UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN INTERCOMPARISON OF LANDSAT LAND SURFACE TEMPERATURE RETRIEVAL METHODS UNDER VARIABLE ATMOSPHERIC CONDITIONS USING IN SITU SKIN TEMPERATURE

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF ARTS

By

EMILY CLAIRE WINDAHL
Norman, Oklahoma
2016

AN INTERCOMPARISON OF LANDSAT LAND SURFACE TEMPERATURE RETRIEVAL METHODS UNDER VARIABLE ATMOSPHERIC CONDITIONS USING IN SITU SKIN TEMPERATURE

A THESIS APPROVED FOR THE
DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL SUSTAINABILITY

BY

_____
Dr. Kirsten de Beurs, Chair

_____
Dr. Renee McPherson

_____
Dr. Petra Klein

Acknowledgements

Research is supported by NASA IDS project NNX12AM89G: Storms, Forms, and Complexity of the Urban Canopy: How Land Use, Settlement Patterns, and the Shape of Cities Influence Severe Weather. I would like to thank Kirsten de Beurs for her excellent advice and guidance. I would also like to thank Braden Owsley for his assistance with data processing; the Oklahoma Mesonet and UCAR's Suominet project for providing data used in this study; Petra Klein and Renee McPherson for their feedback on an early version of this research; the two anonymous reviewers whose recommendations improved this paper.

Table of Contents

List of Tables

List of Figures

Abstract

Land surface temperature (LST) retrieved from Landsat is a valuable resource for understanding land cover change, monitoring the urban heat island effect, and modeling hydrological and carbon cycles, among other applications. However, this dataset is underutilized, in part because it is difficult to accurately correct for atmospheric interference, and in part because it is difficult to validate the resulting LST dataset. As a result, it is often challenging to verify the accuracy of LST calculated from historical data. Currently, three correction methods are commonly used to retrieve LST from single-band Landsat TIR data—the radiative transfer equation (RTE), the mono-window algorithm (MWA), and the generalized single-channel (GSC) method. Based on current research, it is often unclear which method is best applied in different circumstances and what the actual achieved accuracy is—especially when these methods are employed as they would be for actual applications, rather than under validation conditions. This study retrieves LST from two years' worth of clear-sky Landsat 5 TM data using all three methods and validates the results against on-the-ground skin temperature measurements from twenty-five Oklahoma Mesonet stations. Additionally, LST results using both modeled transmittance and transmittance based on precipitable water vapor are assessed, as are results from dates with both high and low precipitable water vapor. Results suggest that the MWA method using modeled transmittance is the most robust, with results statistically indistinguishable from Mesonet skin temperature for the complete dataset and a cloud-free subset, as well as for subsets above and below 2 g/cm^2 precipitable water vapor. The RTE method using modeled atmospheric parameters is also appropriate in some circumstances.

Chapter 1: Introduction

Spaceborne thermal remote sensing relies on the truth that all objects (barring a temperature of absolute zero) emit radiation. Further, using Planck's law, we can relate the amount of radiation emitted at a given wavelength (assuming a black body in thermal equilibrium) to the object's temperature. In theory, this allows us to determine the temperature of the earth's surface based on the amount of electromagnetic energy it emits at a given wavelength, as measured by a sensor in space (Wark et al., 1962). In practice, this process is more complicated; the radiation received by the sensor must be corrected for atmospheric interactions and land surface emissivity (LSE) before true land surface temperature can be determined (Dash et al., 2002; Li et al., 2013; Prata et al., 1995). As a result, practical, accurate atmospheric and emissivity corrections that can be implemented by researchers or incorporated into ready-to-use land surface temperature (LST) products are an essential prerequisite for LST applications.

The most commonly used thermal sensors for land surface applications are: MODIS on satellites Terra and Aqua (1 km resolution, 1–2-day return) (Wan, 2014); ASTER on Terra (90 m resolution, 16-day return) (Palluconi, 1996); TM, ETM+, and TIR on Landsats 4–7 and 8 (60–120 m resolution, 16-day return) (Irons et al., 2012); the SLSTR on the Sentinel-3 satellite(s) (1 km resolution, 1–2-day return) (Donlon et al., 2012); AVHRR on NOAA POES satellites (1.09 km resolution, 24-hour return); (Ouaidrari et al., 2002); and SEVIRI on the Meteosat Second Generation satellites (3 km resolution, 15-minute return) (Jimenez-Munoz et al., 2014a). There are two competing considerations for thermal remote sensing data—spatial resolution and return period (Dash et al., 2002). Sensors with high return frequency, like MODIS, have

coarser resolution, and sensors with relatively fine spatial resolution have a lower return frequency (e.g. TM on Landsat). High-spatial-resolution data and high-temporal-resolution data both provide complementary insights for research applications.

LST retrieved from Landsat, in addition to its relatively fine-scale spatial resolution, is a uniquely valuable dataset due to its regular global coverage and the continuous archive of data dating back to 1983, offered free-of-charge for users since 2008 (Wulder et al., 2012). Only ASTER offers similarly scaled thermal data, but ASTER is not freely available, and does not provide routine, reliable global coverage offered by Landsat. Studies of urban heat, especially, benefit from the longitudinal depth and spatial detail of Landsat thermal data. Intra-urban temperature differences can be as large as or larger than urban-rural differences (Buyantuyev and Wu, 2010), but these intra-urban differences generally can't be detected at the 1-km scale of, for example, MODIS (Sobrino et al., 2012). Additionally, potential solutions are also relatively small-scale—green or high-albedo roofs, for example, or increased urban tree cover (Gober et al., 2010; Stone et al., 2013). Evaluating their effectiveness requires similarly scaled LST data.

Unfortunately, land surface temperature from Landsat Mission thermal data is difficult to calculate accurately because it is limited to a single thermal band (though Landsat 8 TIRS has two thermal bands, band 11 continues to have calibration issues, limiting Landsat 8 to one functional thermal band for the foreseeable future (Montanaro et al., 2014)). Since the thermal infrared (TIR) radiation received by the sensor is a combination of radiance emitted from the earth's surface, radiance emitted upward from the earth's atmosphere, and atmospheric radiance emitted downward toward the earth's

surface and reflected skyward, accurate LST cannot be determined without correcting for these atmospheric effects (Li et al., 2013). Additionally, the earth's surface is not a blackbody; only a fraction of the absorbed radiation is then emitted to be measured by the sensor. Land surface emissivity (LSE)—the ratio of TIR radiation emitted from a given surface compared to what would be emitted by a blackbody at the same temperature and wavelength—must also be estimated and the effects on LST corrected. Numerous factors impact LSE—including chemical composition, moisture, roughness, phenology, and wavelength—making it difficult to predetermine at the pixel scale (Salisbury and D'Aria, 1992). With two or more thermal channels, the differential absorption between the channels (either due to different effective wavelengths or different viewing angles) can be used to correct for atmospheric interference (Zhengming and Dozier, 1996). Single-channel LST retrieval, on the other hand, requires both a priori LSE as well as accurate, ancillary atmospheric profiles concurrent with the satellite overpass and collocated with the study area (Li et al., 2013).

Given these difficulties, there are three commonly used approaches to LST retrieval from Landsat thermal data: the Radiative Transfer Equation (RTE) (Barsi et al., 2003; Barsi et al., 2005), the Mono-Window Algorithm (MWA) (Qin et al., 2001; Wang et al., 2015), and the Generalized Single Channel (GSC) method (Jimenez-Munoz et al., 2009; Jimenez-Munoz et al., 2014b; Sobrino et al., 2004). MWA and GSC are both simplified versions of the RTE, and all three methods requires slightly different atmospheric parameter inputs—transmittance, upwelling radiance, and downwelling radiance for RTE; transmittance and mean atmospheric temperature for MWA; precipitable water vapor (PWV) for GSC—and all require a priori emissivity

estimations. In practice, emissivity and atmospheric inputs for single-channel LST retrieval algorithms are often cobbled together after-the-fact from sources that were not intended for this purpose, introducing additional error into the LST-retrieval process.

Once these corrections are performed, three options exist for validation (Coll et al., 2009; Hakuba et al., 2014; Li et al., 2013): (1) validating against in situ skin temperature measurements, (2) comparing measured top-of-atmosphere (TOA) radiance with TOA radiance produced by a radiative transfer model using the retrieved LST and highly accurate, concurrent atmospheric data as inputs, and (3) cross-validating retrieved LST with another, well-validated LST dataset as close to the same time and place as possible. None of these options is ideal for validating LST results under commonly non-ideal research conditions. The systematic, longitudinal skin temperature measurements across large study areas necessary for in situ skin temperature validation are rare (Li et al., 2013; Sobrino et al., 2004). These point measurements also may not be representative of pixel-scale remote sensing measurements, especially given the high spatial and temporal variability of LST (Prata et al., 1995). Often this method is limited to homogenous land cover like lakes, agricultural fields, or deserts, which is optimal for an initial validation of a new method, but leaves error estimation under non-homogenous conditions unquantified. Radiance-based validation requires high-quality, contemporaneous atmospheric profiles which are unavailable for many, if not most, research applications using historical TIR data. Finally, cross-validation against other satellite-based LST datasets is often limited by the unavailability of concurrent, collocated datasets.

All three Landsat LST retrieval methods, in theory, are validated using a combination of all these methods as accurate at 1–2 K (Barsi et al., 2005; Qin et al., 2001; Sobrino et al., 2004). However, these validations focus on the accuracy of the method applied under ideal conditions; there is little information about which methods work well under the non-ideal conditions often encountered by researchers. As a result, scientists wishing to apply LST in their research understandably default to the simplest retrieval method, which may or may not actually be appropriate. Further, because Landsat LST retrieval for applications is often performed in suboptimal conditions or with imperfect input datasets, the validated accuracies are often much higher than those actually achieved. However, because retrieval error rates have not been quantified under a variety of atmospheric conditions or with commonly used ancillary input datasets (as opposed to, for example, the atmospheric radiosoundings used for validation), the magnitude of the additional error can be difficult to determine and is too often ignored. It is important to keep in mind that, while LST measurements have many potential applications, spaceborne thermal remote sensing is relatively less developed than multispectral or even radar remote sensing (Kuenzer et al., 2013), though recent years have shown an increase in interest (search results for peer-reviewed articles using keywords “land surface temperature” and “remote sensing” from 2015 were double the results from 2010, 204 articles returned compared to 102). As a result, methods for LST retrieval and application are relatively untested, compared to other areas of remote sensing.

Despite the difficulties associated with thermal remote sensing, potential applications are wide-ranging (Kuenzer et al., 2013; Li et al., 2013; Mia and Fujimitsu,

2013). They include: model input for agriculture, climatology, and hydrology; vegetation cover change detection and analysis; analysis of urban climatology, urban heat islands (UHIs), and heat sinks; volcano monitoring and geothermal analysis; fire monitoring (of forest, peat, or coal) and burned area detection; soil moisture retrieval, evapotranspiration estimation, and drought detection; identification of rocks and minerals; and monitoring industrials sites, mining operations, and pipeline security.

Researchers have attempted to use Landsat thermal data for almost all of these applications. Landsat LST applications for drought monitoring, evapotranspiration estimation, and managing water resources are becoming more common. Ghaleb et al. (2015), for example, use NDVI and LST from Landsat to monitor drought conditions in Lebanon from 1982–2014. Gampe et al. (2016) parameterize irrigation modeling using Landsat LST and NDVI in the Mediterranean to improve future agricultural security against drought. Orhan et al. (2014) use Landsat-based vegetation and temperature indices to evaluate the impacts of drought and groundwater use in the Salt Lake Basin Area of Turkey.

Researchers have also found useful applications of Landsat LST data for fire monitoring and burn detection, as well as volcano and geothermal monitoring. Roy et al. (2015) use Landsat LST to detect coal fires in an Indian coal field, while Hongyuan et al. (2014) demonstrated a similar application in China. Quintano et al. (2015) and Vlassova et al. (2014) both use Landsat LST to detect and predict burn severity in Mediterranean forest fires. Volcano and geothermal monitoring are common applications. Mia et al. (2015) used Landsat to calculate the heat discharge rate after a 1995 eruption in Japan. González et al. (2015) use Landsat thermal data to track heat

and mass flux from 2000–2004 for the Lascar volcano in northern Chile. Tian et al. (2015) use Landsat LST in their detection and quantification of potential geothermal resources.

There are also other commercial/industrial applications of Landsat. For instance, Walsh-Thomas et al. (2012) used Landsat LST images from 1984–2011 over a large California wind farm to identify a downwind warming trend. Yan et al. (2014) also used Landsat LST to longitudinally monitor heat at a municipal solid waste disposal facility in Canada.

Landsat LST is commonly used for land use/ land cover change detection in vegetated areas. For example, Ramdani et al. (2014) relate decreases in Indonesian palm oil plantation foliage cover to increases in surface temperature. Similarly, Huang and Anderegg (2014) similarly explore the climatic impacts, including changes in LST, of post-drought Aspen die-off in Colorado.

Most common, however, is the use of Landsat thermal data for urban applications, especially urban-rural and intra-urban patterns of LST. Some studies simply look at vegetation in urban areas, exploring how land cover changes impact urban heat—Rogan et al. (2013), for instance, explores the impact of tree cover loss due to attempts to eradicate an invasive beetle in on local LST Worcester County, Massachusetts. Many of these urban vegetation studies are concerned with the intra-urban structure and distribution of vegetation and the resulting relationship with LST. Zhou et al. (2011), Zhang et al. (2013), Maimaitiyiming et al. (2014), and Asgarian et al. (2015) all use landscape metrics like fractional vegetation cover, edge density, patch density, and percentage of landscape to quantify spatial patterns of green space and

urban hotspots in urban centers (Baltimore; Sydney; Aksu in Northwestern China; and Isfahan, Iran, respectively). The longitudinal effects of urban expansion on LST over time are also a concern, explored by Fu and Weng (2016) and Lu et al. (2015) in Atlanta and Shenyang, China, respectively. Some studies, like Li et al. (2016), also explore intraurban patterns of LST with respect to demographic and socioeconomic factors. Others, like White-Newsome et al. (2013) and Kestens et al. (2011), use Landsat LST to make explicit the connection between variations in urban LST distribution and public health, often highlighting socioeconomic and demographic inequalities.

Many of these applications of Landsat LST offer up intriguing results which require the spatial resolution and multi-decade archive of the Landsat mission. However, because thermal remote sensing is applicable to so many disciplines, many applications of Landsat LST retrieval methods are published in disciplinary journals outside the remote sensing field (Kuenzer et al., 2013). Journals cited here, for example, include *Landscape and Urban Planning*, *Environmental Health Perspectives*, *Applied Geography*, *Climate*, and *Journal of Volcanology and Geothermal Research*. This means that the rigor with which the TIR images are processed and analyzed is not always subject to the scrutiny of the remote sensing community. For Landsat thermal data this is especially problematic, given the large amount of preprocessing required as well as potentially high error rates that often go undetected. Huang and Anderegg (2014), for example, in *Journal of Geophysical Research: Biogeosciences*, simply converted at-sensor TIR radiance to top-of-atmosphere brightness temperature, then applied an assumed uniform transmittance of 0.95, with no apparent emissivity

correction. For summer images, as used in the study, the unaccounted-for variation in atmospheric interference could easily render any LST change detection meaningless.

Ultimately, without a meaningful comparison of different atmospheric correction methods under actual research conditions, there is little reason for a researcher not to use the simplest method available. Similarly, without an understanding of error rates for these methods as applied by researchers, especially under a range of atmospheric conditions, it is difficult to establish a baseline for meaningful change detection.

Chapter 2: An Intercomparison of Landsat Land Surface Temperature Retrieval Methods under Variable Atmospheric Conditions Using in Situ Skin Temperature

2.1. Background

Land surface temperature (LST), calculated based on remotely sensed thermal infrared (TIR) data, is generally accepted as a proxy for the Earth's skin temperature and serves as valuable input for numerous climatic and ecological applications, including climate change, evapotranspiration, vegetation monitoring, hydrological cycle modeling, and urban health and environmental studies (Bindhu et al., 2013; Han and Xu, 2013; Kalma et al., 2008; Maimaitiyiming et al., 2014; Voogt and Oke, 2003; Weng, 2009). For many of these applications, the thirty-year archive of relatively fine spatial resolution (60–120 m) LST retrieved from Landsats 4-5, 7, and 8 promises a uniquely valuable resource. Sobrino et al. (2012), for example, showed that the magnitude of the surface urban heat island (SUHI), is significantly underestimated at coarser resolutions—120 m Landsat imagery would both provide increased detail on the intraurban heat patterns and more accurately quantify the SUHI. Fu and Weng (2015) suggest that utilizing the full Landsat TIR archive (1982-present) offers a singular opportunity to study changes in both inter- and intra- annual LST patterns, with implications for public health and our understanding of the effect of human-environment interactions on thermal regimes.

