In-Situ Transfer Standard and Coincident-View Intercomparisons for Sensor Cross-Calibration

Kurt Thome, Joel McCorkel, and Jeff Czapla-Myers

***Abstract*—There exist numerous methods for accomplishing on-orbit calibration. Methods include the reflectance-based approach relying on measurements of surface and atmospheric properties at the time of a sensor overpass as well as invariant scene approaches relying on knowledge of the temporal characteristics of the site. The current work examines typical cross-calibration methods and discusses the expected uncertainties of the methods. Data from the Advanced Land Imager (ALI), Advanced Spaceborne Thermal Emission and Reflection and Radiometer (ASTER), Enhanced Thematic Mapper Plus (ETM+), Moderate Resolution Imaging Spectroradiometer (MODIS), and Thematic Mapper (TM) are used to demonstrate the limits of relative sensor-to-sensor calibration as applied to current sensors while Landsat-5 TM and Landsat-7 ETM+ are used to evaluate the limits of *in situ* site characterizations for SI-traceable cross calibration. The current work examines the difficulties in trending of results from cross-calibration approaches taking into account sampling issues, site-to-site variability, and accuracy of the method. Special attention is given to the differences caused in the cross-comparison of sensors in radiance space as opposed to reflectance space. The results show that cross calibrations with absolute uncertainties $< 1.5\%$ (1σ) are currently achievable even for sensors without coincident views.**

***Index Terms*— Advanced Land Imager (ALI), Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), cross-calibration, Enhanced Thematic Mapper Plus (ETM+), Moderate Resolution Imaging Spectroradiometer (MODIS), SI-traceability, Thematic Mapper (TM), vicarious calibration.**

I. Introduction

THE GOAL of cross-calibration is to allow accurate intercomparison of sensor data. Note that there is a subtle difference between intercomparison and cross-calibration. A radiance intercomparison implies that spectral radiance derived from two or more sensors can be compared to determine their level of agreement. Ideally, the derived results, when viewing the same source at the same time, will agree within the stated uncertainties for the sensors. Such a comparison is effectively a validation of each sensor's calibration. Cross-calibration uses intercomparison data sets, but the calibration of one sensor is adjusted so that the spectral radiance from both sensors match taking into account view, spectral, and temporal differences. Technically, the term "cross-calibration" is strictly no different than any other calibration, but the descriptive nature of the term is still of use in this context and while strictly not correct, cross calibration and intercomparison are used interchangeably in this paper for simplicity.

Cross-calibration methods are used extensively for both on-orbit and prelaunch characterization and the typical method relies either on knowledge of a source that is common to both sensors or a reference detector that can place multiple sources on the same scale. As described below, one approach to cross-calibration (both on orbit and in the laboratory) uses near-coincident views of the common source. More recent work has emphasized methods that do not require simultaneous data collections but evaluate the temporal nature of the source through laboratory transfer radiometers, *in situ* measurements of ground scenes, or assume the sources to be invariant or characterizable over small time periods.

Accurate radiometric calibration based on well-defined and reproducible protocols allows comparison of data from two sensors calibrated in different facilities with agreement within the stated uncertainties when the sensors view a common source. Such a concept had been demonstrated in the laboratory through traveling transfer radiometers participating in round-robin exercises [1]–[3]. Approaches have been developed for similar comparisons between two satellite-based imagers while on orbit.

One of the first applications of cross-calibration techniques dates to Hovis *et al.* who measured the radiance above a ground target from a high-altitude aircraft to verify the degradation of the response of the Coastal Zone Color Scanner's shorter wavelength bands [4]. A similar approach was used for the calibration of the Advanced Very High Resolution Radiometer (AVHRR) via sensors onboard the ER-2 aircraft [5]. The basic concept is straightforward. The two sensors view the same test site at the same time from the same viewing geometry with identical spectral bands.

The approach has been applied to cases where both sensors are on orbit with data sets that have nearly coincident views of the same test site. The Simultaneous Nadir Overpass (SNO) method is one such approach and it obtains a large number of near-coincident views near the poles for typical sun-synchronous, near-polar orbits [6]–[8]. Such overlapping data sets limit the approach to spectral regions for which the radiance from snow and ice can be well predicted. Uncertainties in the approach tend to be dominated by spectral differences between sensors and bi-directional reflectance effects.

As mentioned, the ideal case is one for which the data from both sensors are coincident in time with identical view and solar geometries. Teillet *et al.* developed a technique to account for small changes in view and solar geometry by using an

Manuscript received February 29, 2012; revised September 7, 2012; accepted October 23, 2012. Date of current version February 21, 2013.

K. Thome and J. McCorkel are with Space and Earth Sciences Directorate, NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA (e-mail: kurtis.thome@nasa.gov; joel.mccorkel@nasa.gov).

J. Czapla-Myers is with College of Optical Sciences, University of Arizona, Tucson, AZ 85721 USA (e-mail: j.czapla-myers@optics.arizona.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2243841

airborne sensor to derive the surface reflectance of a test site both spatially and spectrally and predicts at-sensor radiance relying on coincident atmospheric data [9]. Such an approach allowed the intercomparison of a wide array of sensors viewing the Railroad Valley test site on a single day but at varying times and with varying view angles [10]. The advantage of the method is that the hyperspectral basis of the surface reflectance characterization limits errors due to spectral mismatch between sensors being compared [11].

The characterization can also be based on a model-centric approach in which one sensor is used to understand the relative changes in surface reflectance from solar illumination and view angle effects. The method has been applied quite successfully for the comparison of AVHRR sensors over time based on desert scene data [12]. Alternatively, the characterization can rely completely on *in situ* measurements at the time of both sensors, in which case the *in situ* measurements themselves act as the transfer standard [13].

The current work examines cross-calibration results from a coincident-view pairing as well as from those relying on *in situ* measurements as a transfer standard. Data from the Advanced Spaceborne Thermal Emission and Reflection and Radiometer (ASTER) and Moderate Resolution Imaging Spectroradiometer (MODIS) are used in the coincident view case to demonstrate the uncertainties encountered in the most straightforward situation. The simultaneity of the data should improve the comparisons between the two sensors but differences are still caused by sampling issues, site variability, and accuracy of the method. Special attention is given to the differences caused in the cross-comparison of sensors in radiance space as opposed to reflectance space.

The *in situ* transfer standard uses four sensors, the Advanced Land Imager (ALI), ASTER, Landsat-7 Enhanced Thematic Mapper Plus (ETM+), and Landsat-5 Thematic Mapper (TM). All four have similar spatial resolution with similar spectral bands in the reflective portion of the spectrum. The current work demonstrates that the precision of cross comparisons using the reflectance-based method as the reference approaches 1.4% (1σ) (note that all uncertainties given here are $1-\sigma$ values). The sensors used here are limited to those with moderate resolutions from 15 to 30 m resolution but do not have coincidence in time. Analysis of the results indicates that this uncertainty can be improved through additional field instrument characterization, and higher frequency of observations.

