U.S. Daily Temperatures: The Meaning of Extremes in the Context of Nonnormality

P. HUYBERS, K. A. MCKINNON, AND A. RHINES

Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts

M. TINGLEY

Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts, and Departments of Meteorology and Statistics, The Pennsylvania State University, University Park, Pennsylvania

ABSTRACT

Variations in extreme daily temperatures are explored in relation to changes in seasonal mean temperature using 1218 high-quality U.S. temperature stations spanning 1900–2012. Extreme temperatures are amplified (or damped) by as much as ±50% relative to changes in average temperature, depending on region, season, and whether daily minimum or maximum temperature is analyzed. The majority of this regional structure in amplification is shown to follow from regional variations in temperature distributions. More specifically, there exists a close relationship between departures from normality and the degree to which extreme changes are amplified relative to the mean. To distinguish between intraseasonal and interannual contributions to non-normality and amplification, an additional procedure, referred to as z bootstrapping, is introduced that controls for changes in the mean and variance between years. Application of z bootstrapping indicates that amplification of winter extreme variations is generally consistent with nonnormal intraseasonal variability. Summer variability, in contrast, shows interannual variations in the spread of the temperature distribution related to changes in the mean, especially in the Midwest. Changes in midwestern temperature variability are qualitatively consistent with those expected from decreases in evapotranspiration and are strongly correlated with a measure of drought intensity. The identified patterns of interannual variations in means and extremes may serve as an analog for modes of variability that can be expected at longer time scales.

1. Introduction

There is substantial uncertainty regarding how changes in mean and extreme temperature are related (Alexander and Perkins 2013; Katz et al. 2013). Part of this uncertainty stems from difficulties in disentangling changes in the mean from higher-order moments of the temperature distribution. Hansen et al. (2012), for example, suggested an increase in the variance of summer monthly temperatures, but changes in variance are not discernible after accounting for trends in the mean, among other factors (Rhines and Huybers 2013; Huntingford et al. 2013). Similarly, Donat and Alexander (2012) suggested that the distribution of temperature has become more skewed in recent decades, but it may be that the skew emerges from aggregating across distributions with differing mean values (Karl and Katz 2012; Katz et al. 2013). As a final example, it was suggested that exceptionally warm summers in western Europe in 2003 and Russia in 2010 resulted, in part, from an increase in temperature variability (Schär et al. 2004; Barriopedro et al. 2011), but further analyses were unable to statistically differentiate such events from the consequences of mean warming (Rahmstorf and Coumou 2011; Otto et al. 2012; Tingley and Huybers 2013).

The foregoing examples from the literature illustrate that a shifting mean can lead to the appearance of changes in variance, skew, or other features of the distribution. One method of isolating the effects of changes in the mean is to specifically examine features of the distribution as a function of the mean. Robeson (2002), for example, provides an analysis of the variance of U.S. daily temperatures as

a function of monthly mean values between 1900 and 2000. Such analyses of conditional variance have been developed in detail in econometric applications (Engle 2001).

The functional dependence of variance conditional on the mean would afford a complete description of the distribution if it were normal. But departures from normality are common (Karl 1985; Ruff and Neelin 2012), as can generally be anticipated from the influence of nonlinear phenomena such as those associated with changes in soil moisture, surface albedo, atmospheric stability, and advection (e.g., Seneviratne et al. 2006; Meehl and Tebaldi 2004; Teng et al. 2013). The likely presence of nonnormality thus makes it useful to explore other measures of how the temperature distribution depends on the mean, and we focus on the 5th and 95th percentiles. Allowing for nonnormality also prompts a need to account for inherent dependence between the mean and extremes of a sample. In the following, we analyze interannual variations in mean and extreme U.S. daily temperatures, demonstrate that systematic changes in the relationship between these quantities can result from either nonnormality or changes in the underlying distribution between years, and then introduce a methodology to distinguish between some subsets of nonnormal and nonstationary behaviors.

2. Data

The distributions of minimum (Tn) and maximum (Tx) daily temperature are examined across 1218 stations from version 3.11 of the U.S. Global Historical Climate Network (GHCN) between 1900 and 2012. These stations were selected on the basis of being high quality, well resolved, and well distributed (Menne et al. 2012).

GHCN data from U.S. stations are typically recorded in rounded units of degrees Fahrenheit that are then converted to degrees Celsius and again rounded to the nearest tenth of a degree. One curious result of this double rounding is that no digit should then end in a value of 0.5. Rounding biases certain percentile statistics (Machado and Santos Silva 2005; Zhang et al. 2009), but in section 3 we demonstrate a small sensitivity of our particular analysis when rounding to either 1°F or 0.1°C.

A more substantial issue is the presence of seasonality because it is associated with changes not only in the mean but also in higher-order moments of the temperature distribution (e.g., Huybers and Curry 2006). Seasonality is dealt with in two ways. First, seasonality is removed from Tx and Tn at each station by removing the climatological average seasonal cycle, similar to previous studies (Brown et al. 2008). Climatological seasonal cycles are estimated by taking the average Tx or Tn as

a function of day across the available data points between 1950 and 2000, where this interval is selected as a tradeoff between duration and completeness of the data. Gaps in the data are not filled because they tend to be clustered into long sequences, making such infilling uncertain and impractical. We then low-pass filter the empirically estimated climatological seasonal cycles using a second-order Butterworth filter with a cutoff frequency of once per two weeks.

In a small number of cases individual data points exceed 6 times the sample standard deviation associated with the smoothed climatological seasonal cycle. These outliers are removed and the seasonal climatology reestimated. On average, each record has one such outlier among 35 000 data points, but outliers are clumped such that 10% of the stations account for 70% of all outliers. Repeating the analysis without excluding data yields quantitatively very similar results (see section 4). Robustness to outliers comes, in part, from using 5th and 95th percentiles, as opposed to seasonal minima and maxima, and because there is rarely more than one outlier in a given season. All years are treated as having 365 days. Hereafter Tx and Tn are used to refer to their respective anomalies from the climatological seasonal cycle.

The second control for seasonality is to perform analyses with respect to four 3-month seasons. The year associated with January is assigned to the December–February (DJF) winter season. In section 4, we also describe how monthly resolution analysis affects our results. Although beyond the scope of this study, we note that the relationship between means and extremes likely evolves continuously over the course of the seasonal cycle and that there is substantial temporal (e.g., Stine and Huybers 2012) and spatial (e.g., McKinnon et al. 2013) variability in seasonality, ultimately warranting a more generalized examination of the seasonality of mean and extreme temperature. Seasonal means and quantiles are computed when temperatures on 80% of the days in a given season are recorded. By this criterion, just over 80% of the years between 1900 and 2012 across stations are included for both Tn and Tx for each season. Trials using 70% and 95% seasonal coverage thresholds give quantitatively very similar results, as described in section 4. This follows from the fact that data for most seasons are either complete (~65%) or entirely missing (~15%). Both Tx and Tn have roughly equal data availability across seasons.

3. A mean–extreme metric

In what follows, we make use of the fact that the mean of a sample from a normal distribution is expected to have no covariance with any quantile of the sample, after
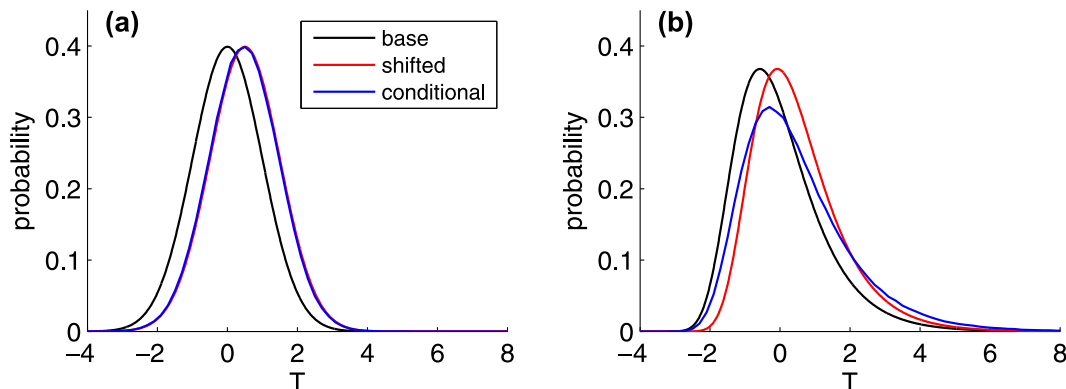
FIG. 1. Illustration of how conditioning a sample upon the mean has different implications for normal and non-normal distributions. (a) A standard normal distribution (black), the original distribution shifted positive by 0.5 (red), and the distribution of samples drawn from the standard normal conditional upon their having a mean of 0.5 (blue). As required by Basu's theorem, the conditional and shifted distributions are equivalent, up to sampling issues associated with our estimation approach. (b) The shifted and conditional distributions differ for a standard Gumbel distribution. In this case, although each has a mean of one-half, the conditional distribution has greater variance. Conditional distributions are approximated by realizing sets with 90 samples and retaining those whose sample mean is within 10% of 0.5.

subtracting the sample mean. That is, $E[\text{COV}(\overline{x}, \hat{x} - \overline{x})] = 0$, where $\overline{x}$ indicates the mean of a sample, $\hat{x}$ is a quantile, COV is the covariance, and the brackets indicate the expectation. This relation follows from a more general theorem regarding the independence of ancillary statistics relative to complete and sufficient statistics (Basu 1958).

A related consequence of Basu's theorem is that the distribution of samples from a normal distribution that is conditioned on a mean value will also be normal and have the same variance as the original distribution. For purposes of illustration, Fig. 1a shows a standard normal distribution and the distribution of realizations conditioned on a specified mean. The conditional distribution is estimated in a simple manner through drawing sets of 90 independent realizations from a standard normal and only accepting those with a sample mean between 0.45 and 0.55, as an approximation to conditioning on a sample mean of 0.5. Sets are generated until 10 000 samples meet the criteria for inclusion and all members of these sets are binned together in estimating the conditional distribution. In this case, the conditional distribution is consistent within sampling uncertainty of having increased the original standard normal distribution's mean by 0.5.

