


## Article

# Improving the Reliability of Mixture Tuned Matched Filtering Remote Sensing Classification Results Using Supervised Learning Algorithms and Cross-Validation

Devin Routh <sup>1,†</sup>, Lindsy Seegmiller <sup>2,†</sup>, Charlie Bettigole <sup>3</sup>, Catherine Kuhn <sup>4</sup>,  
Chadwick D. Oliver <sup>3</sup> and Henry B. Glick <sup>3,\*</sup> 

<sup>1</sup> Department of Environmental Systems Science, ETH Zürich, 8092 Zürich, Switzerland; devin.routh@usys.ethz.ch

<sup>2</sup> Division of Geoinformatics, Department of Urban Planning and Environment, KTH Royal Institute of Technology, SE-114 28 Stockholm, Sweden; lindsy@kth.se

<sup>3</sup> School of Forestry and Environmental Studies, Yale University, New Haven, CT 06511, USA; charlie.bettigole@yale.edu (C.B.); chad.oliver@yale.edu (C.D.O.)

<sup>4</sup> School of Environmental and Forest Sciences, University of Washington, Seattle, WA 98195, USA; ckuhn@uw.edu

\* Correspondence: henry.glick@yale.edu

† Joint first authorship.

Received: 22 August 2018; Accepted: 16 October 2018; Published: 23 October 2018



**Abstract:** Mixture tuned matched filtering (MTMF) image classification capitalizes on the increasing spectral and spatial resolutions of available hyperspectral image data to identify the presence, and potentially the abundance, of a given cover type or endmember. Previous studies using MTMF have relied on extensive user input to obtain a reliable classification. In this study, we expand the traditional MTMF classification by using a selection of supervised learning algorithms with rigorous cross-validation. Our approach removes the need for subjective user input to finalize the classification, ultimately enhancing replicability and reliability of the results. We illustrate this approach with an MTMF classification case study focused on leafy spurge (*Euphorbia esula*), an invasive forb in Western North America, using free 30-m hyperspectral data from the National Aeronautics and Space Administration's (NASA) Hyperion sensor. Our protocol shows for our data, a potential overall accuracy inflation between 18.4% and 30.8% without cross-validation and according to the supervised learning algorithm used. We propose this new protocol as a final step for the MTMF classification algorithm and suggest future researchers report a greater suite of accuracy statistics to affirm their classifications' underlying efficacies.

**Keywords:** mixture tuned matched filtering (MTMF); image classification; accuracy assessment; post-processing automation; linear unmixing; hyperspectral remote sensing; supervised learning; machine learning; leafy spurge

## 1. Introduction

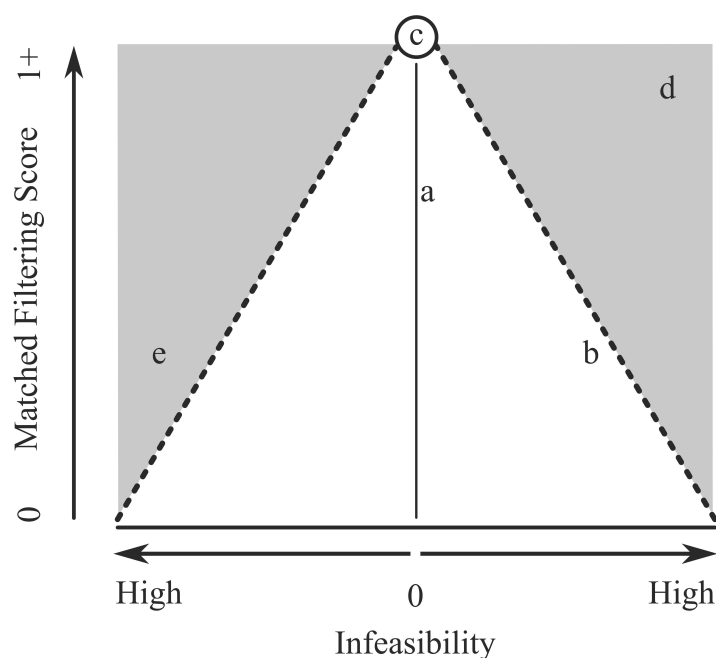
### *Background: MTMF*

Given recent advancements in remote sensing and imagery analysis, the detection of individual plant species and their distributions has become a more realistic goal for land managers with training and access to satellite image data and analysis software. In particular, the mixture tuned matched filtering (MTMF) linear unmixing algorithm has proven an effective tool for identifying the presence

and abundance of specific land cover types and endmembers [1–5]. In contrast to other forms of spectral unmixing (e.g., multiple endmember spectral mixing analysis), MTMF distinguishes itself by requiring the user to supply only the target spectral signature(s) and not the signatures of background features [2]. This feature bypasses previous spectral unmixing hurdles [6], and allows the classification to be easily and rapidly adapted to broad geographic areas that possess suitably uniform target spectra.

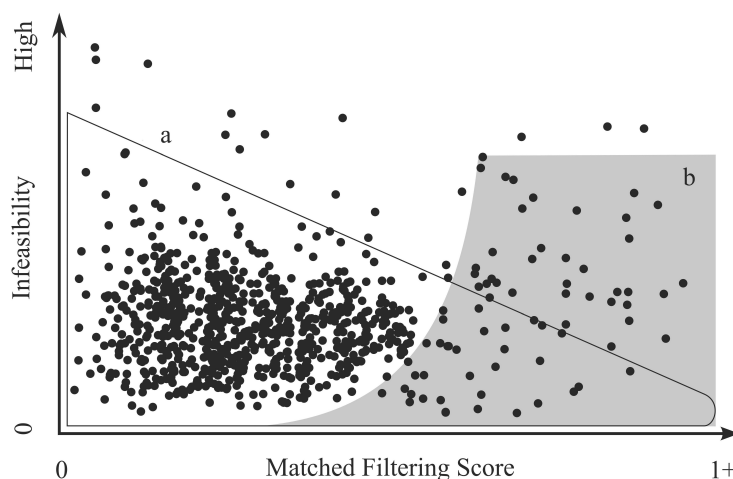
The MTMF algorithm follows three major steps: (1) Reduction of noise in the input image using the minimum noise fraction (MNF) transformation [7]; (2) creation of a matched filtering (MF) score/value image, representing how closely each pixel matches the target signature(s); and (3) creation of a mixture tuning score/value (also termed an infeasibility score/value) to reduce the likelihood of including false-positive pixels (i.e., those improperly assigned to the target class) in the final classified image [8]. The MF score represents how closely a pixel matches the endmember on a scale from approximately 0 to 1, where 0 is least like the endmember and 1 indicates a strong match (though values greater than 1 can be computed mathematically, occurring more commonly in images of low spectral contrast [8]). A standard characteristic of linear spectral mixture analysis methods, the 0 to 1 range for MF values approximately correlate to a 0% to 100% endmember sub-pixel abundance measurement [9–12]. However, analysts should approach this interpretation with caution, for while Mitchell and Glenn [13] found a linear relationship between MF values and field plot abundance measured on the ground, their calculated MF values consistently underestimated field-sampled endmember canopy cover.

The infeasibility score of the MTMF classification represents the multidimensional geometric distance from a pixel's spectra to the target spectra in transformed vector space (i.e., MNF space), and ranges from 0 to an indefinite maximum value [8]. As the MF score increases there is a concomitant narrowing of the range of acceptable infeasibility scores. This phenomenon means that the likelihood of falsely classifying a pixel as containing the cover type or endmember increases as the infeasibility score increases (Figure 1). Therefore, pixels with a large MF score and a small infeasibility score constitute those most likely to contain a large proportion of the endmember.



**Figure 1.** A two-dimensional projection of mixture tuned matched filtering (MTMF) matched filtering (MF) and infeasibility mixture space (after [8]). (a) Large MF value and near 0 infeasibility; (b) smaller MF value and marginally feasible; (c) a perfect MF value and entirely feasible; (d) a very large MF value, but very infeasible; (e) a small MF value and very infeasible.

A crucial part of the MTMF classification process occurs when selecting thresholds for infeasibility scores across the range of MF values, which ultimately determine the final classification of the target spectra at each pixel (i.e., pixel values below the threshold are classified as belonging to the endmember class, while pixel values above the threshold are classified as false positives). This process involves finding the value-specific and measurable infeasibility thresholds (which expand for smaller MF values and contract for large MF values) across the range of MF values (i.e., the dotted lines featured in Figure 1). Published studies have mentioned this aspect of MTMF in passing, noting that optimal infeasibility value thresholds were determined “interactively”, “iteratively”, or “manually” via an MF versus infeasibility value scatterplot (Figure 2; [13–15]). Such approaches have even been used outside of remote sensing applications of MTMF [16]. An artificial example of such a scatterplot is shown in Figure 2, where the shaded region (2b) denotes ground reference plots whose MF values alone theoretically make them likely matches to the target endmember. However, a large proportion of these points fall outside the expected region (2a) for likely endmember pixels. This region, or mixing space, is derived via multidimensional hypercones whose projection in 2-D space becomes triangular (as in Figure 1). The delineation of properly classified reference plots (pixels) should theoretically follow the conical shape (2a). However, previous research (e.g., [13], Figure 5a; [8], Figure 4) has not shown this to be true, with research instead delineating a nearly inverse shape as shown by (2b).



