

A Harmonic Approach for Calculating Daily Temperature Normals Constrained by Homogenized Monthly Temperature Normals

ANTHONY ARGUEZ AND SCOTT APPLEQUIST

National Oceanic and Atmospheric Administration/National Climatic Data Center, Asheville, North Carolina

(Manuscript received 24 August 2012, in final form 5 February 2013)

ABSTRACT

NOAA released the new 1981–2010 climate normals in July 2011. These included monthly and daily normals of minimum and maximum temperature. Monthly normals were computed from monthly temperature values that were corrected for biases (i.e., homogenized) due to changes in observing practices over the course of the normals period (station moves, changes in observation time, and changes in instrumentation). Daily temperature observations, however, are not homogenized, which could lead to inconsistencies between the daily and monthly normals. This study offers a constrained harmonic technique that forces the daily temperature normals to be consistent with the monthly temperature normals. This approach replaces the cubic spline interpolation of monthly temperature normals that was used to compute earlier versions of NOAA's daily temperature normals. It effectively passes the homogenization applied at the monthly scale down to the daily scale, resulting in a smooth annual cycle devoid of day-to-day sampling variability and intermonth discontinuities.

1. Introduction

It is well known that daily temperature normals computed by simply averaging the 30 values for each day of the year (i.e., “raw daily normals”) constitute a fairly noisy time series because of sampling variability (see Fig. 1). This effect is exacerbated if missing values are present in the data record. There are numerous ways to calculate daily temperature normals such that high-frequency noise is suppressed, yielding a smooth representation of the annual cycle.¹ Two common approaches are time series filtering and interpolation. The filtering approach applies a low-pass filter to the raw daily normals. The interpolation approach fits a curve through the 12 monthly normals. This method is extremely useful in situations where the underlying daily data are not available, which is not uncommon when dealing with

international data, as nations are more likely to make their monthly data freely available rather than their daily data (T. Peterson 2013, personal communication).

A factor complicating the computation of daily temperature normals is the need to homogenize temperature time series. The U.S. surface temperature record is known to contain inhomogeneities due to station moves, instrumentation changes, and other changes in observing practices (Menne et al. 2010). For example, the transition from manual observations to Automated Surface Observing System by the National Weather Service in the 1990s is known to have introduced inhomogeneities into those stations' time series (Guttman and Baker 1996). Homogenization is essential for ensuring that climate normals are as representative of the current observing practices as possible at the time of computation. Currently, the U.S. surface temperature record is only homogenized at the monthly (and annual) scale (Menne and Williams 2009; Menne et al. 2009), whereas daily temperature values from the Global Historical Climatology Network–Daily (GHCN–Daily) are not (Menne et al. 2012), even though the mean monthly temperatures are derived from daily observations. Figure 1 illustrates this difference. The raw daily normals for minimum temperature at the Reno, Nevada, station are $\sim 2^{\circ}\text{F}$ ($\sim 1.1^{\circ}\text{C}$) cooler than the homogenized monthly normals for each month of the year. As articulated by Menne et al. (2009),

¹ The term “annual cycle” is often used to refer to a single cosine, sine, or harmonic even though the annual march of temperatures for some locations can be highly asymmetric. Here, we refer to an annual cycle as the annual march of temperatures whose morphology can vary from station to station.

Corresponding author address: Anthony Arguez, NOAA/NCDC, Room 506, 151 Patton Avenue, Asheville, NC 28801.
E-mail: anthony.arguez@noaa.gov

