Remote Sensing of Environment 122 (2012) 41–49

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

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

Thermal infrared radiometric calibration of the entire Landsat 4, 5, and 7 archive (1982–2010)

John R. Schott [a,*], Simon J. Hook [b], Julia A. Barsi [c], Brian L. Markham [d], Jonathan Miller [e], Francis P. Padula [f], Nina G. Raqueno [a]

[a] Rochester Institute of Technology, United States
[b] Jet Propulsion Lab, California Institute of Technology, NASA, United States
[c] Science Systems and Applications, Inc., NASA, United States
[d] GFSC, NASA, United States
[e] USAF, United States
[f] Science Applications International Corporation, United States

ARTICLE INFO

Article history:
Received 11 March 2011
Received in revised form 20 June 2011
Accepted 18 July 2011
Available online 17 February 2012

Keywords:
Infrared
Radiometric calibration
Landsat
Residual uncertainty

ABSTRACT

Landsat's continuing record of the thermal state of the earth's surface represents the only long term (1982 to the present) global record with spatial scales appropriate for human scale studies (i.e., tens of meters). Temperature drives many of the physical and biological processes that impact the global and local environment. As our knowledge of, and interest in, the role of temperature on these processes have grown, the value of Landsat data to monitor trends and process has also grown. The value of the Landsat thermal data archive will continue to grow as we develop more effective ways to study the long term processes and trends affecting the planet. However, in order to take proper advantage of the thermal data, we need to be able to convert the data to surface temperatures. A critical step in this process is to have the entire archive completely and consistently calibrated into absolute radiance so that it can be atmospherically compensated to surface leaving radiance and then to surface radiometric temperature. This paper addresses the methods and procedures that have been used to perform the radiometric calibration of the earliest sizable thermal data set in the archive (Landsat 4 data). The completion of this effort along with the updated calibration of the earlier (1985–1999) Landsat 5 data, also reported here, concludes a comprehensive calibration of the Landsat thermal archive of data from 1982 to the present.

1. Introduction and summary

Landsat's continuing record of the thermal state of the earth's surface represents the only long term (1982 to the present) global record with spatial scales appropriate for human scale studies (i.e., tens of meters). Note, AVHRR data span from 1978 to the present and Modis data span from 2000 to the present with 1 km pixels. Temperature drives many of the physical and biological processes that impact the global and local environment. As our knowledge of, and interest in, the role of temperature on these processes has grown, the value of Landsat data to monitor trends and process has also grown. Areas of study that use Landsat derived thermal data include lake hydrodynamic process (Schott, 1986; Schott et al., 2001), monitoring evapotranspiration (Allen et al., 2008; Anderson & Kustas, 2008), regional water resources (Thenkabail et al., 2009, 2010), and the impact of local climate trends (Schneider et al., 2009). The value of the Landsat thermal data archive will continue to grow as we develop more effective ways to study the long term processes and trends affecting the planet. However, in order to take proper advantage of the thermal data, we need to be able to convert the data to surface temperatures. A critical step in this process is to have the entire archive completely and consistently calibrated into absolute radiance so that it can be atmospherically compensated to surface leaving radiance and then to temperature. This paper addresses the methods and procedures that have been used to perform the radiometric calibration of the earliest sizable thermal data set in the archive (Landsat 4 data). The completion of this effort along with the updated calibration of the earlier (1985–1999) Landsat 5 data, also reported here, concludes a comprehensive calibration of the Landsat thermal archive of data from 1982 to the present. The different methodologies used to accomplish the entire calibration/validation update are reviewed in Section 2. The new results for Landsat 4 and early Landsat 5 as well as the previously reported results for Landsats 5 and 7 (Barsi et al., 2003; Hook et al., 2004) are included in Section 3. In particular, this section includes a discussion of the small residual uncertainty in the data archive associated with each data set. This uncertainty in the collective archive

* Corresponding author. Tel.: +1 585 475 5170; fax: +1 585 475 5988.
E-mail address: schott@cis.rit.edu (J.R. Schott).

0034-4257/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2011.07.022

of approximately 0.6 K means that, with good knowledge of the atmosphere and emissivity, surface temperatures can be retrieved to better than 0.7 K (one sigma) using standard analytical approaches. These results mean that for the first time users can access and analyze the entire 30 year record with confidence in the radiometric integrity of the thermal data.

2. Background

This section will introduce the Landsat sensors from the thermal infrared perspective and then briefly review the approaches that have been used over time for vicarious radiometric calibration of the Landsat thermal instruments.

2.1. Instruments

Landsat 3 was the first Landsat satellite to include a thermal sensing capability. However, the thermal sensor failed quite early in the mission and no significant effort has yet been expended to characterize its post launch performance. Therefore, this paper will focus on the thermal sensors on the Landsat 4 Thematic Mapper (TM4), the Landsat 5 Thematic Mapper (TM5) and the Landsat 7 Enhanced Thematic Mapper plus (ETM+). From the thermal infrared perspective these three instruments were nearly identical with the exception being that ETM+ had eight thermal detectors instead of the 4 on the TM instruments resulting in a 60 m ground instantaneous field of view (GIFOV) instead of 120 m for the TM instruments. Each instrument had a single spectral band nominally covering the 10.5–12.5 μm spectral window (see Fig. 1) with two eight bit gain settings. Fig. 2 shows a generic optomechanical schematic of the instruments highlighting the elements relevant to the thermal band. In particular, note that on the TM instrument 4 detectors are swept across the field of view (FOV) forming 4 120 m lines of data per sweep of the scan mirror and that on ETM+ 8 detectors sweep 8 60 m lines of data. Also note that as the scan mirror is turning around, a calibration shutter is inserted into the line of sight. The shutter has a high emissivity surface with a monitored temperature and a mirror that reflects a cavity blackbody into the line of sight. Thus, as the shutter is swept into the line of sight, the detectors "see" first the known radiance from the shutter and then the known radiance from the blackbody. This provides a two point calibration at the beginning and end of each line of data. These two points are used to calculate the internal gain (g_i) according to

$$g_i = \frac{DN_{BB} - DN_S}{\hat{L}_{BB\lambda} - \hat{L}_{S\lambda}} \tag{1}$$

