Estimation of surface air temperature over central and eastern Eurasia from MODIS land surface temperature

Suhung Shen\textsuperscript{1,2} and Gregory G Leptoukh\textsuperscript{1}

\textsuperscript{1} NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
\textsuperscript{2} George Mason University, USA

E-mail: Suhung.shen@nasa.gov

Received 23 June 2011
Accepted for publication 10 October 2011
Published 1 November 2011
Online at stacks.iop.org/ERL/6/045206

Abstract
Surface air temperature ($T_a$) is a critical variable in the energy and water cycle of the Earth–atmosphere system and is a key input element for hydrology and land surface models. This is a preliminary study to evaluate estimation of $T_a$ from satellite remotely sensed land surface temperature ($T_s$) by using MODIS-Terra data over two Eurasia regions: northern China and fUSSR. High correlations are observed in both regions between station-measured $T_a$ and MODIS $T_s$. The relationships between the maximum $T_a$ and daytime $T_s$ depend significantly on land cover types, but the minimum $T_a$ and nighttime $T_s$ have little dependence on the land cover types. The largest difference between maximum $T_a$ and daytime $T_s$ appears over the barren and sparsely vegetated area during the summer time. Using a linear regression method, the daily maximum $T_a$ were estimated from 1 km resolution MODIS $T_s$ under clear-sky conditions with coefficients calculated based on land cover types, while the minimum $T_a$ were estimated without considering land cover types. The uncertainty, mean absolute error (MAE), of the estimated maximum $T_a$ varies from 2.4$^\circ$C over closed shrublands to 3.2$^\circ$C over grasslands, and the MAE of the estimated minimum $T_a$ is about 3.0$^\circ$C.

Keywords: surface air temperature, land surface temperature, remote sensing, land cover types

1. Introduction
Surface air temperature ($T_a$) is a critical variable in the energy and water cycle of the Earth–atmosphere system and is a key input element for hydrology and land surface models. It is a very important variable in agricultural application and climate change studies. For example, $T_a$ is used to monitor regional evapotranspiration in the biophysical processes and model crop growth for food security assessment [1], local climate and diseases [2, 3] and climate change [4], etc.

Traditionally, $T_a$ is obtained from meteorological stations at $\sim 2$ m height above the ground. In general, meteorological stations are sparsely distributed, especially in rural areas. Information about spatial variation of $T_a$ is often limited. Satellite data have synoptic spatial coverage compared to the discrete spatial distribution of weather stations, thus attracting researchers to use satellite data to fill in the gaps inherent in weather station data. A different variable, the land surface temperature ($T_s$) or surface skin temperature, is measured \textit{in situ} or retrieved from satellite measurements. Observations have shown that $T_a$ and $T_s$ are significantly different in terms of magnitude and diurnal cycle due to the different heat capacities between air and land, and weather conditions [5]. On the other hand, $T_a$ and $T_s$ are correlated due to heat exchange between the land and the air. Recent studies have shown that estimated minimum $T_a$ ($T_{a\text{,min}}$) values from nighttime MODIS land surface temperature ($T_{s\text{,night}}$) values are statistically meaningful over Africa, Mississippi State (USA) and in Alpine areas, but the estimated maximum $T_a$ ($T_{a\text{,max}}$) values from daytime MODIS $T_{s\text{,day}}$ values have large errors in some regions [6–8]. Studying the MODIS $T_a$ and $T_s$ relationship over Africa, Vancutsem \textit{et al} [6] found that the
differences between daytime $T_a$ and maximum $T_a$ ($T_{a,\text{max}} - T_{a,\text{day}}$) vary significantly by region and season. The paper indicated that no good relation was observed between ($T_{a,\text{day}} - T_{a,\text{max}}$) and normalized difference vegetation index (NDVI) or solar zenith angle (SZA) that could be used operationally over Africa.

Over northern Eurasia, especially within the central and eastern Asia region, dramatic climatic, environmental and socioeconomic changes have occurred during the past century. These include significant changes in land cover and land use, floods and droughts, dust storms, desiccation of the Aral Sea, disintegration of the Soviet Union and formation of independent countries in Central Asia. International research programs, such as Northern Eurasian Earth Science Partnership Initiative (NEESPI) and Monsoon Asia Integrated Regional Study (MAIRS), have been created to study the climate, land use and environment changes, and human activities. In supporting the research of the NEESPI and MAIRS programs, the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) has collected relevant satellite remotely sensed and modeled atmosphere, land and ocean products. More recently, 1 km resolution MODIS land products, such as eight-day $T_a$, monthly vegetation index and yearly land cover types over the Asian Monsoon region have been added into their collections. It is extremely useful to have higher resolution, such as 1 km, $T_a$ for local climate and environmental change studies in the central and eastern Asia region to account for significant change in land cover type in recent years.