The paper begins with a brief description of the reflectance-based approach as it applies to this work and the sensors considered. The intercomparison between sensors with coincident-date overpasses are presented followed by an analysis approach between the Landsat-5 and Landsat-7 sensors. The results demonstrate that appropriate handling of *in situ* data leads to results with uncertainty on the order of the empirically-based cross-calibration methods using pseudo-invariant sites.

II. Reflectance-Based Approach

The reflectance-based method uses ground-based measurements to characterize the surface of a test site and the atmosphere over that test site. The results of these characterizations are inputs to a radiative transfer code to predict at-sensor radiance. The approach has been used for a wide range of spatial resolutions by characterizing areas ranging from 104 to 106 m at test sites ranging in sizes from 100-m parking lots to 30 km dry lakes [14], [15]. The work shown here relies on data collected at the Railroad Valley Playa test site in Nevada and Ivanpah Playa in California. Details of the reflectance-based approach and both sites can be found in other sources, so only a brief overview is given here [14].

Test Sites: The Railroad Valley test site is in central Nevada and has an overall size approximately 15 km $\times$ 15 km. The playa's 1.5-km elevation, location in a region with typically clear weather, low aerosol loading, and high surface reflectance makes it a good site for the reflectance-based approach. Typical atmospheric conditions at the site include an average aerosol optical depth at 550 nm of 0.060 [16]. The reflectance of the playa is generally greater than 0.3 and relatively flat spectrally except for the blue portion of the spectrum and an absorption feature in the shortwave infrared. Ground-based measurements of the directional reflectance characteristics of the playa show it to be nearly lambertian out to view angles of 30 $^\circ$ for incident solar zenith angles seen for overpasses of Terra and Landsat [17].

The Ivanpah Playa test site has similar reflectance characteristics but is in general brighter than Railroad Valley Playa. Ivanapah Playa is significantly smaller than Railroad Valley Playa with a width of approximately 3 km and length of 5 km. The surface is much harder and is susceptible to standing water in the winter and after heavy summer rains. The elevation of 0.8 km makes atmospheric effects more important and the site has an average aerosol optical depth at 550 nm of 0.084 [15].

Sensor Overview: Five sensors are used in this work, ALI, ASTER, ETM+, MODIS, and TM. ASTER, which is on the Terra platform, has a 60-km swath width with 14 total bands in the visible and near infrared (VNIR), shortwave infrared (SWIR), and thermal infrared (TIR) [18]. The spatial resolution of the three VNIR bands is 15 m and that of the six SWIR bands is 30 m. The VNIR and SWIR sensors are pushbroom systems. MODIS is also on the Terra platform with a 2330-km swath width and spatial resolutions of 250, 500, and 1000 m [19]. Only those bands with near matches to ASTER spectral bands are considered here and these are restricted to the 250 and 500 m bands.

ETM+ and TM are nearly identical copies of each other. The sensors rely on a whiskbroom scanning approach to allow for the relatively large 185-km swath width. A warm focal plane is used for the four 30-m VNIR bands, and, in the case of ETM+, the 15-m panchromatic band. A cold focal plane is used for the two SWIR bands and also for the single TIR band. The TIR band has 120-m spatial resolution for TM and 60-m resolution for ETM+. ALI was designed to provide imagery with the same aspects of TM and ETM+ such as spatial resolution, swath width, spectral bands, orbit, and overpass time but with an additional three bands, panchromatic resolution improvement, and a higher 12-bit quantization [20]. One fundamental difference between ALI and previous Landsat instruments is that it is a pushbroom system rather than a whiskbroom. Another distinct improvement of ALI is that the more technologically

TABLE I
BANDS AND CENTRAL WAVELENGTHS FOR ALI, ASTER, ETM+, AND TM

	Band Centers (μm)				
Band	ALI	ASTER	ETM+	MODIS	TM
1p	0.443	-	-	-	-
1	0.483	0.554	0.482	0.645	0.485
2	0.565	0.661	0.565	0.858	0.560
3	0.660	0.807	0.660	0.469	0.660
4	0.790	1.652	0.825	0.555	0.830
4p	0.868	-	-	-	-
5p	1.250	-	-	-	-
5	1.650	2.164	1.650	1.240	1.650
6	-	2.204	-	1.640	-
7	2.215	2.259	2.220	2.130	2.215
8	-	2.329	-	0.412	-
9	-	2.394	-	0.443	-

advanced design has about one-fourth the mass, one-fifth the power consumption, and about one-third the volume of Landsat 7.

The spectral bands for each sensor are listed in Table I. Reference is made throughout the paper to the band numbers for each of the sensors. The MODIS bands may appear to be oddly numbered, but it is the result of numbering bands according to spatial resolution and applications rather than center wavelength. The first two bands are 250-m NDVI bands. The next five bands are 500-m bands primarily for land applications. Bands 8 and 9, which are included only to complete the table, are 1-km ocean color bands and are not considered here. The other 27 spectral bands for MODIS are omitted for simplicity. The table shows that many of the bands are similar but with distinct differences. The differences between bands also extend to the bandwidths and these play a role when atmospheric absorption is considered. One critical point to consider related to this work is that cross-calibration of sensors must consider the band differences to provide accurate results [11]. The method described in this work considers these band effects.

There are several key platform parameters that are important for this work. Landsat-7 and Terra are separated by only 30 minutes in their orbits. ALI on the Earth Observing One platform was originally in an orbit within minutes of Landsat-7 but the platform has been allowed to drift since 2005 and no longer coincides with Landsat-7 in either day or time. Landsat-5 is in an orbit that is eight days out of phase from Landsat-7 and Terra. All of the data used in this work has a view angle for the sensors that is $< 30^\circ$ from nadir.

III. COINCIDENT-VIEW RESULTS

As described, the most straightforward cross-calibration approach is one in which the sensors view the same area with the same view angle at the same time. Ideally, the spectral bands would also be identical. The ASTER/MODIS intercomparison case satisfies all but the last, in that while several bands are similar between the sensors, they are not identical. The results shown here have been corrected for spectral differences by using ground-based spectral measurements over a representative region to modify the MODIS-based radiances at the sensor to predict those for the ASTER spectral bands. Only the Railroad Valley test site has surface reflectance measurements on a scale suitable for MODIS and these have been made across an area roughly 1 km $\times$ 1 km. The spectral correction derived for each of the three band pairs for the ASTER VNIR vary from 1–5% depending on date and spectral band related to cross-calibration approaches.

The other advantage to using the Railroad Valley test site is it permits a relatively large number of ASTER data sets to be used. A drawback of the ASTER sensor is that its relatively narrow swath width of 60 km coupled with data rate limitations of the Terra platform requires that the sensor be tasked to collect data over a given area. That is, while MODIS data are continuously collected, ASTER data are not. The Railroad Valley site is one of the highest frequency collections by ASTER due to the site's use as a reflectance-based calibration site. Even so, less than 40 ASTER observations of the Railroad Valley region exist over a five-year period from launch to 2004. Such a small number of observations limit the approaches ability to determine a statistically significant trend in the radiometric calibration. However, it still permits an evaluation of the absolute calibration accuracy of ASTER as well as the precision of cross-calibration.

