



## Article

# Extended Cross-Calibration Analysis Using Data from the Landsat 8 and 9 Underfly Event

Garrison Gross, Dennis Helder and Larry Leigh \*

Image Processing Laboratory, South Dakota State University (SDSU), Brookings, SD 57007, USA

\* Correspondence: larry.leigh@sdstate.edu; Tel.: +1-605-688-5918

**Abstract:** The Landsat 8 and 9 Underfly Event occurred in November 2021, during which Landsat 9 flew beneath Landsat 8 in the final stages before settling in its final orbiting path. An analysis was performed on the images taken during this event, which resulted in a cross-calibration with uncertainties estimated to be less than 0.5%. This level of precision was due, in part, to the near-identical sensors aboard each instrument, as well as the underfly event itself, which allowed the sensors to take nearly the exact same image at nearly the exact same time. This initial calibration was applied before the end of the on-orbit initial verification (OIV) period; this meant the analysis was performed in less than a month. While it was an effective and efficient first look at the data, a longer-term analysis was deemed prudent to obtain the most accurate cross-calibration with the smallest uncertainties. The three forms of uncertainty established in the initial analysis, dubbed “Phase 1”, were geometric, spectral, and angular. This paper covers Phase 2 of the underfly analysis; several modifications were made to the Phase 1 process to improve the cross-calibration results, including a spectral correction in the form of a spectral band adjustment factor (SBAF) and a more robust filtering system that used the statistics of the reflectance data to better include important data compared to the more aggressive filters used in Phase 1. A proper uncertainty analysis was performed to more accurately quantify the uncertainty associated with the underfly cross-calibration. The results of Phase 2 showed that the Phase 1 analysis was within its 0.5% uncertainty estimation, and the cross-calibration gain values in this paper were used by USGS EROS to update the Landsat 9 calibration at the end of 2022.



**Citation:** Gross, G.; Helder, D.; Leigh, L. Extended Cross-Calibration Analysis Using Data from the Landsat 8 and 9 Underfly Event.

*Remote Sens.* **2023**, *15*, 1788.

<https://doi.org/10.3390/rs15071788>

/rs15071788

Academic Editor: David M. Johnson

Received: 14 February 2023

Revised: 20 March 2023

Accepted: 22 March 2023

Published: 27 March 2023



**Copyright:** © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** Operational Land Imager (OLI); Landsat; PICS; EPICS; cross-calibration; spectral band adjustment factor (SBAF); bidirectional reflectance distribution function (BRDF); underfly

## 1. Introduction

After launching on 27 September 2021, Landsat 9 was poised to perform an operation that would ultimately aid in its sub-1% cross-calibration: the Underfly Maneuver. During this maneuver, the spacecraft orbited beneath its virtual twin, Landsat 8, which had been launched over eight years prior on 11 February 2013. 90 days after launch, the two instruments started the week-long maneuver and collected some of the most important images needed for their cross-calibration. The results obtained from analyzing the underfly allowed Landsat 8 and 9 to have near identical calibrations, which would result in consistent measurements between the two instruments. This level of cross-calibration can be accomplished due to the nearly identical sensors aboard each satellite: Operational Land Imager (OLI) and Thermal Infrared Scanner (TIRS) on board Landsat 8 and OLI-2 and TIRS-2 on board Landsat 9 [1]. With an accurate cross-calibration, the two instruments can be considered part of a single system. It takes each satellite 16 days to individually capture images of the entire planet, but the calibration obtained from this maneuver would let Landsat 9 fly 8 days out of phase with Landsat 8. With the instruments being near clones of each other, they can image the planet at different times, combine their image libraries, and double their total temporal resolution.

The design similarities between these two instruments are what contributed most to the promising nature of an underfly cross-calibration analysis. Both versions of OLI have the same spatial resolution (30 m). Landsat 8 and 9 also have the same attitude controllers, onboard radiometric calibrators, and telescope, all of which are discussed more in [2]. The similarities between the focal plane modules and the spectral filters will be touched on more in the spectral uncertainty portion of the introduction. This paper is a second look at the underfly data following a preliminary analysis that was performed during the initial on-orbit verification (OIV) period. Following this introduction, the paper will examine how the new methodology was improved over the preliminary approach. The results generated from that procedure will be presented along with an uncertainty analysis and a discussion of the findings. Finally, the conclusions regarding how this paper expands on the original study will be presented.

### 1.1. Underfly Event

The Underfly event refers to the maneuver in which the new instrument, in this case Landsat 9, orbits underneath the reference sensor, Landsat 8, before moving into its final orbital path. This allows the sensors to see virtually the same image at approximately the same time and location, which reduces many forms of uncertainty. The same operation was performed 29–30 March 2013, after Landsat 8 launched, when it flew underneath Landsat 7 [3], and on 1–4 June 1999 when Landsat 7 orbited beneath Landsat 5 [4]. The data from previous underfly maneuvers granted excellent calibration results, so when the possibility arose to perform the same operation for Landsat 8 and 9, the teams at USGS EROS and NASA immediately took advantage of that opportunity. Since the sensors are virtually identical to each other, along with advances in modern computing to process more images than ever before, this maneuver was even more advantageous than previous ventures.

The Landsat 8 and 9 Underfly Maneuver was executed from 11 to 17 November 2021. As stated, an analysis approach using the data from the underfly was performed before the end of the OIV period [5]. The analysis performed on the underfly data during OIV will be referred to as Phase 1, while the analysis reported in this paper will be referred to as Phase 2. The initial results from the Phase 1 approach were primarily used in the initial calibration of Landsat 9. Since that analysis was performed in less than a month, however, this paper improves upon the methodology in several ways and takes a closer look at the sources of uncertainty associated with the underfly maneuver.

### 1.2. Sources of Uncertainty

A cross-calibration between two sensors can be simply summarized as finding the top of atmosphere (TOA) apparent reflectance of a location observed by each sensor and calculating the ratio of those reflectances. In this case, the ratio would be that of Landsat 8 over Landsat 9, the ratio needed to cross-calibrate Landsat 9 to Landsat 8. However, there are several forms of uncertainty related to a cross-calibration such as this. The initial cross-calibration analysis of Landsat 8 and 9 determined that the main three contributors were geometric, spectral, and angular uncertainty, with the latter in the form of the bidirectional reflectance distribution function (BRDF). The proposed uncertainty budget for the initial analysis was 1%, so individual uncertainty contributions needed to be well under that, with target differences of 0.2%. Differences any less than that would be considered insignificant. In the Phase 1 analysis, geometric uncertainty was split up into pointing error and viewing/illumination geometric error.

#### 1.2.1. Geometric Uncertainty

The first contributor to geometric uncertainty was pointing error. It was considered to be negligible based on discussions with geometry experts from USGS EROS, who stated “that image-to-image errors would be on the order of 5-m”, a difference in geometric positioning that is deemed to be minor enough to ignore [5]. The second form of geometric uncertainty considered was viewing and illumination geometry. Since the sensors were

close to each other during the maneuver, solar geometry was similar enough to be considered the same, while viewing geometry was the more significant issue. The sensors have a maximum view angle of  $\pm 7.5^\circ$ , so the worst possible viewing geometry with image overlap could be  $\pm 15^\circ$ . However, with the sun position consistently in the east and Landsat 9's descending path east of Landsat 8, L9 was tilted to follow the sun and view west in the final days of the underfly maneuver, which increased the difference in view zenith angles (VZA). This operation was performed to extend the number of coincident images but brought the maximum view zenith angle difference (VZAD) up to  $20^\circ$ . Overlap of the sensors varied throughout the week: Landsat 9's ascending path west of Landsat 8 had ~10% overlap on 12 November, ~50% on 13 November, and ~100% overlap on 14 November. The descending path overlap varied further as a result of this tilting maneuver. All these viewing geometries needed to be considered for the geometric uncertainty analysis. The previous analysis stated that the view azimuth geometry is ideally  $98^\circ$  and  $278^\circ$  due to Landsat's orbital position. Since the underfly took place in November, the sensors were closer to the same plane as the sun with respect to a ground target when imaging the northern hemisphere, while viewing a target closer to the plane orthogonal to the sun when in the southern hemisphere. The former example is denoted as the principal plane, and it refers to when the sensor and sun are azimuthally along the same plane with respect to the target. The plane orthogonal to this is considered the cross-principal plane. For this analysis, a cosine correction for the sun angle was not applied to the reflectance, as the main interest of this paper is the difference seen across the dynamic range of the sensors. This approach gave a more "raw" comparison between the two.

### 1.2.2. Spectral Uncertainty

The spectral differences between the sensors were also determined to be minimal in the Phase 1 analysis. OLI and OLI-2 have similar spectral filters, to the point where the filters for OLI-2 came from the original batch developed for OLI. These filters are used with focal plane modules (FPMs), which are modules comprised of photosensitive detectors. OLI and OLI-2 each have fourteen FPMs, which make up a focal plane array (FPA). Even though the respective spectral filters on each sensor cover the same wavelengths as the other, the spectral band passes, also known as Relative Spectral Responses (RSRs), do have small differences, which can result in slightly different reflectance measurements. This assumes view and solar angles are the same, so the only possible reflectance difference could be spectral. These differences can be quantified by looking at hyperspectral reflectance curves of several land cover types and comparing the banded reflectances of each sensor. If the results differ from each other, a spectral band adjustment factor (SBAF) will need to be applied. For more information on how to use SBAF, please look to [6,7].

Since SBAF is a target-dependent function, tests need to be performed with a wide variety of hyperspectral target types. Generally, a higher input signal with few variations in reflectance results in smaller spectral differences. During Phase 1, early estimates of spectral differences between the two sensors determined several land cover types for most of OLI and OLI-2's bands and had less than a 0.2% difference in banded reflectance, which was low enough to be considered negligible for this analysis. However, the green band differed between L8 and L9 by about 0.3% in vegetative land covers. The Phase 1 analysis determined that soil and sand land cover types would be best used for the green band due to its low spectral differences between OLI and OLI-2. TIRS and TIRS-2 (thermal bands 10 and 11) also have minor differences in RSR. The TIRS bands are calibrated solely based on radiance, with water as the primary target for calibration, so tests indicated that the average high emissivity of water made TIRS SBAFs unnecessary. Section 2.4 in the methodology will dive deeper into implementing SBAF into the cross-calibration analysis, and the results section will investigate the uncertainty associated from FPM to FPM between sensors.

The bands in OLI and TIRS with their spatial resolutions and wavelengths are shown in Table 1. For those unfamiliar with Landsat 8's new bands over Landsat 7, the coastal aerosol band is used for aerosol detection and measures ocean color along the coast, which

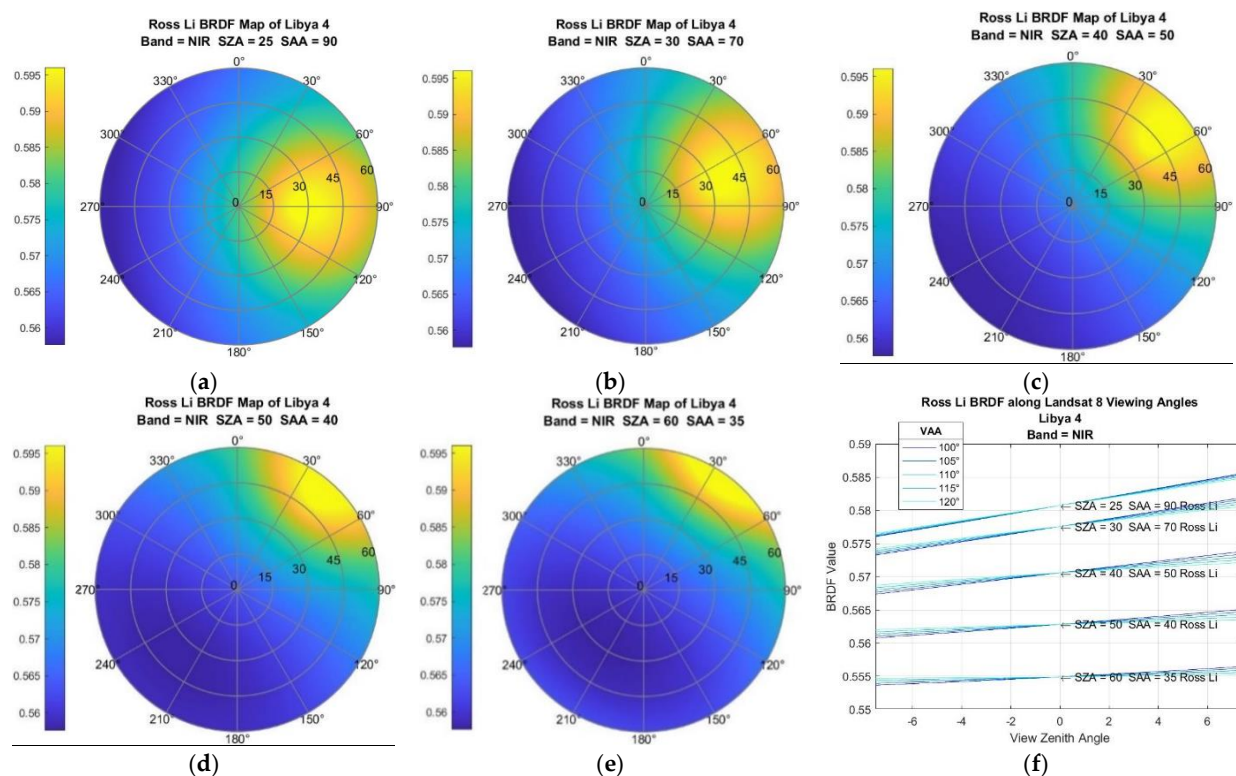
helps determine chlorophyll levels. The cirrus band helps detect clouds by looking at absorption features and measuring energy in the wavelengths where clouds are most discernable. Finally, the TIRS bands are used to measure ground temperature by looking at the emitted thermal radiation from Earth. OLI and OLI-2, as well as TIRS and TIRS-2, have the same bands and bandwidth, with their differences being in the spectral responses.

**Table 1.** Landsat 8 bands.

Bands	Wavelengths (Micrometers)	Resolution (Meters)
Band 1—Coastal Aerosol	0.43–0.45	30
Band 2—Blue	0.45–0.51	30
Band 3—Green	0.53–0.59	30
Band 4—Red	0.64–0.67	30
Band 5—Near Infrared (NIR)	0.85–0.88	30
Band 6—Short Wave Infrared 1 (SWIR1)	1.57–1.65	30
Band 7—SWIR2	2.11–2.29	30
Band 8—Panchromatic Band	0.50–0.68	15
Band 9—Cirrus	1.36–1.38	30
Band 10—Thermal Infrared 1 (TIRS 1)	10.6–11.19	100
Band 11—TIRS 2	11.5–12.51	100

### 1.2.3. Angular Uncertainty

The final type of uncertainty was angular, in the form of the BRDF effect. In Phase 1, it was thought that this was the largest contributor to uncertainty; however, the uncertainty analysis in Section 3.3 will discuss why that is not the case. The Ross–Li models of several different land cover types were analyzed using the BRDF MODIS product MCD43A1, which generates parameters to be used with the Ross–Li model. The use and generation of these BRDF kernels and parameters can be found in [8–10]. The main point derived from that analysis concluded that sensor viewing zenith angle would contribute most to reflectance differences between the instruments. This difference could be minimized if the sensors were closer to the cross-principal plane with respect to a ground target or if the sensors were directly on top of each other, the latter of which results in them having the same view zenith angle. The first suggestion indicated that if the sensors did not have the same azimuth (viewing) angle as the sun, reflectance differences would not be nearly as great due to avoiding the BRDF “hotspot” along the principal plane (Figure 1). The greater the difference in sun to view azimuth angle, or the closer to the cross-principal plane, the less effect the hotspot had, resulting in a smaller disparity in reflectance. The difference between solar illumination azimuth angle and a sensor’s viewing azimuth angle will be referred to as the view azimuth angle difference (VAAD), with each sensor having its own VAAD with respect to a single target.



**Figure 1.** (a–e): BRDF models of grass pasture for different SZA and SAA locations. Note how hot spot follows solar location as the sun moves lower in the sky. (f): Linearity of BRDF versus VZA for entire range of Landsat viewing angles. Note how the slope is greater as the sensor view angle moves closer to the “hot spot”. Image source: [5].