**Figure 2.** An example scatterplot of artificially generated matched filtering (MF) and infeasibility values, like those used during mixture tuned matched filtering (MTMF) post-processing. The delineation of properly classified reference plots (pixels) should theoretically follow the conical shape (a), however, previous research has instead indicated a nearly inverse shape as shown by (b).

Whereas the final classified map accuracy, as computed using confusion matrices, depends on whether a given pixel is assigned to its field-referenced class, this “iterative” assessment of MF and infeasibility value pairings directly determines each pixel’s final class assignment. Furthermore, the interactive, user-specific nature of this final step in the MTMF process hinders the replicability of results in addition to increasing the chances of biased measures of map accuracy from one-time assessments and overfit models. Thus, suggesting augmentations and improvements to this step is the primary focus of our study.

With the above factors in mind, this paper aims to accomplish the following two objectives: (1) Applying a suite of automated, supervised learning algorithms to synthesize the two MTMF results (the MF and infeasibility scores) into a single classified value (i.e., hard, as opposed to fuzzy classification) that can be verified using matrix-based accuracy assessment; and (2) pairing supervised learning algorithms with rigorous cross-validation to reduce artificial inflation of classification accuracies resulting from traditional one-time accuracy assessment. We illustrate how this approach can be used to obtain more representative results than the traditional MTMF post-processing. This is accomplished through an MTMF classification of freely available hyperspectral data using an

endmember of leafy spurge (*Euphorbia esula*), an invasive forb in the Western U.S. that has been examined in many remote sensing studies relying on datasets from various sources (e.g., AVIRIS, HyMap, etc.) [3,8,13,17–29]. This study principally addresses the use of MTMF when performing hard, per-pixel classification, though studies using the MF and infeasibility scores to produce continuous values in a final map could also benefit from synthesizing these metrics using supervised learning approaches with full cross-validation.

## 2. Materials and Methods

### 2.1. Accuracy Assessment via Supervised Learning Algorithms

#### 2.1.1. Overview

The goal for many image classification processes is to maximize the final map accuracy. However, with multistep algorithms, such as MTMF, the complex relationships between MF scores, infeasibility scores, and on-the-ground measurements make it difficult to identify the parameter thresholds that yield optimal results. As mentioned above, the classification of MF and infeasibility images has traditionally relied on a subjective process in which optimal thresholds are chosen for both MF and infeasibility values using an “interactive” or “iterative process” [4,8,13]. Some studies have offered scant descriptions about their chosen protocols for this step of the algorithm [3,27,28,30]. These thresholds ultimately determine how MF and infeasibility values are used for classification and therefore play an important role in the larger mapping process. The question at hand is: Given a set of field-reference plots sampled from a larger study area (i.e., pixels containing MF scores, infeasibility scores, and known endmember abundance values), how can a user objectively determine the relationship between the observed endmember abundance and modelled endmember abundance that optimizes the final map accuracy? As indicated above, here, we are concerned with hard classification, in which each pixel is assigned to a single, mutually exclusive class.

We can address this question by considering an MF versus infeasibility score scatterplot containing all field-reference plot pixels, graphed according to their computed values in these two dimensions (e.g., Figure 2). For any given MF score, pixels containing endmembers should lie closer to an infeasibility value of zero (which is otherwise defined as the geometric altitude of a multidimensional hypercone). Using this scatterplot, a researcher can compute a range of possible map accuracies with any set of field-referenced points by discerning a contiguous area that contains as many of the confirmed endmember points, and as few confirmed non-endmember points, as possible. Provided the field-referenced sample of points accurately reflects the larger distribution of land-covers across a scene—which is determined by the underlying sampling scheme [31]—the boundary of this same contiguous area on the scatterplot can potentially be used to finalize the classification of all pixels in an image. The “interactive” or “iterative” process described above constitutes the process of an analyst “drawing” this enclosed area on the scatterplot, finalizing the classification, inspecting the accuracy results, and then “re-drawing” the area to derive a better set of accuracy statistics [8,13].

However, if an analyst includes their entire field-sampled collection of plots as a guide for “drawing” this area, not only is the final map accuracy a function of scatterplot interpretation and the “re-drawing” process, but the process is difficult to replicate and the analyst runs the risk of artificially inflating the perceived accuracy of the classification across the rest of the study area. To eliminate this iterative and user-directed process, and to eliminate the potential for artificial inflation of map accuracy metrics, we recommend applying automated supervised learning algorithms accompanied by cross-validation procedures to complete the MTMF algorithm [32–34].

#### 2.1.2. Supervised Learning Algorithms and Cross-validation

Though the MTMF algorithm is used as a classification tool, its results do not directly assign each data point a single classified value. MF and infeasibility scores are both continuous metrics that provide

a theoretical basis for the final classification. Given this fact, the final step of the MTMF algorithm is well suited to the application of supervised learning algorithms across these two continuous dimensions.

Approaching the final step of the algorithm in this way means that users will forfeit the direct interpretation of the MF score as a measure of subpixel abundance in the final classification output from the machine learning methods (given that these supervised learning algorithms assign new data points into categories, as opposed to continuous percentage values). Despite this discretization, users can potentially maintain relative levels of subpixel abundance in the machine learning process and in the classified image by a priori binning field-referenced datasets into abundance classes using automated approaches, such as the Jenks natural breaks algorithm [35]. Future applications could bin continuous measures of endmembers in any way that suits relevant ecological contexts or management needs. Our automated protocol seamlessly supports binning into any number of classes, promoting minimal subjective input and matching other standardization protocols added directly within the MTMF algorithm [36]. In conjunction with the field sampling scheme, the binning process itself (whether binary or including multiple classes) determines categorical membership for the training set used as input to the supervised learning algorithms, and therefore ultimately affects classification results. Researchers using our protocol may choose different binning standards according to their needs or landscape properties (e.g., endmember, ecological, or geomorphological characteristics). This study concentrates on presence/absence (i.e., binary) classification to better illustrate issues surrounding cross-validation and model overfitting.

Furthermore, and importantly, despite the discretization in our suggested final step, nothing would prevent an analyst from maintaining the original MF score data at each pixel after the machine learning algorithms have been applied, whether via a multiple class binning process or with a presence/absence machine learning application. In other words, the machine learning algorithms can aid in determining whether a pixel is truly a “false positive” given the field referenced data, and the MF scores of endmember classified pixels will still provide estimates of subpixel abundance.

To date, the scientific literature on machine learning algorithms is rich and expansive. As our study focuses on remote sensing classification, detailed descriptions of particular machine learning algorithms fall outside the purview of this work, and we encourage readers to follow pertinent references for further information. After reviewing a number of established supervised learning approaches (e.g., [37–41]), we chose to evaluate the following algorithms with our case study data: Support vector machines (SVM), naïve Bayes, random forests, single hidden layer back-propagation neural networks, multinomial/logistic regression, and quadratic/linear discriminant analysis. Each of these has been previously used in remote sensing or Geographic Information System (GIS) analysis, including various applications to hyperspectral image classification [42]: support vector machines [43–45], naïve Bayes [46,47], random forests [48–51], neural networks [52–54], multinomial/logistic regression [55,56], and quadratic/linear discriminant analysis [57,58]. However, for the case of hard classification, we have not identified any research that applies these methods to the MTMF classification post-processing workflow.

We analyzed each algorithm’s results using a 10-fold cross-validation procedure, which has been long accepted in computer science, artificial intelligence, and data science circles as a rigorous approach to model validation [32,33,59]. Conceptually, these methods should provide superior results to the traditional MTMF post-processing by maximizing model strength through an iterative, yet replicable, machine learning process, while also producing unbiased approximations of map accuracy obtained through an automated protocol. Each of the above-mentioned supervised learning algorithms relies on a training sample (i.e., our field data), organized in vector form across two dimensions (MF and infeasibility). The issue of systematic bias emanating from a single training set was mitigated through the cross-validation procedure [34]. The algorithm with the greatest accuracies and least variance was considered the most promising.

Supervised learning algorithms and cross-validation are easily scripted and highly repeatable. We implemented both using the R programming language [60] and the ‘caret’ package [61,62]. The

caret package provides a bridge to other machine learning libraries (e.g., nnet, randomForest, kernlab) to provide a consistent framework for training and executing a diverse suite of models. We capitalized on established caret functionality to perform both cross-validation functions and the tuning of model parameters and hyper-parameters. Specific control functions for model training provide a mechanism for parameter tuning, which is conducted using a grid search approach. Grids of possible parameter values are generated by functions specific to each machine learning method, and the training data are then used to determine the optimal parameter values. Duro et al. [33] present a similar use of the caret package, in which they evaluate the strengths of the random forests, SVM, and decision tree algorithms in classifying SPOT-5 imagery.

We tested both binary classification of the training data (presence/absence) as well as multi-level/semi-continuous abundance classification, as has been considered in other studies utilizing MTMF [63,64]. However, due to an imbalance in MF values among our field sampled plots, our data structure prohibited obtaining meaningful results from a multi-level classification. As such, we focus on binary categorization, with conceptual extension to multi-level classification. With respect to the choice of multinomial versus logistic regression, our binary response led us to use logistic regression. With respect to linear versus quadratic discriminant analysis, Chi-square quantile plots indicated that our presence/absence-binned data did not exhibit sufficient multivariate normality across MF and infeasibility values, leading us to use quadratic discriminant analysis.