For nonnormal distributions the sample mean is generally not a complete and sufficient statistic, and samples conditioned on a mean will have higher-order moments that differ from the original distribution. It follows that samples from nonnormal distributions will generally have extremes that are either amplified or damped relative to variations in the mean. For instance, samples from a standard Gumbel distribution that are conditioned on

a mean of 0.5 have 50% greater variance than the original zero-mean distribution (Fig. 1b). Illustrative examples are again generated through realizing 10 000 90-member sets having a mean within 10% of 0.5.

A measure of the amplification or damping of variations in extreme temperature relative to the mean can be computed as

$$S = \text{COV}(\overline{x}, \hat{x} - \overline{x})/\text{VAR}(\overline{x}), \qquad (1)$$

where the sample covariance between the mean and a quantile minus the mean is divided by the sample variance of the mean. This statistic results in a unitless quantity that is a least squares best estimate of the slope between the mean and quantile minus the mean. The quantiles focused on here are the 5th percentile (T05) and the 95th percentile (T95). The measure given by Eq. (1) will be referred to as a mean–extreme slope or, more specifically, the mean–Tx05 slope, mean–Tn95 slope, and so on.

The expected value of the mean–extreme slope is zero if values are drawn from a distribution that is normal and stationary (i.e., the moments of the underlying distribution do not change between years). Departures from a mean–extreme slope of zero indicate nonnormality, changes in the underlying distribution, or both. The converse does not necessarily hold, however, in that certain forms of changes in the underlying distribution do not influence the slope, as will be illustrated. Mean–extreme slopes that are greater than zero indicate variations in extremes that are amplified relative to changes in the mean, whereas negative slopes indicate damping. Note that positive mean–T05 slopes imply shortening of

the lower tail of the distribution, whereas positive mean–T95 slopes imply lengthening of the upper tail.

Percentiles are estimated by constructing the empirical cumulative distribution function, also known as a Kaplan–Meier estimate (Kaplan and Meier 1958), and linearly interpolating for a given percentile. Note that a potentially useful feature of the mean–extreme metric is that it does not require assumptions regarding the parametric form of the underlying distribution. Quantile regression (Koenker 2005) was also explored for purposes of estimating mean–extreme slopes and gave similar results, but a least squares fit to the estimated percentiles is employed here because of its relationship with Basu's theorem.

Six examples of mean–extreme slopes are provided to illustrate the effects of departures from normality and stationarity. A zero mean–extreme slope is expected for the first three examples and a nonzero slope for the last three. Each example has 100 sets of realizations, nominally representing years, and each set comprises 90 samples, nominally representing daily temperatures. The empirical distribution associated with all 9000 samples and the mean–T95 slope are analyzed in each case. Confidence intervals are placed on the sample mean–extreme slope by resampling the realizations of T95 minus the mean with replacement and recalculating slopes 10 000 times. Confidence intervals are two sided and at the 95% level. Analogous results hold for T05 and other percentiles but are not shown.

Example 1: Samples are drawn from a standard normal distribution (Fig. 2, panels 1a and 1b), and the corresponding mean–T95 slope is indistinguishable from zero at the 95% confidence level (Fig. 2, panel 1c), as expected. The y intercept of the mean–extreme regression line indicates the average offset between the mean and 95th percentile of the empirical distribution and equals 1.64 ± 0.02 (1 standard deviation), a value consistent with that expected for the 95th percentile of a standard normal.

Example 2: Samples are again considered from a normal distribution but whose mean value increases linearly with the integer value of the year beginning at year 50 (Fig. 2, panel 2a). The empirical distribution, when considered across all realizations, is mixed normal with a positive skew (Fig. 2, panel 2b) (cf. Karl and Katz 2012). Despite changes in the mean and the empirical distribution appearing nonnormal, the difference between T95 and the mean is invariant, and the estimated mean–T95 slope is consistent with zero, again as expected (Fig. 2, panel 2c).

Example 3: Considering samples from another normal distribution, but with standard deviation increasing linearly with integer values of the year (Fig. 2, panel 3a), leads to an empirical distribution similar to that of Student's t distribution (Fig. 2, panel 3b). As in example 2, the expected mean–max slope again remains zero (Fig. 2, panel 3c), but the slope is more uncertain because the spread of the mean is relatively smaller and the spread of T95 larger.

Example 4: Both the mean and standard deviation of a normal distribution are now made to increase linearly with the integer value of the year (Fig. 3, panel 4a). The empirical distribution considered across all realizations is mixed normal with positive skew (Fig. 3, panel 4b). In this case, a positive mean–T95 slope is expected because of the imposed covariance between the mean and standard deviation, where the latter controls the expected spread between T95 and the mean (Fig. 3, panel 4c).

Example 5: A generalized extreme value distribution is now considered having a scale parameter of 1, shape -0.1, and location selected to give a zero mean. The sign of the distribution is also reversed in order to give negative skew. T95 variations are expected to be damped relative to changes in the mean because of the negative skew, among other higher-order moments, and a negative mean–T95 slope is found (Fig. 3, panels 5a–c). The empirical distribution is similar to that from example 2 in terms of having nonzero skew, but the nonnormal intraseasonal variability in this example leads to the expectation of a negative slope.

Example 6: Finally, given samples from a Student's t distribution, the spread of the distribution leads to extreme values having amplified variability relative to the mean and to a positive mean–T95 slope (Fig. 3, panels 6a–c). (More technically, although T95 $- \overline{T}$ is ancillary to $\overline{T}$, the sample mean is not complete sufficient for Student's t distribution.) Note that the empirical distribution is difficult to distinguish from that in example 3 but that the slopes differ, illustrating how the empirical distribution can have an ambiguous relationship with the mean–extreme slope.

The foregoing examples illustrate a diverse set of relationships between means and 95th percentiles across normal (examples 1–4) and nonnormal (examples 5 and 6) processes as well as between stationary (examples 1, 5, and 6) and nonstationary (examples 2–4) processes. The mean–extreme slope provides a simple description of how the tails of a distribution change with respect to the mean that we apply and interpret in the next section. The fact that the mean–extreme slope can take on similar values despite describing different processes (e.g., examples 3 and 6) is also addressed in section 6 where we introduce
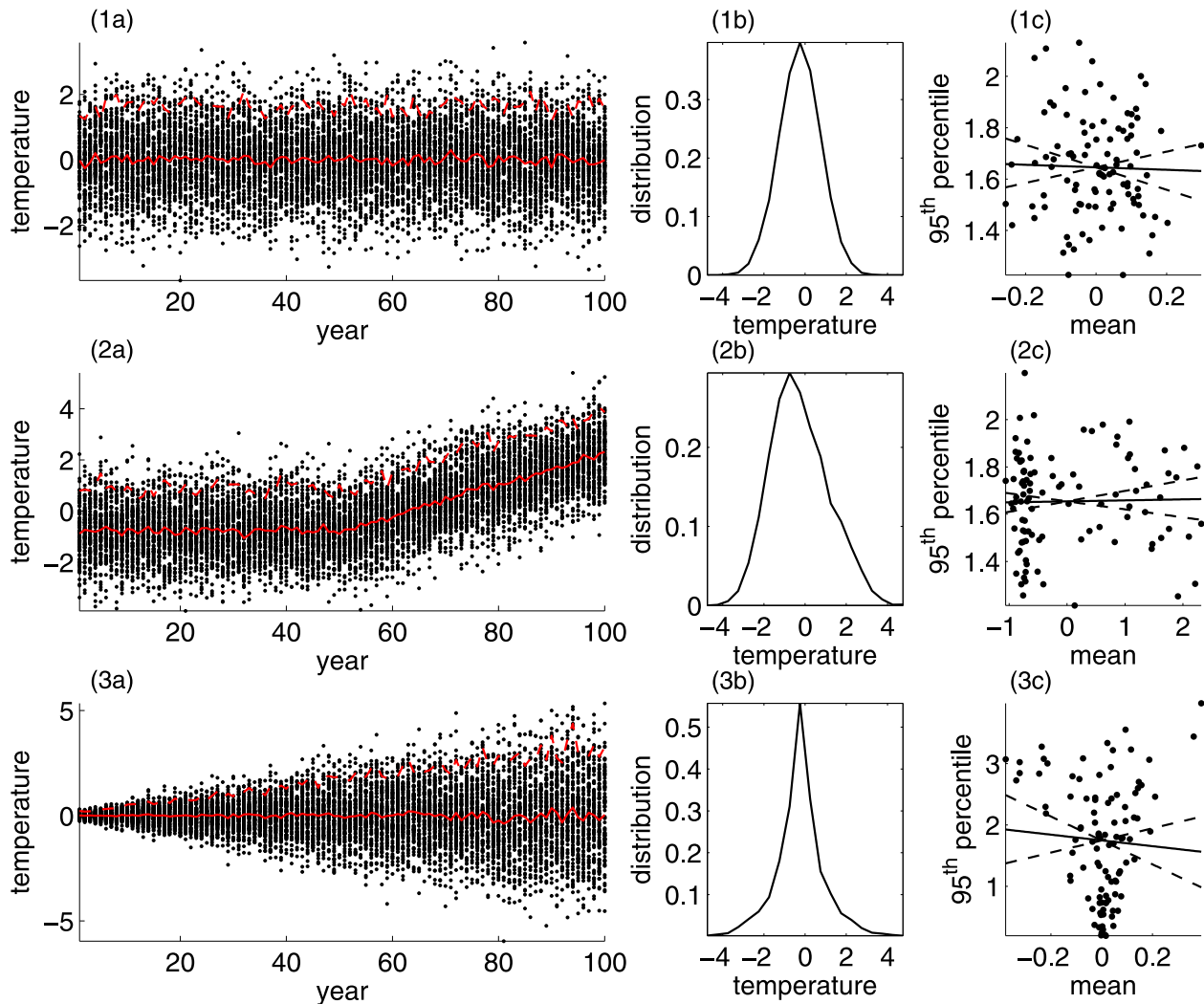
FIG. 2. Illustrative examples of the relationship between sample means and 95th percentiles drawn from normal distributions. (1a)–(3a) Each example has sets of 90 values (nominally labeled temperature) that are drawn 100 times (labeled years) and the mean (solid red) and 95th percentile of each set (dashed red) are indicated. (1b)–(3b) The empirical distributions from the 90×100 samples, and (1c)–(3c) the 95th percentile minus the mean regressed against the mean (black line) along with the 95% confidence interval for the slope (dashed black lines). Distributions correspond to examples in the text: (1a–c) standard normal, (2a–c) standard normal with a mean trend in the latter half, and (3a–c) standard normal with increasing variance.

a bootstrap approach that distinguishes between some subsets of nonstationary and nonnormal behaviors.

