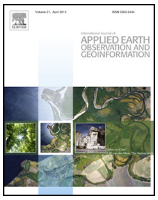
International Journal of Applied Earth Observation and Geoinformation 51 (2016) 11–27

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



An intercomparison of Landsat land surface temperature retrieval methods under variable atmospheric conditions using in situ skin temperature

Emily Windahl*, Kirsten de Beurs

Department of Geography and Environmental Sustainability, University of Oklahoma, 100 East Boyd St. SEC Suite 510, Norman, OK 73019, United States

ARTICLE INFO

Article history:
Received 28 January 2016
Received in revised form 14 April 2016
Accepted 18 April 2016
Available online 3 May 2016

Keywords:
Land surface temperature
Precipitable water vapor
Landsat
Thermal remote sensing
Atmospheric correction
Radiative transfer equation

ABSTRACT

Land surface temperature 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 land surface temperature 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, as well as LST with no atmospheric correction, and validates the results against on-the-ground skin temperature measurements from twenty-five Oklahoma Mesonet stations. Additionally, LST results using both modeled transmittance values and transmittance values 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.

1. Introduction

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 Landsats 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

---

* Corresponding author.
E-mail addresses: ewindahl@ou.edu, ewindahl@gmail.com (E. Windahl), kdebeurs@ou.edu (K.d. Beurs).

http://dx.doi.org/10.1016/j.jag.2016.04.003
0303-2434/

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

Oklahoma Annual Temperature & Precipitation, 2005 & 2006					
	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

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 Landsats 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 an RMSE of 0.9 K, and an RMSE of 1.9 K with PWV-based transmittance; the GSC method achieved an 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 to 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 mis-