Regional Scale Dryland Vegetation Classification with an Integrated Lidar-Hyperspectral Approach

Hamid Dashti 1,*, Andrew Poley 2, Nancy F. Glenn 1, Nayani Ilangakoon 1, Lucas Spaete 3, Dar Roberts 4, Josh Enterkine 1, Alejandro N. Flores 1, Susan L. Ustin 5 and Jessica J. Mitchell 6

1 Department of Geosciences, Boise State University, 1910 W University Dr, Boise, ID 83725, USA; nancyglenn@boisestate.edu (N.F.G.); nayaniilangakoon@u.boisestate.edu (N.I.); joshenterkine@boisestate.edu (J.E.); lejoflores@boisestate.edu (A.N.F.)
2 Michigan Tech Research Institute, 3600 Green Court, Suite 100, Ann Arbor, MI 48105, USA; afpoley@mtu.edu
3 Minnesota Department of Natural Resources, 1201 US-2, Grand Rapids, MN 55744, USA; lspaete@gmail.com
4 Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA; dar@geog.ucsb.edu
5 Department of Land, Air, and Water Resources, University of California, Davis, CA 93106-4060, USA; slustin@ucdavis.edu
6 Montana Natural Heritage Program, University of Montana, Missoula, MT 59812, USA; jessica.mitchell@mso.umt.edu

* Correspondence: hamiddashti@u.boisestate.edu; Tel.: +1-208-991-8748

Received: 20 August 2019; Accepted: 10 September 2019; Published: 14 September 2019

Abstract: The sparse canopy cover and large contribution of bright background soil, along with the heterogeneous vegetation types in close proximity, are common challenges for mapping dryland vegetation with remote sensing. Consequently, the results of a single classification algorithm or one type of sensor to characterize dryland vegetation typically show low accuracy and lack robustness. In our study, we improved classification accuracy in a semi-arid ecosystem based on the use of vegetation optical (hyperspectral) and structural (lidar) information combined with the environmental characteristics of the landscape. To accomplish this goal, we used both spectral angle mapper (SAM) and multiple endmember spectral mixture analysis (MESMA) for optical vegetation classification. Lidar-derived maximum vegetation height and delineated riparian zones were then used to modify the optical classification. Incorporating the lidar information into the classification scheme increased the overall accuracy from 60% to 89%. Canopy structure can have a strong influence on spectral variability and the lidar provided complementary information for SAM’s sensitivity to shape but not magnitude of the spectra. Similar approaches to map large regions of drylands with low uncertainty may be readily implemented with unmixing algorithms applied to upcoming space-based imaging spectroscopy and lidar. This study advances our understanding of the nuances associated with mapping xeric and mesic regions, and highlights the importance of incorporating complementary algorithms and sensors to accurately characterize the heterogeneity of dryland ecosystems.

Keywords: drylands; classification; SAM; MESMA; LiDAR

1. Introduction

Vegetation in semi-arid ecosystems (i.e., drylands) plays an important role in regulating the global carbon balance. As an example, 51% of the global net C sink during 2011 was attributed to three Southern Hemisphere semi-arid regions, with Australia being a global hotspot [1,2]. Yet, differentiating vegetation species and their respective roles in regional scale carbon dynamics in semi-arid and other dryland ecosystem types remain challenging. In many drylands, large environmental gradients (e.g., elevation) and variability in climate over landscapes with complex topography create conditions for
Remote Sens. 2019, 11, 2141 2 of 16
disparate biomes to exist within close proximity to each other [3,4]. Wetter, higher elevation sites and riparian areas are populated with alpine and deciduous vegetation, respectively, while drier, lower elevation landscapes can be dominated by shrubland vegetation. This heterogeneity makes mapping and quantifying vegetation species and structure inherently difficult at regional scales in drylands.

