How can we use MODIS land surface temperature to validate long-term urban model simulations?

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2013JD021101

Key Points:
- Three practical sampling methods are proposed and compared for model validation
- Cloud existence and distribution are the key factors for method selection
- SCM method is recommended for long-term simulations with the large cloud impact

Correspondence to:
L. Hu,
huleiqiu@ku.edu

Citation:
Hu, L., N. A. Brunsell, A. J. Monaghan, M. Barlage, and O. V. Wilhelmi (2014), How can we use MODIS land surface temperature to validate long-term urban model simulations?, *J. Geophys. Res. Atmos.*, *119*, 3185–3201, doi:10.1002/2013JD021101.

Received 28 OCT 2013
Accepted 25 FEB 2014
Accepted article online 3 MAR 2014
Published online 27 MAR 2014

How can we use MODIS land surface temperature to validate long-term urban model simulations?

Leiqiu Hu[1], Nathaniel A. Brunsell[1], Andrew J. Monaghan[2], Michael Barlage[2], and Olga V. Wilhelmi[2]

[1]Department of Geography, University of Kansas, Lawrence, Kansas, USA, [2]National Center for Atmospheric Research, Boulder, Colorado, USA

Abstract High spatial resolution urban climate modeling is essential for understanding urban climatology and predicting the human health impacts under climate change. Satellite thermal remote-sensing data are potential observational sources for urban climate model validation with comparable spatial scales, temporal consistency, broad coverage, and long-term archives. However, sensor view angle, cloud distribution, and cloud-contaminated pixels can confound comparisons between satellite land surface temperature (LST) and modeled surface radiometric temperature. The impacts of sensor view angles on urban LST values are investigated and addressed. Three methods to minimize the confounding factors of clouds are proposed and evaluated using 10 years of Moderate Resolution Imaging Spectroradiometer (MODIS) data and simulations from the High-Resolution Land Data Assimilation System (HRLDAS) over Greater Houston, Texas, U.S. For the satellite cloud mask (SCM) method, prior to comparison, the cloud mask for each MODIS scene is applied to its concurrent HRLDAS simulation. For the max/min temperature (MMT) method, the 50 warmest days and coolest nights for each data set are selected and compared to avoid cloud impacts. For the high clear-sky fraction (HCF) method, only those MODIS scenes that have a high percentage of clear-sky pixels are compared. The SCM method is recommended for validation of long-term simulations because it provides the largest sample size as well as temporal consistency with the simulations. The MMT method is best for comparison at the extremes. And the HCF method gives the best absolute temperature comparison due to the spatial and temporal consistency between simulations and observations.

1. Introduction

Approximately half of the world's population now lives in the metropolitan areas, and urban expansion is expected to continue [*Centre for Health Development World Health Organization*, 2010]. The combined impacts of global climate change, urban heat islands (UHI), urbanization, and demographic trends place more urban residents at risk to heat-related stressors. More accurate information on urban meteorology, high-resolution weather forecasting, and climate change projections for cities are therefore required [*National Research Council*, 2012].

Atmospheric models are helpful for understanding the current climate system and projecting future climate changes. Regional- and local-scale atmospheric model simulations are especially valuable for decision making at the municipal level; however, they necessitate accurate representations of specific urban features (e.g., morphology) and fine-scale climatic background information and ideally are coupled to urban canopy models (UCMs). UCMs have advanced rapidly during recent decades [*Grimmond et al.*, 2010] and have been coupled to a variety of atmospheric models [*Chen et al.*, 2011; *Oleson et al.*, 2008; *Kusaka and Kimura*, 2004] for use in numerous applications. For example, *Chen et al.* [2011] integrated the Weather Research and Forecasting model (WRF) [*Skamarock and Klemp*, 2008] coupled with the UCM of *Kusaka et al.* [2001] at 1 km spatial resolution to assess the impacts of urbanization on Houston's sea breeze circulation and found that the city enhances stagnant air at nighttime, which can exacerbate pollution. Despite the utility of high-resolution urban climate modeling, the heterogeneous surface structures in the urban environment, with multiple sources of energy and water, increase the complexity of models and consequently can increase the uncertainty of model accuracy.

Climate simulations over urban regions are usually validated with in situ measurements [*Grimmond and Oke*, 2002; *Masson*, 2006]. Large collaborative programs that integrate different types of measurements at multiple temporal and spatial scales are becoming popular in urban climate studies [*Mestayer et al.*, 2005;

Grimmond, 2006]. Nevertheless, the expense, location, and duration of in situ measurement programs limit their widespread application for long-term model simulations and validation. Moreover, the heterogeneity of the urban environment cannot be fully represented by the limited footprint of measurements, which complicates attempts to analyze the spatial agreement of observations with model output.

Accurate estimation of radiative temperature is required for any coupled climate modeling system, because it is an integral component of the surface energy budget, water stress evaluation, and soil moisture-climate feedback [*Bodas-Salcedo et al.*, 2008; *Ghent et al.*, 2010]. Land surface temperature (LST) products from satellite remote-sensing platforms are widely available and are potential sources for model validation, having comprehensive spatial coverage, comparable scale with models, and constant periodicity. There are many ways to apply satellite LST data to evaluate and improve atmospheric model simulations at multiple temporal and spatial scales [*Jin et al.*, 1997; *Sohrabinia et al.*, 2012; *Miao et al.*, 2009; *Leroyer et al.*, 2011]. For example, *Jin et al.* [1997] compared skin temperatures simulated with NCAR Community Climate Model (CCM2) coupled with a biosphere-atmosphere transfer scheme with satellite-derived radiative temperature to evaluate model performance globally. LST can also be incorporated into data assimilation, helping to constrain uncertainty in model simulations. For instance, *Ghent et al.* [2010] found that the LST bias of the Joint UK Land Environment Simulator was reduced by two thirds by applying observed LST by means of data assimilation. The limits of satellite remotely sensed LST for improving model performance in the long term are unknown, especially in complex urban environments.

There are several aspects to consider before comparing LST with urban climate model simulations. First, the distribution of clouds across the urban area is heterogeneous and varies seasonally [*Hu and Brunsell*, 2013]. The cloud-screening algorithms for satellite images usually are not perfect for detecting all cloudy pixels, so cloud-contaminated pixels exist in some "cloud-free" products. The data that are unavailable due to clouds would otherwise contain important information for model validation, and the impact of clouds is often larger during daytime than nighttime [*Hu and Brunsell*, 2013]. Second, the satellite LST is a directional radiative temperature, representing the information from a certain view angle [*Voogt and Oke*, 2003]. The anisotropy of LST largely exists due to differences in surface properties, the urban geometric structure, and the relative geometry between the Sun and the sensor. Additionally, the view angle relates to the atmospheric path length corresponding to absorption and reemission of radiation, leading to larger anisotropic effects for large view angles in the remotely sensed LST. Although view angle effects from the atmosphere are considered in LST retrieval algorithms [e.g., *Wan and Dozier*, 1996], the surface structure induced anisotropy of LST still exists [*Voogt and Oke*, 1998; *Rasmussen et al.*, 2010; *Vinnikov et al.*, 2012; *Lagouarde et al.*, 2004]. Third, different sensors have different temporal and spatial resolutions, as well as data acquisition times. Adequately addressing these issues can allow for satellite LST data to be comparable with model simulations of surface radiative temperature. However, little research on this topic has been done previously.

This paper addresses the question of how to use satellite remote-sensing LST to validate surface radiative temperature in an urban canopy model. We propose three practical methods, attempting to decrease the uncertainty of the satellite LST product and to meet a variety of different demands of validations depending on the application. The methods are tested using daily LST from the Moderate Resolution Imaging Spectroradiometer (MODIS) for the decade spanning 2003 to 2012, compared with simulations of surface radiative temperature from the High-Resolution Land Data Assimilation System (HRLDAS) [*Chen et al.*, 2007] over Greater Houston, Texas, U.S. The details of MODIS and the HRLDAS simulations are described in section 2, and the three sampling methods are discussed in section 3. The temperature distributions and spatial analysis of the MODIS data and simulations for each method are presented in section 4, as well as the justification of the assumptions made in the sampling methods. The advantages and disadvantages of the three sampling methods and their scope of applications are discussed in section 5, and conclusions are drawn in section 6.

2. HRLDAS and MODIS Data

2.1. Study Area

We choose Houston, Texas, as the study area, defining the domain from 29.25°N and 96°W (southwest) to 30.5°N and 94.75°W (northeast). Figure 1 shows the HRLDAS model domain and land use categories based on the U.S. Geological Survey (USGS) National Land Cover Database [*Homer et al.*, 2004; *Fry et al.*, 2011]. Three urban types are classified in the domain, including light-intensity urban (LU), heavy-intensity urban

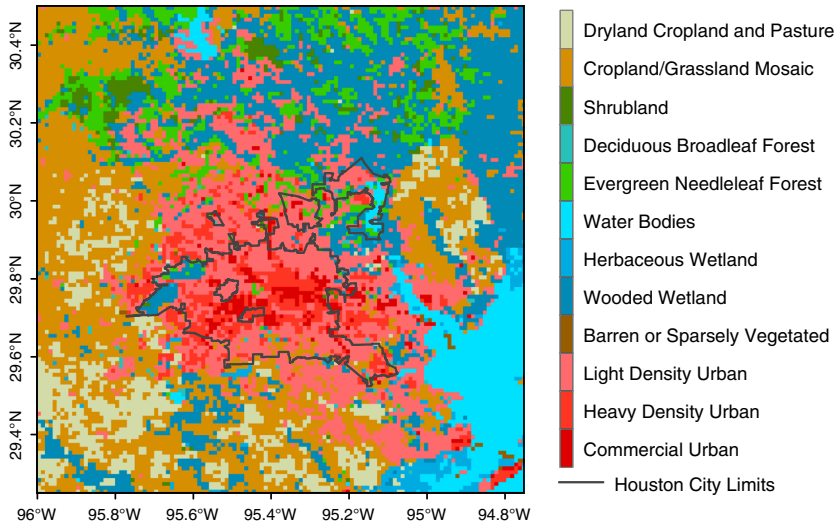

Figure 1. The land use map used in the HRLDAS model simulation.

