation to gridpoint k. This enabled the preservation of pre-1928 PDSI data for regression model validation.

Because the individual stations used in the grid have varying starting years, the first year of data available at each grid point varies over space. All of the grid points have data back to 1913. Prior to 1913, the number of

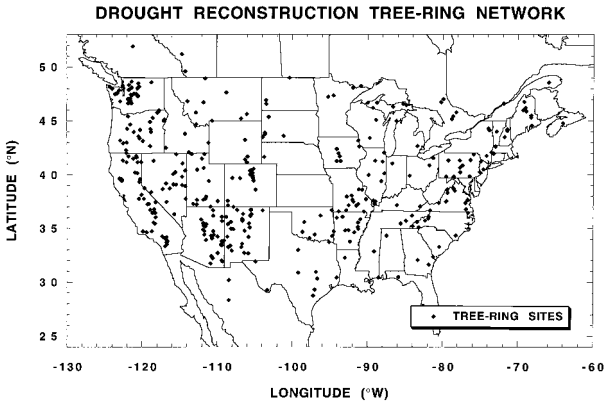

FIG. 2. Map of locations of the 425 annual tree-ring chronologies used in this study. Some of the locations represent more than one chronology. All of the series cover the common time period 1700–1979.

grid points with data declines to 143 by 1903, 97 by 1895, 38 by 1890, and 10 by 1874 due to the changing beginning years of the instrumental PDSI data over the grid. The quality of the grid also varies over space, with the weakest areas located in the northern Great Basin and Rocky Mountains regions (Cook et al. 1996).

The PDSI data were gridded on a monthly basis. Yet, past experience indicates that summer (i.e., June–August) PDSI relates better to tree rings on average than to any single month when comparisons are made over large geographic areas. Cook et al. (1992) found that the peak correlation between PDSIs and tree rings shifted from June for chronologies south of Virginia, to July in the Virginia–New York region, and even to August for some chronologies in northern New England and Canada. This shift was attributed to regional differences in phenology (i.e., the timing of tree growth) associated with the northward march of the growing season in spring and to the time required for evapotranspiration demand to significantly draw down the soil moisture supply. Consequently, we also used summer PDSI here as the best drought index to reconstruct across the United States.

3. The tree-ring network

The tree-ring chronologies used here to reconstruct past drought across the continental U.S. number 425 now, an increase of 177 over those used by Meko et al. (1993) in their examinations of spatial patterns of tree growth. Many of the new chronologies were developed after the start of that study. Others were not included for a variety of reasons. We have chosen to include virtually all available chronologies that begin no later than 1700 and end no earlier than 1979, a criterion previously established by Meko et al. (1993). Figure 2 shows the distribution of sites across the United States. The distribution is highly patchy due, in part, to the uneven distribution of forested land. In addition, the distribution of tree species in the network is highly regional due to the natural distribution of forest communities and species range limits in the U.S. See Cook et al. (1996) for more details. Not all of these chronologies will be useful predictors of drought. However, the screening process described and tested below for retaining candidate tree-ring predictors of PDSI will guard against including those chronologies not statistically related to drought.

Prior to regression analysis, the tree-ring chronologies were put through a process of variance stabilization described in Meko et al. (1993). Variance stabilization was done to reduce the effects of changing sample size on the variance of the tree-ring chronology, especially in the early portions of chronologies below, say, six individual measurement series. This removed a potential artifact from the data that is unlikely to be related to changing climatic variability.

4. Calibration/verification results

The method used to reconstruct the PDSI grid from tree rings is point-by-point regression (PPR). As implemented here, PPR is simply the sequential, automated fitting of single-point principal components regression models of tree rings to a grid of climate variables. The sequential nature of PPR differentiates it from joint space–time methods used to simultaneously relate two fields of variables, such as canonical regression (Glahn 1968; Fritts et al. 1971), orthogonal spatial regression (Briffa et al. 1986; Cook et al. 1994), and singular value decomposition (Bretherton et al. 1992). Mathematical details of the PPR method and tests of its implementation are described in the appendixes at the end of the paper.

The PPR method is based on the premise that only those tree-ring chronologies proximal to a given PDSI grid point are likely to be true predictors of drought at that location, where ''true'' implies a causal relationship between tree rings and drought that is stable through time. The rationale behind this premise is our understanding of drought in the U.S. as a regional- or mesoscale phenomenon (see KK for mapped regions). Consequently, synoptic-scale teleconnections between tree rings and drought, while statistically significant during any given calibration period, may not be stable through time. This ''local control'' over the reconstruction of each gridpoint PDSI is not possible with the joint space–time reconstruction methods mentioned above.

In the PPR method, a fixed search radius around each grid point defines the zone of local control exercised by the method in selecting candidate tree-ring predictors of PDSI. A second level of control is also applied in the form of a screening probability, which eliminates tree-ring chronologies from the candidate pool that are poorly correlated with drought. Here we examine the statistical fidelity of the PDSI reconstructions based a 450-km search radius and a screening probability $\alpha =$

0.10, selected after the extensive testing described in appendix B. The statistics used for this purpose are the same as those described for those tests: the explained variance (R_c^2) over the 1928–78 calibration period, the squared Pearson correlation (R_v^2) over the pre-1928 verification period, the reduction of error (RE) in the verification period, and the coefficient of efficiency (CE) over the verification period. See appendix B for the mathematical definitions of these statistics. All four tests are measures of fractional variance in common between actual and reconstructed PDSI. In this sense, they are comparable. However, they differ markedly in their relative performances. As will be seen, the calibration period R_c^2 always overestimates the true fidelity of the tree-ring estimates of drought, while in the verification period, R_v^2 usually indicates better fidelity in the estimates than RE and CE. Among the verification statistics used here, the CE is the most rigorous (Cook et al. 1994).

A PDSI grid point passed the R_c^2 test if it was successfully calibrated by tree rings following the procedures outlined in appendix A; it passed the R_v^2 test if the Pearson correlation between actual and estimated PDSI in the verification period was significant at the 95% one-tailed confidence level; and it passed the RE or CE if either was >0. See appendixes A and B for details. For the R_c^2 test, 154 of 155 grid points were successfully calibrated. The only grid point not calibrated was the one in southern, subtropical Florida, where there are no tree-ring chronologies. For the R_v^2 test, 147 grid points of the remaining 154 were successfully verified. In contrast, the number of RE and CE tests that passed dropped from 147 to 133 and 124 grid points, respectively. So, the verification test results indicate that the tree-ring reconstructions of drought are significantly related to actual PDSI over most of the grid. The median R_c^2, R_v^2, RE, and CE fractional variances are 0.55, 0.36, 0.31, and 0.22, respectively. The decreasing trend in these test statistics follows exactly the expected level of rigor of the regression model calibration and validation tests.

As good as these verification results are, they are probably understating the true skill of the tree-ring reconstructions at some grid points. The instrumental PDSI data in the verification period are not likely to be as homogeneous as those in the calibration period due to widely variable station record lengths and fewer station observations per year. In addition, some of the single station records used in the grid may be inhomogenous. Consider, for example, two stations used in the PDSI grid in Nevada: Battle Mountain Airport (40°37′N, 116°52′W, elevation 1381 m) and Elko Federal Aviation Administration (FAA) Airport (40°50′N, 115°47′W, elevation 1548 m). These stations are only 53 and 63 km from the grid point in north-central Nevada (see Fig. 1) and are weighted most heavily in this five-station gridpoint average. Over the 1928–78 calibration period, the summer average PDSIs of these nearby stations have a correlation of 0.67. However, in the 1894–1927 precalibration period common to both stations, their correlation drops catastrophically to 0.05. Most of the loss of fidelity is in the pre-1910 data. If the comparison is made over the 1910–27 period only, the correlation improves somewhat to 0.27, but in either case the correlations are not statistically significant ($p < 0.05$). With regard to the tree-ring estimates of PDSI, the calibration R_c^2 is 0.69, while the verification period RE and CE are -0.67 and -0.95, respectively. In this case, there is little doubt that the poor quality of the pre-1928 instrumental PDSI data is contributing strongly to the negative tree-ring verification. Given the extremely high quality and drought sensitivity of the tree-ring chronologies in the Great Basin, it is almost certain that the trees are doing a better job at estimating regional drought and wetness than are the meteorological stations in northern Nevada prior to 1928. It is likely that similar instrumental data problems exist elsewhere in the grid, but we have not yet quantified the impact of this source of error on the verification statistics.

5. Spatial patterns of calibration and verification

The calibration/verification results provided above were summaries for the entire drought grid. We now take a detailed look at the spatial patterns of the fractional variance statistics to see how homogeneous the results are across space.

Figure 3 shows the contoured maps of these statistics. The R_c^2 map indicates that the regression models for large areas of the U.S. explain 50%–70% of the gridpoint PDSI variance. The weakest calibration areas are in the upper Midwest and in northern New England. The R_v^2 map indicates that the drought estimates over virtually all of the United States covary significantly with the actual PDSI data in the verification period. That is, any R_v^2 value in excess of 0.10 (i.e., a simple $r > 0.32$) is statistically significant using a one-tailed test and $\alpha = 0.05$. The RE and CE maps reveal more clearly some weak areas in the reconstruction grid. Specifically, there are some regions (e.g., the Great Basin, the upper Midwest, and the central Great Plains) with RE and CE both <0. Since these regions verified reasonably well in the R_v^2 map, the loss of fidelity is probably due mainly to differences in mean level between the actual and estimated PDSIs over the verification period. This suspicion was largely verified by a series of equality-of-means tests of the verification data. Figure 4 shows maps of the mean differences between the actual and estimated PDSIs (upper map) and the corresponding *t*-test probabilities for those differences (lower map). Most of the differences fall in the range of ± 0.50 PDSI units, which are rarely significant even for $\alpha = 0.20$. The largest region of significant differences ($p < 0.10$) is located in the Great Basin and Wyoming–Montana areas where the RE and CE statistics are conspicuously negative.

