

Technical Note

Observations and Recommendations for Coordinated Calibration Activities of Government and Commercial Optical Satellite Systems

Dennis Helder ^{1,*} , Cody Anderson ¹, Keith Beckett ², Rasmus Houborg ², Ignacio Zuleta ², Valentina Boccia ³, Sebastien Clerc ⁴ , Michele Kuester ⁵, Brian Markham ⁶ and Mary Pagnutti ⁷

¹ U.S. Geological Survey, Earth Resources Observation and Science Center, 47914 252nd Street, Sioux Falls, SD 57198, USA; chanderson@usgs.gov

² Planet, Inc., San Francisco, CA 94103, USA; keith@planet.com (K.B.); rasmus@planet.com (R.H.); ignacio@planet.com (I.Z.)

³ European Space Agency, Science, Applications and Future Technologies Department, Directorate of Earth Observation Programmes, ESRIN, Via Galileo Galilei snc, I-00044 Frascati, Italy; valentina.boccia@esa.int

⁴ ACRI-ST, 260 Route du Pin Montard, 06904 Sophia Antipolis, France; sebastien.clerc@acri-st.fr

⁵ Maxar (previously Digital Globe), Remote Sensing Sciences, 1300 W. 120th Ave, Westminster, CO 80234, USA; michele.kuester@digitalglobe.com

⁶ NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA; brian.l.markham@nasa.gov

⁷ Innovative Imaging and Research, Building 1103, Suite 140C, Stennis Space Center, Hancock County, MS 39529, USA; mpagnutti@i2rcorp.com

* Correspondence: dhelder@usgs.gov; Tel.: +1-605-594-2629

Received: 24 June 2020; Accepted: 24 July 2020; Published: 31 July 2020



Abstract: One of the biggest changes in the world of optical remote sensing over the last several years is the sheer increase in the number of sensors that are imaging the Earth in moderate to high spatial resolution. With respect to the calibration of these sensors, they are broadly classified into two types, namely government systems and commercial systems. Because of the differences in the design and mission of these sensor types, calibration approaches are often substantially different. Thus, an opportunity exists to foster discussion between calibration teams for these sensors with the goal of improving overall sensor calibration and data interoperability. The approach used to accomplish this task was a one-day workshop where team members from both government and commercial sensors could share best practices, discuss methods for collaboration and improvement, and make recommendations for continuing activities. Five major recommendations were developed from the event that focused on coordinated activities using pseudo invariant calibration sites (PICS), broader and more consistent communication, collaboration on specific cross-calibration opportunities, developing a reference sensor for all optical systems, and encouraging the coordinated development of surface reflectance products. Workshop participants concluded that regular interactions between these teams could foster a better calibration of all sensor systems and accelerate the improved interoperability of surface products.

Keywords: calibration; geometric; radiometric; Landsat; Sentinel; Dove; Worldview

1. Introduction

One of the biggest changes in the world of optical remote sensing over the last several years has been the sheer increase in the number of sensors that are imaging the Earth in moderate to high resolution. These systems are categorized in several ways, but one major distinction is to classify them as large government systems as opposed to smaller commercial systems. As a result, this increased

number of sensors from both types of systems provides the community with a tremendous opportunity to combine data sets. However, ensuring interoperability between data sets is clearly based on how well the sensors are calibrated. From a calibration perspective, this classification is also convenient because these two types of systems generally have substantially differing calibration systems and procedures. Government systems often include major onboard calibration capability that can be costly to develop, while commercial systems often resort to vicarious calibration methods as a cost savings approach. Because the missions of these two types of sensors can differ as well, the capabilities and limitations of the respective calibration methodologies can be substantially different.

To investigate these differences, with their corresponding capabilities and limitations, and how these calibration methodologies can potentially complement one another, the annual U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Cal/Val Center of Excellence (ECCOE) workshop was convened with a panel of experts to focus on the issue of calibration and opportunities to interact. The workshop was held at the National Oceanic and Atmospheric Administration (NOAA) Center for Weather and Climate Prediction in College Park, MD, USA, on September 20, 2018. The question presented to the panel for discussion was “Optical Satellite Calibration and Quality Comparison: How Good are Smallsat Data?”

Panelists represented three separate aspects of the question under discussion, namely government satellite calibration, commercial satellite calibration, and independent calibration and applications specialists. Panelists, along with their affiliations and expertise, are listed in Table 1.

Table 1. List of 2018 ECCOE Workshop Panel Members.

Panel Member	Affiliation	Expertise
Brian Markham	NASA	Prelaunch Calibration
Cody Anderson	USGS EROS	Postlaunch Calibration
Sebastien Clerc	ACRI/DSA	Geometric Calibration
Valentina Boccia	ESA	Sentinel 2 Quality Assurance
Ignacio Zuleta	Planet	Imaging/Payload Operations
Keith Beckett	Planet	Principal Scientist
Michele Kuester	Maxar (formerly Digital Globe)	Radiometric Calibration Lead
Mary Pagnutti	Innovative Imaging Research	Vicarious Calibration
Rasmus Houborg	South Dakota State University (with Planet at publication)	Physical Geography

To provide a framework for the discussions, the panelists were asked to address four key questions:

1. What are the calibration capabilities of government systems?
2. What are the calibration capabilities of commercial systems?
3. With respect to these capabilities, what are the limitations of each type of system and the differences between them?
4. What activities can be implemented to optimize the calibration of both sets of systems?

For each panelist to state their initial position on these questions, 20-min blocks of time for presentations were provided. Following this introduction, the rest of the time was devoted to discussions among the panel members with periods when the audience was also invited to interact with the panel.

Finally, a set of recommendations was drawn from the presentations and discussions that can be used as guidance for the remote sensing calibration community to improve the quality and coordination of our calibration activities. The following three sections of this paper highlight the perspective given to each of the key questions by each of the three segments of the panel. These sections are followed by highlights of the panel’s discussions and panel recommendations followed by a short conclusions section.

2. Calibration Capabilities and Limitations of Large Government Systems

Characteristics and performance of government-based calibration systems were provided from a Landsat and Sentinel 2 perspective covering pre-launch and post-launch calibration as well as radiometric and geometric calibration. The following discussion, while not exhaustive in nature, is representative of the current state-of-the-art for the calibration of large government systems. Due to the makeup of the panel, the radiometric calibration perspective is developed from the Landsat program, while Sentinel 2 is used to develop the geometric perspective.

2.1. Radiometric Calibration of Government Systems

One key consideration to the overall radiometric calibration of a sensor is the contribution to uncertainties from the spectral shaping components of an imaging system. The major moderate spatial resolution government systems in development today are of the pushbroom design and consist of a broad linear array of detectors. Generally, these focal planes are designed as an array of modules each requiring a set of spectral bandpass filters. Thus, one of the major concerns for calibration of these systems is uniformity across the modules. In the case of the Landsat 9 Operational Land Imager Two (OLI-2), the spectral bandpasses tend to vary on the order of 0.4 nm from one module to the next. The key concern is the effect that this variation would have on typical types of surfaces. Figure 1 illustrates the impact the variation would have on vegetative surfaces and bare desert surfaces across the OLI-2 focal plane [1]. In the case of vegetation, the impact can be as large as 0.15% while for deserts it is considerably less, i.e., below 0.05%. For comparison purposes, Sentinel-2 differences in at-sensor reflectance due to spectral band differences can be as large as 5% in the red edge bands that were designed to be especially responsive to vegetation. This is due to the very large reflectance gradient that vegetation exhibits at these wavelengths.

