Abstract—The Thermal Infrared Sensor-2 that will be on-board Landsat 9 has undergone a pre-launch testing campaign to characterize its radiometric, spectral, and spatial performance and demonstrate compliance to its requirements. This work reviews key elements of the instrument-level radiometric testing using an SI traceable source to derive its uncertainties. Those arising from on-orbit calibration using the TIRS-2 on-board blackbody are also discussed. We use a Monte Carlo approach to propagate the uncertainties through a non-linear calibration equation and address both random and systematic uncertainty terms. Achieving the required performance demonstrates the instrument’s potential for enhancing our understanding of the Earth’s environment.

Index Terms—Landsat 9, Thermal Infrared Sensor 2, calibration, uncertainty, spectral response, pre-launch testing

I. TIRS-2 INSTRUMENT AND REQUIREMENTS OVERVIEW

The Thermal Infrared Sensor-2 (TIRS-2), expected to launch on Landsat 9 in 2021, will continue the Landsat Program’s legacy of providing moderate resolution thermal imagery over almost four decades. Scientists use thermal imagery provided by the Landsat satellites for a wide variety of environmental applications such as cloud detection and masking [1], [2], evapotranspiration studies [3], [4], water use assessments [5], urban heat fluxes mapping [6], [7], burnt area mapping [8], and vector-borne illness potential identification [9], [10]. Many of these applications are enabled through land surface temperature retrievals from the thermal imagery. TIRS-2, which is a functional copy of TIRS, its predecessor on Landsat 8 [11], has two bands at 10.8 µm and 12.0 µm, enabling more accurate retrievals than previous Landsat thermal sensors where only a single channel was available [12]. A comprehensive survey of applications enabled by land surface temperature retrievals can be found in Ref. 13.

TIRS-2 will produce radiometrically SI-traceable, geo-located thermal imagery of the Earth with the same parameters as Landsat 8 TIRS: 16-day repeat, 100-m spatial sampling (resampled to 30 m in the final product), 185-km swath width, and 70 frames-per-second operation. The instrument is a pushbroom sensor with a 15° cross-track field of view with the same basic architecture as TIRS but with some improvements such as increased electrical redundancy and improved stray light suppression [14]. It has a f/1.6 four-lens telescope that focuses onto three quantum well infrared photodetector (QWIP) arrays usually termed sensor chip assemblies (SCAs) [15], an on-board blackbody for calibration, and a scene-select mirror for switching between Earth view, blackbody, and space views (Fig. 1). The on-board calibrator (OBC), the same design as TIRS, is modeled after the blackbodies used as part of other heritage systems: the Visible Infrared Radiometer Suite (VIIRS) and the Moderate Resolution Imaging Spectroradiometer (MODIS) [16].

The three QWIP arrays are 512 rows by 640 columns each. A filter for each channel covers ~70 rows of each SCA with an opaque section in between for dark count monitoring (Fig. 1). In its operational mode, only two rows are read out from each channel and dark region. A combination of these detectors from the two rows that meet performance requirements form an effective field of view that spans the 185 km swath width and are transmitted to the ground. Those two rows are chosen based on component- and instrument-level radiometric and spectral characterization tests (discussed in the next section). The primary and redundant rows are combined in the ground processing system to produce the standard Landsat scene image product. Any under-performing individual detector from the primary row can be replaced with the corresponding detector in the redundant row to produce a “perfect” row of data for the image product. The two rows may be changed in flight should it become necessary, although the rows have been stable throughout the Landsat 8 mission and similar performance is expected for Landsat 9.

This work addresses the absolute radiometric uncertainty requirement, interpreted as uncertainty. Table I shows the accuracy and noise requirements for TIRS-2 (and TIRS) and the TIRS on-orbit performance, which is discussed further in

| Parameter | TIRS/TIRS-2 requirement | TIRS performance |
|-----------|-------------------------|------------------|
| NEdT @ 300 K (K) | <0.4 | 0.05 [17] |
| NEdL (W/m²/sr/µm) | <0.059, 0.049 | 0.008 [17] |
| Absolute radiometric accuracy (%) | < 2.4 | 1 [18] * |

NEdL values denote the 10.8 µm and 12.0 µm channels, respectively and radiometric accuracy values refer to the nominal brightness temperature range of (260-330 K) and extended range (240-260, 330-360 K), respectively.

*With stray light correction implemented
discussing the uncertainty budget, we first present an overview of the relevant pre-launch testing methodologies and results, followed by the uncertainty methodology.

