

Quality Control for USDA NRCS SM–ST Network Soil Temperatures: A Method and a Dataset*

QI HU AND SONG FENG

*Climate and Bio-Atmospheric Sciences Group, The School of Natural Resource Sciences, University of Nebraska at Lincoln,
Lincoln, Nebraska*

GARRY SCHAEFER

National Water and Climate Center, U.S. Department of Agriculture, Portland, Oregon

(Manuscript received 1 June 2001, in final form 22 January 2002)

ABSTRACT

In 1991, the U.S. Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS) established its Soil Moisture–Soil Temperature (SM–ST) Pilot Network consisting of 21 stations in 19 states in the contiguous United States. At each station, soil temperatures were measured at up to six different depths from 5.08 to 203.20 cm (or 2–80 in.) below the surface. Before 1997, the observations were made every 6 h, and they increased to hourly beginning in 1997. The goal of this network is to provide near-real time soil temperature and soil moisture observations in different regions across the United States for agricultural and water use management as well as for climate research. To improve the usefulness and increase the value of both the data and this network, a quality-control method for the soil temperature data was developed. The method used a soil heat diffusion model and its solution at individual sites to screen and distinguish erroneous soil temperature data and to provide their estimates. Evaluation of the quality-control method showed its accuracy and reliability, particularly when it was applied to hourly data. Application of this method to the data has yielded a high-quality, high-resolution soil temperature database from 1994 to 1999 for the network, which is accessible at the USDA National Water and Climate Center's Web site.

1. Introduction

Soil temperature and soil moisture specify the soil environment for plant growth. In spring, warm soil temperature and adequate soil moisture nurture seed germination and root development, but cold soil temperatures can retard these processes. In summer, soil moisture becomes a particularly sensitive parameter to monitor in agricultural practices of irrigation and chemical application. It has been included in various indices, such as the Palmer drought severity index, to quantify drought conditions (Palmer 1965; Karl and Kosciely 1982; Hubbard 1993; Hu and Willson 2000) and to assist irrigation scheduling and drought mitigation (e.g., Miller 2000).

Soil temperature and moisture also interact and affect regional atmospheric circulation, weather, and climate.

As shown in many modeling experiments, at different temperature and wetness, soils can partition surface sensible heat and latent heat differently and the resulting spatial heterogeneity of the heat fluxes can alter local-to regional-scale circulations and initiate storm development (Charney 1975; Pielke and Zeng 1989; Pielke and Avissar 1990). Thermodynamic effects of soils on regional climate variations at interannual to decadal scales have been revealed in an intriguing scale analysis in Tang (1989). Using long-term soil temperature data from a station in St. Paul, Minnesota, Tang showed that the amplitude of variation in average annual soil temperature in a soil column from the surface to 5 m below is 11.3 K. This temperature change corresponds to an annual variation in heat storage of $1.1 \times 10^4 \text{ J cm}^{-2}$ in the soil column [using average soil density 2 g cm^{-3} and specific heat capacity 1.0 J (g K)^{-1}], which is similar to the variation in annual enthalpy of the entire atmospheric column $1.5 \times 10^4 \text{ J cm}^{-2}$, thus indicating a significant role of soil–atmosphere heat exchange in atmospheric circulations in land areas. In addition, he showed that the interannual variations in soil heat storage in a column from the surface to 12 m below is 710 J cm^{-2} , which can well account for interannual enthalpy

* Contribution Number 13330, University of Nebraska Agricultural Research Division.

Corresponding author address: Dr. Qi Hu, School of Natural Resource Sciences, 237 L.W. Chase Hall, University of Nebraska at Lincoln, Lincoln, NE 68583-0728.
E-mail: qhu2@unl.edu

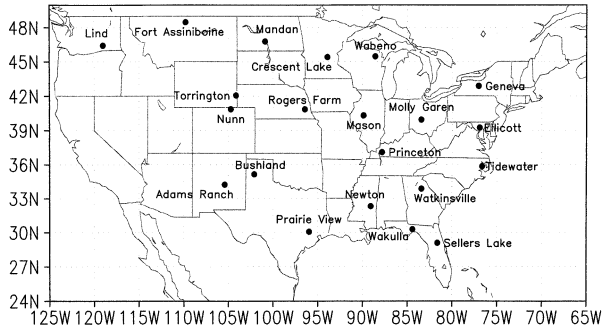


FIG. 1. Station sites and distribution of the USDA NRCS SM-ST Network.

variations in the middle latitude atmosphere $\sim 500 \text{ J cm}^{-2}$.

The value of soil temperature and moisture data to research and applications was reemphasized recently by the U.S. Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS) in its effort to develop a soil temperature and moisture database for the United States, which has been the least-developed among all climate datasets. This effort consisted of a plan to recover existing historical soil temperature data and an implementation in 1991 of the Soil Moisture–Soil Temperature (SM–ST) Pilot Network to measure soil temperature and moisture from 21 stations in 19 states in the United States (Fig. 1 and Table 1). From this network, the NRCS provided near-real time soil temperature and moisture observations in different agricultural regions across the United States and high-resolution data for a quality soil temperature and moisture dataset for climate research.

The stations in the network measured soil temperatures up to six “standard” depths at 5.08, 10.16, 20.32, 50.80, 101.60, and 203.20 cm (or 2, 4, 8, 20, 40, and 80 in.) below bare ground surface. Several of the stations also measured soil temperatures at 30.48 (12 in.), 101.60, and 203.20 cm below the surface at a companion site with short grass cover. Before 1997, the measurement frequency was every 6 h, and it increased to hourly beginning in 1997. Station data were transmitted using meteor burst telemetry to the central data control at the USDA National Water and Climate Center (WCC) in Portland, Oregon, where they were archived and made available on the WCC Web site. Requests for, and access to, the online soil temperature data have increased dramatically over the years from about 700 in October 1998 to near 8000 in January 2000 (see the report online at <http://www.wcc.nrcs.usda.gov/scan/page2.pdf>).

To assure data quality from this network and to accommodate the NRCS effort to develop a high-quality soil temperature dataset, we have developed a quality-control method for the SM–ST Network soil temperatures. This method is based on physical principles of heat transfer in soil media while taking into account “irregular effects” of weather and climate on soil tem-

perature variations. It identifies random instrumentation and human errors in the data and provides physically coherent estimates for missing and erroneous data, when necessary. Details of the method are described in section 3 after the types of error in the data are discussed in section 2. Application procedure of the method is presented in section 4 followed, in section 5, by test results of the method under various error scenarios. These results help quantify the accuracy and reliability of this method. Section 6 contains a summary of the method and its products.

2. Error in the network data

We first inspected the soil temperature data from all the depths at individual stations from 1994 to 1999¹ to find the “modes” of errors in the soil temperature data and to help us design the quality-control method for this network. The following six types summarize all the instrumentation and measurement errors in the data.