To explore the extent to which rounding may influence our results, we conduct paired synthetic experiments using each of the foregoing six examples, where one member of the pair uses full machine precision in estimating mean–extreme slopes and the other has data rounded first to units of degrees Fahrenheit and then, after unit conversion, to tenths of a degree Celsius (see section 2). Cross correlations between paired mean–extreme slopes, calculated from 1000 realizations, always exceed 0.97 and average 0.99 across the six examples. These high correlations indicate that noise contributions from rounding are small.

Regression between the twin realizations of slopes also results in a relationship that is within 5% of unity and passes within ± 0.02 of zero at the y intercept, indicating no appreciable bias in the results.

4. U.S. mean–extreme slopes

Applying the mean–extreme analysis to each of the 1218 stations results in maps of slopes across the United States that show coherent spatial structures depending on season and variable (Figs. 4 and 5). Counting across all seasons and variables, 46% of mean–extreme slopes differ from zero at the 95% confidence level using the
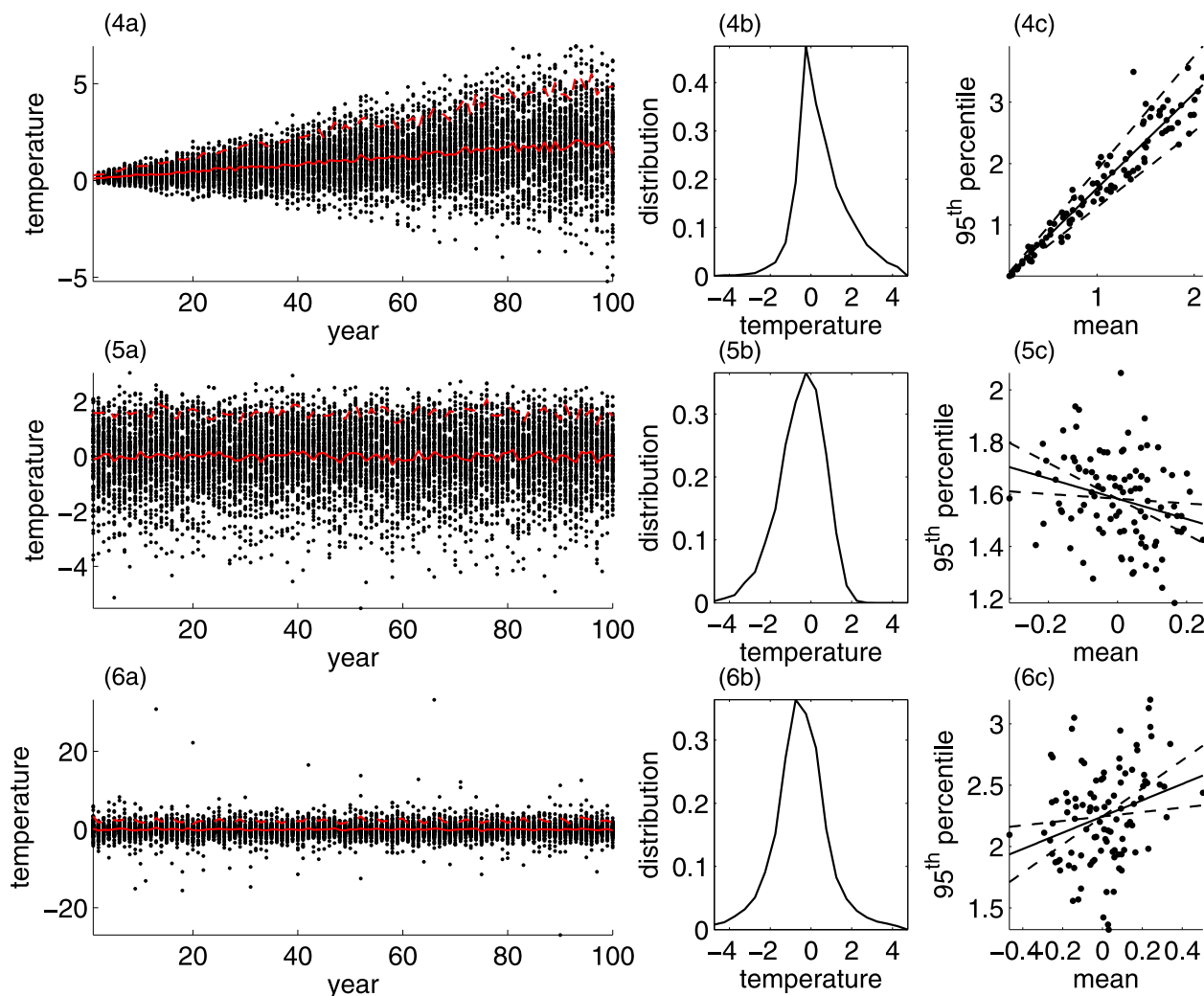
FIG. 3. As in Fig. 2, but for examples (4a–c) using a normal distribution with increasing mean and standard deviation, (5a–c) using a generalized extreme value distribution, and (6a–c) using a Student's *t* distribution.

same two-sided test relied upon in the foregoing examples. See Table 1 and Figs. 4 and 5 for summary statistics and maps of mean–extreme slopes.

Notable features of the mean–extreme slopes include a band of winter Tn slopes that are positive for Tn05 and negative for Tn95 with magnitudes near 0.5°C/°C arcing from the northwest to the northeast and extending southward to Texas. Wettstein and Mearns (2002) examined the influence of the northern annular mode (NAM) on the northeastern segment of this jetlike structure and found that low NAM indices correspond to warmer average Tn and smaller variance, consistent with our findings of a decrease in the spread of Tn with warmer temperatures. Similar results and correspondence with the results of Wettstein and Mearns (2002) hold for spring. Furthermore, our results indicate that these regional variations connect into larger domainwide

patterns, as might be expected from the NAM being a hemisphere-scale descriptor of atmospheric circulation.

During spring through fall, stations adjacent to the Pacific show slopes that are positive for Tx95 and negative for Tx05 with magnitudes of 0.5°C/°C. The proximity of these stations to the ocean suggests sea breeze dynamics that generally suppress values of Tx but that periodically give way to advection of high temperature from more interior continental conditions (e.g., Hughes and Hall 2010). A third feature is that summer Tx95 slopes exhibit negative values in the west and positive values in the east with magnitudes of about 0.25°C/°C. This pattern is similar to the May–June Tx warming trends described by Portmann et al. (2009), who evaluate this east–west dichotomy in the context of differences in precipitation. We return to the influence of precipitation on these variations in more detail later.
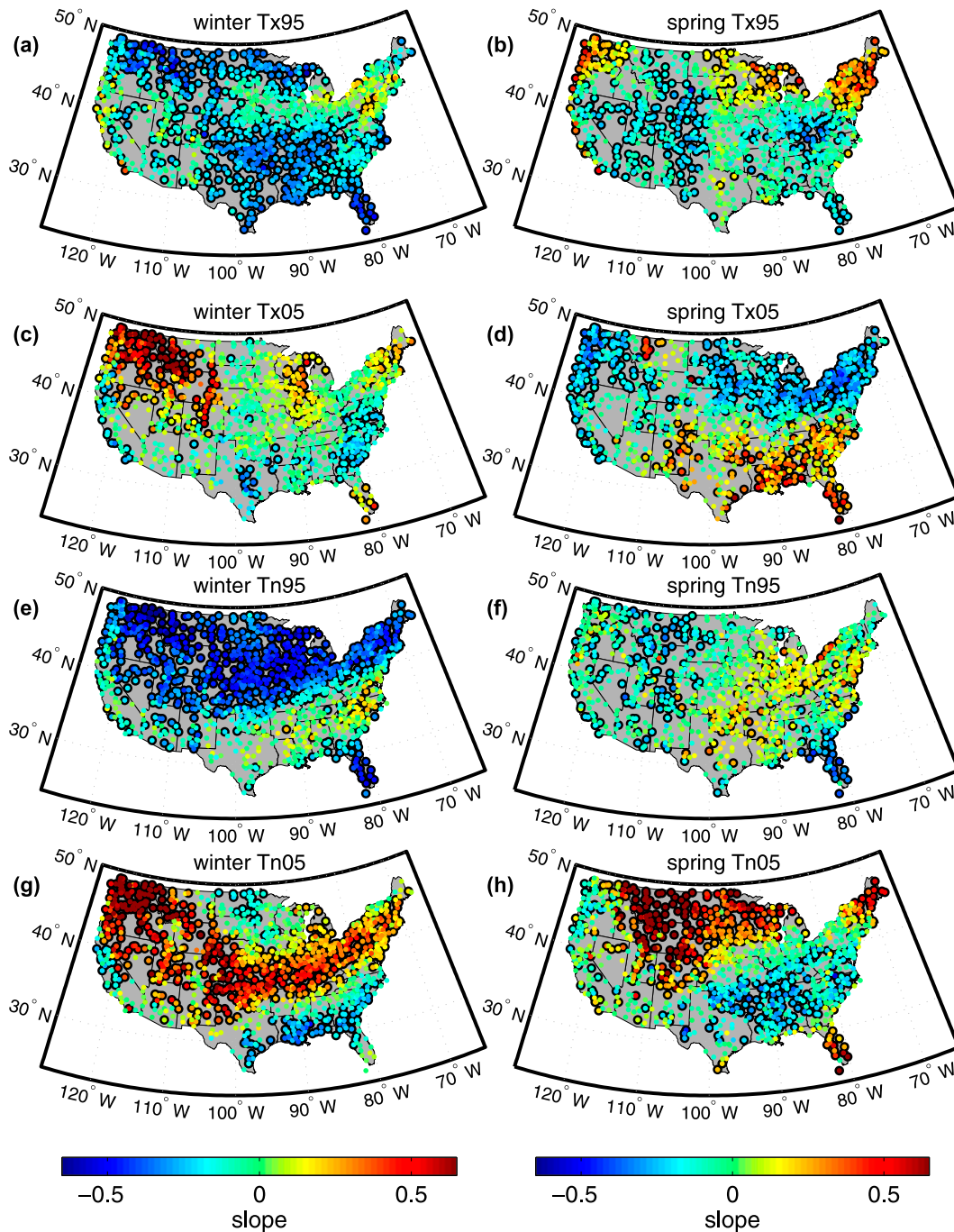
FIG. 4. Mean–extreme slopes for winter and spring. Means are computed as the seasonal average for a given year and slopes are computed with respect to the 5th or 95th percentiles computed from the corresponding seasonal data minus the sample mean. Points outlined in black have a slope differing significantly from zero at the 95% level.

