# NOAA's 1981-2010 Climate Normals
# Methodology of Temperature-related Normals

*Created: June 29, 2011*
*Last Modified: June 30, 2011*

Anthony Arguez[1], Scott Applequist[1], Russell Vose[1], Imke Durre[1], Michael Squires[1], and Xungang Yin[2]

(1) NOAA's National Climatic Data Center, Asheville, NC, 28801
(2) NOAA/STG, Inc., Asheville, NC 28801

## Overview

This report describes the methodology used to compute daily, monthly, seasonal, and annual normals for numerous temperature-related variables at about 7,500 weather stations for the 1981-2010 Normals period. A climate normal is typically defined as a 30-year average of an atmospheric quantity, such as maximum temperature. However, advanced statistical techniques are used to account for missing data values, inhomogeneities, station moves, etc. and therefore the normals presented here are much more than 30-year averages. This report offers a preliminary description of all procedures used to compute the new normals for temperature-related variables. We intend to submit a journal article on this matter which, if and when accepted, would replace this report as the authoritative reference for the computations done on temperature-related variables for the 1981-2010 Normals. For information regarding precipitation-related normals or hourly normals (including hourly temperature normals), please review the accompanying documentation.

## Source Data

The underlying values used to compute the 1981-2010 Normals come from the Global Historical Climatology Network - Daily (GHCN-Daily) dataset (Menne et al., submitted). As its name suggests, this dataset contains daily observations for many atmospheric variables worldwide, and is the most comprehensive set of daily climate data for the United States. The data values have undergone extensive quality control (QC) as described by Durre et al. 2010. The backbone of the stations used in the 1981-2010 Normals come from the U.S. Cooperative Observer Network. First Order stations as well as the U.S. Climate Reference Network are also included, however we do not report climate normals for CoCoRaHS stations.

As described by Menne and Williams (2009) and Menne et al. (2009), NCDC provides monthly temperature data values that have undergone robust quality control and standardization at the monthly timescale. For the 1981-2010 Normals, the approaches described in these papers were applied to monthly maximum and minimum temperature values that were in turn computed from GHCN-Daily values. Monthly values were computed for station-months for which no more than nine missing or suspect daily values were present in GHCN-Daily. The standardization procedures account for both documented and undocumented station moves and other changes in observing practices. Therefore, we give precedence to normals computed from monthly temperature data.

Product Portfolio

The temperature-related products in the 1981-2010 Normals are listed in Table 1. Normals of maximum, average, and minimum temperature; diurnal temperature range; and heating and cooling degree days are provided at the daily, monthly, seasonal, and annual timescales. Part of the standardization performed on monthly temperature values involves application of a time of observation (TOB) adjustment that strives to make it as if the observations at a particular station had been taken at local midnight. We adjust the Normals back to local observation time, but also provide the midnight observing time offsets. Standard deviations of monthly mean temperatures as well as daily temperature values are also reported. Finally, we also provide "count" normals at the monthly, seasonal, and annual timescales. These are parameters such as the normals of the number of days in July where the maximum temperature exceeds 90F.

Table 1. **Temperature Normals Released July 1, 2011**

| | |
|---|---|
| Daily | Tmax, Tmin, Tavg, DTR |
| | Heating and Cooling Degree Days |
| | Standard Deviations |
| Monthly | Tmax, Tmin, Tavg, DTR |
| | Midnight observing time offsets |
| | Heating and Cooling Degree Days |
| | Count Normals |
| | Standard Deviations |
| Seasonal | Tmax, Tmin, Tavg, DTR |
| | Heating and Cooling Degree Days |
| | Count Normals |
| Annual | Tmax, Tmin, Tavg, DTR |
| | Heating and Cooling Degree Days |
| | Count Normals |

Table 1 only shows the normals that will be released July 1, 2011. Later installments of the 1981-2010 Normals will provide several other product classes. This includes agricultural related climate normals such as frost/freeze dates and growing degree days; all normals that require gridding or aggregation including at the climate division level; as well as any climate normals that involve population data such as our population-weighted monthly heating and cooling degree day product.

For the vast majority of stations, we compute the normals using a "traditional approach" that uses 30 years of data wherever possible. However, for about 1100 short-record stations, we employ a "pseudonormals" approach as described by Sun and Peterson (2005). These pseudonormals are based on linear combinations of the normals from neighboring stations computed using the traditional approach. In this report, we focus on the traditional approach. For more information about the pseudonormals approach, please consult the Sun and Peterson (2005) paper.