- 1) *Missing data and obviously incorrect data values.* Missing values was the most frequent problem in the station data. Although measurements were improved dramatically after 1994, missing data remained because of either missing measurement due to equipment or “sensor” failure, missing communication, or both. The other obvious erroneous data were those with values outside the range $\pm 50^\circ\text{C}$, which were physically impossible at all the sites. Figure 2 shows an example of such an error on 17 July 1998 at the station in Newton, Mississippi.
- 2) *Displaced diurnal variation.* In this case, existing hourly or 6-hourly soil temperature data showed a reasonable “diurnal cycle” but the values were substantially higher or lower than the average soil temperature at the depth for the time of year, which was estimated from available data for the time but in different years. Another error in this category was the unrealistically large diurnal amplitude of soil temperature at deep depths. An example of this error is shown in Fig. 3e for the data in 1996.
- 3) *Constant daily temperature over prolonged periods.* Some station data showed no diurnal cycle but constant hourly soil temperatures on different days at various depths for a month or longer. Although some of those data values were not very different from the daily mean temperatures on those days, they were suspicious because of their constant daily values for such a long period.
- 4) *Displaced annual variation.* This error category is similar to that in type 2 but for a period of month or longer. An example of such error is shown in Figs. 3a–d in early 1997.
- 5) *Incorrect annual data.* These errors showed near-

¹ Soil temperature data taken at the stations before 1994 were not continuous and are excluded from this quality control.

TABLE 1. Parameter values used in (2) for the network stations.

Station No. (Name)	Lat (°N), Lon (°W)	T_y (°C)	d_d (m)	d_s (m)	A_y (°C)
2001 (Rogers Farm, NE)	(40.85, 96.47)	9.889	0.102 18	1.952 15	14.264
2002 (Crescent Lake, MN)	(45.42, 94.00)	5.667	0.140 45	2.683 24	11.648
2003 (Wabeno, WI)	(45.47, 88.58)	5.463	0.110 30	2.107 32	14.691
2004 (Mason, IL)	(40.32, 89.90)	11.091	0.100 96	1.928 89	13.288
2005 (Princeton, KY)	(37.10, 87.83)	13.485	0.104 13	1.989 47	11.436
2006 (Bushland, TX)	(35.17, 102.10)	12.879	0.099 91	1.908 76	11.629
2007 (Ellicott, MD)	(39.58, 76.92)	10.824	0.106 97	2.043 73	12.240
2008 (Tidewater, NC)	(36.68, 76.77)	15.393	0.103 51	1.977 53	9.878
2009 (Wakulla, FL)	(30.30, 84.42)	18.636	0.173 71	3.318 67	8.542
2010 (Newton, MS)	(32.33, 89.08)	17.064	0.104 63	1.999 02	9.585
2011 (Geneva, NY)	(42.88, 77.03)	8.724	0.098 08	1.873 88	11.844
2012 (Sellers Lake, FL)	(29.10, 81.63)	20.819	0.100 91	1.927 79	6.841
2013 (Watkinsville, GA)	(33.88, 83.43)	15.679	0.100 67	1.923 38	10.205
2014 (Molly Garen, OH)	(39.95, 83.45)	10.109	0.101 86	1.946 05	12.062
2015 (Adams Ranch, NM)	(34.25, 105.42)	11.577	0.101 27	1.934 75	9.840
2016 (Prairie View, TX)	(30.08, 96.00)	20.694	0.102 85	1.964 86	9.441
2017 (Nunn, CO)	(40.87, 104.73)	7.677	0.096 45	1.842 71	11.817
2018 (Torrington, WY)	(42.07, 104.13)	8.308	0.104 56	1.997 71	12.890
2019 (Ft. Assiniboine, MT)	(48.48, 109.80)	5.778	0.119 48	2.282 73	17.296
2020 (Mandan, ND)	(46.77, 100.92)	4.899	0.109 79	2.097 58	17.600
2021 (Lind, WA)	(47.01, 118.57)	9.505	0.134 06	2.561 24	12.204

constant soil temperature for many months in a year. It often occurred in deep depths. Some examples are shown in Fig. 3e for the data before 1996 and in Fig. 3f for data before 1997.

- 6) *Random error.* Random error is shown by data values substantially larger or smaller than their neighboring values. Examples of this kind of error are shown in Fig. 3d in late 1999, in Fig. 3g in late 1994 and 1998, and Fig. 3e in 1996.

3. A quality-control method

The quality-control method used heat transfer in soils to determine the “reference soil temperatures” and used them to identify erroneous data and calculate the estimated temperatures. In soil heat transfer and budget, the

minor effects of lateral heat exchange and heat generation by chemical and biological processes inside soil was neglected, and the vertical heat flux at the ground surface was the sole source driving soil temperature variations. Variations in soil temperature in turn affected the flux intensity. Heat flux interaction with temperature led to an equilibrium state with a soil temperature profile, the reference soil temperatures, although such a state was hardly reachable in reality because of constant variation in heat flux due to the never-resting weather.

Heat transfer in soil was achieved by conduction and diffusion. Their rates were affected by thermodynamic property of the soil and its vertical variation. For a uniform layer of soil with thermal diffusivity D , we solved the vertical heat diffusion equation with the following pair of boundary conditions:

$$T(0, t) = \bar{T} + A_0 \sin(\omega t), \quad z = 0; \quad \text{and}$$

$$T(z_\infty, t) = \bar{T}, \quad z = z_\infty$$

and obtained (see Hillel 1980)

$$T(z, t) = \bar{T} + A_0 e^{-z/d} \sin(\omega t - z/d + \phi_0). \quad (1)$$

The boundary condition at the top of the soil column ($z = 0$, the ground surface) was an idealized sinusoidal annual cycle of frequency ω ($\omega = 2\pi/360$) and amplitude A_0 around an annual mean temperature of the soil profile \bar{T} . The lower boundary ($z = z_\infty$, positive downward) was assumed at the depth where heat flux across it was negligible and $T = \bar{T}$. In (1), $T(z, t)$ is the soil temperature at depth z and time t , and the parameter d is the “damping depth,” which describes both the amplitude reduction and phase shift of a forced soil temperature variation at some depth relative to the variation of the forcing at the surface. At $z = d$, the amplitude of soil temperature variation is e^{-1} of the amplitude of

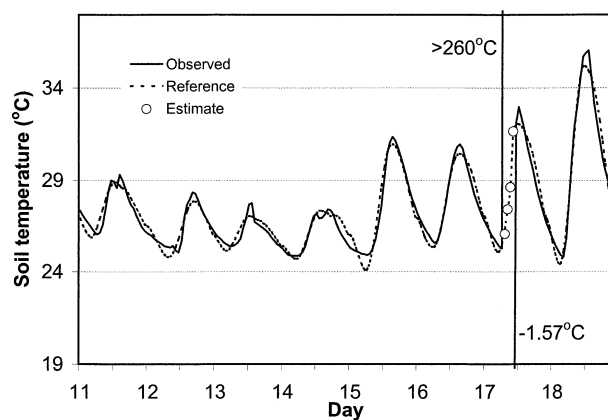


FIG. 2. Hourly ST observations (solid line) at 2 in. 11–18 Jul 1998 at the station in Newton, MS. The dashed line shows the reference temperature at the depth from (4) with the four estimates for the erroneous data on 17 Jul.

