

Article

Development of an Operational Calibration Methodology for the Landsat Thermal Data Archive and Initial Testing of the Atmospheric Compensation Component of a Land Surface Temperature (LST) Product from the Archive

Monica Cook *, John R. Schott, John Mandel and Nina Raqueno

Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology,
54 Lomb Memorial Drive, Rochester, NY 14623, USA; E-Mails: schott@cis.rit.edu (J.R.S.);
jjm4687@rit.edu (J.M.); nina@cis.rit.edu (N.R.)

* Author to whom correspondence should be addressed; E-Mail: mxc7441@rit.edu;
Tel.: +1-585-766-2141.

External Editors: Brian Markham and Prasad S. Thenkabail

*Received: 5 August 2014; in revised form: 21 October 2014 / Accepted: 28 October 2014 /
Published: 13 November 2014*

Abstract: The Landsat program has been producing an archive of thermal imagery that spans the globe and covers 30 years of the thermal history of the planet at human scales (60–120 m). Most of that archive’s absolute radiometric calibration has been fixed through vicarious calibration techniques. These calibration ties to trusted values have often taken a year or more to gather sufficient data and, in some cases, it has been over a decade before calibration certainty has been established. With temperature being such a critical factor for all living systems and the ongoing concern over the impacts of climate change, NASA and the United States Geological Survey (USGS) are leading efforts to provide timely and accurate temperature data from the Landsat thermal data archive. This paper discusses two closely related advances that are critical steps toward providing timely and reliable temperature image maps from Landsat. The first advance involves the development and testing of an autonomous procedure for gathering and performing initial screening of large amounts of vicarious calibration data. The second advance discussed in this paper is the per-pixel atmospheric compensation of the data to permit calculation of the emitted surface radiance (using ancillary sources of emissivity data) and the corresponding land surface temperature (LST).

Keywords: Landsat; calibration; land surface temperature

1. Introduction and Summary

Radiometric calibration of the thermal sensors on the Landsat satellites has traditionally relied on ground truth from a combination of buoy mounted radiometers and thermistors on Lake Tahoe and the Salton Sea, operated by the Jet Propulsion Laboratory (JPL) [1], and flight and field surface temperature measurements obtained by the Rochester Institute of Technology (RIT) [2]. More recently RIT has augmented its field campaigns with data from a small number of NOAA moored buoys [3]. The NOAA buoy approach was shown to be nearly as accurate as the traditional field campaigns [4] and, because of their number and long term deployment, a larger number of samples was obtained, improving the overall confidence in the calibration. Therefore, in recent years, a combination of data from the JPL sites and approximately a half dozen NOAA buoys has been used to monitor and update the calibration of Landsats 5 and 7 [5]. In all cases, radiative transfer models are used to propagate surface radiance to sensor reaching radiance for comparison to Landsat observations.

In anticipation of the launch of Landsat 8, a program to operationalize the use of the NOAA buoys was initiated. This was motivated by the desire to know, as soon as possible after launch, the calibration state of the new Thermal Infrared Sensor (TIRS) [6] and update it if necessary (note that with Landsat 7, it took over a year to be confident there was a calibration issue and nearly a year and a half to implement a correction). This largely automated or operational approach was designed to utilize a large portion of the approximately 100 buoys [7] whose records would be accessed on each Landsat overpass in conjunction with local weather data and nearby radiosonde data, all of which would be automatically combined and processed using MODTRAN [8] to generate sensor reaching radiance values. The goal was to acquire sufficient data within a month or two of TIRS operation on orbit to characterize both the gain and bias calibration of the TIRS thermal bands (note the initial calibration of Landsat 7, described above, after a year of operation was based on only about 30 points). This operational buoy approach was successful and the methodology and results are discussed in Sections 2 and 3 respectively.

For many users, a thermal sensor that provides calibrated sensor reaching radiance is only the beginning of a long struggle to get to the surface temperature values needed for their studies. The proper use of radiative transfer codes, accessing the appropriate atmospheric sounding data, incorporation of sensor relative spectral response, adjusting for surface emissivity and converting surface radiance to temperature can be daunting tasks even for sophisticated users. To aid in this process, the United States Geologic Survey (USGS) is considering providing a land surface temperature (LST) product for Landsat data. This product is being jointly designed and tested by RIT and JPL; RIT will be providing the atmospheric compensation component and JPL the emissivity correction. In this paper, we will focus on the approach and testing of the atmospheric compensation component as it uses a similar radiative transfer process as the calibration methodology, and the NOAA buoys were used to test the algorithm performance.

Unlike the calibration process which needs only single point answers for a hundred or so well defined points, the LST product seeks to provide an LST solution for every pixel everywhere on the globe

(though this initial implementation/validation will be restricted to North America). To achieve this, North American Regional Reanalysis (NARR) data are used to characterize the atmosphere [9], MODTRAN is used to perform the radiative transfer calculations and the data are interpolated in time, elevation and pixel location to provide per pixel atmospheric compensation values. These will then be corrected for emissivity and converted to temperature using ASTER derived emissivity values produced by JPL [10]. For this study, which uses water targets for validation of the atmospheric compensation process, well known emissivity values for water were used along with a radiance to temperature look up table (LUT). The goal of the validation experiment was to test the overall atmospheric compensation procedure over a range of geographic locations, elevations and atmospheric conditions. Validation over land will be incorporated with the emissivity provided by our collaborators at JPL. Over 250 cloud free Landsat 5 scenes over buoys were selected and analyzed yielding very low errors. Landsat 5 was selected for initial testing because it is very well calibrated and no interpolation for scan line corrector (SLC) gaps was required. Based on the very good Landsat 5 results, an initial testing on Landsat 8 data was also conducted to demonstrate the potential for Landsat 8 and to help in characterizing ongoing calibration issues with Landsat 8. The methodology used in developing the LST product is discussed in Section 4 and the results of the validation are covered in Section 5.

Previous studies indicate that the atmospheric compensation will be the largest error source in a LST algorithm so that these errors are likely to be the limiting factor in the final LST error [11]. Results from the atmospheric compensation error assessment reported here will be combined with an assessment of the errors from the emissivity estimation process being conducted by JPL to generate the overall errors in the LST product.

2. Calibration Methodology

This section will address the methodology associated with using the NOAA buoys to calibrate the Landsat thermal bands. The same method is used for all the Landsat instruments with only the instrument specific relative spectral response (RSR) functions changing.

2.1. Sub Surface to Skin Temperature Adjustment

The NOAA buoys measure the water temperature between 0.6 m and 1.5 m below the surface depending on the buoy type. The Landsat sensors “see” the surface radiance being emitted from the top microns of the water surface. Therefore, in order to use the buoy temperatures measured at some depth for calibration, this subsurface temperature must first be corrected to surface or bulk temperature (*i.e.*, temperature change due to mixing from the depth of the temperature measurement to a few millimeters below the surface) and then to the skin temperature (*i.e.*, temperature change due to cooling by radiation from the top microns) that the satellites see. The near surface water temperature varies on a diurnal cycle driven predominately by insolation. This roughly sinusoidal cycle is dampened as a function of depth and wind speed. In addition, there is a temporal lag in the phase that increases with depth and decreases with wind speed. Zeng *et al.* (1999) [12] did an extensive study of this phenomenon including the development of coefficients characterizing the dependency of the diurnal damping and phase delay with depth and wind speed. Using the diurnal temperature and wind speed history from the buoy data for the 24 h preceding the satellite overpass, the subsurface temperatures can be converted to surface

temperatures using the coefficients derived by Zeng *et al.* (1999) [12]. Note that this process breaks down for low wind speeds where there is little mixing between the subsurface and near surface water resulting in poor correlation of any type between the temperatures. As a result, data for conditions with wind speeds below 0.2 m/s are rejected from the analysis. At high wind speeds (above 8 m/s) the surface waters are well mixed and no subsurface to surface correction is applied (see Figure 1). However, for all conditions, a skin temperature correction is applied. This is a small correction needed to account for the radiative cooling of the top microns of the water which are constantly trying to achieve radiative equilibrium with the “cold” sky (note that during the clear sky conditions used for calibration the sky is always considered “cold”). The process consists of three steps.

The first step is to find the average skin temperature over a 24-h period from

$$\langle T_s \rangle = \langle T_z \rangle - az - d \quad (1)$$

where z is the depth of the temperature measurement for the buoy (m), $\langle T_z \rangle$ is the average temperature from the buoy measurement at this depth z over a 24 h period in Kelvin (K), $\langle T_s \rangle$ is the average skin temperature over a 24 h period (K), and d is the surface cool skin effect (K). Donlon *et al.* (2002) [13] showed that for a range of clear sky conditions the surface to skin correction (d) could be approximated with a constant value of 0.17 K (*i.e.*, the skin is 0.17 K cooler than the surface) and this value is used for this study. Finally, a is the thermal gradient represented by

$$a = 0.05 - \frac{0.6}{\mu_m} + 0.03 \ln(\mu_m) \quad (2)$$

where μ_m is the 24 h average wind speed at a height of 10 m above sea level (m/s) (available from sensors on the buoy).

The second step in the process acknowledges that the surface temperature will vary with time (diurnal variation), and can be solved empirically using

$$f(t - cz) = \frac{T(z, t) - \langle T_z \rangle}{e^{-bz}} \quad (3)$$

where $T(z, t)$ is the buoy temperature at depth z and time t , cz is a phase term, and e^{-bz} accounts for dampening of the amplitude with depth. This second step is accomplished by solving for $f(t)$ through the interpolation of $f(t - cz)$ around the specific time of interest (t). The phase constant (hr/m) was empirically derived by Zeng *et al.* (1999) [12] to be:

$$c = 1.32 - 0.64 \ln(\mu_m) \quad (4)$$

and the damping constant with depth (m) to be:

$$b = 0.35 + 0.018e^{0.4\mu_m} \quad (5)$$

Graphically this step can be seen in Figure 2. This can be viewed as first correcting the diurnal temperature variation for the dampening effect of depth and then correcting for the temporal phase shift between the peak temperature at the surface and the peak temperature at depth.

