Article

Validation of Abundance Map Reference Data for Spectral Unmixing

McKay D. Williams [1],*, **Robert J. Parody** [2], **Alexander J. Fafard** [1], **John P. Kerekes** [1] **and Jan van Aardt** [1]

1 Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY 14623, USA; ajf4163@rit.edu (A.J.F.); kerekes@cis.rit.edu (J.P.K.); vanaardt@cis.rit.edu (J.v.A.)
2 School of Mathematical Sciences, Rochester Institute of Technology, Rochester, NY 14623, USA; rjpeqa@rit.edu
* Correspondence: mdw9466@rit.edu; Tel.: +1-435-764-6102

Abstract: The purpose of this study is to validate the accuracy of abundance map reference data (AMRD) for three airborne imaging spectrometer (IS) scenes. AMRD refers to reference data maps ("ground truth") that are specifically designed to quantitatively assess the performance of spectral unmixing algorithms. While classification algorithms typically label whole pixels as belonging to certain ground cover classes, spectral unmixing allows pixels to be composed of fractions or abundances of each class. The AMRD validated in this paper were generated using our previously-proposed remotely-sensed reference data (RSRD) technique, which spatially aggregates the results of standard classification or unmixing algorithms from fine spatial-scale IS data to produce AMRD for co-located coarse-scale IS data. Validation of the three scenes was accomplished by estimating AMRD in 51 randomly-selected 10 m × 10 m plots, using seven independent methods and observers. These independent estimates included field surveys by two observers, imagery analysis by two observers and RSRD by three algorithms. Results indicated statistically-significant differences between all versions of AMRD. Even AMRD from our two field surveys were significantly different for two of the four ground cover classes. These results suggest that all forms of reference data require validation prior to use in assessing the performance of classification and/or unmixing algorithms. Given the significant differences between the independent versions of AMRD, we propose that the mean of all (MOA) versions of reference data for each plot and class is most likely to represent true abundances. Our independent versions of AMRD were compared to MOA to characterize error and uncertainty. Best case results were achieved by a version of imagery analysis, which had mean coverage area differences of 2.0%, with a standard deviation of 5.6%. One of the RSRD algorithms was nearly as accurate, achieving mean differences of 3.0%, with a standard deviation of 6.3%. Further analysis of statistical equivalence yielded an overall zone of equivalence between $[-7.0\%, 7.2\%]$ for this version of RSRD. The relative accuracy of RSRD methods is promising, given their potential to efficiently generate scene-wide AMRD. These results provide the first known validated abundance level reference data for airborne IS data.

Keywords: reference data; ground truth; imaging spectroscopy; hyperspectral; unmixing; classification; abundance map; subpixel

1. Introduction

Since the widespread introduction of aerial imaging spectrometer (IS) data (hyperspectral imagery) in the 1990s [1,2], researchers have been developing processing algorithms to analyze and interpret the vast data produced by IS sensors [3,4]. Often, these algorithms are designed to produce labeled

maps where each pixel is assigned to one of a finite number of ground cover classes. Common ground cover classes for ecological research are photosynthetic vegetation (PV), non-photosynthetic vegetation (NPV), bare soil (BS), rock, etc. This type of processing, where each whole pixel is assigned to a class, is called classification [3,5].

Spectral unmixing is another method of processing imaging spectrometer data that is similar to classification, except that each pixel is allowed to be a mixture of the pure classes [6,7]. Instead of classifying whole pixels, each pixel is assigned to be a fraction or abundance of each class. For example, a given pixel could be 15% PV, 50% NPV and 35% BS. Since the ground sample distance (GSD) of imaging spectrometer data is often 10 m or greater [1,8], most of the pixels in a scene are composed of a mixture of classes. Spectral unmixing accounts for these mixed pixels and produces abundance maps, where each class has a corresponding abundance map that estimates the fraction of that class in each pixel. An example abundance map is shown in Figure 1. Abundances for each pixel range from zero (0%) to one (100%), represented by black and white, respectively.

(**a**)   (**b**)
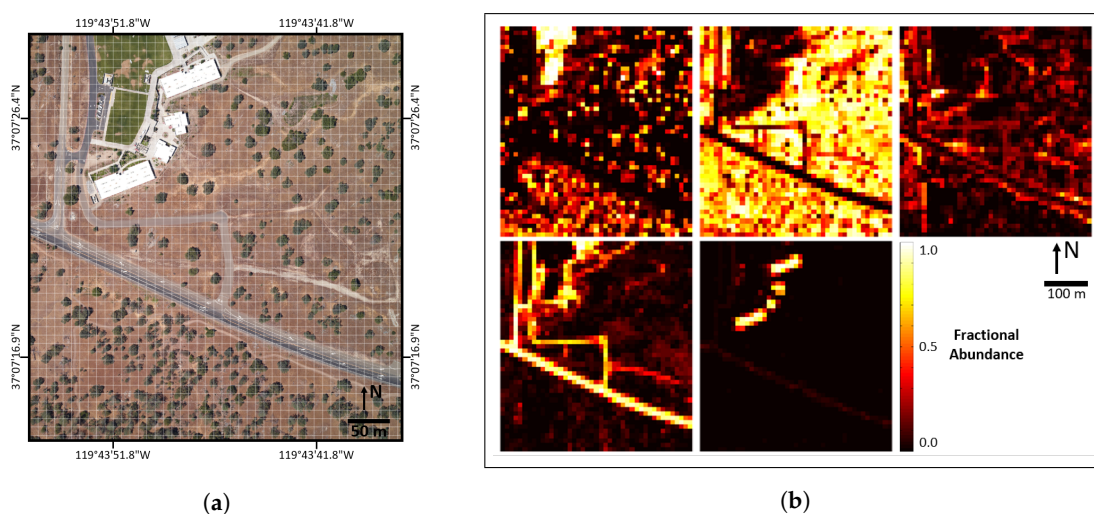
Figure 1. (**a**) RGB imagery of a remote sensing scene overlaid by a 10 m × 10 m grid representing generic imaging spectrometer pixels; (**b**) example abundance maps for for the following classes (left-to-right and top-to-bottom): PV, NPV, BS, rock, other (roof).

Reference data, or reference data maps, represent a specific type of map that is used to quantitatively assess how well a classification or spectral unmixing algorithm performs [9]. Reference data maps are usually created through field surveys, imagery analysis, algorithms or a mixture of these methods [10–12]. Reference data maps have often been called "ground truth" maps within the remote sensing community, but recently, there has been an effort to use reference data terminology [9]. We strongly support this change, because the term ground truth is overused and implies that the maps are free of error.

Due to the large spatial extent of aerial imaging spectrometer scenes, generating new reference data maps is expensive and time consuming [11,13]. As a result, few reference data maps have been created, even though imaging spectrometers have been operational for several decades. Cuprite [10], Indian Pines [14], Salinas [15] and Pavia [12] are examples of commonly-used reference data products. Despite being widely used for several decades to assess the performance of new classification algorithms, we have not been able to locate validation reports that characterize the error or uncertainty in these reference data. Additionally, these products label whole pixels as belonging to classes, making them of little use for assessing the performance of spectral unmixing algorithms.

Researchers have used various methods to assess the performance of spectral unmixing algorithms. Qualitative assessments have visually compared unmixing results to well-known reference data, such as Cuprite, Indian Pines and Pavia [7,16,17]. Quantitative assessments have used synthetic

data [7,16,17], average per-pixel residual error [7,18], taken the mean of several algorithms as the ideal and computed residuals from the ideal [18], etc. Furthermore, there are several studies that have created sub-pixel reference data using high-resolution RGB videography [19], RGB imagery [20], multi-spectral imagery [21] and IS imagery [22]. However, although many of these studies alluded to the need for the validation of sub-pixel reference data, none implemented an assessment approach of reference data, nor did they expand on which reference data development approach is best suited to this challenge. This research therefore further explores the topic of sub-pixel reference data generated via high-resolution imagery.

In summary, the main challenges with reference data are that new reference data maps are prohibitively expensive to produce; existing reference data products have not been validated for accuracy; and we are not aware of any commonly-used reference datasets that contain abundance level detail.

Our previous work [23] proposed a new technique for efficiently generating abundance map reference data (AMRD). This technique is called remotely-sensed reference data (RSRD). RSRD aggregates the results of standard classification or spectral unmixing algorithms on fine-scale imaging spectrometer data (e.g., 1-m GSD National Ecological Observatory Network (NEON) IS data) to generate scene-wide AMRD for co-located coarse-scale imaging spectrometer data (e.g., 15-m GSD Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) IS data), thereby enabling quantitative assessment of spectral unmixing algorithms. An example AVIRIS pixel grid pattern, overlaid on NEON imagery, is shown in Figure 2.
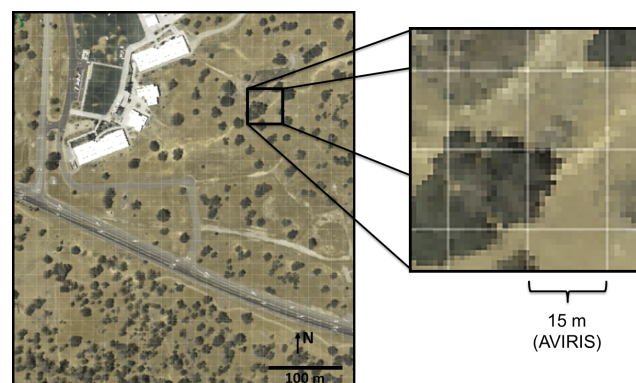
Figure 2. AVIRIS pixel grid pattern overlaid on a pseudo-RGB representation of NEON IS data.

