A Daily Soil Temperature Dataset and Soil Temperature Climatology of the Contiguous United States

Qi S. Hu and Song Feng

Climate and Bio-Atmospheric Sciences Group
School of Natural Resource Sciences
University of Nebraska-Lincoln
Lincoln, Nebraska 68583-0728

September 2002

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

Submitted to ***Journal of Applied Meteorology***, September 23, 2002

Abstract

Although affected by atmospheric circulation, variations in soil temperature result primarily from the radiation and sensible and latent heat exchanges at the surface and heat transfer in the soils of different thermal properties. Thus soil temperature and its variation at various depths are unique parameters useful in understanding both the surface energy processes and regional environmental and climate conditions. Yet, despite the importance, long-term quality data of soil temperatures are not available for the United States. The goal of this study is to fill this data gap and develop a soil temperature dataset from the historical data of U.S. cooperative stations. Cooperative stations' soil temperatures at various depths from 1967 to 2002 are collected and examined by a set of quality checks, and erroneous data of extended periods are estimated using methods constructed in this study. After the quality control, the data are used to describe the soil temperature climatology as well as soil temperature change in the continuous United States. The 35-year climatology shows that the annual soil temperature at 10cm depth, where most stations have soil temperature measurements, decreases gradually from 297˚K in the coastal areas along the Gulf of Mexico to below 281˚K in the U.S.-Canadian boarder. In seasonal variation, the largest seasonal change occurs from spring to summer when soil temperatures are adjusted from the cold season to the warm season particularly in snow cover regions. Mild changes are observed from autumn to winter when the soil heat storage still dominates the soil temperature variations. An analysis of the soil temperature variation reveals a warming trend in soil temperatures in most of the stations in the north and northwestern United States and a large cooling trend in some stations in the southeast United States. Significant warming is found in the winter and spring season. Potential effects of these trends on regional agriculture and environment are discussed.

1. Introduction

Among the processes affecting regional weather and climate in land areas are exchanges of energy and water between the atmosphere and the Earth's surface. The rates of these exchanges are dependent on several factors including soil temperature and moisture. Variations in soil temperature and moisture alter the partitioning of sensible and latent heat from the surface and affect atmospheric boundary-layer processes and regional circulation (Pan and Mahrt 1987; Peters-Lidard et al. 1998). Although influenced by atmospheric circulation anomalies, variations in soil temperature result primarily from the radiation and sensible and latent heat exchanges at the surface and heat transfer in the soils of different thermal properties in the vertical direction. Thus soil temperatures at various depths are unique parameters useful to describe both the surface energy processes and regional environmental and climate conditions.

Soil temperature affects the surface heat flux at various time scales. Because heat conduction in soils is a very slow process (Hillel 1980), soil heat anomalies of daily or weekly time scales in shallow layers near the surface are released to the atmosphere before being distributed to the deeper layers. Only persistent long-term, such as interannual and decadal scale, anomalies in surface heat budget can propagate to deep soil layers and affect temperature variations in those layers (Lachenbruch and Marshall 1986; Beltrami and Harris 2001; Beltrami 2002). This is shown by the relationship $z_m = c \cdot \tau^{1/2}$, which is obtained by solving the heat transfer equation in soils for the depth, z_m, at which a specific time scale variation in the surface heat process, e.g., annual solar cycle, τ, has the largest amplitude in the soil temperature variation (Tang 1989). The parameter c is a constant and proportional to the soil thermal conductivity. As this relationship indicates, soil temperature at different depths records surface temperature variations of different time scales, and effects of longer time-scale temperature and heat variations at the surface and atmosphere are eventually shown in temperature records of deeper soil layers. Because heat conduction in soil is always down-gradient, a positive heat anomaly in deep soil layers after a persistent warm period, for example, would gradually release the extra heat to shallow layers, where the heat anomaly would produce anomalies in surface heat as well as moisture fluxes through effects on landcover and vegetation conditions. By affecting the surface heat and moisture fluxes at interannual and decadal time-scales, deep soil temperature anomalies could be one of the sources affecting regional climate (Tang et al. 1982; Tang and Reiter 1986; Retnakumari et al. 2000). Thus, the potential use of soil temperature is attractive for prediction of regional circulation and climate, if interactions of soil temperature and atmospheric variations are understood.

In additional to affecting the atmospheric circulation, soil temperature anomalies at various depths also directly affect growth and yield of agricultural crops. For example, cool spring season soil temperature in shallow layers delays corn development and, on the other hand, warm spring season soil temperature contributes to increase in corn yield (Bollero et al. 1996; Hu and Buyanovsky 2002). Warm soil temperature encourages plant leaf growth, which further affects the evapotranspiration and soil water budget in crop fields (Bollero et al. 1996; Wraith and Ferguson 1994). Warm soil temperature in early spring also allows planting at earlier dates, and an extended growing season would open the possibility to grow high-yield varieties that usually take a longer time to mature. Furthermore, soil temperature anomalies affect the crop root growth; cool soil temperatures limit root expansion (Kaspar and Bland 1992) and significantly affect the root system and hence yield (McMichael and Burke 1998). Because of these effects of soil temperature on agriculture and economy, it is important to understand soil temperature variations so that both farming technology and management strategies can be improved to reduce climate risks and sustain grain production and agriculture in the changing environment.

In order to understand these roles of soil temperature in land surface processes affecting weather and climate and in production agriculture, a reliable dataset of soil temperatures is required. Existing pieces of short-term soil temperature data series are in different archives with various formats. Because of their variable quality they are difficult to use in research and application. This situation imposes a strong need for developing a quality dataset of soil temperatures from the existing sources. A part of this effort was recently completed and resulted in a high quality dataset of hourly soil temperatures at multiple depths using measurements from 21 automated stations in the USDA NRCS SM-ST Network (Hu et al. 2002). As a continuation of that previous work, this study extends the quality assurance analysis to the U.S. National Weather Service (NWS) cooperative stations daily soil temperature data and produces a quality dataset for the contiguous United States from 1967 to early 2002. This quality dataset can be used, as demonstrated in some of the previously cited studies, in analysis and understanding of the soil temperature variations and their interactions with regional circulation and their impacts on agriculture and natural resources.