The MODIS and ASTER geolocation information are used to register the two data sets. The full playa is used to determine the cross-calibration relying on areas greater than 10 km^2 for all data shown that corresponds to more than 40 MODIS pixels and 4000 ASTER pixels. The varying number of pixels used from case to case is a result of fewer pixels for dates where the view angle is off nadir and those dates when pointing of the sensor was such that the playa was not imaged in the center of the ASTER scene.

Fig. 1 shows the calibration coefficient derived for the VNIR bands of ASTER based on radiances derived from MODIS relying on its onboard calibration method. Calibration coefficient is defined here in terms of counts per unit spectral radiance where counts are the digital output from the sensor for the pixels of interest. The uncertainty, as discussed below, in the derived coefficients is slightly larger than the vertical size of the symbols in the graph. Similar results are obtained for the SWIR bands but are somewhat complicated by the presence of an optical crosstalk in the ASTER channels which has not been corrected in the data used here. The corresponding band pairs are ASTER band 1 to MODIS band 4, ASTER band 2 to MODIS band 1, and ASTER band 3 to MODIS band 2. The requisite figures showing the relative spectral responses of each band are shown in Fig. 2 and show that the spectral bands are very similar in location but the band widths and shapes are not identical.

The approach begins with a spectral reflectance measured on the ground for a large area of Railroad Valley that was collected on the nearest date matching the cross-calibration date. The surface is assumed to be lambertian. At-sensor reflectance is computed based on average atmospheric conditions derived from on-site, ground-based solar radiometer measurements and the sun-sensor geometry for the Terra platform. The at-sensor, predicted hyperspectral reflectance is band averaged using the best estimate for the ASTER and MODIS relative spectral responses providing a correction factor specific to the individual date. The reported MODIS at-sensor reflectance is multiplied by the correction factor to give an equivalent ASTER at-sensor

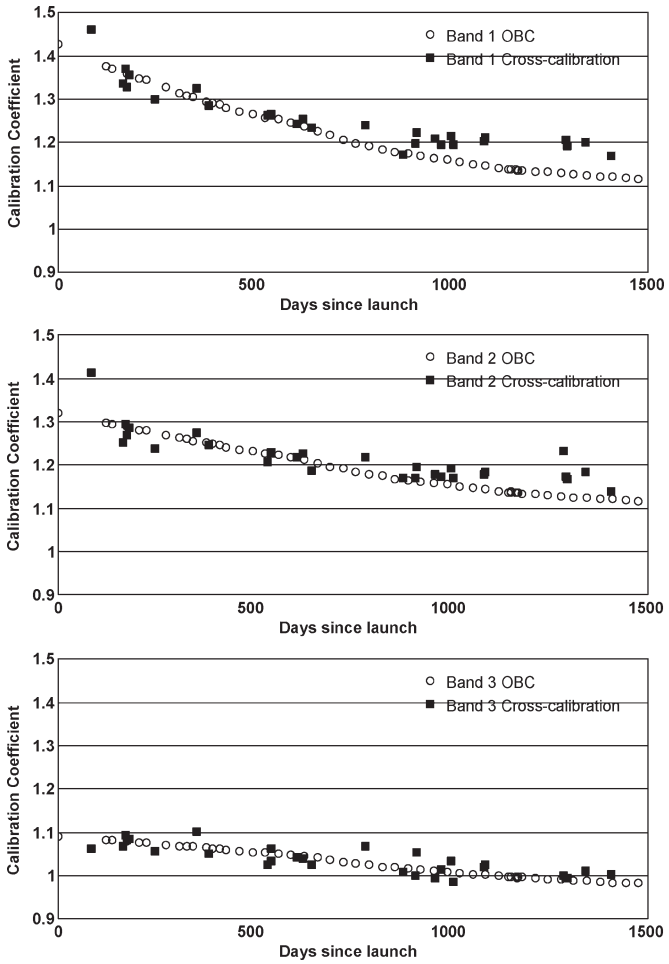

Fig. 1. ASTER Level 1A calibration coefficients derived from similar MODIS spectral band data over Railroad Valley Playa. Also shown are data from lamp-based onboard calibrator data.

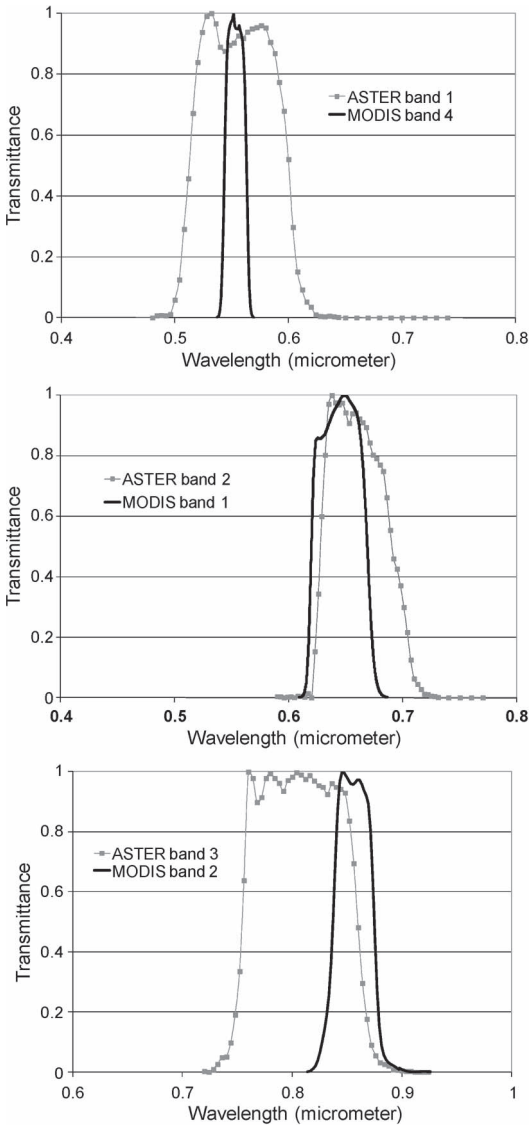

Fig. 2. Relative spectral response data for MODIS/ASTER band pairs used in the current work.

reflectance which is converted to an at-sensor radiance using the sun angle for the data set and an appropriate solar irradiance model.

The calibration coefficient (G) for ASTER is computed using the offset (DN_{offset}) for the given band, reported digital numbers (DN_{image}), and the at-sensor radiance $(L_{\text{at-sensor}})$ according to

$$G = \frac{(DN_{\text{image}} - DN_{\text{offset}})}{L_{\text{at-sensor}}}.$$

The first notable feature in Fig. 1 is that the onboard calibrator data show degradation with time for all three bands with Band 1 showing the largest degradation of 28% over the data shown and 18% and 11% degradation for the other two bands. The cross-calibration data show statistically identical degradation for all three bands over ASTER's first 600 days on orbit (10%, 6%, 3% for the onboard data) and for band 3 there is no statistical difference between the onboard calibrator and the cross-calibration results for the entire length of time shown. The onboard calibrator for bands 1 and 2 shows statistically larger degradation than that obtained from the cross-calibration. Similar results have been obtained from direct application of the reflectance-based data sets [21] giving confidence that the onboard calibrator data are suffering degradation. One conjecture is that the onboard calibrator's relay optics are degrading over time.