Figure 1. Illustration of parameters used in Equations (1) and (2).

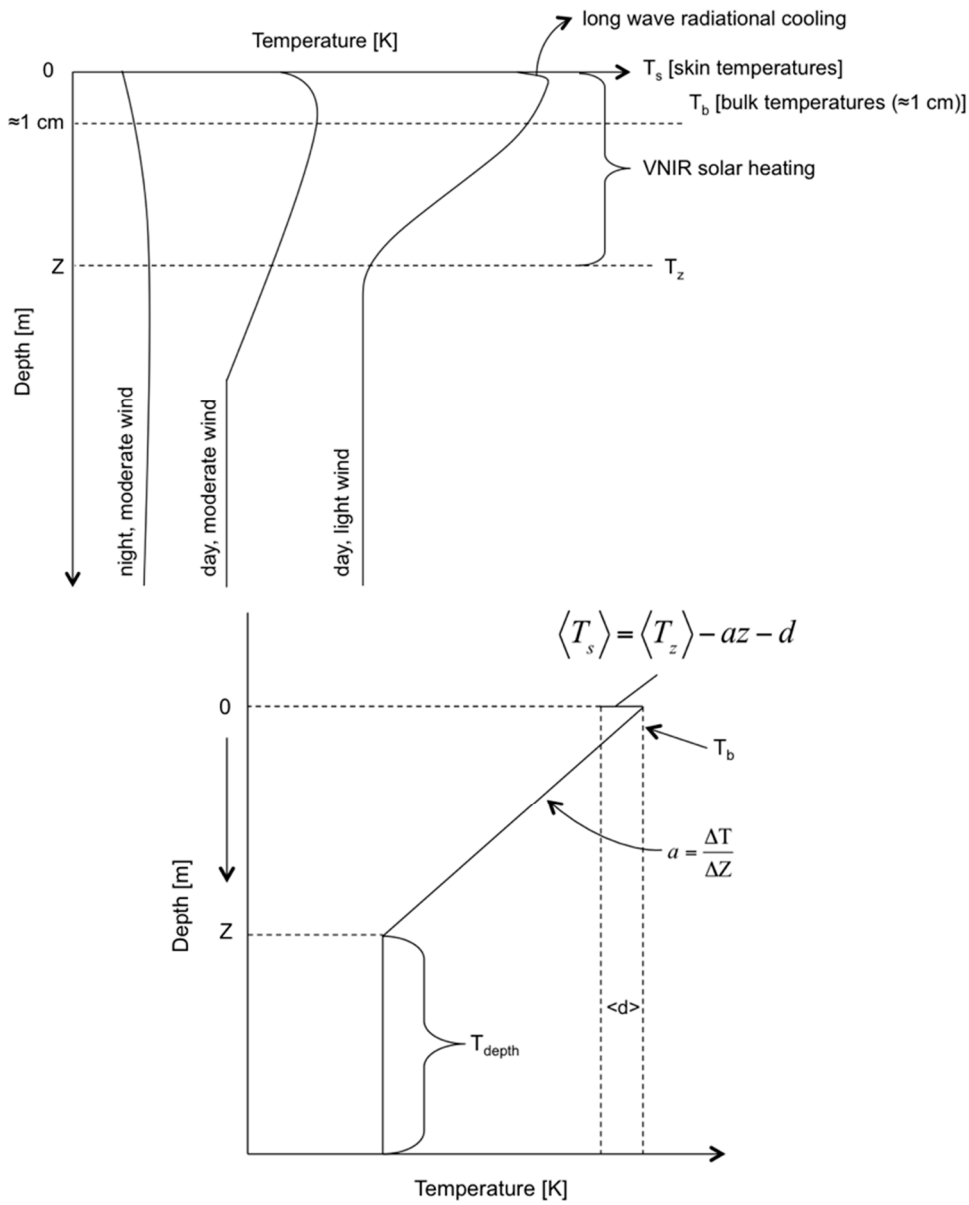
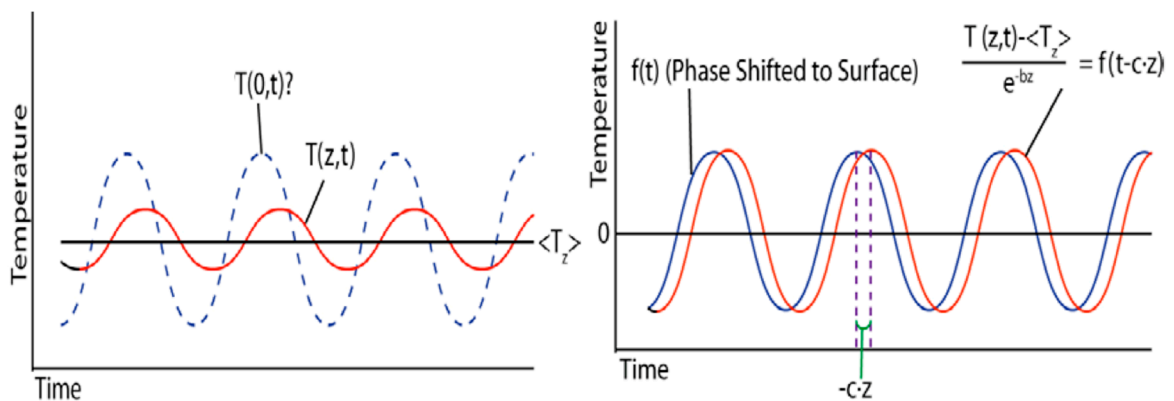


Figure 2. Graphic representation of the unknown diurnal surface temperature at time t (left), and the Zeng empirical correction method (right).



The final step in estimating the skin temperature is completed using the 24 h average skin temperature ($\langle T_s \rangle$) and $f(t)$ solved for in steps 1 and 2 respectively to yield

$$T_s = T(0, t) = \langle T_s \rangle + f(t) \quad (6)$$

$T(0, t)$ represents the skin temperature at the time of interest, or in other words the skin temperature at the time of the spacecraft collection. Padula *et al.* (2011) [3] applied this approach to a small set of buoys with large fetch (open water upwind from the buoy site) and clear sky conditions (*i.e.*, Landsat calibration conditions) and demonstrated its applicability to a wider range of conditions than the South Pacific where Zeng *et al.* (1999) [12] gathered the data used to develop the coefficients. Note that this entire process when applied to a wide range of buoys typically amounts to corrections of only a few tenths Kelvin (large corrections would only be needed at low wind speeds and are rejected as untrusted); however, because we are attempting to achieve an overall calibration with a residual error of that same order, this level of correction is necessary.

2.2. Use of Radiative Transfer Models to Estimate Top of Atmosphere (ToA) Radiance from Skin Temperature

The next step in the calibration process is to estimate a top of atmosphere (ToA) radiance from the skin temperature to directly compare to the radiance observed by the Landsat thermal instruments. This is accomplished using the MODTRAN radiative transfer code [8] with careful characterization of the atmospheric inputs to the model. The ToA or effective in band sensor reaching spectral radiance can be expressed as [14]

$$L_{ToA\lambda} = \frac{\int \{ \varepsilon(\lambda) L_{T\lambda} \tau(\lambda) + (1 - \varepsilon(\lambda)) L_{d\lambda} \tau(\lambda) + L_{u\lambda} \} RSR(\lambda) d\lambda}{\int RSR(\lambda) d\lambda} \quad (7)$$

$L_{T\lambda}$ is the Planckian spectral radiance ($\text{W} \cdot \text{m}^{-2} \cdot \text{sr}^{-1} \cdot \mu\text{m}^{-1}$) emitted from the surface due to the skin temperature (T) solved for in Section 2.1, $\tau(\lambda)$ is the wavelength (λ) dependent transmission, $L_{d\lambda}$ is the downwelled spectral radiance which will be reflected ($1 - \varepsilon(\lambda)$) from the water, and $L_{u\lambda}$ is the path radiance along the sensor to ground line of sight. $RSR(\lambda)$ is the known relative spectral response of the particular thermal band of the Landsat instrument under calibration and $\varepsilon(\lambda)$ is a lab measured spectral emissivity of water.

MODTRAN is used to solve for the unknowns τ , L_u , and L_d in Equation (7) based on estimates of the atmospheric profile between the sensor and the ground (the temperature and water vapor profiles are particularly critical in the Landsat bands). MODTRAN radiative transfer code has been used in conjunction with ground truth data to support all Landsat calibration and it has been shown that errors in the process are dominated by lack of knowledge of atmospheric variables rather than the radiative transfer code itself. The residual error in the process is on the order of a few tenths Kelvin [3].

The inputs to MODTRAN come from the nearest radiosonde which typically has measured the temperature, pressure and relative humidity up to 20–30 km. A standard atmosphere is appended to the radiosonde data to characterize the more stable atmosphere up to space. Because the radiosonde may be tens of kilometers (or even up to 100 km) from the buoy and have sampled the atmosphere hours before the satellite over pass, the lowest and most critical atmospheric layers (surface to 1–2 km) are updated

using local surface air temperature and relative humidity data (e.g., from the buoy) interpolated from the surface to the thermal inversion layer, or to 1 km if no inversion layer exists [15].

The ToA spectral radiance obtained from the MODTRAN model can be directly compared to the observed radiance (Lobs) obtained by converting the sampled Landsat digital number (DN) to spectral radiance using the metadata in the Landsat header. Any difference is described as the error in the calibration. By taking many samples over a range of temperatures a sound estimate of the calibration state of the instrument can be made.