However, accurately calculating LST from a single thermal band, as is the case with Landsat 4-5, 7, and 8 (as long as band 11 continues to have calibration issues), is difficult. At-sensor thermal radiance is a combination of radiance emitted from the earth's surface, radiance emitted upward from the earth's atmosphere, and atmospheric

radiance emitted downward toward the earth's surface and reflected skyward. In order to determine LST accurately, the emitted surface radiance must be isolated—primarily by correcting for the effects of water vapor in the atmosphere—and corrected for land surface emissivity (LSE). With thermal data from Landsat TM and ETM+ limited to a single band, it is impossible to correct for atmospheric interference or LSE without ancillary data (Li et al., 2013). Three different LST retrieval methods are commonly used with Landsat TM and ETM+ (and increasingly with Landsat 8 TIRS band 10): the Radiative Transfer Equation (RTE), the Mono-Window Algorithm (MWA), and the Generalized Single Channel (GSC) method. Each requires slightly different atmospheric parameter inputs—transmittance, upwelling radiance, and downwelling radiance for RTE; transmittance and mean atmospheric temperature for MWA; precipitable water vapor (PWV) for GSC—and all require a priori emissivity estimations.

Reported accuracy of these different methods vary, but is generally cited as below 2 K: the GSC method has expected errors between 1 and 2 K when PWV is between 0.5 and 2 g/cm^2 (Jimenez-Munoz et al., 2009); the RTE approach used by Barsi et al. (2005) has a global expected accuracy of about 2 K; and Qin et al. (2001) estimate error for LST retrieval with MWA to be between 1.0 and 1.5 K when transmittance is above 0.8. A few different approaches have been taken to validate LST retrieval methods. One approach, used by Barsi et al. (2005) and Cook et al. (2014), validates retrieved LST using ground measurements of the surface temperature of water, which is easier to measure than land surface temperature and has a known and constant emissivity. Alternatively, as in Qin et al. (2001), radiative transfer code can be used to simulate atmospheric conditions to test the theoretical accuracy of a retrieval method.

Similarly, Sobrino et al. (2004) and Jimenez-Munoz et al. (2009) paired MODTRAN radiative transfer code with concurrent radiosounding data to validate their LST retrieval method. In situ measurements of skin temperature over land are not often used, primarily because measurements gathered on a large enough spatial scale to be useful validating satellite data are rare (Brabyn et al., 2014; Qin et al., 2001; Sobrino et al., 2004).

These validation methods generally provide the best possible theoretical error rate for the method in question, which is essential for establishing the fundamental validity of each approach. However, errors reported from these validations are underestimations of error rates incurred through the actual use of a method as employed for applications, when cloud contamination, non-optimal atmospheric data, and estimated emissivity are at play. Further, studies suggest that atmospheric correction is likely to contribute the largest share of error to LST retrieval methods (Hook et al., 2007; Qin et al., 2001). Yet LST retrieval methods are often validated under optimal atmospheric conditions, when precipitable water vapor in the atmosphere is relatively low, minimizing the impact of atmospheric interference (Barsi et al., 2005; Qin et al., 2001; Sobrino et al., 2004). Additionally, it is understood that for all single-channel retrieval methods, when PWV is higher than 2 g/cm^2, the accuracy of LST retrieved using transmittance values based on PWV decreases significantly (Jimenez-Munoz et al., 2009). However, global PWV routinely rises above this 2 g/cm^2 threshold, especially in summer months—a problem for research projects involving urban heat islands, for example, which often focus on summer LST patterns. Relatedly, cloud detection, especially of cirrus clouds, also poses a problem for LST retrieval from

Landsat 4–7. Landsats 4–7, unlike Landsat 8, do not have a SWIR band that is specifically optimized for detecting optically thin cirrus clouds (1.36–1.38 μm). The consequence is that, even using cloud masking software, cloudy pixels are identified as clear land, with artificially low LST measurements (Zhu et al., 2015). Previous research found that without any cloud masking, mean error rates in Landsat LST estimation can approach -9 K, while visually confirming cloud-free images can bring mean error rates close to zero (Cook et al., 2014). This suggests that there is no automated way to confirm that Landsat 4-5 and 7 images are cloud-free, and that undetected cloud contamination can introduce a negative bias into LST results.

Ultimately, the reality of LST retrieval under non-ideal research conditions raises questions about both the accuracy and suitability of different LST retrieval methods under different conditions. A few intercomparisons have been performed for LST retrieval methods for Landsat 5 TIR data, with mixed results. Sobrino et al. (2004) compared their GSC method with the MWA method for one July Landsat image over Valencia, Spain. They compared transmittance values based both on radiosounding data and PWV; the RTE using concurrent radiosounding data as input was treated as ground truth. They found that MWA with transmittance based on radiosounding data had a RMSE of 0.9 K, and a RMSE of 1.9 K with PWV-based transmittance; the GSC method achieved a RMSE of 1.0 K using PWV as input. Zhou et al. (2012) performed an intercomparison of the MWA and GSC methods in an arid region of northwestern China during March and April 2008, validating the methods against (1) a limited dataset of in situ skin temperature measurements, (2) skin temperature simulated with radiosoundings and MODTRAN 4.0, and (3) skin temperature calculated using the RTE

and radiosounding data. Similar to Sobrino et al. (2004), they also calculated LST using all three methods with both radiosounding data and PWV-based inputs. They found that for one study site, GSC was most accurate, and all methods had accuracies within 2–3 K, while for the other two study sites, MWA performed better. Because their study region featured extremely low PWV, their results also suggested atmospheric correction may not have been necessary. They also noted that the limited amount of in situ skin temperature measurements restricted their capacity to fully evaluate the three methods. Most recently, Vlassova et al. (2014) performed a study of 13 Landsat images in Central Spain from 2009–2011, using skin temperature simulated with MODTRAN 5 to validate the GSC and MWA methods as well as the RTE method employed using NASA's online atmospheric correction parameter calculator (as done in this paper). Only three dates featured PWV marginally above 2.0 g/cm^2. Based on one sample point per image, they found that GSC resulted in RMSD of 0.5 K, RTE of 0.85 K, and MWA of 2.34 K.

Though these intercomparisons are useful, they are focused on relatively limited Landsat datasets, study regions featuring low PWV, and they rely heavily on modeled validation datasets with limited in situ validation data. While not commonly used, in large part because ground measurements of skin temperature are relatively rare (Li et al., 2013), studies suggest that in situ measurements are appropriate for validating satellite LST measurements. The biggest concern when using this sort of validation is the mismatch in scale—given a certain degree of spatial variability, how representative is a point measurement of a parameter like LST of the spatially aggregated value as measured by a satellite (Román et al., 2009)? Rigo et al. (2006) found a strong

correlation between TIR emissions measured by satellites with a range of spatial resolutions (they examined data from AVHRR, MODIS, and Landsat 7 ETM+) and by ground stations in both urban and rural areas. Further, Hale et al. (2011) successfully used 90-meter ASTER TIR data to determine the spatial variability of LST at the 1-km MODIS scale in order to predict how representative point LST measurements were of the 1-km MODIS LST. This suggests that Landsat 5 TM LST, which has a similar spatial resolution to ASTER, is sufficiently sensitive to the spatial variability of LST to justify the use of in situ point measurements directly for validation. Finally, Mesonet site standards state that stations should be placed in locations where the physical characteristics (soil composition, land cover, etc.) are representative of as large an area as possible (Brock et al., 1995). This suggests that Mesonet stations are generally typical of their surroundings, making a comparison with the encompassing 120-m Landsat 5 TM TIR pixel appropriate.

The goal of this paper is to compare the accuracy of the three most commonly applied atmospheric correction methods for retrieving LST from Landsat 5 TM TIR data—as they are actually applied—using both modeled transmittance values and transmittance estimated using PWV, over a range of atmospheric conditions. Importantly, retrieved LST datasets are validated using in situ skin temperature measurements from twenty-five Oklahoma Mesonet stations on fifteen dates in 2005 and 2006. A conservative, cloud-free subset was also analyzed, as were subsets from low-PWV scenes (≤ 2 g/cm^2) and high-PWV scenes (> 2 g/cm^2).

2.2. Study Area

This study focuses on the two Landsat 5 TM scenes that encompass Oklahoma City and the surrounding region (path 28, rows 35 and 36) (Figure 1).The majority of the study area is Central Great Plains, composed primarily of grassland and cropland. The southeast region, however, is Cross Timbers, with a mix of grassland/rangeland and oak trees. The largest urban center in the study area is the Oklahoma City Metropolitan Area, with a 2006 population estimated by the U.S. Census Bureau of approximately 1.2 million, and a footprint of approximately 16,000 km^2. Climatic variables in Oklahoma generally demonstrate a gradient from northwest to southeast, with temperatures and precipitation increasing (Costa et al., 2008). Additionally, Oklahoma demonstrates large intra-annual PWV and temperature gradients, making it ideal for exploring the impact of these changes on LST retrieval.

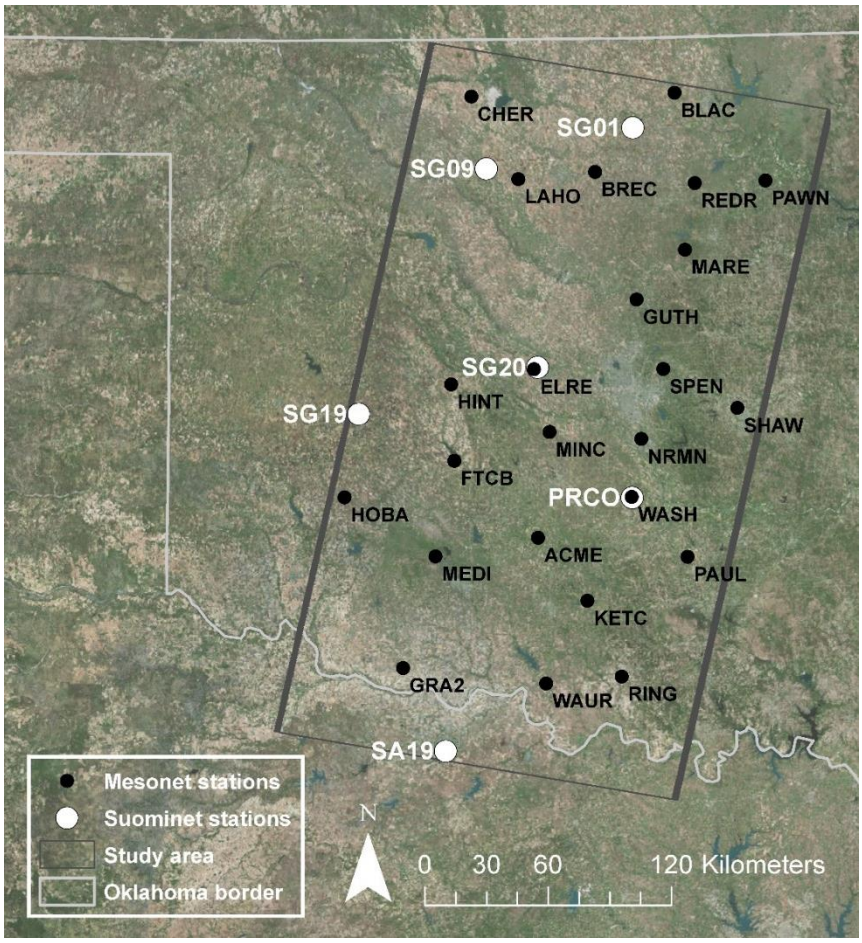

Figure 1. Map of the study area. The Landsat footprint is in dark gray, the Suominet stations used to generate atmospheric parameters are in white, and the Mesonet stations, which gathered skin temperature, are in black.

2.3. Data and Preprocessing

2.3.1 Landsat 5 TM data

We identified Landsat 5 TM data from the USGS EarthExplorer website for all dates in 2005 and 2006 in our study area with approximately less than 10% cloud cover across both scenes. These years were chosen because the Mesonet skin temperature sensors were phased out in 2007, and because there was a preponderance of relatively cloud-free dates. As shown in Table 1, Oklahoma was drier and slightly warmer than the 1901–2000 means in both 2005 and 2006 (NOAA, 2015).

We ultimately downloaded and processed twenty-two images on eleven dates in 2006 and twelve images on six dates in 2005, all with low or no cloud cover according to the image metadata (Tables 2–3). For each of the thirty-four scenes, we (1) radiometrically corrected the Band 6 thermal infrared data (TIR) using the bias and offset values from each scene's metadata and (2) used the Fmask v3.2 algorithm to mask clouds, cloud shadow, water, and snow in all bands, using a conservative ten-pixel buffer for clouds and cloud shadow and a 0.50% cloud probability threshold (Zhu et al., 2015). It is worth noting that while the Landsat metadata may show little or no cloud cover, Fmask is much more conservative when identifying clouds. We also removed pixels with top-of-atmosphere radiance below 1.15 W m−2 sr−1 μm−1, which appeared to signal invalid data on the edges of scenes far below the overall distribution of pixels in the images. After all data were processed, we excluded two dates due to highly variable PWV, leaving fifteen final dates that were analyzed further.

Table 1. Oklahoma annual temperature and precipitation. 2005 and 2006 compared to 1901–2001 means. Both years were similar, demonstrating higher temperatures and less precipitation than the long-term mean (NOAA, 2015).

Measurement	1901–2000 mean	2005	2006	2005 difference from mean	2006 difference from mean
Avg Min Temperature (°C)	8.7	9.3	9.6	0.6	0.9
Mean Temperature (°C)	15.4	16.0	16.8	0.6	1.4
Avg Max Temperature (°C)	22.1	22.8	24.0	0.7	1.9
Precipitation (mm)	859.5	700.0	757.2	-159.5	-102.3

Table 2. Landsat TM scene details and modeled atmospheric correction parameters used to calculate LST.

Scene	Date	Cloud cover (%)	Used in conservative analysis	τ	Upwelling Radiance	Downwelling Radiance
LT50280352005020PAC01	1/20/2005	0	N	0.84	1.10	1.81
LT50280352005116PAC01	4/26/2005	1	N	0.85	1.00	1.69
LT50280352005180PAC01	6/29/2005	2	N	0.65	2.99	4.69
LT50280352005212PAC01	7/31/2005	0	Y	0.70	2.54	4.09
LT50280352005244GNC01	9/1/2005	0	N	0.59	3.41	5.23
LT50280352005308PAC01	11/4/2005	0	N	0.80	1.51	2.49
LT50280352006023PAC01	1/23/2006	1	N	0.91	0.58	0.97
LT50280352006103PAC01	4/13/2006	0	Y	0.73	2.28	3.64
LT50280352006167PAC01	6/16/2006	5	N	0.61	3.30	5.08
LT50280352006183PAC04	7/2/2006	0	N	0.62	3.24	5.06
LT50280352006199PAC01	7/18/2006	0	N	0.46	4.59	6.89
LT50280352006263PAC01	9/20/2006	0	N	0.80	1.58	2.60
LT50280352006279PAC01	10/6/2006	0	Y	0.63	2.92	4.50
LT50280352006295PAC01	10/22/2006	0	Y	0.89	0.70	1.17
LT50280352006311PAC01	11/7/2006	0	N	0.70	2.04	3.26
LT50280352006327PAC01	11/23/2006	0	Y	0.85	1.13	1.84
LT50280352006359PAC01	12/25/2006	0	Y	0.96	0.24	0.41
LT50280362005020PAC01	1/20/2005	0	Y	0.84	1.14	1.87
LT50280362005116PAC01	4/26/2005	1	N	0.84	1.12	1.88
LT50280362005180PAC01	6/29/2005	0	Y	0.71	2.52	4.05
LT50280362005212PAC01	7/31/2005	0	Y	0.68	2.66	4.23
LT50280362005244GNC01	9/1/2005	0	N	0.65	2.96	4.59
LT50280362005308PAC01	11/4/2005	0	Y	0.80	1.56	2.57
LT50280362006023PAC01	1/23/2006	7	N	0.92	0.48	0.80
LT50280362006103PAC01	4/13/2006	0	Y	0.72	2.37	3.74
LT50280362006167PAC01	6/16/2006	9	N	0.61	3.30	5.07
LT50280362006183PAC04	7/2/2006	11	N	0.61	3.26	5.07
LT50280362006199PAC01	7/18/2006	1	N	0.53	4.16	6.34
LT50280362006263PAC01	9/20/2006	0	N	0.81	1.56	2.55
LT50280362006279PAC01	10/6/2006	0	N	0.62	2.98	4.56
LT50280362006295PAC01	10/22/2006	11	N	0.86	0.90	1.49
LT50280362006311PAC01	11/7/2006	19	N	0.70	2.09	3.32
LT50280362006327PAC01	11/23/2006	0	Y	0.86	1.05	1.72
LT50280362006359PAC01	12/25/2006	0	Y	0.93	0.38	0.65

Table 3. Suominet atmospheric parameters. Parameters used to calculate LST, by Landsat scene.