The primary objective of this paper is to validate three AMRD scenes that were produced using the RSRD technique. The secondary objective of this paper is to promote the understanding that every method of generating reference data is vulnerable to certain types of error; reference data should not be used to assess classification or unmixing performance without a validation report that characterizes error and uncertainty in the reference data itself.

2. Methodology

2.1. Airborne Data and Cover Classes

The three scenes used throughout this paper are located on or adjacent to NEON D17 research sites near Fresno, CA [24]. Overhead images of the three scenes, collected by NEON's co-mounted framing RGB camera, are shown in Figure 3. Imagery over NEON D17 research sites was chosen for this study because NEON and AVIRIS conducted a joint campaign in June 2013, collecting imagery over the same areas on the same days [25]. The three specific scenes were chosen to be representative of a wide variety of remote sensing scenes. The selected scenes include asphalt roads, dirt roads, concrete, buildings, grass fields, dry valley grasslands, high mountain forests, large rock outcroppings, etc.

As such, the specific validation methods and conclusions in this paper should be generally applicable to many remote sensing scenes, especially environments dominated by rural or suburban features. It is worth noting that the intent of this research was to introduce the fine spatial resolution approach to coarser spatial resolution imaging spectroscopy pixel unmixing. Although we attempted to include both natural and man-made environments, we would not expect these scenes to be representative of dense urban environments.
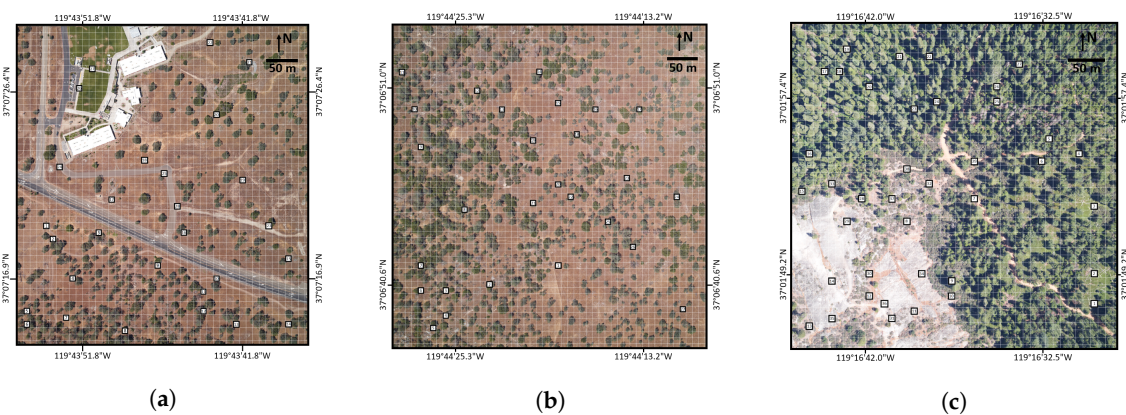
Figure 3. Overhead imagery of the three remote sensing scenes validated in this paper: (**a**) SJER High School; (**b**) SJER 116; and (**c**) SOAP 299. A generic 10 m×10 m grid is overlaid on the RGB photographs, showing the size of validation sampling plots. Randomly selected sampling plots are marked.

During the June 2013 campaign, NEON collected 0.25-m GSD RGB and 1-m GSD IS data and AVIRIS collected 15-m GSD IS data. Imagery was provided to us in the orthorectified reflectance data format, including nearest neighbor resampling to a north-south-oriented grid. Further data details are listed in Table 1. It is important to note that while we intend to produce AMRD specifically for AVIRIS data, the validation study in this paper was done for a generic 10-m GSD coarse-scale grid, allowing specific application for any imagery product with GSD larger than 10 m (i.e., AVIRIS, HyspIRI, Hyperion, EnMap, Landsat, etc.)

Table 1. NEON data details. AtmCal and OrthoRect refer to proprietary atmoshperic calibration and orthorectification respectively.

Data	GSD	Altitude	Scan Type	Spectra	Bands	Processing	Date
NEON RGB	0.25 m	1 km	Framing	0.4–0.7 μm	3	N/A	13 June 2013
NEON IS	1 m	1 km	Push broom	0.4–2.5 μm	428	AtmCal/OrthoRect	13 June 2013

Through analysis of RGB imagery of the three scenes, we chose a classification scheme similar to common ecological studies [26,27]. The following is a list of the ground cover classes we chose for our three scenes:

- Photosynthetic vegetation (PV): live green trees, bushes, weeds, grass, etc.
- Non-photosynthetic vegetation (NPV): dead brown trees, bushes, weeds, grass, etc.
- Bare soil (BS): undisturbed soil, sand, dirt trails, dirt roads, etc.
- Rock: rocks, concrete, asphalt, etc.
- Other: roof, metal, road paint, etc.

These classes represent the main ground constituents of our scenes, without over-complicating reference data generation and validation, i.e., they serve to illustrate the approach in this ecological context.

Historically, reference data maps have been generated using field surveys, imagery analysis, algorithms or a combination of multiple methods [15]. Each of these methods for generating reference

data has strengths and weaknesses. For example, field surveys provide the best situational awareness and ability to closely examine ground cover. However, they are subject to human error, such as bias, fatigue, mistakes, etc. Field surveys are also subject to perspective and positional error [9]. Imagery analysis eliminates perspective and positional error, but reduces situational awareness and retains human error. Algorithms eliminate perspective and positional error, and reduce human error, but remove human intuition and situational awareness. Taking this information into account, the optimal method of generating reference data is not clear and will be further explored in this paper.

An important question for our study is how to validate and characterize the error in reference data, when reference data are already the most accurate estimates available. Our approach is to produce multiple independent versions of reference data for the same plots on the ground using traditional methods and RSRD and then to compare the results. To our knowledge, this is the first validated reference data presented at abundance-level spatial detail.

Time and budget constraints permitted the compilation of seven independent versions of reference data for 51 10 m×10 m plots, spread across the three scenes. Sample plots were selected using a pseudo-random number generator. The reference data methods and observers are listed below (Supplementary S1):

1.　Field surveys by Observer-A (Field-A)
2.　Field surveys by Observer-B (Field-B)
3.　Imagery analysis by Observer-A on NEON IS, aided by NEON RGB (Analyst-A)
4.　Imagery analysis by Observer-B on NEON IS, aided by NEON RGB (Analyst-B)
5.　RSRD by Euclidean distance (ED) on NEON IS (RSRD-ED)
6.　RSRD by non-negative least squares (NNLS) spectral unmixing on NEON IS (RSRD-NNLS)
7.　RSRD by maximum likelihood (ML) on NEON IS (RSRD-ML)

The observers were both graduate students in imaging science, i.e., a degree of familiarity with imaging products and image interpretation was assured.

Validation and characterization of error was accomplished by comparing each of these data to each other, as well as comparing each version of reference data to the mean of all (MOA) data. Through this analysis, we examined whether certain datasets were statistically different from one another, estimated the mean and standard deviation of differences from MOA and used equivalence hypothesis tests to determine statistical similarity to MOA.

2.2. Reference Data Collection

Validation of RSRD as a credible technique for producing AMRD is dependent on compiling multiple independent versions of reference data that are each of the highest quality. As such, our team compiled seven independent versions of reference data for 51 of the selected 10 m×10 m plots shown in Figure 3. All but eight of the 51 plots were randomly selected, discarding plots that were located on busy roads, rooftops, etc. Eight non-randomly selected plots targeted under-represented cover types, such as asphalt, sidewalks and dirt roads. Due to time and budget constraints, private property issues, etc., the 51 plots were a subset of the original randomly selected plots.

The goal of reference data collection was to carefully estimate the abundance of ground cover classes in each of the 51 10 m×10 m plots, using field surveys, imagery analysis and RSRD methodologies. The following sections explain the details for each of these collection methodologies.

2.2.1. Field Surveys

Observers traveled to Fresno, CA, in summer 2016 and spent 12 days in the field measuring the abundance of ground cover classes in 51 10 m×10 m plots. The plots were spread nearly evenly across three scenes:

- San Joaquin Experimental Range High School (SJERHS): the area surrounding Minarets High School, which lies adjacent to SJER on the north site of the range.

- San Joaquin Experimental Range 116 (SJER116): the area surrounding NEON SJER field site #116.
- Soaproot Saddle 299 (SOAP299): the area surrounding NEON SOAP field site #299.

Potential plots were chosen prior to the trip and are shown in Figure 3. All of the potential plots were accessible, with the exception of SJERHS plots south of the main road, which are located on private, actively grazed, ranch land.

The landscape in SJERHS and SJER116 experienced little apparent change between 2013 imagery collection and 2016 field surveys; however, SOAP299 experienced a significant pine beetle infestation, resulting in the death of many trees in the area [28]. This forced us to abandon most of the forest-dominated plots in SOAP299, instead focusing on plots with little apparent change, such as those dominated by dirt roads, bush, rock and soil.

Approximate navigation to plot locations was accomplished using GPS-enabled smartphone mapping software and a compass. Fine placement of measurement grids was accomplished using detailed NEON RGB images of each plot, similar to Figure 4a. In other words, our 10 m×10 m sampling grids were positioned by matching in-scene objects, rather than relying on the unknown precision of field GPS readings and imagery georeferencing; we assumed that this approach limited geospatial inaccuracies.