In the next section, various data sources for soil temperatures from cooperative stations in the contiguous United States are discussed. Major sources of error in individual station's soil temperature data series are discussed in section 2 along with methods to detect the erroneous data and calculate their estimates. Because the data were measured at various time schedules, e.g., some stations measured daily maximum and minimum at midnight and some measured soil temperatures two times each day, one at the seventh hour local time and the other at nineteenth hour local time, and still the others measured once on a day, a method is proposed in section 3 to calculate the daily average soil temperatures which can be used for comparison and research purposes. In section 4, the new dataset of daily average soil temperatures at multiple depths from 1967 to early 2002 are used to describe the climatology of soil temperatures in the contiguous United States. Additional features, such as trend and major changes over the 36 years of record for the soil temperatures at various depths, are examined and the potential impact of these changes on agriculture are discussed in section 5. Section 6 contains a summary of this study and remarks on the significance of major results.

2. Daily soil temperature data and quality control

a. Daily soil temperature data sources

Daily soil temperature data for the contiguous United States are from three major sources: i) the TD-9639 soil temperature data archive for the period from January 1967 to December 1981, ii) the TD-3200 daily data archive from January 1982 to December 1993, and iii) stations daily soil temperature data from January 1994 to March 2002 available online at http://cdo.ncdc.noaa.gov/plclimprod/plsql/poemain.poe. All three datasets were produced at the National Climatic Data Center (NCDC) from reports of individual stations. Additional soil temperatures from cooperative stations not contained in these datasets, for example, some stations in individual states, were found from archives at state climate offices and included in this analysis. Soil temperatures in all these archives were not previously subjected to quality control although the air temperature and precipitation data in TD-3200 and the online archives were quality controlled using the method developed at NCDC (Reek et al. 1992).

According to these data sources, there were 337 stations in the contiguous United States with soil temperature measurements between 1967 and 2002. Of these 337, only 292 stations continued their measurements for longer than 5 years for at least one depth. The spatial distribution of these 292 stations is shown in Fig. 1a. The spatial distribution of the stations is quite irregular; more stations were in the Ohio and Mississippi River valleys and the states of Iowa and the Dakotas, and only a few in the southwestern and northeastern U.S. and none in Nevada and Wisconsin. Among these stations, the number of stations in service also changed over the years. As shown in Fig. 1b, more stations were installed after 1978, after a dry period in the central United States, and the number of stations peaked in the early 1980s but has since been decreasing. Figure 1b also shows that most of the stations measured soil temperature at 10cm depth, and only about 50 stations measured temperature at 100cm depth. The number of stations in service also shows an "annual cycle" in Fig. 1b because some stations, particularly in the northern states, only measured soil temperature in warm season or during the growing season for agricultural crops from mid April to late September.

At these individual stations, soil temperatures were often measured under different ground covers. A total of seven different covers were found in the records: bare ground, fallow, grass, brome grass, sod, straw muck, grass muck, and bare muck. Except for some stations not specifying their cover types, most stations have measurements under bare ground, and these measurements often have the longest record length. The measurements under other cover types at a station were sometimes made for specific experimental purposes, such as soil temperature effect on vegetation health and crop growth, and continued only for certain seasons or short time periods. So, in developing the dataset, we used the soil temperatures measured under bare ground cover, and for stations without measurements under bare ground, we used the soil temperatures under the specific ground cover that had the longest record.

At most stations, soil temperatures were measured at multiple depths, and for various reasons both English and metric units were used to define the depths at the stations. Stations using the English units measured soil temperatures at depths of 2, 4, 8, 20, and 40 inches below the surface, and stations using the metric units measured the temperature at depths of 5, 10, 20, 50, and 100 centimeters below the surface (only a few stations had measurements at 80 inches or 200cm). Although each pair of these English and metric units, e.g., 2in vs. 5cm, 4in vs. 10cm, etc, is not identical, they are very close to one another. The largest difference is 1.6cm between 40in and 100cm. Temperature differences between comparable depths are small because depth differences are small, particularly in shallow depths, and the slightly bigger differences at deeper depths will have little effect on temperature because of small temperature gradient across them. Therefore, we accepted the temperatures at depths in English unit as comparable to temperatures at the corresponding depths in metric unit and obtained soil temperatures at five metric depths: 5, 10, 20, 50, and 100cm, at each station.

Soil temperatures at different stations were measured at three different "times-of-observations" (Karl et al. 1986). Some stations measured the maximum and minimum soil temperatures at various depths at midnight; some stations measured soil temperatures twice a day, one in the early morning hours at either 7:00am or 8:00am local time and the other in the late afternoon hours at local time either 5:00pm or 6:00pm; and the remaining stations measured soil temperature only once a day fixed at one of these local hours: 7:00am, 8:00am, 12:00noon, 5:00pm, and 6:00pm.