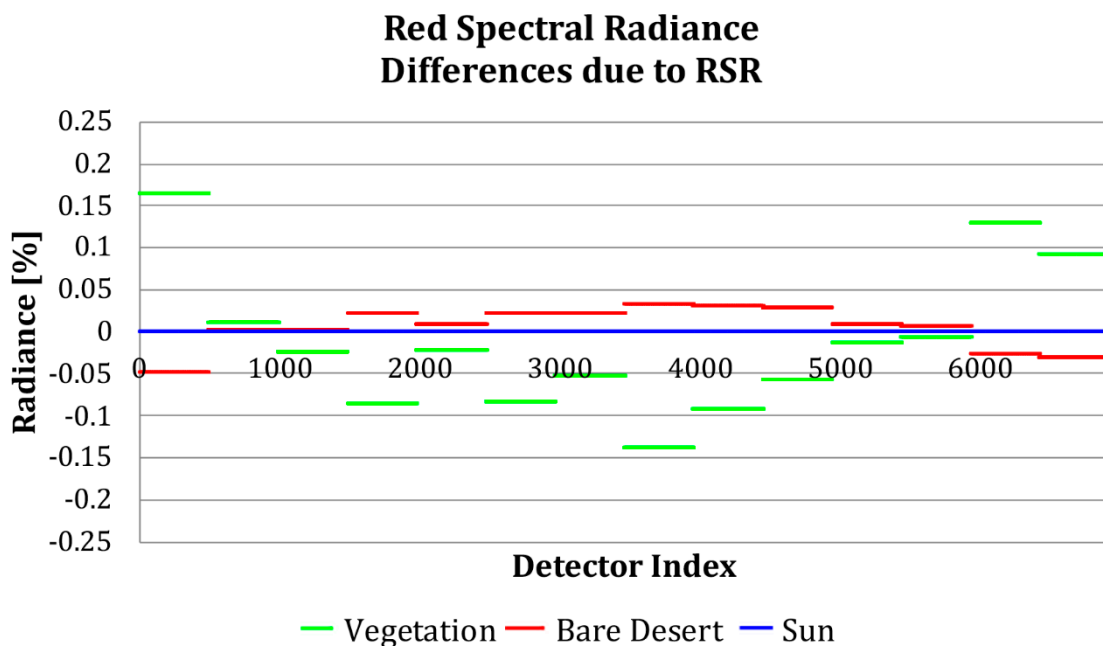


Figure 1. Changes in radiance from a constant sun source that are expected from vegetation and desert targets due to differences in relative spectral response (RSR) across the OLI-2 focal plane.

A second critical aspect is how well the spectral uncertainties are known at the system level as a simple sum of component uncertainties often does not capture additional interactions when systems are assembled. An example of this is shown for the OLI-2 SWIR1 Band in Figure 2, which shows that component level spectral responses can change by more than an order of magnitude when system level issues such as crosstalk are factored into the overall uncertainty estimates.

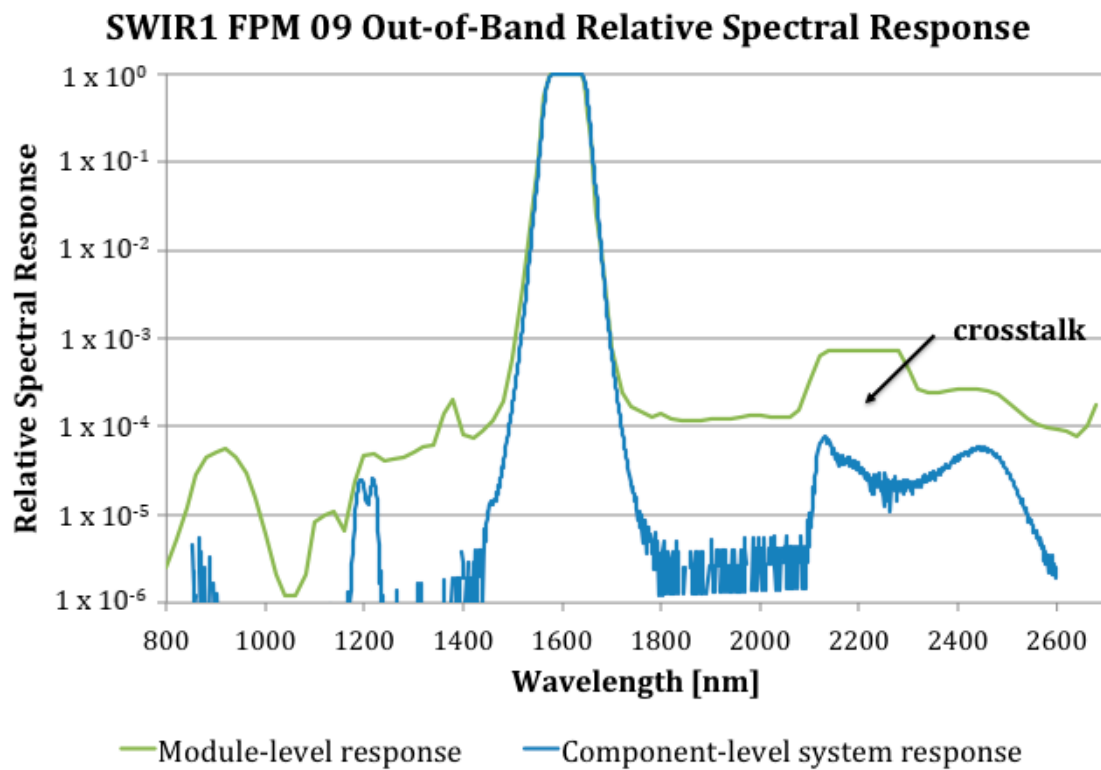


Figure 2. Relative spectral response (RSR) for OLI-2 SWIR1 Band and Focal Plane Module (FPM) 9. Blue curve shows component level result, green curve shows the RSR at the module level. Note the substantial loss in out-of-band rejection due to interaction between FPM components.

The overall uncertainties due to spectral differences are small. For example, typical values would be $\pm 0.5\%$ for differences in spectral filter relative spectral responses, $\pm 0.5\%$ for uncertainties in overall system level spectral response, and $\pm 0.5\%$ for out-of-band response uncertainty. However, these uncertainties start to become major contributors now that absolute pre-launch radiometric calibration is beginning to approach the 1% uncertainty level.

Radiometric calibration is often broken down into two components, namely relative and absolute. Relative radiometric calibration refers to the uniformity in detector response across the focal plane of the sensor. Absolute radiometric calibration describes the ability to place the digital numbers obtained from the sensor on an absolute radiance or reflectance scale.

Uniformity (sometimes called flat-fielding) across a focal plane has become an increasingly difficult problem as the number of detectors has grown into the thousands and radiometric resolution is in the 12–16 bit range. Current Landsat requirements for uniformity are 0.5% or less across the field of view (FOV) and the same for any individual detector. This becomes especially difficult for low-level signals such as over water. Measurements by Pahlevan et al. [2] suggest that uniformity is being accomplished for Landsat 8 OLI.

Historically, absolute radiometric calibration for sensors was done as a function of radiance. A radiance-based approach takes advantage of the fact that the sensor is directly measuring radiance at its aperture. However, more recently sensors are being calibrated as a function of reflectance. There are two key advantages to this approach: (1) uncertainties in the estimate can be reduced as a result of removing the need for a solar irradiance model, and (2) users of remote sensing data prefer working in the reflectance domain as opposed to a radiance domain. As an example of the uncertainties inherent in the reflectance approach, Table 2 shows the reflectance-based calibration uncertainty for Landsat 8 OLI. Note that uncertainties for this instrument are about 2%. Landsat 9 OLI-2 is expected to reduce

these uncertainties even further through additional prelaunch spectral characterizations (as alluded to previously) and the use of tunable laser-based calibration instrumentation [3].