2.3. Data Sources and Web Access to Data Needed for Operational Calibration

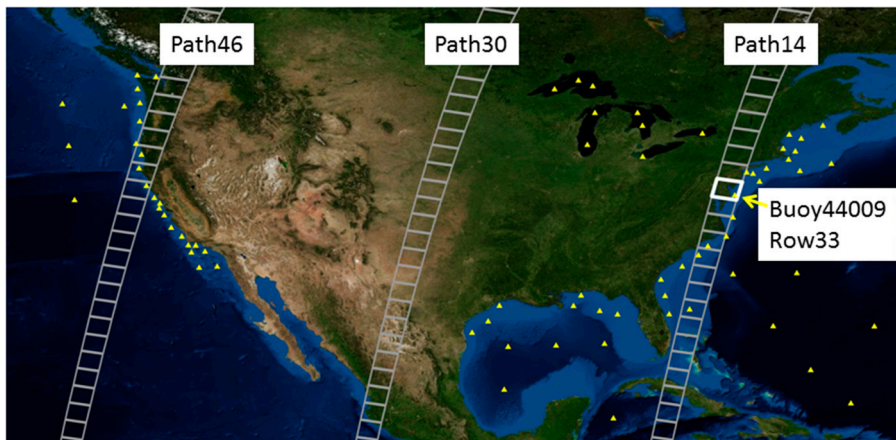
The process described in Sections 2.1 and 2.2 for calibrating the Landsat thermal bands is relatively straight forward for an expert with the appropriate software tools and a day to gather and combine the data to yield one calibration point. The effort to operationalize (or semi automate) the process did not involve any major changes to the analytical process. Rather, it involved a major effort to define a database of necessary inputs for each buoy as a function of time (the buoys can move from year to year) and tools to automatically obtain, sample, organize, format, and filter the data from the internet for input to the existing analytical tools.

The operational buoy calibration approach leverages several existing IDL analytical tools for the core radiometric calibration by linking them with automated input retrievals in one cohesive Python workflow. The entire workflow initiates with the selection of a specific date to be processed as it was initially designed to continuously run, checking daily for any new Landsat scenes with buoy data for that current day. However, a simple batch process can be used to process many dates in cases of historic investigations or when the Landsat archive has been reprocessed with updated calibration. Given a specific date and which instrument to evaluate (Landsat 4, 5, 7, or 8), the workflow identifies a list of all possible paths imaged on that day. This is possible because, once each satellite achieves its final orbit each path is on a predictable 16 day repeat. The Python workflow steps through all possible paths for that given day and all buoys within each path.

This brings us to the real driving force behind the process, the buoy meteorological database that describes all the necessary inputs for each specific buoy, such as, operational start and stop date, location, watch radius, depth of temperature measurement and height of wind measurement. The watch radius of the buoy is the distance a buoy can drift on the anchor chain from the nominal anchor location; this will be important in screening buoys later in the process. The corresponding radiosonde and weather station identification codes and the Landsat scene's WRS-2 path and row are also stored within this database. It is also necessary for the database to maintain a temporal record of the buoy locations over the years as the deployment position may change slightly. Given a specific path, the workflow queries this database to generate a list of all the possible buoys to be processed within that path. For example, Figure 3 illustrates the possible paths for Landsat 8 for a specific date.

Once a specific path and buoy has been identified, the row is selected for that buoy location, which identifies the specific scene to automatically download directly from the USGS Earth Explorer [16] using a customized wget retrieval. Next all the necessary inputs are identified from the buoy meteorological database and automatically downloaded from diverse web servers.

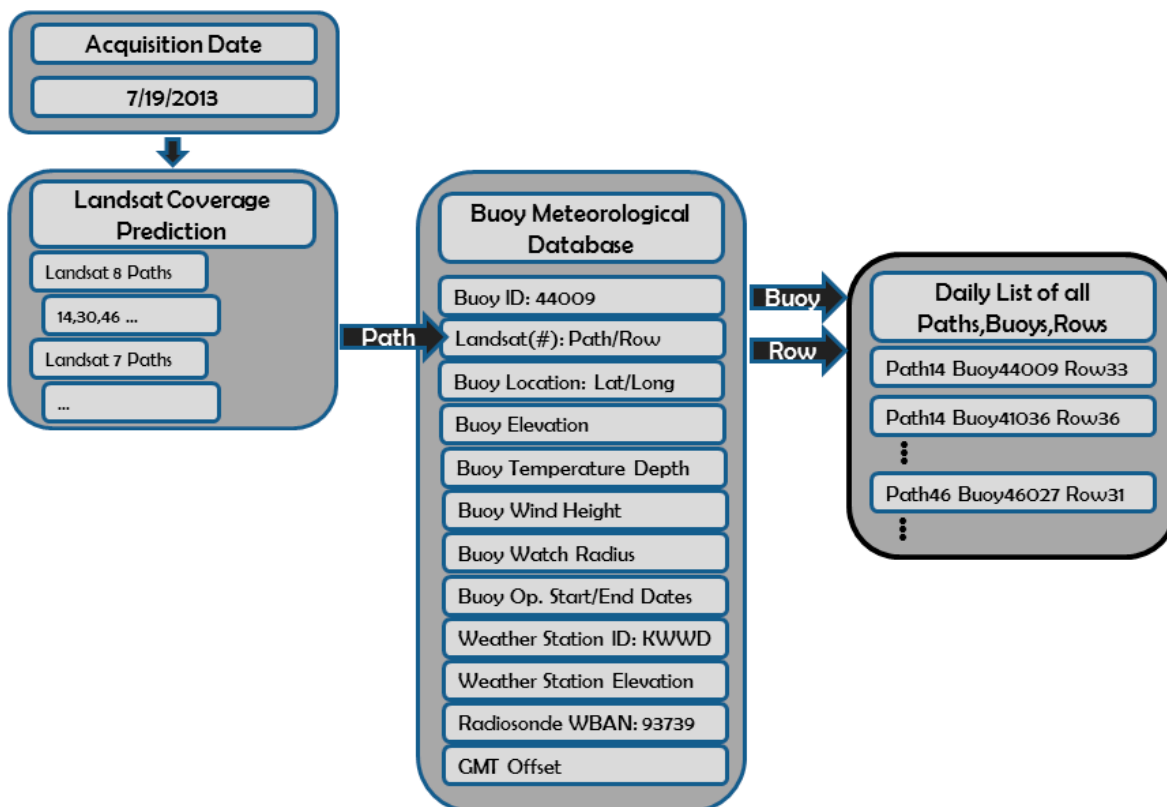
Figure 3. Example of possible Landsat 8 paths, buoys, and specific scene rows for 19 July 2013; triangles indicate buoy location.



- The atmospheric profile data is obtained through a PHP script call to NOAA’s ESRL Radiosonde Database [17].
- The surface weather data is extracted from Weather Underground [18].
- The buoy data is downloaded from NOAA’s National Data Buoy Center [7].

This process is summarized in Figure 4.

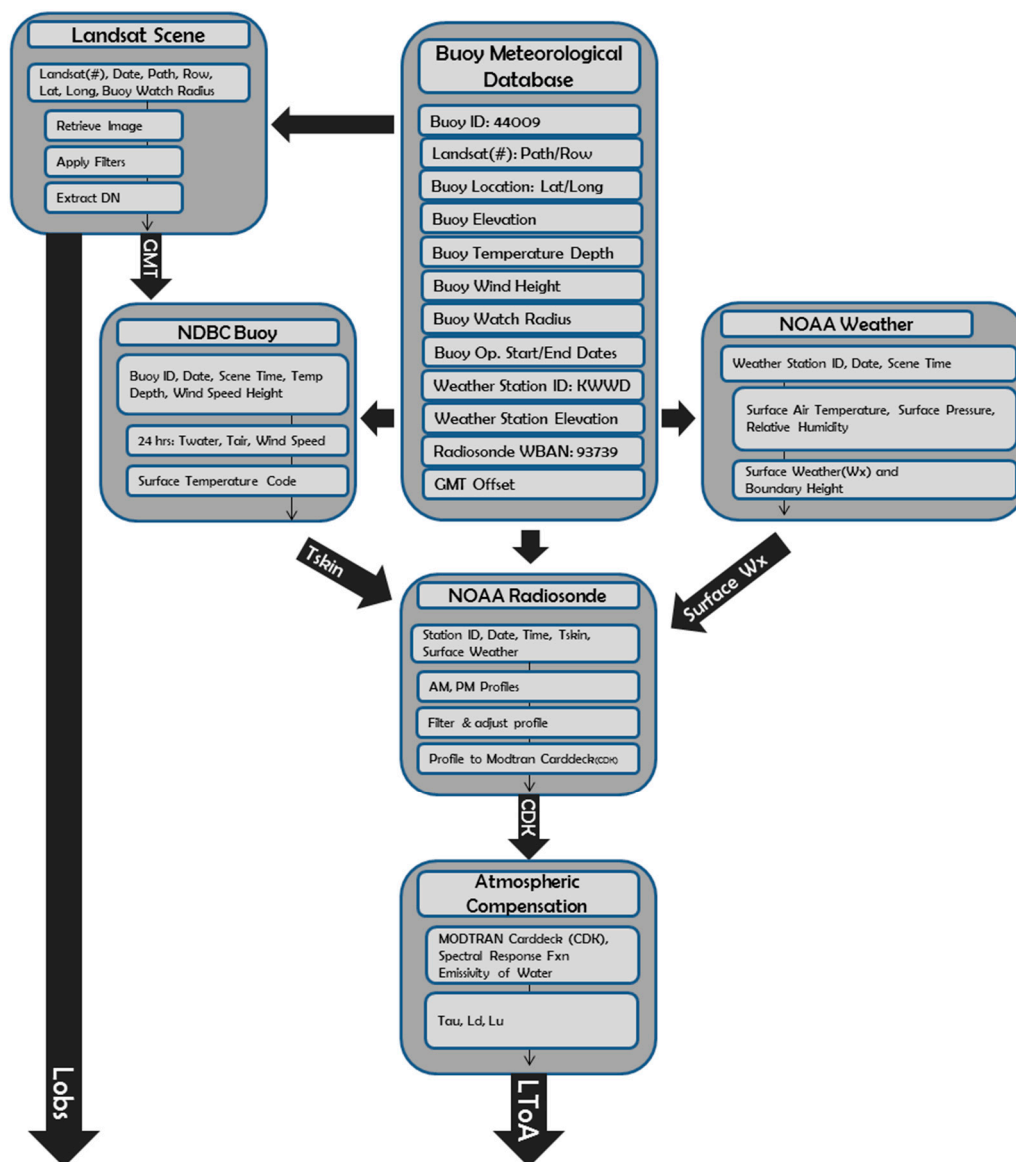
Figure 4. Selection of specific Landsat scenes and corresponding buoys for a given date using the buoy meteorological database.