(HU), and commercial urban (CU, including industrial area), covering approximately 20.6%, 5.0%, and 1.9% of total land area, respectively. Other majority land cover types include cropland and grassland mosaic (27.9%), wooded wetland (23%), and dryland cropland and pasture (8.9%).

2.2. HRLDAS Simulations

Here we briefly describe the methodology for simulating radiative temperature, T_{rad}, with HRLDAS. For additional details, the reader is referred to a full description of the modeling approach in A. J. Monaghan et al. (Evaluating the impact of urban morphology configurations on the accuracy of urban canopy model temperature simulations with MODIS, *Journal of Geophysical Research-Atmospheres*, in revision, 2014). HRLDAS is an offline (i.e., uncoupled to an atmospheric model) version of the Noah Land Surface Model (Noah LSM) [*Chen and Dudhia*, 2001; *Chen et al.*, 2007]. The Noah LSM simulates the fluxes of energy and water from the land surface as a function of the underlying land surface and soil properties while also maintaining stores of water and energy in four soil layers to a depth of 2 m. In order to better represent the physical processes involved in the exchange of heat, momentum, and water vapor in the urban environment, the UCM of *Kusaka et al.* [2001] is employed within the Noah LSM. The UCM provides a more realistic description of the urban lower boundary conditions and morphology and therefore more accurate simulations for urban regions. As mentioned above, the urban land use categories (low-intensity, heavy-intensity, and commercial) for Houston were characterized in the UCM with the 30 m resolution 2001 USGS National Land Cover Database (NLCD) [*Homer et al.*, 2004]. The urban-versus-vegetated (i.e., nonurban) percentage within each grid cell was specified by the National Urban Database and Access Portal Tool (NUDAPT) [*Ching et al.*, 2009; *Burian and Ching*, 2009], a two-dimensional data set of influential urban properties. The use of NUDAPT leads to a more heterogeneous UHI characterization in the HRLDAS simulations.

Half-hourly, 1 km resolution HRLDAS simulations were performed over metropolitan Houston for 2003–2012, preceded by a 1 year (2002) "spin-up" period to allow the soil temperature and moisture states to equilibrate [*Chen et al.*, 2007]. The upper boundary conditions for HRLDAS were derived from one-eighth degree hourly meteorological data from the North American Land Data Assimilation Phase 2 (NLDAS-2) [*Cosgrove et al.*, 2003; *Xia et al.*, 2012] forcing data set. It was assumed that watering and irrigation was applied to all vegetated surfaces, a process represented in HRLDAS by setting in the upper soil layer a lower threshold of 30% water by volume.

In order to ensure that the temperature fields being compared were as similar as possible, we generated a diagnostic radiative temperature (T_{rad}) variable in HRLDAS such that it is similar to the MODIS LST product. At each time step, the energy balances for the roof, wall, and road components for each grid cell containing urban areas were simulated by the UCM subroutine within the Noah LSM. The simulated outward longwave radiation values for the roof, wall, and road were then combined into a bulk urban longwave radiation value for each grid cell, based on the weighted average of the roof and canyon (wall + road) longwave

components. A bulk urban emissivity value was calculated in the same manner for each grid cell. Concurrently, the longwave radiation was calculated in the LSM for the vegetated portion of each grid cell. Then, the bulk urban and vegetated longwave radiation and emissivity values were averaged by weighting the fraction of urban-versus-vegetated area in each grid cell. Finally, these resulting "integrated" longwave radiation and emissivity values for each grid cell were used to calculate T_{rad} using the Stefan-Boltzmann law. The T_{rad} is most representative of nadir for all cases if we do not include wall proportion.

2.3. MODIS Data

The MODIS sensors onboard the Terra and Aqua satellites provide global coverage of LST four times per day. The satellite Sun-synchronous orbits of MODIS are designed to cross the equator at 10:30 and 22:30 for Terra and at 13:30 and 1:30 for Aqua (local solar time). The high frequency of global coverage is at the expense of a large range of view angles spanning from −65° to +65° from nadir. The retrieved spatial resolution for LST is as high as 1 km, which coincides with the scale of the HRLDAS output. The 1 km resolution LST data are produced using the split-window method described by *Wan and Dozier* [1996] and filtered with the MODIS Cloud Mask products (MOD35 L2). The LST MOD11A1 (Terra) and MYD11A1 (Aqua) Version 5 [*Wan*, 2008] products from 2003–2012 summers (June-July-August, JJA) across the Houston area were collected and processed for this study.

3. Sampling Strategies

To ensure that the samples from MODIS and HRLDAS are comparable, we attempt to minimize the biases caused by factors other than the model. Three major aspects must be considered and properly addressed before the MODIS-based model evaluation: satellite overpass time, sensor view angle, and cloud distributions and cloud-contaminated pixels. Among them, the cloud distributions are the primary issue affecting sampling strategies, and we describe the different sampling methods to account for this issue in this section.

3.1. Satellite Overpass Time

The overpass times provided by MODIS LST product are in local solar time, which is defined as the MODIS observation time in coordinated universal time (UTC) plus longitude in degrees divided by 15 [*Williamson et al.*, 2013]. MODIS overpass times are converted from local solar time to local standard time or UTC to enable comparison with model simulations. Each overpass time for the Houston area is slightly different but is usually within 2 h. For example, the daytime overpass data are distributed between 10:00 and 12:10 for Terra and between 12:30 and 14:30 for Aqua over the 10 year summer record. The median local standard times of satellite overpasses for the data sets are about 11:16 (Terra day), 13:34 (Aqua day), 22:34 (Terra night), and 02:16 (Aqua night). We selected the HRLDAS outputs at 11:30, 13:30, 22:30, and 2:30 as the four daily times; therefore, we could conduct the comparison at the nearest half hours to the median MODIS overpass times.

3.2. View Angle

Thermal angular anisotropy is a widely observed phenomenon and has been reported in soil, grass, savanna [*Pinheiro et al.*, 2006], forest canopy [*Smith et al.*, 1997], and urban area by modeling [*Soux et al.*, 2004], surface observations [*Zhan et al.*, 2012], and airborne thermal sensors [*Voogt and Oke*, 1998] with temperature biases of several degrees or even larger. The accuracy of LST related to view angle depends largely on the correction of atmospheric effects, which is a function of wavelength, local meteorological conditions, time of day, season, and location. Although the MODIS LST products are corrected for atmospheric effects related to view angles, the anisotropy bias effect can still be detected due to the generalized parameters used for the atmosphere correction algorithm, as well as the heterogeneous surface structures of various surface components.

Generally, there are two primary solutions to minimize the impact of different view angles between the observations and model output: correct the MODIS LST view angles to nadir or adjust the HRLDAS T_{rad} to the same view angle of MODIS. The use of the split-window atmospheric correction does not adequately remove the effect of view angle from the MODIS data. We can potentially incorporate additional local observations to conduct a more accurate atmosphere correction. However, most correction processes are complex and require knowledge of the atmospheric conditions (e.g., an atmospheric sounding) near the time of satellite overpass. This ancillary information is sometimes unavailable for a specific location and a certain period of time [*Vinnikov et al.*, 2012]. On the other hand, *Lagouarde et al.* [2004, 2012] demonstrated