Table 2. Prelaunch reflectance-based calibration uncertainty for Landsat 8 OLI.

Term	Uncertainty								
	CA	Blue	Green	Red	NIR	SWIR1	SWIR2	Pan	Cirrus
Initial Diffuser BRDF	1.4%	1.3%	1.1%	1.0%	1.0%	1.7%	1.4%	1.1%	1.7%
Diffuser Geometry	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%
Diffuser Stray Light	0.7%	0.7%	0.7%	0.7%	0.7%	0.7%	0.7%	0.7%	0.7%
Pristine -> Wkg	0%	0%	0%	0%	0%	0%	0%	0%	0%
Non-Linearity									
Wkg -> Scene	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%
Non-Linearity									
FFOV Non-Uniformity	0.3%	0.2%	0.3%	0.3%	0.3%	0.3%	0.4%	0.3%	0.3%
Long Term Stability	0.1%	0.1%	0.1%	0.1%	0.0%	0.3%	0.2%	0.1%	0.6%
Total	2.1%	1.9%	1.7%	1.7%	1.7%	2.2%	2.0%	1.7%	2.3%

Several important strengths exist in the post-launch calibration of large government sensors. Because of the efforts placed in controlling the environment of the instrument, long-term stability can be excellent. Examples include the bias and gain of Landsat 8, which was launched in February 2013, and has exhibited little change in either radiometric gain or bias over five years as shown in Figure 3. The left-hand plot shows that the bias for the coastal aerosol band has a temporal uncertainty of only 0.07%. On the right-hand side of Figure 3, the change in gain for the coastal aerosol band is on the order of 1% after 5 years of life. The rest of the bands have changed far less over time. Despite the stability exhibited by these instruments, calibration is monitored regularly to ensure that any change that does occur is detected as early as possible.

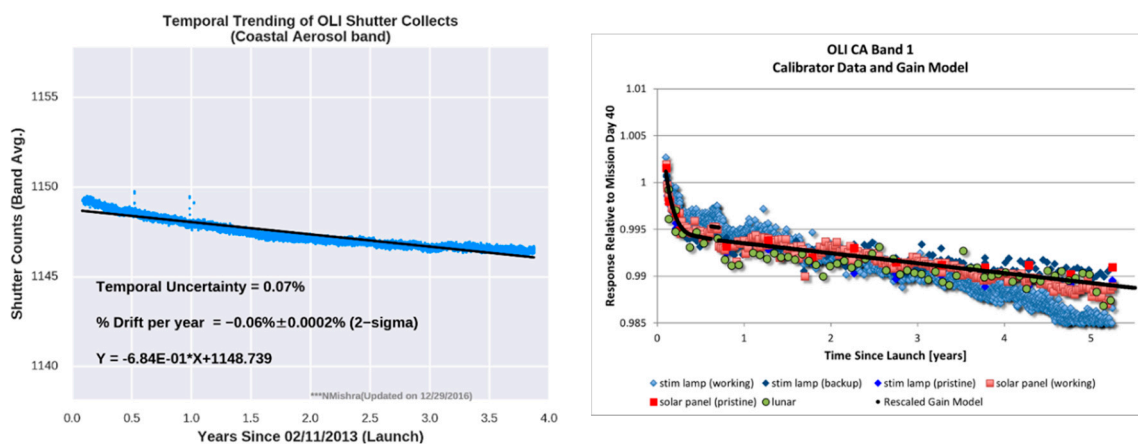


Figure 3. Radiometric bias and gain trends for Landsat 8 Operational Land Imager Coastal Aerosol Band over approximately 5 years of operation.

Vicarious calibration is also regularly performed post-launch using a variety of sources including both terrestrial and lunar targets. These methodologies, while not as accurate or precise as pre-launch and onboard calibration methods, nevertheless provide an independent assessment of the calibration of the instruments. As an example of what is possible using vicarious methods, Figure 4 shows vicarious calibration of Landsat 8 thermal infrared scanner (TIRS) using buoys, which measure water temperature. Figure 4 shows that results from two teams, Rochester Institute of Technology and the NASA Jet Propulsion Laboratory, not only agree well, but span a broad range of radiance/temperature values. Uncertainties associated with this method of calibration are on the order of 0.5K.

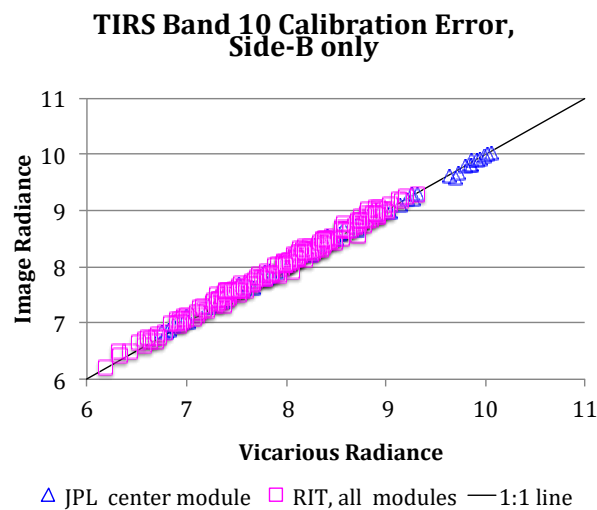


Figure 4. Calibration of Landsat 8 TIRS Band 10 using buoys that measure water temperature. The red squares indicate data from buoys deployed around the coast of the United States as collected by Rochester Institute of Technology and the blue triangles are data from buoys deployed on the Salton Sea and Lake Tahoe in California, USA, by the NASA Jet Propulsion Laboratory. Standard units of radiance are used ($W/m^2/sr/\mu m$).

A second terrestrial vicarious calibration approach uses pseudo-invariant calibration sites (PICS) to determine the stability of sensors over time. PICS are normally located in desert regions such as the Sahara Desert in North Africa because they must be as immune as possible to any change in surface reflectance. Thus, arid regions with no vegetation are primarily targeted. By integrating information obtained from multiple PICS, reasonably precise estimates of sensor drift are obtained. Figure 5 shows an example using four PICS compared to estimates based from observations of an onboard diffuser panel for Landsat 8 OLI. In this example, the weighted average from PICS is indistinguishable from the estimate obtained from an onboard diffuser. This is especially important for smallsats since stability is determined from vicarious measurements when the onboard methodology is not feasible due to size or cost constraints.

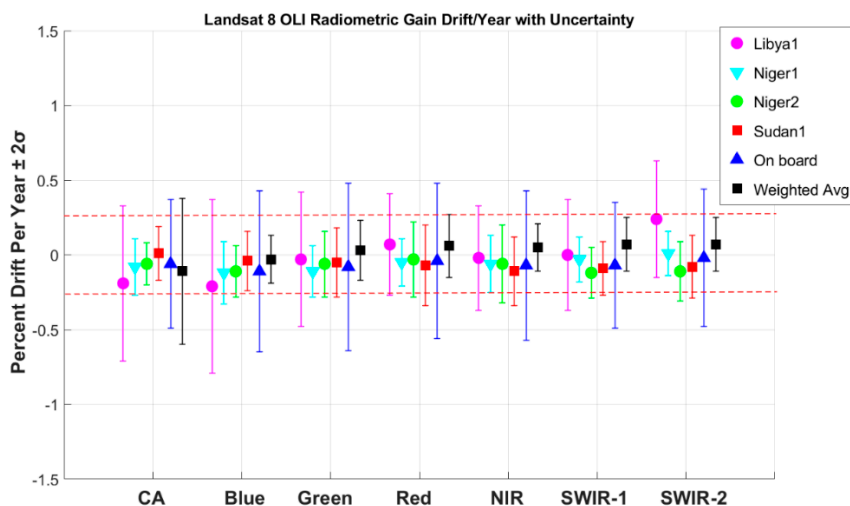


Figure 5. Estimates of long-term drift for Landsat 8 OLI as measured by observation of PICS and an onboard diffuser. By averaging several PICS together, results are statistically indistinguishable from onboard methods. Compare the weighted PICS averages (black squares), inversely weighted by their uncertainty, to the onboard diffuser (blue triangle). Figure courtesy of South Dakota State University Image Processing Laboratory.

