Seasonally Resolved Distributional Trends of North American Temperatures Show Contraction of Winter Variability

Seasonally Resolved Distributional Trends of North American Temperatures Show Contraction of Winter Variability

ANDREW RHINES

Department of Atmospheric Sciences, University of Washington, Seattle, Washington

KAREN A. MCKINNON

National Center for Atmospheric Research, Boulder, Colorado

MARTIN P. TINGLEY

Departments of Statistics and Meteorology, The Pennsylvania State University, State College, Pennsylvania

PETER HUYBERS

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts

(Manuscript received 9 May 2016, in final form 20 October 2016)

ABSTRACT

There is considerable interest in determining whether recent changes in the temperature distribution extend beyond simple shifts in the mean. The authors present a framework based on quantile regression, wherein trends are estimated across percentiles. Pointwise trends from surface station observations are mapped into continuous spatial fields using thin-plate spline regression. This procedure allows for resolving spatial dependence of distributional trends, providing uncertainty estimates that account for spatial covariance and varying station density. The method is applied to seasonal near-surface temperatures between 1979 and 2014 to unambiguously assess distributional changes in the densely sampled North American region. Strong seasonal differences are found, with summer trends exhibiting significant warming throughout the domain with little distributional dependence, while the spatial distribution of spring and fall trends show a dipole structure. In contrast, the spread between the 95th and 5th percentile in winter has decreased, with trends of $-0.71°$ and $-0.85°\mathrm{C\,decade^{-1}}$, respectively, for daily maximum and minimum temperature, a contraction that is statistically significant over 84% of the domain. This decrease in variability is dominated by warming of the coldest days, which has outpaced the median trend by approximately a factor of 4. Identical analyses using ERA-Interim and NCEP-2 yield consistent estimates for winter (though not for other seasons), suggesting that reanalyses can be reliably used for relating winter trends to circulation anomalies. These results are consistent with Arctic-amplified warming being strongest in winter and with the influence of synoptic-scale advection on winter temperatures. Maps for all percentiles, seasons, and datasets are provided via an online tool.

1. Introduction

Trends of increasing large-scale mean temperature are unequivocally contributing to an increasing frequency of warm temperature extremes and decreasing frequency of cold temperature extremes in many regions (Meehl et al. 2009; Rahmstorf and Coumou 2011), and would do so even in the absence of distributional changes. However, other changes in temperature distributions and their

Denotes Open Access content.

Supplemental information related to this paper is available at the Journals Online website: http://dx.doi.org/10.1175/JCLI-D-16-0363.s1.

Corresponding author e-mail: Andrew Rhines, arhines@atmos.uw.edu

DOI: 10.1175/JCLI-D-16-0363.1

contribution to the observed changes in extremes are less well understood (Alexander and Perkins 2013). Indeed, changes in the variance and higher-order moments can be of leading-order importance (Katz and Brown 1992; Ruff and Neelin 2012).

Previous studies have examined changes in temperature extremes and variability using a variety of data, models, and metrics. In observation-based studies, limitations of data availability lead to analyses being performed primarily on spatially aggregated (Hansen et al. 2012), zonally averaged (Huntingford et al. 2013; Screen 2014), or reanalysis-derived samples (Screen 2014). Hansen et al. (2012) used regional standardization to aggregate monthly data and found a large increase in temperature variability during winter and summer. Rhines and Huybers (2013) noted that there is no statistically significant change in these aggregated monthly observations when several biases are accounted for, and Huntingford et al. (2013) used some of these corrections to show that zonally averaged temperature variability has generally remained unchanged on monthly time scales, albeit with regional differences. Further studies have identified decreases in variability in both models and observations (Screen 2014; Schneider et al. 2015). Determining whether these disagreements arise from data or model limitations, methodological distinctions, or qualitative differences in interpretation is important for assessing the role of distributional changes relative to that of simple shifts in the mean.

Discrepancies between observational results stem in part from differences in methodology (Alexander and Perkins 2013). Many previous studies have employed averaging or aggregation methods that are prone to biases and have low resolving power for detecting distributional changes (Tingley 2012; Rhines and Huybers 2013). For example, variance-based metrics in common use (Hansen et al. 2012; Huntingford et al. 2013) are best suited to detecting changes in normally distributed temperatures and have limited interpretability in view of the nonnormal characteristics of daily temperature distributions (Cavanaugh and Shen 2014). Further distortions of the distribution arise from statistical procedures used to interpolate or aggregate samples across space and time. For example, it is well known (Haylock et al. 2008; Zhang et al. 2011; Kennedy et al. 2011; Rhines and Huybers 2013; Director and Bornn 2015; Cavanaugh and Shen 2015) that interpolating or taking spatial averages alters the variance and higher-order moments relative to point estimates. Similarly, monthly averages (as employed by, e.g., Hansen et al. 2012; Huntingford et al. 2013) mask nonnormality that dominates the tails of daily temperature distributions (Sardeshmukh and Sura 2009), as do filters designed to isolate specific time scales (Proistosescu et al. 2016). Finally, the analysis of temporal blocks of anomalies defined relative to a fixed subset of the full time period (Caesar et al. 2006; Brown et al. 2008; Simolo et al. 2011, 2012; Hansen et al. 2012; Cavanaugh and Shen 2014) artificially influences trends and variability (Tingley 2012; Hawkins and Sutton 2016). Little signal remains once appropriate corrections are performed to account for these estimation biases (Rhines and Huybers 2013), pointing to the utility in employing different methods that minimize the use of parametric assumptions.

We argue that many of these issues can be obviated through several specific methodological choices. First, we suggest that in situ observations are necessary at least as a complement to reanalyses. Second, the use of daily surface station observations wherever possible in the initial analysis will reduce the tendency to limit distributional resolution through spatial or temporal averaging. Third, existing regression techniques designed specifically to assess distributional change should be employed, as opposed to the more ad hoc combination of methods such as ordinary least squares and empirical quantiles. Finally, the full spatial field of distributional trends should be estimated using a spatial model to account for spatial variations in observation density, and to retain uncertainty information from the pointwise regression. This order of operations permits the pooling of information across space when needed, for example, when an area average of trends is desired, while also distinguishing temporal changes from spatial variability.

Here we combine a set of methods satisfying these criteria to estimate distributional trends in the densely sampled 25°–55°N region of North America. Focusing on this region serves the dual purpose of demonstrating that there is important distributional dependence to the observed climate changes in the region, while also permitting for an unambiguous comparison between surface observations and reanalysis-based estimates.

We use quantile regression (QR; Koenker and Hallock 2001) to estimate local distributional trends in temperatures from 3220 weather stations in the Global Historical Climatology Network–Daily database (GHCN-D; Menne et al. 2012). Quantile regression has been successfully applied to other climate data; however, the present analysis entails two orders of magnitude more observations than previous quantile regression analyses of station data from South Korea (Lee et al. 2013) and Europe (Barbosa et al. 2011; Matiu et al. 2016) and has not been previously possible using raw daily observations because of a recently resolved issue with quantization of the data (Rhines et al. 2015). Unlike approaches employed elsewhere (Brown et al. 2008; Simolo et al. 2012; Mannshardt et al. 2013), QR does not require assumptions regarding the parametric form of the underlying probability

distribution. As a result, the method does not yield biased estimates of distributional spread when data are nonnormal (see the appendix and Fig. A3).

We estimate the full spatial field of distributional trends and their uncertainties using thin-plate spline regression (Furrer et al. 2006). This mapping allows for quantification of the combined uncertainty due to variations in station density, spatial covariance of the climate field, and uncertainties in the individual pointwise trend estimates. We also map the GHCN-D-based observational estimate to the uniform grids employed by both the European Centre for Medium-Range Weather Forecasts interim reanalysis (ERA-Interim; Dee et al. 2011) and National Centers for Environmental Prediction's Reanalysis-2 (NCEP-2; Kanamitsu et al. 2002), intercomparing them to evaluate their consistency.

2. Data and methods

a. Station data

The GHCN-D station data are screened by location and temporal coverage and for quality control indicators. All observations with negative quality control flags are removed, and only stations in North America between 25° and 55°N are retained. We then screen for stations that are 80% complete for a given season over the duration of the 1979–2014 interval. Each pentad (5 yr) is then examined, and stations are excluded if they have fewer than 75% completeness in more than three years per pentad, ensuring that the maximum of 20% missing data are not overly concentrated in a particular span of time. The screening leaves 3220 total stations and an average of approximately 3000 stations reporting on a given day within the region.

Daily minimum T_n and maximum T_x temperatures are examined separately without subtracting a climatology, in contrast to some previous studies that estimate daily mean temperature anomalies using averaging and smoothing (Caesar et al. 2006; Alexander et al. 2006; Cavanaugh and Shen 2014), as the two extrema are often governed by different physical processes—for example, by longwave radiative cooling in stable, clear, dry nighttime conditions for T_n, and by boundary layer mixing with the free troposphere during the day for T_x (Misra et al. 2012; McNider et al. 2012). Even if higher-order moments of extrema were to cancel in averaging, this need not lead to unbiased estimates of the true daily mean (Wang 2014).

The seasons used here are DJF, MAM, JJA, and SON, though we note that we obtained similar results when using shorter 2-month seasons. Seasonal subsets are used instead of the full annual time series to avoid aliasing of the entire annual cycle into the modeled distribution, and to permit for each season having trends that are divergent as a result of their being controlled by different processes. Spring and fall generally contain most of the distributional structure related to seasonal transitions, a factor that bears consideration in interpreting our results for those seasons.

b. Quantile regression

Characterization of the distribution of a random variable X can be couched in terms of its cumulative distribution function,

$$F(x) = P(X \leq x), \tag{1}$$

or equivalently in terms of the quantile function,

$$Q_X(\tau) = F^{-1}(\tau), \tag{2}$$

where $0 < \tau < 1$ so that τ is the probability of X exceeding $Q_X(\tau)$. For independent, identically distributed (IID) temperature observations, $T = \{T_1, T_2, \ldots, T_n\}$, the quantile function $Q_T(\tau)$ can be estimated by optimization (Koenker and Bassett 1978),

$$\arg\min_{\xi \in \mathbf{R}} \sum_{j=1}^{n} \rho_\tau(T_i - \xi), \tag{3}$$

where ξ is the quantile, and the piecewise-linear function,

$$\rho_\tau(x) = \begin{cases} \tau x, & \text{if } x \geq 0 \\ (\tau - 1)x, & \text{otherwise}, \end{cases} \tag{4}$$

yields a tilted absolute value loss function whose angle depends on the chosen value of τ. For values of τ above the median ($\tau > 1/2$) the tilt in the loss function results in positive residuals contributing more to the loss function than negative residuals, and vice versa for values below the median ($\tau > 1/2$). In contrast, ordinary least squares (OLS) regression is based on minimizing a symmetric, quadratic loss function. QR extends this approach to dependent random variables by performing a similar minimization but with ξ replaced by basis functions in the independent variable (in our case time), $t = \{t_1, t_2, \ldots, t_n\}$, to estimate the conditional quantile function $Q_T(\tau \mid t)$. Because the 1979–2014 interval exhibits climatic trends that have been relatively constant, we use a linear model with intercept α and temporal slope β. Thus the problem becomes one of minimization (Koenker and Bassett 1978; Koenker and Hallock 2001; Cade and Noon 2003):

$$\arg\min_{\beta,\alpha} \sum_{j=1}^{n} \rho_\tau\{T_i - [\alpha(\tau) + \beta(\tau)t_i]\}. \tag{5}$$

The QR minimization problem is solved via a linear programming algorithm (Koenker and Bassett 1978), and we use a MATLAB implementation (Koenker 2014).