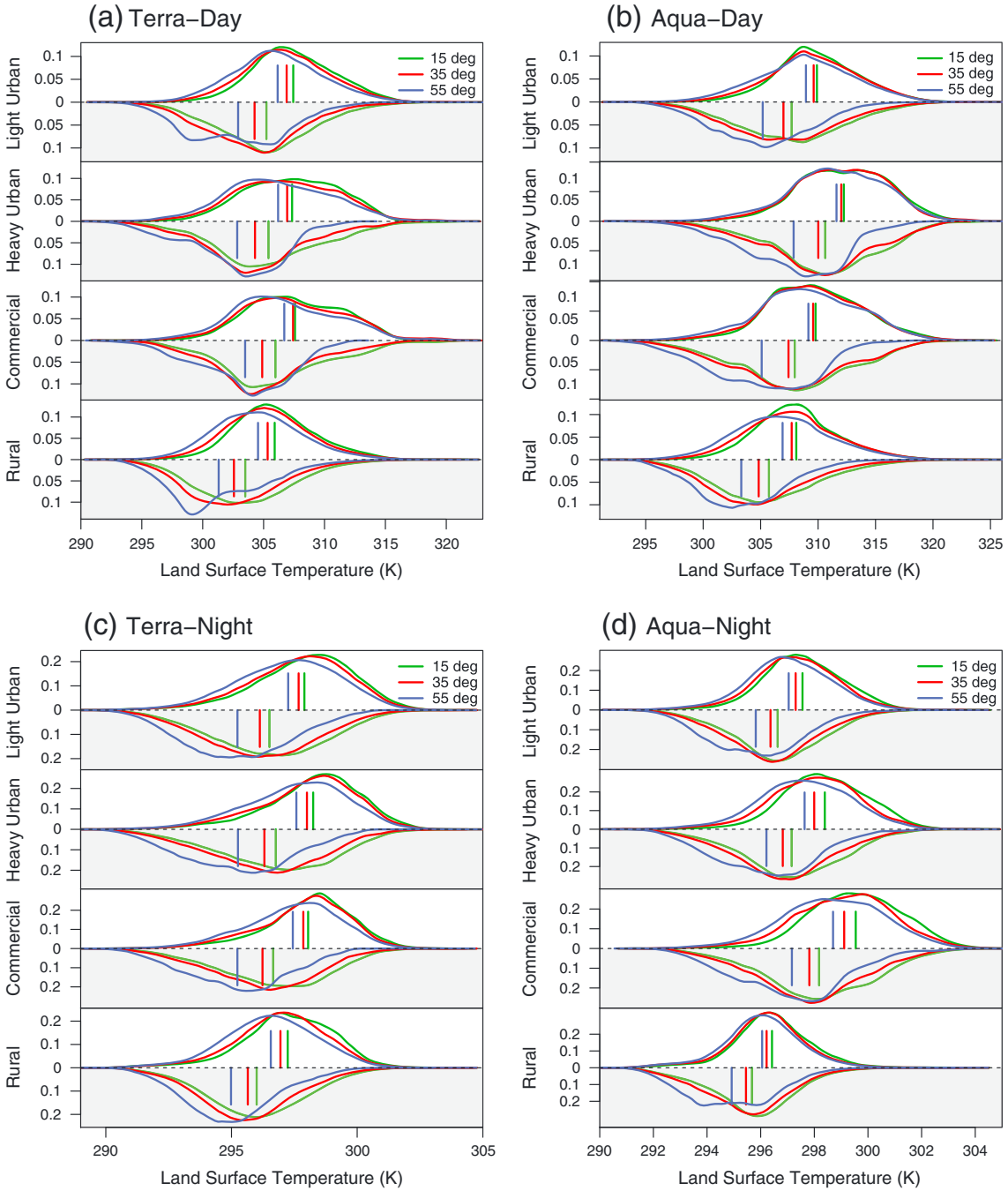

Figure 2. The LST probability density distributions for view angle thresholds of 15° (green), 35° (orange), and 55° (blue), respectively. The upper half of the distribution plots are the pixels with view angles smaller than the given threshold, and the lower shaded half indicates the LST distributions with larger view angles. The short segment line indicates the mean temperature of each distribution.

with airborne measurements that angular impacts on land surface temperature exist over urban areas. Corrections for this effect require knowledge of the angular anisotropy of urban surfaces, which is even more complex than that of atmospheric attenuation. The magnitude of biases induced by view angle effects related to the urban surface structures is unclear at the 1 km scale.

There is also the option to adjust the view angle for simulated T_{rad} based on the morphology of simulated urban canyon structure due to the availability of each urban component temperature (T_{roof}, T_{wall}, and T_{road}) and corresponding emissivity for each grid cell within the model. However, the difficulty of simulating directional vegetation temperatures prohibits the estimation of the biases in mixed urban-rural and rural-only

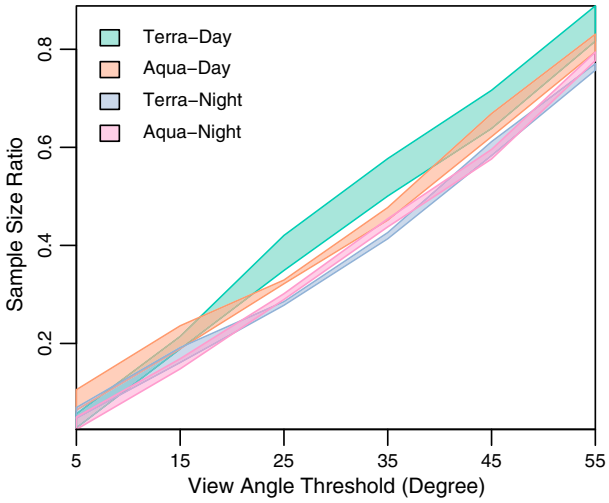

Figure 3. The ratio of sampled pixels with a view angle smaller than the threshold to the total number of available pixels for the four times of day corresponding to MODIS overpass times (with four different colors). The shaded region indicates the range of ratios for the three urban land use types and the rural classification.

pixels, especially for a heavily vegetated city like Houston (there are no pure urban pixels). Therefore, we applied the view-angle correction to the temperature of only the urban fraction in each grid cell and tested the results only for the daytime. The mean adjusted daytime T_{rad} is about 1 K lower than the unadjusted T_{rad} (which is equivalent to the temperature at the nadir); however, the distributions of adjusted T_{rad} at different view angles are similar.

Figure 2 illustrates the impact of view angle on MODIS LST. View angle thresholds of 15°, 35°, and 55° were applied, and LST distributions with larger and smaller view angles than the threshold were analyzed. Generally, the LSTs with smaller view angles tend to be warmer. The distributions differ with land cover type and time of day for LSTs with larger and smaller view angles. For example, the daytime mean LST differences between pixels with smaller and larger view angles are about 2.4 K, 2.8 K, and 3.4 K for the view angle thresholds of 15°, 35°, and 55° over the commercial urban area. The nighttime mean LST differences for the three view angle thresholds are 1.3 K, 1.3 K, and 1.7 K for Terra and about 0.5 K lower for Aqua. Moreover, we tested the view angle effect on water bodies. It showed the existence of anisotropy, indicating a large impact from the atmospheric interaction rather than the surface morphology. Consequently, correcting the view angle of either MODIS LST or adjusting HRLDAS T_{rad} does not seem practical. In order to minimize the biases caused by the view angle, we have chosen to only compare HRLDAS T_{rad} at nadir with MODIS LST that has a small view angle.

However, we should be aware that there are some potential biases comparing the temperature at nadir and off-nadir. As we mentioned before, the atmospheric attenuation is the major view-angle-dependent bias for MODIS LST. The LST at off-nadir angles generally will have a lower bias compared with temperature at nadir due to the longer atmospheric path. From the urban surface morphology perspective, the LST at off-nadir may further decrease during the daytime due to viewing less road proportion and more wall proportion in the urban canyon where usually $T_{road} > T_{wall} > T_{roof}$. There is a higher T_{wall} than T_{road} and T_{roof} at night, so LST bias can be offset at a certain level (please find further details about the urban component temperature in Monaghan et al., manuscript in revision, 2014). By excluding LST with large view angles, the bias of the sampled LST at off-nadir angles can be effectively controlled.

In order to maintain a sufficient sample size, we cannot set the view angle threshold too small. Figure 3 shows the ratio of sampled pixels to the total available pixels (excluding outliers). The ratio of sampled versus total pixels for a given threshold exhibits a nearly linear relationship, indicating a relatively uniform distribution of view angles. At a view angle threshold of 35°, more than 40% of pixels are selected for analysis. The total number of available pixels is dependent on the overpass time due to the variable cloud coverage throughout the diurnal cycle. The Terra daytime data usually have the fewest clear-sky pixels over the urban area during summer. With consideration of sample size and view angle impacts on LST, we decide to use 35° as the view angle threshold in this study.

3.3. Cloud Distributions and Cloud-Contaminated Pixels

Cloud-contaminated pixels, i.e., pixels with cloud influences that are not removed during quality control, are also an issue found in the LST product [*Williamson et al.*, 2013]. In general, the cloud edges and thin clouds are the major sources of contamination [*Platnick et al.*, 2003]. Thick aerosol content from pollutants over the urban area also contributes to pixel contamination, which may partially explain why fewer daytime pixels meet the clear-sky criteria over urban areas compared to adjacent rural areas. It is unclear how advection of pollution plumes affects outlying suburban and rural areas. When the pixels contaminated by clouds or high

Table 1. The Statistics of the Outliers for Four Times of Day Corresponding to MODIS Overpass Times, Including the Mean and Standard Deviation (STD) of Hot and Cold Outliers, and the Ratio of Outliers to the Total Number of Original Data

	Terra Day		Aqua Day		Terra Night		Aqua Night	
	Hot	Cold	Hot	Cold	Hot	Cold	Hot	Cold
Mean (K)	316.2	293.7	318.3	294.5	302.6	289.9	299.8	290.2
STD (K)	3.0	3.6	3.7	5.6	2.6	1.5	0.7	1.8
Ratio	1.16%	0.75%	1.22%	0.92%	0.04%	1.46%	0.12%	3.58%

aerosol content during the daytime remain in the analysis, the MODIS temperature is biased low relative to the "true" temperature. Most cloud-contaminated pixels have been removed in the Version 5 LST product [*Wan*, 2008]. However, some extremely low temperatures were found in the LST product during summer daytime, presumably due to cloud contamination.

Cloud-contaminated pixels (those that remained in the Version 5 LST product) were removed first from MODIS LST by retaining only the data within 1.5 times the interquartile range (Q3–Q1) at each pixel point and time of day. Using a time series to identify the outliers is better, because the detection through spatial analysis may bias the temperature variations caused by land cover type and may also fail to detect all the contaminated pixels due to the existence of many contaminated pixels in one scene. Table 1 shows the statistics related to the outliers. After masking the outliers, the lower end temperature during 10 years increases from 248.3 K and 238.6 K to 290.8 K and 291.5 K for Terra and Aqua daytime images, respectively, and from 276.2 K and 273.8 K to 289.3 K and 289.9 K threshold for Terra and Aqua night, respectively. The maximum temperature deceases by about 5 K for daytime, 11 K for Terra night, and 2 K for Aqua night after filtering the outliers. This process may potentially exclude extreme heat days. However, it is only a small portion (about 1.2% of the total original pixels). Moreover, the extreme days were excluded from MODIS, as well as the corresponding data in HRLDAS, so it would not impact the sampling methods or the comparison for the long term.