As noted in section 2, 80% seasonal coverage was required for inclusion of a given year in estimating a mean–extreme slope. If instead this criterion is loosened to 70% or tightened to 95% coverage, the cross-correlations between the mean–extreme slopes reported here (Figs. 4 and 5) and the alternate values

respectively average 0.997 and 0.992. The lowest correlation across the 16 season–variable pairs and both threshold criteria is 0.98. Similarly, the cross correlations between the results reported here and mean–extreme slopes computed without removing outliers from the data are uniformly high, averaging
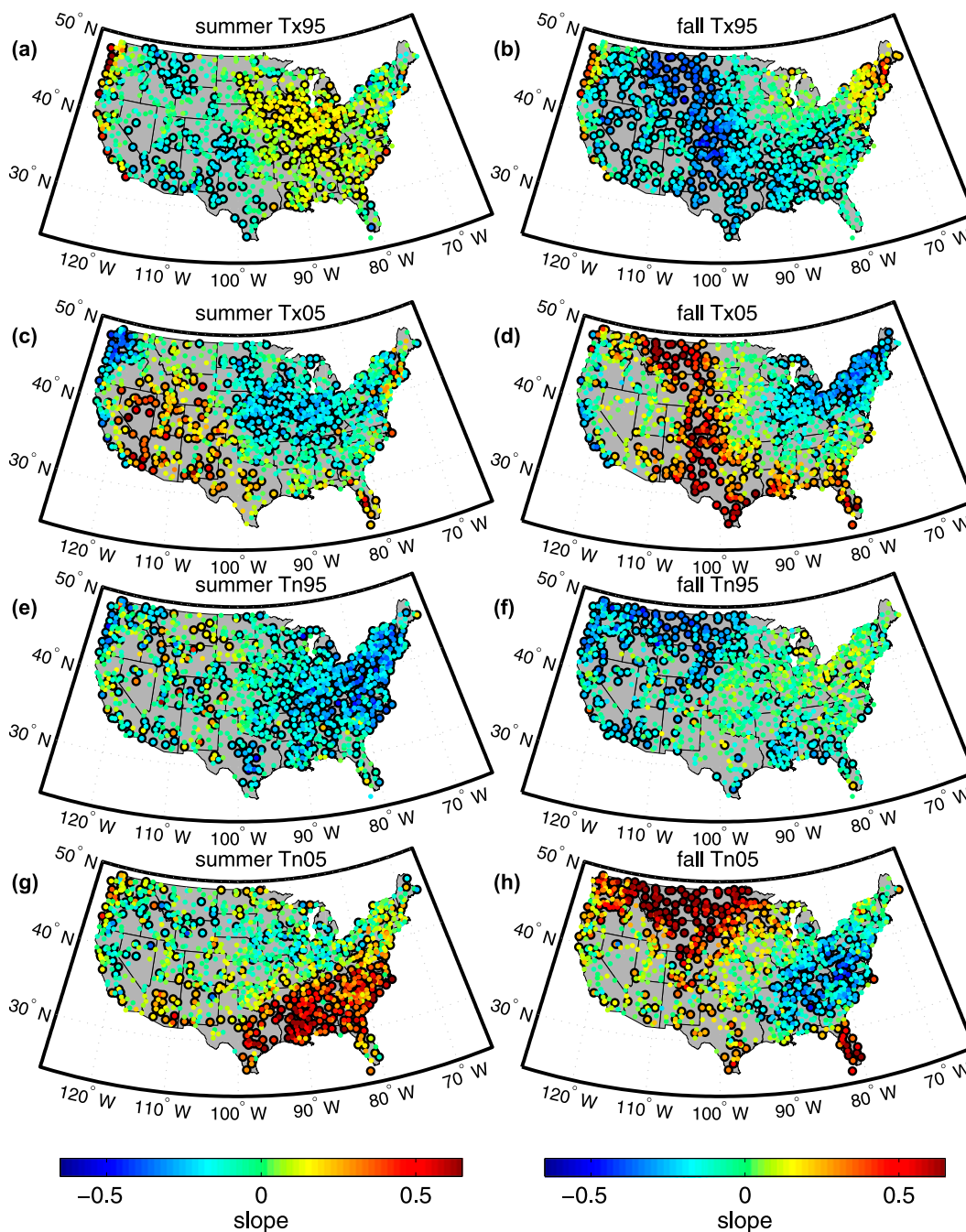
FIG. 5. As in Fig. 4, but for summer and fall.

0.991 across each of the 16 season–variable pairs, with the lowest correlation being 0.96. Regression between the variable pairs, either changing the coverage or outlier treatments, also results in slopes that are within 2% of unity in all cases and *y* intercepts within 0 ± 0.005. These additional tests demonstrate that our results are not sensitive to reasonable changes in how the data is processed.

A general interpretation of how changes in the lower and upper tail relate to each other across the United States can be obtained from plotting the mean–T95 slopes against the mean–T05 slopes for each season and variable (Fig. 6). Four quadrants can be defined with respect to mean–T95 and mean–T05 slopes: quadrant 1 (+T05, +T95) indicates that a higher mean is associated with shortening of the lower tail and lengthening of the upper tail, consistent with

TABLE 1. Percentage of mean–extreme slopes that significantly differ from zero (first set of columns), from the full bootstrapping results (second set), and from the "z bootstrapping" results (last set). Significance is assessed at the 95% confidence level using two-sided tests.

	Relative to zero				Full bootstrapping				z bootstrapping			
	Tx95	Tx05	Tn95	Tn05	Tx95	Tx05	Tn95	Tn05	Tx95	Tx05	Tn95	Tn05
Winter	59	40	39	53	26	18	16	15	36	24	13	35
Spring	30	53	42	43	14	17	11	14	32	20	21	20
Summer	70	25	54	29	16	26	21	26	61	37	40	28
Fall	63	46	40	45	18	13	12	15	42	19	15	26

increasing skew; quadrant 2 (−T05, +T95) indicates that a higher mean is associated with an increase in the spread of the distribution, consistent with greater variance; quadrant 3 (−T05, −T95) indicates decreased skew; and quadrant 4 (+T05, −T95) indicates decreased variance. Example distributional changes are also shown in Fig. 6 using a generalized extreme value distribution that is fit to the average summer Tx distribution and is characterized by a negative skew. More thorough analysis of extreme temperatures involving the use of generalized extreme value distributions has been presented elsewhere (e.g., Brown et al. 2008; Zwiers et al. 2011), but here this parametric fit is used only for the purposes of illustration. Whether generalized extreme distribution would accurately capture the mean–extreme covariance in the data is unclear, and we instead rely on nonparametric estimates.

Each of the eight T05–T95 trend comparisons has a negative correlation, indicating that the primary axis of U.S. interannual variability is oriented along decreasing and increasing spread with changes in the mean. For example, the Midwest tends to show greater spread in Tx during warm summers, whereas the Southwest shows less spread. There is also a greater representation in the third than first quadrant, particularly for Tx during the spring and fall in the northeastern and northwestern United States, indicating that anomalously warm seasonal averages are generally associated with a shift toward more negative skew.

Results can be compared to those of Robeson (2002), who analyzed the relationship between monthly means and standard deviations of U.S. Tn and Tx. Regions showing a trend toward lower variance with increasing mean generally correspond to those identified by Robeson (2002), including for winter Tn in the northern United States, summer Tx in the West, and fall Tx in the Great Plains. With respect to variance increasing with the mean, however, there is a mismatch in the Midwest. We find midwestern summer Tx95 to be amplified by about 25% relative to the mean and Tx05 to be damped by about 25%, whereas Robeson (2002) finds no appreciable increase in standard deviation. (See Fig. 5 for which stations have significant amplification at the 95% confidence level.) If, however, the analysis procedure of

Robeson (2002) is applied to the entire summer season—and not only monthly intervals—increasing midwestern variance with the mean is found. Conversely, applying our mean–Tx analysis to monthly intervals shows essentially no increase in variance. This suggests that the processes governing amplification of midwestern temperature extremes have time scales that exceed a month. This issue of time scale separation is addressed in more quantitative detail in section 6.

5. Consistency between slopes and empirical distributions

A block bootstrap method is used to quantify the degree to which U.S. mean–extreme slopes can be directly inferred from the empirical distribution. As a specific example, values of Tx are sampled with replacement across all summer data in order to construct new realizations of temperature. Blocks of 15 contiguous days are sampled so as to preserve synoptic-scale autocorrelation. More specifically, 15 days exceed estimates of decorrelation times for daily U.S. temperatures, which generally grade from about 5 days in the east to 10 days in the west (Király et al. 2006).

Different block realizations are selected from across years and, therefore, give results consistent with the assumption that the empirical distribution reflects a stationary process. Summer means and extremes are calculated from the resampled observations for each year, and the mean–extreme slope is computed using a number of realizations equal to the numbers of years in the original station data. This process is repeated 1000 times to construct a distribution of mean–extreme slopes for summer Tx and is likewise applied to all other season–variable pairs. Significant divergence of the observed slope from the bootstrapped slopes is indicative of nonstationarity.