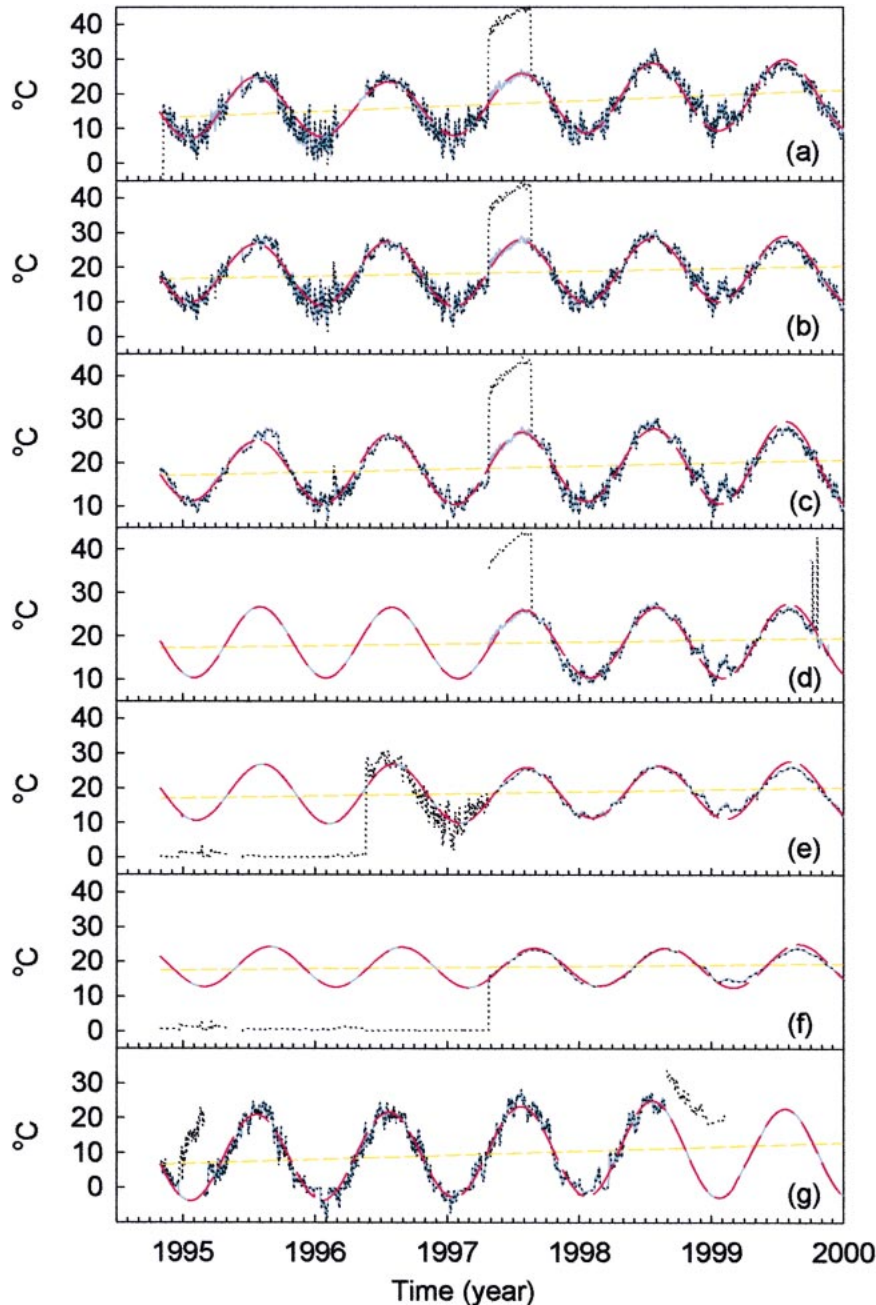


FIG. 3. Observed (dotted line), calculated annual soil temperature variation (red, dashed line), and reference soil temperature (blue, solid line) at six depths from 1994 to 1999 at the station in Newton, MS for (a)–(f) 5.08, 10.16, 20.32, 50.80, 101.60, and 203.20 cm, respectively. (g) A similar plot for 5.08-cm (2 in.) soil temperature variation at the station in Rogers Farm, NE, illustrates data error type 6 and the accuracy of quality-controlled results. The yellow, dashed line in each panel shows a trend of soil temperature.

the variation at surface, and the phase of the variation shifts by 1 rad from that at the surface. The form of d is $d = (2D/\omega)^{1/2}$, which shows that a higher-frequency wave forcing (larger ω) at the surface, such as the diurnal cycle, damps more quickly in soils than a forcing of lower frequency, such as the annual cycle. The con-

stant phase ϕ_0 aligns soil temperature variation with the forcing.

Although the diurnal solar forcing damps quickly in the soil, it significantly affects the soil temperature variations in shallow soil layers. After this effect was included in the problem, the solution of (1) became

$$T(z, t) = \bar{T} + A_y e^{-z/d_y} \sin(\omega_y t - z/d_y + \phi_y) + A_d e^{-z/d_d} \sin(\omega_d t - z/d_d + \phi_d), \quad (2)$$

where the subscripts y and d are for annual and daily variables, respectively, and ϕ_y and ϕ_d are the constant phase of average annual and diurnal variation of soil temperature, respectively.

In (2), $T(z, t)$ is an idealized value of the temperature corresponding to the annual and diurnal forcing in soils of known thermodynamic properties. It can serve as the anchor temperature at given time and depth. After the irregular weather effects to this anchor temperature are added, it becomes the theoretical reference soil temperature at depth z and time t . These reference temperatures can then be used to screen and identify erroneous data from measurements and calculate estimates for the erroneous temperatures. This is the theoretical framework of this quality-control method. Accordingly, the steps in developing this quality control are: (i) calculate the parameters in (2) for a particular site, (ii) include irregular weather effects in (2), (iii) calculate the reference temperature at various depths from (2), and (iv) apply the reference temperatures to screen the observed soil temperature, identify erroneous data, and calculate estimated temperature. Details of these steps are as follows.

a. Calculating parameters in (2)

In (2), the damping depths of annual and diurnal solar forcing were calculated from (Campbell 1985)

$$d_y = (\gamma_y K / c \pi)^{0.5}, \quad \text{and} \quad d_d = (\gamma_d K / c \pi)^{0.5}, \quad (3)$$

where $\gamma_d = 86\,400$ s, and $\gamma_y = 365\gamma_d$; K is soil thermal conductivity ($\text{W m}^{-1} \text{K}^{-1}$); and

$$K = A + B(\theta/100) - (A - E) \exp[-C(\theta/100)^4],$$

wherein $A = 0.65 - 0.78b + 0.60b^2$, $B = 1.06b(\theta/100)$, $C = 1 + 1.26(\text{percentage of clay in soil}/100)^{-0.5}$, and $E = 0.03 + 0.1b^2$. In (3), the volumetric specific heat capacity, c ($\text{J m}^{-3} \text{K}^{-1}$), is determined by $c = [2\,390\,000b/2.65 + 4\,180\,000(\theta/100)]/2$, where b is average bulk density of soil and is determined from soil texture and composition and θ is the average soil volumetric water content ($\text{m}^3 \text{m}^{-3}$). Because of θ , the damping depths are functions of soil moisture. In our calculation, the soil texture and composition for the 21 stations in the network were obtained from the USDA Soil Survey Center at Lincoln, Nebraska. Average soil moisture value was used to calculate a station's d_y and d_d . The values of d_y and d_d for the stations are given in Table 1. Effects on these values from fluctuation of soil moisture due to irregular weather and climate events will be discussed and included in section 3b.