These considerable irregularities among stations data in observation period, ground cover, depths where data were taken, and time-of-observations make it difficult, if not possible, to develop a soil temperature dataset with a consistent format for all the stations. A practical approach is to apply quality control methods to individual stations and their various time series of soil temperatures at specific time-of-observation, depths, and ground cover, and obtain a mosaic yet useful dataset of the soil temperatures. A quality control method to obtain such a dataset is described in the next subsection.

b. Data quality control

Although many methods have been developed to identify erroneous air temperatures (Hanssen-Bauer and Forland 1994; Menne and Duchon 2001; Peterson et al. 1998a, b; Vincent et al. 2002), little attention has been given to quality control of soil temperature data. Hu et al. (2002) developed a quality control method to examine the soil temperature data from the USDA NRCS SM-ST Network. In that approach, a soil temperature model was established to calculate annual and diurnal reference soil temperatures at various depths for each station. These reference temperatures are used in evaluation of observed soil temperatures to identify erroneous data and provide their estimates. In developing quality control method for the cooperative stations, we use some of the tools developed in Hu et al. (2002) and also expand their method to include additional evaluation tools. Several evaluation methods introduced in quality control for air temperatures also are used. Details of each of these evaluation tools and their application procedure to the soil temperatures are presented below.

b.1. Internal consistency check

Reek et al. (1992) outlined eight rules to identify erroneous data of all meteorological variables resulting from data reporting and digitizing, from typos, unit differences, and use of different based values in data reporting (see their Table 1). We used three of their eight rules for temperatures to check the soil temperatures for a) *internal inconsistency*, which includes cases with daily maximum temperature smaller than daily minimum, and unrealistically large or small temperature values, b) *excessive diurnal range*, which includes cases with extraordinarily large daily temperature range (T_{max}-T_{min}) even though T_{max} and T_{min} may be within their reasonable ranges, and c) *spike in temperature series*, which are defined "as the smallest absolute result from the comparison of singular differences of three consecutive days, centered on the day in question" (Reek et al. 1992). Spikes greater than 28˚C are flagged for deletion, and spikes smaller than 28˚C but greater than 22˚C are marked as suspect. Applications of these rules to individual station soil temperature series singled out the suspected data totaling 0.10% of the entire soil temperature data. These suspected data values were flagged.

b.2. Variation range check

The above internal consistency checks identified some obvious outliers in the data series, but they cannot detect erroneous data that have the following problems: the data values are much larger than neighboring values but not larger than the threshold for being detected by the consistency check in b1, data that create substantially large changes in soil temperature from the previous day or days, and data that show absence of any change over some extended period. To identify these erroneous data, we used two methods based on the two rules proposed and used by O'Brien and Keefer (1985) in their identification and elimination of erroneous data in real-time hydrological data. These methods define a) high/low bounds for soil temperature at given depths (LIM), and b) the limit for the rate of change of soil temperature at given depths (ROC). By defining the high/low bounds, the LIM method can help identify the data outside the variation range of soil temperatures at each depth. Similarly, defining the bounds for the rate of change of soil temperatures, the ROC detects data inconsistent with temporal variation in soil temperatures.

Similar methods derived from these rules also were used in quality control of daily air temperatures in Meek and Hatfield (1994; 2001).

To derive the high/low bounds in LIM for a given soil temperature data series, for example, the daily maximum soil temperature or soil temperature measured at an early morning hour at a depth, we first identified both the highest and the lowest values of soil temperatures on each calendar day of a year in the data history of up to 36 years. After getting these pairs of "extreme" values for each day, we used functions of form, , to describe the envelope of annual variation of these extreme temperatures. In this function, t is the day of year, is the frequency (=2 /365), z is depth, and and are annual mean of these "extreme" soil temperatures and their amplitude of variations, respectively. Because of the potential for some of these extreme values to be erroneous, this envelope was determined by visual inspection and embraced 98% of those daily extreme values.

After the envelope functions are derived at each station and for each observation time series, they were denoted as the upper and lower bounds of soil temperature variations, T_U and T_L, respectively, for those series and applied to compare with the individual data values. Data satisfying, , passed the LIM test, and data that failed it were flagged as suspicious. A typical result is shown in Fig. 2a for minimum soil temperature at 20cm from 1994 to 1997 at Auburn, Alabama. The two thin lines are T_U and T_L and the thick line is the soil temperature. The open circles indicate erroneous data identified by this LIM test.

To check the daily rate of soil temperature change, we used a method in Meek and Hatfield (1994; 2001) to determine the bound of the changes, similar to finding the bounds in LIM. At each station and for each time series, the soil temperature change between two consecutive days was calculated from $\Delta T(z,t) =$ and then the maximum daily rate of change was identified among all the values for the same day from the data of up to 36 years. Again, a fitting function was constructed to describe the annual variation in those "extreme" values of the rate of daily soil temperature change. However, because the daily rate of change for soil temperatures is very small in winter, particularly when a station has snow cover (Beltrami 2001), and large during spring, it was difficult to describe the variation of the extreme values except for a step function. For simplicity, we took a different approach and used a constant value as the ROC bound (no bound for minimum change). In deriving this bound, we first calculated the standard deviation (SD) of the daily soil temperature change for a data series and denoted the values larger than 8 SD as suspect. The time series of soil temperature change without the suspect values was examined and used to recalculate the SD, and the new 8 SD value was used as the ROC bound. In this method, the first SD analysis was necessary because a few large outliers in a data series can inflate the SD of the series large enough to cause acceptance of some erroneous data. Such erroneous data can be detected by the ROC criterion determined in the second SD analysis on the new data series, which has the large outliers removed.

An example of the ROC test is shown in Fig. 2b for the minimum soil temperature at 20cm from 1994 to 1997 at Auburn, Alabama. The dashed line is the ROC bound. Many of the erroneous data identified by the ROC also were found erroneous by LIM.

Both the ROC and LIM checks were applied to the entire soil temperature dataset and caused flags on 0.17% of the data (see Table 1).

b.3. Check drift in soil temperatures

After these consistency and variation range checks, the remaining possible errors in a soil temperature data series would be those embedded in the variations of the series, including changes in “average” and variance (amplitude) of soil temperatures in a period of a few months or years. In this subsection, we use “drift” to refer to such changes of short-term average soil temperatures. Sources causing such drifts are unclear due to lack of documentation. In this and the next subsection, we describe the methods that identify these two kinds of errors in the data and also provide estimates of soil temperatures in those erroneous periods.

