REMOTE SENSING TIME-SERIES ANALYSIS, MACHINE LEARNING, AND K-MEANS CLUSTERING IMPROVES DRYLAND VEGETATION AND BIOLOGICAL SOIL CRUST CLASSIFICATION

by

Joshua Enterkine

A thesis
submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Geosciences
Boise State University

May 2019
BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Joshua Enterkine

Thesis Title: Remote Sensing Time-Series Analysis, Machine Learning, And K-Means Clustering Improves Dryland Vegetation and Biological Soil Crust Classification

Date of Final Oral Examination: 09 November 2018

The following individuals read and discussed the thesis submitted by student Joshua Enterkine, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Nancy Glenn, Ph.D.  Chair, Supervisory Committee
Jodi Brandt, Ph.D.  Member, Supervisory Committee
Jennifer Forbey, Ph.D.  Member, Supervisory Committee

The final reading approval of the thesis was granted by Nancy Glenn, Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.
DEDICATION

To science, and my family.
ACKNOWLEDGEMENTS

To everyone who has contributed to me being here: family, friends, mentors, funding sources, and those who came before me and on whose work I build on. I can hardly thank you all enough, and could add many more pages saying so. Thank you all for the support, assistance, understanding, and love (as applicable).
ABSTRACT

Dryland and semi-arid vegetation communities, although appearing to the casual observer as relatively simplistic and homogeneous, are in fact the opposite. Upon further inspection, semi-arid vegetation is highly complex and heterogeneous at almost any scale. The same holds true for biological soil crust. Growing concern about global changes in climate, nutrient cycles, and land use have required increasing scrutiny of our understanding of these communities and all of their constituents, as we seek to improve forecasting models and inform land management decisions. This thesis aims to provide insight to the paradigm of how we create and interpret vegetation classifications in a semi-arid ecosystem.

In the first chapter, I examine the potential of new remote sensing imaging platforms in combination with machine learning algorithms and cloud computing as they apply to time-series analyses for vegetation classification. The results of this indicate that sinusoidal approximations (“Harmonic Models”) of vegetation indices are able to predict vegetation cover with nearly the same accuracy as monthly composites, and that a combination of both perform no better than either. Additionally, I examine how assigning classes to training data (e.g. species-level, plant functional type) influence the classification accuracy, interpretability, and potential uses. Stricter class membership requirements at increasingly aggregated scales (e.g. PFT) lead to greater accuracies.

Finding the implication of this conclusion unsatisfactory – at the extreme, everything was either cheatgrass or bare, the shrub class succumbing almost entirely to
errors of omission – I investigated approaching other methods to assign classes to field data that captured more of the realities of semi-arid vegetation within our study area. To this extent, k-means clustering was used to determine what community classes were present in the field data. The outcome of this approach is a class where each potential constituent cover has a known distribution. Overall accuracies were found to be lower for this approach. However, the classification outcomes quantify overlapping distributions of cover types (e.g. ‘sagebrush’ or ‘shrub’) between classes. These accuracies are assessed using ‘fuzzy’ confusion matrices. This enables more information to be preserved through the remote sensing classification process, and reserves more interpretation for the map user than a typical ‘hard’ classification. Importantly the distributions of cover types are likely most representative of field conditions and thus more useful to land managers making holistic decisions about restoration or fuel management.

For the second chapter, I delved deeper into the potentials of new remote sensing and computing platforms to predict biological soil crust cover. The growing field of research on biological soil crust points to potentially significant implications for nutrient and water cycling, in addition to positive effects on native vascular vegetation. However, spatial data are lacking due to remote sensing limitations. Using time-series of multispectral imagery (from Chapter 1) and data fusion of radar and geophysical parameters, I developed a map of biocrust cover for the study area with high accuracy. This outcome allows us to examine important predictor variables (e.g. particular vegetation indices, soil type) and their relationship to plot-scale processes related to biological soil crust while also providing the spatial data needed for biological soil crust to be included in studies at the landscape scale.
TABLE OF CONTENTS

DEDICATION ........................................................................................................ iv

ACKNOWLEDGEMENTS ...................................................................................... v

ABSTRACT ........................................................................................................... vi

LIST OF TABLES .................................................................................................. xi

LIST OF FIGURES ............................................................................................... xii

LIST OF ABBREVIATIONS ................................................................................... xiv

CHAPTER ONE: TIME-SERIES, SENTINEL-2, VEGETATION MAPS IN SEMI-ARID ECOSYSTEMS ......................................................................................... 1

Abstract ............................................................................................................... 1

Background and Introduction ............................................................................... 2

Semi-Arid Ecosystems & Remote Sensing Thereof .............................................. 2

Phenology and Time-Series ............................................................................... 3

The Mixed Pixel Problem and Hard, Soft, and Fuzzy Classifications .............. 7

Sentinel-2 ........................................................................................................... 10

Research Question ............................................................................................ 15

Materials and Methods ..................................................................................... 16

Study Area ......................................................................................................... 16

Field and Training Data .................................................................................. 18

Assigning Hierarchical Levels of Dominant Cover ......................................... 22

Assigning Ecological Classes Using K-Means Clustering ............................... 23
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imagery and Predictor Variables</td>
<td>29</td>
</tr>
<tr>
<td>Classification &amp; Random Forest Model</td>
<td>34</td>
</tr>
<tr>
<td>Accuracy Assessment</td>
<td>34</td>
</tr>
<tr>
<td>Results</td>
<td>35</td>
</tr>
<tr>
<td>Discussion</td>
<td>38</td>
</tr>
<tr>
<td>Predictor Variable Importance</td>
<td>39</td>
</tr>
<tr>
<td>Selecting Categorical Levels of Response Variables Is Important</td>
<td>42</td>
</tr>
<tr>
<td>Soft and Fuzzy Classification Schemes</td>
<td>43</td>
</tr>
<tr>
<td>Issues, Improvements, and Future Steps</td>
<td>46</td>
</tr>
<tr>
<td>Implications</td>
<td>48</td>
</tr>
<tr>
<td>Conclusion</td>
<td>49</td>
</tr>
</tbody>
</table>

CHAPTER TWO: REMOTE SENSING OF BIOLOGICAL SOIL CRUST ..................51

Abstract .............................................................................51

Background and Introduction ..............................................52

Semi-Arid Ecosystems and Biocrusts ....................................52

What are Biological Soil Crusts? ........................................53

Remote Sensing of Biocrusts ..............................................55

Research Objectives and Importance ....................................60

Materials and Methods .....................................................61

Study Area .........................................................................61

Field Data ..........................................................................63

Imagery & Predictor Variables ..........................................64

Random Forest and Validation ...........................................68
Results .........................................................................................................................70

Biocrust Cover ........................................................................................................70

Regression Results ...............................................................................................71

Comparison with Ecotype Classification Models .................................................72

Discussion ...............................................................................................................73

Random Forest Model Assessment ........................................................................75

Predictor Variable Importance and Inferences .......................................................76

Validation between Ecotype/Community Classification Model and Biocrust Regression ...........................................................................................................78

Issues, Improvements, and Future Studies ..............................................................79

Conclusion ...............................................................................................................80

REFERENCES ...........................................................................................................82

APPENDIX A ............................................................................................................95

Comparing Biocrust Cover with Bare Cover ..........................................................96
LIST OF TABLES

Table 1: Sentinel-2 Satellite Constellation and Sensor Characteristics ..................12
Table 2: Selected Vegetation and Cover Types..................................................19
Table 3: Class Assignment Levels........................................................................24
Table 4: K-Means Species-Level (L1) Cluster Mean Centers and Number of Plots per Class. Each cluster is comprised of plots sharing common proportions of one or more cover types, but are not required to have commonalities of all cover types. .................................................................25
Table 5: K-Means PFT-Level (L4) Cluster Mean Centers and Number of Plots per Class. Each cluster is comprised of plots sharing common proportions of one or more cover types, but are not required to have commonalities of all cover types. .................................................................27
Table 6: Spectral Indices and Formulas Used in this Study ......................................31
Table 7: Cloud Control Dates (year, month, day)....................................................32
Table 8: Predictor Variables..................................................................................33
Table 9: Overall Accuracy and Kappa Coefficients for Predictor and Response Variables ..................................................................................................................36
Table 10: Deterministic Confusion Matrix: Community Classes (L1 k = 15).........37
Table 11: Cheatgrass Fuzzy (>25% overlap) Confusion Matrix: Community Classes (L1 k = 15)......................................................................................................37
Table 12: Sentinel-1 SAR Summary Statistics.......................................................66
Table 13: Re-Ordered Confusion Matrix with Fuzzy Confusion of 5% Difference in Mean Centers ....................................................................................................73
LIST OF FIGURES

Figure 1: Harmonic Analysis Coefficients. The curve (in red) shows an example of a sinusoid fitted to a time series of NDVI values (blue). The phase and amplitude components relate the phenological components of start of growing season and the range of photosynthetic activity as observed from NDVI, and the residuals relate to how well the growing phenology is represented by the sinusoid.........................................................6

Figure 2: Sentinel-2 Spectral Characteristics Compared to Landsat 7 and 8. Sentinel-2 has higher spatial resolution than Landsat in addition to several narrow bands in the Red-Edge portion of the electromagnetic spectrum (Bands 5-7, 8a). [NASA, 2015].................................................................13

Figure 3: Study Area ......................................................................................................................17

Figure 4: Distributions of Cover Types (species-level). Distributions of each species or cover type used for the L1 and K-means L1 classification schemes, for all field plots. Widths of boxplots represent the number of field plots with the cover type. This figure illustrates that there are very few species that can be found covering more than half of a plot. BRTE (cheatgrass) and BARE (bare, non-biocrust) both have many observations and a large range of percent cover. In comparison, ARTR (sagebrush) has a much smaller percent cover range.................................................................21

Figure 5: Schematic of Field Plot Design.....................................................................................22

Figure 6: K-Means Species-Level (L1) Cluster Community Distributions. Each cluster (1-15) represents a group of field plots sharing similar proportions of cover types. The vertical axis of each cluster represents the density of cover type within the cluster. The horizontal axis represents percent cover of each type. .........................................................................................26

Figure 7: K-Means PFT-Level (L4) Cluster Community Distributions. Each cluster (1-8) represents a group of field plots sharing similar proportions of cover types. The vertical axis of each cluster represents the density of cover type within the cluster. The horizontal axis represents percent cover of each type. .........................................................................................28

Figure 8: Example Illustration of Percent Overlap in BRTE Distributions. This figure illustrates that two different classes (2 & 3 in this example) may
have similar distributions of a particular cover type. Therefore, if BRTE is the species of interest to the map user, confusion between these classes is acceptable. This commutability forms the basis for ‘fuzzy’ confusion. The level of acceptable overlap is additionally determined by the map user.

Figure 9: NDVI Harmonic Coefficients. Image a) illustrates the three predictor coefficients (b-d, normalized to 1) in Saturation-Hue-Value, respectively. The background of a) shows the number of Sentinel-2 scenes used for the harmonics, where overlap in delivered MGRS tiles and the data-take (diagonal) artefacts are apparent. The RMSE band b) contains the majority of calculation artefacts.

Figure 10: Precipitation and B8A (Red Edge 865 nm) for a Sagebrush Plot (-116.3656650619256, 43.23953563635724), Vertical lines and arrows illustrate the correlated lag concept.

Figure 11: Gini-ranked Variable Importance

Figure 12: Imputed Biological Soil Crust Percent Cover in the Morley Nelson Snake River Birds of Prey, 2016

Figure 13: Predicted vs. Observed Biological Soil Crust Cover, R-Squared 0.74

Figure A.1: Imputed Bare Cover in the Morley Nelson Snake River Birds of Prey, 2016

Figure A.2: Predicted vs. Observed Bare Cover

Figure A.3: Bare Cover (red) vs. Biocrust Cover (blue) in the Morley Nelson Snake River Birds of Prey, 2016
LIST OF ABBREVIATIONS

AVHRR  Advanced Very High Resolution Radiometer
B1    Band 1 (also Bands 2-12)
BOA   Bottom of Atmosphere
BOP   Birds of Prey
BSC   Biological Soil Crust
BSS   Between Sum-of-Squares
CHIRPS Climate Hazards Group InfraRed Precipitation with Station data
EM    Endmembers
ESA   European Space Agency
GEE   Google Earth Engine
GPS   Global Positioning System
GSM   General Soil Map (of the United States)
HA    Harmonic Analysis
HC    Harmonic Coefficient
IDANG Idaho Army National Guard
IQR   Inner-Quartile Range
MAE   Mean Absolute Error
MC    Monthly Composite
MHAFB Mountain Home Air Force Base
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGRS</td>
<td>Military Grid Reference System</td>
</tr>
<tr>
<td>MODIS</td>
<td>Moderate Resolution Imaging Spectroradiometer</td>
</tr>
<tr>
<td>MPP</td>
<td>Mixed-Pixel Problem</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized Difference Vegetation Index</td>
</tr>
<tr>
<td>NIR</td>
<td>Near Infrared</td>
</tr>
<tr>
<td>OA</td>
<td>Overall Accuracy</td>
</tr>
<tr>
<td>OCTC</td>
<td>Orchard Combat Training Center</td>
</tr>
<tr>
<td>PFT</td>
<td>Plant Functional Type</td>
</tr>
<tr>
<td>RF</td>
<td>Random Forest</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root-Mean-Squared Error</td>
</tr>
<tr>
<td>RTK GPS</td>
<td>Real-Time-Kinematic Global Positioning System</td>
</tr>
<tr>
<td>S1</td>
<td>Sentinel-1</td>
</tr>
<tr>
<td>S2</td>
<td>Sentinel-2</td>
</tr>
<tr>
<td>SMAP</td>
<td>Soil Moisture Active Passive</td>
</tr>
<tr>
<td>SP</td>
<td>SamplePoint</td>
</tr>
<tr>
<td>SPOT</td>
<td>Satellite Pour l’Observation de la Terre</td>
</tr>
<tr>
<td>SSURGO</td>
<td>Soil Survey Geographic Database</td>
</tr>
<tr>
<td>STATSGO2</td>
<td>State Soil Geographic dataset</td>
</tr>
<tr>
<td>TOA</td>
<td>Top of Atmosphere</td>
</tr>
<tr>
<td>TLA</td>
<td>Three-Letter Acronym</td>
</tr>
<tr>
<td>TSS</td>
<td>Total Sum-of-Squares</td>
</tr>
<tr>
<td>USDA</td>
<td>United States Department of Agriculture</td>
</tr>
<tr>
<td>Code</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td>VH</td>
<td>Vertical-Horizontal (polarization)</td>
</tr>
<tr>
<td>VV</td>
<td>Vertical-Vertical (polarization)</td>
</tr>
<tr>
<td>WSS</td>
<td>Within Sum-of-Squares</td>
</tr>
</tbody>
</table>
CHAPTER ONE: TIME-SERIES, SENTINEL-2, VEGETATION MAPS IN SEMI-ARID ECOSYSTEMS

Abstract

As a modification of the common phrase “garbage in, gospel out” conveys, the success of a remote sensing vegetation classification relies not on complex or seemingly magical statistical and physical models using neural networks in abstract ‘feature space’, but rather on the quality of its input data. Semi-arid ecosystems are no exception; they contain very few types of land cover that can be found in patch sizes larger than the pixel with which they are represented - even at 10-meter scale pixels. Creating a vegetation map that assigns a single species or plant functional type to each pixel has limitations and may provide as little information as “this category is here in some proportion, likely”. But how can field observations of nature be translated into a form compatible with scientific investigation? Is the output data meaningful for land managers or ecosystem modelers? Using robust field data of vegetative cover combined with a time-series of high-resolution multispectral imaging, this paper examines the relationship between classification levels of observed land cover data and the amount of predictors needed to generate a vegetation classification that is representative and informative for land management and scientific research.
Background and Introduction

Semi-Arid Ecosystems & Remote Sensing Thereof

Dryland environments occupy nearly 40% of Earth’s terrestrial surface [Millennium Ecosystem Assessment, 2005] and are increasingly affected by changes in climate and land use (Li et al. 2015). While our understanding of their function and global impact is enough to know that they are important, our knowledge falls short of adequate [Millennium Ecosystem Assessment, 2005]. It would be difficult to understate the importance of dryland environments. Approximately 38% of the global population inhabits drylands, of which 250 million people in the developing world are directly impacted by degradation of such environments [Reynolds et al., 2007]. Semi-arid ecosystems have been observed to be a driving factor in the interannual variability of the global carbon cycle [Poulter et al., 2014]. Dynamics of such ecosystems are reflective of changes in climate (interannually and as part of larger trends), vegetation communities (as a result of invasive species), land use (from many past uses and management approaches), and fire (as a result of the previous three).