Cross-calibration of sensors is also performed using vicarious PICS-based methods. Because PICS are stable for extended periods of time, sensors do not have to observe them simultaneously. Observations can be several days apart provided they are at essentially the same viewing/illumination angles. An example of this is shown for Landsat 8, Sentinel 2A, and Sentinel 2B in Figure 6. In this plot, all three instruments observed the PICS commonly known as Libya 4 over a period of three years. The calibration of these instruments can be determined with a precision of 2%. Furthermore, there was no statistical difference between comparisons made from coincident collects and comparisons made up to six days apart [4]. This methodology is especially useful for both large government sensors and small commercial sensors since it is agnostic to sensor type, requires little or no time and effort, and thus is low cost.

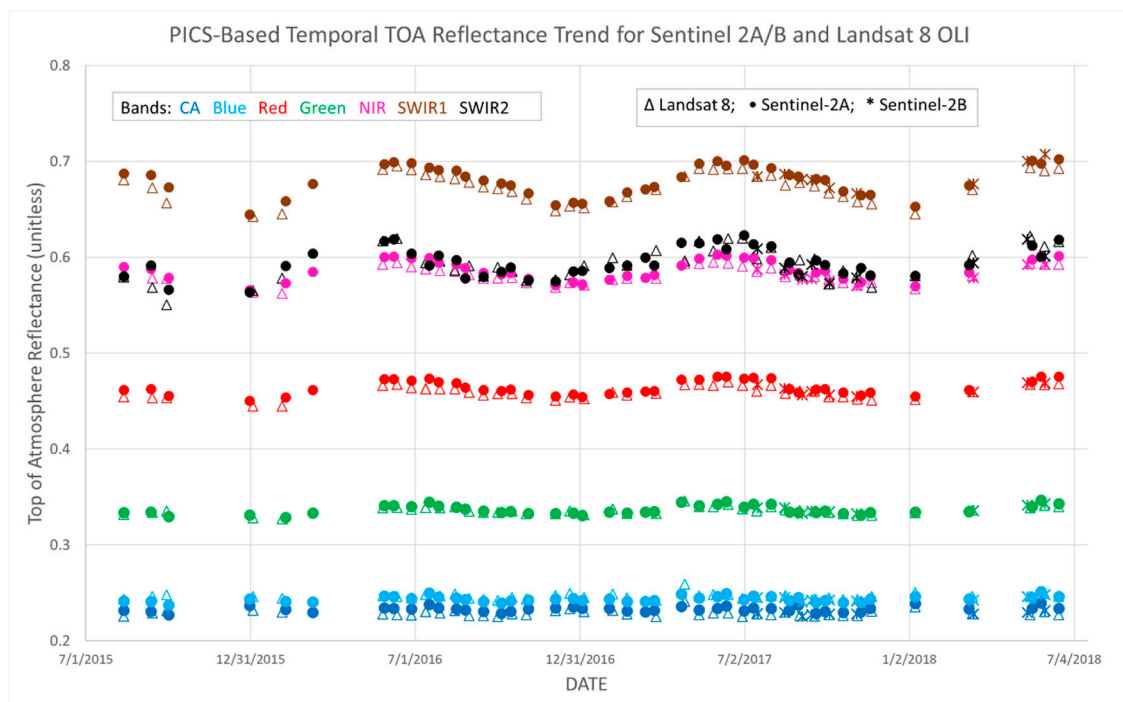


Figure 6. Temporal trends of observations of the Libya 4 PICS by Landsat 8 OLI, Sentinel 2A and Sentinel 2B. These trends allow for cross-calibration of sensors, after accounting for spectral and viewing/illumination geometry differences (however, these data do not include corrections for annual solar elevation angle changes). Observations up to 6 days apart provide the same level of precision as simultaneous observations due to the stability of these sites. Figure courtesy of South Dakota State University Image Processing Laboratory.

Vicarious methodologies provide opportunities for comparisons and the inter-calibration of large government sensors and small commercial sensors. However, inherent in performing these comparisons is the need to share key information about the respective sensors, such as relative spectral responses, point spread functions, ground sample distance, signal-to-noise ratio, uniformity, etc. With this type of information available, it then becomes key to ensure that those who are performing the sensor comparisons use agreed upon processing approaches that ensure a consistent apples-to-apples comparison. Once these items are in place, it will be possible to better understand the strengths and weaknesses of all sensors, which is essential as scientists and commercial users blend data together from multiple sensors to increase the temporal resolution of the phenomena they wish to monitor. Lastly, understanding the calibration of a sensor generally improves over the sensor lifetime as more information from it becomes available. Thus, it is also key that, as calibration improves over time, the historical data acquired by the sensor must be recalibrated so that the data are as accurate as possible.

2.2. Geometric Calibration of Government Systems

Geometric calibration is arguably more important than radiometric calibration. If one does not know where the pixels are on the ground, then radiometric accuracy is not useful. Thus, a great deal of effort is expended to ensure the geometric accuracy of large government sensors. The greatest need for most users is geometric consistency. That is, relative geometric calibration (consistent location image to image) is more important than absolute geometric accuracy (making sure the image is absolutely located on the Earth's surface). However, it is obvious that relative accuracy implies good absolute accuracy as well. A goal for government systems is accuracy of 0.3 pixel for temporal analysis.

Geometric calibration begins with a good understanding of where the satellite is pointing based on the attitude and orbital control system on the satellite. The pointing error can be characterized at first order by constant biases between the estimated and observed line of sight. An example of this is shown for Sentinel 2B for its first two years in orbit (Figure 7). Here, the pitch and yaw angles are stabilizing, but the roll angle continues to change. Thus, it is important to monitor these angles on a continuous basis to understand the behavior of the satellite and ensure it remains within the control box so further correction is not necessary.

To reduce the need for continuous calibration updating, the Copernicus Sentinel 2 program developed a new approach for absolute geometric calibration called the global reference image (GRI). The idea is to develop a set of reference images with calibrated geometry. The GRI was constructed by selecting a set of data strips that were as cloud-free as possible with sufficient overlap between adjacent strips [5]. Spatio-triangulation using a global set of ground control points (GCPs) was used to correct the viewing models. By performing these steps at a continental scale, block level accuracy was ensured across multiple data strips. Finally, global adjustment was done to ensure consistency across blocks. The result is a global reference with accuracy requirement of 10 m that performs with accuracy of 7 m (95% confidence).

To ensure the greatest degree of interoperability between data sets, other missions could adopt or incorporate the GRI with their current reference. Major advantages include global coverage, including high latitudes and islands where GCPs are scarce. This is the case for the Landsat 8 sensor. The Landsat program maintains their data archive on a collection-based reprocessing schedule. The Collection 2 update, currently scheduled for 2020, also incorporates the GRI as its absolute geometric calibration basis. If multiple missions were to adopt the same reference for geometric calibration, data interoperability would be greatly enhanced. This concept is important for interoperability between government sensors and commercial smallsat sensors. However, better than 7 m accuracy may be required for high spatial resolution sensors.

In addition to using a common reference for geometric calibration, a common digital elevation model, or DEM, is also strongly encouraged. The Copernicus program recently acquired two reference DEMs at 90 and 30 m. The specifications aim at a consistent accuracy across all latitudes and continents (including Antarctica), in order to alleviate limitations in currently available DEMs. The 90 m version will be released as open data. An overall goal for the community is the adoption of a common reference for absolute geolocation and a common DEM for altitude reference.