In order to more accurately estimate plot level abundances, a rope grid was positioned on top of the 10 m×10 m plot, dividing the plot into 25 2 m×2 m samples. This is illustrated in Figure 4a. With the grid in place, the two observers independently estimated the abundance of ground cover classes in each of the 2 m×2 m samples. If the observers' estimates differed by more than 15% cover area for a given class, the sample was re-examined by both researchers to reduce errors. Estimates were recorded on data collection sheets, as shown in Figure 4b. In addition to abundance estimates for each 2 m×2 m sample, photos were captured showing the entire plot, near nadir photos were captured of each 2 m×2 m sample, GPS coordinates of plot center locations were measured, and a library of spectral samples was compiled using a Spectra Vista Corporation (SVC) field portable spectrometer. Paper data collection sheets and other data were compiled and recorded in a spreadsheet each evening so that errors and ambiguities could be resolved before forgetting important information.


| (a) | (b) |
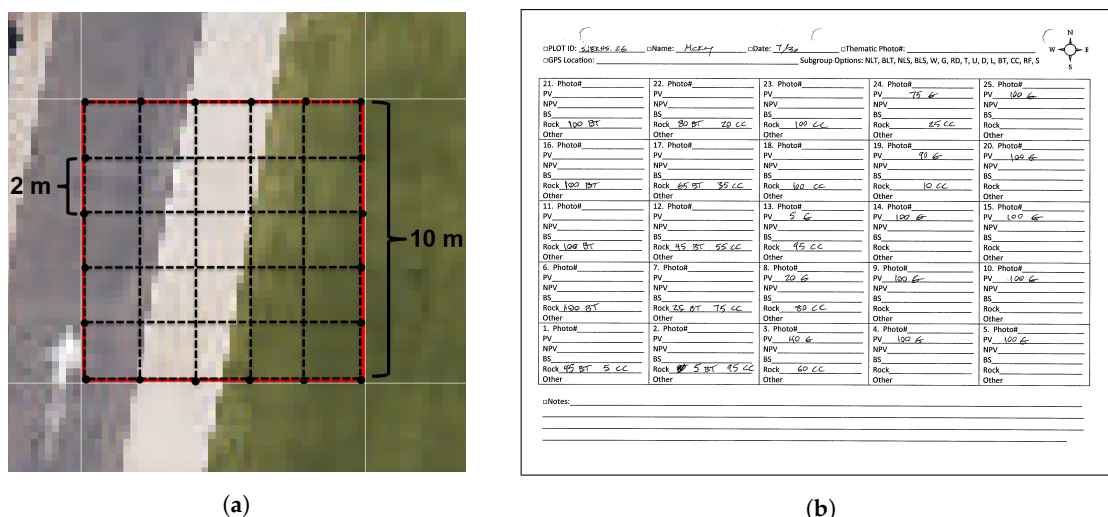
Figure 4. (**a**) An illustration of the sampling grid setup for one of the 10 m×10 m plots; (**b**) an example data collection sheet; one of these was filled out by each observer for each of the 51 10 m×10 m plots. Note that abundances were estimated by each observer for 25 2 m×2 m samples within each plot.

2.2.2. Imagery Analysis

Imagery analyst-derived reference data were generated, corresponding to the same 51 plots estimated by the observers in the field. Instead of estimating fractional abundances for 25 2 m×2 m

samples, as was done in the field, analyst reference data classified 100 1 m×1 m NEON IS whole pixels within each plot to one of the cover classes. Imagery analysis was performed several months after completing the field surveys, and field survey data were not consulted during imagery analysis. Analysis was accomplished in the ENVI 5.3 environment using pseudo-RGB visualizations of NEON IS data and NEON RGB imagery. Figure 5 illustrates what the analysts saw as they classified the plots. This figure shows a hybrid view, where a pseudo-RGB view of NEON IS is 20% transparent on top of NEON RGB. The analyst was able to toggle between the hybrid view shown here, NEON IS and NEON RGB in order to make the best classification decision for the NEON IS pixels.

(**a**)　　　　　　　　　　　　　　　　　　　　　　(**b**)
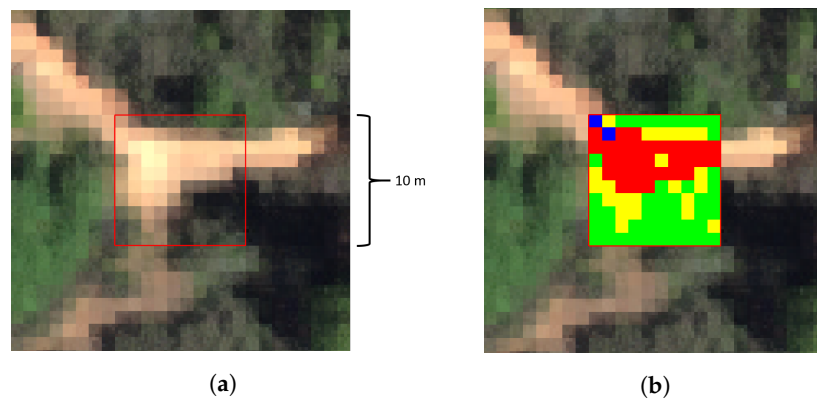
Figure 5. (**a**) An example 10 m×10 m NEON IS plot to be classified by imagery analysis with the aid of NEON RGB; (**b**) a classified 10 m×10 m plot (PV, green; NPV, yellow; BS, red; rock, blue). Note that 100 whole 1 m×1 m pixels were classified by the analyst per 10 m×10 m plot.

2.2.3. Remotely-Sensed Reference Data

Euclidean distance (ED) [3], NNLS unmixing [29] and ML [3] were selected to estimate scene-wide AMRD using the RSRD technique. We chose these algorithms because of their widespread adoption, established reputation, and relative simplicity. This study was not intended to identify the best possible algorithm, i.e., we opted to use established methods to evaluate reference data outcomes.

Endmembers were extracted for each scene separately by selecting ten exemplar pixels for each cover class and taking their spectral mean. We opted for this approach, rather than algorithmic endmember extraction techniques, in order maintain human analyst control and interpretative intent over class representative spectra. The same endmembers were used for ED, NNLS and ML within each scene.

For ED, which is equivalent to a nearest-neighbor classifier, NEON IS pixels were evaluated using Equation (1), where x represents the spectrum of a pixel, s is an endmember, and n is the number of spectral bands. Pixels were assigned to the class whose endmember yielded the smallest distance. For NNLS, NEON IS pixels were represented by the linear mixture model in Equation (2), where a are abundances, n is noise, and M is the number of classes. Lawson's implementation of non-negative least squares was used to solve for a [29]. It should be noted that while the NNLS algorithm does not impose a sum-to-one constraint, we later imposed this constraint on resulting abundances, such that ED, NNLS and ML methods all provided sum-to-one constrained solutions. For ML, initial prior probabilities, $\hat{P}(s)$, were obtained via spectral angle mapper (SAM) classification (Equation (3)), then results were processed for ten iterations using Equation (4), where s represents the various ground cover classes. Final posterior probabilities were used as abundances. RSRD was performed on the entire SJERHS, SJER116 and SOAP299 scenes, with results corresponding to the 51 10 m×10 m plots extracted for comparison to field and analyst estimated reference data.

$$r_{ED}(\mathbf{x}) = ||\mathbf{x} - \mathbf{s}|| = \sqrt{(x_1 - s_1)^2 + (x_2 - s_2)^2 + ... + (x_n - s_n)^2} \qquad (1)$$

$$x_{NNLS} = \sum_{m=1}^{M} a_m s_m + n, \quad \text{and} \quad a_m >= 0 \tag{2}$$

$$r_{SAM}(\mathbf{x}) = -cos^{-1}\left(\frac{\mathbf{s}^T\mathbf{x}}{\sqrt{(\mathbf{s}^T\mathbf{s})(\mathbf{x}^T\mathbf{x})}}\right) \tag{3}$$

$$x_{ML} = \hat{P}(s|x) = \frac{\hat{P}(s)\hat{p}(x|s)}{\sum_{s=1}^{S} \hat{P}(s)\hat{p}(x|s)} \quad \text{and} \quad p(x|s) \sim N(\mu_s, \Sigma_s) \tag{4}$$

Figure 6 shows an example of fine-scale abundances generated using ED and NNLS. Both results are presented in abundance map format, even though ED assigns whole pixels to one class or another. For this reason, ED abundance maps are black and white, representing 0 (0%) or 1 (100%) coverage, while NNLS abundance maps have values ranging continuously from 0 to 1.

(**a**)   (**b**)
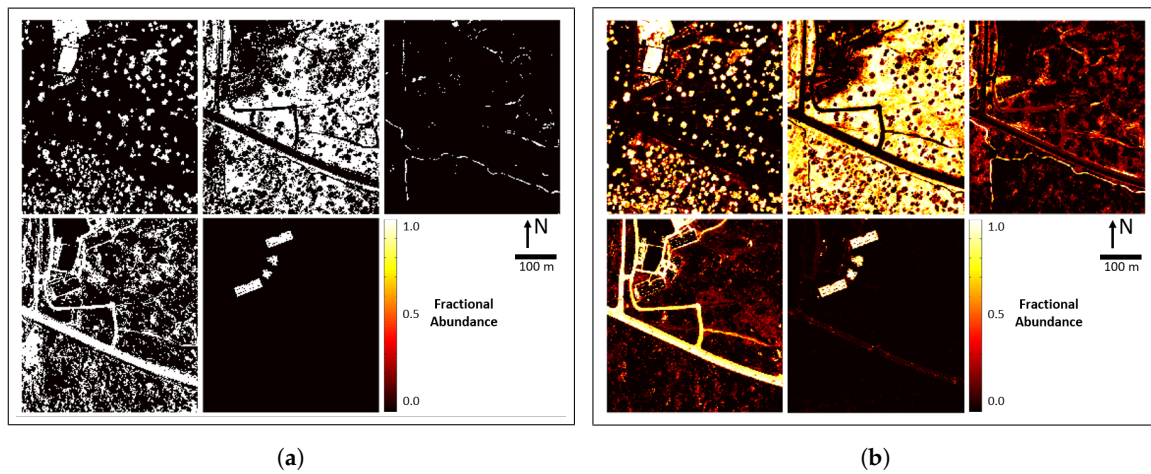
Figure 6. Example fine-scale SJERHS scene-wide abundance map estimation of 1-m GSD NEON IS data for classes (left-to-right and top-to-bottom): PV, NPV, BS, rock and other (roof) via: (**a**) RSRD-ED (note that data are either 0 or 1, hence the binary color scale); and (**b**) RSRD-NNLS.