The second suggestion indicated that the closer the sensors were to each other with respect to VZA when looking at the same target, the smaller the difference in reflectance would be. This difference in view zenith angles will be referred to as the view zenith angle difference (VZAD). The initial cross-calibration analysis determined that plotting VZAD against the cross-calibration gain (the reflectance ratio of Landsat 8 to Landsat 9) and finding the intercept at  $VZAD = 0^\circ$  assumed the sensors were directly on top of each other. When the Underfly Event occurred, there were very little data at  $VZAD = 0^\circ$ , so a linear fit equation was used to estimate the intercept. This intercept essentially interpolated the value where the sensors had the same view zenith angle and became the basis for the cross-calibration gain value estimation. Section 3.1.2 will investigate the effect VZAD has on the VZAD intercept, and the results section will have an uncertainty analysis regarding the Ross–Li model of BRDF. Using the VZAD intercept value as the cross-calibration estimate was one of the strongest advantages when examining the initial Phase 1 underfly approach, as will be discussed more in depth in the methodology section; the other benefits from that analysis are discussed in the next section.

### 1.3. Phase 1 Retrospection

#### 1.3.1. Advantages

Several key observations and advantages resulted from the Phase 1 analysis: understanding the small effect of geometric and spectral uncertainty; the introduction of the VZAD intercept as the cross-calibration estimate; the relatively quick turnaround following the underfly event; and the precision and accuracy of the actual estimate. As stated in the previous subsection, the geometric differences were deemed to be negligible, so that form of uncertainty no longer needed to be considered. Spectral differences were heavily studied, so any uncertainty there could be controlled by choosing specific land cover types for certain bands if the need arose. This left angular differences as the primary contributing

factor to uncertainty. As mentioned before, the view zenith angle of each sensor was included with each downloaded image. This meant the reflectance and VZA of each scene could be plotted against each other. Inherently, this also meant the reflectance ratio of Landsat 8 to Landsat 9 could be plotted against the VZA difference, or VZAD, of each scene. By fitting a line to the data and finding the intercept when VZAD corresponded to 0 degrees, the cross-calibration ratio when the sensors were in line with each other could be estimated, effectively interpolating the result as accurately as possible and minimizing the uncertainty of BRDF with respect to VZA between the sensors.

The VZAD intercept also had other advantages when it was chosen for the cross-calibration estimate, such as its resistance to statistical outliers. The overall mean of the cross-calibration ratios was considered for the estimation; however, any outliers in the data could skew the estimate. Another reason the VZAD intercept was used was due to the lack of actual data when VZAD was  $0^\circ$ . As previously stated, the time when the sensors were perfectly in line with each other was brief, and unfortunately this time occurred mostly over the ocean, with only a few scenes collected over Australia. This meant that an estimate using the median ratio value could not be used, since the value when VZAD was  $0^\circ$  quite possibly may never have occurred. This left the VZAD intercept as the best candidate for the cross-calibration estimate.

The Phase 1 underfly analysis resulted in an uncertainty significantly lower than other cross-calibration methods, and the resulting cross-calibration at the end of OIV was based primarily on the estimates produced by that effort. This was also, in part, due to the quick turnaround of the analysis, all of which was completed before the end of OIV. Therefore, it was deemed prudent to perform a detailed analysis of the underfly data to validate, or even improve, the initial results.

### 1.3.2. Shortcomings

Phase 1 consisted of downloading the real-time images, processing them, and analyzing the statistics of each scene as a function of land cover type, band, and VZAD. Since all of this was performed in just a few short weeks following the actual underfly event, several shortcomings have been noted. The first of these has been stated, which was that of spectral correction. Even though extensive research was conducted regarding how each sensor's RSRs behaved with several land cover types, no actual SBAF was applied to the final result. As mentioned, the Landsat 9 green band had a 0.3% difference compared to Landsat 8 when looking at vegetation. In Phase 1, the sand/soil gains were weighted more heavily knowing this difference; however, a proper SBAF correction would provide a more accurate result. Spectral correction was not applied due to time constraints, but another reason why it was not prioritized in Phase 1 was due to uncertainty regarding how pixels were sorted into specific land cover types. Several optimizations needed to be applied to the categorization process to have a high enough confidence that each pixel was properly classified.

To determine the spectral profile of several land cover types, the MODIS Land Cover Type product MCD12Q1 was used as reference. This product classified every one of its pixels into different land cover types based on the International Geosphere-Biosphere Programme's (IGBP) system. Product MCD12Q1 was cross-referenced with the MODIS BRDF product MCD43A1 to determine BRDF properties of each land cover type, which was needed when performing the BRDF analysis. Product MCD43A1 was also used to determine the reflectance profiles of each IGBP land cover type in Phase 1, but Phase 2 instead uses MODIS TOA reflectance product MOD02HKM as it is more appropriate for a TOA cross-calibration. By better separating pixels into these land cover types, spectral correction would be a lot more accurate.

Another shortcoming of Phase 1 was that the VAAD was not considered to be a component of uncertainty, mostly since variation in VAAD did not change the VZAD intercept; this will be covered more in the methodology section. The reflectance of individual scenes and cover types was also not prominently covered in Phase 1, as the cross-calibration

ratio between the two sensors was the most important statistic. Nevertheless, an analysis regarding reflectance properties would help better understand the data. This analysis might include studying the signal to noise ratio and examining the reflectance differences in vegetation across the globe, considering that during the underfly, the northern hemisphere was in its winter season and the southern hemisphere was in its summer season. Finally, the panchromatic band was constrained to image pairs where resampling was not required, as resampling often introduced extra pixels that broke the image processing workflow. This was not changed in Phase 2 but should be noted as a small limitation of the analysis. By considering the shortcomings of Phase 1, the methodology of Phase 2 could be greatly improved.

## 2. Methodology

The methodology of Phase 1 was developed in the months leading up to the Underfly Event. Spectral, geometric, and BRDF analysis was performed to properly understand the uncertainty involved with the cross-calibration before even looking at the actual underfly data. The image processing workflow was also constructed in the months prior to the maneuver, while the actual processing occurred in just a few weeks during OIV. Several changes that could be used to optimize the procedure were discovered during both OIV and in the months after the event; however, these could not be implemented until a proper second analysis was performed. This section will look at the modifications made to the original procedure, starting with the optimized pixel sorting of land cover types.

### 2.1. Spectral Characterization of Land Cover Types

#### 2.1.1. MODIS Land Cover Product

As stated in the introduction, the MODIS products were used to help characterize pixels into different IGBP land cover types. MODIS annually characterizes each of its global pixels into land cover groups, which can be found in the IGBP product MCD12Q1. By using this product in conjunction with another MODIS product, average properties of each IGBP type can be calculated. The BRDF product MCD43A1 was used for a BRDF analysis of each IGBP land cover type, as seen in Table 2 [11]. This product consists of Ross–Li parameters that can be used to generate BRDF models for each MODIS band; as such, it was of great use in the BRDF analysis during the months leading up to the underfly event. Because the product was familiar to the authors, it was also used to help determine the average spectral profile of each IGBP land cover by finding the nadir reflectance of the BRDF models. However, it was noted after Phase 1 that since this cross-calibration would be using TOA reflectance values, another product should be used to determine spectral profiles for the land cover types. The TOA reflectance product MOD02HKM was found to be more suitable for this analysis. Data from November using this product were utilized to find the average TOA reflectance for all the IGBP land cover types, since the underfly took place in November 2021. These spectral profiles were needed since a direct pixel comparison is difficult between Landsat and MODIS. This is because MODIS has a much larger spatial resolution than Landsat, with 250 m resolution for bands 1 and 2 and 500 m resolution for bands 3 through 7.

Because the reflectance values found were from MODIS, an SBAF was needed to spectrally correct the values to what Landsat 8 would measure. SCanning Imaging Absorption spectroMeter for Atmospheric CHartography (SCIAMACHY) has TOA hyperspectral profiles which were used to spectrally convert the MODIS TOA IGBP reflectance values to Landsat 8, which can be used to sort Landsat pixels more accurately [12]. Since the MODIS to Landsat 8 spectral correction was not performed in Phase 1, the resulting pixel sorting process should increase in accuracy from this step alone, but this was only the first modification to the process. The Landsat 8 average banded TOA reflectance values for each IGBP type are shown in Table 3.

**Table 2.** IGBP land cover types.

Name	Value	Description
Evergreen Needleleaf Forests	1	Dominated by evergreen conifer trees (canopy > 2 m). Tree cover > 60%.
Evergreen Broadleaf Forests	2	Dominated by evergreen broadleaf and palmate trees (canopy > 2 m). Tree cover > 60%.
Deciduous Needleleaf Forests	3	Dominated by deciduous needleleaf (larch) trees (canopy > 2 m). Tree cover > 60%.
Deciduous Broadleaf Forests	4	Dominated by deciduous broadleaf trees (canopy > 2 m). Tree cover > 60%.
Mixed Forests	5	Dominated by neither deciduous nor evergreen (40–60% of each) tree type (canopy > 2 m). Tree cover > 60%.
Closed Shrublands	6	Dominated by woody perennials (1–2 m height) > 60% cover.
Open Shrublands	7	Dominated by woody perennials (1–2 m height) 10–60% cover.
Woody Savannas	8	Tree cover 30–60% (canopy > 2 m).
Savannas	9	Tree cover 10–30% (canopy > 2 m).
Grasslands	10	Dominated by herbaceous annuals (<2 m).
Permanent Wetlands	11	Permanently inundated lands with 30–60% water cover and >10% vegetated cover.
Croplands	12	At least 60% of area is cultivated cropland.
Urban and Built-up Lands	13	At least 30% impervious surface area including building materials, asphalt, and vehicles.
Cropland/Natural Vegetation Mosaics	14	Mosaics of small-scale cultivation 40–60% with natural tree, shrub, or herbaceous vegetation.
Permanent Snow and Ice	15	At least 60% of area is covered by snow and ice for at least 10 months of the year.
Barren	16	At least 60% of area is non-vegetated barren (sand, rock, soil) areas with less than 10% vegetation.
Water Bodies	17	At least 60% of area is covered by permanent water bodies. Has not received a map label because of missing inputs.

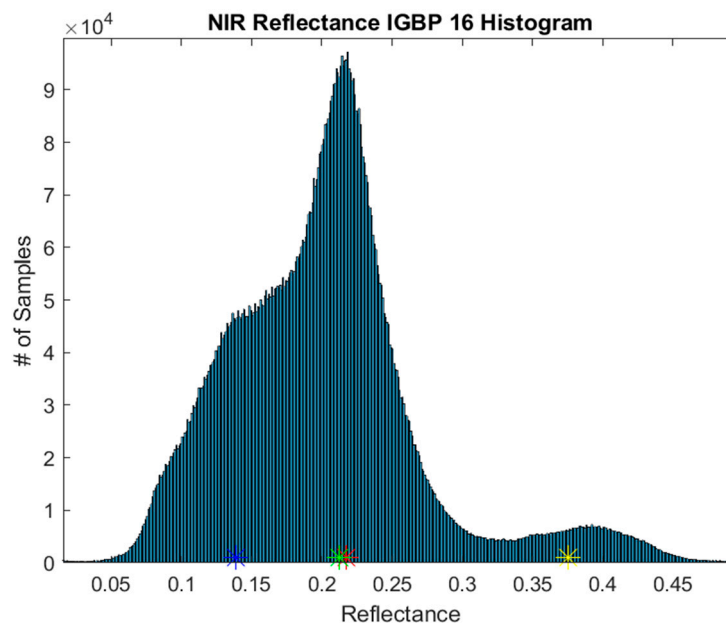
**Table 3.** TOA reflectance profiles for IGBP types.

IGBP	1	2	3	4	5	6	7	8	9	10	12	14
Blue	0.103	0.147	0.072	0.129	0.120	0.097	0.127	0.121	0.129	0.120	0.131	0.126
Green	0.076	0.151	0.053	0.109	0.113	0.086	0.128	0.114	0.129	0.111	0.111	0.131
Red	0.058	0.181	0.048	0.103	0.121	0.107	0.185	0.124	0.147	0.119	0.101	0.161
NIR	0.179	0.291	0.075	0.209	0.228	0.195	0.285	0.231	0.242	0.201	0.198	0.264
SWIR1	0.116	0.339	0.037	0.170	0.216	0.261	0.372	0.233	0.267	0.221	0.167	0.278
SWIR2	0.055	0.262	0.021	0.117	0.148	0.195	0.306	0.164	0.199	0.157	0.099	0.220

In Phase 1, the Barren IGBP type was split into three separate groups. Since it was noted that the IGBP system favors vegetative types, the Barren type was originally left as a broad class that encapsulates sand, soil, minerals, rocks, and mountains. By separating the Barren class into succinct types, a more accurate sorting procedure could be produced. In Phase 1, the groups were split by looking at the spectral histograms of the Barren class, specifically the NIR band since it has little atmospheric interference. There were three peaks in the histogram, so it was determined that the Barren class could be split up into three groups. The peaks in the histogram were continent dependent, so the three groups could be easily sorted into three groups known as Dark Soil, Light Soil, and Sand. In Phase 2, a more statistical approach was taken, and the Barren class was separated using



the k-means algorithm. Using the TOA MODIS product, nominally three or four different peaks in the histogram were observed, as seen in Figure 2. The k-means algorithm was trained to four groups, which gave the four distinct Barren reflectance profiles, as seen in Table 4. Now all that was needed was a way to properly identify Landsat pixels as each of these land cover types during the underfly.



**Figure 2.** Global NIR reflectance distribution of barren land cover. Note that the colored stars are the reflectance values for each group calculated by the k-means algorithm.

**Table 4.** Separated barren reflectance profiles.

IGBP	Barren1	Barren2	Barren3	Barren4
Blue	0.097	0.098	0.169	0.137
Green	0.095	0.084	0.201	0.124
Red	0.105	0.067	0.276	0.132
NIR	0.139	0.213	0.375	0.217
SWIR1	0.178	0.142	0.457	0.264
SWIR2	0.153	0.071	0.398	0.201

### 2.1.2. Global Classification of Pixels during Underfly

Similar to Phase 1, the Extended Pseudo Invariant Calibration Site (EPICS) project at South Dakota State University was used to sort Landsat pixels based on the average IGBP profiles. “EPICS was developed to identify every potential Pseudo Invariant Calibration Site (PICS) on the planet” [5]. Its high levels of accuracy and stability are explored further in [13–15]. The algorithm works by looking at every pixel ever captured by Landsat 8. The EPICS project analyzed all the Landsat 8 data in the underfly month of November and sorted all the pixels into 500 classes using an unsupervised k-mean algorithm; this was the same approach the Barren cover type separation took. The k-means algorithm gave each of the 500 groups an average reflectance and standard deviation for each band. These, like Phase 1, were used to classify the 500 clusters as vegetation or soil using the normalized difference vegetation index (NDVI) and the bare soil index (BSI) [16,17]. These two Equations (1) and (2), are ratios that use the banded reflectance values of each cluster to determine if they are soil or vegetation. The same minimum values as Phase 1 were used, with an NDVI threshold of 0.2 and a BSI threshold of 0.021, indicating the smallest values considered to be vegetation or soil, respectively. The total number of clusters considered

either of vegetation or soil was 358, which, as a subset, would reduce processing time compared to all 500. The NDVI formula is shown as

$$\text{NDVI} = \frac{R_{\text{NIR}} - R_{\text{Red}}}{R_{\text{NIR}} + R_{\text{Red}}} \quad (1)$$

where R is the reflectance of the red and NIR spectral bands for Landsat 8. This index focuses on the reflectance change between the red and NIR bands to determine if the pixel is vegetation. The BSI equation is

$$\text{BSI} = \frac{(R_{\text{SWIR2}} + R_{\text{Red}}) - (R_{\text{NIR}} + R_{\text{Blue}})}{(R_{\text{SWIR2}} + R_{\text{Red}}) + (R_{\text{NIR}} + R_{\text{Blue}})} \quad (2)$$

where R is the reflectance of the blue, red, NIR, and SWIR2 bands of Landsat 8.