Similar verification problems in the RE and CE maps

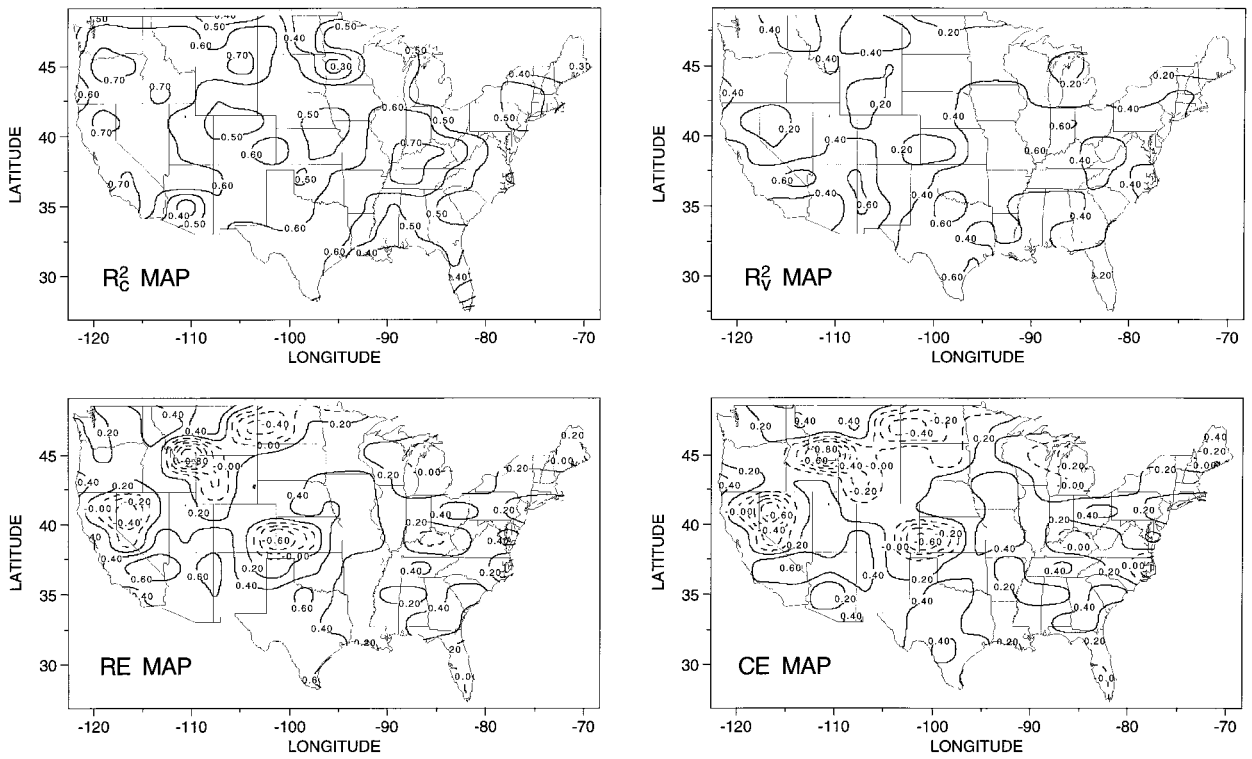

FIG. 3. Contour maps of the R_c^2, R_v^2, RE, and CE gridpoint fractional variance statistics. The calibration map shows that 50%–70% of the PDSI variance was calibrated by tree rings over most of the U.S. The R_v^2, RE, and CE maps indicate, as expected, that there has been some loss of fidelity in the verification period. However, PDSI estimates over most of the U.S. is also verify significantly. The areas of negative RE and CE are mostly in areas where the mean levels of the actual and estimated PDSIs differ significantly ($p < 0.10$; see Fig. 4).

around the Great Basin were also found by Cook et al. (1996) in their earlier tests of the PPR method. Their analyses of the gridded instrumental PDSI data indicated that the regions of weak verification in Fig. 3 could be due in part to decreasing quality of the instrumental data, the Great Basin being the most likely case in point (Cook et al. 1996). However, it is also true that the quality and coverage of tree-ring data is somewhat poor in the other areas that do not verify well. Therefore, the gridpoint reconstructions in these regions should be used with more caution than those where the verification tests are all significant. The quality of these reconstructions should improve if new tree-ring chronologies can be developed in areas of poor coverage.

Finally, we examined the degree to which the reconstructions preserved the pattern of PDSI variability across the U.S. For example, it is known that drought variability is generally greater in the western half of the U.S. However, there is no guarantee that the reconstructions have preserved the spatial pattern of drought variability because of variance lost by regression. Yet, such information is important to retain if one is to make meaningful comparisons of drought variability over different time periods in the reconstructions. Figure 5 shows the contoured maps of gridpoint standard deviations for the actual and estimated PDSIs over the calibration period. There appears to be a high degree of similarity between the two maps, even down to the local level of detail. Indeed, the maps have a correlation of 0.91. The preservation of the standard deviation pattern is due to the reasonably homogeneous pattern of calibrated variance shown in the R_c^2 map. Thus, the reconstructions ought to be quite useful for studying changing patterns of drought variability across the U.S. over the past three centuries.

6. Additional spatial comparisons

The calibration/verification results for the PPR method indicate that the PDSI reconstructions are generally valid expressions of drought over the United States. However, the PPR method does not contain any explicit spatial component other than that related to the search radius. The spatial relationships of drought in the United States (sensu KK) extend over regions that are generally larger than the 450-km search radius used here. Thus, while the individual gridpoint reconstructions are reasonably accurate in a temporal sense, they are not necessarily guaranteed to be valid in a spatial sense beyond the scale of the search radius used here. For this reason, we will describe here a number of analyses that collectively indicate that the spatial patterns of drought have also been well reconstructed by the PPR method.

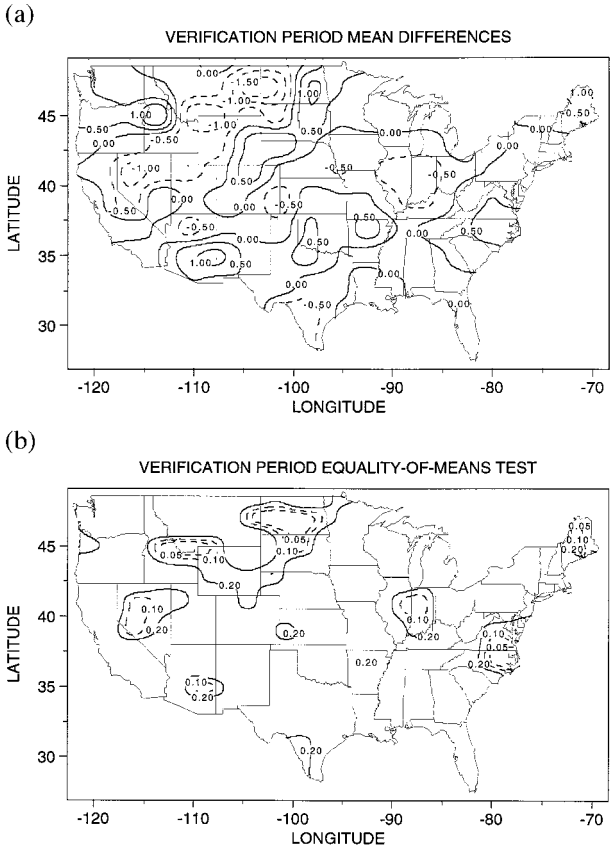

FIG. 4. Equality-of-means tests in the verification period. The contoured mean differences in PDSI units are shown in the upper map. The contoured *t*-test probabilities of those mean differences are shown in the lower map for regions where $p < 0.20$, $p < 0.10$, and $p < 0.05$. Most of the regions that have significantly different ($p < 0.10$) means also have negative REs and CEs (see Fig. 3).

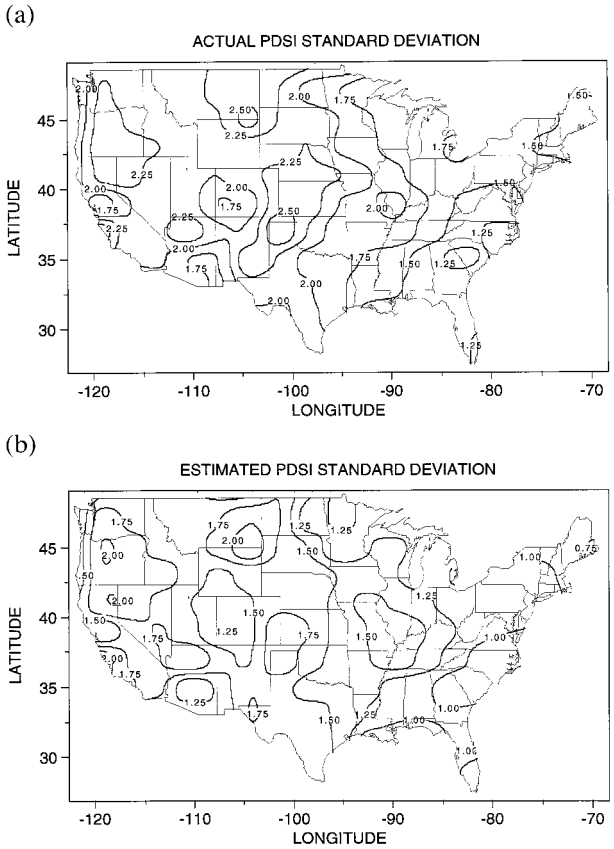

FIG. 5. Contour maps of the standard deviations of the actual and estimated PDSIs. These maps show the excellent degree to which the tree-ring estimates have preserved the spatial patterns of drought variability over the U.S.

a. Mean field, correlation, and congruence analyses of yearly PDSI maps

First, we will describe the degree to which the yearly maps (or spatial patterns) of reconstructed PDSI agree with those based on the gridded instrumental data. For this purpose, we will compare (a) the mean PDSI fields over time, (b) the relative spatial patterns using the Pearson correlation coefficient, and (c) the absolute spatial patterns using the congruence coefficient. The reconstructed patterns will be compared to two gridded instrumental PDSI datasets: the original one used for calibration purposes and a new one based on climate division PDSI data not directly used for calibration purposes. Comparisons using the former will be made over the 1874–1981 time period for which there are at least 10 grid points of actual and reconstructed PDSI data, while the latter has data for all grid points back to 1895 only. The rationale for using the gridded climate division data for additional tests is discussed next.

As described earlier, the single-station instrumental PDSI records used in the grid vary in length over space. Thus, some grid points have much longer records than others. The ending years of the tree-ring chronologies also constrain the end year of analysis. This is revealed in Fig. 6a, which shows the number of grid points with actual and reconstructed PDSI data as a function of time. The maximum number (154) covers the period 1913–78. After 1978, the number of reconstructed grid points declines precipitously to 10 by 1981 due to the ending years of the tree-ring chronologies used. Prior to 1913, the number of grid points declines to 143 by 1903, 97 by 1895, 38 by 1890, and 10 by 1874, all due to the changing beginning years of the instrumental PDSI data over the grid. The quality of the instrumental PDSI grid declines back in time as well because the number of single-station records used to estimate the gridpoint values for each year decreases as shorter records drop out of the gridding process. For these reasons, some degradation in the map comparisons ought to be expected, which is unrelated to the true fidelity of the tree-ring estimates.