Though invaluable, the buoy data presents its own set of challenges since the data format, missing data flags, and download location changes depending on length of time since the buoy measurement. There are at least three different formats: data measured in the past 45 days, data more than a year old, data taken within the current year but more than 45 days old. In the operational scenario this workflow would be constantly running on the past 45 day product. However, historic studies or cases when the entire Landsat archive has been reprocessed require that the workflow handles all buoy format nuances. Buoy data availability requires further verification as many buoys are retrieved for the winter months or may be offline due to instrument maintenance.

Once all the input data are acquired, they are screened for missing values and formatted for acceptance into the sub surface to skin temperature adjustment and atmospheric compensation as described previously in Sections 2.1 and 2.2. This portion of the workflow is illustrated in Figure 5.

Figure 5. Workflow illustrating: assembling the necessary inputs, Landsat scene filtering, buoy to surface temperature adjustment, and atmospheric compensation.



2.4. Automated Screening of Landsat Data

Every time a Landsat satellite passes over a NOAA buoy a potential calibration point is generated. However, most of these points are not appropriate for use in generating high quality calibration data. In order to isolate only high quality data in which we have high confidence, a number of data filters were developed to exclude inappropriate or questionable points. This starts at the data ingest process where missing data or ill formatted data will cause buoy, radiosonde or local weather data to be rejected. Any of these errors will cause the calibration point to be rejected. In addition, certain buoys consistently yield untrustworthy results; for example, they may have excessively large watch radii (*i.e.*, for example, at deep water sites, the watch radius could be as large as a kilometer, so we do not have confidence in where the buoy is within a Landsat scene). These buoys are culled from the database to avoid unnecessary processing. Once the data passes the basic data integrity test, the Landsat image can be downloaded and additional filtering to isolate calibration quality data begins. At this point, the majority of the data are filtered by a series of image based tests that involve thresholds on the data variability around the nominal buoy location. The standard deviation of all points within the watch radius and within 0.22 km of the nominal buoy location is calculated. The local window is used to estimate the observed spectral radiance for comparison to the buoy predicted radiance. This window has a small threshold ($0.039 \text{ W}\cdot\text{m}^{-2}\cdot\text{sr}^{-1}\cdot\mu\text{m}^{-1}$) because variation within that window would suggest local variation in water temperature or local cloud contamination, either of which should cause the point to be rejected. A slightly larger threshold ($0.044 \text{ W}\cdot\text{m}^{-2}\cdot\text{sr}^{-1}\cdot\mu\text{m}^{-1}$) is used for the larger window; this allows for some small variation in the water temperature but screens for any variation that might be introduced by clouds in the surround. To screen for uniform clouds between the buoy and the sensor, a threshold is used to reject points where the difference between the surface air temperature and the sensor observed apparent temperature is too large (cloud apparent temperatures are typically much colder than surface air temperature). Additional checks include filters for excessive water vapor in the air column (as calculated by MODTRAN), too many moist layers (as estimated by low dew point depression) and low lapse rates (temperature decrease per 100 meters of altitude increase). Failure to pass through any filter will result in rejection of a point. As a result of this screening process, only about 20% of the points that pass the initial data contamination test get through the image based and atmospheric condition calibration data quality filters.

At this point, the satellite spectral radiance averaged over the region encompassed by the watch circle is compared to (differenced from) the spectral radiance propagated from the buoy skin temperature to ToA using the MODTRAN radiative transfer process to produce a single point in the calibration assessment. To test this methodology, hundreds of points were produced for the well calibrated Landsat 5 and Landsat 7 instruments [3,5] to confirm that the results agreed with the traditional calibration. Only then, if appropriate—*i.e.*, a successful comparison—would the method be applied to the Landsat 8 calibration for which it was designed. The results of these successful comparisons are discussed in the following section.

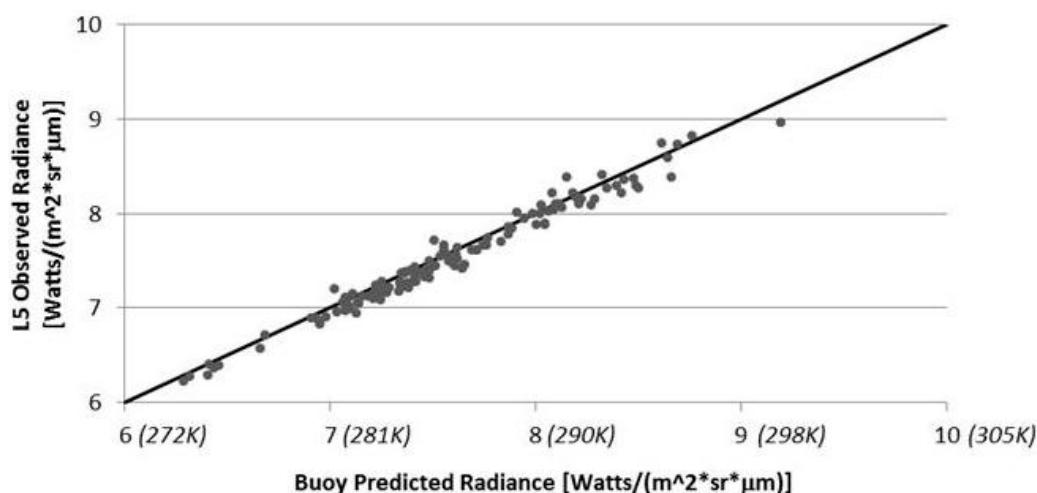
3. Calibration Results

In this section, the results of the automated/operational calibration approach are first compared to the well understood calibration of Landsats 5 and 7 (Section 3.1). Based on the nearly identical results of the automated and more traditional approaches the method could then be confidently applied to Landsat 8 data (see Section 3.2).

3.1. Comparison of Operational Results to Current Landsat 5 and Landsat 7 Calibration

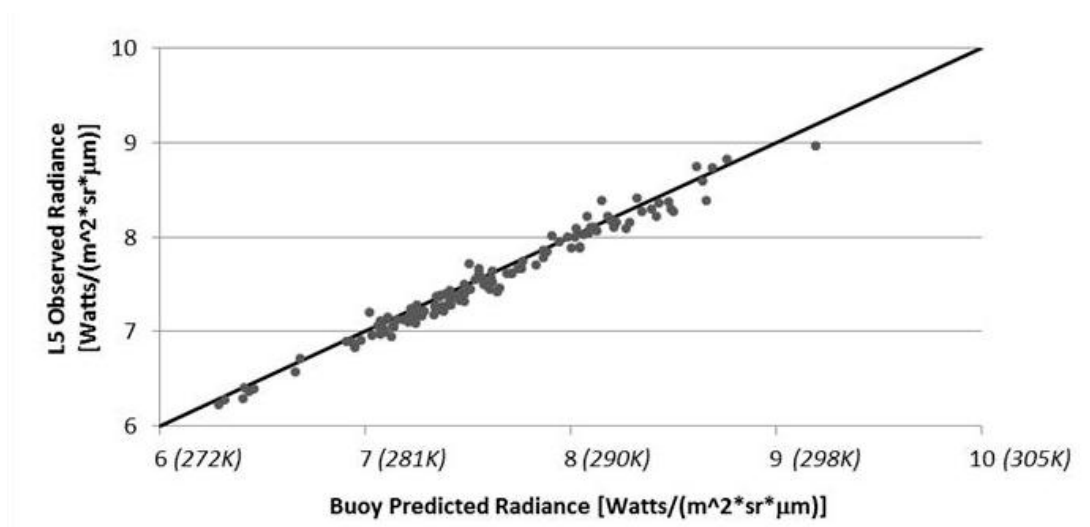
To test the operational calibration approach described in Section 2, the methodology was applied to 603 sites in Landsat 5 data covering primarily the period 2006–2012. The sites were chosen for convenience because buoy data were available and the Landsat scenes were on the active discs at USGS Earth Resource Observation System (EROS) data center (*i.e.*, they did not need to be ordered). After the screening process discussed in Section 2.4, 131 points remained. Seven points with large errors (>2 K) were visually assessed and six of the seven were shown to be cloud contaminated. This suggests that future work should include improved filters for subtle cloud contamination. The errors results for the remaining points are plotted in Figure 6; these points show a mean bias of -0.52 K \pm 0.72 K (*versus* the data set before removal of the six points, which had a mean bias of 0.57 K \pm 0.084 K). This is in good agreement with the current calibration estimate for Landsat 5 that was generated by the standard methods; personal communication with the Landsat Calibration team indicates that Landsat 5 is currently slightly miscalibrated with a bias of -0.33 K \pm 0.54 K (at 300 K).