2.3. Data Aggregation

Field-A and Field-B data estimated abundances at the 2-m scale. Analyst-A, Analyst-B, RSRD-SAM and RSRD-ED classified whole NEON IS pixels at the 1-m scale. RSRD-NNLS and RSRD-ML estimated abundances at the 1-m scale. In order to directly compare field survey data with the other methods, 1-m scale data were aggregated to the 2-m scale through simple averaging of four 1-m samples. Further aggregation yielded 6 m and 10 m-scale data. Data aggregation is illustrated in Figure 7. Note that aggregation at the 6-m scale excluded edge data.
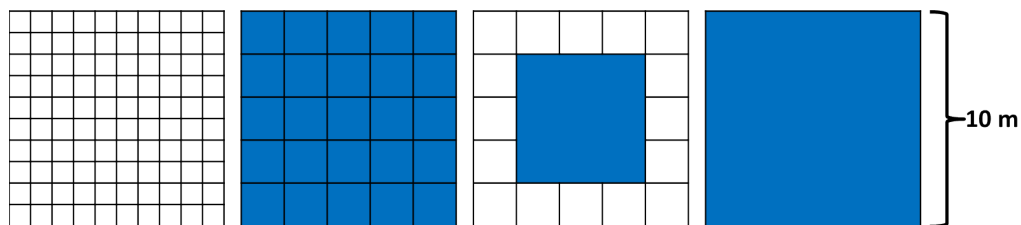
Figure 7. An illustration of how fine-scale samples were aggregated to produce coarse scale samples (left-to-right): 100 unaggregated 1-m samples within a 10 m×10 m plot, 1-m samples aggregated to the 2-m scale, 1-m and 2-m samples aggregated to the 6-m scale, 1-m and 2-m samples aggregated to the 10-m scale.

Aggregating data and comparing results allowed us to build confidence that the RSRD methodology can be utilized to provide quality AMRD for any medium to large GSD imagery. As individual samples were aggregated to the 2-m, 6-m and 10-m scales, the variance of abundance estimates decreased, showing that the RSRD methodology, applied to 1-m NEON IS data, can be used to generate AMRD for any imagery with GSD greater than or equal to 10 m.

2.4. Methods of Data Comparison

Abundance map reference data estimates from the seven methods discussed above (Field-A, Field-B, Analyst-A, Analyst-B, RSRD-ED, RSRD-NNLS, RSRD-ML) were compared at the 2-m, 6-m and 10-m spatial aggregation scale. At the 2-m scale, there were 1275 data points for each of the seven versions of reference data, while 6-m and 10-m scale data each had 51 data points. Pairwise comparisons were made to evaluate how close each version was to the other versions. Each version was also compared to the mean of all (MOA) versions per plot and class. In making these comparisons, data were often subtracted from each other to produce difference data. Difference data were evaluated using histograms, statistical measures such as mean and standard deviation of differences, *t*-tests and equivalence tests.

When using a *t*-test, the intention is to reject the null hypothesis, with the null hypothesis being that all data or treatments are the same. In other words, *t*-tests are intended to demonstrate differences between datasets or treatments [30]. The purpose of using *t*-tests in this paper was to show that independent versions of reference data exhibit statistically significant differences from one another and, as such, need to be validated rather than simply assuming "ground truth".

Failing to reject the null hypothesis does not prove similarity, but provides inconclusive results. Equivalence tests are similar to *t*-tests, but are used to prove similarity. When using an equivalence test, the intention is still to reject the null hypothesis, but in this case, the null hypothesis is that data are different. Rejecting the null hypothesis in this case proves statistical equivalence. Equivalence tests require establishing a zone of equivalence, $[\theta_0 - \delta, \theta_0 + \delta]$. An equivalence test then finds the probability p that $\mu \notin [\theta_0 - \delta, \theta_0 + \delta]$. If $p < \alpha$, the data are proven equivalent with $100(1 - 2\alpha)\%$ confidence [31]. The purpose of using equivalence tests in this paper was to define equivalence zones wherein versions of reference data were statistically equivalent [31] to MOA.

In practice, equivalence tests are performed by calculating confidence intervals for the difference between two datasets; if the entire confidence interval falls with the zone of equivalence, the two datasets are deemed statistically equivalent. It should also be noted that we want reference data to have both small mean and small standard deviation differences from one another; however, it is easier to "pass" statistical tests when data have a large standard deviation. For this reason, the mean and standard deviation of difference data are presented along with *t*-tests and equivalence tests.

3. Results

3.1. Comparison of Field Survey Data

Histograms of the difference between Field-A and Field-B data at 2-m aggregation are presented in Figure 8. These histograms highlight the frequency with which the observers differed by more than 15%, despite the careful collection procedures, which mandated re-assessment of any such samples. These histograms also demonstrate that difference data between the two field surveys were relatively Gaussian in shape, allowing for comparison using metrics such as the mean and standard deviation.

The mean and standard deviation of the data in Figure 8 are presented in Figure 9, along with the same data aggregated to 6 m and 10 m. Mean differences between samples at 2-m and 10-m aggregation ranged from −3.4 to 2.3%, depending on cover class. Mean differences only changed at the 6-m scale because edge samples were excluded in 6-m aggregation, as shown in Figure 7. On average, Field-B data overestimated NPV and underestimated PV and BS when compared to Field-A data. Differences for rock were near zero mean.
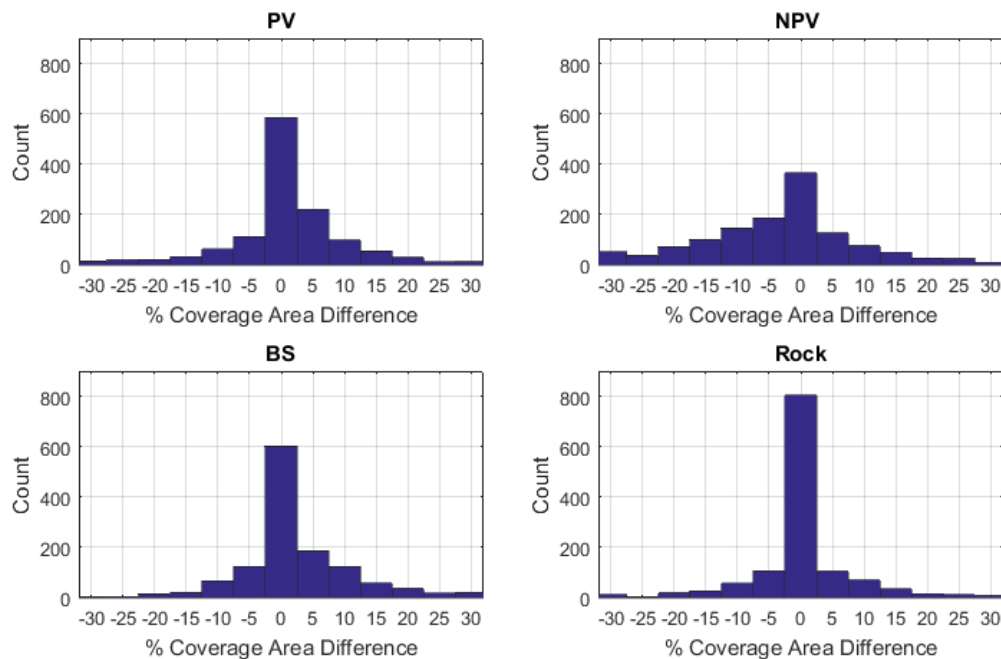
Figure 8. Histograms of differences between abundance estimates from Field-A and Field-B observers for PV, NPV, BS and rock (Field-A 2 m samples − Field-B 2 m samples).

| (**a**) | (**b**) |
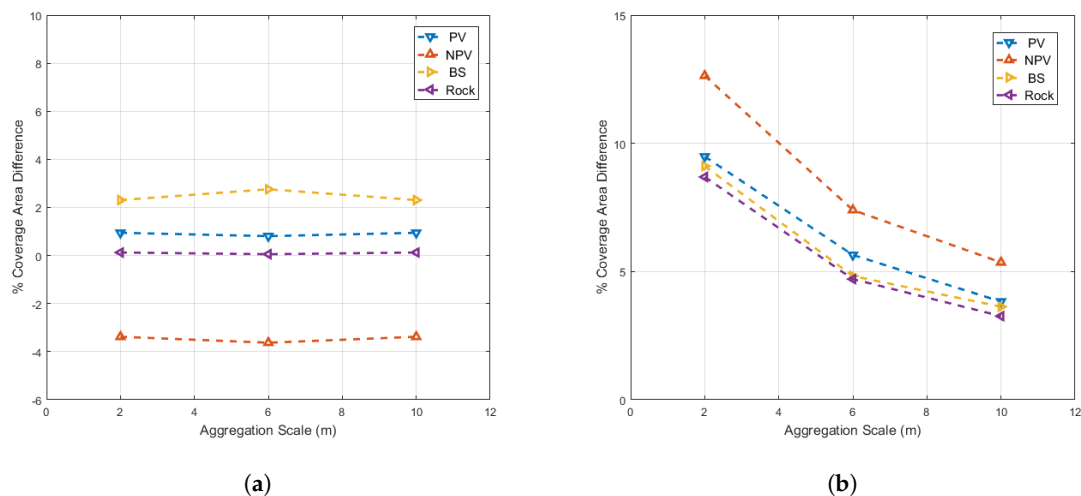
Figure 9. (**a**) Mean and (**b**) standard deviation of differences between Field-A and Field-B data for PV, NPV, BS and rock for 2-m, 6-m and 10-m aggregation scales (Field-A data − Field-B data).