Quantile regression shares a first-order equivalence with least squares regression on empirical quantiles calculated from blocks of data segmented by the predictor variable (Bassett et al. 2002). Indeed, past studies have taken the latter approach (Caesar et al. 2006; Robeson et al. 2014; Huybers et al. 2014), but improving the resolution of distributional tails with the empirical quantile method requires pooling of data across two or more years to increase the number of observations available in each block. As our interest is in trends in the tails of seasonal temperature distributions at annual and greater time scales, QR has distinct advantages in terms of being more efficient and less susceptible to biases introduced by gaps in observations. QR also yields a clearer representation of uncertainties in that the estimated trends are asymptotically normal (Koenker and Bassett 1978), with deviations appearing only for extremal quantiles that are sparsely represented in the data (Chernozhukov 2005), whereas to our knowledge no comparable properties have been identified for OLS performed on running blocks of either empirical quantiles (Caesar et al. 2006) or moments (Cavanaugh and Shen 2014). We provide several pedagogical examples of QR applications in the appendix.

Standard analytical treatments of QR uncertainties assume continuity of the conditional density function. Continuity facilitates asymptotic approaches to estimating functional forms for standard errors and normal approximations to the QR estimator (Koenker and Bassett 1978; Simpson et al. 1987; Knight 1998). The rounded nature of the GHCN-D data and the expectation that daily temperature observations are autocorrelated renders standard approaches impractical, and we instead quantify uncertainty using a block residual bootstrap procedure (Barbosa et al. 2011; Lee et al. 2013). Residuals about the fits for each percentile are resampled with replacement in blocks of entire years and then added back to the estimated trend. Quantile regression is then repeated for 1000 different resampled time series for the percentile in question, forming a bootstrap estimate of the slope and intercept terms. For trend differences (e.g., the 95th − 5th percentile), the same bootstrap samples are used simultaneously for both percentiles. Slopes and differences for which the corresponding 95% central confidence intervals exclude zero are considered significant. We note that, because the data are from 3-month seasons, annual blocks yield the same results as any block size choice between seasonal and annual time scales.

c. Precision decoding

Quantile regression relies on the assumption that data are continuously distributed rather than having finite precision, leading to significant—but correctable—biases in estimating temperature trends from discrete observations (Koenker and Hallock 2001; Rhines et al. 2015). Specifically, nonsmooth distributions lead to nondifferentiable objective functions and, consequently, biased estimates of conditional quantiles (Machado and Silva 2005). In practice all data are discrete to some numerical precision, but the bias becomes especially severe with the coarse sampling present in the GHCN-D data. If the rounding methods used to record the observations are known it is possible to correct for this bias (Reich and Smith 2013; Machado and Silva 2005). Accordingly, jitter drawn from a uniform distribution is added to all GHCN-D observations, with amplitude that is inferred using precision decoding (Rhines et al. 2015). Inferring precision is nontrivial because the majority of observations have been rounded in Fahrenheit, converted to Celsius, and rerounded. Further, observations are generally without metadata to indicate the original precision but can be inferred to have followed a variety of reporting conventions that differ both temporally and spatially. We also correct for small offsets that occur as a result of Fahrenheit observations having been converted to 0.1°C precision in the GHCN-D database, though the effect of these double-rounding errors on the results is small. However, absent the procedure of jittering the data, the majority of QR trend estimates would be erroneously reported as being close to machine-precision zero (see the appendix and Fig. A1 for an instructive example).

To assess the additional uncertainty due to finite precision of the observations and the jitter-based correction procedure, we calculate the variances of the slope estimates under 1000 independent realizations of the jittering procedure. The overall uncertainty is formed from the sum of the jitter and annual block bootstrap variances, and is strongly dominated by the bootstrap uncertainty. Bootstrap variance contributed by jittering is typically an order of magnitude smaller than the total, and jittering is important for avoiding what would otherwise be large biases.

d. Reanalyses

Reanalyses that blend multiple data sources with dynamical models in a formal data assimilation procedure are an alternative to purely observational datasets. Their advantages include spatial and temporal completeness, and the fact that long-term statistics at a given point are determined only by the physical model, rather than by the density of observations. Indeed, a number of studies (e.g., Huntingford et al. 2013; Screen 2014; Schneider et al. 2015) have used reanalyses to examine variability trends. However, evidence that purely

observational datasets more reliably capture surface temperatures (Bengtsson et al. 2004; Hanson et al. 2007; Hofstra et al. 2010; Donat et al. 2014), as well as known issues with data homogeneity due to changing observational networks and bias corrections (e.g., Dee et al. 2011; Bracegirdle and Marshall 2012), suggests that a careful comparison is warranted. Such caveats naturally also apply to analyses of individual or ensemble climate model integrations when observationally based bias corrections are used.

Near-surface temperature estimates from ERA-Interim and NCEP-2, both available from 1979 to the present, are used for comparison with station data over the same 1979–2014 interval. For ERA-Interim, the 6-hourly cumulative 2-m minimum and maximum temperature fields are used to derive T_n and T_x at each grid point using the 3-h forecast output. This forecast period was used as it has been found to produce slightly better estimates of near-surface temperature relative to the 12-h forecast within Europe (Cornes and Jones 2013). We provide results for three resolutions for ERA-Interim: 2.5° × 2.5°, 1.5° × 1.5°, and 0.75° × 0.75° (close to the effective resolution of the models dynamical core). The regridded samples are averaged using area-conservative remapping of the high-resolution grid, which is in principle necessary as the default regridded datasets provided by ECMWF are more representative of higher-resolution output than gridcell averages. The highest resolution (0.75° × 0.75°) is considered most representative of point observations (Cornes and Jones 2013), though the use of subgrid-scale parameterizations renders the distinction between point and area-averaged processes less distinct at the native resolution. Regridding is performed prior to extracting the daily extrema. The NCEP-2 6-hourly 2-m temperature field is used without altering the standard 2.5° × 2.5° grid, using each day's samples to select the maximum and minimum temperature. The 6-hourly data underestimate the true variability of T_n and T_x but are a useful proxy given that NCEP-2 does not archive daily extrema directly.

e. Spatial smoothing and uncertainty quantification

To assess overall uncertainty and to facilitate spatial averaging and intercomparison with other datasets, we estimate a smooth spatial field from the station-level results using thin-plate spline regression and map them onto the different regular latitude–longitude grids from the NCEP-2 and ERA-Interim reanalyses (Fig. 1). Thin-plate spline regression is performed using fastTPS within the R (Ihaka and Gentleman 1996) package fields (Furrer et al. 2013). The fastTPS method differs from standard kriging in that it uses the compactly supported, isotropic, and stationary Wendland covariance functions,

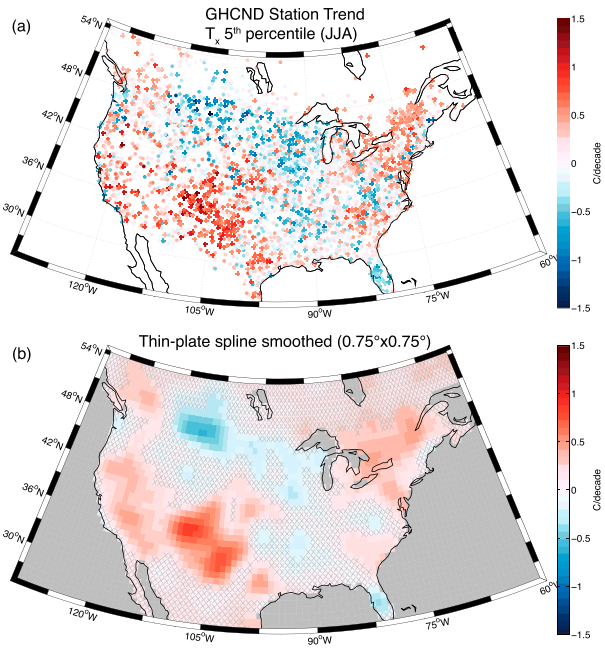

FIG. 1. Example of spatial smoothing via thin-plate spline regression. (a) Trends (1979–2014) in the 5th percentile of summer T_x using only pointwise GHCN-D time series. (b) As in (a), smoothed to the 0.75° × 0.75° ERA-Interim grid. In (a), crosses indicate a trend that is significant with respect to the 95% central confidence interval of the combined block bootstrap and jitter variances, while circles are insignificant but nevertheless used in the smoothing. In (b), insignificant trends (hatching) are identified by comparing the trend with the standard errors inferred from the thin-plate spline regression. When the trend differs from zero by at least 1.96 standard errors it is deemed significant.

permitting smoothing to be performed efficiently for large datasets. We use a range of $\theta = 400$ miles (great circle distance) so that the estimated value at each location is affected by all values within that radius, though they are naturally dominated by nearby stations; results are similar for larger (1000 miles) and smaller (200 miles) values of θ. A maximum likelihood estimate of the smoothing parameter λ that determines the partitioning of variance between the smooth interpolating surface and the spread of the observations around the smooth surface is used by fastTPS. The uncertainty in the estimate of the smoothed field at each location is determined by the data availability in the surrounding area, the QR bootstrap uncertainties, and also by the overall partitioning of variance controlled by λ.

The standard errors of the estimated smoothed surface are used to quantify spatial dependence of the uncertainty, and we consider as significant those grid boxes with values that are more than 1.96 standard errors from zero (equivalent in the context of the spatial model to the central 95% uncertainty estimate used for the

stationwise trends). These uncertainties represent the combination of sampling limitations and small-scale spatial variability that is independent of sampling and estimation issues. This approach to uncertainty quantification displays robustness to the presence of inhomogeneities known to exist (Menne et al. 2012) in some of the original station-level time series. Step function inhomogeneities, such as those arising from changes in instrument location, type, time of observation, or other observation protocol, will generally increase the variance of the original series (Figs. 4 and 5 of Menne and Williams 2009) and will therefore increase variance about any climatic trends. Inasmuch as these changes have been introduced at different times for different stations, these differences will then also influence spatial variance. Though the GHCN-D archival process does not involve any homogenization, examination of the distribution of residuals from the thin-plate spline regression suggests that artificial inhomogeneities related to sampling do not significantly influence our conclusions. Furthermore, examination of hourly observations in tandem with daily extrema from nearby stations suggests that the systematic biases induced by time of observation changes are on the order of 0.01°C decade^{-1} (McKinnon et al. 2016), which is small compared with the signals on the order of 0.1°–1.0°C decade^{-1} that we resolve. Furthermore, we note that a subset of the analyses performed using only stations from the higher-quality U.S. Historical Climatology Network yields similar results.

The uncertainty of spatially averaged fields is estimated using a parametric bootstrap procedure, wherein 1000 realizations of the field estimated by thin-plate spline regression are generated on the same uniform grid. Each sample is spatially averaged to produce a bootstrap distribution that is analogous to the spatially dependent standard error estimates, but which accounts for the effect of spatial covariance upon the estimate of the mean.

A second procedure was also considered wherein we reversed the order of operations, using thin-plate splines to estimate the temperature field as a function of time and then estimating regression quantiles as the final step. We tested the sensitivity of our results to this methodological choice on the GHCN-D data and found only minor differences relative to the primary method (see the online visualization tool at http://qrmaps.earthto.me, where we also provide results with no spatial smoothing). Though the spatial patterns are very similar when using the alternative approach, applying the spatial model as the final step is advantageous in that it incorporates spatial covariance in its estimates of significance. Furthermore, there is empirical evidence that distributional properties independent of the mean vary more smoothly in space than the raw temperature field (Cavanaugh and Shen 2014), and prematurely imposing smoothness may lead to biases in regions with large small-scale variability such as those dominated by topographic gradients.