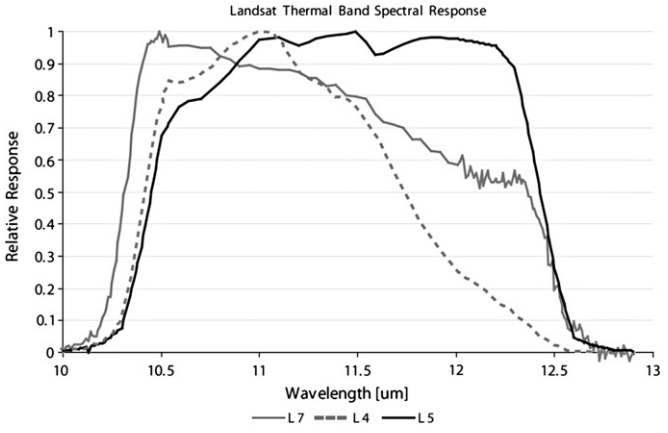

Fig. 1. Plots of the relative spectral response of the TM4, TM5 and ETM+ instruments.

where DN_{BB} and DN_S are the average digital number from a sample of image data taken from the blackbody and the shutter respectively and $\hat{L}_{BB\lambda}$ and $\hat{L}_{S\lambda}$ are the spectral radiance in the passband associated with a blackbody at the temperature of the blackbody (BB) and shutter (S) respectively, i.e.

$$\hat{L}_{BB\lambda} = \frac{\int L_{BB\lambda} R(\lambda) d\lambda}{\int R(\lambda) d\lambda} \tag{2}$$

where $L_{BB\lambda}$ is the spectral radiance from a blackbody at temperature BB [W m^{-2} sr^{-1} μm^{-1}] and $R(\lambda)$ is the relative spectral response of the sensor with wavelength λ.

The actual sensor output (DN) can then be related to the radiance from the scene (L_λ) as

$$DN = g_f(g_i L_\lambda + b_i) + b_f \tag{3}$$

where b_i is an internal bias level related to the shutter radiance (i.e. shutter temperature) and g_f and b_f are gain and bias terms associated with multiplicative and additive effects not captured by the internal calibration process (e.g. transmission losses from the optics forward of the shutter and additive radiance from the optics forward of the shutter). Note that the actual equations used in the software for the different instruments vary slightly from instrument to instrument but can be generalized to the form shown in Eq. (3).

Values for the remaining calibration coefficients (i.e. b_i, g_f and b_f in this representation) were determined from pre-launch laboratory calibration with external radiance sources. Given these calibration coefficients, the sensor reaching radiance can be calculated from the recorded DN for each pixel by inversion of Eq. (3). Some of these coefficients are dependent on the operating temperatures of the forward optics. In practice, coefficients appropriate for the expected operating temperatures were calculated empirically pre-launch, however, no trusted fore-optic calibration model was developed which would allow adjustment of the coefficients based on the recorded temperatures of the optical elements. Thus, when significant changes to the instrument's operating temperatures occur (either planned or unplanned), or if there is degradation of the forward optics, new values for the calibration coefficients must be determined and applied to the calibration equation (Eq. 3) to maintain the calibration of the data. Because there is no onboard system to adequately monitor any radiometric changes in the forward optics, vicarious calibration procedures have been regularly used to verify the stability of the calibration after launch and then (at least since 1999) to monitor the calibration on orbit.

2.2. Vicarious calibration approaches

Four different vicarious calibration techniques have been used to support the calibration of the Landsat thermal bands. Each of these will be introduced in this section. All of the methods use water as a target. Water is an ideal target because of its high thermal inertia, high emissivity and as a liquid it is hard to maintain thermal gradients, so large regions of uniform temperature are common (Schott et al., 2004; Tonoka et al., 2005). All of the methods also use the radiation propagation code MODTRAN (Berk et al., 1999) to propagate the radiation to the sensor. The governing radiometry equations and the use of MODTRAN will therefore be introduced in Section 2.2.1 before the discussion of the different measurement techniques.

2.2.1. Atmospheric propagation

The governing equation for radiation propagation to the sensor can be expressed as

$$\hat{L}_\lambda = \frac{\int [(\varepsilon(\lambda) L_{T\lambda} + (1-\varepsilon(\lambda)) L_{d\lambda}) \tau(\lambda) + L_{u\lambda}] R(\lambda) d\lambda}{\int R(\lambda) d\lambda} \cong \hat{L}_{obs\lambda} \tau + \hat{L}_{u\lambda} \tag{4}$$

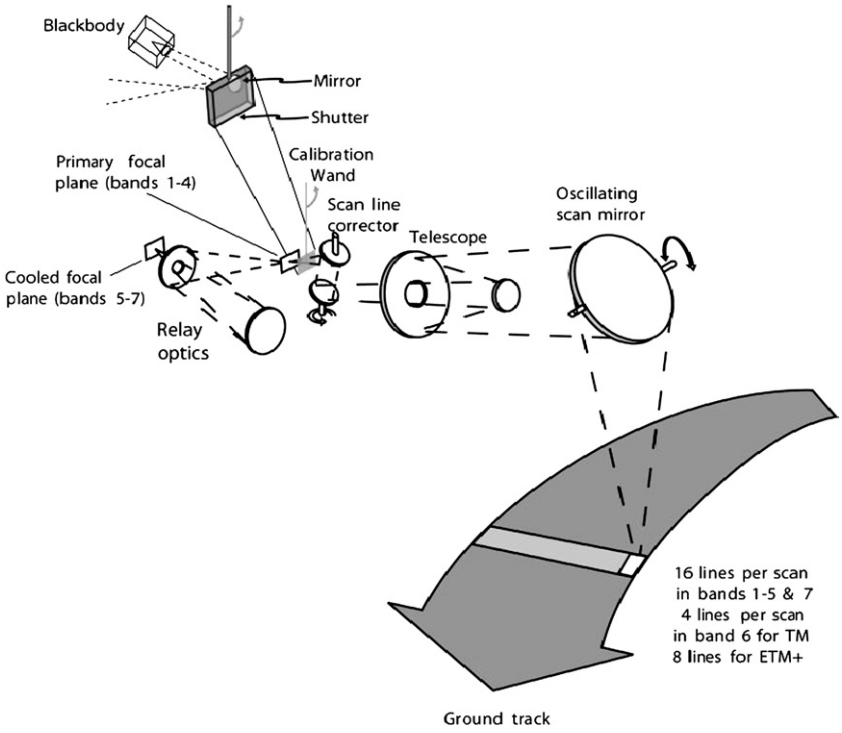

Fig. 2. Optomechanical schematic of TM and ETM+ instruments showing the thermal features.

where ε is the emissivity of the target, $L_{T\lambda}$ is the spectral blackbody radiance associated with a target at temperature T, $L_{d\lambda}$ is the spectral downwelled radiance, $L_{u\lambda}$ is the spectral upwelled radiance, $\tau(\lambda)$ is the transmission from the target to the sensor and in general all the terms are a function of wavelength (λ). In practice the emissivity of water is essentially constant over the Landsat bandpass at approximately 0.986. In addition, the emissivity is independent of water condition and surface roughness for the Landsat fields of view (Schott et al., 2004). Note that $L_{obs\lambda}$ is the effective spectral radiance in the Landsat band observed at the ground or at the aircraft altitude for the aircraft underflight calibration approach, τ is the effective transmission over the Landsat passband from the observation location to the sensor and $L_{u\lambda}$ is the effective spectral upwelled radiance in the passband for the path between the observation and the sensor. These effective passband values are used when surface or aerial radiometric measurements or estimates of the target are available.