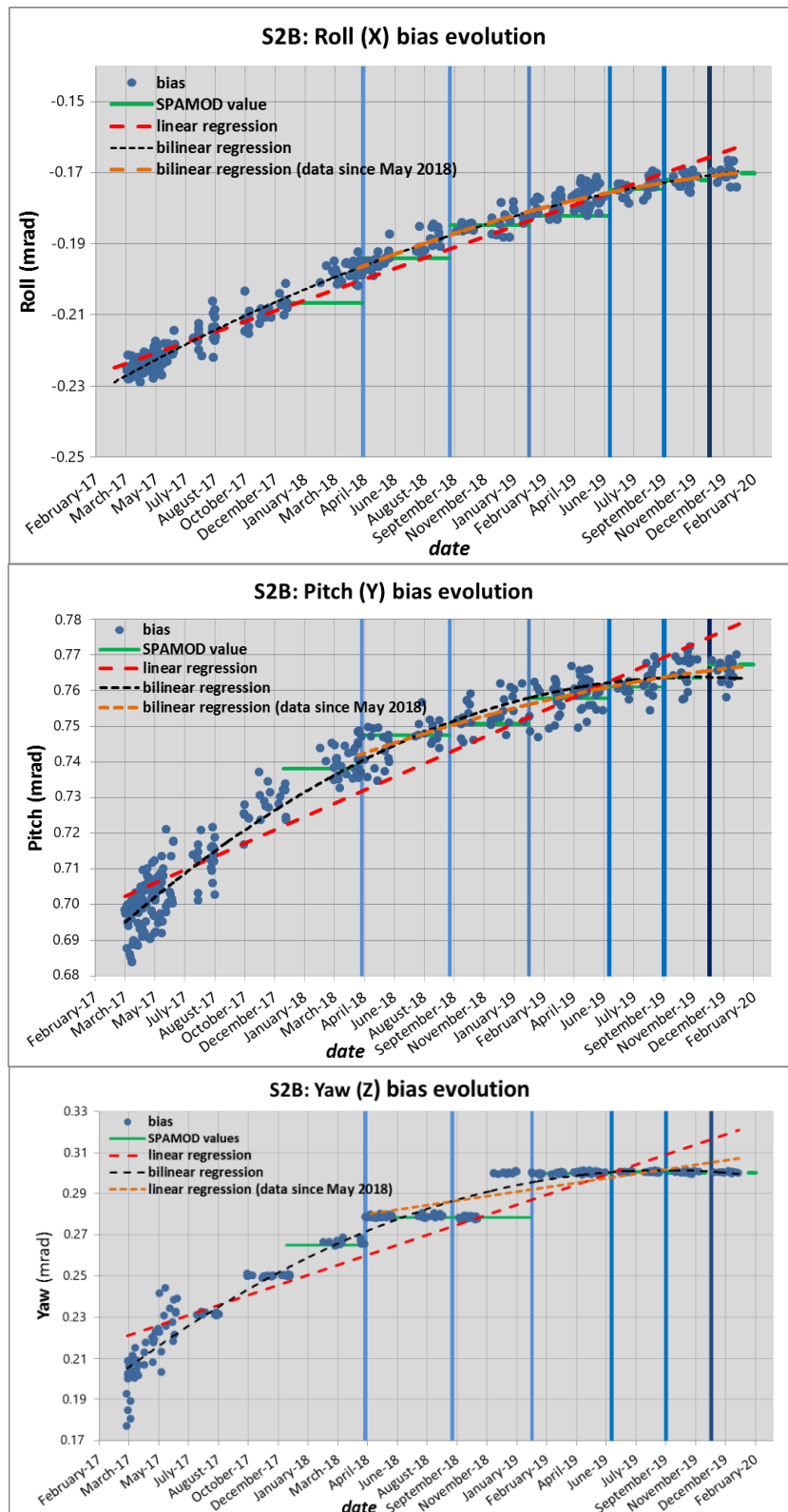


Figure 7. Roll, pitch and yaw angle measurement biases for Sentinel 2B for the first 550 days following launch. Note that pitch and yaw angle bias is stabilizing, while roll continues to drift.

2.3. Limitations to Calibration of Government Systems

Due to the large investment that is used in the development of government remote sensing satellites, there have been major improvements in calibration capabilities over the past several decades. Many instruments contain onboard calibration devices that have the potential for high levels of accuracy and precision. Multiple calibration/validation (cal/val) methods are often implemented. Extensive and quality data are available from the various missions. Often there are dedicated calibration teams monitoring the performance of their respective instruments. Data are often cross-validated with other similar missions. Extensive descriptions exist of the technology and processing algorithms that are used. Often these procedures are internationally agreed upon and there is coordination across government agencies.

However, limitations do exist. There is often disagreement concerning the procedures used to perform calibration of instruments. As a result, when data from instruments are compared, there can be uncertainty about whether differences are due to sensors or processing methods. Broad agreement on procedures remains elusive. Opportunities to improve in this area include broad adoption of RadCalNet across missions [6] and standardization in the use of PICS. These represent the two key terrestrial vicarious calibration methodologies. Other difficult areas of calibration that are not fully developed include sensor response non-linearity and low-level radiance calibration. These limitations are amplified by the development of sensors with finer radiometric resolution and users who are monitoring portions of the Earth with very low signal levels such as water bodies. The Committee on Earth Observation Satellites Working Group on Calibration and Validation (CEOS WGCV) is working on coordinating many of these efforts internationally.

In addition to cal/val, activities for surface-based products such as surface reflectance and temperature are becoming increasingly important for calibration teams. Limitations here include a limited set of ground truth measurements that do not represent the various combinations of surface type and atmosphere on our planet. Inherent in the validation of surface products is the understanding of radiative transfer codes (RTCs). Many RTCs exist and standardization to date remains limited.

To improve interactions amongst the government community, several opportunities exist. The Committee on Earth Observation Satellites (CEOS) provides an international forum for addressing many of these issues. Better engagement with this organization is needed. Related to CEOS is the atmospheric comparison inter-comparison exercise (ACIX) and the Cloud masking inter-comparison exercise (CMIX). These venues provide additional opportunities to improve understanding and coordination of the atmospheric issues that limit development of surface products. All of these limitations suggest that more collaboration amongst governmental agencies would be beneficial, as well as exploitation of advanced technologies such as cloud computing and artificial intelligence, to improve our global capabilities for the calibration of government sensors.

3. Calibration Capabilities and Limitations of Commercial Systems

3.1. Cubesat Calibration

Cubesats for optical remote sensing have been on an explosive growth curve over the past decade with over 1000 launched as of the time of this writing [7]. Cubesats are generally built in multiples of a standard size, referred to as 3 U units ($10 \times 10 \times 11.35$ cm) and have a mass of less than 1.33 kg per unit. Because of their small size, deployment to space is low-cost and cubesats are often secondary payloads on launch vehicles. A major opportunity of using cubesats for remote sensing is a marked increase in temporal coverage. Daily observation of the planet at high spatial resolution (<10 m) is now possible.

Because of their small size, however, cubesats have unique limitations that impact data obtained from them. Due to small aperture size, signal-to-noise ratio (SNR) is often a problem, but it can be offset by time delay integration (TDI) for some sensors. Stray light is a more significant problem due to the reduced size of the optical system, and the modulation transfer function, or MTF, will be lower than large government sensors. Geometric calibration is more difficult due to less knowledge

of the spacecraft attitude, but is generally mitigated to some degree through use of ground control points and reference to government sensors. Spectral filters are also a greater challenge, but recently sensors were developed with spectral bandpasses based on government sensors such as Sentinel 2 and Landsat 8. These were available from vendors that had provided the filters to the government. Radiometric calibration is more challenging since onboard calibration systems are generally not possible; so vicarious calibration methods are relied upon heavily for these commercial systems. All of these factors lead to limited interoperability between cubesats and large government sensors and need addressing.