The constant phase in the annual variation term in (2) was calculated as the phase difference between the variation described by the term without ϕ_y and the av-

eraged variation from the data. The same method was used to derive the constant phase ϕ_d in the diurnal variation term. Because the soil temperature data were not quality checked at the time of calculation, the values of ϕ_y and ϕ_d were used only as "first guess" of actual phase. They were improved later using an iterative method described in 3b(1).

The average temperature \bar{T} was composed of both annual and diurnal contributions, as well as irregular effects of weather events; that is, $\bar{T} = T_y + T_a + T_d$, where $(T_y + T_a)$ described the average annual temperature of the soil profile, and T_d is daily effect on the average. In the average annual temperature, the two components were mean surface air temperature at the site (T_y) and a correction (T_a) that converted T_y to be the mean annual temperature of the soil profile. This "complication" resulted because 1) the soil temperature data series from the stations were still too short to yield a meaningful "mean value," and 2) this mean value also varied from year to year because of climate variations. In this calculation, T_y was obtained from 1) using daily surface air temperatures from four nearest National Weather Service cooperative observing (coop) stations that had quality-controlled 30-yr daily data, 2) interpolating the daily air temperature to the soil station site through an objective analysis, and 3) using the data to calculate T_y . The derived value of T_y and average amplitude of surface air temperature variation A_y at the stations are given in Table 1, and the coop stations selected in calculation of these values at each soil site are listed in Table 2. Again, these values were different from the mean temperatures of the soil profile. This difference at individual sites was eliminated by T_a , which was calculated in a way described in 3b(3).

Last, T_a , the daily temperature effect on \bar{T} , was determined as the difference of actual average daily soil temperature and the projected value by the mean annual wave [the first two terms on the right-hand side of (2)]. Details of the calculation are described in 3b(6).

b. Including effects of weather/climate anomalies

Weather events, such as frontal passage with extended cloudiness and precipitation, alter the solar forcing on soil temperature and result in departures of daily soil temperature from a quasi-sinusoidal variation. In addition, precipitation or extended dryness can cause fluctuations of soil moisture and thermal diffusivity, both of which affect the soil damping rate and generate anomalies in amplitude and phase of temperature variations in the soil profile. Similar soil temperature departures also can result from annual climate fluctuations, such as abnormally cool and wet summers or warm and dry winters.

To account for effects of such irregular weather and climate on heat conduction and diffusion (thermal damping) in soil, we replaced $e^{-z/d}$ with $e^{-(z/d+\Delta)}$, where Δ was a moisture correction to d because of the irreg-

TABLE 2. A list of National Weather Service (NWS) cooperative stations whose long-term air temperature data were used to determine T_y in (4), and the mean distance from the set of cooperative stations to the target soil station.

Soil station	NWS cooperative stations (Nos.) used to obtain T_y in (4)				Mean distance (km)
2001	250375	254795	254815	255362	19.92
2002	217294	217502	211107	211691	24.55
2003	471875	473174	474523	474582	19.92
2004	115413	116711	113940	115079	31.04
2005	156110	156580	156595	143295	38.30
2006	410211	411000	411430	414098	28.26
2007	181862	189750	180465	185111	12.97
2008	310674	316135	311458	310375	47.71
2009	080211	088758	087025	089566	61.61
2010	226308	225776	223107	226894	21.77
2011	301152	303184	306510	307842	8.34
2012	082158	087982	082229	081978	39.84
2013	099466	090435	098950		19.76
2014	334189	334681	334979	339552	12.35
2015	296687	292096 (292095)*	291918		25.94
2016	411048	418160	411956	411889	41.69
2017	480270	481547	481675	053553	17.14
2018	488995	489925	486852	255590	30.11
2019	243110	243996	247620	243530	27.80
2020	320819	321052	325479	329455	8.34
2021	454679	457059	456039	453515	62.54

* This station was substituted for station 292096 in 1999. It is close to station 292096 but did not start collecting data until 1990.

ularities. Instead of analytically determining Δ , we calculated $e^{-\Delta}$ as a correction to the damping effect. In a similar consideration, we added correction factors to the phase constants in (2) to account for effect on phase shift from soil moisture variations resulting from weather and climate irregularities. With these corrections, (2) was rewritten as

$T(z, t)$

$$= (T_y + T_a + T_d) + \alpha_y A_y e^{-z/d_y} \sin(\omega_y t - z/d_y + \phi_y + \phi_{ya}) + \alpha_d A_d e^{-z/d_d} \sin(\omega_d t - z/d_d + \phi_d + \phi_{da}), \quad (4)$$

where $\alpha_y = e^{-\Delta_y}$ and $\alpha_d = e^{-\Delta_d}$ are the amplitude corrections and ϕ_{ya} and ϕ_{da} the phase corrections to annual and diurnal variations. Procedures determining these corrections, and the correction terms T_a and T_d in (4), are described below.

1) ANNUAL PHASE CORRECTION (ϕ_{ya})

The sequence in calculating these correction terms had to start at phase correction and then amplitude correction because the latter could not be determined when the variations had different phases. When aligning the phase, we first aligned the annual variation phase. Following this sequence, we inspected the data for a period of 12 months or longer and set any observed soil temperatures outside the range $\pm 50^\circ\text{C}$ as missing and ignored them in calculations. We then calculated the provisional reference soil temperatures for the same period from (4) with α_y and α_d set to unity, and compared them with the observed temperatures on a daily basis. Lagged correlations between the reference temperature and the

observed soil temperature variations at different depths were computed, and ϕ_{ya} for each depth was determined such that it maximized the lagged correlation. Results of this procedure for the station in Newton, Mississippi, at 5.08–203.2 cm (2–80 in.) depth are plotted in Figs. 3a–f (red, dashed line) against the observed variations in soil temperature (dotted line). They show a nearly perfect match of the phases of soil temperature variations over the years. Similar annual phase match between the adjusted reference and the observed soil temperatures was also achieved at the other stations of the network (figures are not shown).