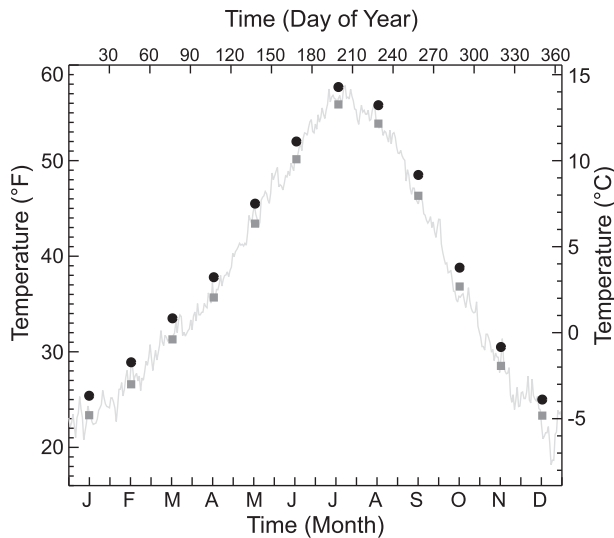


FIG. 1. Raw daily averages of minimum temperatures for Reno, NV (light gray curve): monthly averages computed from the raw daily averages (light gray squares) and homogenized monthly normals (black circles).

this station's minimum temperature time series was homogenized to account for an apparent urbanization signal, resulting in warmer temperature values over 1981–2010 and a substantially reduced upward trend over the last three decades that is more in line with the trends at neighboring stations (see their Fig. 8 for more information).

Thus, we developed a technique for computing daily normals with the following properties:

- 1) Day-to-day sampling variability is removed.
- 2) The monthly average of the daily temperatures exactly equals the corresponding monthly normal for each month, thereby “passing through” the monthly homogeneity adjustments to the daily scale.
- 3) There are no intermonth discontinuities (which can be evident in other methods when monthly varying homogeneity adjustments are applied).
- 4) The shape of the annual cycle is influenced by the daily data.

Prior to the release of the 1981–2010 climate normals, daily climate normals computed at the National Oceanic and Atmospheric Administration's (NOAA's) National Climatic Data Center (NCDC) were not computed directly from daily data values. The 1971–2000 monthly normals were homogenized following the procedures outlined by Peterson and Easterling (1994) and Easterling and Peterson (1995). Next, the daily temperature normals were interpolated using a cubic spline through the 12 homogenized monthly normals. Endpoint issues were

addressed by applying the spline to a once-repeating 24-point July–June series (Owen and Whitehurst 2002). The monthly means anchored the middle of the month; for example, the 16 July value is set to the July normal value. Finally, an ad hoc postspline correction was utilized in an attempt to make the monthly and daily normals consistent, as well as to make the temperature normals consistent with the heating and cooling degree-day normals. However, this ad hoc approach could not guarantee consistency for all stations (R. Heim Jr. 2012, personal communication). Therefore, this cubic spline approach does not satisfy property 2 listed above, or property 4 as we show in section 4.

A method that satisfies all four properties is an approach we call the constrained harmonic fit, which was utilized to compute NOAA's 1981–2010 daily temperature normals (Arguez et al. 2012). Harmonic analysis (without constraints) has been shown to be a useful tool for estimating an annual cycle (see examples 8.8–8.11 in Wilks 2006). By incorporating constraints, we can ensure that the average of the resulting daily temperature normals for each month exactly equals the corresponding monthly temperature normal. The constrained harmonic fit approach is described in greater detail in section 2. Examples are shown in section 3, followed by a summary in section 4.

2. Data and methodology

Daily maximum temperatures (t_{\max}) and minimum temperatures (t_{\min}) were extracted from GHCN-Daily for 2041 U.S. stations. The data values in GHCN-Daily have undergone extensive quality control as described by Durre et al. 2010, resulting in some values that are flagged as erroneous. Each station has at least 25 non-flagged values available for each Julian day (t), for both t_{\max} and t_{\min} , over the 1981–2010 time period. The raw daily normals $y(t)$ are calculated as the simple arithmetic mean of the valid values for each Julian day. The homogenized monthly normals of t_{\min} and t_{\max} are also used. The generalized approach presented here does not depend on the manner in which monthly temperature normals are calculated. For the purposes of this study, we assume that accurate monthly temperature normals have been computed a priori for use in calculating daily temperature normals.

