A Comparison of Daily Temperature-Averaging Methods: Spatial Variability and Recent Change for the CONUS

JASE BERNHARDT
Department of Geology, Environment and Sustainability, Hofstra University, Hempstead, New York

ANDREW M. CARLETON
Department of Geography, and Polar Center, The Pennsylvania State University, University Park, Pennsylvania

CHRIS LAMAGNA
Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 13 February 2017, in final form 27 October 2017)

ABSTRACT

Traditionally, the daily average air temperature at a weather station is computed by taking the mean of two values, the maximum temperature (Tmax) and the minimum temperature (Tmin), over a 24-h period. These values form the basis for numerous studies of long-term climatologies (e.g., 30-yr normals) and recent temperature trends and changes. However, many first-order weather stations—such as those at airports—also record hourly temperature data. Using an average of the 24 hourly temperature readings to compute daily average temperature has been shown to provide a more precise and representative estimate of a given day’s temperature. This study assesses the spatial variability of the differences in these two methods of daily temperature averaging [i.e., (Tmax + Tmin)/2; average of 24 hourly temperature values] for 215 first-order weather stations across the conterminous United States (CONUS) over the 30-yr period 1981–2010. A statistically significant difference is shown between the two methods, as well as consistent overestimation of temperature by the traditional method [(Tmax + Tmin)/2], particularly in southern and coastal portions of the CONUS. The explanation for the long-term difference between the two methods is the underlying assumption for the twice-daily method that the diurnal curve of temperature is symmetrical. Moreover, this paper demonstrates a spatially coherent pattern in the difference compared to the most recent part of the temperature record (2001–15). The spatial and temporal differences shown have implications for assessments of the physical factors influencing the diurnal temperature curve, as well as the exact magnitude of contemporary climate change.

1. Introduction

a. Background

Essential to the accurate assessment of contemporary climate change involving the near-surface (i.e., 2 m) air temperature is the reliability of the observational database and the method used to derive the mean temperature. Traditionally, this daily average mean air temperature is computed from two values: the maximum temperature (Tmax) and the minimum temperature (Tmin) for a given 24-h period. This method of daily temperature averaging has been used to depict the short-term weather conditions and longer-term climatology of a given location; it comprises the 30-yr climate “normal” recommended by the World Meteorological Organization (World Meteorological Organization 2011). Further, the traditional method is used to derive other temperature-based climate indices, such as heating degree days, cooling degree days, and growing degree days, important for determining residential heating and cooling needs, and agricultural activity, along with their spatiotemporal variations. Long-term temperature climatologies (i.e.,
30-yr normals) in the United States are derived from data collected at 262 first-order weather stations (Applequist et al. 2012), primarily concentrated in and around medium and large cities or metropolitan areas. Despite acquiring temperature data at hourly intervals, these weather stations utilize the traditional method of daily temperature averaging, due to observing practices that date to the start of organized record-keeping in the United States during the nineteenth century (e.g., Conner 2006; Fiebrich 2009). Nevertheless, first-order stations likely should utilize the 24 hourly temperature measurements (e.g., Zeng and Wang 2012; Barry and Blanken 2016, 11–12), hereafter referred to as the hourly method of temperature averaging, to determine daily average temperature, given the following principles:

1) A greater number of data points are being considered (24 vs 2).

2) The daily temperature curve does not typically follow a normal (i.e., bell-curve type) distribution, an important assumption made by the traditional method, but is instead skewed toward the right (i.e., Tmin) on average.

3) This skew of the daily temperature curve varies spatially, seasonally, and temporally, with recent variability potentially being enhanced by anthropogenic climate change (Shiu et al. 2009).

4) The hourly method lessens or eradicates artificial effects on the temperature from the “time of observation bias,” for example, the difference between determining Tmax, Tmin, and thus the average daily temperature from a midnight-to-midnight period versus a morning-to-morning period. In the United States, Tmax, Tmin, and the average daily temperature are determined from midnight local time to midnight local time, occasionally introducing a “reset” high or low temperature, which is not representative of the true Tmax or Tmin from the diurnal curve. (McKinnon et al. 2016; Karl et al. 1988; Schaal and Dale 1977)

To help motivate such a shift in temperature-averaging methodology, it is necessary to determine the differences between the two methods of temperature averaging and to quantify the spatial and seasonal variations in these differences. These are the objectives of the present study.

b. Motivation

A number of studies have compared the two methods of daily temperature averaging for either single stations or localized areas. For example, Gough and He (2015) compared both methods for Churchill, Manitoba, located adjacent to Hudson Bay. The authors selected Churchill owing to its unique microclimate: fog is a frequent occurrence during the Bay’s ice-free season, typically the summer and autumn months. There was a monthly averaged difference between the methods of up to 1.5°C, with strong seasonality exhibited. The traditional method overestimated temperature during the late spring through early autumn and underestimated temperature during the cold-season months, compared to the hourly method, and showed a close correspondence with fog frequency. Earlier, Weiss and Hays (2005) analyzed the difference between the two daily temperature-averaging methods at seven stations in different climatological subregions across the conterminous United States (CONUS), for periods of 6 to 10 years. Average differences between the traditional and hourly methods ranged from −0.52°C in Bishop, California, to 0.43°C in Tampa, Florida. The large negative difference in Bishop was attributed to the large diurnal temperature range (DTR) of that station, greater than at any of the other stations studied.