The purpose of this study is to develop a method for estimating $T_a$ from satellite retrieved $T_a$ over two Eurasian regions: northern China and the former Union of Soviet Socialist Republics (fUSSR). The accuracy or statistical errors of the estimated $T_a$ have been investigated by validating the results against the data from ground measurements.

2. Data

2.1. Station data

In this study, $T_a$ come from ground meteorological stations over two northern Eurasia regions: (a) dry and semiarid northern China (75°E–132°E, 33°N–52°N), and (b) fUSSR (50°E–75°E, 55°N–60°N) as marked on figure 1. The daily maximum and minimum $T_a$ values of year 2009 from 75 World Meteorological Organization (WMO) stations over northern China are downloaded directly from the China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn/). The data from fUSSR are three-hourly measurements at 58 stations for year 2000, which were obtained from the US National Climate Data Center (NCDC), NOAA (www.ncdc.noaa.gov). The daily maximum and minimum $T_a$ of the fUSSR stations are defined as the maximum and minimum values from the valid three-hourly $T_a$ values during a day. Both datasets are highly quality controlled.

2.2. MODIS data

The study used MODIS daily $T_a$, monthly vegetation index (VI), and yearly land cover type (LCT) from USGS Land Processes Distributed Active Archive Center (LP DAAC). MODIS data for 2009 over northern China and for 2000 over fUSSR were extracted at the nearest point to a station location. The MODIS $T_a$ data have a large amount of missing data due to cloud, heavy aerosols and gaps between satellite swath, etc. Some previous similar studies have used MODIS eight-day $T_a$ products [6, 8]. Considering that the value of the eight-day $T_a$ was calculated by averaging all valid points under clear-sky only and the number of data points participating in the average could vary from 1 to 8, it may introduce uncertainty due to sampling if comparing eight-day $T_a$ (clear-sky) and eight days averaged $T_a$ (all-sky). Therefore, we have decided to use daily $T_a$ product for comparing clear-sky days only instead of using time averaged $T_a$. There are no year 2000 LCT data available, therefore the 2001 LCT data were used here, assuming no significant LCT change from 2000 to 2001.

MODIS daily $T_a$ at 1 km resolution of Collection-5 from Terra (MOD11A1.005) is a Level-3 global product. The MODIS $T_a$ is derived from two thermal infrared bands: 31 (10.78–11.28 μm) and 32 (11.77–12.27 μm) using a split-window algorithm that corrects for atmospheric effects and emissivity using a look-up table based on global land surface emissivity in the thermal infrared bands [11]. Some validation studies have shown that the accuracy of MODIS $T_a$ is less than 1 K at 1 km resolution under clear-sky conditions with known emissivity, higher errors may occur at large view angles and in semiarid regions [12, 13]. Another validation study indicated that nighttime MODIS $T_a$ are underestimated by 2–3 K at some locations [14]. Higher errors in arid regions may exist due to effects of heavy dust aerosols and greater uncertainties in classification-based surface emissivities. The error in $T_a$ contaminated by clouds and heavy aerosols can be very large (4–11 K or even larger). The $T_a$ that are severely contaminated by clouds and heavy aerosols were removed from MODIS Collection-5 products using empirical constraints on temporal variations in clear-sky $T_a$. However, it is very difficult for
the MODIS cloud-mask to discriminate all pixels affected by clouds and heavy aerosols from clear-sky pixels, particularly cirrus, near cloud edges, and/or with sub-pixel clouds [12]. The accuracy of \( T_s \) is affected by the wind speed and at highly reflective surfaces in the visible/near-infrared range, such as snow/ice cover, as well [11].

The MODIS-Terra (MOD13A3.005) monthly vegetation index product at the same spatial resolution as \( T_s \) is used. The product contains both the normalized difference vegetation index (NDVI) and the enhanced vegetation index (EVI), retrieved from blue, red, and near-infrared reflectance, centered at 469 nm, 645 nm, and 858 nm channels, respectively [15]. The MODIS NDVI and EVI products are computed from atmospheric corrected bi-directional surface reflectance values that have been masked for water, clouds, heavy aerosols, and cloud shadows. The monthly NDVI are used in this study.

