Article

Landsat-8 Thermal Infrared Sensor (TIRS) Vicarious Radiometric Calibration

Julia A. Barsi [1],*, John R. Schott [2], Simon J. Hook [3], Nina G. Raqueno [2], Brian L. Markham [4] and Robert G. Radocinski [3]

[1] Science Systems and Applications, Inc., NASA/GSFC Code 618, Greenbelt, MD 20771, USA
[2] Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Dr., Rochester, NY 14623, USA; E-Mails: schott@cis.rit.edu (J.R.S.); nina@cis.rit.edu (N.G.R.)
[3] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA; E-Mails: simon.j.hook@jpl.nasa.gov (S.J.H.); Robert.G.Radocinski@jpl.nasa.gov (R.G.R.)
[4] Biospheric Sciences Branch, Code 618, NASA/GSFC, Greenbelt, MD 20771, USA; E-Mail: Brian.L.Markham@nasa.gov

* Author to whom correspondence should be addressed; E-Mail: julia.barsi@nasa.gov; Tel.: +1-301-614-6667; Fax: +1-301-614-6695.

Abstract: Launched in February 2013, the Landsat-8 carries on-board the Thermal Infrared Sensor (TIRS), a two-band thermal pushbroom imager, to maintain the thermal imaging capability of the Landsat program. The TIRS bands are centered at roughly 10.9 and 12 μm (Bands 10 and 11 respectively). They have 100 m spatial resolution and image coincidently with the Operational Land Imager (OLI), also on-board Landsat-8. The TIRS instrument has an internal calibration system consisting of a variable temperature blackbody and a special viewport with which it can see deep space; a two point calibration can be performed twice an orbit. Immediately after launch, a rigorous vicarious calibration program was started to validate the absolute calibration of the system. The two vicarious calibration teams, NASA/Jet Propulsion Laboratory (JPL) and the Rochester Institute of Technology (RIT), both make use of buoys deployed on large water bodies as the primary monitoring technique. RIT took advantage of cross-calibration opportunity soon after launch when Landsat-8 and Landsat-7 were imaging the same targets within a few minutes of each other to perform a validation of the absolute calibration. Terra MODIS is also being used for regular monitoring

of the TIRS absolute calibration. The buoy initial results showed a large error in both bands, 0.29 and 0.51 W/m^2·sr·μm or −2.1 K and −4.4 K at 300 K in Band 10 and 11 respectively, where TIRS data was too hot. A calibration update was recommended for both bands to correct for a bias error and was implemented on 3 February 2014 in the USGS/EROS processing system, but the residual variability is still larger than desired for both bands (0.12 and 0.2 W/m^2·sr·μm or 0.87 and 1.67 K at 300 K). Additional work has uncovered the source of the calibration error: out-of-field stray light. While analysis continues to characterize the stray light contribution, the vicarious calibration work proceeds. The additional data have not changed the statistical assessment but indicate that the correction (particularly in band 11) is probably only valid for a subset of data. While the stray light effect is small enough in Band 10 to make the data useful across a wide array of applications, the effect in Band 11 is larger and the vicarious results suggest that Band 11 data should not be used where absolute calibration is required.

Keywords: Landsat-8; TIRS; thermal infrared sensor; thermal imaging; radiometric calibration

1. Introduction

Launched in February 2013, Landsat-8 is the latest in the series of Landsat satellites. It continues the 40+-year mission of acquiring global, moderate resolution images of the Earth's surface every 16 days. Unlike prior Landsat instruments where the thermal and reflective band images were acquired with the same sensor, the Landsat-8 satellite carries two imaging sensors, the Operational Land Imager (OLI) which images in the visible to short-wave infrared (0.4–2.5 μm) portion of the spectrum, and the Thermal Infrared Sensor (TIRS), which images in the thermal region (10–12.5 μm). Though the instruments are different, Landsat-8 maintains the swath width, scene framing, radiometric and geometric accuracy and precision and general spectral bandwidths of prior Landsat missions. Details of the OLI are covered in other papers [1–3]; this paper will focus only on TIRS and the vicarious calibration of the sensor.

The long term record provided by the Landsat thermal sensors has grown in importance as our understanding of how temperature drives many physical and biological processes that impact the global and local environment has grown. Studies of lake hydrology [4,5], evapotranspiration [6], regional water resources [7] and the impact of local climate trends all make use of Landsat derived thermal data. The value of the archive will continue to grow as more effective ways to study long-term thermal processes are developed.

In order to make use of the thermal data for long-term studies, the whole Landsat archive must be consistently calibrated. Teams have been monitoring the Landsat-5 Thematic Mapper (TM) and Landsat-7 Enhanced Thematic Mapper+ (ETM+) thermal calibration since 1999 and have made several updates to the calibration to correct for both errors in gain and bias [8]. The ETM+ thermal band is calibrated to within 0.48 K and the Landsat-5 TM thermal band to within 0.53 K (at 300 K). TIRS was rigorously characterized and calibrated pre-launch [9] and an on-board calibration system allows for continuous characterization now that it's on-orbit [10]. This paper addresses the methods used to validate the

on-orbit calibration using ground targets and other satellites, to confirm that the calibration is consistent with the Landsat historical record.

1.1. The TIRS Instrument

The TIRS instrument is a departure from prior Landsat thermal imagers in multiple ways: it is a pushbroom instrument rather than a whiskbroom; it has two spectral channels rather than one; and it has 100 m spatial resolution rather than 60 m (Table 1).

Table 1. TIRS Salient Characteristics. The pushbroom design allows for longer dwell time, thereby greatly improving the Noise Equivalent Delta Temperature (NEΔT) of TIRS as compared to ETM+ and TM. The TIRS results are based on a 1-min long acquisition of the blackbody set at 280 K. The ETM+ and TM results are based on shutter data [8]. Note that TIRS NEΔT is not typically given at 280 K but is provided here to be consistent with the other instruments. See [10] for the standard performance levels.

Instrument	Band (#)	Center Wavelength (µm)	Bandwidth (µm)	Spatial Resolution (m)	Active Detectors (#)	NEΔT (K@280K)
TIRS	10	10.9	0.6	100	1920 (3 × 640)	0.05
TIRS	11	12.0	1.0	100	1920	0.06
ETM+	6	11.3	2.0	60	8	H: 0.22 L: 0.28
TM	6	11.4	2.0	120	4	0.17–0.30

The instrument features a four-element refractive optics telescope, with three germanium (Ge) elements and one zinc selenide (ZnSe) element, which directs the incoming energy onto the focal plane (Figure 1). A flat mirror at the front of the telescope, the Scene Select Mechanism (SSM), switches the field of view between the earth and the two internal calibration positions for view of deep space and the on-board blackbody.

The pushbroom focal plane consists of three separate Sensor Chip Assemblies (SCAs), each 512 × 640 pixels. Figure 2 shows the layout of the chips on the focal plane; two SCAs are slightly offset from the third in the along-track direction by about 300 rows in Band 10 and 200 rows in Band 11. The SCAs overlap in the across-track direction by 28 pixels. The temperature of the focal plane is controlled by the cryocooler to ~40 K and maintained to within ±0.01 K [9]. The spectral interference filters lay on top of the SCAs, covering about 30 rows of the 512-row chip. Of these 30 rows, only one row per-band is read out to generate the standard image product. A second row per-band is designated as backup, in case a detector in the primary row fails. Several rows outside of the spectral filter are completely blocked from incoming energy and are used to characterize the internal instrument dark signal.

The spectral interference filters were designed to provide the optimal band combination for use in a split-window atmospheric correction algorithm [11]. Figure 3 shows the final band-average TIRS relative spectral responses (RSRs) for each band.