II. PRE-LAUNCH CALIBRATION AND CHARACTERIZATION

A. Pre-launch Testing Overview

The pre-launch testing – with characterization at various integration phases – follows best practices established by the remote sensing community [22]. The integration phases include component-level, sub-assembly-level, instrument-level, and spacecraft-level testing to address the radiometric requirements (including the ones introduced earlier), as well as spatial, spectral, and geometric requirements [19], [23]. Fig. 2 shows some of the key measurements in all testing phases. Many tests are repeated at different phases to ensure the validity of their results. For instance, the relative spectral response (RSR) is derived independently through the component-level filter transmittance and detector relative response measurements, as well as through instrument-level measurements to ensure their consistency [24]. Spectral characterization was also performed at sub-assembly level in order to refine test setup and sampling strategy in advance of the instrument-level characterization (Fig. 2) [25]. The instrument-level spectral measurements provided lower uncertainties in establishing the channel average spectral responses, so they are expected to be used in operations rather than those derived via component-level measurements. Thus, the uncertainties of the instrument-level RSR will be discussed thoroughly here (and those derived via component-level measurements, considered as a validation, are discussed in Ref. 24).

The radiometry (calibration) measurements were done only at instrument-level, since this is the only configuration where the reference blackbody source could be installed and the final instrument electronics were integrated. This full aperture reference blackbody source called the flood source used for instrument-level testing is the basis for the TIRS-2 radiometric calibration. Its SI traceability was first established for TIRS testing in 2012 and re-characterized for use for the TIRS-2 program in 2018 at the Space Dynamics Laboratory (SDL) [26]. This was done through comparison to a standard blackbody source called the Long-Wave Infrared Calibration Source (LWIRCS) via an SDL transfer radiometer [27]. The LWIRCS spectral radiance scale is tied to both the NIST-calibrated temperature sensors and cavity model as well as through the NIST Transfer Radiometer (TXR) [28]. The effective emissivity of 0.9920 was determined with an uncertainty of 0.0023, which corresponds to the uncertainty in the spectral radiance scale including the TXR transfer uncertainty. The emissivity results showed no significant changes between the 2012 and 2018 measurements. This TIRS-2 response to the flood source illumination is compared to the TIRS-2 on-board blackbody illumination to establish an on-orbit calibration capability (Fig. 2)

The instrument-level testing occurred at Goddard Space Flight Center in two phases with vibration/acoustic testing performed in-between. Together, these phases verified the stability of the hardware and repeatability of its calibration

Fig. 1: TIRS-2 schematic with scene select mirror (SSM), telescope lenses, and Sensor chip assembly (SCA). The lower figure shows the SCA layout with three quantum well infrared photodetector (QWIP) arrays and approximate filter positions for the 10.8 µm and 12.0 µm channels and dark regions. (Note that the on-board calibrator (OBC) is not included in the figure.)
and $RSR$ results in a simulated on-orbit environment inside a thermal vacuum chamber. The spatial response, stray light, focus, and bright target recovery tests were also included in this phase. The focus test is done to verify that the image is focused on the detector arrays; the bright target recovery test is done to verify that the detector arrays do not suffer from image memory artifacts. The scatter and spatial response characterize potential image artifacts due to spatial scene non-uniformities. As mentioned, the analysis here will address radiometric uncertainties arising from uniform scenes only, where the results from the radiometry and $RSR$ tests are most relevant. The stray light tests including far-field stray light and near field ghosting artifacts are outside the scope of this paper [29].

B. Instrument-Level Calibration

During thermal vacuum radiometry (calibration) testing, the flood source illuminates the TIRS-2 aperture at varying blackbody temperature values in the range of 200 to 360 K to capture the full dynamic range and characterize any non-linear response. During a calibration sequence, the scene select mirror switches between the space view, OBC, and flood source (nadir) view. During each 1-minute view of this sequence, two operational detector rows from each band (10.8 µm, 12.0 µm, and dark) are read out at 70 Hz to generate 4200 samples per detector. Additional diagnostic modes are available where almost every detector of each array is read out. In this way, calibration data are obtained for the flood source and OBC for all potential operational detectors. We will discuss one potential approach to on-orbit calibration using the OBC but the final approach will be determined on-orbit. The space view during testing consists of a cold (~100 K) target inside the thermal vacuum chamber.

The calibration process described here is similar to that of TIRS discussed in Ref. 30. There are three basic steps in the calibration: linearization, TIRS-2 flood source signal retrieval, and calculation of the flood source effective spectral radiance. The linearization process is required because the electronic gain is not constant through its range of digital counts. It has two distinct gains with a transition region between them that are revealed if the integration time is swept while illuminated by a stable radiance (Fig. 3). The signals are linearized so that the difference in background between the blackbody and space view:

$$\epsilon L + L_{bk} = q\Delta C^2 + m\Delta C + b$$  \hspace{1cm} (1a) \\
$$\epsilon L = q\Delta C^2 + m\Delta C$$  \hspace{1cm} (1b)

The background term present in Eq.1a can justify an offset term in the calibration equation. We neglect this term, however, since the background is small (derived as 0.0039 ± 0.0128 W/m²/sr/µm and 0.0593 ± 0.0135 W/m²/sr/µm across all detector columns for the 10.8 µm, 12.0 µm channels, respectively using Eq.1a) and could deviate on-orbit from the thermal vacuum test conditions, to obtain Eq.1b — our primary calibration equation used in this work.