MODIS yearly LCT at 500 m resolution (MCD12Q1.005) is produced by combining the inputs from both MODIS-Terra and -Aqua. The product includes five different land cover classification schemes, derived through a supervised decision-tree classification method. This study uses the primary land cover scheme that identifies 17 land cover classes defined by the International Geosphere Biosphere Programme (IGBP), including 11 natural vegetation classes, three developed and mosaicked land classes, and three non-vegetated land classes [16].

The standard MODIS 1km resolution \( T_s \), VI and 500 m resolution LCT data files are \( 10^3 \times 10^3 \) tile-based and gridded in the sinusoidal projection. In this study, for easy use, the MODIS data were stitched for the study region and reprojected with the nearest point method onto cylindrical equidistant projection by using the MODIS Reprojection Tool (MRT) [17] provided by LP DAAC. As shown in figure 1, stations over northern China and fUSSR are located in different land surface environments. Most stations over northern China are near barren and sparsely vegetated areas, grasslands, and croplands, while most stations over fUSSR are close to crop and natural mosaic, mixed forest, and croplands.

3. Time series comparison

For each station, \( T_s \) values are extracted from the 1 km daily files at the nearest grid point for the daytime and nighttime, respectively. The data have a significant amount of missing points due to cloud, heavy aerosols, gaps between satellite swaths, and failed retrieval under certain conditions. On average, the daytime \( T_s \) has only about 44% valid points over northern China, and about 38% valid points over fUSSR; the nighttime \( T_s \) has about 56% valid points over north China, and about 40% over fUSSR.

Time series of \( T_{s,\text{max}} \) and \( T_{s,\text{min}} \) (black lines) are plotted together with MODIS \( T_{s,\text{day}} \) and \( T_{s,\text{night}} \) (blue dots) for each station. For most stations, \( T_{s,\text{day}} \) is higher than \( T_{s,\text{max}} \) during the summer time, but is opposite during the winter time. Similarly, \( T_{s,\text{night}} \) is higher than \( T_{s,\text{min}} \) during the summer time and is reversed during the winter time. The amount of difference \( (T_{s,\text{day}} - T_{s,\text{max}}) \) varies from one station to another. It is noticed that \( (T_{s,\text{night}} - T_{s,\text{min}}) \) is smaller than \( (T_{s,\text{day}} - T_{s,\text{max}}) \) over most stations.

Figure 2 shows plots of sample time series at three stations over different LCTs; two stations are in northern China and one is in fUSSR. It is clear that \( (T_{s,\text{day}} - T_{s,\text{max}}) \) are very different at the three stations. Station A (ID = 51644 at 83.07°E, 41.71°N) is located over the barren or sparsely vegetated land, with maximum monthly NDVI of 0.14. \( T_{s,\text{day}} \) at station A is significantly higher than \( T_{s,\text{max}} \) during the summer and the largest difference exceeds 15°C. The time series of station B (ID = 54027 at 119.40°E, 43.98°N) looks different from that over station A. \( T_{s,\text{day}} \) and \( T_{s,\text{max}} \) have no significant differences in the summer at this station. Station B is over the urban area surrounded by grasslands and croplands, with maximum monthly NDVI of 0.69. The third time series is at station C (ID = 22200028264 at 65.40°E, 58.10°N), which is located in the evergreen needleleaf forest, with maximum monthly NDVI of 0.73. \( T_{s,\text{day}} \) at this station is similar to \( T_{s,\text{max}} \) during the summer, while it is colder than \( T_{s,\text{max}} \) during the winter. There are no \( T_s \) data in the first two months at station C as the MODIS daily data are available since 5 March 2000. In all three cases, the difference between \( T_{s,\text{night}} \) and \( T_{s,\text{min}} \) seems smaller than the difference between \( T_{s,\text{day}} \) and \( T_{s,\text{max}} \) and is not sensitive to the season.

4. Statistical relations between \( T_s \) and MODIS \( T_s \)

Statistical relations between \( T_s \) and MODIS \( T_s \), such as correlation coefficient (\( r \)), mean difference (mean), standard deviation (STD), mean absolute error (MAE), as well as linear regression coefficients (slope, offset) have been calculated for each station. It was found that (a) the correlation coefficients are high, ranging from 0.83 to 0.99 at 0.001 significance level; (b) the linear regression coefficients vary significantly from one station to another between \( T_{s,\text{day}} \) and \( T_{s,\text{max}} \), but not between \( T_{s,\text{night}} \) and \( T_{s,\text{min}} \).

The MODIS LCT data for each station location were used to figure out any connection between the slope change and LCT. Classes of nine grid points centered at the station were extracted from the IGBP classification. The LCT of a station is defined as the class at the nearest grid point to the station location and the other eight points are used to check its surrounding environment.