Opportunities for calibration collaboration between cubesats and large government sensors exist that can substantially improve data quality for these systems. Key to all of these is the development of standard practices that provide a consistent ability to compare data across platforms ranging from cubesats through large government systems. Engaging both commercial and government entities in the development of these standards will ensure standards are created in an open agile manner that reflects the current state-of-the-art and are testable and automated. Adoption of a common standard for geometric calibration is needed. With the Landsat 8 and Sentinel 2 sensors coordinating geometric calibration efforts through use of the GRI and a common DEM as described in the previous section, an obvious approach would be for smallsats to follow this same reference system. For radiometric calibration, the consistent and standardized use of vicarious calibration sites and methodologies, both terrestrial and lunar, are recommended.

The remote sensing community would benefit from greater interoperability between commercial sensors and government sensors. Thus, promoting greater interaction between calibration teams to work with data from multiple sensors systems in a holistic approach is a positive step towards validating the ability to merge data sets. Building cubesats with similar spectral bandpasses to government sensors will facilitate this concept and has already been implemented by Planet.

3.2. Smallsat Calibration

Since the beginning of the 21st century, high spatial resolution (1 m–10 m) commercial smallsats have been collecting imagery of targeted areas of the Earth's surface. Fitting in a niche between cubesats and large government optical satellites, these sensors are highly agile and able to take advantage of multiple looks at a single target or, conversely, single looks at many targets at a given location. Since smallsats are larger systems than cubesats, they are normally launched one at a time and flown in constellations of no more than a handful of satellites. Because these systems have a larger aperture and telescope, with larger power capabilities than cubesats, pointing knowledge and radiometric stability are significantly enhanced and become comparable to large government systems. Typically, because of the commercial focus of smallsats, onboard radiometric calibration systems are usually absent, and the instrument is calibrated vicariously using manned sites, automated sites, or PICS. Often spectral coverage is broader than cubesats extending into the SWIR wavelengths (1.6–2.4 μm). These instruments are multispectral often covering wavelengths similar to government systems such as Landsat, Sentinel 2, and moderate resolution imaging spectroradiometer (MODIS).

Commercial teams that use manned vicarious sites and the surface reflectance-based method of vicarious calibration have similar uncertainties and traceabilities as the government calibration teams—typically on the order of 3% top of atmosphere (TOA) radiance. Cross-calibration to government satellites using simultaneous nadir overpass (SNO) opportunities also show excellent agreement. Figure 8 is an example of the cross-calibration of the Digital Globe CAVIS instrument to Landsat 8. Excellent agreement is obtained, on the order of 3% or better, which is due in large part to an emphasis on similar spectral bandpasses.

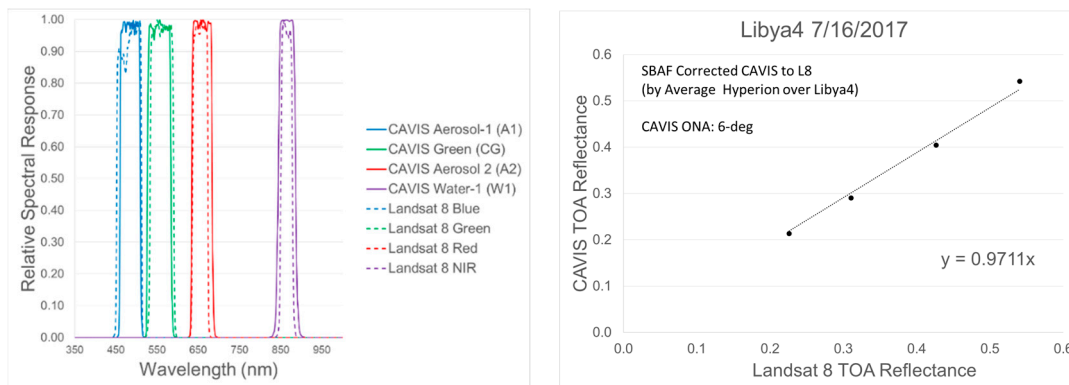


Figure 8. The left panel illustrates cross-calibration of Digital Globe’s CAVIS instrument to Landsat 8 using the Libya 4 PICS. Excellent agreement is obtained due in large part to the similarity of the spectral bandpasses between the two instruments as illustrated in the right panel.

Because of the agility available with commercial sensors, capabilities often exist that are not present with government sensors. One example is an enhanced bidirectional reflectance distribution function (BRDF) model of the Libya 4 PICS that was produced by Digital Globe’s constellation of sensors [8]. This type of information can substantially complement the calibration work done by both commercial and government teams and strongly suggests that all can benefit by increased collaborations.

Throughout the workshop presentations, increased collaboration between government and commercial calibration teams was strongly encouraged. These types of activities could include round-robin activities where multiple teams compare measurements and algorithms for consistency, interacting with freely available calibration systems such as RadCalNet, and extending efforts from TOA calibration to the validation of surface reflectance products that are becoming the primary product for both government and commercial sectors.

4. Independent Calibration and Applications Perspective

After presentations by representatives from the government and commercial optical remote sensing satellite sectors, the morning session of the workshop was concluded by considering input from independent calibration experts with close ties to the user community. The closing presentations provided not only an independent perspective, but also an opportunity to begin integrating ideas and developing recommendations from the workshop participants.

Perhaps one of the most important aspects of calibration, previously alluded to but in this part of the workshop explicitly addressed, is the competition that exists between user applications/needs and the cost of performing calibration. While all parties would agree that calibration can be improved with more effort/cost, the defining criteria for what should be accomplished/achieved needs to be based primarily on users’ needs and applications. As an example, is it worth the effort to improve absolute radiometric calibration to 1% accuracy and geometric accuracy to 0.25 pixel if there is not a set of user needs driving it? Conversely, there is also an unknown relationship between what user needs are today and what science can be done tomorrow if accuracies of this order were available. Thus, calibration needs to continue to follow a path of improved accuracy and precision, and the only open question is how fast and at what cost it should be pursued.

A second important point made by the independent presenters was the need for a gold standard. The user community deeply desires a reference of some type upon which all data are compared. Historically, the standard for optical systems has been the Landsat series of satellites that extend back into the 1970s. More recently, at coarser resolutions, the MODIS sensors are providing similar capabilities. In the past five years, the Sentinel-2 satellites have also exhibited accuracy and stability rivaling Landsat and MODIS. Thus, the independent calibration community strongly recommends the adoption of a standard, either one of those listed above or perhaps a spaceborne radiometer such as

represented by the CLARREO or TRUTHS projects [9,10]. The independent calibration community have also strongly suggested that any such instrument should be hyperspectral in nature so that it can be a standard for all optical remote sensing systems.

With respect to large government systems, observations were presented that calibration systems on these satellites represent a large investment in both time, money and are also generally complex systems that can be simplified. As an example of this, it was noted that the calibrator on the Landsat 8 thermal infrared sensor (TIRS) almost doubled the size of the instrument [11] and comments about the size and cost of the Landsat OLI calibration system have also been raised. As a suggestion for improvement, current onboard calibration lamp systems could be replaced with LED-based systems. LEDs are smaller and more rugged than lamps. They tend to be much more stable, degrade more slowly over time, and can last many more hours than a lamp. Multiple LEDs can be combined to obtain a desired emissive spectrum.