C. Instrument-Level Relative Spectral Response

The calibration equation uses the channel average $RSR$ to calculate effective spectral radiance. The measurement setup for $RSR$ consists of the setup depicted in Fig. 2 and custom calibration ground support equipment (GSE) inside the thermal vacuum chamber. The calibration GSE has capabilities to test radiometric (via the integrated flood source), geometric, and spectral performance of a sensor under thermal vacuum conditions. There is a special mode designed for spectral measurements, where a beam from the monochromator-based setup outside the thermal vacuum chamber can propagate through the calibration GSE and onto the designated TIRS-2 detectors. The blackbody-monochromator setup consists of a 1000ºC blackbody source filling the entrance slit of a monochromator (with 50 l/mm grating with blaze wavelength of 12 µm, and reciprocal dispersion 78.3 nm/mm). The monochromator output is collimated and directed into the chamber through a ZnSe window and to the TIRS-2 aperture via the calibration GSE optics. The monochromator wavelength is swept through a range of wavelengths covering the desired channel, and at each spectral interval, detector samples are acquired with the monochromator shutter open and closed. The latter signal is subtracted from the former to remove the background and obtain the TIRS-2 signal ($dn_{TIRS}$). A similar wavelength sweep is conducted subsequently with the monochromator output beam directed to a reference detector (liquid nitrogen-cooled HgCdTe detector) with a calibrated relative responsivity ($R_{ref}$). To separate the blackbody signal from the large ambient background, the blackbody signal is chopped to enable lock-in amplification of the optical signal. The $RSR$ is derived as

$$SR_{TIRS}(\lambda, pix) = \frac{dn_{TIRS}(\lambda, pix)}{\tau_{TIRS path} \frac{R_{ref}}{V_{ref}}} \times \frac{\tau_{ref path}}{R_{ref}}$$  \hspace{1cm} (2a) \\
$$RSR_{TIRS}(\lambda, pix) = \frac{SR_{TIRS}(\lambda, pix)}{\max_{\lambda}(SR_{TIRS}(\lambda, pix))}$$  \hspace{1cm} (2b)

The $dn_{TIRS}$ with a correction for different optical path transmittances to the reference detector ($\tau_{ref path}$) and to the TIRS-2 aperture ($\tau_{TIRS path}$) is combined with a reference
Fig. 2: Overview of pre-launch testing including key measurements at component, sub-assembly, instrument, and spacecraft-level testing. The calibration uncertainty is derived from instrument-level measurements of $RSR$ and radiometry tests. The SI traceable flood source and on-board calibrator shown on the bottom left establish the pre-launch calibration and on-orbit calibration capability. The blackbody-monochromator setup shown on the right is used to measure the $RSR$ at both sub-assembly and instrument-level test phases.

Further details of the measurements are described in Refs. [31], [32] and an uncertainty budget is included in Ref. 32, which we will review in a later section. Once the $RSRs$ were measured, they were used to derive effective spectral radiance used in the calibration equation.

III. Uncertainty Budget Methodology

This work addresses the TIRS-2 absolute radiometric uncertainty and SI traceability derived from the measurements described previously. Such uncertainty evaluations have been conducted for heritage and operational sensors such as VIIRS, MODIS, Geostationary Operational Environmental Satellite (GOES) Imager, and GOES-R series Advanced Baseline Imager (ABI). These sensors generally follow the best practices for establishing SI-traceability and uncertainty analysis [33] and have undergone extensive pre-launch testing to characterize their behavior in operational conditions [34]–[37]. The uncertainty budgets account for the major uncertainty contributors in their calibration equations. Similar parameters were explored (with some added/omitted) based on their unique design attributes and addressed in the calibration. We also follow such best practices by assessing all major uncertainty contributors and include them in the calibration equation. The quality of the fit using the TIRS-2 calibration equation (Eq. 1b) give us confidence in this model that forms the basis of the uncertainty budget. The parameters defined in the equation include the fit coefficients, measured signal, and effective spectral radiance of the blackbody given by its temperature sensors, emissivity values, and $RSR$. These parameters along with additional terms to account for systematic differences from repeated measurements form a comprehensive uncertainty budget.

The parameter uncertainties are propagated through the measurement (calibration) equation (Eq. 1b) to derive the uncertainty contribution to radiometric uncertainty expressed with a coverage factor $k = 1$. There are several approaches to propagating uncertainties through a calibration equation: analytically, computationally via perturbation, and computationally via a Monte Carlo approach [38], [39]. The analytical approach refers to using the propagation of uncertainty formula. The perturbation method refers to perturbing the mean measured value by its uncertainty and propagating through the measurement equation to find the resulting radiometric uncertainty. Both of these methods give equivalent results and are used for VIIRS and MODIS uncertainty analyses, respectively [35], [40].