Hyperspectral remote sensing has been used to classify vegetation species within different biomes across the globe [5–9]. Upcoming imaging spectroscopy sensors that are satellite-based (e.g., PRISMA (recently launched), EnMAP, SBG, and CHIME) or on the International Space Station (ISS) (e.g., DESIS, HISUI, and EMIT) will allow large area mapping. While hyperspectral remote sensing provides the advantage of narrow spectral bands across important regions for vegetation analysis, optical classification over large areas of drylands faces several challenges [10,11]. The first challenge is spectral mixing, which is the contribution of different endmembers (EMs) to the total optical signal of a pixel received by the sensor [12,13]. The spectral mixing problem can be severe in drylands, and in particular, xeric areas where the vegetation cover is typically a mix of shrubs, soil, and grasses. Further, soil can be a mixture of inorganic and organic matter including litter and cryptobiotic crust. The influence of one EM may make the detection of other contributing classes challenging. For example, a recent study shows soil can contribute up to 95% to the total radiation in shrublands in the Great Basin, U.S. [12]. The second challenge is in the spectral heterogeneity of these complex ecosystems over a large spatial extent. The spectral variability of vegetation is mainly a function of canopy structure (e.g., height and leaf area) and biochemistry, such as chlorophyll content [14–18]. For example, different tree and shrub species might have similar spectral signatures due to complex interactions between canopy structure, canopy biochemistry, and light [19,20]. Moreover, drylands are typically subject to environmental (e.g., precipitation, temperature, soil moisture, etc.) gradients that occur over fine spatial and temporal scales. In particular, the spectral heterogeneity becomes more complex at finer spatial resolution (e.g., 1 × 1 m pixel size). These environmental changes are mostly a function of the topography of the watershed. For example, in Great Basin semi-arid ecosystems in the western U.S., higher elevations typically receive more precipitation than lower elevations which favor denser vegetation composition. Even a small change in environmental conditions can have a large influence on vegetation traits (i.e., biochemistry and structure) and composition, which ultimately affects the optical signature. Thus, to fully solve the classification problem in drylands we need to consider canopy biochemistry, structure, and environmental variables. Such a task cannot be addressed using a single method or sensor and requires a framework that incorporates multiple methods and sensors. The main goal of this study is to develop a framework to map vegetation in drylands over a large landscape. Due to different optical properties we break down the study area into xeric and mesic regions. Following that, we discuss different remote sensing techniques that inform approaches that can be used for the classification of different xeric and mesic vegetation and soil.

In xeric regions, traditional pixel-based approaches may result in unreliable classifications due to the mixed pixels [13,21,22]. An alternative is to use sub-pixel classification techniques (for a review refer to [23]). The main idea behind the unmixing approaches is to retrieve the fractional contribution of each EM in the total radiation. There are three different approaches for sub-pixel classification: (1) spectral mixture analysis (SMA), (2) regression methods, and (3) soft classification methods based on fuzzy algorithms [13]. In this study, our focus is on spectral mixture analysis. The SMA methods are based on the inversion of a mixture model to retrieve the relative abundance of each endmember present in a pixel. Depending on assumptions such as linear or nonlinear spectral mixture and algorithm constraints, different SMA methods have been proposed [13,24–28]. One of the widely adopted unmixing methods is multiple endmember spectral mixture analysis (MESMA), developed as an approach to deal with complex systems [22]. MESMA is sensitive to changes in the spectral albedo of an EM rather than its shape. In other words, MESMA is able to detect small changes in the magnitude of the spectra. This property makes it a suitable choice for drylands where the spectral variability of xeric areas are generally low.
In mesic sites where vegetation is dense, the problem of spectral mixing is not as severe as in xeric regions and pixel-based methods can be used for classification. A common pixel-based classification approach is the spectral angle mapper (SAM). The SAM compares image spectra with reference spectra, or endmembers, by mapping each spectrum as a vector in an n-dimensional space and measuring the angle between the two vectors. SAM only considers the shape and is insensitive to the magnitude of the spectra [29]. SAM has been successfully used to classify evergreen and deciduous forest [30], juniper expansion [31], and over mountainous terrain [8]. However, mesic classification becomes increasingly challenging over larger landscapes (e.g., several hundred km$^2$) in heterogeneous systems. For example, several species can coexist in close proximity in riparian zones. Moreover, riparian vegetation along the same stream corridor can be significantly different dependent upon elevation. This environmental gradient leads to increased spectral variability when a sufficiently large study area is observed on a single acquisition date (e.g., large-swath airborne applications and satellite-based observations). The high spectral variability of mesic regions makes classification a challenging task using just optical information.