To provide context and to illustrate the benefits of the cross-validated supervised learning classification results, we also implemented the traditional scatterplot “drawing” approach [8,13] for binary classification. We used an iterative drawing-re-drawing process that attempted to follow the conceptual logic illustrated in Figure 2. Both of the frameworks shown in Figure 2a,b yielded unsatisfactory results, though better results were obtained using a shape represented here (Figure 5) by a five-knot cubic spline. Note that the spline’s positive, increasing trend with increasing MF scores more closely follows Figure 2b than Figure 2a, indicating general agreement with Mitchell and Glenn’s ([13], Figure 5a) and Mundt et al.’s ([8], Figure 4) findings. The unsatisfactory results led us to extend the “drawing” strategy from simple delineations (as in Figure 2 or the cubic spline in Figure 5) to an “overfit” model wherein classification thresholds were neither regular nor predictable (Figure 5 “overfit” model). In this model, each field reference data point that contained leafy spurge (here, defined as  $\geq 5\%$  cover) and that also demonstrated the greatest infeasibility score among points within a range of 0.01 MF score units (i.e., an MF bin of width, 0.01), defined the threshold of presence and absence. All reference data points within the same MF bin, but with a lesser infeasibility (i.e., below the line), were defined as containing some leafy spurge (present); all other points (i.e., above the line) were defined as lacking leafy spurge (absent).

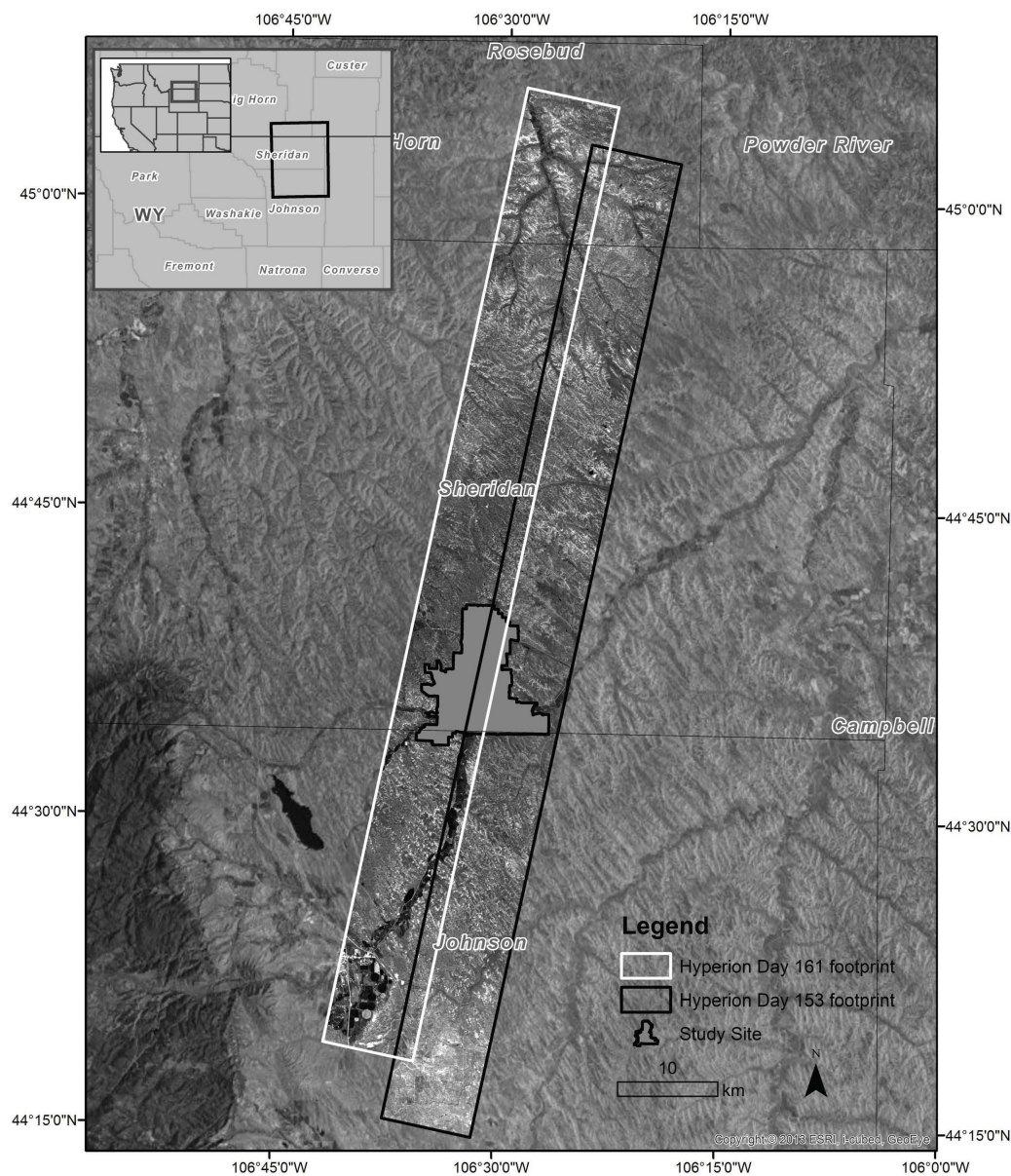
While grossly overfit and violating some basic linear mixing principles, we created this model because: (1) It is theoretically possible to manually draw the designated form on a scatterplot (though practically impossible using the ENVI (Environment for Visualizing Images) software [65]); and (2) it should aid in maximizing map accuracy for a given dataset without violating all of the logic of scatterplot “drawing” (e.g., classifying reference data points within the same MF bin as an interspersed mixture of presence and absence).

## 2.2. Study Site, Image Acquisition and Processing, Field Sampling, and Classification Overview

### 2.2.1. Study Site

We tested our conceptual framework by seeking to classify the distribution and abundance of leafy spurge (*Euphorbia esula*)—a forb whose vegetative propagation displaces native North American grasses typically grazed by livestock. The study site is a  $\sim 95\text{-km}^2$  cattle ranch in northern Wyoming, USA (Figure 3) composed largely of upland sagebrush-steppe (*Artemisia tridentata*) with two rivers (Piney Creek and Clear Creek) cutting through the lowlands. Land cover across the site is heterogeneous across small extents, and consists primarily of semi-arid rangeland, including areas of

perennial/annual grasses, sagebrush, exposed soil and rock, small reservoirs, and a collection of ranch buildings and irrigated fields in the lowlands. Though leafy spurge has yet to establish dominance over significant portions of the ranch, it has existed in this area for decades and is commonly found in large vegetatively connected patches that displace native grasses serving as livestock forage [66–68]. The range of spurge areal coverage in 15 m radius buffered sample points was 0% to 66%.



**Figure 3.** A map of the study site showing the Hyperion flight lines for Days 153 and 161 of 2014.

### 2.2.2. Hyperion Data and Tasking

We used two images from the Hyperion sensor on board the National Aeronautics and Space Administration's (NASA) EO-1 satellite for this study (Figure 3). We chose Hyperion images because they offer high spectral resolution (242 bands) and moderate spatial resolution (30-m pixels) from a publicly available source [69]. In coordination with NASA, we tasked images of our study site at a time when the leafy spurge (*Euphorbia esula*) presented its distinctive yellow-green flower bracts [70] and at a time that ensured the smallest temporal window between each image and ground-collected spectral measurements. We received two sets of L1R images of the study site—one covering the eastern

portion (Year 2014, Day 153) and the other covering the western portion (Year 2014, Day 161). We used the L1R images for the classification to maintain the integrity of the data and to match field-collected signatures; we used the corresponding L1T images only for post-classification georeferencing. Each L1R image is 185-km long by 7.7-km wide with 242 bands at  $\approx 10\text{--}11$  nm each. Bands 1–7, 58–76, and 224–242 contained no data.

### 2.2.3. Data Pre-Processing

Unless otherwise noted, all image processing and analysis took place in ENVI 5.1 [65].

Though cloud presence was negligible for both days, we produced a cloud mask for each L1R image before atmospherically correcting the images using Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH; [71]). The push-broom sensor on board the EO-1 satellite produces striping where sensors are broken or improperly calibrated [72,73]. Most commonly, stripe pixels contain values much lower than those on the adjacent sides. We examined all bands for both L1R images and created stripe masks to correct the values of stripe pixels for all bands that otherwise contained valuable data [72]. We used cross-track illumination correction to mitigate the impact of spectral “smile”, a phenomenon in which pixels artificially increase in value across the width of the image, or sensor array [74,75]. This effect is best observed in the first MNF band after a minimum noise fraction forward transformation.