Understanding the interactions between climatic changes and vegetative response is therefore important at the global scale. At regional scales, understanding the dynamics of semi-arid vegetation community composition is important for fire risk, mitigation, and restoration [Shinneman et al., 2015], and for management efforts to balance multiple uses [Knick and Rotenberry, 1997].

There have been many approaches using remote sensing data to classify semi-arid vegetation in order to further understanding of these important and dynamic ecosystems. Common challenges among them are the relatively ‘weak’ signals of semi-arid vegetation
in relation to soil background, and high seasonal variability. These challenges are often met with high spectral and spatial resolution (as in airborne hyperspectral imaging), or high temporal resolution (as in MODIS or AVHRR time-series). Airborne hyperspectral imaging is expensive to collect and requires significant data processing, and consequently is not collected many times over the same area. While MODIS imagery is collected every one to two days in 36 spectral bands, the spatial resolution (250m-1.1km) is not commensurate with the scale of semi-arid vegetation (classically described as ‘patchwork’, varying through the landscape on the scale of 10s of meters or less).

It would be unfair to characterize these two approaches as the only options. The Landsat, SPOT, and RapidEye satellite constellations offer good compromises between temporal, spectral, and spatial resolution. They have all been used successfully to map semi-arid vegetation by leveraging their strengths, sometimes in concert with other platforms. New platforms such as the ESA’s Sentinel-2 satellites offer advances in optical remote sensing beneficial to vegetation mapping in dryland ecosystems, such as increased spatial and temporal resolution.

**Phenology and Time-Series**

Semi-arid vegetation closely follows seasonal and climatic patterns, and is responsive to weather events [Sonnenschein et al., 2011; Bansal and Sheley, 2016]. However, relatively sparse cover and long periods of senescence dampens this signal [Sonnenschein et al., 2011]. Therefore repeated observations over time are beneficial for the purposes of vegetation classification. Although many time-series methods exist, they can be broadly grouped into two approaches. The first - *phenometrics* - uses metrics such as the scale of minimums and maximums, growing season length the slope between them,
and when such occur in a growing season [Braget, 2017]. The second - *phenology* - examines temporal signals by matching observed with reference temporal signals, or comparing fitted mathematical curves (“reconstructed time-series”) to the signals [Zhou et al., 2016].

Phenometric time-series use metrics such as the beginning and end of a growing season and the date of maximal photosynthetic activity (observed from spectral indices such as NDVI) to preserve some temporal component of remote sensing data. In fact, retaining the maximum value of a spectral index can be sufficient for classification (such as in the Maximum-Value Composite procedure, Holben 1986). However, such approaches rely on selecting the appropriate time window to capture maximal variation in vegetative responses, while being sufficiently large to reduce atmospheric and anomalous values, and small enough to minimize the likelihood that neighboring pixels are collected at increasingly different times [Vancutsem et al., 2007; Zhou et al., 2016]. Additionally, the relative and absolute thresholds used to determine the beginning and end of the growing seasons can be difficult to establish (e.g. what point of the curve is the ‘start’), and the frequency component of vegetative responses are not captured [Huesca et al., 2015]. This is problematic for many types of semi-arid vegetation, which respond strongly to precipitation events. However, this methodology can be applied with platforms that have higher spatial resolution but generally less temporal resolution (such as Landsat).

Finely resolving patterns and changes in vegetative growth requires high temporal resolution and statistical methods to observe patterns in vegetation responses [Huesca et al., 2015]. Such approaches are diverse and varied. They generally rely on platforms such
as MODIS which trade high revisit frequency with spatial and spectral resolution [Llamas, 2010]. Other methods relate the temporal resolution of MODIS with the spatial and spectral qualities of Landsat [Gallagher, 2018]. Yet others ‘reconstruct’ time-series data by filling gaps between observations or by fitting mathematical functions to the observations [Zhou et al., 2016]. Such temporal data can therefore leverage methods that examine vegetation’s ‘signal’ more directly in the frequency domain.

One such approach is Harmonic Analysis (HA), first proposed by Menenti et al. (1993, referenced in Roerink et al. 2000). This approach reconstructs time series data by fitting sinusoids to the data, thus both reducing the effects of anomalous observations (such as cloud cover) in addition to indirect representations of phenology such as the amplitude and phase of the observed signal [Roerink et al., 2000; Menenti et al., 2010; Zhou et al., 2016]. Figure 1 illustrates this method. Such phenological parameters can be more easily related to climatic variables [Roerink et al., 2000]. Noted problems with this technique are due to the potential variability of the signals representing a land cover or vegetation class, due to the added climactic information (e.g. growth of the same class begins at slightly different times), and due to the mixed-pixel effect (e.g. pixels of the same class are assumed to represent the same on the ground; Parmesan 2007; Zhou et al. 2016). Such issues can be addressed with study areas sufficiently small to experience similar climatic conditions over a growing season, and by methods addressing the mixed-pixel problem.
Figure 1: Harmonic Analysis Coefficients. The curve (in red) shows an example of a sinusoid fitted to a time series of NDVI values (blue). The phase and amplitude components relate the phenological components of start of growing season and the range of photosynthetic activity as observed from NDVI, and the residuals relate to how well the growing phenology is represented by the sinusoid.
**The Mixed Pixel Problem and Hard, Soft, and Fuzzy Classifications**

The mixed-pixel problem (MPP) is well known in remote sensing classifications. For the purposes of this paper, a mixed-pixel is a pixel that represents some proportion of two or more ‘classes’. The notion of a ‘class’ that can be represented with a rectangular shape is the root of this problem; a pixel is assigned to a single class such as ‘tree’ when it is only ‘mostly tree’ or more specifically, the remotely-sensed signal is more consistent with ‘tree’ than it is other classes. This belies the issue that the observed signal from training locations called ‘tree’ are assumed to be a pure signal of ‘tree’, when in fact some may be 60% tree and others 85%, and some may have a background of dark soil as the remaining portion, and others with a verdant meadow as an understory. Should a pixel containing mostly ‘tree’ have enough of a different composition to fall outside of the distribution of the training ‘tree’ pixels, the classification may decide that it would be better called ‘shrub’, or perhaps ‘building’. The classifier may not have any clear choice, since it likely has no training pixels that match the pixel in question. Or, should the pixel contain 50% tree and 50% building, the resulting signal may happen to be more closely aligned with signals of ‘shrub’ pixels than those of either the ‘tree’ or ‘building’ classes. These are the classification issues that can arise from the mixed-pixel problem [Gebbinck, 1998].

Consequently, a class may be ‘over classified’ if its training pixels represent a large distribution of cover combinations (errors of commission). Radiative transfer, ray tracing models, and other physical models caution that the observed signal cannot be linearly-related to the proportion of cover [Disney, 2016]. As a result, one cover type may be over-represented in the observed signal. This cuts two ways. First, field observations
may describe the plot as 40% soil and 60% shrub, but the observed signal may be closer to the sum of a plot of 60% soil and 40% shrub - a complete change of class. Second, the signal of one cover type of interest may not be adequately observed due to the portion in which it naturally occurs, or at the date in which the image is collected. Even if a short, delicate bunchgrass is the dominant ‘class’, it may be difficult to observe its signal apart from the soil. This does not necessarily cause confusion if the signal is representative (i.e. ‘bare’ as a proxy for the bunchgrass), but should ‘bare’ be its own class of interest then significant under classification may occur (errors of omission).

There are several remote sensing approaches designed to address the mixed-pixel problem. The first is to simply acquire imagery of higher spatial resolution. A similar scale allows the pixel to better represent the subjects of interest [Fraser et al., 2014]. However, this can introduce ‘new’ spectral classes that were not previously observable at a coarse scale [Campbell and Wynne, 2011], and can result in a similar classification outcome - this time just at a finer spatial scale.

Fuzzy classification schemes accept that some confusion is less bad than others. This approach works well with continuous data that has been binned into categories (0-10%, 10-20%, and so on), where confusion between neighboring classes is easily explained. This method can also work for categorical classifications, but ascribing acceptable misclassifications is subjective. The producer may be comfortable with more confusion between ‘bare’ and ‘playa’ than a user interested in slickspot peppergrass. Accuracy metrics for these approaches typically report the deterministic confusion matrix in conjunction with a fuzzy matrix [Congalton and Green, 2009]. Additional fuzzy classification approaches ascribe partial membership to pixels as a way to describe
proportional composition. This has been done by assessing confidence levels per class, but is still dependent in some fashion on a single class per pixel (i.e. 100% confidence).

Another solution uses very high spectral resolution data to ‘unmix’ pixels. More comprehensive spectral signatures, especially at a high spatial resolution ($\approx 1m$) allow subtle distinctions between pixels to be observed. Known spectral endmembers (‘pure’ pixels, either from known locations in the images, or a reference signal) can then inform unmixing of the pixel into its likely constituents. This process has been used with remarkable success, even in shrub-steppe environments [Poley, 2017].

However, until satellite imaging spectroscopy data such as from the future NASA Surface Biology and Geology [National Academies of Sciences, Engineering, 2018] are available, such imagery is expensive to acquire and therefore not typically collected at multiple times over a growing season. Consequently, the ability to classify ground cover types is dependent on the date that the image was collected. Dormant deciduous trees species may be under-represented or confused with dead stands of trees, for example. For landscapes characterized by long periods of senescence punctuated by sharp changes in phenology - such as semi-arid ecosystems - the imagery collection date becomes crucial. It may not be possible to pick a singular date that can minimize confusion between all classes since phenologies of semi-arid vegetation can be disparate [Gallagher, 2018].

Ustin and Gamon [Ustin and Gamon, 2010] proposed that “Optical Types” may be a solution to this by assigning classes in a similar way as with Plant Functional Types, but also including information on optical properties and physical interactions between photons and canopy structure. This, they posit, is more effective in linking remote sensing observations with plant traits and environmental conditions. This framework has
the advantage of accounting for physical interactions with photons at the canopy scale, and seeks to relate such insight into more accurate classifications. However, it does little to address the MPP directly.

Yet other approaches leverage the changes in a pixel over time as in phenology (see above section). This additional order of dimensionality enables greater distinction between pixels, and potentially the ability to group pixels that are ‘more similar’ over time. Huesca et al. [2015] apply the “Optical Types” framework in this way. This approach effectively links Optical Types with temporal dynamics. Their concluding remark: “Thus, exploration of temporal dynamics presents a promising opportunity to further explore vegetation composition within mixed pixels.”

The majority of approaches outlined so far have sought to decompose the signal into its constituent parts (fuzzy membership, spectral unmixing), collect higher resolution imagery (spatial, spectral, or temporal) with the hopes of discerning better between classes, or to justify confusion errors after classification. Few, if any, have attempted to perform classification on pixels assumed to be mixed (soft) when they are input into the classification model.

**Sentinel-2**

The European Space Agency's Sentinel-2 constellation (opposing twin satellites, A and B) is directed at monitoring the land surface and coastal waters with high-resolution multi-spectral images in three- to five-day intervals (Table 1: Sentinel-2 Characteristics). Sentinel-2A was launched on 23 June 2015 and Sentinel-2B on 7 March 2017. The Sentinel-2 (S2) mission is designed to provide continuity to the SPOT and LANDSAT missions, with some improvements. Notable among Sentinel-2s’ 13 spectral
bands are three narrow-bandwidth spectral bands in the “Red Edge” between visible red band (≈660 nm) and the near infrared band (≈840 nm). These Red Edge (RE) bands are important in monitoring vegetation health because this range is where chlorophyll is highly reflective [Adam et al., 2014]. Additionally, RE spectral information have been observed to be important for vegetation in semi-arid environments [Schumacher et al., 2016].
<table>
<thead>
<tr>
<th>Table 1: Sentinel-2 Satellite Constellation and Sensor Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Satellites</strong></td>
</tr>
<tr>
<td><strong>Orbit Altitude</strong></td>
</tr>
<tr>
<td><strong>Swath Width &amp; Data-take Length</strong></td>
</tr>
<tr>
<td><strong>Revisit Time</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Spectral Instrument</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Ground Sampling Distance</strong></td>
</tr>
<tr>
<td><strong>Products Levels</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Figure 2: Sentinel-2 Spectral Characteristics Compared to Landsat 7 and 8. Sentinel-2 has higher spatial resolution than Landsat in addition to several narrow bands in the Red-Edge portion of the electromagnetic spectrum (Bands 5-7, 8a). [NASA, 2015]

The increased spatial resolution (10-20 m, compared to 30-60 m) is markedly closer to scales important for land management and planning decisions [Cihlar, 2000; Franklin and Wulder, 2002], and closer to the scale of human influences [Wulder et al., 2012]. Higher spatial resolution means more representative pixels (i.e. less mixed pixels), especially in areas where the vegetative cover is patchy or highly variable. Monitoring the change in position of an ecotone, for example, is difficult if the footprint of the pixel straddles the entire transition. Other less immediately-recognized benefits of higher spatial resolution pixels are better cloud screening (as clouds are less-often included within a pixel), and better image coregistration and georeferencing.

Although the newer S2 offers advantages compared with previous platforms, it does have several drawbacks including the limited historical imagery. S2 satellites are on a different orbital path than Landsat and MODIS, so cannot readily be used in the same
way that information from these platforms are often used together. While S2’s blue, green, red, and near-infrared bands were designed to capture similar wavelengths as Landsat 7 and 8 (Figure 2), the small differences in spectral responses requires bandpass adjustment in order to be used together [Claverie and Masek, 2016]. The ESA delivers S2 scenes at Level-1C, which includes top-of-atmosphere (TOA) reflectance in cartographic geometries which are registered to be within 3, 6, or 18 m per pixel for the 10, 20, and 60 m bands, respectively (Table 1: Sentinel-2 Characteristics). While ESA’s freely-available SNAP toolbox can perform atmospheric corrections to bottom-of-atmosphere (BOA) reflectance (Level-2A), it still requires that the user download and process each scene on a personal computer [Baillarin et al., 2012]. Using TOA values for temporal comparisons can be problematic as a result of changing atmospheric conditions. Unfortunately, several of the important spectral bands are sensitive to atmospheric interferences; for example TOA NDVI is lower than BOA NDVI [Beck et al., 2006].