Analyses have indicated that the slope of the linear regression between \( T_{s,\text{day}} \) and \( T_{s,\text{max}} \) is closely associated with the LCT where a station is located. The value of slope is relatively low if a station is near barren land, while the value is higher if a station is near forest or croplands. Table 1 lists the statistical values for stations grouped by LCT. Over northern China, among 75 stations, about a half of the stations are located in urban and built-up areas (urban). The other stations are mainly located in croplands, grasslands, barren land, and closed shrublands. There are only two or fewer stations over other LCTs, and therefore, their statistical values were not computed. The differences of the statistics among groups are significant for the daytime, but not for the nighttime. The averaged \( (T_{s,\text{day}} - T_{s,\text{max}}) \) varies from 6.1°C over the barren
land to 0.6 °C over the croplands. The MAEs vary from 6.7 °C over barren land to 3.6 °C over closed shrublands.

Over urban stations, the slope varies widely from 0.61 to 0.96. The urban stations over northern China are surrounded by different LCTs. It is noticed that the slope is relatively smaller at urban stations near dryer and less vegetated areas. To find if there is a good relation between the slope and vegetation condition, NDVI values at the nearest point are extracted from MODIS 1 km monthly data. It is found that, over urban stations, the slopes and the maximum monthly mean NDVI have positive correlations. Similar relationships are found between slope value and yearly accumulated NDVI. Thus, the urban stations over northern China are divided further as urban-1 (NDVI < 0.4) and urban-2 (NDVI ≥ 0.4).

Over fUSSR, 58 stations are mainly distributed over six LCTs: evergreen needleleaf forest, mixed forest, woody savannas, croplands, urban and cropland/natural vegetation mosaic. There are only two or fewer stations over other types and no statistics were computed. Differently from northern China, in general, \( T_{\text{day}} \) is lower than \( T_{\text{max}} \) over fUSSR. The
largest mean difference occurs over the evergreen needleleaf forest. The differences of statistical values among groups are smaller for the nighttime than the daytime, which is similar to what was found over northern China.

Both \( T_a \) and \( T_s \) have strong seasonal cycles and their seasonal cycles are very similar, as shown in figure 2. This may be due to strong energy exchange between the land surface and the near surface air while \( T_a \) and \( T_s \) depend highly on the seasonal change of the solar irradiance. The \( T_a \) is also influenced by weather conditions, such as winds, humidity, and cloud, etc. What is the correlation if the annual cycle is removed? To examine it, we have selected two stations over different LCTs. Station #1 (ID = 50949 at 124.87°E, 45.08°N, croplands), has a higher slope value (0.95) between \( T_a \) and \( T_s \). Station #2 (ID = 51644 at 83.07°E, 41.72°N, barren), has a lower slope value (0.78). The lower slope value indicates that the difference between \( T_a \) and \( T_s \) depends more on the season. The daily \( T_a \) climatology was calculated by using 30 years of data from 1980 to 2009. The correlation coefficient at station #1 is 0.78 with 154 degrees of freedom (DF) for \( T_a \) and \( T_s \), and \( T_a \) and \( T_s \) anomalies, and 0.61 with 205 DF for \( T_a \) and \( T_s \) anomalies, respectively. For station #2, the correlation coefficient is 0.31 with 190 DF for \( T_a \) and \( T_s \) anomalies, and 0.24 with 237 DF for \( T_a \) and \( T_s \) anomalies. At both stations, the correlations between \( T_a \) anomaly and \( T_s \) anomaly pass the 0.001 significance level, although the values are smaller compared to the original data (0.95–0.98), especially for stations at barren land. When performing the correlation significance test, \( T_a \) and \( T_s \) anomalies are assumed in normal distribution and made null hypothesis. This indicates that the high correlations between \( T_a \) and \( T_s \) are composed by the low frequency annual cycle and high frequency signals. To better estimate \( T_a \) from \( T_s \), the input data sample group should be considered to have both an annual cycle and high frequency signals. Thus, the statistical relationships between \( T_a \) and \( T_s \) should be evaluated by using data from an entire year, not their anomaly or from a season only.