Other studies of the differences between temperature-averaging methods include those that have relied on data that utilize both observed and modeled data to interpolate areas having no observations. Wang and Zeng (2013) developed an hourly climatology of 2-m air temperature from four reanalysis datasets, the Modern-Era Retrospective Analysis for Research and Applications (MERRA), 40-yr ECMWF Re-Analysis (ERA-40), ECMWF interim reanalysis (ERA-Interim), and Climate Research Unit Time Series, version 3.10 (CRU TS3.10), using linear and temporal interpolation to create a full diurnal curve from 3- and 6-hourly data. They found that the twice-daily averaging of surface data generally overestimated daily average temperature compared to hourly averaging of the adjusted reanalysis datasets. Moreover, Wang and Zeng (2014) blended 2-m air temperature data from two reanalysis datasets, CRU TS3.10 and MERRA. These authors found that for the period 1979–2009, the hourly averaging approach slightly overestimated the daily averaged temperature compared to the traditional method, and those differences hinted at a latitude variation; that is, there were greater differences in areas at and poleward of 40 degrees latitude and also seasonally (greatest in July) in the Northern Hemisphere. At the global scale, Li et al. (2016) compared the hourly and traditional methods on monthly time scales for the years 2000–13, to derive a bias correction for stations lacking hourly data. A linear-regression model explained the difference between the two averaging methods as a function of day length, DTR (determined from hourly measurements), and the twice-daily averaged temperature
(i.e., traditional method). Unlike the Wang and Zeng (2013) study, Li et al. (2016) determined that, globally, the traditional method overestimates station temperature compared to the hourly method by between 0° and 1°C.

Similar to Li et al.’s (2016) study, except using observational data, Wang (2014) analyzed approximately 5600 weather stations globally from the NCDC Integrated Surface Database and found an average difference between the two temperature-averaging methods of 0.2°C, with the traditional method overestimating the hourly average temperature. The study also determined that this difference was largest during the cold season, much like that of Li et al. (2016). Wang (2014) attributed these differences in temperature averages to two factors: an asymmetry in the daily temperature curve, which results in a systematic bias, and a more random sampling bias caused by twice-daily observations that undersample weather events such as frontal passages, which can greatly influence the temperature during a given 24-h period. None of the foregoing studies, however, analyzed the differences between methods on sub-seasonal (i.e., monthly) scales, whereas such an analysis could reveal the climatic variables influencing those patterns. Moreover, only Wang and Zeng (2013, 2014) considered an entire 30-yr period commensurate with that used to obtain climatic normals, yet their studies utilized reanalysis temperature data.

The time of day when observations are taken also influences the differences between daily temperature-averaging methods. Schaal and Dale (1977) determined that the time of observation led to bias when computing the daily average temperature at Indianapolis International Airport. They concluded that a warming trend up to that date could simply be attributed to a change in observation time from local morning to local midnight. To further clarify the time-of-observation bias, Karl et al. (1988) developed a linear regression model that could be applied to any location across the CONUS to determine the magnitude of this bias, based on latitude, time of observation, and longitudinal location within the time zone. Together, these studies revealed that the traditional method of daily temperature averaging results in a bias compared to the “true” average daily temperature (i.e., that determined as the 24-hourly average).

Despite the insights provided by previous research, there has not been an attempt to utilize solely observational data to compare temperature-averaging methods across a large (subcontinental) region for assessing the monthly patterns of the differences and the seasonal progression of the associated spatial patterns. Such an assessment for a region also containing a relatively dense network of weather stations across a range of climatic zones and for which the data are quality controlled, such as the CONUS, is highly desirable. Accordingly, the present research utilizes hourly air temperature observations from 215 first-order CONUS weather stations for the most recent 30-yr climate normals period (1981–2010). From determining the daily difference between the two temperature-averaging methods at those stations, we present the spatial and seasonal patterns of the differences along with a single-station case study to highlight certain physical climate factors (i.e., land use/land cover, air masses) influencing the bias in temperature between methods. Last, we assess the most recent change in the difference between these temperature-averaging methods, by comparing the 1981–2010 climate normals period with that characterizing the most recent period of 2001–15, to determine—in a preliminary way—if the ongoing climatic warming may have influenced the bias.

2. Data and methods

Hourly temperature data from the full complement of 262 first-order weather stations across the CONUS were originally selected for study. These stations, typically located at airports, all have periods of record of several decades or longer and are professionally maintained and routinely quality controlled (Applequist et al. 2012; Fig. 1). Stations from this initial dataset were eliminated where at least 5% of all hourly observations were missing. Thus, the final dataset comprises 215 CONUS first-order weather stations having complete records. To match the latest decennial U.S. climate normals period, we used daily temperature data at these 215 stations for 1981–2010. All data were obtained from the Northeast Regional Climate Center’s CLIMOD data system (Northeast Regional Climate Center 2016;
To assess the difference in means generated using the two temperature-averaging methods [i.e., \((T_{\text{max}} + T_{\text{min}})/2\) for 24 hourly observations] two climatologies were generated for each station, one using each method. Then, the station climatology generated using the hourly averaging method was subtracted from that of the traditional method, to determine the difference. Those differences are presented at the monthly scale and also averaged for the whole year. The unit of temperature used for our analysis and throughout this study was degrees Fahrenheit, because it is the original unit of surface temperature observation in the United States.

The hourly temperature data were also utilized to determine the average (1981–2010) shape of the diurnal (i.e., daily) temperature curve at each station. We determine the interquartile range (IQR) of the 24 hourly measurements by subtracting the value of the 16th largest hourly observation from the value of the 8th largest. Moreover, the skewness of the average daily temperature curve at each station was determined. The skewness indicates the amount of asymmetry about the mean distribution of the average 24 hourly temperature readings at a station. Skewness was computed for each station using the Yule–Kendall index (Wilks 1995), which compares the spread between the second quartile (i.e., median) and third quartile with the spread between the first and second quartiles. Thus, a positive Yule–Kendall index indicates right-skewed data (i.e., more hours spent near \(T_{\text{max}}\) than \(T_{\text{min}}\)), negative indicates left-skewed data (i.e., more hours spent near \(T_{\text{max}}\) than \(T_{\text{min}}\)), and zero implies a symmetrical daily temperature curve. To further quantify the symmetry—or lack thereof—in the daily temperature curve, we determined the number of hours spent in the four “quarters” (Q1–Q4) of each station’s temperature distribution. The boundaries of these quarters are as follows: one at the mean of the 24 hourly temperature values, one halfway between the mean and the minimum of the 24 readings, and another between the mean and the maximum of the 24 readings.