3. Results

Here we provide estimates of distributional trends for each season and percentile for both T_x and T_n. Summary plots illustrating the spatially averaged trends, the spatial deviation about those averages, and their corresponding uncertainties are shown for all cases in section 3a. Spatial maps for all percentiles, seasons, and datasets are also provided via an online visualization tool (http://qrmaps.earthto.me).

a. *Spatial averages of trends*

Although our analysis specifically aims to retain spatial information about distributional trends, the spatial average of distributional changes is a question of broad interest (Alexander and Perkins 2013). By analyzing the variability changes locally prior to spatial aggregation we avoid many issues that artificially affect the inferred average distributional change (Rhines and Huybers 2013). We compare the spatially averaged trends, the spatial variability about that average, and the uncertainty estimates by season, variable, and percentile in Fig. 2 (also see Table S1 in the supplemental material). The spatial standard deviation of the trends,

$$\sigma_{\mathrm{sp}} = \sqrt{\frac{N}{N-1}\sum_{j=1}^{N} w_j(\beta_j - \overline{\beta})^2}, \tag{6}$$

where w_j are area weights proportional to the size of the corresponding grid box and are normalized such that the sum equals unity, and N is the number of grid cells, summarizes the level of spatial variability relative to the spatial-mean trend $\overline{\beta}$. The mapping of the quantile trends shown in Fig. 2 onto moments such as variance and skewness depends on the base distribution; however, interpreting the changes in variability from quantile trends is arguably more straightforward than moments when distributions are nonnormal (Koenker and Hallock 2001; McKinnon et al. 2016). Negative slopes of the quantile trends (e.g., Figs. 2a,c,e) indicate contraction of the entire distribution, whereas positive slopes (the lower half of the distribution in Fig. 2b) indicate expansion of that part of the distribution. Trends that have little dependence on τ (e.g., Fig. 2d) suggest a uniform shift in the distribution, consistent with a simple trend in the mean.

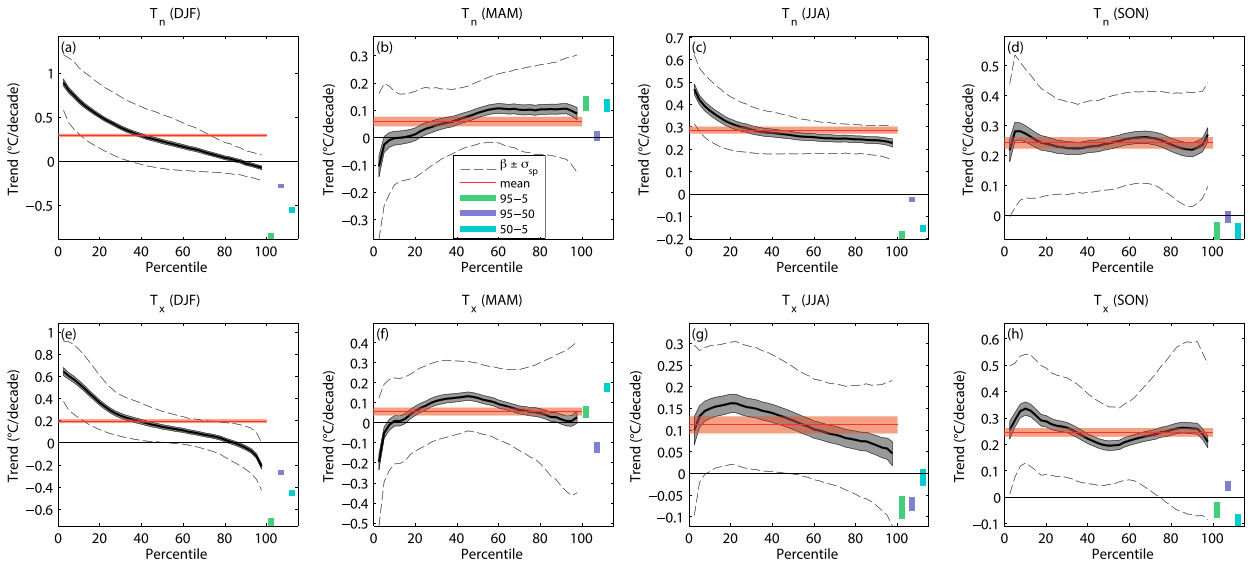

FIG. 2. Daily (a)–(d) T_n and (e)–(h) T_x trends (1979–2014) by percentile for each season for the GHCN-D station data. The domain-averaged of the temporal trend β is plotted in black with bootstrap 95% confidence intervals indicated by the gray band. The domain-average OLS trend in the mean its 95% bootstrap confidence interval is shown in red. For reference, the spatial variability about the mean is illustrated by the $\pm\sigma_{sp}$ dashed lines about the mean. The 95% confidence interval for trend differences of the 95th − 5th, 95th − 50th, and 50th − 5th are shown at the right in each panel. Note the different vertical scales.

In interpreting the spatial-average results (Fig. 2) we highlight that trends in the median, upper tail, and lower tail of the distribution can all differ substantially from each other and from the OLS trend in the mean. Furthermore, in each season and for each percentile, the combined signal from the spatial mean trend and the spatial standard deviation σ_{sp} is substantially larger than the mean standard error, indicating a large signal-to-noise ratio.

The spatial variability of the trends is usually at least as large in magnitude as the spatial mean trend, with the exception of summer where trends are broadly consistent with a simple increase in the mean, and low percentiles in winter. This underscores that spatial averages often provide an incomplete description of distributional changes. We find that, with the exception of spring, the distributions of both T_x and T_n are decreasing in variability, as measured by the difference between the 95th and 5th percentile trends (green box-and-whisker markers in Fig. 2, and Table S1 in the supplemental material). This trend toward a reduction in variability is dominated by winter, with trends of −0.71° and −0.85°C decade^{-1} for T_x and T_n, respectively, and is not sensitive to the particular choice of 95th and 5th percentiles as opposed to, for example, the 90th and 10th. These findings provide purely observational evidence against the claim that temperature variability is increasing substantially over Northern Hemisphere land (Hansen et al. 2012), at least in the densely sampled North American region.

b. *Spatial patterns of distributional trends*

Winter temperatures show strong warming, with a spatial pattern of polar amplification in the lower half of the distribution for both T_x and T_n (Figs. 3 and 2a,e). In the upper half of the distribution, warming gives way to a dipole pattern with cooling in the mountain west and warming in the southeast. By the 95th percentile the warming signal is mostly absent, while the cooling is strong and spatially widespread. In eastern Canada, warming is still present in T_n as high as the 80th percentile, while that signal dissipates in T_x near the median, contributing to the wider spatial distribution (dashed lines, Figs. 2a,e). The competing sign of trends at the low and high tails of the distribution leads to a significant contraction in daily temperature variability over most of the domain, with distributional dependence that is easily distinguishable from a shift in the mean. We discuss potential mechanisms for the winter results in detail in section 5.

For the spring season we find generally stronger spatial signals in T_x than in T_n (Fig. 4), with isolated warming in the southwest at low percentiles giving way to broader warming south of the Canadian border by the median. A strengthened north–south dipole emerges toward higher percentiles in the center of the domain, with magnitudes becoming pronounced by the 95th percentile (reflected as increased spatial variability by dashed lines in Fig. 2f). Averaged over the domain, the

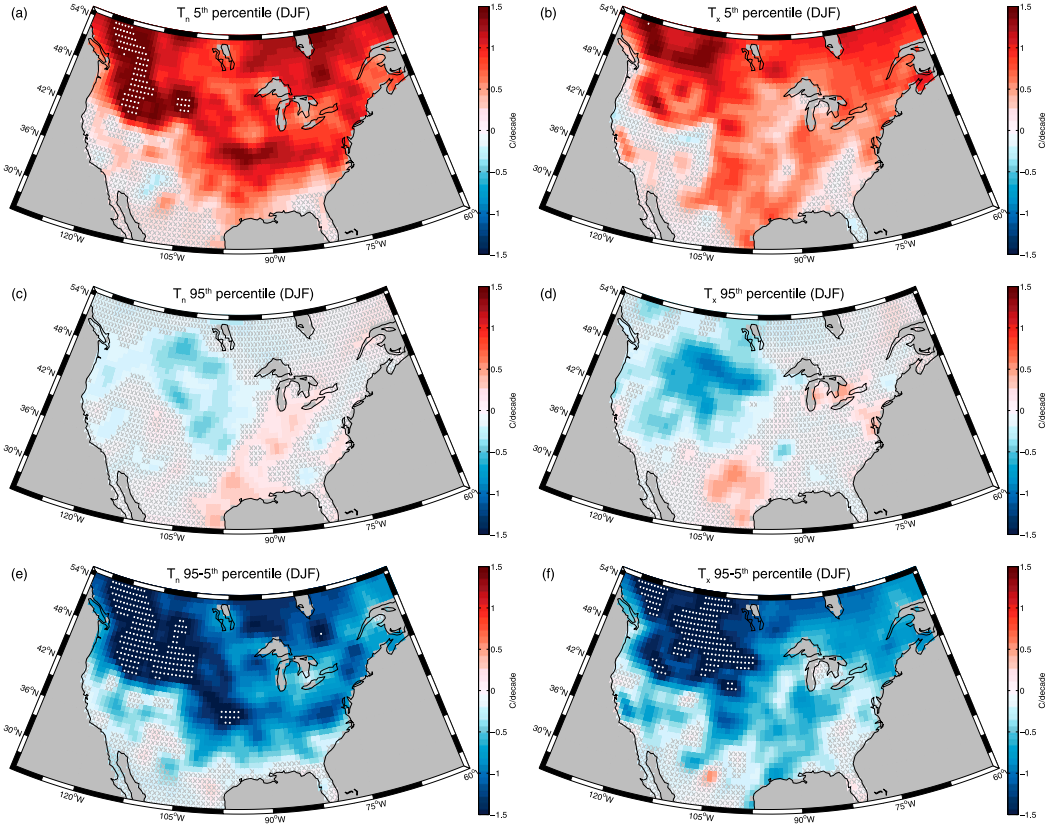

FIG. 3. Trends (1979–2014) in winter (left) T_n and (right) T_x (a),(b) 5th, (c),(d) 95th, and (e),(f) 95th − 5th percentile for the GHCN-D station data. Broad warming is seen in the lower half of the distribution, particularly at high latitudes. The pattern changes gradually with increasing percentile to one of cooling in the central and mountain United States. Stippling indicates locations where the trends exceed the color scale.

effect on the spread of the distribution is insignificant for T_x, and a weak expansion of the distribution for T_n—in contrast to most other seasons, in which on average the distribution has contracted.

Summer months have a generally increasing trend in T_n across all percentiles that is strongest in the lower tail of the distribution (Fig. 2c, and Table S1 in the supplemental material). Specifically, while T_n shows broad patterns of warming across all percentiles (Figs. 2c and 5a,c), the pattern of 95th minus 5th percentile trends is uniformly negative (Fig. 5e). Despite the reduction of distributional spread, the general warming of summer T_n is also of particular importance because of the impact of nighttime temperature minima on heat stress (Luber and McGeehin 2008). The pattern for T_x differs in two key respects. First, larger trends relative to the median are evident in the upper tail (Fig. 5d—similar but weaker trend pattern for the median not shown). Second, the pattern at the lower tail of the distribution is weaker and less similar to the median in its spatial pattern, leading to a 95th minus 5th pattern with less large-scale spatial structure.