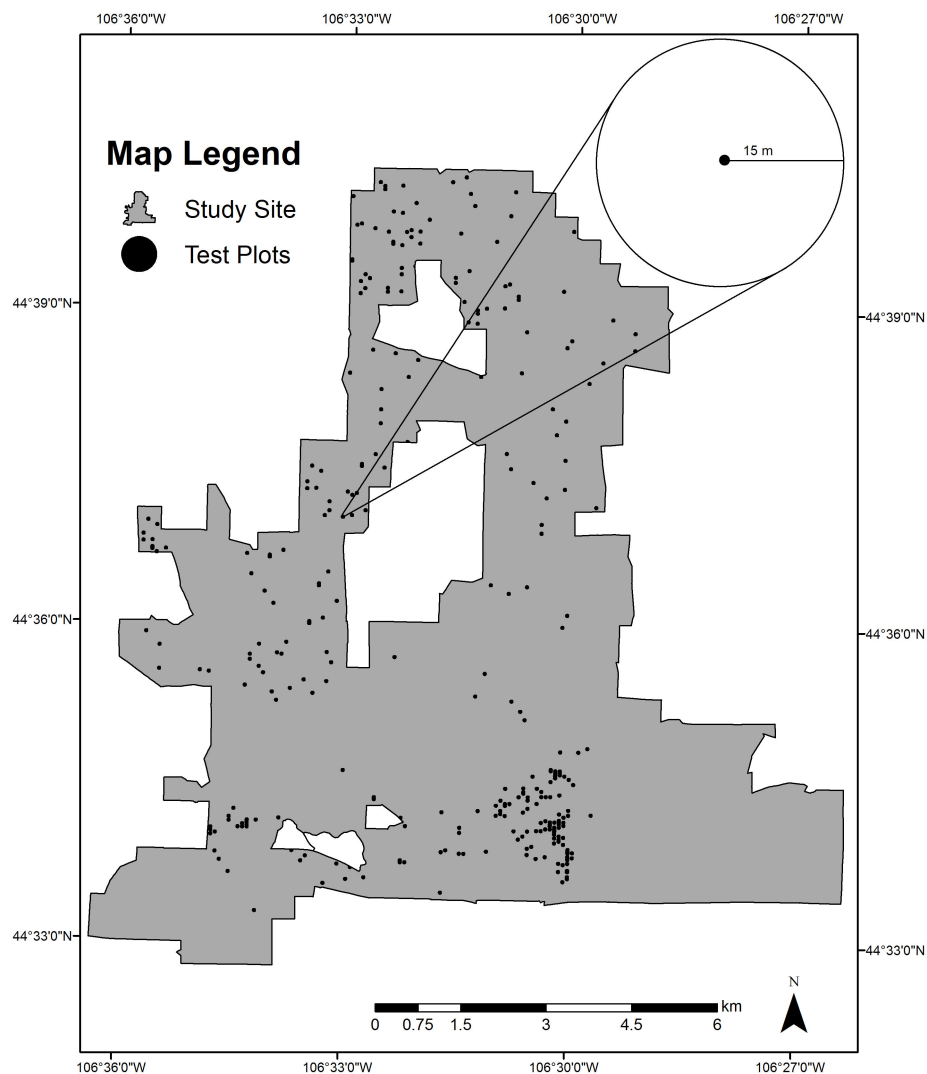
We spectrally subset our images to eliminate bands that contained no data, significant striping, or were too noisy to contain any useful information. After this process, the western image (Year 2014; Day 161) contained 143 bands and the eastern image (Year 2014; Day 153) contained 132 bands. Because we processed each image independently, this discrepancy had no effect on the analysis. We spatially subset each image by trimming the border pixels that contained extreme values. For interior pixels with extreme outlying values, we used a localized kernel mean to obtain more realistic pixel values. Finally, we normalized the images to the reflectance range (0–1).

### 2.2.4. Target Spectra and Reference Plot Collection

During June and July of 2014, we collected spectral signatures of leafy spurge using a portable field spectrometer (Field Spec Pro; Analytical Spectral Devices, Inc., Boulder, CO, USA). As noted above, MTMF does not require spectra from background surfaces [2], which theoretically allows for broad geographic application. To ensure that our target spectral signature was representative of our landscape, we gathered field signatures at eight separate spurge infested locations across the study area. We took hundreds of readings at each location and then averaged all field collected signatures to form a final signature. We scheduled spectral signature collection on the same days and at the approximate times when the tasked Hyperion images were captured. We did not observe unusual variability in the signatures we captured; they generally showed agreement with published signatures and those in the U.S. Geological Survey spectral library available through the ENVI software. The variability of target endmembers is a function of both the endmembers and the geographic domain over which they are being mapped. Not all target spectra will be so consistent across one’s study area, and it is important to acknowledge that variable signatures would limit one’s ability to estimate relative sub-pixel abundance through undesirable spectral mixing with background signatures.

To aid in the identification of field plot locations, we first generated preliminary MF values to produce rough estimates of spurge abundance across the study area, and then binned the results into five uniform abundance classes:  $\leq 0.0\text{--}0.2$ ,  $0.2\text{--}0.4$ ,  $0.4\text{--}0.6$ ,  $0.6\text{--}0.8$ ,  $0.8\text{--}\geq 1.0$ . For each spatially explicit abundance class, we generated at least 50 field-reference plots, each with a 15 m radius buffer, using a spatially explicit form of stratified random sampling ([31]; total  $n = 325$ , see Figure 4). While based on a preliminary classification, we hoped that the allocation of field plots to random locations within each abundance stratum would ensure a balanced distribution of spurge abundance values in our final dataset. This, in turn, would help ensure that the classification accuracy metrics we computed for each abundance class would have comparable precision. We visited these reference sites between late

May and mid-June 2014, and confirmed peak spurge bract coloration at the time of the Hyperion data collection. At each site, we noted the presence/absence of leafy spurge as well as an ocular estimate of leafy spurge canopy cover. These data were subsequently managed in an aspatial environment (i.e., script-based, tabular).



**Figure 4.** Map of field reference plots collected over the study site. These plots were used as test plots in the post-classification and accuracy assessment process. The centroid of each 15 m radius circle was found using a GPS with a self-tested maximum 95% circular error probable (CEP95) of  $\pm 2.28$  m.

#### 2.2.5. MTMF Classification

We selected a set of “shift difference” regions for each image to establish quantitative estimates of image noise, which are required for the minimum noise fraction (MNF) transformation. ‘Shift difference’ refers to the relative difference in variability statistics between (1) a homogeneous set of contiguous pixels defined by some bounding window, and (2) the set of contiguous pixels defined by that same bounding window after it has been shifted one or more cells in a specified direction. Shift difference regions from both images were selected according to a visual assessment of homogeneity using true-color display (bands 29, 21, and 16 in these images). The standard deviation for each band in each region was computed and compared to that from the other regions of the same band. Within each of our two images, the region with the lowest total standard deviation value (computed by summing the standard deviation values across all bands) was selected for the final shift difference region. We selected a 234-pixel region for the Day 161 image, and a 270-pixel region for the Day 153 image. To

further our interest in automating the human-guided portions of MTMF classification, we successfully automated a shift difference region selection method using Google Earth Engine [36].

We performed an MNF transformation on each image using its respective shift difference region to estimate the underlying noise. Based on eigenvalue scree plots, we selected the first 11 MNF bands of Day 161 and the first nine bands of Day 153 for image classification [1]. We then imported the target endmember's spectral signature (leafy spurge), transformed the spectral signature into MNF space to match the images, and applied the MTMF classification, producing the final MF and infeasibility scores/images.

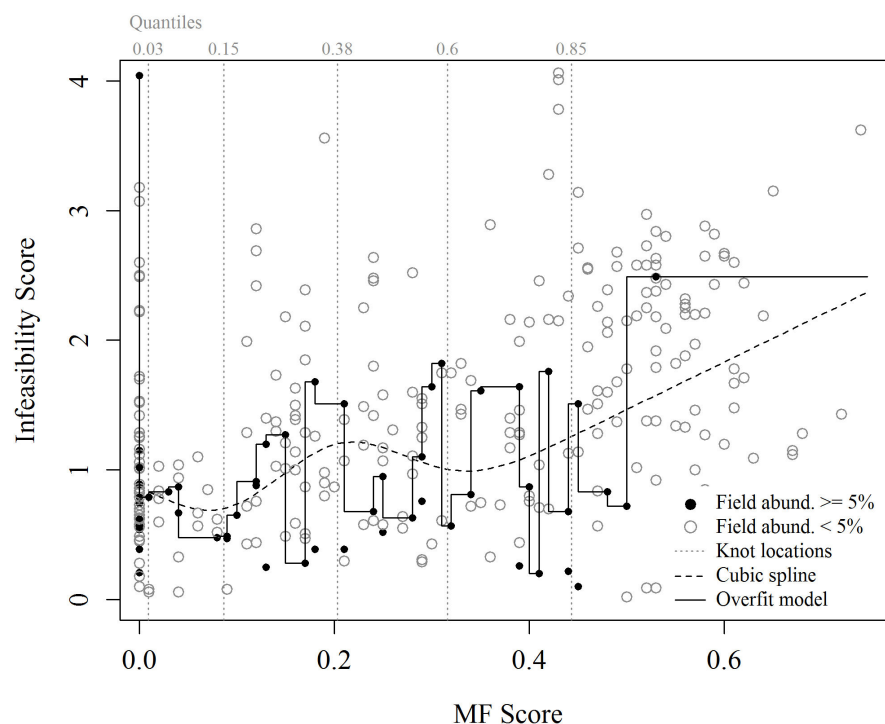
### 2.2.6. Georeferencing and Post-Processing

The MTMF classification does not rely on pixel- or field-based target spectra in their original spectral space, but rather it relies on spectral endmember target(s) that exist solely in transformed MNF space. For this reason, we chose not to initially georeference our initial L1R images to minimize the effects of spatial resampling artifacts during the shift difference region selection and MNF transformations. After the MTMF classification, however, we georeferenced each MF and infeasibility image using an image-to-image georeferencing approach to facilitate accuracy assessment.