A representative slope for each station is estimated by taking the average of the corresponding bootstrapped slopes. The squared cross correlation between the mean bootstrapped and observed slopes averaged across each season–variable pair is 0.71, indicating close correspondence between the empirical distribution and the observed
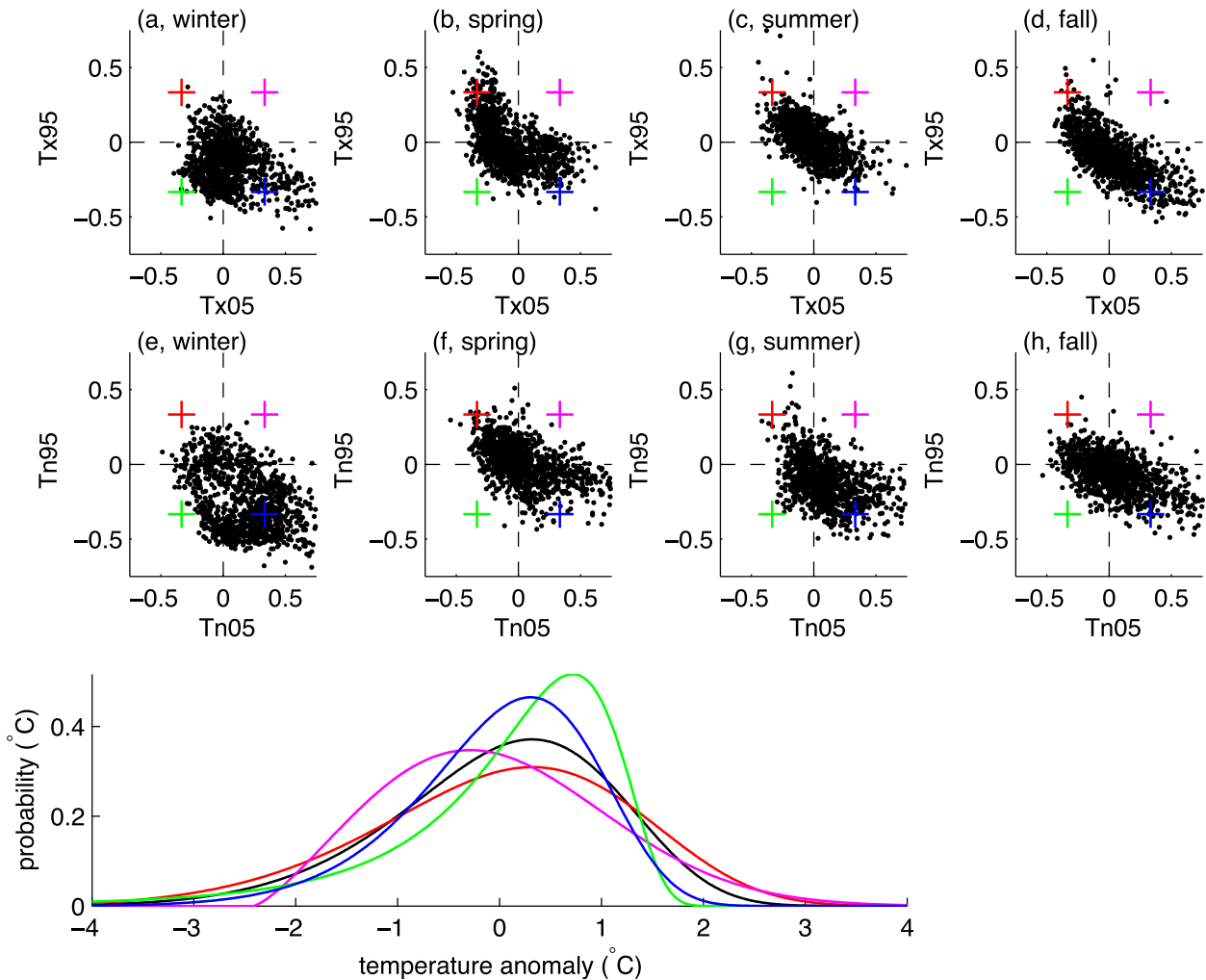
FIG. 6. Mean–T05 slopes vs mean–T95 slopes for each season and for (a)–(d) Tx and (e)–(h) Tn. Dashed lines demark quadrants that, moving from top right counterclockwise, indicate that mean warming is associated with more positive skew, greater variance, more negative skew, and less variance. U.S. stations are generally aligned along the axis of more or less variance. At bottom is a generalized extreme value probability distribution (black) with coefficients chosen to be approximately consistent with average summer Tx. The distribution has its orientation reversed from the usual definition such that it has negative skew. Also plotted are alterations of the base distribution that have values of T05 and T95 values consistent with a 1°C mean warming for the various mean–extreme slopes indicated by the colored crosses in the scatterplots. The red distribution, for example, has a 0.67°C increase in spread between the 5th and 95th percentiles. To aid in visual comparison, all distributions are centered on zero.

slope (Table 2). Cross correlation does not, however, account for mean offsets or differences in scaling between the data, and a reduction in variability statistic provides a complementary description, calculated here as

$$ f = \sum_i (s_i - \hat{s}_i)^2 \bigg/ \sum_i (s_i^2), $$

where s_i are the slopes and $\hat{s}_i$ are the mean of the corresponding bootstrapped estimates. A value of $f = 0$ indicates a perfect prediction, whereas $f > 1$ indicates that the prediction introduces relatively more noise variance

than it explains signal variance and is interpretable as an overall lack of predictive skill. All of the 16 values of f are less than one except for summer Tx05, which has $f = 1.3$. Winter values of f are generally the smallest, averaging only 0.29, consistent with synoptic winter variability largely controlling these mean–extreme slopes.

Bootstrapped samples of the slopes also permit for examining the probability with which an observed slope differs from a stationary process described by the empirical distribution. As opposed to the 47% of estimates that show significantly nonzero slopes, only 17% of mean–extreme slopes significantly differ from these

TABLE 2. Correspondence between mean–extreme slopes and those predicted from the empirical distribution. Upper rows are the cross correlation between mean–extreme slopes and the slopes expected from bootstrapping (left columns) and z bootstrapping (right columns). Lower rows are similar but for the reduction in variance statistic.

	Full bootstrapping				z bootstrapping			
	Tx95	Tx05	Tn95	Tn05	Tx95	Tx05	Tn95	Tn05
Cross correlation								
Winter	0.85	0.83	0.93	0.92	0.74	0.71	0.87	0.87
Spring	0.84	0.82	0.79	0.90	0.73	0.77	0.61	0.84
Summer	0.82	0.79	0.75	0.83	0.46	0.58	0.50	0.70
Fall	0.87	0.88	0.68	0.91	0.73	0.78	0.50	0.83
Variance fraction								
Winter	0.33	0.58	0.70	0.17	0.58	0.56	0.10	0.37
Spring	0.49	0.62	0.53	0.29	0.130	0.58	0.130	0.35
Summer	0.69	1.30	0.68	0.85	4.00	2.10	1.20	0.94
Fall	0.39	0.32	0.51	0.27	0.110	0.45	0.74	0.44

bootstrapped expectations at the 95% confidence level. The general agreement between observed slopes and those derived from the bootstrap indicates that nonnormality in the empirical distributions largely explains the observed nonzero slopes.

Some insight as to why 17% of the observed slopes do not correspond with the bootstrapped results, instead of the expected 5% of false positives, can be gained by again appealing to the six examples provided in section 3. A comparison of mean–Tx95 slopes is made between those obtained from the standard regression across years [i.e., Eq. (1)] and those obtained from bootstrapping (e.g., Fig. 7) using 10 000 data realizations according to each example. Both techniques yield nearly identical distributions of mean–Tx95 slopes in cases where the process is stationary (i.e., examples 1, 5, or 6) but discordant results when the distribution is nonstationary in mean or variance (i.e., examples 2, 3, or 4). Differences arise because bootstrapping gives an estimate of the mean–extreme slope that is consistent with the empirical distribution being stationary, whereas changes in mean and variance generate the appearance of nonnormality in examples 2 and 3 and impose covariance between means and extremes in example 4. These examples suggest that the discordance between the observed and bootstrapped mean–extreme results derives from interannual changes in the underlying distribution.

6. z bootstrapping and soil moisture

To isolate mean–extreme contributions associated with changes in the mean or variance between years, a modification of the bootstrapping procedure is introduced and applied. In this modified form, samples associated with each season and year are normalized to zero mean and unit variance prior to bootstrapping. This procedure is referred to as z bootstrapping in analogy with a z test. The use of z-bootstrapped samples drawn from a normal distribution is expected to yield mean–extreme slopes near zero, regardless of whether the mean or variance of the distribution changes from year to year. Normalizing these leading two moments does not account for interannual changes in the higher-order moments of the temperature distribution, but is at least plausibly sufficient in that the primary axis of interannual variability diagnosed in the initial mean–extreme analyses is oriented along changes in the spread of the distribution (Fig. 6).

Examples 1–6 are again used for purposes of illustration (Fig. 7). The empirical distribution associated with all realized data for each example ($10^4 \times 10^2 \times 90 = 9 \times 10^7$) is shown in Fig. 7 (panels 1a–6a). Empirical distributions are sampled from distributions that are normal (Fig. 7, panel 1a); normal mixtures over means (Fig. 7, panel 2a), variances (Fig. 7, panel 3a), and both means and variances (Fig. 7, panel 4a); the generalized extreme value distribution (Fig. 7, panel 5a); and Student's t distribution (Fig. 7, panel 6a). The z bootstrapping procedure has negligible effect on the samples drawn from a stationary normal distribution (Fig. 7, panel 1a), and converts the normal mixture distributions (Fig. 7, panels 2a–4a; red curves) to approximately standard normal distributions (blue curves) through the suppression of interannual changes in mean and variance. Finally, z bootstrapping has only minor effects on the stationary samples drawn from nonnormal distributions (Fig. 7, panels 5a and 6a).