The other notable feature in the data is the variability in the derived calibration coefficient for ASTER. Examination of the data after day 800 for which the calibration coefficient of ASTER appears reasonably stable with time shows variability of 1.2%, 1.6%, and 1.7% for Bands 1, 2, and 3, respectively. These values are used as surrogates for the relative uncertainty of the method and on the graph correspond to being slightly larger than the size of the square symbols. Absolute uncertainty is dominated by the absolute uncertainty of the MODIS sensor which is constant in time. The results shown here are on par with those obtained over desert sites but those cases were not coincident view situations [12]. The causes of the scatter cannot be due to surface directional reflectance nor atmospheric effects because of the shared platform. The only sources of error must be 1) the spectral correction; 2) registration of the common areas used for the cross-calibration; and 3) temporal variability in the sensors relative to one another.

The data shown in Fig. 1 have been corrected for the spectral reflectance of the playa based on surface reflectance data collected at the site close in time to the cross calibration. Two effects may still be present. The first is that the area measured on the ground does not correspond exactly to the region used for the cross-calibration. The effect is that spatial non-uniformity of the spectral reflectance leads to uncertainties in the spectral correction. The second effect is that the spectral reflectance of the playa is not invariant in time and thus small changes in surface moisture or other temporal changes in the surface can lead to an incorrect spectral correction between the sensors.

Registration errors between the two sensors were examined by shifting the areas for each sensor used for the cross-calibration relative to one another. The relative large number of MODIS pixels used and ensuring that the region of the playa does not include those pixels near the edge of the playa minimizes the registration effects to be nearly negligible. Note that atmospheric adjacency effects should not play a role, except for the small differences that might be caused by the different spectral responses of the sensor. Spatial response of the MODIS sensor coupled with the spatial inhomogeneity of the playa could play a role and will be investigated in future work, but again it is expected that this effect is small because more than 5 MODIS pixels in the cross-track direction are included in the averaging.

The last factor, temporal variability of the sensors, is the ultimate goal of cross-calibration work. The results shown here show the difficulty in determining temporal variability of the sensor response when the calibration approach itself has significant variability. One approach to solving this problem is to increase the number of data points available for analysis which has the effect of making the determination of outliers more straightforward, improves overall understanding of the cause of scatter in the data sets, and allows for improved analysis of means and trends.

Such an increase in the number of data sets is straightforward for wide-swath, 100% duty cycle sensors, but is not quite as simple for on-demand sensors such as ASTER. One must keep in mind that the ultimate use of the sensors is for science applications. Thus, tasking the sensor for calibration purposes leads to the loss of science data. Fig. 3 demonstrates this to some extent. The data shown in Fig. 3 are a repeat of the Band 1 data from Fig. 1 but also includes cross-calibration data points from White Sands and from multiple African desert sites all of which have seen use in past vicarious calibration efforts [12]. The specifics of these sites are not critical for this work except that White Sands cannot be imaged by ASTER without the loss of data from a long-term ecological monitoring site and the African test sites do not have ground-based data to supply band corrections for cross-calibration approaches. The competition between scheduling science collects and cross-calibration collections, coupled with the fact that White Sands saturates ASTER band 1 in late spring to early fall means that only a relatively small number of data points are available. The clustering of data sets early in the sensor lifetime demonstrates the shift in priorities from sensor understanding early in the mission to operational science collections later.

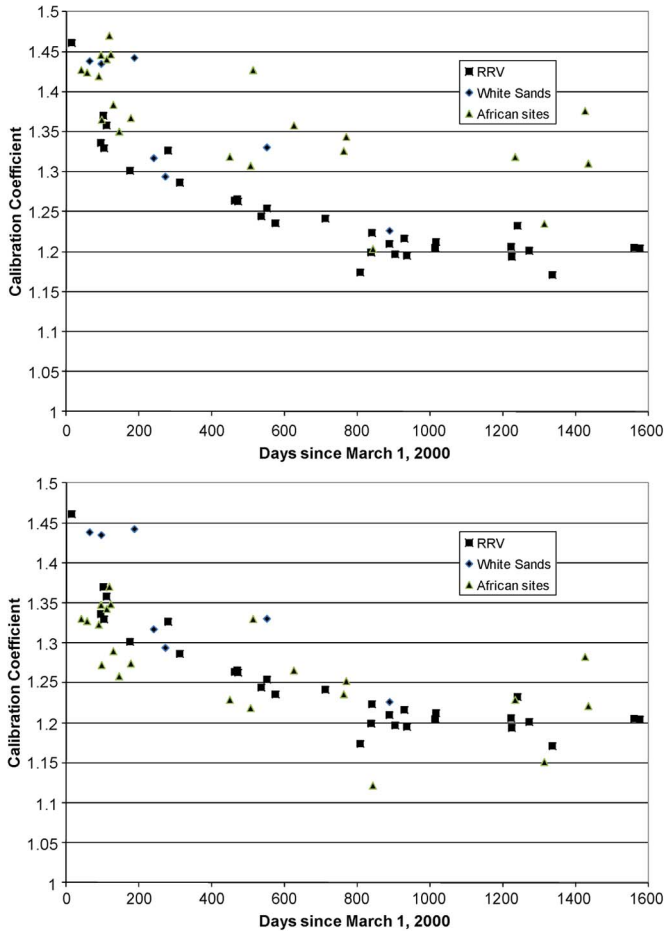

Fig. 3. ASTER Band 1 cross-calibration results for multiple test sites as compared to reflectance-based results.

The 1σ relative uncertainty for a single data point caused by registration is slightly larger than the size of the data points. The uncertainty due to spectral differences is minimal for the White Sands and Railroad Valley data, but is significant for the African data. The upper graph of Fig. 3 shows the results obtained by using the best-estimated spectral reflectance of the surface. In the case of the multiple African sites this relies on laboratory spectra with a secondary correction forcing agreement around day 100 for which data from the multiple sites were available near in time to each other. It is clear from the upper graph that either there are significant scene-based effects between the test sites as evidenced by the lack of agreement later in the mission lifetime or an additional bias is present. The most likely bias is caused by the spectral differences between ASTER and MODIS not being properly corrected. That is, the knowledge of the spectral reflectance is not sufficient in this case.