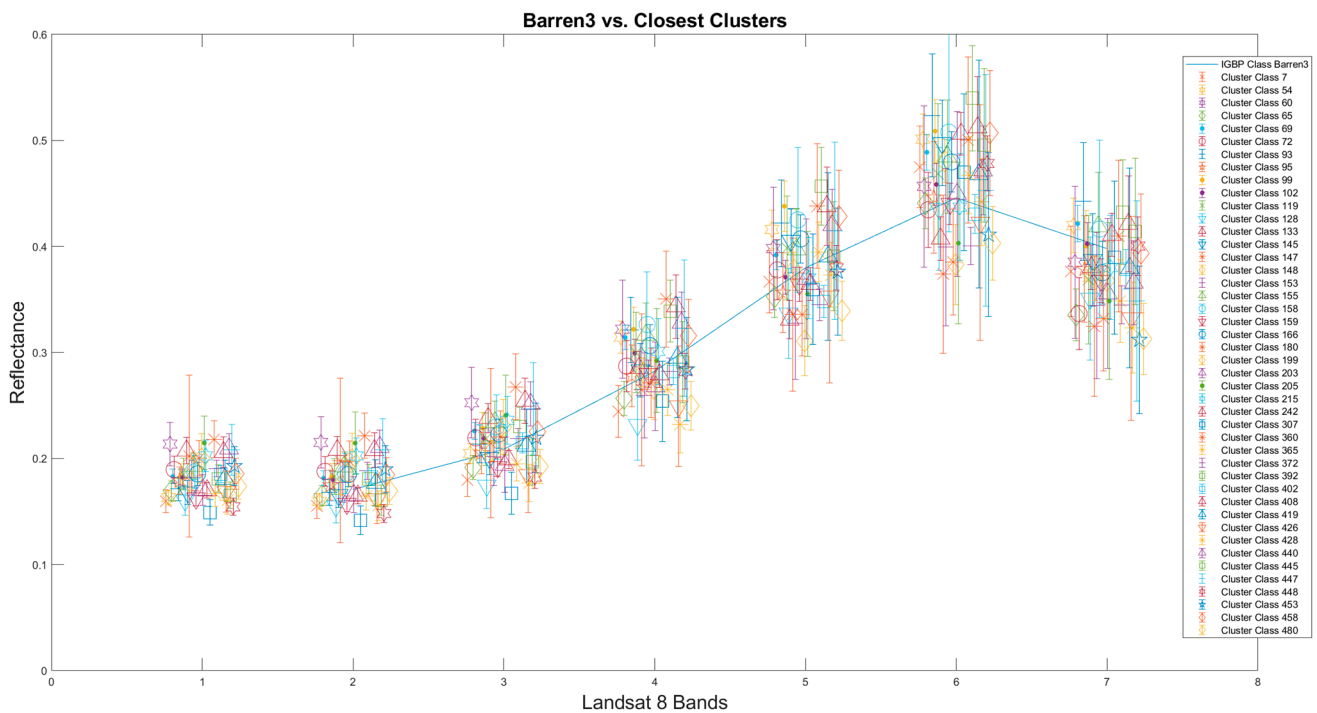
To determine which of the 358 clusters resembled each of the IGBP spectral profiles the best, some analysis needed to be conducted. In Phase 1, the simple average banded absolute difference was used to sort the clusters into IGBP types, seen in Equation (3). All 358 clusters were sorted into a land cover type, regardless of whether it was a good fit or not; however, Phase 2 took a more dynamic approach. The averaging expression in Phase 1 was

$$\frac{1}{\text{Band}} \sum_{\text{Band}} |\text{IGBP}_{\text{Band}} - \text{Cluster}_{\text{Band}}| \quad (3)$$

where IGBP is the banded reflectance of the IGBP type in a particular band and Cluster is the reflectance of an EPICS cluster in a particular band. This summation was then divided by the number of bands. This approach used a new difference formula, seen in Equation (4). This equation is a multi-dimensional distance formula that takes all six bands from the IGBP profiles (Blue, Green, Red, NIR, SWIR1, and SWIR2) and finds the difference with its corresponding cluster banded profile, which gives a 6-dimensional distance. MODIS does not have a CA band equivalent, so it was not used in this comparison. The IGBP land cover that returned the smallest 6-dimensional distance was the classification that the cluster was identified as. Because Equation (4) squares the differences rather than averaging them as in Equation (3), it is more sensitive to changes, which made it a more ideal candidate for classifying clusters. This new expression is shown as

$$\sqrt{\sum_{\text{Band}} (\text{IGBP}_{\text{Band}} - \text{Cluster}_{\text{Band}})^2} \quad (4)$$

where IGBP and Cluster are the banded reflectances of their respective source. To alleviate the concern whether a cluster is a good match for a specific IGBP type, the standard deviation associated with each cluster band was employed. If an IGBP banded value was outside three standard deviations (sigma) away from its cluster banded value, that cluster was removed. Three sigma was used since it statistically covers about 99.7% of the data, and any clusters outside that range should not be considered as that classification. With this new dynamic sorting process, the total number of clusters retained was reduced to 209. This left only the clusters that matched up within three sigma with a land cover type. The clusters filtered out of the process were low-confidence clusters, which can be defended by looking at the pixel counts of each group: there were over 93 billion pixels associated with all 358 clusters, and over 66 billion pixels in with the filtered 209 clusters. This left 58% of the total vegetation and soil clusters, but about 70% of the overall pixels. By filtering out the “edge case” clusters, higher confidence in sorting can be attained. In Phase 1, only a subset of the IGBP types had cluster matches; in Phase 2, all the IGBP types had at least one matching cluster, except Barren4, which alone was removed. For an example that presents the effectiveness of the classification process, Figure 3 below shows the clusters sorted into Barren3.



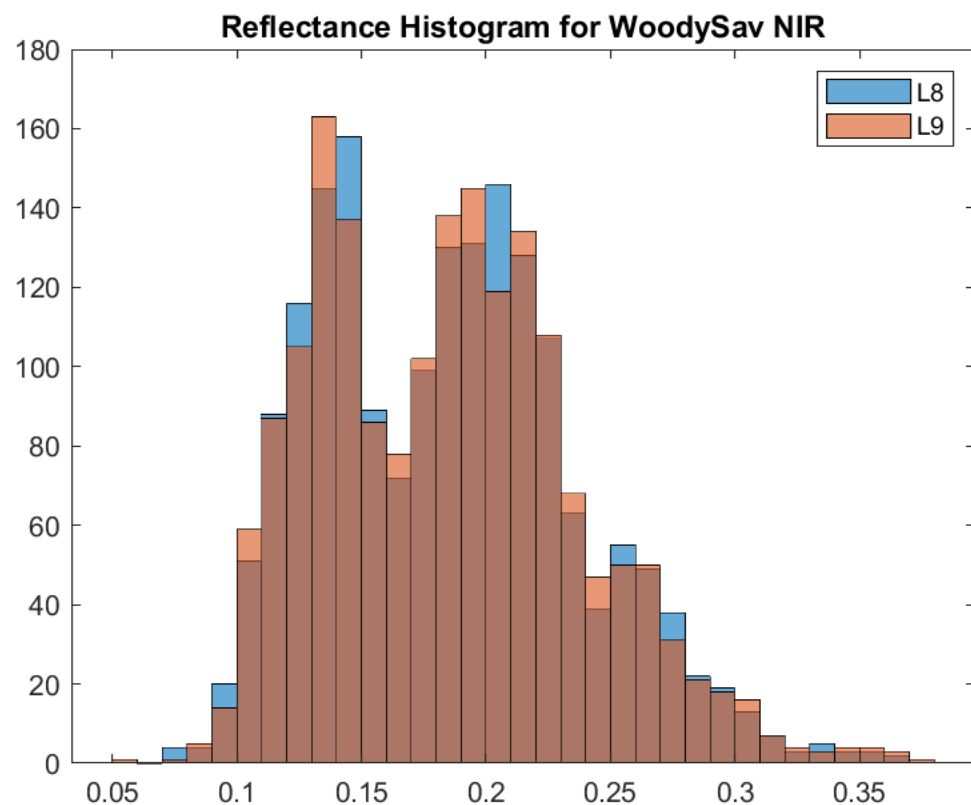
**Figure 3.** Barren3 reflectance profile plotted against matching clusters. Note the clusters grouping around the banded IGBP profile. The CA band does not have an IGBP value since MODIS does not have a CA band equivalent.

With the clusters sorted into specific IGBP land cover types, the same image processing workflow from Phase 1 was implemented [5]. To summarize the procedure, overlapping regions in corresponding Landsat 8 and 9 images were identified. These overlapping zones were split up by VZAD into  $0.25^\circ$  slices, so a particular scene could have up to six VZAD observations. VZAD itself is a continuous value but, for the underfly analysis, has been binned up into  $0.25^\circ$  bins. The digital numbers (DNs) of each scene were converted to reflectance and radiance using physical unit conversion formulas found in the Landsat 8 Data Users Handbook [18]. The statistics of all the pixels in a specific IGBP type in an individual VZAD slice were calculated. These statistics were minimum, maximum, mean, and standard deviation of each instrument's TOA reflectance, as well as the same statistics for the reflectance ratio of Landsat 8 to Landsat 9, which was performed for every band. Just like Phase 1, band 8 was processed using only scene pairs that did not require reprojection. Its higher spatial resolution would require resampling the images for reprojection, but the resampling process often introduced extra pixel artifacts. This limited VZAD in these scenes to within  $\pm 4^\circ$ . Band 9 was similarly removed from analysis due to its high atmospheric absorption which made ground pixel comparison impossible. TIRS bands 10 and 11 were cross calibrated in radiance space only. These bands used water for calibration; however, there were no pure ocean scenes collected, as the only water pixels available were those near land, such as lakes, ocean coasts, and rivers. Other than these deviations, the methodology for bands 8, 10, and 11 is identical to the other bands. Instead of the Real-Time images from Landsat 9 Collection 2, the data used were the Tier 1 images updated with the values from Phase 1 in order to find any differences between Landsat 8 and Landsat 9 after the initial OIV cross-calibration. Landsat Collection 2 is the second significant reprocessing endeavor of the Landsat archive, which improved the data products found in Collection 1. Once all the global data were completely processed, it was analyzed.

## 2.2. Data Analysis

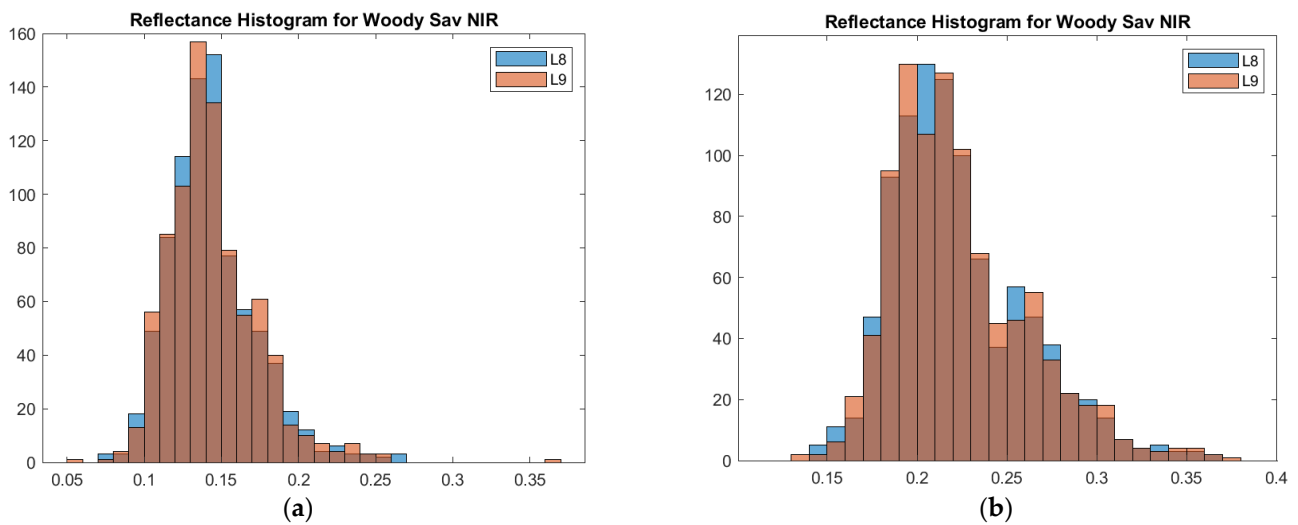
### 2.2.1. Seasonality

As mentioned in the introduction, reflectance investigation was a larger part of Phase 2. Reflectance differences in vegetation, especially their seasonality, was a factor that helped improve the accuracy of the underfly cross-calibration analysis. With vegetation, the NIR band is crucial in determining whether the pixel represents healthy vegetation. Healthy vegetation tends to have a high signal in the NIR wavelengths, while the signal is a lot lower when the vegetation is senescent. Since the NIR band is not greatly influenced by atmospheric effects, it was also a great candidate for TOA reflectance inspection. When investigating the global pixels for vegetation IGBP land cover types, a clear bimodal distribution was observed in the NIR band histogram (Figure 4).



**Figure 4.** Woody Savanna IGBP global NIR band reflectance distribution. Note the two peaks in the distribution.

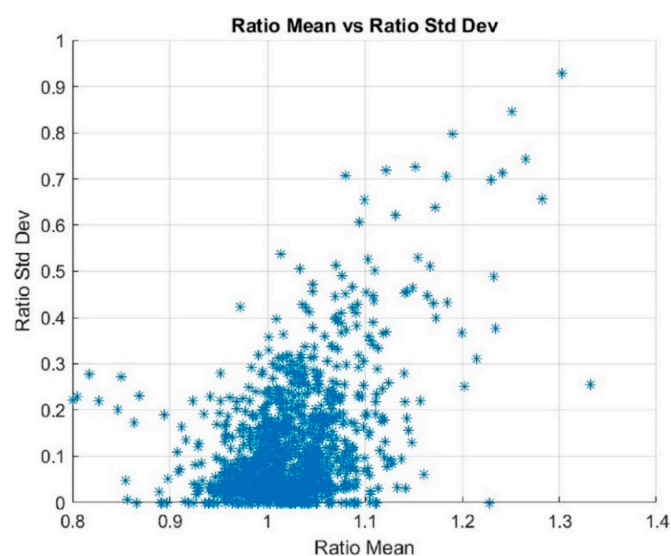
By separating the pixels by latitude, the source of this bimodality was determined. When the scenes were split at the Tropic of Cancer ( $23.5^{\circ}\text{N}$ ), the two peaks in the histogram were separated, giving two similar distributions (Figure 5). This is likely due to the seasonality of vegetation during the underfly: the northern hemisphere was in its winter during November while the southern hemisphere was in its summer during that time. The vegetation in the north was reaching the end of its summer cycle, so it was either dying or dead. This is represented in the histogram where the reflectance in the NIR band is lower in the northern latitudes, since the chlorophyll-like shape of vegetation's spectral curve is lowered dramatically in the NIR wavelengths when it is dying. Since the tropical climate between the Tropic of Cancer and Capricorn allows vegetation to stay green all year round, the bimodal split at the Tropic of Cancer makes physical sense. This discovery was used to filter data in the cross-calibration analysis, which will be discussed in the next section.



**Figure 5.** (a) Woody Savanna global NIR band reflectance distribution. This plot corresponds to reflectances above the Tropic of Cancer latitude. (b) This plot corresponds to reflectances below Tropic of Cancer. Note the peaks split up after separating by latitude.