Discrete return small footprint lidar has been successfully used in various vegetation mapping studies (e.g., [32–35]); however, the data are typically insufficient for characterizing sparse semi-arid vegetation cover at high spatial resolutions [36,37]. Most recently, full-waveform lidar features such as pulse width, rise time, and derived height have been shown to improve discrimination of dryland components such as bare ground, grasses, and shrubs, but may need to be resampled to a coarser spatial resolution (e.g., 10 m × 10 m cells; [38]). Moreover, techniques for deriving full-waveform features are generally sophisticated and computationally expensive which limits their practical applications. However, in paying attention to environmental variability of drylands and changes in vegetation structure, one can observe that some simple lidar metrics provide complimentary information that can improve optical classifications. For example, previous studies have demonstrated that canopy structure provides a distinguishing difference between vegetation cover in riparian zones and other dense vegetation [39,40], and canopy height is a particularly influential variable [41]. The ability of lidar metrics such as canopy height to characterize three-dimensional canopy structure provides complimentary information that improves land cover classifications [42], [43]. Moreover, lidar data can be used to derive a high spatial resolution digital elevation model (DEM) of the terrain. Important environmental variables that influence vegetation distribution patterns can be derived from a DEM. These variables include slope, aspect, and elevation. In addition, DEMs support flow accumulation modeling that can be combined with topographical information to delineate riparian zones. Topographic-based riparian delineations can be integrated into classification schemes to improve the accuracy of riparian vegetation mapping [44,45]. In this study, we focus on the use of state-of-the-art airborne lidar for improving confusion between riparian and other mesic vegetation classes. A combination of sensors and methods is recommended to properly address the limitations of hyperspectral and lidar datasets and the challenges of mesic and xeric regions in drylands [46–49].

The goal of this study is to develop a straightforward framework to classify vegetation cover at a watershed scale (several hundred km$^2$) in a dryland ecosystem. Our primary objective is to develop two reliable mapping products for a watershed located in the drylands of the western U.S. One product is an abundance map for xeric area components (soil, grass, and shrub) and the second product is a hard classification map for mesic areas (aspen, Douglas fir, juniper, and riparian). The main assumptions are: 1) hyperspectral information is sufficient for classification of xeric areas and 2) canopy structural information and riparian zone delineation from lidar improves optical classification in mesic areas. To accomplish our goal, we integrate the optical classification techniques SAM and MESMA, and lidar-derived products, to map soil, grass, shrub, aspen, Douglas fir, juniper, and riparian vegetation across a watershed with a significant environmental gradient. This work builds on previous remote sensing of vegetation studies in the study area where the focus was on local scales or species-specific analyses (e.g., [36–38,50,51]). This work also builds on lessons learned from previous efforts to map vegetation structure (e.g., shrub biomass; [52]) and biochemical information (i.e., sagebrush foliar
N content; [12]) across the watershed by developing a framework that combines the lidar-derived structural variable canopy height with a hyperspectral unmixing approach to addressing fine scale combinations of soil with upland grasses and shrubs.

2. Materials and Methods

This study was conducted in the Reynolds Creek Experimental Watershed (RCEW), located in the Owyhee Mountains, Idaho, U.S. (Figure 1). The RCEW encompasses 23,900 ha of land in a semi-arid ecosystem with an elevation gradient of 900–2200 m. Annual mean precipitation ranges from 250–1100 mm and increases linearly with elevation. The RCEW can be divided into three major biomes with boundaries that are defined by both sharp and gradual ecotones. Alpine vegetation dominates higher elevations in the southern portion of the watershed and consists primarily of quaking aspen (*Populus tremuloides*), Douglas fir (*Pseudotsuga menziesii*), and western juniper (*Juniperus occidentalis*) [53]. There is a gradual transition into the lower elevations in the northern portion of the watershed from alpine vegetation to primarily Wyoming big sagebrush (*Artemisia tridentata* ssp. *wyomingensis*), low sagebrush (*Artemisia arbuscula*), rabbitbrush (*Ericameria nauseosa*), and bitterbrush (*Pushia tridentata*). This ecotone contains western juniper that is expanding into lower elevation shrub communities. Common species in lower elevation grass communities include bluebunch wheatgrass (*Pseudoroegneria spicata*), needle and thread (*Hesperostipa comata*), western wheatgrass (*Pascopyrum smithii*), tapertip hawksbeard (*Crepis acuminata*), and yarrow (*Achillea millefolium*) [54]. The third biome within RCEW is defined as the riparian areas of the watershed. Common riparian vegetation in the region includes black cottonwood (*Populus trichocarpa*), coyote willow (*Salix exigua*), and peachleaf willow (*Salix amygdaloides*) [55]. Riparian regions transition gradually into alpine vegetation in the higher elevation areas; whereas the transition between riparian regions and shrublands is much sharper in the lower elevations and is defined by mesic areas (e.g., streams).