Landsat Scene	Date	Suominet Station Name	Pressure (kPa)	Air Temp (degC)	Relative Humidity	PWV (g/cm^2)	Est. τ
LT50280352005020PAC01	1/20/2005	SG20	968.6	16.3	47.9	1.05	0.88
LT50280352005116PAC01	4/26/2005	SG20	961.2	19.9	31.1	1.15	0.87
LT50280352005180PAC01	6/29/2005	SG20	964.3	32.7	43.8	2.43	0.75
LT50280352005212PAC01	7/31/2005	SG20	971.2	32.4	38.8	2.72	0.72
LT50280352005244GNC01	9/1/2005	SG20	969.5	30.7	53.9	2.23	0.77
LT50280352005308PAC01	11/4/2005	SG20	959.0	25.9	34.3	1.49	0.84
LT50280352006023PAC01	1/23/2006	SG20	976.1	7.3	65.7	0.66	0.92
LT50280352006103PAC01	4/13/2006	SG20	969.6	29.3	48.0	1.72	0.83
LT50280352006167PAC01	6/16/2006	SG20	961.4	30.8	50.5	2.99	0.69
LT50280352006183PAC04	7/2/2006	SG20	969.7	32.9	42.3	3.01	0.68
LT50280352006199PAC01	7/18/2006	SG09	967.1	38.6	30.0	3.97	0.57
LT50280352006263PAC01	9/20/2006	SG09	963.9	27.5	31.6	1.98	0.80
LT50280352006279PAC01	10/6/2006	SG09	975.5	25.5	52.5	2.84	0.65
LT50280352006295PAC01	10/22/2006	SG09	977.9	10.1	49.5	0.97	0.89
LT50280352006311PAC01	11/7/2006	SG09	963.9	20.0	31.1	1.99	0.77
LT50280352006327PAC01	11/23/2006	SG09	967.5	15.5	55.3	0.87	0.90
LT50280352006359PAC01	12/25/2006	SG09	973.3	2.7	51.7	0.42	0.94
LT50280362005020PAC01	1/20/2005	PRCO	979.6	15.1	51.9	1.21	0.87
LT50280362005116PAC01	4/26/2005	PRCO	971.1	20.6	33.0	1.15	0.87
LT50280362005180PAC01	6/29/2005	SG19	958.9	33.9	36.4	2.62	0.73
LT50280362005212PAC01	7/31/2005	PRCO	981.6	31.3	45.3	2.83	0.70
LT50280362005244GNC01	9/1/2005	PRCO	979.0	30.1	53.8	2.12	0.79
LT50280362005308PAC01	11/4/2005	SA19	973.8	25.5	35.2	1.62	0.82
LT50280362006023PAC01	1/23/2006	PRCO	986.5	8.6	47.6	0.78	0.91
LT50280362006103PAC01	4/13/2006	SA19	984.0	27.5	54.3	2.15	0.78
LT50280362006167PAC01	6/16/2006	SA19	973.9	31.0	50.2	3.18	0.66
LT50280362006183PAC04	7/2/2006	SG19	964.2	31.5	47.4	3.12	0.67
LT50280362006199PAC01	7/18/2006	PRCO	976.7	39.4	29.1	2.98	0.69
LT50280362006263PAC01	9/20/2006	SA19	978.7	28.0	32.6	1.52	0.85
LT50280362006279PAC01	10/6/2006	PRCO	986.1	24.4	60.2	2.83	0.65
LT50280362006295PAC01	10/22/2006	PRCO	988.3	11.3	44.7	1.02	0.88
LT50280362006311PAC01	11/7/2006	PRCO	976.4	18.3	61.5	1.89	0.79
LT50280362006327PAC01	11/23/2006	PRCO	979.3	19.3	44.5	1.08	0.88
LT50280362006359PAC01	12/25/2006	PRCO	981.6	5.0	48.2	0.52	0.93

2.3.2. Mesonet data

The Oklahoma Mesonet provides a network of 120 environmental monitoring stations across Oklahoma, with observations for many stations extending back to 1994 (McPherson et al., 2007). Each station has instruments on or near a 10-meter tower that provide observational readings every five minutes. From 1999 to 2007, 89 Oklahoma Mesonet stations hosted infrared temperature (IRT) sensors (Fiebrich et al., 2003)—25 of these stations were located within our study area (Figure 1). The IRT sensors, installed at a height of 1.5 m and with a 0.5 m diameter circle field of view, had an accuracy of approximately ±0.2°C from 15° to 35°C and ±0.3°C from 5° to 45°C. Initial temperatures were calculated using the Stefan-Boltzmann law and assumed land surface emissivity (LSE) of 1.0, and final skin temperatures were calculated as a function of the difference between the initial measured temperature and the temperature of the sensor body. There are two drawbacks to the IRT sensors used to measure skin temperature (Fiebrich et al., 2003). First, the skin temperature sensor and calculation method used overestimates skin temperatures above 40°C by an average of 2.1°C. In Oklahoma summer months, surface temperatures routinely rise above this threshold. To address this overestimation, we introduced a -2.1°C offset for any IRT sensor measurements above that 40°C threshold—36 measurements in total were offset, occurring between late April and early October. Second, despite careful siting of Mesonet stations, the sensor's limited field of view potentially means that a given station may not be representative of the surrounding landscape. To ensure that we relied on data from stations that were representative of the surrounding landscape (to make the use of point data to validate 120-meter raster data more robust), we performed regressions for all

data by station to identify any stations which showed unusually weak correlations between skin temperature and Landsat LST datasets. As a result, we removed the Tipton station from the analysis due to an unusually low correlation; the remaining stations showed satisfactorily high ($r > 0.8$) correlations.

2.3.3. Suominet atmospheric parameters

Suominet is a real-time GPS network that uses phase delays in GPS signals to calculate precipitable water vapor in the atmospheric column (accuracy better than 2 mm), in addition to gathering ground-level atmospheric data (Ware et al., 2000). For each Landsat scene, we acquired air temperature, air pressure, relative humidity, and precipitable water vapor (PWV) from a centrally located Suominet station (three stations per scene were located in our study area and period), as close to concurrent with Landsat 5 TM scene center time as possible (within three minutes) (Figure 1, Tables 2–3). These variables were then used to approximate atmospheric parameters needed to atmospherically correct the Landsat 5 TM thermal data—transmittance, upwelling and downwelling radiances, and mean atmospheric temperature.

2.3.4. MODIS precipitable water vapor

To provide global context for the impact of PWV levels on Landsat LST retrieval, we used the mean maximum precipitable water vapor measurements from the Level-3 MODIS Atmosphere Eight-Day Global Product (MOD08_E3). Generally speaking, MODIS has a comparable overpass time to Landsat. This dataset aggregates cloud-free daily PWV data from MODIS, which have an estimated accuracy of 5-10%, to return a global, 1° grid of eight-day maximum PWV measurements. We calculated

the global average of these PWV measurements in g/cm^2 for the entire time series, as well as seasonal averages.

2.4. Methods

2.4.1. Land surface temperature retrieval

Four land surface temperature datasets were calculated using the three methods—RTE, using modeled atmospheric parameters; GSC, which relies on PWV as a direct input for atmospheric correction; and MWA, using both modeled transmittance and transmittance estimated based on PWV. Additionally, we calculated a fifth LST dataset using no atmospheric correction. The LST retrieval process is summarized in the flow chart in Figure 2.

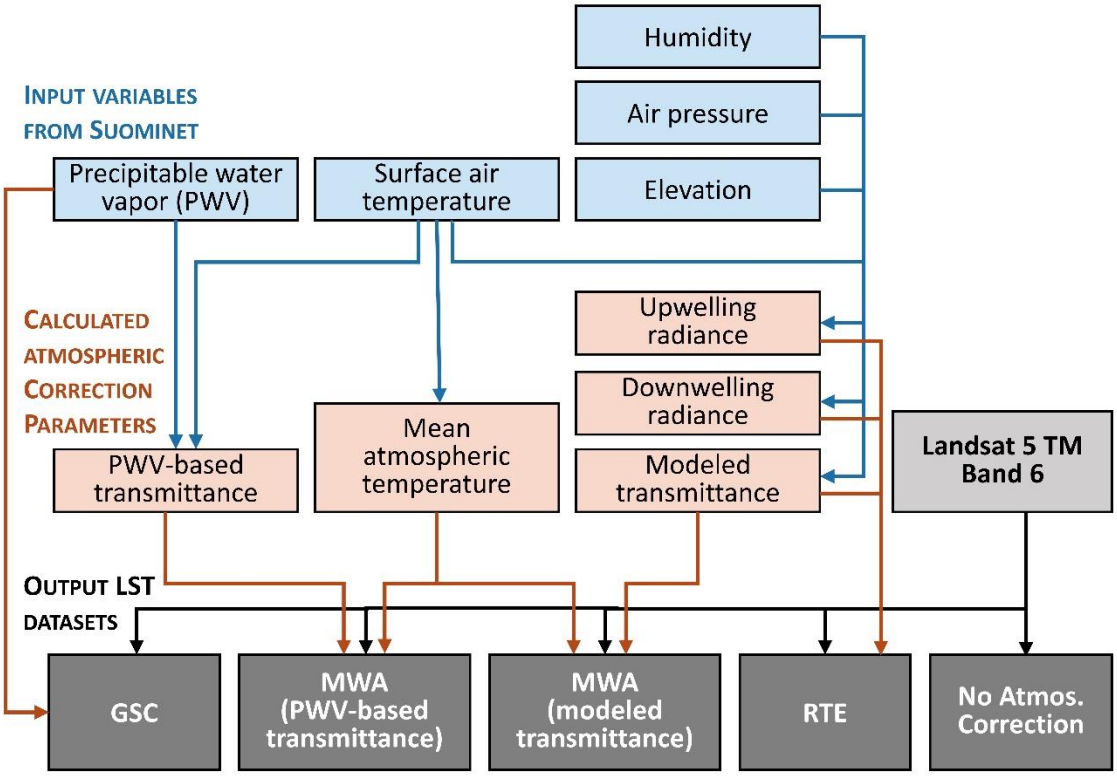

Figure 2. Data processing flow chart. Traces the input variables from Suominet station data, the resulting calculated atmospheric correction parameters, the raw Landsat 5 TM TIR data, and the final LST datasets.

2.4.1.1. Radiative Transfer Equation

The radiative transfer equation (RTE)—which describes the at-sensor radiance as a function of surface, upwelling, and downwelling emission terms for a given band i— can be used to retrieve LST assuming one knows or can approximate band-specific land surface emissivity, atmospheric transmittance, upwelling radiance, and downwelling radiance (Li et al., 2013; Ottle and Stoll, 1993). The at-sensor thermal radiance can be described as (Barsi et al., 2005; Zhou et al., 2012):

$$L_i = \varepsilon_i \tau_i B_i(T_s) + (1 - \varepsilon_i)\tau_i L_i^{\downarrow} + L_i^{\uparrow} \quad (1)$$

where L_i is the top-of-atmosphere radiance (W m^{-2} sr^{-1} μm^{-1}); τ_i is atmospheric transmittance; ε_i is land surface emissivity; $L_i^{\uparrow}$ is upwelling radiance (W m^{-2} sr^{-1} μm^{-1}); $L_i^{\downarrow}$ is downwelling radiance (W m^{-2} sr^{-1} μm^{-1}); and $B_i(T_S)$ is the blackbody radiance at temperature T_S (K). Based on this equation, the emitted blackbody radiance at ground-level for Landsat TM Band 6 can be expressed:

$$B_6(T_s) = \frac{L_6 - (1-\varepsilon_6)\tau_6 L_6^{\downarrow} - L_6^{\uparrow}}{\varepsilon_6 \tau_6} \quad (2)$$

The inverse of the Planck function (adapted for Landsat, specifically) can then be used to calculate LST (K) from the ground-level blackbody radiance (Chander et al., 2009):

$$T_S = \frac{K_2}{\ln\left(\frac{K_1}{B_6(T_S)} + 1\right)} \quad (3)$$

For the TM sensor, the calibration constants are K_1= 607.76 W m^{-2} sr^{-1} μm^{-1} and K_2= 1260.56 K. Atmospheric transmittance, upwelling radiance, and downwelling radiance were modeled as discussed in Section 2.4.1.4.

2.4.1.2 Mono-Window Algorithm

The Mono-Window Algorithm (MWA) is a simplification of the RTE proposed by Qin et al. (2001) to determine LST from Landsat 5 TM Band 6 data using only three input parameters: land surface emissivity, atmospheric transmittance, and effective mean atmospheric temperature (used to estimate upwelling and downwelling radiance terms using a linearized version of Planck's function). The MWA method applies the following equations:

$$T_S = \frac{[a_6(1-C_6-D_6)+(b_6(1-C_6-D_6)+C_6+D_6)T_6-D_6T_a]}{C_6} \quad (4)$$

$$C_6 = \varepsilon_6\tau_6 \quad (5)$$

$$D_6 = (1-\tau_6)[1+(1-\varepsilon_6)\tau_6] \quad (6)$$

where T_a is the effective mean atmospheric temperature (K); coefficients a_6 and b_6 (determined by linearizing the Planck function using Taylor's expansion) equal -67.355351 and 0.458606 respectively, given brightness temperatures between 0 and 70°C; and T_6 is the at-sensor brightness temperature (K) (calculated by applying the inverse of the Planck function to the top-of-atmosphere radiance measured by Landsat 5 TM using Equation 3). More information about the derivation of formulas 4–6 can be found in Qin et al. (2001). Mean atmospheric temperature can be determined as a function of near-surface air temperature for different atmospheric profiles—we used the Suominet air temperature as input for the mid-latitude summer and mid-latitude winter equations as described in Qin et al. (2001). Transmittance was estimated using both the methods described in Section 2.4.1.4.

2.4.1.3. Generalized Single Channel method

The Generalized Single Channel (GSC) method was intended to be an easily operationalized LST retrieval method requiring minimal input. It relies on estimated atmospheric functions dependent on total precipitable water vapor as determined from a variety of atmospheric radiosounding databases and MODTRAN-4 radiative transfer code (Sobrino et al., 2004). The GSC method calculates LST using the equation:

$$T_S = \gamma \left[\frac{1}{\varepsilon} [\psi_1 L_6 + \psi_2] + \psi_3 \right] + \delta \quad (7)$$

where parameters γ and δ are determined as follows:

$$\gamma \approx \frac{T_6^2}{b_\gamma L_6} \quad (8)$$

$$\delta \approx T_6 - \frac{T_6^2}{b_\gamma} \quad (9)$$

using at-sensor brightness temperature (T_6) as calculated using Equation 3, and where b_γ is equal to 1256 K for Landsat 5 band 6, and L_6 is equal to at-sensor radiance in W m^{-2} sr^{-1} μm^{-1}. ψ_1, ψ_2, and ψ_3 are three PWV-based atmospheric functions which relate transmittance, upwelling radiance, and downwelling radiance to PWV. These functions are derived from databases of atmospheric profiles, with slightly different coefficients depending on the database used (equations 10–12). In this case, we used the atmospheric function coefficients derived from the TIGR61 database because it included atmospheric profiles with the widest range of precipitable water vapor, and so was most representative of our study area and period (Jimenez-Munoz et al., 2009). The equations used are:

$$\psi_1 = 0.08735\omega^2 - 0.09553\omega + 1.10188, \quad (10)$$

$$\psi_2 = -0.69188\omega^2 - 0.58185\omega - 0.29887, \quad (11)$$

$$\psi_3 = -0.03724\,\omega^2 + 1.53065\omega - 0.45476, \quad (12)$$

where ω is precipitable water vapor in g/cm^2, as measured by a central Suominet station. More detail about the derivation of this method can be found in Sobrino et al. (2004) and especially Jimenez-Munoz et al. (2009).