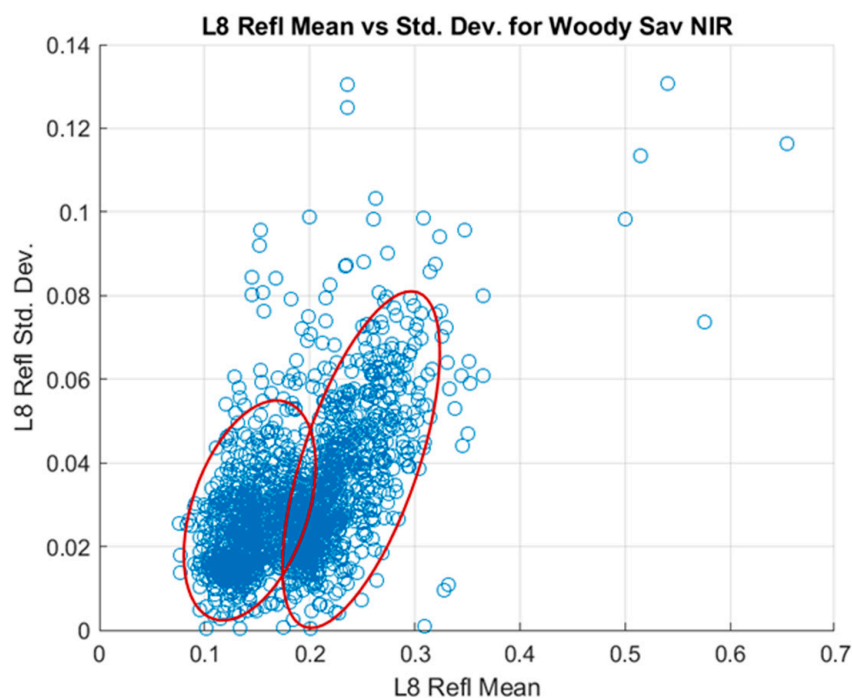
### 2.2.2. Signal to Noise Relationship

During Phase 1, several filters were applied to the underlying data to remove outliers and produce a concise cross-calibration gain. These filters included a pixel threshold, which filtered out scenes with fewer than 10,000 pixels of a specific IGBP type; a signal floor, which removed scenes with a reflectance less than 0.01; a VZAD filter, which removed scenes where the sensors were so far apart that the BRDF effect became observably nonlinear (this linearity over a short range will be described more in the following sections); and a ratio standard deviation filter, which filtered out scenes with a ratio standard deviation greater than 0.2, as a linear trend was observed which was thought to bias the data (Figure 6). All these filters in conjunction gave a cross-calibration result with a small overall standard deviation, but they also removed a lot of scenes with usable data, which could have biased the final results. The approach in Phase 2 was to use as much data as possible, so a single filter was designed to take the place of all the previous ones.



**Figure 6.** Ratio mean vs. ratio standard deviation. Note the discernable positive linear trend above a ratio standard deviation of 0.2 as the data drift to larger values. Since the ratio is derived by taking the reflectance of Landsat 8 to 9, this behavior could be a real and observable phenomenon. Image source: [5].

The signal-to-noise ratio (SNR) of the data was an aspect that was not covered in Phase 1, so a greater focus was brought to it in Phase 2. By better understanding this relationship, a well-rounded filter could be developed. The average reflectance of an observation (which in this case was a scene binned by  $0.25^\circ$  VZAD) was plotted against its standard deviation, giving a signal to noise relationship plot (Figure 7). In vegetative land covers, especially in the NIR band, two clusters can be visually observed (red ellipses in the figure). This behavior comes from the data's seasonality, as mentioned previously. Outliers can also be observed in this plot, which could be filtered out by taking a statistical look at these clustering data.



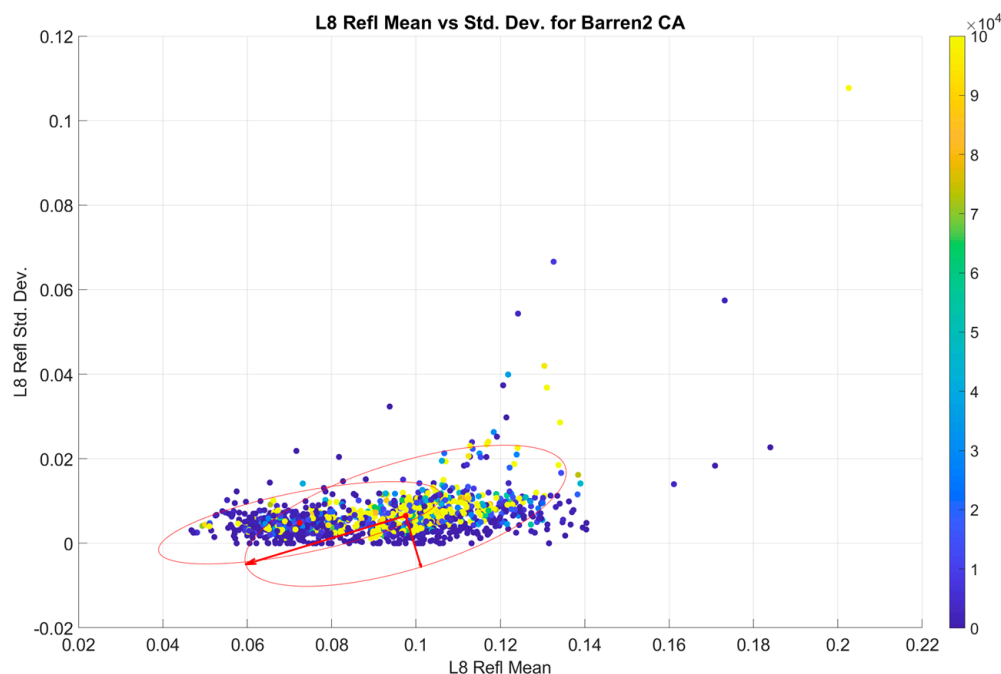
**Figure 7.** Signal to noise relationship of Woody Savanna NIR band. Note the two clusters of data points, which can be discerned by latitude.

### 2.2.3. Data-Driven Filtering

By plotting the data as a heat map, where hot spots indicate scenes with higher pixel counts, it can be seen that scenes with low pixels counts make up most of the outliers. Taking this into account, an ellipse filter was developed to encompass the clusters and remove outliers. These ellipses have a centroid weighted by  $n$ , the number of pixels in the scene. All weighted calculations are with respect to  $n$ . The seasonality of the data was considered by assuming there are two overlapping data distributions: one in the northern hemisphere and one in the south. The ellipse itself is constructed by taking the weighted covariance of the data, Equation (5), where  $w$  is the weight of the observation,  $\mu$  is the weighted centroid,  $x$  is the reflectance mean, and  $y$  is the reflectance standard deviation. Eigenvectors and eigenvalues can then be calculated from that. The eigenvalues determine the radius of the ellipse in the  $x$  and  $y$  directions, as well as the "tilt" of the ellipse if there is one. The eigenvalues scaled the size of the ellipse, along with sigma. It was decided to use three sigma, which encompassed about 99.7% of the weighted data. Anything outside of the ellipse was filtered out. There were several scenes filtered out by this algorithm that had fairly high pixel counts (over 100,000 pixels). Individual scene investigation determined these scenes were too cloudy to be useful for this analysis, so they were manually filtered out to reduce their influence in the ellipse filters. Figure 8 shows an example for the Barren2 land cover type. This helped the algorithm to incorporate more scenes overall. The weighted covariance equation is calculated by

$$\sigma^2 = \frac{1}{1 - \sum_{i=1}^N w_i^2} \sum_{i=1}^N w_i^2 (x_i - \mu_x)(y_i - \mu_y) \quad (5)$$

where  $w$  is the weight of the data point,  $x_i$  and  $y_i$  are the location of the data point, and  $\mu_x$  and  $\mu_y$  are the location of the weighted centroid.



**Figure 8.** Weighted ellipse filter for Barren2 CA band, with the eigenvectors dictating the shape of the ellipse. Note that in this example, the high pixel counts of the outliers “pull” the slope of the ellipse up. The points are color coded by pixel count, with the scale on the right of the plot. The arrows indicate the eigenvectors that create the boundaries of the ellipse.

The ellipse filter has several advantages over the filters applied in Phase 1. It makes several of the previous filters essentially obsolete, especially the pixel threshold. Since the ellipse centroid and the eigenvectors were weighted by  $n$ , filtering by a minimum number of pixels was no longer required. The same conclusion can be reached regarding the noise floor, as the ellipse filter is driven by the signal to noise relationship of the data. The scene ratio standard deviation filter was dropped for Phase 2. It was thought that the new pixel sorting process would remove the linear trend in this relationship; however, the trend was still present in Phase 2. Because of this, it was determined that the behavior might be a legitimate phenomenon, so filtering it out might bias the data. Finally, the most important advantage of the ellipse filter approach was that it incorporated more scenes overall as opposed to the filters used in Phase 1. By choosing to keep three sigma-worth of the data, about 99.7% of the data were kept. However, the ellipse filter did not remove the effects of BRDF.

### 2.3. BRDF Observations

As stated in the uncertainty section, the BRDF effect was the driving source of uncertainty in both Phase 1 and Phase 2 analyses. The biggest factor of BRDF was alleviated in Phase 1 with the introduction of the VZAD intercept estimator. By plotting the ratio between Landsat 8 and 9 against their VZA differences, a linear trend could be observed, at least over a restricted scale. In Phase 1, the VZAD was constrained to  $\pm 10^\circ$ , because over that short of a range, the behavior in the data was less scattered, resulting in a more noticeably linear trend. This response made physical sense when compared to the Ross–Li models analyzed during Phase 1, which had a similar linear behavior over such a short range. Since a linear equation was later fit to the data to determine the VZAD intercept

at  $0^\circ$ , this constraint could lead to a higher  $R^2$  value, or coefficient of determination, for the linear model, and a more accurate VZAD intercept. The same constraint was applied to Phase 2; as such, that was the second filter applied in this analysis. An improvement brought to Phase 2, however, was to weigh the VZAD linear fit equation by  $n$  to maintain consistency with the other parts of the methodology. This, in turn, led to a more consistent intercept value.

The other factor of BRDF examined was the VAAD. As formerly mentioned, the VAAD was not a driving factor of analysis in Phase 1, but an investigation in Phase 2 was conducted to determine if it had an effect on the cross-calibration gain. As expressed in the introduction, the Ross–Li BRDF model shows that if the sensors are viewing targets closer to the cross-principal plane, their differences are much smaller than if they were in the “hot spot” of the principal plane. If the VZAD intercept drastically changed as VAAD changed, then the data would need to be constrained close to the cross-principal plane. For the investigation, the VAAD was constrained to  $20^\circ$  “slices” in increments of  $10^\circ$ . These parameters were chosen since the sensors were usually within  $10^\circ$  VAA of each other. For this analysis, the BRDF model was assumed to be symmetrical on either side of the principal plane, reducing the scale from  $360^\circ$  down to  $180^\circ$ . Since the point of the analysis was to determine how close the sensors were to the principal plane, the scale was reduced further to  $90^\circ$ , since any VAADs over  $90^\circ$  would be closer to the principal plane again. This analysis was to be tested regardless of VZA, so if the worst-case range of  $0\text{--}90^\circ$  VAAD had no effect on the VZAD intercept, then the best case range from  $90\text{--}180^\circ$  would have the same if not more consistent behavior. The results section will present how much of an effect the VAAD of each sensor has on the final results.

#### 2.4. SBAF Correction

The introduction covered the small differences between the spectral filters of OLI and OLI-2, but the Phase 1 analysis never actually corrected for them. The green band was the driving factor for spectral discrepancies, especially in vegetative cover types. Because of this, soil cover types were weighted more heavily in the green band in the Phase 1 analysis. However, measurement differences in the green band seemed to show up in the months following the calibration. Users of the Landsat data found small discrepancies in this band, which were thought to be calibration errors. It was unknown whether the inconsistencies were target-dependent, so a spectral calibration was deemed necessary to accurately see what was happening in this band.

As mentioned in the introduction, spectral analysis of various land cover types was performed in the months leading up to the underfly. This was accomplished by taking spectroradiometer-measured, surface-level hyperspectral signatures and banding them using the OLI and OLI-2 RSRs, shown in Equation (6), calculated as

$$\rho_8 = \frac{\int_{-\infty}^{\infty} \rho(\lambda) \text{RSR}_8(\lambda) d\lambda}{\int_{-\infty}^{\infty} \text{RSR}_8(\lambda) d\lambda} \quad (6)$$

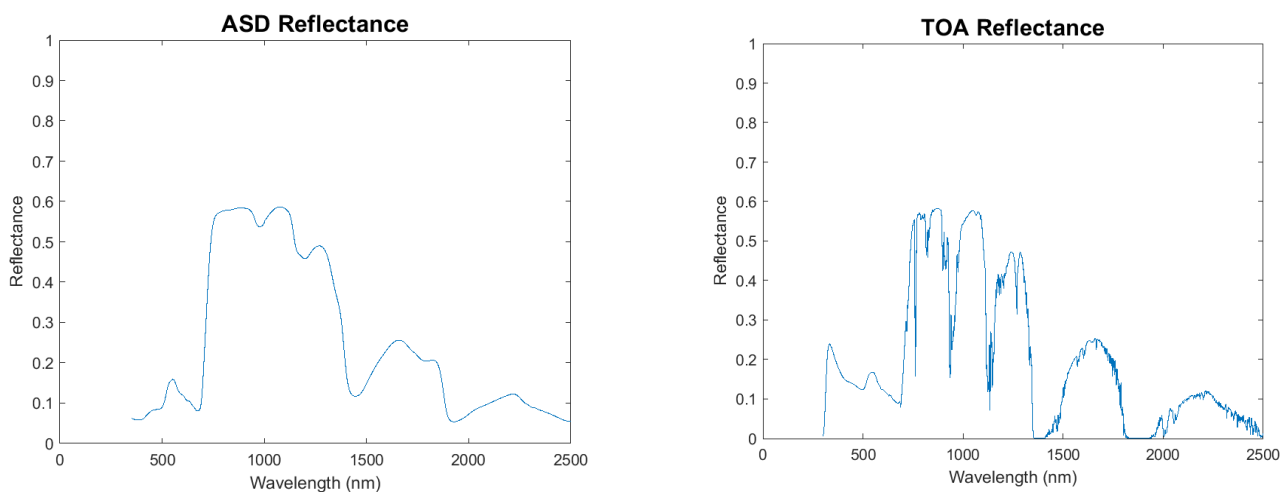
where  $\rho_8$  is the banded reflectance of Landsat 8,  $\rho(\lambda)$  is the hyperspectral reflectance with respect to wavelength, and  $\text{RSR}_8(\lambda)$  is the spectral response of Landsat 8 with respect to wavelength. These surface-level spectra were compiled into a comprehensive library, which included several spectral databases: cropland data came from the GHISA library [19], forest materials from the ECOSTRESS database [20,21], sand and soil information was found in the ICRAF-ISRIC Library [22], and snow data were comprised from the National Research Council of Italy Institute of Polar Sciences [23]. Most of these surface level signatures were measure using an Analytical Spectral Devices (ASD) spectroradiometer. However, the underfly cross-calibration is inherently a top of atmosphere (TOA) analysis, so an atmospheric correction would need to be applied to the spectra to accurately convey what they look like from space. The MODerate resolution atmospheric TRANsmission (MODTRAN) computer code is an algorithm that is often used to “remove” atmosphere from data to simulate what the data look like on a surface level. However, it can also “add”



atmospheric properties to hyperspectral surface data to determine how it appears from space, shown in Figure 9. For this analysis, a 1976 atmospheric model was applied to the surface-level spectra in MODTRAN, due to its globally generic properties [24]. By using the same ASD-measured hyperspectral libraries as used in the Phase 1 spectral uncertainty analysis, SBAF correction on a TOA level was produced:

$$\text{SBAF} = \frac{\int_{-\infty}^{\infty} \rho(\lambda) \text{RSR}_8(\lambda) d\lambda}{\int_{-\infty}^{\infty} \text{RSR}_8(\lambda) d\lambda} \frac{\int_{-\infty}^{\infty} \rho(\lambda) \text{RSR}_9(\lambda) d\lambda}{\int_{-\infty}^{\infty} \text{RSR}_9(\lambda) d\lambda} \quad (7)$$

where SBAF is the ratio of banded reflectance for Landsat 8 to Landsat 9.



**Figure 9.** Surface level reflectance of a white fir tree vs. MODTRAN-simulated TOA reflectance. Note the water absorption in the longer wavelengths and the scattering due to aerosols in the shorter wavelengths.

The Phase 2 analysis introduced several enhancements over Phase 1: a more optimized pixel sorting of land cover types; a deeper look at reflectance behavior including seasonality and signal to noise ratio; a more robust BRDF analysis; and finally, a proper SBAF correction. Based on the improvements implemented for data processing as described in this section, a more accurate result was anticipated.

### 3. Results

After evaluating the methodology of Phase 1 and finding several ways to improve it, the actual calculation to determine the cross-calibration ratio gains remained relatively untouched. That is, the same estimator was used as Phase 1: the VZAD intercept at  $0^\circ$ . The only adjustment for Phase 2 was using number of pixels ( $n$ ) to weigh the linear fit equation (Figure 10). Each point is an observation that indicates the average ratio of a VZAD slice within a scene. The figure shows just how much of the data are actually along the line. Using the heat legend as reference, each data point is color coded as a function of how many pixel pairs were obtained from that observation. Briefly, dark blue points indicate scenes with 10,000 or fewer pixels, while bright yellow points designate observations with 100,000 or more pixels. The same VZAD constraint of  $\pm 10^\circ$  was used to avoid the non-linearity of the BRDF effect as angles increase out to  $\pm 20^\circ$ . The cross-calibration gain values were found in both reflectance and radiance space, with the TIRS bands only calculated in radiance units.