We can model the daily temperature normals function as a linear combination of harmonics. As described by Wilks (2006), a single harmonic can sometimes provide a reasonable representation of the annual cycle, but additional harmonics are needed in order to account for features that deviate from a single sinusoidal shape, such as asymmetries between summer and winter, or between

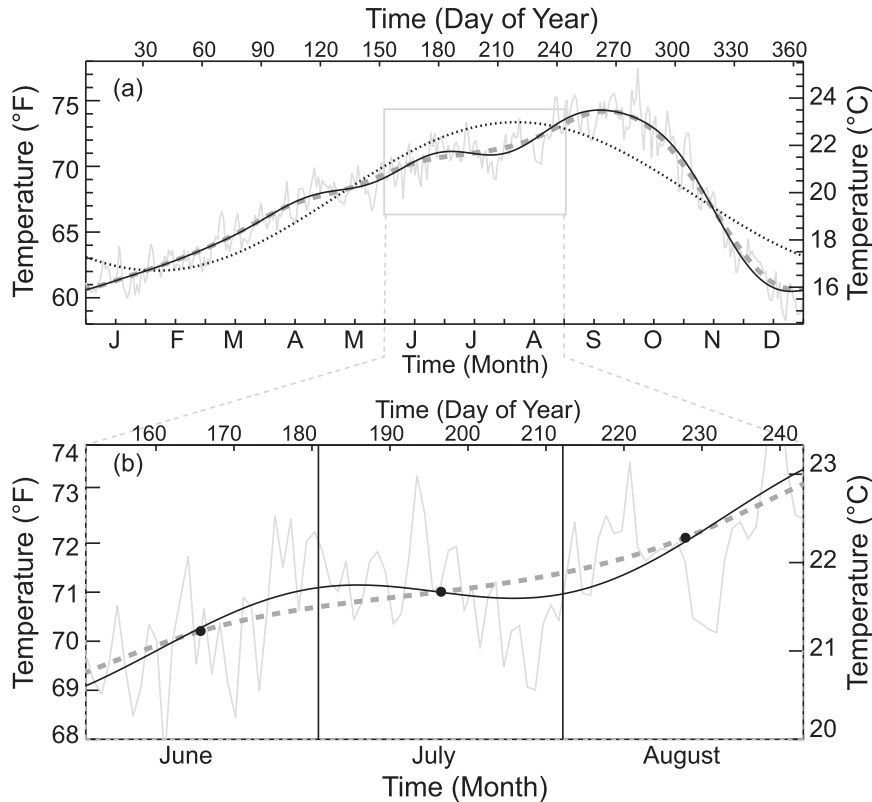


FIG. 2. (a) Raw daily averages of maximum temperatures for Salinas, CA (light gray curve). Bold black curve shows constrained harmonic fit for this time series. Dashed gray curve shows the cubic spline interpolation. Dotted black line indicates the (unconstrained) harmonic fit using only one harmonic. (b) As in (a), but for the June–August period; black circles indicate the homogenized monthly normals.

transition seasons. The equation for the harmonic fit $h(t)$ takes the familiar form

$$h(t) = A_0 + \sum_{k=1}^M [A_k \cos(\omega_k t) + B_k \sin(\omega_k t)], \quad (1)$$

where t ranges from 1 to $N = 365$, $\omega_k = 2\pi k/N$, and M is the number of harmonics used. If $M = N/2$, then $h(t) = y(t)$; the curve $y(t)$ has been reconstructed exactly via the superposition of sines and cosines. If $M \ll N/2$, then $h(t)$ represents a heavily smoothed version of $y(t)$, equivalent to a low-pass filter. If no constraints are applied, then we can solve for the coefficient vectors \mathbf{A} and \mathbf{B} in (1) via least squares minimization of the following cost function:

$$J_u(\mathbf{A}, \mathbf{B}) = \sum_{t=1}^N \left[y(t) - A_0 - \sum_{k=1}^M [A_k \cos(\omega_k t) + B_k \sin(\omega_k t)] \right]^2. \quad (2)$$

This system of linear equations ($2M + 1$ equations and $2M + 1$ unknowns) can be solved fairly easily using singular value decomposition (SVD). For this unconstrained case, the coefficient vectors \mathbf{A} and \mathbf{B} are more commonly calculated using a Fourier transform.