## 3. Results

### Overall Accuracy and Kappa Accuracy Results

Using both the supervised learning and scatterplot “drawing” and “overfit” approaches (Figure 5), we achieved a range of accuracy results for each model and validation trial (Table 1). Here, we present several common accuracy assessment metrics, including Cohen's [76] variant of kappa, which is computed as the overall accuracy less the expected accuracy, divided by one minus the expected accuracy. In a spatial context, Cohen's kappa generally seeks to ascertain map accuracy having considered the fact that any given classification is likely to contain some proportion of properly classified pixels by chance alone.



**Figure 5.** Possible thresholds for MTMF post-processing image classification attainable with an iterative “drawing” approach. Compare with Figure 2b.

**Table 1.** Overall, kappa, user's, and producer's accuracies of MTMF-classified maps in which the presence/absence of leafy spurge was matched to field reference data using cross-validated supervised learning algorithms. SD = standard deviation of overall and kappa accuracies obtained through 10-fold holdout cross-validation. A = Absent, P = Present.

Presence/Absence Trials	Overall Accuracy	Overall SD	Kappa Accuracy	Kappa SD		User's Accuracy	Prod. Accuracy
Support Vector Machines	65.6%	9.4%	0.28	0.19	Mapped Class	A	67.5%
						P	38.9%
Naïve Bayes	64.9%	7.3%	0.27	0.15		A	63.6%
						P	85.1%
Quadratic Discriminant Analysis	56.3%	3.7%	0.03	0.10		A	63.2%
						P	46.5%
Random Forests	66.2%	5.2%	0.31	0.11		A	78.5%
						P	53.8%
Neural Networks	68.7%	9.6%	0.35	0.20		A	57.5%
						P	86.7%
Logistic Regression	64.6%	7.0%	0.25	0.14		A	63.5%
						P	60.4%
Manual Drawing	87.1%	–	0.23	–		A	69.7%
						P	72.4%
Over-fit Drawing	77.1%	–	0.43	–		A	65.3%
						P	53.5%
						A	77.3%
						P	77.3%
						A	73.9%
						P	35.4%
						A	90.1%
						A	88.4%
						P	97.9%
						A	57.1%
						P	18.2%
						A	76.5%
						P	96.4%
						A	80.0%
						P	41.7%

A = Absent, P = Present.

A neural network model provided the greatest overall and kappa accuracies (68.7% and 0.35) across all supervised learning approaches (range of 56.3–68.7% for overall and 0.03–0.35 for kappa). Random forests, support vector machines, naïve Bayes, and logistic regression all demonstrated a minor reduction in overall and kappa accuracies compared to neural networks (2.5–4.1% and 0.04–0.1, respectively), but they also demonstrated a concomitant reduction in standard deviation (0.2–4.4% and 0.01–0.09, respectively). In terms of the standard deviations of kappa values among supervised learning models, neural networks and support vector machines provided comparable levels of variability, as did naïve Bayes and logistic regression, as well as random forests and quadratic discriminant analysis. As kappa values may not be as informative as commonly believed [77,78], we also present user's and producer's accuracies. These results lend support to the neural networks and random forests models, both of which reflect relatively balanced, low to moderate errors of commission and errors of omission. Other models provide stronger classifications from either the user's or producer's point of view, but are less balanced. Overall, the random forests model might be considered the superior model in this analysis, as it has the second highest accuracy values, the lowest measures of accuracy variability, and balanced user and producer accuracies of moderate strength. Of particular note, the manual drawing and overfit drawing approaches led to high, but ultimately unreliable, accuracies, as indicated by our cross-validated supervised learning results.

In considering the application of our conceptual framework to other image classification problems, our actual accuracy values (Table 1) serve mainly to illustrate our approach. We emphasize that the ability to easily apply supervised learning modeling frameworks to classification problems, and the ability to easily compare results across a set of models, are more important than the specific values associated with our study site and our particular model parameters.

## 4. Discussion

### 4.1. Inflation of Map Accuracy

To our knowledge, the strategy we have described here has not been tested prior to this study, though we identified several studies that relied on regression modeling approaches for distilling MF and infeasibility values into meaningful results. For example, as a component of a study on Eastern

hemlock (*Tsuga canadensis*) decline, Pontius et al. [79] used MF and infeasibility from MTMF to calibrate the field-estimated percent basal area. The final regression model was extended to all MTMF output pixels to estimate species abundance. Pontius et al. [80] used a similar approach to map ash trees (*Fraxinus* L.) in an urban context, employing MF and infeasibility as predictive variables in logistic regression. Their approach produced a map illustrating the probability that a given pixel contains ash. Gudex-Cross et al. [63] used MF and infeasibility to predict the percent basal area of different tree species at reference plots using a form of stepwise regression. As part of a hierarchical mapping strategy, the resulting percent basal area rasters served as input to an object-based image analysis that ultimately led to hard classification. These studies support our work by indicating a broader interest in improving traditional MTMF post-processing by employing replicable modeling strategies for integrating MF and infeasibility values.

Recent use of supervised machine learning algorithms in image classification also lends support to our use of these models, particularly considering their strong performance. Rodriguez-Galiano et al. [48], for example, found that pixel-wise random forests classification of multi-season, multi-texture Landsat TM imagery improved the overall classification accuracy by 31% over traditional maximum likelihood classification. Pal and Mather [81] found that SVM improved the overall accuracy of classified Landsat ETM+ data by 5% over maximum likelihood and 2.8% over neural network algorithms. Duro et al. [33] compared three supervised learning methods applied to SPOT-5 imagery: Random forests, SVM, and decision trees. They reported pixel-wise overall accuracies of 89.7%, 89.3%, and 87.6%, respectively, with even higher accuracies achieved using object-based image analysis. Brenning [82] evaluated eleven algorithms in the classification of Enhanced Thematic Mapper Plus (ETM+) imagery and elevation model terrain derivatives. He found penalized linear discriminant analysis to produce significantly lower error rates compared to the other algorithms. Huang et al. [34] compared SVM, neural networks, decision trees, and maximum likelihood classifiers. They found that SVM and neural networks generally yielded superior accuracies in comparison with decision trees and maximum likelihood, though model superiority varied based on the way each model was trained. That the best performing algorithm varies between these studies and between the results we present above, is indication that data dimensionality, image heterogeneity, analytical framework (e.g., multi-temporal versus high resolution versus MTMF), and other landscape attributes influence classification accuracy.

With respect to the traditional strategy of manually drawing a region of properly classified cells on a two-dimensional scatterplot of MF by infeasibility, we achieved comparable or slightly lower overall map accuracies. Mundt et al. [8], for instance, computed an overall accuracy of 82% using a presence/absence classification when applying their “iterative” approach on HyMap scenes at 3.5 m pixel resolution. Mitchell and Glenn [13] also classified HyMap imagery varying from 3.2 m to 3.3 m resolution and produced overall accuracy values ranging from 67% to 85% via their “interactive” scatterplot approach. Parker-Williams and Hunt [27] achieved an overall accuracy of 95% when classifying leafy spurge for presence/absence and used custom flown Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data at a 20 m resolution. Depending on the supervised learning algorithm used, our protocol suggests potential overall accuracy inflation between 18.4% and 30.8% without cross-validation—the differences between the overall accuracy from our manual drawing method and that from the cross-validated models.

Importantly, similar MTMF studies (e.g., [3,8,13]) do not report one or more of kappa accuracy, user’s accuracy, producer’s accuracies, or the results of cross-validation procedures along with their classification accuracies. Thus, comparing results across studies may not be appropriate. Indeed, we too achieved overall accuracies as great as 87.1% using the traditional iterative “drawing” approach applied to binary delineation (Table 1). However, our kappa of 0.23 indicates that this classification is quite weak and inflated by random chance. When overfitting a presence-absence model, we still achieved an overall accuracy of 77.1%, though in this case, we also obtained a kappa of 0.43, which together suggest a more successful, albeit mediocre, classification. If we compare this result with the

cross-validated supervised learning methods, we see that the overfit model's accuracy figures exceed all other presence-absence trials. This is because one-time accuracy assessment can lead to biased results [78], which is why we recommend using cross-validation. Other studies of image classification have made use of cross-validation strategies [33,34,79], indicating support for the extension of these methods to MTMF.

However, the more significant results of our models are the kappa accuracy values ( $\text{kappa} \leq 0.35$ ), all of which consistently demonstrate that these models offer little more predictive ability than random assignment according to class proportions. This is what we might expect to see, as the cross-validated supervised learning methods help to correct systematic bias that appears in the drawn and overfit models, yielding instead more realistic (and less misleading) classifications. Table 1 illustrates how supervised learning algorithms can be used to maximize accuracy, while simultaneously minimizing artificial inflation through cross-validation. For our case study, our classified map is not 87.1% accurate, as basic "drawing" would lead us to believe, nor is it 77.1% accurate, as an overfit model might lead us to believe. In the case of presence-absence binned data, we can see that our mapped results are ~35% better than chance, with moderate user's and producer's accuracies. Some might generously consider these figures to reflect 'fair' agreement between the field data and classified map [83].