![Figure 1. Elevation range and location of plots at Reynolds Creek Experimental Watershed, Idaho, U.S.](image)

2.1. Data

Field data were collected in 2014–2016 that support vegetation cover type mapping for RCEW. Shrub, grass, and soil cover was measured during September–November of 2014 and May and June
of 2015 at 48 plots (10 × 10 m) distributed across a range of elevation, cover, and species. Sample point photo analysis [56] was used to classify each plot using a series of 20 photos taken every 2 m across each plot. We derived percent cover of soil and vegetation from the photo analysis, and used this information in endmember derivation and accuracy assessment for the xeric classification (Section 2.2.2). A full description of the field data collection is presented in [12,38]. Dominant vegetation cover for the collected plots included sagebrush, bitterbrush, rabbitbrush, mixed shrub, and grasses. In mesic regions where aspen, Douglas fir, juniper, and riparian classes dominated as patches in the landscape, we collected validation points during July of 2016 using a Topcon HiPer V Real Time Kinematic (RTK) GPS.

Imaging spectroscopy data were collected in June 2015 using NASA’s Airborne Visible and Infrared Imaging Spectrometer Next Generation (AVIRIS-ng). AVIRIS-ng collects 432 spectral bands in the optical range (380–2500 nm) with a bandwidth of about 5 nm. The AVIRIS-ng Level 2 product, consisting of orthorectified surface reflectance atmospherically corrected with ATREM [57], was used for this study. A total of 15 AVIRIS-ng images were used to capture the full extent of RCEW. While maintaining a consistent spatial resolution during the data collection, approximately 17 km² (~6%) of the total watershed was not captured (Figure 1), resulting in data gaps between flight lines. These gaps were larger in the southern portion of the watershed where elevations are higher. Spectral bands were inspected for noise caused by atmospheric water absorption, and 59 of the original bands were removed. These bands are located around the strong water absorption regions (1375 and 1875 nm). Bands less than 400 nm were also removed from the analysis due to low signal to noise in this region.

Full-waveform lidar data were acquired in August 2014, using the NASA Airborne Snow Observatory’s Riegl LMS-Q1560 (Riegl Laser Measurement Systems GmbH, Horn, Austria). The scanner footprint is 20–60 cm with ± 30° angle which results in 10–14 points/m². In total 40 lidar flight lines cover the study area. The hyperspectral data were resampled to 1 × 1 m pixels and co-registered to the lidar flightlines. The relative error of this registration is roughly half a pixel or 0.5 m.

2.2. Methods

2.2.1. Classification

Figure 2 shows the framework for classification of xeric and mesic vegetation in the study area. First, the mesic and xeric classes are classified using SAM and MESMA, respectively. We then use lidar to investigate the difference in canopy height among mesic classes and define a height threshold that can be used to separate mesic vegetation into aspen, Douglas fir, juniper, and riparian classes. We also use a DEM generated by lidar to map riparian zones. In the last step, we assign vegetation to mesic classes based on height thresholds and riparian zones.
The open-source software Visualization and Image Processing for Environmental Research (VIPER) [58] was used to compile and build EM bundles from the extracted spectra for each class from multiple images covering the study area. The VIPER toolbox calculates several statistical metrics including EM average root mean square error (EAR) and minimum average spectral angle (MASA) to highlight inter-class spectral variability and ultimately choose spectral bundles that best represent an entire class (for details, see [59]). In general, smaller EAR and MASA values lead to endmembers that are better representative of the associated class [60]. Spectral libraries were reduced to smaller subsets from the originally extracted spectra through an iterative process of taking the top spectra with the lowest EAR and MASA, running the classification process on several random validation plots, and removing spectra that performed poorly or were unused. This process was repeated until removing spectra decreased classification results. This procedure helped reduce computation time and the number of training samples used and maximized the number of validation samples. Table 1 shows the number of spectra per class that were delineated in the initial spectral library, the number of samples used in the final classification, and the number of samples used for classification validation. EM bundle sizes ranged from 2–5 spectra per class. Field plots with spectra that were not used in the final spectral libraries were used to assess the classification accuracy.

### Table 1. The number of ground validation points (represented by 1 m pixels) used for endmember (EM) derivation and classification accuracy assessment per class.