Figure 1. TIRS optical diagram. The SSM mirror rotates between the nadir viewport, blackbody and deep space viewport to provide calibration at least once an orbit.
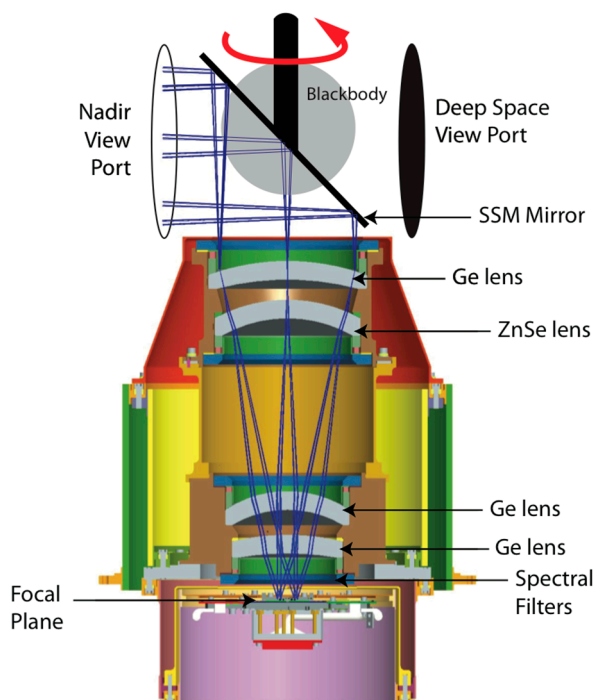

Figure 2. TIRS focal plane diagram. The areas of the three SCAs covered by the spectral filters are indicated, along with the region on each SCA used to measure the dark signal.
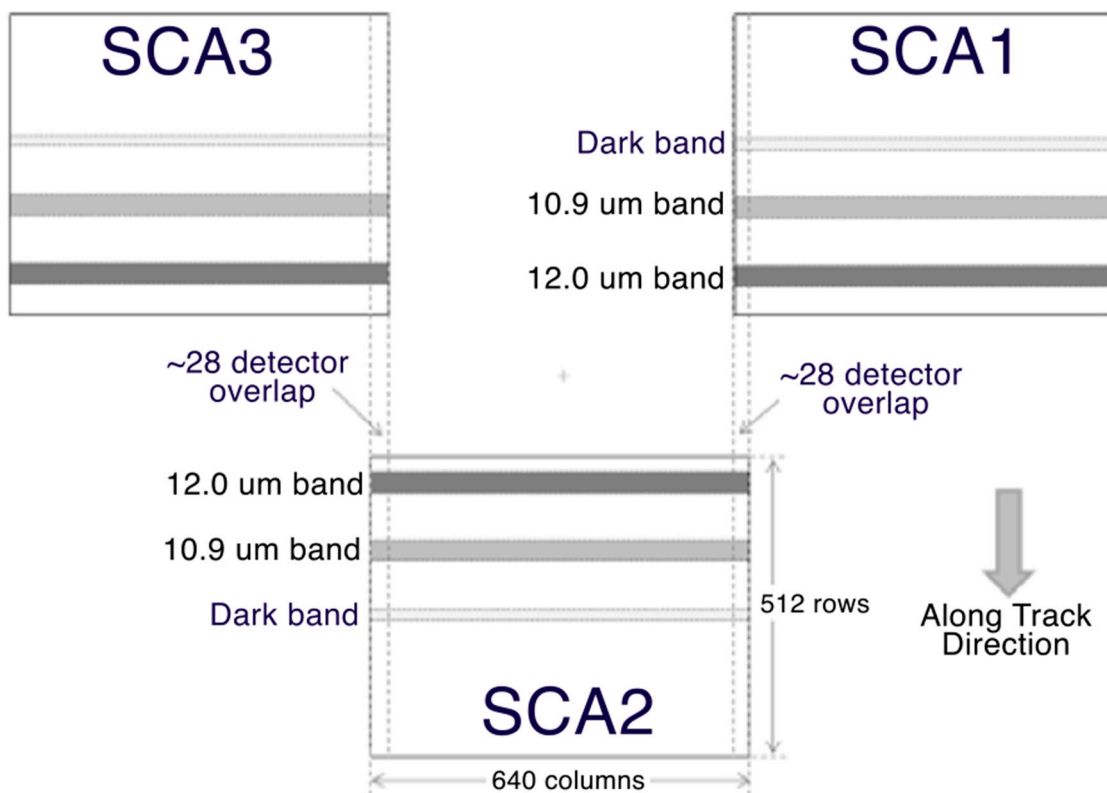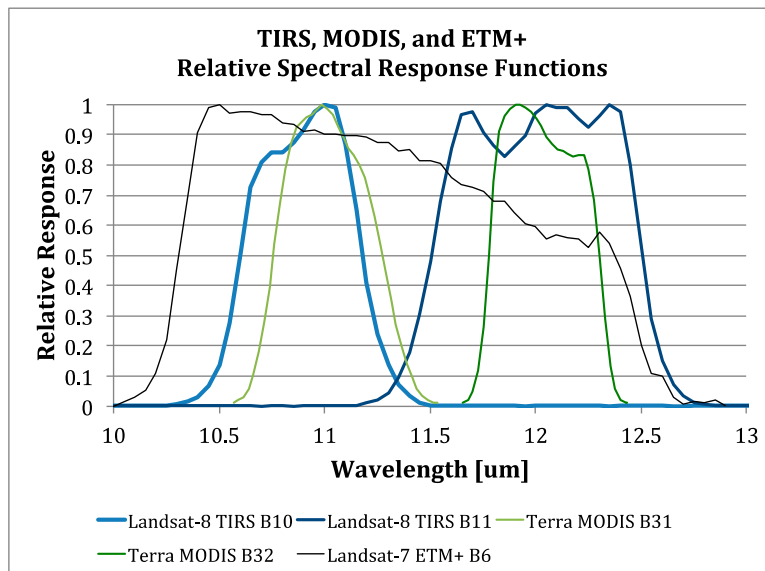
Figure 3. The relative spectral responses (RSR) of the TIRS bands (B10 and 11). Also shown for comparison are the RSRs of ETM+ band (B6) and the equivalent MODIS bands (B31 and 32). The TIRS and ETM+ RSRs are band-average but the MODIS RSRs are each for a specific detector (detector 5 in both cases).


1.2. Internal Calibrator

The internal calibration system consists of a variable temperature blackbody, a port by which the instrument can view deep space, and a Scene Select Mechanism (SSM) that allows the sensor to view the blackbody, deep space or the earth (see Figure 1). The blackbody and deep space views are acquired at the bottom of the descending pass and at the top of the ascending pass every orbit. During these opportunities, the SSM is moved so that the instrument captures a one-minute image of deep space. The mirror flips to the blackbody position for a one-minute image and then back to the deep space for another one minute image. The blackbody is nominally kept at a single temperature (295 K). Details of the on-board calibration methods can be found in [10]. The purpose of discussing them here is to demonstrate that the TIRS instrument appears to be stable based on the on-orbit internal calibrator results. A responsivity metric g, of the on-board stability is calculated from the response to the blackbody and deep space for individual calibration sequences:

$$g = \frac{(Q_{BB} - Q_0)}{(L_{BB} - L_{space})} \tag{1}$$

where Q_{BB} is the bias-subtracted, linearized digital counts extracted from the blackbody image, Q_0 is the instrument offset which incorporates instrument and electronic biases, L_{BB} is the spectral radiance of the blackbody as converted from the monitor thermistor readout and L_{space} is the spectral radiance of deep space, assuming a 4 K background. The metric is calculated per-SCA since all detectors on a single SCA share electronics. Note that this is not the actual gain equation. Instrument gain and bias are covered in detail in [12]. Since launch, the SCA-average metric may have a slowly decreasing trend (maximum of −0.28%/year ± 0.005%/year in the worst case) but the total variability over the lifetime is still only 0.08% (1σ) for the worst case (Figure 4).