Clouds frequently form on summer afternoons in Houston [*Burian and Shepherd*, 2005], coinciding with MODIS daytime overpasses. Figure 4 shows the distribution of the ratio of clear days to the total days (there are 920 total days) during 10 year JJA. The cloud distribution is most heterogeneous during the summer daytime compared with other seasons or with nighttime [*Hu and Brunsell*, 2013]. The deep convection due to the thermal properties of urban surface, urban-rural advection, thick aerosols, etc. all contribute to the large cloudiness differences between urban and rural areas.

After masking the outliers and pixels with larger view angles, the LST data from 2003 to 2012 from four times a day were used as the primary samples from MODIS. Clouds are not explicitly resolved in the HRLDAS forcing data. Instead, clouds are manifested through fluctuations in the surface downwelling longwave and shortwave radiative forcing fields, which makes it difficult to define exactly when cloudy conditions are present in the simulations over a given grid point. We therefore tested the three methodologies described below as a means of minimizing differences between MODIS and HRLDAS due to clouds.

3.4. Sampling Methods for MODIS-HRLDAS Comparisons

In the first sampling method, satellite cloud mask (SCM) method, we assume that the cloud distributions in MODIS and HRLDAS are the same. Therefore, we can apply the masks built from MODIS for the clouds, outliers, and view angles to HRLDAS. As a result, the total sampled pixels for 2003–2012 is 16.8%, 11.6%, 24.2%, and 20.7% of the total pixels in the domain (except the inland water and sea) for the Terra day, Aqua day, Terra night, and Aqua night passes, respectively.

The second method, max/min temperature (MMT) method, is to independently select the warmest days and coolest nights in the model output and MODIS observations. This method relies on the fact that clear-sky LSTs are generally higher (lower) than cloudy or cloud-contaminated pixels during the daytime (nighttime) because of the tendency of clouds to dampen daytime high and nighttime low temperatures [*Dai et al.*, 1999]. We applied the common masks from MODIS (outliers and view angles) to both MODIS and HRLDAS data. It is difficult to determine whether clouds are concurrently impacting both MODIS and HRLDAS fields at every grid point, and therefore we sampled the warmest days and coolest nights separately for each data set, in order to minimize the possibility of clouds impacting either data set. Assuming that neither MODIS

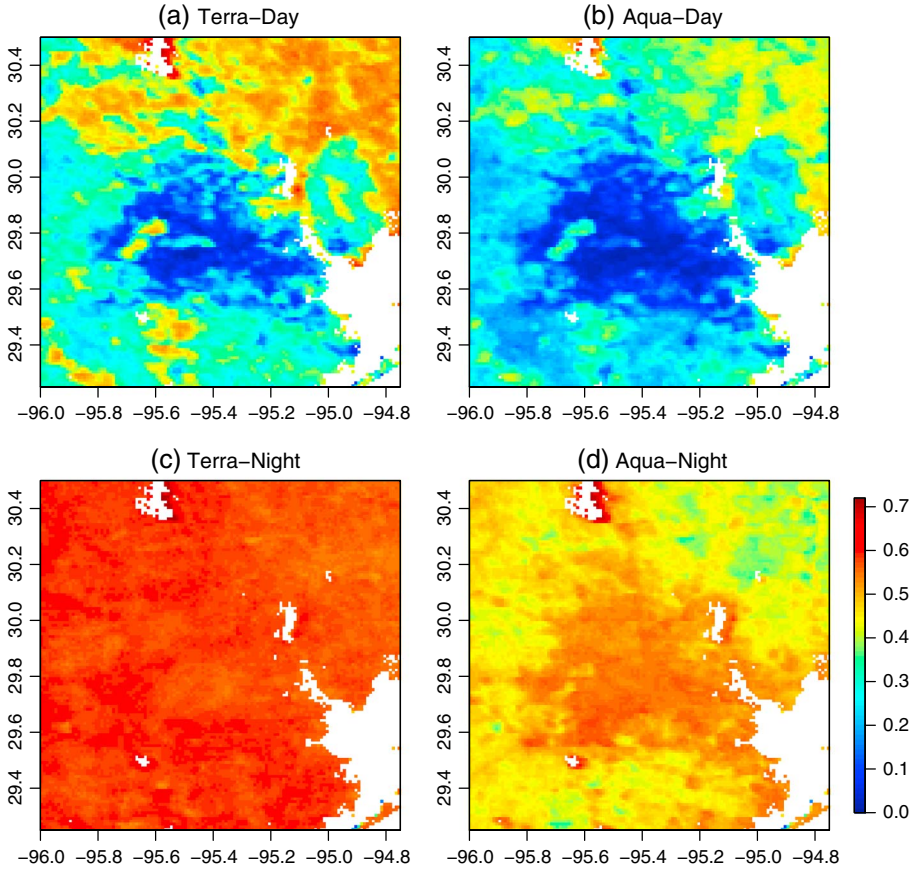

Figure 4. The ratio of selected pixels to total days available for four times of day during 2003–2012 JJA. Inland water and sea pixels are masked, which are the white regions in the figure.

nor the HRLDAS simulations perfectly characterize clouds in space and time, this approach necessitates that the comparisons will not be temporally consistent.

There are many missing data in the MODIS fields, especially during daytime when convective clouds are common (see Figure 4). One of our objectives is to map spatial patterns of the heat island, which ideally requires a consistent number of days available for each grid point. Despite the large amount of missing data, we found that for most grid points, there are at least 50 days with available ("clear-sky") data for a given daytime pass during 2003–2012 summers. We therefore sampled the warmest 50 days and the coolest 50 nights from the HRLDAS and MODIS data for each overpass time, respectively, representing about 5.4% of each set of data. The percentage of available data was slightly lower than 5.4% for daytime due to some grid cells having less than 50 cloud-free days. There are 2993, 733, and 278 grid cells with LU, HU, and CU, respectively, and therefore, despite using only a small fraction of the overall total days for the comparison, there remains an adequate sample size for each land use category to draw statistically meaningful conclusions.

In the third method, high clear-sky fraction (HCF) method, we only choose the daytime and nighttime MODIS overpasses with a high percentage of clear-sky coverage in a scene after excluding the outliers and pixels with larger view angles. The same times are selected from the HRLDAS simulations. Approximately 25% of the total pixels in the domain is urban land cover type in this study. We suggest the clear-sky coverage ratio be set to 90% or more, so that we can guarantee at least more than half of urban pixels are available (clear sky) in each sampled overpass. This method helps eliminate problems caused by cloud mismatches and also maintains the temporal and spatial consistency between observations and simulations. However, based on our selection criteria, there are limited number of days available, due to the prevalence of clouds. About 28 and 8 cloud-free days and 132 and 97 cloud-free nights are sampled from Terra day, Aqua day, Terra night, and Aqua night, respectively, for MODIS during the study period.

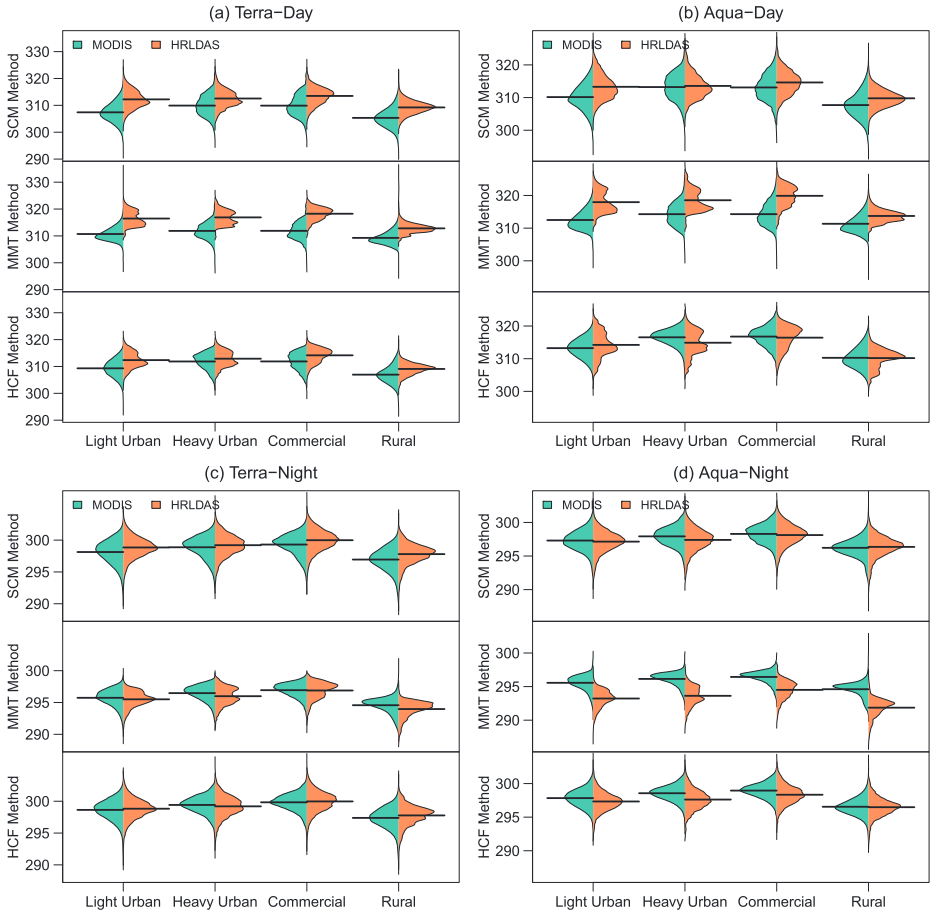

Figure 5. The probability density distribution of MODIS LST (K) and HRLDAS T_{rad} (K) for the different sampling methods. The black horizontal line indicates the mean temperature. The satellite cloud mask (SCM) method uses all pixels after quality control; the max/min temperature (MMT) method uses the 50 warmest days/coolest nights for each grid; and the high clear-sky fraction (HCF) method uses all days/nights with a spatial clear-sky coverage ratio over 90%.