A second improvement is the development of onboard calibration radiometers to replace large diffuser panels [12]. Since the calibrator has much coarser resolution than the primary instrument, it can be designed for improved thermal management, improved stability, and will allow a simplified implementation, thus reducing size/weight/power by an order of magnitude or more. Instruments of this type could be built for all wavelengths from visible through thermal, could be flown on all large government sensors, and could be built with common traceability to national institute standards.

Similar to the concept presented by the commercial sector, the independent calibration experts also recommended an open access calibration infrastructure. This was especially emphasized as the government tends to be the largest customer for commercial data. Thus, it is in the best interest of all parties to develop a cal/val infrastructure and make the data produced by it freely available to the broader calibration community.

Specific ideas for building an open cal/val infrastructure include building upon existing instrumented sites. The most obvious opportunity for pursuing this effort would be to expand the capabilities of the RadCalNet program. Examples could include expanding BRDF knowledge, increasing areal coverage, and increasing number of sites. Expanding the types of sites that are used for cal/val was also discussed. Ideas here included different types of sites. Historically, cal/val teams have used large, high reflectance homogeneous sites for calibration. However, the use of heterogeneous sites, such as a small bright target surrounded by a lower reflectance region, would provide enhanced opportunities for the calibration of high-resolution sensors. Lastly, sites should be developed for radiometric, geometric, and spatial characterization/calibration. It was pointed out that, currently, the United States does not maintain a site, such as an edge target, for spatial characterization.

From the user perspective, it is extremely clear that their desire is for smallsats to augment and complement largesat systems. The huge advantage of smallsats is their greater spatiotemporal resolution. These properties can lead to enhanced time series of surface phenology through leveraging the greater calibration properties of large government satellites. Methodologies have already been developed using currently available systems. One such method, CESTEM, combines data from Landsat 8, MODIS, and Planet Doves [13]. In Figure 9, an example is shown of improved temporal resolution of alfalfa in Saudi Arabia. This plot clearly delineates the potential that exists when smallsat data are anchored by large government satellite measurements. It also suggests the potential that exists through improved calibration of smallsats.

Summarizing the user perspective points clearly to the need for large government systems to maintain, and even improve, their high standards for calibration accuracy and precision, ensure consistency and continuity of observations, and improve coordination across agencies and platforms. For the smallsat community, their contribution in terms of adding observational capabilities and closing temporal gaps is critical to the advancement of the industry. As this young segment continues to grow and develop over the next decade, efforts to improve the spatial, temporal, spectral, and radiometric accuracy and precision should continue.

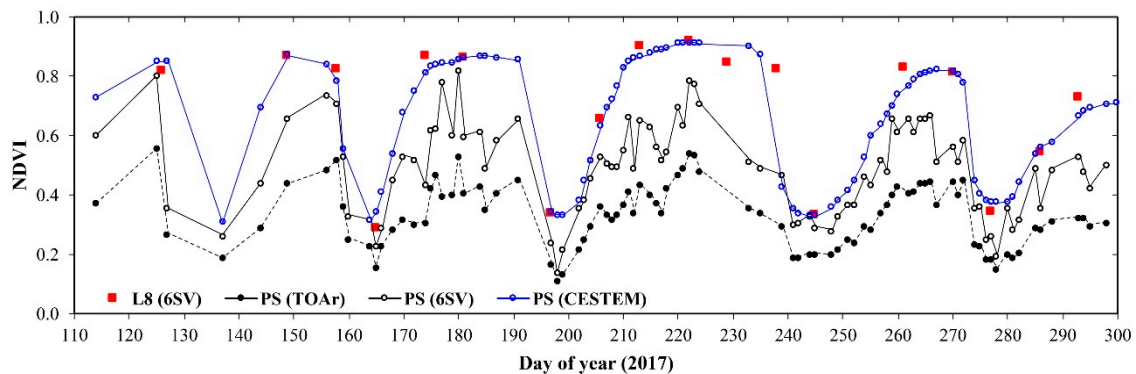


Figure 9. Temporal plot of Normalized Difference Vegetation Index (NDVI) measurements of alfalfa in Saudi Arabia. Measurements by Planet Doves (black circles) show poor calibration but rich temporal density. Combining these measurements with Landsat 8 (red squares) through CESTEM provides an accurate and spatially dense NDVI estimate (blue line).

5. Discussion

Following the presentations, the remainder of the workshop focused on panel and audience discussions. Over three dozen ideas for improved calibration and interaction between government satellites and smallsats were captured from the presentations. These ideas were discussed, prioritized, downsized, and then developed into specific workshop recommendations.

The first area of discussion centered around concepts for improved calibration of sensors. One key concern throughout the workshop was the limitations on and difficulties surrounding spectral calibration. If common spectral bandpasses could be agreed upon, cross-calibration of multispectral sensors could be improved substantially. Even more fundamental were ideas related to consistent sensor reporting and the adoption of similar calibration algorithms, which would lead to a better understanding of the strengths and weaknesses of sensors allowing much better cross comparisons. It was noted that adoption of the GRI as a geometric reference is improving the consistency between government sensors and has the potential for extension into the commercial arena.

In particular, workshop participants centered on the use of PICS for improved calibration between the largesats and smallsats. PICS were identified as a calibration methodology that is readily available for both groups of sensors and can be employed at low cost with significant potential for calibration improvement. However, there are still many areas of PICS-based calibration that need to be addressed. First among these is agreement on a common region of interest (ROI). The most commonly used site is Libya 4. However, even within that site, there are multiple ROIs used by various groups. Standardization on a common ROI, likely the one recommended by CEOS, is suggested. However, the significantly differing spatial resolutions between government and commercial sensors need to be considered. In addition, it was strongly suggested that a second, dark PICS be developed with a surface reflectance in the 2–3% range. Both government and commercial sensor systems have some weakness in calibration at the low end of the dynamic range and this is a major step forward to collect consistent data to address this concern.

Because of the substantial differences that can exist in viewing geometries between sensors, both viewing and illumination angles will need consideration in the development of a PICS, which strongly suggests the need for a standardized BRDF model for both Libya 4 and the dark PICS. Coupled with any inter-calibration effort, it is necessary for the relative spectral responses of the instrument to be known as well as the spectral model of the target. Thus, a spectral model for the Libya 4 and proposed dark PICS would need to be standardized as well. It was also suggested that the models developed for PICS be done in both TOA radiance and reflectance spaces.

A second focus area was improved communications between the government calibration teams and the smallsat calibration teams. Communications is an area that everyone agrees is needed but all

too easy to overlook on a day-to-day basis. Thus, it is important that this recommendation include specific examples for implementation with specific achievable goals so that the benefits are visible and encourage continued interaction. The first suggestion in this topic area was greater participation by commercial entities in CEOS. Commercial participation is strongly encouraged in the CEOS WGCV subgroup on infrared visible and optical sensors (IVOS). This subgroup meets at least annually and is addressing many of the topic areas of interest to this panel, including PICS as discussed in the preceding paragraphs.

Other key areas for improved communications include attendance at yearly workshops held on quality assessment of commercial systems. These have been hosted annually by Planet for two years and provide a growing venue for solid interaction across the optical remote sensing community. The Joint Agency Commercial Imagery Evaluation (JACIE) workshop has also been held annually for the past 20 years and provides an excellent venue for government, commercial, and university interaction. As calibration continues to depend on vicarious methodologies, a plan for regular round-robin field campaigns would help to not only improve communications, but also serve as a vehicle for standardization of measurements and algorithms.