To depict the spatial patterns (i.e., on maps) of the differences between the two methods of temperature averaging and the asymmetry in the average daily temperature curve, the station-level differences—both annually and separately for each month—were interpolated using ordinary kriging, which accounts for spatial variation at the regional level when predicting values at unknown points over a larger area (Burrough 1986). This technique has been used successfully in many studies analyzing the spatial variations of climate variables, including temperature and precipitation (e.g., Aalto et al. 2013; Wu and Li 2013; Verdin et al. 2016).

### 3. Results and discussion

**a. Station-average daily temperature curve**

The descriptive statistics (i.e., mean, median, standard deviation, and extreme values) resulting from the analysis of the daily temperature curves for the 215 stations are given in Table 1. A number of statistical and spatial patterns emerge from this data analysis regarding the shape of the average daily temperature curve. In particular, the Yule–Kendall index is positive for all 215 CONUS stations, indicating that the daily temperature curve is right skewed; that is, more hourly observations occur near the bottom of the distribution of hourly temperatures (i.e., around \(T_{\text{min}}\)) than near the top of the distribution (around \(T_{\text{max}}\)). For example, Figs. 2a and 2b are a histogram and box plot, respectively, of the daily temperature distribution (1981–2010) at Albany, New York, showing a rightward skew in the data (Yule–Kendall index = 0.14). Figure 2c shows the daily temperature curve for Albany with the boundaries of the four quarters superimposed.

We used ordinary kriging in the present study to estimate the values at unknown points (i.e., locations between stations) based on the values at surrounding stations and the spatial relationships among them.

<table>
<thead>
<tr>
<th>Hours spent in quarter</th>
<th>Yule–Kendall skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Q2</td>
</tr>
<tr>
<td>Mean</td>
<td>8.54</td>
</tr>
<tr>
<td>Median</td>
<td>9.00</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.68</td>
</tr>
<tr>
<td>Maximum</td>
<td>10.00</td>
</tr>
<tr>
<td>Minimum</td>
<td>4.00</td>
</tr>
</tbody>
</table>

Table 1. Descriptive statistics of the annually averaged, daily temperature curve averaged for all 215 stations.
afternoon thunderstorms, especially during the long warm season, which accompany increased convective cloud cover, precipitation, and evaporative cooling of the air. These events cause Tmax to occur earlier in the day, followed by only gradual cooling overnight due to increased amounts of atmospheric humidity (e.g., at New Orleans; Figs. 3c,d; Yule–Kendall = 0.30). Conversely, the lowest skewness values annually (i.e., exhibiting a minimally rightward skewed daily temperature curve) occur in the CONUS’s Desert Southwest (e.g., Ely, Nevada; Figs. 3e,f; Yule–Kendall = 0.22). These lower skewness values are due to a greater DTR, as the interquartile range of the daily temperature distribution serves as the denominator in the Yule–Kendall formula.

There is a distinct regionalization to the length of time the daily temperature curve spends in each quarter of the distribution. On average, for all 215 first-order stations averaged together, the daily temperature curve spends the most time in the area between Tmin and halfway between Tmin and the mean, spending a mean amount of time of 8.55 h and median amount of time of 9 h. The relative dominance of the time spent in this portion of the distribution is consistent with the right-skewness of the daily temperature curve; the temperature spends more time near its minimum value, on average, than near the maximum (i.e., the upper part of the range, between Tmax and halfway to the mean; mean amount of time spent = 7.30 h and median = 7 h). Further, more time is spent
on average in the second quarter of the distribution (i.e., between the mean and halfway between the mean and Tmin) than the third quarter (between the mean and halfway between the mean and Tmax), also consistent with the rightward skew in the long-term (1981–2010) hourly average temperature data averaged across stations.

There is also spatial variation present in the time spent in each quarter of the daily temperature curve, related to both the station physical geography and CONUS time zone. The daily temperature curve spends the most hours in the first quarter of the curve at stations in the eastern Great Lakes, the central and northern High Plains, the Southeast, and the Pacific Northwest (Fig. 4a). These regions, especially the first two named, are similar to the areas of high skewness in the daily temperature curve. Stations in the southeast CONUS experience more time in the lowest quarter of their daily temperature distribution due to higher amounts of atmospheric moisture (e.g., given by the specific humidity) and the fact that moister air warms more slowly than drier air. Further, the Great Lakes, central and northern

FIG. 3. The 30-yr-averaged (1981–2010) daily temperature curve and box plot, for (a),(b) North Platte, Nebraska, (c),(d) New Orleans, Louisiana, and (e),(f) Ely, Nevada. In (a), the graph begins at 0600 local time in order to start at the average minimum hourly temperature for the day, while in (c) and (e), the graphs begin at 0500 for the same reason.
High Plains, and Pacific Northwest are located in the western portions of their respective time zones, which can allow the hourly measurements to miss the timing of Tmin by a wider margin and instead cause more hourly readings to occur in the lowest portion of the distribution. A similar result was found by Karl et al. (1986) when developing a time-of-observation bias for estimating mean monthly temperature from hourly data. The average duration that the daily temperature curve spends in the second quarter of the distribution (Fig. 4b) demonstrates the inverse amount of time spent in the first quarter. For example, stations in the northern High Plains switch from having the most time spent in the first quarter to the least time spent in the second quarter. This result is logical, as more time spent in the bottom quarter would imply less time spent in the next quarter of the distribution. These results expand upon previous work suggesting that the daily temperature curve varies spatially across the CONUS (e.g., Qu et al. 2014). Moreover, our results suggest the underlying physical factors (e.g., cloud cover, atmospheric humidity) contributing to these variations, quantify the variations by using the Yule–Kendall index, and determine the number of hours spent in each quarter of the average daily temperature distribution.