Other metrics, covered in detail in [10], also show both bands of the TIRS instrument to be stable: over 36 min the background signal is stable to within about 0.01 W/m²·sr·µm (1σ) and over 36 min the gain is stable to within 0.1% (1σ).

Figure 4. The SCA-average per-calibration sequence responsivity metric for both TIRS bands along with the per-SCA lifetime average. Based on this and other metrics, the TIRS instrument is internally stable.
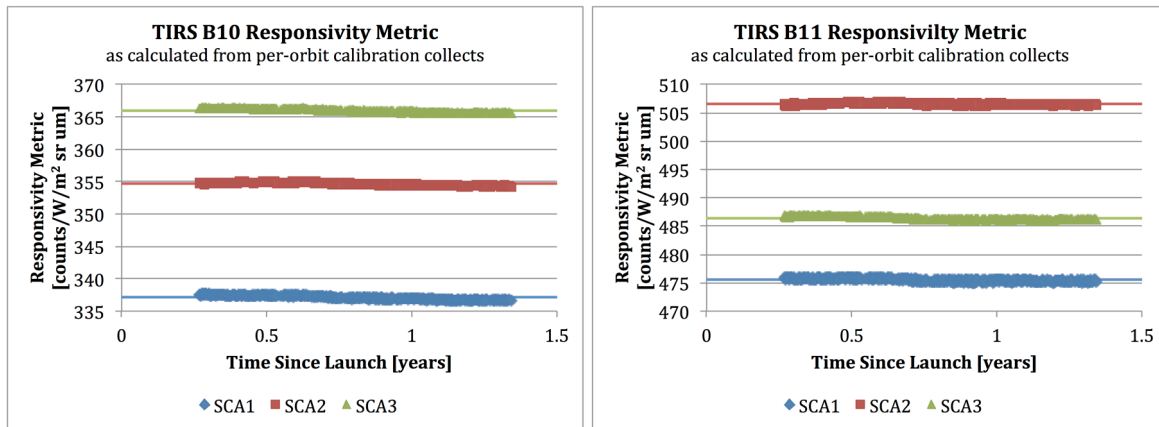
2. Vicarious Calibration Approaches

Water has long been used as the primary target for vicarious calibration of the Landsat thermal bands: it is uniform in composition, has a high and known emissivity and often exhibits low surface temperature variation (less than 1 °C) over large areas. Land targets can provide a higher range of temperatures but they are generally more difficult to characterize. Vicarious calibration is performed by teams at the NASA/Jet Propulsion Laboratory (JPL) and the Rochester Institute of Technology (RIT). They do their work on various large water bodies, over a range of temperature from about 4 °C to 35 °C.

Two different methods have been used to support the calibration of TIRS. Each will be introduced here.

2.1. Buoy Methods

The governing equation for radiation propagation from the Earth's surface to the sensor can be expressed as

$$\hat{L}_{TOA\lambda} = \frac{\int \left[\left(\varepsilon(\lambda)L_{T\lambda} + (1 - \varepsilon(\lambda))L_{d\lambda}\right)\tau(\lambda) + L_{u\lambda}\right]R(\lambda)d\lambda}{\int R(\lambda)d\lambda} \tag{2}$$

where $\hat{L}_{TOA\lambda}$ is the predicted top-of-atmosphere (TOA) reaching radiance, 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 downwelling radiance, $L_{u\lambda}$ is the spectral upwelling radiance, $\tau(\lambda)$ is the transmission from the target to the sensor, and $R(\lambda)$ is the relative spectral response of the band. All terms are a function of wavelength (λ). For the Landsat bandpasses, Equation (2) can be approximated as

$$\hat{L}_{TOA\lambda} = \varepsilon\hat{L}_{T\lambda}\tau + \hat{L}_{u\lambda} + \hat{L}_{d\lambda}(1 - \varepsilon)\tau \tag{3}$$

where all terms are integrated over the appropriate spectral response. The surface-leaving radiance, $\varepsilon\hat{L}_{T\lambda}$, is the effective spectral radiance in the Landsat band observed at the ground. The emissivity for water is

essentially constant over the Landsat bandpasses. The upwelling and downwelling radiances, along with the atmospheric transmission, can be estimated using the radiation propagation code, MODTRAN [13], given knowledge of the atmosphere. Local atmospheric data are available in radiosonde collections or from assimilated weather products. Both the RIT and JPL teams make use of a buoy technique for validating the calibration of the TIRS instrument, but the methods are slightly different and are described below.

The estimated TOA radiance as predicted from the surface measurement can be compared to the radiance measured by the TIRS instrument to provide a validation of the absolute calibration. By building up a long history of cloud-free vicarious calibration measurements, trends over time and/or surface temperature can emerge.

2.1.1. JPL Lake Tahoe and Salton Sea

The JPL has operated four instrumented buoys on Lake Tahoe on the California/Nevada boarder since 1999 [14] and a similarly instrumented platform on the Salton Sea in Southern California since 2008 for the purpose of thermal calibration. The high altitude Lake Tahoe is an ideal thermal calibration target; there is little atmosphere above the lake, the lake is extremely deep so it does not freeze in the winter and it has an annual temperature range from about 4 °C to 20 °C. The Salton Sea is a less ideal target because its surface is below sea level and the atmosphere is generally quite thick. But the water can get as hot as 35 °C in the summer, so it extends the range of temperatures over which calibration can be performed.

The instrumentation on each platform includes near surface contact thermistors, near-nadir viewing calibrated radiometers and weather stations. The 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 bandpass [15–17]. Because the radiometers are not filtered to match the Landsat spectral bandpass, 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 downwelling radiance computed from MODTRAN [14,15].

Data from the buoys are acquired every 2 to 5 min and transmitted to JPL for processing. The output from the processing system 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 (as in Equation (2)). 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 [17]. The uncertainty in the modeled radiance is within 0.41 K and this method has been used to calibrate Landsat-7 ETM+ thermal band to within 0.48 K and the Landsat-5 TM thermal band to within 0.73 K [8].

Lake Tahoe and the Salton Sea are acquired every opportunity with TIRS during the day passes as part of the standard Landsat-8 acquisition strategy. Starting immediately after launch, special requests were made to acquire the two water bodies at night to increase the number of images available for vicarious calibration. After a few months on-orbit, special night pointing acquisitions were scheduled to view the water bodies from off-nadir, with the spacecraft being rotated to view the lakes from one path over. This increased the number of images available for calibration. For these acquisitions, the spacecraft was pointed such that the lake appeared in the center of the image, falling in SCA2, regardless of the pointing

angle. As the lakes naturally fall in SCA2 during the day acquisitions, the JPL data are heavily concentrated in SCA2 (Figure 5).