In contrast, the gridded climate division data are available for all 154 grid points back to 1895 (Fig. 7a), and the number of single-station records represented in the climatic division summaries is generally greater than

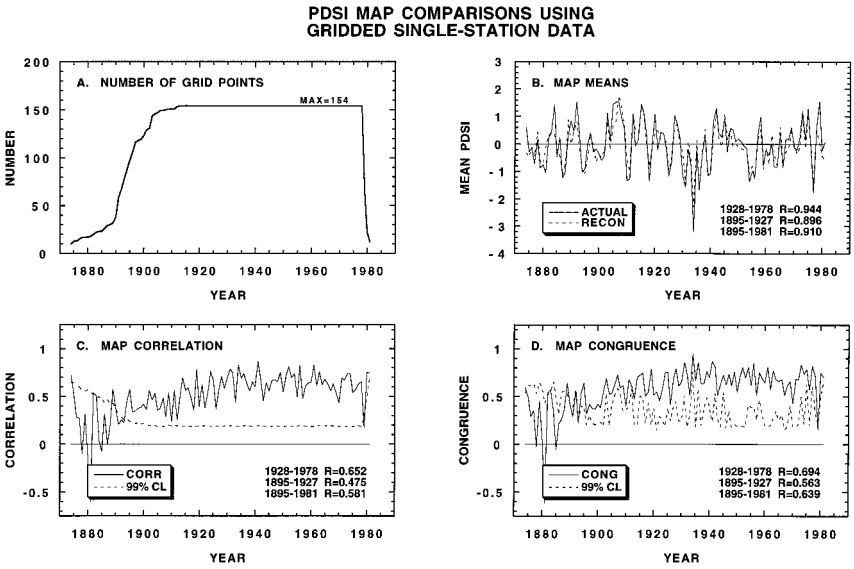

FIG. 6. Annual map comparisons of the actual and estimated PDSIs. The actual PDSI data are from the single-station grid used for calibration and verification. The number of grid points available for comparison each year is shown in (a). Note the steep drop-off in the number of available grid points prior to 1903. The grand mean PDSI averaged over all available grid points for each year is shown in (b). Note the very high correlations between the actual and estimated grand means in both the calibration and precalibration periods. The series of annual map correlations and congruences are shown in (c) and (d), respectively. The 99% confidence limits for each were determined by randomization tests. Most of the annual map comparisons exceed the 99% confidence level by wide margins.

that available from the Historical Climate Network alone. Consequently, the map comparisons (especially before 1928) ought to be more robust using the gridded climatic division data. Finally, the relationships between the actual and reconstructed data over the 1928–78 calibration period have been optimized in a least squares sense only for the single-station grid. The gridded climate division data, while highly related to the single-station grid, do not suffer from this constraint. Therefore, these additional spatial comparisons should pro-

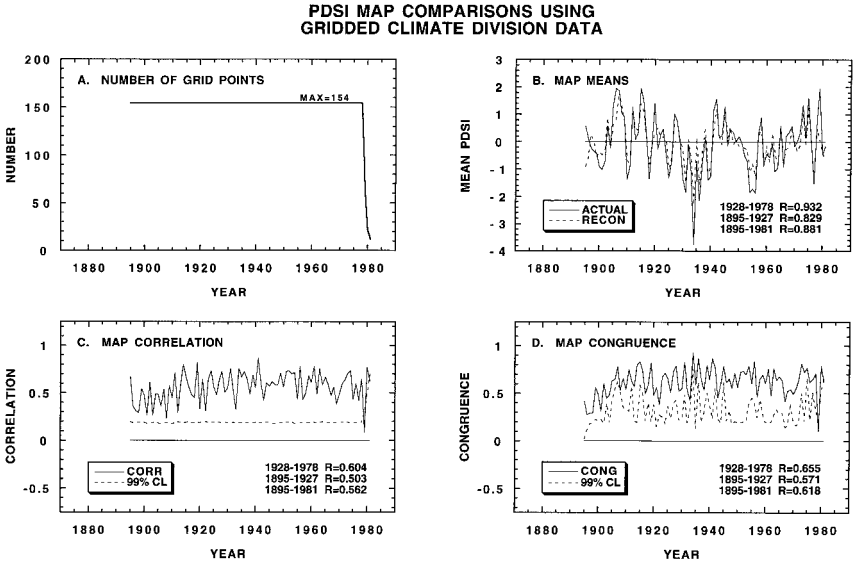

FIG. 7. Annual map comparisons of the actual and estimated PDSIs. This time the actual data are gridded climate division PDSI data, which all begin in 1895. These data avoid the steep pre-1903 loss of grid points in the single-station grid (Fig. 6a). See the Fig. 6 caption for more details.

vide a less biased evaluation of the true fidelity of the reconstructed PDSI maps.

1) MEAN FIELD COMPARISONS

A simple first-order comparison of the actual and reconstructed PDSI maps can be made by simply comparing the mean fields over time. That is, for each year all of the available gridpoint values are averaged together, separately for the actual and reconstructed data. The resulting grand mean time series are shown in Figs. 6b and 7b for the single-station and climate division grids, respectively. Although the single-station data go back to 1874, correlations between the mean actual and reconstructed series have only been calculated back to 1895 for comparison with the climatic division grid results.

There is a high degree of similarity between the mean series in each case. First, for the single-station grid, the correlation between actual and reconstructed PDSI is 0.944 over the 1928–78 calibration period. The correlation then drops to 0.896 in the 1895–1927 precalibration (i.e., verification) period, but is still highly significant. And over the entire 1895–1981 period, the correlation is 0.889. For the climate division grid, the correlations are comparable: 0.932 for the 1928–78 calibration period, 0.829 for the 1895–1927 precalibration period, and 0.881 overall. Clearly, the mean PDSI fields are well estimated by tree rings back to 1895 at least and even where the number of grid points drops to 10 in the 1874–94 period (see Fig. 6b). The similarity of the calibration period correlations based on the different grids indicates that, at this spatial scale of comparison, there is no significant least squares fitting bias in the reconstructions. These results are highly encouraging. However, this analysis does not reveal how well the spatial patterns that make up the mean fields have been replicated each year by the tree rings. To determine this, we will use map correlation and congruence analyses.

2) MAP CORRELATION

To test the relative agreement between actual and reconstructed PDSI maps, we used the Pearson product-moment correlation coefficient. First, we normalized each gridpoint reconstruction using its calibration period mean and standard deviation. This was done to avoid any regional bias in the map correlations due to spatial variations in the PDSI standard deviation (see Fig. 5). Then, for each year i of the n-yr overlap period between actual and reconstructed PDSI, we calculated

$$r_{pq} = \frac{\sum (p_k - \overline{p})(q_k - \overline{q})}{\left[\sum (p_k - \overline{p})^2 \sum (q_k - \overline{q})^2\right]^{1/2}}, \qquad (2)$$

where the summation extends over the $k = 154$ PDSI grid points, p_k and q_k are the actual and reconstructed PDSI values at grid point k, respectively, and $\overline{p}$ and $\overline{q}$ are the mean fields for each year i, for example, those shown in Fig. 6b. Because the mean fields are subtracted from the gridpoint values, the information contained in them, which is a real component of drought over the coterminous United States, does not contribute to the estimate of r_{pq}.

Figure 6c shows the time series of map correlations (solid line) for the single-station grid. Also included is the one-tailed 99% confidence level (dashed line) estimated by randomizing the reconstructed PDSIs and calculating the correlation of this randomized field with the actual data. This was done 5000 times for each year. The resulting significance levels are very close to that which would have been obtained by the standard t test of the correlation coefficient. Note that back to 1889, where the number of grid points exceeds 30, the map correlations all exceed the 99% level. On average, the highest correlations occur during the 1928–78 calibration period ($\overline{r} = 0.652$), a result that is probably related to the least squares optimization. Prior to 1928, the correlations decline systematically (e.g., $\overline{r} = 0.475$ over the 1895–1927 precalibration period), a result that is consistent with the difference in the pointwise calibration/verification statistics reported earlier. Prior to 1889, the map correlations are mostly nonsignificant, in some cases catastrophically so. This result is probably a combination of a number of things that are at least partly unrelated to the overall quality of the PDSI reconstructions. Some of the longest (i.e., pre-1889) instrumental PDSI records come from the eastern U.S., where the calibration/verification statistics are relatively weak. So, a decline in spatial correlation is not too surprising when that restricted area contributes the most to the correlation analyses. And it is also likely that the quality of the actual gridded data declines back in time as the number of individual station records used in the grid declines.

The map correlations based on the climate division grid are shown in Fig. 7c. Except for 1979, when the number of grid points is small, all correlations exceed the 99% significance level. The average map correlation over the 1928–78 calibration period ($\overline{r} = 0.604$) is somewhat weaker than that based on the single-station grid ($\overline{r} = 0.652$), a result that is probably related to the lack of prior least squares fitting bias here. However, there is less evidence of the systematic decay in the map correlations prior to 1928 in this case, and the average correlation over the 1895–1927 period ($\overline{r} = 0.503$) actually exceeds that based on the single-station grid ($\overline{r} = 0.475$).

3) MAP CONGRUENCE

The map correlations are very useful for describing how well the spatial patterns of drought covary in a relative sense. However, because the mean fields are a

true component of drought variability over time (see Figs. 6b and 7b), it is desirable to include this information in the spatial analyses as well. To do this, we have used the congruence coefficient, which was originally developed as a measure of the similarity between two factor patterns in multivariate research (Richman 1986; Broadbrooks and Elmore 1987). The congruence coefficient is computed for each year of the overlap period as

$$c_{pq} = \frac{\sum p_k q_k}{\left[\sum p_k^2 \sum q_k^2\right]^{1/2}}, \quad (3)$$

where p_k and q_k are the actual and reconstructed PDSI values at grid point k. Note that the only difference between c_{pq} and r_{pq} is the lack of $\overline{p}$ and $\overline{q}$ here. Thus, the means are not removed in computing c_{pq}, leading some earlier studies to describe the congruence coefficient as an unadjusted correlation coefficient (Broadbrooks and Elmore 1987). The theoretical range that c_{pq} may take is the same as r_{pq}. However, the presence of $\overline{p}$ and $\overline{q}$ in the calculation of congruence means that c_{pq} tends to be biased toward 1.0 relative to r_{pq} (Richman 1986). There is no theoretical sampling distribution for testing the significance of c_{pq} because of its partial dependence on $\overline{p}$ and $\overline{q}$, which are random variables in their own right. Several studies have used Monte Carlo methods to generate empirical limits for c_{pq} (e.g., Broadbrooks and Elmore 1987). Here, we have generated our own significance levels using the randomization procedure described for the correlation coefficient.