b. Monthly, seasonally, and annually averaged differences between the two daily temperature-averaging methods

To identify seasonal variations in the difference between temperature-averaging methods, those differences were mapped for all stations at monthly time scales. Table 2 shows descriptive and spatial statistics of the difference between the two temperature-averaging methods for all stations combined over each month of the 30-yr period of study, as well as the annual averages.
(i.e., averages of all 12 months) of those differences. When considering the annual average of the differences, there is large variation across the dataset [standard deviation (SD) = 0.38°F]. The mean difference between temperature-averaging methods for all 215 stations combined is 0.16°F, with a median of 0.10°F. However, these differences range from a maximum of 1.6°F at Santa Maria, California, to a minimum of −0.9°F—the traditional method underestimates daily average temperature compared to the hourly method—at Winnemucca, Nevada. There is an absolute difference of at least 0.1°F at 190 of the 215 stations. Moreover, the traditional method overestimates the daily average temperature at 134 stations, underestimates it at 76 stations (35.4%), and shows no difference at only 5 stations (2.3%).

Spatial patterns in the annually averaged differences between the two temperature-averaging methods are readily apparent (Fig. 5a). Stations having the largest positive differences are located in the Southeast, Gulf Coast, coastal California, and southern and central Great Plains (i.e., Kansas, Oklahoma, and the Texas Panhandle). Stations with the largest negative differences between methods occur in the Great Basin (i.e., Nevada), the upper Midwest, and New England. These station difference values are significantly spatially autocorrelated with a Moran’s I score = 0.440 (indicative of a Z score = 12.4 and p value < 0.001) on a scale of −1.0 (dispersed) to +1.0 (clustered). This confirms that there is a spatial coherence (i.e., organization of differences by sign and magnitude over space) to the differences between the two temperature-averaging methods at regional scales.

The regions of greatest difference between the traditional and hourly temperature-averaging methods resemble previously defined climatic zones in the CONUS (e.g., Bukovsky 2011). The Bukovsky subregions are defined from climatological (i.e., temperature, precipitation, and their seasonality) and ecological data (i.e., National Ecological Observatory Network ecoregions) as distinct climatic zones in the CONUS. To most effectively analyze regional patterns in the difference between temperature-averaging methods, we group the 17 numbered Bukovsky climate subregions defined for the CONUS (Fig. 5b) into five broader regions, based on the distinct spatial patterns evident in our analysis. These broader regions are the Pacific West Coast (consisting of Bukovsky subregions 3 and 4), Intermountain West (5–9), Midwest (10, 13, and 14), South (11, 12, 16, and 17), and Northeast (15, 19, and 20). Table 3 shows the mean differences between the two methods for those five regional groupings. Regions with the greatest differences visually correspond with those previously identified from the map in Fig. 5a. Of the five regions, the Pacific West Coast and South show the

Table 2. Descriptive and spatial statistics of the mean annual and monthly differences between the two temperature-averaging methods averaged for all 215 stations. Boldface indicates a significant Z score.

<table>
<thead>
<tr>
<th></th>
<th>Annual</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>−0.1</td>
<td>−0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>−0.1</td>
</tr>
<tr>
<td>Median</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>−0.1</td>
<td>−0.1</td>
<td>0.0</td>
<td>0.2</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>−0.1</td>
</tr>
<tr>
<td>SD</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.5</td>
<td>0.7</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Max</td>
<td>1.6</td>
<td>1.4</td>
<td>1.0</td>
<td>1.0</td>
<td>1.2</td>
<td>1.5</td>
<td>1.7</td>
<td>2.2</td>
<td>2.2</td>
<td>2.0</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Min</td>
<td>−0.9</td>
<td>−1.3</td>
<td>−1.3</td>
<td>−1.1</td>
<td>−1.2</td>
<td>−1.5</td>
<td>−2.1</td>
<td>−2.3</td>
<td>−2.1</td>
<td>−1.0</td>
<td>−0.5</td>
<td>−0.9</td>
<td>−1.5</td>
</tr>
<tr>
<td>Moran’s I</td>
<td>0.4</td>
<td>0.6</td>
<td>0.5</td>
<td>0.2</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.6</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Z score</td>
<td>12.4</td>
<td>15.8</td>
<td>13.3</td>
<td>4.4</td>
<td>4.1</td>
<td>7.9</td>
<td>12.5</td>
<td>12.8</td>
<td>11.3</td>
<td>10.9</td>
<td>15.6</td>
<td>11.4</td>
<td>15.1</td>
</tr>
</tbody>
</table>

FIG. 5. (a) Average annual (all months of the year) difference between temperature-averaging methods [(Tmax + Tmin)/2 − 24 hourly observations], 1981–2010, with the 215 first-order weather stations used in this study superimposed on the map and (b) map of the 17 Bukovsky ecoclimatological subregions, of which 17 total subregions, 3–17 and 19–20, cover the CONUS. Image source: Prein et al. (2016, their Fig. 1b).
greatest positive differences between the two methods. Further, the Midwest and Intermountain West show slight negative differences between methods, while the Northeast shows a slight positive difference. Thus, there is a significant difference, as determined by a Mann–Whitney U test, between both the Pacific West Coast and South regions with the Midwest and Intermountain West regions. The Northeast and South regions were significantly different from all other regions. The general broad similarity between our regional variations of temperature difference and the Bukovsky climate regions implies that climate variables at the regional level (e.g., their synoptic climatology) may also influence the difference between methods.