Figure 6. Plot of top of atmosphere (ToA) buoy predicted radiance (apparent temperature) vs. satellite observed radiance for Landsat 5 after removal of six cloud contaminated points (mean error -0.52 K \pm 0.72 K).



A similar study looked at 570 Landsat 7 points across its ongoing operation life; 129 points remained after the currently implemented screening process and showed a mean bias of -0.24 K \pm 0.81 K, as shown in Figure 7. This is in excellent agreement with the calibration observed by the traditional methods of -0.26 K \pm 0.42 K (for data processed after 1 October 2013, this bias has been corrected to 0.0 K \pm 0.42 K). Note that much of the data included in the study was from prior to reprocessing (*i.e.*, before the bias correction was applied).

Figure 7. Plot of ToA buoy predicted radiance vs. satellite observed radiance for Landsat 7 for all data processed automatically (mean error $-0.24 \text{ K} \pm 0.81 \text{ K}$).



These automated results have slightly higher standard deviations than the handpicked best condition data used by JPL and RIT to generate the data on which the current calibration is based ($\sigma \approx 0.5$). However, we expect to be able to generate four or five times as many points (N) with significantly less effort. The standard error in the results, which is proportional to $\frac{1}{\sqrt{N}}$, will remain the same or even improve which means our confidence in the estimate of bias error will remain the same or improve; the automated results should also be available and updated on much shorter time intervals as desired. Furthermore, as more data becomes available, the filters can be made more restrictive reducing the number of points while improving data quality.

3.2. Landsat 8 Results Using Operational Approach

Based on the excellent performance of this approach, the methodology was applied to a Landsat 8 data set consisting of 33 Image-Buoy derived ToA radiance comparisons. The results shown in Figure 8 for band 10 show a bias of $-0.01 \text{ K} \pm 0.90 \text{ K}$ (note this is on data that were reprocessed to correct for an initial bias of approximately 2 K observed shortly after launch [6]). Similar results are shown for band 11 in Figure 9, which shows a bias of $-0.91 \text{ K} \pm 1.28 \text{ K}$ (again after an initial bias correction of approximately 4 K was applied to the entire database by USGS). These data show that the initial correction is performing reasonably well in band 10 but has overcorrected band 11 by approximately 1 K for this data set. In addition, the standard deviations about the mean bias are significantly higher for Landsat 8, particularly in band 11, even though the instrument signal to noise ratios (SNRs) for Landsat 8 are higher than those for earlier instruments. Since the methodology is identical and the SNR for Landsat 8 is higher, this large variation is likely due to actual changes in the calibration; the correlation of error with scene radiance [19] suggests there is a variable source of calibration error that is not fully accounted for by the combined use of on board [20] and vicarious calibration described here. This residual error has been attributed to a ghosting or stray light problem associated with radiation from outside the detectors nominal point spread function (PSF) impinging on the detector (see Section 6). Note that because the ghosting error is dependent on the radiance of the surround (particularly in

band 11), the bias errors will be dependent on the data set [19]. The data set used to generate the initial calibration included some much higher radiance scenes (from the JPL sites) resulting in a larger band 11 correction than these data would suggest is necessary.

Figure 8. Plot of ToA buoy predicted radiance vs. satellite observed radiance for Landsat 8 band 10 for all data processed automatically (mean error $-0.01 \text{ K} \pm 0.90 \text{ K}$).

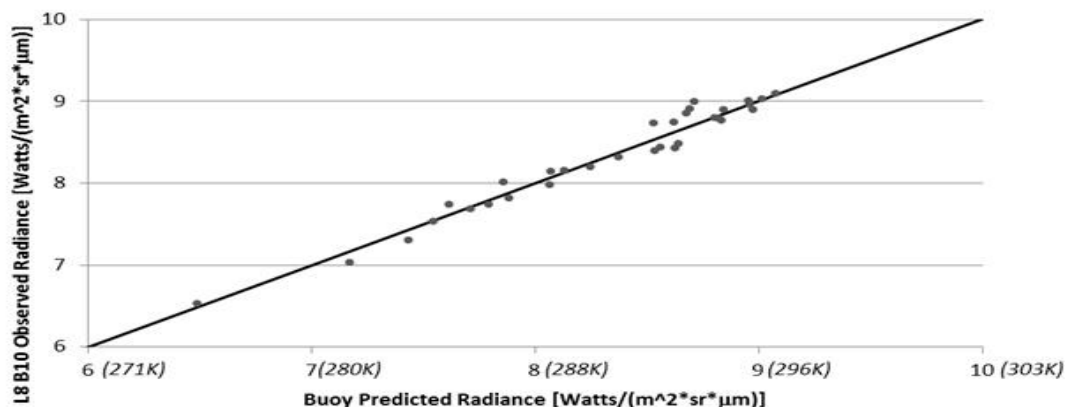
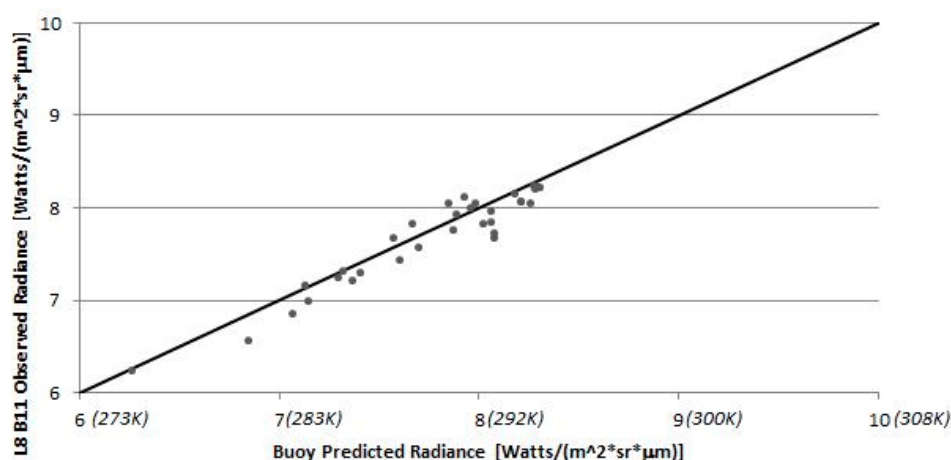


Figure 9. Plot of ToA buoy predicted radiance vs. satellite observed radiance for Landsat 8 band 11 for all data processed automatically (mean error $-0.91 \text{ K} \pm 1.28 \text{ K}$).



4. Atmospheric Compensation Methodology for LST

This section will address the methodology associated with the atmospheric compensation component of the Landsat land surface temperature product. Similar to the calibration methodology discussed in Section 2.2, the goal is to generate radiative transfer parameters using atmospheric profiles, but using a modified approach with different input data for the LST product. This is not the first tool developed to atmospherically correct Landsat data; Barsi *et al.* [21] have an online tool which generates the necessary radiative transfer parameters for a given location in a Landsat scene. The methodology presented uses a similar radiative transfer approach, but utilizes more finely sampled reanalysis data and includes a number of differences in the process. Most importantly, the approach presented here is designed to generate a per pixel product over the entire scene including adjustments for elevation, in addition to the space and time adjustment included in the earlier product. Though the main advantage of

the approach presented here is an image wide operational solution, the extensive testing also shows some performance improvements over the currently available solution [21]. Our methodology was developed and validated using Landsat 5 scenes and further tested on Landsat 8 as discussed in Section 5. As discussed in Section 3.1, personal communication with the Landsat Calibration team indicates that Landsat 5 is currently slightly miscalibrated, based on traditional methods, with a bias of -0.33 K at 300 K; this bias is corrected for in all scenes processed using the LST methodology. Given the correction, the methodology assumes the data from the archive are well calibrated using the traditional approach (to be augmented in the future using the automated approach described above).

4.1. MODTRAN Radiative Transfer Using NARR Database

Similar to the methodology discussed in Section 2.2, MODTRAN is used to solve for the atmospheric radiative transfer parameters: transmission, upwelled radiance, and downwelled radiance. As mentioned, the temperature and water vapor profiles are critical in the Landsat bands. However, because of the spatial and temporal coverage desired for the LST product, the use of radiosondes and local surface data is not feasible due to both the spatial availability and desired automation. Various alternative data sources were considered. Retrospective analysis, or reanalysis, uses a variety of observed data as input to numerical models to generate a regularly sampled spatial and temporal grid of variables that are difficult to individually measure [22]. The North American Regional Reanalysis (NARR) data was selected because it has the desired temporal and spatial span and resolution. Data is available eight times daily beginning 1 January 1979 with plans for continuing coverage and covers all of North America at 0.3° samples (roughly 32 km at the lowest latitude) [9]. Geopotential height (gpm) (converted to geometric height (km)), temperature (K), and specific humidity (kg/kg) (converted to relative humidity (%)) are used at 29 pressure levels as inputs to MODTRAN. Like the methods discussed in Section 2.2, standard atmospheres are appended to the top of these profiles to create complete characterizations from ground to space.

The necessary radiative transfer parameters are not explicit outputs of MODTRAN. A boundary (surface) temperature (corresponding to some spectral radiance due to temperature, L_T , based on the spectral response function of the sensor) and surface albedo (corresponding to some spectral emissivity, ϵ) are input into MODTRAN and a spectral observed radiance, $L_{obs\lambda}$, is output. There are a number of different methods for determining the effective radiative transfer parameters for a particular atmosphere given the MODTRAN outputs. This method was chosen for both accuracy and computational efficiency. If $\epsilon = 1$, Equation (7) simplifies to the linear equation shown in Equation (8), where the λ_{eff} subscript indicates integration over the spectral response function. Effective spectral values will be used (and assumed so explicit notation will be dropped) throughout this treatment of the methodology.