Figure 5. Distribution of vicarious calibration data across the TIRS focal plane. Dashed red lines indicate the boundaries between SCAs; SCA1 consists of detectors 1–640, SCA2 contains detectors 641–1280, and SCA3 is detectors 1281–1920. Note that all the JPL day data since April 2013 falls in SCA2 and for most of the night acquisitions, the satellite has been pointed such that Tahoe falls in SCA2. The RIT acquisitions are based on eight different buoys and are distributed across the focal plane.
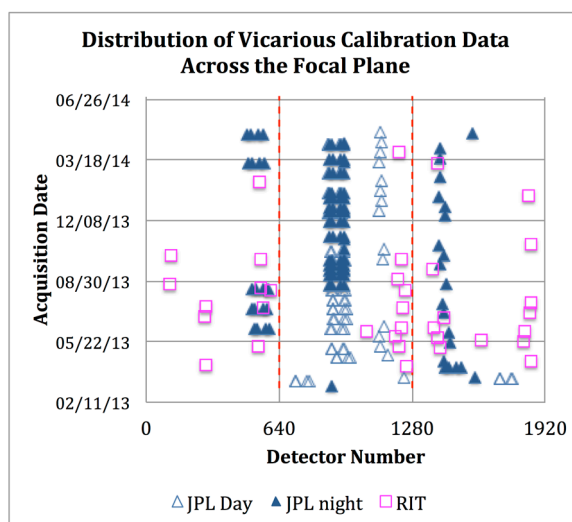
2.1.2. RIT NOAA Ocean and Great Lakes

The RIT team makes use of the fleet of moored buoys operated by the National Oceanic and Atmospheric Administration (NOAA), which are distributed in open water around the United States [8]. To date, RIT has made use of data originating from buoys in the Great Lakes, the Atlantic and Pacific Oceans and the Gulf of Mexico. With the variation in location and season, the temperature ranges from about 3 to 30 °C. While not providing as consistent a dataset as Lake Tahoe and the Salton Sea, the sheer number of buoys available to work with means that loss of precision due to varying targets can be reduced by increased numbers of measurements. The NOAA buoy method has been found to be nearly as accurate as the JPL buoy method, at 0.46 K [8].

When operational, each buoy in the network records hourly subsurface temperatures (0.6 m or 1.5 m) as well as weather data, and archives it in the National Data Buoy Center (NDBC). The NDBC database can be queried to access the recorded temperatures and meteorological data. Because the buoys do not make measurements of skin temperature, a correction needs to be made to estimate the surface-leaving radiance based on the subsurface temperature. Using 24 h of temperature measurements before the satellite overpass along with meteorological data, the surface temperature can be estimated from the subsurface values [18]. The method accounts for the diurnal cycle, the temporal phase shift in the diurnal cycle with depth, thermal gradients with depth that are a function of wind speed and the cool skin effect [8]. The derived surface temperature is used along with emissivity, local weather data, and MODTRAN to estimate sensor-reaching radiance as in Equation (2). This method has been used to

calibrate Landsat-7 ETM+ thermal band to within 0.59 K and the Landsat-5 TM thermal band to within 0.60 K [8].

The buoy data are acquired every opportunity during the day passes, primarily as part of the standard Landsat-8 acquisition strategy. They are scattered throughout the coastal waters of the United States and they are distributed across the TIRS focal plane. Figure 5 shows the position of each buoy on the focal plane for every cloud-free acquisition used for vicarious calibration analysis.

2.2. Inter-Satellite Top-of-Atmosphere Comparison

Many instruments are making measurements in the thermal region and if they are acquired close enough in time, a calibrated sensor can be used to monitor another sensor's calibration. However, since the bands being compared between the instruments rarely have the same relative spectral response (RSR) functions, this method provides a means for monitoring changes in behavior over time between instruments rather than absolute calibration. TIRS benefited from a special collect just after launch when Landsat-8 under flew Landsat-7 on its way to its permanent orbit. During this maneuver, ETM+ and TIRS acquired near-coincident data for three days.

In its permanent orbit, Terra/MODIS is 8-days offset from Landsat-8 with its nadir view but the wide swath width of MODIS means that it can view the same targets as TIRS nadir view within about 30 min. Over large water bodies the diurnal warming in this short time period is typically small and at night can be very small. This provides an opportunity to compare the measurements from the instruments. Nonetheless it is not the same as the buoy radiometer measurements, which are made within a few minutes of the satellite overpass. A permutation of this approach that is being explored but is not presented in this study involves using a split-window algorithm to calculate the surface skin temperature with MODIS and then propagate this surface skin temperature to the at-sensor radiance with MODTRAN and convolve the result to the RSR of Landsat or any other sensor. This approach enables the absolute calibration to be evaluated though it does depend on the accuracy of the satellite-derived skin temperature used in the forward calculation.

The three sensors, TIRS, ETM+ and MODIS, cover the same spectral regions but don't have identical RSRs (Figure 3). Sensor reaching radiance can be converted to top-of-atmosphere brightness temperature using the Planck function.

A study simulating the difference between spectral radiance and brightness temperature was performed to verify the ability to compare data between two different calibrated sensors. Using MODTRAN to perform the radiometric propagation, four North American Regional Reanalysis (NARR) atmospheres [19] were processed to test the difference between brightness temperature for ETM+ or MODIS and the TIRS bands. The NARR atmospheres were selected to cover a wide range of atmospheric conditions, from hot and dry to cold and wet. When compared in terms of apparent temperature, Landsat-7 Band 6 temperatures and Landsat-8 Band 10 or 11 temperatures over water should agree to within ±0.5 K for most realistic conditions and within ±1 K for extreme conditions (e.g., very warm moist air over very cold water) (Figure 6). The agreements are better between MODIS Band 31 and TIRS Band 10 and MODIS Band 32 and TIRS Band 11, where even the worst cases should be within ±0.5 K.

Figure 6. The estimated root-mean-squared error (RMSE) error in comparing the ETM+ (**left**) or MODIS (**right**) to TIRS brightness temperatures over a range of surface temperatures and through a variety of atmosphere types. The largest errors are due to less likely conditions, *i.e.*, a very warm surface in a very cold atmosphere.
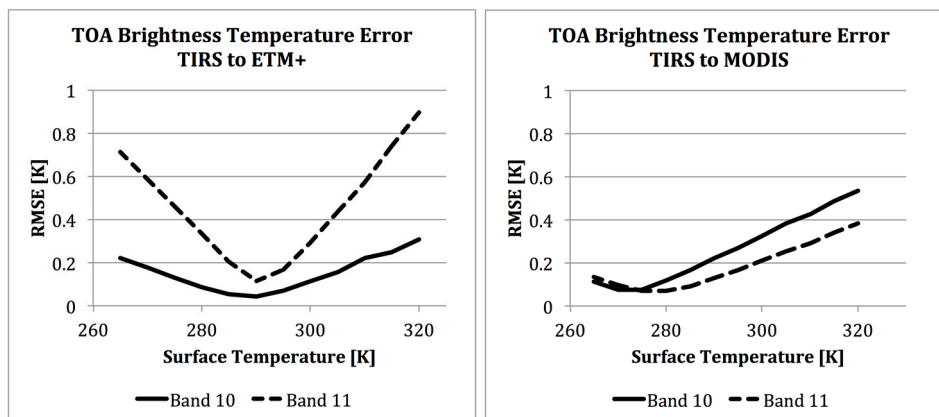
2.2.1. JPL MODIS and Landsat-8

The MODIS instrument aboard the Terra spacecraft has a much wider field of view than the Landsat instruments, thus, even though the nadir passes of Landsat-8 and Terra are eight days offset from each other, MODIS can see Lake Tahoe and the Salton Sea on the same days as TIRS. The overpasses are generally acquired within 20 to 35 min of each other. Comparisons have been made for every cloud-free pair of images, a total of 63 pairs with MODIS band 31 compared to Landat-8 band 10 and MODIS band 32 compared to Landsat-8 band 11.

2.2.2. RIT Landsat-7 and Landsat-8

During the early weeks of the Landsat-8 mission, while the satellite was being maneuvered into its permanent orbital location, Landsat-8 was in roughly the same orbit as Landsat-7 for three days, 29–31 March 2013. The two instruments acquired earth images nearly simultaneously. The time difference shifted over the three days but in general, the images were acquired within 2 and 20 min of each other. For well-mixed water bodies, that is sufficiently close to compare the brightness temperatures from the two instruments.

On 30 March 2013, Landsat-7 and -8 acquired data in a single pass, from the Hudson Bay to the Gulf of Mexico. Water temperatures ranged from −3 to 17 °C. The time difference between the two sets of images is 2.5 min. Forty-eight regions were extracted for comparison from the two instruments' imagery.