MODTRAN can be used to solve for all of the terms in Eq. (4) if the atmosphere has been characterized and the surface temperature and emissivity are supplied. In practice, for the Rochester Institute of Technology (RIT) analysis (Sections 2.2.3 and 2.2.5), the atmosphere is characterized using the closest available radiosonde data typically adjusted in the boundary layer (0 to approximately 0.5 km) for observed surface conditions (temperature and relative humidity). This surface correction accounts for variation in the lowest layers of the atmosphere that may exist because of spatial and temporal offsets between the target location and sample time and the radiosonde/upper-air-sampling location and time (Padula, et al., 2011). The Jet Propulsion Laboratory (JPL) analysis uses an atmospheric profile from the nearest atmospheric reanalysis grid point (see Section 2.2.4).

In order to take advantage of Eq. (4) to predict the expected sensor reaching radiance we must either know the temperature and emissivity of the target (left form of Eq. 4) or the radiance from the target (right form of Eq. 4). In the next subsections we will briefly review four methods that have been used to measure the temperature or radiance (recall that emissivity is known and constant for the Landsat observation conditions over water bodies).

2.2.2. Underflight approach

Schott and Volchok (1985) describe the method that was used in the mid 1980s in an attempt to calibrate TM4 and TM5 immediately after they were launched. The method involved flying a well-calibrated infrared line scanner underneath the spacecraft and imaging water targets simultaneously (i.e. $\pm$30 min) with the Landsat acquisition (see Fig. 3a). The infrared line scanner was spectrally filtered to match the Landsat spectral response and calibrated such that the effective spectral radiance could be calculated for any pixel in the image. This yields values for $\hat{L}_{obs\lambda}$ which can be propagated to the spacecraft using MODTRAN and the righthand side of Eq. (4).

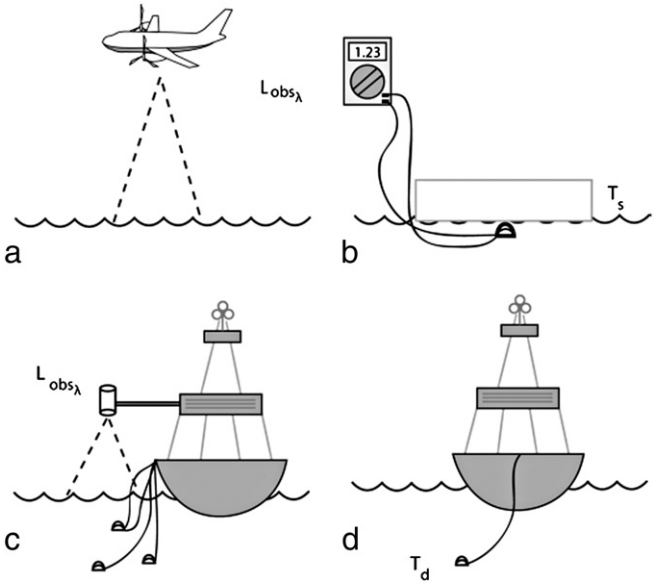

Fig. 3. Illustration of methods used to measure surface temperature/radiance. a. Aircraft approach — Section 2.2.2, b. surface kinetic temperature approach — Section 2.2.3, c. surface temperature approach using radiometers — Section 2.2.4 and d. NOAA buoy approach — Section 2.2.5.

Because the aircraft was typically flown at 3–5 km above ground level the observed radiance incorporated the largest and most variable effects due to the atmosphere, thereby reducing any errors due to less than perfect atmospheric propagation. MODTRAN was only used to correct for the atmosphere above the aircraft. Any small error in these small corrections results in very small errors in the predicted sensor reaching radiance.

In addition, because the aerial system has a much smaller ground sample distance (GSD) than the satellite, many pixels of a uniform region of the water can be averaged to estimate $\hat{L}_{obs\lambda}$. This reduces noise in the estimate by the square root of the number of samples (typically 25 to 100) resulting in very precise estimates of $\hat{L}_{obs\lambda}$. The uncertainty of this method is typically limited only by the calibration uncertainty of the aerial instrument (see Section 3.4).

The obvious drawbacks to this approach are the cost and logistical difficulties associated with acquiring significant amounts of aerial data coincident with the spacecraft as well as difficulties associated with maintaining the calibration knowledge of the airborne instrument. On the positive side, one can often fly over different water regions with some range of temperatures and thereby obtain several calibration points from a single successful flight. Nevertheless, this method is difficult to justify operationally and so other methods must be considered.

2.2.3. Surface kinetic temperature measurement approach

To augment, or in lieu of the aerial approach, surface kinetic temperature measurements can be employed. To avoid near surface gradients this approach should be limited to well mixed waters and measurements should be taken very close to the surface. RIT uses thermistors attached under small blocks of Styrofoam that are floated on the surface on the windward side of the boat (see Fig. 3b). The targets have usually been the waters of Lake Erie or Lake Ontario which have very long fetch (i.e. a long stretch of open water in the upwind direction) and are rarely still. This method has been extensively used by RIT since 2001 to support calibration of ETM+ and TM5 (Barsi et al., 2003). The temperatures obtained by this method along with the emissivity of water and the radiative transfer values from MODTRAN allow the use of Eq. (4) to predict the sensor reaching radiance. While this method has a slightly higher error associated with each individual measurement (see Section 3.4) than the aerial approach, the larger number of points that can be acquired tends to compensate in the overall instrument calibration uncertainty.