During fall we obtain distinct patterns for T_n and T_x (Fig. 6) that are not evident in the spatially averaged case (Figs. 2d,h). Trends in T_n have a similar pattern across all percentiles with warming in the west and throughout Canada, leading to little large-scale structure in the 95th minus 5th percentile trends. Meanwhile, T_x trends are nearly spatially orthogonal between the 95th (having warming isolated to Canada) and the 5th (having warming concentrated through most of the western and central United States). The result is a widespread contraction of the distribution in the Southwest and Great Plains, and weaker expansion of the distribution in isolated parts of southern Canada and the eastern United States.

4. Comparison of trends from surface observations and reanalyses

a. Pattern correlations

We compare the estimated quantile trends from each of the datasets—the smoothed GHCN-D surface observations, and the ERA-Interim and NCEP-2

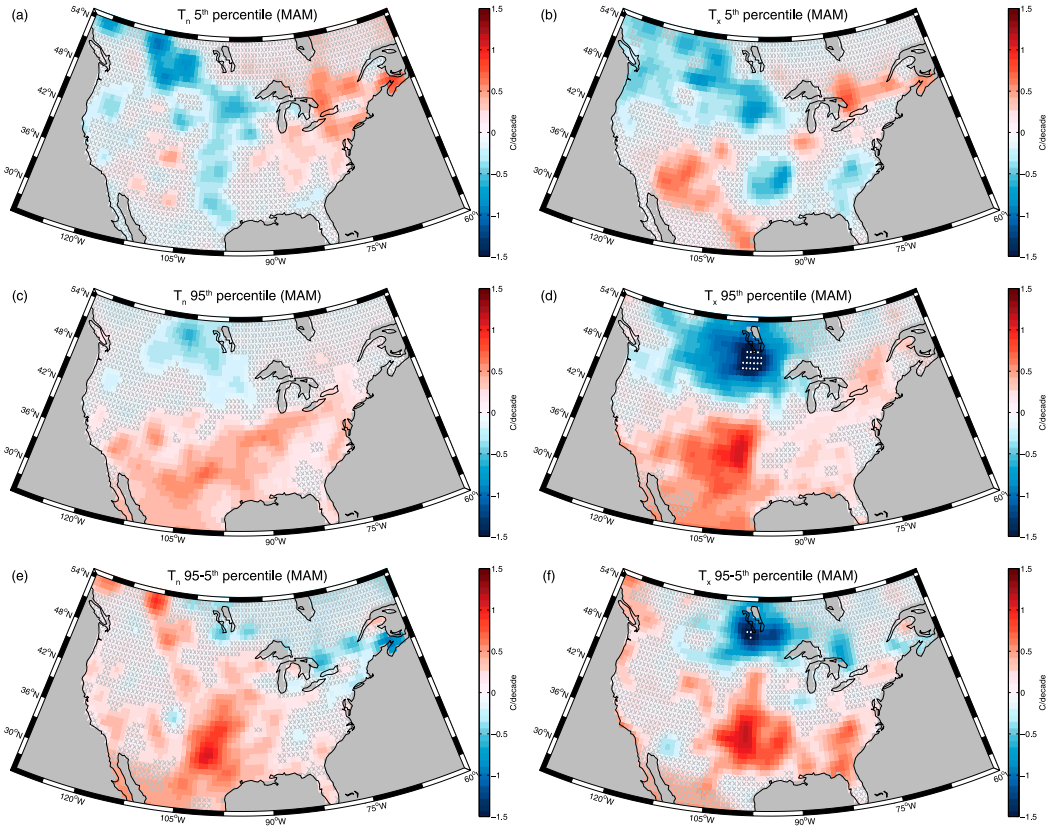

FIG. 4. As in Fig. 3, but for spring. The color scale, hatching, and stippling are as in Figs. 1–3.

reanalyses—by computing area-weighted pattern correlations (i.e., centered anomaly correlations where the spatial mean is removed prior to analysis; Von Storch and Zwiers 1999) and root-mean-squared errors (RMSE) for each pair having the same spatial resolution. These results are summarized in Figs. 7 and 8 and Tables S2–S4 in the supplemental material.

The reanalysis fields are in some cases broadly similar to the GHCN-D results, while in others there are substantial discrepancies. Notably, the reanalysis patterns are in all cases detectably different from those of the GHCN-D results (Fig. 7). Even when pattern correlations are high, differing magnitudes or pattern means also often lead to large absolute differences that are not always captured by pattern correlations alone (Fig. 8). In winter, T_x (Fig. 7e) trend patterns are relatively similar across percentiles in the different datasets. Meanwhile, winter T_n has distributional tails that are not as accurately represented; discrepancies here are particularly evident with NCEP-2, for which the 95th percentile has patches of warming on the order of 1°C decade^{-1} centered on Colorado, New Mexico, and east of Hudson Bay that are not present in the station data. The relative skill in T_x may stem from its dependence on midtropospheric temperatures that are diurnally mixed into the boundary layer (Misra et al. 2012; McNider et al. 2012). In contrast, T_n may be dominated by nocturnal cooling that is sensitive to stability very close to the surface and thus may not be resolved in the reanalyses. ERA-Interim and NCEP-2 also perform poorly during summer, again perhaps due to unresolved small-scale processes involving the interface between the land surface and atmosphere (Alapaty et al. 2001). The limited agreement between observations and reanalyses during summer is in agreement with Cornes and Jones (2013), who found that within Europe the largest discrepancies between ERA-Interim and station time series were for trends in summer 90th percentile temperatures.

We also provide a tabulation comparing spatial-mean trends of the GHCN-D-derived fields, ERA-Interim, and NCEP-2 for several quantiles and quantile differences (Tables S2 and S3 in the supplemental material).

b. *Limitations of reanalysis surface fields*

Near-surface observations can provide a powerful constraint on the planetary boundary layer of data assimilation systems (Alapaty et al. 2001; Hacker and

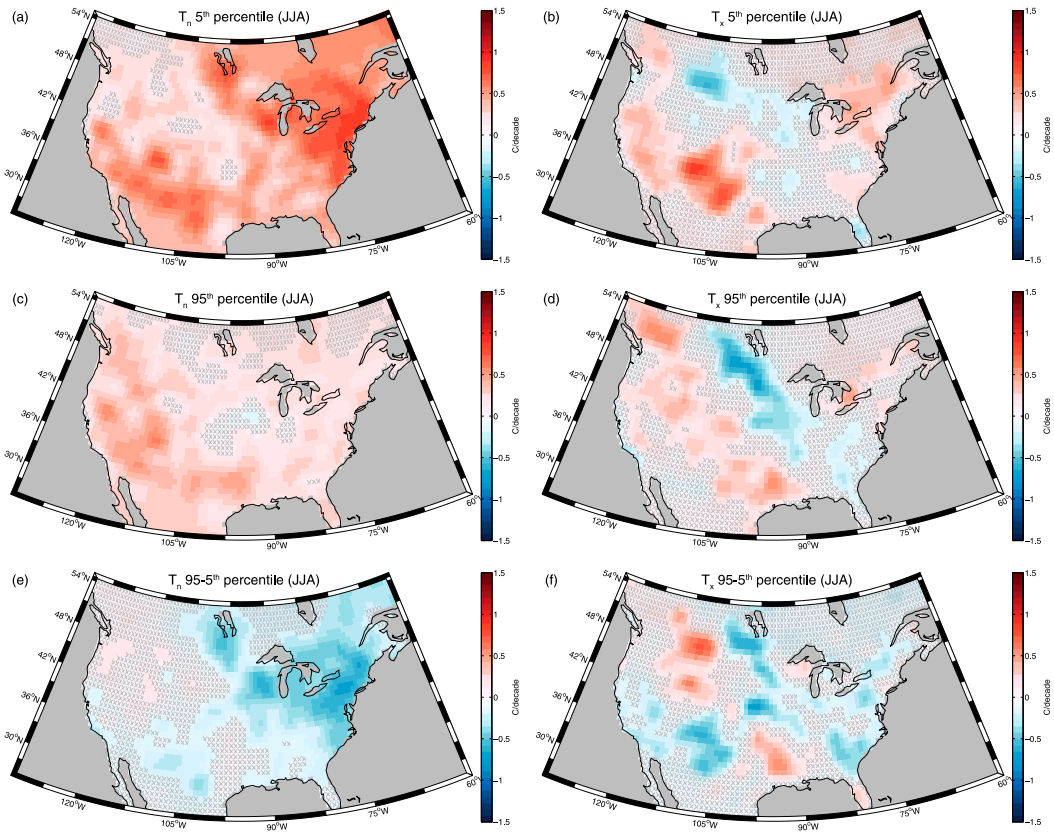

FIG. 5. As in Fig. 3, but for summer. For T_n, broad warming is seen at all percentiles but is dominated by the lower tail of the distribution as a consequence of increasing skew. For T_x, the pattern visible near the median strengthens toward the upper tail of the distribution, showing enhanced high extremes in a large swath of the west and southwest, while central Canada, the Corn Belt, and the southeast have all seen cooling of the hottest days. Color scale and hatching are as in Fig. 1.

Snyder 2005); however, surface temperature observations are typically not directly assimilated because they can render variational assimilation systems unstable, particularly in regions of variable topography or when small-scale convection is important (e.g., Stauffer and Seaman 1990). In contrast, surface pressure observations integrate over the atmospheric column and can produce a reliable estimate of the atmospheric state even when no other data sources are used (Compo et al. 2011). That NCEP-2 does not assimilate surface temperature observations is known to have adversely affected its representation of interannual variability and trends of monthly average near-surface temperatures relative to observations (Kalnay and Cai 2003). ERA-Interim uses an optimal interpolation procedure, also implemented in ERA-40 (Simmons et al. 2004), to bias correct the near-surface temperature and relative humidity by examining surface observations within 300 m of the model's surface elevation. These fields are also used to drive an offline land surface model, but ultimately have no impact on the state estimate of the atmosphere. The optimal interpolation procedure excludes any surface observations that differ from the model state at nearby grid points by more than a fixed threshold, leading to only ~40% of surface observations being used during a given time step (ECMWF 2007). This tendency to exclude not only outliers but fully ~60% of observations suggests that extremes may be represented particularly poorly, especially when they are influenced by small-scale land surface or boundary layer processes.