Figure 6d shows the map congruence coefficients over time for the single-station grid, along with the empirical 99% confidence levels. The results are qualitatively similar to that found by correlation alone. The highest average congruence occurs in the 1928–78 calibration period ($\overline{c} = 0.694$), followed by declining but still significant congruence from 1895 to 1927 ($\overline{c} = 0.563$). Prior to 1895, congruence remains significant back to 1889 and then becomes very poor. A close comparison of the correlation and congruence plots reveals that, as expected, there is a small positive bias in the estimates of the latter (e.g., $\overline{r} = 0.581$ and $\overline{c} = 0.639$, respectively, over the 1895–1981 period). Perhaps the greatest difference in the two analyses is in the estimates of the confidence levels. The dependence of c_{pq} on $\overline{p}$ and $\overline{q}$ has resulted in a highly variable and erratic series of confidence levels that are clearly related to the variability between the mean fields shown in Fig. 6b.

The map congruences based on the climate division grid are shown in Fig. 7d. These results are again qualitatively similar to the correlation results. As before, there is greater long-term stability in map congruence using this grid for comparison with the reconstructed maps, and, except for 1979, all congruences exceed the 99% confidence level. All other details concerning the map correlations above are similar here; for example, the average map congruence for the 1928–78 calibration period ($\overline{c} = 0.655$) is weaker than that based on the single-station grid ($\overline{c} = 0.694$), but the average congruence over the 1895–1927 period ($\overline{c} = 0.571$) actually exceeds that based on the single-station grid ($\overline{c} = 0.563$). None of these results indicate any serious deficiency in the annual PDSI maps produced by the tree-ring estimates.

4) Discussion

The analyses shown in Figs. 6 and 7 indicate that the spatial patterns of PDSI in the reconstructions have captured those in the actual data with high ($p < 0.01$) statistical fidelity. Prior to 1895 some of the differences between the actual and reconstructed PDSI clearly arise from the deterioration of the instrumental PDSI coverage in time and space. The mean fields of the reconstructions have been estimated extremely well back as far as it is possible to test them, while the more detailed reconstructed annual maps are significantly related to the actual PDSI fields back to 1895 at least. The correlations between the actual and reconstructed mean fields are higher than the mean calibration and verification results of the individual grid points using PPR. This result indicates that the gridpoint reconstructions contain more ''local noise'' than do larger-scale averages of PDSI, with the mean fields studied here being the limiting case of averaging over all grid points (i.e., a continental-scale average). Averaging gridded climate reconstructions spatially to improve the signal-to-noise ratio has been used with considerable success (e.g., Fritts 1991; Mann et al. 1998) and is clearly justified if the main interest is in climate variability at spatial scales larger than that initially reconstructed. However, how one averages the reconstructions spatially to preserve insights into the physical climate system needs to be carefully considered. One approach to this problem is to use rotated principal component analysis (Richman 1986) to objectively define the natural regional drought climatologies in the United States (sensu KK). In the next section, we will use rotated principal component analysis to demonstrate that the regional drought patterns found by KK in their instrumental data can be reproduced well in most cases by our PDSI reconstructions.

b. Rotated principal component analysis

Rotated principal component analysis (RPCA) is a powerful tool for objectively decomposing spatial arrays of climate data into natural regional clusters or patterns (e.g., Barnston and Livezey 1987). Richman (1986) reviewed this topic in detail and performed a number of Monte Carlo experiments to test the performance of a number of rotation methods. We have used two of the recommended rotation methods here: orthogonal vari-

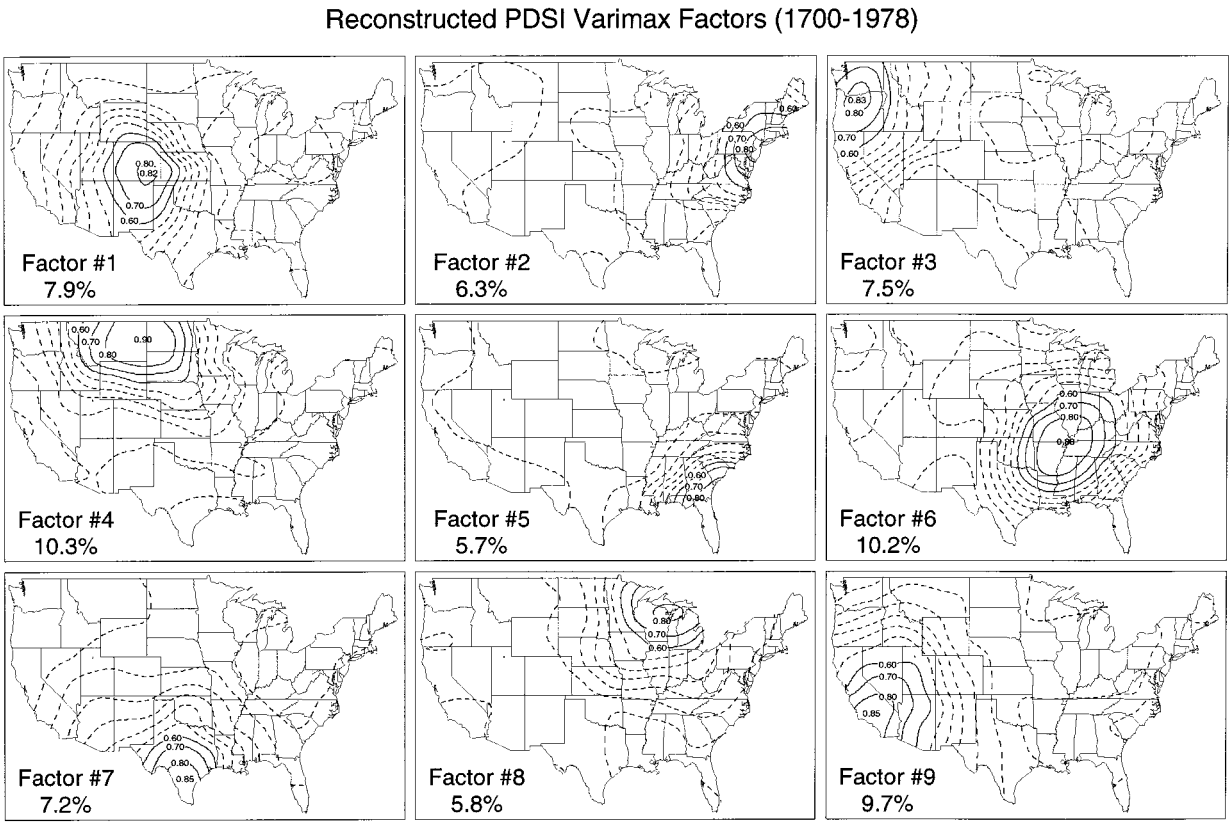

FIG. 8. Reconstructed summer drought varimax factors for the U.S. (1700–1978). The percent variance accounted for by each is indicated in each map. Most of these factors agree well with those of KK based on monthly instrumental PDSIs. The largest discrepancy is with our factor 1, located mainly over Colorado, which has replaced the ''southwest'' factor of KK, located mainly over Arizona. This discrepancy is investigated further in Fig. 9.

max rotation (Kaiser 1960) and oblique promax rotation (Hendrickson and White 1964). The former is perhaps the most popular orthogonal rotation method in use today, while the latter was shown by Richman (1986) to be one of the best oblique rotation methods for recovering the underlying true ''simple structure'' in his synthetic data tests. (Following the recommendation of Richman, the promax method was applied using a power $k = 2$.) Each method rotates the axes of a retained subset of unrotated principal components in order to achieve some degree of simple structure among variables. The resulting rotated factor loadings should be near one or near zero ideally and each variable should load heavily on one or, at most, a small number of factors only (Reyment and Jöreskog 1993). The loadings themselves can also be interpreted as simple correlations between each factor and the original variables.

In the context of this study, the variables are the PDSI grid points and the observations are the associated time series of PDSIs. The varimax rotation maintains orthogonality between the resulting factors, while promax rotation allows for intercorrelations between the factors to emerge as part of the simple structure solution. In the context of RPCA of climate data fields, it is arguable that orthogonal factors are physically unrealistic. Therefore, an oblique solution might be preferred.

We used RPCA here to see how well the regional drought patterns in our reconstructions agree with those of KK. In their analyses, KK applied both varimax and oblique (direct quartimin) rotation to a 60-point grid of monthly instrumental PDSI data (1895–1981) across the continent. In so doing, they identified nine regional drought factors that they were able to associate with distinct, regional precipitation climatologies. We will assume this subspace dimension in our study here to simplify the analyses, but do not expect 1:1 congruence between our factors and those of KK. Unlike our summer PDSIs, the factors produced by KK were based on monthly data. The difference in grid resolution (154 vs 60) between our two studies could also affect the location of the factor boundaries. However, we should expect to see some strong similarities because the regional precipitation climatologies described by KK should contribute strongly to our summer PDSI estimates as well.

Figure 8 shows the reconstructed drought factor maps over the full 1700–1978 reconstruction period, based on the varimax rotation method. Most of the regional