There are clear monthly and seasonal (i.e., four groups of three adjacent months and two groups of six months) variations in the difference between the two temperature-averaging methods (Fig. 6). In winter (December, January, February) the average difference between methods is well below the annual mean difference of 0.16°F, and at a majority of CONUS stations in December (58%) the traditional method underestimates the hourly averaged temperature. In spring (March, April, May) the difference between temperature-averaging methods is also less than the annual average. Both April and May have a majority of stations underestimating the daily average temperature (67% and 55%, respectively), and April has the lowest monthly average difference of −0.09°F. Conversely, summer (June, July, August) shows a greater positive difference from the annual average between the two methods, with over 75% of stations overestimating temperature in July and August and with August having the highest monthly averaged difference between the two methods of 0.54°F. The autumn (September, October, November) also has higher than the annual average positive difference between the two methods, with over 75% of the stations overestimating temperature compared to the hourly method. Additionally (Table 2), the variability in the difference between temperature-averaging methods across stations within a given month varies substantially (i.e., standard deviation > 0.50°F), with notably higher difference during the summer months and the lowest during spring. Also, the temperature differences (both positive and negative) between the two methods show significant spatial autocorrelation in all months, with the strongest spatial autocorrelation in winter, and the weakest in spring (Table 2). These differences between methods are due to higher variability (at each individual station, spatially) in the shape of the daily temperature curve during the spring transition season. Moreover, there is a seasonal progression in the differences between the two temperature-averaging methods, evident when analyzing monthly maps of these differences. In winter, an underestimation of temperature by the twice-daily method occurs over much of the northern tier states of the CONUS, with greatest magnitude over the upper Midwest and New England (Fig. 7). Conversely, the daily average temperature is overestimated over much of the southern CONUS in this season, attaining highest magnitude across the southern Plains and coastal California. As mentioned above regarding Table 2, and as shown in Fig. 7b, the spring months show less spatial coherence.

The mean change in the two temperature-averaging methods between adjacent months can be depicted by difference maps. Thus, when considering the map showing the March differences minus those for February (Fig. 7c), the underestimation by the twice-daily averaging method across the northern CONUS decreases to almost zero, while much of the Desert Southwest and Intermountain West sees a change from overestimation to underestimation. In contrast, for March subtracted from April, nearly all the CONUS shows an underestimation by the traditional method, with the exception of coastal California and the lower Southern Plains, a spatial pattern similar to that observed during winter. In late spring (April subtracted from May), the area of temperature overestimation by the traditional method expands again, returning to the southeastern CONUS, while the magnitude of the difference increases along both the California and Gulf coasts. However, the magnitude of underestimated of the average daily temperature by the traditional method also increases in May across the Desert Southwest and Intermountain West.

<table>
<thead>
<tr>
<th>Region</th>
<th>No. of stations</th>
<th>Mean</th>
<th>Max</th>
<th>Min</th>
<th>SD</th>
<th>Significant regional diff (Tukey)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pacific West Coast</td>
<td>17</td>
<td>0.49</td>
<td>1.6</td>
<td>−0.1</td>
<td>0.45</td>
<td>Intermountain West, Midwest, Northeast</td>
</tr>
<tr>
<td>Intermountain West</td>
<td>30</td>
<td>−0.05</td>
<td>0.9</td>
<td>−0.9</td>
<td>0.35</td>
<td>Pacific West Coast, South, Northeast</td>
</tr>
<tr>
<td>Midwest</td>
<td>51</td>
<td>−0.11</td>
<td>0.5</td>
<td>−0.7</td>
<td>0.22</td>
<td>Pacific West Coast, Northeast, South</td>
</tr>
<tr>
<td>South</td>
<td>53</td>
<td>0.48</td>
<td>1.2</td>
<td>0.0</td>
<td>0.24</td>
<td>Intermountain West, Midwest, Northeast, Northeast</td>
</tr>
<tr>
<td>Northeast</td>
<td>64</td>
<td>0.11</td>
<td>0.7</td>
<td>−0.4</td>
<td>0.28</td>
<td>Pacific West Coast, Midwest, South, Intermountain West</td>
</tr>
</tbody>
</table>
These results indicate that the shape of the average daily temperature curve changes considerably during the year, while also showing spatial coherence, likely a result of shifting air mass climatologies, a phenomenon that is discussed more in the following subsection.

The expansion of the area of overestimation by the traditional temperature-averaging method continues through summer, with an increase of the differences occurring in the Southeast, along the California coast, and in the Southern Plains (Fig. 7d). By August, much of the CONUS shows the daily mean temperature overestimated by the traditional method, except for the Great Basin and upper Midwest. During autumn, two distinct spatial patterns emerge in the difference between temperature-averaging methods (Figs. 7e,f). First, the area of daily mean temperature overestimation by the traditional method retreats southward in the eastern half of the CONUS, with underestimation returning to much of the Northeast by October. In contrast, the large swath of mean temperature underestimation by the traditional method, centered on the Great Basin and persisting since April, drastically erodes in September and completely disappears in October to be replaced by mean temperature overestimation (Fig. 7g). Moreover, in November (not shown) over the northern tier states of the CONUS, a transition back to temperature underestimation occurs, similar to that observed during winter.