The drawback of these methods is that they may be less accurate when the measurement equation is non-linear. Thus, we instead computed the uncertainty contributions via a Monte Carlo computational approach, which does not have such a limitation and is consistent with established standards for uncertainty analysis [39]. Here we choose a probability distribution, a Gaussian distribution typically, and generate random numbers with the mean measured value and standard deviation corresponding to its $k = 1$ uncertainty and propagate them through the calibration equation to establish an output distribution of radiance values. The standard deviation of the resulting radiance values is the radiometric
Fig. 3: (a) Integration time sweep showing the lower, transition, and upper regions. The raw counts are converted to match the slope of the upper region (b) The curve shows the conversion over the entire range of digital counts.

Fig. 4: (a) Average RSR over all SCAs obtained from instrument-level testing for each channel. The shading represents the standard deviation across all locations.

uncertainty. Note that the RSR measurement equations are linear (Eq. 2), so propagating their uncertainties followed the perturbation approach through the RSR calculation and then converting to effective spectral radiance assuming blackbody illumination. Note that the Monte Carlo approach avoids any inaccuracy due to non-linearities and gives a convenient way of addressing correlations between parameters by simply replacing the Gaussian-distributed random numbers with bivariate-Gaussian-distributed numbers as is done for the highly-correlated fitting coefficients [41]. For comparison, the VIIRS analysis in Ref. 35 handled correlations by calculating covariance terms in the uncertainty propagation formula or estimating their upper bounds.

Once the radiometric uncertainties were determined for all parameters, the combined uncertainty $u_c$ is calculated as the root sum squared of the uncertainty components:

$$u_c^2(y) = \sum_{n=1}^{N} u_i^2$$

The uncertainties can not be combined strictly according to this approach, however, due to some terms involving uncorrected biases. This leads to some complications in combining uncertainties. To our knowledge, there have been no other attempts to address uncorrected bias in previous pre-launch uncertainty analyses for satellite sensors. Such biases should be corrected if possible, but this is not realizable for some parameters; thus, such uncorrected biases are often neglected until post-launch. By including the uncorrected biases, the results can include these terms but remain independent of the correction methodology applied post-launch. We follow the method described in Ref. 42 referred to as the sum uncertainties method (SUMU) to express and combine uncertainties by adding biases to obtain the combined bias ($\delta_c$), taking the root sum squared of random uncertainties including the bias uncertainties to obtain the combined random uncertainty $u_c$, and forming a confidence interval corresponding to the total combined $k = 1$ uncertainty:

$$U_+ = ku_c - \delta_c$$

$$U_- = ku_c + \delta_c$$

$$y \begin{cases} U_+ \\ -U_- \end{cases}$$

Most of the uncertainty terms are derived based on thermal vacuum test data at nominal (expected on-orbit) conditions. The exceptions are in evaluating calibration reproducibility, where data are taken at instrument temperatures outside the nominal on-orbit conditions to obtain worst case values. The
uncertainties will be expressed for both the nominal and extended brightness temperature ranges as defined in the requirements.

IV. Uncertainty Results and Discussion

The uncertainty budget is roughly organized into pre-launch and on-orbit uncertainties. The pre-launch uncertainties are associated with the flood source-based calibration and RSR measurements. Most of the on-orbit uncertainties are tied to the calibration using the OBC (although some flood-source related terms are included in the combined on-orbit uncertainty as will be discussed). The on-orbit uncertainty total refers to an implementation of the OBC calibration without attempting to make corrections to match the flood source-based pre-launch calibration.

The flood source calibration radiance scale uncertainty was previously discussed in Section II. An additional flood source-related uncertainty is its temperature uncertainty given by the standard deviation of the mean of its two temperature sensor values in the 200-360 K range. A Monte Carlo method was used to propagated the temperature uncertainty through the calibration equation: The effective spectral radiance values are calculated using Gaussian-distributed temperature values and fitted with the corresponding ΔC to obtain a set of calibration curves (radiance versus ΔC). The uncertainty is calculated by taking the average standard deviation of the radiance at each ΔC and dividing by the radiance. A wide range of radiance values are included to cover the brightness temperature ranges for comparison to the requirements. The maximum of the nominal and extended ranges are compared to the requirements for operational detectors in all array columns. The average uncertainty (over all detectors) is 0.067 % in the nominal range and 0.082 % in the extended temperature regions for the 10.8 μm channel and 0.059 % and 0.072 % in the respective ranges for the 12.0 μm channel.

The calibration using the flood source is also affected by the TIRS-2 count noise, which is approximately the noise (NEDL) divided by the square root of the number of samples. This turns out to be negligible since the instrument noise is low (10.8 μm channel: 0.005-0.010 W/m²/sr/μm, 12.0μm channel: 0.006-0.010 W/m²/sr/μm — similar to TIRS performance (Table I)) and a large number of samples are taken (4200 in each flood source and space view). This yields 0.0046 % in the nominal range and 0.0054 % in the extended temperature regions for the 10.8 μm channel and 0.0071 % and 0.0084 % in the respective regions for the 12.0 μm channel.