<table>
<thead>
<tr>
<th>Classes</th>
<th>Image Extracted EM</th>
<th>EM Used in Final Classification</th>
<th>Validation Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspen</td>
<td>1004</td>
<td>3</td>
<td>4816</td>
</tr>
<tr>
<td>Douglas fir</td>
<td>90</td>
<td>3</td>
<td>3947</td>
</tr>
<tr>
<td>Juniper</td>
<td>187</td>
<td>3</td>
<td>1409</td>
</tr>
<tr>
<td>Riparian</td>
<td>1316</td>
<td>5</td>
<td>3271</td>
</tr>
<tr>
<td>Shrub</td>
<td>328</td>
<td>7</td>
<td>3400</td>
</tr>
<tr>
<td>Grass</td>
<td>464</td>
<td>2</td>
<td>3400</td>
</tr>
<tr>
<td>Soil</td>
<td>100</td>
<td>3</td>
<td>3400</td>
</tr>
</tbody>
</table>

MESMA was used to map the xeric classes of shrub, grass, and soil. The output of MESMA is the relative abundance of each of these classes in each pixel. The MESMA parameters of minimum and maximum allowable EM fraction, minimum and maximum allowable shade fraction, and maximum allowable root mean square error (RMSE) were set as 0.0, 1.0, 0.0, 0.8, and 0.025, respectively. Model complexity, defined by the allowable number of EMs per pixel, was set to three; this includes the combination of two class EMs from the spectral library and a shade component, which was set to photometric (zero) for this study. Running MESMA with a three-EM model constraint was found to best represent the heterogeneity of the landscape that was observed in the field. The absence of either grass or soil EM in the prediction of shrub abundance led to poor results in initial testing. A total of 56 models were used during the final unmixing process.

SAM was used to classify mesic classes: juniper, Douglas fir, aspen and riparian vegetation. Endmembers derived using VIPER were used for the SAM classification. The maximum allowable spectral angle for SAM was set to 0.1 radians. A post-classification 3 × 3-pixel moving average filter was performed to enhance spatial consistency of the SAM results.

Backscattered full waveform lidar signals were Gaussian decomposed to derive point clouds. A pulse width threshold was used to distinguish ground returns from vegetation returns. A thin plate smoothing function was applied to the ground returns to derive the DEM. The maximum height at each pixel (1 m × 1 m) of the vegetation returns was used as the maximum vegetation height. The canopy height was extracted by subtracting the ground elevations from the maximum height using BCAL Lidar Tools [61]. We used the method developed by [62] to extract the stream network from the DEM. Stream network extraction was implemented using the TauDEM toolbox in the ESRI ArcGIS software (Environmental Systems Research Institute, 2017). We delineated the riparian zone using the buffering technique based on the Strahler order of the stream [63]. Following [64] for semi-arid
ecosystems, we employed a buffer width of 15 m for stream order one, 35 m for stream orders two and three, and 70 m for stream orders larger than three. The canopy height distribution of each mesic class was extracted from the lidar data (Figure 3a) across the watershed. Results indicated that most misclassification was between riparian vegetation and aspen areas. Using the ksdensity function in MATLAB (MathWorks, Inc; 2017), the probability density function (PDF) of height for these two classes was estimated and plotted (Figure 3b).  

The two classes can be reasonably distinguished using a ~7 m height threshold. Thus, in this study we determined the riparian class based on the SAM classification and locations within the riparian boundary and lower than 7 m height.

2.2.2. Classification Accuracy Assessment

MESMA produces abundances of multiple classes in a single pixel, a traditional confusion matrix cannot be directly applied to assess the accuracy. Silván-Cárdenas and Wang, 2008 [65], developed a specialized confusion matrix that evaluates the performance of linear spectral unmixing techniques, termed sub-pixel confusion-uncertainty matrix (SCM). The SCM calculates the accuracy of sub-pixel estimates with the use of field measured class cover instead of determining if a location was correctly classified or not, as in a traditional confusion matrix. The method uses a composite operator, $L_{nij}$, to calculate the SCM:

$$
L_{nij} = \begin{cases}
    \text{MIN}(c_{ni}, f_{nj}), & i = j \\
    (c_{ni} - L_{nii}) \times \frac{f_{nj} - L_{nij}}{\sum_{p=1}^{k} (f_{nj} - L_{nj})}, & i \neq j
\end{cases}
$$

where $L$ is the specific cell in the SCM for pixel $n$, $c$ is the class cover produced from the linear mixture model for pixel $n$, $f$ is the field measure class cover for pixel $n$, $k$ is the total number of classes, and $i$ and $j$ represent the row and column of the SCM, respectively. The SCM was calculated for each plot, then the resulting SCMs were averaged to produce a final SCM. Overall accuracy, user’s and producer’s accuracies were produced from the final SCM for the shrub, grass, and soil percent cover.