A limitation of the surface temperature approach is that the temperature measured is very slightly below the surface and the infrared sensors measure the true surface temperature i.e., the skin temperature. Under most circumstance when calibration data would be acquired (clear skies), radiational surface (top microns) cooling lowers the surface temperature slightly (up to a few tenths Kelvin) when compared to the temperature immediately below the surface (few mm). Generally skin temperature is cooler for well mixed waters, however, if the water is still then the skin temperature can be warmer than the temperature immediately below the surface. Surface radiometers can be used to avoid this limitation.

2.2.4. Surface temperature measurement approach using field radiometers

Since 1999, NASA's Jet Propulsion Lab (JPL) has operated four instrumented buoys in Lake Tahoe CA/NV, and since 2008 a similarly instrumented platform in the Salton Sea CA. Note, the Salton Sea site was added to allow acquisition of higher temperature/radiance targets to allow better assessment of small gain errors in the satellites being calibrated. The instrumentation on each platform includes near surface contact thermistors, near nadir viewing calibrated radiometers and weather stations. The JPL suite of field sensors has been used to perform thermal calibration assessment of a number of sensors including MODIS and ASTER and therefore uses radiometers with a wide passband (Hook et al., 2003, 2005, 2007). Because the radiometers are not filtered to match the Landsat spectral passband, the surface temperature corrected for the cool skin effect is computed using a combination of the observed radiometric temperature, the near surface contact temperature and the downwelled radiance computed from MODTRAN (Hook et al., 2003; Hook et al., 2004). The data are acquired every 2 to 5 min and transmitted to JPL for processing. The output from the processing is an estimate of the surface kinetic temperature which can be combined with the surface emissivity and MODTRAN generated radiative transfer parameters to generate the predicted sensor reaching radiance (see Eq. 4). The atmospheric profile data used for input to MODTRAN come from the nearest National Center for Environmental Prediction (NCEP) reanalysis point interpolated to the Landsat acquisition time (Hook et al., 2007).

Regrettably none of the methods discussed so far could be used to assess temporal gaps in the calibration of the Landsat instruments during a period when NASA was not funding Landsat thermal band vicarious calibration programs (1985–1999). Thus, a method was needed to fill this long knowledge gap.

2.2.5. Estimation of surface temperature from subsurface measurements by NOAA buoys

The National Oceanic and Atmospheric Administration (NOAA) operates a fleet of moored buoys in U.S. coastal waters and in the Great Lakes. These buoys record hourly subsurface temperatures (0.6 m or 1.5 m) as well as weather data and archive it in the National Data Buoy Center (NDBC). The archive includes buoy records spanning the entire period of interest (1982–present). Temperature from buoys has been used previously to validate atmospheric retrieval algorithms and instrument calibration for the Advanced Very High Resolution Radiometer (AVHRR) (Emery et al., 2001; Walton et al., 1998). Padula and Schott (2010) describe an improved technique for estimation of surface temperature from the subsurface buoy temperatures. The method utilizes the 24 h of temperature measurements before the satellite overpass along with surface meteorological data to compute surface temperature from the subsurface values. The method accounts for the diurnal temperature cycle, the temporal phase shift in the diurnal cycle with depth, thermal gradients with depth that are a function of wind speed and finally the cool skin effect. The derived surface temperature was then used along with emissivity values, local weather data and radiosonde data as input to MODTRAN to predict sensor reaching radiance as described in Eq. (4). By comparison to well-calibrated ETM+ and TM5 data post 1999, the sensor reaching radiance values predicted by this method were shown to be in good agreement (mean bias differences less than 0.2 K) with the predictions made using the JPL and RIT surface temperature/radiance methods described above. Therefore, this technique was the source of data for the period 1982–1999 and is the basis for the TM4 and early TM5 calibration results.

3. Results

This section reviews how the methods described in the previous section have been used to calibrate the complete archive of TM and ETM+ thermal data. The formation of the calibration team, that led to the effort to assess and update the calibration of the archive, began with the launch of Landsat 7. Therefore, we had the discussion begin there and work back in time through the instruments.

Because Padula and Schott (2010) and Barsi et al. (2003) show that all the methods introduced in the previous section predict sensor reaching radiance values that are in agreement with each other to within the errors in the methodology (see Section 3.4), all of the methods will be considered as trusted unbiased sources of sensor reaching radiance.

3.1. Landsat 7 — ETM+

At launch RIT and JPL were independently charged with assessing the thermal calibration of Landsat 7. RIT used aerial underflights and surface temperature measurements on the Great Lakes, and JPL used the automated sites at Lake Tahoe and later at the Salton Sea. Both teams quickly identified a significant constant bias in the data of 3.1 (W m^{-2} sr^{-1} μm^{-1}) or 0.31 K at 300 K, where bias will be expressed as radiance observed by the instrument minus at sensor radiance predicted from the ground truth measurements. The bias was calculated as the weighted average of all of the individual estimates available at the time. The weights were proportional to one over the standard error about the mean bias estimate obtained by each of the teams. The bias correction was implemented by USGS EROS on 12/20/2000, and like all other corrections discussed here it was applied to all impacted data such that any data processed after 12/20/2000 should be unbiased, including all data acquired before that date. Note this bias was eventually determined to be the result of an error in coefficients in the processing system and not due to any changes to the instrument on launch.

The calibration team continued to monitor the ETM+ thermal performance adding more data each year with an emphasis on high and low temperature values as the large amount of data began to indicate that a slight error in gain had been present since launch. Because the error was small it could only be identified with confidence when significant amounts of warm and cold data were added to the data set. In 2009 with data ranging from 6.3 to 9.6 [W m^{-2} sr^{-1} μm^{-1}] or 275 to 305 K it was determined that a 5.8% error in gain existed. The gain error is expressed as

$$S_g = \left(1 - \frac{\Delta\hat{L}_{Sat\lambda}}{\Delta\hat{L}_{\lambda}}\right) \cdot 100 \tag{5}$$

where $\Delta\hat{L}_{Sat}/\Delta\hat{L}_{\lambda}$ is the slope of the satellite observed spectral radiance versus the predicted radiance values for all of the calibration points. Fig. 4 shows the calibration data for the two teams that indicated the need for a small gain correction. Note that for most of the radiance range the correction causes only small changes in temperature. However, for hot targets (300 K) the change could be as large as 0.8 K. The gain correction was implemented in the USGS processing system on 1/1/2010.

Based on the early success of the ETM+ vicarious calibration campaigns, NASA tasked the calibration team to assess the radiometric calibration of the TM5 instrument.