Parallel results hold for the distribution of mean–extreme slopes shown in Fig. 7 (panels 1b–6b). The distribution of mean–extreme slopes associated with a standard normal distribution is centered on zero, regardless of what bootstrapping approach is applied (Fig. 7, panel 1b). Standard bootstrapping from sample distributions involving mixed normals, however, results in mean–extreme slopes centered away from zero (Fig. 7, panels 2b–4b; red curves). One way to understand this result is that samples drawn across years from a normal mixture follow a nonnormal distribution, as noted in section 5. Application of z bootstrapping in examples 2–4 suppresses the nonstationarity associated with interannual changes in means and variances and yields mean–extreme slopes centered on zero (Fig. 7, panels 2b–4b; blue curves). Finally, application of either bootstrapping or z bootstrapping in example 5 and 6 does not appreciably shift mean–extreme slopes toward zero because they are derived from stationary samples whose nonnormality is inherent at the annual level (Fig. 7, panels 5b and 6b).
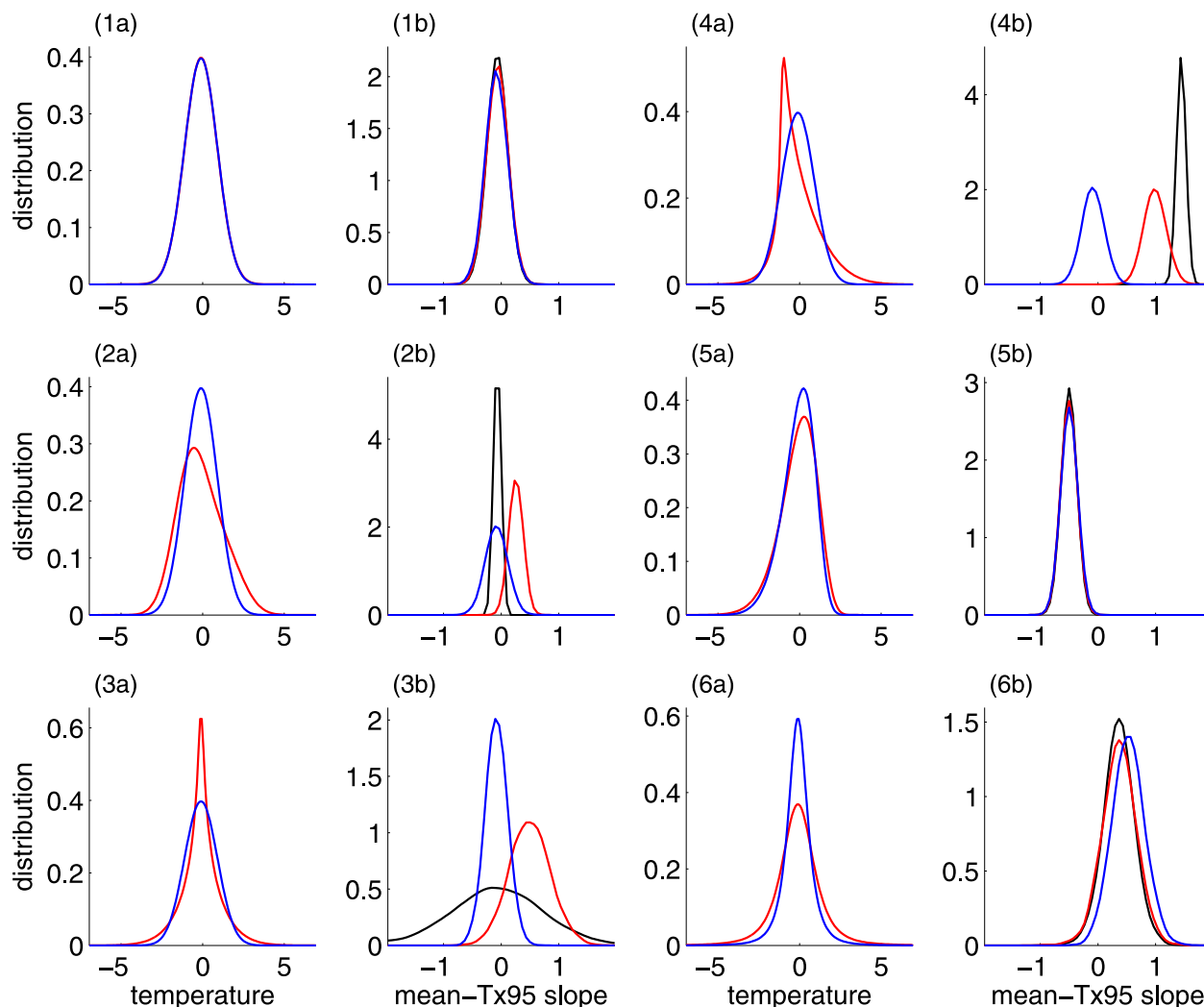
FIG. 7. Approximate distributions for observed, bootstrapped, and z bootstrapped slopes. Numbers refer to the examples described in section 3. (1a)–(6a) The empirical distribution before (red) and after (blue) normalizing each set of observations to zero mean and unit variance. (1b)–(6b) The empirical distribution of mean–extreme slopes (black) and those derived from full bootstrapping (red) and z bootstrapping (blue). Note that in example 4, where trends in the mean and variance are imposed, z bootstrapping (blue) gives values centered nearly on zero that are quite distinct from the observed mean–extreme slopes (black).

A single rescaling of the mean and variance of the empirical distribution across all years would not influence the mean–extreme slopes because the magnitude of this slope is a measure of relative, not absolute, changes. The z bootstrapping procedure, however, acts on each year individually and tends to slightly decrease certain higher-order moments. For instance, the annually normalized empirical distribution associated with examples 1–4 has a kurtosis of 2.8, as opposed to the value of 3 expected for a normal distribution. Relatedly, the average z bootstrapped slopes in examples 1–4 are each −0.05 to within one significant figure. Similar minor distortions of the underlying distribution take place for examples 5 and 6, with normalization leading to

a decrease of skew and kurtosis. As can be seen in Fig. 7, however, the suppression of intrinsic sample variability in mean and variance has a minor influence relative to that associated with the example interannual trends in mean and variance.

Application of the z bootstrap to all U.S. stations and season–variable pairs leads to a squared cross correlation of 0.52 between observed slopes and the average of the z bootstrapped results. This value is lower than the 0.71 average for the full bootstrapping procedure because interannual contributions are no longer accounted for. The reduction in variability statistic again has the smallest values for winter at $f = 0.36$ (Table 2, Fig. 8). Higgins et al. (2002) find interannual variations in the mean and
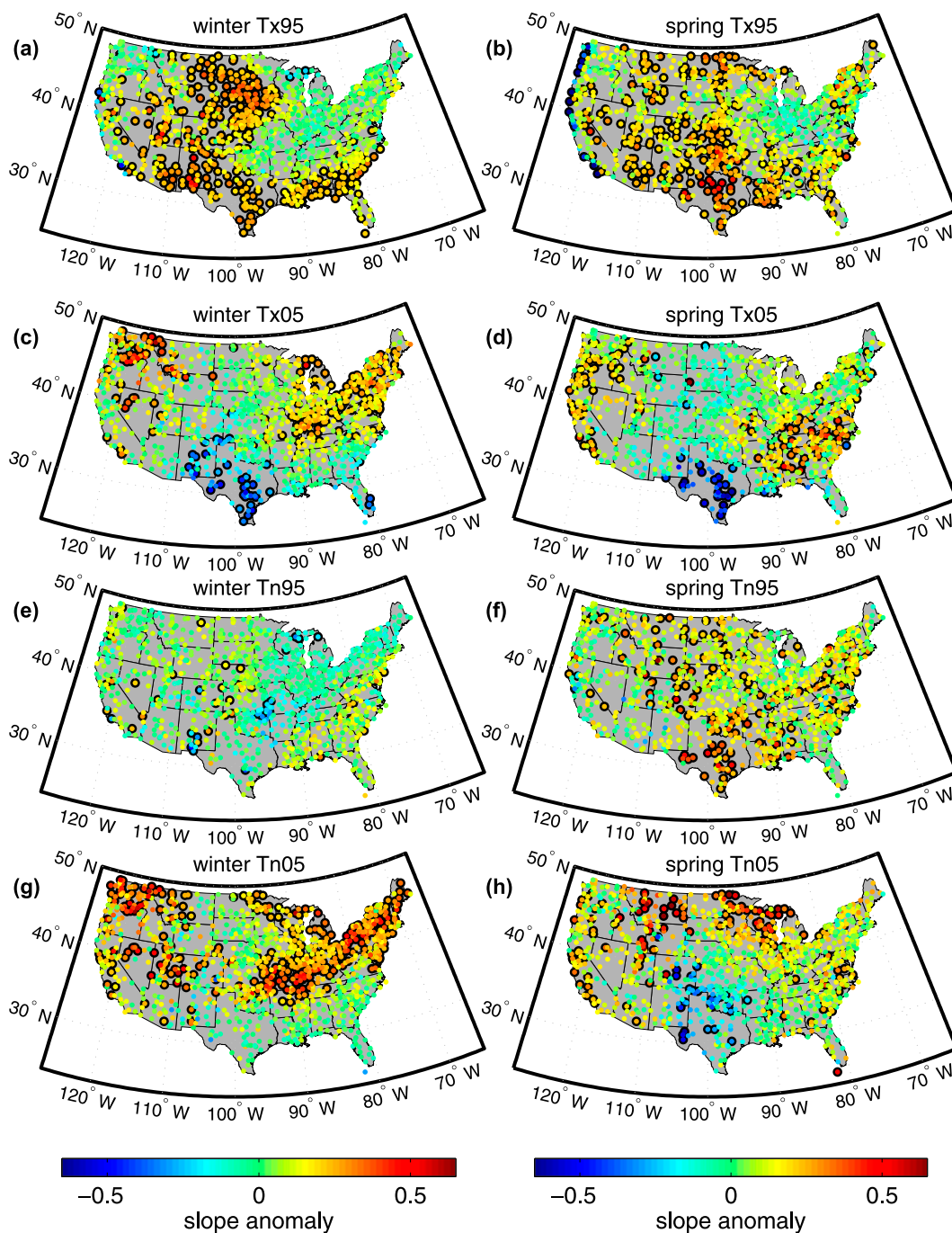
FIG. 8. The difference between observed and z bootstrapped slopes. Points outlined in black have mean–extreme slope that are outside the 95% confidence interval derived from the z bootstrapping results.

skewness of daily winter temperatures associated with La Niña and El Niño conditions that may account for some of this residual structure.