Imagery from 2016 is additionally challenging. Sentinel-2B was not launched until March 2017, so the full temporal resolution was not realized in 2016. While the temporal resolution of a single S2 satellite (A or B) would be ten days, it can vary from two to twelve as the orbit of Sentinel-2A was modified several times. Spatial misalignment of delivered scenes has also been an issue, but ESA have generally reprocessed imagery as issues are found and algorithms enhanced. Google Earth Engine’s collection is regularly updated to capture these changes. The tiling of delivered S2-L1C imagery is UTM-projected, and aligned with the Military Grid Reference System (MGRS) and delivered in 10,000 km² tiles. There are several issues with the delivery system of tiles, where not all tiles delivered are a full ‘square’; images falling partially
inside one MGRS tile are delivered with the label of the tile but are only a sliver of the image. The processing baseline for imagery has also been updated several times, as well as some of the metadata format. These issues complicate spatial and temporal compositing of Sentinel-2 imagery. Figure 9 illustrates this issue.

**Research Question**

Remote sensing of vegetation in semi-arid ecosystems and drylands is inherently challenging. Phenology in these ecosystems is highly dynamic both in response to weather and climatic changes, and through the growing season. Fractional vegetative cover is low and community composition is heterogeneous at scales from tens to hundreds of meters. These difficulties lead to two questions: 1) are a cohort of spectral predictors at regular time intervals better than a generalized phenology curve, or are a few ‘snapshots’ equivalent? and 2) what is a meaningful way to characterize vegetation on a scale commensurate with remote sensing imagery?

I investigate the first question by comparing classification accuracies using imagery from discrete time-series (monthly) composites, sinusoidal approximations of time-series data (“harmonic” models), and both in concert. I contrast these results with a control set of predictor variables consisting of all dates of cloud-free imagery covering the study area. Keeping the spatial resolution the same (10 meter pixels) and using the same spectral predictors with each approach highlights how the differences in temporal representation correspond to classification accuracies.

I find answers to the second question through examining methods for assigning ‘classes’ to field data, and by evaluating the classification results. First, I investigate per-pixel classification methods of designating the ‘class’ as that of the largest cover
At several hierarchical levels from species-level to Plant Functional Types (PFTs). I also explore the effects of barriers to class membership (majority cover, and without strong co-dominant cover) on classification accuracy. Next, I use a clustering algorithm to determine ecologically meaningful (‘soft’) classes. Finally, I propose a method to evaluate classification accuracies of the soft classes that reserves interpretation for the vegetation map user. Taken together, insights from these questions inform remote sensing classifications of semi-arid vegetation, and preserves interpretation for land management and scientific inquiry.

Materials and Methods

Study Area

The study area for this paper includes the Morley Nelson Snake River Birds of Prey National Conservation Area (BOP, approximately 2430 km², managed by the Bureau of Land Management) including the enclosed Orchard Combat Training Center (OCTC, approximately 580 km², managed by the Idaho Army National Guard), and the Mountain Home Air Force Base and Small Arms Range (MHAFB) The location of the study area and constituent public and government management agencies are shown in Figure 3: Study Area. The area for this study was determined by buffering the BOP perimeter by 4 km in order to include additional field data collected as part of the same field campaign but outside of the BOP boundary, for a total of approximately 4,140 km². The ecology of this study area is broadly characterized as a semi-arid ecotype composed of communities of Wyoming big sagebrush (Artemisia tridentata ssp. wyomingensis),
salt-desert shrubs (primarily members of the Chenopodioidae subfamily), and bunchgrasses (Poa secunda, Pseudoroegneria spicata, and Agropyron cristatum).

Figure 3: Study Area

There has been a large variety of land uses within the study area over the last century, principally livestock grazing, recreation, and military training [Reynolds et al., 2007]. Natural gradients in the heterogeneity of the environment are exaggerated with free ranging grazing [Adler et al., 2001]. Extensive ecosystem degradation is observable in many parts of the study area, where invasive annual grasses (such as Bromus tectorum, or ‘cheatgrass’) and secondary weeds such as Russian thistle (Salsola kali) and annual mustards (Descurainia spp., Sisymbirum ssp.) are now the dominant land cover [US Department of Interior (USDI), 2008]. As a result of the invasive species and a long
history of changing land uses and practices, the BOP is a ‘patchwork’ of many different vegetative communities. Balancing the current uses of the study area requires managing agencies and stakeholders to be informed of yearly changes in vegetative cover, impacts of rangeland fires, and how restoration and remediation efforts are progressing. To this end, increasing the accuracy of vegetation cover data and frequency of its creation is propitious.

**Field and Training Data**

Field data plots used in this study (n = 378) were collected during March-August 2016. Field data plots were selected in the field as homogenous communities of each vegetation and land cover type (Table 2: Selected Vegetation and Cover Types), and representative of different community combinations and variations among spatial gradients. A field survey was conducted for each plot, and five nadir-pointing images were taken approximately two meters above the ground surface using a 16-megapixel all-weather camera (Nikon COOLPIX AW120). A RTK GPS recorded the imager’s location simultaneously (Figure 5: Field Plot Design). Vegetation and groundcover were quantified using SamplePoint software (v1.59, Booth et al. 2006) using 100 points per image for a total of 500 points characterizing each field plot. Figure 4 illustrates the cover distributions for each of the main cover types classified.
<table>
<thead>
<tr>
<th>Scientific Name(s)</th>
<th>Species Abbreviation (Finest-Level Class Assigned)</th>
<th>Common Name</th>
<th>Growth Type &amp; Plant Functional Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Agropyron cristatum</em></td>
<td>AGCR</td>
<td>crested wheatgrass</td>
<td>perennial, bunchgrass</td>
</tr>
<tr>
<td><em>Artemisia tridentata</em></td>
<td>ARTR</td>
<td>sagebrush</td>
<td>perennial, shrub</td>
</tr>
<tr>
<td><em>Atriplex confertifolia</em></td>
<td>ATCO</td>
<td>shadscale saltbush</td>
<td>perennial, shrub</td>
</tr>
<tr>
<td><em>Bassia prostrata</em></td>
<td>BAPR</td>
<td>forage kochia</td>
<td>perennial, subshrub</td>
</tr>
<tr>
<td><em>Bromus tectorum</em></td>
<td>BRTE</td>
<td>cheatgrass</td>
<td>annual, grass</td>
</tr>
<tr>
<td><em>Ceratocephala testiculata</em></td>
<td>(EXAN)</td>
<td>bur buttercup</td>
<td>annual, forb</td>
</tr>
<tr>
<td><em>Lepidium perfoliatum</em></td>
<td>(EXAN)</td>
<td>claspers pepperweed</td>
<td>annual, forb</td>
</tr>
<tr>
<td><em>Bassia scoparia</em></td>
<td>(EXAN)</td>
<td>weed kochia</td>
<td>annual, forb</td>
</tr>
<tr>
<td><em>Krascheninnikovia lanata</em></td>
<td>KRLA</td>
<td>winterfat</td>
<td>perennial, shrub</td>
</tr>
<tr>
<td>Species</td>
<td>Code</td>
<td>Common Name</td>
<td>Life Form</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------</td>
<td>------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td><em>Descurainia spp.</em></td>
<td>(MSTD)</td>
<td>mustards</td>
<td>annual, forb</td>
</tr>
<tr>
<td><em>Sisymbirum ssp.</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Poa secunda</em></td>
<td>POSE</td>
<td>Sandberg’s bluegrass</td>
<td>annual, bunchgrass</td>
</tr>
<tr>
<td><em>Pseudoroegneria spicata</em></td>
<td>PSSP</td>
<td>bluebunch wheatgrass</td>
<td>annual, bunchgrass</td>
</tr>
<tr>
<td><em>Chrysothamnus nauseosus</em></td>
<td>(RABB)</td>
<td>gray rabbitbrush</td>
<td>perennial, shrub</td>
</tr>
<tr>
<td><em>Chrysothamnus viscidiflorus</em></td>
<td>(RABB)</td>
<td>green rabbitbrush</td>
<td>perennial, shrub</td>
</tr>
<tr>
<td>-</td>
<td>(BARE)</td>
<td>bare ground</td>
<td>n/a, bare</td>
</tr>
<tr>
<td>-</td>
<td>(BSC)</td>
<td>biological soil crusts</td>
<td>bacteria, moss, and lichen; groundcover</td>
</tr>
<tr>
<td>-</td>
<td>(NPSV)</td>
<td>non-photosynthetic vegetation</td>
<td>dead plant matter</td>
</tr>
</tbody>
</table>
Figure 4: Distributions of Cover Types (species-level). Distributions of each species or cover type used for the L1 and K-means L1 classification schemes, for all field plots. Widths of boxplots represent the number of field plots with the cover type. This figure illustrates that there are very few species that can be found covering more than half of a plot. BRTE (cheatgrass) and BARE (bare, non-biocrust) both have many observations and a large range of percent cover. In comparison, ARTR (sagebrush) has a much smaller percent cover range.
Assigning Hierarchical Levels of Dominant Cover

Several orders of classes were assigned to each plot determined by dominant cover. The principal level of classification ("Dominant Cover L1") designates classes based on the largest percent cover at the plot, regardless of it representing more than 50% (majority cover). For example, a plot containing 16% ARTR, 43% BRTE, and 18% NPSV would be assigned to the “BRTE” class. At the L1 level, classes with few or no members (e.g. ROCK, FORB) were removed or combined minimally (e.g. RABB class represents grey and green rabbitbrush, because only one plot contained a majority of grey rabbitbrush). Increasing levels of aggregation were based on two philosophies. The first sought to balance the number of members of each class by combining similar but smaller classes (e.g. KRLA and ATCO combined into salt-desert shrubs, Levels 1 and 3) or
splitting classes with many members into smaller classes (i.e. BRTE, Level 2), up to the level of Plant Functional Types (PFTs, Level 4). The second philosophy sought to increase the distinction of each PFT class by imposing criteria for membership (Levels 4-6). Table 3 lists the differences between each dominant cover level.

**Assigning Ecological Classes Using K-Means Clustering**

Ecologically-meaningful classes were determined using k-means clustering at two of the aggregation levels from the dominant cover classes (L1 species-level, and L4 PFT). K-means is a vector quantization method that is often used for clustering data in multi-dimensional space. This method was chosen to separate the field data into ecological classes without bias. The number of k-means clusters was preliminarily constrained using diminishing reductions in within-cluster sum of squares, and further refined subjectively to the fewest \( \kappa \) clusters that retained at least one cluster per cover type (e.g. a cluster representing KRLA communities). Multiple iterations and random starting sets of data ensure convergence of cluster centers. Each clustering maintained a ratio of Between Sum-of-Squares to Total Sum-of-Squares (BSS/TSS, equivalent to \( R^2 \)) of approximately 0.8. Figures 6 and 7 represent the distribution of vegetation cover types contained in each community class determined from the k-means classification scheme. Tables 4 and 5 list cluster centers (means) of the cover type for each class and the number of field plots contained in each class (cluster).
<table>
<thead>
<tr>
<th>Class Assignment Level</th>
<th>Description</th>
<th>Number of Classes</th>
<th>Number of Training Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant Cover L1</td>
<td>Original (Species) Levels of Interest (Table 2)</td>
<td>18</td>
<td>378</td>
</tr>
<tr>
<td>L2</td>
<td>Same as L1, except two classes of BRTE (&lt;50%, &gt;= 50%)</td>
<td>19</td>
<td>378</td>
</tr>
<tr>
<td>L3</td>
<td>Same as L2, except Perennial Grasses = AGCR + PSSP Salt-Desert Scrub = ATCO + KRLA</td>
<td>14</td>
<td>378</td>
</tr>
<tr>
<td>L4</td>
<td>Plant and Cover Functional Types (PFT): Shrub, Perennial Bunchgrasses, Annuals, Bare, Non-Photosynthetic Vegetation, and Biocrust</td>
<td>6</td>
<td>378</td>
</tr>
<tr>
<td>L5</td>
<td>Same as L4, excluding plots where $PFT_{\text{secondary}} &lt; PFT_{\text{primary}} - (\mu + \sigma)$</td>
<td>6</td>
<td>294</td>
</tr>
<tr>
<td>L6</td>
<td>Same as L4, excluding $PFT_{\text{primary}} &lt; 50%$</td>
<td>3</td>
<td>150</td>
</tr>
<tr>
<td>K-Means Community L1</td>
<td>Clustered L1-level cover, $k = 15$</td>
<td>15</td>
<td>378</td>
</tr>
<tr>
<td>L4</td>
<td>Clustered L4-level cover, $k = 8$</td>
<td>8</td>
<td>378</td>
</tr>
</tbody>
</table>
Table 4: K-Means Species-Level (L1) Cluster Mean Centers and Number of Plots per Class. Each cluster is comprised of plots sharing common proportions of one or more cover types, but are not required to have commonalities of all cover types.

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>ARTR</th>
<th>RABB</th>
<th>KRLA</th>
<th>ATCO</th>
<th>BAPR</th>
<th>POSE</th>
<th>BRTE</th>
<th>AGCR</th>
<th>PSSP</th>
<th>MSTD</th>
<th>NPSV</th>
<th>EXAN</th>
<th>BSC</th>
<th>BARE</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.2</td>
<td>0.5</td>
<td>0.0</td>
<td>1.9</td>
<td>3.6</td>
<td>0.8</td>
<td>0.4</td>
<td>0.0</td>
<td>0.6</td>
<td>6.2</td>
<td>25.7</td>
<td>11.4</td>
<td>42.3</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>2.9</td>
<td>4.7</td>
<td>2.1</td>
<td>1.5</td>
<td>0.3</td>
<td>1.7</td>
<td>15.2</td>
<td>1.5</td>
<td>1.5</td>
<td>1.3</td>
<td>8.5</td>
<td>0.3</td>
<td>6.3</td>
<td>35.5</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
<td>25.5</td>
<td>38.9</td>
<td>0.0</td>
<td>0.0</td>
<td>8.1</td>
<td>0.3</td>
<td>11.4</td>
<td>10.0</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>30.3</td>
<td>0.1</td>
<td>0.4</td>
<td>0.5</td>
<td>0.0</td>
<td>6.9</td>
<td>2.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.4</td>
<td>14.8</td>
<td>5.7</td>
<td>12.6</td>
<td>21.2</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>2.0</td>
<td>1.3</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.9</td>
<td>81.5</td>
<td>1.0</td>
<td>0.0</td>
<td>1.8</td>
<td>2.3</td>
<td>0.0</td>
<td>1.6</td>
<td>4.6</td>
<td>58</td>
</tr>
<tr>
<td>6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>33.4</td>
<td>2.4</td>
<td>1.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>7.3</td>
<td>2.8</td>
<td>19.4</td>
<td>29.3</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>8.5</td>
<td>1.5</td>
<td>1.8</td>
<td>0.0</td>
<td>49.7</td>
<td>6.7</td>
<td>0.6</td>
<td>6.2</td>
<td>23.0</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>1.8</td>
<td>0.2</td>
<td>0.5</td>
<td>1.0</td>
<td>0.8</td>
<td>44.4</td>
<td>0.6</td>
<td>0.1</td>
<td>0.0</td>
<td>1.4</td>
<td>4.1</td>
<td>1.3</td>
<td>8.2</td>
<td>29.1</td>
<td>37</td>
</tr>
<tr>
<td>9</td>
<td>1.7</td>
<td>1.1</td>
<td>9.1</td>
<td>1.5</td>
<td>0.1</td>
<td>12.7</td>
<td>4.3</td>
<td>0.1</td>
<td>5.0</td>
<td>2.2</td>
<td>8.1</td>
<td>0.9</td>
<td>25.1</td>
<td>8.6</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>1.2</td>
<td>0.0</td>
<td>1.8</td>
<td>25.5</td>
<td>0.0</td>
<td>0.4</td>
<td>0.6</td>
<td>0.0</td>
<td>0.0</td>
<td>1.3</td>
<td>5.6</td>
<td>7.1</td>
<td>19.7</td>
<td>33.7</td>
<td>13</td>
</tr>
<tr>
<td>11</td>
<td>3.0</td>
<td>4.3</td>
<td>0.8</td>
<td>2.3</td>
<td>0.0</td>
<td>3.2</td>
<td>47.8</td>
<td>0.8</td>
<td>1.5</td>
<td>0.8</td>
<td>6.4</td>
<td>0.3</td>
<td>7.1</td>
<td>14.0</td>
<td>54</td>
</tr>
<tr>
<td>12</td>
<td>0.4</td>
<td>0.5</td>
<td>1.6</td>
<td>1.1</td>
<td>2.3</td>
<td>4.0</td>
<td>1.1</td>
<td>0.6</td>
<td>0.0</td>
<td>1.8</td>
<td>3.9</td>
<td>4.4</td>
<td>4.3</td>
<td>56.1</td>
<td>32</td>
</tr>
<tr>
<td>13</td>
<td>0.5</td>
<td>3.2</td>
<td>0.7</td>
<td>2.7</td>
<td>0.0</td>
<td>11.9</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>3.5</td>
<td>43.5</td>
<td>5.8</td>
<td>22.5</td>
<td>11.5</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.0</td>
<td>0.1</td>
<td>20.0</td>
<td>0.2</td>
<td>0.0</td>
<td>1.5</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>1.1</td>
<td>4.6</td>
<td>5.6</td>
<td>18.2</td>
<td>44.3</td>
<td>29</td>
</tr>
<tr>
<td>15</td>
<td>0.4</td>
<td>27.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>17.9</td>
<td>0.4</td>
<td>0.0</td>
<td>0.2</td>
<td>0.6</td>
<td>10.5</td>
<td>2.1</td>
<td>17.9</td>
<td>19.0</td>
<td>20</td>
</tr>
</tbody>
</table>
Figure 6: K-Means Species-Level (L1) Cluster Community Distributions. Each cluster (1-15) represents a group of field plots sharing similar proportions of cover types. The vertical axis of each cluster represents the density of cover type within the cluster. The horizontal axis represents percent cover of each type.
Table 5: K-Means PFT-Level (L4) Cluster Mean Centers and Number of Plots per Class. Each cluster is comprised of plots sharing common proportions of one or more cover types, but are not required to have commonalities of all cover types.