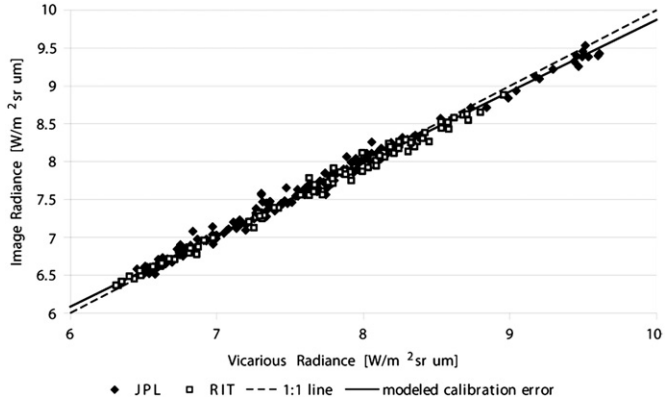

Fig. 4. Plot of predicted versus observed at sensor radiance for ETM+ showing the need for a small gain correction.

3.2. Landsat TM5 results

Schott and Volchok (1985) and Schott et al. (1987), using the RIT aerial technique, obtained a small number of TM5 calibration points shortly after launch in 1985. They reported a small bias of 0.58 K, which was within the approximately 1 K uncertainty of the vicarious instrumentation available in the early 1980s, and no adjustment was recommended. There was no significant effort to evaluate the TM5 calibration from 1985 until the RIT and JPL teams were tasked to investigate it in approximately 2001.

By using its automated sites, the JPL team was able to go back to 1999 when they were initialized. In 2006, based once again on the combined JPL automated site data and RIT surface temperature data, it was determined that a small bias error of −0.092 [W $m^{-2}sr^{-1}$ μm^{-1}] or −0.68 K at 300 K was present in all of the available data from 1999 on. On 4/1/2007 USGS, based on the recommendation of the Landsat calibration team, implemented a correction to the data processing to remove the bias from all data processed after that date applicable to imagery acquired after 4/1/1999. Because no RIT or JPL calibration data were available for the period 1985–1999 it was not clear how, or if, to correct data prior to April 1999 so no change was made to the earlier data at that time.

O'Donnell et al. (2002) describe an effort to evaluate the (1985–1999) calibration gap using cold water temperatures, from the center of the Great Lakes in winter, as ground truth. The assessment indicated that there was no discernable calibration error over the 1985–1999 period. However, this was limited by uncertainty in the knowledge of the lake temperatures to 1 to 2 K uncertainty in apparent temperature at the sensor.

In the early 2000s RIT had begun to investigate whether data from the NOAA buoys could be used to accurately determine surface temperature. Padula and Schott (2010) describe a pair of experiments using data from a few buoys (Great Lakes and off the Delmarva Peninsula) propagated to sensor reaching radiance. They first compared the predicted radiance to the calibrated radiance observed by the ETM+ sensor for 32 points from 2000 to 2007. The gain and bias values estimated from the buoy data (near unity and zero, respectively, as expected) were shown to be statistically the same as the values estimated using the accepted RIT and JPL measurement techniques. The second experiment compared the buoy-based calibration results for TM5 to the results obtained by the surface temperature approach for the time period after field campaigns had resumed (1999). The results were similar to the ETM+ results leading to the conclusion that the NOAA buoy archive could be used to provide a source of "ground truth" for calibration of the data during the 1985–1999 gap.

Padula and Schott (2010) report that using 7 buoys, 198 calibration points were obtained spanning the period 1984–2007. These results suggested that a small but statistically significant gain error had been present in the data since launch and a bias shift of approximately −0.69 K occurred sometime in the late nineties (consistent with the results reported above for the 1999+ data). By adding additional buoys, more data were added in an attempt to better estimate the date when the bias change occurred (see Fig. 5). In addition, significant amounts of new TM5 data from all the methodologies were combined to estimate the calibration error for the entire TM5 data set. Using the composite data set it was determined that a bias shift of −0.11 [W m^{-2} sr^{-1} μm^{-1}] occurred in the early part of 1997 (no source for the shift in bias has been determined) and that a small gain error of 5.0% was present in all of the TM5 data. This small deviation from a unity slope in the predicted to observed radiance plots had been suggested by the earlier data, however, the addition of a larger volume and range of data increased the statistical confidence to a point where correction was warranted. Based on the recommendation of the calibration team, USGS implemented changes to the TM5 data processing to implement these calibration updates for data processed after 4/1/2010.

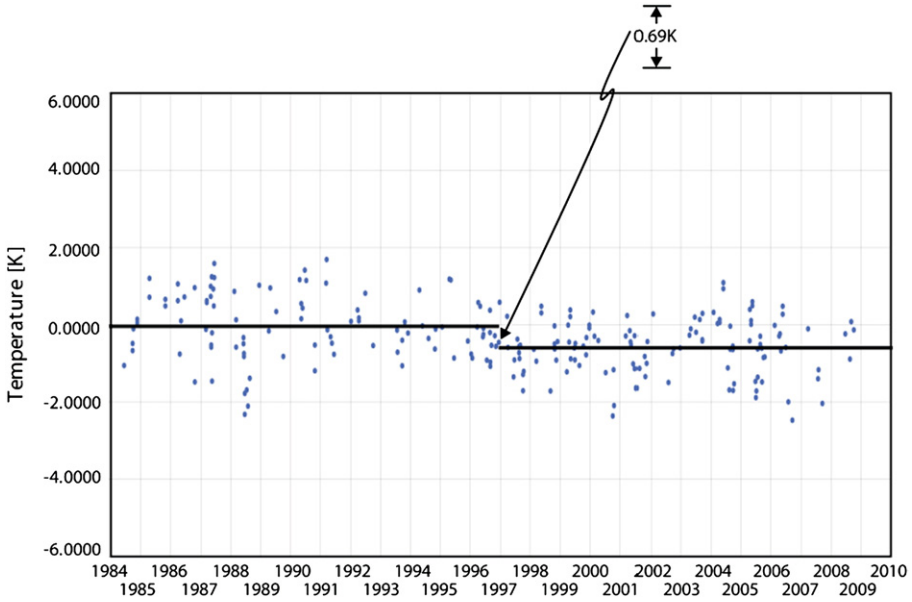

Fig. 5. TM5 temporal plot showing bias estimates by date as well as the mean bias line for the periods pre and post 1/1/1997.