3. Stray Light Effect on Imagery

Soon after launch, it was clear from the vicarious calibration by JPL and RIT that there was a significant calibration error (see Section 4.1). Multiple investigations, including detailed reassessment of pre-launch calibration equipment, lunar scans and optics modeling, determined that the error was due to radiance entering the telescope from far out-of-field. Research continues on the source of and correction for this stray light effect, but it has a significant effect on image data, including the data to be

used for vicarious calibration. This intent of this section is to detail the effect of the stray light on the vicarious calibration image data.

Using special scans of the moon, it has been shown that energy is reaching the focal plane from a ring of about 15° outside the center of the field of view. Because the source of the stray light is unique for each detector and significantly different between SCAs, the location of the calibration target on the focal plane and what is in the 15° ring outside the image frame (surface type and/or cloud) makes a difference to the calibration results. The current status of the stray light investigation from an instrument perspective is covered in [20]. This section describes how the stray light effect was identified and the observed magnitude on the on-orbit calibration validation.

The first hint that there was a problem with TIRS was in the discontinuity between SCAs over what should be a uniform target. It was clear from the images that the overlap between adjacent SCAs was not smooth, manifesting itself as a discrete step in the image (Figure 7). Also, the difference appeared to change over time, as the satellite traveled though space. In Figure 7, this is apparent as at the top of the lake, SCA3 is warmer than SCA2 but further south in the image, closer to the peninsula, the contrast flips and SCA2 is warmer.

Figure 7. TIRS Band 11 image of Lake Superior (47.5N, 88W) illustrating the discontinuities between the Sensor Chip Assemblies (SCAs) and time-varying nature of the difference. The edges of the SCAs are clearly defined (red arrows) and the differences between the SCAs change from north to south in the image. In a stable system, even with a calibration error, the differences between the SCAs should remain constant for the length of the lake. However, in this example, SCA3 is warmer than SCA2 by 0.2 K at the region marked 1 but is cooler by 0.8 K at region 2. SCA1 is warmer than SCA2 by 0.2 K at region 3 but cooler than SCA2 by 0.7 K at region 4.
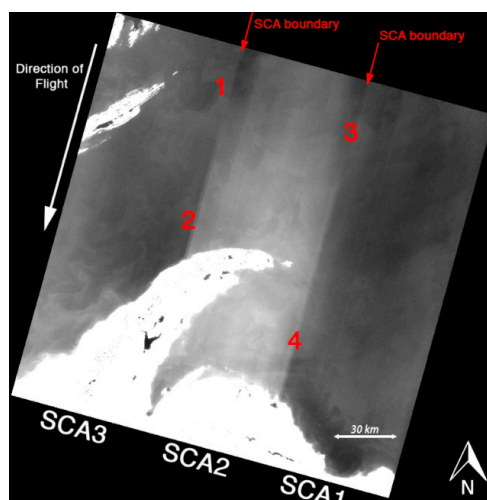
The extent of the stray light contribution was defined by lunar scans so the source of the stray light in the vicarious calibration imagery can at least be examined. Lake Tahoe and the Salton Sea are both relatively small water bodies, surrounded by land. The out-of-field contribution for these targets comes in large part from the arid desert regions of Nevada and southern California (Figure 8). For many of the RIT sites, the buoys are largely surrounded by water, though some, as in Figure 7, are inland buoys.

The stray light will thus originate from a combination of land and water. This makes for a significantly different out-of-field radiance contribution than for the Tahoe and Salton Sea images.

While the geographic extent of the stray light contribution is known, the total radiance contribution is not. Work is on-going to develop a technique to estimate the out-of-field radiance but none have been implemented yet. Thus the current vicarious calibration results do not account for any knowledge of the external source of energy.

Figure 8. The 185 km wide scene boundaries of the standard Lake Tahoe, located at 39N, 120W (**left**), and Salton Sea, located at 33.3N, 115.8W (**right**), image frames are indicated by the green box. The blue circle indicates the 15° ring source of stray light (though the stray light does not necessarily come from the whole circle). In both cases, the source of the stray light is primarily from land (given that no snow or clouds are covering the surface) outside the area observed by Landsat-8.

4. Vicarious Calibration Results

Since the knowledge and understanding of the calibration error continues to evolve, this section is presented in chronological order, the order in which we responded to the analyses. In that way, this paper explains the logic by which decisions were made. The first section covers the initial look at the calibration, in the first eight months after launch, when a stray light problem had been hypothesized but no studies had yet been done. As work progressed, both on the stray light study and the vicarious calibration, more detailed analyses could be done. The second section details the understanding of the vicarious calibration as the dataset grew and how the stray light assessment complements the vicarious calibration results.

4.1. Initial Results

Starting with the very first vicarious calibration campaigns, there was a hint that there was a problem with the absolute calibration results. The ETM+/TIRS cross-calibration results from March 2013 were showing average bias errors of −1.84 and −1.94 °C for bands 10 and 11, respectively. Over the first two months on-orbit, the bias error determined by the Lake Tahoe and Salton Sea appeared to be growing. And although the instrument appeared to be internally stable [10], there was odd structure visible in the earth imagery that could not be explained by a stable imager (see Figure 7). Investigations continued and thanks to special TIRS scans of the moon, the source of the problems could be traced to out-of-field stray light [20].

While the research was ongoing to determine the per-detector source and magnitude of the stray light, the decision was made to make an initial correction to the calibration to account for the additional radiance impinging on the focal plane. Regardless of the actual source of the stray light, it was apparent from the vicarious calibration that the instrument was predicting too high; the out-of-field energy made the surface appear warmer than it was. The vicarious results were much noisier than was expected as well, which is the result of not knowing the source of the stray light.

The initial calibration error was calculated in November 2013, based on all the available daytime buoy data from JPL and RIT. The TOA-predicted radiance (vicarious radiance) was compared to the TIRS estimated radiance (image radiance) point for point. The data were spread amongst all three SCAs, though all of the JPL data falls on SCA2. All data from both teams and the three SCAs were treated as one data set and the slope and offset were assessed to determine the calibration error (Figure 9). The slope of the trend was not statistically significant, indicating that there was not likely an error in the calibration gain. However, the data were all above the 1:1 line, indicating that the instrument is predicting a radiance that is too high. This bias error was calculated as the average difference between the vicarious radiance and the TIRS image radiance for all data points for each band.

Figure 9. The initial vicarious calibration results for both TIRS bands, based on the day JPL and RIT buoy data. If the instrument were perfectly calibrated, the data would fall scattered about the 1:1 line. All results for both bands are above the 1:1 line indicating that the instrument is predicting too high.
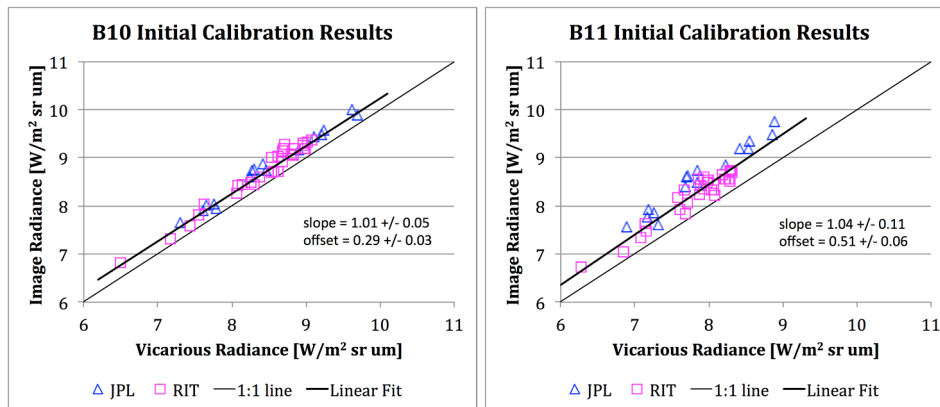

Figure 10 and Table 2 show the calculated calibration error for just the JPL data, broken down by day and night, for all data acquired before November 2013. This difference between day and night acquisitions is also thought to be an effect of the stray light; the solar loading results in the temperature difference between our water targets and the land being greater during the day than at night. Since there is a statistically significant difference between night and day populations and Landsat primarily acquires data during the day, it was decided to determine the bias error from just the day population of data.