To compare the results of this study with the findings of Wang (2014) for comparable periods, we averaged the difference between methods for two six-month periods, a cold season (November–April) and a warm season (May–October). For the cold season, the present study determined that the traditional method underestimates temperature across much of the northern tier of the CONUS, with strongest underestimation across the upper Midwest (Fig. 8a). In contrast, there is an overestimation in most of the southern CONUS, with strongest overestimation across California, the southern Great Plains, and Florida. During the warm season, there is strong overestimation by the traditional method along the Pacific coast, southern Great Plains, and southeastern CONUS (Fig. 8b). Conversely, underestimation occurs in the Great Basin and upper Midwest, strongest in the former region. The warm season result from the present
FIG. 7. Monthly averaged differences between the two temperature-averaging methods for representative mid-season months, (a) January, (b) April, (d) July, and (f) October and (c), (e), (g) change from the previous month, with the 215 first-order weather stations used in this study superimposed on the maps.
study is similar to that found by Wang (2014), despite the different time periods used (1981–2010 vs 2004–13). The cold season results, however, differ considerably. Wang (2014) determined that during the cold season, the traditional method overestimates temperature across the entire CONUS, with strongest overestimation in the southern Great Plains. Thus, the difference between cold and warm season (i.e., cold minus warm season) also contrasts with the increased overestimation (highest in the Great Basin) during the cold season found by Wang (2014). However, the present study shows that overestimation by the traditional method only increases in the Great Basin during the cold season (Fig. 8c). Across the rest of the CONUS, the difference between methods shows a decrease in magnitude, with the underestimation increasing in the upper Midwest and Northeast, and overestimation decreasing in all other regions.

Last, a strong annual variation in the differences between the two temperature-averaging methods is also readily apparent at individual stations, with a rise or fall maintained over several consecutive months, and a sign change between adjacent seasons, such as spring to summer. Accordingly, Fig. 9a depicts spatially the standard deviation of the 12 monthly averaged temperature differences between the two averaging methods. Arid stations, especially in Nevada, Arizona, and Utah, show the largest variation in daily mean temperature difference for the year, as exemplified by Elko, Nevada (Fig. 9b).

c. Potential climatic controls on the difference between the two daily temperature-averaging methods: The example of Tucson, Arizona

Tucson, Arizona, was selected to better understand the various climatic controls—climate variables such as cloud cover, dewpoint temperature, and synoptic climatology—influencing the difference in daily mean temperature between the two averaging methods for a region having strong seasonality in the sign of those
differences. In particular, this station is representative of the marked change in climatic conditions occurring between the spring and summer associated with the North American monsoon (NAM) (Bryson and Lowry 1955; Adams and Comrie 1997). During the dry “pre-monsoon” late winter and spring months, the traditional method underestimates the daily mean temperature at Tucson relative to the hourly method, by 0.19°F, 0.75°F, 1.19°F, and 0.95°F for March, April, May, and June, respectively. The underestimation in June is similar to that observed over the remainder of the Interior West and Southwest (e.g., Fig. 10a). However, during July in Tucson and also at adjacent stations of Arizona, the traditional method of averaging shifts to overestimating the daily mean temperature, despite areas to the north and west in the Sierra Nevada continuing to underestimates it (e.g., Fig. 10b).

The change in sign of the difference between temperature-averaging methods at Tucson and nearby stations in Arizona is likely a response to the onset of the NAM in early July, which occurs as increased atmospheric humidity and frequent scattered afternoon and evening thunderstorms on the Colorado Plateau and adjacent desert and mountain regions (e.g., Carleton 1985; Carleton et al. 1990; Grantz et al. 2007). During June, the distribution of the 24 hourly temperature observations is skewed left (e.g., Fig. 10a), with a Yule–Kendall index = −0.20 (i.e., the daily temperature spends more time closer to Tmax than Tmin). Conversely, during July, the Yule–Kendall index changes to a positive number (=0.13) as the distribution is skewed right (Fig. 10b); more time is spent near Tmin than Tmax, similar to the findings for the CONUS in general in this month. Thus, the initiation and progression of the NAM over the Desert Southwest region has a clear influence on the daily temperature curve and, as a result, the difference between the two temperature-averaging methods.

Supporting the role of the NAM onset in the observed change from underestimation to overestimation of the daily mean temperature at Tucson by the two temperature-averaging methods, is the change in the mean daily frequency of the moist tropical (mT) air mass given by the spatial synoptic classification (SSC) of Sheridan (2002). The monthly average value of mT air mass for June is 8.2%. For the same month, the cloud cover percentage at Tucson, determined from a recent climatology (Wilson and Jentz 2016), is 11.7%; the average 2-m specific humidity according to North American Regional Reanalysis (NARR) data (Mesinger et al. 2006) is 5.30 g kg⁻¹; and the total precipitation, derived from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) reanalysis dataset (PRISM Climate Group 2016), is 4.83 mm. Thus, during June, the typically very dry air warms up readily, allowing more of the daily temperature distribution to occur near Tmax than Tmin. As a result, the daily temperature curve is skewed left, and the traditional method underestimates the daily average temperature. For July, the NAM onset is clearly evident from those same variables: the frequency of mT air increases to 23.4% of days, along with increases in cloud cover (39.5%), specific humidity (9.37 g kg⁻¹), and precipitation (48.77 mm). Thus, when Tucson comes under the influence of the NAM during July, the moister air means that the temperature increases more slowly in the morning, and more time is spent close to Tmin than to Tmax, allowing for a rightward-skewed daily

![Figure 10a](http://journals.ametsoc.org/jcli/article-pdf/31/3/979/4703469/jcli-d-17-0089_1.pdf)
temperature curve and an overestimation of daily average temperature by the traditional method.

d. Difference between temperature-averaging methods for the most recent period (2001–15)