4. Results

4.1. Distribution Analyses

The probability density distributions for the three proposed sampling methods were compared between MODIS observations and HRLDAS simulations at each time of day and each land cover type (Figure 5). If the conditions are mostly clear sky, the MODIS and HRLDAS samples selected by the SCM and HCF methods should be similar spatially and temporally, which is the most ideal case for comparison. For example, the LST distributions of the SCM and HCF methods are analogous during the nighttime in most land cover types. However, the high frequency of clouds during the daytime in Houston, especially over the urban area, dramatically decreases the sample size under the HCF method. The daytime distribution is narrower in the HCF method compared to the SCM method due to fewer sampled days. The distributions are less coincident in HU and CU, which is probably due to the smaller sample size compared with rural and LU. The MMT method is somewhat different in that it scrutinizes the ability of HRLDAS to simulate the two extreme ends of the temperature distribution, and therefore the distributions are less smooth with much smaller LST ranges.

T_{rad} in daytime HRLDAS simulations are overestimated in most cases (Figure 5). An exception is the relatively good match between the HRLDAS and MODIS distributions for rural and CU areas in the HCF method for Aqua daytime overpasses, despite small sample sizes. In general, the nighttime comparisons are in better agreement. For example, the Terra nighttime comparison in the MMT method indicates good agreement with differences less than 0.6 K. The largest differences result from using the MMT method which necessarily compares the coolest nights in both data sets; HRLDAS is about 1.9 K ~ 2.7 K cooler for different land cover types. HRLDAS simulates a fast decrease in temperature from 10:30 pm (Terra night) to 2:30 am (Aqua night),

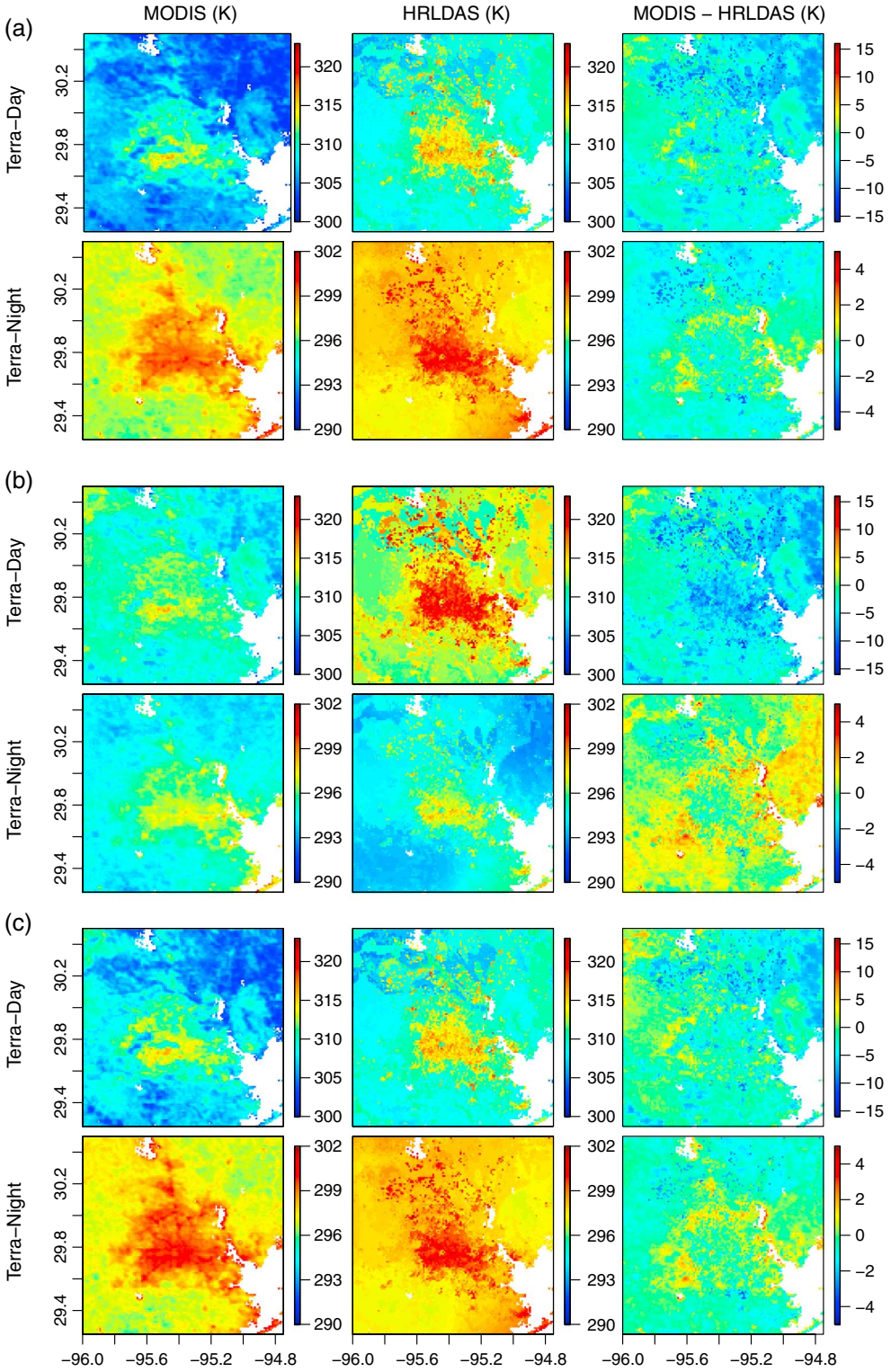

Figure 6. The temporally averaged LST determined from the three sampling methods and the difference map between MODIS and HRLDAS. The sampling methods are demonstrated: (a) satellite cloud mask (SCM) method, (b) max/min temperature (MMT) method, and (c) high clear-sky fraction (HCF) method.

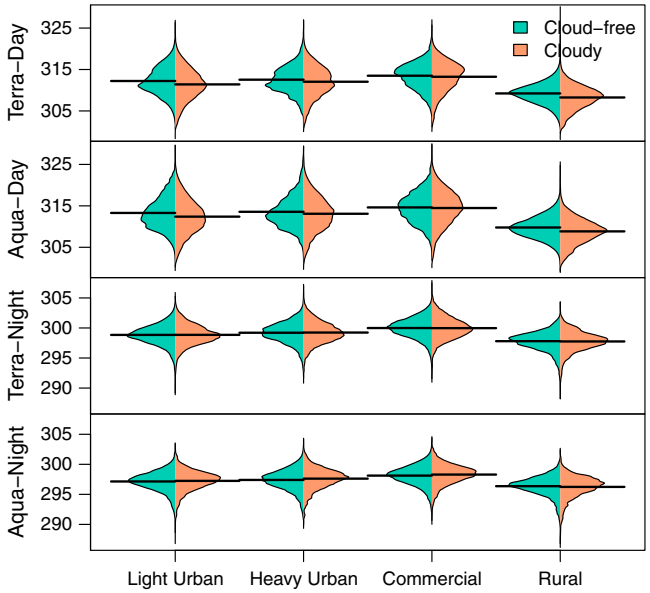

Figure 7. The comparison between the "cloud-free" and "cloudy" pixels of HRLDAS T_{rad} (K) by using the cloud mask from MODIS in the SCM method at four times of day and four land cover types.

which is not observed in the MODIS data, suggesting it underestimates the urban heat island. HRLDAS is an offline model, and therefore it is not coupled with the atmosphere, which can be a source of bias. For example, it is possible that a coupled model would mitigate large temperature gradients between the urban area and atmosphere. Also, the forcing data may have some urban signature but likely not at the scale we are using the model. Overall, the temperature patterns among different land cover types are generally comparable. Rural areas are cooler than urban areas, and the "most rural" urban type, LU, is cooler than other urban types for all times and sampling methods in both data sets.

4.2. Spatial Analyses

All the sampled LST and T_{rad} through time were averaged at each pixel, and the resultant spatial distributions of the averaged LST and T_{rad} maps from the three sampling methods are shown in Figure 6 (Terra only), indicating different spatial patterns due to the different sampling strategies. The spatial patterns among the sampling methods are more similar during nighttime than daytime. The HCF method matches well with the SCM method spatially, with correlations between HRLDAS and MODIS of 0.98 and 0.99 for daytime and nighttime, respectively. The MODIS LST distribution in the MMT method has correlations of 0.90 and 0.88 with SCM and HCF methods within the urban area and slightly higher correlations in rural areas for most of the time during day and night. The nighttime urban boundary is better defined among the methods; however, the UHI is still observed in the mean of warmest 50 pixels during the daytime. The difference maps between MODIS and HRLDAS show a similar pattern among the methods for daytime and nighttime. The heterogeneous distribution of bias is related to the land cover type. For example, the wetland temperature was overestimated at the northeast corner of the domain by about 5 K according to the SCM and MMT methods and 3.9 K for the HCF method, while the urban areas were overestimated by about 3.6 K, 5.6 K, and 2.5 K for the three methods for Terra daytime data. The range of differences during the daytime is similar among the three methods, about 20 K, while the range during the nighttime is smaller, about 8 K. The overall domain-average temperature differences are about 4 K and 3 K for Terra and Aqua daytime for the MMT method, which is about 1 K $\sim$ 2 K larger than the other two methods, suggesting that HRLDAS has comparatively lower skill at simulating extremes. We will discuss potential reasons in the next section.