To identify the drift in a soil temperature series, we displayed each station’s data and compared the variation of the target time series with time series at different depths. When there was no observation at other depths at the same station, the comparison was made between observations at the station and two to three neighboring stations within 200km radius. From this comparison, we identified the segment in the data series with sudden changes of the trend. A similar inspection method also was used in examining soil moisture data (Robock et al., 2000) and in automated weather station data quality control (Ashcroft et al., 1990). A result of this inspection is shown in Fig. 3a for the data at the station in Brawley, California. From 1989 to 1992, a drift of the station’s soil temperature at 20cm is clearly shown. No similar decrease was observed at the other depths of the station. Another example is shown in Fig. 3b for Estherville, Iowa. In this case, an unknown source caused the station’s soil temperature at 100cm to decrease continuously from 1979 to 1986. This drift of the soil temperature was inconsistent with temperature variations at other depths for this station (see upper panel of Fig. 3b). These and similar erroneous data identified using this method were flagged to indicate their error type. This procedure flagged 1.48% of the data with this drift error.

Because data with the drift usually continued for a period of a few months or years, as in the previous two examples, it is desirable to provide estimates of soil temperatures for some of those drift periods. Such an effort is justifiable for those periods when only the average of soil temperatures was incorrect but the temperature variance was similar to that of the neighboring periods, such as in the case at Estherville, Iowa from 1979 to 1986 (Fig. 3b). For period of drift resulting from sensor problems when both drift and error in variance occurred, such as in the case at Brawley, California from 1989 to 1992 (Fig. 3a), no estimate was attempted.

There are several ways to correct the drift. For example, we could use soil temperatures at the same depth for the period excluding the drift segment to get a mean temperature and use it as an estimate of the average temperature in the drift segment. This method may miss fluctuations in the mean temperatures in the segment, however, especially if the segment contains several years. Another way is to use average from soil temperatures at different depths observed at the same time. As showed in Hu et al. (2002, see their figure 3), the trend in soil temperatures is consistent between the shallow layers. In addition, Zhang et al. (2001) used 100-year soil temperature records from a station in Irkutsk, Russia and showed that the long-term soil temperature trend at 40cm and 80cm are nearly identical. These results suggest that the soil temperatures of different depths at the same station vary with the same trend. Thus the trend at one depth can be used as a good reference to estimate the temperatures in a drift segment at a different depth. Based on these results, we constructed the following equation to estimate the soil temperatures in drift segments:

. (1)

In the above, and are the estimated and candidate soil temperatures in the drift segment, respectively, t is time in day, and and are linear regressions of the candidate and reference average soil temperature for the drift segment, with , , and as the regression coefficients. The reference temperatures are from a different depth. The difference of average soil temperatures between candidate and reference series excluding the drift segment, , is included in (1) to account for soil temperature variation with depth.

This method was used at 16 stations where drift of soil temperatures at some depths was identified. As an example, the estimated soil temperatures for the station at Estherville, Iowa are shown in Fig. 3b. The estimated values at 100cm from 1979 to 1986 describe a consistent variation as the observed in the other depths at the station.

b.4. Check changes in soil temperature variance

A problem parallel to the drift is the change or drift of variance or amplitude of the soil temperatures. In detecting this kind of error, we used the method of "scale cumulative sum" (SCS) proposed by Peterson et al. (1998b) to measure changes in amplitude (variance) of a soil temperature time series. An example of this error and SCS test result is shown in Fig. 4 for the station in Salt Lake City, Utah. In this case, the SCS revealed significant changes in variance of the soil temperatures at 10 and 20cm depths after December 1995. Applying the SCS to the entire dataset, we found about 0.54% of the data with such errors.

Again, because such errors often continued for a period of a few months or years (Fig. 4), it is useful to provide estimated soil temperatures for those erroneous data periods. To estimate the soil temperatures with the variance error, we took an approach similar to that in b3 and used the soil temperatures at a different depth of the same station as the reference temperatures. The candidate soil temperature series at one depth was compared to the soil temperature series at other depths, or from neighboring stations if data at only one depth was observed at the target station, for their variance in the periods *excluding* the erroneous segment to select the soil temperature series that yielded the least difference in both the variance and average with the candidate station series. After identifying the reference time series, the following was used to calculate the estimate of soil temperatures in the erroneous segment:

, (2)

where, and are candidate (erroneous) soil temperature and its estimate, respectively, is the average candidate soil temperature and and are standard deviations of the candidate and reference temperature series, respectively, for the erroneous data segment, and and are standard deviation of candidate and reference temperature series for the data period excluding the erroneous data segment(s). The difference of the average soil temperatures between candidate and reference series excluding the erroneous data segment, , is included in (2) to account for soil temperature variation with depth. The estimate temperatures from (2) will have a consistent variance with the rest of the data series, as shown in Fig. 4 by the result from applying this method to the data at the station in Salt Lake City, Utah.

b.5. Procedure of the quality control

These previously described individual procedures are used in the following sequence to perform the quality control of a station's soil temperature data. The internal consistency check (b1) was used to flag inconsistent data. Then, variation range check (b2) is used to identify data outliers. Finally, errors from drift and inconsistent variance are examined (b3 and b4) in sequence to assure consistent variations in soil temperatures at a depth and observation time.

This procedure has been applied to the soil temperature dataset from the NWS cooperative stations in the contiguous United States. Erroneous data are flagged and their estimates also are calculated. The total percentage of erroneous data detected is 2.29%, and those for individual error categories are listed in Table 1. In Table 1, the percentage for ROC error was small because ROC check was applied after LIM check which flagged many erroneous data.

3. A daily mean soil temperature dataset

a. Why use a daily mean soil temperature?