The stability of the ASTER response after day 1200 allows for a simple correction of the African data to match the results from Railroad Valley Playa. Such an empirical approach is commonly used for cross-calibration of narrow-swath sensors based on the assumption that uncertainty in the knowledge of the spectral reflectance dominates the error budget. Such an approach does not lead to traceable calibration and one of the purposes of the current work is to discourage such empirical-only methods. The resulting correction is still of interest to

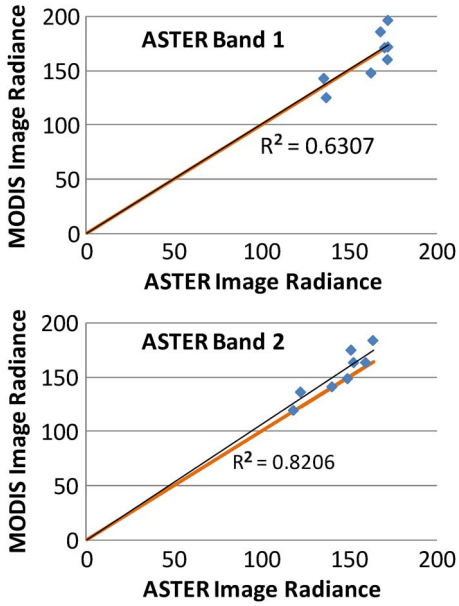

Fig. 4. Radiance to radiance comparison between ASTER and MODIS at Railroad Valley Playa using subset of data shown in Fig. 1

highlight day-to-day variability and is shown in the lower graph and demonstrates the improved agreement between the African and Railroad Valley results. No correction has been made to the White Sands data due to the lack of data after day 1200. An improved correction could be achieved through the recent availability of on-orbit hyperspectral data sets, but the scatter in the Railroad Valley data still demonstrates the difficulty of applying the spectral correction when working with ground sites.

Note that large size of the African test sites allows the entire 60-km ASTER scene to be averaged for the cross calibration. Reflectance-based results from ASTER [21] do not show a dependency of the ASTER calibration across the swath, but it is not possible to rule out completely that the large extent of the high-reflectance desert sites could be responsible for some of the bias seen in Fig. 3 between the smaller test sites and the larger test sites. The test sites are also quite uniform. The cause of the scatter seen in Fig. 3 is a result that each individual African site has its own unique spectral reflectance creating slightly different spectral band difference effects at each site. The scatter seen in Figs. 1 and 3 demonstrates the difficulties in performing cross-calibration even under the ideal conditions of identical view and simultaneous collections.

An alternate method for comparing ASTER and MODIS is through radiance intercomparisons. The results shown here purposely do not include a spectral correction to emphasize the philosophical issues with making such a correction on at-sensor radiance. Fortunately, the spectral differences do not affect the specific points that are to be made for the cases shown here. Fig. 4 shows the radiance intercomparisons using a subset of the dates from Fig. 1 at Railroad Valley. Bands 1 and 2 of ASTER compared to Bands 4 and 1 of MODIS are shown. The radiance values used for both sensors are based on each sensor's Level 1B data sets. The narrow line is the linear, least-squares fit and the thicker line is the one-to-one line. A deviation from the one-to-one line indicates a bias in the intercomparison, which in

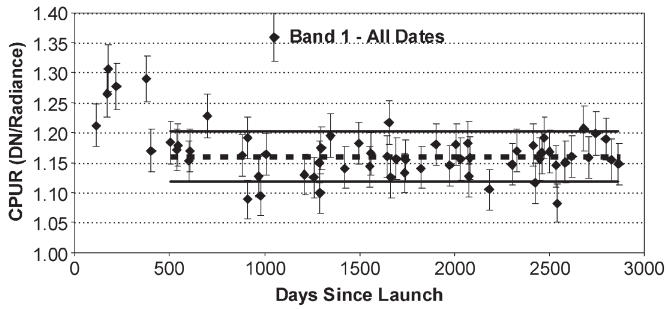

Fig. 5. Calibration coefficients for ASTER Band 1 based on reflectance-based results at multiple test sites.

this case could be a spectral-difference effect or a difference in calibration.

Of interest in Fig. 4 is that ASTER band 1 appears to agree well with MODIS band 4 but with a relatively large amount of scatter. In reality, there is a significant bias seen in the ASTER band 1 results relative to MODIS, but this is masked by the fact that saturated ASTER data still provide a time varying radiance, and also reflectance, in the Level 1B data products that appears to be valid. Such results are straightforward to recognize, but as larger data sets at multiple sites with multiple sensors are used for cross-calibration, and more automated methods are developed, such nuances can be easily overlooked. The Band 2 results for ASTER show a good correlation, indicating the quality of the intercomparison, but the deviation from the one-to-one line is statistically significant and larger than the absolute uncertainties for both sensors.

IV. *In-Situ* Transfer Standards

An alternate approach to cross-calibration is to use the reflectance-based method as a transfer standard. Fig. 5 shows the results of the reflectance-based method for Band 1 from ASTER. The dashed line in the figure represents the average of the calibration derived after day 500 and the solid lines are $\pm$ one standard deviation from the average. Day 500 was chosen as the point for which the degradation in seen in the data appeared to stabilize. The data shown in Fig. 5 and in Fig. 1 have similar degradation giving confidence in both sets of results. The temporal frequency of the data and the precision of the reflectance-based approach prevent trending of the data with any statistical confidence, but the change in coefficients is readily apparent. The results later in the lifetime of ASTER show the scatter that is one of the main issues with the reflectance-based approach. No single cause has been found to date to explain the scatter in the data. The notable outlier near day 1000, for instance, was from a cloud-free day with good surface conditions. Examination of ancillary information from that date indicates the reflectance data were collected with a borrowed spectrometer operated by an inexperienced user. Such outliers are removed from calculations of calibration coefficients, but only when a just cause is found beyond the fact that the data does not match the average.

Similar graphs can be produced for other bands of ASTER and similarly for ALI, ETM+, and TM. The results of the work for all four sensors show that the reflectance-based method is

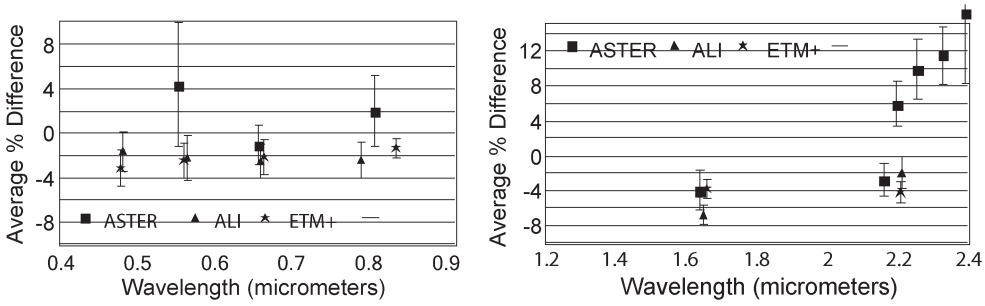

Fig. 6. Cross-comparison results between ASTER and ETM+ allowing the determination of a cross-calibration between the two sensors. Data are based on Level 1B data from ASTER processed with post-2006 calibration.

useful in detecting sensor degradation and anomalies. This is evident in the results for ASTER which show degradation in the VNIR bands, an optical crosstalk effect in the shortwave infrared, and the change in offset due to a cooler problem for the SWIR bands. Trending of degradation is difficult due to the variable nature of the results but use of the reflectance-based results as a transfer standard is feasible.