The fitting uncertainty was originally calculated with the perturbation method. The results showed, however, that this method overestimated the uncertainty due to the non-linear calibration equation. To generate a more accurate fitting uncertainty, we switched to a Monte Carlo methodology while also accounting for the fitting coefficient correlations (calculated using detectors across all columns). The quadratic fit with confidence intervals for each coefficient corresponding to a k = 1 coverage factor is calculated. Random fitting coefficients are generated with a bivariate Gaussian distribution based on these confidence intervals and correlation coefficients. These coefficients are used to generate calibration curves (radiance as a function of ΔC). This process is illustrated in Fig. 5. The standard deviation of these radiance curves gives the uncertainty. This process is repeated for all operational row detectors and shown in Fig. 6 (a). The maximum uncertainty in the nominal and extended brightness temperature regions are then averaged over the detectors to obtain the final uncertainty values. For the 10.8 μm, the uncertainty is 0.37 % (nominal) and 0.41 % (extended), and for 12.0 μm channel, 0.22 % (nominal) and 0.27 % (extended).

The uncertainty of the linearization process is captured by evaluating its reproducibility. We derive a linearization based on two different sets of calibration data taken at nominal and non-nominal conditions. The difference between their impact on radiance gives this additional uncertainty. We found the difference between them to be 0.30 % at nominal and 0.70 % at extended temperature ranges for the the 10.8 μm channel, and 0.04 % at nominal and 0.10 % at extended temperature ranges for the the 12.0 μm channel. The latter channel is affected less because the higher counts levels in this channel lie on the upper section of the linearization curve, where the impact is negligible.

The calibration reproducibility in nominal conditions was also characterized through repeated measurements at selected flood source temperatures. The fit to the eleven flood source temperature points were used to generate the original quadratic fit. This fit was then applied to the signal at the repeated measurements at temperatures of 260, 270, 290, and 300 K. The radiometric difference between the original and repeated points shows a clear positive mean bias of 0.22 % and 0.24 % over all detectors and temperatures with an uncertainty of 0.24 % and 0.10 % for the 10.8 and 12.0 μm bands, respectively. The uncertainty of the bias is simply the standard deviation because we are evaluating each detector individually (the standard deviation of the mean does not tell us how well we know the bias). Note that this term was not evaluated at extended temperatures, since these points were not repeated for this test.

To review the RSR uncertainty as derived in Ref. 32, we observe all the terms in Eq. 2a and 2b, which includes reference detector signal uncertainty and TIRS-2 noise. We also included the monochromator wavelength calibration uncertainty, which was determined to be 1 nm by measuring two reference absorption lines of a NIST standard reference material 1921B [43] closest to the TIRS-2 spectral channels and therefore considered negligible. (Note that a third spectral line at 9.352 μm showed a 15 nm bias but this was neglected since its wavelength was far outside the TIRS-2 bands). Another source of wavelength uncertainty, however, is a consequence of the dispersion across the monochromator slit. The RSR is calculated using the TIRS-2 pixels near the maximum signal, which does not exactly correspond to the monochromator wavelength setting. Since this setting corresponds to the center of the slit image, a wavelength correction is applied equal to the distance between the location with the maximum signal and the center of the slit in wavelength. The uncertainty of this correction is 1.5 pixel rows or 15 nm in wavelength. We incorporated the uncertainty contributions of the
Fig. 5: Illustration of Monte Carlo process with correlations. (a) An example of a quadratic radiance versus linearized background subtracted counts fit for one detector is shown. (b) The range of fits generated with random values of fitting coefficients with a correlation coefficient of 0.85 and with their $k = 1$ uncertainties. The inset shows the histogram of the coefficients used to illustrate their bivariate Gaussian distribution.

Fig. 6: Results for the Monte-Carlo-derived fit uncertainties for the (a) flood source-based calibration and (b) on-board calibration in the nominal and extended brightness temperature ranges for each channel. The correlation coefficients of the fitting coefficients ($r$) are responsible for either increasing or reducing the uncertainties. Note that the uncertainty for all SCAs is included (refer to Fig. 1 for SCA layout)

- Reflectance/transmittance spectra of the optical components or reference detector response by using 10 % of their maximum change within the channels (since these uncertainties were not provided by the vendors).
- The TIRS-2 noise term is calculated as the standard deviation of the mean for a typical pixel over its samples calculated per wavelength. Incorporating the RSR non-uniformity across all locations gives the uncertainty introduced by using an average RSR per band to represent all detectors as is planned operationally (Figure 8). This effect and the wavelength uncertainty dominate the spectral uncertainty.
- The radiometric uncertainties from the combined spectral uncertainty are 0.05 % (nominal) and 0.06 % (extended) for the 10.8 $\mu$m channel and 0.10 % (nominal) and 0.15 % (extended) for the 12.0 $\mu$m channel.