2.4.1.4. Estimating atmospheric parameters

For LST retrieval, we used both modeled parameters and parameters estimated as functions of PWV and surface air temperature; all relied on Suominet station data. Parameters required by the RTE to atmospherically correct remotely sensed TIR data—transmittance, upwelling radiance, and downwelling radiance—can be estimated using an atmospheric radiative transfer model applied to an atmospheric profile concurrent with satellite data. In our case, we used NASA's web-based Atmospheric Correction Parameter Calculator (Barsi et al., 2003), which applies the atmospheric radiative transfer model MODTRAN 4.0 to atmospheric profiles generated by the National Centers for Environmental Protection (NCEP), accounting for local conditions input by the user—in this case, we used the Suominet station values to maintain consistency with transmittance values derived from Suominet-based precipitable water vapor. The model takes the point location, time, and atmospheric data from the weather station as input and interpolates the NCEP profiles in time and space for that point—the result is estimated point values for transmittance, upwelling radiance, and downwelling radiance. Those atmospheric parameters are then assumed to be representative for an entire Landsat 5 TM scene.

Transmittance for band 6 is primarily determined by precipitable water vapor in the atmosphere, and thus can be estimated as a function of PWV as described in Qin et

al. (2001). Qin et al. (2001) used LOWTRAN 7 software to simulate the relationship between PWV and transmittance for low and high temperature profiles (centered around surface air temperatures of 18°C and 35°C, respectively) for the PWV range of 0.4-4.0 g/cm^2—given the same PWV, transmittance will be higher at higher temperatures. The result is four linear models relating transmittance to PWV for low and high PWV measurements (0.4-1.6 g/cm^2 and 1.6-3.0 g/cm^2, respectively) at both low and high temperatures. We used measurements from the most central Suominet station with valid precipitable water vapor data available for each date along with concurrent surface air temperature to calculate one transmittance value for each scene using these four equations, assuming a cutoff of 1.6 g/cm^2.

2.4.2. LST validation

In order to validate the LST datasets retrieved from Landsat 5 TM, we used skin temperature measurements from Mesonet IRT sensors as ground truth. We first performed linear regressions on each full LST dataset, examining LST as a function of skin temperature. We also subset each LST dataset into data from scenes with PWV below 2 g/cm^2 and data from scenes with PWV above 2 g/cm^2—since this is an oft-cited threshold above which single-band LST retrieval becomes less accurate—and performed linear regression analysis on each subset. It is reasonable to assume that perfect LST retrieval would result in a one-to-one regression line with skin temperature. To test this, we performed a t-test to determine if each regression line's slope and intercept were significantly different from one and zero, respectively, at the $p=0.05$ level. We also calculated root mean squared error (RMSE) and mean absolute error

(MAE) from skin temperature to determine error rates from ground truth for each method.

Additionally, we were interested whether the regression lines for data above and data below the PWV threshold were the same, or if the relationship between skin temperature and LST changed as PWV increased. To test this, we used analysis of covariance (ANCOVA) testing, introducing a grouping factor to determine whether the two data subsets were statistically distinct. If t-testing showed that the coefficient for the grouping factor's interaction with the regression slope was indistinguishable from zero ($p \geq 0.05$), the null hypothesis that the two subsets have the same slope cannot be rejected. The implication is that the same relationship between LST and skin temperature can be used for the complete dataset.

Cloud detection in Landsat 4–7 data is problematic; unmasked cirrus clouds especially can have an impact on LST results. To determine if cirrus cloud contamination impacted our results, we also applied the same regression and ANCOVA analyses to a more conservative subset of our data, using only images that were visually clear of apparent cloud contamination. Scenes with large dark areas or unexplained streaking were excluded to create a dataset that was as likely to be free of cloud contamination as possible. Figure 3 is a an example of an image that Landsat metadata identified as cloud-free, and for which Fmask provided only minimal masking, yet is likely marked by extensive cloud contamination as evidenced by splotches and streaks of unexplained low temperatures. There may be other explanations for some of the visual artifacts we excluded, but for our more conservative analysis we erred on the side of excluding some good data in order to ensure a dataset free of cloud contamination.

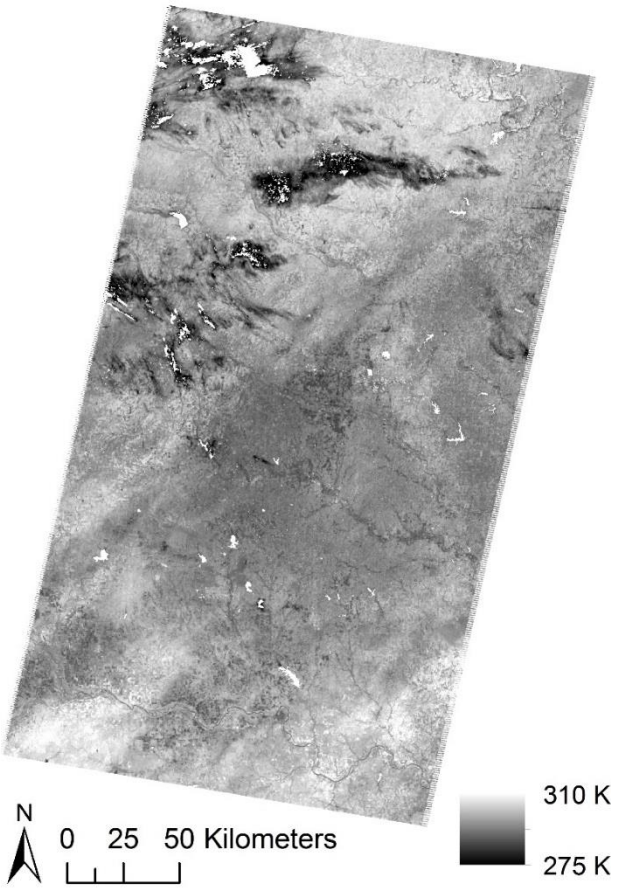

Figure 3. Example of undetected cloud contamination (Sept. 20, 2006). The Landsat 5 TM metadata reports cloud cover for both scenes on this date as 0.00%. Fmask has clearly masked some clouds, but large areas of unusually low temperatures suggest unmasked cloud contamination.

Finally, emissivity estimations pose a significant problem for single-channel LST retrieval (Li et al., 2013). However, because the skin temperature data is provided with an assumed emissivity of one, emissivity corrections for our LST datasets are unnecessary and were not applied. As a result, we avoid the additional uncertainty introduced by emissivity estimation and instead focus on the accuracy of the atmospheric correction methods. However, it also means that all skin temperature and LST retrievals are slightly underestimated, since the actual emissivity of the land

surface is less than one; additionally, that underestimation is distributed unevenly across land cover types.

2.5 Results

2.5.1 Land surface temperature retrieval

LST was calculated from Landsat 5 TM data for seventeen dates using RTE, MWA (using both modeled transmittance and transmittance estimated using PWV), GSC, and no atmospheric correction—five LST datasets in all. Due to variable precipitable water vapor across the study area, two dates (September 1, 2005 and July 18, 2006) were excluded from further analysis. Mean temperature maps based on the four atmospherically corrected datasets for a high water vapor date (July 7, 2005), a medium water vapor date (April 13, 2006), and a low water vapor date (December 25, 2006) can be seen in Figure 4. Histograms of LST results for all methods for the same two dates can be seen in Figure 5—LST variability both within each method and among correction methods clearly increases as PWV increases. Figures 6–8, which illustrate the variation among methods in finer detail, further support this observation. Much higher variability is observable on the high water vapor date, both spatially and among the different methods. While the GSC method produces results comparable to the other methods on the low-temperature, low-PWV date (Figure 6), LST retrieval from GSC appears increasingly warm relative to LST retrieved using the other methods as temperatures and PWV increase (Figures 7 and 8). Similarly, the increase in spatial variability as temperatures and PWV increase is evidenced by the increased contrast between vegetated areas and the surrounding landscape, with vegetated areas staying

relatively cool (the green streaks in Figures 7 and 8) as the rest of the landscape heats up.

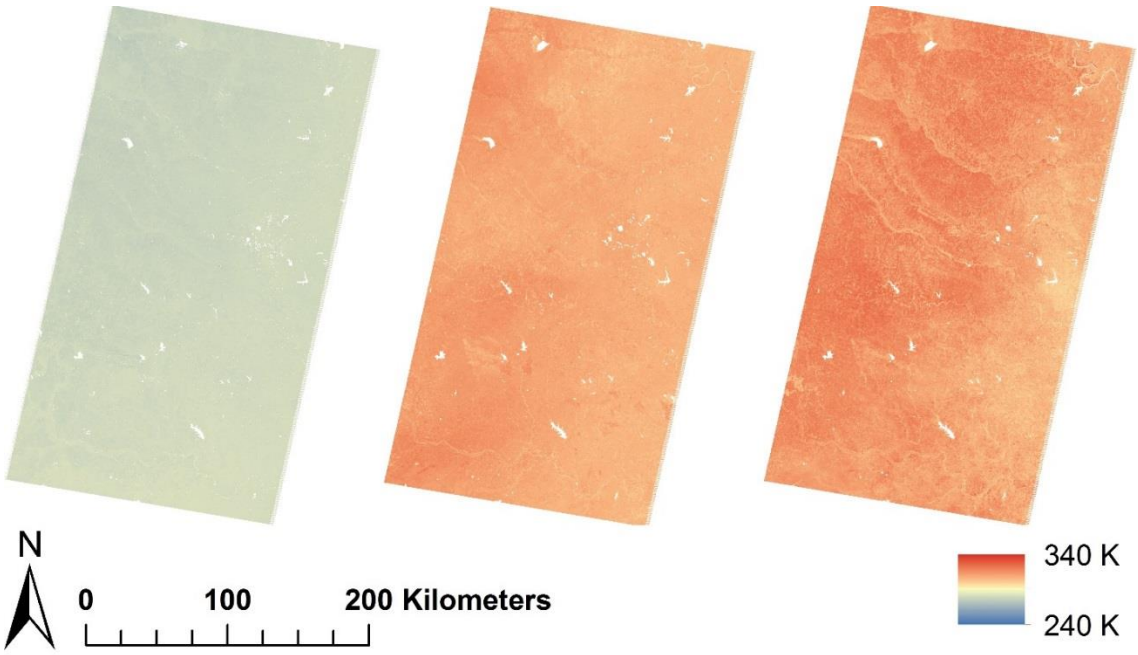

Figure 4. Mean temperature maps. Temperature maps averaging all atmospheric correction methods for (A) low-, (B) medium-, and (C) high-PWV days.

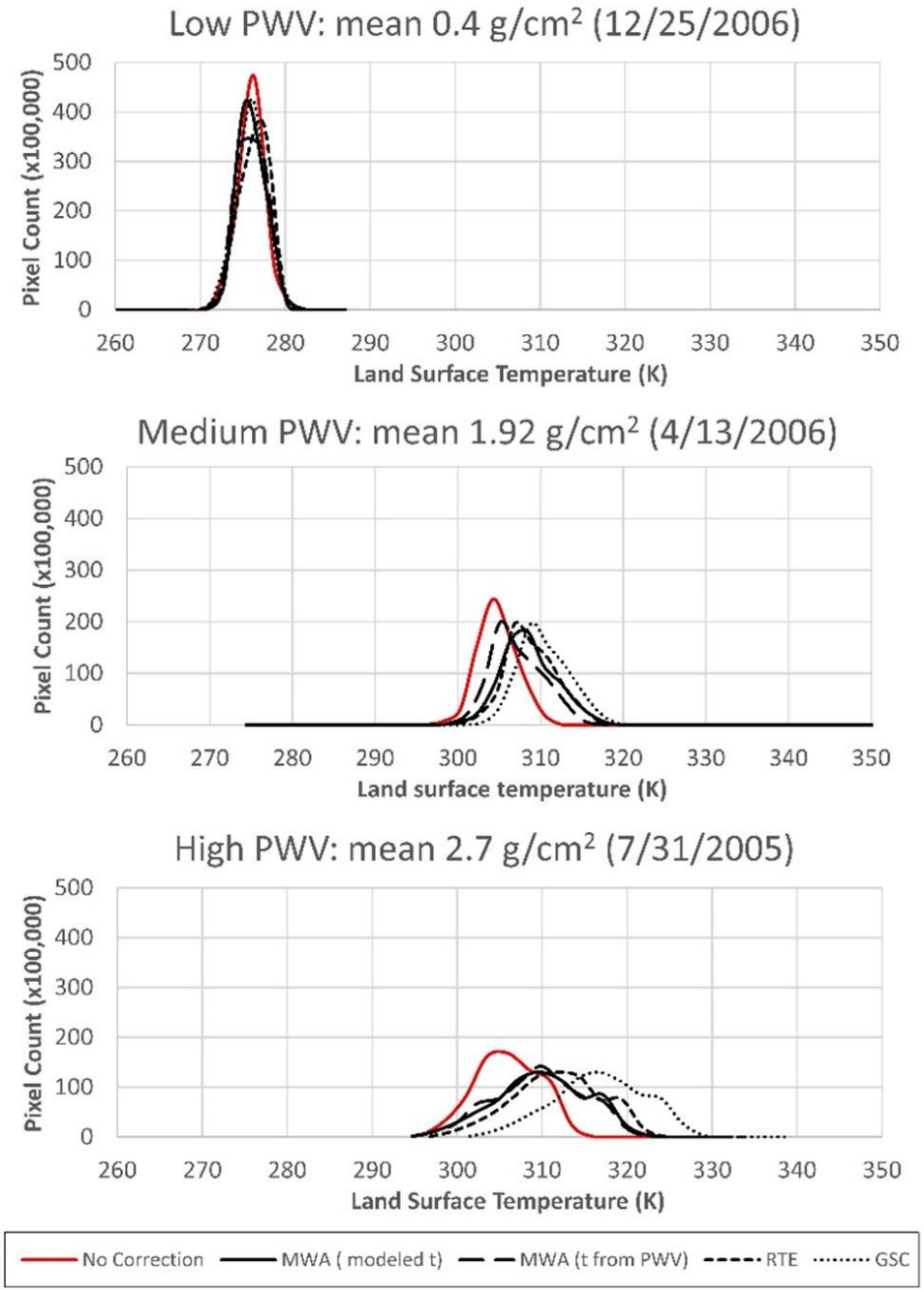

Figure 5. Land surface temperature histograms. Histograms of LST results for all LST datasets for low-, medium-, and high-PWV days. As PWV increases, temperature results become increasingly variable both within each LST dataset and among them. For the low-PWV date, correction method makes little difference overall; for the high-PWV date, it has a large impact on overall results.

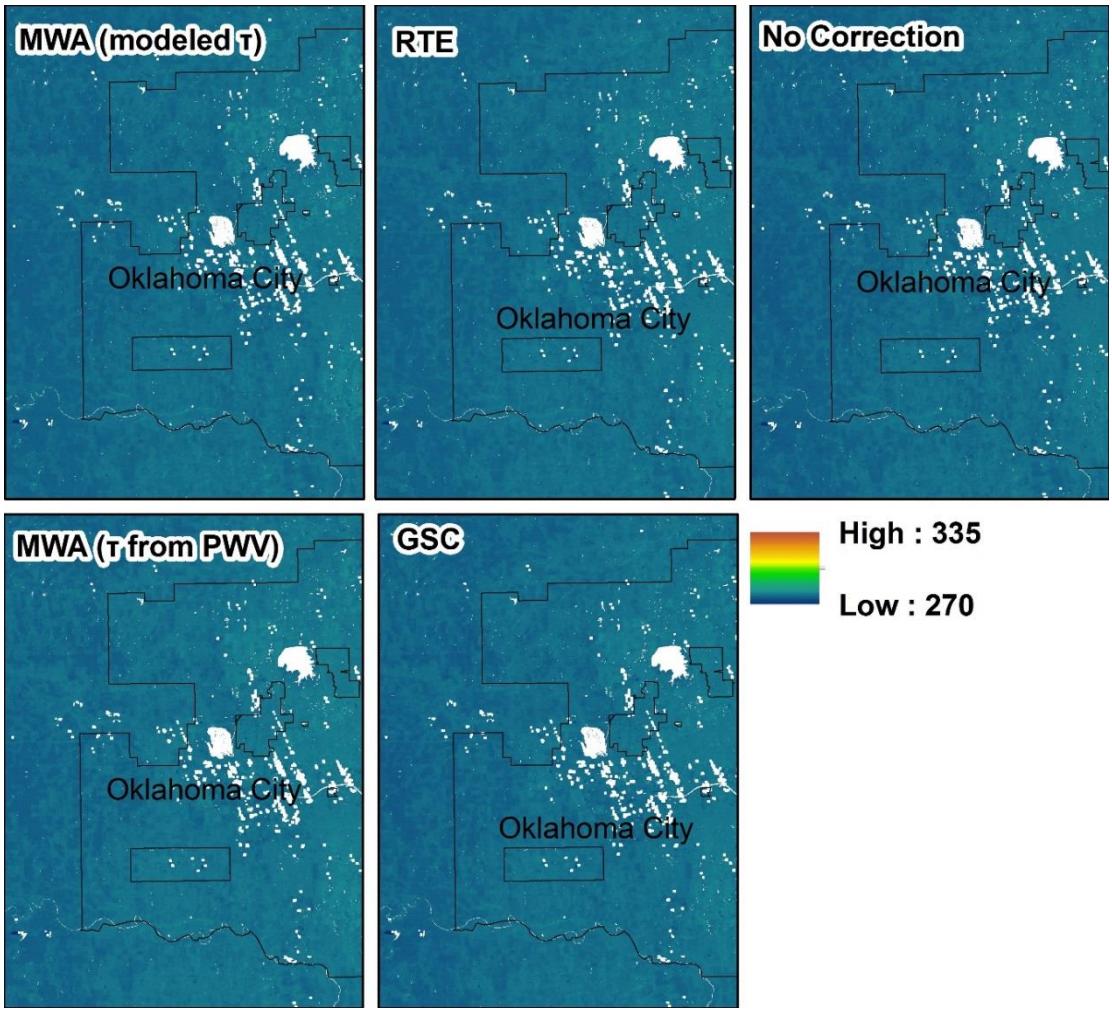

Figure 6. LST detail maps for all methods for a low-PWV day (December 25, 2006). The low variability across space and among methods, relative to the medium- and high-PWV dates illustrated in Figures 7 and 8, is evident. The temperature scale is the same for Figures 6, 7 and 8.