Using the reflectance-based method as a transfer standard relies on comparing the sensor calibrations relative to the *in situ* data. Likewise, one could compute the calibration coefficients for each individual sensor using the reflectance-based approach and the consistency and traceability of the method would inherently provide consistency between the sensors of interest. That is, using a calibration coefficient determined from sensor output and vicarious prediction provides a set of coefficients that cross-calibrates the sensors using the vicarious results as a transfer standard.

Fig. 6 follows the idea of developing a set of calibration coefficients for a single sensor that allows it to compare well with a different sensor. Consider the case where a user already has a set of ETM+ data and they wish to include ALI and ASTER data in the study. Fig. 6 shows the average percent difference between the reported radiance from a given sensor using its best calibration coefficient and the radiance predicted by the reflectance-based approach. The averages represent at least five data points for each sensor covering similar periods of time for which the calibration of a given sensor is known to be stable. The error bars for each sensor relate to a single standard deviation of the average.

Then comparisons of the ALI and ASTER results against the reflectance-based method shows Band 1 of ASTER disagrees by 6.8% with ETM+ band 2 and a similar value for Band 2 of ALI. The user should then adjust all of their ASTER Band 1 data by 6.8% to gain agreement between the two sensors. Such an approach is usually needed when similar bands are being compared, but the approach becomes more difficult to apply when the desire is to use all of the ASTER bands in the study and the hope is to have a consistent ETM+/ALI/ASTER data set. In that case, the SWIR bands 7–9 could be corrected to agree with the corresponding ALI or ETM+ band, but this may not necessarily be the best approach. The recommended method in such a case would be to correct both data sets relative to the 0% line in both graphs.

The other issue with this approach is that it does not readily lend itself to cases where there are changes in the calibration of the sensor with time. The solution to these situations is to use whatever methods are needed to determine the temporal changes in the radiometric calibration and then use the cross-calibration approach to anchor that radiometric calibration curve. The TM sensor is a good example of a system that has suffered significant degradation with time. TM is also a good test bed for this approach since a cross-calibration with TM and ETM+ using typical methods is not trivial due to the eight-day difference in orbit. In addition, the stability of TM and ETM+ in recent years allows for an evaluation of the limits of the precision of this approach.

The first step in the Landsat-5 TM calibration determination is to find the best dates from the reflectance-based method at all sites. The period 2004 to 2005 was chosen as the cross-calibration period since ground data collected during that period showed good agreement between the *in situ* results and the preflight calibration of ETM+ [17], [18]. The two-year time frame also corresponds to a time period for which sensor degradation is negligible for both TM and ETM+ is minimal. The sensor stability allows averages to be determined for the full two years giving a sufficient number of dates to evaluate accuracy and precision. Later dates are not included due to the issues with the Landsat-5 platform beginning in December 2005 causing complications in scheduling field collections and a lack of ground data.

Table II shows eight Landsat-5 TM dates in 2004 and 2005 for which there are *in situ* results. An additional nine dates of collection were attempted during the period with poor weather or poor surface conditions/snow preventing data collections in January, February, March, and June 2004, and January, February, March, April, and November 2005.

The goal of this work is to provide the Landsat-5 TM calibration with the highest confidence, thus it was decided to use only the "best" days of reflectance-based results for the period. Many options for selecting the dates are available ranging from selecting those dates with the lowest atmospheric optical depths, the highest surface reflectance, a specific aerosol size distribution, a certain season, etc. The process applied here is to scale each of the calibration coefficients in a band relative to the average for that band. The scaled coefficients on a given date are then averaged across the bands and a standard deviation is computed. Dates with the smallest standard deviations are kept and these are shown in Table III.

The standard deviations of the averages for each of the eight dates ranged from 0.9 to 2.1%. An arbitrary cut off of 1% was

TABLE II
TM RESULTS FOR EIGHT DATES (WITH ASSOCIATED DAYS SINCE LAUNCH USED TO DETERMINE CROSS-CALIBRATION PARAMETERS FOR ETM+. STD DEV. IS THE STANDARD DEVIATION OF THE AVERAGE OF THE EIGHT DATES GIVEN BOTH AS PERCENT (%) AND IN TERMS OF CALIBRATION COEFFICIENT

Date	DSL	Site	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
13-May-04	7249	RRV	1.237	0.644	0.916	1.102	7.930	14.980
23-Jun-04	7290	Ivan.	1.224	0.651	0.908	1.095	7.922	14.863
16-Dec-04	7466	Ivan.	1.176	0.639	0.906	1.096	8.180	14.435
17-Jun-05	7649	RRV	1.185	0.642	0.915	1.113	8.073	15.185
12-Jul-05	7674	Ivan.	1.271	0.638	0.911	1.106	7.965	14.905
19-Jul-05	7681	RRV	1.171	0.634	0.902	1.094	7.906	14.876
13-Aug-05	7706	Ivan.	1.184	0.622	0.885	1.076	7.712	14.074
23-Oct-05	7777	RRV	1.213	0.655	0.927	1.104	8.081	14.718
Average	---	---	1.208	0.641	0.909	1.098	7.971	14.755
Std. Dev.	---	---	0.035	0.010	0.012	0.011	0.142	0.349
% Std. Dev.	---	---	2.9	1.6	1.4	1.0	1.8	2.4

TABLE III
SELECTED FOUR DATA SETS OF TM RESULTS SCALED RELATIVE TO BAND 3. STD DEV. IS THE STANDARD DEVIATION OF THE AVERAGE OF THE FOUR DATES GIVEN BOTH AS PERCENT (%) AND IN TERMS OF CALIBRATION COEFFICIENT

Date	DSL	Site	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
13-May-04	7249	RRV	1.228	0.639	0.909	1.094	7.869	14.866
23-Jun-04	7290	Ivan.	1.225	0.652	0.909	1.096	7.931	14.879
13-Aug-05	7706	Ivan.	1.216	0.639	0.909	1.105	7.921	14.456
23-Oct-05	7777	RRV	1.189	0.642	0.909	1.083	7.924	14.432
Average	---	---	1.215	0.643	0.909	1.094	7.911	14.658
Std. Dev.	---	---	0.017	0.006	---	0.009	0.028	0.248
% Std. Dev.	---	---	1.4	0.9	---	0.9	0.4	1.7

selected since this was both a natural break point in the data set as well as leaving four days of data to be used. This number of dates is important as statistical analysis of the expected accuracies shows that four data sets are sufficient to estimate the calibration coefficient to within 2.5% at a 95% confidence interval [18]. The dates omitted from further analysis and their standard deviations are 16 December 2004 (1.9%), 17 June 2005 (1.6%), 12 July 2005 (2.1%), and 19 July 2005 (1.3%).

It is of interest to understand the causes of the larger scatter on certain dates, though the cause is not critical for this analysis. Likely sources of band-to-band scatter on a given date are spectral-spatial variations in the surface reflectance, errors in knowledge of atmospheric conditions including the use of an incorrect aerosol model, processing errors, and larger than normal noise in the reflectance measurements. The average values for the calibration coefficients for the full data set and the four selected data points changed by only a small amount with the largest differences for bands 5 and 7 which both decreased by 0.7%. Small increases in the standard deviations are seen in all bands.