Table 1. Statistical relations for A: \( (T_a^{max} \) and \( T_a^{min} \), and B: \( (T_s^{max} \) and \( T_s^{min} \) over (a) northern China, (b) USSR.

| IGBP groups | Stn num | \( r \) | Mean diff (°C) | STD (°C) | MAE (°C) | Slope (a) | Offset (b) |
|-------------|---------|-------|---------------|---------|---------|-----------|------------|
|             | A       | B     | A       | B       | A       | B       | A        | B        |
| (a)         |         |       |         |         |         |         |     |         |  
| Shrub (6)   | 2       | 0.96  | 0.96    | -2.77  | -0.38  | 3.54     | 3.50     | 3.58     | 2.74     | 0.87     | 0.96     | 0.31     | -0.23    |
| Grasslands (10) | 10   | 0.94  | 0.95    | -4.74  | 0.42   | 5.33     | 3.93    | 5.96     | 3.08     | 0.76     | 0.90     | 0.45     | 0.30      |
| Croplands (12) | 21   | 0.97  | 0.96    | -0.58  | 0.00   | 4.65     | 3.98    | 3.62     | 3.11     | 0.85     | 0.91   | 2.14     | 0.08      |
| Urban (13)   | 36    | 0.96  | 0.97    | -2.87  | -0.74  | 5.12     | 3.82    | 4.65     | 3.11     | 0.80     | 0.94     | 1.13     | -0.72     |
| Barren (16)  | 3     | 0.93  | 0.94    | -6.05  | -0.08  | 5.71     | 4.21    | 6.73     | 3.28     | 0.71     | 0.91     | 2.71     | 0.24      |
| Urban-1      | 18    | 0.96  | 0.96    | -4.18  | -1.18  | 5.23     | 3.81    | 5.40     | 3.23     | 0.78     | 0.94     | 0.79     | -1.17     |
| Urban-2      | 18    | 0.96  | 0.96    | -1.57  | -0.22  | 4.67     | 3.76    | 3.91     | 2.96     | 0.84     | 0.94     | 1.35     | -0.20     |
| All          | 75    | 0.95  | 0.96    | -2.64  | -0.32  | 5.27     | 3.93    | 4.63     | 3.11     | 0.80     | 0.93     | 1.39     | -0.28     |
| (b)          |       |       |         |         |         |         |     |         |  
| Tree         | 3     | 0.97  | 0.96    | 3.74   | -0.41  | 3.93     | 4.09    | 4.15     | 3.18     | 0.87     | 0.92     | 4.83     | -0.35     |
| Mixed tree   | 10    | 0.97  | 0.95    | 3.60   | -0.88  | 4.04     | 4.32    | 4.16     | 3.34     | 0.87     | 0.88     | 4.75     | -0.73     |
| Savannas     | 3     | 0.96  | 0.95    | 1.65   | -0.73  | 5.14     | 4.49    | 3.97     | 3.65     | 0.79     | 0.88     | 4.02     | -0.67     |
| Croplands    | 8     | 0.96  | 0.95    | 1.04   | 0.24   | 5.07     | 3.91    | 3.88     | 2.96     | 0.79     | 0.88     | 3.58     | 0.42      |
| Urban        | 6     | 0.97  | 0.95    | 0.62   | -0.44  | 5.07     | 4.16    | 3.80     | 3.12     | 0.79     | 0.89     | 3.24     | -0.18     |
| Crop         | 24    | 0.96  | 0.95    | 2.05   | 0.16   | 4.81     | 4.16    | 3.91     | 3.12     | 0.81     | 0.89     | 4.10     | 0.31      |
| All          | 58    | 0.96  | 0.95    | 2.21   | -0.20  | 4.80     | 4.24    | 3.98     | 3.21     | 0.82     | 0.89     | 4.15     | -0.04     |

Figure 3 shows scatter plots of \( T_a \) and \( T_s \) for three LCT groups. The slope values are (a) smallest for the barren group, (b) medium for the croplands group, and (c) largest for the mixed forest group. The correlation between \( T_a \) and \( T_s \) is high for all groups, varying from 0.93 for the barren group to 0.97 for the croplands and the mixed forest group. Over the barren land, the \( T_a \) values are higher than the \( T_s \) values when \( T_a \) is more than about 10°C. The \( T_a \) values are much higher than the \( T_s \) values in the summer over the barren land, while \( T_a \) is slightly lower than \( T_s \) in the winter. The mixed forest group is very different from the barren group. The \( T_a \) is lower than \( T_s \) for most data points in the mixed forest group, and the \( T_a \) is higher than \( T_s \) only when the temperature reaches about 40°C.

5. Estimation of surface air temperature

As documented in section 4, the linear correlations between \( T_a \) and MODIS \( T_s \) are significant. Therefore, it is reasonable to estimate \( T_a \) with a linear regression equation:

\[
T_a' = aT_s + b,
\]

where \( T_a' \) is the estimated \( T_a \), \( a \) and \( b \) are coefficients. For estimating \( T_s \), \( a \) and \( b \) were calculated from all stations in each LCT group as shown in table 1. For estimating \( T_s \), \( a \) and \( b \) were calculated from all stations in each geographic region as they are independent of LCT within the geographic region.