In comparisons between ERA-Interim and GHCN-D results we find evidence of biases resulting from these postprocessing procedures. For example, a sharp discontinuity along the Colorado–Utah border appears in many season–percentile pairs. Examination of 2-m temperature time series from the region shows a large shift in ERA-Interim around 1992 that is not identifiable in nearby GHCN-D station observations, suggesting that the variational bias corrections applied to

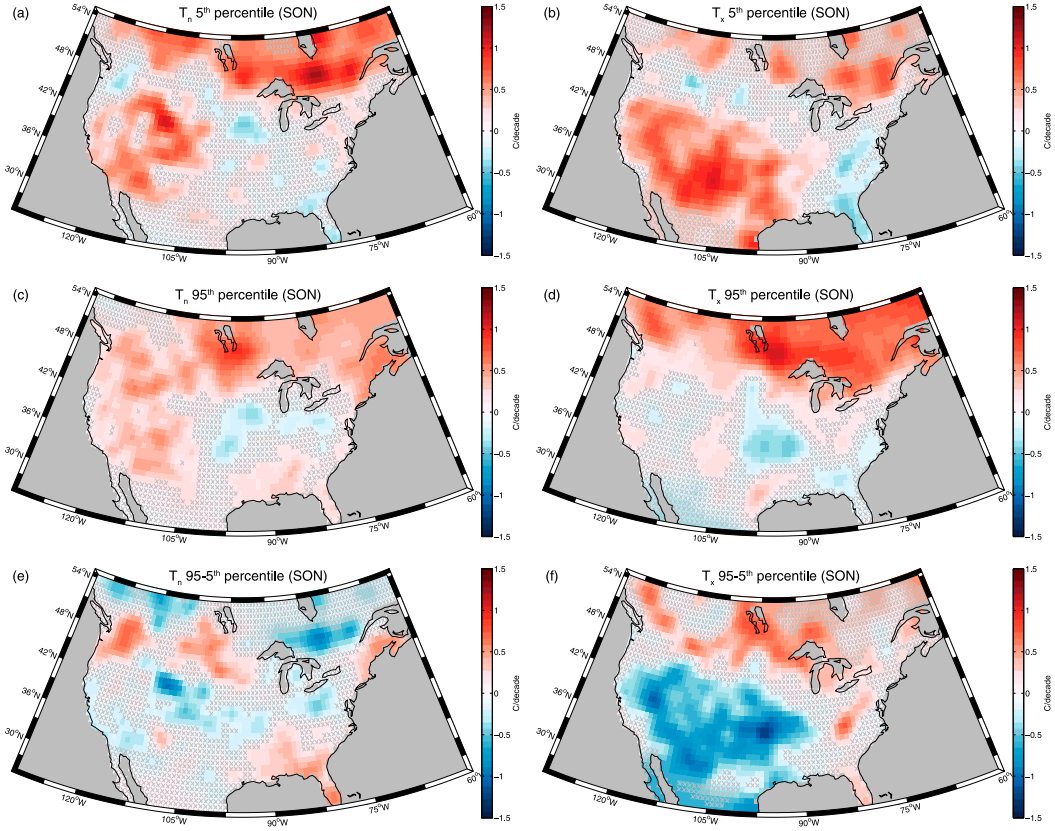

FIG. 6. As in Fig. 3, but for fall. The color scale, hatching, and stippling are as in Figs. 1–3.

the satellite and radiosonde data may have led to some important surface observations being newly included or excluded beginning at that time. The presence of discontinuities in ERA-Interim around 1992 has been previously noted in precipitation fields and may result from changes in the SSM/I constellation (Dee et al. 2011). Though not as persistent across seasons and percentiles, similar features appear elsewhere and may point to further assimilation, bias correction, or post-processing issues. Interestingly, these sharp features are generally less prominent in difference plots (e.g., 95th − 5th), suggesting that such breakpoints may affect the mean without substantially altering the shape or spread of the distribution.

That the same sharp features are not seen as often in the NCEP-2 maps is not entirely surprising: when ERA-Interim is regridded to the lower 2.5° × 2.5° resolution of NCEP-2, the discontinuities are less visually obvious. Additionally, the lack of a separate postprocessing step in NCEP may reduce the tendency for spurious sharp spatial gradients in trends to occur, except where ongoing issues in assimilated satellite snow-cover observations have directly impacted surface conditions (Kanamitsu et al. 2002).

5. Discussion

Previous analyses of variability changes have come to conflicting conclusions (Alexander and Perkins 2013), and the results presented in the foregoing section are intended to unambiguously assess changes in the spread of the temperature distribution in the densely sampled North American region. We found small contractions in the temperature distribution during summer and fall, a small increase during spring, and a large, spatially extensive decrease in variability during winter. With respect to annual extremes, summer trends consistent with an increase in the mean imply that hot summer events in midlatitudes may not be as sensitive as cold winter events to feedbacks on mean warming.

We assess that the spatial mean of the trend toward decreasing winter variability is statistically significant in the GHCN-D data and both reanalysis datasets, with consistent spatial patterns and magnitudes between all three datasets. Pointwise variability trends are significantly negative over 84% of the domain for the station data. These changes are at least partly consistent with the physical argument that the coldest winter days should warm faster than the warmest days as a result of

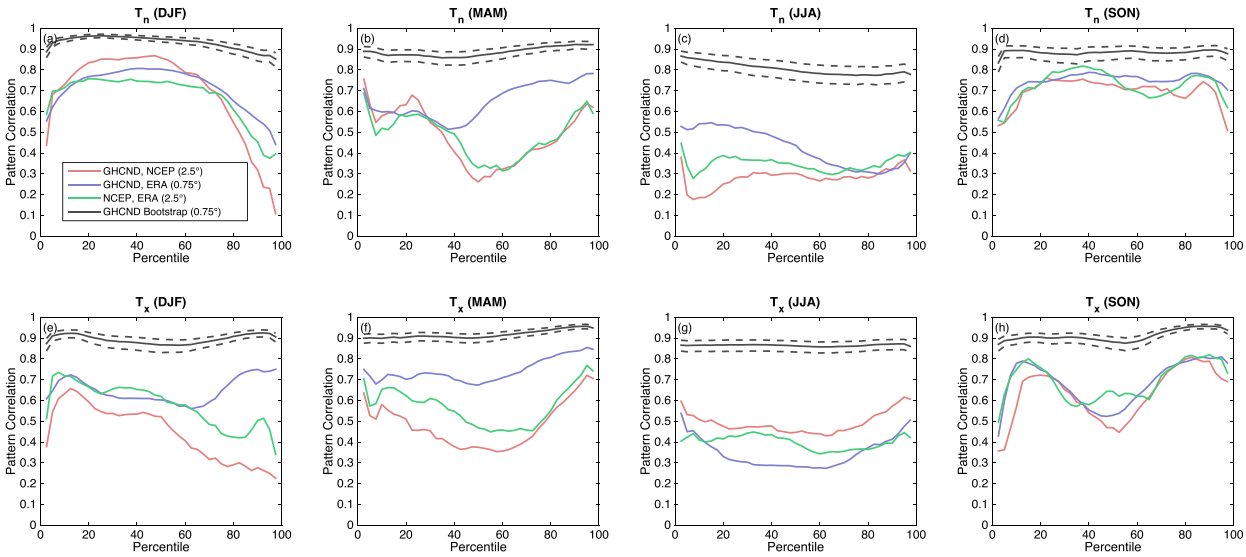

FIG. 7. Daily (a)–(d) T_n and (e)–(h) T_x pattern correlations of quantile trend maps for several combinations of reanalysis and station observations at different resolutions, all for the 1979–2014 interval. Comparisons are performed for one season in each panel, with lines indicating the area-weighted pattern correlation between the two fields. A surrogate measure of uncertainty is indicated in gray, where pattern correlations are computed for the smoothed GHCN-D-based estimate against the same 1000 parametric bootstrap realizations used to estimate the uncertainty in the spatial mean (Fig. 2), and where the dashed lines indicate the central 95% bootstrap confidence interval. That the gray line significantly exceeds the colored lines in all cases indicates that the reanalysis patterns still generally contain significant differences when accounting for sampling uncertainties of the station data.

Arctic amplification (Screen 2014; Schneider et al. 2015; Holmes et al. 2016).

A widespread, statistically significant signal of reduced winter variability associated with warming of the coldest days has not been previously identified in observations for this region. However, its magnitude is consistent with results using the variance of ERA-Interim temperatures, averaged zonally in decadal bins over land as in Screen (2014). Examining DJF ERA-Interim T_n over the North American land region from 42° to 55°N (covering the peak of the cooling pattern in Fig. 3), we do find a spatial mean variance reduction of 9.8°C^2 for 1997–2014 relative to 1979–96, similar to the zonally averaged values of Screen (2014) at those

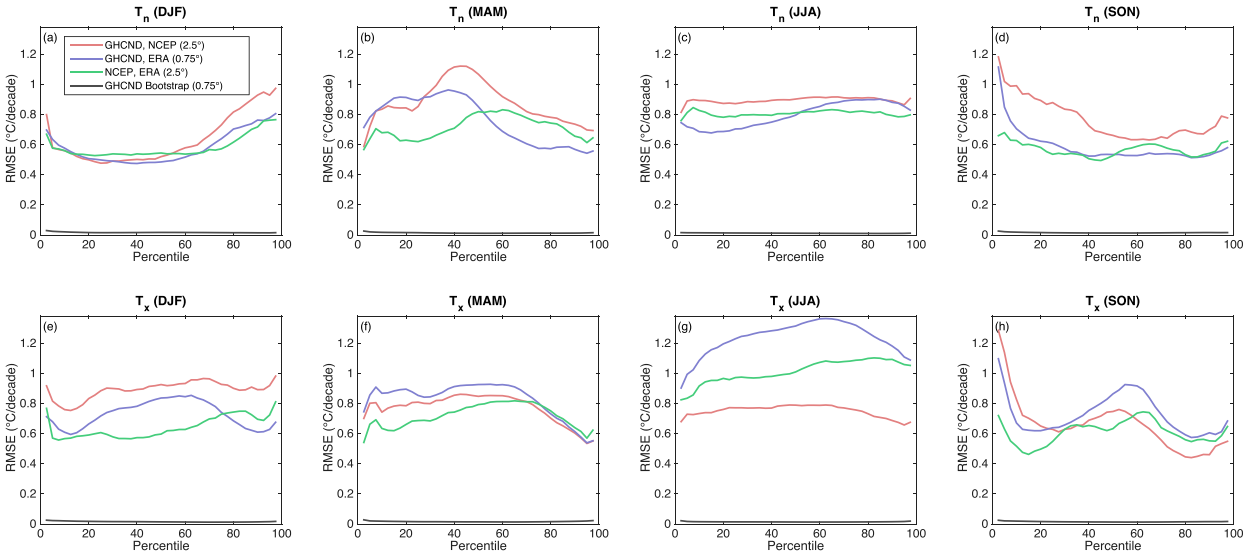

FIG. 8. As in Fig. 7, but for the RMSE of trend maps between different datasets. Here, the gray parametric bootstrap line indicates the baseline uncertainty of the observations.

latitudes and equivalent to a 9.7%, or a 0.7°C reduction in standard deviation. That we find a statistically significant signal can be attributed to the limited extent of our domain, the separate spatial model, and the use of quantile regression rather than ordinary least squares applied to binned sample variances. Significance in this context is not a statement of anthropogenic attribution, but rather that there is low-frequency structure inconsistent with annual samples that are independent of one another. Though the 1979–2014 interval is long compared with the time scales of established modes of interannual variability, we cannot exclude that it plays some role.

Although polar amplification in the absence of mean warming should lead to both northerly and southerly winds being important for reducing variability, Screen (2014) has noted that reduced cold-air advection by northerly winds would lead to an asymmetry in the trends for the upper and lower tails of the distribution. The fact that changing cold-air advection dominates the overall changes in variability can be inferred more directly using QR, where we see strong warming in the 5th percentile giving way to weak trends at the 95th percentile (Figs. 2a,e). Because mixing length scales are on the order of 1000 km in midlatitude (Schneider et al. 2015)—and perhaps longer over continental regions where equilibration with the surface is generally slower than over the ocean (Swanson and Pierrehumbert 1997)—trends in these distant source regions can influence changes in variability. Yet other processes may be necessary to fully explain the magnitude and spatial pattern of the trends; the strongest amplification of warming is in the eastern sector of Canada, while the western part of the domain has experienced the greatest declines in variability. Given that cold events in the west are often associated with northerly or northwesterly advection, and rarely with northeasterly advection, it appears unlikely that these specific patterns can be explained by advection of the basic state alone. Furthermore, the rate of warming of median ERA-Interim 850-mb temperatures is between ~0.2° and 0.7°C decade^{-1} in most of the Canadian Arctic, smaller than the peak or even the mean rate of 5th percentile warming in the 25°–55°N domain (2° and 0.8°C decade^{-1}, respectively).