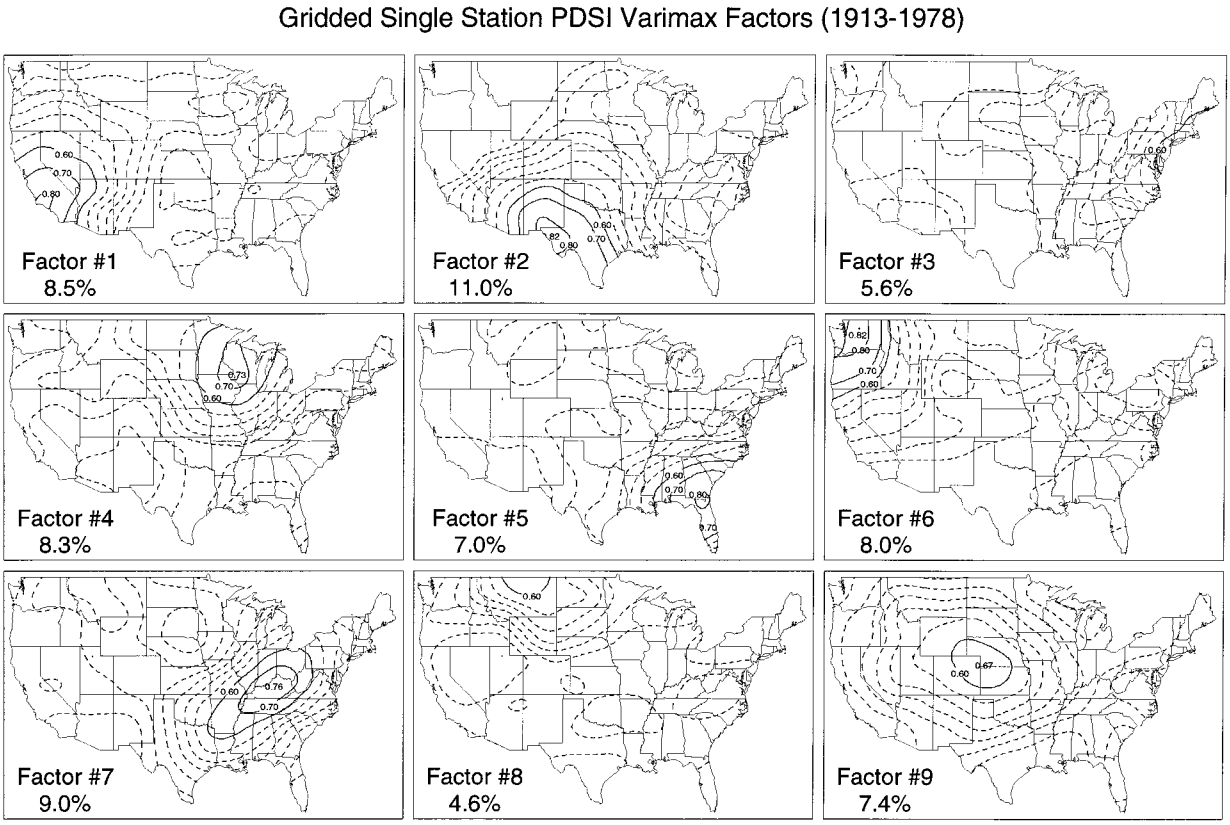

FIG. 9. Actual summer drought varimax factors for the U.S. (1913–78) based on the single-station grid. These maps agree very well with those in Fig. 8, even to the extent that the ''Colorado'' factor in our reconstructions is present in the actual data as well. Thus, the southwest factor of KK, calculated over the 1895–1981 period, may be more related to nonsummer PDSI variability.

drought patterns described in KK are clearly evident here. For example, factors 2–9 correspond closely to their northeast, northwest, west–north-central, southeast, central, south, east–north-central, and west factors (KK, their Fig. 4). The biggest difference between the factor analyses is between our factor 1 (located mainly over Colorado) and their southwest factor (located mainly over Arizona). To see if this difference was an artifact in our reconstructions, we performed RPCA on the single-station summer PDSI grid used for calibration and verification. In this case, the analysis period was 1913–78, common to all grid points. The nine varimax factors of the instrumental data are shown in Fig. 9. A visual comparison of the equivalent varimax factors in Figs. 8 and 9 (e.g., factor 9 with factor 1) indicates that the tree-ring estimates have indeed captured the regional summer drought climatologies in the U.S. very well. In particular, factor 9 in Fig. 9 (located equally over Colorado, Kansas, and Nebraska) is very similar to factor 1 in the reconstructions. So, it appears that the southwest factor of KK is not evident when only summer PDSI is evaluated. Rather, it is replaced by a drought factor located mainly over the eastern Colorado region.

Oblique promax rotation was applied next to the drought reconstructions. Although there were some apparently significant correlations between some of the oblique factors (the four largest correlations are $r_{1,4} = 0.28$, $r_{1,7} = 0.33$, $r_{1,9} = 0.32$, and $r_{2,6} = 0.27$), the varimax and promax factors (not shown) were essentially identical. Therefore, the orthogonal varimax solution appears to be an adequate representation of the underlying regional summer drought factors in the U.S.

Figure 10 shows the factor scores estimated from the Fig. 8 varimax factors. These nine time series provide histories of relative drought and wetness for the identified regions. Some of the well-known droughts of the twentieth century are indicated in these histories. For example, the serious drought that struck the Northeast in the 1960s (Cook and Jacoby 1977) is indicated in the factor 2 scores. The drought that occurred during the ''Dust Bowl'' years of the 1930s (Stockton and Meko 1983) shows up in the scores of factors 3 and 4 in the Pacific Northwest and northern Great Plains. And the drought that struck the Texas region in the 1950s (Stahle and Cleaveland 1988) is revealed in the scores of factor 7. In contrast, a notable wet period in the early 1900s is indicated in several of the factors. Prior to 1900, the factor scores reveal periods of drought and wetness that in some cases appear to be unprecedented in the record. The factor scores also show varying degrees of inter-

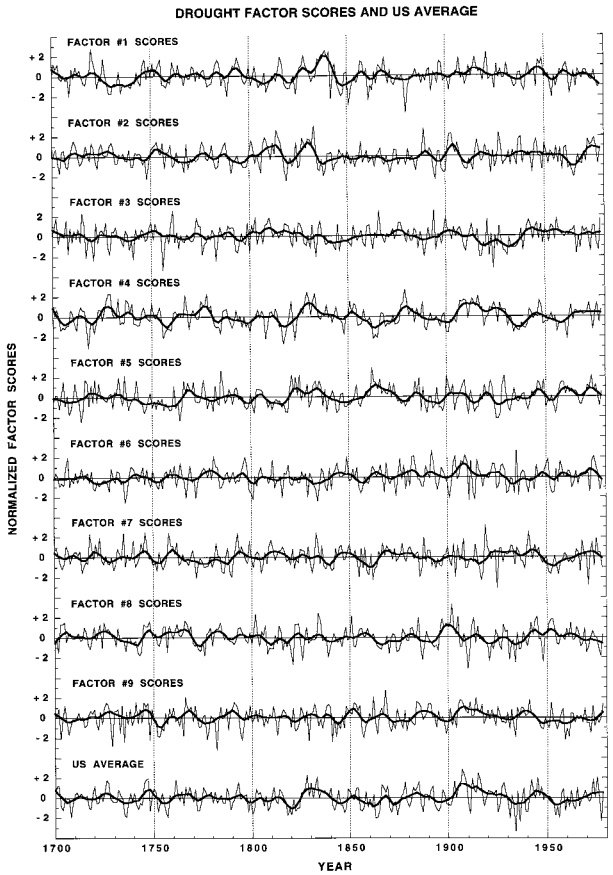

FIG. 10. The nine varimax factor scores of the loading maps shown in Fig. 8. These scores represent the regional histories of relative drought and wetness in the US since 1700. They have also been smoothed to emphasize the interdecadal (>10 yr) timescale of variability. The bottom time series is a weighted average of relative drought and wetness for the U.S. based on the scores of the first unrotated principal component of the reconstructions. It shows that the 1930s Dust Bowl drought was easily the most severe such event to hit the U.S. since 1700.

decadal variability, with factor 4 in the northern Great Plains showing the most multiyear power. Clearly, there is much to be learned about U.S. drought from the reconstructions produced here.

Also included in Fig. 10 is a plot of the scores of the first unrotated principal component of drought (U.S. average), scaled to have variance comparable to the varimax scores. This component accounts for 22% of the total PDSI variance. This series is effectively a (weakly) centrally weighted average of the actual drought reconstructions. Based on the very high correlations (~0.90) between the actual and estimated mean fields shown in Figs. 6b and 7b, this record is a highly accurate history of drought for the U.S. as a whole. It shows that the Dust Bowl drought of the 1930s was clearly the most severe event to strike the U.S. since 1700. Indeed, the two most extreme years in this series are 1934 and 1936. The third most severe drought year occurred in 1977. Other notable dry periods occurred around 1820, 1860, and 1950. Two unusual wet periods also stand out in the 1820–40 and 1900–20 intervals.

c. Discussion

The results of our spatial comparisons between instrumental PDSIs and the tree-ring estimates indicate that the reconstructions have captured the inherent spatial variability of drought across the U.S. with a high degree of fidelity. This has been accomplished using PPR, which does not contain any explicit spatial component in its design, except for the selected search radius. The ability of PPR to reconstruct the spatial features of drought is undoubtedly due to the underlying regionality of climate in the U.S. that determines patterns of drought and wetness and, ultimately, patterns of tree growth (e.g., Fritts 1965; LaMarche and Fritts 1971).

7. Concluding remarks

The U.S. drought reconstruction grid produced here is a significant new application of dendroclimatic techniques to reconstruct past climate. The quality of the gridpoint reconstructions is generally quite good and spatially homogeneous. As noted earlier, the automated nature of the PPR method virtually guarantees that the gridpoint reconstructions produced here will not necessarily be the best possible at any given grid point. However, the PPR method does guarantee that each gridpoint model is developed in a consistent manner. Its specific implementation here also guarantees a near-global optimum in terms of model verification.

There is little doubt that the joint space–time methods of climate reconstruction (e.g., canonical regression or singular value decomposition) would have resulted in better calibration statistics here (e.g., a higher global R_c^2). These techniques are constrained to find the optimal linear association between two fields of variables. However, they would almost certainly lead to poorer verification results (see the verification surfaces shown in Appendix B, Fig. B1) because they would allow long-range, temporally unstable, tree-ring teleconnections to influence the estimation of the gridpoint PDSIs. The joint space–time methods would also be mathematically ill-conditioned if the original data were used directly here because the number of grid points (155) greatly exceeds the number of years (51) used for calibration. This problem can be ameliorated by reducing the dimensions of the problem through principal component analysis of the predictor and/or predictand fields prior to regression (Fritts 1976; Cook et al. 1994; Mann et al. 1998) but at the loss of higher-order dimensionality in the resulting field of reconstructions. The reconstructions produced by PPR do not have this limitation.

These reconstructions are freely available from the National Geophysical Data Center (NGDC), World Data Center-A for Paleoclimatology, in Boulder, Colorado.

(The Web site addresses for these data are http://www.ngdc.noaa.gov/paleo/pdsiyear.html for the annual PDSI maps and http://www.ngdc.noaa.gov/paleo/usclient2.html for the summer PDSI time series.) We hope that future analyses of these reconstructions will lead to an improved understanding of drought variability in the U.S., especially on the interdecadal timescale (e.g., Karl and Riebsame 1984), which is very difficult to investigate with instrumental climate data alone.

We expect to improve the drought reconstructions for the United States in the future through the use of improved statistical methods and, more importantly, the development of new tree-ring chronologies in areas that are now weakly modeled. In addition, efforts are now under way to extend the reconstructions back in time to enable analyses of drought variability at longer timescales. Consequently, the PDSI reconstructions on the National Oceanic and Atmospheric Administration Web sites will be updated periodically as these improvements are made.

Acknowledgments. This research is supported by the National Oceanic and Atmospheric Administration, Office of Global Programs, Paleoclimatology Program through Grants NA36GP0139 and NA66GP0251, and is an outgrowth of previous support by the National Science Foundation, Grant ATM 88-14675 and the U.S. Geological Survey, Biological Resources Division, Agreement CA-8012-2-9001. The PDSI data were kindly obtained from Ned Guttman of NOAA. Most of the tree-ring chronologies used here came from the International Tree-Ring Data Bank (ITRDB) at the NGDC in Boulder. Others not in the ITRDB were kindly contributed by a number of scientists, including Hal Fritts, Lisa Graumlich, Tom Swetnam, Connie Woodhouse, Laura Haston, Joel Michaelsen, and Glen MacDonald.