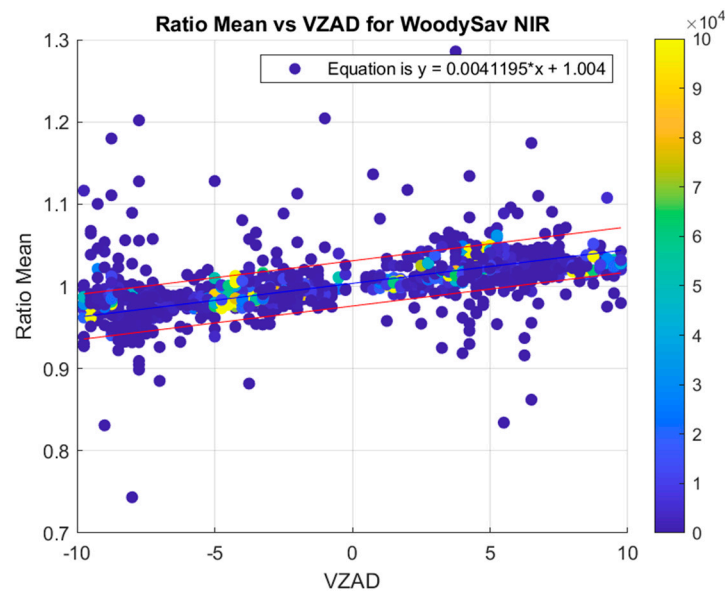


Figure 10. VZAD vs. ratio mean weighted linear fit. Note the colored data points which indicate the number of pixels per observation, with yellow points being weighted higher in the linear fit equation.

### 3.1. Cross-Calibration Gains—Reflectance

#### 3.1.1. Weighted Variance Estimator

The cross-calibration gain values were found in reflectance space using a  $\pm 10^\circ$  VZAD constraint and an ellipse filter covering three sigma of the signal to noise relationship of the data. The reflectance ratios of Landsat 8 to Landsat 9 were plotted against their VZAD, and a linear fit equation weighted by number of pixels in the observation was calculated. The intercept was taken as the cross-calibration gain for that IGBP type and band. A 68% confidence interval was calculated for each linear fit, which was used as 1 sigma of the uncertainty measurement. Table 5 presents the values before SBAF correction.

Table 5. Reflectance cross-calibration gain estimates for each IGBP type before SBAF correction.

Cross-Cal Gains	Barren1		Barren2		Barren3		ClosedShrub		Crops		DecBroad		DecNeed		EvBroad	
Band	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma
CA	1.000	0.014	0.999	0.014	0.997	0.021	0.999	0.012	0.998	0.016	0.999	0.013	1.000	0.014	1.000	0.015
Blue	1.001	0.016	1.000	0.015	0.998	0.026	1.000	0.014	1.000	0.018	1.001	0.016	1.001	0.017	1.002	0.019
Green	0.996	0.020	0.993	0.021	0.997	0.033	0.995	0.017	0.994	0.028	0.995	0.025	0.992	0.024	0.999	0.024
Red	1.002	0.029	1.001	0.027	0.997	0.032	0.998	0.019	0.999	0.036	1.001	0.034	1.000	0.038	1.001	0.031
NIR	1.003	0.031	1.005	0.028	0.998	0.024	0.999	0.023	1.002	0.031	1.001	0.029	1.008	0.054	1.001	0.026
SWIR1	1.010	0.045	1.014	0.037	0.997	0.026	1.002	0.021	1.008	0.055	1.007	0.040	1.020	0.069	1.002	0.027
SWIR2	1.009	0.040	1.018	0.054	0.997	0.032	1.000	0.029	1.009	0.060	1.007	0.089	1.028	0.082	1.001	0.035
Pan	1.002	0.019	0.996	0.021	1.005	0.038	1.000	0.015	0.993	0.022	1.000	0.015	0.968	0.034	1.005	0.023
	EvNeed		Grass		MixedFor		NatVeg		OpenShrub		Savanna		WoodySav			
Band	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma
CA	0.999	0.014	1.000	0.013	1.000	0.015	0.999	0.013	0.999	0.019	0.999	0.012	1.000	0.015		
Blue	1.001	0.014	1.001	0.014	1.002	0.018	1.001	0.015	1.001	0.021	1.000	0.020	1.002	0.018		
Green	0.992	0.021	0.996	0.017	0.996	0.028	0.998	0.020	1.000	0.022	0.996	0.022	0.997	0.024		
Red	0.999	0.031	1.000	0.022	1.003	0.026	1.000	0.030	1.000	0.022	1.000	0.022	1.003	0.032		
NIR	1.004	0.039	1.001	0.024	1.004	0.024	1.001	0.022	1.000	0.017	1.001	0.022	1.004	0.027		
SWIR1	1.011	0.052	1.005	0.027	1.011	0.034	1.002	0.027	0.999	0.019	1.003	0.025	1.009	0.034		
SWIR2	1.016	0.051	1.003	0.032	1.012	0.041	1.001	0.034	0.999	0.024	1.002	0.031	1.010	0.043		
Pan	0.984	0.024	1.001	0.019	1.001	0.019	1.003	0.017	1.006	0.018	1.003	0.016	1.004	0.016		

In Phase 1, the Barren land cover types were separated from the vegetative types, but results were ultimately combined to give one set of values. In Phase 2, the extra step of separating them was deemed unnecessary. The process used to combine the values was an inverse weighted variance mean, Equation (8), with the standard deviation of the estimate being a similar weighted formula, Equation (9). These formulas inversely weigh each value

by their variances; the higher the variance, the lower the weight. The same process was applied to Phase 2 since it is a robust method for combining all the values. The formulas are effective in using all the data available by giving the values with lower confidence less weight than the higher confidence values. The weighted variance estimator values for reflectance before spectral correction are shown in Table 6. The inverse weighted variance mean formula is given by

$$\hat{y} = \frac{\sum_i \frac{y_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}} \quad (8)$$

where  $y_i$  is the ratio mean of each IGBP type and  $\sigma_i$  is the sigma of each IGBP type. The standard deviation of this inverse weighted mean is

$$\text{Var}(\hat{y}) = \frac{1}{\sum_i \frac{1}{\sigma_i^2}} \quad (9)$$

where the same variables are given [25].

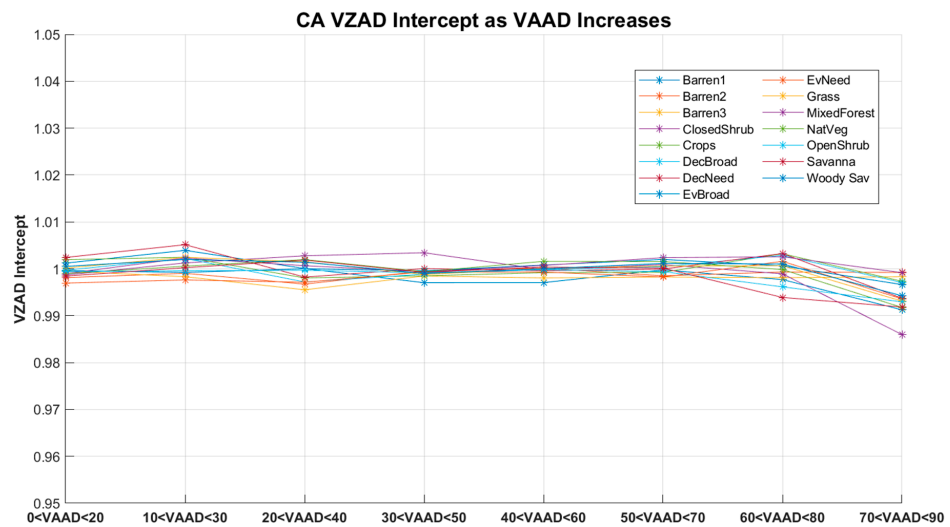
**Table 6.** Inverse-weighted mean variance estimation before SBAF correction.

		Weighted Variance Estimator
Band	Mean	Std Dev
CA	0.999	0.004
Blue	1.001	0.004
Green	0.996	0.006
Red	1.000	0.007
NIR	1.001	0.007
SWIR1	1.004	0.008
SWIR2	1.004	0.010
Pan	1.000	0.005

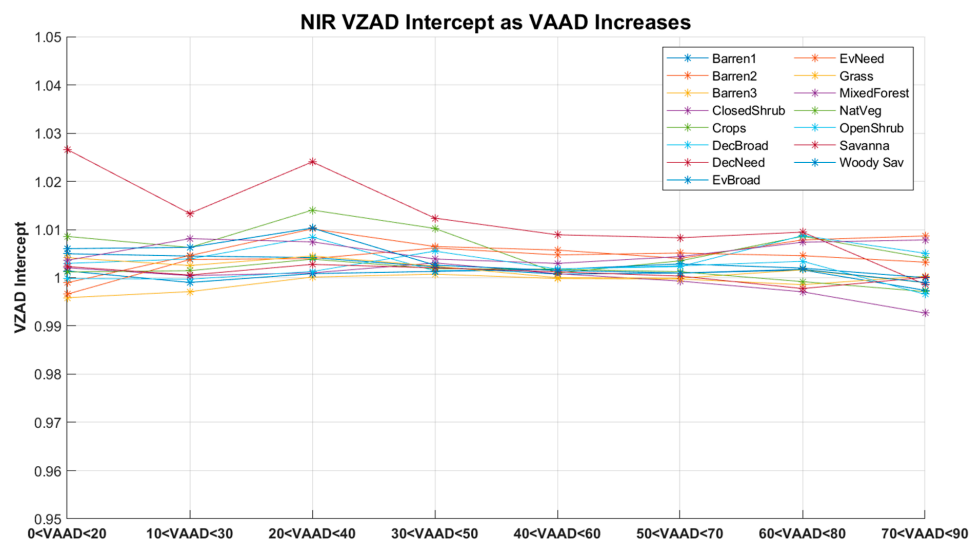
These values are within the 0.5% uncertainty budget suggested in Phase 1 of the analysis; however, this table shows the results before spectral correction has been applied or BRDF effects have been analyzed.

### 3.1.2. VAAD Influence on Results

As stated in the previous section, the VAAD of each sensor has an effect on the cross-calibration gains. If there was a large difference in ratio means as the sensor views moved closer to the cross-principal plane, then the largest VAADs would need to be used to reduce the effect of the “hot spot” of the BRDF model. To test this, the same linear fit equation was calculated only for data with specific 20° VAAD slices incremented by 10°. Figure 11 shows the VZAD intercept as VAAD increases from 0° to 90° for the CA band, the wavelength most influenced by atmosphere effects. Figure 12 shows the same trend in the NIR, albeit with slightly more spread, likely due to the atmospheric effects being lessened at longer wavelengths. The lone outlying IGBP type that appears to be varying outside the others is Deciduous Needleleaf. Its behavior is likely caused by its low signal value, where all seven bands have average reflectance values of around 0.10. This results in a low signal-to-noise ratio (SNR), which was likely the cause of Deciduous Needleleaf’s atypical performance.



**Figure 11.** The VZAD intercept for all IGBP types as VAAD increases from 0–90° in the CA band. Note how the intercepts are about ±0.5% for all VZAD slices, except at the 70 < VAAD < 90 range, which is likely driven by number of samples; there are very few underlying scenes at that range.



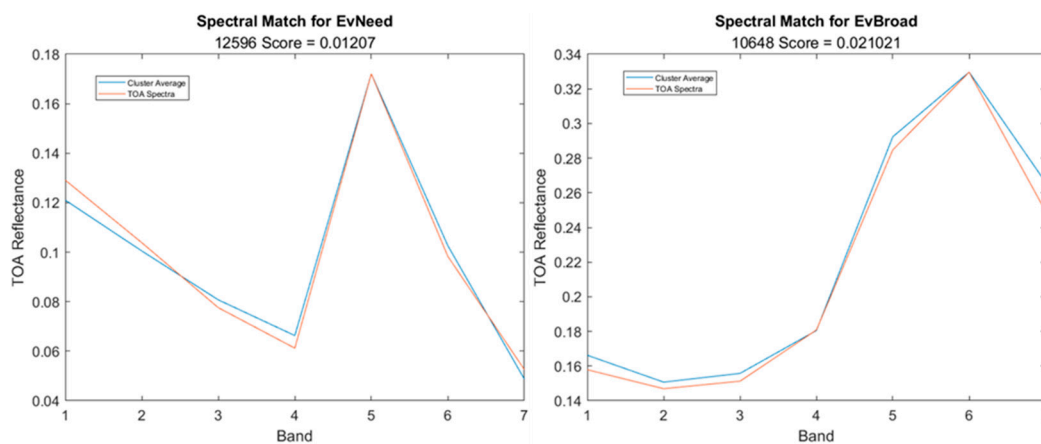
**Figure 12.** The VZAD intercept for all IGBP types as VAAD increases from 0–90° in the NIR band. Note how the intercepts are about ±1% for all VZAD slices, except for the Deciduous Needleleaf class, which has a very low SNR, which likely drives its atypical performance. The “tightest” VAAD slice appeared to be the 50 < VAAD < 70 range, which had the greatest number of samples during the underfly, so a *t*-test was performed to test if the other VAAD slices were statistically similar to it.

Briefly, the results appear to show each VAAD slice is similar to the next, with each IGBP behaving about the same. The most consistent behavior appears in VAAD slices 40 < VAAD < 60 and 50 < VAAD < 70. This behavior seems to be driven by the number of samples, since these VAAD slices have about three or four times the samples of the other slices. These VAADs occur when the instruments are between World Reference System (WRS) rows 30 and 55, which are all above the equator. There is a larger amount of physical land mass in the northern hemisphere versus the southern hemisphere, so these VAADs having a larger number of samples makes sense. In contrast, VAAD slices close to the principal plane and cross-principal plane have much smaller sample sizes. For reference, the instruments consistently have around a 98° viewing azimuth angle with respect to a target, so again the more consistent VAADs in the northern hemisphere, where the solar azimuth angle with respect to the Earth is further south, can be confirmed when discussing the problem geometrically.

To prove that each of the VAAD slices are statistically unchanging, a *t*-test was performed, which compared all the VAADs to  $50 < \text{VAAD} < 70$  in each IGBP type and band using an alpha of 99%, where any VAAD slice with a value greater than one percent means it is statistically the same as  $50 < \text{VAAD} < 70$ . The results of this *t*-test are shown in Appendix A, with tables for the CA and NIR bands given as examples, showing that the worst performing VAAD slice is  $0 < \text{VAAD} < 20$ . These results make sense given that there are a low number of samples in this slice, and the sensors are in the same azimuth plane as the sun with respect to their target; this increases BRDF uncertainty, as stated in the introduction. The *t*-test also showed that as wavelength increases, the effects of atmospheric scattering are decreased, which in turn increases variation as the target becomes more visible to the sensor. Solar irradiance also decreases at these longer wavelengths, which might play a factor in the increased variation. Other than these few aspects, VZAD intercept remains largely unchanged as VAAD increases, so all VAAD slices past  $0 < \text{VAAD} < 20$  are practical for this analysis. The consistent VZAD intercept behavior as VAAD increases was expected after the Phase 1 Ross–Li model analysis, and it was reassuring to observe the same trend in the underlying analysis.

### 3.1.3. SBAF Correction

The spectral library compiled at SDSU was used to find the best spectral match for each of the IGBP types. After applying a 1976 MODTRAN model with default settings to each of the hyperspectral signatures, Equation (4) was used to find the best match. The Landsat 8 banded spectra from SDSU were scaled to each of the IGBP types to find the best shape that matched, since that mattered more for SBAF than the magnitude. This normalization process more accurately matches hyperspectral signatures to multispectral profiles since the shape of the profile matters much more than the scale of it with respect to SBAF. Figure 13 shows a couple examples indicating the small distance between the multispectral profile and its matching hyperspectral counterpart.