Computation of Tmax, Tmin, Tavg, and DTR Normals and Standard Deviations

As described earlier, we give precedence to monthly temperature values and therefore first compute monthly normals of maximum and minimum temperature (see the flowchart in Figure 1). Missing or suspect monthly values are filled using a regression technique based on index of agreement with neighboring values. For more information on the filling analysis, please consult the accompanying precipitation methodology. Once we arrive at filled monthly values for all 30 years, the monthly normals of maximum and minimum temperature are computed as the simple averages of the 30 values for each station-month. We report a completeness flag with each normals value describing the relative completeness of each data record (before filling). The completeness criteria are an extension of the guidelines provided by the World Meteorological Organization (WMO 1989). The monthly average temperature (diurnal temperature range) normal is computed as the mean (difference) of the monthly maximum temperature normal and the monthly minimum temperature normal. We also compute the standard deviations across the 30-year period for the four variables.

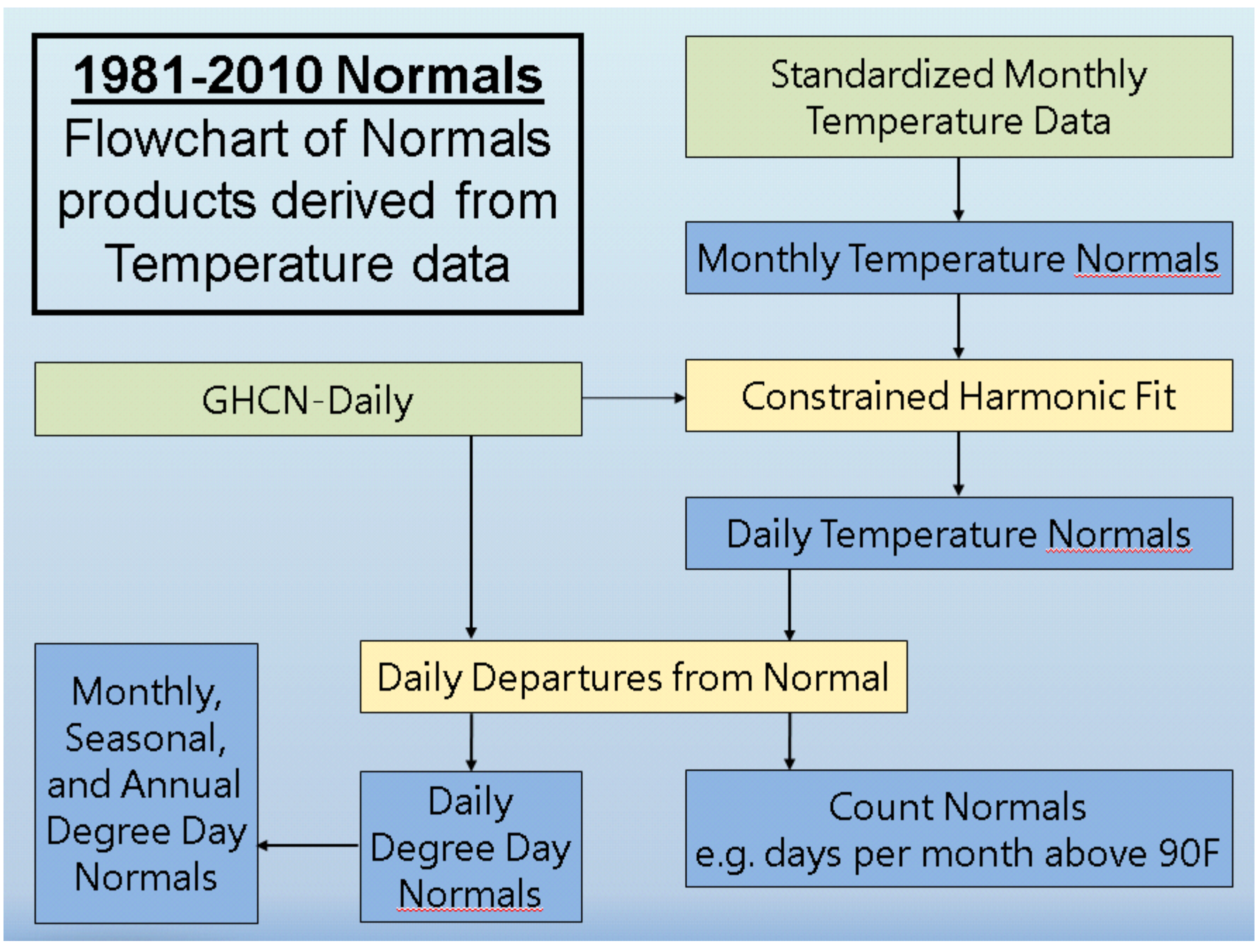

Figure 1. Flowchart of temperature-derived climate normals. Monthly temperature values are given precedence. Datasets are shown in green, methods are shown in orange, and products are shown in blue.