However, we need to constrain the coefficient values such that the means of the daily normals for a particular month are consistent with the monthly normal. For example, we need to guarantee that the average of the 31 daily tmax normals in January will equal the monthly tmax normal for January. The cost function with 12 monthly constraints is

$$J_c(\mathbf{A}, \mathbf{B}, \boldsymbol{\lambda}) = \sum_{t=1}^N [y(t) - h(t)]^2 + \lambda_{\text{jan}} \left[T_{\text{jan}} - \frac{1}{31} \sum_{t=1}^{31} h(t) \right] + \lambda_{\text{feb}} \left[T_{\text{feb}} - \frac{1}{28} \sum_{t=32}^{59} h(t) \right] \dots + \lambda_{\text{dec}} \left[T_{\text{dec}} - \frac{1}{31} \sum_{t=335}^{365} h(t) \right], \quad (3)$$

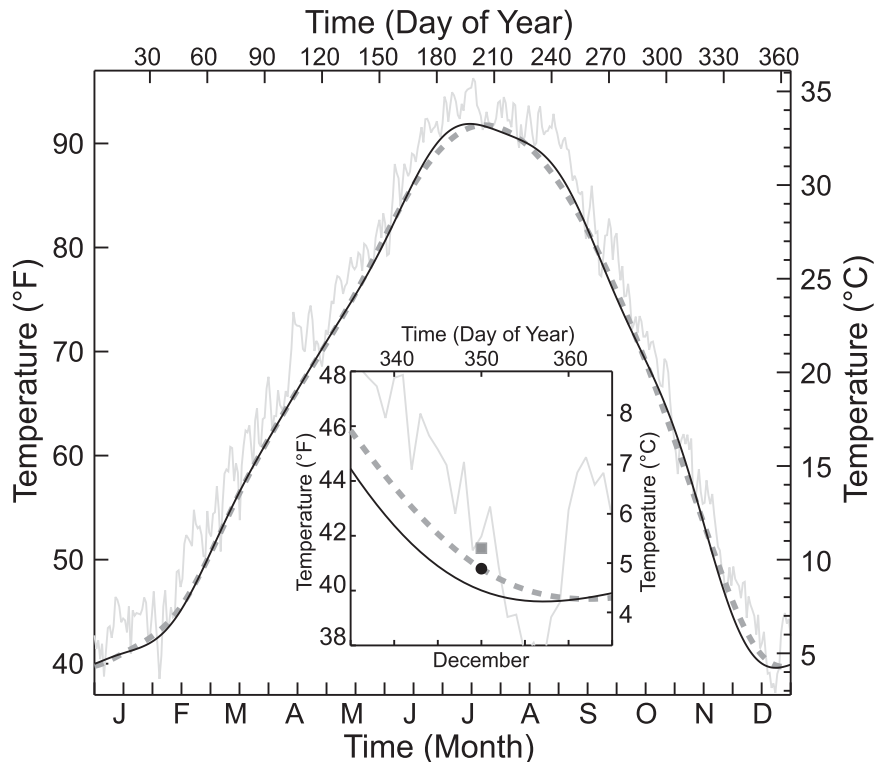


FIG. 3. As in Fig. 2, but for maximum temperatures in Beaver City, NE. December values are highlighted in the inset; the gray square indicates the average of the December values of the cubic spline interpolation and the black circle indicates the homogenized December normal.

where T_{jan} is the monthly temperature normal for January, T_{feb} is the monthly temperature normal for February, etc. The λ terms are Lagrange multipliers that impose the constraints. The above formulation follows the “earlier tradition” or the “Lagrangian function approach” described by Kalman (2009). Once again, we can solve this linear system of equations using SVD.

If the number of harmonics M is ≥ 6 , then the constraints may be met exactly. Otherwise, the linear system is overdetermined and, though SVD can yield a solution for the coefficients, the constraints will not be met, as we will illustrate in the next section. After experimenting with $M \geq 6$ and visually inspecting the smoothing provided on several samples, we chose to set M equal to 6 for this exercise, although $M > 6$ can be a viable alternative for other applications so long as the potential for overfitting is accounted for. The constrained harmonic fit was calculated for t_{max} and t_{min} for all 2041 stations.