We present the overfit modeling approach as an extension of the manual, iterative drawing approach that has been used in the literature to date. Manual methods of this nature generally rely on a single and complete training set to obtain the classification threshold(s), and are therefore not conducive to rigorous cross-validation. This makes the classification threshold obtained through a complex drawing approach less applicable to other scenes. Yet, recent examples from the literature illustrate that the traditional "scatterplot drawing" approach to MTMF classification threshold determination is still very much in use [14,15,30]. Such projects have limited replicability, may present overly optimistic results, and generally show few signs of optimizing the results beyond comparing a single classification product to reference data (i.e., basic accuracy assessment). Projects of this nature might be strengthened from the framework we have outlined above.

To reiterate, it is important to recognize that MTMF is not limited to cases of hard classification, where the analyst seeks to obtain mutually exclusive class assignments. While we have presented an example of a hard classification, MTMF is also employed in fuzzy, continuous, or otherwise unique frameworks, such that the MF and infeasibility images are never integrated into discrete classes. For example, Franke et al. [84] used a fusion of three MF images of Brazilian cerrado to produce a continuous-scaled map of fire fuel load conditions that was calibrated against field-estimated biomass. Mikheeva et al. [64] developed abundance classes for tundra-taiga ecotone vegetation using MTMF and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) imagery. They relied on high resolution classified QuickBird imagery for accuracy assessment, and performed the binning of abundance data not during the MTMF post-processing workflow, but as a separate step, based on errors in the relationship between the two classified images. With the exception of thresholding MF values to those with infeasibility values less than 10, Ayoobi and Tangestani [85] relied solely on their MF image to map copper abundance. As noted above, Pontius et al. [79,80] and Gudex-Cross et al. [63] also employed MTMF classification in ways that were not initially hard classification.

With non-hard classification aside, we argue that machine learning algorithms provide an efficient, semi-automated approach to classification, and that cross-validation should be considered a critical component when computing the accuracy of any classified map product. Other studies within the remote sensing and geospatial literature base have already demonstrated the importance of cross-validation (e.g., [33,34,86–88]), and our study strongly suggests extending this standard procedure to future MTMF algorithm use.

#### 4.2. Hyperion, Automation

Our results do not demonstrate the highly successful use of Hyperion data for the classification of leafy spurge. This is visible in our computed accuracies and in Figure 5. When considering the relatively

low estimates of kappa in conjunction with the variability visible in Figure 5 and Table 1, we might question the spectral fidelity of our target endmember (compared to the background) or the quality of the imagery. The many previous studies focused on this particular plant species [3,8,13,17–29,89] collectively suggest that spectral fidelity of the endmember relative to the scene background signatures is not the issue. With respect to image quality, the Hyperion sensor is among those known to exhibit variable signal-to-noise ratios. Kruse et al. [90] and Ayoobi and Tangestani [85] note that the noise levels for a given sensor are generally fixed, but that the strength of the signal is dependent on external factors, such as solar zenith angle, atmospheric interference, or surface reflectance, among others. Kruse et al. [90] demonstrate that the signal in Hyperion imagery is sensitive to acquisition conditions, and that superior signal-to-noise ratios are obtained from periods with high solar zenith angles. Our imagery was collected during optimal conditions—within roughly one month of the summer solstice (Northern hemisphere), at mid-day, and with low cloud cover (NASA ratings of 0–9% and 20–29%). When paired with our methods for identifying shift difference regions and applying noise reduction transformations (MNF), we feel confident that the signal-to-noise ratios for our two images could not be markedly improved. This may indirectly indicate that Hyperion imagery is not suitable for making estimates of low abundance leafy spurge, or that our ocular field estimates were not reliable. Additional studies are needed to explore the application of unmixing algorithms to Hyperion scenes of heterogeneously vegetated landscapes.

While our results do not demonstrate a highly successful use of Hyperion data for leafy spurge, our results help to highlight the precise reason why reliable threshold selection and cross-validation are so essential. Our one-time manual accuracy assessment led us to an inflated estimate, and the cross-validation methods we employed provided the rigor to challenge these inflated accuracies and improve the reliability of our results. In general, automating as many aspects of image classification and other relevant protocols as possible (e.g., [36]) will help to ensure that the remote sensing and land-management communities maintain common ground in dialogues concerning their respective disciplines. Automation may also help in sidestepping limitations in the underlying theoretical bases of various imagery analysis protocols. In particular, the consistently underestimated subpixel abundance estimates illuminated by Mitchell and Glenn [13] detract from the interpretability and reliability of MTMF outputs. As an alternative, investigators might consider coarser metrics for assessing subpixel abundance (e.g., discretizing the continuous abundance values into bins using a clustering algorithm) that offer accuracy at scales that land managers will still find applicable to their needs.

## 5. Conclusions

This study has highlighted a subjective element of the mixture tuned matched filtering (MTMF) classification process and has drawn attention to published MTMF-based map accuracies that may be overly optimistic, ultimately calling into question traditionally post-processed MTMF results. We have proposed a way to reduce the subjective human input during post-processing workflows. Cross-validated supervised learning algorithms, as implemented using the caret package [61,62] in R, provide a robust, repeatable framework for maximizing map accuracies while simultaneously reducing artificial inflation of those accuracies. Through a case study of a common endmember—the forb leafy spurge (*Euphorbia esula*)—we illustrate an automated post-processing workflow and present map accuracy values alongside measures of their variability that we have produced using rigorous cross-validation. Our approach can be easily extended from binary classification to multi-class problems and those with continuous value outputs.

We recommend that future MTMF scholarship report a full suite of accuracy statistics on models that have been subjected to thorough cross-validation protocols. Additional research is also needed in evaluating our approach with respect to different strategies for binning MF values prior to abundance-based classification. As well, it appears critical that supervised learning methods be applied to balanced reference data.

Lastly, it is important to emphasize that the techniques described in this study are not solely relevant to our target plant species. They can be readily adapted for use in monitoring a wide variety of natural and man-made sub-pixel cover types [91] from a variety of hyperspectral sensors. With an increasing number of semi-automated, open-access tools for accessing and manipulating these algorithms and their associated datasets, analysts have an expanding ability to apply these techniques to new geospatial challenges [36,92].

**Author Contributions:** Conceptualization, D.R., L.S., C.B. and H.B.G.; Formal analysis, D.R., L.S. and H.B.G.; Funding acquisition, C.D.O.; Investigation, D.R., L.S., C.B., C.K. and H.B.G.; Methodology, D.R., L.S., C.B. and H.B.G.; Project administration, H.B.G.; Resources, C.D.O.; Supervision, H.B.G. and C.D.O.; Validation, D.R., L.S., C.B., C.K. and H.B.G.; Visualization, L.S. and H.B.G.; Writing—original draft, D.R., L.S. and H.B.G.; Writing—review & editing, D.R., L.S., C.B., C.K., C.D.O. and H.B.G.

**Funding:** This research was made possible by the Ucross High Plains Stewardship Initiative and a Google Earth Engine Research Award.

**Acknowledgments:** The authors would like to thank Stuart Frye at the NASA Goddard Space Flight Center, Apache Foundation management, Bauer Land and Livestock, and the Ucross Foundation for technical and logistical support, and four anonymous reviewers for constructive feedback on the manuscript. This study would not have been possible without the support of R. Plank and the Ucross High Plains Stewardship Initiative.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Boardman, J.W.; Kruse, F.A. Analysis of imaging spectrometer data using N-dimensional geometry and a mixture-tuned matched filtering approach. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4138–4152. [\[CrossRef\]](#)
2. Boardman, J.W. Leveraging the high dimensionality of AVIRIS data for improved sub-pixel target unmixing and rejection of false positives: Mixture tuned matched filtering. In *Proceedings of the Summaries of the Seventh JPL Airborne Geoscience Workshop*; JPL: Pasadena, CA, USA, 1998; pp. 55–56.
3. Parker Williams, A.E.; Hunt, E.R., Jr. Estimation of leafy spurge cover from hyperspectral imagery using mixture tuned matched filtering. *Remote Sens. Environ.* **2002**, *82*, 446–456. [\[CrossRef\]](#)
4. Glenn, N.F.; Mundt, J.T.; Weber, K.T.; Prather, T.S.; Lass, L.W.; Pettingill, J. Hyperspectral data processing for repeat detection of small infestations of leafy spurge. *Remote Sens. Environ.* **2005**, *95*, 399–412. [\[CrossRef\]](#)
5. Wang, J.-J.; Zhou, G.; Zhang, Y.; Bussink, C.; Zhang, J.; Ge, H. An unsupervised mixture-tuned matched filtering-based method for the remote sensing of opium poppy fields using EO-1 Hyperion data: An example from Helmand, Afghanistan. *Remote Sens. Lett.* **2016**, *7*, 945–954. [\[CrossRef\]](#)
6. Boardman, J.W. Automating spectral unmixing of AVIRIS data using convex geometry concepts. In *Proceedings of the 4th JPL Airborne Geoscience Workshop*, Pasadena, CA, USA, 25–29 October 1993; pp. 11–14.
7. Green, A.A.; Berman, M.; Switzer, P.; Craig, M.D. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosci. Remote Sens.* **1988**, *26*, 65–74. [\[CrossRef\]](#)
8. Mundt, J.T.; Streutker, D.R.; Glenn, N.F. Partial unmixing of hyperspectral imagery: Theory and methods. In *Proceedings of the American Society of Photogrammetry and Remote Sensing*, Tampa, FL, USA, 7–11 May 2007.
9. Chang, C.-I. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*; Springer Science & Business Media: New York, NY, USA, 2003; Volume 1.
10. Heinz, D.C. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 529–545. [\[CrossRef\]](#)
11. Okin, G.S.; Roberts, D.A.; Murray, B.; Okin, W.J. Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. *Remote Sens. Environ.* **2001**, *77*, 212–225. [\[CrossRef\]](#)
12. Settle, J.J.; Drake, N.A. Linear mixing and the estimation of ground cover proportions. *Int. J. Remote Sens.* **1993**, *14*, 1159–1177. [\[CrossRef\]](#)