2) ANNUAL AMPLITUDE ADJUSTMENT (α_y)

After the annual phase was adjusted, we calculated the annual amplitude adjustment α_y accounting for effects of irregular weather and climate variations on the amplitude of the annual soil temperature wave, not on the average value around which the annual wave fluctuated. The latter was described by $(T_y + T_a)$. Our inspection of data indicated that the amplitude of the reference temperature without α_y was usually larger (smaller) than the observed at shallow (deep) layers. Their difference also varied over time. Based on these observations, we derived α_y from the iteration formula

$$S(z) = \left\{ \sum_{t=a}^b [\Delta T_i(z, t) - \Delta T_i^R(\alpha_y, z, t)]^2 \right\}^{1/2}. \quad (5)$$

In (5), ΔT_i is the difference of T_y and observed annual mean soil temperature at z , and ΔT_i^R is the difference between T_y and calculated annual mean soil temperature

TABLE 3. Values of α_y for the station in Newton, MS.

Cover type	Depth [cm (in.)]	1994	1995	1996	1997	1998	1999
Bare soil	5.08 (2)	0.970	0.970	0.830	0.970	1.050	1.120
	10.16 (4)	0.940	0.940	1.080	1.050	1.000	1.080
	20.32 (8)	0.820	0.820	0.970	0.970	0.950	1.160
	50.80 (20)	1.127	1.127	1.127	1.090	1.090	1.200
	101.60 (40)	1.435	1.435	1.600	1.370	1.250	1.520
	203.20 (80)	1.707	1.707	1.707	1.690	1.530	1.900
Short grass	30.48 (12)	0.922	0.922	1.022	1.010	0.997	1.173
	50.80 (20)	1.127	1.127	1.127	1.090	1.090	1.200
	101.60 (40)	1.435	1.435	1.600	1.370	1.250	1.520

at z from (4) using an α_y . The limits of the summation are chosen such that if we used 1-yr data from 1 January to 31 December, a was 1 July in the year before the working year and b the 30 June in the following year. The extra months added to the working year helped reduce the edge effect on α_y . The final α_y was selected such that it minimized S and yielded the reference temperature variation that best resembled the observed annual soil temperature variation. In iterating (5), we used only the observed soil temperatures within $\pm 50^\circ\text{C}$ and the observed values whose difference from the reference temperature was within $\pm 10^\circ\text{C}$. Soil temperature data with a difference beyond $\pm 10^\circ\text{C}$ were considered suspicious and were not used in deriving the parameter. For a depth where there were large chunks of missing data in a year, its α_y was obtained from vertical linear interpolation of α_y for the depth above and below it. The values of α_y obtained using this method for all depths at the Newton station are given, as an example, in Table 3.

3) ANNUAL TEMPERATURE CORRECTION (T_a)

After ϕ_{ya} and α_y were computed for a station, the annual temperature correction T_a was calculated from the difference between the mean reference soil temperature obtained from (4) with ϕ_{ya} and α_y , and observed average daily soil temperature. The T_a for different years and depths for the station in Newton, Mississippi, are listed in Table 4. These values helped to eliminate bias in T_y [arising from using long-term (30-yr) average surface air temperature to represent the mean temperature of soil profile] and also restored the trend in annual mean

soil temperature erased by the long-term averaging. With these corrections, the reference soil temperature depicted the observed annual variations in soil temperatures very well (Fig. 3). They also captured a warming trend of the soil temperatures at the station, particularly at the shallow depths.

4) DAILY PHASE CORRECTION (ϕ_{da})

After (4) captured the effects of weather and climate irregularity on annual and monthly soil temperature variations, we developed methods to include the effects of diurnal variation anomalies in (4). Some examples of these anomalies were cold temperatures in early morning hours, or warm temperature in late night hours because of passage of a cold or warm front. The irregular diurnal cycles were described in (4) by ϕ_{da} , the diurnal cycle, or daily phase correction. Its value on each day was calculated in a way similar to the calculation of ϕ_{ya} in 3b(1). We first used (4) to calculate a provisional hourly reference soil temperature without ϕ_{da} for a target day and padded the same diurnal variation to either side of the day. This padding created an adequate number of sample elements, particularly for 6-h data, to calculate lagged correlations of the reference soil temperature versus the observed and to determine ϕ_{da} such that it maximized the lagged correlations.

5) DIURNAL AMPLITUDE CORRECTION (α_d)

The amplitude correction to diurnal variation in soil temperature at various depths was determined from an

TABLE 4. Values of T_a at different depths for the station in Newton, MS.

Cover type	Depth [cm (in.)]	1994	1995	1996	1997	1998	1999
Bare soil	5.08 (2)	-1.313	-1.313	-1.566	-0.486	1.817	2.317
	10.16 (4)	1.061	1.061	0.590	1.181	2.115	2.092
	20.32 (8)	0.781	0.781	0.983	1.495	2.390	2.439
	50.80 (20)	1.277	1.277	1.277	0.814	1.425	1.592
	101.60 (40)	1.393	1.393	0.799	0.942	1.898	1.931
	203.20 (80)	1.203	1.203	1.203	0.921	1.322	1.367
Short grass	30.48 (12)	0.946	0.946	1.081	1.268	2.068	2.156
	50.80 (20)	1.277	1.277	1.277	0.814	1.425	1.592
	101.60 (40)	1.393	1.393	0.799	0.942	1.898	1.931

iteration method similar to (5) in 3b(2). The correction was made to a value of A_d in the second term on the right-hand side of (4), which was calculated as average amplitude of daily temperature variation using the coop stations data described in 3a. To calculate α_d , the hourly or 6-hourly soil temperature data for a target day were examined first. Data beyond the range of $\pm 50^\circ\text{C}$ were assigned “missing,” and the hourly reference soil temperature was calculated for the day using (4). Difference between the amplitude of the reference and observed soil temperature was calculated for the hours without missing data. If the absolute value of this difference at a particular hour and depth was larger than a threshold for that depth, given in Table 5, the observed value at the hour was marked as suspicious and removed in the next round of calculation. For all the data within the threshold, α_d was determined iteratively such that it minimized the amplitude difference of diurnal soil temperature variation.

6) DAILY MEAN TEMPERATURE IN SOIL PROFILE (T_d)

After obtaining the diurnal corrections, ϕ_{da} and α_d , we recalculated T_d as the difference of the average daily soil temperature between that observed and that calculated. Comparisons between the reference and observed diurnal soil temperature variations at various depths at the station in Newton, Mississippi, are shown in Fig. 4. At most of the depths, the reference soil temperature nearly coincides with the observed soil temperature.

7) CALCULATE REFERENCE SOIL TEMPERATURE

After we included in (4) both the mean variations in soil temperatures at various depths and their effects by irregular weather and climate, it became a formula that can be used to screen soil temperature data and to provide physically coherent estimates for suspicious and/or missing soil temperatures when necessary. This capacity has been shown in Figs. 2–4. In these figures, the calculated hourly soil temperatures at 5.08 cm (2 in.) match the observed temperatures well, at the same depth. In Fig. 3, the variation in calculated daily soil temperature over the course of several years also closely matches the observed soil temperature variations. The method discriminated the unrealistically hot temperature at the shallow depths in the summer days in 1997 and provided good estimates. The reliability of these estimated values is examined and discussed in section 5 after the quality-control procedure is detailed in the next section.