3.3. Landsat TM4 results

The TM4 instrument was launched in 1982 and operated for nearly a year before being put in on-orbit storage because of solar array and direct downlink failures. It was returned to operation three years later, using the Tracking and Data Relay Satellite System (TDRSS) to downlink data, and successfully operated for another 6 years. During its short early life no rigorous validation of the calibration was accomplished nor was a validation attempted when it was returned to operation. Based on the successful use of the NOAA buoys in filling the calibration gap in the TM5 calibration history, a similar effort was undertaken for TM4. The problem with the TM4 database was that it was more sparsely populated because of the short period of initial use and somewhat limited use over the U.S. after the storage period.

To deal with this issue, a significantly larger set of buoys and corresponding radiosonde sites were used to find simultaneous buoy, radiosonde and clear TM4 image dates. In all, 17 buoy sites were used in the TM4 calibration resulting in 9 calibration points from 1982 to 1983 and 19 points from 1987 to 1992. The results showed that there was no significant bias error apparent in the early TM4 data but that a large consistent bias of −0.43 [W m^{-2} sr^{-1} μm^{-1}] or −3.3 K at 300 K was present post storage (see Fig. 6).

The post storage data showed no significant gain error (less than 0.5%) so all the error was attributed to and corrected with a bias adjustment to be implemented in 2011 and applied to all TM4 post storage data processed after that date. In analyzing the TM4 data the variability of the bias estimates pre storage was observed to be significantly larger (factor of two) than post storage and also significantly larger than the variability in the bias estimates observed for TM5 or ETM+. In an effort to understand/explain the source of this variability and the bias change post storage, the instrument operating temperatures were analyzed by plotting some telemetry records of the forward optical element temperatures available from the NASA Landsat Image Assessment System (IAS) (see Fig. 6). These data indicate two things. The first is the dramatic change in operating temperature post storage which is a likely source of the bias shift (recall that no accepted fore optics radiation model exists so no model-based test can be run to assess this assumption) and the second is the relatively large variation in operating temperatures early in the lifetime (i.e. pre storage) which is a likely source of the larger than expected bias variation pre storage (note: this large variation in temperature is only partially captured in the small sample shown here for the scan line corrector (SLC)). This discussion leads us to an assessment of the uncertainty remaining in the calibration of the data now coming from the USGS archive.

3.4. Residual uncertainty in the calibrated database

By implementing the calibration adjustments described above, all known systematic errors associated with the Landsat thermal bands should be reduced to insignificant levels compared to the precision errors associated with the calibration procedures and the instrument noise. Thus, in theory the overall uncertainty in the radiance values generated from the data in the archive can be expressed as

$$S_L = \left[S_p^2 + S_i^2\right]^{\frac{1}{2}} \tag{6}$$

where S_L is the uncertainty in the "calibrated" data, S_p is the uncertainty in the radiance predicted at the sensor through the various calibration procedures (assumed to be unbiased (i.e. accurate) at this point) and S_i is the noise in the instrument measurements (note: all uncertainties are expressed as 1 standard deviation values). Assuming all

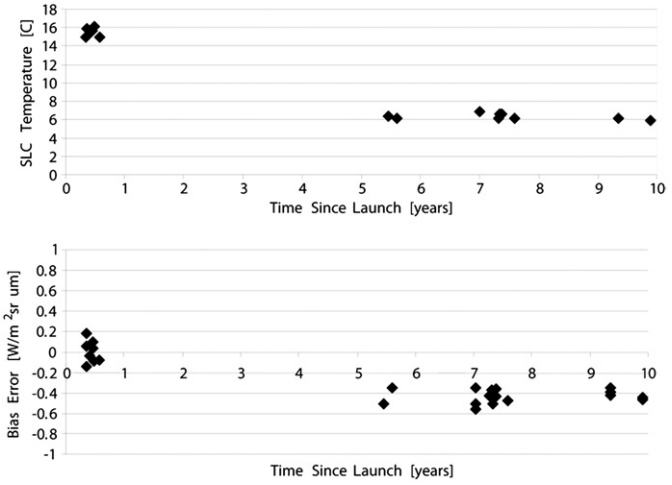

Fig. 6. A time series of TM4 bias estimates and corresponding temperatures of the scan line corrector (SLC) during the same period from the IAS.

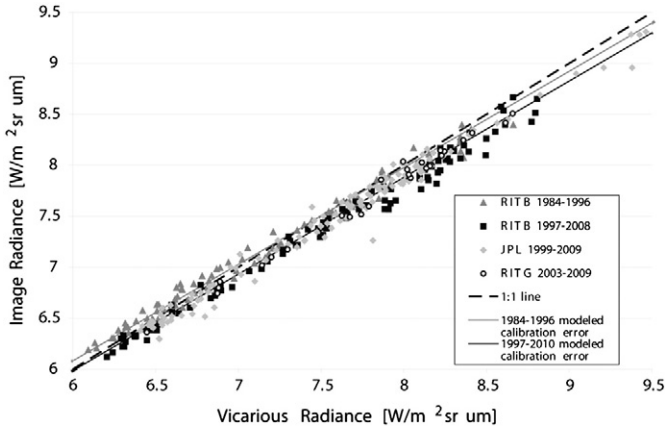

Fig. 7. Plot of radiance predicted at the sensor versus radiance measured at the sensor for TM5 data for the two time periods (pre and post 1/1/1997). The RMS deviation about the best fit lines provides an estimate of the residual uncertainty about the data in the archive.

systematic variation in the instrument signals is accounted for through the onboard calibration processes, S_i should just be the standard deviation of the signal about the mean when observing a constant flux (e.g. from the onboard blackbody). The overall uncertainty (S_L) can also be measured empirically by calculating the root mean square error (S_{RMS}) in the observed radiance values from the best fit or final calibration line generated from the calibration data (see for example, Fig. 7). This empirical measurement is the best estimate of the residual uncertainty in the data from the archive. Furthermore any difference between the modeled uncertainty (Eq. 6) and the observed uncertainty (S_{RMS}) suggests that our estimates of S_p or S_i are in error or that there are unaccounted sources of variability in the process.

Table 1 includes estimates of the expected uncertainty: in sensor reaching radiance for each methodology (S_p), due to noise in each instrument (S_i), as well as the modeled (S_L) and measured (S_{RMS}) estimates of the overall uncertainty. In some cases where significantly different methods were used for calibration or different uncertainties measured, multiple results are shown per instrument. Note that all the radiance uncertainties are shown in equivalent apparent temperature (typically at 300 K to make them more intuitive).