A specific workshop recommendation that has strong potential for improving collaboration between the commercial and government sectors is based on the development of a Planet Dove cubesat that was built with the same spectral filters as Landsat 8. Because of identical RSRs, the cross-calibration of these sensors is greatly enhanced. Collaboration between Planet and Landsat calibration teams would provide a benefit to Planet through greater characterization of the Dove cubesat and to the Landsat team through greater understanding of the potential of commercial sensors. Ultimately, this collaboration could serve as a model for future characterization and collaboration opportunities between government and commercial systems.

All the panelists strongly agreed to recommend the adoption of a reference sensor for radiometric calibration. While the de facto standard for decades has been the Landsat series of sensors, today's environment is replete with many sensors that exhibit a high degree of calibration including Landsat, Sentinel 2, MODIS, and VIIRS. In addition, there is important work in progress worldwide for the development of essentially a spaceborne absolute radiometer with sub-1% radiometric accuracy [14]. Ideally, this sensor would be hyperspectral in nature and placed in an orbit such that both government and commercial systems could take advantage of cross-calibration through simultaneous nadir overpass (SNO) opportunities on a regular basis. Ultimately, a decision must be made as soon as possible so that a reference exists for the optical remote sensing community, as well as an agreed upon approach for updates as improvements in technology occur.

The final discussion area centered around the growing movement in the community for standardization of surface products for distribution as opposed to traditional TOA products. Standardization of these products is more difficult as the propagation of a TOA measurement back down to the surface incurs greater uncertainties due to atmospheric interference. It was noted that there are international efforts across government agencies to develop standard approaches for generating these products, and that the commercial sector is also working in the same direction. Perhaps the most comprehensive effort at standardization of surface measurements is the fiducial reference measurements for vegetation (FRM4VEG) that is being led by the European Space Agency. This project is open to both government and commercial entities and provides an excellent opportunity for greater collaboration as well as development of standardized procedures for generating surface-based products.

6. Recommendations

Five specific recommendations were developed from the workshop. All of them can be implemented immediately and can substantially improve the remote sensing community's ability to synthesize government and commercial data sets into improved products.

1. Standardize and broadly share models and data obtained from bright and dark PICS.

2. Improve communications between the government and commercial sectors through involvement in CEOS, key workshops such as JACIE, and focused calibration campaigns/meetings.
3. Collaborate on the characterization and calibration of the Planet Dove equipped with Landsat 8 spectral filters as a model for enhanced government/commercial sensor cross-calibration.
4. Agree upon a standard reference sensor with accepted traceability and spectral coverage now as well as plan for future modification as technology continues to improve.
5. Encourage government and commercial participation in the FRM4VEG project for a broader voice in development of standards for enhanced validation of vegetative products.

7. Conclusions

It is clear from the workshop that both the government and commercial sectors share many of the same concerns regarding calibration of their instruments with the overriding goal to provide value to the remote sensing user community. Methodologies to achieve these goals differ, but to a large degree, can complement each other to the benefit of both groups. Both types of sensor systems provide data with unique aspects and combining these data sets is to everyone's advantage. It is also clear that calibration teams from both communities can benefit from greater interaction.

As a result of the workshop, five specific recommendations were generated. These recommendations can be implemented easily, yet can provide a long-term basis for enhanced collaboration, higher quality data, and increased benefit to users. It is hoped that the output of this workshop will provide the beginnings of a roadmap that will lead to an improved understanding of instrument performance resulting in better calibration and ultimately improved data for users.

Author Contributions: Introduction, D.H.; Calibration Capabilities and Limitations of Large Government Systems, B.M., S.C., D.H., C.A.; Calibration Capabilities and Limitations of Commercial Systems, M.K., I.Z.; Independent Calibration and Applications Perspective, M.K., M.P.; Discussion, C.A., K.B., R.H., I.Z., V.B., S.C., M.K., B.M., M.P.; Recommendations, C.A., K.B., R.H., I.Z., V.B., S.C., M.K., B.M., M.P.; Conclusions, D.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

1. Barsi, J.A.; Lee, K.; Kvaran, G.; Markham, B.L.; Pedelty, J.A. The Spectral Response of the Landsat-8 Operational Land Imager. *Remote Sens.* **2014**, *6*, 10232–10251. [[CrossRef](#)]
2. Pahlevan, N.; Schott, J.R.; Franz, B.; Zibordi, G.; Markham, B.; Bailey, S.; Schaaf, C.; Ondrusek, M.; Greb, S.; Strait, C.M. Landsat 8 remote sensing reflectance (Rrs) products: Evaluations, intercomparisons, and enhancements. *Remote Sens. Environ.* **2017**, *190*, 289–301. [[CrossRef](#)]
3. McCorkel, J.; McAndrew, B.; Barsi, J.; Markham, B.; Pharr, J.; Rodriguez, M.; Shuman, T.; Sushkov, A.; Zukowski, B. First Results from Laser-Based Spectral Characterization of Landsat 9 Operational Land Imager-2. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 27 July–2 August 2019.
4. Barsi, J.; AlHammoud, B.; Czapla-Myers, J.; Gascon, F.; Haque, O.; Kaewmanee, M.; Leigh, L.; Markham, B. Sentinel-2A MSI and Landsat-8 OLI radiometric cross comparison over desert sites. *Eur. J. Remote Sens.* **2018**, *51*, 822–837. [[CrossRef](#)]
5. Bouvet, M.; Thome, K.; Berthelot, B.; Bialek, A.; Czapla-Myers, J.S.; Fox, N.P.; Goryl, P.; Henry, P.; Ma, L.; Marcq, S.; et al. RadCalNet: A Radiometric Calibration Network for Earth Observing Imagers Operating in the Visible to Shortwave Infrared Spectral Range. *Remote Sens.* **2019**, *11*, 2401. [[CrossRef](#)]
6. Dechoz, C.; Poulain, V.; Massera, S.; Languille, F.; Greslou, D.; De Lussy, F.; Gaudel, A.; L'Helguen, C.; Picard, C.; Tremas, T. Sentinel 2 global reference image. *Image Signal Process. Remote Sens.* **XXI** **2015**, 9643, 96430. [[CrossRef](#)]
7. Kulu, E. Nanosatellite & CubeSat Database. Available online: <https://www.nanosats.eu/> (accessed on 12 November 2019).

8. Ochoa, T.; Kuester, M. Empirical BRDF Model for Absolute Radiometric Calibration Targets. In Proceedings of the JACIE Workshop, College Park, MD, USA, 17–19 September 2018.
9. Clarreo Pathfinder. Available online: <https://clarreo/larc.nasa.gov> (accessed on 15 October 2019).
10. TRUTHS: A New Potential ESA Earth Watch Mission. Available online: https://www.esa.int/Applications/Observing_the_Earth/TRUTHS_a_new_potential_ESA_Earth_Watch_mission (accessed on 27 July 2020).
11. Barsi, J.A.; Schott, J.R.; Hook, S.J.; Raqueno, N.G.; Markham, B.; Radocinski, R.G. Landsat-8 Thermal Infrared Sensor (TIRS) Vicarious Radiometric Calibration. *Remote Sens.* **2014**, *6*, 11607–11626. [[CrossRef](#)]
12. Pagnutti, M.R.; Burch, K.R. Compact Thermal Imaging Calibrator (CTIC) for Landsat-like Missions. In Proceedings of the ASPRS Pecora 20, Sioux Falls, SD, USA, 14 November 2017.
13. Houborg, R.; McCabe, M.F. A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. *Remote Sens. Environ.* **2018**, *209*, 211–226. [[CrossRef](#)]
14. Hewison, T.; Fox, N.; Wielicki, B.; Kopp, G.; Hu, X. SI-Traceable Space-based Climate Observing System—A CEOS and GSICS Workshop Report. In Proceedings of the SI-Traceable Space-based Climate Observing System—A CEOS and GSICS Workshop, National Physics Laboratory, London, UK, 9–11 September 2019; In Progress.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).