APPENDIX A

The Point-by-Point Regression Method

Principal components regression analysis is the foundation of PPR and is described in Briffa et al. (1986) and Cook et al. (1994) for the multiple-predictor–multiple-predictand case. Here we give a brief description of this method, restricting it to the multiple-predictor–single-predictand case that is appropriate to PPR.

Let

$$\mathbf{y}_k = \mathbf{UB} + \mathbf{e}_k, \tag{A1}$$

where $\mathbf{y}_k$ is the vector of standardized (i.e., zero mean, unit standard deviation) instrumental PDSIs at grid point k, $\mathbf{U}$ is the matrix of orthogonal tree-ring principal component scores, $\mathbf{B}$ is the matrix of regression coefficients, and $\mathbf{e}_k$ is the vector of regression model errors. The actual tree-ring series used as predictors are related to their scores as

$$\mathbf{U} = \mathbf{XF}, \tag{A2}$$

where $\mathbf{X}$ is the matrix of standardized tree-ring chronologies used as predictors and $\mathbf{F}$ is the orthonormal matrix of column eigenvectors calculated from the correlation matrix of $\mathbf{X}$. Each of the k PPR regression models is developed here over the *calibration* time period 1928–78 common to the predictors and predictands, with the pre-1928 instrumental PDSI data reserved for regression model *verification* tests of the tree-ring model estimates. See Fritts (1976) and Cook et al. (1994) for more details concerning the calibration–verification procedures commonly used in dendroclimatology.

Once the regression coefficients in $\mathbf{B}$ have been estimated for the calibration period, they can be applied to the precalibration period tree-ring scores after projecting the early tree-ring data onto the relevant eigenvectors in $\mathbf{F}$. The resulting augmented tree-ring scores in $\mathbf{U}$ are then used to produce a series of standardized PDSI estimates back in time as

$$\hat{\mathbf{y}}_k = \mathbf{UB}, \tag{A3}$$

after which they are back-transformed into original PDSI units. Here we restrict our PDSI reconstructions to cover the period 1700–1978, which is the time interval common to all tree-ring chronologies in our network.

The description of the principal components regression model given in (A1)–(A3) is generic in that any tree-ring data can be used to form the principal component scores in $\mathbf{U}$. However, we have found that the reconstructions of PDSI from tree rings can be significantly improved through the careful use of autoregressive (AR) prewhitening of the both the tree rings and PDSI data prior to regression analysis. This procedure is used to correct for the sometimes large differences in short-lag autocorrelation between climate and tree rings that are believed to be due to physiological and stand dynamics effects on annual ring widths, which are unrelated to climate. In this case, low-order AR(p) models are fit to the time series used at each grid point as

$$\mathbf{z_t} = \sum_{i=1}^{p} \phi_i \mathbf{z_{t-i}} + a_t, \tag{A4}$$

where z_t is the PDSI or tree-ring series used at grid point k, ϕ_i is the AR coefficient at lag-i yr, and a_t is the resulting series of ''white noise'' (i.e., serially random) residuals (Box and Jenkins 1970). The order p is objectively determined using the Akaike information criterion (AIC; Akaike 1974), with a correction for small sample bias (Hurvich and Tsai 1989). The original and corrected AICs are calculated as

$$\mathrm{AIC} = N \ln\sigma_e^2 + 2(m + 1) \tag{A5}$$

and

$$\mathrm{AIC}_c = \mathrm{AIC} + \frac{2(m + 1)(m + 2)}{N - m - 2}, \tag{A6}$$

respectively, where N is the number of observations,

σ_e^2 is the residual variance of the model, and m is the number of explanatory variables. In the case of AR modeling, $m = p$.

The a_t of PDSI and tree rings are then used as $\mathbf{y}_k$ and $\mathbf{X}$ in (A1) and (A2), respectively, instead of the original variables. The serially random property of the AR residuals simplifies tests of association between tree rings and drought because the degrees of freedom do not need to be corrected for persistence. In addition, the identification of lagged responses between tree rings and drought are also simplified by first prewhitening the time series prior to testing for lead-lag associations (Haugh and Box 1977).

After the $\hat{\mathbf{y}}_k$ are produced for grid point k, any autocorrelation in the PDSI that had been modeled and removed must be added back into the reconstruction. This operation usually involves adding some ''redness'' (i.e., positive autocorrelation) to the reconstructed PDSIs because most of the instrumental PDSI records behave as moderate ''red noise'' processes (sensu Gilman et al. 1963). This procedure is accomplished by substituting $\hat{\mathbf{y}}$ for $\mathbf{z}$ in (A4) and using the ϕ_i of the grid point k PDSI series, with the necessary p starting values estimated by backcasting. It closely follows the ''random shock model'' method of Meko (1981) for reconstructing precipitation from tree rings.

The PPR method allows for precise control over which tree-ring chronologies and their principal component scores enter into the regression equation for reconstructing PDSI at each grid point. This control is exercised in four sequential stages of model development, which culminate in the selection of the final tree-ring predictors of drought. These stages result in the creation of four ''pools'' of tree-ring variables that seek to concentrate the common drought signal and winnow out the nondrought noise.

The level-1 pool of tree-ring variables contains those chronologies that are believed to be well related to drought at a given grid point due to their proximity to it alone. As noted earlier, the PPR method assumes that only those tree-ring chronologies proximal to a given PDSI grid point are likely to be true predictors of drought. Here we operationally define ''proximal'' to mean those tree-ring chronologies located within a given search radius around a PDSI grid point. This radius should be small enough to preserve the local and regional character of PDSI at each grid point but also large enough to include most or all of the ''true'' tree-ring predictors of drought.

The ideal search radius for this purpose would seem to be the same as that used for gridding the single-station PDSI records, that is, 150 km. However, this distance is generally impractical for many areas of the PDSI grid because the tree-ring network is much less dense and more patchy than the original single-station PDSI network. In addition, the central Great Plains is largely devoid of tree-ring chronologies, making a 150-km search radius clearly impractical there. Finally, regional drought anomalies ought to exceed 150 km in radius on average given the size of the regional drought climatologies in the U.S. (cf. the drought factor maps in KK). This suggests that tree-ring chronologies from some greater distance (i.e., $>$150 km) ought to be useful predictors of PDSI at a grid point. The inherent patchiness of the tree-ring network also means that many areas of the grid will require a relatively small fixed search radius to find enough tree-ring chronologies, while other areas will require a larger search radius. This indicates the need for a dynamic search radius that will enlarge until a minimum number of tree-ring chronologies has been found for a given grid point. The minimum number used here is five, a compromise between locating a reasonable number of tree-ring chronologies per grid point versus the desire to minimize the size of the eventual search radius to preserve the meanings of proximal and true provided above. So, given a prescribed minimum search radius, the search for tree-ring chronologies is conducted as follows. If five or more series are found within the minimum radius from a PDSI grid point, the search is considered successful and terminated. If not, the search radius is expanded by 50-km increments until at least five chronologies are found.

The level-2 pool contains those tree-ring variables from the level-1 pool that are well correlated with PDSI. The level-1 search procedure only deals with finding candidate tree-ring chronologies within a given radius of each PDSI grid point. However, there is no guarantee that these candidates will be significantly correlated with PDSI at a grid point. For this reason, they are next subjected to statistical screening prior to use in regression analysis. This is accomplished by correlating the prewhitened candidate tree-ring variables with the prewhitened PDSIs over the calibration period 1928–78. The correlations are calculated using both year t and $t + 1$ tree-ring residuals as candidate predictors of year t PDSI residuals to allow for a 1-yr lag response to climate found in some tree-ring chronologies. Thus, for m candidate tree-ring chronologies found within a given search radius, there are actually $2m$ candidate predictors of PDSI at a grid point.

The screening criterion used is the two-tailed hypothesis test of the Pearson correlation coefficient (say $\alpha = 0.05$) with $n - 2$ degrees of freedom, in this case 49 for the 1928–78 calibration period. Depending on the α-level probability used for screening the level-1 candidate pool and the strength of the PDSI signal in the chronologies, the number of retained candidate predictors may be $\ll 2m$. This reduced set of m' tree-ring variables is the level-2 pool [matrix $\mathbf{X}$ in (A2)] that is subjected to principal components analysis in the next stage of PPR.

The level-3 pool contains the tree-ring principal components that are retained as candidate predictors of PDSI in multiple regression analysis. Principal components analysis is used to reduce the size of the level-2 pool and concentrate common drought signal(s) further. Its

main virtues are the way in which it orthogonalizes the intercorrelated set of predictors and reduces the dimensions of the regression problem through the elimination of higher-order eigenvectors that account for very little variance. We use the objective Kaiser–Guttman eigenvalue-1 rule (Guttman 1954; Kaiser 1960) to eliminate those higher-order variables. This eigenvalue cutoff criterion typically reduces the dimension of the level-3 pool m'' to $<0.3m'$.

The level-4 pool of tree-ring predictor variables is the order of each gridpoint regression model, that is, the rank of **U** in (A1). This is accomplished by first correlating the PDSIs with the orthogonal tree-ring scores over the 1928–78 calibration period. The correlations are then ranked in order of decreasing magnitude and entered into the model until the minimum AIC_c criterion is achieved. Because the entered variables are orthogonal, the square of each variable's correlation with PDSI is its partial R^2, so the sum of those squared correlations is the final model R^2.

APPENDIX B

Optimizing the Search Radius and Screening Probability Criteria

As implemented above, the level-3 and level-4 pooling procedures of PPR are fully automatic and objective. However, the level-1 and level-2 procedures contain somewhat ill-defined and subjective elements, these being the choice of the search radius and screening probability, respectively. There is no guarantee that a single, optimal search radius–screening probability combination exists for all points on the grid, where ''optimal'' means in this case the best possible reconstruction of PDSI at each location. To produce such optimal reconstructions would probably require modeling the 155 grid points as a series of completely independent regression problems, a very time-consuming and tedious option. However, one combination or narrow range of combinations of search radius–screening probability may produce on average, or *globally*, the best reconstructions across the grid. If a global optimum could be found for the search radius and screening probability, then the PPR method could be made fully objective and automatic.

To see to what degree a joint search radius–screening probability optimum exists for the PDSI grid, we conducted a number of experiments of PPR in which we varied the search radius and screening probability over a wide range of values. Specifically, we varied the search radius from 200 to 3000 km and the screening α-level probability from 0.05 to 0.40. The search radius upper limit was chosen to allow for possible transcontinental drought teleconnections to enter into the model, while the screening probability upper limit forced most variables into the principal components analysis.