**Figure 13.** IGBP to Banded hyperspectral signature matched examples. Note the score in the subtitles of the plots, as it is the resulting value from Equation (4). The lower the value, the closer the spectral match of the IGBP profile to the spectral library profile. The results show the spectra are about one to two reflectance units different from their corresponding match. The number to the left of the score signifies the sample in the spectral library used to match.

After the hyperspectral signature was chosen, it was banded using the Landsat 9 RSRs. These values were used to calculate the SBAFs for each of the IGBP types, as shown in Table 7.

**Table 7.** SBAFs for each of the IGBP land cover types.

IGBP	CA	Blue	Green	Red	NIR	SWIR1	SWIR2	Pan
EvNeed	0.998	0.999	0.998	1.000	1.000	1.001	1.004	1.005
EvBroad	0.999	1.000	1.000	1.001	1.000	1.001	1.001	0.997
DecNeed	0.998	0.999	0.998	1.000	1.000	1.001	1.003	1.006
DecBroad	0.998	0.999	0.998	1.000	1.000	1.001	1.002	1.004
Mixedfor	0.999	0.999	0.998	1.000	1.000	1.002	1.002	1.005
ClosedShrub	0.999	1.000	0.999	1.001	1.000	1.001	1.002	1.000
OpenShrub	0.999	0.999	1.001	1.001	1.000	1.000	1.000	0.992
WoodySav	0.998	0.999	0.998	1.000	1.000	1.001	1.002	1.003
Savanna	0.999	1.000	0.999	1.001	1.000	1.001	1.001	0.999
Grass	0.998	0.999	0.999	1.001	1.000	1.001	1.001	0.999
Crops	0.999	0.999	0.999	1.000	1.000	1.002	1.003	1.003
NatVeg	0.999	1.000	0.999	1.001	1.000	1.001	1.002	1.000
Barren1	0.998	0.999	0.999	1.001	1.000	1.001	1.001	0.999
Barren2	0.998	0.999	0.998	0.999	1.000	1.002	1.003	1.009
Barren3	1.000	1.000	1.002	1.001	1.000	1.000	1.000	0.992

The results of spectral correction show negligible differences in most of the bands across IGBP types, with a couple of notable exceptions in the pan band, which has differences up to 0.9%. The purpose of the pan band is focused more on spatial accuracy than spectral, so this result is not surprising. The NIR band is virtually unaffected by SBAF, while the green and SWIR2 bands show differences between soil and vegetative targets. Since the green band has some SBAFs above and below unity depending on land cover type, the cluster classification step in the methodology was confirmed to be incredibly important for spectral correction. These SBAFs were applied to the cross-calibration gains to find the SBAF-corrected results (Table 8).

**Table 8.** SBAF-corrected cross-calibration gains. The inverse-weighted mean variance estimation results were the values recommended to USGS EROS for final reflectance cross-calibration. Note how the values are within 0.5% of unity, which further supports how effective Phase 1 was.

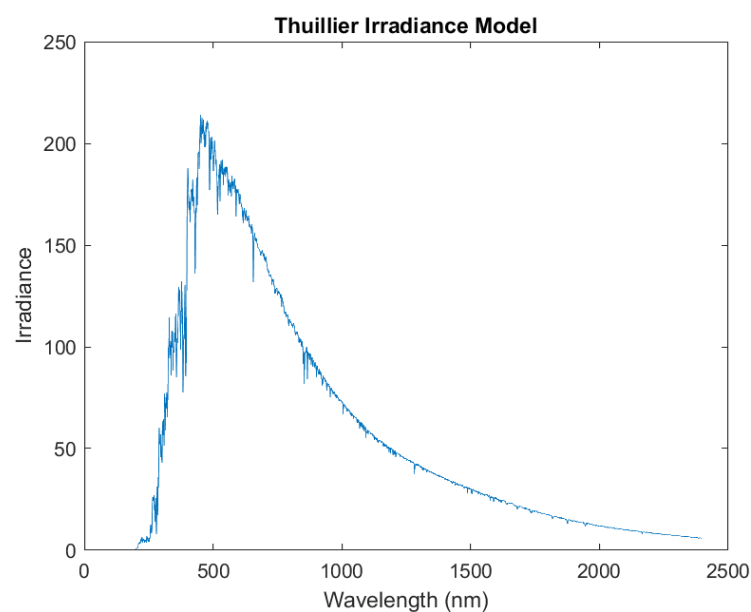
SBAF Corrected Gains	Barren1		Barren2		Barren3		ClosedShrub		Crops		DecBroad		DecNeed		EvBroad	
	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma
CA	1.001	0.014	1.001	0.015	0.997	0.021	1.001	0.012	1.000	0.016	1.001	0.013	1.002	0.014	1.001	0.015
Blue	1.002	0.016	1.002	0.015	0.998	0.026	1.000	0.014	1.001	0.018	1.002	0.016	1.002	0.017	1.003	0.019
Green	0.997	0.020	0.995	0.021	0.995	0.033	0.996	0.017	0.995	0.028	0.996	0.025	0.993	0.024	0.999	0.024
Red	1.001	0.029	1.001	0.027	0.996	0.032	0.998	0.019	0.999	0.036	1.001	0.034	1.000	0.038	1.000	0.030
NIR	1.003	0.031	1.005	0.028	0.998	0.024	0.999	0.023	1.002	0.031	1.001	0.029	1.008	0.054	1.001	0.026
SWIR1	1.009	0.045	1.011	0.037	0.997	0.026	1.001	0.021	1.006	0.055	1.005	0.040	1.018	0.069	1.001	0.027
SWIR2	1.008	0.040	1.015	0.054	0.996	0.032	0.999	0.029	1.006	0.060	1.005	0.089	1.026	0.082	1.000	0.035
Pan	1.003	0.019	0.988	0.021	1.013	0.038	1.000	0.015	0.989	0.022	0.996	0.014	0.962	0.034	1.009	0.023

SBAF Corrected Gains	EvNeed		Grass		MixedFor		NatVeg		OpenShrub		Savanna		WoodySav		Inverse-Weighted Variance Estimator	
	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	±Sigma	Mean	Std Dev
CA	1.000	0.015	1.001	0.013	1.001	0.015	1.001	0.013	1.000	0.019	1.000	0.012	1.002	0.015	1.001	0.004
Blue	1.002	0.014	1.001	0.014	1.002	0.018	1.002	0.015	1.002	0.021	1.001	0.020	1.002	0.018	1.002	0.004
Green	0.994	0.021	0.996	0.017	0.998	0.028	0.998	0.020	0.999	0.022	0.997	0.022	0.998	0.024	0.996	0.006
Red	0.999	0.031	0.999	0.022	1.003	0.026	1.000	0.030	0.999	0.022	0.999	0.022	1.002	0.032	1.000	0.007
NIR	1.004	0.039	1.001	0.024	1.004	0.024	1.000	0.022	1.000	0.017	1.000	0.022	1.003	0.027	1.001	0.007
SWIR1	1.009	0.052	1.004	0.027	1.009	0.034	1.000	0.027	0.999	0.019	1.002	0.025	1.008	0.034	1.003	0.008
SWIR2	1.013	0.050	1.002	0.032	1.009	0.041	1.000	0.034	0.999	0.024	1.001	0.031	1.008	0.042	1.002	0.010
Pan	0.978	0.023	1.002	0.019	0.997	0.019	1.004	0.017	1.015	0.018	1.004	0.016	1.001	0.016	0.999	0.005

### 3.2. Cross-Calibration Gains—Radiance

The radiance cross-calibration followed the same core steps as the reflectance process, with the main difference being the addition of the TIRS bands. These bands were calibrated using only water pixels, and these pixels were extracted from the same underfly images as the rest of the IGBP land covers. This meant that the water pixels were not from deep-ocean images, but instead coastlines, rivers, and lakes. Because water's emissivity is so high, the TIRS bands did not require a spectral correction, while the other bands did. This SBAF correction was calculated by applying the Thuillier irradiance model to the same land cover hyperspectral signatures used in the reflectance process to determine the radiance of each IGBP type, as shown in Figure 14. The Thuillier model was chosen due to its wide use in the remote sensing community [26]. However, the model does have a mean uncertainty around 3%. The ratios of these radiances were then taken to calculate the radiance SBAFs for each land cover. The SBAF-corrected radiance cross-calibration gain estimates are shown in Table 9.



**Figure 14.** Thuillier irradiance model. This irradiance was applied to the reflectance curves in the SDSU-compiled spectral library to determine the radiance of each sample.

**Table 9.** Radiance SBAF-corrected cross-calibration gain estimates. These values are further from unity than the reflectance values, which was the main focus of Phase 1.

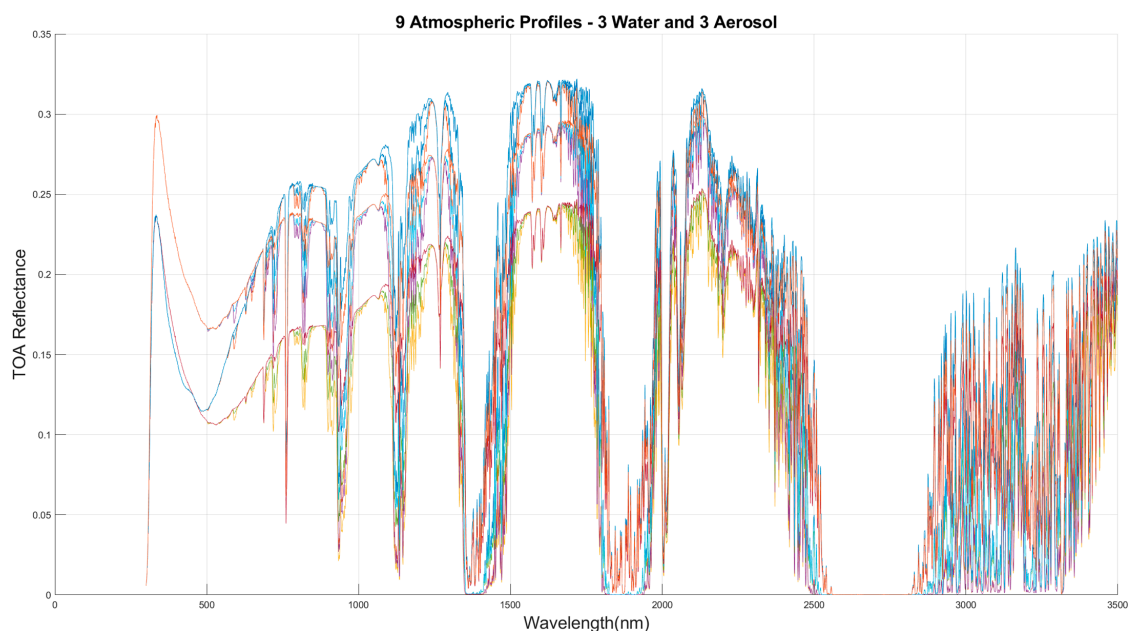
Band	Inverse-Weighted Variance Estimator	
	Mean	Std Dev
CA	1.001	0.004
Blue	0.999	0.004
Green	0.997	0.006
Red	0.996	0.007
NIR	0.996	0.007
SWIR1	0.992	0.008
SWIR2	0.990	0.010
Pan	0.996	0.005
Band 10	0.998	0.008
Band 11	1.010	0.008

### 3.3. Uncertainty Analysis

The Phase 1 Underfly analysis did not include an exhaustive uncertainty analysis, primarily due to time constraints. Instead, it relied on rough estimates from spectral target analysis and the use of the VZAD intercept approach to reduce the uncertainty due to BRDF effects, which resulted in an estimated uncertainty within 1%. In this effort, a comprehensive study of uncertainty due to spectral, geometric, and BRDF effects was undertaken and is presented in the following sections.

#### 3.3.1. Spectral Uncertainty

As stated in the introduction, spectral uncertainty came in the form of band pass differences between the sensors as well as the spectral differences between targets. Each sensor has 14 focal plane modules (FPM) for each band. The SBAF correction for Phase 2 was calculated using the band averages of each sensor, so there was not an FPM-to-FPM correction, resulting in some uncertainty. The SBAF correction step in Phase 2 corrected for several generic land cover types across a single generic atmospheric model; this model did not account for different water absorption levels or aerosol profiles, which again resulted in some uncertainty. To quantify all these uncertainties, different atmospheric profiles were simulated using MODTRAN at three different water levels and using three different aerosol profiles. An example including all nine combinations with the land cover Barren3 is shown in Figure 15. By taking the SBAF of each of Landsat 8's 14 FPMs and L9's 14 FPMs, the sensors uncertainty can be quantified as well. The total number of simulations used in the spectral uncertainty analysis can be calculated by taking 14 L8 FPMs, 14 L9 FPMs, 15 IGBP land cover types, and nine atmospheric profiles, which resulted in 26,460 simulations. The SBAF for each simulation was calculated, with the average SBAF and standard deviation being shown for each band in Table 10. The standard deviation can be considered the total spectral uncertainty associated with Phase 2 since this is an exhaustive SBAF analysis. As can be seen below, the uncertainties were all less than a quarter of a percent, with the largest being the pan band, which is the broadest spectral band.



**Figure 15.** Barren3 atmospheric profiles for spectral uncertainty analysis. Note there are three aerosol profiles and three water absorption profiles, which, combined, give a total of nine combinations. This large range of atmospheric profiles gave the spectral uncertainty analysis enough simulations to cover all possible scenarios during the underfly.



**Table 10.** Spectral uncertainty table. Uncertainty units are reflectance.

Band	CA	Blue	Green	Red	NIR	SWIR1	SWIR2	Pan
Ratio Mean	0.9989	0.9995	0.9992	1.0004	1.0001	1.0011	1.0013	1.0009
Uncertainty	0.0012	0.0007	0.0010	0.0007	0.0008	0.0017	0.0015	0.0022

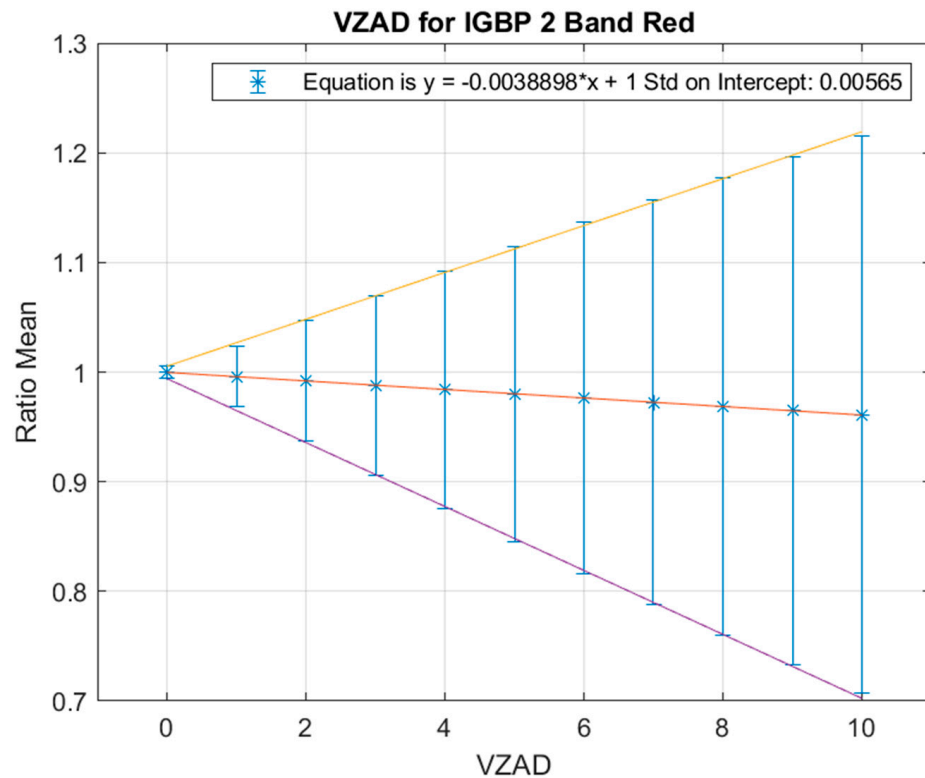
### 3.3.2. BRDF Uncertainty

Since there was no BRDF correction in the underfly analysis, the BRDF uncertainty study focused more on the uncertainty associated with the VZAD intercept estimation approach. For the BRDF uncertainty analysis, VZAD was binned up into  $1^\circ$  bins to save on processing time that non-integer values would have required. Each VZAD angle had several sensor viewing angles that could be observed in it. For example, a VZAD of  $1^\circ$  could have Landsat 8 viewing a target at a VZA of  $1^\circ$  and Landsat 9 viewing at  $0^\circ$  where the reflectance ratio could be a set value, or L8 could be at  $2^\circ$  VZA and L9 at  $1^\circ$  VZA, with a slightly different reflectance ratio. VZAD of  $1^\circ$  has up to 14 different observations, all with slightly different reflectance ratios between them, where VZA  $-6^\circ$  and  $-7^\circ$  exhibited a ratio up to several percent different than VZA  $7^\circ$  and  $6^\circ$ . This disparity only increased as VZAD increased out to  $10^\circ$ , which was the cutoff for the VZAD intercept estimator. These differences were at their greatest when the sensor was in the principal plane, that is, when the sensor and sun have the same azimuth angle with respect to a target. All BRDF analysis took place in the principal plane to determine the worst possible outcome.