Comparing regions of northern China and USSR, besides those of class ‘urban’, ‘croplands’ is the only class type that appears in both regions. The regression coefficients of this class differ slightly for the two regions. To have a clear picture of the difference in \( T_s \) when using linear relationships trained from two different regions, a set of \( T_s \) is calculated from \( T_s' \). In general, the differences between the \( T_s' \) values calculated by using coefficients from the two regions are less than 10%. \( T_s' \) is lower than \( T_s \) during the summer time by about 4°C when
The estimated T_a is 40.0°C, but is higher than T_a during the winter time by 7.0°C when T_a is −30.0°C.

T_a was calculated daily for each station where the LCT is listed in Table 1. If the LCT at a station is not listed in Table 1, T_a at this station was not computed and was set as undefined.

Validation of estimated T_a was carried out for each LCT group in the two regions, as listed in Table 2. The mean (T_a – T_a') is zero. Over northern China, the STD of (T_a – T_a') varies from 3.1°C over closed shrublands to 4.1°C over grassland; the corresponding MAE varies from 2.4 to 3.2°C. The STD of (T_{a,max} – T_{a,max}') over fUSSR varies from 3.4°C over evergreen needleleaf forest to 3.7°C over crop and natural mosaic lands, and the corresponding MAE is from 2.5 to 2.9°C. The STDs of (T_{a,min} – T_{a,min}') are 3.8°C over northern China and 4.0°C over fUSSR, and their MAEs are about 3.0°C. The estimated T_a is plotted against observed T_a for all stations (Figure 4). The data points are distributed nicely along the diagonal line, indicating that the mean error is zero. The averaged uncertainty for all stations is about 3°C (STD) and 3.0°C (MAE). The yearly averaged uncertainties of the estimated maximum T_a and minimum T_a are similar. The uncertainty is slightly smaller in the winter than in the summer for the estimated maximum T_a, while it has no seasonal difference of uncertainty for the estimated minimum T_a.

6. Conclusions and discussion

The statistical relationships between station-measured T_a and MODIS remotely sensed T_a were examined over two different geographic regions in central and east Eurasia: dry and semiarid northern China, and boreal fUSSR. 75 stations over northern China and 58 stations over fUSSR were used in this study. The results are summarized as follows.

(a) MODIS-Terra measured T_{a,day} and T_{a,night} have high correlations with T_{a,max} and T_{a,min}, respectively. The correlation between T_{a,night} and T_{a,min} is slightly higher than that between T_{a,day} and T_{a,max}.

(b) Linear regression between T_{a,max} and T_{a,day} shows that it depends significantly on LCT. The slope varies from 0.71 over barren land to 0.87 over forest. A lower slope value indicates that the difference between T_{a,day} and T_{a,max} is more season dependent. For example, at 40.0°C T_{a,day}
The estimated $T_{a,\text{max}}$ is $31.0^\circ$C over barren land ($T_{a,\text{day}} - T_{a,\text{max}} = 9.0^\circ$C), but is $39.6^\circ$C over forest ($T_{a,\text{day}} - T_{a,\text{max}} = 0.4^\circ$C). For $-40.0^\circ$C $T_{a,\text{day}}$, the estimated $T_{a,\text{max}}$ is $-25.6^\circ$C over barren land ($T_{a,\text{day}} - T_{a,\text{max}} = -14.4^\circ$C), but is $-30.1^\circ$C over forest ($T_{a,\text{day}} - T_{a,\text{max}} = -9.9^\circ$C).

(c) There is no significant difference among linear coefficients of $T_{a,\text{day}}$ and $T_{a,\text{min}}$ for different LCTs. Thus, to estimate minimum $T_{a}$, we have used linear regression coefficients derived from all stations for each geographic region. However, for the maximum $T_{a}$ estimation, we have used linear regression coefficients derived from stations grouped by LCT. With such a method, the uncertainty of the estimated minimum $T_{a}$ is about $3.9^\circ$C (STD) or $3.0^\circ$C (MAE), and the uncertainty of the estimated maximum $T_{a}$ varies from $3.1^\circ$C (STD) or $2.4^\circ$C (MAE) over closed shrublands to $4.1^\circ$C (STD) or $3.2^\circ$C (MAE) over grasslands. Averaging all stations, there is no significant difference in the uncertainty of the estimated minimum and maximum $T_{a}$.