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Shrubs</th>
<th>Perennial Bunchgrasses</th>
<th>Annual Grasses and Forbs</th>
<th>Bare</th>
<th>Non-Photosynthetic Vegetation (NPV)</th>
<th>Biological Soil Crust (BSC)</th>
<th>Cluster Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.8</td>
<td>3.9</td>
<td>40.4</td>
<td>37.6</td>
<td>6.1</td>
<td>6.9</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>2.6</td>
<td>36.5</td>
<td>9.7</td>
<td>41.2</td>
<td>4.2</td>
<td>5.4</td>
<td>33</td>
</tr>
<tr>
<td>3</td>
<td>5.4</td>
<td>50.6</td>
<td>8.0</td>
<td>13.8</td>
<td>6.0</td>
<td>15.5</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>5.2</td>
<td>3.6</td>
<td>12.0</td>
<td>70.5</td>
<td>3.8</td>
<td>4.4</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>31.2</td>
<td>11.6</td>
<td>5.0</td>
<td>17.8</td>
<td>13.6</td>
<td>19.3</td>
<td>59</td>
</tr>
<tr>
<td>6</td>
<td>22.2</td>
<td>2.9</td>
<td>10.2</td>
<td>42.0</td>
<td>4.9</td>
<td>16.3</td>
<td>75</td>
</tr>
<tr>
<td>7</td>
<td>12.5</td>
<td>11.4</td>
<td>50.3</td>
<td>10.0</td>
<td>6.3</td>
<td>8.1</td>
<td>55</td>
</tr>
<tr>
<td>8</td>
<td>2.9</td>
<td>2.2</td>
<td>84.8</td>
<td>5.4</td>
<td>2.2</td>
<td>1.7</td>
<td>57</td>
</tr>
</tbody>
</table>
Figure 7: K-Means PFT-Level (L4) Cluster Community Distributions. Each cluster (1-8) represents a group of field plots sharing similar proportions of cover types. The vertical axis of each cluster represents the density of cover type within the cluster. The horizontal axis represents percent cover of each type.
Imagery and Predictor Variables

Imagery Selection & Preprocessing

Level-1C Sentinel-2 imagery covering the study area from 01 January 2016 to 28 December 2016 were selected to span the growing season corresponding to the field data collection. Images were preliminarily filtered to remove small image footprints (ESA processing artefacts as discussed in Background: Sentinel-2), images that failed radiometric or general quality control, and scenes with greater than 50% cloud cover (from metadata, for speed of processing). The resulting image collection was comprised of 117 images.

Cloud & Shadow Masking

Cloud and shadow masking algorithms were applied to each image using tiered thresholds. This method is a modification of a community-developed algorithm implemented in GEE. The cloud masking algorithm scores each pixel with threshold values of the atmospherically-sensitive bands (B1, B9, and B10) and masks the pixel if the sum of the band thresholds exceeds a global threshold. Thresholds using the blue band (B2) and RGB sums (B2+B3+B4) were not used due to frequent confusion of playa with clouds. Cloud shadows are masked by identifying possible shadow paths based on the solar zenith and azimuth (from metadata delivered with each S2 image), combined with a tiered threshold dark pixel algorithm similar to the cloud masking algorithm (B2+B3+B4, and B8+B11+B12) and a global threshold. Intersections of dark areas and possible cloud paths are then masked as shadows. Threshold values for each masking step were established qualitatively. Masks were applied at the 10 m pixel scale. Imagery from January was not cloud masked due to differences in metadata resulting in errors in the
cloud masking algorithm. The January mosaic was visually inspected to ensure that no significant clouds or shadows were visible in the final mosaicked scene. The February mosaic was discarded due to the excessive proportion of masked pixels, which resulted in all field observations having null values for this time.

**Spectral Indices & Predictor Variables**

Spectral indices were calculated for each image following cloud and shadow masking. Table 6 lists the spectral indices and their formulas, informed by previous scientific work for their predictive ability and that take advantage of S2-specific bands [Frampton et al., 2013], or for their novelty in using the several red-edge bands of S2. Sentinel-2 bands were also used directly as predictors in addition to spectral indices. Following this, the image collection was forked in order to generate the different sets of predictor variables. Monthly (30-day) composites were created based on pixel quality (cloud threshold values) of all of the images within the temporal window. The temporal steps were initialized beginning on the 4th day of the year (corresponding to the first S2 image) in order to ensure approximately three collection opportunities per 30-day window.
Table 6: Spectral Indices and Formulas Used in this Study

<table>
<thead>
<tr>
<th>Index or Ratio</th>
<th>Abbreviation</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthocyanin Reflectance Index</td>
<td>ARI</td>
<td>( \frac{1}{B_3} - \frac{1}{B_5} )</td>
</tr>
<tr>
<td>Canopy Chlorophyll Content Index</td>
<td>CCCI</td>
<td>( \frac{B_8A - B_7}{B_8A + B_7} ) ( \frac{B_8A - B_4}{B_8A + B_4} )</td>
</tr>
<tr>
<td>Enhanced Vegetation Index</td>
<td>EVI</td>
<td>( 2.5 \times \frac{B_8 - B_4}{B_8 + (6 \times B_4) - (7.5 \times B_2) + 1} )</td>
</tr>
<tr>
<td>Inverted Red-Edge Chlorophyll Index</td>
<td>IRECI</td>
<td>( \frac{B_7 - B_4}{B_5/B_6} )</td>
</tr>
<tr>
<td>Normalized Difference Near Infrared Red-Edge</td>
<td>NDMI</td>
<td>( \frac{B_8A - B_{11}}{B_8A + B_{11}} )</td>
</tr>
<tr>
<td>Normalized Difference Vegetation Index</td>
<td>NDVI</td>
<td>( \frac{B_8 - B_4}{B_8 + B_4} )</td>
</tr>
<tr>
<td>Sentinel-2 Red-Edge Position</td>
<td>S2REP</td>
<td>( 0.705 + 0.035 \times \frac{(B_7 + B_4)/2 - B_5}{B_6 - B_5} )</td>
</tr>
<tr>
<td>Soil Composition Index</td>
<td>SCI</td>
<td>( \frac{B_{11} - B_8}{B_{11} + B_8} )</td>
</tr>
</tbody>
</table>

Harmonic coefficients were calculated from the masked imagery (i.e. retaining original temporal information) for each of the bands and spectral indices used in the monthly composites. A period of \( 2\pi \) was used, as preliminary results demonstrated that it best modeled typical phenology of semi-arid ecosystems. The phase, amplitude, and
residuals of the harmonic calculation were used as predictors. Figure 1 illustrates the metrics derived from each harmonic model, calculated for each pixel.

As a control for issues caused by cloud and shadow masking, temporal compositing, and harmonic coefficient calculations, S2 collections with little to no cloud cover were also evaluated for predictive ability using the same bands and spectral indices as the 30-day composites. Eight dates had imagery covering the entire study area where all but one tile had less than six percent cloud cover. Table 7 lists dates and cloud cover for these images. Table 8 summarizes predictor variables.

**Table 7: Cloud Control Dates (year, month, day)**

<table>
<thead>
<tr>
<th>Date Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-03-30</td>
</tr>
<tr>
<td>2016-04-19</td>
</tr>
<tr>
<td>2016-05-09</td>
</tr>
<tr>
<td>2016-06-28</td>
</tr>
<tr>
<td>2016-07-18</td>
</tr>
<tr>
<td>2016-07-28</td>
</tr>
<tr>
<td>2016-09-16</td>
</tr>
<tr>
<td>2016-11-05</td>
</tr>
<tr>
<td>Predictor Data Sets</td>
</tr>
<tr>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Cloud Control: Bands and Spectral Indices</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Monthly (30-Day) Composites: Bands and Spectral</td>
</tr>
<tr>
<td>Indices</td>
</tr>
<tr>
<td>Harmonic Coefficients</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Monthly Composites + Harmonic Coefficients</td>
</tr>
</tbody>
</table>

*see Cloud & Shadow Masking section above for explanation*
Classification & Random Forest Model

The random forest (RF) classifier was chosen to evaluate the different predictor and response variable combinations due to the high dimensionality and multicollinearity of the predictor variables, and its insensitivity to overfitting [Breiman, 2001; Belgiu and Dragut, 2016]. The RF classifier can handle unbalanced data [Pal, 2005], which is important for the field data in this study, as some classes of interest are relatively under-represented in comparison with others (e.g. number of shadscale plots verses cheatgrass plots). It is also able to handle missing values when imputing the classification, which is advantageous for portions of the predictor data that contain masked areas after mosaicking. As implemented in GEE, training data must be free of missing values to grow a RF model.

Field data points were buffered using a 10 m radius, and the mean value of the intersected pixels in each band were extracted as the predictor variables for the RF model. The RF model was implemented in Google Earth Engine using 500 trees and out-of-bag internal sampling, with a random seed value of zero. Data was divided into training data (67%) and validation data (33%) using a random number. The validation data set was used to calculate a confusion matrix at each level of predictor imagery.

Accuracy Assessment

Accuracy assessment was performed for each of the classification schemes (L1-L6) against each predictor set by calculating the overall accuracy of a deterministic confusion matrix generated through the internal OOB validation of RF. Kappa coefficients were also calculated for each combination to illustrate accuracy accounting for chance classification.
Accuracy of the k-means vegetation classification schemes were assessed as above, with the addition of a matrix representing “fuzzy” confusion for select classes. Determining acceptable fuzzy confusion between ecotype classes was calculated for each of the cover types by examining the percent of overlap of the density distributions per class. If the mean value of each cover type was greater than 25%, the intersection of each distribution was summed. This overlap threshold was then applied, and the combination of clusters was flagged as an acceptable fuzzy confusion for that class if the overlapping percentage exceeded the threshold. Additional overlap thresholds were examined (75% and 50%) but are not presented here.

**Results**

Table 9 shows the results of the varying class aggregations and predictor variable selection combinations. The harmonic coefficients (HC) predictors had generally lower overall accuracy than the Monthly composite (MC) time-series predictors for all class. The combination of HC and MC imagery showed little-to-no improvement over the MC alone. The control predictors consistently yielded higher accuracies compared to HC predictors, and were comparable to the results from the MC and HC+MC predictors. An example confusion matrix for the L1 community class and MC is presented in Table 10.
Table 9: Overall Accuracy and Kappa Coefficients for Predictor and Response Variables

<table>
<thead>
<tr>
<th>Classification Schemes</th>
<th>PREDICTOR VARIABLES</th>
<th>Control</th>
<th>Response Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Harmonic Coefficients</td>
<td>Monthly Composites</td>
<td>Harmonic + Monthly</td>
</tr>
<tr>
<td></td>
<td>Overall Accuracy</td>
<td>Kappa</td>
<td>OA</td>
</tr>
<tr>
<td>(Species) L1</td>
<td>0.58</td>
<td>0.40</td>
<td>0.59</td>
</tr>
<tr>
<td>L2</td>
<td>0.50</td>
<td>0.32</td>
<td>0.53</td>
</tr>
<tr>
<td>L3</td>
<td>0.53</td>
<td>0.36</td>
<td>0.55</td>
</tr>
<tr>
<td>(PFT) L4</td>
<td>0.62</td>
<td>0.43</td>
<td>0.66</td>
</tr>
<tr>
<td>L5</td>
<td>0.65</td>
<td>0.45</td>
<td>0.71</td>
</tr>
<tr>
<td>L6</td>
<td>0.86</td>
<td>0.69</td>
<td>0.89</td>
</tr>
<tr>
<td>k-means</td>
<td>L1 k = 15</td>
<td>0.40</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>L4 k = 8</td>
<td>0.47</td>
<td>0.37</td>
</tr>
</tbody>
</table>

As Table 9 illustrates, results show that overall accuracy (OA) for dominant cover classes (L1-L4) increased with increasing levels of aggregation (L1-L4). Similarly, more selective criteria generally increased levels of overall accuracy (L4-L6). The mean increase from the species-level class (L1) to the PFT (L4) was 8.25%, and the mean increase from PFT (L4) to stringent-membership PFT (L6) was 25%. Kappa values followed similar patterns, and illustrate that the certainty of each classification improves with increasingly broad classes.

The results of the k-means-clustered data were similar with several notable exceptions. The control, MC, and HC+MC predictors returned roughly equivalent OA and K, while OA and K for the HC model alone using L1 k-means clusters was lower (Table 9). The k-means classes based on the coarser aggregation (L4 PFT) had slightly lower accuracies than the k-means classes determined from the finer-level cover classes (L1). Table 11 presents an example fuzzy classification with a focus on BRTE cover; Figure 8 illustrates one overlap distribution between two classes.
### Table 10: Deterministic Confusion Matrix: Community Classes (L1 k = 15)

<table>
<thead>
<tr>
<th>CLASS</th>
<th>Reference Plots</th>
<th>Consumers accuracy:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUM</td>
<td>24</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: outlined boxes denote classes with >25% overlap in BRTE distributions

### Table 11: Cheatgrass Fuzzy (>25% overlap) Confusion Matrix: Community Classes (L1 k = 15)

<table>
<thead>
<tr>
<th>CLASS</th>
<th>Reference Plots</th>
<th>Consumers accuracy:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUM</td>
<td>24</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: outlined boxes denote classes with >25% overlap in BRTE distributions
Figure 8: Example Illustration of Percent Overlap in BRTE Distributions. This figure illustrates that two different classes (2 & 3 in this example) may have similar distributions of a particular cover type. Therefore, if BRTE is the species of interest to the map user, confusion between these classes is acceptable. This commutability forms the basis for ‘fuzzy’ confusion. The level of acceptable overlap is additionally determined by the map user.