Interestingly, the future emergence of a significant midlatitude signal in the future was anticipated by Screen (2014) in a complementary analysis of the CMIP5 model ensemble, with a spatial structure similar to our results appearing by the 2070–99 interval using the representative concentration pathway 8.5 (RCP8.5) forcing scenario (their Fig. S8d). Despite methodological differences leading to contrasts in significance assessment, the similarity suggests the pattern may be a robust response to anthropogenic forcing. Inasmuch as moist or radiative processes are necessary to explain the full signal of distributional contraction in the historical period, it will be important to determine whether global climate models adequately resolve these effects (e.g., Pithan et al. 2014) when considering their applicability to extremes in multidecadal projections.

6. Conclusions

We present a framework for assessing observed changes in the distribution and variability of near-surface temperatures and apply it in an analysis of daily T_n and T_x for the 25°–55°N region of North America, resolving significant seasonal distributional trends during the 1979–2014 period. By using a spatial model that employs thin-plate spline regression we also provide a mapping of station-level results from GHCN-D onto latitude–longitude grids to facilitate analysis and intercomparison with other data sources, while retaining meaningful uncertainty estimates that represent the combined errors in station-level trends and those due to their variable spatial coverage. Changes in the distribution are found to depend strongly on season and can differ between T_n and T_x, indicating that it is useful to resolve spatial, seasonal, and diurnal dependence separately. The spatial signals we identify are generally substantially larger than the uncertainties in their estimation; however, we also show that collapsing the spatial information by averaging over the domain—or, similarly, the distributional information by assessing only changes in the mean—can lead to the perception of insignificant trends purely as a result of cancellations. Examples of this sensitivity to averaging include the shoulder seasons of spring and fall, for which small spatially averaged trends contrast with large-amplitude meridional dipoles that reflect strong regional trends. Summer has seen large-scale warming, particularly in the lower tail of the distribution for T_n, while changes in variability are also predominantly regional and appear to occur on smaller spatial scales than those of spring and fall.

The winter temperature distribution has contracted significantly over most of the North American region. We provide evidence that the physical mechanisms driving the contraction are distributionally dependent, with the reduction in variability dominated by warming of cold days relative to other parts of the distribution. This distributional dependence is consistent with cold days relying on northerly advection from regions that have experienced greater rates of warming (Screen 2014; Schneider et al. 2015). However, several factors may lead to deviations from predictions of the advective

model on local scales. Cold-air advection in the western United States is predominantly continental, while warm-air advection is sourced from more maritime regions, implying that the orientation of the relevant directional temperature gradient will differ between warm and cold conditions. Differences in the strength of coupling at the lower boundary between maritime and continental conditions (Swanson and Pierrehumbert 1997) would also lead to a difference in effective length scales, and potentially to an asymmetric sensitivity to moist feedbacks. Whether it is necessary to consider higher-order moments of advection statistics or fundamentally different physical processes related to radiation, clouds, or boundary layer dynamics will be explored in future work.

We also compared all trends from the station data with the NCEP-2 and ERA-Interim reanalyses, finding strengths and weaknesses in each that depend on season, percentile, and location. While the general pattern of reduced variability during winter is found in both of the reanalysis datasets, discrepancies at small spatial scales suggest that time-varying biases in the ERA-Interim data assimilation system may lead to abrupt changes in the subset of surface observations blended with its output during postprocessing. The seasonal differences in ERA-Interim's representation of surface trends also mirrors Cornes and Jones (2013), who found that extreme-based indices are least accurately reproduced for hot days during summer. Nevertheless, the decoupling between ERA-Interim's assimilation of upper-air observations and its inclusion of surface observations during postprocessing suggests that it may still be reliably used to examine tropospheric conditions associated with surface temperature extremes inferred from the GHCN-D observations.

Our use of 3-month seasons is most appropriate for summer and winter, for which seasonal curvature is small and the observations are most closely exchangeable. The large-scale trend dipoles evident during spring and fall may be partly a consequence of long-term variability or trends in circulation during the seasonal transition. While it is possible to use shorter seasons, this choice would further reduce sample sizes; extensions of QR such as quantile periodograms (Li 2012) may lead to a more robust assessment of the role of seasonality.

Further analysis will be necessary to determine the extent to which the reduced winter variability results from anthropogenic forcing. Decadal variability associated with, for example, the Pacific decadal oscillation (Mantua and Hare 2002) or the Pacific–North America pattern (Wallace and Gutzler 1981) cannot be ruled out as playing a role in generating this pattern because of the relatively short 1979–2014 interval; trends in tropically forced teleconnections (Ding et al. 2014) are also potentially important, given the low-frequency variability associated with ENSO teleconnections during North American winter. However, inasmuch as anthropogenic forcing leads to Arctic amplification, it appears likely that further warming will result in additional contraction of winter temperature variability in midlatitudes. These relative influences of internal versus externally forced variability can be assessed by performing a similar analysis using climate model ensembles. Other natural extensions of this work include expanding the analysis to regions with less dense observation networks, and the use of a nonlinear basis for QR that would permit continuous resolution of distributional changes over longer time periods for which linear trends are less applicable.

Acknowledgments. The authors thank three anonymous reviewers and the editor for helpful comments. This study was supported by NSF Grant AGS-1304309. AR acknowledges support from James S. McDonnell Foundation Grant 220020421. Computations were performed on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University.

APPENDIX

Quantile Regression: Synthetic Examples

Several key differences between QR and more traditional techniques such as ordinary least squares regression on the mean can be elucidated using the following synthetic examples.

Example 1: A synthetic realization is generated as unitless Gaussian white noise $\mathcal{N}(at, \sigma^2)$ with a weak linear temporal trend a equaling $0.026\,\mathrm{yr}^{-1}$ and constant unitless variance of 10^2. The length of the interval is 36 years, with 90 samples per year, approximating the sampling characteristics for each GHCN-D station when analyzed for a specific season. QR slopes are then estimated for the original data; after rounding all observations to the nearest integer; and after then adding independent $\mathcal{U}(-0.5, 0.5)$ draws of jitter to each rounded observation (Fig. A1a). When this procedure is repeated for 1000 realizations, a two-sample Kolmogorov–Smirnov test comparing the distribution of slopes from the original and rounded data clearly rejects the null hypothesis that the samples follow the same distribution, with a p value less than 0.001. However, the use of jitter to restore the distribution produces a distribution of slopes that is

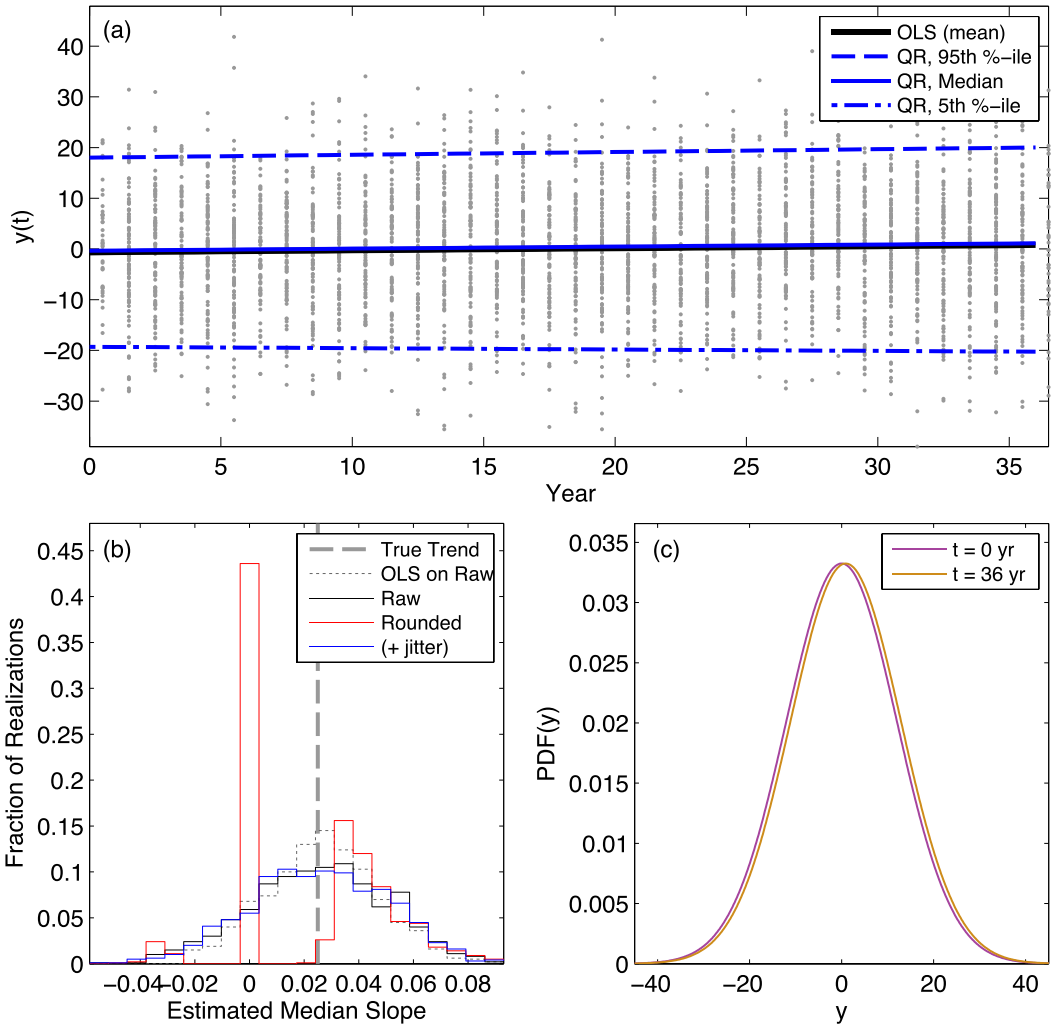

FIG. A1. Example of quantile and OLS regression on normally distributed synthetic data with and without discrete rounding. (a) Time series with 90 samples per year for 36 years are drawn from a normal distribution with a linear trend in the mean, and QR estimates of the median, 5th, and 95th percentiles are shown alongside the OLS estimate of the mean trend for one realization's raw values. The same procedure is then repeated for 1000 realizations, and the data are first rounded and then approximately restored through the addition of uniform jitter. (b) Histograms of the QR trend estimates of the median are shown for the raw, rounded, and jitter-restored realizations, illustrating how the severe bias present when data are rounded can be corrected through jittering. (c) The initial and final distributions.

indistinguishable from that of the original observations (Fig. A1b), with the Kolmogorov–Smirnov test p value of 0.69 giving no indication that the distributions of the inferred trends are different. One might venture that the zero-inflation in trend estimates of the rounded data would simply yield some metric of significance; however, the expected value of the rounded slope estimates is biased, in this case equaling $0.019\,\mathrm{yr}^{-1}$, in comparison with $0.025\,\mathrm{yr}^{-1}$ for the unrounded or jittered cases using either QR or OLS. In this case, OLS and QR are essentially equivalent as all central moments are fixed with only the mean of the distribution changing.