The largest discrepancies with the z bootstrapped results are found for summer, where observed and z-bootstrapped slopes have $f = 4.2$ for Tx95 and $f = 2.1$ for Tx05. A locus of positive residuals in excess of 0.5 appear across the Midwest for Tx95 and negative values in the range of -0.5 for Tx05 across the Midwest, South, and East (Table 2, Fig. 9). These discrepancies are substantially larger than the magnitude of the original mean–extreme slopes (Fig. 5). Station USC00122149,
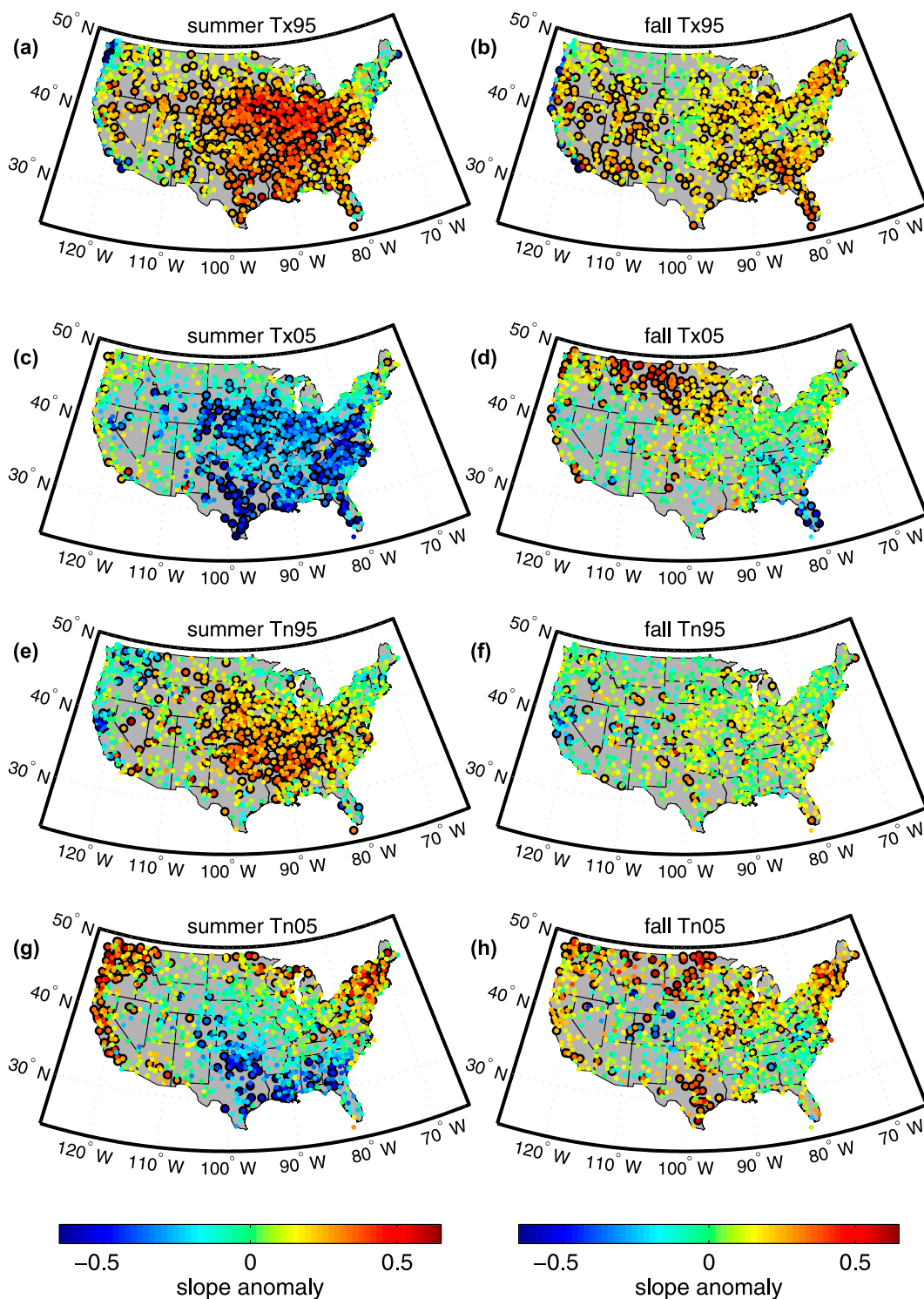
FIG. 9. As in Fig. 8, but for summer and fall.

for example, is located in northwestern Indiana and has a summer mean–Tx95 slope of 0.26, but a z bootstrapped slope of −0.16, giving a residual of 0.42. It can be inferred that interannual nonstationarity at station

USC00122149 gives a positive mean–Tx95 slope, whereas the intraseasonal variations isolated by z bootstrapping are associated with a negative mean–Tx95 slope. This scenario is a hybrid of examples 4 and 5, where the
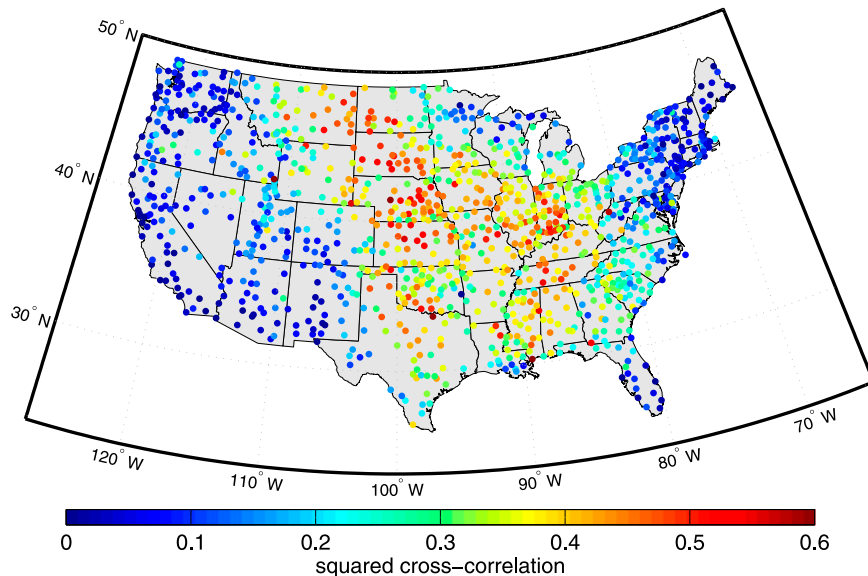
FIG. 10. The squared cross correlation between the self-calibrated Palmer drought severity index and Tx95 for summers between 1900 and 2012. The average squared cross correlation across all stations is 0.25, reflecting the much higher values in the Midwest and parts of the U.S. South. Similar analyses using any other season-variable combination result in average squared cross correlations that are always smaller by at least a factor of 2 and that average 0.04.

generalized extreme value distribution used in example 5 provides a plausible fit to the normalized summer Tx data. Other trials (not shown) demonstrate that a systematic increase in the mean and variance of this distribution, similar to that found in example 4, readily leads to a positive slope when regressing across year but a negative slope from z bootstrapping. The apparent duality in the behavior of the extremes across intraseasonal and interannual time scales suggests another reason for the confusion in the literature regarding whether extremes are changing relative to the mean (Alexander and Perkins 2013; Katz et al. 2013).

A similar distinction was observed with respect to the analysis procedure of Robeson (2002), where increased midwestern variance is observed when using seasonal intervals but not monthly intervals. Evidently, a process with an intrinsic time scale longer than a month is of basic importance for controlling the summer temperature distribution in the Midwest. A prime candidate for increasing the variance of summer Tx is loss of soil moisture. Karl et al. (2012), for example, discuss how evapotranspiration suppresses maximum temperatures when soil moisture is available, and how loss of this latent release of heat translates into higher sensible temperatures with respect to the 2012 U.S. drought. The importance of soil moisture for regulating temperature variability has been established in a large number of other observational and model analyses (Seneviratne et al. 2010) and, given seasonal persistence in soil

moisture properties (e.g., Palmer 1965; Huang et al. 1996), it follows that the shape of the underlying distribution of summer temperature can change between years.

To further examine this relationship, each station and season–variable pair was regressed against a measure of drought, the self-calibrated Palmer drought severity index with the Penman–Monteith formulation of evapotranspiration (PDSI) (Dai 2011). PDSI is provided at monthly resolution on a $2.5° \times 2.5°$ grid, and after averaging to seasonal resolution, comparisons are made between interannual variations in PDSI at each grid box and the 5th and 95th percentiles in station data. Although temperature is itself used in estimating PDSI, the cross correlation with the 5th and 95th temperature percentiles is generally weak, with a median value across stations and season–variable pairs of 0.12. The major exception is for summer Tx95, for which correlations average 0.46 and are as high as 0.80 in the Midwest and parts of the South (Fig. 10). The strongest correlations between drought and temperature correspond in region, season, and variable to the largest discrepancy from the z bootstrapped estimates—namely for Tx in the Midwest during summer—further substantiating a link between increased high temperature extremes and drought. The present analysis does not of itself provide evidence for causality, although other studies have demonstrated the importance of antecedent soil moisture conditions in controlling extreme summer

temperatures (e.g., Durre et al. 2000; Hirschi et al. 2011). A more thorough assessment of causality in this mean–extreme framework is deferred for later work.

7. Summary and conclusions

A simple shift in the distribution is sometimes suggested as the default assumption for how the temperature distribution will change with warming (e.g., Rhines and Huybers 2013; Tingley and Huybers 2013), but such a relationship is not necessarily to be expected when the underlying temperature distribution is nonnormal. Nonnormality implies the presence of an intrinsic relationship between the mean and higher-order moments of the distribution. Indeed, 46% of the 19 488 tested station–variable pairs show significant amplification or damping of extreme values in relation to the mean at the 95% confidence level. But when taking into account the nonnormality indicated by the empirical distribution of each variable at each station, only 17% of station–variable pairs significantly differ from the expected slope.

Contributions to nonnormality in an empirical distribution can come from both intraseasonal variability that is inherently nonnormal and from interannual changes in mean, variance, or higher-order moments (e.g., Karl and Katz 2012; Katz et al. 2013; Otto et al. 2012; Rhines and Huybers 2013; Huntingford et al. 2013; Tingley and Huybers 2013). A second examination of baseline variability is undertaken using a procedure referred to as z bootstrapping that controls for interannual changes in mean and variance. Winter mean–extreme slopes are well explained as a consequence of nonnormal intraseasonal variability, indicating that synoptic variability controls most of the observed wintertime amplification.