The question is why does the difference between MODIS $T_{a,\text{day}}$ and $T_{a,\text{max}}$ vary with LCT? $T_{a}$ and $T_{s}$ are different in principle [5]. The satellite remotely sensed $T_{a}$ is a measure of the surface radiation. $T_{s}$ was calculated from surface emissivities, which are sensitive to LCT, especially during daytime. The heat capacity or specific heat of air is about $1.0 \text{ kJ kg}^{-1} \text{ K}^{-1}$, and varies only slightly with air temperature. However, the specific heat varies significantly from one LCT to another. For example, the specific heat is $0.8 \text{ kJ kg}^{-1} \text{ K}^{-1}$ for dry soil, $1.48 \text{ kJ kg}^{-1} \text{ K}^{-1}$ for wet soil, $0.84 \text{ kJ kg}^{-1} \text{ K}^{-1}$ for stone, and $4.19 \text{ kJ kg}^{-1} \text{ K}^{-1}$ for fresh water. The variation of the difference between $T_{a,\text{day}}$ and $T_{a,\text{max}}$ may due to the different heat capacities or specific heats of LCTs. For the extreme cases, the barren land has much lower heat capacity than the air, while the forest may have higher heat capacity than the air. The heat capacity changes with temperature, which may result in different relations in the winter and summer over the same LCT.

With the method in this study, it is possible to estimate daily 1 km resolution $T_{a}$ by using MODIS $T_{s}$ data under clear-sky conditions. The data are useful for application studies at local scales, such as temperature related diseases, crop growth, etc. The uncertainty of the estimated $T_{a}$ is about $3.8^\circ$C (STD) or $3.0^\circ$C (MAE), averaged from all stations, which may be considered large for some applications. More research on reducing uncertainty is needed. The following two paragraphs address some sources of estimated $T_{a}$ uncertainty. The calculation of $T_{a}$ takes into account the surface–atmosphere interactions and energy fluxes between the atmosphere and the ground [18, 19]. The accuracy of calculating $T_{a}$ depends on a number of conditions, such as solar zenith angle, accuracy of atmospheric correction, surfaces of high reflectance, such as snow/ice and desert, surface wind, and humidity, etc. In addition, the observation times at a location have a window of about 2–3 h that may contribute to the uncertainty of ($T_{a} - T_{s}$) due to diurnal variations of the temperature. The Terra’s nadir passes the Equator at about 10:30 am and 10:30 pm local solar time. For a location that is away from the Equator, the observation time varies. For example, at the location (59.5°N, 70°E) the MODIS-Terra observation times are between 10:10 am–1:10 pm and 8:20 pm–11:20 pm. In general, the temperature changes rather quickly during these time periods, especially in the morning. The $T_{a}$ could increase by up to about $13 \text{ K}$ over a desert area from 10:00 am to 1:00 pm [20].

Another important source of the uncertainty may come from the cloud and aerosol contaminated data points. The error in $T_{a}$ contaminated by clouds and heavy aerosols can be very large (4–11 K or even larger). The $T_{a}$ values that are severely contaminated by clouds and heavy aerosols were removed from Collection-5 Level-3 MODIS $T_{a}$ products using empirical constraints on temporal variations in clear-sky $T_{a}$. However, it is very difficult for the MODIS cloud-mask to discriminate all pixels affected by clouds and heavy aerosols from clear-sky pixels, particularly near cloud edges, cirrus and/or with sub-pixel clouds. Higher errors in some arid regions are apparent due to the effects of heavy dust aerosols and greater uncertainties in classification-based surface emissivities [12].

One limitation of estimating $T_{a}$ from MODIS $T_{s}$ is that the MODIS $T_{s}$ product is retrieved by using thermal infrared data.
that are available in clear-sky conditions only. The large gaps can be partly filled by using data from multiple sensors, such as MODIS-Aqua with clear understanding of the Terra–Aqua time difference contributing to $T_a$ uncertainty. The microwave techniques work for the all-sky condition. However, it is not possible to retrieve global land surface temperature at an accuracy of 1–2 K by using microwave techniques alone due to the much higher variations in the surface emissivities of land surfaces in the microwave range and the dependence of microwave brightness temperatures on surface roughness and structures [11]. An alternative method may be to use MODIS $T_a$ as an input to an assimilation model for areas under clear-sky conditions and the model forecast $T_a$ values for areas under cloudy atmospheric conditions.

Acknowledgments

This work was supported by NASA ROSES 2008 NNH08ZD-A001N-LCLU. The iUSSR station data are from NOAA NCDC provided by Dr Pavel Y Groisman. The authors are grateful to the NASA GES DISC S4PA team for preprocessing the data used in this research. Part of the visualization and initial analysis in this study was conducted by using the NASA Giovanni online data system [21, 22].