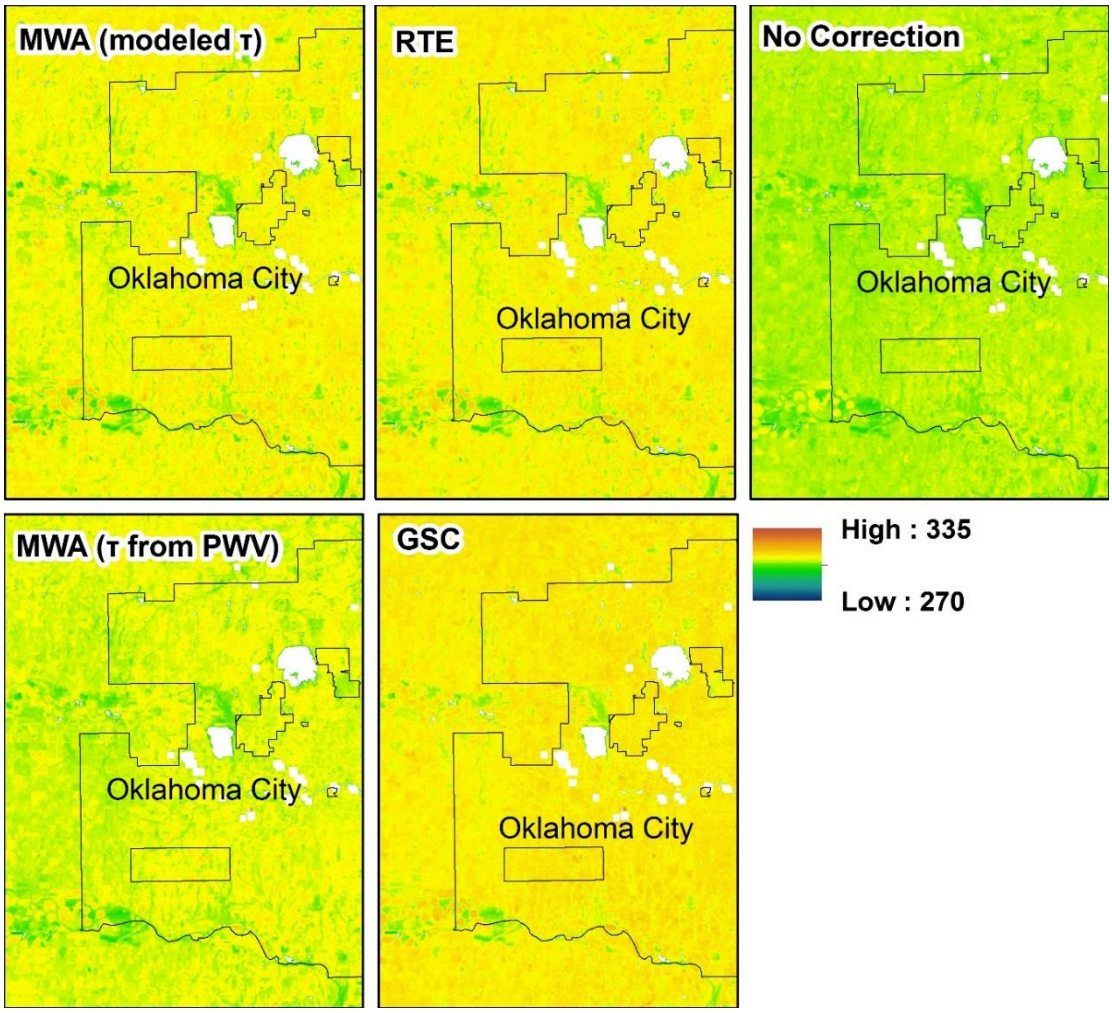

Figure 7. LST detail maps for all methods for a medium-PWV day (April 13, 2006). The increase in variability across space and among methods, relative to the low-PWV date illustrated in Figure 6, is evident, with GSC producing relatively warmer results, and the cooling effect of vegetation apparent in the green streaks across the various LST datasets.

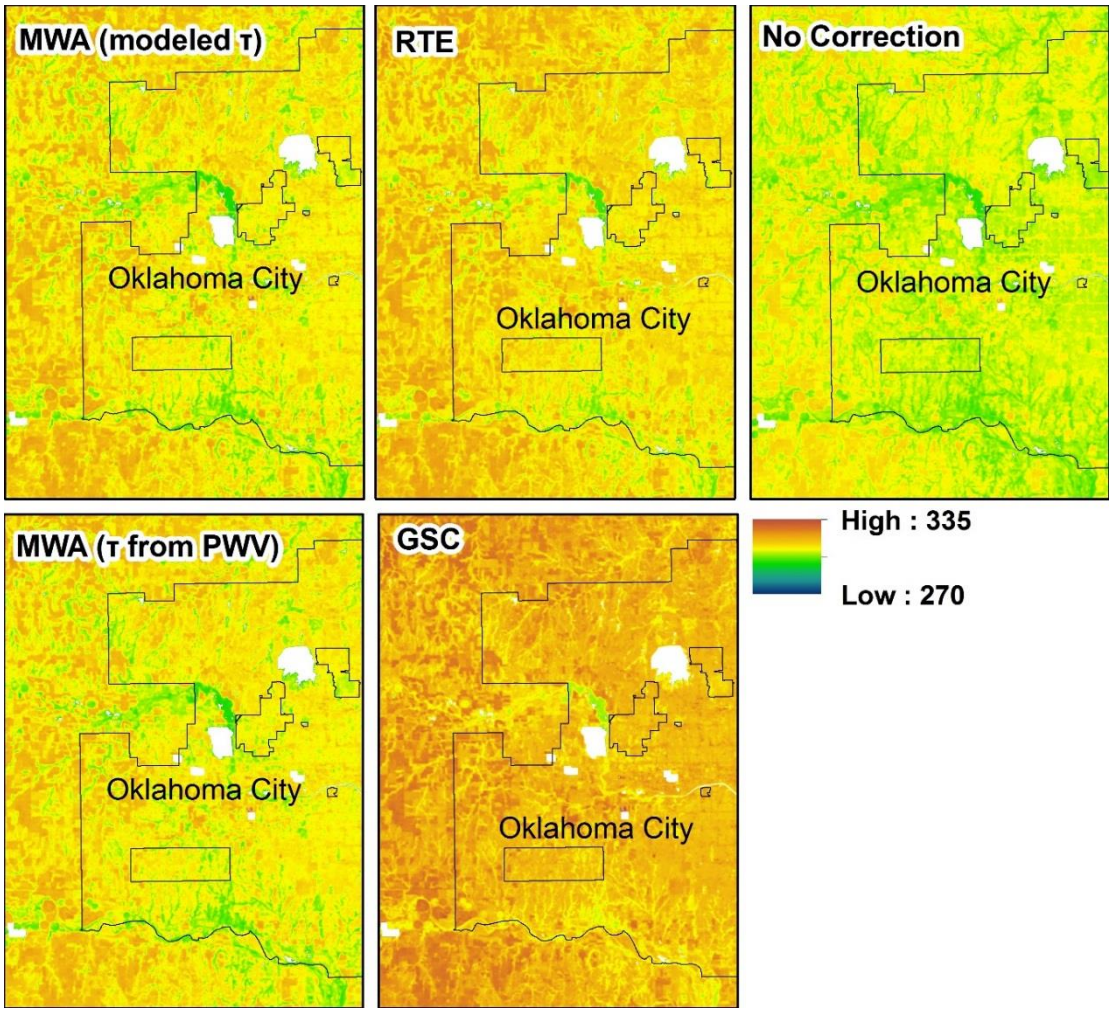

Figure 8. LST detail maps for all retrieval methods for a high-PWV day (July 31, 2005). The increase in variability across space and among methods, relative to the low- and medium- PWV date illustrated in Figures 6 and 7, is evident. GSC produces noticeably warmer results, and the contrast between vegetation and surrounding land cover is increased.

The atmospheric parameters we employed to retrieve LST from Landsat 5 TM—estimated using PWV and modeled using NCEP atmospheric profiles with MODTRAN 4.0 code—can be seen in Tables 2–3. Overall, transmittance from Suominet PWV was overestimated compared to modeled values (Figure 9)—differences ranged from -0.01 to 0.10, excluding the two dates with variable PWV—with the differences increasing as PWV increased. Additionally, the distribution of the

PWV measurements from the six Suominet stations in the study area can be seen in Figure 10 (not all stations recorded measurements for all dates). As seen in the boxplots, PWV measurements for each date are relatively consistent across the study area, with the exception of September 1, 2005 and July 18, 2006 (indicated with black arrows). Actual stations and PWV values used to retrieve LST can be found in Tables 2–3.

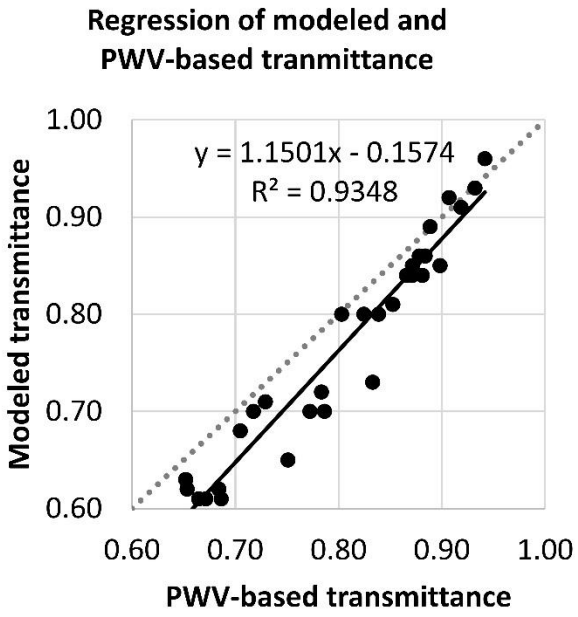

Figure 9. Regression of modeled and PWV-based transmittance. PWV-based transmittance is generally underestimated compared to modeled transmittance, with the degree of underestimation increasing as transmittance decreases.

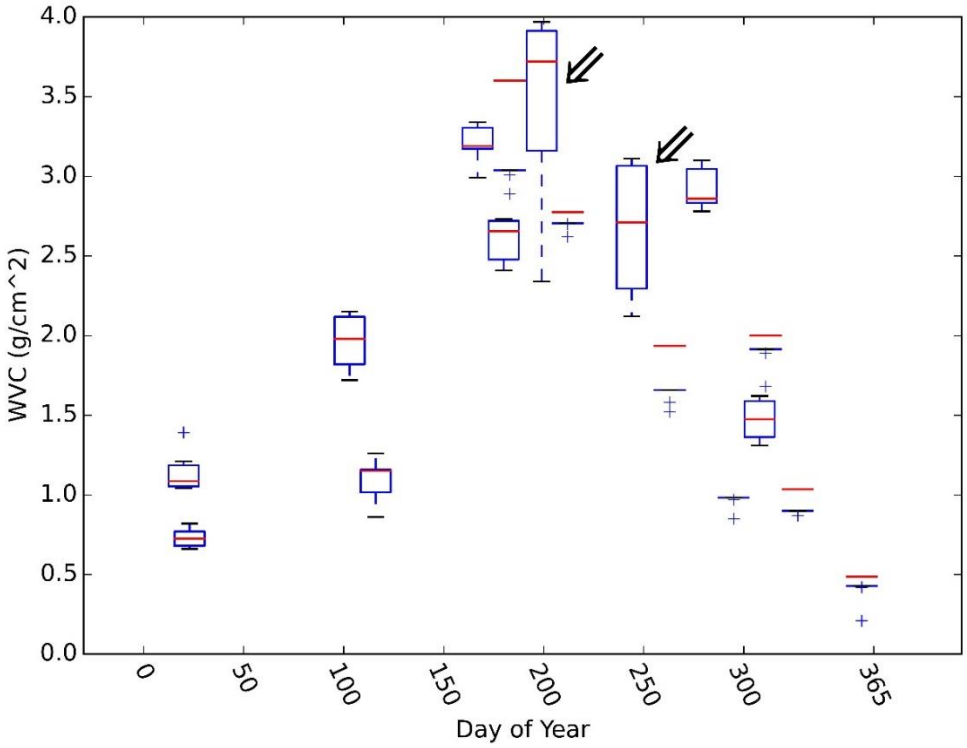

Figure 10. Distributions of Suominet PWV. PWV distributions across the study area for each date studied, by day of the year. While variability in PWV across the study area is generally low, September 1, 2005 and July 18, 2006 (identified with black arrows), with PWV ranges of approximately 1 g/cm^2, show that this is not always the case.

2.5.2. Land surface temperature validation

Two analyses were performed—one comparing skin temperature with all processed data, and a more conservative analysis utilizing only scenes that were visually confirmed to be free of cloud contamination. Regression results for all five methods using all data are listed in Table 4 and are illustrated in Figure 11; regression results using the more conservative, cloud-contamination-free data subset are listed in Table 5 and are illustrated in Figure 12. As expected, all atmospheric correction methods for all subsets in all analyses improved on LST results using no atmospheric correction, as evidenced by lower RMSEs and MAEs when compared with skin

temperature. Overall, LST retrieved using MWA with modeled transmittance and RTE outperformed methods using PWV-based transmittance.