As of November 2013, the calibration error appeared to be strictly an error in bias; the data did not indicate a statistically significant error in gain. Table 3 shows the calculated calibration error as well as the variability in the buoy data results.

Figure 10. The initial vicarious calibration results for Band 10, based on the JPL data only, displayed as difference between the predicted vicarious radiance minus the image radiance. The data are split into the day and night series. There is a statistically significant difference between the day and night results so only the results for the day data were used to calculate the bias error.
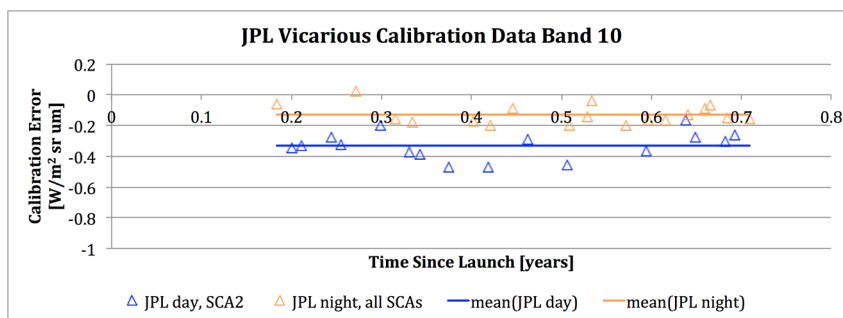

Table 2. Calibration error for both TIRS bands from the JPL data, spilt into day and night acquisitions. There is a statistically significant difference between the day and night results.

Band Number	# Acquisitions	Calibration Error ± 95% Confidence (W/m² sr μm)	
		Band 10	Band 11
Day	16	−0.33 ± 0.05	−0.71 ± 0.08
Night	18	−0.13 ± 0.03	−0.27 ± 0.04

Table 3. TIRS Bias Correction, where the TIRS measurement was too hot, and the RMS variability in the bias, implemented in the USGS/EROS processing system on 2 February 2014.

TIRS Band	Bias Correction ± 1σ (W/m² sr μm)	Variability (K at 300 K)
10	−0.29 ± 0.12	−2.1 ± 0.80
11	−0.51 ± 0.20	−4.4 ± 1.75

On 2 February 2014, the calibration parameters in the USGS/EROS Landsat-8 processing system were changed to account for the bias error. The correction is a constant for each band, which does not account for SCA-to-SCA differences much less detector-to-detector differences. The bias correction adjusts for an average stray light contribution, regardless of season, location or clouds. This was implemented with the knowledge that work was ongoing to characterize and model the stray light effect, but in the hopes that the data would be incrementally better while waiting for a more appropriate correction algorithm.

4.2. Current Status

While the investigation into the stray light is ongoing, the vicarious calibration data collection continues. Buoy data are acquired at every opportunity, increasing the number of points in the dataset and increasing the confidence in the results. The vicarious data now cover all four seasons, which begins to indicate a flaw in the use of a constant bias correction for all conditions.

All the RIT and JPL data used to generate the bias correction were reprocessed with the updated calibration parameters (47 points) and new data continues to be collected so the updated dataset includes

63 points. The residual error is the average difference between the vicarious radiance and the image radiance for the reprocessed dataset.

The comparison with Landsat-7 was updated to account for the calibration change and an analysis comparing to Terra/MODIS was added. The data in this section have all been processed with the updated calibration parameters.

4.2.1. Seasonal Calibration Error

The calibration correction that was implemented in February 2014 was based on data acquired between April and October 2013, just seven months of data and all in the Northern Hemisphere. Since the error has been attributed to out-of-field radiance, the out-of-field radiance will generally be cooler in the Northern Hemisphere winter than it is in the Northern Hemisphere summer. The calibration correction that was implemented is a constant for all scenes thus it will over-correct scenes where the surrounding is cooler than the April through October average. This is apparent in the Band 11 JPL data.

Figure 11 shows the residual bias error over the year for the JPL day data. The out-of-field surround is very consistent for both the JPL targets: the dataset only includes Lake Tahoe and Salton Sea data and the water bodies are consistently on the same place on the focal plane. Therefore, the surfaces from which the out-of-field contribution is originating are always the same (barring the presence of clouds or snow).

Figure 11. Seasonal effect of the residual bias error for the Band 11 JPL data (includes both Lake Tahoe and Salton Sea). Data are plotted versus day of year so the change in the residual error over the year is apparent.
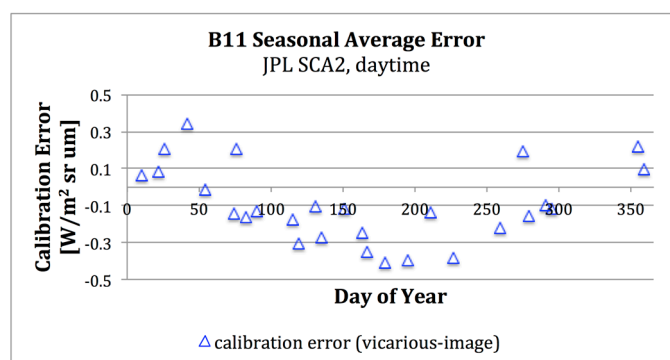

The same seasonal pattern is not currently apparent in the RIT data, but that is likely due to the larger variation in location of targets. The out-of-field contribution for each buoy originates from a different surround and there are not enough points from any one location to see the seasonal trend.

4.2.2. Inter-Satellite Comparison

The RIT ETM+ to TIRS comparison relies on data from a single day early in the mission when the instrument hadn't reached its final operating conditions yet. Additionally, the images were from March, outside of the time period over which the bias error was determined. The 48 data points from the ETM+ to TIRS comparison, which ranged in temperature from -3 to $28\,°C$, have a mean error of $0.78\ \text{W/m}^2\!\cdot\!\text{sr}\!\cdot\!\mu\text{m}$ in Band 10 and $2.85\ \text{W/m}^2\ \text{sr}\ \mu\text{m}$ in Band 11. This error falls within the range of the RIT buoy results.

Figure 12 shows the residual error for the buoy data over day of year. The ETM+ comparison has been added as a single point and it falls within the distribution of data for that time of year.

Figure 12. Seasonal effect of the residual bias error for the RIT data, including the average error for the ETM+ comparison. Data are plotted versus day of year but the seasonal effect is not as apparent in the RIT data as in the JPL data (Figure 11). The ETM+ comparison data point sits within the residual errors of the buoy data.
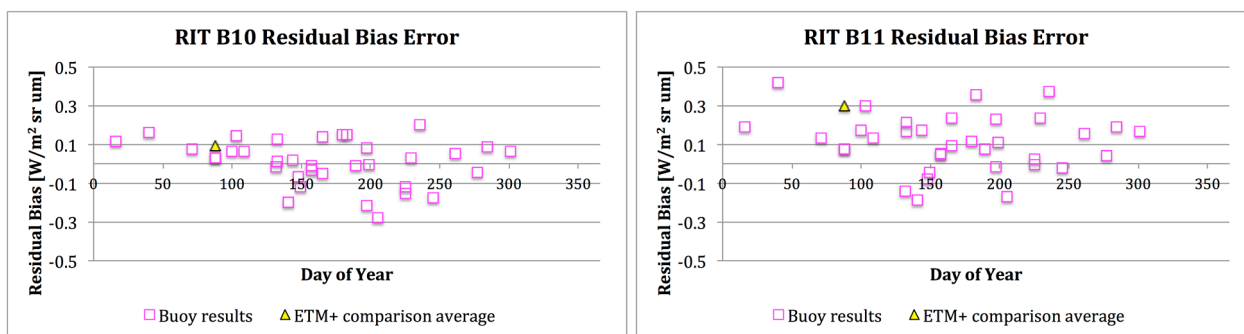
The JPL MODIS to TIRS comparison illustrates how the biases are different day and night, how there is more scatter in the day than night data and how there is more scatter with Landsat-8 band 11 than with Landsat-8 band 10. The fact that there is a larger scatter than with the buoy comparisons can be attributed to mismatch in the acquisition times between the two sensors (Figure 13) and highlights why these sensor-to-sensor comparisons do not provide a substitute for validation against in-situ data. As noted in the introduction, the next step in this analysis will be to derive the surface skin temperature from the MODIS data and use that in a similar manner to the way the buoy data are used to remove the RSR differences.