Accuracy assessment of the SAM classification was performed using the independent field validation data generated as part of the EM selection process in MESMA (Table 1). The SAM classification was evaluated using standard classification metrics from the confusion matrix (overall accuracy and user’s and producer’s accuracies). The producer’s accuracy shows how often an observed class on the ground is correctly labeled in the final classification map and the user’s accuracy indicates how often a class labeled on the map is found on the ground [65,66].
3. Results

Classification

The final SCM for the MESMA accuracy results were produced using 44 field-measured cover estimates. Overall accuracy derived from the final SCM for the MESMA shrub, grass, and soil cover was 0.67 (Table 2). User’s and producer’s accuracies were primarily above roughly 60%. The producer’s accuracy of the soil class was unusually low; whereas the shrub and grass classes were mapped with higher accuracy. Figure 4 shows the percent cover distribution of xeric classes over the region. The proportion of watershed with dominant percent cover (class cover in pixel > 50%) for grass, shrub and soil is 22.4%, 38.2%, and 16.9%, respectively.

Table 2. User’s and producer’s accuracy for multiple endmember spectral mixture analysis (MESMA) classification.

<table>
<thead>
<tr>
<th>Class</th>
<th>User’s Accuracy</th>
<th>Producer’s Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shrub</td>
<td>0.59</td>
<td>0.99</td>
</tr>
<tr>
<td>Grass</td>
<td>0.76</td>
<td>0.79</td>
</tr>
<tr>
<td>Soil</td>
<td>0.99</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Overall accuracy = 0.67

The results for the mesic classification using the SAM method alone and with the lidar-derived information are presented in Table 3. The SAM classification without the lidar resulted in confusion between the aspen and riparian classes. However, incorporating the lidar-derived canopy height and riparian zones significantly improved the accuracy of these classes. The overall accuracy increased from 60% to 89%. The final mesic and xeric classification maps over the 23,900 ha region were then produced from the 1 m × 1 m pixels from our approach (Figure 5). The proportion of aspen, Douglas fir, juniper and riparian classes are respectively, 2.3%, 0.60%, 0.1%, and 1.5%, of the watershed.

![Figure 4. Distribution of xeric percent cover over the region.](image-url)
Table 3. Confusion matrix for spectral angle mapper (SAM) results alone and with lidar-derived information.

<table>
<thead>
<tr>
<th>Classified</th>
<th>Ground Reference</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ground Reference</td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td>Aspen</td>
<td>2015</td>
</tr>
<tr>
<td></td>
<td>Riparian</td>
<td>2411</td>
</tr>
<tr>
<td></td>
<td>Douglas fir</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Juniper</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>4528</td>
</tr>
</tbody>
</table>

Overall accuracy = 0.60

| Classified incorporating lidar | Ground Reference | Accuracy |
|                                | Aspen            | 4298     | 128     | 398     | 66      | 4890    | 0.87  | 0.94  |
|                                | Riparian         | 129      | 2718    | 130     | 5       | 2982    | 0.91  | 0.95  |
|                                | Douglas fir      | 94       | 13      | 2014    | 100     | 2221    | 0.90  | 0.77  |
|                                | Juniper          | 7        | 0       | 46      | 636     | 689     | 0.92  | 0.78  |
|                                | Total            | 4528     | 2859    | 2588    | 807     | 10782   | —     | —     |

Overall accuracy = 0.89

Figure 5. Final vegetation cover map of RCEW using SAM, MESMA, and lidar-derived products; (left) mesic classes and (right) xeric classes.