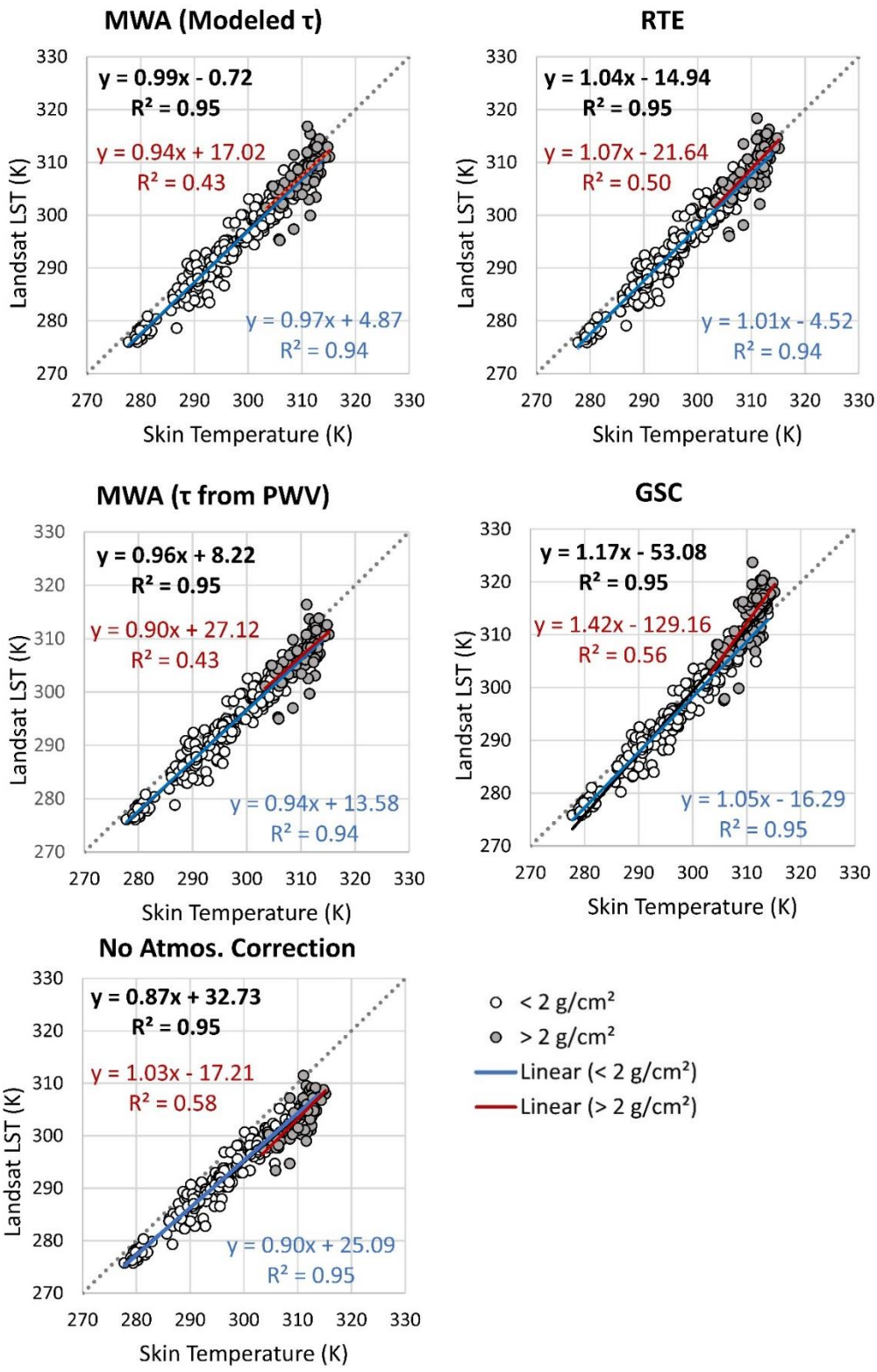

Figure 11. Regressions comparing LST results to skin temperature, using the full Landsat dataset. Regression lines and equations for all data; data from scenes below 2 g/cm^2 PWV; and data from scenes above 2 g/cm^2 PWV are shown in black, blue, and red, respectively.

Table 4: Summary of regression results using the full Landsat dataset. Results for all five methods generated are included for all data, and for data above and below the 2 g/cm^2 threshold. Included are model equations and errors as well as errors compared to skin temperature, which serves as ground-truth. We determined if the regression results were statistically significantly different than skin temperature; p-values greater than 0.05 suggest that the model slope or intercept is statistically indistinguishable from one or zero, respectively. Finally, we compared the slopes of regressions above and below the 2 g/cm^2 threshold; p-values above 0.05 suggest that the slopes are statistically indistinguishable.

Regression Models: LST as a function of skin temperature							Slope difference above and below PWV threshold (p-value)	Compared to 1-to-1 line (skin temperature)			
Correction method	Transmittance method	Est. scene PWV (g/cm^2)	Slope	Intercept	Adj. R^2	Model RMSE		RMSE from skin temp (K)	MAE from skin temp (K)	Slope different than 1 (p-value)	Intercept different than 0 (p-value)
MWA	Modeled	All	0.99	-0.73	0.95	2.34	0.697	3.67	3.05	0.59	0.85
MWA	Modeled	≤ 2	0.97	4.86	0.94	2.11		3.59	3.06	0.12	0.32
MWA	Modeled	> 2	0.94	17.02	0.42	2.75		3.82	3.03	0.57	0.62
MWA	from PWV	All	0.96	8.22	0.95	2.26	0.651	4	3.43	0.00252	0.0302
MWA	from PWV	≤ 2	0.94	13.59	0.94	2.06		3.84	3.33	0.000532	0.00486
MWA	from PWV	> 2	0.9	27.11	0.43	2.62		4.34	3.66	0.35	0.41
RTE	Modeled	All	1.04	-14.94	0.95	2.35	0.536	3.12	2.51	0.001073	0.000167
RTE	Modeled	≤ 2	1.01	-4.53	0.94	2.13		3.2	2.7	0.67	0.36
RTE	Modeled	> 2	1.07	-21.62	0.5	2.68		2.93	2.11	0.54	0.52
GSC	from PWV	All	1.17	-53.09	0.95	2.81	0.000242	3.37	2.71	<2e-16	<2e-16
GSC	from PWV	≤ 2	1.05	-16.3	0.95	2.09		2.92	2.37	0.003562	0.000935
GSC	from PWV	> 2	1.42	-129.22	0.56	3.17		4.19	3.46	0.0012	0.00147
None	from PWV	All	0.87	32.73	0.95	2.03	0.1028	5.6	5.08	<2e-16	<2e-16
None	from PWV	≤ 2	0.9	25.09	0.95	1.87		4.77	4.32	1.22E-10	2.66E-08
None	from PWV	> 2	1.03	-17.16	0.57	2.24		7.09	6.73	0.71	0.54

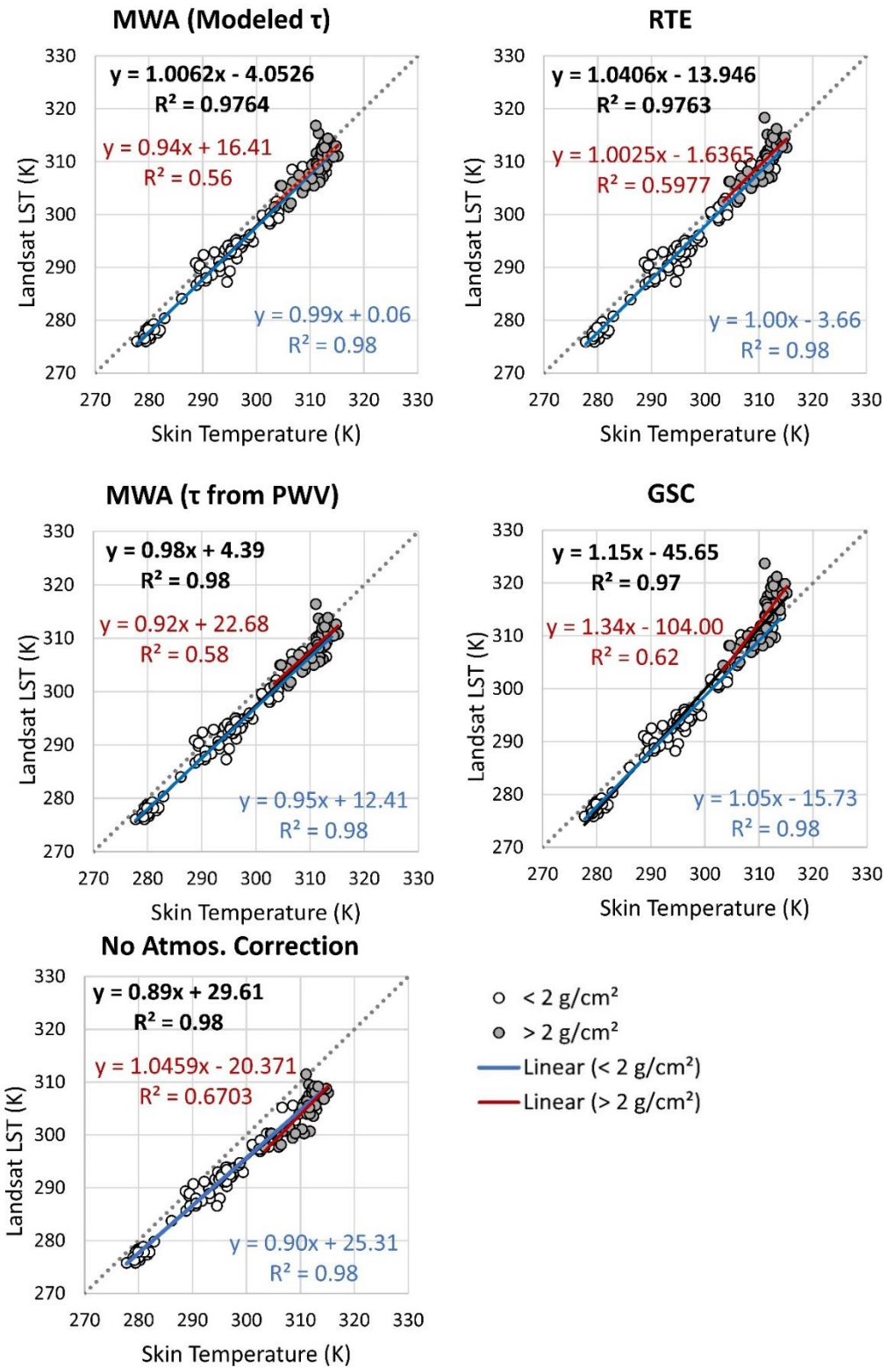

Figure 12. Regressions comparing LST results to skin temperature, using the cloud-free Landsat dataset. Regression lines and equations for all data, data from scenes below 2 g/cm^2 PWV, and data from scenes above 2 g/cm^2 PWV are shown in black, blue, and red, respectively.

Table 5. Summary of regression results using the cloud-free Landsat dataset. Results for all five methods generated are included for all data, and for data above and below the 2 g/cm^2 threshold. Included are model equations and errors as well as errors compared to skin temperature, which serves as ground-truth. We determined if the regression results were statistically significantly different than skin temperature; p-values greater than 0.05 suggest that the model slope or intercept is statistically indistinguishable from one or zero, respectively. Finally, we compared the slopes of regressions above and below the 2 g/cm^2 threshold; p-values above 0.05 suggest that the slopes are statistically indistinguishable.

Regression Models: LST as a function of skin temperature							Slope difference above and below PWV threshold (p-value)	Compared to 1-to-1 line (skin temperature)			
Correction method	Transmittance method	Est. scene PWV (g/cm^2)	Slope	Intercept	Adj. R^2	Model RMSE		RMSE from skin temp (K)	MAE from skin temp (K)	Slope different than 1 (p-value)	Intercept different than 0 (p-value)
MWA	Modeled	All	1.01	-4.06	0.98	1.84	0.6031	2.86	2.5	0.63	0.3
MWA	Modeled	≤ 2	0.99	0.05	0.98	1.53		2.79	2.5	0.6	0.99
MWA	Modeled	> 2	0.94	16.44	0.55	2.3		2.99	2.48	0.63	0.67
MWA	from PWV	All	0.98	4.4	0.98	1.8	0.752	3.19	2.84	0.07	0.25
MWA	from PWV	≤ 2	0.95	12.42	0.98	1.5		3.11	2.81	0.000906	0.00589
MWA	from PWV	> 2	0.92	22.71	0.57	2.18		3.34	2.89	0.49	0.53
RTE	Modeled	All	1.04	-13.94	0.98	1.91	0.9804	2.65	2.26	0.003344	0.000829
RTE	Modeled	≤ 2	1	-3.66	0.98	1.6		2.75	2.45	0.76	0.44
RTE	Modeled	> 2	1	-1.6	0.59	2.27		2.43	1.88	0.98	0.97
GSC	from PWV	All	1.15	-45.66	0.97	2.46	0.009862	3.05	2.44	1.03E-14	8.77E-15
GSC	from PWV	≤ 2	1.05	-15.73	0.98	1.56		2.35	1.99	0.002929	0.000937
GSC	from PWV	> 2	1.34	-104.05	0.61	2.9		4.08	3.32	0.0297	0.0338
None	from PWV	All	0.89	29.61	0.98	1.66	0.09737	5.02	4.57	< 2e-16	5.70E-14
None	from PWV	≤ 2	0.9	25.32	0.98	1.37		4.13	3.78	1.45E-10	1.09E-08
None	from PWV	> 2	1.05	-20.33	0.66	2.03		6.42	6.1	0.67	0.54

2.5.2.1. All LST data

Model RMSEs for all atmospheric correction methods were comparable, though slightly higher for GSC overall and slightly lower with no atmospheric correction (Figure 11, Table 4).

2.5.2.1.1. All LST data: Full dataset

For all data in the full dataset, MWA using modeled transmittance performed the best—regression analysis suggests that LST retrieved with this method was overall statistically indistinguishable from skin temperature; no other method achieved this result for all data. RMSE compared to skin temperature using this method was 3.67 K, and MAE was 3.05 K (Table 4, 1:1 line). MWA using PWV-based transmittance underestimated skin temperature overall, with the degree of underestimation increasing as both skin temperature and PWV increase (Figure 11). This method also resulted in the highest errors of all correction methods, with a 4.00 K RMSE and a 3.43 MAE (Table 4, 1:1 line). High regression coefficients suggest that GSC, which relies heavily on PWV for LST extraction, was the worst at generating a one-to-one relationship with skin temperature (Table 4, Regression Models). Interestingly, RMSE and MAE using GSC were slightly lower than both MWA datasets; however, with a slope of 1.17 and an intercept of -53.09, regression analysis suggests that this is because the model crosses the one-to-one line, rather than because GSC does a better job of accurately predicting LST in all atmospheric and climatic contexts (Table 4, Regression Models; Figure 11, GSC). The very high slope means that GSC underestimates LST at low skin temperatures, and overestimates LST at high temperatures. Because the relationship between skin temperature and retrieved LST changes as atmospheric and climatic

conditions change, this method produces results that may not be intercomparable when retrieved under different conditions.

2.5.2.1.2. All LST data: Low-PWV-subset

Both methods using modeled transmittance, MWA and RTE, provided results that were statistically indistinguishable from skin temperature measurements for data under the 2 g/cm^2 PWV threshold. RTE did a marginally better job predicting skin temperature for dates with low PWV, with higher p-values for model slope and intercept and lower RMSE and MAE than MWA with modeled transmittance (3.20 K and 2.70 K compared to 3.59 K and 3.06 K, respectively) (Table 4, 1-to-1 line). As with the full dataset, GSC had the lowest errors of all methods—RMSE of 2.92 and MAE of 2.70—but the model coefficients suggest that this method does not produce LST results equivalent with skin temperature. MWA using transmittance from PWV performed the worst of all correction methods, with the highest RMSE and MAE (3.84 K and 3.33 K, respectively), and regression coefficients indicating the results were not equivalent to skin temperature (Table 4).

2.5.2.1.3. All LST data: High-PWV subset

Adjusted r-squared values ranged from 0.42 to 0.57 for all methods, relatively low compared to regressions using all data or only data below the PWV threshold (Table 4, Regression Models). Because the dataset being analyzed is smaller, and the temperature and PWV ranges are relatively smaller, only LST retrieved using GSC was determined to be statistically different from skin temperature. ANCOVA analysis also suggested that the high-PWV and low-PWV regressions using GSC were statistically significantly different from each other; this could not be determined for any of the other

methods (Table 4, Slope difference). Error rates using GSC were also relatively high; RMSE was 4.19 K and MAE was 3.46K (Table 4, 1-to-1 line). Based on RMSE and MAE, RTE outperformed MWA with modeled transmittance—both measurements of error were almost 1 K lower with RTE. MWA with PWV-based transmittance had the highest errors of all correction methods—4.34 K RMSE and 3.66 K MAE.

5.2.2. Cloud-free subset

Results from the conservative analysis, which used data from scenes with no likely cloud contamination based on visual inspection, were similar to the analysis incorporating all data. Overall, model errors for the cloud-free subset are lower than when all data are included, with RMSE ranging from 1.37 to 2.9 K for all conservative regressions (Table 5), compared to 1.87 to 3.17K for regressions with all data (Table 4).

5.2.2.1. Cloud-free subset: Full dataset

As with the analysis of all LST data, MWA using modeled transmittance again performed best for all data being analyzed, with regression coefficients statistically indistinguishable from a one-to-one line and error rates that were second-lowest (RMSE of 2.86 K and MAE of 2.50 K) (Table 5, 1-to-1 line). In this case, MWA using PWV-based transmittance also produced LST results that were statistically indistinguishable from skin temperature, though error rates were the highest of all correction methods (RMSE of 3.19 K and MAE of 2.84 K). RTE produced the lowest errors—RMSE of 2.65 and MAE of 2.26—but the overall regression suggested that LST retrieved using RTE was statistically different from skin temperature measurements, as was LST retrieved using GSC.