Contrasting results are obtained from the z bootstrap for summer temperature variability, particularly in the eastern half of the United States. Amplification of summer Tx relative to the mean only reaches about 10% over most of the eastern United States (Fig. 5), but this apparently represents the residual between contributions from nonnormal intraseasonal variability that would lead to a damping and interannual changes in variance that lead to amplification (Fig. 9). In agreement with other findings (e.g., Durre et al. 2000; Seneviratne et al. 2010; Hirschi et al. 2011; Karl et al. 2012), we speculate that the increase in temperature variance is associated with drying of soils and an attendant loss of evapotranspirative cooling. A strong relationship between variations in the 95th percentile of temperatures and PDSI also supports the suggestion of drought-induced increases in the temperature spread (Fig. 10).

Results indicate that the majority of interannual variability in extreme temperatures follows from the nonnormality of the seasonal temperature distribution (section 5; see Table 2), but that a distinct interannual component can also be identified, especially for summer Tx95 in the eastern United States (section 6). A similar approach could be followed for exploring longer-term variations in the temperature distribution with respect to consistency with the distribution of interannual variability. Insomuch as certain nonlinear processes generate nonnormality at one time scale, they may be expected to contribute a similar coupling at other time scales. Analogous with the concept of forced variations projecting onto "natural modes" of variability, it would be helpful to better understand the degree to which decadal-scale changes project onto "natural moments" of the temperature distribution.

REFERENCES

Alexander, L., and S. Perkins, 2013: Debate heating up over changes in climate variability. *Environ. Res. Lett.,* **8,** 041001, doi:10.1088/1748-9326/8/4/041001.

Barriopedro, D., E. Fischer, J. Luterbacher, R. Trigo, and R. García-Herrera, 2011: The hot summer of 2010: Redrawing the temperature record map of Europe. *Science,* **332,** 220–224, doi:10.1126/science.1201224.

Basu, D., 1958: On statistics independent of sufficient statistics. *Sankhyā,* **20,** 223–226.

Brown, S., J. Caesar, and C. Ferro, 2008: Global changes in extreme daily temperature since 1950. *J. Geophys. Res.,* **113,** D05115, doi:10.1029/2006JD008091.

Dai, A., 2011: Characteristics and trends in various forms of the Palmer Drought Severity Index during 1900–2008. *J. Geophys. Res.,* **116,** D12115, doi:10.1029/2010JD015541.

Donat, M. G., and L. V. Alexander, 2012: The shifting probability distribution of global daytime and night-time temperatures. *Geophys. Res. Lett.,* **39,** L14707, doi:10.1029/2012GL052459.

Durre, I., J. M. Wallace, and D. P. Lettenmaier, 2000: Dependence of extreme daily maximum temperatures on antecedent soil moisture in the contiguous United States during summer. *J. Climate,* **13,** 2641–2651, doi:10.1175/1520-0442(2000)013<2641:DOEDMT>2.0.CO;2.

Engle, R., 2001: GARCH 101: The use of ARCH/GARCH models in applied econometrics. *J. Econ. Perspect.,* **15,** 157–168, doi:10.1257/jep.15.4.157.

Hansen, J., M. Sato, and R. Ruedy, 2012: Perception of climate change. *Proc. Natl. Acad. Sci. USA,* **109,** E2415–E2423, doi:10.1073/pnas.1205276109.

Higgins, R., A. Leetmaa, and V. Kousky, 2002: Relationships between climate variability and winter temperature extremes in the United States. *J. Climate,* **15,** 1555–1572, doi:10.1175/1520-0442(2002)015<1555:RBCVAW>2.0.CO;2.

Hirschi, M., and Coauthors, 2011: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nat. Geosci.,* **4,** 17–21, doi:10.1038/ngeo1032.

Huang, J., H. M. Van den Dool, and K. P. Georgarakos, 1996: Analysis of model-calculated soil moisture over the United States (1931–1993) and applications to long-range temperature forecasts. *J. Climate,* **9,** 1350–1362, doi:10.1175/1520-0442(1996)009<1350:AOMCSM>2.0.CO;2.

Hughes, M., and A. Hall, 2010: Local and synoptic mechanisms causing Southern California's Santa Ana winds. *Climate Dyn.,* **34,** 847–857, doi:10.1007/s00382-009-0650-4.

Huntingford, C., P. D. Jones, V. N. Livina, T. M. Lenton, and P. M. Cox, 2013: No increase in global temperature variability despite changing regional patterns. *Nature,* **500,** 327–330, doi:10.1038/nature12310.

Huybers, P., and W. Curry, 2006: Links between annual, Milankovitch and continuum temperature variability. *Nature,* **441,** 329–332, doi:10.1038/nature04745.

Kaplan, E. L., and P. Meier, 1958: Nonparametric estimation from incomplete observations. *J. Amer. Stat. Assoc.,* **53,** 457–481, doi:10.1080/01621459.1958.10501452.

Karl, T. R., 1985: Intraseasonal variability of extremely cold and warm months in the contiguous United States. *J. Appl. Meteor.,* **24,** 215–227, doi:10.1175/1520-0450(1985)024<0215:IVOECA>2.0.CO;2.

——, and R. W. Katz, 2012: A new face for climate dice. *Proc. Natl. Acad. Sci. USA,* **109,** 14 720–14 721, doi:10.1073/pnas.1211721109.

——, and Coauthors, 2012: U.S. temperature and drought: Recent anomalies and trends. *Eos, Trans. Amer. Geophys. Union,* **93,** 473–474, doi:10.1029/2012EO470001.

Katz, R. W., P. F. Craigmile, P. Guttorp, M. Haran, B. Sansó, and M. L. Stein, 2013: Uncertainty analysis in climate change assessments. *Nat. Climate Change,* **3,** 769–771, doi:10.1038/nclimate1980.

Király, A., I. Bartos, and I. M. Jánosi, 2006: Correlation properties of daily temperature anomalies over land. *Tellus,* **58A,** 593–600, doi:10.1111/j.1600-0870.2006.00195.x.

Koenker, R., 2005: *Quantile Regression.* Cambridge University Press, 349 pp.

Machado, J. A. F., and J. M. C. Santos Silva, 2005: Quantiles for counts. *J. Amer. Stat. Assoc.,* **100,** 1226–1237, doi:10.1198/016214505000000330.

McKinnon, K. A., A. R. Stine, and P. Huybers, 2013: The spatial structure of the annual cycle in surface temperature: Amplitude, phase, and Lagrangian history. *J. Climate,* **26,** 7852–7862, doi:10.1175/JCLI-D-13-00021.1.

Meehl, G., and C. Tebaldi, 2004: More intense, more frequent, and longer lasting heat waves in the 21st century. *Science,* **305,** 994–997, doi:10.1126/science.1098704.

Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the global historical climatology network-daily database. *J. Atmos. Oceanic Technol.,* **29,** 897–910, doi:10.1175/JTECH-D-11-00103.1.

Otto, F. E. L., N. Massey, G. J. V. Oldenborgh, R. G. Jones, and M. R. Allen, 2012: Reconciling two approaches to attribution of the 2010 Russian heat wave. *Geophys. Res. Lett.,* **39,** L04702, doi:10.1029/2011GL050422.

Palmer, W. C., 1965: Meteorological drought. U.S. Department of Commerce, Weather Bureau, Research Paper 45, 58 pp.

Portmann, R., S. Soloman, and G. Hegerl, 2009: Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States. *Proc. Natl. Acad. Sci. USA,* **106,** 7324–7329, doi:10.1073/pnas.0808533106.

Rahmstorf, S., and D. Coumou, 2011: Increase of extreme events in a warming world. *Proc. Natl. Acad. Sci. USA,* **108,** 17 905–17 909, doi:10.1073/pnas.1101766108.

Rhines, A., and P. Huybers, 2013: Frequent summer temperature extremes reflect changes in the mean, not the variance. *Proc. Natl. Acad. Sci. USA,* **110,** E546, doi:10.1073/pnas.1218748110.

Robeson, S. M., 2002: Relationships between mean and standard deviation of air temperature: Implications for global warming. *Climate Res.,* **22,** 205–213, doi:10.3354/cr022205.

Ruff, T. W., and J. D. Neelin, 2012: Long tails in regional surface temperature probability distributions with implications for extremes under global warming. *Geophys. Res. Lett.,* **39,** L04704, doi:10.1029/2011GL050610.

Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature,* **427,** 332–336, doi:10.1038/nature02300.

Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär, 2006: Land–atmosphere coupling and climate change in Europe. *Nature,* **443,** 205–209, doi:10.1038/nature05095.

——, T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth Sci. Rev.,* **99,** 125–161, doi:10.1016/j.earscirev.2010.02.004.

Stine, A. R., and P. Huybers, 2012: Changes in the seasonal cycle of temperature and atmospheric circulation. *J. Climate,* **25,** 7362–7380, doi:10.1175/JCLI-D-11-00470.1.

Teng, H., G. Branstator, H. Wang, G. Meehl, and W. Washington, 2013: Probability of US heat waves affected by a subseasonal planetary wave pattern. *Nat. Geosci.,* **6,** 1056–1061, doi:10.1038/ngeo1988.

Tingley, M. P., and P. Huybers, 2013: Recent temperature extremes at high northern latitudes unprecedented in the past 600 years. *Nature,* **496,** 201–205, doi:10.1038/nature11969.

Wettstein, J. J., and L. O. Mearns, 2002: The influence of the North Atlantic–Arctic Oscillation on mean, variance, and extremes of temperature in the northeastern United States and Canada. *J. Climate,* **15,** 3586–3600, doi:10.1175/1520-0442(2002)015<3586:TIOTNA>2.0.CO;2.

Zhang, X., F. W. Zwiers, and G. Hegerl, 2009: The influences of data precision on the calculation of temperature percentile indices. *Int. J. Climatol.,* **29,** 321–327, doi:10.1002/joc.1738.

Zwiers, F. W., X. Zhang, and Y. Feng, 2011: Anthropogenic influence on long return period daily temperature extremes at regional scales. *J. Climate,* **24,** 881–892, doi:10.1175/2010JCLI3908.1.