In a preliminary attempt to assess the extent to which the magnitude, sign, and spatial variability of the station differences between the two temperature-averaging methods (traditional, 24-hourly) may be stable temporally, we performed a similar analysis to that for the 1981–2010 period, only updated for the most recent 15-yr period (2001–15) of generally rapid temperature increase (e.g., Hansen et al. 2012). The difference between methods was computed for the same 215 first-order weather stations and averaged annually at each station for the 15-yr period 2001–15. A shift toward underestimation by the traditional method is evident in the more recent period, with the mean and median differences between the two methods averaged for all stations being, respectively, 0.16°F and 0.10°F for 1981–2010 and 0.07°F and 0.02°F for 2001–15. Moreover, the difference between methods is underestimated at 100 stations during the recent period, compared to just 76 stations during the earlier period (1981–2010). The absolute magnitude of the difference (i.e., irrespective of sign) increased slightly during the more recent period, changing from 0.25°F to 0.32°F. These results demonstrate that a difference between temperature-averaging methods continues in the most recent period of enhanced warming, although a shift toward underestimation is present. Comparing the hourly temperature-averaging climatologies for the two periods further confirms that temperatures in general have risen across the CONUS. The hourly averaged temperature at all 215 stations increased by 0.5°F from 1981–2010 to 2001–15. At 18 stations, the hourly averaged temperature increased by at least 1.0°F, with the largest increases occurring at the arid stations of Reno, Nevada; Bishop, California; Las Vegas, Nevada; Helena, Montana; and Boise, Idaho. Conversely, a decrease in the mean temperature occurred at only nine stations, and with magnitude of 0.3°F or less.

Maps of the difference between the two temperature-averaging methods for the 2001–15 period, again utilizing ordinary kriging (Fig. 11a), display broadly similar spatial patterns to those observed for 1981–2010 (Fig. 5a), including the large area of overestimation by the traditional method in the south-central and southeast CONUS, as well as coastal California. However, a new area of underestimation occurs in the Intermountain West, Midwest, and Northeast CONUS for the most recent period. There has been a spatial shift from moderate overestimation (0°F–0.15°F) in the earlier period to slight underestimation (0°F–0.1°F) in the recent period by the traditional method in the central Great Plains, an increased underestimation in the central Rocky Mountains and Midwest and reduced overestimation in the upper South (e.g., Tennessee Valley) (Fig. 11b). The temporal shift from over- to underestimation in the Great Plains represents the largest magnitude change between the two overlapping time periods of any CONUS region. This shift coincides with modest land-use/land-cover (LULC) change in that region, specifically a continued transition from irrigated agriculture to open grassland (Sleeter et al. 2013). The Third National Climate Assessment determined that agricultural land coverage decreased by 1.6% between 1973 and 2000 (Melillo et al. 2014). Such a change in LULC contributes to a change in the distribution of daily temperature values, by reducing irrigation and thus the evaporation of water into the atmosphere. This drying pattern in the more recent period is evident from the NARR data, as the average relative humidity at all 215 stations in the dataset decreased by about 0.1% between 1981–2010 and 2001–15, from 69.6% to 69.5%. Between the same two periods, average specific humidity at
all 215 stations increased slightly from 7.64 to 7.75 g kg\(^{-1}\). These results are consistent with the overall warming of the atmosphere observed between the two periods.

An additional factor capable of introducing artificial changes over time to the difference between temperature-averaging methods involves station relocation and the introduction of new equipment. While weather stations occasionally move, nearly all first-order weather stations in the CONUS relocated during the 1990s or early 2000s due to the installation of Automated Surface Observing System (ASOS) equipment, replacing manual observation of weather variables. Several studies have documented a bias in the observation of temperature and other variables accompanying the conversion to ASOS systems (Sun et al. 2005; McKee et al. 2000; Guttman and Baker 1996), and an influence on the daily temperature curve is therefore likely as well. According to the National Centers for Environmental Information’s (NCEI) Historical Observing Metadata Repository (NCEI 2017; https://www.ncdc.noaa.gov/homr), all stations analyzed in the present research converted to ASOS equipment and simultaneously experienced at least a slight shift in location during the full climate normals period (1981–2010). The majority of stations (80.9%) experienced the switch to ASOS between 1994 and 1996, with the earliest occurring in 1992 and the latest in 2001, with a median year in 1995. Thus, the 1981–2010 period was split roughly into two when considering observation type, half ASOS and half manual observations. Conversely, all stations had changed to ASOS by the start of the second period studied, 2001–15, with the exception of two stations that installed ASOS during February and July 2001, respectively. Thus, the shift to ASOS equipment and associated station move could be one factor helping explain differences in the two temperature-averaging methods between the earlier and recent periods. For example, the aforementioned arid stations that featured changes of at least 1.0°F between the two periods experienced instrumentation moves of a considerable distance across airport grounds and, in some cases, into locations having a different LULC type. These changes are summarized in Table 4.

Moreover, in addition to the drying trend mentioned above, the conversion to ASOS during the 1981–2010 period studied may also help explain the difference in results between this study and that by Wang (2014). Wang (2014) considered hourly data from the NCEI

TABLE 4. Details of conversion to ASOS at select stations exhibiting a large change in the difference between methods from 1981–2010 to 2001–15 time periods.

<table>
<thead>
<tr>
<th>Station</th>
<th>Move date</th>
<th>Move distance</th>
<th>Old LULC</th>
<th>New LULC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Las Vegas, NV</td>
<td>Sep 1995</td>
<td>0.4 miles</td>
<td>Rooftop</td>
<td>Center field of airport</td>
</tr>
<tr>
<td>Reno, NV</td>
<td>Sep 1995</td>
<td>0.6 miles</td>
<td>Airport taxiway</td>
<td>Airport taxiway</td>
</tr>
<tr>
<td>Boise, ID</td>
<td>Dec 1995</td>
<td>1.3 miles</td>
<td>Next to buildings</td>
<td>Airport taxiway</td>
</tr>
<tr>
<td>Helena, MT</td>
<td>Nov 1998</td>
<td>1.5 miles</td>
<td>Next to buildings</td>
<td>Airport taxiway</td>
</tr>
</tbody>
</table>
integrated surface dataset, which contains stations where temperature is collected via automated observation, and which had already switched to ASOS given the period studied (2004–13). There is no consistent difference, however, in the seasonal bias of the ASOS conversion, as that is most strongly controlled by siting changes, which vary by location. Thus, further study—especially when considering entire 30-yr climatology periods—is needed to help explain why there is a considerable difference in the cold-season results between Wang (2014) and the present study but only a negligible difference for the warm season.