As expected, aggregation to 6 m and 10 m data significantly reduced the variance of difference data. This reduction in variance occurred similarly for field survey, imagery analysis and RSRD-derived reference data, as they were aggregated from fine to coarse scale. The standard deviations of 2-m aggregation samples were 8.7 to 12.6%, depending on cover class, with NPV having the largest standard deviations. Not coincidentally, NPV is also the most abundant cover class in the scenes. When 2-m samples were aggregated to 6-m and 10-m plot sizes, the standard deviation decreased consistently. At 10-m aggregation, standard deviations dropped to 3.3 to 5.4%.

A *t*-test rejects the null hypothesis when $p < \alpha$. α, representing type I error, is typically set to 0.05 [32]. In this case, the null hypothesis was that difference data had a zero mean. Table 2 displays paired *t*-test results for Field-A and Field-B data, along with the standard deviation of differences (σ).

The results indicate that statistically significant differences exist between the field observers' estimates for all spatial levels of NPV and BS and PV at the 2-m scale.

Table 2. *t*-test and standard deviation of differences between Field-A and Field-B data. Results are bold when $p < 0.05$, representing rejection of the null hypothesis, i.e., significant differences between reference data versions.

Difference Data	PV (p)	NPV (p)	BS (p)	Rock (p)	PV (σ)	NPV (σ)	BS (σ)	Rock (σ)
Field-A/Field-B (2 m)	**4e-4**	**6e-21**	**6e-19**	0.61	9.48	12.64	9.11	8.69
Field-A/Field-B (6 m)	0.31	**1e-3**	**2e-4**	0.93	5.65	7.40	4.83	4.71
Field-A/Field-B (10 m)	0.08	**4e-5**	**4e-5**	0.78	3.83	5.36	3.63	3.26

3.2. Pairwise Comparisons of Reference Data Versions

In the previous section, Field-A and Field-B data were compared in some detail, including Table 2, which showed *t*-test results and standard deviations between the two datasets. In this section, we expand our pairwise comparisons to all seven versions of reference data and produce a similar table of *t*-test results and standard deviations for pairwise comparisons at the 10-m aggregation scale. Table 3 displays these results, with bold values indicating $p < 0.05$, where the *t*-test rejected the null hypothesis, indicating significant difference between data.

Table 3. *t*-test results (p) and standard deviation (σ) of the differences between all versions of reference data at the 10-m aggregation scale. Results are bold when $p < 0.05$, indicating rejection of the null hypothesis, i.e., significant differences between reference data versions.

Difference Data	PV (p)	NPV (p)	BS (p)	Rock (p)	PV (σ)	NPV (σ)	BS (σ)	Rock (σ)
Field-A/Field-B	0.08	**4e-5**	**4e-5**	0.78	3.83	5.36	3.63	3.26
Field-A/Analyst-A	0.06	**2e-5**	**2e-5**	0.34	13.17	18.55	11.43	8.49
Field-A/Analyst-B	**2e-3**	0.36	**0.01**	0.33	11.86	12.72	8.09	5.48
Field-A/RSRD-ED	0.23	0.72	**6e-9**	**0.01**	15.84	17.90	9.89	16.26
Field-A/RSRD-NNLS	0.06	0.35	**2e-6**	**5e-4**	11.71	13.27	9.78	11.56
Field-A/RSRD-ML	0.12	**1e-3**	**7e-4**	0.51	15.15	19.17	9.85	11.71
Field-B/Analyst-A	**6e-3**	**2e-7**	**1e-7**	0.30	11.20	18.44	11.18	8.66
Field-B/Analyst-B	**3e-5**	**1e-3**	0.41	0.37	9.90	10.43	6.64	4.99
Field-B/RSRD-ED	0.07	0.27	**7e-8**	**2e-3**	14.04	15.83	8.35	15.35
Field-B/RSRD-NNLS	**4e-3**	**1e-3**	**6e-5**	**9e-5**	9.57	10.69	8.21	10.18
Field-B/RSRD-ML	0.22	**0.03**	**0.03**	0.56	13.88	18.36	8.38	11.70
Analyst-A/Analyst-B	0.02	**2e-5**	**6e-8**	0.11	7.73	15.92	11.75	8.28
Analyst-A/RSRD-ED	0.54	**6e-6**	**3e-14**	0.05	9.76	18.35	11.62	17.51
Analyst-A/RSRD-NNLS	0.67	**1e-4**	**1e-10**	**6e-3**	7.21	17.61	12.95	11.99
Analyst-A/RSRD-ML	**2e-5**	**3e-9**	**3e-9**	0.23	10.55	21.00	12.30	13.03
Analyst-B/RSRD-ED	0.07	0.19	**8e-8**	**3e-3**	10.59	13.85	7.55	15.43
Analyst-B/RSRD-NNLS	**8e-3**	0.92	**2e-4**	**2e-5**	6.08	8.73	7.49	10.04
Analyst-B/RSRD-ML	**1e-5**	**6e-5**	0.08	0.82	12.93	17.40	7.53	10.54
RSRD-ED/RSRD-NNLS	0.68	0.11	0.05	0.94	6.99	11.86	8.42	13.25
RSRD-ED/RSRD-ML	**3e-6**	**3e-4**	**4e-4**	**8e-4**	8.32	15.05	8.86	14.41
RSRD-NNLS/RSRD-ML	**9e-6**	**3e-5**	0.09	**1e-3**	9.36	16.70	9.85	14.55

Each reference data version was compared to six other versions with four ground cover classes, resulting in 24 *t*-tests. The null hypothesis was rejected for at least half of the 24 *t*-tests for all reference data versions. Comparing field survey and imagery analysis methods only, the null hypothesis was rejected for 13 of the 24 *t*-tests. These results showed that our independent versions of reference data were significantly different from one another and that no single method stood out as the clear winner.

3.3. Comparison of Reference Data Versions to the Mean of All

Results from pairwise comparisons showed that various forms of carefully estimated reference data were significantly different. Furthermore, it was not clear which version of reference data was superior. Given these findings, we propose that for each plot and class, the mean of all (MOA) versions of reference data is more likely to represent true ground cover class abundances than any single method or observer. In this section, all reference data versions are compared to MOA, in order to determine which reference data approach was closest to the best available estimate of true abundances.

In order to better visualize the MOA concept, we present Figure 10, which shows an example plot, along with abundance estimates from each version of reference data and MOA for each class. The variance among abundance estimates in Figure 10 was near the median of our 51 sample plots. Difference from MOA data was calculated by subtracting MOA for each plot and class from the various reference data versions.


(**a**)　　　　　　　　　　　　　　　(**b**)
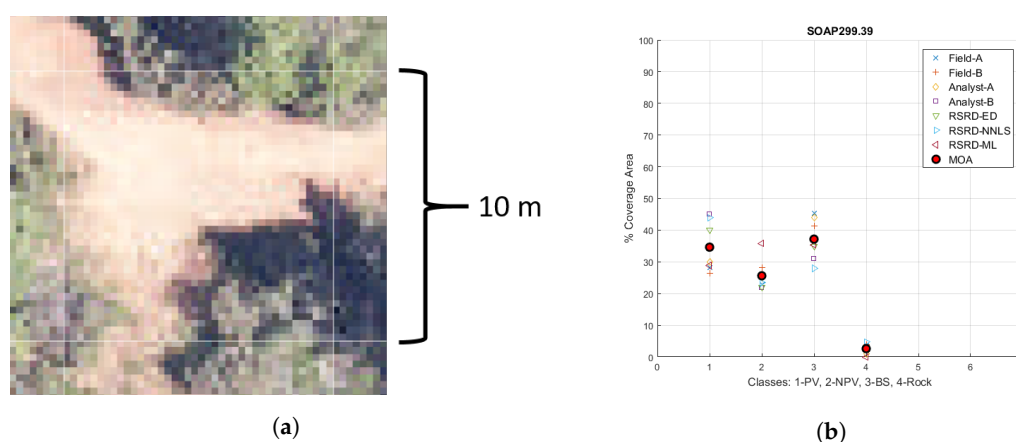
Figure 10. (**a**) NEON RGB view of plot SOAP299.39; (**b**) reference data estimates and MOA for plot SOAP299.39.

One way to explore which reference data method is closest to MOA would be to display difference data histograms, similar to Figure 8, for each combination of different reference data methods and aggregation scales. However, we can also represent much of the information in a histogram by finding the center (mean) and spread (standard deviation) of difference data. To further streamline data presentation, we can combine mean and standard deviation data across classes. These data are presented in Figure 11. Figure 11a displays the result of calculating the mean of difference data across 51 plots (histogram distribution center locations per class), then calculating the absolute value (otherwise, classes would cancel each other out) and then extracting the mean across classes. Figure 11b displays the results of calculating the standard deviation of difference data across 51 plots (histogram distribution spread per class), followed by extracting the mean across classes.