Example 2: The same procedure as in example 1 is repeated, but for a nonnormal time series having a trend in the shape and width of the distribution. Samples are generated from a generalized extreme value distribution with shape parameter ξ, scale parameter σ, and location parameter μ, varying linearly in time such that the distribution shifts from positive to slightly negative skew while shifting in mean toward increasing values, such that the 95th percentile is expected to remain approximately

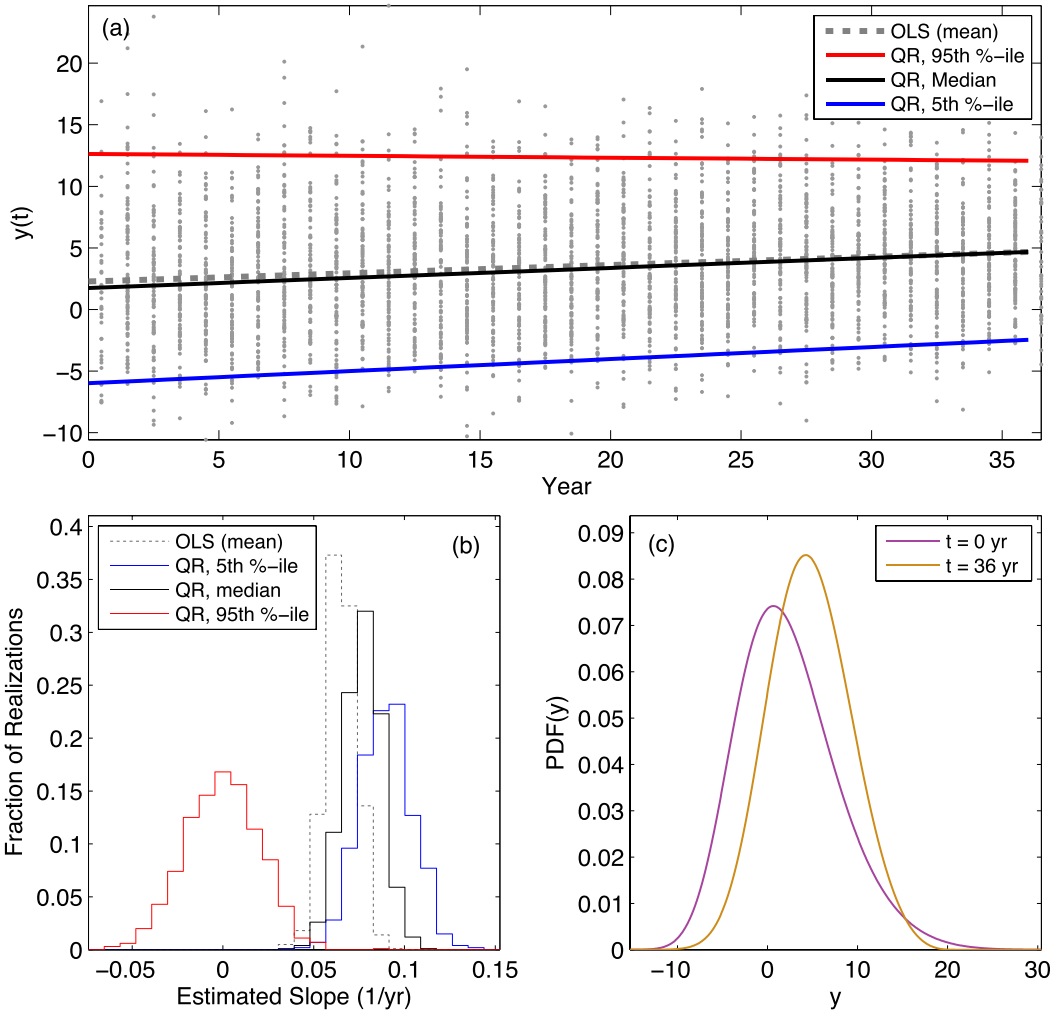

FIG. A2. Example of quantile and OLS regression on nonnormal data. (a) Time series with 90 samples per year for 36 years are drawn from a generalized extreme value distribution with a linear trend in the location, shape, and scale parameters, and QR estimates of the median, 5th, and 95th percentiles are shown alongside the OLS estimate of the mean trend for one realization's raw values. The same procedure is then repeated for 1000 realizations, and (b) histograms of the trend estimates in each percentile and for the OLS mean illustrate how distributional changes are not captured by OLS or any one percentile. (c) The initial and final distributions.

constant with time. The spread in the distribution is on average decreasing as the center of the distribution increases, with OLS failing to capture this distributional dependence (Fig. A2).

In Fig. A3, we provide examples using a wider variety of distributional changes and using alternative estimators. Three different methods are used:

1) Differences of block variances (Fig. A3, left center), similar to Hansen et al. (2012) but with the addition of explicit trend estimates. The first and second half of the time series are assumed to be two independent samples, and trends in the sample mean and sample variance are inferred directly from these using differences.
2) Maximum likelihood estimates of trends in the mean and variance (Fig. A3, right center). This method also implicitly assumes that the data are normally distributed with mean and variance that can change in time. This method is similar to estimates based on trends in moving block variances (e.g., Screen 2014; Huntingford et al. 2013). The maximum likelihood parameter estimates have a small sample-size-dependent bias for which we do not apply a correction here as it has only a minor influence on the inference in these examples.
3) Quantile regression (Koenker and Hallock 2001) (Fig. A3, right) with a linear slope and intercept term being estimated for each percentile.

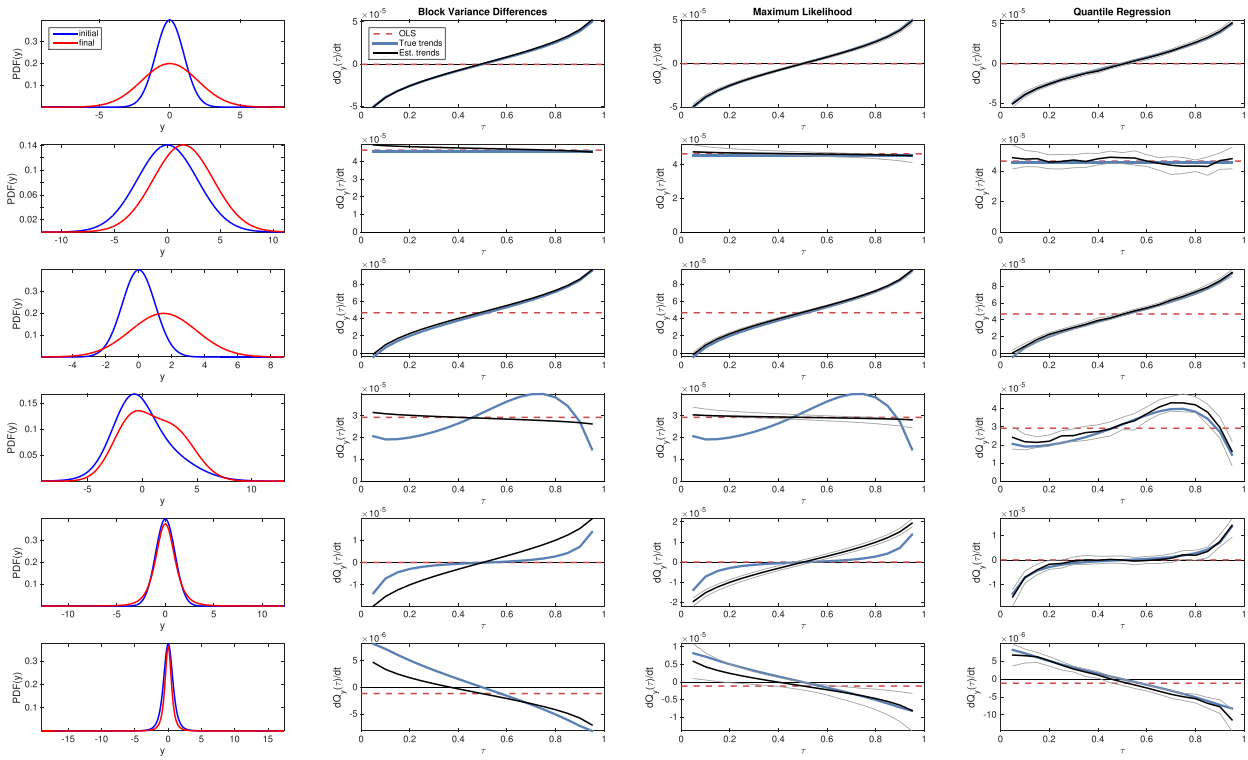

FIG. A3. Synthetic examples of inferred distributional change. (top)–(bottom) Random realizations are drawn from a parametric distribution whose parameters vary linearly with time (left; blue indicates the initial distribution and red indicates the final distribution). Distributional trends are estimated using one of three methods: (left center) block variance, (right center) maximum likelihood with linear mean and variance trends, and (right) quantile regression. In each case, the true quantile trends (blue) are compared with the estimator (black) and the results when using an ordinary least squares trend in the mean only (red dashed line). Approximate 95% confidence intervals are shown by thin gray lines for maximum likelihood and quantile regression.

Approximate 95% bootstrap confidence intervals are estimated for maximum likelihood and quantile regression, with the samples assumed to be conditionally independent given time t. This differs from the treatment used for temperature observations, where dependence necessitates the use of a block bootstrap.

For each example, a single realization of 2000 observations is generated according to a prescribed parametric distribution with time-dependent parameters. The first three distributional examples are normally distributed with imposed linearly varying mean, variance, and both mean and variance, respectively. The fourth example is a Gaussian mixture with two components and parameters such that the variance is constant in time. The fifth example is a t distribution, and the sixth is a standardized t distribution such that the variance is identically constant despite changes in higher-order moments.

In each case quantile regression robustly recovers the underlying distributional trends. While block variance differences and maximum likelihood estimates assuming normally distributed data perform relatively well when the data are indeed normal, biases inherent to maximum likelihood estimation are still evident when using smaller sample sizes. Block variances also alias temporal trends as variability, leading to sensitivities to whether detrending is first applied to the full time series, to individual blocks, or not at all (Rhines and Huybers 2013). Bias corrections and a model selection procedure to distinguish would improve the robustness of the maximum likelihood procedure, albeit at the cost of added model complexity. Furthermore, the two parametric methods perform poorly when higher-order moments are present (Fig. A3, rows 4–6), even in the relatively simple case of a two-component Gaussian mixture (Fig. A3, row 4). Quantile regression trades a small amount of efficiency in the normal case—evident in the wider confidence intervals—for robustness to nonnormality.

REFERENCES

Alapaty, K., N. L. Seaman, D. S. Niyogi, and A. F. Hanna, 2001: Assimilating surface data to improve the accuracy of atmospheric boundary layer simulations. *J. Appl. Meteor.*, **40**, 2068–2082, doi:10.1175/1520-0450(2001)040<2068:ASDTIT>2.0.CO;2.

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.

——, and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05109, doi:10.1029/2005JD006290.

Barbosa, S., M. Scotto, and A. Alonso, 2011: Summarising changes in air temperature over central Europe by quantile regression and clustering. *Nat. Hazards Earth Syst. Sci.*, **11**, 3227–3233, doi:10.5194/nhess-11-3227-2011.

Bassett, G. W., Jr., M.-Y. S. Tam, and K. Knight, 2002: Quantile models and estimators for data analysis. *Metrika*, **55**, 17–26, doi:10.1007/s001840200183.

Bengtsson, L., S. Hagemann, and K. I. Hodges, 2004: Can climate trends be calculated from reanalysis data? *J. Geophys. Res.*, **109**, D11111, doi:10.1029/2004JD004536.