4.3. Justification of Sampling Method Assumptions

For the SCM method, we assumed that the cloud distribution in HRLDAS is coincident with MODIS observations, so that we could apply the MODIS cloud mask to HRLDAS. Figure 7 shows the probability density distributions for HRLDAS T_{rad} for clear-sky and cloudy pixels defined by the mask in the MODIS LST product in the SCM method. HRLDAS mean T_{rad} is about 0.4 K (about 1 K for LU and rural areas) warmer in the cloud-free pixels over the cloudy pixels during daytime, which is expected because clear days are warmer

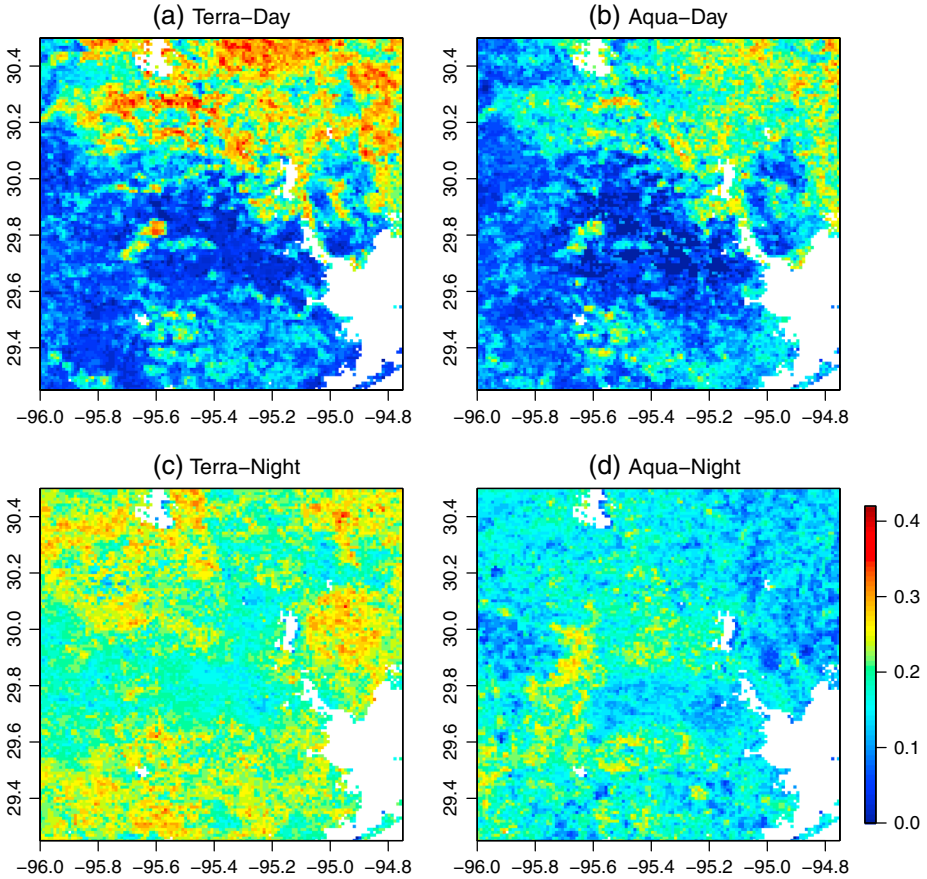

Figure 8. The ratio of the temporally consistent days to the total selected 50 days for four times a day.

than cloudy days. The nighttime differences are smaller, probably due to the lack of interaction of strong external energy sources (e.g., solar radiation) and also the accuracy of cloud maps. Generally speaking, cloud detection is more difficult during the nighttime [*Platnick et al.*, 2003] due to less information in the optical regions of the spectrum. In summary, the assumption we made for SCM method—that cloud distributions are similar between HRLDAS and MODIS—appears to be true for daytime and inconclusive for nighttime.

The assumption we made for the MMT method was that the warmest (daytime)/coolest (nighttime) pixels from MODIS and HRLDAS should be temporally consistent at a certain level. In other words, if the cloud masks of two data sources are comparable and the model is able to simulate the temperature variability similar to the MODIS LST observations, the distributions of sampled pixels that are cloud free should share many of the same hot days and cool nights in common. The ratio of the number of temporally consistent days to the total sampled days for each grid is shown in Figure 8. The temporal agreement in the urban area is very low (ratios of about 0.1 (Terra) and 0.09 (Aqua)) compared with northeast-corner rural areas (wetland and forest land cover types with ratios of about 0.23 (Terra) and 0.18 (Aqua)) during the daytime. The spatial distribution of the ratios is less variable during the night with an average of 0.20 and 0.17 in urban areas for Terra and Aqua, respectively. The spatial patterns in Figure 8 are closely related to the spatial distribution of the available-to-total pixels after excluding the outliers and the pixels with large view angles (not shown here, it is similar to Figure 4), especially during the daytime with correlation coefficients of 0.8 (Terra) and 0.7 (Aqua). The correlation drops to 0.2 for Terra night when the cloud-free ratio is high and with less spatial variability. Consequently, the cloud distributions and frequency may directly impact this method, i.e., it is so cloudy during most days that insufficient samples are available over the urban areas. For example, there are less than 20 clear-sky days in some urban pixels during the 2003–2012 JJA periods. On the other hand, the poor temporal agreement may also be attributed to model defects, such as the mismatch of clouds due to the forcing data, or simplified assumptions (we evaluate model performance in the companion paper by A. J. Monaghan et al., manuscript in revision, 2014). In summary, the MMT method may be better

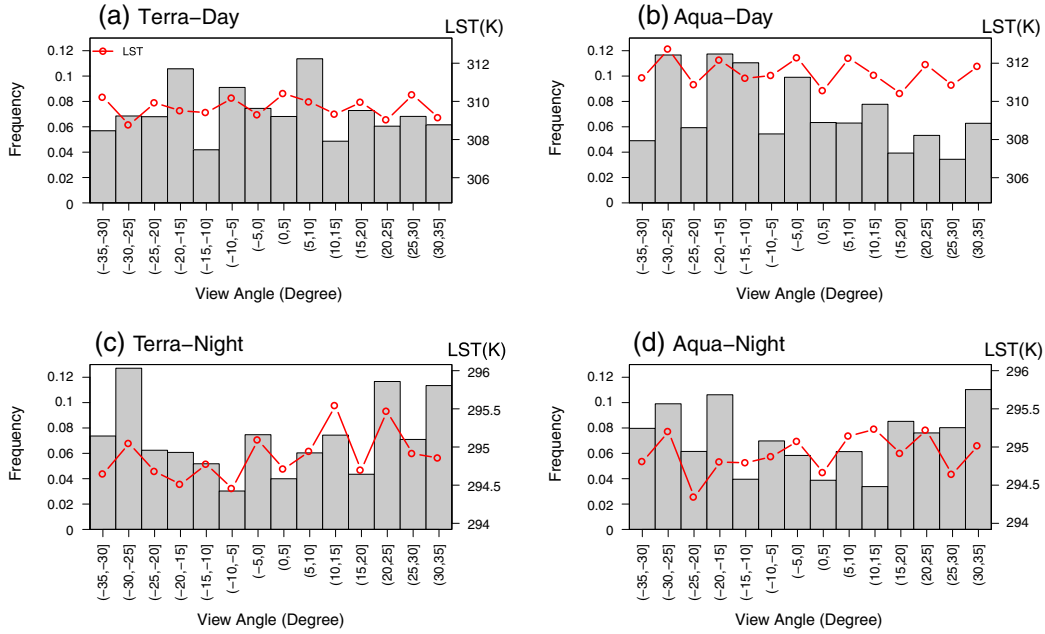

Figure 9. The view angle histogram of sampled MODIS LST by the MMT method for four times a day (left *y* axis, gray bar) with the mean LST to the corresponding view angle ranges (right *y* axis, red line).

suited for regions where daytime cloudiness is not as persistent so that a larger distribution of pixels can be drawn from.

The MMT method is designed to sample the two ends of the temperature distribution from MODIS and the model. For the observations, potentially it enhances the impact of view angle-induced temperature biases. In order to minimize this effect, prior to employing any of the sampling methods, we masked all of the pixels with view angles larger than 35°. Figure 9 shows the view angle distributions and the corresponding mean MODIS LST of sampled pixels by the MMT method. The nighttime mean LST indicates that warmer pixels slightly coincide with larger view angles, which may be attributed to the warmer radiative temperature of the walls over the roof and roads in the urban area [*Kusaka et al.*, 2001; *Voogt and Oke*, 2003; *Lagouarde et al.*, 2012]. However, the temperature trend here is not statistically significant. The nighttime temperature ranges from 294.5 K to 295.5 K, which is less than the associated daytime temperature variation (about 2 K). Further, the similar statistics for each urban land cover type and rural area (not shown) show that LU and rural types share similar LST trends related to view angles with Figure 9, while HU and CU have larger LST variations due to insufficient pixels for some view angles. In summary, the strategy of excluding the larger view angles efficiently decreases the view angle impact on LST.