Figure 13. JPL buoy and MODIS comparison results plotted versus target brightness temperature. The trends do not overlap but they both indicate that the residual error is not dependent on target temperature. The trend in the Band 11 data is a function of the seasonal effect of the stray light.
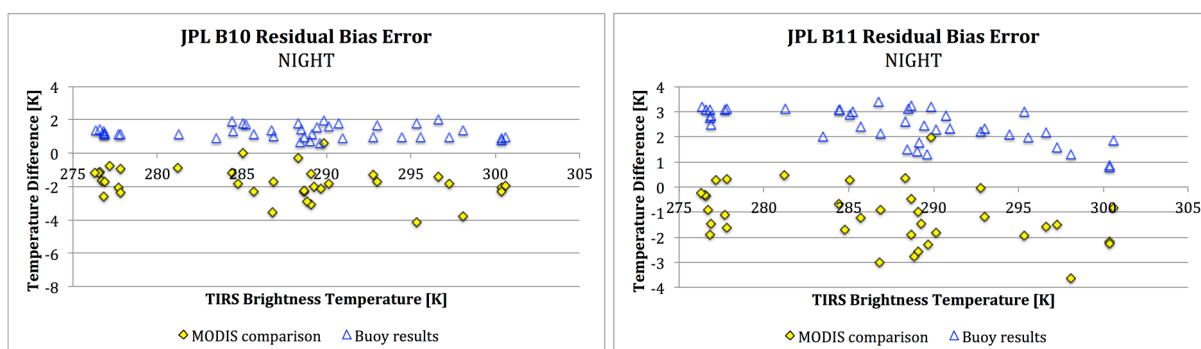
4.2.3. Current Residual Bias Error

The additional data since November 2013 has allowed for more statistical analysis and confidence in the buoy vicarious calibration results. Even given the understanding that the implemented bias correction was more applicable for the Northern Hemisphere growing season, the new data do not change the statistical results. The bias error has been removed by the updated calibration coefficients and processing

on the collected dataset now does not have a statistically significant bias error (Figures 14 and 15 and Table 4).

The night data continue to be statistically different than the day data but help to illustrate how stable the sensor is under stable conditions. In the absence of solar loading, the RMS variability is as low as 0.43 K (Band 10) and 0.66 K (Band 11) suggesting that if a model can be developed to account for the stray light, the day data RMS error could be reduced to that level or better.

Figure 14. Current vicarious calibration results for the two TIRS bands, including both RIT and JPL data for all SCAs, but only displaying day data. The data are scattered about the 1:1 line, indicating that the residual error has been removed. Neither the slope nor the offset is statistically significant.
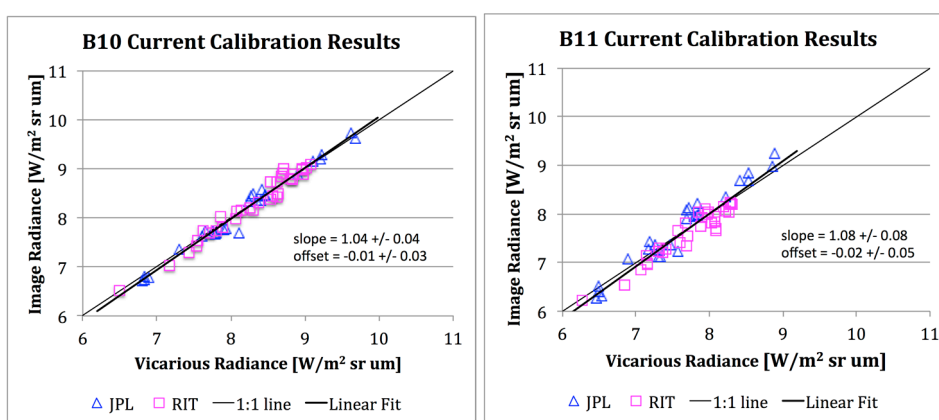

Figure 15. Current vicarious calibration results for the two TIRS bands over time. The residual difference between the vicarious radiance and the image radiance is scattered about the zero axis.
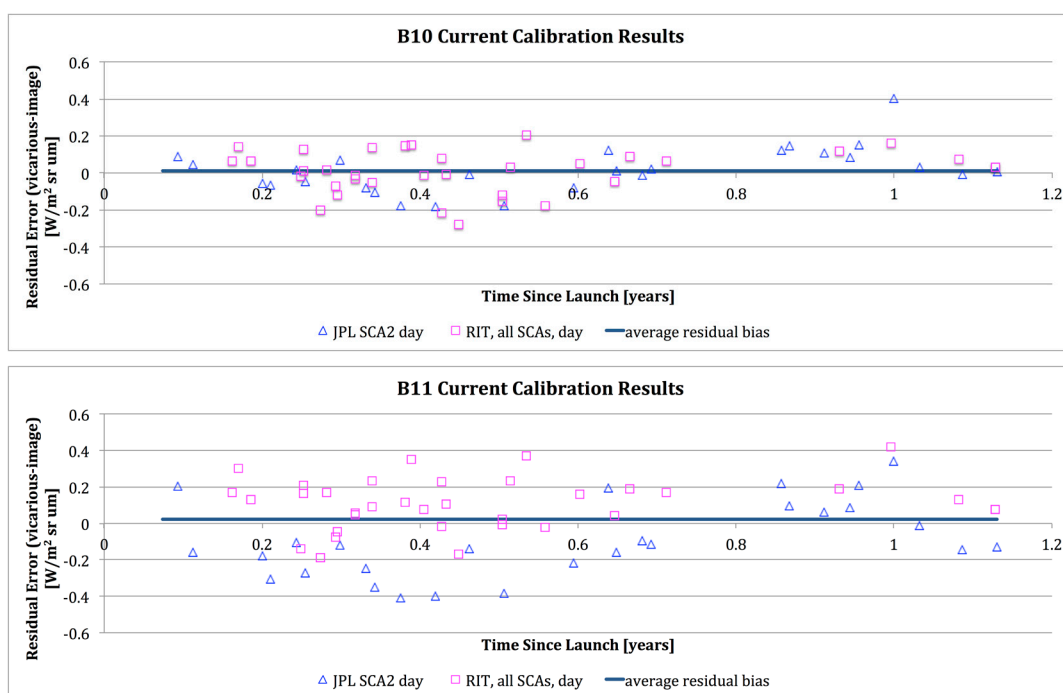
Table 4. Current TIRS calibration status, the residual bias error and the RMS variability for the day data after all vicarious calibration imagery was reprocessed with the corrected calibration parameters.