The S_i values in Table 1 are measured values for each instrument and in all cases they show no significant change over the lifetime of the instruments. However, the values for TM4 and 5 do change within an outgassing cycle because the internal gain decays as ice builds up on the filters resulting in worse error due to electronic/quantization noise as the transmission decreases. The uncertainties in the predicted radiances (S_p) for the NOAA buoy and surface temperature methods (RIT) are drawn from the results of an extensive error propagation study that cascades the uncertainties in the temperature measurements with the uncertainties in temperature propagation to the surface skin temperature, uncertainties in the meteorological and radiosonde measurements and uncertainties in the radiation propagation models (Padula et al., 2011). The uncertainties in the predicted radiances for the aerial technique are drawn from a similar analysis reported in Schott et al. (2004). Finally the errors in predicted radiance for the surface radiometer/near surface thermistor technique (JPL) are estimated from surface temperature uncertainty estimates drawn from Hook et al. (2007) and estimates of the radiation propagation uncertainties based on the automated use of unfiltered NCEP data. Note, all the data leading to the estimates of S_L in Table 1 are assuming single pixel measurements. If multiple uniform points are averaged, the instrument noise can be significantly reduced such that S_L should approach S_p.

Analysis of Table 1 indicates that the S_L and S_{RMS} values are in general agreement with the exception of the early TM4 results. Solving for the magnitude of the uncertainty unaccounted for by the error model as

$$S_u = \left(S_{RMS}^2 - S_L^2\right)^{\frac{1}{2}} \tag{7}$$

yields a value of 0.86 K. As indicated, above we believe that much of this unaccounted variability is due to the large variations in instrument operating temperatures during its first year of operation. It is also possible that the much smaller variability in operating temperatures for TM4 after storage and for the other instruments is still a significant contributor to the smaller unaccounted uncertainties from these instruments.

Table 1
Residual uncertainties in the data from the USGS Landsat archive expressed in apparent temperature [K]. Values in parenthesis are the number of points included in the analysis.

	Uncertainly in predicted radiance S_p	Instrument noise S_i	Modeled uncertainty in sensed radiance S_L	Observed variability about best fit calibration line S_{RMS}	Observed variability unaccounted uncertainty S_u
Aerial (A)	0.31				
Surface temperature (RIT)	0.34				
Surface radiometers and thermistors (JPL)	0.35				
Subsurface temperature (NOAA buoys)	0.41				
Landsat 7 (composite) (324)		0.21[a]	0.41	0.48[b]	0.25
RIT (51)			0.40	0.32	
JPL (234)			0.41	0.48	
NOAA buoys (39)			0.46	0.59	
Landsat 5		0.17–0.3			
1984–1998 NOAA buoy (102)			0.44–0.51	0.53[b]	0.24
1997–2010 composite (285)			0.41–0.48	0.66[b]	0.49
RIT (29)			0.38–0.45	0.48	
JPL (149)			0.39–0.46	0.73	
NOAA buoy (107)			0.44–0.51	0.60	
Landsat 4		0.22–0.32			
1982–2983 NOAA buoy (9)			0.47–0.52	0.98[b]	0.86
1987–1992 NOAA buoy (19)			0.47–0.52	0.43[b]	

[a] $NE\Delta T$ for the low gain is (0.26 K).
[b] These are the best values to use for the expected uncertainty in the radiance values (i.e. they are based on the average of all available points for the respective data era).

4. Impact on user community

With the implementation by USGS of the Landsat 4 results reported here, all of the Landsat 4, 5 and 7 thermal data produced from the archive after April 1, 2011 should be free of first order systematic errors. This, for the first time, will allow users to process any of these data with confidence in their integrity. These data now represent one of the longest well-calibrated records of the thermal history of the earth's surface. This will allow the science community to study the temporal trends in thermal behavior of relatively small landscape features (Schneider et al., 2009). Coupled with the parallel work to radiometrically calibrate the reflective data in the Landsat archive (Markham & Helder, 2012), this calibration effort allows the Landsat archive to truly live up to its goal to serve as the long-term record of the planet at human scales.

From a quantitative standpoint, the residual uncertainty in the thermal data is approximately 0.6 K when the radiance uncertainty is expressed as a change in apparent temperature at 300 K (see Table 1 for a detailed breakdown by instrument and time period). If a user were to convert the radiance values to surface temperatures using similar procedures (MODTRAN, clear atmospheric conditions and known emissivities) to those used in the calibration process, this should result in only slightly larger uncertainties in the estimated temperatures of approximately 0.7 K (Padula, 2008).

Finally, while the results reported here indicate that the Landsat archive is well calibrated with quite small residual uncertainty, they also suggest that there are small sources of uncertainty that are not accounted for by the error models. If these sources can be identified and are systematic rather than random it may be possible to drive the errors down even further (to less than 0.5 K). At present several possible sources for this unaccounted variability are under evaluation. The most intriguing systematic source is variation in observed radiance due to changes in the operating temperature of the instrument. A satisfactory model for this systematic process might serve to reduce the relatively small error reported here even further.

5. Lessons learned and future directions

Having summarized here the calibration/validation of three decades of Landsat thermal instrument data, it seems appropriate to review some lessons learned. First, having two independent teams was invaluable for quickly identifying issues and confirming problems with the spacecraft. When the team's results weren't in agreement, it prompted a thorough review of procedures that could identify a processing issue. More importantly, when both teams were in agreement that the data were out of calibration by the same amount, then a quick fix could be implemented. Second, it is important to maintain a continuous monitoring of the calibration of the instruments. In the Landsat case, Landsats 4 and 5 appeared nominal at launch, but 7 needed an immediate calibration update. However, the calibration of 4 and 5, which would have appeared stable, changed abruptly during their lifetimes. Furthermore, it took a significant amount of data over a wide temperature range to establish with confidence that small but significant gain corrections were needed in Landsats 5 and 7. Third, at the permanent monitoring sites, it was found that acquiring day and night data was useful as the nighttime data showed reduced uncertainty. This is most likely due to the greater thermal stability of the lakes at night. Finally, operating the instrument in a consistent fashion over time (i.e., avoiding radical changes in duty cycle) could reduce uncertainty in the data and possibly calibration changes (see, in particular, the discussion of Landsat 4TM).

The successful use of the NOAA buoy data for calibration of early TM5 data and, particularly, the very consistent results reported here for TM4 when many buoys were employed, suggests that many buoys used in an operational mode (i.e., in a semi-automated fashion) could provide ongoing monitoring of Landsats 5 and 7. In addition, the use of many NOAA buoys augmented with the JPL buoy data could quickly provide many data points for calibration of LDCM immediately after launch. To take advantage of the opportunity, RIT plans to develop tools to use the NOAA buoys in an operational mode.