5. Discussion

Before sampling data for model validation, we must be clear of several important facts of using satellite remotely sensed data. First, the overpass time of the satellite is not constant. There is approximately a 2 h range for a given MODIS overpass, so we choose a median time for which to make the comparison of MODIS LST to HRLDAS-simulated T_{rad}. Consequently, biases may be introduced by temporal gaps. The impact will be larger for the daytime than nighttime because the components of the surface energy balance are generally larger and more variable during daytime. Also, the periodicity of Sun-synchronous satellites is not frequent enough for determining diurnal variation. As a result, the diurnal temperature range calculated from the MODIS day/night difference is unlikely to match the traditional definition, defined as the difference between daily maximum and minimum temperatures. Less "diurnal" information is available due to cloud distributions varying at the pair of overpass times. Second, the thermal angular anisotropy should be considered because of atmospheric correction, surface geometry, solar zenith and azimuth angles, surface materials, etc. Correcting MODIS LST or simulated T_{rad} view angle over large areas is not practical at this stage. Consequently, we only selected the MODIS data within view angles ± 35° from nadir, which we demonstrated was an effective approach for view-angle-dependent biases in MODIS LST. These facts will be

Table 2. The Advantages and Disadvantages of the Three Proposed Sampling Methods

Strategies	Advantages	Disadvantages
SCM method	The largest possible sample size	Unequal sample size in each grid box
	Temporally consistent	Possible cloud contamination
	Diurnal temperature comparison	Possible cloud mismatch
	Long-term comparison	
MMT method	All samples are cloud free	Temporal inconsistency
	Equal sample size from each grid point	Only apply to long-term data
		Unsuitable for year-to-year variability or time series analysis
	Sampling extremes	Unable to do diurnal temperature comparison
HCF method	Temporally consistent	Small sample size
	Spatially consistent	Unsuitable for year-to-year variability
	Likely equal sample size from each grid point	or time series analysis
	Suitable for case studies	Hard to do diurnal temperature comparison

also valid for MODIS LST product Version 6, because the difference between Version 5 (used in this study) and Version 6 products is very small in all regions but hot and warm bare soil zones [*Wan*, 2014]. For other narrow-swath satellite remote sensors, view angle may be less influential. Additionally, we had the advantage of having 10 years of data and therefore retained a large sample of MODIS data for comparison, even after omitting the passes with large view angles. Lastly, clouds can strongly impact the accuracy and applicability of thermal remote-sensing data. The sampling methods we proposed are aimed at minimizing cloudy pixels in both the MODIS LST data and simulations.

By selecting only cloud-free pixels for comparison, there is the likelihood of biasing the comparison toward warmer daytime conditions [*Hu and Brunsell*, 2013]. However, the validation of a clear-sky case could be the most robust test of how the model surface energy balance performs since it is expected to maximize the differences with variations in surface conditions in the clear-sky cases. Due to the limitation of remote-sensing systems, the land surface temperature cannot be directly measured under the partial-cloudy or cloudy conditions. *Bisht and Bras* [2010] developed the remote-sensing-based estimation of net radiation under all sky conditions. However, the surface temperature and other parameters under cloudy conditions are the result of global model analyses or regressions based on the in situ measurements. The model validation for long-term simulations is still challenging under cloudy conditions.

Three sampling methods for validating model simulations versus MODIS LST are proposed and compared. Table 2 summarizes the major advantages and disadvantages of each sampling strategy. The length of model simulation and the climate of the study area will affect selection of the optimal method. Generally, if the study area is a cloudy region, such as Houston, we recommend using the SCM method to maximize the sample size of observed data. The HCF method is also an acceptable choice but may lack representativeness due to small sample size. If the study area is predominantly sunny, such as Phoenix, the difference between the SCM and HCF methods should be modest. The MMT method may be complimentary to the SCM and HCF methods, in that it can be used to evaluate model performance for extreme cases, and it provides an implicit test of the agreement of the simulated-versus-observed cloud distributions. The HCF method is a better choice when simulations are only available for a short period. In summary, each sampling strategy has its own best applicable situations. If the length of simulation or cloud conditions are not limiting factors in the study area, it is also a plausible approach to employ all three approaches for model validation, because each has different strengths.

The thresholds for the processing as well as for the MMT and HCF methods can be adjusted for the specific application. The choice of view angle threshold requires balance between the sample size and LST accuracy. If modelers are interesting in extreme events, for the outlier exclusion, it is also reasonable to keep the tail of temperature distribution on the side that is unaffected by clouds or keep a larger portion of the tails.

For the MMT method, the number of warmest days/coolest nights depends on the minimum number of nearly cloud-free days/nights in the study area, so that we can have relatively large coverage for spatial comparisons. Besides the MMT method, there is another option to select certain high and low quantiles for days and nights in each pixel. This strategy guarantees spatial coverage but sacrifices the equality of

sample size in each grid point. We tested 50% of the warmest days and coolest nights (not shown), and the LST distributions for each land cover types are similar to the case for the MMT method presented above.

We recommend setting the spatial coverage ratio threshold for the HCF method to be larger than the rural coverage ratio in the domain in this study. For example, we set 90% for the Houston area which is with 25% of urban pixels in the domain, due to the cloud patterns over the metropolitan areas. As shown in Figure 4, clouds occur more frequently over the urban area. The decrease in the spatial coverage threshold may result in enlarging the rural sample size only, which does not benefit the urban temperature comparison.

The SCM method masks all pixels based only on whether clouds exist in the MODIS data. However, it does not mask pixels based on whether clouds exist in the HRLDAS data, because cloud fields are not present in the HRLDAS forcing fields (only fluctuations of longwave and shortwave radiation that may indicate the presence of clouds, though not accurately). Therefore, it is likely that some proportion of HRLDAS pixels which are designated as "clear sky" by the SCM method may actually be influenced by cloudy or partly cloudy conditions. The effect of this influence is that the HRLDAS simulations that represent clear sky in the SCM method may be cooler (warmer) during daytime (nighttime) than they otherwise would if they represented clear-sky conditions perfectly. It is therefore recommended that if a user does have access to a cloud mask for their simulations, they apply it, in addition to the MODIS cloud mask, to both the simulations and the MODIS data to ensure that all pixels represent clear-sky conditions.

The land cover map used for analyses is based on the 30 m resolution 2001 USGS NLCD. Compared with the MODIS annual land cover product (MCD12Q1) from 2003 to 2011, which are resampled to 1 km resolution, the land cover types and their distributions have some variability. For instance, the total urban pixels in MODIS are about 300 less than the urban pixels defined from NLCD. The major disagreement is with the LU class outside the Houston core urban area, perhaps due in part to having to resample both data sets from their native resolution to the 1 km HRLDAS domain. For rural land cover types, the biggest gap is the wetland in NLCD (25% of total land area in the domain), which is classified as grassland in MODIS (9 year average about 8.2%) in the northeast corner. Similarly, cropland and natural vegetation mosaics from NLCD are defined as savanna in MODIS. For the land surface temperature, the major impact of this discrepancy in classification is with the emissivity defined according to the land cover type for MODIS, while the related parameters specified for emissivity in the model also differ by land cover type, affecting the radiative temperature. The land cover dynamics and their disagreements are not directly related to the sampling methods, but these differences may be an important contribution to the error for model validation.

Only summer season was discussed in this paper, due to the negative impact of extreme summer heat on human health. *Hu and Brunsell* [2013] discussed the seasonal cloud distributions during 2000 to 2010 in Greater Houston. The cloud impact is largest during the summer with few cloud-free days and the most heterogeneous clear-sky ratio distribution. Consequently, if we apply these three methods to evaluate the other three seasons, the samples will be more representative for the comparison due to less cloud impact.

We must emphasize that the validation only shows the model performance for radiative temperature. For urban climate modeling applications aimed at characterizing urban extreme heat events and their impacts on human health, air temperature observations must also be considered. Radiative temperature and 2 m air temperature are highly correlated [*Prihodko and Goward*, 1997] but determined by different surface processes. Radiative temperature is determined by the radiation budget, which is the part of the energy balance directly affecting the atmospheric temperature from below. *Jin et al.* [1997] compared the radiative temperature and air temperature at the global scale, finding that the differences are scale dependent and they are more agreeable at larger scales (i.e., monthly scale and hemispheric spatial scales). *Voogt and Oke* [2003] summarized the relationship between remotely observed radiometric surface temperature and the actual air temperature of the urban-atmosphere interface. The HRLDAS-simulated 2 m air temperatures are compared to air temperature observations throughout Houston in our companion paper (A. J. Monaghan et al., manuscript in revision, 2014), and the results are similar to those presented here for LST, although the model biases of air temperature are smaller. Therefore, LST-based comparisons may exaggerate the magnitude of model temperature biases because they are based on skin temperatures as opposed to air temperatures, which typically exhibit smaller diurnal cycles and are less sensitive to changes in the surface energy balance.

6. Conclusions

Thermal remotely sensed data are good sources of observations for model validation, with comparable spatial scale, temporal consistency, large coverage, and long-term archives. This study has highlighted several important aspects of satellite LST data that can easily be neglected by modelers for model validation, such as cloud distribution and sensor view angle. From a 10 year case study over Greater Houston during 2003–2012, three sampling strategies were applied to compare MODIS LST with HRLDAS-simulated T_{rad}, temporally and spatially. The assumptions made by each of the strategies were discussed. We assessed the strengths and weakness of the three sampling methods and suggested their best applications. The SCM method gives the best statistical climatological comparison conditioned on the assumed agreement between MODIS clouds and forcing clouds (the longer the time series, the more appropriate this may be). The MMT method gives the best comparison between model simulation and MODIS LST at the extremes in a climatological application. The HCF method gives the best absolute temperature comparison because it is the most consistent, both spatially and between model and observations.

Further research for the view angle correction related to satellite remotely sensed data or model simulations appears warranted, which may further increase the consistency of data properties for validation. Also, the magnitude of impact from temporal gaps between observations and simulations could be tested by modeling exactly at satellite overpass time for each image (beyond the scope of this study). Additional evaluations of the three methods could be performed for other seasons and less cloudy regions. Also, there remains large potential for using other remotely sensed variables to conduct various model validations, such as soil moisture, radiation, precipitation, vegetation indices, evapotranspiration, etc. at multiple scales. Researchers would need to be aware of the accuracy and limitations of the remotely sensed products before proceeding when comparing different variables.

Acknowledgments
This work was supported under the National Aeronautics and Space Administration program (NNX10AK79G). NCAR is funded in part by the National Science Foundation.