Both plots in Figure 11 are scaled from 0 to 25% coverage area to emphasize that mean differences from MOA (histogram distribution center locations) were relatively small compared to standard deviation differences. Examination of Figure 11b suggests that Field-B, Analyst-B and RSRD-NNLS reference data estimates were more precise than Field-A, Analyst-A, RSRD-ED and RSRD-ML.
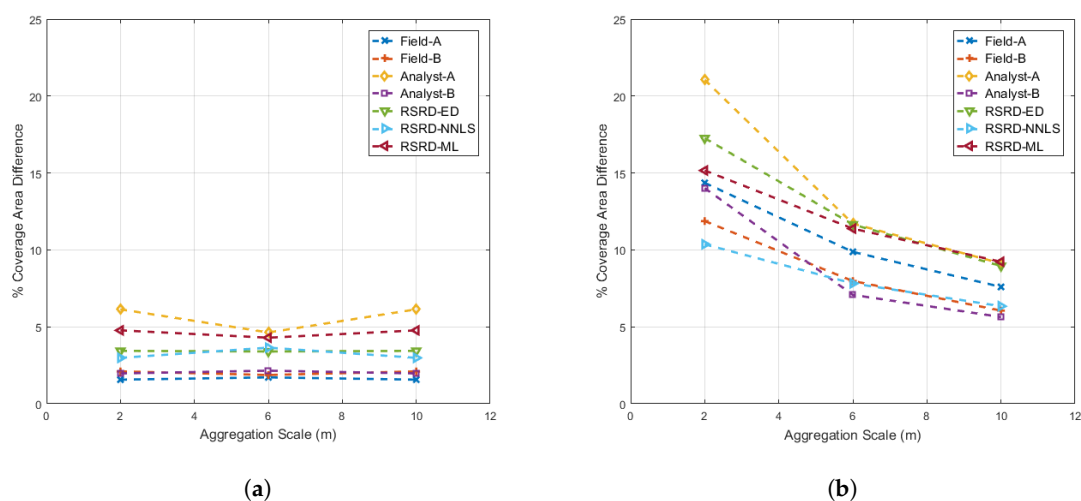
Figure 11. (**a**) Mean and (**b**) standard deviation of absolute differences from MOA for 2-m, 6-m and 10-m aggregation scales. Note that spatial aggregation of data do not affect the mean difference from MOA, but greatly reduce the standard deviation.

3.4. Statistical Equivalence Tests

MATLAB R2016a's ttest function provides confidence intervals that could be used for equivalence testing; however, we modeled all versions of our reference data within the Statistical Analysis Software (SAS OnDemand) general linear model (GLM) environment to produce more robust confidence intervals. The results of this analysis are provided in Figure 12. It should be noted that for modeling purposes in the SAS GLM environment, we compared each reference data version to a linear combination of all other versions. This "leave-one-out" approach is slightly different than comparison to MOA, as MOA includes the data being compared. For this reason, SAS GLM confidence intervals are not perfectly centered on mean distance from MOA results displayed in subsequent tables.

The ideal confidence interval result would be narrow intervals centered on zero. Our results showed that the accuracy of reference data versions varied considerably across classes, e.g., Analyst-A data were the best version of reference data for rock, but were the worst version for NPV and BS. Among RSRD versions, RSRD-NNLS had the overall tightest zone of equivalence, with RSRD-ED a close second. Tabulated confidence interval data and the resulting equivalence zone are provided for RSRD-NNLS data in Table 4. Note that the overall equivalence zone is simply the extent of extreme lower and upper CI values.

Table 4. Confidence intervals and zone of equivalence for RSRD-NNLS, expressed in terms of percent coverage area difference from MOA. Bold CIs defined the limits of the zone of equivalence.

Class	Lower CI (%)	Upper CI (%)
PV	−0.1	3.8
NPV	−4.4	1.1
BS	**−7.0**	−3.7
Rock	3.0	**7.2**
Equivalence Zone	−7.0	7.2

(**a**)

(**b**)

(**c**)

(**d**)
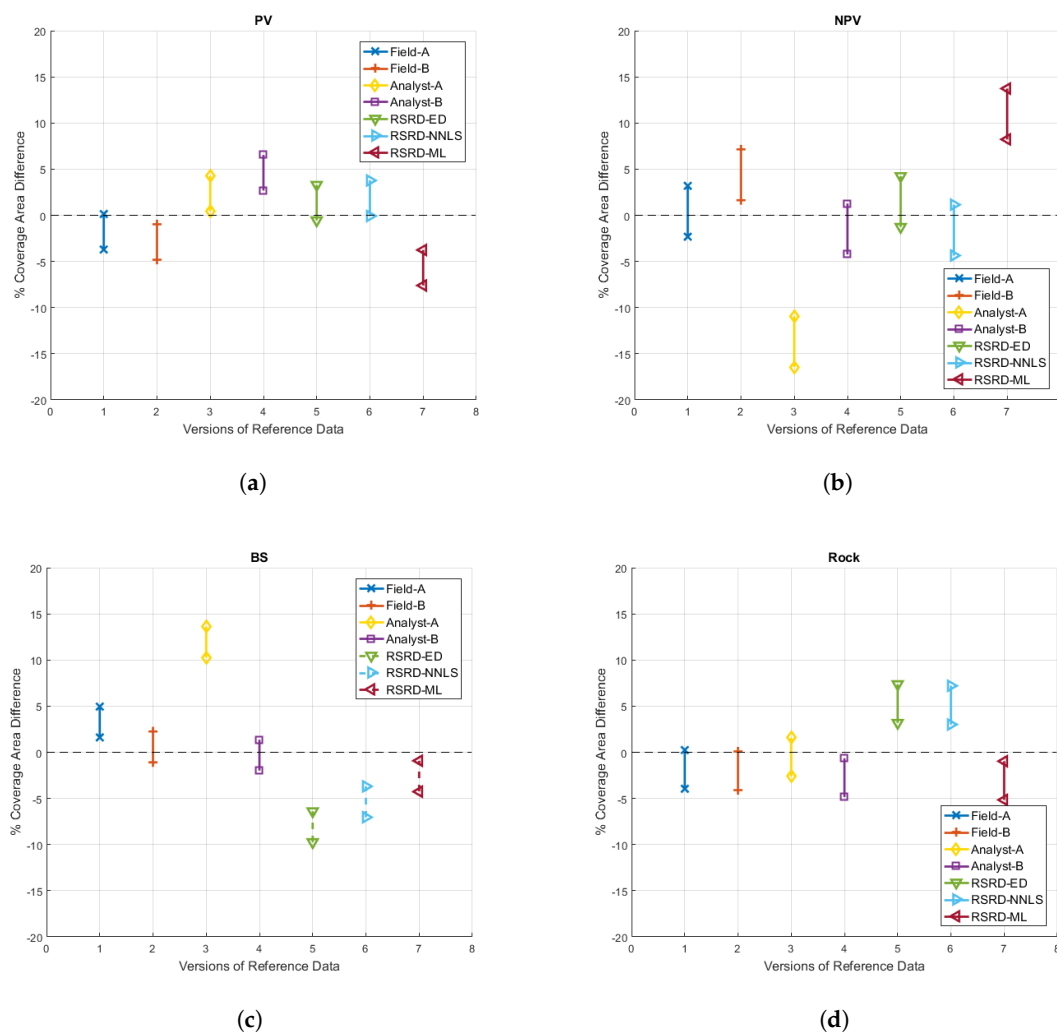
Figure 12. Confidence intervals for the mean difference between individual versions of reference data and a linear combination of all other versions of reference data for: (**a**) photosynthetic vegetation (PV); (**b**) non-photosynthetic vegetation (NPV); (**c**) bare soil (BS); and (**d**) rock.

4. Discussion

As noted before, both field observers were graduate students in imaging science, with an emphasis in remote sensing, who carefully and independently estimated reference data for 1275 2 m×2 m samples, within 51 10 m×10 m plots. Their estimates were collected at the same time and compared in the field after every fifth sample to reduce the chances of human error. If any percent coverage estimate varied by 15% or more for any class, they each examined the 2 m×2 m sample again. Data were transferred to electronic form each evening with errors being corrected on the same day as data collection. Our field surveys were assumed to be of the highest quality that can reasonably be collected, and procedures were in place to reduce human error. Yet, despite this careful estimation of reference data, statistically significant differences existed between Field-A and Field-B data at 2-m, 6-m and 10-m spatial aggregation scales. Further examination revealed statistically significant differences between all our versions of reference data, and traditional methods of field surveys and imagery analysis faired no better than RSRD methods.

It should be noted that the standard deviation of differences between Field-A and Field-B data were smaller than for the other pairwise comparisons. This may indicate that field surveys were the most dependable way to produce accurate reference data. However, it should also be noted that the field survey data may have been less independent than the other versions. The observers worked in

tandem throughout the field surveys. While they each independently estimated all 1275 2 m×2 m samples, the collection procedures included comparing results every five samples in an effort to reduce errors and mistakes. This process of checking answers inevitably had some effect on abundance estimation decisions.

The *t*-test results in this paper are a reminder that even carefully generated reference data are not perfectly accurate and should be validated prior to use in assessing other algorithms. In the past, remote sensing algorithm papers generally have not discussed the accuracy of reference data and often have not cited how reference data were created. Given the results in this paper, we encourage a more robust treatment of reference data in the remote sensing community.