While this difference could be measured using BRDF models, it changes from cover type to cover type, and includes uncertainty within a specific land cover type. This means the VZAD uncertainty increases as VZAD increases from  $1^\circ$  to  $10^\circ$ . This effect also appears in negative VZADs on the left side of the intercept, so the effect is an absolute function, resulting in a “bowtie” shape of uncertainty. This bowtie effect is most prominent when measuring the BRDF uncertainty in the principal plane, which is the worst-case scenario. Because most of the underfly data takes place in the VAAD slice  $50 < VAAD < 70$ , the bowtie effect is not visibly observed in VZAD versus reflectance mean plot; however, an uncertainty analysis should look at the worst possible case for study to determine how large uncertainty could be. By isolating and assuming the only differences between the sensors is viewing angle in this analysis, the BRDF uncertainty can be expected to be mirrored across the intercept, meaning all analysis could be performed using positive VZAD values. The uncertainty at the VZAD intercept was interpolated by finding the uncertainty at every integer VZAD angle and fitting a linear trend to the error bars, which can be seen in Figure 16. The resulting intercept from the error bars would be considered the BRDF uncertainty at VZAD =  $0^\circ$ .

The uncertainty within a single VZAD was calculated using the MODIS Ross–Li BRDF parameters studied during Phase 1 of the analysis. 100,000 random models of each IGBP type were used, along with three different sun angles, to simulate a wide range of data. The sun angles chosen were the angles at the northernmost point of the underfly, the southernmost point, and a sun position near the equator to cover most scenarios. As previously mentioned, VZAD =  $1^\circ$  could have a total of 14 different observations, while VZAD =  $10^\circ$  could only have five observations, with a VZA cutoff of  $7^\circ$ . This meant that with 100,000 models at three sun positions, VZAD =  $1^\circ$  would have 4.2 million ratio means in each band and IGBP type, while VZAD =  $10^\circ$  has 1.5 million of these samples. The median and median absolute deviation (MAD) of all the samples in each VZAD were found to create the type of plot found in Figure 16, where the points are the median and the error bars are the MAD. Each MODIS band was modelled, which meant the Landsat coastal aerosol band and panchromatic band did not have MODIS equivalents. When results were simulated, uncertainties for these bands were estimated using linear extrapolation from the green and blue bands out to the CA band, while a root mean square (RMS) of red, green, and blue was used to determine the pan band results. The intercept uncertainties of each

IGBP type in each band were found and combined across IGBP types using their RMS values; results are shown in Table 11.



**Figure 16.** BRDF uncertainty at the VZAD intercept. Note that the “bowtie” effect is prominent when looking at the BRDF uncertainties while exclusively in the principal plane, which gave the worst-case scenario for uncertainty. The error bars on the VZAD intercept are the estimated BRDF uncertainty at VZAD = 0° based on the uncertainties of VZADs = 1–10°. This value was averaged with the VZAD intercept uncertainties of every IGBP type in each band to find the overall BRDF uncertainty for all the bands.

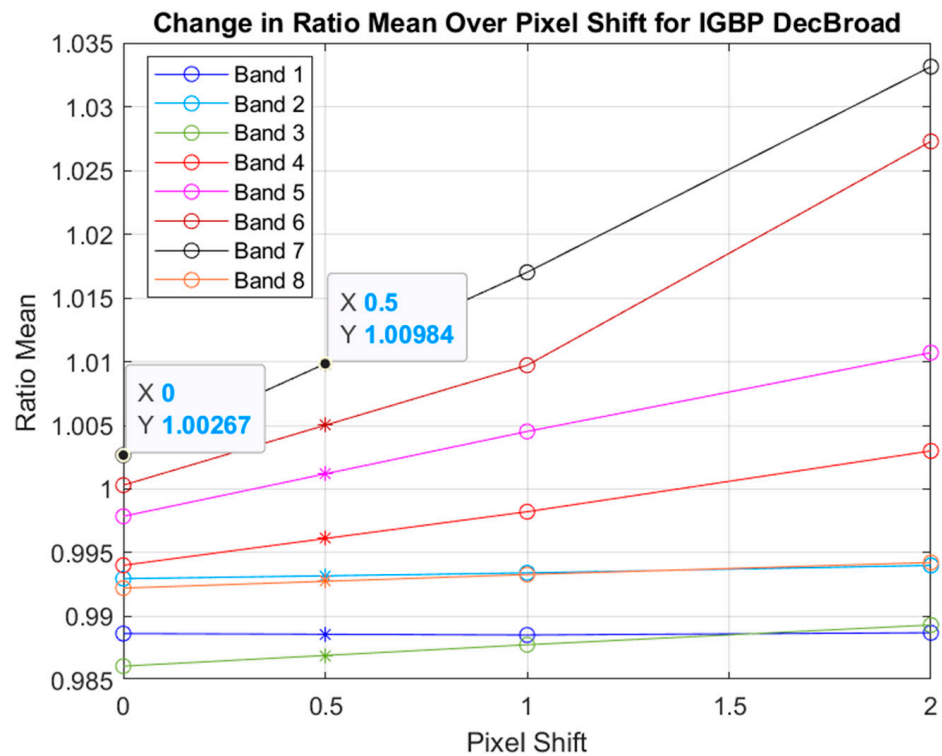
**Table 11.** BRDF uncertainty table.

Band	CA	Blue	Green	Red	NIR	SWIR1	SWIR2	Pan
Uncertainty	0.0007	0.0011	0.0024	0.0015	0.0026	0.0017	0.0009	0.0017

### 3.3.3. Geometric Uncertainty

The geometric uncertainty analysis focused on pixel shift error. As stated by USGS, the geometric alignment of Landsat products is “expected to be consistent to within 12 m”; as such, a sub-pixel geometric analysis was deemed necessary [18]. The analysis was performed by taking the original underfly image-overlapping script and shifting the Landsat 9 image up by one pixel. The L8 to L9 reflectance ratio of the image pair was compared to the original to determine the pixel shift error. This same process was conducted for a two pixel shift. The ratio mean consistently increased as pixel shift increased, which was due to the way ratio mean was calculated in a scene. Assume in a normal distribution that a reflectance ratio is just as common as its inverse. If those two values were averaged, their mean would be greater than one. For example, a ratio of 2 and ½ are inverses of each other and their average is 1.25. The ratio mean distribution also becomes skewed right, with a longer tail on the right. This overpowers the values on the left of the distribution and forces the mean greater than 1. As pixel shift increases, the ratio mean always increases, which can be seen as a statistical bias. The spatial resolution of bands 1 through 7 is 30 m, so the geometric bias at 12 m was estimated. This was accomplished by interpolating the

ratio mean value at pixel shift 0.5, which is a slight overestimation at 15 m shift, shown in Figure 17. The difference between the interpolated value and the value at pixel shift 0 was considered the introduced geometric bias at about 12 m. Because the pan band spatial resolution is 15 m, the interpolation step was skipped and the difference between pixel shift 1 and 0 was the estimated geometric bias. The bias across IGBP types and bands were combined using their RMS values; results are shown in Table 12.



**Figure 17.** Geometric bias plot. Pixel shift 0 has an uncertainty of ~1/3 pixel, so the bias introduced at that value needed to be determined. The ratio value between 0 and 1 at 0.5 was interpolated, denoted by \*. Since the bias at pixel 0.5 is likely larger or equal to the bias introduced at ~1/3, that value was used to overestimate the bias as a worst-case scenario. The difference between 1.00984 and 1.00267 was determined to be the geometric bias on the SWIR2 band, being 0.00717.

**Table 12.** Geometric bias table.

Band	CA	Blue	Green	Red	NIR	SWIR1	SWIR2	Pan
Uncertainty	0.0001	0.0002	0.0006	0.0020	0.0055	0.0080	0.0081	0.0007

### 3.3.4. Total Uncertainty

Once all three main aspects of uncertainty were quantified in each band, Equation (10) was used to combine them into a total uncertainty:

$$\sigma_{total} = \sigma_{Geometric} + \sqrt{\sigma_{spectral}^2 + \sigma_{BRDF}^2} \tag{10}$$

where  $\sigma$  is the corresponding uncertainty contributions for spectral, BRDF, and geometric, respectively. The spectral and BRDF contributions are random uncertainties, so they could use the root sum of squares (RSS) formula. The geometric contribution was a bias due to it always being a positive additive value, so it had to be added directly into the total. The total uncertainty for each band is shown in Table 13. When compared to the inverse-weighted estimator standard deviations in Table 8, the total uncertainty values follow a similar trend, with the shorter wavelengths having the highest precision and the SWIR bands having the lowest precision. The largest contribution to total uncertainty is the geometric bias, so the

main driving factor to the standard deviation on each IGBP type in Table 8 may very well be the geometric bias.

**Table 13.** Total uncertainty table.

Band	CA	Blue	Green	Red	NIR	SWIR1	SWIR2	Pan
Uncertainty	0.0014	0.0013	0.0027	0.0026	0.0062	0.0084	0.0083	0.0029

#### 4. Discussion

The results of the Phase 2 analysis provided insights into the advantages that are possible with an underfly maneuver. The main drawback with the Phase 1 analysis was the short amount of time available during OIV to process the data, analyze it, and provide an accurate set of values for the cross-calibration. The main benefits of the Phase 2 process were its more consistent and physics-driven processing and statistical analysis of the data, which came from a better understanding of the underfly data set. Since the Phase 2 analysis used Collection 2 Tier 1 images that had the Phase 1 cross-calibration results applied to them, any deviations from unity would be considered the corrections to the current cross-calibration gains.

##### 4.1. Results Analysis

The first point that needed addressing in Phase 1 was the spectral uncertainty. Since the analysis period of Phase 1 was so short, a proper spectral correction was not applied. The Phase 2 effort attempted to solve this problem by finding hyperspectral matches to each of the derived IGBP land cover type profiles. This was accomplished with the use of the TOA reflectance MODIS product MOD02HKM to derive the IGBP reflectance profiles. These profiles were used to sort the SDSU EPICS clusters to their closest IGBP land cover type. Integrating the TOA MODIS product into the methodology instead of the surface level MODIS BRDF product utilized in Phase 1 removed a potentially significant source of error of said profiles. The algorithm used to sort the clusters was also improved to remove any clusters that were not within three standard deviations of the MODIS profile, in order to retain the best pixels for this analysis.

This spectral matching was designed to group all like pixels, which could then be SBAF corrected once the underfly data were processed. In Phase 1, a weighted mean was used instead of SBAF correcting Landsat 9 to look like Landsat 8. The soil land cover type was weighted more heavily in the combined values due to its low spectral uncertainty in the green band. Phase 2, alternatively, found the best hyperspectral match to each of the IGBP land cover types in order to SBAF correct Landsat 9. These hyperspectral signatures came from surface-level spectroradiometer measurements and were converted to TOA reflectance using MODTRAN simulations, which again would improve overall accuracy since this was a TOA cross-calibration. The SBAF values shown in Table 7 showed that the spectral discrepancies between L8 and L9 were minimal, with the largest differences found in the pan band. These small spectral differences were due to the engineering effort to make OLI and OLI-2 near clones of each other.

The next improvements were found in the data analysis and filtering. Looking at the reflectance values across the globe, a phenomenon was discovered: the reflectance values of vegetation were darker in the northern latitudes compared to the south, especially in the NIR band. This was caused by the seasonality effects of vegetation. Since the northern hemisphere was in its winter season in November, the month of the underfly, the reflectance of vegetation was significantly darker when compared to that of the southern hemisphere, which was in its summer season. This observation led to better understanding of the underfly data set, which in turn allowed for more refined filtering when it came to removing outliers. The Phase 2 analysis focused on a more data-driven outlier rejection approach that, in the end, kept significantly more data when compared to that of Phase 1. Phase 1 used aggressive filters such as a noise floor filter and a minimum pixel count

threshold, which were both made obsolete with the statistic-based ellipse filter. This filter weighted each observation by its pixel count and used a 3-sigma ellipse to encompass about 99.7% of the total data for each IGBP type. This was not conducted to separate the data into distinct sets, but rather to retain as much good data as possible by understanding the bimodality of the data to remove outliers. This approach led to larger error bars with lower precision. However, Phase 2 should have a higher accuracy when compared to Phase 1 because more data were used in its analysis. The aggressive filters in Phase 1 could have introduced biases or artifacts to the results, while Phase 2 gave a more comprehensive look at the underfly data.

After filtering the data, the cross-calibration gain was determined using the same VZAD intercept method used in Phase 1, which accounted for the BRDF effect with respect to view zenith angle. The uncertainty regarding how the VAADs of each sensor affects the cross-calibration gain was determined to be minimal, as the value did not significantly change as VAAD increased. Once the cross-calibration gain for each IGBP type in every band was determined, an SBAF was applied to each one. Finally, the results were combined using an inverse-weighted variance mean formula. The reflectance results after SBAF correction showed that the Phase 1 calibration was a good initial calibration, as the values in Bands 1 through 7 showed differences of less than 0.5%. SBAF correction did not change the final results significantly, so the spectral uncertainty analysis performed during Phase 1 was accurate. The radiance results were slightly less encouraging when compared to the reflectance results, but this is likely due to the additional 3% uncertainty introduced in the solar irradiance model that reflectance space does not have.

The final results showed that the Phase 1 cross-calibration was a good first attempt, since all of the differences from unity in reflectance space were less than 0.5%. The radiance results, however, had a maximum difference of 1% in the SWIR2 band. The values given in Tables 8 and 9 were recommended to USGS EROS for a calibration update of Landsat 9. Since these values came after the initial calibration and include SBAF corrections, they represented the real differences between the sensors following the Phase 1 calibration. The standard deviations in both radiance and reflectance space were as high as 1%, so it is difficult to determine the significance of these changes. Uncertainty analysis was performed to determine the actual uncertainty introduced spectrally by looking at FPM-to-FPM differences, the BRDF uncertainty by analyzing Ross–Li models of each IGBP land cover types, and the geometric pointing uncertainty by performing a pixel shift analysis. The combined uncertainties of these analyses were shown in Table 13 and demonstrate that the cross-calibration was well within the proposed 1% uncertainty budget.

#### *4.2. Future Underfly Maneuvers*

The methodology constructed in Phase 1 and Phase 2 could easily be used as the foundation for the process to cross-calibrate future Landsat missions or any other spacecraft pair that is designed to perform an underfly maneuver during orbital insertion. The driving factor of uncertainty for the Landsat 8 and 9 underfly was the BRDF effect, especially with respect to the view zenith angle differences between the sensors. This was accounted for by taking the reflectance ratio at the VZAD intercept at  $0^\circ$ , and this technique can be used in any future mission. Spectral uncertainty was the next biggest factor, thus spectral correction would likely need to be considered. OLI and OLI-2 were nearly identical sensors spectrally, so an SBAF might have more of an effect in another mission depending on how spectrally similar the sensors are. Compared to Landsat 8 and 9, whose RSRs were essentially the same, missions like the Landsat 7 and 8 Underfly had a much larger spectral disparity between the sensors. The small total uncertainties obtained in the Landsat 8 and 9 analysis were due, in large part, to the minimized spectral differences. However, if one sensor was a hyperspectral sensor, SBAF would not need to be considered at all. The Landsat 8 and 9 underfly used the SDSU EPICS clusters to classify global pixels as specific land cover types, which turned almost every part of the world into a region of interest (ROI); as such, a similar system is recommended for future endeavors. A different set of global

IGBP land cover type profiles would likely need to be derived, specifically in the month the subsequent underfly takes place. This would allow for the greatest amount of global spectral accuracy. Teams leading similar underfly maneuvers may use this paper to help develop their process for cross-calibration.