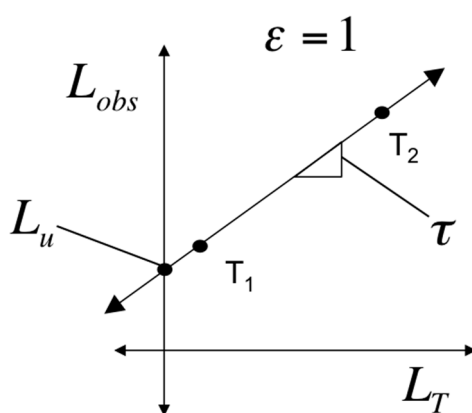
$$L_{obs\lambda_{eff}} = L_{T\lambda_{eff}}\tau_{\lambda_{eff}} + L_{u\lambda_{eff}} \quad (8)$$

Therefore, after two MODTRAN runs with different surface temperatures (T), a linear regression can be used to determine the effective transmission and upwelled radiance as shown in Figure 10. Boundary temperatures of 273 K and 310 K were used for these MODTRAN runs in order to encompass most temperatures that would be encountered in the process. Given the transmission and upwelled radiance, a third MODTRAN run can be used with any $\epsilon < 1$ (*i.e.*, 0.9) to solve the rearranged governing equation

for the downwelled radiance as shown in Equation (9). To ensure that the boundary temperature is relatively close to the land surface temperatures that these variables will be used to solve for, the temperature of the lowest layer of the atmosphere is used as the boundary temperature for this MODTRAN run. These operations will generate the necessary radiative transfer parameters for a specific NARR point location at a specific ground altitude. The input atmospheric profiles (and ground altitudes) can be adjusted in MODTRAN so that this process can be used to generate radiative transfer parameters for different heights at the same NARR point location.

$$L_d = \frac{\frac{L_{obs} - L_u - L_T \epsilon}{\tau}}{1 - \epsilon} \tag{9}$$

Figure 10. Illustration of regression to calculate transmission and upwelled radiance.



4.2. Temporal, Elevation, and Spatial Interpolators

Because the NARR data and Landsat data are not sampled at the same temporal or spatial resolution, various interpolations are required to generate accurate answers for each pixel in the Landsat scene. The NARR profiles at time samples before and after the Landsat acquisition time are converted to the appropriate variables as discussed above and then linearly interpolated to the Landsat acquisition time. For example, NARR profiles from 12 Z and 15 Z would be linearly interpolated to the Landsat acquisition time of 14.3 Z. These appropriately timed profiles are then input into MODTRAN.

As mentioned above, the methodology described in Section 4.1 can be used at various heights at the same NARR point location, so the radiative transfer parameters are calculated at nine fixed heights for every NARR point location within a scene. Once the radiative transfer parameters have been generated at various heights, interpolations in elevation and location are necessary in order to determine the radiative transfer parameters associated with each pixel. The radiative transfer parameters for each pixel, along with the associated emissivity and ToA radiance, can then be used to solve the governing equation (shown in Equation (9)) for the surface radiance L_T , which can be inverted to temperature.

For each pixel, the elevation is determined from the USGS supplied digital elevation model and the four surrounding NARR points are identified. At each of these four NARR points, only the radiative transfer parameters from the heights closest above and closest below the elevation of the current pixel are piecewise linearly interpolated to the elevation of the current pixel. This results in radiative transfer parameters at the appropriate elevation associated with the four most relevant NARR locations.

Finally, these four values for each radiative transfer parameter are spatially interpolated to the location of the pixel. Equations (10)–(12) illustrate Shepard’s method, an inverse distance weighting method. Equation (10) calculates a distance (d_i) from each point (NARR location, x_i, y_i) to the interpolated point (pixel location, x, y). Equation (11) calculates a weight function (w_i) for each point based on these distances and the number of points, $n = 4$. The weighting exponent p is an arbitrary real number and generally set to 2. Finally, Equation (12) calculates the interpolated value ($F(x, y)$) as a weighted summation of the original points (f_i). This method is used to interpolate transmission, upwelled radiance, and downwelled radiance to the location of each pixel [23].

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (10)$$

$$w_i = \frac{d_i^{-p}}{\sum_{j=1}^n d_j^{-p}} \quad (11)$$

$$F(x, y) = \sum_{i=1}^n w_i f_i \quad (12)$$

These per pixel radiative transfer parameters are written out as additional bands in the Landsat image. One more band will be added to provide the ASTER based emissivity value for each pixel using a methodology being developed by our colleagues at JPL [24]. Since our focus here is only the atmospheric compensation component of the LST product, the validation of the methodology will focus on a single target class (water) with known spectral emissivity that is effectively constant over the range of view angles observed by Landsat [25].

4.3. Comparison to Water Surface Temperatures for Cloud Screened Data Set

It is difficult to directly measure the land surface temperature of many surfaces without altering the temperature of the surface being measured. Water temperatures are more easily measured by submerging and acclimating a temperature transducer; also, it is exceedingly helpful for this validation that the emissivity of water is well known because we can estimate the water temperature for any water pixel in a Landsat scene using the atmospheric compensation methodology described above. A subset of the NOAA buoys described above, along with validation data from the JPL sites at Lake Tahoe and the Salton Sea, were used to validate the LST atmospheric compensation methodology.

Radiometers in Lake Tahoe and Salton Sea are operated and maintained by JPL. They provided various validation points that were corrected to skin temperature [11]. A number of NOAA buoys were also selected and points were corrected to skin temperature as described in Section 2.1. Locations across the great lakes were used, in Lake Ontario, Lake Huron, and Lake Superior, along with several ocean buoys including locations off the Delaware/Maryland Coast, Georgia, and two locations in California near Santa Monica and Santa Barbara. Surface temperatures for corresponding Landsat 5 pixels were sampled from the LST product and compared to the skin temperatures from the buoy measurements. This validation dataset contains 827 points. Errors were calculated as shown in Equation (13); negative values indicate that our method underestimated the temperatures.

$$Error(K) = Predicted\ LST(K) - Ground\ Truth\ Temperature(K) \quad (13)$$

A visual cloud screening was also performed for the validation data set. Clouds are named and classified based on height above the ground and texture; for this classification, distinctions were important in texture (stratus or cumulus) but not height (cirro- or alto-). Cumulus clouds (cumulo-prefix meaning heap) are clouds with individual or cellular elements that can line up in streets or rows; for this categorization, cumulus clouds also include cirrocumulus and altocumulus. Stratus clouds (strato-prefix meaning layer) are more uniform, widespread and layered. This classification of stratus clouds includes cirrostratus, altostratus, stratocumulus, and nimbostratus. Finally, cirrus clouds are wispy, feathery clouds. These are included with stratus clouds in this classification because of their wider spread, rather than defined edges like cumulus clouds. Each Landsat scene, and corresponding buoy location, was classified on a level from 0–5 as shown in Table 1. Because of this visual cloud screening, error can be analyzed at varying and known cloud conditions and therefore can be better understood. This is motivated by the increasingly prevalent use of cloud screened image mosaics to utilize clear pixels from multiple images. Thus, we would like to provide users with an LST product and error estimate not only for clear conditions but also for conditions with clouds in the surround (which will have a somewhat higher expected error).

Table 1. Description of visual cloud screening categories.

Category	Description	Number of Scenes	Percent of Scenes
0	Cloud Free	259	31.3%
1	Cumulus in vicinity	98	11.9%
2	Stratus or Cirrus in vicinity	158	19.1%
3	Cumulus over buoy	60	7.3%
4	Stratus or Cirrus over buoy	202	24.4%
5	Totally cloudy image	50	6.0%

5. Validation of Atmospheric Compensation for LST

In this section, LST retrievals using the methodology described in Section 4 are compared to buoy derived surface temperatures for both Landsat 5 and Landsat 8. This comparison to ground truth is used to validate the LST atmospheric compensation methodology and to further study the initial calibration of Landsat 8.

Comparison of LST Retrievals to Buoy Derived Surface Temperatures for Landsat 5 and Landsat 8

The initial validation data set, using traditionally calibrated Landsat 5 imagery, contained 827 points. As mentioned in Section 4, personal communication with the Landsat Calibration team indicates that Landsat 5 is currently slight miscalibrated, based on traditional methods, with a bias of -0.33 K at 300 K; this bias is corrected for in all scenes processed using the LST methodology. A histogram of errors for the entire dataset is shown in Figure 11. As expected, there are a large number of scenes with large negative errors due to clouds. With all scenes included, 432 scenes (52% of included scenes) fall in the center three bins (-1.5 K, 1.5 K). Figure 12 shows a histogram of the samples categorized into cloud categories 0–2, which includes scenes with clouds in the vicinity but not scenes with clouds over the buoy. There are 515 scenes in these three cloud categories; 395 scenes (77% of the included scenes) fall

in the center three bins. These figures are included to illustrate the implications of clouds in the scene and in the surround; we expect this to become important in the future work to generate expected errors for the user based on the type and proximity of clouds. Finally, Figure 13 shows a histogram containing only scenes without clouds near the buoy (categorization 0). There are 259 scenes included in this histogram, and 233 scenes (90% of the included scenes) fall within the center three bins. This figure is a better illustration of the validation of our implemented process and methodology.

Figure 11. Histogram of errors for all Landsat 5 scenes in validation data set.