Having used *t*-tests to demonstrate significant differences between versions of reference data, in order to reinforce the importance of accuracy validation, we now turn our attention toward determining the degree of similarity among reference data. We accomplished this by examining the mean and standard deviation of differences, as well as confidence intervals and resulting zones of equivalence.

We interpreted mean differences as absolute bias between different forms of reference data, i.e., Observer-A "saw" less NPV on average than Observer-B. These biases between reference data versions and MOA were relatively similar across reference data versions. Standard deviation of differences from MOA, on the other hand, were interpreted as the steadiness or reliability of each reference data estimate, i.e., Observer-B's estimates were reliably close to MOA and there were few outliers. These mean and standard deviation differences can be equated to accuracy and precision, respectively. Spatial aggregation of reference data from 2 m to 10 m did not decrease mean differences; however, aggregation significantly and predictably reduced standard deviation differences.

This reduction in variance with spatial aggregation was a positive trend, showing the power of the RSRD concept of aggregating many fine-scale pixels to produce AMRD for coarse-scale imagery. We expect that aggregation to coarse imagery with GSD >10 m would result in little change to mean differences, but would further reduce standard deviation differences. As such, we consider the validation in this paper to be valid for any coarse imagery with GSD >10 m.

Statistical analysis of difference data revealed that Field-B, Analyst-B and RSRD-NNLS versions of reference data were the closest versions to MOA. Analyst-B appeared to represent the best overall version of reference data when compared to MOA, with a mean difference of 2.0% and a standard deviation of 5.6% at 10-m aggregation. RSRD-NNLS was nearly as accurate and precise, with a mean difference of 3.0% and standard deviation of 6.3%. Furthermore, RSRD-NNLS also had the tightest overall zone of equivalence among RSRD versions, $[-7.0\%, 7.2\%]$.

It should be noted here that the generation of MOA required significant effort for just 51 plots of data and would be unrealistic across large spatial extents. Similarly, the generation of scene-wide AMRD using traditional field survey or imagery analysis methods would be unsustainable. As such, the MOA concept is used to validate the accuracy of scene-wide AMRD generated using the RSRD technique. Based on standard deviation differences and zones of equivalence, we selected RSRD-NNLS to serve as validated scene-wide AMRD.

We considered using the best case version of RSRD for each ground cover class in an effort to reduce the expected error and uncertainty in the final reference data. However, constructing scene-wide AMRD from different algorithms would introduce sum-to-one error (an assumption in many unmixing algorithms) in the final abundances, and correcting this would introduce another source of error.

Using the NNLS implementation of RSRD to produce scene-wide AMRD for our three scenes yields Figures 13–15. The accuracy of these reference data has been validated in this paper, resulting in the data provided in Table 5.
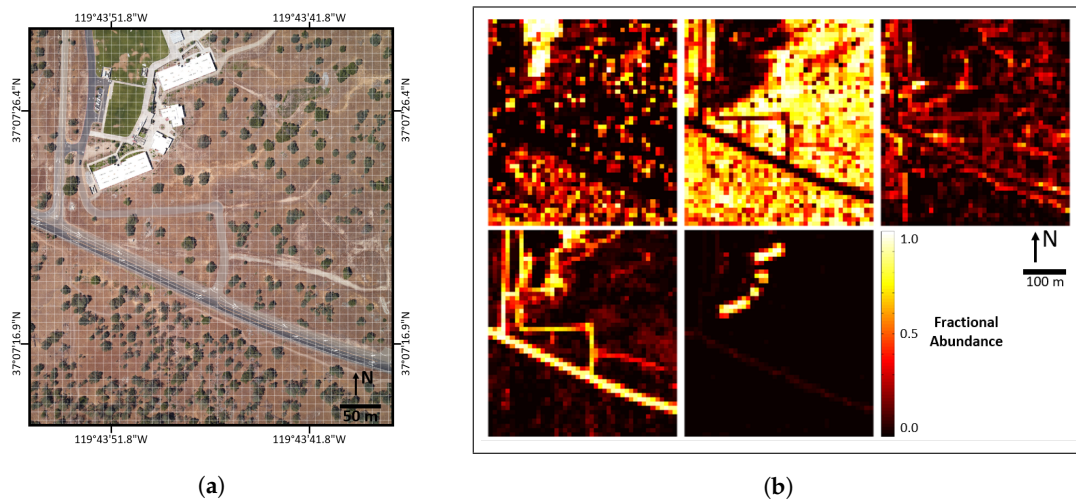
Figure 13. (**a**) RGB imagery SJERHS; (**b**) validated abundance map reference data (AMRD) at the 10 m×10 m aggregation scale for SJERHS, with ground cover classes (left-to-right and top-to-bottom): PV, NPV, BS, rock and other (roof).
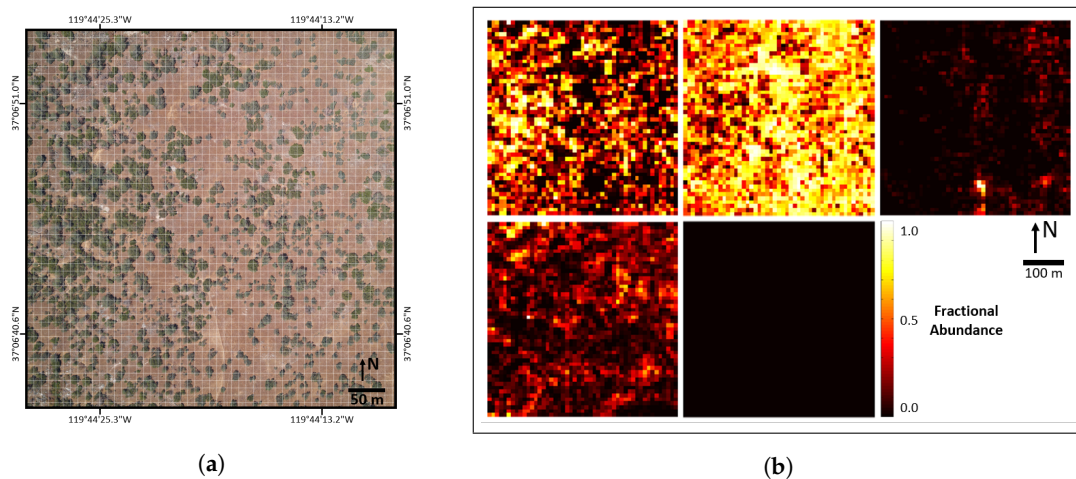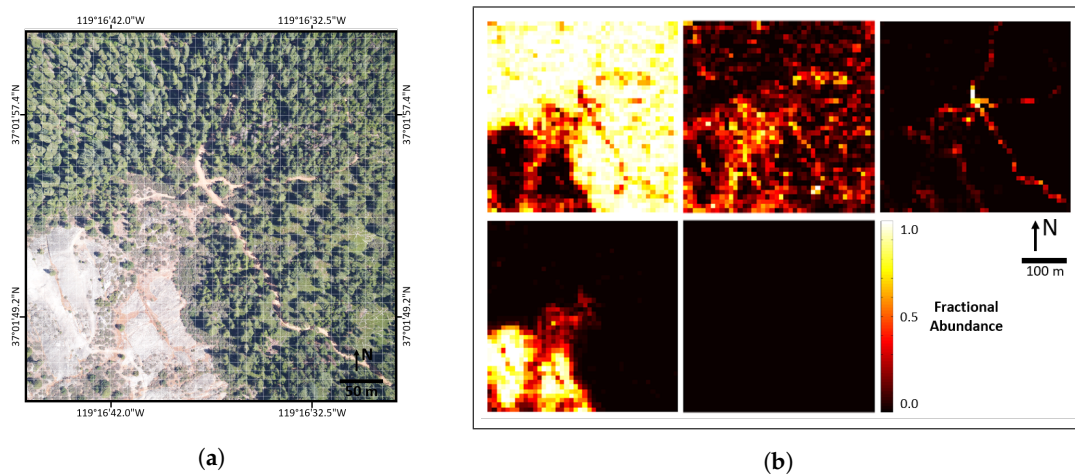

Figure 14. (**a**) RGB imagery of SJER116; (**b**) validated abundance map reference data (AMRD) at the 10 m×10 m aggregation scale for SJER116, with ground cover classes (left-to-right and top-to-bottom): PV, NPV, BS, rock and other.

Table 5. Accuracy validation data for scene-wide AMRD generated using RSRD-NNLS. All data are expressed in terms of percent coverage area differences.

Class	Mean Difference from MOA	Lower CI	Upper CI	SD Difference from MOA
PV	1.6	−0.1	3.8	3.8
NPV	−1.5	−4.4	1.1	7.2
BS	−4.5	−7.0	−3.7	6.4
Rock	4.3	3.0	7.2	8.0

Figure 15. (**a**) RGB imagery of SOAP299; (**b**) validated abundance map reference data (AMRD) at the 10 m×10 m aggregation scale for SOAP299, with ground cover classes (left-to-right and top-to-bottom): PV, NPV, BS, rock and other.

5. Conclusions

The purpose of this paper was to validate abundance map reference data (AMRD) generated using the remotely-sensed reference data (RSRD) technique, which aggregates classification or unmixing results from fine-scale imaging spectrometer (IS) data to produce AMRD for co-located coarse-scale IS data. This validation effort included three separate remote sensing scenes. The scenes were specifically chosen to be representative of many remote sensing environments. They contain a variety of common ground cover, including asphalt roads, dirt roads, concrete, buildings, grass fields, dry valley grasslands, high mountain forests, large rock outcroppings, etc. Therefore, we expect the conclusions of this paper to be generally applicable to similar rural and suburban scenes. We recommend that additional validation studies focus on more diverse and even densely-mixed land cover types.