In this “quality-controlled” dataset of soil temperatures, nearly all the stations’ data are for bare ground cover and at a depth from 5 to 100cm under the surface. But the stations still have different times-of-observations, i.e., daily maximum and minimum, observations at morning *and* afternoon hours, and observation at a single hour on a day. For the purpose of data quality control, it is necessary for each station to keep its original data series. For research purposes, however, these various stations data series are least desirable because they offer little common ground for direct comparison and analysis of soil temperature variation between stations and regions. On the other hand, although the station data were measured in different formats, e.g., daily maximum vs. observation at a specific time, and at different time on a day, it may still be possible to derive one measure of “daily soil temperature.” From analysis of the stations’ data series, it was apparent that a “daily mean soil temperature” at each depth is a reasonable way to “synchronize” the data. Accordingly, we defined a daily mean soil temperature at each depth and developed a method to calculate this mean temperature using stations data series.

b. Calculate daily mean soil temperatures

In this method, for stations measuring the daily maximum and minimum soil temperatures, the daily mean soil temperature is obtained as an arithmetic average of the maximum and minimum temperatures. In fact, this has been the definition for daily means (Karl et al. 1995). For stations measuring temperatures two times a day, one in an early morning hour and the other in afternoon hour, the daily mean is defined as the arithmetic average of the two temperatures. For stations measuring temperatures only once a day, no similar mean can be defined. However, because soil temperatures at depths below 50cm have very small amplitude in daily variation, these single daily measurements at some particular hours can be reasonable representations of daily mean temperatures at those depths. In the following, we present the analysis results that verify the use of the second and third definitions.

We examined the differences between the "standard" daily mean, i.e., the arithmetic average of daily maximum and minimum temperatures (T_{mm}), and the averages calculated from the second and third definitions. The data used in the evaluation were quality-controlled hourly soil temperatures from 1994-2002 at the 21 automated soil temperature stations in the USDA NRCS SM-ST Network (Hu et al. 2002). At each station, the data are at depths from 5 to 100cm under the surface. After calculating T_{mm} and the average from the two observations in the morning and afternoon hours (T_{ap}) at each depth, we examined the average difference between T_{ap} and T_{mm}, , and the average root mean square error (RMSE) of the difference, , where *m* is the number of observations at a depth and *N* is the total number of stations. Two results of these differences, the largest one at 5cm and a small one at 50cm depth, are shown in Fig. 5. The two dashed lines in each panel mark the one standard deviations of variations in the differences among individual stations. Figure 5 shows that T_{ap} calculated from temperatures at those specific morning and afternoon hours is a reasonably accurate measure of T_{mm} in both the shallow and deep depths. The largest average RMSE between T_{ap} and T_{mm} is less than 0.25˚C. At depths below 50cm the difference reduces to a few tenth of one degree of Celsius. Another result in Fig. 5 is that the T_{ap}, calculated from temperatures at those different morning and afternoon hours, approximates T_{mm} with a similar and small difference, showing a steadiness of T_{ap} in representing the daily mean temperature.

Comparisons of T_{mm} and the single measurement of soil temperature at various hours (T_{sm}) are shown in Fig. 6. Figures 6c and 6d show that at 5cm and other shallow depths at 10 and 20cm, the differences of T_{sm} and T_{mm} are large, and the largest average RMSE at 5cm is about 1.3˚C. These large differences clearly indicate that T_{sm} cannot be used to represent daily mean at those shallow depths. However, at deep depths below 50cm, because daily temperature cycle has very small amplitudes (Hu et al., 2002), T_{sm}, particularly at the two morning hours, can represent T_{mm}. For example, at 50cm depth the largest average difference between T_{mm} and T_{sm} is 0.12˚C, and the difference further reduces to a few tenths of one degree Celsius at 100cm.

In summary, these results indicate that T_{ap} calculated from the pair of morning and afternoon hour observations at the stations consistently describe the daily mean soil temperature, T_{mm}, at all depths with the maximum average error of 0.25˚C at 5cm. A single daily measurement, T_{sm}, can only represent T_{mm} at depths below 50cm with an error smaller than 0.12˚C. An explanation for T_{ap} and deep depths T_{sm} to describe T_{mm} is the soil's thermal damping on temperature variations so that the amplitude of diurnal cycle of soil temperature is small particularly in deep soil layers. In addition, the phase shift of diurnal cycle with the depth (e.g., Hillel 1980) also results in shift of the daily maximum and minimum temperatures and helps reduce the difference between T_{ap} and T_{mm}. These methods were applied to the soil temperatures at all the stations and a summary of their application is in Table 2.

4. Soil temperature climatology

From the calculated daily mean soil temperatures, we derived the annual and seasonal soil temperatures at various depths averaged over the period 1980-2002, because of more stations in service after 1980 (see Fig. 1b), and analyzed seasonal variations of soil temperatures in the contiguous United States. The average and standard deviation of annual soil temperatures at five depths, 5, 10, 20, 50, and 100cm, are shown in Fig. 7 at individual stations and also by contour lines, because of the low density and highly irregular distribution of the stations. The number of stations is highest at 10cm and reduced considerably at the other depths (see Fig. 1b).

Figure 7 shows that the annual average soil temperature decreases northward as the climate changes from subtropical humid or dry in the southern U.S. to temperate continental climate in the northern territories. At 5 and 10cm below the surface, the average soil temperature and temperature distribution are similar; the mean temperature ranges from 24˚C in the coastal areas along the Gulf of Mexico to below 8˚C in the U.S.-Canadian boarder. At 20cm, the annual soil temperature is the coolest among temperatures at all five depths. The coolest annual temperature at 20cm depth results from a combination of the warm and cold season temperature variations in the soil column. As shown in Fig. 8, the vertical profile of cold season temperature has a minimum at 20cm. In the warm season, the temperature is warmer in the shallow layers and cooler below 20cm than that at 20cm, and the decrease of temperature below 20cm is at a rate smaller than the increase of soil temperature in those depths in the cold season. A collective result of these variations yields the minimum temperature at 20cm. At 50 and 100cm depths, the warmer soil temperature in the cold season contributed to annual average soil temperatures warmer than those at 5 and 10cm depths.