The computation of daily normals of maximum and minimum temperature involves a constrained harmonic least squares fit. We begin by first computing the "raw" daily normals from GHCN-Daily. Let $y(t)$ represent the *raw* daily normals:

$$y(t) = \frac{1}{30}\sum_{k=0}^{29} x(t + 365k) \tag{1}$$

Here, $t$ ranges from 1 to 365. For Julian days where there are less than 10 non-missing and non-suspect values from 1981-2010, we use a windowing technique as needed to yield at least 10 values for the average. We can model the daily temperature normals function as a linear combination of harmonics. As described in 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 transition seasons. On the other hand, over-fitting must be guarded against as well. The equation for the harmonic fit, $h(t)$, is as follows:

$$h(t) = A_0 + \sum_{k=1}^{M} [A_k \cos(\omega_k t) + B_k sin(\omega_k t)] \tag{2}$$

where t ranges from 1 to $N$=365 and $\omega_k$=2π$k$/$N$. If $M$=$N$, then $h(t)$=$y(t)$; if $M$<$N$ then $h(t)$ represents a smoothed version of $y(t)$. If no constraints are applied, we can solve for the coefficients in (2) via least squares minimization of the following cost function:

$$J_u(\vec{A}, \vec{B}) = \sum_{t=1}^{N} \left[ y(t) - A_0 - \sum_{k=1}^{M} [A_k \cos(\omega_k t) + B_k sin(\omega_k t)] \right]^2 \tag{3}$$

Setting partial derivatives to zero, we have 2$M$+1 equations and 2$M$+1 unknowns. This system of linear equations can be solved fairly easily using singular vector decomposition (SVD) to arrive at the coefficients ($A_0$, $A_1$, $B_1$, etc.). The coefficients are then plugged into (2) to define the unconstrained harmonic fit. However, because we give precedence to the monthly temperature values, we need to constrain the coefficient values such that the mean monthly normals are consistent with the means of the daily normals for a particular month. Therefore, we need to impose 12 constraints, one for each month:

$$J_c(\vec{A}, \vec{B}, \vec{\lambda}) = \sum_{t=1}^{N} \left[ h(t) - A_0 - \sum_{k=1}^{M} [A_k \cos(\omega_k t) + B_k sin(\omega_k t)] \right]^2 \tag{4}$$

$$+\lambda_{jan}\left[T_{jan} - A_0 - \frac{1}{31}\sum_{t=1}^{31}\sum_{k=1}^{M}[A_k\cos(\omega_k t) + B_k sin(\omega_k t)]\right]$$

$$+\lambda_{feb}\left[T_{feb} - A_0 - \frac{1}{28}\sum_{t=32}^{59}\sum_{k=1}^{M}[A_k\cos(\omega_k t) + B_k sin(\omega_k t)]\right]$$

$$\ldots +\lambda_{dec}\left[T_{dec} - A_0 - \frac{1}{31}\sum_{t=335}^{365}\sum_{k=1}^{M}[A_k\cos(\omega_k t) + B_k sin(\omega_k t)]\right]$$

$T_{jan}$ is the monthly temperature normal for January, $T_{feb}$ is the monthly temperature normal for February, etc. The $\lambda$ terms are Lagrange multipliers that impose the constraints. Now, setting partial derivatives to zero, we have $2M$+13 equations and $2M$+13 unknowns. Once again, we can solve this linear system of equations using SVD.

In order for the constraint to be imposed exactly, $M$ must be greater than or equal to 6. Otherwise, there are more constraints than coefficients. However, to guard against over-fitting, we need to restrict the number of harmonics. Therefore, **we set *M* to 6 for all computations.** Daily maximum and minimum temperature normals are computed in this fashion. As before, the average temperature and DTR normals are derived from these.

To compute the standard deviations of daily temperature values, we use a 15-day window about the centered Julian day. In practice, this results in a time series of at least 100 good data values. We then simply take the standard deviation of these values. To smooth out considerable noise in these estimates, we employ a running 29-day equal-weight filter. Note that standard deviations (both monthly and daily) are not computed for pseudonormal stations.