A further conditioning step is applied by recognizing that all bands have a similar bias from the average for a given date. Such a correlated effect implies that the reflectance-based results on a given date have a consistent bias. The most likely cause is a bias from the surface reflectance measurements since the spectral effect is small. The data conditioning is to scale each of the data points by an amount determined from the bias in a selected band from its average. The average of Band 3 remained the same between the eight- and four-date data sets and Band 3 was selected as the reference band. Ratios of the difference between the calibration coefficient for band 3 on a given date to the average are computed and used to scale all other coefficients. The scaled coefficients are shown in the Table III. It should be noted that these two scaling steps described above are actually quite similar in concept to other data conditioning methods used for intercomparison studies and lunar calibration [22].

The results shown in Table III for Band 3 are the same since it was the band scaled relative to its own average. The results for other bands show effectively no change in the average with dramatic improvements in the standard deviation. The results shown here are believed to be the best estimates for the absolute radiometric calibration coefficients for Landsat-5 TM. Note that the average calibration coefficients do not change significantly from the full data set to the re-normalized and scaled coefficients.

A similar process has been applied to ETM+ data. Key differences are that a total of 17 data sets were available for ETM+ for the 2004–2005 period. Using a 1% standard deviation screening for scatter left seven dates in the period giving the averages and standard deviations shown in Table IV, and scaling relative to band 3 gives the last two rows of the table. Additionally, results after scaling relative to Band 4 have been included for reference.

Users of ETM+ data will most likely use preflight calibration information for ETM+. It makes sense then, to normalize the TM calibration coefficients relative to the ETM+ preflight calibration. This is done by multiplying the TM coefficients by the ratio of the ETM+ preflight calibration to the scatter-screened, band-normalized ETM+ results. This gives the final, best estimate values for the TM calibration coefficients that will produce ETM+ equivalent radiances based on preflight calibration of ETM+. The values are shown in Table V.

TABLE IV
AVERAGE CALIBRATION COEFFICIENTS AND STANDARD DEVIATIONS FOR 17 DATES FOR THE SAME TIME PERIOD AS THE EIGHT TM DATES SHOWN IN TABLE III. AVERAGES ARE ALSO COMPUTED FOR SELECTED SEVEN ETM+ DATES AS WELL AS NORMALIZED RESULTS TO BANDS 3 AND 4

Date		Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Average	Full data 2004-2005 data set	1.191	1.174	1.567	1.529	7.451	21.243
%Std. Dev.		2.2	4.1	3.7	3.7	4.2	6.2
Average	Scatter-screened seven dates	1.185	1.167	1.554	1.509	7.388	21.187
%Std. Dev.		1.0	1.3	1.2	1.5	1.2	1.1
Average	Band 3 normalized	1.185	1.167	1.554	1.509	7.388	21.187
%Std. Dev.		0.6	0.7	---	0.8	0.8	0.9
Average	Band 4 Normalized	1.186	1.167	1.554	1.509	7.389	21.188
%Std. Dev.		1.0	1.4	0.7	---	0.5	0.6

TABLE V
TM CALIBRATION COEFFICIENTS BASED ON CROSS-CALIBRATION TO ETM+ USING BAND 3 NORMALIZED REFLECTANCE-BASED RESULTS FROM TABLES III AND IV

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
L5 TM	1.250	0.650	0.884	1.095	8.127	15.048

V. CONCLUSION

The results above demonstrate that even the "simplest" cross-calibration problem can suffer from difficult to overcome uncertainties. Likewise, the "precision-poor" reflectance-based data can be conditioned much like desert data are culled for solar-view geometry to improve the precision of the method to that approaching the cross-calibration data sets. It is the breadth of the vicarious data available and the combination of these approaches along with the prelaunch characterization and calibration and the onboard calibration information that are critical in developing the most accurate and precise approaches to sensor radiometric calibration.

The most important aspect of combining calibration approaches is that it readily allows the inclusion of SI-traceable approaches and this will permit determining biases in a single given approach relative to the others. Doing this for multiple types of vicarious approaches makes it possible to compare results from different methods. For example, the coincident view results from ASTER showed a large level of scatter. The reflectance-based results showed similar levels of degradation as the MODIS cross-calibration but both methods disagree with the onboard calibrator data. This is critical to allowing the vicarious results to be used to evaluate the sensors at an unprecedented level.

The key conclusion from this work is that it is feasible to cross-calibrate sensors at a 1–2% level. The SI-traceable nature of the *in situ* transfer standard allows this level of cross-calibration even in situations when there are gaps in the data sets. Results are improved if the ground measurements are made in a consistent fashion but the ultimate goal should be to make measurements in a traceable fashion with high precision. The power of such a method should be clear when considering the lifetime records of Landsat and MODIS. The goal is to continue these data records through the Landsat Data Continuity Mission and subsequent MODIS-like sensors into the future. At the time of this writing, TM has already failed and it is possible that ETM+ may not be operational by the launch of LDCM. Likewise, both Aqua and Terra MODIS could fail prior to the launch of their follow-on missions. The *in situ* transfer standard allows for a consistent data record across any gaps in data collections permitting the future sensors to be put on the same radiometric scale as the current sensors.

Combining the *in situ* transfer standard with the desert site intercomparisons naturally leads to the attempt to place the intercomparison data sets on an absolute scale with SI traceability. Doing so would provide an approach with the gap-tolerance of the *in situ* method with the trending capabilities of the intercomparison data sets. The added benefit is that improvements to the on-orbit sensors would go hand in hand with improved understanding of surface reflectance characterization, atmospheric measurement, and radiative transfer code improvements.

ACKNOWLEDGMENT

The author is indebted to the many people assisting with the field data collections that are the basis of this work as well as the Bureau of Land Management offices in Needles, Calif. and Tonopah, Nevada for their assistance and permission in using the Ivanpah and Railroad Valley test sites.

REFERENCES

[1] B. C. Johnson, P. Y. Barnes, T. R. O'Brian, J. J. Butler, C. J. Bruegge, S. Biggar, P. R. Spyak, and M. M. Pavlov, "Initial results of the bidirectional reflectance characterization round-robin in support of EOS," *Metrologia*, vol. 35, no. 4, pp. 609–613, 1998.

[2] J. J. Butler, S. W. Brown, R. D. Saunders, B. C. Johnson, S. F. Biggar, E. F. Zalewski, B. L. Markham, P. N. Gracey, J. B. Young, and R. A. Barnes, "Radiometric measurement comparison on the integrating sphere source used to calibrate the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Landsat-7 Enhanced Thematic Mapper Plus," *J. Res. Nat. Inst. Std. Technol.*, vol. 108, no. 3, pp. 199–228, May 2003.