Discussion

This study assessed the relative importance of predictors and response variables with regards to creating a vegetation classification. The capabilities of cloud-based computing remote sensing analysis combined with the Sentinel-2 platform enables exploration of these data dimensions. Broadly speaking, the results of this chapter illustrate that the increased capabilities from new satellites and computation platforms (e.g. “Landcover 2.0”, Wulder et al. 2018) allows us to reevaluate how field data are described as “classes”, and how the results are interpreted.
A faithful representation of pixel-scale vegetative and land cover can be created without using advanced unmixing techniques of hyperspectral data. Vegetation classification using ecologically-meaningful communities as classes (as discussed in Assigning Ecological Classes and in Figures 6 and 7), combined with user-specific fuzzy confusion matrices (Tables 4 and 5) is a better approach than classifications using majority cover as classes and a single deterministic error matrix. This new method requires robust field observations and inclusion of some temporal remote sensing data, but the results are worthwhile for two reasons. First, community-classes are a more faithful representation of vegetative communities in semi-arid and dryland ecosystems, since very few species or types exist as 100% cover at the 10-meter-by-10-meter pixel scale. Second, the overlapping distributions of each species or type between clusters allows for fuzzy confusion between classes based on the question at hand. Although this approach requires more engagement from the map user, it enables greater flexibility in how the classification is used and most likely increases the probability that the information in the map is used to inform management or planning decisions.

**Predictor Variable Importance**

The results of this study indicate that cloud and shadow masking, temporal compositing, and harmonic coefficient calculation pains may not be worthwhile. The least cloudy dates that fully cover the study area appear to yield comparable results to the masked monthly composites, and to the harmonic models. It may be that the benefits of the additional time intervals (spanning the season, and 11 instead of 8) were countered by the reduction in temporal certainty because of the temporal compositing. In other words, the monthly composites can introduce up to a 30-day disparity in spectral values
representing areas of the same cover type. The 30-day window was chosen for this study due to the inconsistent nature of Sentinel-2A’s 2016 orbit pattern (revisit period ranges from 2-20 days). Studies using 2018 data will be able to take full advantage of the consistent 5-day revisit period, and can therefore choose smaller compositing windows.

The comparable success of the control dates may also be due to a fortuitous eight particular days (Table 7) that happened to capture enough differences in phenology, and therefore this result is not transferrable to other study areas or years. Iteratively removing several of the cloud-free control dates and observing decreases in accuracy will provide further insight to the influence of particular dates, or number of temporal observations needed. Increasing the monthly compositing window to 60 days and observing the change in classification accuracies could also elucidate the effects of compositing.

**Harmonic Model Performance**

The harmonic models alone have roughly equivalent overall accuracies to the MC, HC+MC, and control sets. If the volume of data is an issue (e.g. outside of GEE or similar cloud computing environment), the significance of comparable accuracies is important: the HC predictors were able to classify nearly as well as using 34% of the data size of the MC predictors (15.1 GB vs 44.3 GB).

The lower kappa values of the HC models indicate a higher portion of classification accuracy is attributable to chance. Although cursory exploration indicated that the RMSE value (‘goodness-of-fit’ of harmonic sinusoid to the data) was equivalent in predictive ability to the amplitude and phase, it could be related to some of the classification error. Further examination of the individual bands demonstrates that the RMSE band has obvious artefacts resulting from the Sentinel-2 tiling scheme (Figure 9:
Harmonic NDVI Separate Bands). The overlapping areas of the imagery are caused either by delivering the same pixel values with the two adjacent tiles, or the same area is imaged a day apart (the diagonal stripe from NNE to SSW), or both. This mosaicking issue requires further scrutiny.

Figure 9: NDVI Harmonic Coefficients. Image a) illustrates the three predictor coefficients (b-d, normalized to 1) in Saturation-Hue-Value, respectively. The background of a) shows the number of Sentinel-2 scenes used for the harmonics, where overlap in delivered MGRS tiles and the data-take (diagonal) artefacts are apparent. The RMSE band b) contains the majority of calculation artefacts.
Selecting Categorical Levels of Response Variables Is Important

Plant Functional Types More Reliable than Species Levels

The results exploring measures of dominant cover class distinctions demonstrate that aggregation at the PFT level are more accurately classified compared to species-level class assignments, and that the threshold of class membership significantly influences classification outcome.

Splitting the cheatgrass class (‘BRTE’) into two classes representing majority cover (50-100%) cover and dominant cover (less than 50%, but still the largest fraction) led to a decrease in accuracy largely due to confusion between the two. This result demonstrates the strength of the temporal signature of BRTE [Clinton et al., 2010] as dominant cover is only mildly influenced by the other codominant signals. It could also be an effect of the intrusion of BRTE into all community types; one third of field plots have greater than 20% BRTE cover. Additionally, grouping large perennial bunchgrasses (AGCR + PSSP => BUNCH) and salt-desert shrubs (KRLA + ATCO => SALT) for the purposes of balancing class sizes decreased overall accuracy for all predictor variable trials. This could be due to increasing the variability of signals defining the class, to the point where they begin to overlap with other classes. For example, the predictors defining the BUNCH class may begin to overlap with those of the POSE class, another perennial bunchgrass. This confusion may be compounded by the use of multi-temporal data, where similar phenologies may be offset in time caused by small differences in climate (e.g. higher elevation) can lead to confusion between classes. However at the level of PFTs, similar signals are ascribed to classes more representative of more general
phenologies such as annuals and perennial grasses. This level had the highest overall accuracies due to these broader similarities.

**Stringent Requirements for Class Membership Increases Accuracy, at a Cost**

The largest increase in overall classification accuracy was observed in the differences between the dominant PFT plots as training data (L4) compared with only selecting plots containing more than 50% of the PFT cover (L6). Selecting training data from plots whose second/codominant PFT was significantly less than the primary/dominant PFT ($\text{PFT2} < \text{PFT1} - (\mu + \sigma)$; L5) increased overall classification accuracy slightly between all predictor variables. This is analogous to selecting endmembers (EMs) in the process of spectral unmixing.

However, it is more likely that the increase in overall accuracies (as well as kappa coefficients) are attributable to a reduction in the number of classes, as higher standards for membership begin to diverge from the realities of semi-arid vegetative ecotypes. For example, only three classes remain after imposing thresholds for L6: shrub ($n = 4$), bare ($n = 50$), and annuals ($n = 96$). Such a classification could not be used for land management purposes even if these were the only classes of interest; all of the shrub plots were classified as annuals. Should it have been accurate, errors of omission would likely be greater for shrubs since the threshold for PFT class membership ($> 50\%$ cover) is not representative of most shrub communities in the study area.

**Soft and Fuzzy Classification Schemes**

The community classes followed the opposite trend in accuracy that we observed with the dominant cover class assignments (L1-L4); as the level of aggregation increased, overall accuracy decreased. This is attributable to the process of clustering data: more
dimensions (i.e. greater numbers of variables) better distinguish between classes. Thus, there are more ways to distinguish between k-means clusters generated at the species-level than at the PFT-level. Pursuant to the overarching goal of producing a representative vegetation map, this is a desirable corollary: k-means clusters at the species level more accurately represent higher-level descriptions of land cover that are of interest for land managers and ecosystem demography modelers.

However, the process of determining the number of clusters (k) and their centers is problematic as it can bias the resulting classes. While a larger k improves cluster distinction (within sum-of-squares), resulting clusters can be overly-specific by grouping a handful plots with very similar cover distributions. Fewer clusters can result in poor distinction, where one or more clusters may have loose cohesion (occasionally called a ‘garbage bin’ cluster) or represent large distributions of one or more cover types (e.g. plots with ≈50% BRTE but no other commonalities). The cluster centers for this study were not set with a priori information in order to avoid imposing preconceptions of community types, and instead relied on qualitative assessment to ensure that each plant species of interest was represented in a cluster. However, the success of the cluster assignments is confounded by two cover types, BARE and BRTE. These cover types have a relatively large range of fractional cover (e.g. 0% to >85%) and many plots. Figure 4 illustrates the pervasive cover of BARE and BRTE: nearly all training plots have some proportion of one or the other. BARE has an interquartile range (IQR) of 31% cover, and BRTE of 64%. As a result, clusters with some portion of BARE or BRTE are often confused. Table 11 shows this confusion, particularly between clusters 5 and 11, which are both defined by large percentages of BRTE.
Further examination of the clusters and their constituent data point to likely sources of this confusion. The Within-Sum-of-Squares (WSS) had a Pearson’s’ Correlation Coefficient of 0.88 compared with the commission errors, and a Pearson’s’ Correlation Coefficient of 0.65 when compared with the omission errors. This indicates that the cluster is sufficiently large to near (or intersect) the distribution of other clusters in one or more dimensions. The L1 k = 15 classes 5, 11, 2, 9, and 4 all had most of the highest WSS. This implies that reducing the number of outliers within each cluster (reducing the WSS) or of the field data would likely reduce confusion between classes.

The degree to which this source of error can be reduced likely depends on the ecology of the area in question; vegetative types that exist on a broad continuum and amongst a variety of other community types may require many classes to describe them, or not be able to be separated into distinct groups at all.

Cheatgrass (BRTE) fits this description, and the classes with the highest confusion errors have high proportions of BRTE as their constituents. This indicates that the pervasive distribution of BRTE is a significant source of confusion. Closer examination of the distributions of BRTE in each class would better qualify which confusions are acceptable. Repeating this for each class and repeating the ‘focus’ of the confusion matrix would allow for more effective end use. Should the map user be primarily interested in ARTR, referencing the ARTR-based confusion matrix would be more informative. In this way, the imposition of important classes on the data is preserved for the user, not the producer.
Issues, Improvements, and Future Steps

Remote Sensing Imagery

Computational limits due to the spatial resolution and study area extent necessitated pre-filtering S2 tiles of excessive cloud cover. Geography outside of the study area but within the same tiles (i.e. the mountains) means that much of a tile can be cloudy, but the portion covering the scene is cloud-free. Methods to calculate the relevant cloudy portion of the study area prior to cloud masking were abandoned due to time constraints, but would potentially allow the inclusion of additional dates.

The predictive ability of the harmonic models will likely improve with increased temporal resolution. Studies from mid-2017 will benefit directly from the increased temporal coverage from Sentinel-2B. Increasing temporal coverage in 2016 could be achieved several ways. Further investment in including all possible S2 scenes (as above) would be one method. Other methods that incorporate other imaging platforms such as the ‘harmonized’ S2 and Landsat data (Claverie et. al 2016), or using MODIS and STARFM to ‘hallucinate’ S2 imagery [Gallagher, 2018] could be used. More specifically, I could have investigated the threshold for the effects of aberrant (i.e. cloud or shadow) values on the fit of the sinusoid, potentially informing the masking threshold value.

Increasing image quality would also likely benefit classification accuracies. Although L-1C images are geometrically corrected and delivered to within 1/3 of a pixel, co-registering all images may be beneficial in ensuring that the field data is properly represented by the same signal over time and not neighboring pixels which may be similar in composition, but have different proportions of cover types. Using L-2A
(Bottom-of-Atmosphere reflectance) would likely be a significant improvement especially for a semi-arid ecosystem where observed reflectance’s can be reduced [Beck et al., 2006].

Including other imagery of different types may increase accuracy of these classifications, as the results of Chapter 2 (below) indicate. For example, including indices from radar (e.g. Sentinel-1) can add information about the physical ‘texture’ or soil moisture of each community type. A time-series of radar data would potentially be able to observe physical changes over a growing season, such as the rapid growth of cheatgrass in the spring or inflorescence of sagebrush later in the fall. Similarly, adding other sources of predictor information such as precipitation, soil type, or land use history (such as grazing and fire) may further improve classification accuracies. In order to preserve the clarity of the research question, such data was omitted for this study.

**Training Data**

Several aspects of this study can be improved regarding training data. The k-means method to determine community classes allows the data to ‘speak for itself’ but it may be beneficial to examine each member of each class to remove possible outliers. Other more complicated ecosystem demography models may create more ecologically meaningful community classes. An iterative process of determining community classes in tandem with field data collection could create more homogenous cohorts, or inform the proportion and rate of change between two community classes. For example, there may be little gradient between rabbitbrush communities compared with shadscale communities, but a steady gradient between shadscale and winterfat communities. Performing a regression of each cover type (at the species level or PFT level) to calculate
percent cover (as in Chapter 2) may lead to a similar outcome with greater degrees of
certainty and more easy interpretability, but may not account for the complex interactions
of optical wavelengths due to plant canopy structures [Disney, 2016].

**Model Development**

Although the RF models are able to relate the training data and satellite imagery,
a focused effort to examine the relationship between the ‘optical type’ concept proposed
by Ustin and Gamon (2010) or similar measurements and the community classes would
add surety to the classification model [Cingolani et al., 2008; Huesca et al., 2015]. Other
data fusion techniques (such as in Chapter 2) may also improve classification. Further
refining of important predictor variables and RF model tuning performed external to GEE
but implemented therein has anecdotally been observed to improve model performance in
prior iterations of this project, meriting further investigation. Although this study
attempts to preserve interpretability throughout the process, more advanced machine-
learning algorithms or neural networks may be able to provide better classification
outcomes and more robust metrics of interrelations of classes (measures of fuzziness).

**Implications**

This project reaffirms the paradigm of matching the spatial scales of subjects of
interest with the imagery used to observe them. Increased spatial, spectral, and temporal
remote sensing imagery such as Sentinel-2 prompt a re-examination of what schemes are
appropriate for classifying field observations. Additionally, the availability of cloud-
computing platforms allows study areas to increase in size and across time; this requires
more careful consideration of land cover gradients. For example, this study initially
included the entire MHAFB Range Complex extending the southern range of the study
area by approximately 63km SSE and the elevation by an additional 400 meters. This introduced a significant amount of variability in phenology between communities of similar composition, and was therefore excluded.

The soft and fuzzy classification model presented here has the potential to better inform land management and scientific research. It preserves interpretation for the user by the use of fuzzy confusion matrices tailored to the class in question. However, it has been noted that a single map summarizing most land cover information is preferred [Cingolani et al., 2004].

**Conclusion**

In general terms, this chapter illustrates that the limiting factor of classification accuracy is due to the training data, not the quantity of predictors. While other approaches using even higher spectral, spatial, or temporal resolutions may not face this same limitation, the Sentinel-2 platform offers a new freely-available compromise of resolutions. This, combined with cloud computing, raise a reminder to evaluate the paradigm of resolution and the scale of subjects.

Corollary to this and specific to semi-arid ecosystems (and those with similar scales of ‘patchiness’) is the possibility to use ecologically-meaningful classes instead of majority-cover classes. Such classes may more closely align with observed remote sensing observations over time, due to the mixed pixel effect and the interaction of light with canopy structure.

Monthly temporal composites might sacrifice some temporal signal, and it may be that a careful selection of cloud-free days is equivalent in accuracy and avoids the headaches of cloud and shadow masking. Harmonic models are also approximately...
equivalent in accuracy, but are an additional level of processing and likely not warranted unless there are limitations in data storage and processing speed.