4. Discussion

4.1. Xeric Classification

Using high spatial resolution imaging spectroscopy to classify tree and shrub species across ecotones in drylands is effective but has limitations where there is a complex intergrowth of spectrally-similar vegetation species and where there is significant soil exposure. The 67% overall accuracy of xeric classes shows that MESMA was successful in classifying sparse vegetation cover in this dryland watershed. The low producer’s accuracy of the soil class indicates that MESMA did not perform well in classifying the reference soil class; whereas the high user’s accuracy indicates that the majority of classified soil pixels are indeed soil on the ground. These results are likely representative of
our high soil cover and the spectral confusion of soil that is often mixed with litter (non-photosynthetic vegetation). Regardless, our soil percent cover class can be dependably used. The high producer’s accuracy and corresponding lower user’s accuracy of the shrub class indicates an over-classification of shrub within the study area. This may be attributed to the woody biomass of the shrubs, which results in a similar optical response to the mixture of litter and soil. Thus some soil plots with higher amounts of litter may be classified as shrub. This interpretation may also be used to understand the misclassification of the soil class (i.e., low producer’s accuracy). We conclude that the misclassification in xeric areas likely occurs where shrubs have relatively lower photosynthetically-active biomass (i.e., higher woody biomass) and soil has high litter cover. This conclusion is consistent with a thorough physical analysis of the same area using physical radiative transfer modeling [12]. Grasses were mapped surprisingly well at the sub-pixel level with the SCM accuracies. The strong results are likely due to the timing of the imagery (June) when grasses were still photosynthetically-active and thus differentiable from the litter and soil. Figure 4 demonstrates most pixels in the study area include shrub as an endmember, with a normal distribution and mean cover of ~45%. The latter is consistent with our overclassification of shrub cover. The distribution of soil and grass cover is more uniform than shrub cover in the classification, including pixels with low and high soil and grass cover.

4.2. Mesic Classification

As expected, mesic classes cover a small portion of the watershed and are located mostly in higher elevations and along streams where environmental conditions such as precipitation, temperature, and soil moisture favor riparian and denser vegetation cover. SAM was able to distinguish classes that are both spectrally and structurally different (i.e., juniper and Douglas fir) but failed when vegetation is structurally different but spectrally similar (i.e., riparian vs. aspen). Consequently, the main source of classification error in this study was the confusion between aspen trees and riparian vegetation in mesic areas. Vegetation species present within the riparian zones are driven by the surrounding environmental conditions including elevation, precipitation, and soil composition [55,67,68]. The favorable conditions of the riparian areas lead to the presence of multiple species in close proximity. Due to large environmental gradients observed within our study site, it is likely that the composition of vegetation species between riparian areas was also highly varied throughout the study area. At coarser scales (e.g., landscape scale) the influence of canopy structure on spectral variability may be greater than canopy biochemistry [19,20,69]. Thus, the complex structure of vegetation in riparian zones is the main reason for the spectral variability within this class in the region. Significant improvement of classifications using canopy height indicates that the structure of the riparian zone can be explained to a large extent by canopy height which is consistent with the findings of other studies as well [70–72]. Moreover, mapping riparian areas using lidar and incorporating it into classifications not only contributes to the improvement of the classification but also leads to ecologically meaningful results. For example, inspection of the SAM classification (alone) indicated that some pixels classified as riparian were distant from streams and on closer inspection were clearly not riparian vegetation. We should also note that there are multiple approaches for delineating riparian zones in semi-arid ecosystems (refer to [63] for a review), and the method and its performance should be evaluated with respect to the goals of the research. The buffering technique used in this study was appropriate for our mapping goals. Comparable results can be achieved using other lidar-hyperspectral integration schemes such as the deep convolutional neural network [73,74], support vector machines [75], or graph-based methods [76]. However, these methods are not always straightforward and may need more calibration to build a robust classification model. There is no doubt that integrated lidar-hyperspectral classification leads to better results [74,77–79]. The framework and approach described here are straightforward to implement and interpret, and are consistent with our ecological understanding.

Lidar can also be used to improve spectral unmixing, such as in [80]. This study demonstrated that lidar-derived height distributions used for endmember selection and as additional fractional constraints reduced the confusion between spectrally similar classes in an urban area. We did not use a
similar approach because the main source of sub-pixel classification error was between shrub and soil classes, two classes that are not always easily distinguished in point cloud lidar data. In general, lidar do not provide the information to map sparse semi-arid vegetation cover at high spatial resolution. The point density (10–14 pts/m$^2$) in this study was not able to fully distinguish ground from shrubs at 1 × 1 m grid sizes. Airborne lidar requires point aggregation to achieve sufficient density (e.g., 10 × 10 m cells in [38]) whereas spaceborne lidar (such as GEDI) require even larger cells to achieve point density for similar classification purposes. Thus, a combination of sensors and methods is recommended to properly address the limitations of hyperspectral and lidar datasets and the challenges of mesic and xeric regions in drylands [46–49,81].