Variations in the soil temperature are shown by the standard deviation in Fig. 7. Primarily because of soil's thermal damping property, the annual fluctuations are smaller than 4˚C at all five depths, with relatively large fluctuations in the central and north-central United States.

Seasonal average and standard deviation of soil temperatures are shown in Fig. 9, and they describe seasonal changes in the soil temperatures. At 5 and 10cm, freezing occurs in winter season, from December through January, in northern plains north of the 40˚N parallel from west Montana to the west of the Great Lakes and also in the northeastern U.S. Freezing disappears at most stations in central and southern U.S. at depths below 100cm (figure not shown). The warmest soil temperature occurs in summer when the soil temperatures at 10cm are as high as 34˚C in southern part of Texas. Seasonal changes in soil temperatures show the largest from spring to summer, when the 20˚C contour line "jumped" over more than 20 degrees of latitude from the southern U.S. to the U.S.-Canadian boarder. In contrast, the soil temperature change in the transition from summer to fall and from fall to winter is mild.

The standard deviations of seasonal soil temperatures in Fig. 9 show the largest standard deviations in spring and the smallest standard deviations in fall. In spring, the large standard deviations, indicating large variance of soil temperatures, has been related to fluctuations of winter snow amount in high latitude regions of the north-central U.S. but affected by alternations of contrasting weather regimes in the central U.S. Our analyses showed that winters with more snow accumulation in the northern U.S. caused cooler soil temperatures in springs due to spring melting, whereas considerably warmer soil temperatures were observed in springs following winters with less snow and snow accumulation. This snow effect on soil temperatures also contributed to the large winter season soil temperature variations in the same region (Fig. 9). In the central U.S., late snowstorms in April and early May during the warming course of temperatures caused large soil temperature variations and large soil temperature deviations in that region. In contrast, lack of similar processes in autumns leaves small temperature fluctuations in that season.

5. Discussions

In addition to be useful to describing the average conditions and spatial and seasonal variations of the soil temperatures, the developed soil temperature dataset can be examined to reveal soil temperature variations and understand changes in land surface processes as well as possible effects of these changes in agriculture and environment. Two of these features are discussed in this section.

a. Soil temperature trend

Although data length for most stations is only 35 years (1967-2002), a trend analysis of the data could still provide insight into the soil temperature variation and its relationship with changes in air temperature and precipitation. We show in Fig. 10a the linear trend of soil temperature at 10cm for 38 stations that have more than 30 years of data, and Fig. 10b the temporal variation in the temperature averaged over the 38 stations in Fig. 10a. Because the number of stations with 35-year temperature records at 100cm depth is much fewer than that at 10cm, we plotted the station average 100cm soil temperature variation and trend in Fig. 10c.

These results show that the soil temperature at 10cm depth has been warming at most stations over the 35 years with an average rate of 0.31˚C$(10yr)^{-1}$. This rate, though larger than the average of 0.10˚C$(10yr)^{-1}$ for the air temperature, is comparable to the rate of change in air temperatures in the recent decades (Karl et al. 1986). Among the stations with warming soil temperatures, the ones in the northern half of the U.S. have the largest rates. The warming trend at those stations is similar to the warming trend of their air temperature. On the other hand, among the few stations with cooling trend, the ones with largest cooling rate are in the southeast, and the cooling of soil temperatures at those stations also was consistent with decrease of the air temperature at the stations (see Fig. 4 in Karl et al., 1996). At 100cm depth, the few stations in the north-central United States all showed a warming trend at an average rate of 0.30˚C$(10yr)^{-1}$.

Superimposed on the trend of the soil temperatures in both the shallow and deep soil layers are interannual variations in the temperatures. The station-averaged variations of temperature deviation are shown in Figs. 10b and 10c (solid lines). The variations have similar phase and large amplitude from 5 to 100cm depths over the years. These variations could result from various sources including air temperature and precipitation variations, particularly winter snow anomalies and snow melting in spring. Details of these effects on interannual soil temperature variations are currently examined in a separate study.

b. Soil temperature change and agriculture

Soil temperature under bare ground is often used as a measure of the thermal condition in cropland before planting and is one of the indicators for spring planting time, because warm soil temperature is essential to seed germination and early development (Bollero et al. 1996). As shown in Figs. 11a and 11b, both the averaged winter and spring soil temperature at 5 and 10cm depths increased in the last 35 years in the contiguous United States (significant above the 95% confidence level). Such a warming trend in spring soil temperature would favor planting in earlier dates, especially in the northern half of the U.S. (e.g., Meyer and Dutcher 1998), and allow growth of high yield varieties of certain crops, e.g., corn, which require a longer growing season. Warmer soil temperature in early spring season also would encourage both leaf and root growth of crops and higher yield (Bollero et al. 1996; McMichael and Burke 1998). The soil temperature dataset developed in this work can be used to quantify changes of average planting dates for various crops and additional statistical features of these new planting dates, e.g., probability of temperature deviations from their mean. Moreover, analysis of the soil temperature data with local air temperature and precipitation could be used to evaluate the feasibility of growing new variety of crop as well as different crops.

Warming soil temperatures also can have negative impacts on agriculture. The rising winter soil temperatures would be favorable for insects to survive the winters and thus large insect populations in the following growing season. Additionally, warmer soil and air temperatures in winter and spring in the mid- and higher-latitude regions could assist fast melting of snow to reduce winter season snowpack and hence stream flow and water resources in growing season. The soil temperature dataset produced from this study can be used to analyze these various effects of soil temperature on agriculture and assist in development of practical methods to maintain and improve agricultural production.