- The on-board calibration uncertainties are derived with a similar approach as the flood source-based calibration uncertainties. The main difference is that we establish the traceability of the OBC calibration by relating the on-orbit OBC-based calibration to the flood source-based calibration. Using the same calibration equation, we derive the coefficients and their uncertainties independently with the OBC-based calibration. The blackbody operates in a more limited range of temperatures 270-320 K than the flood source. The fitting uncertainties derived through Monte Carlo analysis are 0.27 % (nominal) and 0.36 % (extended) for the 10.8 $\mu$m channel and 0.19 % (nominal) and 0.35 % (extended) for the 12.0 $\mu$m channel. The low uncertainties for the latter channel are due to the strong anti-correlation between the quadratic
Fig. 7: The reproducibility of the radiometric flood source-based calibration expressed as a percent difference is shown for all detector columns in an operational row for the (a) 10.8 µm and (b) 12.0 µm channels.

Fig. 8: Spectral uniformity for cold (240 K), nominal (300 K), and high (360 K) brightness temperature blackbody targets for the (a) 10.8 µm and (b) 12.0 µm channels shown for an operational detector in each column across all SCAs.

and linear coefficients (correlation coefficient = −0.99) as compared to the strong positive correlation in the former channel (correlation coefficient = 0.78) as shown in Fig. 6 (b).

The flood-source based and OBC-based calibration curves are compared for all detectors (Fig. 9). By taking the maximum differences in the nominal and extended ranges, we derived biases of 0.42 % and 0.49 % in the nominal and extended regions, respectively for the 10.8 µm channel and 0.32 % and 0.40 % in the nominal and extended regions, respectively for the 12.0 µm channel. Note that the biases over the range of blackbody temperature values for the 12.0 µm channel have more variability than for the 10.8 µm channel. The sharp discontinuities are associated with a change in SCA (Fig.1), since detectors on different arrays have larger response differences than the high frequency differences normally observed between columns within a SCA. These small response differences generate fits with slightly different uncertainties. The uncertainties of the biases are 0.20 % (nominal) and 0.22 % (extended) for the 10.8 µm channel and 0.31 % (nominal) and 0.32 % (extended) for the 12.0 µm channel. Note that during operations, these biases can potentially be corrected, and we briefly discuss one approach to making this correction, but leave them uncorrected in the uncertainty budget to show a conservative estimate. For instance, consider the background term in Eq. 1a for both the flood source and OBC views. The radiance difference can be used as a correction term representing the background difference between these two views. This difference varies with OBC/flood source temperature and detector column. Optimally, this background
difference would correspond to the same blackbody temperature difference for each channel, but we found that using different background temperature differences for each channel optimized the correction. This could be due to different view factors or emissivity values from the background source in each channel, which can lead to an apparent temperature difference. If we derive these differences for all detectors and channels, the OBC calibration equation can be corrected to obtain the flood source-based calibration. This correction may be sensitive to the actual on-orbit thermal conditions, so further on-orbit analysis would be recommended before implementing such a correction.

The OBC has four temperature sensors that are averaged to determine its temperature and then converted to effective spectral radiance. The radiometric uncertainty due to the temperature uncertainty is calculated using the same approach as the flood source. The uncertainties are 0.21 % (nominal) and 0.36 % (extended) for the 10.8 μm channel and 0.19 % and 0.35 % for the 12.0 μm channel. To characterize any potential impacts to a small change of OBC view angle, a special test was conducted viewing the OBC at its nominal view angle and through an angle of ±0.5° its nominal angle. We found that this effect introduces a negligible uncertainty of 0.02 % indicating a highly Lambertian source in this range of view angles.

Like the flood source-based calibration reproducibility term, the OBC calibration was characterized through repeated measurements. Since all temperature values were repeated, we compared the calibration curves generated with two independent sets of acquisitions. These fits were then compared through their radiometric differences. The results show a slight positive bias through most of the brightness temperature range: mean bias of 0.071 % and 0.075 % over all detectors and temperatures with an uncertainty of 0.34 % and 0.54 % for the 10.8 and 12.0 μm bands, respectively.