3. Results

The raw daily t_{max} normal curve for Salinas, California (Fig. 2a), clearly demonstrates an annual cycle that deviates from a classic sinusoidal curve. In fact, using a

single harmonic ($M = 1$), the constrained harmonic fit would place the annual peak in early August, whereas the daily data indicate the warmest time of year in Salinas is September–early October. Salinas is representative of many stations in the western United States that have a steeper transition during fall and a more gradual transition during spring. Both the cubic spline interpolation and the constrained harmonic fit ($M = 6$) do a respectable job of capturing the annual cycle in Salinas [the maximum absolute difference between the two is $\sim 0.5^\circ\text{F}$ ($\sim 0.3^\circ\text{C}$)], although the cubic spline interpolation is incapable of resolving intramonth fluctuations (Fig. 2b). For example, the raw daily normals clearly indicate the presence of a local minimum in late July, a feature that is captured by the constrained harmonic fit, yet the cubic spline interpolation increases monotonically from June through August.

The raw daily t_{max} normal curve for Beaver City, Nebraska (Fig. 3), exhibits strong continentality, in sharp contrast to the much narrower annual range of t_{max} in coastal California as seen in Salinas. The raw daily normals are noticeably warmer ($\sim 2.5^\circ\text{F}$; $\sim 1.4^\circ\text{C}$) than the cubic spline interpolation or the constrained harmonic fit, as the monthly homogenization procedure had the net

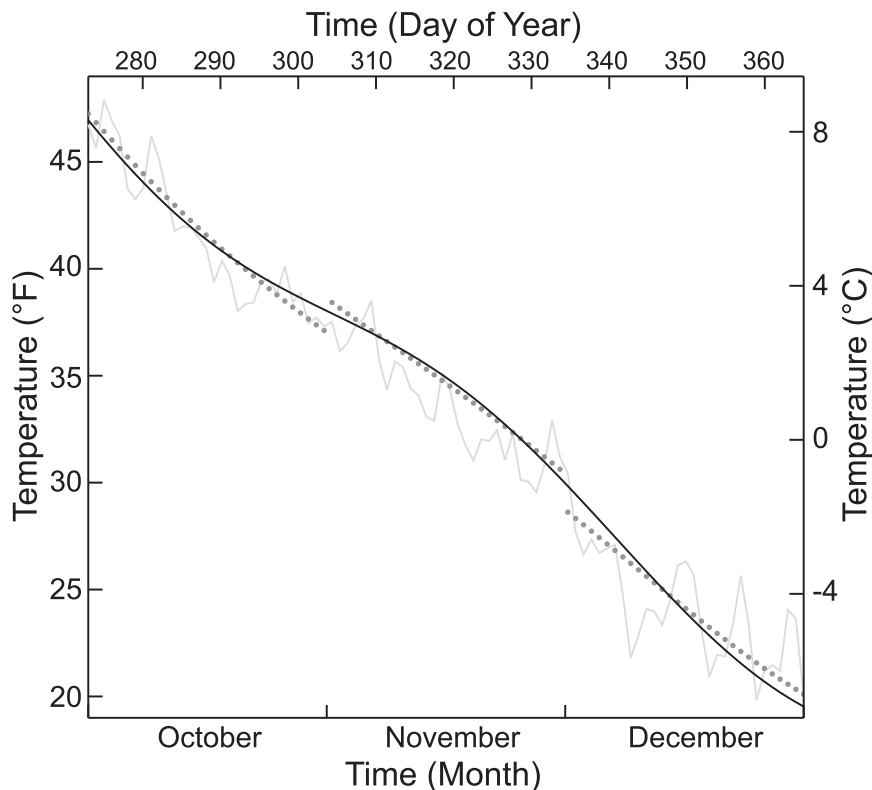


FIG. 4. Raw daily averages of t_{min} for Hartford, CT, for the October–December period (light gray curve). Solid black curve shows the constrained harmonic fit for this time series. Dark gray circles indicate an unconstrained harmonic fit ($M = 6$), subsequently adjusted such that the average of the daily values in a month matches the corresponding, homogenized monthly normal.

effect of lowering the monthly t_{max} values for Beaver City (in contrast to Reno t_{min} ; Fig. 1). The constrained harmonic fit is shifted toward cooler t_{max} values to accommodate this, whereas the cubic spline interpolation is simply fit through the cooler homogenized normals. The two methods result in daily temperature normals that differ by as much as $\pm 1.5^{\circ}\text{F}$ ($\pm 0.8^{\circ}\text{C}$) for specific days, particularly in the warmest and coldest months, although the annual means differ by less than 0.1°F (0.06°C).