Validation was accomplished by estimating AMRD in 51 randomly selected 10 m×10 m plots, using seven different methods or observers, and comparing the results. These independent versions of reference data included field surveys by two observers, imagery analysis by two observers and RSRD by three algorithms. Given that *t*-test comparisons showed statistically significant differences between all seven versions of reference data, we proposed that the best estimate of actual ground cover abundance fractions within the 51 plots is found by taking the mean of all (MOA) independent versions of reference data for each plot and class. Generating MOA for a limited number of random plots is labor intensive, but once generated, MOA can then be used to validate AMRD generated using the RSRD technique, which efficiently produces scene-wide reference data and can serve as a validated baseline for all future algorithm developments.

At the 10-m GSD aggregation scale, mean differences between versions of reference data and MOA were 1.6 to 6.1%, with standard deviations of 5.6 to 9.2%. The closest version of reference data to MOA was a version of imagery analysis, with mean differences of 2.0% and a standard deviation of 5.6%. The RSRD algorithm based on NNLS spectral unmixing was nearly as close to MOA, achieving mean differences of 3.0% and a standard deviation of 6.3%. Equivalence testing yielded a zone of equivalence between $[-7.0, 7.2\%]$ for RSRD-NNLS. Considering the efficiency of RSRD in producing scene-wide AMRD, these results are promising. Validated scene-wide AMRD generated using RSRD-NNLS are available for use by the remote sensing community.

The novel contributions to the pixel unmixing branch of remote sensing research include: (1) a documented validation of the accuracy of abundance map reference data (AMRD) itself; (2) the inclusion of five reference data classes, which is similar to many contemporary classification/unmixing efforts; and (3) the creation and validation of AMRD for three new study scenes using different IS sensors. Secondary goals are to make the larger remote sensing community more aware of the need

to validate reference data itself and to establish methods to quantitatively assess the performance of spectral unmixing algorithms.

This validation effort centered around using generic 10-m GSD coarse-scale imagery, which was designed such that the reference data accuracy validation in this paper would be valid for use with any coarse imagery with GSD >10 m. The next step of our research is to apply the validated AMRD to examples of real coarse-scale imagery, such as the AVIRIS data mentioned previously. Applying AMRD to real imagery requires addressing various important topics, including the sub-pixel spatial alignment of fine- and coarse-scale imagery (critical to overlaying imagery with different spatial resolutions), comparing aggregation strategies, such as simple rectangular pixel aggregation and sensor point-spread function (PSF) aggregation, and the introduction of methods to compare coarse unmixing results to AMRD.

Supplementary Materials: The following are available online at www.mdpi.com/2072-4292/9/5/473/s1, Supplementary S1: A spreadsheet containing all seven versions of estimated reference data for 1275 2-m samples in 51 10-m plots. Further study data is available upon request to the authors.

Author Contributions: M.D.W. conceived of and designed the experiments, performed the experiments, analyzed the data and wrote the paper. R.J.P. analyzed the data, edited the paper and provided statistical expertise. A.J.F. performed experiments and edited the paper. J.P.K. reviewed results, edited the paper and provided remote sensing expertise. J.v.A. conceived of and designed the experiments, reviewed the results, edited the paper and provided remote sensing expertise.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AMRD	Abundance map reference data
AVIRIS	Airborne Visible-Infrared Imaging Spectrometer
BS	Bare soil
ED	Euclidean distance
GSD	Ground sample distance
IS	Imaging spectrometer
ML	Maximum likelihood
MOA	Mean of all reference data
NEON	National Ecological Observatory Network
NNLS	Non-negative least squares
NPV	Non-photosynthetic vegetation
PV	Photosynthetic vegetation
RSRD	Remotely-sensed reference data
SJER	San Joaquin Experimental Range
SOAP	Soaproot Saddle

References

1. Vane, G.; Green, R.O.; Chrien, T.G.; Enmark, H.T.; Hansen, E.G.; Porter, W.M. The airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sens. Environ.* **1993**, *44*, 127–143.
2. Green, R.O.; Eastwood, M.L.; Sarture, C.M.; Chrien, T.G.; Aronsson, M.; Chippendale, B.J.; Faust, J.A.; Pavri, B.E.; Chovit, C.J.; Solis, M.; et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sens. Environ.* **1998**, *65*, 227–248.
3. Eismann, M.T. *Hyperspectral remote sensing*; SPIE Press: Bellingham, WA, USA, 2012.

4. Goetz, A.F.H.; Vane, G.; Solomon, J.E.; Rock, B.N. Imaging spectrometry for earth remote sensing. *Science* **1985**, *228*, 1147–1152.

5. Chang, C.I. *Hyperspectral data exploitation: Theory and applications*; John Wiley & Sons: Hoboken, NJ, USA, 2007.

6. Keshava, N.; Mustard, J.F. Spectral unmixing. *IEEE Signal Process. Mag.* **2002**, *19*, 44–57.

7. Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 354–379.

8. Pearlman, J.S.; Barry, P.S.; Segal, C.C.; Shepanski, J.; Beiso, D.; Carman, S.L. Hyperion, a space-based imaging spectrometer. *IEEE Tran. Geosci. Remote Sens.* **2003**, *41*, 1160–1173.

9. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC press: Boca Raton, FL, USA, 2008.

10. Swayze, G.; Clark, R.N.; Kruse, F.; Sutley, S.; Gallagher, A. *Ground-Truthing AVIRIS Mineral Mapping at Cuprite, Nevada*; NASA: Denver, CO, USA, 1992.

11. Baumgardner, M.; Biehl, L.; Landgrebe, D. 220 band aviris hyperspectral image data set: June 12, 1992 Indian pine test site 3. *Purdue Univ. Res. Repos.* **2015**, *10*, doi:10.4231/R7RX991C.

12. Gamba, P. A collection of data for urban area characterization. In Proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004.

13. Swayze, G.A. The Hydrothermal and Structural History of the Cuprite Mining District, Southwestern Nevada: An Integrated Geological and Geophysical Approach. Ph.D. Thesis, University of Colorado at Boulder, Boulder, CO, USA, 1997.

14. Landgrebe, D. Indian Pines Aviris Hyperpsectral Reflectance Data: 92av3c, 296. Available online: http://makalu.jpl.nasa.gov/ (accessed on 10 May 2017).

15. Hyperspectral Remote Sensing Scenes. Available online: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (accessed on 17 November 2016).

16. Winter, M.E. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In Proceedings of the SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, San-Diego, CA, USA, 9–14 July 1999.

17. Nascimento, J.M.; Dias, J.M. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Tran. Geosci. Remote Sens.* **2005**, *43*, 898–910.

18. Zare, A.; Ho, K. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Process. Mag.* **2014**, *31*, 95–104.

19. Powell, R.L.; Roberts, D.A.; Dennison, P.E.; Hess, L.L. Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sens. Environ.* **2007**, *106*, 253–267.

20. Van de Voorde, T.; Jacquet, W.; Canters, F. Mapping form and function in urban areas: An approach based on urban metrics and continuous impervious surface data. *Landsc. Urban Plan.* **2011**, *102*, 143–155.

21. Walton, J.T. Subpixel urban land cover estimation. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 1213–1222.

22. Schwieder, M.; Leitão, P.J.; Suess, S.; Senf, C.; Hostert, P. Estimating fractional shrub cover using simulated EnMAP data: A comparison of three machine learning regression techniques. *Remote Sens.* **2014**, *6*, 3427–3445.

23. Williams, M.D.; van Aardt, J.; Kerekes, J.P. Generation of remotely sensed reference data using low altitude, high spatial resolution hyperspectral imagery. In Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXII, Baltimore, MD, USA, 17 April 2016.

24. Kampe, T.U.; Johnson, B.R.; Kuester, M.; Keller, M. NEON: The first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. *J. Appl. Remote Sens* **2010**, *4*, doi:10.1117/12.825697.

25. Kampe, T.; Leisso, N.; Musinsky, J.; Petroy, S.; Karpowiez, B.; Krause, K.; Crocker, R.I.; DeVoe, M.; Penniman, E.; Guadagno, T.; et al. The NEON 2013 Airborne Campaign at Domain 17 Terrestrial and Aquatic Sites in California. Available online: https://www.researchgate.net/profile/Thomas_Kampe/publication/264862121_The_NEON_2013_Airborne_Campaign_at_Domain_17_Terrestrial_and_Aquatic_Sites_in_California/links/53f409e70cf2155be354f1cc.pdf (accessed on 1 December 2016).

26. Asner, G.P.; Lobell, D.B. A biogeophysical approach for automated SWIR unmixing of soils and vegetation. *Remote Sens. Environ.* **2000**, *74*, 99–112.

27. Asner, G.P.; Heidebrecht, K.B. Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: Comparing multispectral and hyperspectral observations. *Int. J. Remote Sens.* **2002**, *23*, 3939–3958.

28. Soaproot Saddle—SOAP. Available online: http://www.neonscience.org/science-design/field-sites/soaproot-saddle (accessed on 1 December 2016).

29. Lawson, C.L.; Hanson, R.J. *Solving Least Squares Problems*; SIAM: Philadelphia, PA, USA, 1995.

30. Paired Comparisons. Available online: https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_ttest_sect011.htm (accessed on 1 December 2016).

31. Equivalence and Noninferiority Testing. Available online: https://support.sas.com/resources/papers/proceedings15/SAS1911-2015.pdf (accessed on 1 December 2016).

32. Ott, R.L.; Longnecker, M.T. *An Introduction to Statistical Methods and Data Analysis*; Nelson Education: Toronto, ON, Canada, 2015.