Bracegirdle, T. J., and G. J. Marshall, 2012: The reliability of Antarctic tropospheric pressure and temperature in the latest global reanalyses. *J. Climate*, **25**, 7138–7146, doi:10.1175/JCLI-D-11-00685.1.

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.

Cade, B. S., and B. R. Noon, 2003: A gentle introduction to quantile regression for ecologists. *Front. Ecol. Environ.*, **1**, 412–420, doi:10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2.

Caesar, J., L. Alexander, and R. Vose, 2006: Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *J. Geophys. Res.*, **111**, D05101, doi:10.1029/2005JD006280.

Cavanaugh, N. R., and S. S. Shen, 2014: Northern Hemisphere climatology and trends of statistical moments documented from GHCN-Daily surface air temperature station data from 1950 to 2010. *J. Climate*, **27**, 5396–5410, doi:10.1175/JCLI-D-13-00470.1.

——, and ——, 2015: The effects of gridding algorithms on the statistical moments and their trends of daily surface air temperature. *J. Climate*, **28**, 9188–9205, doi:10.1175/JCLI-D-14-00668.1.

Chernozhukov, V., 2005: Extremal quantile regression. *Ann. Stat.*, **33**, 806–839, doi:10.1214/009053604000001165.

Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, doi:10.1002/qj.776.

Cornes, R. C., and P. D. Jones, 2013: How well does the ERA-Interim reanalysis replicate trends in extremes of surface temperature across Europe? *J. Geophys. Res. Atmos.*, **118**, 10 262–10 276, doi:10.1002/jgrd.50799.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

Ding, Q., J. M. Wallace, D. S. Battisti, E. J. Steig, A. J. Gallant, H.-J. Kim, and L. Geng, 2014: Tropical forcing of the recent rapid Arctic warming in northeastern Canada and Greenland. *Nature*, **509**, 209–212, doi:10.1038/nature13260.

Director, H., and L. Bornn, 2015: Connecting point-level and gridded moments in the analysis of climate data. *J. Climate*, **28**, 3496–3510, doi:10.1175/JCLI-D-14-00571.1.

Donat, M. G., J. Sillmann, S. Wild, L. V. Alexander, T. Lippmann, and F. W. Zwiers, 2014: Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *J. Climate*, **27**, 5019–5035, doi:10.1175/JCLI-D-13-00405.1.

ECMWF, 2007: Cy31r1—Operational implementation 12 September 2006. Part IV: Physical processes. IFS Documentation, 155 pp. [Available online at http://www.ecmwf.int/sites/default/files/elibrary/2007/9221-part-iv-physical-processes.pdf.]

Furrer, R., M. Genton, and D. Nychka, 2006: Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Stat.*, **15**, 502–523, doi:10.1198/106186006X132178.

——, D. Nychka, and S. Sain, 2013: Fields: Tools for spatial data, version 6.8. R package. [Available online at http://www.image.ucar.edu/Software/Fields.]

Hacker, J. P., and C. Snyder, 2005: Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized PBL. *Mon. Wea. Rev.*, **133**, 3260–3275, doi:10.1175/MWR3022.1.

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.

Hanson, C., and Coauthors, 2007: Modelling the impact of climate extremes: An overview of the MICE project. *Climatic Change*, **81**, 163–177, doi:10.1007/s10584-006-9230-3.

Hawkins, E., and R. Sutton, 2016: Connecting climate model projections of global temperature change with the real world. *Bull. Amer. Meteor. Soc.*, **97**, 963–980, doi:10.1175/BAMS-D-14-00154.1.

Haylock, M., N. Hofstra, A. Klein Tank, E. Klok, P. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res.*, **113**, D20119, doi:10.1029/2008JD010201.

Hofstra, N., M. New, and C. McSweeney, 2010: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dyn.*, **35**, 841–858, doi:10.1007/s00382-009-0698-1.

Holmes, C. R., T. Woollings, E. Hawkins, and H. de Vries, 2016: Robust future changes in temperature variability under greenhouse gas forcing and the relationship with thermal advection. *J. Climate*, **29**, 2221–2236, doi:10.1175/JCLI-D-14-00735.1.

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., K. McKinnon, A. Rhines, and M. Tingley, 2014: U.S. daily temperatures: The meaning of extremes in the context of nonnormality. *J. Climate*, **27**, 7368–7384, doi:10.1175/JCLI-D-14-00216.1.

Ihaka, R., and R. Gentleman, 1996: R: A language for data analysis and graphics. *J. Comput. Graph. Stat.*, **5**, 299–314.

Kalnay, E., and M. Cai, 2003: Impact of urbanization and land-use change on climate. *Nature*, **423**, 528–531, doi:10.1038/nature01675.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643, doi:10.1175/BAMS-83-11-1631.

Katz, R. W., and B. G. Brown, 1992: Extreme events in a changing climate: variability is more important than averages. *Climatic Change*, **21**, 289–302, doi:10.1007/BF00139728.

Kennedy, J., N. Rayner, R. Smith, D. Parker, and M. Saunby, 2011: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.*, **116**, D14104, doi:10.1029/2010JD015220.

Knight, K., 1998: Limiting distributions for L_1 regression estimators under general conditions. *Ann. Stat.*, **26**, 755–770, doi:10.1214/aos/1028144858.

Koenker, R., 2014: Quantile regression. [Available online at http://www.econ.uiuc.edu/~roger/research/rq/rq.html.]

——, and G. Bassett Jr., 1978: Regression quantiles. *Econometrica*, **46**, 33–50, doi:10.2307/1913643.

——, and K. Hallock, 2001: Quantile regression. *J. Econ. Perspect.*, **15**, 143–156, doi:10.1257/jep.15.4.143.

Lee, K., H.-J. Baek, and C. Cho, 2013: Analysis of changes in extreme temperatures using quantile regression. *Asia-Pac. J. Atmos. Sci.*, **49**, 313–323, doi:10.1007/s13143-013-0030-1.

Li, T.-H., 2012: Quantile periodograms. *J. Amer. Stat. Assoc.*, **107**, 765–776, doi:10.1080/01621459.2012.682815.

Luber, G., and M. McGeehin, 2008: Climate change and extreme heat events. *Amer. J. Prev. Med.*, **35**, 429–435, doi:10.1016/j.amepre.2008.08.021.

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

Mannshardt, E., P. Craigmile, and M. Tingley, 2013: Statistical modeling of extreme value behavior in North American tree-ring density series. *Climatic Change*, **117**, 843–858, doi:10.1007/s10584-012-0575-5.

Mantua, N. J., and S. R. Hare, 2002: The Pacific decadal oscillation. *J. Oceanogr.*, **58**, 35–44, doi:10.1023/A:1015820616384.

Matiu, M., D. P. Ankerst, and A. Menzel, 2016: Asymmetric trends in seasonal temperature variability in instrumental records from ten stations in Switzerland, Germany and the UK from 1864 to 2012. *Int. J. Climatol.*, **36**, 13–27, doi:10.1002/joc.4326.

McKinnon, K. A., A. Rhines, M. P. Tingley, and P. Huybers, 2016: The changing shape of Northern Hemisphere summer temperature distributions. *J. Geophys. Res. Atmos.*, **121**, 8849–8868, doi:10.1002/2016JD025292.

McNider, R., and Coauthors, 2012: Response and sensitivity of the nocturnal boundary layer over land to added longwave radiative forcing. *J. Geophys. Res.*, **117**, D14106, doi:10.1029/2012JD017578.

Meehl, G., C. Tebaldi, G. Walton, D. Easterling, and L. McDaniel, 2009: Relative increase of record high maximum temperatures compared to record low minimum temperatures in the U.S. *Geophys. Res. Lett.*, **36**, L23701, doi:10.1029/2009GL040736.

Menne, M. J., and C. N. Williams Jr., 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700–1717, doi:10.1175/2008JCLI2263.1.

——, 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.

Misra, V., J.-P. Michael, R. Boyles, E. Chassignet, M. Griffin, and J. O'Brien, 2012: Reconciling the spatial distribution of the surface temperature trends in the southeastern United States. *J. Climate*, **25**, 3610–3618, doi:10.1175/JCLI-D-11-00170.1.

Pithan, F., B. Medeiros, and T. Mauritsen, 2014: Mixed-phase clouds cause climate model biases in Arctic wintertime temperature inversions. *Climate Dyn.*, **43**, 289–303, doi:10.1007/s00382-013-1964-9.

Proistosescu, C., A. Rhines, and P. Huybers, 2016: Identification and interpretation of nonnormality in atmospheric time series. *Geophys. Res. Lett.*, **43**, 5425–5434, doi:10.1002/2016GL068880.

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.

Reich, B. J., and L. B. Smith, 2013: Bayesian quantile regression for censored data. *Biometrics*, **69**, 651–660, doi:10.1111/biom.12053.

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.

——, M. P. Tingley, K. A. McKinnon, and P. Huybers, 2015: Decoding the precision of historical temperature observations. *Quart. J. Roy. Meteor. Soc.*, **141**, 2923–2933, doi:10.1002/qj.2612.

Robeson, S. M., C. J. Willmott, and P. D. Jones, 2014: Trends in hemispheric warm and cold anomalies. *Geophys. Res. Lett.*, **41**, 9065–9071, doi:10.1002/2014GL062323.

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.

Sardeshmukh, P. D., and P. Sura, 2009: Reconciling non-Gaussian climate statistics with linear dynamics. *J. Climate*, **22**, 1193–1207, doi:10.1175/2008JCLI2358.1.

Schneider, T., T. Bischoff, and H. Potka, 2015: Physics of changes in synoptic midlatitude temperature variability. *J. Climate*, **28**, 2312–2331, doi:10.1175/JCLI-D-14-00632.1.

Screen, J. A., 2014: Arctic amplification decreases temperature variance in northern mid- to high-latitudes. *Nat. Climate Change*, **4**, 577–582, doi:10.1038/nclimate2268.

Simmons, A. J., and Coauthors, 2004: Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *J. Geophys. Res.*, **109**, D24115, doi:10.1029/2004JD005306.

Simolo, C., M. Brunetti, M. Maugeri, and T. Nanni, 2011: Evolution of extreme temperatures in a warming climate. *Geophys. Res. Lett.*, **38**, L16701, doi:10.1029/2011GL048437.

——, ——, ——, and ——, 2012: Extreme summer temperatures in Western Europe. *Adv. Sci. Res.*, **8**, 5–9, doi:10.5194/asr-8-5-2012.

Simpson, D. G., R. J. Carroll, and D. Ruppert, 1987: *M*-estimation for discrete data: Asymptotic distribution theory and implications. *Ann. Stat.*, **15**, 657–669, doi:10.1214/aos/1176350367.

Stauffer, D. R., and N. L. Seaman, 1990: Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. *Mon. Wea. Rev.*, **118**, 1250–1277, doi:10.1175/1520-0493(1990)118<1250:UOFDDA>2.0.CO;2.

Swanson, K. L., and R. T. Pierrehumbert, 1997: Lower-tropospheric heat transport in the Pacific storm track. *J. Atmos. Sci.*, **54**, 1533–1543, doi:10.1175/1520-0469(1997)054<1533:LTHTIT>2.0.CO;2.

Tingley, M., 2012: A Bayesian ANOVA scheme for calculating climate anomalies, with applications to the instrumental temperature record. *J. Climate*, **25**, 777–791, doi:10.1175/JCLI-D-11-00008.1.

Von Storch, H., and F. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812, doi:10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

Wang, K., 2014: Sampling biases in datasets of historical mean air temperature over land. *Nat. Sci. Rep.*, **4**, 4637, doi:10.1038/srep04637.

Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 851–870, doi:10.1002/wcc.147.