4. Procedure of the quality control

In applying (4) to “raw” soil temperature data, we first set data with type-1 error defined in section 2 to -99.99 , and checked the 6-hourly data in years before 1997 for type-5 error (which happened often in those years), and assigned -99.99 to those erroneous data.

TABLE 5. Threshold temperature difference at various depths for station in Newton, MS.

Depth (cm)	5.08	10.16	20.32	50.80	101.60	203.20	30.48	50.80	101.60
(cover type)	(Bare soil)	(Bare soil)	(Bare soil)	(Bare soil)	(Bare soil)	(Bare soil)	(Short grass)	(Short grass)	(Short grass)
Threshold ΔT ($^\circ\text{C}$)	3.0	1.5	1.0	0.5	0.5	0.5	0.8	0.5	0.5

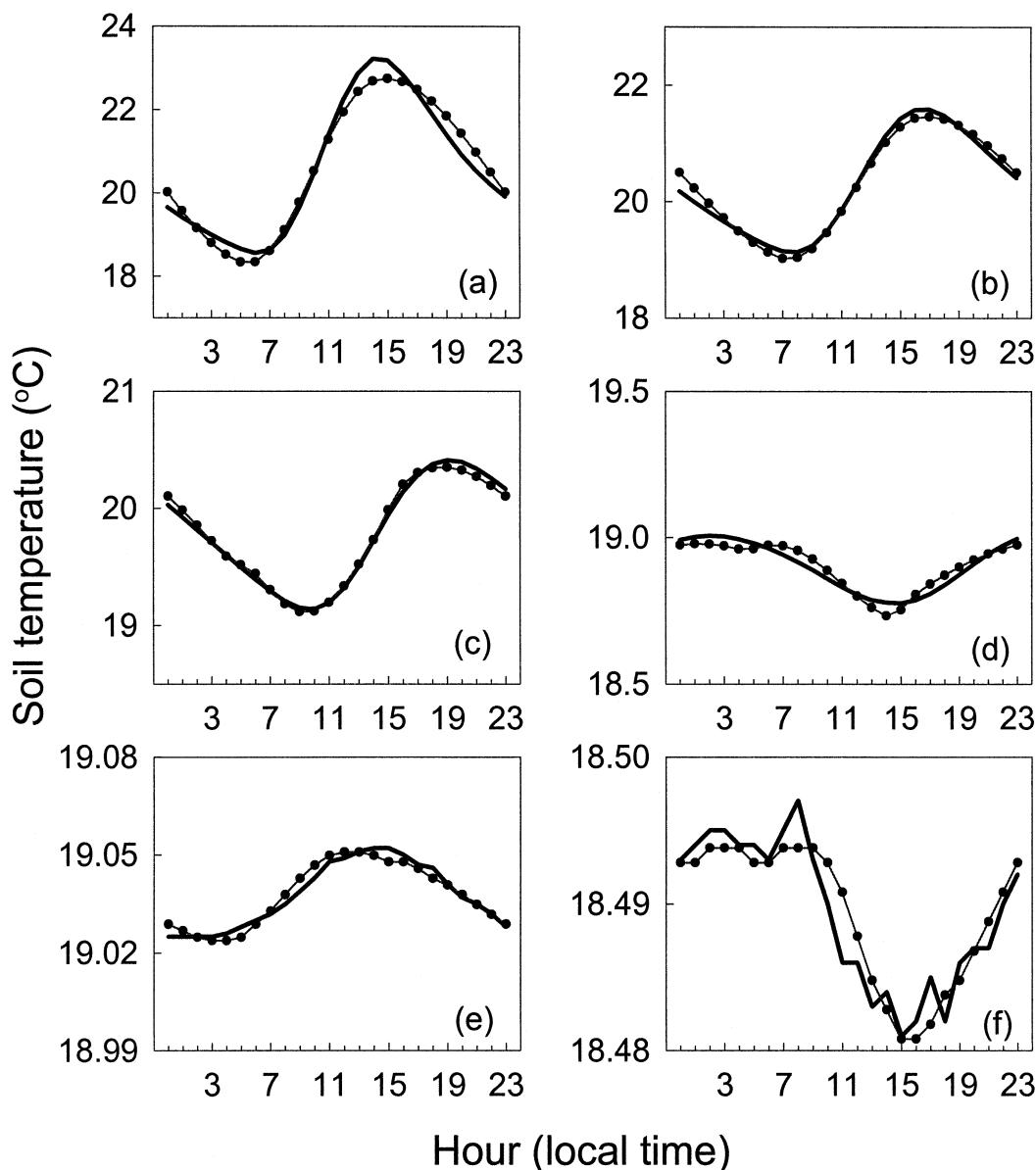


FIG. 4. Observed (solid line) and calculated reference (thin dotted line) hourly soil temperature for the station in Newton, MS: (a)–(f) are for depth 5.08, 10.16, 20.32, 50.80, 101.60, and 203.20, respectively.

After the preparation, the procedures detailed in section 3 were followed to calculate the daily reference soil temperature and construct the annual variation (wave) at the standard or any required depths. When we were calculating the annual wave, the diurnal term in (4) was not used. The daily temperature variation following the annual wave was compared with the observed temperature and their difference was computed for each day. If this difference (δ) was within $\pm 7^\circ\text{C}$ ($\pm 5^\circ\text{C}$ at deep depths), the observed daily soil temperature was accepted for that day, otherwise the observed data were considered questionable. If the questionable data continued for a period longer than a month, some systematic bias (for example, data shift by a constant value) was suspected and the questionable data

in that period were further examined. In this examination, the standard deviation of the δ series in that time segment was computed. If it was greater than 2°C the data were considered erroneous (error type 5) and were replaced by the reference temperatures. Otherwise, the data were classified as having error type 4, and we then subtracted the shift or the constant value in the data and adjusted them to be consistent with the daily and annual variations. Examples of these results are shown in Figs. 3a–d for the data in early 1997. Similar erroneous daily data, but for a period of shorter than a month, were considered as random error (error type 6). Their estimates were made using the reference temperature. Examples of such corrections are shown in Fig. 3d in several cases in late 1999.

TABLE 6. Probability (P , %) of estimated soil temperature (4) to be off from the measured soil temperature in the given ranges. The last two columns are averages of absolute and accumulated differences of estimated and measured soil temperature in kelvin.