We did not use MESMA for a mesic subpixel classification due to the high spectral variability in these regions. Our preliminary analysis indicated that one set of EMs selected for a tree class were very similar to another set of EMs for a different tree class (results not shown in this paper), which ultimately resulted in poor classification. Endmember variability is the main source of error in unmixing methods including MESMA [22,82,83]. This problem is exacerbated at 1 × 1 m pixel sizes. Alternatively, the simple classification method of SAM informed by lidar produced accurate mesic maps. Due to the strong influence of canopy structure in the total reflected radiation, the spectral variability in dense vegetation cover is mostly in magnitude of the spectrum rather than its shape [14,16,19,84]. Thus lidar provides complementary information for SAM’s sensitivity to shape (and insensitivity to magnitude) of the spectra.

The large number of AVIRIS-ng images used in this study is noteworthy. The 15 airborne images were collected over approximately two and a half hours to capture the full extent of the study area. During this time solar zenith angle changed from ~22° to ~28° for the first and last flight lines, respectively, and it is likely that illumination differences occurred. Hall et al. [85] reported changes in canopy reflectance due to solar illumination can occur during the sub-hourly scale. Additionally, illumination of vegetation was affected by shading caused by topographic features, surrounding vegetation and the distance of pixels from nadir. These shading factors would have also varied in intensity throughout the image collection. The sum of the soil, grass, and shrub fractions was more than unity in several pixels. The majority of which were located along flightline edges. While these issues are unavoidable when analyzing airborne imaging spectroscopy data at the landscape scale, EM extraction from multiple images for a single class should be sufficient to capture the spectral variability present. In addition, in this study we assumed that the effect of the timing difference between imagery (lidar, imaging spectroscopy) was negligible in regards to changes in vegetation.

The framework developed in this study can be extended to the integration of new spaceborne hyperspectral (e.g., PRISMA) and laser altimetry (e.g., GEDI, ICESat-2). Furthermore, studies such as ours provides opportunities for precursor studies for new or proposed spaceborne hyperspectral sensors (e.g., DESIS, HISUI, EMIT on ISS or EnMAP, CHIME, SBG). The spatial resolution of these sensors are coarser, in comparison to data used in this study. For example, the PRISMA data are 30 m pixels and GEDI’s footprint is roughly 25 m [86]. Thus, integration of these coarse resolution sensors using our framework is a promising approach for large scale (e.g., global scale) vegetation classification of drylands. The xeric and mesic vegetation maps may be used as baseline information for applications including restoration [87] and habitat suitability [88] studies. Mesic vegetation mapping is particularly relevant for drylands where mesic resources drive habitat quality [89]. Importantly, a significant use of vegetation maps in this study is expected from scientific communities seeking to understand carbon and water fluxes and storage within the study area (e.g., [90,91]).

5. Conclusions

Using a combination of optical and lidar data we were able to classify semi-arid vegetation types at high spatial resolution. The main source of misclassification was in mesic regions where riparian and aspen vegetation were confused. Incorporating canopy height and riparian zone delineation significantly improved the mesic classification. MESMA classification of shrub-grass-soil mixed pixels
leads to suitable accuracies; however, some misclassification between soil and shrub was observed. In order to reduce uncertainty in the classification of spectrally diverse areas such as drylands, an approach that employs spectral and structural observations along with requisite data processing methods is required. Different configurations of vegetation biochemical and biophysical characteristics may lead to similar spectrums across optical wavelengths. Our straightforward method provides accurate classification maps of xeric and mesic vegetation in drylands across large heterogeneous areas.


Funding: This research was funded by NASA Terrestrial Ecology, grant number NNX14AD81G, and Department of the Interior Northwest Climate Adaptation Science Center graduate fellowship.

Acknowledgments: Field access and support were provided by USDA ARS Northwest Watershed Research Center, NSF EAR 1331872, and NSF EAR 1665519.

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, or in the decision to publish the results.

References


44. Solomons, A.G.; Mikhailova, E.A.; Post, C.J.; Sharp, J.L. LiDAR-based predictions of flow channels through riparian buffer zones. Water Sci. 2015, 29, 123–133. [CrossRef]


90. Flerchinger, G.N.; Fellows, A.W.; Seyfried, M.S.; Clark, P.E.; Lohse, K.A. Water and Carbon Fluxes Along an Elevation Gradient in a Sagebrush Ecosystem. *Ecosystems 2019, 1–18. [CrossRef]*

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).