Input classes are generally better with coarser levels of aggregation (i.e. PFT) and thresholds of membership, but lose levels of ecological meaning in trade. Management needs and research can be better served with more representative classes. For example, there is a key uncertainty in land surface models with the use of PFTs [Hartley et al., 2017]. There are many classification algorithms for relating field observations with remote sensing data, but I posit that more effort should be paid to the methods by which we assign classes to the field data. These approaches should seek to balance the scale of the target subject (e.g. ‘drylands/forest/prairie’ vs. ‘sagebrush/invasive annuals/bunchgrasses’) with the capabilities of the sensing platform. In such a fashion, veracity of what field data describes is preserved to its maximal extent.
CHAPTER TWO: REMOTE SENSING OF BIOLOGICAL SOIL CRUST

Abstract

Biological soil crusts (or ‘biocrusts’) are an important but under-studied component of dryland and semi-arid ecosystems. These ecosystems have large influences on global carbon and nitrogen fluxes, in addition to much of the world’s population health and well-being. Biocrusts are understood to play a significant role in carbon and nitrogen fixation, preserving the health and stability of dryland ecosystems. But this understanding is informed by plant- and plot-scale studies, and refining estimates of global impact remains difficult due to a deficiency of spatial data of biocrust cover. In addition to nutrient fluxes, biocrusts play large roles in soil stability and micronutrient capture, and increase soil moisture by increasing infiltration and decreasing evaporation. Understanding the impacts of these functions on the landscape scale is also hindered by lack of spatial data.

This chapter builds on the methods of Chapter 1, adding more biocrust-specific spectral indices, structural information (from radar), and soil predictor variables. The random forest model is run in regression mode to determine the most significant predictors for biological soil crust cover. This both improves model accuracy, and allows for interpretation of reasons for predictor significance. Biocrust cover within the study area is calculated with an observed 74% coefficient of determination, in addition to validation using the outputs of Chapter 1.
Background and Introduction

Semi-Arid Ecosystems and Biocrusts

Scientific papers with study areas in dryland environments appear to be required to point out three things about drylands by the end of their first page: 1) that they occupy nearly 40% of Earth’s terrestrial surface [Millennium Ecosystem Assessment, 2005], 2) that they are fragile and increasingly affected by changes in climate and land use [Li et al., 2015], and 3) that our understanding of their function and global impact is enough to know that they are important but that our knowledge falls short of adequate [Millennium Ecosystem Assessment, 2005; Huang et al., 2016].

Biocrusts are the quintessence of these points. Biocrusts are communities with foundations of cyanobacteria and algae, also containing bacteria, lichens, mosses, and microfauna in varying proportions that live on and within the top few centimeters of soil surfaces. Although biocrusts can be found on all continents [Elbert et al., 2009; Weber et al., 2016] and are one of the most dominant community types on Earth [Weber et al., 2016], their study as an ecosystem component is relatively young – coming of age within the last four decades [Weber et al., 2016, chap.2]. Biocrusts (also called biological soil crust, or occasionally cryptogamic-, cryptobiotic-, microphytic-, mycrobial-, or microbiotic crust) are most commonly found in the interspaces between vascular plants and under their canopies in arid and semi-arid ecosystems. In some arid ecosystems, they are more than 60% of the land cover [Chen et al., 2005].

Because biocrusts live on the margins of what is livable, they are very sensitive to changes in climate and disturbance [Belnap et al., 2006; Elbert et al., 2012]. Scientific investigation of biocrusts has greatly increased in the last several decades, with fewer
than 20 publications in 2000 but over 160 in 2014 (as discussed in Weber et al. 2016). There has been a growing realization that biocrusts play a notable role in global carbon and nitrogen pools and fluxes, but a dearth of spatial understanding is a constraint in quantifying biocrusts’ contribution to natural processes at all scales.

What are Biological Soil Crusts?

Biocrust as an ecological unit is a relatively recent stand-alone field [Belnap and Lange, 2003], and is growing rapidly [Lange and Belnap, 2016, chap.2; Weber et al., 2016, chap.12]. A significant portion of new research is focused on understanding the biological processes and components of these remarkably diverse and highly complex communities. There is also a growing body of scientific literature exploring the ecological functions and impacts of biocrusts from site- to global-scales.

Although demure in appearance, biocrusts play a critical role in arid and semi-arid ecosystems. Filamentous cyanobacteria and algae (along with other structural elements of lichens and bryophytes) form a matrix within the top 1-2 cm of soil [Weber et al., 2016, chap.1]. As this structure develops from several millimeters in thickness, the diversity of organisms increase and the ecological impacts of biocrusts typically increase [Weber et al., 2016, chap.1]. As a physical construction and a biologically-active community, biocrusts influence nearly all transfers of gases, nutrients, and water between the land and atmosphere [Weber et al., 2016, chap.1]. Biocrusts have additional impacts on ecosystem processes, such as carbon and nitrogen fixation in forms beneficial to vascular plant growth [Weber et al., 2016, chap.19].

Depending on community composition and environmental conditions, a biocrust can take anywhere from 20-250 years to mature [Belnap, 1993; Hilty et al., 2004].
Although there is a fair amount of uncertainty in recovery rates effectuated by being a young field of study, evidence of WWII-era military exercises in the Mojave Desert are still visible through differences in biological soil crust development where heavy vehicles travelled [Belnap and Warren, 2002]. Recovery from severe disturbance begins with the re-establishment of the cyanobacteria (or algal) matrix [Hilty et al., 2004]. Although slow-developing as a community, biocrusts can by spry; photosynthetic activity can begin early in the season (before vascular plants begin photosynthesis), and virtually instantaneously with exposure to precipitation [Rodríguez-Caballero et al., 2017]. Biocrusts are also sensitive to long term changes in precipitation and temperature [Ferrenberg et al., 2015]. Observing changes in biocrust cover and composition spatially and through time can be used to assess ecosystem health, disturbance, and changes in climate [Belnap et al., 2001; Kirol et al., 2012; Blay et al., 2017].

Biocrust can constitute nearly 70% of the groundcover in some dryland ecosystems [Belnap and Lange, 2003], and can be found on all continents [Elbert et al., 2009; Weber et al., 2016, chap.3]. Remembering that dryland ecosystems cover approximately 40% of terrestrial land surface [Millennium Ecosystem Assessment, 2005] contextualizes the importance of understanding biocrusts’ spatial component on landscape and global-scales. The high variability of biocrust composition - even within similar ecosystems - means that extrapolating small-scale observations or inferring spatial cover is problematic. Chamizo et al. [Weber et al., 2016, chap.17] conclude that in order to understand the impacts of biocrust on hydrologic processes and clear up contradictory findings that “studies at larger spatial scales (hillslope, landscape, catchment)
incorporating temporal variability are an essential component of future biocrust research,” [Weber et al., 2016, chap.17].

The variability and the ‘patchy’ nature of biocrusts within a landscape that make scaling up observations problematic also make traditional ground-based mapping techniques difficult [Weber et al., 2016, chap.12]. There are two remote sensing approaches that may be employed to detect biocrust directly: 1) use distinguishing spectral characteristics of biocrusts, or 2) exploit the phenological differences between biocrusts and vascular plants.

Remote Sensing of Biocrusts

Important Physiogeny of Biocrusts Regarding Remote Sensing

Phycobiliprotein and carotenoid signatures in the blue-green region (≈430-544nm) are also unique features that have been observed in cyanobacteria-dominated biocrusts [Weber et al., 2016, chap.12]. Additionally, overall reflectance of cyanobacteria-dominated biocrusts have been observed to be lower when compared to bare soil [Weber et al., 2008]. Spectral characteristics of moss- and lichen-dominated biocrusts may not present similarly in this portion of the spectrum, but there is still insufficient research comparing spectra amongst different biocrust communities [Weber et al., 2016, chap.12].

Biocrusts’ Response to Water

As mentioned previously, water has an immediate influence on the spectral properties of biocrust. Within minutes of watering, the absorption feature from chlorophyll (≈680nm) deepens, remaining distinguishable up to five weeks of drying following a 10-day wetting event [Weber et al., 2016, chap.12]. Following a two-week
rainfall event, biocrusts were observed to have a reflectance profile similar to vascular plants: a local maxima around 550nm, a strong absorption feature around 680nm, and a sharp rise around 700nm (known as the “red-edge”), and a plateau from 700-1000nm [Karnieli et al., 2002]. The same study observed that by the end of the dry season the reflectance of biocrust closely matched that of bare soil [Karnieli et al., 2002]. Several general observations of note are 1) the overall spectral profile of biocrusts are lowered following a watering event (i.e. lower albedo; Karnieli & Sarafis 1996) and is also lower than vascular vegetation although similar in profile [Karnieli and Tsoar, 1995], 2) longer watering events yield more robust spectral signatures [Karnieli et al., 1999, 2002; Rodríguez-Caballero et al., 2015], 3) disturbance leads to increased overall reflectance (i.e. higher albedo; Ustin et al. 2009; Chamizo et al. 2012), and 4) that there are spectral variations according to soil substrate, biocrust composition, the phenological state of the biocrust, and vascular plant phenological states [Rozenstein and Adamowski, 2017].

History of Remote Sensing Approaches to Detect Biocrust

Spectral indices have been used to detect and map biocrust since the early days of spaceborne multispectral imagery [Wessels and van Vuuren, 1986]. Since then, the Normalized Difference Vegetation Index (NDVI = (Near Infra Red - Red) / (NIR + R)) has been used due to the chlorophyll contained in both vascular plants and biocrust, with moderate success [Karnieli et al., 1996, 1999, Burgheimer et al., 2006c, 2006a; Zaady et al., 2007]. In 1997 Karnieli proposed the Crust Index (CI = 1 - (R - B) / (R + B)) based on the observations of biocrusts’ higher reflectance in the blue region relative to vascular chlorophyll-bearers [Karnieli, 1997]. The CI struggles with biocrusts that are not dominated by cyanobacteria [Li et al., 2014]. As an improvement to the CI, Chen et al.
[2005] proposed the Biological Soil Crust Index (BSCI = \(1 - L * |R - G| \) / (Average(G + R + NIR))) where L is an empirical adjustment parameter between 2 and 4. The BSCI and has been used successfully, but struggles when vascular plants are also conducting photosynthesis [Chen et al., 2005; Weber et al., 2008; Potter and Weigand, 2016]. Other indices have been used to discriminate between vegetation and bare ground with success in arid and semi-arid ecosystems (such as the Soil-Adjusted Vegetation Index, Enhanced Vegetation Index, Water Index, and their derivatives). The development of multispectral indices has been limited to using the spectral bands of the Landsat and SPOT satellites, which do not collect data in the red-edge portion of the spectrum. Although the difference between NIR and R is effective in discriminating between photosynthetic and non-photosynthetic pixels, the additional information that exists in-between is important when the relative changes may be small - as in the case of biocrust. The addition of the red-edge band in spectral indices has been demonstrated to greatly increase the accuracy of biomass estimations in sparsely-vegetated and semi-arid environments [Schumacher et al., 2016]. Okin et al. review several caveats in using spectral data in semi-arid environments: desert plants often lack a strong signal in the red-edge, and that shrubs in arid and semi-arid are often highly variable in spectral profiles possibly due to rapid phenological changes [2001]. In concluding remarks discussing remote sensing of biocrusts, Weber and Hill state that, “...further studies are needed to evaluate the overall explanatory power of spectral data,” [Weber et al., 2016, chap.12].

Exploiting the temporal variability of biocrusts is an additional method to circumvent some of the difficulties in separating the signatures of biocrust from both soil (when dry) and vegetation (when wet) and to overcome difficulties previously discussed.
Burgheimer et al. observed the changes in spectral reflectance of biocrusts over two consecutive years, and demonstrated that biocrusts and annual plants could be differentiated using phenology [2006a]. They determined that the window in which biocrusts responded to moisture - but before the development of annual plants - was approximately two weeks [Burgheimer et al., 2006b]. Elste et al. used multi-temporal RapidEye imagery over two seasons (images ordered every 10 days; 9 used for the first year and 13 the second) to differentiate between biocrust, annuals, and perennials in the Northern Negev Desert in Israel [2015]. Elste et al. [2015] evaluated the inclusion of RapidEye’s red-edge band by using the Normalized Difference Red Edge Index (NDRE = ((RE - R) / (RE + R)). Barnes et al. [2000] found that it was able to detect the changes associated with perennial vegetation where NDVI and Modified-NDVI (MNDVI = (NIR - R) / (NIR + R + 2*B)) were not. They concluded that MNDVI was better at predicting areas of biocrusts, and NDVI was the best predictor for annual vegetation; however, the authors did not present biocrust-specific indices in their study. The importance of using multi-temporal imagery is emphasized in Weber and Hill’s concluding remarks:

“The high variability of different biocrust types (i.e., cyanobacteria-, lichen-, or moss-dominated), which frequently will be found associated to or mixed with each other, adds additional complications which render also the exploitation of such multi-temporal observation schemes anything else than trivial.” [Weber et al., 2016, chap.12]

Furthermore, Weber and Hill state that:

“Improving the production of reliable maps of biocrust cover further depends on the availability of imaging systems which provide not only adequate spatial and spectral resolution but are also capable of collecting images sufficiently
frequent…. This primarily qualifies airborne imaging or, alternatively, space systems which build on multi-satellite constellations for implementing sufficiently dense repetition rates, such as RapidEye…” [Weber et al., 2016, chap.12]

While Weber and Hill’s suggestions of both the need to explore additional spectral indices and use multi-temporal imagery both have sound bases, they offer no suggestions on how to use both in concert. While calculating spectral indices is relatively straightforward, determining which possible new spectral indices can detect biocrust or evaluating multi-temporal data would be a difficult and time-consuming process using traditional methods.

Because biocrust is not often the ‘dominant’ cover at scales in the 10s to 100s of meters (i.e. commensurate with satellite remote sensing pixels), it is important to examine other spatial components of communal or climatic associates. In an overview of the distribution patterns of biocrust, Bowker et al. [Weber et al., 2016, chap.10] describe the influences on biocrust distribution as biogeographic, climatic, edaphic, topographic, and biotic. Biogeographic (such as distance between land masses) and climatic influences act on the larger biome scale. Edaphic, topographic, and biotic factors are more dominant when examined on ecoregional and smaller scales [Weber et al., 2016, chap.10]. When observing biocrust patterns in sites from 1 ha down to several m² (ecoregional and smaller scales), the differences in community composition beneath shrub canopies and those in the plant interspaces becomes more noticeable. In a study of functional diversity of heterogeneous biocrust communities, Li et al. [2014] state the need for further studies into how to quantitatively measure the spatial distribution of biocrusts on the scale of several hectares.
While the relationship between biocrust types and vascular plants is a complex two-way relationship, it is understood that certain biocrust communities promote vascular plant growth (i.e. vascular plants are correlated with biocrust cover; discussed in Weber et al. 2016, chap.19) and that plants can provide microhabitats for biocrust (i.e. biocrusts are correlated with areas of shrub cover; reviewed in Weber et al. 2016, chap.10). Zhao et al. observed that when the effects of erosion or grazing are high, biocrusts may ‘refuge’ near shrubs [2010, 2011]. This may be the case in the Morley Nelson Snake River Birds of Prey National Conservation Area and surrounding public lands, where grazing and recreation are nearly ubiquitous and the spatial distribution of biocrust is not well understood. In a study of the variation in bacterial composition of biocrust in the Reynolds Creek Critical Zone Observatory in southwestern Idaho, Blay et al. [2017] found a strong correlation (P < 0.001) between community composition and elevation, and posited that some variation may be attributable to land use changes. While Blay et al. did note the difference in shrub composition with elevation, they did not discuss any implications between relative shrub cover and biocrust composition. Additionally, biocrust community structures have been found to correspond with soil type and climatic gradients [Bowker and Belnap, 2008; Blay et al., 2017], which may be correlated to vascular plant associations.

Research Objectives and Importance

Remote sensing technologies (imaging platforms and computational methods) have progressed significantly since the early experiments that included biological soil crust as a landcover type. Scientific understanding of biocrust has also progressed, and its ecological roles are understood to be an integral component to many other fields of earth
surface processes on varying scales. Yet there exists a dearth of spatial information about biocrust cover.