Depth (cm)	P ($\Delta T > 2.0$)	P ($2.0 \geq \Delta T > 1.5$)	P ($1.5 \geq \Delta T > 1.0$)	P ($1.0 \geq \Delta T > 0.5$)	P ($0.5 \geq \Delta T > 0.2$)	P ($\Delta T \leq 0.2$)	def	def
0000 local time (LT)								
5.08	0.16	0.47	0.78	3.72	62.02	32.87	0.279	-0.265
10.16	0.00	0.15	0.58	1.02	46.57	51.68	0.214	-0.205
20.32	0.00	0.00	0.00	1.31	8.14	90.55	0.118	-0.108
50.80	0.00	0.00	0.00	0.17	0.66	99.17	0.027	0.014
101.60	0.00	0.00	0.00	0.00	1.17	98.83	0.018	-0.002
203.20	0.00	0.00	0.00	0.00	1.78	98.22	0.015	0.002
0600 LT								
5.08	0.16	0.00	0.00	0.16	9.92	89.77	0.103	-0.029
10.16	0.00	0.29	0.00	0.00	3.07	96.64	0.073	0.027
20.32	0.00	0.00	0.00	0.00	25.73	74.27	0.145	-0.145
50.80	0.00	0.00	0.00	0.00	0.17	99.83	0.052	-0.048
101.60	0.00	0.00	0.00	0.00	0.44	99.56	0.016	-0.010
203.20	0.00	0.00	0.00	0.00	1.48	98.52	0.008	-0.001
1200 LT								
5.08	0.62	0.31	1.40	14.42	39.53	43.72	0.304	0.168
10.16	0.00	0.00	0.00	0.29	7.59	92.12	0.083	0.008
20.32	0.00	0.00	0.00	0.15	0.44	99.42	0.041	-0.019
50.80	0.00	0.00	0.00	0.00	0.00	100.00	0.020	0.014
101.60	0.00	0.00	0.00	0.00	0.58	99.41	0.011	-0.003
203.20	0.00	0.00	0.00	0.00	1.63	98.37	0.013	0.005
1800 LT								
5.08	0.00	0.16	0.31	10.08	37.98	51.47	0.236	-0.194
10.16	0.00	0.00	0.15	0.44	5.84	93.58	0.087	-0.041
20.32	0.00	0.00	0.00	0.00	0.44	99.56	0.069	0.062
50.80	0.00	0.00	0.00	0.00	0.00	100.00	0.020	0.004
101.60	0.00	0.00	0.00	0.00	0.88	99.12	0.014	0.005
203.20	0.00	0.00	0.00	0.00	1.63	98.37	0.015	-0.006

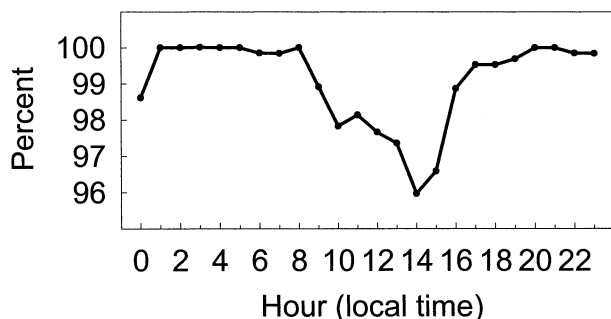


FIG. 5. Distribution of probability (%) of estimated soil temperature at 5.08 cm (2 in.) to be smaller than $\pm 1.0^{\circ}\text{C}$ from the measured soil temperature at each hour.

The next step in the procedure was to examine and estimate, when necessary, the observed hourly and 6-hourly soil temperatures on individual days. In this step, data of a 3-day segment (this period could be adjusted depending on application interest) at each depth were used to derive the parameters and construct the diurnal temperature variation term in (4), which was then used to compute hourly or 6-hourly reference soil temperatures. One caveat in this step was that when erroneous data of type 3 existed the daily phase correction ϕ_{da} could not be obtained and had to be set to zero in the calculation; otherwise it was calculated. After the diurnal term was constructed, (4) was used to calculate the reference soil temperature as estimates of the individual erroneous hourly and 6-hourly data. In Fig. 2, we show the estimated soil temperatures at 5.08 cm at the station in Newton, Mississippi, on 17 July 1998. The four estimates show consistency with the observed diurnal variation inferred by observations at the neighboring hours.

5. Result and reliability

The quality-control method was applied to the soil temperature data collected from the 21 stations in the USDA NRCS SM-ST Network from 1994 to 1999, and the quality-controlled dataset contained continuous hourly (and 6-hourly for years before 1997) soil temperatures for the multiple depths.

To quantify the reliability of this method and the confidence level of the quality-controlled soil temperature dataset, we conducted a series of tests of the method and compared the estimated soil temperatures with observed temperatures. The tests took the data selected from observation periods that had no missing or known erroneous data. These periods include both the 6-hourly observation time and hourly observation time. In these tests, we created missing data and erroneous data for selected hours and examined whether this quality-control method could detect them and the accuracy of the estimated soil temperatures from the method.

Results of the tests using hourly data are given in Table 6. The four groups in the table contain the results for the 0th, 6th, 12th, and 18th hour of a day at six "regular"

depths for the 21 stations. In each group, the second-seventh columns contain the probability for an estimated soil temperature to be different from the observed by the indicated magnitude. The last two columns show averages of absolute and accumulated difference between the estimated and observed soil temperatures, respectively.

These results show that, at the shallow depth with most varying soil temperature, 5.08 cm (2 in.), nearly 98% of the estimated (reference) soil temperature is within $\pm 1^{\circ}\text{C}$ from the observed soil temperature at the 0, 6th, 12th, and 18th hour of a day. As supported by the results in Fig. 5, this quality of estimated data also holds for the other hours of the day. In addition, in Fig. 5, the variation of accuracy shows that the estimated soil temperatures are more accurate for the evening to midmorning hours from 2100 to 0900 local time than for noon and afternoon hours. More than 98% of the estimated 5.08-cm soil temperatures are within $\pm 1.0^{\circ}\text{C}$ in the former period, and about 96% for the latter hours (more than 90% of the estimated 5.08-cm soil temperatures are within $\pm 0.5^{\circ}\text{C}$ in the former period, and more than 83% for the latter hours). This diurnal distribution of reliability could well result from the fact that more frequent storms develop in the afternoon hours and add an extra fine-scale irregularity to the temperature variation in those hours. Another possible source for this variability is that both the diurnal phase and amplitude variation in (4) cannot be accurately derived when temperature value at the normally warmest hours in a day is unavailable. At deep depths from 10.16 to 203.2 cm (4–80 in.), where soil damping alleviates the effects of these extra irregularities on soil temperature, nearly 100% of the estimated temperatures are within $\pm 1^{\circ}\text{C}$ (Table 6). Additional tests have further shown that a majority of the estimated soil temperatures at various depths are within $\pm 0.2^{\circ}\text{C}$ from the observed, except around the local noon at the two shallowest depths.

Test results of the reliability of estimates for 6-hourly data are summarized in Table 7. They are not as good as that for the hourly data; only 64% of the estimated values are within $\pm 1^{\circ}\text{C}$ between the evening to early morning hours, about 55% in the late morning, and 40% in the noon and afternoon hours. The decreasing reliability of the estimates for the 6-hourly data is primarily because of a lack of detailed information on diurnal variation in the coarse-resolution observation. Differences in the results of Tables 6 and 7 indicate that high-resolution data are essential for both accurate description and reconstruction of diurnal variations in soil temperatures.