6. Summary

A dataset of soil temperatures is developed for the continuous United States. The soil temperature data were from the U.S. NWS cooperative stations. Each station has soil temperatures at up to five depths of 5, 10, 20, 50, and 100cm below the surface and up to 35 years from 1967 to 2002. These station data were examined for consistency in both temporal variations and variations at various depths. In addition, the erroneous data continuing for extended periods were provided with estimates using methods developed in this study. This dataset not only extends the climatic data array to include the soil temperatures but also opens opportunities for acquiring knowledge of soil temperature variations and their relationship with changes in air temperature and precipitation and effect of soil temperature change on regional agriculture and environment.

Taking this advantage, we explored soil temperature climatology using a daily mean soil temperature derived from the "quality-controlled" dataset. The climatology shows both the soil temperatures at various depths average over each station's record history (up to 35 years) and the dynamic aspects of the temperatures during transitions between seasons. The former described a north-south gradient across the contiguous U.S., and the latter revealed a large change of soil temperature from spring to summer. In the winter months from December to February, soil temperature is below freezing in the north-central U.S. and the northern Great Plains in shallow layers above 100cm. At 100cm depth, the winter temperature is below freezing only at a few stations in high latitude of the northern U.S. In the vertical direction across the soil column at each station, the annual average soil temperature is coolest at 20cm below the surface as a result of different annual soil temperature variations in the shallow and deep soil layers. Soil temperatures at depths below 20cm are as warm as and at many stations warmer than that above 20cm because of the heat storage capacity and slow heat release at the deeper layers.

Soil temperature at most stations has shown a trend of warming in the last 35 years, a result similar to the observed in the air temperature changes. The average warming rate is 0.3˚C $(10yr)^{-1}$ and comparable to the warming rate of air temperature. A few stations in southeastern U.S. show a large trend of cooling. These trends in soil temperatures are significant in both the shallow and deep layers. Potential impacts of the changing soil temperatures on agriculture are outlined based on relationships of warming soil temperature in spring and earlier planting dates of various crops.

In addition to these applications, the developed soil temperature dataset can be used in studies leading to improved understanding of, for example, the thermodynamic process in soils and the relationship of soil temperature and surface heat flux variations. This relationship could be used to examine the soil temperature variations described by regional circulation or climate models and validate and also improve models' ability in describing the surface energy processes.

Acknowledgements

We thank Dr. Kenneth G. Hubbard for useful discussions during the work. This work was partially supported by the U.S. National Climatic Data Center and by USDA Cooperative Research Project NEB-40-008.

References

Ashcroft, G.L., G.D. McCurdy, G.E. Bingham, and E. Malek, 1990: Quality control of automated-climate-station data. ASAE Paper No. 904052, 6pp.

Beltrami, H., 2002: Earth's long-term memory. *Science*, **297**, 206-207.

Beltrami, H., 2001: On the relationship between ground temperature histories and meteorological records: a report on the Pomquet station. *Global Planet. Change*, **29**, 327-348.

Beltrami, H., and R.N. Harris, 2001: Foreword: inference of climate change form geothermal data. *Global Planet. Change*, **29**, 149-152.

Bollero G. A., D.G. Bullock and S. E. Hollinger, 1996: Soil temperature and planting date effects on corn yield, leaf area, and plant development. *Agron. J.*, **88**, 385-390.

Hu, Q., S. Feng, and G. Schaefer, 2002: Quality control for USDA NRCS SM-ST network soil temperatures: A method and a dataset. *J. Appl. Meteor.,* **41**, 607-619.

Karl, T.R., C.N. Williams, P.J. Young, and W.M. Wendland, 1986: A model of estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States. *J. Climate Appl. Meteor*., **25**, 145-160.

Karl, T.R., R.W. Knight, and N. Plummer, 1995: Trends in high frequency climate variability in the twentieth century. *Nature*, **377**, 217-220.

Karl, T.R., R.W. Knight, D.R. Easterling, and R.G. Quayle, 1996: Indicies of climate change for the United States. *Bull. Amer. Meteor. Soc*., **77**, 279-292.

Kaspar, T.C., and W.L. Bland, 1992: Soil temperature and root growth. *Soil Science*, **154**, 290-299.

Lachenbruch, A., and B. V. Marshall, 1986: Changing climate: geothermal evidence form permafrost in the Alaskan Arctic. *Science*, **234**, 689-696.

McMichael B.L., and J.J. Burke, 1998: Soil temperature and root growth. *HortScience*, **33**, 947-951.

Meek, D.W., and J. L. Hatfield, 1994: Data quality checking for single station meteorological databases. *Agric. For. Meteor.*, **69**, 85-109.

Meek, D.W., and J. L. Hatfield, 2001: Single station quality control procedures. In proceedings of *Automated weather stations for applications in agriculture and water resources management: current use and future perspectives.* Eds. K.G. Hubbard and M.V.K. Sivakumar, Lincoln, Nebraska, USA, 6-10 March 2000, 123-132.

Meyer, S.J., and A.L. Dutcher, 1998: Soil Temperatures and Spring Planting Dates. UNL NebGuide G98-1362. (http://ianrwww.unl.edu/pubs/fieldcrops/g1362.htm).

Pan, H.-L., and L. Mahrt, 1987: Interaction between soil hydrology and boundary-layer development. *Bound.-Layer Meteor.*, **38**, 185-202.

Perters-Lidard, C.D., E. Blackburn, X. Liang, and E.F. Wood, 1998: The effect of soil thermal conductivity parameterization on surface energy fluxes and temperatures. *J. Atmos. Sci.*, **55**, 1209-1224.

Peterson, T.C., and 20 coauthors, 1998a: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.,* **18**, 1493-1517.

Peterson, T.C., R. Vose, R. Schmoyer, and V. Razuvaev, 1998b: Global historical climatology network (GHCN) quality control of monthly temperature data. *Int. J. Climatol.,* **18**, 1169-1179.