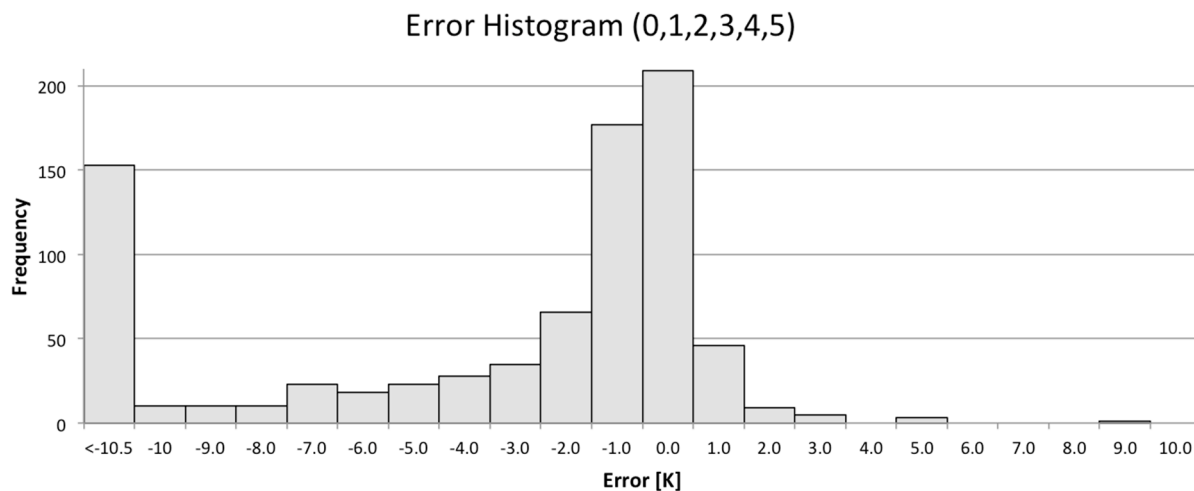


Figure 12. Histogram of errors for Landsat 5 scenes with possible clouds in the vicinity but not over the buoy.

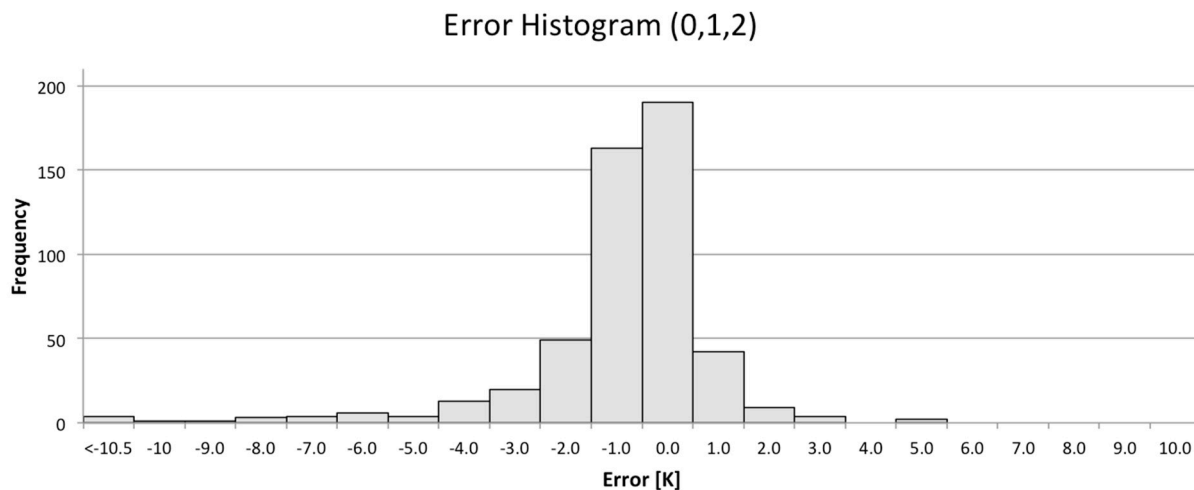


Table 2 contains a summary of results for the validation using Landsat 5 imagery. The mean error and standard deviation are for the distribution of comparisons between the predicted and buoy temperatures. As expected, errors are quite large when all scenes are included, but results are very good for cloud free scenes. For cloud free scenes, the magnitude of the mean error is less than 0.3 K; these results are very encouraging for the validation of our methodology and this magnitude of errors is well matched to the targeted applications. Because the atmospheric data comes from numerical models, rather

than direct observations, the slight negative bias can likely be attributed to small amounts of water vapor that are not accounted for in the atmospheric input data.

Figure 13. Histogram of errors for Landsat 5 scenes without clouds near the buoy.

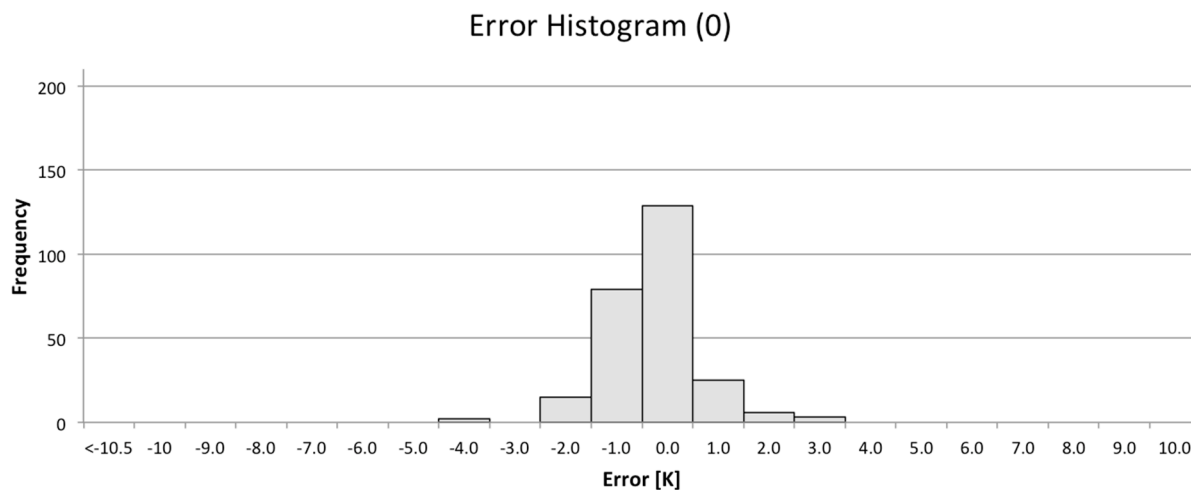
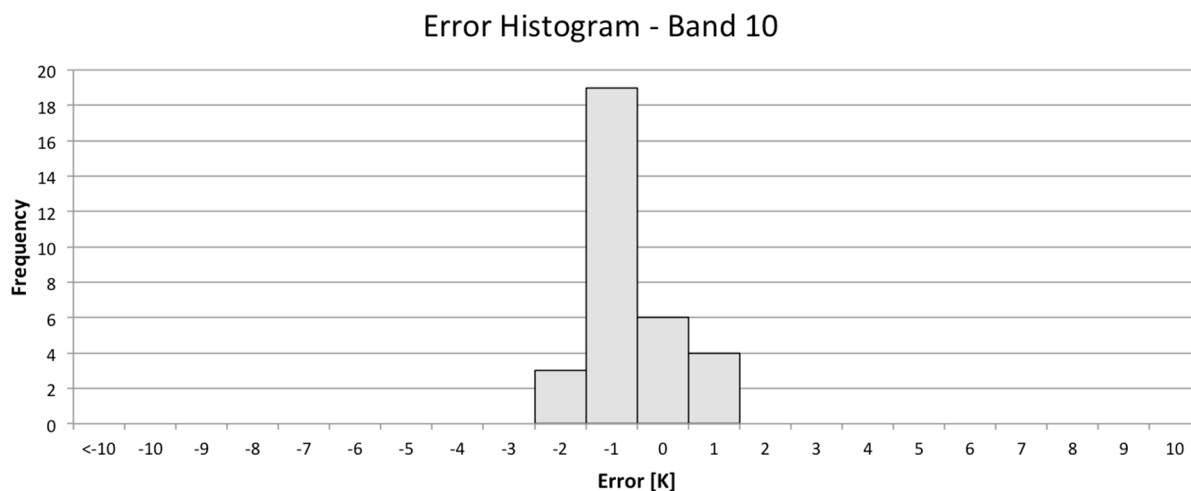


Table 2. Summary of results for Landsat 5 validation.

Cloud Category	Mean Error	Standard Deviation
(0,1,2,3,4,5)	-8.47 K	19.31 K
(0,1,2)	-0.93 K	2.46 K
(0)	-0.27 K	0.89 K

Figure 14. Histogram of errors for Landsat 8 band 10 including only cloud free scenes.



After validation of the methodology using Landsat 5 imagery, a set of cloud free Landsat 8 imagery with corresponding buoy data was also processed. Both bands of Landsat 8 were analyzed separately using a 32 scene cloud free dataset. Figure 14 is a histogram of errors for band 10; of the 32 scenes, 28 scenes have errors within $[-1.5\text{ K}, 1.5\text{ K}]$. A histogram of errors for band 11 is shown in Figure 15; of the same 32 scenes, only 17 scenes have errors within $[-1.5\text{ K}, 1.5\text{ K}]$.