All uncertainty terms without bias are combined according to Eq. 3. The bias terms are added separately and the final $k = 1$ confidence intervals are formed using Eqs. 4 and 5 and are shown in Table II. For uncorrected bias terms, there is both a bias listed as well as the bias uncertainty. The biases cause an asymmetric uncertainty interval under the requirements for the nominal and extended regions for both channels. The combined pre-launch uncertainties include terms associated with flood source and spectral measurements. The on-orbit calibration includes OBC measurement terms as well as several terms included in the pre-launch uncertainty: linearization, RSR, and TIRS-2 noise. The uncertainty terms contributing to the pre-launch and on-orbit totals are depicted on the figure next to the table. The on-orbit total uncertainty can be considered worst case, since in operations, there will likely be a method for correcting the bias between the on-board and pre-launch calibration, which is not included here but will be a subject of future investigation using on-orbit data. Thus, these combined uncertainties can be thought of as giving the likely range of uncertainties during operations. For the worst case (on-orbit uncertainty), the largest edge of the interval is 1.3 % (nominal) and 1.7 % (extended) – well under the required 2 % and 4 % values for the on-orbit values. Note that even if we had considered non-uniform scenes in the uncertainty budget, which would include both far-field and near-field stray light (ghosting) artifacts, these values would increase to 1.9 % (nominal) and 2.2 % (extended). This includes contributions from far field stray light of 0.57 % and 0.83 % for the nominal and extended temperature range, respectively. This assumes a simple correction using an out-of-field temperature of 293 K. More sophisticated corrections, such as those used for TIRS, could also be employed to reduce these uncertainties further [18]. The uncertainty contributions from ghosting are 0.50 % (nominal) and 0.88 % (extended). As mentioned, stray light artifacts will be discussed in more detail in a separate work.

Although this uncertainty analysis is meant to be comprehensive, it neglects several parameters. For instance, the electronics and detector temperature dependence of the calibration have not been discussed here, although it has been characterized in case TIRS-2 is operated in non-nominal
conditions. The on-board calibration methodology may be changed to include corrections to match the flood source-based calibration more closely. The blackbody degradation over the lifetime of the mission is neglected here, since it cannot be accurately estimated; the TIRS OBC (with the same design as TIRS-2) radiance has experienced on-orbit degradation of 0.2 % and 0.1 % per year for the 10.8 and 12.0 \( \mu \text{m} \) channels, respectively, but this could be due to reduction in detector sensitivities or blackbody degradation [17]. This paper also does not consider more complex calibration models in order to simplify the analysis and interpretation of the results.

V. CONCLUSION

Through extensive pre-launch testing, TIRS-2 has been characterized and calibrated and has demonstrated the performance necessary to meet its radiometric uncertainty requirements. TIRS-2 has established SI traceability through its pre-launch testing and has demonstrated an accurate on-board calibration capability. The pre-launch TIRS-2 radiometric uncertainty is less than 0.80 % for 10.8 \( \mu \text{m} \) and 0.58 % for 12.0 \( \mu \text{m} \) channels over source temperatures of 260-330 K corresponding to a brightness temperature uncertainty of about 0.5 K and 0.4 K for the two channels. The uncertainty analysis method implemented provides a general framework for addressing calibration non-linearity, correlated parameters, and uncorrected biases — commonly encountered in pre-launch calibration of remote sensing systems. The expected performance of TIRS-2 is established to give Landsat data users the confidence to continue using Landsat thermal imagery for many important environmental applications after its launch and will also serve as a baseline for future on-orbit validation work. It is expected to meet its users’ needs for a variety of environmental applications to continue Landsat’s legacy of providing high quality thermal imagery.

ACKNOWLEDGMENT

The authors would like to thank the TIRS-2 Goddard Space Flight Center team for testing support, especially Jeffery Love for software support. We would also like to thank Brian Markham and Raviv Levy of the Landsat Cal/Val team for helpful discussions and the Space Dynamics Laboratory including Alan Thurgood for blackbody characterization.