6. Summary

In this study, we developed and tested a quality-control method for soil temperatures from 21 stations in the USDA NRCS SM-ST Network measured at six standard depths from 5.08 to 203.2 cm (2–80 in.), and three additional depths at 30.48, 50.8, and 101.6 cm (12, 20, and 40 in.) at companion sites with a different surface cover. The method used a heat diffusion model and its solution at

TABLE 7. Same as Table 6 but for 6-hour data.

Depth (cm)	P		P		P		P		P		def	def
	$(\Delta T > 2.0)$	$(2.0 \geq \Delta T > 1.5)$	$(1.5 \geq \Delta T > 1.0)$	$(1.0 \geq \Delta T > 0.5)$	$(0.5 \geq \Delta T > 0.2)$	$(\Delta T \leq 0.2)$						
0000 local time (LT)												
5.08	11.32	9.30	16.12	24.96	22.79	15.50	0.938	0.036				
10.16	1.61	4.23	12.26	30.51	25.69	25.69	0.587	-0.128				
20.32	0.15	2.62	14.39	35.17	26.31	21.37	0.588	0.386				
50.80	0.00	0.00	0.17	12.25	39.90	47.68	0.259	0.076				
101.60	0.00	0.00	0.00	0.29	25.15	74.56	0.141	-0.001				
203.20	0.00	0.00	0.00	1.04	8.28	90.68	0.107	-0.002				
0600 LT												
5.08	12.40	10.39	21.71	27.13	15.66	12.71	1.057	0.034				
10.16	4.23	6.42	17.08	28.76	23.65	19.85	0.748	0.065				
20.32	0.15	0.73	6.40	29.36	35.90	27.47	0.451	0.050				
50.80	0.00	0.00	0.00	1.99	38.91	59.11	0.187	-0.06				
101.60	0.00	0.00	0.00	0.73	27.05	72.22	0.150	0.037				
203.20	0.00	0.00	0.00	0.15	6.95	92.90	0.089	-0.001				
1200 LT												
5.08	37.52	10.85	13.64	20.47	10.23	7.29	1.777	0.958				
10.16	8.18	10.66	22.63	29.05	17.66	11.82	0.961	-0.067				
20.32	0.00	1.02	3.20	17.44	43.02	35.32	0.350	-0.112				
50.80	0.00	0.00	0.00	0.33	33.44	66.23	0.159	-0.102				
101.60	0.00	0.00	0.00	0.00	9.65	90.35	0.076	-0.033				
203.20	0.00	0.00	0.00	0.15	1.78	98.08	0.029	0.010				
1800 LT												
5.08	11.78	8.37	15.97	26.51	22.02	15.35	0.991	0.185				
10.16	3.21	8.91	22.92	32.12	19.85	12.99	0.826	0.601				
20.32	0.00	0.44	3.20	28.34	35.17	32.85	0.386	0.192				
50.80	0.00	0.00	0.00	3.15	13.25	83.61	0.129	-0.041				
101.60	0.00	0.00	0.00	0.00	1.17	98.83	0.042	0.013				
203.20	0.00	0.00	0.00	0.00	2.07	97.93	0.028	-0.003				

individual sites to identify erroneous soil temperature data and calculated the estimates for the erroneous data. Its application to the observed soil temperature data has resulted in a high-quality and reliable soil temperature dataset for the network. The dataset is accessible from the USDA National Water and Climate Center's Web site.

This method can be adapted to other soil temperature datasets, for example, the U.S. cooperative station soil temperature data, from 1950s to the present, archived at the U.S. National Climatic Data Center at Asheville, North Carolina. It can help to develop high-quality soil temperature datasets that have been least developed among the current list of climate data sources. Such datasets have been desired for multiple purposes in agricultural and industrial production practices and in climate research.

High-quality near-real time soil temperature data from this USDA NRCS SM-ST network can be used to assist decision making in agricultural production and in facility designs by utility and construction industries. In research, such data have been underutilized, but they could be used to understand soil temperature variation and the effect of annual and interannual heat storage/release anomalies in the soil column on regional- and continental-scale climate and precipitation variations. In fact, the results in Fig. 3 suggest a warming of soil temperature occurring in recent years at the station in Newton, Mississippi, and several other sites in the United States. The warming also occurred at deep soil layers, indicating a persistent trend in interannual and longer-scale variations in climate because only long-term variation signals can be recorded at those depths because of soil's natural filtering capacity. These local and regional soil temperature anomalies could significantly affect the soil water and moisture movement by changing the soil moisture potential field. Thus, both soil moisture and temperature data are needed in our understanding of land-atmosphere interaction and feedback and their role in variations in regional weather and climate.

Acknowledgments. We thank Dr. Philip Pasteris at USDA WCC for his collaboration during this work, Ms. Cynthia Hays for organizing data and her programming work in early phases of this study, Dr. Ronald Paetzold at the USDA National Soil Survey Center in Lincoln, Nebraska, for providing us data of soil type at each of the 21 sites in the network and his consulting in our calculation of soil thermal properties and parameters in the soil temperature model, and Dr. Kenneth Hubbard for useful discussions. The comments of Dr. David Kristovich and three anonymous reviewers helped to improve the clarity of this manuscript. This work was supported by USDA through Cooperative Agreement 68-7482-9-459Y with the University of Nebraska-Lincoln and by USDA Cooperative Research Project NEB-40-008.

REFERENCES

- Campbell, G. S., 1985: *Soil Physics with Basic Transport Models for Soil Plant Systems*. Elsevier, 150 pp.
- Charney, J. G., 1975: Dynamics of desert and drought in the Sahel. *Quart. J. Roy. Meteor. Soc.*, **101**, 193–202.
- Hillel, D., 1980: *Fundamentals of Soil Physics*, Academic Press, 413 pp.
- Hu, Q., and G. D. Willson, 2000: Effects of temperature anomalies on the Palmer drought severity index in the central United States. *Int. J. Climatol.*, **20**, 1899–1911.
- Hubbard, K. G., 1993: Monitoring regional drought conditions. *Drought Assessment, Management, and Planning: Theory and Case Studies*, D. A. Wilhite, Ed., Kluwer Academic, 17–30.
- Karl, T. R., and A. J. Koscienczy, 1982: Drought in the United States: 1895–1981. *J. Climatol.*, **2**, 313–329.
- Miller, K. A., 2000: Managing supply variability: The use of water banks in the western United States. *Drought: Global Assessment*, D. A. Wilhite, Ed., Routledge, 70–86.
- Palmer, W. C., 1965: Meteorological drought. U. S. Weather Bureau Res. Paper 45, 58 pp.
- Pielke, R. A., Sr., and X. Zeng, 1989: Influence on severe storm development of irrigated land. *Nat. Wea. Dig.*, **14**, 16–17.
- , and R. Avissar, 1990: Influence of landscape structure on local and regional climate. *Landscape Ecol.*, **4**, 133–155.
- Tang, M. C., 1989: *An Introduction to Theoretical Climatology*. Meteorology Press, 299 pp.