Remote sensing of semi-arid vegetation is difficult for several reasons (as discussed in Chapter 1), and the influence of ‘background soil’ is well-known [Dashti et al., accepted]. Understanding when the background component of the observed signal is exposed soil and when it is biocrust may increase the accuracy of remote-sensing vegetation mapping projects. The growing body of scientific literature on biocrust is primarily at the plant scale, and the ability to scale this knowledge up to ecosystem scales is an acknowledged research gap [Weber et al., 2016, chap.12]. Thus, refining remote sensing methods to map biocrust cover at the landscape scale are of interest. The ability to repeat this mapping annually (e.g. straightforward, programmatically) is also beneficial in that change in cover over time can inform land management decisions, and aid in the understanding of the interactions of biocrust with global climate and nutrient cycling. Applying new satellite platforms and computational methods to classification methods open new avenues to explore novel approaches of classifying a difficult cover type.

This study builds on the results of the previous chapter, examining accuracy assessments in finer detail, and investigates the addition of additional ecological predictor variables with a focus on creating a map of biological soil crust.

**Materials and Methods**

**Study Area**

The study took place in the Morley Nelson Snake River Birds of Prey National Conservation Area (BOP, including enclosed and neighboring military installations). This study area represents diverse semi-arid vegetation communities of Wyoming sagebrush;
green and grey rabbitbrush; salt-desert shrubs such as winterfat and shadscale saltbush; native and non-native perennial bunchgrasses grasses such as Sandberg Bluegrass, crested wheatgrass, and bluebunch wheatgrass; and invasive annual grasses and forbs such as cheatgrass, bur buttercup, and some mustards. Undisturbed communities of native vegetation are typically composed of sparse shrub cover intermixed with bunchgrasses, and an interspace of biocrust. Types of shrubs, grasses, and biocrusts can vary with soil types and gradient in climate [Compagnoni and Adler, 2014; Blay et al., 2017; Schwabedissen et al., 2017]. Further discussion of the study area and species can be found in Chapter 1.

This area is a good candidate for the study of biocrusts due to its multifarious cover as a byproduct of its many land uses over time. Due to biocrusts’ fragility and long regrowth period, areas that have experienced heavy use (e.g. overgrazing, recreation) typically have very little biocrust cover. The added factor of invasive species within the study area mean that biocrust is often replaced by Bromus tectorum (cheatgrass) or other invasive species. B. tectorum cover has a negative relationship with biocrust cover [Bansal and Sheley, 2016; Raymondi, 2017]. The other end of the spectrum - communities of native vegetation with robust biocrust cover - owe their thanks to the history of scientific study in the BOP and to the military installations in the area. Numerous ‘exclosures’ have been established within the BOP, some dating back to the 1960s. Military installations also have restricted access to the public and to grazing, and have therefore preserved pristine community types. Raymondi [2017] also found a positive relationship between sagebrush cover and distance to the nearest grazing allotment with biocrust cover for unburned areas, and with native non-sagebrush shrubs
in burned areas. This gradient of biocrust cover with its known cover associations betoken the BOP as a suitable study location for exploring remote sensing approaches.

**Field Data**

This chapter uses the same field data as used in Chapter 1 (n = 378). Field data points were qualitatively selected as homogeneous communities of each vegetation type (Table 2: Selected Vegetation and Cover Types), and representative of different community combinations and variations among spatial gradients. Each plot is comprised of a field survey and five nadir-pointing images taken from a one-meter boom attached perpendicular to a two-meter tall pole. These images were taken along with an RTK GPS location at a central point and at seven-meters in the four cardinal directions (Figure 5: Schematic of Field Plot Design). Only the percent biocrust cover class (BSC) was used for this study.

Vegetation cover was quantified using SamplePoint software [Booth et al., 2006]. The SamplePoint software ("SP" or "SamplePoint") overlays a regularly-spaced 10x10 grid of crosshairs over the image, where a technician identified vegetation or land cover (including biocrust) indicated by each crosshair. Technicians with minimal training are able to discern biocrust types and developmental stages (e.g. ‘incipient cyanobacteria-dominated’, ‘mature lichen-dominated’) from images of equivalent resolution to this study [Belnap et al., 2008]. Creating classes for ‘biocrust’, ‘bare’, ‘rock’, and ‘unknown’ ensured that technicians had sufficient choices to describe observations, and were instructed to err on the side of ‘bare’ if it was uncertain if it was bare soil or nascent biocrust in order to increase confidence in the ‘biocrust’ observations. Levels of
development and community composition of biocrust (e.g. ‘incipient cyanobacteria-dominated’, ‘mature lichen-dominated’) were all classified as ‘biocrust’.

**Imagery & Predictor Variables**

Several sources of remote sensing and predictor data were included in this study. Multiple sources of data can provide insight to how small-scale understanding of biocrust properties relate to satellite-based observations (and other landscape-scale predictors such as precipitation and soil type, as described below). Knowing these relationships can inform future classifications of biocrust cover. An additional benefit of including multiple sources of data as predictor variables is increasing the accuracy of the model. For example, random forest models can handle multi-source datasets and generate high classification accuracies [Van Beijma et al., 2014].

**Optical Predictors**

This paper is informed by Chapter 1, and uses similar predictors from the same imagery, in addition to outputs used for validation. Google Earth Engine (GEE) was used to preprocess Sentinel-2 (S2) imagery from 2016 into monthly (30-day) composites, which included bands and spectral indices. Atmospheric bands (B1, B9, and B10) were excluded as predictors in order to reduce potential sources of error in the classification process. Harmonic models from Chapter 1 were also used to capture seasonal trends of biological soil crust. In addition to removing atmospheric bands, 23 bands identified as important predictors of biocrust were removed from the monthly composites. This reduced set of predictors was used to generate a species-level community classification as in Chapter 1. See Random Forest and Validation for additional discussion of predictor variables.
Radar Predictors

Sentinel-1 radar imagery was also used to add structural information as predictors. The two Sentinel-1 (S1) satellites are C-band synthetic aperture radars by the European Space Agency. Although the S1 satellites can collect radar data in several modes, interferometric-wide imagery was used for this study due to greater availability over the study area during 2016, and because this mode occasionally collects Vertical transmit/Horizontal receive (VH) polarization in addition to Vertical transmit/Vertical receive (VV). Coverage of the study area was inconsistent for 2016, necessitating the use of summary statistics (Table 12: Sentinel-1 SAR Summary Statistics). The summary statistics calculations also addressed the need to de-speckle each image, essentially replicating the multi-look processing approach [Wynne and Campbell, 2011]. This approach typically uses several images taken in short temporal steps to reduce spatial variability of aberrant values. This study used imagery for the growing season in a similar fashion. GEE pre-processed S1 imagery using the ESA’s Sentinel-1 Toolbox procedures of thermal noise removal, radiometric calibration, terrain correction, and conversion to 16-bit decibel values in 10 meter pixels. Imagery covering the study area was separated into ascending and descending before summary statistic calculations.
Table 12: Sentinel-1 SAR Summary Statistics

<table>
<thead>
<tr>
<th>Name</th>
<th>Orbit Path</th>
<th>Description or Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>VVmin, VVmax, VVmean</td>
<td>Ascending</td>
<td>Minimum, maximum, mean of collection</td>
</tr>
<tr>
<td>VVmin_d, VVmax_d, VVmean_d</td>
<td>Descending</td>
<td>Min, max, mean</td>
</tr>
<tr>
<td>VHmin, VHmax, VHmean</td>
<td>Ascending</td>
<td>Min, max, mean</td>
</tr>
<tr>
<td>NDVHV Vmin, NDVHV Vmax, NDVHV Vmean</td>
<td>Ascending</td>
<td>$</td>
</tr>
</tbody>
</table>

Geophysical Predictors

Two soil data layers (SSURGO and STATSGO2) were also added as predictors. The SSURGO is produced by the Natural Resource Conservation Service and is a collection of soil surveys collected at the 1:12,000 to 1:63,360 scales of ‘map units’ having similar properties. STATSGO2 (also known as the General Soil Map of the United States, or GSM) is a small-scale (1:250,000) dataset of soils and non-soil areas of general soil association units. Each layer was rasterized to 10m pixels to match the Sentinel-1 and Sentinel-2 pixels.

Climate Predictors

Correlation of vegetation response with precipitation was calculated with a seven-day lag in order to investigate the hypothesis that photosynthetic activity in biocrust responds quickly to precipitation. The Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), a gridded rainfall time-series dataset, was used for precipitation due to its relatively high spatial resolution (0.05°) and temporal (daily) resolution (Funk et al. 2015). All S2 images of 2016 covering the study area were used without cloud masking (the harmonic function mitigates extraneous values, Roerink et al. ...
2000; Roerink et al. 2003). A lag of seven days was chosen to capture the photosynthetic response of biocrust [Weber et al., 2016, chap.12] without confusing the response signal with additional S2 observations more likely indicating vascular plant responses. The correlated lag is calculated from precipitation events with the response (Figure 10: Precipitation and B8A). Five spectral responses were chosen to calculate the lag (NDVI, EVI, CCCI, ARI, and B11) due to their representation in the ranked variable importance (Random Forests and Validation, below).
Figure 10: Precipitation and B8A (Red Edge 865 nm) for a Sagebrush Plot (-116.3656650619256, 43.23953563635724), Vertical lines and arrows illustrate the correlated lag concept.

Random Forest and Validation

The random forest (RF) machine learning method [Breiman, 2001] was chosen for its ability to model non-linear interactions between the predictors and response variables, especially for complex interactions and without overfitting the data [Belgiu and Dragut, 2016]. Random forest models can handle multi-source datasets [Van Beijma et al., 2014], and can help identify important predictor data for remote sensing [Schumacher et al., 2016].

For this study, the RF algorithm was implemented in R (randomForest v4.6-14, Liaw & Wiener 2002) as well as in Google Earth Engine. All models were run in regression mode with 500 trees with out-of-bag sampling enabled, and using ⅓ of the
number of predictors sampled at each node. The top predictors for biocrust cover (BSC) were selected using a Gini Index threshold (Figure 11: Gini-ranked Variable Importance). These predictors were then used in a RF model in GEE using the same parameters to impute biocrust cover for the BOP. In addition to the RF internal out-of-bag error assessment, a linear model was fit between the observed biocrust cover and the imputed biocrust cover. The imputed cover was calculated using the mean pixel values contained within the buffered point, reflecting the method used to extract the predictor data.

![Figure 11: Gini-ranked Variable Importance](image)

Several approaches were used for validation of the biocrust cover predictions. We did not have an independent field sample validation dataset. Thus, we performed an intercomparison of two models. First, the Level 1 ecotype classes (species-level, 15 clusters) model from Chapter 1 was re-generated using the monthly optical composites.
In order to avoid using the same predictor data for this model and our RF biocrust cover model, the top predictors from the biocrust regression were removed prior to developing the ecotype classification model. Then, the RF calculated biocrust cover was compared to the biocrust cover distribution in the ecotype classification. As a final layer of validation, a biocrust-specific fuzzy confusion matrix of the Chapter 1 species-level community classification (L1 \( k = 15 \) using monthly composites) is presented.

**Results**

**Biocrust Cover**

Spatial extent and proportional cover of biological soil crust in the Birds of Prey NCA in 2016 (displayed in Figure 12) was calculated using field observations of known cover (n = 378) and predictors from optical and radar satellite imagery (Sentinel-2 and Sentinel-1), soil data (U.S. General Soil Map/STATSGO2 and SSURGO), and seven-day lag coefficients of optical predictor variables with precipitation (CHIRPS dataset).
Figure 12: Imputed Biological Soil Crust Percent Cover in the Morley Nelson Snake River Birds of Prey, 2016

Regression Results

The overall accuracy of the random forest model implemented in R using monthly optical, radar, geophysical, and climate predictors had a pseudo-$R^2 = 0.35$, MAE = 6.49%, and a RMSE = 8.50%. The RF model using only important predictor variables with a Gini Index greater than 200 ($p = 33$, Table 11: Gini-ranked Variable Importance) yielded a pseudo-$R^2 = 0.40$, MAE = 6.06%, and a RMSE = 8.12%. The observed biocrust cover compared to the imputed biocrust cover had an adjusted $R$-squared of 0.74. Figure 13 shows predicted versus observed percent cover biocrust using our model.
Comparison with Ecotype Classification Models

The overall deterministic accuracy (OA) of the ecotype cover map generated without the biocrust-specific predictor variables was 0.46 and a kappa of 0.40. Removing the biocrust-specific predictors did not significantly affect the performance of the original model (of monthly predictors, from Chapter 1), which had an OA = 0.49 and kappa = 0.43. The mean overlap percent of the RF imputed biocrust cover compared to biocrust cover of each ecotype was 0.75. Using the specific fuzzy confusion matrix method (discussed in Chapter 1) with a threshold of mean-centers within 5% cover, overall fuzzy accuracy was 0.65 (Table 13).
Table 13: Re-Ordered Confusion Matrix with Fuzzy Confusion of 5% Difference in Mean Centers

<table>
<thead>
<tr>
<th>Class</th>
<th>Reference</th>
<th>Mean Centers</th>
<th>Consumers</th>
<th>Class Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Biocrust</td>
<td></td>
<td>accuracy</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>25.1</td>
<td></td>
<td>0.29</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>19.7</td>
<td></td>
<td>0.46</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>19.4</td>
<td></td>
<td>0.35</td>
<td>17</td>
</tr>
<tr>
<td>14</td>
<td>18.2</td>
<td></td>
<td>0.55</td>
<td>29</td>
</tr>
<tr>
<td>15</td>
<td>17.9</td>
<td></td>
<td>0.55</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>12.6</td>
<td></td>
<td>0.43</td>
<td>24</td>
</tr>
<tr>
<td>1</td>
<td>11.4</td>
<td></td>
<td>0.27</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>11.4</td>
<td></td>
<td>0.00</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8.2</td>
<td></td>
<td>0.46</td>
<td>37</td>
</tr>
<tr>
<td>11</td>
<td>7.1</td>
<td></td>
<td>0.46</td>
<td>54</td>
</tr>
<tr>
<td>2</td>
<td>6.3</td>
<td></td>
<td>0.10</td>
<td>23</td>
</tr>
<tr>
<td>7</td>
<td>6.2</td>
<td></td>
<td>0.31</td>
<td>13</td>
</tr>
<tr>
<td>13</td>
<td>5.8</td>
<td></td>
<td>0.00</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>4.2</td>
<td></td>
<td>0.66</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>1.8</td>
<td></td>
<td>0.77</td>
<td>58</td>
</tr>
<tr>
<td>SUM</td>
<td>19 11 11 33 27 17 24 0 35 65 10 10 3 46 57</td>
<td>0.46</td>
<td>0.65</td>
<td></td>
</tr>
</tbody>
</table>

| Producers accuracy: | 0.26 0.55 0.55 0.48 0.41 0.59 0.29 0.46 0.37 0.20 0.40 0.00 0.46 0.65 |

Discussion

These methods demonstrate that biocrust cover can be mapped with 65% accuracy. The use of time-series techniques and the random forest model are able to predict biocrust cover, therefore justifying its inclusion as a cover type in field surveys. The results are a meaningful step toward addressing the dearth of spatial information of biocrust cover, and open the door to understanding the effects and processes of biocrust at the landscape scale.