Consider how the cubic spline interpolation performs in December for Beaver City t_{max} . The December t_{max} normal is 40.8°F (4.9°C), and the average of the constrained harmonic fit values for December is also 40.8°F (4.9°C), given the hard constraint. The cubic spline interpolation is anchored at 40.8°F (4.9°C) on 16 December (the middle of the month), but early December values range from 41° to 46°F (5.0° – 7.8°C), whereas late December values bottom out near 40°F (4.4°C). As a result, the mean of the cubic spline interpolation for December is over 0.7°F (0.4°C) warmer than the December normal (see inset in Fig. 3). This highlights the inability of the cubic spline

interpolation to preserve the homogenized monthly mean. This leads the cubic spline interpolation to underestimate peaks in summer and overestimate minima in winter, resulting in a slight underestimation of total variance with respect to the raw daily normals as well as the constrained harmonic fit approach, which does not result in lowered total variance (not shown).

Overall, the mean differences between the constrained harmonic fit and the cubic spline interpolation results for the 2041 stations considered is essentially zero for both t_{max} and t_{min} . However, individual values can vary to some degree, as seen in Figs. 2 and 3. The average absolute difference between the two approaches is 0.3°F (0.2°C) for both t_{max} and t_{min} . In terms of maintaining consistency between the monthly temperature normals and the resulting daily temperature normals, the constrained harmonic fit approach with ($M = 6$) has zero discrepancy, by design, for all station months. For both t_{max} and t_{min} , the spline fit interpolation is typically off by 0.2°F (0.1°C) in an absolute sense from maintaining consistency, and the discrepancy can be as high as 1.0°F (0.6°C). As intimated by Fig. 3, the largest

TABLE 1. RMSEs in °F (°C) between the homogenized monthly normal and the monthly average of the daily normals calculated using the constrained harmonic fit as a function of M .

	$M = 1$	$M = 2$	$M = 3$	$M = 4$	$M = 5$	$M = 6+$
Reno, NV, tmin	1.73 (0.96)	0.45 (0.25)	0.32 (0.18)	0.27 (0.15)	0.23 (0.13)	0.00
Salinas, CA, tmax	1.69 (0.94)	0.71 (0.39)	0.27 (0.15)	0.14 (0.08)	0.02 (0.01)	0.00
Beaver City, NE, tmax	2.03 (1.13)	0.82 (0.46)	0.58 (0.32)	0.41 (0.23)	0.19 (0.11)	0.00
Hartford, CT, tmin	1.11 (0.62)	1.09 (0.61)	0.57 (0.32)	0.42 (0.23)	0.07 (0.04)	0.00

inconsistencies tend to occur in summer and winter months near the maxima and minima for the year.

An alternative to these methods is to apply a low-pass filter to the raw daily normals, such as successive passes of the 1–2–1 filter (see von Storch and Zwiers 1999). However, this simple approach cannot guarantee preservation of the mean. Further, low-pass filtering typically results in reduced variance (e.g., see Arguez et al. 2008), which can lead to severe underestimation and overestimation of highs and lows, respectively.

Other potential options involve a multistep approach. For example, one could pass the monthly adjustments down to either the daily data or to the raw daily normals series (e.g., Vincent et al. 2002). Then, an unconstrained harmonic fit could be used to derive the daily temperature normals. Again, this method does not guarantee preservation of the monthly normals. Another approach is to compute the unconstrained harmonic fit of the raw daily normals, and then adjust the values in each month up or down to match the corresponding monthly normal. However, this introduces intermonth discontinuities, as seen when it is applied to the raw daily tmin normals for Hartford, Connecticut (Fig. 4).