References

[1] De Wit A J W and Van Diepen C A 2008 Crop growth modeling and crop yield forecasting using satellite-derived meteorological inputs Int. J. Appl. Earth Observ. Geoinform. 10 414–25
[2] Lowen A C, Mubareka S, Steel J and Palese P 2007 Influenza virus transmission is dependent on relative humidity and temperature PLoS Pathogens 3 1470–6
[3] Fuhrmann C 2010 The effects of weather and climate on the seasonality of influenza: what we know and what we need to know Geogr. Compass 4 718–30
[4] Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt K B, Tignor M and Miller H L (ed) 2007 Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge: Cambridge University Press)
[5] Jin M and Dickinson R E 2010 Land surface skin temperature climatology: benefitting from the strengths of satellite observations Environ. Res. Lett. 5 044004
[6] Vancutsem C, Ceccato P, Dinku T and Connor S J 2010 Evaluation of MODIS and surface temperature data to estimate airm temperature in different ecosystems over Africa Remote Sensing Environ. 114 449–65
[7] Colombi A, Michele C, Pepeand M and Rampini A 2007 Estimation of daily mean air temperature from MODIS LST in Alpine areas EARSeL eProceedings vol 6, pp 38–46
[8] Mostouey G V, King R L, Reddy K R, Kakani V G and Filippova M G 2006 Statistical estimation of daily maximum and minimum air temperatures from MODIS LST data over the state of Mississippi J. Geophys. Res. 101 73–80
[9] Leptoukh G, Csiszaz I, Romanov P, Shen S, Loboda T and Gerasimov I 2007 NASA NEESPI data center for satellite remote sensing data and services global and planetary change Environ. Res. Lett. 2 045009
[10] Shen S, Leptoukh G, Loboda T, Csiszaz I, Romanov P and Gerasimov I 2009 The NASA NEESPI data portal to support studies of climate and environmental changes in non-boreal Europe Regional Aspects of Climate–Terrestrial–Hydrologic Interactions in Non-boreal Eastern Europe ed Y P Groisman and S Ivanov (Berlin: Springer)
[11] Wan Z 1999 MODIS land-surface temperature algorithm theoretical basis document http://modis.gsfc.nasa.gov/data/atbd/atbd_land11.pdf
[12] Wan Z and Li Z-L 2008 Radiance-based validation of the V5 MODIS land-surface temperature product Int. J. Remote Sens. 29 5373–95
[13] Coll C, Wan Z and Galve J M 2009 Temperature-based and radiance-based validations of the V5 MODIS land surface temperature product J. Geophys. Res. 114 D20102
[14] Wang W, Liang S and Meyers T 2008 Validating MODIS land surface temperature products using long-term nighttime ground measurements Remote Sens. Environ. 112 623–35
[15] Huete A, Justice C and Leeuwen W 1999 MODIS vegetation index (MOD13) Algorithm Theoretical Basis Document, Version 3 (http://modis.gsfc.nasa.gov/data/atbd/atbd_m13.pdf)
[16] Strahler A, Muchoney D, Borak J, Friedl M, Gopal S, Lambin E and Moody A 1999 MODIS Land Cover Product Algorithm Theoretical Basis Document (ATBD), Version 5.0 (http://modis.gsfc.nasa.gov/data/atbd/atbd_Lc12.pdf)
[17] Land Processes DAAC 2008 MODIS Reprojection Tool User’s Manual (http://lpdaac.usgs.gov/tools/modisreprojectiontool)
[18] Mannstein H 1987 Surface energy budget, surface temperature and thermal inertia Remote Sensing Applications in Meteorology and Climatology (NATO ASI Ser. C: Math. Phys. Sci. vol 201) ed R A Vaughan (Dordrecht: Reidel) pp 391–410
[19] Sellers P J, Hall F G, Asrar G, Strebel D E and Murphy R E 1988 The first ISLSCP field experiment (FIFE) Bull. Am. Meteorol. Soc. 69 22–7
[20] Pinker R T, Sun D, Miller M and Robinson G J 2007 Diurnal cycle of land surface temperature in a desert encroachment zone as observed from satellites Geophys. Res. Lett. 34 L11809
[21] Acker J and Leptoukh G 2007 Online analysis enhances use of NASA earth science data EOS Trans. AGU 88 14–7
[22] Berrick S W, Leptoukh G, Farley J D and Rui H 2009 Giovanni: a web service workflow-based data visualization and analysis system IEEE Trans. Geosci. Remote Sens. 47 106–13