13. Mitchell, J.J.; Glenn, N.F. Subpixel abundance estimates in mixture-tuned matched filtering classifications of leafy spurge (*Euphorbia esula* L.). *Int. J. Remote Sens.* **2009**, *30*, 6099–6119. [[CrossRef](#)]
14. Carrino, T.A.; Crósta, A.P.; Toledo, C.L.B.; Silva, A.M. Hyperspectral remote sensing applied to mineral exploration in Southern Peru: A multiple data integration approach in the chapi chiara gold prospect. *Int. J. Appl. Earth Observ. Geoinf.* **2018**, *64*, 287–300. [[CrossRef](#)]
15. Dos Reis Salles, R.; De Souza Filho, C.R.; Cudahy, T.; Vicente, L.E.; Monteiro, L.V.S. Hyperspectral remote sensing applied to uranium exploration: A case study at the Mary Kathleen metamorphic-hydrothermal U-REE deposit, NW, Queensland, Australia. *J. Geochem. Explor.* **2017**, *179*, 36–50. [[CrossRef](#)]
16. Calin, M.-A.; Coman, T.; Parasca, S.V.; Bercaru, N.; Savastru, R.S.; Manea, D. Hyperspectral imaging-based wound analysis using mixture-tuned matched filtering classification method. *J. Biomed. Opt.* **2015**, *20*, 046004. [[CrossRef](#)] [[PubMed](#)]
17. Anderson, G.L.; Everitt, J.H.; Escobar, D.E.; Spencer, N.R.; Andrascik, R.J. Mapping leafy spurge (*Euphorbia esula*) infestations using aerial photography and geographic information systems. *Geocarto Int.* **1996**, *11*, 81–89. [[CrossRef](#)]
18. Casady, G.M.; Hanley, R.S.; Seelan, S.K. Detection of leafy spurge (*Euphorbia esula*) using multirate high-resolution satellite imagery. *Weed Technol.* **2005**, *19*, 462–467. [[CrossRef](#)]
19. Dudek, K.B.; Root, R.R.; Kokaly, R.F.; Anderson, G.L. Increased spatial and temporal consistency of leafy spurge maps from multirate AVIRIS imagery: A modified, hybrid linear spectral mixture analysis/mixture-tuned matched filtering approach. In Proceedings of the 13th JPL Airborne Earth Sciences Workshop, Tampa, FL, USA, 7–11 May 2007.
20. Everitt, J.H.; Anderson, G.L.; Escobar, D.E.; Davis, M.R.; Spencer, N.R.; Andrascik, R.J. Use of remote sensing for detecting and mapping leafy spurge (*Euphorbia esula*). *Weed Technol.* **1995**, *9*, 599–609. [[CrossRef](#)]
21. Hunt, E.R., Jr.; Parker Williams, A.E. Comparison of AVIRIS and multispectral remote sensing for detection of leafy spurge. In Proceedings of the 13th JPL Airborne Earth Sciences Workshop, Tampa, FL, USA, 7–11 May 2007.
22. Hunt, E.R., Jr.; Parker Williams, A.E. Detection of flowering leafy spurge with satellite multispectral imagery. *Rangel. Ecol. Manag.* **2006**, *59*, 494–499. [[CrossRef](#)]
23. Hunt, E.R., Jr.; Everitt, J.H.; Ritchie, J.C.; Moran, M.S.; Booth, D.T.; Anderson, G.L.; Clark, P.E.; Seyfried, M.S. Applications and research using remote sensing for rangeland management. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 675–693. [[CrossRef](#)]
24. Kokaly, R.F.; Root, R.R.; Brown, K.; Anderson, G.L.; Hager, S. Mapping the invasive species leafy spurge (*Euphorbia esula*) in Theodore Roosevelt National Park using field measurements of vegetation spectra and imaging spectroscopy data. In Proceedings of the 2002 Denver Annual Meeting, Denver, CO, USA, 18–20 February 2002.
25. Kokaly, R.F.; Anderson, G.L.; Root, R.R.; Brown, K.E.; Mladinich, C.S.; Hager, S.; Dudek, K.B. Mapping leafy spurge by identifying signatures of vegetation field spectra in compact airborne spectrographic imager (CASI) data. In Proceedings of the Society for Range Management 57th Annual Meeting, Salt Lake City, UT, USA, 10–14 February 2004; pp. 249–260.
26. Lawrence, R.L.; Wood, S.D.; Sheley, R.L. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sens. Environ.* **2006**, *100*, 356–362. [[CrossRef](#)]
27. Parker Williams, A.E.; Hunt, E.R., Jr. Accuracy assessment for detection of leafy spurge with hyperspectral imagery. *J. Range Manag.* **2004**, *57*, 106–112. [[CrossRef](#)]
28. Parker Williams, A.E. Biological Control and Hyperspectral Remote Sensing of Leafy Spurge (*Euphorbia esula* L.) an Exotic Plant Species in North America. Ph.D. Thesis, University of Wyoming, Laramie, WY, USA, 2001.
29. Root, R.R.; Anderson, G.L.; Hager, S.N.; Brown, K.E.; Dudek, K.B.; Kokaly, R.F. Application of advanced remote sensing and modeling techniques for the detection and management of leafy spurge: Challenges and opportunities. In Proceedings of the Society for Range Management 57th Annual Meeting, Salt Lake City, UT, USA, 10–14 February 2004; pp. 24–30.
30. Babu, M.J.R.; Rao, E.N.D.; Kallempudi, L.; Chandra, D.I. Mapping of Aluminous Rich Laterite Depositions through Hyper Spectral Remote Sensing. *Int. J. Geosci.* **2018**, *9*, 93–105. [[CrossRef](#)]
31. Gregoire, T.G.; Valentine, H.T. *Sampling Strategies for Natural Resources and the Environment*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2004.

32. Mannel, S.; Price, M.; Hua, D. Impact of reference datasets and autocorrelation on classification accuracy. *Int. J. Remote Sens.* **2011**, *32*, 5321–5330. [CrossRef]
33. Duro, D.C.; Franklin, S.E.; Dubé, M.G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens. Environ.* **2012**, *118*, 259–272. [CrossRef]
34. Huang, C.; Davis, L.S.; Townshend, J.R.G. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **2002**, *23*, 725–749. [CrossRef]
35. Jenks, G.F. The data model concept in statistical mapping. *Int. Yearbook Cartogr.* **1967**, *7*, 186–190.
36. Routh, D.; Seegmiller, L. *Mixture Tuned Match Filtering (MTMF) Classification (MTMF\_V1.0): Raw Google Earth Engine Algorithm Code*. Ucross High Plains Stewardship Initiative. 2015. Available online: <https://code.earthengine.google.com/c66847b2b0ed5c332a912c94bdb8f05c> (accessed on 23 October 2018).
37. Aspinall, R.J. Use of logistic regression for validation of maps of the spatial distribution of vegetation species derived from high spatial resolution hyperspectral remotely sensed data. *Ecol. Model.* **2002**, *157*, 301–312. [CrossRef]
38. Caruana, R.; Niculescu-Mizil, A. An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 161–168.
39. Rogan, J.; Franklin, J.; Stow, D.; Miller, J.; Woodcock, C.; Roberts, D. Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote Sens. Environ.* **2008**, *112*, 2272–2283. [CrossRef]
40. Srivastava, P.K.; Han, D.; Rico-Ramirez, M.A.; Bray, M.; Islam, T. Selection of classification techniques for land use/land cover change investigation. *Adv. Space Res.* **2012**, *50*, 1250–1265. [CrossRef]
41. Vink, J.P.; De Haan, G. Comparison of machine learning techniques for target detection. *Artif. Intell. Rev.* **2015**, *43*, 125–139. [CrossRef]
42. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [CrossRef]
43. Camps-Valls, G.; Gómez-Chova, L.; Calpe-Maravilla, J.; Martín-Guerrero, J.D.; Soria-Olivas, E.; Alonso-Chordá, L.; Moreno, J. Robust support vector method for hyperspectral data classification and knowledge discovery. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1530–1542. [CrossRef]
44. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
45. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]
46. Erol, H.; Tyoden, B.M.; Erol, R. Classification performances of data mining clustering algorithms for remotely sensed multispectral image data. In Proceedings of the 2018 Innovations in Intelligent Systems and Applications (INISTA), Thessaloniki, Greece, 3–5 July 2018; pp. 1–4.
47. Siedliska, A.; Baranowski, P.; Zubik, M.; Mazurek, W.; Sosnowska, B. Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging. *Postharvest Biol. Technol.* **2018**, *139*, 115–126. [CrossRef]
48. Rodriguez-Galiano, V.F.; Chica-Olmo, M.; Abarca-Hernandez, F.; Atkinson, P.M.; Jeganathan, C. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens. Environ.* **2012**, *121*, 93–107. [CrossRef]
49. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
50. Pullanagari, R.; Kereszturi, G.; Yule, I. Integrating airborne hyperspectral, topographic, and soil data for estimating pasture quality using recursive feature elimination with random forest regression. *Remote Sens.* **2018**, *10*, 1117. [CrossRef]
51. Son, N.-T.; Chen, C.-F.; Chen, C.-R.; Minh, V.-Q. Assessment of Sentinel-1A data for rice crop classification using random forests and support vector machines. *Geocarto Int.* **2018**, *33*, 587–601. [CrossRef]
52. Atkinson, P.M.; Tatnall, A.R.L. Introduction neural networks in remote sensing. *Int. J. Remote Sens.* **1997**, *18*, 699–709. [CrossRef]
53. Paola, J.D.; Schowengerdt, R.A. A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *Int. J. Remote Sens.* **1995**, *16*, 3033–3058. [CrossRef]

54. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [\[CrossRef\]](#)
55. Debella-Gilo, M.; Etzelmüller, B. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vestfold County, Norway. *Catena* **2009**, *77*, 8–18. [\[CrossRef\]](#)
56. Afshar, F.A.; Ayoubi, S.; Jafari, A. The extrapolation of soil great groups using multinomial logistic regression at regional scale in arid regions of Iran. *Geoderma* **2018**, *315*, 36–48. [\[CrossRef\]](#)
57. Bandos, T.V.; Bruzzone, L.; Camps-Valls, G. Classification of hyperspectral images with regularized linear discriminant analysis. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 862–873. [\[CrossRef\]](#)
58. Eikvil, L.; Aurdal, L.; Koren, H. Classification-based vehicle detection in high-resolution satellite images. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 65–72. [\[CrossRef\]](#)
59. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, CA, USA, 20–25 August 1995; pp. 1137–1145.
60. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016.
61. Kuhn, M. Package ‘Caret’, v6.0-80. The Comprehensive R Archive Network (CRAN). 2018, p. 213. Available online: <https://cran.r-project.org/> (accessed on 23 October 2018).
62. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **2008**, *28*, 1–26. [\[CrossRef\]](#)
63. Gudex-Cross, D.; Pontius, J.; Adams, A. Enhanced forest cover mapping using spectral unmixing and object-based classification of multi-temporal Landsat imagery. *Remote Sens. Environ.* **2017**, *196*, 193–204. [\[CrossRef\]](#)
64. Mikheeva, A.I.; Tutubalina, O.V.; Zimin, M.V.; Golubeva, E.I. A Subpixel Classification of Multispectral Satellite Imagery for Interpretation of Tundra-Taiga Ecotone Vegetation (Case Study on Tuliok River Valley, Khibiny, Russia). *Izvestiya Atmos. Ocean. Phys.* **2017**, *53*, 1164–1173. [\[CrossRef\]](#)
65. *ENVI, v. 5.1*; Exelis Visual Information Solutions: Boulder, CO, USA, 2015.
66. Lym, R.G. The biology and integrated management of leafy spurge (*Euphorbia esula*) on North Dakota rangeland. *Weed Technol.* **1998**, *12*, 367–373. [\[CrossRef\]](#)
67. Mooney, H.A.; Cleland, E.E. The evolutionary impact of invasive species. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5446–5451. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Walker, J.W.; Kronberg, S.L.; Al-Rowaily, S.L.; West, N.E. Comparison of sheep and goat preferences for leafy spurge. *J. Range Manag.* **1994**, 429–434. [\[CrossRef\]](#)
69. Pearlman, J.S.; Barry, P.S.; Segal, C.C.; Shepanski, J.; Beiso, D.; Carman, S.L. Hyperion, a space-based imaging spectrometer. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1160–1173. [\[CrossRef\]](#)
70. Hunt, E.R.; McMurtrey, J.E.; Parker Williams, A.E. Spectral characteristics of leafy spurge (*Euphorbia esula*) leaves and flower bracts. *Weed Sci.* **2004**, *52*, 492–497. [\[CrossRef\]](#)
71. Perkins, T.; Adler-Golden, S.; Cappelaere, P.; Mandl, D. High-speed atmospheric correction for spectral image processing. In Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, Baltimore, MD, USA, 23–27 April 2012; p. 83900V.
72. Goodenough, D.G.; Dyk, A.; Niemann, K.O.; Pearlman, J.S.; Chen, H.; Han, T.; Murdoch, M.; West, C. Processing Hyperion and ALI for forest classification. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1321–1331. [\[CrossRef\]](#)
73. Scheffler, D.; Karrasch, P. Destriping of hyperspectral image data: An evaluation of different algorithms using EO-1 Hyperion data. *J. Appl. Remote Sens.* **2014**, *8*, 083645. [\[CrossRef\]](#)
74. Aktaruzzaman, A. *Simulation and Correction of Spectral Smile Effect and Its Influence on Hyperspectral Mapping*; International Institute for Geo-Information Science and Earth Observation: Enschede, The Netherlands, 2008.
75. Dadon, A.; Ben-Dor, E.; Karnieli, A. Use of derivative calculations and minimum noise fraction transform for detecting and correcting the spectral curvature effect (smile) in Hyperion images. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2603–2612. [\[CrossRef\]](#)
76. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [\[CrossRef\]](#)
77. Pontius, R.G., Jr.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [\[CrossRef\]](#)

78. Glick, H.B.; Routh, D.; Bettigole, C.; Oliver, C.D.; Seegmiller, L.; Kuhn, C. Modeling the effects of horizontal positional error on classification accuracy statistics. *Photogramm. Eng. Remote Sens.* **2016**, *82*, 789–802. [[CrossRef](#)]
79. Pontius, J.; Hallett, R.; Martin, M. Using AVIRIS to assess hemlock abundance and early decline in the Catskills, New York. *Remote Sens. Environ.* **2005**, *97*, 163–173. [[CrossRef](#)]
80. Pontius, J.; Hanavan, R.P.; Hallett, R.A.; Cook, B.D.; Corp, L.A. High spatial resolution spectral unmixing for mapping ash species across a complex urban environment. *Remote Sens. Environ.* **2017**, *199*, 360–369. [[CrossRef](#)]
81. Pal, M.; Mather, P.M. Support vector machines for classification in remote sensing. *Int. J. Remote Sens.* **2005**, *26*, 1007–1011. [[CrossRef](#)]
82. Brenning, A. Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection. *Remote Sens. Environ.* **2009**, *113*, 239–247. [[CrossRef](#)]
83. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)] [[PubMed](#)]
84. Franke, J.; Barradas, A.C.S.; Borges, M.A.; Costa, M.M.; Dias, P.A.; Hoffmann, A.A.; Orozco Filho, J.C.; Melchiori, A.E.; Siegert, F. Fuel load mapping in the Brazilian Cerrado in support of integrated fire management. *Remote Sens. Environ.* **2018**, *217*, 221–232. [[CrossRef](#)]
85. Ayoobi, I.; Tangestani, M.H. Evaluation of subpixel unmixing algorithms in mapping the porphyry copper alterations using EO-1 Hyperion data, a case study from SE Iran. *Remote Sens. Appl. Soc. Environ.* **2018**, *10*, 120–127. [[CrossRef](#)]
86. Castillo-Santiago, M.A.; Ricker, M.; De Jong, B.H. Estimation of tropical forest structure from SPOT-5 satellite images. *Int. J. Remote Sens.* **2010**, *31*, 2767–2782. [[CrossRef](#)]
87. Crowther, T.W.; Glick, H.B.; Covey, K.R.; Bettigole, C.; Maynard, D.S.; Thomas, S.M.; Smith, J.R.; Hintler, G.; Duguid, M.C.; Amatulli, G.; et al. Mapping tree density at a global scale. *Nature* **2015**, *525*, 201–205. [[CrossRef](#)] [[PubMed](#)]
88. Hengl, T.; De Jesus, J.M.; Heuvelink, G.B.M.; Gonzalez, M.R.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e0169748. [[CrossRef](#)] [[PubMed](#)]
89. Anderson, G.L.; Prosser, C.W.; Haggard, S.; Foster, B. Change detection of leafy spurge infestations using aerial photography and geographic information systems. In Proceedings of the 17th Annual Biennial Workshop on Color Aerial Photography and Videography in Natural Resource Assessment, Medora, ND, USA, 29 June 1999; pp. 5–7.
90. Kruse, F.A.; Boardman, J.W.; Huntington, J.F. Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1388–1400. [[CrossRef](#)]
91. Underwood, E.; Ustin, S.; DiPietro, D. Mapping nonnative plants using hyperspectral imagery. *Remote Sens. Environ.* **2003**, *86*, 150–161. [[CrossRef](#)]
92. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]