5.2.2.2. Cloud-free subset: Low-PWV subset

Regressions showed that, for scenes with PWV under the 2 g/cm^2 threshold, only LST datasets retrieved using RTE and MWA with modeled transmittance were statistically indistinguishable from skin temperature (Table 5). Both had had similar errors—2.75 K RMSE and 2.45 K MAE with RTE, and 2.79 K RMSE and 2.50 K MAE with MWA. Again, GSC had the lowest errors—2.35 K RMSE and 1.99 K MAE—but the regression coefficients indicated that the results produced were not statistically comparable to skin temperature, even when only cloud-free data below the PWV threshold were included (Table 5, 1-to-1 line).

5.2.2.3. Cloud-free subset: High-PWV subset

R-squared values for regressions with cloud-free data above the 2 g/cm^2 threshold were higher in the conservative analysis (Table 5) than when all high-PWV data were included (Table 4), suggesting that cloud-free LST datasets better accounted for the variability in the skin temperature data. Again, for the high-PWV subset, only LST retrieved using GSC was determined to be statistically different from skin temperature; all other methods could not be distinguished statistically from skin temperature. Similarly, ANCOVA analysis suggested that the high-PWV and low-PWV regressions using GSC were statistically significantly different from each other (the null hypothesis that the high- and low-PWV regressions were the same could not be rejected for any of the other methods) (Table 5, Slope difference). Compared to skin temperature, error rates with GSC were only slightly lower than with all data: RMSE was lowered to 4.08 K and MAE to 3.32 K. Though the regressions could not be statistically distinguished from a one-to-one relationship with skin temperature, RTE

came very close to an actual one-to-one relationship, with a slope of 1.00, an intercept of -1.60, and p-values compared to 1 and 0 of 0.98 and 0.97, respectively (Table 5). For high- PWV data, RTE also provided the lowest errors compared to skin temperature of all methods, with RMSE of 2.43 K and MAE of 1.88 K. And while most of the methods did not have statistical evidence that the slopes above and below the 2 g/cm^2 threshold were different, a very high p-value for RTE ANCOVA analysis—0.98—suggests that regression slopes above and below the PWV threshold are especially similar.

2.6. Discussion

2.6.1. Method selection

Overall, MWA using modeled transmittance provided the best results, with regression analysis suggesting that the full LST dataset and all subsets were statistically indistinguishable from skin temperature. This method also was the most consistent across all datasets, with the relationship between skin temperature and LST remaining relatively constant as skin temperature and PWV increased. Overall errors were high, however, with RMSE relative to skin temperature remaining at 2.86 K even with the cloud-free subset. Some of this error is attributable to problems inherent when comparing point data to 120m raster data. However, it suggests that the bar for statistically significant change detection is fairly high.

While MWA using modeled transmittance performed best overall, RTE arguably performed well enough to be useful in some contexts. For data with atmospheric PWV below 2 g/cm^2, this method is as or more accurate than MWA. Additionally, regression analysis using the cloud-free dataset suggests that it may also be effective for data with higher PWV content with good quality control. And while the

regressions for both the full LST dataset and the full cloud-free subsets suggest that LST retrieved using this method is not statistically equivalent to skin temperature, ANCOVA analysis cannot confirm that the regressions above and below the PWV threshold have different slopes. Error rates using this method were also consistently lower than with MWA using modeled transmittance, dropping to 2.65 K RMSE for the full, cloud-free subset. Additionally, because this method does not require PWV content as input, it is logistically much simpler to implement than MWA with modeled transmittance, which requires both modeled and PWV-based input parameters.

Both methods that relied on PWV-based transmittance were less accurate than the methods with modeled transmittance, suggesting that basing transmittance values for an entire scene on one PWV measurement is not the most effective option. GSC in particular performed poorly, with the relationship between skin temperature and LST retrieved using GSC changing dramatically as both temperature and PWV increased. GSC also failed to produce a regression equation that was statistically equivalent to a one-to-one relationship with skin temperature even when data was limited to scenes with PWV less than 2 g/cm^2. However, further analysis shows that the GSC method can produce results statistically indistinguishable from skin temperature measurements under a 1.6 g/cm^2 PWV threshold, with p-values for slope (compared to one) and intercept (compared to zero) of 0.36 and 0.20, respectively. It follows that this method may be appropriate for study areas and periods which are not likely to have PWV above 1.6 g/cm^2. This result is similar to that found by (Jimenez-Munoz et al. (2009)) showing that both error rates and bias with GSC increase as PWV approaches 2.0 g/cm^2.

2.6.2. Precipitable water vapor distribution

Two issues with PWV arise with Landsat LST retrieval methods as employed here—first, is the PWV measurement used for atmospheric correction representative of the entire scene? Second, can that PWV measurement be used to accurately atmospherically correct the Landsat thermal data? Based on Suominet PWV measurements, the spatial variability of PWV is generally small across our study area, but not always. When the variability in PWV across the study area is greater than 1 g/cm^2, for instance, the choice of one value over another would have a significant impact on transmittance and thus LST retrieval. Further, for at least some dates, the variation in PWV across the study area probably means one transmittance value per scene may not accurately reflect the spatial variability of atmospheric conditions, suggesting that monitoring the spatial variability of this parameter may be an important initial quality control measure for LST retrieval from Landsat.

Regarding accuracy, it is clear that error rates increase and the effectiveness of different retrieval methods decreases as PWV rises above 2.0 g/cm^2. The relevant question, then, is where and when will PWV content be high enough to introduce additional error to Landsat LST retrieval, and make the choice of retrieval method relevant? Mean 8-day maximum global PWV from MODIS can be seen in Figure 13, with values ranging from essentially zero to 6.78 g/cm^2. These estimates are probably slightly inflated due to the aggregation method, but still provide useful global context. The fraction of all observations that were above the 2 g/cm^2 threshold for the entire dataset and for seasonal subsets are in Figures 14 and 15; above 3 g/cm^2 for summer in Figure 16. The takeaway is that global PWV content varies quite a bit seasonally and

spatially, and is often above the 2 g/cm^2 threshold, impacting both expected error rates as well as the appropriate retrieval method. The United States, for example, has PWV below the 2 g/cm^2 threshold during winter (with the exception of the Gulf Coast), but PWV increases throughout the south and central U.S. in spring and fall, with summer values across most of the U.S. peaking above 2 g/cm^2 (and in many cases above 3 g/cm^2). This suggests that error rates for LST retrieval in summer will be higher than in winter, and that retrieval methods that rely on PWV content for atmospheric correction may not be appropriate for much of the year. Further, it also suggests that even when using an LST-retrieval method that does not require PWV measurements, like RTE, it is important to gather ancillary PWV data for quality control and error estimation purposes.

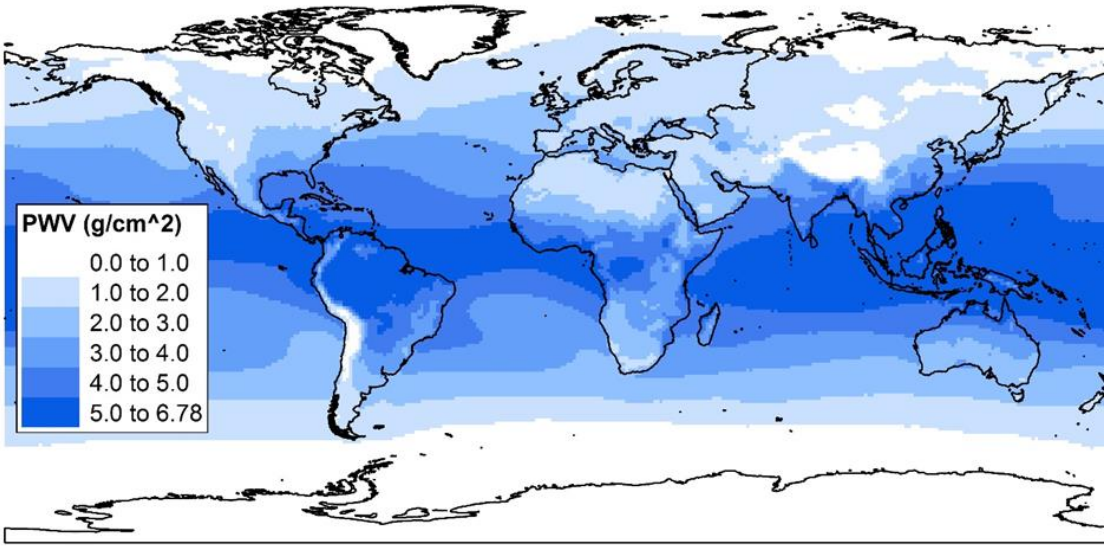

Figure 13. Mean global PWV distribution (g/cm^2). Large swaths of the globe average greater than 2 g/cm^2 PWV annually, including much of the southeastern United States.

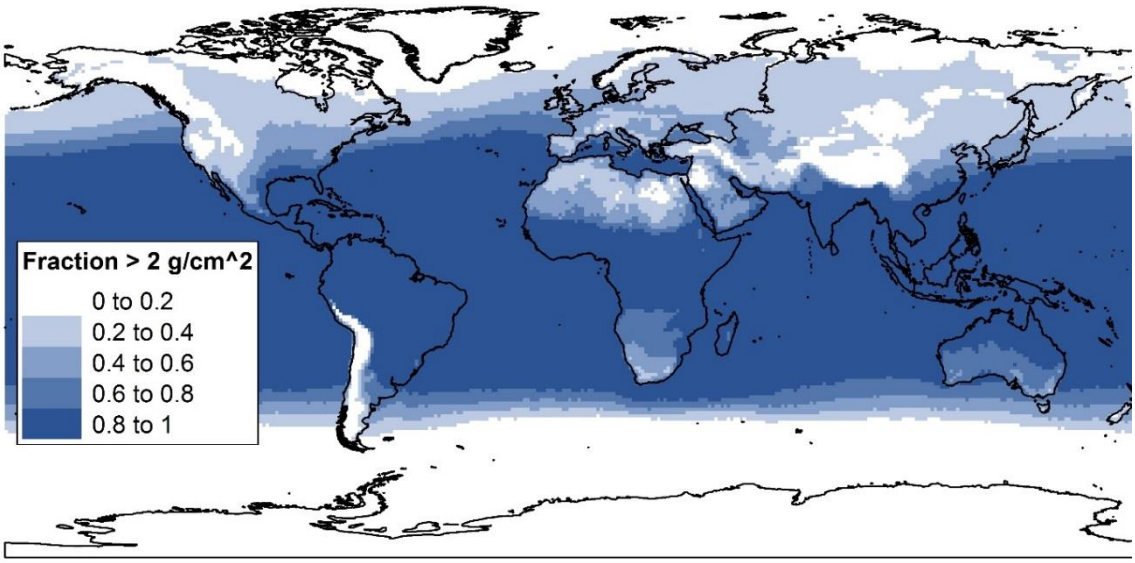

Figure 14. Fraction of the annual global PWV data greater than 2 g/cm^2. Much of the globe is subject to atmospheric conditions with more than 2 g/cm^2 PWV for at least part of the year.

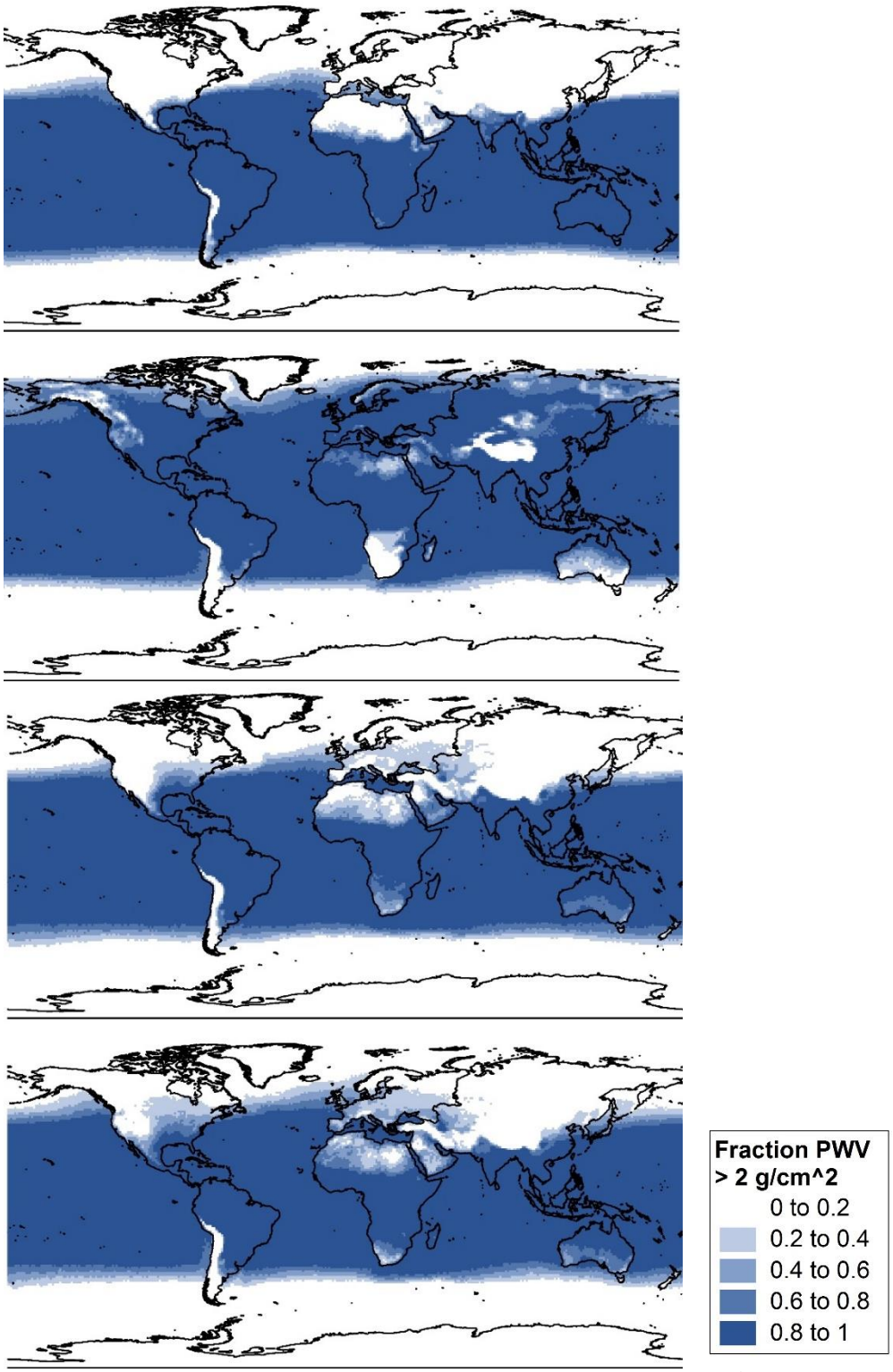

Figure 15. Fraction of the global PWV data greater than 2 g/cm^2, by season. Top to bottom: winter, summer; spring, fall. While much of the northern hemisphere is below the 2 g/cm^2 threshold during winter months, with only 0–20% of data above the threshold, most of the globe's land cover is above this threshold at least 80% of the time during summer months.

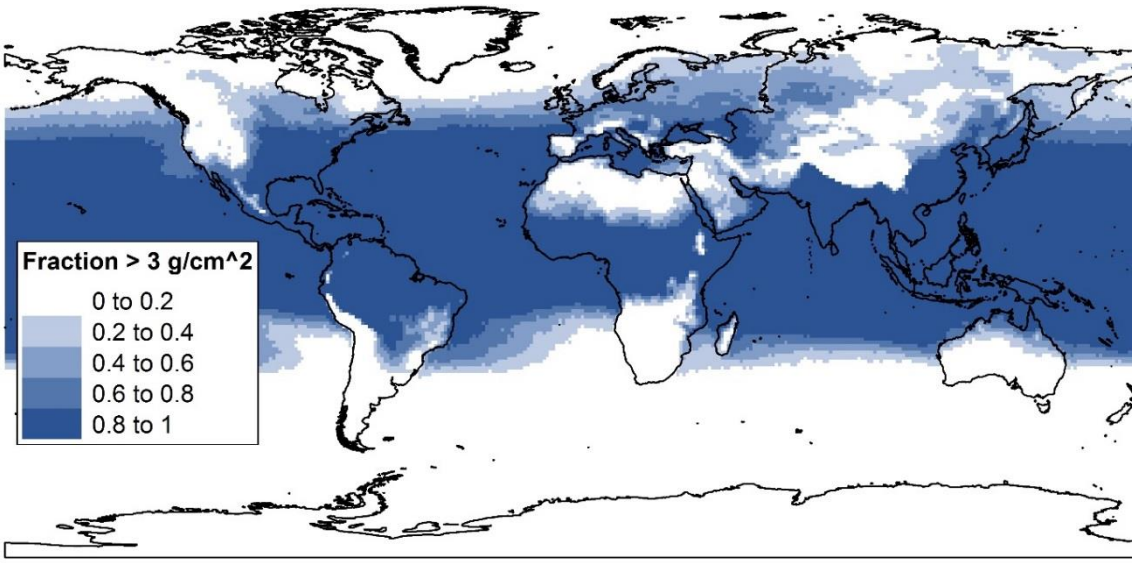

Figure 16. Fraction of the global PWV data greater than 3 g/cm^2 in summer months. Many regions, including the southeastern United States, India, and Southeast Asia, report PWV greater than 3 g/cm^2 at least 80% of the time during summer months.