REFERENCES

[1] Steve Foga, Pat L. Scaramuzza, Song Guo, Zhe Zhu, Ronald D. Dilley, Tim Beckmann, Gail L. Schmidt, John L. Dwyer, M. Joseph Hughes, and Brady Laue, “Cloud detection algorithm comparison and validation for operational Landsat data products,” Remote Sensing of Environment, vol. 194, pp. 379–390, 2017.
[2] Michael J. Wilson and Lazaros Oreopoulos, “Enhancing a simple MODIS cloud mask algorithm for the Landsat data continuity mission,” IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 2, pp. 723–731, 2013.
[3] Martha C. Anderson, Richard G. Allen, Anthony Morse, and William P. Kustas, “Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources,” Remote Sensing of Environment, vol. 194, pp. 379–390, 2017.
[4] Richard G. Allen, Boyd Burnett, William Kramber, Justin Huntington, Jeppe Kjærsgaard, Ayse Kilic, Carlos Kelly, and Ricardo Trezza, “Automated calibration of the METRIC-Landsat evapotranspiration process,” JAWRA Journal of the American Water Resources Association, vol. 49, no. 3, pp. 563–576, May 2013.
[5] C. Santos, I. J. Lorite, R. G. Allen, and M. Tasumi, “Aerodynamic parameterization of the satellite-based energy balance (METRIC) model for ET estimation in rainfed olive orchards of Andalusia, Spain,” Water Resources Management, vol. 26, no. 11, pp. 3267–3283, June 2012.
[6] Xiao-Ling Chen, Hong-Mei Zhao, Ping-Xiang Li, and Zhi-Yong Yin, “Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes,” Remote Sensing of Environment, vol. 190, pp. 133–146, Sept. 2006.
[7] M. El-Hattab, Amany S.M., and Lamia G.E., “Monitoring and assessment of urban heat islands over the southern region of Cairo Governorate, Egypt,” The Egyptian Journal of Remote Sensing and Space Science, vol. 21, no. 3, pp. 311–323, Dec. 2018.
[8] Carmen Quintano, Alfonso Fernandez-Manso, and Dar A. Roberts, “Burn severity mapping from Landsat MESMA fraction images and land surface temperature,” Remote Sensing of Environment, vol. 190, pp. 83–95, Mar. 2017.
[9] I. Ogashawara, L. Li, and M. J. Moreno-Madrid, “Spatial-temporal assessment of environmental factors related to Dengue outbreaks in São Paulo, Brazil,” GeoHealth, vol. 3, no. 6, pp. 202–217, Aug 2019.
[10] D.P. Roy, M.A. Waldner, T.R. Loveland, Woodcock C.E., R.G. Allen, M.C. Anderson, D. Helder, J.R. Irons, D.M. Johnson, R. Kennedy, T.A. Scambos, C.B. Schaaf, J.R. Schott, Y. Sheng, E.F. Vermote, A.S. Belward, R. Bindschadler, W.B. Cohen, F. Gao, J.D. Hipple, P. Hostert,
Aaron Pearlman received the B.S. in electrical engineering in 2001 from Tufts University and the M.S. and Ph.D. in electrical engineering from the University of Rochester, in 2003 and 2006, respectively, and became an IC postdoctoral research fellow at the National Institute of Standards and Technology to conduct research on single photon generating and detecting technologies for quantum information applications.

Dr. Pearlman has since provided pre-launch testing support and technical oversight and coordination for a validation field campaign for National Oceanic and Atmospheric Administration’s (NOAA’s) new generation of satellite sensors, and in 2017, joined the NASA Goddard Space Flight Center’s Calibration Team to support pre-launch testing for the Landsat Program’s Thermal Infrared Sensor-2. He is the Chief Scientist of GeoThinkTank LLC. His current work focuses on characterizing remote sensing systems before launch and developing novel methods for validating their post-launch performance.

Matthew Montanaro holds degrees in Physics (BS, 2005) and Imaging Science (PhD, 2009) from the Rochester Institute of Technology (RIT). He is currently a Senior Research Scientist involved in the calibration of thermal infrared imaging instruments for NASA Goddard Space Flight Center and the US Geological Survey. He specialized in the calibration of the Thermal Infrared Sensor (TIRS) onboard Landsat 8, both pre-flight and on-orbit. He now serves as the Deputy Calibration Lead for the upcoming Landsat 9/TIRS-2 instrument and serves on the Landsat Calibration & Validation team. Additionally, he has supported and advised a number of Imaging Science graduate and undergraduate students through RIT.

Boryana Efremova has been working in the field of remote sensing instrument calibration since 2011, supporting the prelaunch and/or on-orbit calibration, instrument monitoring and validation of sensors such as Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi NPP spacecraft, the Advanced Baseline Imager (ABI) onboard the geostationary GOES-16 and GOES-17 satellites, and the prelaunch calibration and testing of the Landsat 9 Thermal Infrared Sensor 2 (TIRS-2). She received her PhD in Astronomy and Astrophysics from Sofia University in Bulgaria in 2009, and holds a M.S. in Physics from the same university.

Joel McCorkel holds degrees in Optical Engineering (BS, 2005) and Optical Sciences (PhD, 2009) from the University of Arizona. He currently is a physical research scientist with the Biospheric Sciences Laboratory at NASA’s Goddard Space Flight Center where his current work involves development and characterization of the next-generation environmental monitoring and Earth observing sensors. He serves as the GOES-R Flight Project Scientist, the Landsat 9 TIRS-2 Deputy Instrument Scientist leading the prelaunch instrument characterization, and the principle investigator of the Goddard Laser for Absolute Measurement of Radiance (GLAMR) calibration facility.

Brian N. Wenny received a physics B.S. degree from Ursinus College in Collegeville, PA in 1991, and M.S. and Ph.D. degrees in atmospheric science from North Carolina State University, Raleigh, in 1996 and 2000, respectively. He is currently a Chief Research Scientist with Science Systems and Applications, Inc. (SSAI) in Lanham, MD working primarily on the RadCalNet project. Previously he worked with the Moderate-Resolution Imaging Spectroradiometer Characterization Support Team (MCST), pre-launch calibration for the Landsat 8 Thermal Infrared Sensor (TIRS) and Landsat 9 TIRS-2 instruments and the Stratospheric Aerosol and Gas Experiment III (SAGE III).