Finally with the demonstration reported here, that the Landsat thermal archive is calibrated in radiance to acceptable levels, the opportunity exists to consider development of land surface temperature (LST) maps from the Landsat radiance data. USGS NASA Goddard, JPL, and RIT are initiating a proof of concept effort to demonstrate that LST maps can be operationally produced from the Landsat thermal archive.

References

Allen, R. G., Hendricks, J. M. H., Toll, D., Anderson, M., Kustas, W., & Kleissl, J. (2008). From high overhead-ET measurement via remote sensing. *Southwest Hydrology*, *7*, 30–32.

Anderson, M. C., & Kustas, W. P. (2008). Thermal remote sensing of drought and evapotranspiration. *EOS Transactions*, *89*(26), 233.

Barsi, J. A., Schott, J. R., Palluconi, F. D., Helder, D. L., Markham, B. L., & Chander, G. (2003). Landsat TM and ETM+ thermal band calibration. *Canadian Journal of Remote Sensing*, *29*(2), 141–153.

Berk, A., Anderson, G. P., Bernstein, L. S., Acharya, P. K., Dothe, H., Mathew, M. W., et al. (1999). MODTRAN4 radiative transfer modeling for atmospheric correction. *Optical spectroscopic techniques and instrumentation for atmospheric and space research III* (pp. 348–353). : SPIE.

Emery, W. J., Baldwin, D., Schlussel, P., & Reynolds, R. (2001). Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *Journal of Geophysical Research*, *106*, 2387–2405.

Hook, S. J., Chander, G., Barsi, J. A., Alley, R. E., Abtahi, A., Palluconi, B. L., et al. (2004). In-flight validation and recovery of water surface temperature with Landsat-5 thermal infrared data using an automated high-altitude lake validation site at Lake Tahoe. *Transactions on Geoscience and Remote Sensing*, *42*(12), 2767–2776.

Hook, S. J., Clodius, W. B., Balick, L., Alley, R. E., Abtahi, A., Richards, R. C., et al. (2005). In-flight validation of mid and thermal infrared data from the Multispectral Thermal Imager (MTI) using an automated high altitude validation site at Lake Tahoe CA/NV, USA. *IEEE on Transactions Geoscience and Remote Sensing*, *43*, 1991–1999.

Hook, S. J., Prata, A. J., Alley, R. E., Abtahi, A., Richards, R. C., Schladow, S. G., et al. (2003). Retrieval of lake bulk and skin temperatures using Along-Track Scanning Radiometer (ATSR-2) data: A case study using Lake Tahoe, California. *Journal of Atmospheric and Oceanic Technology*, *20*(4), 534–548.

Hook, S. J., Vaughan, R. G., Tonooka, H., & Schladow, S. G. (2007). Absolute radiometric in-flight validation of mid infrared and thermal infrared data from ASTER and MODIS on the terra spacecraft using the Lake Tahoe, CA/NV, USA. *Transactions on Geoscience and Remote Sensing*, *45*(6), 1798–1807.

Markham, B., & Helder, D. (2012). Forty-year calibrated record of earth-surface reflected radiance from Landsat: A review. *Remote Sensing of Environment, Landsat special edition*.

O'Donnell, E. M., Schott, J. R., & Raqueño, N. G. (2002 June). Calibration history of Landsat thermal data. *Proceedings at IGARSS, IEEE, Vol. 1*. (pp. 27–29) Toronto, Canada.

Padula, F. P., Schott, J. R., Barsi, J. A., Raqueno, N. G., & Hook, S. J. (2011). Calibration of Landsat 5 thermal infrared channel: Updated calibration history and assessment of the errors associated with the methodology. *Canadian Journal of Remote Sensing*, *36*(No. 5).

Padula, F. P., & Schott, J. R. (2010). Historic calibration of the thermal infrared band of Landsat-5 TM. *Photogrammetric Engineering and Remote Sensing*, *76*(11), 1225–1238.

Padula, F. P. Historic Thermal Calibration of Landsat-5 TM through an Improved Physics Based Approach. Master's thesis, Rochester Institute of Technology, 2008.

Schneider, P., Hook, S. J., Radocinski, R. G., Corlett, G. K., Hulley, G. C., Schladow, S. G., et al. (2009). Satellite observations indicate rapid warming trend for lakes in California and Nevada. *Geophysical Research Letters*, *36*.

Schott, J. R. (1986). The role of remotely sensed data in studies of the thermal bar. *Remote Sensing Reviews*, *2*, 341–358.

Schott, J. R., Barsi, J. A., Nordgren, B. L., Raqueño, N. G., & de Alwis, D. (2001). Calibration of Landsat thermal data and application to water resource studies. *Remote Sensing of Environment*, *78*, 108–117.

Schott, J. R., Brown, S. D., & Barsi, J. A. (2004). *Thermal remote sensing in land surface processes: Calibration of thermal infrared (TIR) sensors* (2nd edition). CRC-Press.

Schott, J. R., & Volchok, W. J. (1985 September). Thematic Dapper thermal infrared calibration. *Photogrammetric Engineering and Remote Sensing*, *51*(9), 1351–1357.

Schott, J. R., Volchok, W. J., & Biegel, J. D. (1987 January). *Radiometric analysis of the longwave infrared channel of the Thematic Mapper on Landsat 4 and 5, prepared for NASA/Goddard Space Flight Center, Greenbelt, MD, Report RIT/DIRS 86/87-51-112*.

Thenkabail, P. S., Biradar, C. M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., et al. (2009). Global Irrigated Area Map (GIAM), derived from remote sensing, for the end of the last millenniums. *International Journal of Remote Sensing*, *30*(14), 3679–3733.

Thenkabail, P. S., Hanjra, M., Dheeravath, V., & Gumma, M. (2010). A holistic view of global croplands and their water use for ensuring global food security in the 21st century through advanced remote sensing and non-remote sensing approaches. *Remote Sensing*, *2*(1), 211–261.

Tonoka, H., Palluconi, F. D., Hook, S. J., & Matsunagi, T. (2005). Vicarious calibration of ASTER thermal infrared bands. *Transactions on Geoscience and Remote Sensing*, 2733–2746.

Walton, C., Pichel, W., Sapper, J., & May, D. (1998). The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. *Journal of Geophysical Research*, *103*, 999–1028.