Computation of Heating and Cooling Degree Day Normals

The 1971-2000 climate normals of monthly heating and cooling degree days (HDD/CDD) were originally computed for all stations using a modification of the Thom Method (Thom 1954; Thom 1966), which is based on monthly means and standard deviations. Daily degree day normals were computed as a spline fit through the monthly degree day values. After receiving feedback from NWS and industry, it was decided that HDD/CDD calculations for 1971-2000 would be done 'directly' for the first-order stations for which relatively complete daily records were available. For 1981-2010, we compute degree days in a 'more direct' fashion for all stations, leveraging off of the improvements to the daily temperature normals.

For computation of degree days, we utilize the daily mean temperature normals for a particular station, which is ultimately derived from the constrained harmonic fit analysis described above. The key step is to estimate the spread of daily temperature values about the daily normal. This is critical since the definition of HDD/CDD is constructed as an asymmetric sum. Let us first consider this definition. Suppose $T_j(t)$ represents 30 mean temperature values, one for each year from $t$=1981 to $t$=2010, for a particular Julian day $j$, which ranges from 1 to 365. $H_j$ and $C_j$ represent the daily heating and cooling degree days, respectively, as follows:

$$H_j(t) = \left\{\frac{0 \; if \; T_j(t) > 65F}{65F - T_j(t) if \; T(t)_j \leq 65F}\right\} \tag{5}$$

$$C_j(t) = \left\{\frac{0 \; if \; T_j(t) < 65F}{T_j(t) - 65F \; if \; T(t)_j \geq 65F}\right\} \tag{6}$$

The most direct way to compute daily normals of $H_j$ and $C_j$ would be to average the 30 annual values for each $j$. That poses three major issues: (1) the normals would be quite noisy, (2) missing values would have to be accounted for somehow, and (3) the HDD/CDD normals would not be consistent with the daily mean temperature normals which are consistent with the monthly values. Since we already know the daily mean temperature normal from the harmonic analysis, we just need an estimate of the distribution about this average to estimate the daily HDD/CDD normals, preferably in such a way that smoothes out sampling variability.

Analogous to the approach for daily standard deviations, we use a 15-day window centered on $j$ (and allowing it to extend across the beginning/end of the year) across all 30 years, for a maximum distribution of 450 values. The anomalies with respect to the distribution mean are computed for all non-missing values, and then the corresponding daily mean temperature normal is added to each of these anomaly values. Then, the individual HDD/CDD values are computed following the equations above. Finally, these values are averaged (over the number of non-missing values in the window) to arrive at a normal value of HDD/CDD for that Julian day. This process is repeated for all Julian days. These daily normals are smoothed lightly using up to 11 passes of a 1-2-1 filter (the number of passes is based on the time series). Monthly HDD/CDD normals are computed by summing up the corresponding daily normals. Annual HDD/CDD normals are computed as the sums of the 12 monthly values.

Computation of Count Normals

Count normals, such as the number of days per month in which the minimum temperature drops below 32F, is computed using an analogous windowing strategy as that used for standard deviations and HDD/CDD, except that instead of a centered window we use the days in a given month. All available daily values are adjusted such that the daily average for each day is equal to the relevant daily temperature normal. The percentage of values that meet the particular criterion (e.g., tmax greater than 70F) is calculated, and that percentage is scaled to account for missing values and arrive at the count normal for that month. From the monthly values, seasonal and annual count normals are computed.

References


Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *Journal of Applied Meteorology and Climatology,* **8***, 1615-1633.*

Menne, M., I. Durre, B. G. Gleason, T. G. Houston, and R. S. Vose, 2011: An overview of the Global Historical Climatology Network-Daily database. Submitted to the Journal of Atmospheric and Oceanic Technology.

Menne, M. J., and C. N. Williams, Jr., 2009: Homogenization of Temperature Series via Pairwise Comparisons. *J. Clim*., **22,** 1700-1717.

Menne, M. J., Williams, Jr., C. N., and R. S. Vose, 2009: The United States Historical Climatology Network Monthly Temperature Data Version 2. *Bull. Amer. Meteor. Soc*., **90**, 993-1007.

Thom, H.C.S., 1954: The rational relationship between heating degree days and temperature, *Mon. Wea. Rev.*, **82**, 1-6.

Thom, H.C.S., 1966: Normal degree days above any base by the universal truncation coefficient, *Mon. Wea. Rev.*, **94**, 461-465.

Sun, B., and T. C. Peterson, 2005: Estimating temperature normals for USCRN stations. *International Journal of Climatology,* **25***, 1809-1817*.

Wilks, D.S., 2006. Statistical Methods in the Atmospheric Sciences, 2nd Ed. International Geophysics Series, Vol. 59, Academic Press, 627 pp.

World Meteorological Organization (WMO), 1989: Calculation of monthly and annual 30-year standard normals. WMO-TD 341, Geneva.