4. Conclusions

An analysis of the spatial and temporal (monthly, seasonal, annual) differences in mean daily temperature between the traditional \( [(T_{max} + T_{min})/2] \) and 24 hourly observations at 215 CONUS first-order weather stations sampling a wide range of climate types is undertaken for the recent climate normals period (1981–2010). Such an analysis is predicated on the need to identify any statistical biases present in near-surface air temperature data derived utilizing the traditional method that is used routinely in studies of recent climate trends and changes. Moreover, the possibility that spatial patterns of the temperature differences between the two methods may not be stable temporally is assessed in a preliminary way by comparison with the more recent period of rapid temperature increase (2001–15).

The results reveal statistically significant differences—both over- and underestimation—between the two methods of daily temperature averaging on monthly and seasonal time scales, and these differences show considerable spatial coherence. The fundamental reason for the difference between the two methods is the assumption inherent in the traditional method, of a symmetrical rise and fall of the daily temperature. However, at all of the stations studied, there is on average a rightward skew in the daily temperature curve; that is, the temperature rises more quickly in the morning than it falls in the evening and at night. There is variation in the shape of the daily temperature curve among stations, including the DTR and the number of hours spent in each quarter of the daily distribution of temperature. Our results support previous work indicating an important difference between the two methods of temperature averaging—agreeing with Wang (2014) and Li et al. (2016) that, on average, the traditional method overestimates the average daily temperature—while expanding upon Gough and He’s (2015) finding for a single station of a strong seasonality in the difference between temperature-averaging methods. Regarding the seasonality of these differences, our work agrees with Wang (2014) for the warm season but disagrees for the cold season. More particularly, mapping the differences at high spatial resolutions in the present study discloses new features of the annual and monthly climatologies of near-surface air temperature for the CONUS. On average, the traditional method overestimates the daily average temperature compared to hourly averaging by approximately 0.16°F, though there is strong spatial variability. For some stations, mainly those located in New England, the upper Midwest, and Great Basin, the traditional method underestimates temperature, while the largest overestimation of average daily temperature by the traditional method occurs in the Southeast, southern Great Plains, and along the California coast. Monthly and seasonal variations in the differences between the two temperature-averaging methods include the greatest overestimation by the traditional method occurring, on average (i.e., at all 215 stations studied), during the late summer and early fall, but with little difference, or a slight underestimation, during winter and spring. At Tucson, Arizona, this seasonality of the differences between methods manifests itself as a sign change from June to July and is most readily explained by a strong increase in the frequency of the moist tropical airmass type accompanying onset of the NAM. This assertion is supported by concomitant increases in the climatological cloud cover and precipitation amounts.

Comparing spatially the differences between the two temperature-averaging methods for the most recent climate normals period (1981–2010) with the last 15 years of the most rapid temperature increase (2001–15) reveals a shift, on average, toward underestimation by the traditional method. This result strongly suggests that the shape of the daily temperature curve is changing, such that more hours per day are spent closer to Tmin than Tmax during the more recent (2001–15) period versus the base period (1981–2010). A physical driver of this shift appears to be an overall increase in specific humidity at all 215 stations studied, aligned with the observed warming. Air with more moisture warms more slowly, thus allowing for more time to be spent closer to the Tmin than to Tmax. Of course, clarification of this possible shift requires the continued forward extension of temperature records at the 215 first-order stations used here, culminating in the next climate normals period of 1991–2020, and consideration of the impact of the changeover to ASOS data.

As part of the continuing inquiry into local-scale climate variations, it should be noted that several states and regions have implemented “mesonets” in recent years. These mesonets typically feature a dense network of well-maintained automated weather stations, which measure weather and climate variables, including temperature, at subhourly time scales (Mahmood et al. 2017).
These attributes make mesonets an ideal data source for future study of the differences between temperature-averaging methods, especially as more systems are put in place, and the length of record for existing systems increases. In particular, the use of data from multiple mesonets could more robustly address the question of whether averaging the temperature at subhourly scales significantly differs from the two methods considered here. There are numerous instances of climatic phenomena strongly influencing temperature at such time scales, including evaporative cooling from convection and frontal passages and downdrafts initiated by deep convection or terrain.

Accordingly, the importance of continuing to collect, quality control, analyze, and predict future changes in the hourly mean temperature is underscored. Anthropogenic and ecological systems are sensitive to the shape of the daily temperature curve, and given that it is evidently changing through time (e.g., the present study; Thorne et al. 2016) the change in the overall daily temperature curve, and not just the two values of Tmax and Tmin, must be considered in the context of recent and future climatic changes. In addition, temperature-based indices useful to both climate researchers and society, such as heating and cooling degree days, growing degree days, and measures of extreme temperatures (e.g., determination of “heat wave” episodes), should be calculated utilizing the hourly temperature data whenever possible. Last, future work should include improved evaluation of the implications of the present research for possible. Last, future work should include improved quality control, analyze, and predict future changes in

**Acknowledgments.** This paper represents a portion of the first author’s Ph.D. dissertation at Penn State University. We are grateful for the guidance and insight provided by dissertation committee members Drs. Brent Yarnal, Jon Nese, and Martin Tingley. Partial funding for this research was provided by the Pennsylvania Space Grant Consortium as part of its Graduate Research Fellowship Program at Penn State.

**REFERENCES**