Band Number	# Acquisitions	Calibration Error ± 95% Confidence		RMS Variability (1σ)	
		(W/m^2 sr μm)	(K at 300 K)	(W/m^2 sr μm)	(K at 300 K)
Band 10	63	−0.01 ± 0.03	−0.09 ± 0.21	0.12	0.87
Band 11	63	−0.02 ± 0.05	−0.20 ± 0.42	0.20	1.67

5. Conclusions

The TIRS instrument has proven itself to be internally stable, based on the on-board calibration results, though the vicarious calibration results have revealed instability in the system calibration (particularly for band 11). This result highlights the necessity for vicarious calibration of all space borne sensors; without the vicarious calibration the blackbody data would suggest the instrument was well calibrated. The presence of stray light in the instrument means that the pre-launch calibration did not appropriately characterize the radiometric calibration and highlights the need to characterize stray light. Characterizing stray light with ground measurements is challenging and this instrument design is particularly susceptible.

In order to minimize the calibration error induced by the stray light effect in the imagery, an update was made to the calibration parameters in the processing system. On 3 February 2014, the USGS/EROS implemented a bias correction for both TIRS thermal bands and reprocessed all imagery so that users downloading data after that date would receive only data processed with the updated calibration. Users can check the metadata file to see when their data was processed to ensure they are working with data processed with the latest calibration. The FILE_DATE field indicates the date on which the image data were processed.

The vicarious calibration team continues to analyze data and refine the data set. The buoy datasets and the comparison with ETM+ and MODIS show the bias correction has improved the calibration, but leaves a seasonal error and is probably only valid for the Northern Hemisphere summer. Though the calibration won't truly be correct until the stray light contribution is removed, the TIRS Band 10 calibration is within ±0.12 W/m^2 sr μm (0.87 K) and Band 11 within ±0.20 W/m^2·sr·μm (1.67 K). While this is a larger error than was available for the Landsat-7 ETM+ (0.48 K), the hope is that Band 10 is still usable for most applications as a single band for thermometry while work is underway to improve the calibration of both bands. The cause for the larger bias and scatter in the Landsat-8 band 11 data is still under investigation.

Author Contributions

Julia Barsi was the primary author for this text. John Schott and Simon Hook contributed with details on the analysis specific to their team's techniques and results. Nina Raqueno and Robert Radocinski did the bulk of the data reduction for the vicarious teams. Brian Markham shaped the discussion and provided valuable direction in his role as Landsat Calibration Scientist.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Knight, E.J.; Kvaran, G. Landsat-8 Operational Land Imager design, characterization, and performance. *Remote Sens.* **2014**, *6*, 10286–10305.
2. Barsi, J.A.; Lee, K.; Markham, B.L.; Kvaran, G.; Pedelty, J.A. The spectral response of the Landsat-8 Operational Land Imager. *Remote Sens.* **2014**, *6*, 10232–10251.
3. Markham, B.L.; Barsi, J.A.; Kvaran, G.; Ong, L.; Kaita, E.; Biggar, S.; Czapla-Myers, J.; Mishra, N.; Helder, D.L. Landsat-8 Operational Land Imager radiometric calibration and stability. *Remote Sens.* **2014**, in press.
4. Schott, J.R.; Barsi, J.A.; Nordgren, B.L.; Raqueno, N.G.; de Alwis, D. Calibration of Landsat thermal data and application to water resource studies. *Remote Sens. Environ.* **2001**, *78*, 108–117.
5. Steissberg, T.E.; Hook, S.J.; Schladow, S.G. Measuring surface currents in lakes with high spatial resolution thermal infrared imagery. *Geophys. Res. Lett.* **2005**, *32*, doi:10.1029/2005GL022912.
6. Allen, R.G.; Hendricks, J.M.H.; Toll, D.; Anderson, M.; Kustas, W.; Kleissl, J. From high overhead-ET measurement via remote sensing. *Southwest Hydrol.* **2008**, *7*, 30–32.
7. Thenkabail, P.S.; Biradar, C.M.; Noojipady, P.; Dheeravath, V.; Li, Y.; Velpuri, M.; Gumma, M.; Gangalakunta, O.R.P.; Turral, H.; Cai, X.L.; *et al.* Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. *Int. J. Remote Sens.* **2009**, *30*, 3679–3733.
8. Schott, J.R.; Hook, S.J.; Barsi, J.A.; Markham, B.L.; Miller, J.; Padula, F.P.; Raqueno, N.G. Thermal infrared radiometric calibration of the entire Landsat 4, 5, and 7 archive (1982–2010). *Remote Sens. Environ.* **2012**, *122*, 41–49.
9. Reuter, D.C.; Richardson, C.; Pellerano, F.; Irons, J.R.; Allen, R.; Anderson, M.; Jhabvala, M.; Lunsford, A.; Montanaro, M.; Smith, R.; *et al.* The Thermal Infrared Sensor (TIRS) on Landsat 8: Design overview and pre-launch characterization. *Remote Sens.* **2014**, in press.
10. Montanaro, M.; Levy, R.; Markham, B. On-orbit radiometric performance of the Landsat 8 Thermal Infrared Sensor. *Remote Sens.* **2014**, submitted for publication.
11. Caselles, V.; Rubio, E.; Coll, C.; Valor, E. Thermal band selection for the PRISM instrument: 3. Optimal band configurations. *J. Geophys. Res. Atmos.* **1998**, *103*, 17057–17067.
12. Montanaro, M.; Lunsford, A.; Tesfaye, Z.; Wenny, B.; Reuter, D. Radiometric calibration methodology of the Landsat 8 Thermal Infrared Sensor. *Remote Sens.* **2014**, *6*, 8803–8821.

13. Berk, A.; Anderson, G.P.; Bernstein, L.S.; Acharya, P.K.; Dothe, H.; Matthew, M.W.; Adler-Golden, S.M.; Chetwynd, J.H., Jr.; Richtsmeier, S.C.; Pukall, B.; *et al.* MODTRAN4 radiative transfer modeling for atmospheric correction. In *Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research III*; Larar, A.M., Ed.; SPIE-Int. Soc. Optical Engineering: Bellingham, WA, USA, 1999; pp. 348–353.

14. Hook, S.J.; Chander, G.; Barsi, J.A.; Alley, R.E.; Abtahi, A.; Palluconi, F.D.; Markham, B.L.; Richards, R.C.; Schladow, S.G.; Helder, D.L. 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. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 2767–2776.

15. Hook, S.J.; Prata, F.J.; Alley, R.E.; Abtahi, A.; Richards, R.C.; Schladow, S.G.; Palmarsson, S.O. Retrieval of lake bulk and skin temperatures using Along-Track Scanning Radiometer (ATSR-2) data: A case study using Lake Tahoe, California. *J. Atmos. Ocean. Technol.* **2003**, *20*, 534–548.

16. Hook, S.J.; Clodius, W.B.; Balick, L.; Alley, R.E.; Abtahi, A.; Richards, R.C.; Schladow, S.G. 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 Trans. Geosci. Remote Sens.* **2005**, *43*, 1991–1999.

17. Hook, S.J.; Vaughan, R.G.; Tonooka, H.; Schladow, S.G. 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, automated validation site. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1798–1807.

18. Padula, F.P.; Schott, J.R.; Barsi, J.A.; Raqueno, N.G.; Hook, S.J. Calibration of Landsat 5 thermal infrared channel: Updated calibration history and assessment of the errors associated with the methodology. *Can. J. Remote Sens.* **2010**, *36*, 617–630.

19. NCEP North American Regional Reanalysis: NARR. Available online: http: //www.esrl.noaa.gov/ psd/data/gridded/data.narr.html (accessed on 10 September 2014).

20. Montanaro, M.; Gerace, A.; Lunsford, A.; Reuter, D. Stray light artifacts in imagery from the Landsat 8 Thermal Infrared Sensor. *Remote Sens.* **2014**, *6*, 10435–10456.