[3] J. J. Butler, B. C. Johnson, and R. A. Barnes, "Radiometric measurement comparisons at NASA's Goddard Space Flight Center: Part 1. The GSFC sphere sources," *Earth Observer*, vol. 14, pp. 3–7, 2002.
[4] W. A. Hovis, J. S. Knoll, and G. R. Smith, "Aircraft measurements for calibration of an orbiting spacecraft sensor," *Appl. Opt.*, vol. 24, no. 3, pp. 407–410, Feb. 1985.
[5] P. Abel, B. Guenther, R. N. Galimore, and J. W. Cooper, "Calibration results for NOAA-11 AVHRR channels 1 and 2 from congruent path aircraft observations," *J. Atmosp. Ocean. Technol.*, vol. 10, no. 4, pp. 493–508, Aug. 1993.
[6] P. Ardanuy, B. Bergen, A. Huang, G. Kratz, J. Puschell, C. Schueler, and J. Walker, "Simultaneous overpass off nadir (SOON): A method for unified calibration/validation across IEOS and GEOSS system of systems," in *Proc. SPIE*, San Diego, 2006, vol. 6301, p. 630 101.
[7] A. K. Heidinger, C. Cao, and J. T. Sullivan, "Using Moderate Resolution Imaging Spectrometer (MODIS) to calibrate Advance Very High Resolution Radiometer reflectance channels," *J. Geophys. Res.*, vol. 107, no. D23, pp. 4702–4704, Dec. 2002.
[8] A. Wu, X. Xiong, and C. Cao, "Terra and Aqua MODIS inter-comparison of three reflective solar bands using AVHRR onboard the NOAA-KLM satellites," *Int. J. Remote Sens.*, vol. 29, no. 7, pp. 1997–2010, Apr. 2008.
[9] P. M. Teillet, G. Fedosejevs, R. P. Gauthier, N. T. O'Neill, K. J. Thome, S. F. Biggar, H. Ripley, and A. Meygret, "A generalized approach to the vicarious calibration of multiple Earth observation sensors using hyperspectral data," *Remote Sens. Environ.*, vol. 77, no. 3, pp. 304–327, Sep. 2001.
[10] K. J. Thome, S. F. Biggar, and W. T. Wisniewski, "Cross-comparison of EO-1 sensors and other Earth resources sensors to Landsat-7 ETM+ using railroad valley playa," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1180–1188, Jun. 2003.
[11] P. M. Teillet, G. Fedosejevs, K. Thome, and J. L. Barker, "Impacts of spectral band difference effects on radiometric cross-calibration between satellite sensors in the solar-reflective spectral domain," *Remote Sens. Environ.*, vol. 110, no. 3, pp. 393–409, Oct. 2007.
[12] E. F. Vermote and N. Z. Saleous, "Calibration of NOAA-16 AVHRR over a desert site using MODIS data," *Remote Sens. Environ.*, vol. 105, no. 3, pp. 214–220, Dec. 2006.
[13] K. J. Thome, "In-flight intersensor radiometric calibration using vicarious approaches," in *Post-Launch Calibration of Satellite Sensors*, S. A. Morain and A. M. Budge, Eds. Philadephia, PA, USA: Balkema Publ., 2004, pp. 93–102.
[14] K. J. Thome, D. L. Helder, D. Aaron, and J. D. Dewald, "Landsat-5 TM and Landsat-7 ETM+ absolute radiometric calibration using the reflectance-based method," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 12, pp. 2777–2785, Dec. 2004.
[15] K. J. Thome, J. Geis, and C. Cattrall, "Comparison of ground-reference calibration results for Landsat-7 ETM+ for large and small test sites," in *Proc. SPIE Conf.*, 2005, vol. 5882, p. 58820A.
[16] K. Thome, C. Cattrall, J. D'Amico, and J. Geis, "Ground-reference calibration results for Landsat-7 ETM+," in *Proc. SPIE Conf.*, 2005, vol. 5882, p. 58820B.
[17] P. Nandy, K. Thome, and S. Biggar, "Characterization and field use of a CCD camera system for retrieval of bi-directional reflectance distribution function," *J. Geophys. Res.*, vol. 106, no. D11, pp. 11 957–11 966, Jan. 2001.
[18] Y. Yamaguchi, A. Kahle, H. Tsu, T. Kawakami, and M. Pniel, "Overview of Advanced Spaceborne Thermal Emission and Reflection radiometer (ASTER)," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1062–1071, Jul. 1998.
[19] W. L. Barnes, T. S. Pagano, and V. V. Salomonson, "Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1088–1100, Jul. 1998.
[20] D. E. Lencioni, C. J. Digenis, W. E. Bicknell, D. R. Hearn, and J. A. Mendenhall, "Design and performance of the EO-1 Advanced Land Imager," in *Proc. SPIE*, 1999, vol. 3870, p. 269.
[21] K. J. Thome, K. Arai, S. Tsuchida, and S. Biggar, "Vicarious calibration of ASTER via the reflectance-based approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 3285–3295, Oct. 2008.
[22] R. A. Barnes, R. E. Eplee, Jr., G. M. Schmidt, F. S. Patt, and C. R. McClain, "Calibration of SeaWiFS. I. Direct techniques," *Appl. Opt.*, vol. 40, no. 36, pp. 6682–6700, Dec. 2001.

Kurt Thome received the B.S. degree in meteorology from Texas A&M University, College Station, TX, and the M.S. and Ph.D. degrees in atmospheric sciences from the University of Arizona, Tucson, AZ.

He then joined what is now the College of Optical Sciences and became a Full Professor in 2006, where he served as Director of the Remote Sensing Group from 1997 to 2008. In 2008, he became a Physical Scientist in the Biospheric Sciences Branch of NASA's Goddard Space Flight Center. He has been a member of the Landsat-7, ASTER, MODIS, and EO-1 Science Teams providing vicarious calibration results for those and other imaging sensors.

Dr. Thome is a Fellow of SPIE, is the Instrument Scientist for the Visible Infrared Imaging Radiometer Suite on the Joint Polar Satellite System, and is serving as the Calibration Lead for the Thermal Infrared Sensor on the Landsat Data Continuity Mission. He is the Deputy Project Scientist for CLARREO for which he is also the instrument lead for the Reflected Solar Instrument.

Joel McCorkel received the B.S. degree in optical engineering and the Ph.D. degree in optical sciences from the University of Arizona, Tucson, AZ.

In 2009, he joined the National Ecological Observatory Network as a staff scientist for the Airborne Observation Platform. He is currently a Physical Scientist in the Biospheric Sciences Laboratory at NASA's Goddard Space Flight Center, Greenbelt, MD, USA.

Jeff Czapla-Myers received the B.S. degree in optical engineering from the University of Arizona, Tucson, AZ, in 1997, the M.Sc. degree in earth and space science from York University, Toronto, ON, Canada, in 2000, and the Ph.D. degree in optical sciences from the University of Arizona in 2006.

He is an Assistant Research Professor in the College of Optical Sciences at the University of Arizona. His research interests include remote sensing, radiometry, ground-based vicarious calibration of airborne and satellite systems, and the design, development, and laboratory characterization of radiometers.