2.6.3 Cloud contamination

Detecting cloud contamination, especially cirrus clouds, remains a concern with Landsat 4–7 data. Results here suggest that cloud contamination (likely cirrus clouds) not reported in the Landsat metadata and not masked by Fmask does introduce error into LST retrieval. At least some of this contamination can be detected visually in the thermal data, but it is difficult to mask or automate detection. Further, a manual inspection of data points occurring in these contaminated areas indicates that the contamination may only decrease LST by 5–10 K below skin temperature measurements; this has a material effect on LST retrieval, but may be difficult to detect statistically. In this analysis, limiting regression to scenes visually confirmed to be cloud-free resulted in a mean decrease in RMSE of 0.62 K and in MAE of 0.42 K for all atmospherically corrected methods and subsets when compared to skin temperature. While this decrease is not necessarily due to the elimination of cloud contamination

alone, it does seem reasonable to conclude that undetected clouds increase error in LST retrieval from Landsat 4–7.

2.7 Conclusion

Given the spatial and intra-annual variation in PWV seen in Figures 10–13, an LST-retrieval method that is not only accurate, but also maintains a uniform relationship with actual skin temperature even as temperatures and PWV in the atmosphere increase, is essential for producing results that can be compared across time and space. LST retrieval methods that used PWV-based transmittance struggled to accomplish this; modeled transmittance rates provided more robust results. Additionally, from this perspective the mono-window algorithm was the most effective overall, though the radiative transfer equation as implemented here provided results that were almost as good, and is more easily employed. Caveats remain, however. Undetected cloud contamination and spatially variable PWV remain of concern, and suggest that quality control measures targeting these problems should be utilized when retrieving LST from Landsat. Finally, though emissivity estimation is not addressed here, it remains an important issue for accurate LST retrieval, and will likely increase error beyond what is quantified here. This research provides more accurate error assessment for applied Landsat LST retrieval methods that will aid in change assessment through both time and space, and will allow users to better strike a balance between accuracy and ease of implementation for various applications.

References

Asgarian, A., Amiri, B., Sakieh, Y., 2015. Assessing the effect of green cover spatial patterns on urban land surface temperature using landscape metrics approach. *Urban Ecosystems* 18, 209-222.

Barsi, J.A., Barker, J.L., Schott, J.R., 2003. An Atmospheric Correction Parameter Calculator for a single thermal band earth-sensing instrument, Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International, pp. 3014-3016 vol.3015.

Barsi, J.A., Schott, J.R., Palluconi, F.D., Hook, S.J., 2005. Validation of a web-based atmospheric correction tool for single thermal band instruments, pp. 58820E-58820E-58827.

Buyantuyev, A., Wu, J., 2010. Urban heat islands and landscape heterogeneity: linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. *Landscape Ecology* 25, 17-33.

Chander, G., Markham, B.L., Helder, D.L., 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment* 113, 893-903.

Coll, C., Wan, Z., Galve, J.M., 2009. Temperature-based and radiance-based validations of the V5 MODIS land surface temperature product. *Journal of Geophysical Research: Atmospheres* 114.

Costa, G.C., Wolfe, C., Shepard, D.B., Caldwell, J.P., Vitt, L.J., 2008. Detecting the influence of climatic variables on species distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient. *Journal of Biogeography* 35, 637-646.

Dash, P., Göttsche, F.M., Olesen, F.S., Fischer, H., 2002. Land surface temperature and emissivity estimation from passive sensor data: Theory and practice-current trends. *International Journal of Remote Sensing* 23, 2563-2594.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.H., Féménias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., Nieke, J., Rebhan, H., Seitz, B., Stroede, J., Sciarra, R., 2012. The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. *Remote Sensing of Environment* 120, 37-57.

Fu, P., Weng, Q., 2016. A time series analysis of urbanization induced land use and land cover change and its impact on land surface temperature with Landsat imagery. *Remote Sensing of Environment* 175, 205-214.

Gampe, D., Ludwig, R., Qahman, K., Afifi, S., 2016. Applying the Triangle Method for the parameterization of irrigated areas as input for spatially distributed hydrological modeling — Assessing future drought risk in the Gaza Strip (Palestine). *Science of The Total Environment* 543, Part B, 877-888.

Ghaleb, F., Mario, M., Sandra, A., 2015. Regional Landsat-Based Drought Monitoring from 1982 to 2014. *Climate* 3, 563-577.

Gober, P., Brazel, A., Quay, R., Myint, S., Grossman-Clarke, S., Miller, A., Rossi, S., 2010. Using Watered Landscapes to Manipulate Urban Heat Island Effects: How Much Water Will It Take to Cool Phoenix? *Journal of the American Planning Association* 76, 109-121.

González, C., Inostroza, M., Aguilera, F., González, R., Viramonte, J., Menzies, A., 2015. Heat and mass flux measurements using Landsat images from the 2000–2004 period, Lascar volcano, northern Chile. *Journal of Volcanology and Geothermal Research* 301, 277-292.

Hakuba, M.Z., Folini, D., Sanchez-Lorenzo, A., Wild, M., 2014. Spatial representativeness of ground-based solar radiation measurements—Extension to the full Meteosat disk. *Journal of Geophysical Research: Atmospheres* 119, 11,760-711,771.

Hongyuan, H., Xiaoguang, J., Xianfeng, S., Zhao-Liang, L., Zhuoya, N., Caixia, G., 2014. Detection of Coal Fire Dynamics and Propagation Direction from Multi-Temporal Nighttime Landsat SWIR and TIR Data: A Case Study on the Rujigou Coalfield, Northwest (NW) China. *Remote Sensing* 6, 1234-1259.

Huang, C.Y., Anderegg, W.R.L., 2014. Vegetation, land surface brightness, and temperature dynamics after aspen forest die-off. *Journal of Geophysical Research: Biogeosciences* 119, 1297-1308.

Irons, J.R., Dwyer, J.L., Barsi, J.A., 2012. The next Landsat satellite: The Landsat Data Continuity Mission. *Remote Sensing of Environment* 122, 11-21.

Jimenez-Munoz, J.C., Cristobal, J., Sobrino, J.A., Soria, G., Ninyerola, M., Pons, X., Pons, X., 2009. Revision of the Single-Channel Algorithm for Land Surface Temperature Retrieval From Landsat Thermal-Infrared Data. *Geoscience and Remote Sensing, IEEE Transactions on* 47, 339-349.

Jimenez-Munoz, J.C., Sobrino, J.A., Mattar, C., Hulley, G., Gottsche, F.-M., 2014a. Temperature and Emissivity Separation From MSG/SEVIRI Data. *Geoscience and Remote Sensing, IEEE Transactions on* 52, 5937-5951.

Jimenez-Munoz, J.C., Sobrino, J.A., Skokovic, D., Mattar, C., Cristobal, J., 2014b. Land Surface Temperature Retrieval Methods From Landsat-8 Thermal Infrared Sensor Data. *Geoscience and Remote Sensing Letters, IEEE* 11, 1840-1843.

Kestens, Y., Brand, A., Fournier, M., Goudreau, S., Kosatsky, T., Maloley, M., Smargiassi, A., 2011. Modelling the variation of land surface temperature as determinant of risk of heat-related health events.(Research)(Report). *International Journal of Health Geographics* 10, 7.

Kuenzer, C., Guo, H., Ottinger, M., Zhang, J., Dech, S., 2013. Spaceborne Thermal Infrared Observation – An Overview of Most Frequently Used Sensors for Applied Research, in: Kuenzer, C., Dech, S. (Eds.), *Thermal Infrared Remote Sensing: Sensors, Methods, Applications*. Springer Netherlands, Dordrecht, pp. 131-148.

Li, X., Li, W., Middel, A., Harlan, S., Brazel, A., Turner, B., 2016. Remote sensing of the surface urban heat island and land architecture in Phoenix, Arizona: Combined

effects of land composition and configuration and cadastral-demographic-economic factors. *Remote Sensing of Environment* 174, 233-243.

Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., Sobrino, J.A., 2013. Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment* 131, 14-37.

Lu, D., Song, K., Zang, S., Jia, M., Du, J., Ren, C., 2015. The Effect of Urban Expansion on Urban Surface Temperature in Shenyang, China: an Analysis with Landsat Imagery. *Environ Model Assess* 20, 197-210.

Maimaitiyiming, M., Ghulam, A., Tiyip, T., Pla, F., Latorre-Carmona, P., Halik, Ü., Sawut, M., Caetano, M., 2014. Effects of green space spatial pattern on land surface temperature: Implications for sustainable urban planning and climate change adaptation. *ISPRS Journal of Photogrammetry and Remote Sensing* 89, 59-66.

Mia, M., Nishijima, J., Fujimitsu, Y., 2015. Monitoring heat flow before and after eruption of Kuju fumaroles in 1995 using Landsat TIR images. *Acta Geodaetica et Geophysica* 50, 295-305.

Mia, M.B., Fujimitsu, Y., 2013. Landsat Thermal Infrared based Monitoring of Heat Losses from Kuju Fumaroles Area in Japan. *Procedia Earth and Planetary Science* 6, 114-120.

Montanaro, M., Gerace, A., Lunsford, A., Reuter, D., 2014. Stray Light Artifacts in Imagery from the Landsat 8 Thermal Infrared Sensor. *Remote Sensing* 6, 10435-10456.

Orhan, O., Ekercin, S., Dadaser-Celik, F., 2014. Use of Landsat Land Surface Temperature and Vegetation Indices for Monitoring Drought in the Salt Lake Basin Area, Turkey. *The Scientific World Journal* 2014.

Ottle, C., Stoll, M., 1993. Effect of atmospheric absorption and surface emissivity on the determination of land surface temperature from infrared satellite data. *International Journal of Remote Sensing* 14, 2025-2037.

Palluconi, F.D., 1996. Validation of the ASTER thermal infrared surface radiance data product, *Proc. SPIE 2820, Earth Observing System*, pp. 97-104.

Prata, A.J., Caselles, V., Coll, C., Sobrino, J.A., Ottlé, C., 1995. Thermal remote sensing of land surface temperature from satellites: Current status and future prospects. *Remote Sensing Reviews* 12, 175-224.

Qin, Z., Karnieli, A., Berliner, P., 2001. A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel-Egypt border region. *International Journal of Remote Sensing* 22, 3719-3746.

Quintano, C., Fernández-Manso, A., Calvo, L., Marcos, E., Valbuena, L., 2015. Land surface temperature as potential indicator of burn severity in forest Mediterranean ecosystems. *International Journal of Applied Earth Observations and Geoinformation* 36, 1-12.

Ramdani, F., Moffiet, T., Hino, M., 2014. Local surface temperature change due to expansion of oil palm plantation in Indonesia. *Climatic Change* 123, 189-200.

Rogan, J., Ziemer, M., Martin, D., Ratick, S., Cuba, N., DeLauer, V., 2013. The impact of tree cover loss on land surface temperature: A case study of central Massachusetts using Landsat Thematic Mapper thermal data. *Applied Geography* 45, 49-57.

Román, M.O., Schaaf, C.B., Woodcock, C.E., Strahler, A.H., Yang, X., Braswell, R.H., Curtis, P.S., Davis, K.J., Dragoni, D., Goulden, M.L., Gu, L., Hollinger, D.Y., Kolb, T.E., Meyers, T.P., Munger, J.W., Privette, J.L., Richardson, A.D., Wilson, T.B., Wofsy, S.C., 2009. The MODIS (Collection V005) BRDF/albedo product: Assessment of spatial representativeness over forested landscapes. *Remote Sensing of Environment* 113, 2476-2498.

Roy, P., Guha, A., Kumar, K.V., 2015. An approach of surface coal fire detection from ASTER and Landsat-8 thermal data: Jharia coal field, India. *International Journal of Applied Earth Observations and Geoinformation* 39, 120-127.

Salisbury, J.W., D'Aria, D.M., 1992. Emissivity of terrestrial materials in the 8–14 μm atmospheric window. *Remote Sensing of Environment* 42, 83-106.

Sobrino, J.A., Jiménez-Muñoz, J.C., Paolini, L., 2004. Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of Environment* 90, 434-440.

Sobrino, J.A., Oltra-Carrió, R., Sòria, G., Bianchi, R., Paganini, M., 2012. Impact of spatial resolution and satellite overpass time on evaluation of the surface urban heat island effects. *Remote Sensing of Environment* 117, 50-56.

Stone, B., Vargo, J., Liu, P., Hu, Y., Russell, A., 2013. Climate Change Adaptation Through Urban Heat Management in Atlanta, Georgia. *Environmental Science & Technology* 47, 7780-7786.

Tian, B., Wang, L., Kashiwaya, K., Koike, K., 2015. Combination of Well-Logging Temperature and Thermal Remote Sensing for Characterization of Geothermal Resources in Hokkaido, Northern Japan. *Remote Sensing* 7, 2647-2667.

Vlassova, L., Perez-Cabello, F., Mimbrero, M., Lloveria, R., Garcia-Martin, A., 2014. Analysis of the Relationship between Land Surface Temperature and Wildfire Severity in a Series of Landsat Images. *Remote Sensing* 6, 6136-6162.

Walsh-Thomas, J.M., Cervone, G., Agouris, P., Manca, G., 2012. Further evidence of impacts of large-scale wind farms on land surface temperature. *Renewable and Sustainable Energy Reviews* 16, 6432-6437.

Wan, Z., 2014. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment* 140, 36-45.

Wang, F., Qin, Z., Song, C., Tu, L., Karnieli, A., Zhao, S., 2015. An Improved Mono-Window Algorithm for Land Surface Temperature Retrieval from Landsat 8 Thermal Infrared Sensor Data. *Remote Sensing* 7, 4268-4289.

Wark, D.Q., Yamamoto, G., Lienesch, J.H., 1962. Methods of Estimating Infrared Flux and Surface Temperature from Meteorological Satellites. *Journal of the Atmospheric Sciences* 19, 369-384.

White-Newsome, J.L., Brines, S.J., Brown, D.G., Dvonch, J.T., Gronlund, C.J., Zhang, K., Oswald, E.M., O'neill, M.S., 2013. Validating Satellite-Derived Land Surface Temperature with in Situ Measurements: A Public Health Perspective. *Environmental Health Perspectives* 121, 925-931.

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment* 122, 2-10.

Yan, W., Mahendrarajah, P., Shaker, A., Faisal, K., Luong, R., Al-Ahmad, M., 2014. Analysis of multi-temporal landsat satellite images for monitoring land surface temperature of municipal solid waste disposal sites. *Environmental Monitoring and Assessment* 186, 8161-8173.

Zhang, Y., Odeh, I.O.A., Ramadan, E., 2013. Assessment of land surface temperature in relation to landscape metrics and fractional vegetation cover in an urban/peri-urban region using Landsat data. *International Journal of Remote Sensing* 34, 168-189.

Zhengming, W., Dozier, J., 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. *Geoscience and Remote Sensing, IEEE Transactions on* 34, 892-905.

Zhou, J., Li, J., Zhang, L.X., Hu, D.Y., Zhan, W.F., 2012. Intercomparison of methods for estimating land surface temperature from a Landsat-5 TM image in an arid region with low water vapour in the atmosphere. *International Journal of Remote Sensing* 33, 2582-2602.

Zhou, W., Huang, G., Cadenasso, M.L., 2011. Does spatial configuration matter? Understanding the effects of land cover pattern on land surface temperature in urban landscapes. *Landscape and Urban Planning* 102, 54-63.