## 5. Conclusions

The underfly maneuver for Landsat 9 was performed in mid-November 2021, during which the instrument collected images alongside its virtual twin, Landsat 8. These instruments were nearly identical, with their sensors designed to be as similar as possible. By imaging the same areas at the same time and at nearly identical locations, an optimal calibration opportunity was created. By considering three sources of uncertainty (spectral, angular, and pointing), an initial cross-calibration was performed before the end of OIV. The results from the Phase 1 analysis of the Landsat 8 and Landsat 9 underfly were a good first attempt at calibrating Landsat 9, but several shortcomings were discovered in the months following that first analysis, which Phase 2 was designed to improve upon. These improvements included more robust land cover type spectral matching, SBAF correction, and physics-based, statistically-driven data filtering.

The underfly data filtering was improved by splitting the data into two groups based on latitude, specifically separating the northern and southern hemispheres. The underfly calibration effort needed a proper understanding of vegetative seasonality, so that approach is recommended for any global analysis. This change over the Phase 1 analysis led to a better understanding of the underfly maneuver and allowed for more data to be used. The larger amount of data gave a more comprehensive look of the underfly event as a whole. The elliptical filtering process was an effective method that let the data drive the filters based on physical properties and statistics compared to other ad-hoc filtering approaches that might force artifacts and biases into the data. The BRDF analysis proved that almost all VAAD slices in the underfly were viable for the cross-calibration analysis. The difference in azimuth angle between the sun and a sensor with respect to a target showed no significant effect on the cross-calibration gain ratio, except when viewing along the principal plane.

The enhancements in the methodology led to the results in Tables 8 and 9, which were recommended to USGS EROS for the cross-calibration update to Landsat 9. The 0.4% change to the green band was the most drastic of the changes in reflectance space, while the bands in radiance space saw changes up to 1% in SWIR2. The green band's large change was due to the spectral approach in Phase 1, which weighted barren land cover types more heavily in the green band based on spectral uncertainty. With a proper SBAF correction, the green band's spectral uncertainty was mostly accounted for when compared to Phase 1. The green band is very sensitive to the difference between barren land covers and vegetation; this makes it is one of the most important bands, so the accuracy supplied by these results is vital for the community. Since the changes in reflectance space were all within 0.5%, the Phase 1 analysis can be seen as a success that was only improved upon further with Phase 2.

The uncertainty analysis performed in this paper further proved how robust the VZAD intercept was as cross-calibration ratio estimate. With BRDF uncertainties estimated within 0.26% for all bands, the VZAD intercept essentially interpolated the ratio gain for the case where the sensors were directly on top of each other at the same time to near perfect precision, which almost completely accounts for several aspects of BRDF. With near identical RSRs between the sensors, the spectral uncertainties added very little possible error to the analysis. There was a maximum spectral uncertainty of 0.22%, which did not contribute much to the overall uncertainty. The major contribution to the uncertainty was the bias introduced to the ratio mean calculation from geometric uncertainty. Even with that bias, the total uncertainty was well under the 1% budget suggested in Phase 1. A calibration with this level of precision has not been accomplished before; this is in part due to the similarities between the instruments, but also from the approach described in this paper. This process should not be thought of as exclusive to Landsat 8 and 9, so anyone

planning a future underfly maneuver can look to this paper to derive a methodology to cross-calibrate those instruments.

**Author Contributions:** G.G. from SDSU wrote the original and final drafts, processed and analyzed the underfly data, refined the methodology, as well as performed the uncertainty analysis on SBAF, BRDF, and geometry. D.H. from SDSU oversaw the project, reviewed the paper, and provided valuable insights into the conceptualization of the methodology and uncertainty analyses. L.L. from SDSU helped with processing and downloading the data, as well as other logistical contributions needed for the underfly event. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by NASA Radiometric Calibration grant number SA22000091 from the Landsat Project Science Office and by USGS EROS Landsat 8-9 grant number SA2000371.

**Data Availability Statement:** Landsat 8 and Landsat 9 courtesy of the US Geological Survey Collection 2 Landsat 8-9 OLI/TIRS Digital Object Identifier (DOI) number: /10.5066/P975CC9B.

**Acknowledgments:** The authors would like to thank the Landsat Calibration/Validation Team at USGS EROS Data Center for their insight regarding geometric uncertainty as well as other advice and feedback on this project. The authors would also like to thank NASA Goddard Space Flight Center for their support during this project.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** CA band VAAD *t*-test table. Note that any values less than 0.01, or 1%, are statistically different from the VAAD slice  $50 < VAAD < 70$ . In this case, all slices were viable for analysis.

CA	Barren1	Barren2	Barren3	ClosedShrub	Crops	DecBroad	DecNeed	EvBroad
0–20	0.8716	0.9649	0.8868	0.8026	0.7616	0.9333	0.7833	0.7882
10–30	0.6414	0.9524	0.9948	0.9676	0.7198	0.8416	0.5831	0.7561
20–40	0.9746	0.8418	0.7587	0.8336	0.8953	0.7153	0.8502	0.8235
30–50	0.7601	0.9079	0.9814	0.7747	0.9324	0.8577	0.9910	0.7841
40–60	0.7594	0.8536	0.9819	0.9045	0.9923	0.9471	0.9684	0.8920
50–70	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
60–80	0.8184	0.9280	0.9871	0.8269	0.6558	0.7863	0.7117	0.8789
70–90	0.3640	0.9385	0.9956	0.1166	0.8301	0.7092	0.5255	0.5539
CA	EvNeed	Grass	MixedFor	NatVeg	OpenShrub	Savanna	WoodySav	
0–20	0.8844	0.9605	0.7001	0.7712	0.9993	0.8468	0.9430	
10–30	0.9443	0.8349	0.9932	0.9116	0.9840	0.9934	0.9249	
20–40	0.9003	0.9207	0.8473	0.9623	0.9742	0.8485	0.9760	
30–50	0.9028	0.8395	0.7358	0.8078	0.9788	0.9203	0.8097	
40–60	0.9033	0.9533	0.8518	0.9962	0.9639	0.9811	0.8986	
50–70	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
60–80	0.7074	0.9948	0.9832	0.8494	0.7176	0.4844	0.9893	
70–90	0.6281	0.4169	0.7273	0.3087	0.4939	0.3638	0.4562	

**Table A2.** NIR band VAAD *t*-test table. Note that any values less than 0.01, or 1%, are statistically different from the VAAD slice  $50 < \text{VAAD} < 70$ . In this case, most slices were viable for analysis, except for  $0 < \text{VAAD} < 20$ . This slice is closest to the principal plane, which results in the highest amount of BRDF uncertainty and does not have a lot of samples in the underfly data.

NIR	Barren1	Barren2	Barren3	ClosedShrub	Crops	DecBroad	DecNeed	EvBroad
0–20	0.0052	0.0004	0.0020	0.0423	0.0123	0.6946	0.0458	0.2458
10–30	0.0393	0.7505	0.1070	0.4638	0.4542	0.4155	0.8374	0.0046
20–40	0.1207	0.0043	0.9258	0.3504	0.0133	0.0047	0.1800	0.4647
30–50	0.6059	0.1606	0.4763	0.0838	0.0260	0.5023	0.5692	0.5833
40–60	0.9068	0.2603	0.9398	0.4909	0.1145	0.7847	0.8391	0.2597
50–70	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
60–80	0.7786	0.3516	0.7036	0.3136	0.1068	0.0331	0.8417	0.7532
70–90	0.2552	0.2167	0.8661	0.0037	0.8733	0.6044	0.4770	0.2837
NIR	EvNeed	Grass	MixedFor	NatVeg	OpenShrub	Savanna	WoodySav	
0–20	0.0085	0.0006	0.7018	0.9596	0.0003	0.0837	0.0180	
10–30	0.8709	0.1364	0.0433	0.8758	0.0003	0.8795	0.0010	
20–40	0.8567	0.2009	0.1961	0.1051	0.3255	0.1354	0.0000	
30–50	0.8462	0.1614	0.7879	0.5485	0.3377	0.1343	0.3131	
40–60	0.8743	0.9424	0.2594	0.5539	0.6783	0.5187	0.7324	
50–70	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
60–80	0.9102	0.3674	0.3017	0.5841	0.8025	0.2013	0.8368	
70–90	0.7138	0.7657	0.4444	0.0025	0.2192	0.8652	0.7935	

## References

- Masek, J.G.; Wulder, M.A.; Markham, B.; McCorkel, J.; Crawford, C.J.; Storey, J.; Jenstrom, D.T. Landsat 9: Empowering open science and applications through continuity. *Remote Sens. Environ.* **2020**, *248*, 111968. [\[CrossRef\]](#)
- Markham, M.B.; McCorkel, J.; Montanaro, M.; Morland, E.; Pearlman, A.; Pedelty, J.; Wenny, B.; Barsi, J.; Donley, E.; Efremova, B.; et al. Landsat 9: Mission Status and Prelaunch Instrument Performance Characterization and Calibration. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5788–5791. [\[CrossRef\]](#)
- Mishra, N.; Haque, M.O.; Leigh, L.; Aaron, D.; Helder, D.; Markham, B. Radiometric Cross Calibration of Landsat 8 Operational Land Imager (OLI) and Landsat 7 Enhanced Thematic Mapper Plus (ETM+). *Remote Sens.* **2014**, *6*, 12619–12638. [\[CrossRef\]](#)
- Teillet, P.M.; Markham, B.L.; Barker, J.L.; Storey, J.C.; Irish, R.R.; Seiferth, J.C. Landsat sensor cross-calibration using nearly coincidental matching scenes. In *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*; SPIE: Bellingham, WA, USA, 2000; Volume 4049.
- Gross, G.; Helder, D.; Begeman, C.; Leigh, L.; Kaewmanee, M.; Shah, R. Initial Cross-Calibration of Landsat 8 and Landsat 9 Using the Simultaneous Underfly Event. *Remote Sens.* **2022**, *14*, 2418. [\[CrossRef\]](#)
- Teillet, P.M.; Fedosejevs, G.; Thome, K.J.; Barker, J.L. Impacts of spectral band difference effects on radiometric cross-calibration between satellite sensors in the solar-reflective spectral domain. *Remote Sens. Environ.* **2007**, *110*, 393–409. [\[CrossRef\]](#)
- Chander, G.; Mishra, N.; Helder, D.L.; Aaron, D.; Choi, T.; Angal, A.; Xiong, X. Use of EO-1 Hyperion data to calculate spectral band adjustment factors (SBAF) between the L7 ETM+ and Terra MODIS sensors. In Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, USA, 25–30 July 2010.
- Wanner, W.; Li, X.; Strahler, A.H. On the derivation of kernels for kernel-driven models of bidirectional reflectance. *J. Geophys. Res. Atmos.* **1995**, *100*, 21077–21089. [\[CrossRef\]](#)
- Wanner, W.; Strahler, A.H.; Hu, B.; Lewis, P.; Muller, J.-P.; Li, X.; Schaaf, C.L.B.; Barnsley, M.J. Global retrieval of bidirectional reflectance and albedo over land from EOS MODIS and MISR data: Theory and algorithm. *J. Geophys. Res. Atmos.* **1997**, *102*, 17143–17161. [\[CrossRef\]](#)
- Lucht, W.; Schaaf, C.B.; Strahler, A.H. An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 977–998. [\[CrossRef\]](#)



11. Strahler, A.H.; Muchoney, D.; Borak, J.; Friedl, F.; Gopal, S.; Lambin, L.; Moody, A. MODIS Land Cover Product Algorithm Theoretical Basis Document (ATBD). 1999, 72. Available online: [http://modis.gsfc.nasa.gov/data/atbd/atbd\\_mod12.pdf](http://modis.gsfc.nasa.gov/data/atbd/atbd_mod12.pdf) (accessed on 13 October 2021).
12. Bovensmann, H.; Burrows, J.P.; Buchwitz, M.; Frerick, J.; Noël, S.; Rozanov, V.V.; Chance, K.V.; Goede, A.P.H. SCIAMACHY: Mission objectives and measurement modes. *J. Atmos. Sci.* **1999**, *56*, 127–150. [[CrossRef](#)]
13. Leigh, L.; Shrestha, M.; Hasan, N.; Kaewmanee, M. Classification of North Africa for Use as an Extended Pseudo Invariant Calibration Site for Radiometric Calibration and Stability Monitoring of Optical Satellite Sensors. In Proceedings of the CALCON 2019, Utah State University, Logan, UT, USA, 19–21 September 2019.
14. Shrestha, M.; Leigh, L.; Helder, D. Classification of North Africa for use as an extended pseudo invariant calibration site (EPICS) for radiometric calibration and stability monitoring of optical satellite sensors. *Remote Sens.* **2019**, *11*, 875. [[CrossRef](#)]
15. Rueda, J.F.; Leigh, L.; Pinto, C.T.; Kaewmanee, M.; Helder, D. Classification and Evaluation of Extended PICS (EPICS) on a Global Scale for Calibration and Stability Monitoring of Optical Satellite Sensors. *Remote Sens.* **2021**, *13*, 3350. [[CrossRef](#)]
16. Remote Sensing Phenology. NDVI, the Foundation for Remote Sensing Phenology. USGS. 2018. Available online: <https://www.usgs.gov/special-topics/remote-sensing-phenology/science/ndvi-foundation-remote-sensing-phenology> (accessed on 25 October 2021).
17. Diek, S.; Fornallaz, F.; Schaepman, M.E.; De Jong, R. Barest Pixel Composite for Agricultural Areas Using Landsat Time Series. *Remote Sens.* **2017**, *9*, 1245. [[CrossRef](#)]
18. Landsat 8 Data Users Handbook. USGS. 2019. Available online: <https://www.usgs.gov/landsat-missions/landsat-8-data-users-handbook> (accessed on 25 October 2021).
19. Thenkabail, P.; Anece, I. Global Hyperspectral Imaging Spectral-library of Agricultural crops for Conterminous United States V001 [Data Set]. NASA EOSDIS Land Processes DAAC. 2019. Available online: <https://lpdaac.usgs.gov/products/ghisaconusv001> (accessed on 20 March 2021).
20. Meerdink, S.K.; Hook, S.J.; Roberts, D.A.; Abbott, E.A. The ECOSTRESS spectral library version 1.0. *Remote Sens. Environ.* **2019**, *230*, 111196. [[CrossRef](#)]
21. Baldrige, A.M.; Hook, S.J.; Grove, C.I.; Rivera, G. The ASTER Spectral Library Version 2.0. *Remote Sens. Environ.* **2009**, *113*, 711–715. [[CrossRef](#)]
22. World Agroforestry (ICRAF); International Soil Reference and Information Centre (ISRIC). ICRAF-ISRIC Soil VNIR Spectral Library. In *World Agroforestry—Research Data Repository, V1*; World Agroforestry (ICRAF): Nairobi, Kenya, 2021. [[CrossRef](#)]
23. Salvatori, R.; Salzano, R.; Franco, S.D.; Fontinovo, G.; Plini, P. *SISpec 2.0 Snow-Ice Spectral Library*; National Research Council of Italy, Institute of Polar Sciences: Rome, Italy, 2020.
24. U.S. Standard Atmosphere. National Oceanic and Atmospheric Administration. 1976. Available online: [https://www.ngdc.noaa.gov/stp/space-weather/online-publications/miscellaneous/us-standard-atmosphere-1976/us-standard-atmosphere\\_st76-1562\\_noaa.pdf](https://www.ngdc.noaa.gov/stp/space-weather/online-publications/miscellaneous/us-standard-atmosphere-1976/us-standard-atmosphere_st76-1562_noaa.pdf) (accessed on 14 March 2022).
25. Hartung, J.; Knapp, G.; Singa, B.K. *Statistical Meta-Analysis with Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2008; pp. 44–46.
26. Thuillier, G.; Hers, M.; Simon, P.C.; Labs, D.; Mandel, H.; Gillotay, D.; Foujols, T. The solar spectral irradiance from 200 to 2400 nm as measured by the SOLSPEC spectrometer from the ATLAS 1-2-3 and EURECA missions. *Sol. Phys.* **2003**, *214*, 1–22. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.