To evaluate the test results, we calculated four statistics as measures of goodness of fit between the actual and estimated PDSI. These tests used the actual PDSI data and the tree-ring estimates after autocorrelation had been added back into them, as described in appendix A. So, for all 155 PDSI grid points we calculated the following.

1) Average explained variance (R_c^2) over the 1928–78 calibration period is

$$R_c^2 = 1.0 - \left[\frac{\sum (x_i - \hat{x}_i)^2}{\sum (x_i - \overline{x}_c)^2}\right], \tag{B1}$$

where x_i and $\hat{x}_i$ are the actual and estimated data in year i of the calibration period and $\overline{x}_c$ is the mean of the actual data. This is a direct measure of the least squares goodness of fit of the regression model that is achieved here by the minimum AIC_c criterion. However, R_c^2 is known to be a very poor, biased measure of true goodness of it when the regression model is applied to data not used for calibration (Cramer 1987; Helland 1987), hence the need for regression model verification tests.

2) Average squared Pearson correlation (R_v^2) over the pre-1928 verification period is

$$R_v^2 = \frac{\left[\sum (x_i - \overline{x}_v)(\hat{x}_i - \overline{\hat{x}}_v)\right]^2}{\sum (x_i - \overline{x}_v)^2 \sum (\hat{x}_i - \overline{\hat{x}}_v)^2}, \tag{B2}$$

where x_i and $\hat{x}_i$ are the actual and estimated data in year i of the verification period and $\overline{x}_v$ and $\overline{\hat{x}}_v$ are the means of the actual and reconstructed data in the verification period. This is a useful measure of covariance between the actual and estimated PDSIs in the verification period of withheld actual data. However, it is not sensitive to differences in mean level between the covariates. Therefore, it is the least rigorous of the three verification tests used here. The Pearson correlation itself also allows for negative relationships between variables, which would be nonsensical in the context of its application here. This never occurred in any of our PPR experiments.

3) Average reduction of error (RE) in the verification period is

$$\mathrm{RE} = 1.0 - \left[\frac{\sum (x_i - \hat{x}_i)^2}{\sum (x_i - \overline{x}_c)^2}\right], \tag{B3}$$

where x_i and $\hat{x}_i$ are the actual and estimated data in year i of the verification period and $\overline{x}_c$ is the mean of the actual data in the calibration period. This statistic was first introduced by Lorenz (1956) to meteorology as a measure of forecast skill, and has been extensively used in dendroclimatology to verify reconstructions of climate from tree rings (Fritts 1976; Kutzbach and Guetter 1980). Here RE has a theoretical range of $-\infty$ to $+1$. An RE > 0 indicates

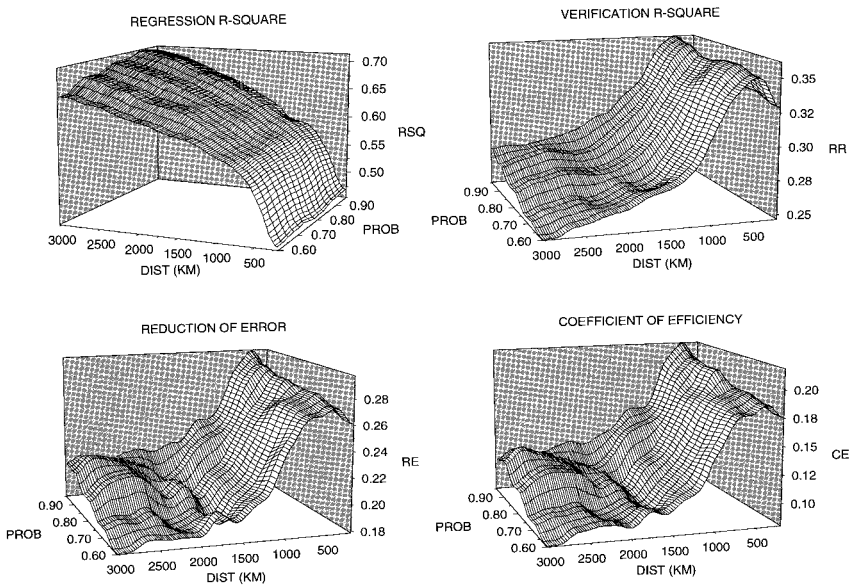

FIG. B1. The regression model calibration and verification surfaces as a function of search radius and screening probability. The screening probability is expressed as $1 - \alpha$, where α is the null hypothesis test probability. The R_c^2 surface increases without limit as the search radius enlarges, although the rate of increase diminishes beyond about 500 km. In contrast, the R_v^2, RE, and CE surfaces all show general maxima in the form of a ridge along the 400–500-km search radius axis. Beyond that distance, the verification tests all do more poorly. The ridge of maximum verification is also not strongly dependent on screening probability.

hindcast or reconstruction skill in excess of climatology (i.e., $\overline{x}_c$); an RE < 0 indicates less skill than climatology.

4) Average coefficient of efficiency (CE) over the verification period is

$$\text{CE} = 1.0 - \left[\frac{\sum (x_i - \hat{x}_i)^2}{\sum (x_i - \overline{x}_v)^2}\right], \qquad \text{(B4)}$$

where x_i and $\hat{x}_i$ are the actual and estimated data in year i of the verification period and $\overline{x}_v$ is the mean of the actual data in the verification period. This statistic was first described in the hydrology literature as an expression of the true R^2 of a regression equation when it is applied to new data (Nash and Sutcliffe 1971). Like RE, CE has a theoretical range of $-\infty$ to $+1$, but here the benchmark for determining skill is the verification period mean. Thus, a CE > 0 indicates skill in excess of the verification period climatology (i.e., $\overline{x}_v$); a CE < 0 indicates less skill than verification period climatology.

Among the verification statistics used here, the CE is the most rigorous. The only difference between the RE and CE lies in the denominator term. However, this difference generally makes the CE more difficult to pass (i.e., CE > 0). When $\overline{x}_v = \overline{x}_c$, CE = RE. But when $\overline{x}_v \neq \overline{x}_c$, RE will be greater then the CE by a factor related to that difference. This follows by noting that for the CE, the sum of squares in the denominator is fully corrected because $\overline{x}_v$ is the proper mean. However for the RE, the denominator sum of squares will not be fully corrected unless the calibration period mean is fortuitously identical to the verification period mean. When this is not the case, the denominator sum of squares of the RE will be larger than that of the CE resulting in RE $>$ CE.

As described in section 2, the verification data available at each grid point varies in length depending on the lengths of the original station records used in gridding the PDSI. The median length of the pre-1928 data is 35 yr, with a minimum of 15, an interquartile range of 30–37, and a maximum of 90. The different lengths make verification across the network potentially inhomogeneous for comparative purposes. For this reason, we restricted the verification tests to data in the period 1893–1927. Thus, a maximum of 35 yr of data was used for verification at each grid point.

A total of 24 search radii ranging from 200 to 3000 km and five screening probability levels ranging from 0.05 to 0.40 were used, resulting in a the total of 120 PPR test runs. The results of these runs are succinctly summarized in four surface plots shown in Fig. B1. For plotting purposes, the screening probabilities (i.e., α levels) are expressed as $1 - \alpha$ probabilities. The calibration R_c^2 surface shows a nearly monotonic increase in explained variance with increasing distance. This effect occurs regardless of the screening probability used. However, the R_c^2 surface does show a clear break in slope at a distance of 400–500 km. At longer distances, the rate of increase in R_c^2 noticeably declines. Thus, most

of the explained variance comes from tree-ring chronologies located within about 500 km of the PDSI grid points on average. In contrast, the R_v^2, RE, and CE surfaces are radically different. Each of these verification statistics shows that PDSI reconstruction fidelity increases as the search radius increases from 200 to 400–500 km for each grid point on average. At greater distances there is a clear loss of fidelity. Interestingly, reconstruction fidelity rebounds somewhat at distances in excess of 2000 km, especially for the more stringent screening probabilities. This result probably reflects true synoptic-scale teleconnection patterns of drought and wetness in the continental U.S. Presumably, the stricter screening probabilities also winnowed out many spurious tree-ring chronologies, resulting in a noticeable increase in long-range verification performance. However, this rebound never regains the fidelity lost by increasing the search radius beyond ~500 km. Consequently, nothing is truly gained by searching for long-range tree-ring teleconnections with drought in the U.S., except excessively inflated R_c^2.

The verification surfaces in Fig. B1 indicate the presence of a ridge of maximum reconstruction fidelity that consistently falls in the 400–500-km search radius band. This ridge is only weakly dependent on screening probability. Thus ''search radius,'' which determines the size of the level-1 pool of candidate tree-ring predictors, is clearly the most important free variable to select properly in the PPR method. Based on the results here, we concluded that a search radius of 450 km was the best global search radius to use. As the second free variable in this test, ''screening probability'' had a much weaker effect on the PPR method. Presumably, the use of principal components analysis on the level-2 pool of screened tree-ring chronologies served to protect the subsequent regression results from the inclusion of weakly correlated or spurious tree-ring predictors. Even so, there is some indication of improved verification with increasing $1 - \alpha$ probability along the 400–500-km ridge. Therefore, we chose an α-level probability of 0.10 for screening the tree-ring chronologies entering into the level-2 pool. With these two free variables fixed, the PPR method can be applied automatically for reconstructing drought across the continental U.S.

With regard to our selected 450-km search radius, the median level-1 pool size was 48 tree-ring variables, which equates to 24 tree-ring chronologies for years t and $t + 1$. This initial pool of candidate variables was reduced to a median size of 18 after screening with $\alpha = 0.10$. After principal components analysis of the 18 retained variables, a median of six eigenvectors passed the Kaiser–Guttman test. Finally, the minimum AIC_c test entered a median of three principal components into the regression model. Thus, the PPR method reduced the median predictor variable space from 48 to 3 variables, a 93.7% reduction in the variable pool size, resulting in a large conservation of degrees of freedom in the final regression model. Since the expected value for R_c^2 increases with the number of variables in a regression model (Morrison 1990) and also increases with the number of candidate predictors available for regression (Rencher and Pun 1980), the sequential reduction of predictor variable pool size should also reduce the inflation of R_c^2.

REFERENCES

Akaike, H., 1974: A new look at the statistical model identification. *IEEE Trans. Autom. Control,* **AC-19,** 716–723.

Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality, and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.,* **115,** 1083–1126.

Bergman, K. H., C. F. Ropelewski, and M. S. Halpert, 1986: The record Southeast drought of 1986. *Weatherwise,* **39,** 262–266.