Reek, T., S.R. Doty, and T.W. Owen, 1992: A deterministic approach to the validation of historical daily temperature and precipitation data form the cooperative network. *Bull. Amer. Meteor. Soc.*, **73**, 753-762.

Retnakumari, K., G. Renuka and G. Rao, 2000: Relation between pre-monsoon soil temperature and monsoon rainfall in a tropical station. *Mausam*, **51**, 365-366.

Robock, A., and 8 coauthors, 2000: The global soil moisture data bank. *Bull. Amer. Meteor. Soc.*, **81**, 1281-1299.

Tang, M., and E.R. Reiter, 1986: The similarity between the maps of soil temperature in U.S. and precipitation anomaly of the subsequent season. *Plateau Meteorology*, **5**, 293-307.

Tang, M., Q. Zhong, and S. Wu, 1982: Brief introduction to a long-rang forecasting methods for rainfall anomalies in China. *Arch. Met. Geophys. Biocl. (ser A)*, **31**, 349-359.

Tang, M., T. Li, J. Zhang, and C. Li, 1989: The operational forecasting of total precipitation in flood seasons (April to September) of 5 years (1983-87). *Adv. Atmos. Sci.*, **6**,289-300.

Wraith, J.M., and A.H. Ferguson, 1994: Soil temperature limitation to water use by field-grown winter wheat. *Agron. J.*, **86**, 974-979.

Zhang, T., R.G. Barry, D.Gilichinsky, S.S. Bykhovets, V.A. Sorokovikov, and J. Ye, 2001: An amplified signal of climatic change in soil temperatures during the last century at Irkutsk, Russia. *Climatic Change*, **49**, 41-76.

Tables

Table 1: Percentages of erroneous data by categories.

Inconsistency data	0.10%
Violating LIM rule	0.15%
Violating ROC rule	0.02%
Drift in average soil temperature	1.48%
Drift in soil temperature variance	0.54%
Total	2.29%

Table 2: Total number of daily mean soil temperatures at five depths and percentages of the daily mean soil temperatures calculated using T_{mm}, T_{ap}, and T_{sm}.

Depth	Number of observations (daily mean × stations)	T_{mm} (%)	T_{ap} (7am+ 5pm) (%)	T_{ap} (7am+ 6pm) (%)	T_{ap} (8am+ 5pm) (%)	T_{ap} (8am+ 6pm) (%)	T_{sm} (7am) (%)	T_{sm} (8am) (%)	T_{sm} (12Noon) (%)	T_{sm} (5pm) (%)	T_{sm} (6pm) (%)
5cm	692074	66.16	2.86	4.48	22.59	4.92	--	--	--	--	--
10cm	1991490	87.73	1.04	1.89	7.64	1.71	--	--	--	--	--
20cm	824914	71.15	2.40	4.21	18.10	4.14	--	--	--	--	--
50cm	634032	2.16	1.06	0.81	2.94	0.06	4.59	34.57	0.01	33.51	14.19
100cm	515452	1.69	1.30	0.99	3.60	0.07	5.65	37.97	0.00	31.86	11.86

Figure Legends

Figure 1: (a) Distribution of the 292 U.S. cooperative stations (dots). Open cycles are USDA SM-ST Network soil temperature stations whose data are used in this study. (b) Variations of number of cooperative stations measuring soil temperatures.

Figure 2: (a) Minimum daily soil temperature variation at 20cm (solid line) and its T_U and T_L derived from the LIM method (gray thin lines), and (b) daily rate of change of the daily minimum temperature (solid line) and the ROC bound (dashed line). The open circles in (a) and (b) are the suspected data points. The data are from Auburn Agronomy Farm station (32.6˚N, -85.5˚W) in Auburn, Alabama.

Figure 3: (a) Soil temperatures at 10 and 20cm depths at the station (33.0˚N, -115.5˚W) in Brawley, California. Compared to the temperature at 10cm, the temperature at 20cm shows a drift from 1989 to 1992. (b) Soil temperatures at 50 and 100cm depths at the station (43.5˚N, -94.8˚W) in Estherville, Iowa. A drift of soil temperature is shown in 100cm data from 1979 to 1987. The estimated values of soil temperature for the drift period are shown by the dashed-line.

Figure 4: Soil temperatures at 10 and 20cm from a station (40.8˚N, -112.0˚W) in Salt Lake City, Utah, showing a change in variance or amplitude of the soil temperatures after 1995. The estimated values of soil temperature for the drift period are shown by the dashed-line.

Figure 5: (a) Stations average difference between 5cm T_{mm} and T_{ap}, which is calculated as arithmetic average from observations at the two local hours indicated in the legend of the abscissa. The two dashed-lines show one standard deviation from the average. (b) The same as in (a) but for average RMSE (see text for details). (c) and (d) are the same as (a) and (b), respectively, but for temperature at 50cm depth.

Figure 6: Same as in corresponding panels in Fig. 5 but for difference between T_{mm} and T_{sm}

measured at different local time indicated in the legend of the abscissa.

Figure 7: Annual mean and standard deviation of soil temperatures at five depths. Contour interval is 2˚C for annual mean temperature and 0.5˚C for standard deviation. Individual stations' values also are shown by the scales marked in the lower-left corner of each panel.

Figure 8: Vertical soil temperature profiles for February, August, and annual average.

Figure 9: Monthly mean and standard deviations of 10cm soil temperature for January, April, July, and October. Individual stations' values also are shown by the scales marked in the lower-left corner of each panel.

Figure 10: (a) Trend in soil temperatures at 10cm depth, in ˚C$(10yr)^{-1}$. The scales are marked in the lower-left corner of the figure. (b) Temporal variations of 10cm soil temperature averaged over the stations in (a). (c) Temporal variations of 100cm soil temperature averaged over stations with record length longer than 30 years. The dashed-lines in (b) and (c) show the trend of the variations.

Figure 11: Trend of average 10cm soil temperature variation for 1967-2002 (see text for details).