References

Bisht, G., and R. L. Bras (2010), Estimation of net radiation from the MODIS data under all sky conditions: Southern great plains case study, *Remote Sens. Environ.*, *114*(7), 1522–1534.

Bodas-Salcedo, A., M. A. Ringer, and A. Jones (2008), Evaluation of the surface radiation budget in the atmospheric component of the Hadley Centre Global Environmental Model (HADGEM1), *J. Clim.*, *21*(18), 4723–4748.

Burian, S. J., and J. K. Ching (2009), Development of gridded fields of urban canopy parameters for advanced urban meteorological and air quality models, *EPA/600/R-10/007*, U.S. Environmental Protection Agency, Washington, D. C.

Burian, S. J., and J. M. Shepherd (2005), Effect of urbanization on the diurnal rainfall pattern in Houston, *Hydrol. Processes*, *19*(5), 1089–1103.

Centre for Health Development World Health Organization (2010), Hidden cities: Unmasking and overcoming health inequities in urban settings, World Health Organization.

Chen, F., and J. Dudhia (2001), Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 Modeling System. Part I: Model implementation and sensitivity, *Mon. Weather Rev.*, *129*(4), 569–585.

Chen, F., et al. (2007), Description and evaluation of the characteristics of the NCAR High-Resolution Land Data Assimilation System, *J. Appl. Meteorol. Climatol.*, *46*(6), 694–713.

Chen, F., et al. (2011), The integrated WRF/Urban Modelling System: Development, evaluation, and applications to urban environmental problems, *Int. J. Climatol.*, *31*(2), 273–288.

Ching, J., et al. (2009), National urban database and access portal tool, *Bull. Am. Meteorol. Soc.*, *90*(8), 1157–1168.

Cosgrove, B. A., et al. (2003), Real-time and retrospective forcing in the North American land data assimilation system (NLDAS) project, *J. Geophys. Res.*, *108*(D22), 8842, doi:10.1029/2002JD003118.

Dai, A., K. E. Trenberth, and T. R. Karl (1999), Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range, *J. Clim.*, *12*(8), 2451–2473.

Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold, and J. Wickham (2011), Completion of the 2006 National Land Cover Database for the conterminous United States, *Photogramm. Eng. Remote Sens.*, *77*(9), 858–864.

Ghent, D., J. Kaduk, J. Remedios, J. Ardo, and H. Balzter (2010), Assimilation of land surface temperature into the land surface model JULES with an ensemble Kalman filter, *J. Geophys. Res.*, *115*, D19112, doi:10.1029/2010JD014392.

Grimmond, C. S. B. (2006), Progress in measuring and observing the urban atmosphere, *Theor. Appl. Climatol.*, *84*(1-3), 3–22.

Grimmond, C. S. B., and T. R. Oke (2002), Turbulent heat fluxes in urban areas: Observations and a Local-Scale Urban Meteorological Parameterization Scheme (lumps), *J. Appl. Meteorol.*, *41*(7), 792–810.

Grimmond, C. S. B., et al. (2010), The international urban energy balance models comparison project: First results from phase 1, *J. Appl. Meteorol. Climatol.*, *49*(6), 1268–1292.

Homer, C., C. Q. Huang, L. M. Yang, B. Wylie, and M. Coan (2004), Development of a 2001 national land-cover database for the United States, *Photogramm. Eng. Remote Sens.*, *70*(7), 829–840.

Hu, L., and N. A. Brunsell (2013), The impact of temporal aggregation of land surface temperature data for surface urban heat island (SUHI) monitoring, *Remote Sens. Environ.*, *134*, 162–174.

Jin, M. L., R. E. Dickinson, and A. M. Vogelmann (1997), A comparison of CCM2-BATS skin temperature and surface-air temperature with satellite and surface observations, *J. Clim.*, *10*(7), 1505–1524.

Kusaka, H., and F. Kimura (2004), Coupling a single-layer urban canopy model with a simple atmospheric model: Impact on urban heat island simulation for an idealized case, *J. Meteorol. Soc. Jpn.*, *82*(1), 67–80.

Kusaka, H., H. Kondo, Y. Kikegawa, and F. Kimura (2001), A simple single-layer urban canopy model for atmospheric models: Comparison with multi-layer and slab models, *Boundary Layer Meteorol.*, *101*(3), 329–358.

Lagouarde, J. P., P. Moreau, M. Irvine, J. M. Bonnefond, J. A. Voogt, and F. Solliec (2004), Airborne experimental measurements of the angular variations in surface temperature over urban areas: Case study of Marseille (France), *Remote Sens. Environ.*, *93*(4), 443–462.

Lagouarde, J. P., A. Henon, M. Irvine, J. Voogt, G. Pigeon, P. Moreau, V. Masson, and P. Mestayer (2012), Experimental characterization and modelling of the nighttime directional anisotropy of thermal infrared measurements over an urban area: Case study of Toulouse (France), *Remote Sens. Environ.*, *117*, 19–33.

Leroyer, S., S. Belair, and J. Mailhot (2011), Microscale numerical prediction over Montreal with the Canadian external urban modeling system, *J. Appl. Meteorol. Climatol.*, *50*(12), 2410–2428.

Masson, V. (2006), Urban surface modeling and the meso-scale impact of cities, *Theor. Appl. Climatol.*, *84*(1-3), 35–45.

Mestayer, P. G., et al. (2005), The urban boundary-layer field campaign in Marseille (UBL/CLU-Escompte): Set-up and first results, *Boundary Layer Meteorol.*, *114*(2), 315–365.

Miao, S. G., F. Chen, M. A. Lemone, M. Tewari, Q. C. Li, and Y. C. Wang (2009), An observational and modeling study of characteristics of urban heat island and boundary layer structures in Beijing, *J. Appl. Meteorol. Climatol.*, *48*(3), 484–501.

National Research Council (2012), *Urban Meteorology: Forecasting, Monitoring, and Meeting Users' Needs*, The National Academies Press, Washington, D. C.

Oleson, K. W., G. B. Bonan, J. Feddema, M. Vertenstein, and C. S. B. Grimmond (2008), An urban parameterization for a global climate model. Part I: Formulation and evaluation for two cities, *J. Appl. Meteorol. Climatol.*, *47*(4), 1038–1060.

Pinheiro, A. C. T., J. L. Privette, and P. Guillevic (2006), Modeling the observed angular anisotropy of land surface temperature in a savanna, *IEEE Trans. Geosci. Remote Sens.*, *44*(4), 1036–1047.

Platnick, S., M. D. King, S. A. Ackerman, W. P. Menzel, B. A. Baum, J. C. Riedi, and R. A. Frey (2003), The MODIS cloud products: Algorithms and examples from Terra, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 459–473.

Prihodko, L., and S. N. Goward (1997), Estimation of air temperature from remotely sensed surface observations, *Remote Sens. Environ.*, *60*(3), 335–346.

Rasmussen, M. O., A. C. Pinheiro, S. R. Proud, and I. Sandholt (2010), Modeling angular dependences in land surface temperatures from the Seviri instrument onboard the geostationary Meteosat second generation satellites, *IEEE Trans. Geosci. Remote Sens.*, *48*(8), 3123–3133.

Skamarock, W. C., and J. B. Klemp (2008), A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, *J. Comput. Phys.*, *227*(7), 3465–3485.

Smith, J. A., J. R. Ballard, and J. A. Pedelty (1997), Effect of three-dimensional canopy architecture on thermal infrared exitance, *Opt. Eng.*, *36*(11), 3093–3100.

Sohrabinia, M., W. Rack, and P. Zawar-Reza (2012), Analysis of MODIS LST compared with WRF model and in situ data over the Waimakariri River basin, Canterbury, New Zealand, *Remote Sens.*, *4*(11), 3501–3527.

Soux, A., J. A. Voogt, and T. R. Oke (2004), A model to calculate what a remote sensor 'sees' of an urban surface, *Boundary Layer Meteorol.*, *112*(2), 401–424.

Vinnikov, K. Y., Y. Yu, M. D. Goldberg, D. Tarpley, P. Romanov, I. Laszlo, and M. Chen (2012), Angular anisotropy of satellite observations of land surface temperature, *Geophys. Res. Lett.*, *39*, L23802, doi:10.1029/2012GL054059.

Voogt, J. A., and T. R. Oke (1998), Effects of urban surface geometry on remotely-sensed surface temperature, *Int. J. Remote Sens.*, *19*(5), 895–920.

Voogt, J. A., and T. R. Oke (2003), Thermal remote sensing of urban climates, *Remote Sens. Environ.*, *86*(3), 370–384.

Wan, Z. (2008), New refinements and validation of the MODIS land-surface temperature/emissivity products, *Remote Sens. Environ.*, *112*(1), 59–74.

Wan, Z. (2014), New refinements and validation of the Collection-6 MODIS land-surface temperature/emissivity product, *Remote Sens. Environ.*, *140*, 36–45.

Wan, Z., and J. Dozier (1996), A generalized split-window algorithm for retrieving land-surface temperature from space, *IEEE Trans. Geosci. Remote Sens.*, *34*(4), 892–905.

Williamson, S. N., D. S. Hik, J. A. Gamon, J. L. Kavanaugh, and S. Koh (2013), Evaluating cloud contamination in clear-sky MODIS Terra daytime land surface temperatures using ground-based meteorology station observations, *J. Clim.*, *26*(5), 1551–1560.

Xia, Y. L., et al. (2012), Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, *117*, D03109, doi:10.1029/2011JD016048.

Zhan, W. F., Y. H. Chen, J. A. Voogt, J. Zhou, J. F. Wang, W. Ma, and W. Y. Liu (2012), Assessment of thermal anisotropy on remote estimation of urban thermal inertia, *Remote Sens. Environ.*, *123*, 12–24.