A qualitative assessment of the predicted cover indicates that it is broadly correct in several aspects. Agricultural fields and water have very uniform predicted values, and although likely not indicative of biocrust at those locations, does suggest that the regression is relatively well constrained to biocrust and associated community types.
Known ‘hotspots’ of exotic annuals (including cheatgrass) from fire and overgrazing have very low biocrust cover. Dirt and gravel roads and parking area also show low biocrust cover, indicating that there is little confusion between this type of bare ground and biocrust. Appendix A compares bare cover and biocrust cover classifications.

Conversely, areas without a recent fire history and in areas of established sagebrush, winterfat, or shadscale cover have relatively high proportions of biocrust cover which confirms the plot-scale understanding of the relationship between biocrust and shrub cover.

Biocrust spatial data can be used in studies of biocrust directly, or to include biocrust as a factor in other studies. As of 2016, there was an identified need of spatial data at landscape scales in order to relate biocrust to other processes such as soil erosion and hydrologic processes [Weber et al., 2016, chap.17; Chamizo et al., 2017].

A map of biocrust cover could be used as a proxy or visible indicator for situations or processes that are not observable directly. For example, biocrust cover could be used as an indicator of grazing impacts related to different regimes (such as allotment timing or animal units), or as an indicator of ‘interspace’ for fire risk assessment. It should be noted here that the methods outlined in this paper do not distinguish directly between biocrust cover underneath a shrub versus interspace. Percent biocrust cover (in-between and under shrub canopies) has been found to be an important factor in microhabitat selection for nesting sage-grouse [Kirol et al., 2012].

On a global scale, the ability to map biocrust cover using remote sensing has implications for understanding carbon and nitrogen cycling. Elbert et al.’s meta-analysis of studies of cryptogamic cover (biological soil crusts, mosses in trees, lichen on rocks,
etc.) found that their global contribution is significant [Elbert et al., 2012]. Such cover constitutes approximately 7% of global terrestrial net primary production (roughly the equivalent of carbon released by burning fossil fuels). For nitrogen fixation, they found that roughly 46% of global biological nitrogen fixation is attributable to cryptogamic cover, 80% of which comes from the North American continent. Furthermore, the largest flux of biological nitrogen fixation is due to ground cover (Elbert et al. 2012). Models of the interannual variability of carbon flux in semi-arid ecosystems have identified an unknown sink [Ahlstrom et al., 2015]. Again, there is understanding that changes in climate have impacts on the functions of biocrust (such as the reduction of nitrogen-fixing collema, [Belnap et al., 2006] but the ability to relate this process to a landscape-, continent-, or global-scale. The knowledge from both plot-scale and global meta-analysis and modeling underscore the potential impacts of biocrust cover data on scales in between. Biocrust cover can further refine such global calculations and inform models. Repeatability of generating biocrust cover data also enables the impact of on carbon stocks with changes in climate (as in Poulter et al. 2014), or in examining the impact of a drought year on global terrestrial net primary production (as in Huang et al. 2016).

Random Forest Model Assessment

Although the results from the random forest regressions (as implemented in R) explain somewhat low percentages of the variability of biocrust cover (34.5% and 39.9% for all- and top-predictors, respectively) the resulting output describes observed biocrust cover well, as evidenced by the R-squared of 0.74 calculated from the linear regression (Figure 12). A k-fold cross validation of RF regression models would more precisely describe the ability of the predictor data to determine biocrust cover.
Interestingly, the RF model always predicted some percent biocrust cover (minimum predicted value of 0.4%). This is likely attributable in some capacity to how RF models ‘average’ predicted values over multiple trees. However, it may also be partially attributable to two other causes. It may be that there is more biological soil crust cover than identified in the field data. This is reasonable because we erred on the side of classifying incipient biocrust as bare, and the fact that our field sampling protocol – while representative – likely do not capture small portions of biocrust that may exist within the plot, or in neighboring pixels. It may also be that the biocrust signals from each pixel are influenced by vascular plants within the pixel, or that the signals may be reflective of one or more covariates of biocrust. Examining the community classification approach and the predictor variable importance ranking can potentially provide insight to this question. For direct use of this data layer, a minimum confidence threshold of biocrust cover should be established from literature or informed by the ecotype classification.

Predictor Variable Importance and Inferences

The importance of particular variables as determined through the RF model’s Gini index (a measure of “node purity”, Table 11) may support some of the foundational hypotheses of remote sensing of biocrust. The most important variable by a respectable margin is the harmonic amplitude of NDVI, as well as the amplitude of the IRECI (inverted red-edge chlorophyll index) as third most important. These show that the chlorophyll in biocrust or in plants co-varying with biocrust has a significantly unique signature; neither of these are in the similar top predictors for sagebrush or cheatgrass. Additionally, the CCCI (canopy chlorophyll content index) value for the second mosaic (≈April) is the fourth most important predictor; March CCCI is the seventh. The April
CCCI has a positive correlation with biocrust cover (Pearson’s Correlation = 0.259), and a negative correlation with cheatgrass cover (-0.286). The March CCCI has a similar relationship (0.22 and -0.27, respectively). This is interesting, because March and April are the primary growing season for cheatgrass. The NDVI of cheatgrass peaks around April in the Great Basin [Boyte et al., 2016]. The slight positive correlation with biocrust and slight negative correlation with cheatgrass likely makes the CCCI an effective value on which to split the two.

The ranking of the seasonal maximum radar metric of Vertical-transmit, Vertical-receive as the second most important predictor indicates that there is a signature structural characteristic of plots with biocrust, likely the ‘flatness’ of the plot. The sixth-ranked predictor is the maximum value of the radar Normalized-Difference Vertical-transmit, Horizontal-receive with the Vertical-transmit, Vertical-receive which further supports this inference. Further investigation is needed to confirm that the importance of the radar predictors is related to structure, as it is also possible that the signal is more indicative of soil moisture.

Also of note is that three B11 predictors (Shortwave Infrared, 1614 nm) were ranked in the top ten most important predictors. This spectral range has been found to correlate to Nitrogen content of plants [Herrmann et al., 2010]. Nitrogen fixation (through diazotrophy) is common to biocrusts, and is likely the dominant source of N in dryland environments [Evans and Ehleringer, 1993; Weber et al., 2016, chap.14]. While this study indicates that B11 may be important for detecting biocrusts, it is important to note that it is controversial whether N can be detected directly using remote sensing techniques [Dashti et al., accepted].
Also of note is that the Blue Wide Dynamic Range Vegetation Index (BWDRVI) is ranked in the top ten most important predictors, which could be due to the noted increase in spectral reflectance in the blue portion of the spectrum in comparison to vascular plants [Weber et al., 2016, chap.12]. The finer-scale soil survey data (SSURGO) was also in the top predictors, which indicates that the soil type is either a predictor for biocrust cover directly, or indirectly in the sense that it is correlated with a vascular plant which is also correlated with biocrust (such as winterfat or shadscale).

Validation between Ecotype/Community Classification Model and Biocrust Regression

The predicted biocrust cover from the regression model falls reasonably well within the community type model (Level 1, 15 clusters). This may be biased by the similarities between many of the communities, however. Three of the clusters have mean centers of 11.4-12.6% biocrust cover, four with 17.8-19.7%, and five with 5.8-8.1%. This underscores the role of biocrust in the vegetation community types within the study area: they are often the background of healthy and stable communities. Although the term ‘interspace’ is often used in the context of native shrub and bunchgrass communities to describe the community, ‘interspace’ might be better described as biological soil crust, bare or exposed soil, or rock.

Overall, the agreement between the regression of biocrust cover and community-type classifications demonstrates that the observed signals of each pixel are representative of the product of its constituents. While not often the ‘dominant’ cover type, biocrust is a component of the community types and its fractional cover is detectable. If biocrust is not detectable directly it can be inferred. A secondary observation from this experimental design is that even with \( \approx 10\% \) of predictors removed
(24 out of 220), the community classification using monthly composites of optical predictors (Sentinel-2 bands and spectral indices) performed nearly as well but with some decrease in overall accuracy.

The fuzzy confusion matrix (Table 13) is an example of the proposed accuracy assessment method in Chapter 1. It shows that while the deterministic overall accuracy appears ‘low’, the cover distributions within each class allow the traditional confusion matrix to be presented differently depending on the scientific or management question at hand.

Issues, Improvements, and Future Studies

The results of this study show that the field data collection methods are able to capture biocrust cover reliably. Future work would be well-served to validate this using a handheld hyperspectral spectroradiometer or by collecting samples for laboratory verification. These efforts would increase confidence in identifying biological soil crust, physical crust, or exposed soil in the field photographs. Because biological soil crusts exist on a continuum, there may be significant confusion between nascent biocrust and soil in the training data. Future studies could investigate what development levels of biocrust are discernable by classifying field data to capture a gradient such as soil, soil-nascent biocrust, mature biocrust, etc. Using a cross-fold validation method using this field data would speak to the homogeneity of the points classified as biocrust. Were the cross-folds to vary accuracy of the model, it might illustrate that the signals are sufficiently different within the ‘biocrust’ class to warrant further investigation of what type is being detected, or if there is confusion between certain types and the bare soil class. Additional investigation into whether discriminating between ‘bare’ and ‘biocrust’
improves the community classification approach would be particularly informative to remote sensing studies in semi-arid ecosystems, where its existence is often overlooked or discussion about possible influences is omitted.

If the field data used in this study collected over the summer of 2017 or 2018, higher temporal resolution of Sentinel-2 imagery would be available. This would likely improve classification accuracy especially for biocrust. Additionally, it would be interesting to investigate different harmonic oscillations (e.g. \(3\pi\) or \(4\pi\) instead of just \(2\pi\)), or more advanced harmonic estimates of phenology such as Double logistic, Asymmetric Gaussian, Whittaker smoother, Savitzky-Golay (such as in Zhou et al. 2016). Investigating different lag coefficients between precipitation and spectral responses would potentially improve detection of biocrust since response windows may be smaller than one week. Soil moisture datasets (such as the NASA-USDA SMAP Global Soil Moisture Data) would also likely improve classification of biocrust if included with optical responses in some capacity. It was not included in this study due to computational limits. Land use data and other geographical parameters might further increase the predictive ability of a biocrust regression model.

**Conclusion**

This chapter has demonstrated additional capabilities of cloud computing and fusion of data from emerging platforms and datasets. Further investigation is needed to validate biocrust cover calculations in a robust manner. However, multiple approaches using community/ecotype-classification and biocrust cover regressions demonstrate that biological soil crust does contribute to signals received by remote sensing platforms.
This map of biocrust cover can be integrated with land management decisions and included in research. Land management decisions regarding grazing allotments and timing can seek to avoid areas of high biological soil crust cover altogether, or time grazing to minimize harm to dry and brittle biocrust cover. Additionally, land swaps between agencies or offset vegetation restoration projects can include measures of biocrust cover as a valuable land cover type. Fire risk assessment can include biocrust cover as a proxy for ‘interspace’. Land use studies can use biocrust cover as a proxy for degradation, as robust biocrust cover can take decades to develop.

Scientific inquiry can use this biocrust layer to examine spatial relationships with vascular plants, as well as animal habitat preferences. Soil studies can investigate biological soil crust as an influencing factor on erosion and deposition, or as a possible factor in water infiltration at the landscape scale. Studies and ecosystem demography models of carbon and nitrogen fluxes can include biocrust cover.

The ability of this cover map to be updated annually (given new or validated field data) enables scientists and land managers to examine changes over time. In such a manner, studies of biological soil crust can investigate annual changes in biological soil crust, or long-term trends in regrowth with changes in climate. Repeating studies such as these require the inclusion of biological soil crust in field data collection protocols.
REFERENCES


Barnes, E. et al. (2000), COINCIDENT DETECTION OF CROP WATER STRESS, NITROGEN STATUS AND CANOPY DENSITY USING GROUND-BASED MULTISPECTRAL DATA, in *Proceedings of the Fifth International Conference on Precision Agriculture*, ASA-CSSA-SSSA, Madison, WI.


Belnap, J., S. L. Phillips, D. L. Witwicki, and M. E. Miller (2008), Visually assessing the level of development and soil surface stability of cyanobacterially dominated
biological soil crusts, *J. Arid Environ.*, 72(7), 1257–1264,


Braget, A. (2017), Time Series Analysis of Phenometrics and Long-Term Vegetation Trends for the Flint Hills Ecoregion using Moderate Resolution Satellite Imagery,


Gallagher, M. (2018), UTILIZING SATELLITE FUSION METHODS TO ASSESS VEGETATION.


Gelbard, J. L., and J. Belnap (2003), Roads as Conduits for Exotic Plant Invasions in a Semiarid Landscape\r\nCaminos como Conductos para Invasiones de Plantas en un


Iiames, J. (2010), Accuracy sampling design bias on coarse spatial resolution land cover data in the Great Lakes Region (United States and Canada), Accuracy 2010 Symp., 389–391.


Millennium Ecosystem Assessment (2005), *Ecosystems and Human Well-Being: desertification synthesis*, World Resources Institute, Washington, DC.


Poley, A. (2017), DERIVING LANDSCAPE-SCALE VEGETATION COVER AND ABOVEGROUND BIOMASS IN A SEMI-ARID ECOSYSTEM USING IMAGING SPECTROSCOPY, Boise State University.


Raymondi, A. M. (2017), THE RELATIVE IMPORTANCE OF FIRE HISTORY, MANAGEMENT TREATMENTS, BIOTIC, AND ABIOTIC FACTORS ON THE ABUNDANCE OF KEY VEGETATIVE COMPONENTS IN AN ENDANGERED SAGEBRUSH-STEPPE ECOSYSTEM.


Comparing Biocrust Cover with Bare Cover

Running the same process as in Chapter 2, but for ‘bare’ cover instead would elucidate whether biocrust is discernable from bare soil. This demonstrates that technicians were able to distinguish between biocrust and bare soil using SamplePoint (v1.59, Booth et al. 2006) on our field imagery, and that there is a significant difference in the signals of bare soil and biocrust observable from satellite remote sensing and predictable using geophysical parameters. To clarify, ‘bare’ here means exposed soil or dirt which has no observable biological soil crust component.

To this end, the predicted variable in the model was changed from percent biological soil crust cover to percent bare cover. Predictor variables from the monthly composites model were used ($\rho = 253$). Figure A.1 shows the results of the bare cover regression. A regression of imputed bare cover compared to field data cover is shown in Figure A.2. The R-squared value of this regression was 0.80, which is similar to that of biocrust (0.74) showing that the model was able to identify areas of bare cover reasonably reliably.
Figure A.1: Imputed Bare Cover in the Morley Nelson Snake River Birds of Prey, 2016

Figure A.2: Predicted vs. Observed Bare Cover
Qualitatively, this output is reasonable: areas of known bare cover (e.g. military training areas within the OCTC that are regularly scraped to remove vegetation, dirt roads) contain pixels with the largest percent cover bare. Conversely, areas known to have little to no proportion of bare soil appear as such (e.g. dense annual grass communities, areas of undisturbed sagebrush). Areas of water notably have the same calculated cover (∼ 40%) and although incorrect, do show that the model is stable.

A direct comparison of bare cover and biocrust cover was created as additional validation by subtracting the calculated bare cover from the calculated biocrust cover. Figure A.3 shows this for the entire study area and a subset (∼ 3 km by 3 km, -116.25316, 43.34776) to elucidate the difference between majority bare areas and majority biocrust areas. Note that roads are classified as heavily bare-dominant, with more biocrust farther from roads. This could be due to roads being known conduits for invasion of exotic plants [Gelbard and Belnap, 2003], combined with the known negative relationship between exotic plants and biocrust cover (e.g. Bansal & Sheley 2016; Serpe et al. 2013).
Figure A.3:  Bare Cover (red) vs. Biocrust Cover (blue) in the Morley Nelson Snake River Birds of Prey, 2016