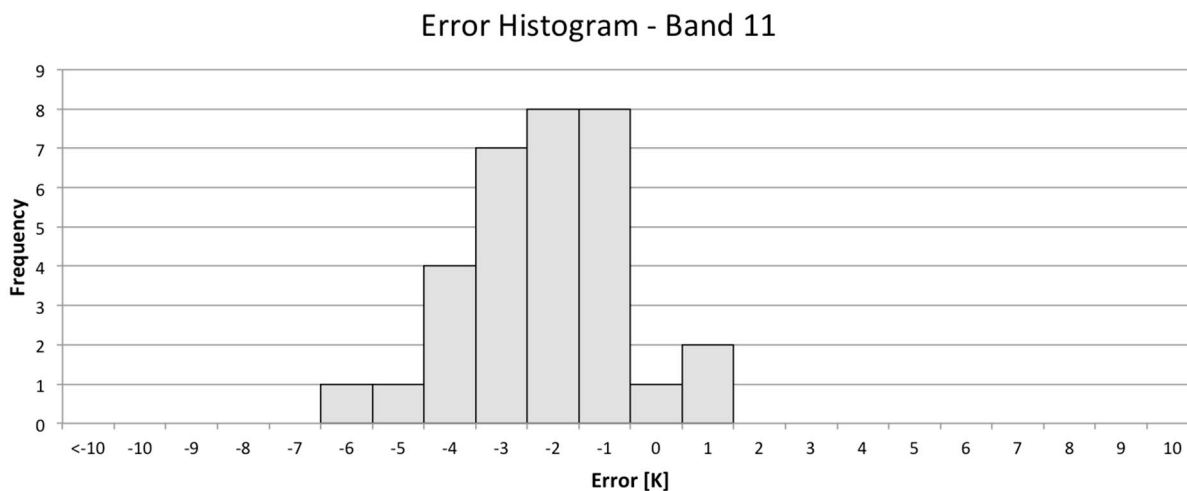
Figure 15. Histogram of errors for Landsat 8 band 11 including only cloud free scenes.

Table 3 contains a summary of Landsat 8 validation results. While both bands still have a negative bias, band 10 has errors with magnitudes greater than Landsat 5 but still less than 1 K and reasonable for the targeted applications. Band 11 has errors with magnitudes greater than 2 K and a larger standard deviation, making these LST results less trustworthy. As discussed in Section 3.2, these results are consistent with the calibration results and will be utilized in the continuing calibration work and solutions to the stray light problem.

Table 3. Summary of results for Landsat 8.

Band	Mean Error	Standard Deviation
10	-0.56 K	0.76 K
11	-2.16 K	1.64 K

6. Conclusions and Next Steps

The use of the NOAA moored buoy data and operational processing has been shown to generate calibration biases that are in good agreement with the traditional approach with differences in the methods of 0.19 K for Landsat 5 and little or no difference for Landsat 7. Based on this successful test of the methodology, the method was used to evaluate the updated calibration of Landsat 8 and showed band 10 to be in good calibration with a bias of $-0.01 \text{ K} \pm 0.90 \text{ K}$; the large standard deviation is indicative of ongoing calibration issues associated with stray light. In band 11, where the stray light issue is approximately 2.5 times worse, the bias was found to be $-0.91 \text{ K} \pm 1.28 \text{ K}$, indicating that band 11 should not be used for quantitative analysis at present.

The validation of the atmospheric compensation component of the LST showed quite small residual errors of $-0.27 \text{ K} \pm 0.89 \text{ K}$ for cloud free data and acceptable errors for some applications of $-0.93 \text{ K} \pm 2.46 \text{ K}$ when clouds were observed in the surround. The method appears to have a slight negative bias that worsens in the presence of clouds. This is most likely due to unaccounted water vapor in the atmosphere. Ongoing efforts are being directed at trying to identify and correct for this small effect as part of the data processing or to determine the impact of applying an overall bias to data with clouds in the surround to reduce the residual bias error in the final product.

Reflecting the increased variance in the Landsat 8 band 10 calibration, due to ongoing stray light issues, the atmospheric compensation component of the LST product for cloud free data showed residual errors of $-0.56 \text{ K} \pm 0.76 \text{ K}$ which should allow for the generation of a band 10 only LST product. The results for band 11 were $-2.16 \text{ K} \pm 1.64 \text{ K}$ and will require compensation for the stray light effects before a band 11 or a combined band 10-band 11 LST product can be generated.

The atmospheric compensation tool described here is now being combined with the JPL emissivity compensation tool to produce an overall LST product with a data quality band that will help users characterize the expected errors in the observed temperatures.

Ongoing efforts for thermal calibration are being primarily focused at understanding the Landsat 8 ghosting issues [19] to attempt to reduce the surround radiance dependent bias and the larger variation in observed calibration errors. In addition, additional filters are being considered to attempt to remove points with clouds in the surround, which appear to introduce small errors in the automated calibration approach.

Acknowledgments

The authors would like to thank the sponsors of this research. This work was funded under NASA grants NNX11AG70G, Radiometric Calibration of Landsat Thermal Data: Development and Test of an Operational Approach, and NMO710627, A Land Surface Temperature Product for Landsat.

Author Contributions

All authors contributed to this work.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Hook, S.J.; Chander, G.; Barsi, J.A.; Alley, R.E.; Abtahi, A.; Palluconi, B.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.
2. Schott, J.R.; Barsi, J.A.; Nordgren, B.L.; Raqueño, N.G.; De Alwis, D. Calibration of Landsat thermal data and application to water resource studies. *Remote Sens. Environ.* **2001**, *78*, 108–117.
3. Padula, F.P.; Schott, J.R.; Barsi, J.A.; Raqueño, 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.* **2011**, *36*, 617–630.
4. Schott, J.R.; Hook, S.J.; Barsi, J.A.; Markham, B.L.; Miller, J.; Padula, F.P.; Raqueño, N.G. Thermal infrared radiometric calibration of the entire Landsat 4, 5, and 7 archive (1982–2010). *Remote Sens. Environ.* **2012**, *122*, 41–49.

5. Barsi, J.A.; Markham, B.L.; Schott, J.R.; Hook, S.J.; Raqueño, N.G. Twenty-five years of Landsat thermal band calibration. In Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, USA, 25–30 July 2010; pp. 2287–2290.
6. Barsi, J.A.; Schott, J.R.; Hook, S.J.; Markham, B.L. TIRS vicarious radiometric calibration. *Remote Sens.* **2014**, pending online.
7. National Oceanic and Atmospheric Administration. Available online: <http://www.ndbc.noaa.gov/> (accessed on 4 August 2014).
8. Berk, A.; Anderson, G.P.; Bernstein, L.S.; Acharya, P.K.; Dothe, H.; Mathew, M.W. MODTRAN4 radiative transfer modeling for atmospheric correction. *Proc. SPIE* **1999**, *3756*, 348–353.
9. Shafran, P. North American Regional Reanalysis. Available online: <http://www.emc.ncep.noaa.gov/mmb/rrean/> (accessed on 20 March 2013).
10. Hulley, G.C.; Hughes, C.G.; Hook, S.J. Quantifying uncertainties in land surface temperature and emissivity retrievals from ASTER and MODIS thermal infrared data. *J. Geophys. Res.* **2012**, *117*, D23113.
11. Hook, S.J.; Vaughan, 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.
12. Zeng, X.; Zhao, M.; Dickinson, R.E.; He, Y. A multiyear hourly sea surface skin temperature data set derived from the TOGA TAO bulk temperature and wind speed over the tropical pacific. *J. Geophys. Res.* **1999**, *104*, 1525–1536.
13. Donlon, C.J.; Minnett, P.J.; Gentemann, C.; Nightingale, T.J.; Barton, I.J.; Ward, B.; Murray, M.J. Toward improved validation of satellite sea surface skin temperature for climate research. *J. Clim.* **2002**, *15*, 353–369.
14. Schott, J.R. *Remote Sensing: The Image Chain Approach*, 2nd ed.; Oxford University Press: New York, NY, USA, 2007.
15. Padula, F.P.; Schott, J.R. Historic calibration of the thermal infrared band of Landsat-5 TM. *PE&RS* **2010**, *76*, 1225–1238.
16. USGS Earth Explorer. Available online: <http://earthexplorer.usgs.gov/> (accessed on 30 October 2014).
17. NOAA/ESRL Radiosonde Database. Available online: <http://www.esrl.noaa.gov/raobs/> (accessed on 7 November 2014).
18. Weather Underground Historical Weather. Available online: <http://www.wunderground.com/history/> (accessed on 7 November 2014).
19. Schott, J.R.; Raqueno, N.; Montanaro, M.; Ientilucci, E.; Gerace, A.; Lunsford, A.W. Chasing the TIRS ghosts: Calibrating the Landsat 8 thermal bands. *Proc. SPIE.* **2014**, doi:10.1117/12.2063236.
20. Montanaro, M.; Tesfaye, Z.; Lunsford, A.; Wenny, B.; Reuter, D.; Markham, B.; Smith, R.; Thome, K. Preliminary on-orbit performance of the Thermal Infrared Sensor (TIRS) on board Landsat 8. *Proc. SPIE.* **2013**, doi:10.1117/12.2025650.
21. Barsi, J.A.; Schott, J.R.; Palluconi, F.D.; Hook, S.J. Validation of a web-based atmospheric correction tool for single thermal band instruments. *Proc. SPIE.* **2005**, doi:10.1117/12.619990.
22. Rienecker, M.; Gass, J. Merra: Modern-Era Retrospective Analysis for Research and Applications. Available online: <http://gmao.gsfc.nasa.gov/research/merra/> (accessed on 20 March 2013).

23. Shepard, D. A two-dimensional interpolation function for irregularly-spaced data. In Proceedings of the 1968 ACM National Conference, New York, NY, USA, 27–29 August 1968; pp. 517–524.
24. Hulley, G.C.; Hook, S.J. The North American ASTER land surface emissivity database (NAALSED) version 2.0. *Remote Sens. Environ.* **2009**, *113*, 1967–1975.
25. Masuda, K.; Takashima, T.; Takayama, Y. Emissivity of pure and sea waters for the model sea surface in the infrared window regions. *Remote Sens. Environ.* **1988**, *24*, 313–329.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).