As stated earlier, the constrained harmonic fit does not preserve the mean monthly normals when fewer than six harmonics are used. Table 1 shows, as a function of M (the number of harmonics used), the root-mean-square errors (RMSEs) between the homogenized monthly normals and the monthly averages of the constrained harmonic fit values for the four time series considered in this investigation. For example, the $M = 1$ case for Salinas tmax, which was displayed graphically in Fig. 2a, has an RMSE value of 1.69°F (0.9°C). The constraints in (3) ensure that the RMSE will equal 0 for every month when $M \geq 6$, which is reflected in Table 1.

4. Summary

The constrained harmonic fit effectively smooths the raw daily normals without introducing discontinuities, and ensures preservation of the mean monthly normals. The cubic spline interpolation approach results in a smooth and continuous daily temperature normals curve. However, it does not always capture intraseasonal

fluctuations as seen in Fig. 2a, and it does not guarantee preservation of the mean monthly normals as documented in Fig. 3.

The constrained harmonic fit described here offers a one-step approach for computing daily temperature normals satisfying the four properties listed in the introduction. It is particularly useful when monthly temperatures are homogenized but the underlying daily data values are not. However, it also accommodates inherent inconsistencies between monthly and daily data caused by other factors, such as separate quality control of monthly and daily temperatures or the manner in which monthly temperature normals are computed. We recommend the constrained harmonic fit approach for computation of daily temperature normals when the underlying daily data are available and monthly temperature data are deemed to be of superior quality versus the daily data.

Acknowledgments. The authors kindly acknowledge I. Durre, R. Vose, M. Squires, X. Yin, J. Nielsen-Gammon, J. Christy, and J. Curtis for providing feedback on this approach. We also thank three anonymous reviewers for providing actionable feedback that helped strengthen the manuscript.

REFERENCES

- Arguez, A., P. Yu, and J. J. O'Brien, 2008: A new method for time series filtering near endpoints. *J. Atmos. Oceanic Technol.*, **25**, 534–546.
- , I. Durre, S. Applequist, R. S. Vose, M. F. Squires, X. Yin, R. R. Heim Jr., and T. W. Owen, 2012: NOAA's 1981–2010 U.S. Climate Normals: An overview. *Bull. Amer. Meteor. Soc.*, **93**, 1687–1697.
- Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *J. Appl. Meteor. Climatol.*, **49**, 1615–1633.
- Easterling, D. R., and T. C. Peterson, 1995: A new method for detecting and adjusting for undocumented discontinuities in climatological time series. *Int. J. Climatol.*, **15**, 369–377.
- Guttman, N. B., and C. B. Baker, 1996: Exploratory analysis of the difference between temperature observations recorded by ASOS and conventional methods. *Bull. Amer. Meteor. Soc.*, **77**, 2865–2873.
- Kalman, D., 2009: Leveling with Lagrange: An alternative view of constrained optimization. *Math. Mag.*, **82**, 186–196.

- Menne, M. J., and C. N. Williams Jr., 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700–1717.
- , —, and R. S. Vose, 2009: The United States Historical Climatology Network monthly temperature data version 2. *Bull. Amer. Meteor. Soc.*, **90**, 993–1007.
- , —, and M. A. Palecki, 2010: On the reliability of the U.S. surface temperature record. *J. Geophys. Res.*, **115**, D11108, doi:10.1029/2009JD013094.
- , 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.
- Owen, T. W., and T. Whitehurst, 2002: United States Climate Normals for the 1971–2000 period: Product descriptions and applications. Preprints, *Third Symp. on Environmental Applications: Facilitating the Use of Environmental Information*, Orlando, FL, Amer. Meteor. Soc., J4.3. [Available online at https://ams.confex.com/ams/annual2002/techprogram/paper_26747.htm.]
- Peterson, T. C., and D. R. Easterling, 1994: Creation of homogeneous composite climatological reference series. *Int. J. Climatol.*, **14**, 671–679.
- Vincent, L. A., X. Zhang, B. R. Bonsal, and W. D. Hogg, 2002: Homogenization of daily temperatures over Canada. *J. Climate*, **15**, 1322–1334.
- von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.
- Wilks, D. S., 2006. *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 59, Academic Press, 627 pp.