Blasing, T. J., and D. N. Duvick, 1984: Reconstruction of precipitation history in North American corn belt using tree rings. *Nature,* **307,** 143–145.

Box, G. E. P., and G. M. Jenkins, 1970: *Time Series Analysis: Forecasting and Control.* Holden-Day, 553 pp.

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

Briffa, K. R., P. D. Jones, T. M. L. Wigley, J. R. Pilcher, and M. G. L. Baillie, 1986: Climate reconstruction from tree rings: Part 2, Spatial reconstruction of summer mean sea-level pressure patterns over Great Britain. *J. Climatol.,* **6,** 1–15.

Broadbrooks, W. J., and P. B. Elmore, 1987: A Monte Carlo study of the sampling distribution of the congruence coefficient. *Educ. Psychol. Meas.,* **47,** 1–11.

Cole, J. E., and E. R. Cook, 1998: The changing relationship between ENSO variability and moisture balance in the continental United States. *Geophys. Res. Lett.,* **25,** 4529–4532.

Cook, E. R., and G. C. Jacoby Jr., 1977: Tree-ring–drought relationships in the Hudson Valley, New York. *Science,* **198,** 399–401.

——, M. A. Kablack, and G. C. Jacoby, 1988: The 1986 drought in the southeastern United States: How rare an event was it? *J. Geophys. Res.,* **93,** 14 257–14 260.

——, D. W. Stahle, and M. K. Cleaveland, 1992: Dendroclimatic evidence from eastern North America. *Climate Since AD 1500,* R. S. Bradley and P. D. Jones, Eds., Routledge, 331–348.

——, K. R. Briffa, and P. D. Jones, 1994: Spatial regression methods in dendroclimatology: A review and comparison of two techniques. *Int. J. Climatol.,* **14,** 379–402.

——, D. M. Meko, D. W. Stahle, and M. K. Cleaveland, 1996: Tree-ring reconstructions of past drought across the coterminous United States: Tests of a regression method and calibration/verification results. *Tree Rings, Environment, and Humanity,* J. S. Dean, D. M. Meko, and T. W. Swetnam, Eds., Radiocarbon, 155–169.

——, ——, and C. W. Stockton, 1997: A new assessment of possible solar and lunar forcing of the bidecadal drought rhythm in the western United States. *J. Climate,* **10,** 1343–1356.

——, J. E. Cole, R. D. D'Arrigo, D. W. Stahle, and R. Villalba, 1999: Tree ring records of past ENSO variability and forcing. *El Niño and the Southern Oscillation: Multiscale Variability and its Impacts on Natural Ecosystems and Society,* H. F. Diaz and V. Markgraf, Eds., Cambridge University Press, in press.

Cramer, J. S., 1987: Mean and variance of R^2 in small and moderate samples. *J. Econometrics,* **35,** 253–266.

Fritts, H. C., 1965: Tree-ring evidence for climatic changes in western North America. *Mon. Wea. Rev.,* **93,** 421–443.

——, 1976: *Tree Rings and Climate.* Academic Press, 567 pp.

——, 1991: *Reconstructing Large-Scale Climate Patterns from Tree-Ring Data.* The University of Arizona Press, 286 pp.

——, T. J. Blasing, B. P. Hayden, and J. E. Kutzbach, 1971: Multivariate techniques for specifying tree-growth and climate rela-

tionships and for reconstructing anomalies in paleoclimate. *J. Appl. Meteor.*, **10,** 845–864.
Gilman, D. L., F. J. Fuglister, and J. M. Mitchell Jr., 1963: On the power spectrum of ''red noise.'' *J. Atmos. Sci.*, **20,** 182–184.
Glahn, H. R., 1968: Canonical correlation and its relationship to discriminant analysis and multiple regression. *J. Atmos. Sci.*, **25,** 23–31.
Graumlich, L. J., 1993: A 1000-year record of temperature and precipitation in the Sierra Nevada. *Quat. Res.*, **39,** 249–255.
Guttman, L., 1954: Some necessary conditions for common-factor analysis. *Psychometrika*, **19,** 149–161.
Guttman, N., 1991: Sensitivity of the Palmer hydrologic drought index to temperature and precipitation departures from average conditions. *Water Res. Bull.*, **27,** 797–807.
Haston, L., and J. Michaelsen, 1994: Long-term central coastal California precipitation variability and relationships to El Niño–Southern Oscillation. *J. Climate*, **7,** 1373–1387.
——, and ——, 1997: Spatial and temporal variability of southern California precipitation over the last 400 yr and relationships to atmospheric circulation patterns. *J. Climate*, **10,** 1836–1852.
Haugh, L. D., and G. E. P. Box, 1977: Identification of dynamic regression (distributed lag) models connecting two time series. *J. Amer. Stat. Assoc.*, **72,** 121–130.
Helland, I. S., 1987: On the interpretation and use of R^2 in regression analysis. *Biometrics*, **43,** 61–69.
Hendrickson, A. E., and P. O. White, 1964: Promax: A quick method for rotation to oblique simple structure. *Br. J. Stat. Psychol.*, **17,** 65–70.
Hurvich, C. M., and C. Tsai, 1989: Regression and time series model selection in small samples. *Biometrika*, **76,** 297–307.
Kaiser, H. F., 1960: The application of electronic computers to factor analysis. *Educ. Psychol. Meas.*, **20,** 141–151.
Karl, T. R., and A. J. Koscielny, 1982: Drought in the United States. *J. Climatol.*, **2,** 313–329.
——, and W. E. Riebsame, 1984: The identification of 10- to 20-year temperature and precipitation fluctuations in the contiguous United States. *J. Climate Appl. Meteor.*, **23,** 950–966.
——, C. N. Williams Jr., and F. T. Quinlan, 1990: United States Historical Climatology Network (HCN) serial temperature and precipitation data. Environmental Sciences Division Publ. 3404, 371 pp. [Available from Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN 37831.]
Kutzbach, J. E., and P. J. Guetter, 1980: On the design of paleoenvironmental data networks for estimating large-scale patterns of climate. *Quat. Res.*, **14,** 169–187.
LaMarche, V. C., Jr., and H. C. Fritts, 1971: Anomaly patterns of climate over the western United States, 1700–1930, derived from principal component analysis of tree-ring data. *Mon. Wea. Rev.*, **99,** 138–142.
Langbein, W. B., and J. R. Slack, 1982: Yearly variations in runoff and frequency of dry years for the conterminous United States, 1911–79. U.S. Geological Survey Open-File Rep. 82-751, 85 pp. [Available from Branch of Distribution, USGS, Box 25425, Denver Federal Center, Denver, CO 80225.]
Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Statistical Forecasting Scientific Rep. 1, Department of Meteorology, Massachusetts Institute of Technology, Cambridge, MA, 57 pp.
Mann, M. E., R. S. Bradley, and M. K. Hughes, 1998: Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, **392,** 779–787.
Matthai, H. F., 1979: Hydrologic and human aspects of the 1976–77 drought. USGS Prof. Paper 1130, U.S. Government Printing Office, 84 pp. [Available from U.S. Government Printing Office, Washington, DC 20402.]
Meko, D. M., 1981: Applications of Box-Jenkins methods of time series analysis to the reconstruction of drought from tree rings. Unpublished Ph.D. dissertation, The University of Arizona, 149 pp.
——, 1992: Dendroclimatic evidence from the Great Plains of the United States. *Climate Since A.D. 1500*, R. S. Bradley and P. D. Jones, Eds., Routledge, 312–330.
——, E. R. Cook, D. W. Stahle, C. W. Stockton, and M. K. Hughes, 1993: Spatial patterns of tree-growth anomalies in the United States and southeastern Canada. *J. Climate*, **6,** 1773–1786.
Mitchell, J. M., Jr., C. W. Stockton, and D. M. Meko, 1979: Evidence of a 22-year rhythm of drought in the western United States related to the Hale solar cycle since the 17th century. *Solar–Terrestrial Influences on Weather and Climate*, B. M. McCormac and T. A. Seliga, Eds., D. Reidel, 125–144.
Morrison, D. F., 1990: *Multivariate Statistical Methods.* 3d ed. McGraw-Hill, 480 pp.
Namias, J., 1955: Some meteorological aspects of drought with special reference to the summers of 1952–54 over the United States. *Mon. Wea. Rev.*, **83,** 199–205.
——, 1966: Nature and possible causes of the northeastern United States drought during 1962–65. *Mon. Wea. Rev.*, **94,** 543–554.
——, 1978: Multiple causes of the North American abnormal winter 1976–77. *Mon. Wea. Rev.*, **106,** 279–295.
Nash, J. E., and J. V. Sutcliffe, 1971: Riverflow forecasting through conceptual models 1, A discussion of principles. *J. Hydrol.*, **10,** 282–290.
Palmer, W. C., 1965: Meteorological drought. Weather Bureau Res. Paper 45, U.S. Department of Commerce, Washington, DC, 58 pp.
Rencher, A. C., and F. C. Pun, 1980: Inflation of R^2 in best subset regression. *Technometrics*, **22,** 49–53.
Reyment, R., and K. G. Jöreskog, 1993: *Applied Factor Analysis in the Natural Sciences.* Cambridge University Press, 371 pp.
Richman, M. B., 1986: Rotation of principal components. *J. Climatol.*, **6,** 293–335.
Roos, M., 1994: Is the California drought over. Proceedings of the Tenth Annual Pacific Climate (PACLIM) Workshop, Interagency Ecological Studies Program for the Sacramento-San Joaquin Estuary, Tech. Rep. 36, 123–128.
Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114,** 2352–2362.
Stahle, D. W., and M. K. Cleaveland, 1988: Texas drought history reconstructed and analyzed from 1698–1980. *J. Climate*, **1,** 59–74.
——, ——, and J. G. Hehr, 1985: A 450-year drought reconstruction for Arkansas, United States. *Nature*, **316,** 530–532.
——, ——, and ——, 1988: North Carolina climate changes reconstructed from tree rings: A.D. 372–1985. *Science*, **240,** 1517–1519.
Stockton, C. W., and D. M. Meko, 1975: A long-term history of drought occurrence in western United States inferred from tree rings. *Weatherwise*, **28,** 244–249.
——, and ——, 1983: Drought recurrence in the Great Plains as reconstructed from long-term tree-ring records. *J. Climate Appl. Meteor.*, **22,** 17–29.
Warrick, R. A., 1980: Drought in the Great Plains: A case study of research on climate and society in the U.S.A. *Climatic Constraints and Human Activities*, J. Ausubel and A. K. Biswas, Eds., Pergamon, 93–124.
Woodhouse, C., and D. Meko, 1997: Number of winter precipitation days reconstructed from southwestern tree rings. *J. Climate*, **10,** 2663–2669.