

*Correlating Remotely Sensed Vegetative Indices Against On-the-ground
Rangeland Monitoring Data*

Dylan Cicero
Yale School of Forestry and Environmental Studies
Master of Environmental Management, 2019
December 15, 2018

Background and Objectives

My partner for this project was *Grasslands, LLC*- an organization based out of Bozeman, Montana whose mission is “the skillful stewardship of grasslands, via Holistic Planned Grazing, toward ecological regeneration, solar based profitability, and thriving rural communities.” *Grasslands, LLC* manages over 200,000 acres of rangeland across five ranches in the United States and New Zealand. Their approach uses pioneering science in rangeland ecology, including rigorous on-the-ground data collection for site monitoring. Ecological characteristics that are monitored intensively across their landholdings include land cover type, plant diversity, soil carbon, forage productivity, and others (*Grasslands, LLC*).

For this study, I used ecological monitoring data from one *Grasslands, LLC* property (not named due to confidentiality agreements). My objectives were (1) to determine the relationship between on-the-ground measures of total forage biomass and remotely sensed vegetative indices and (2) to consider steps for future work.

Literature Review

The literature assessing remotely sensed vegetation indices for rangeland monitoring applications is extensive. Studies have assessed such issues including land cover, net primary productivity, land degradation, fire-burn severity, fuel moisture load, and others, typically using Landsat, MODIS, or SPOT sensors (Reeves et al. 2015). Studies that specifically assess the relationships of vegetative indices and vegetative biomass have also been conducted (Rouse et al. 1974; Todd et al. 1998; Kawamura et al. 2005; Yu et al. 2011; Al-Bukhari et al. 2018).

Dozens of vegetation indices have been used in these studies, each conferring a slightly different set of advantages and limitations. The normalized difference vegetation image (NDVI) is probably most well-known, though as several authors have pointed out, NDVI is sensitive to the effects of soil brightness, atmosphere, soil color, cloud and cloud shadow, and leaf canopy

shadow (Xue & Su 2017). Indices like the soil adjusted vegetation index (SAVI) and the modified secondary soil adjusted vegetation index (MSAVI2) have been developed to better withstand the effects of underlying soil variance (Xue & Sue 2017). All three of these indices, however, and many others, are only sensitive to green vegetation. A separate group of vegetation indices, including the normalized difference senescent vegetation index (NDSVI) and the dead fuel index, are sensitive to browned or “senescent” vegetation (Li & Guo 2015). These indices are particularly relevant in arid or semi-arid rangelands where vegetation lies dormant in a browned state for long periods of time awaiting moisture.

In 2006, Marsett et al. proposed the soil adjusted total vegetation index (SATVI), specifically for arid rangeland monitoring, which captures both green and senesced vegetation in a single index. In that paper, the authors assessed the validity of SATVI for predicting total vegetative cover by comparing SATVI values (derived from the Landsat TM sensor) at a given sample site and a ground-based assessment of vegetative cover for the same sample site. The authors found that SATVI explained 80% of the variation in their ground-based vegetative cover assessment. The authors did not correlate SATVI on vegetative biomass (Marsett et al. 2006).

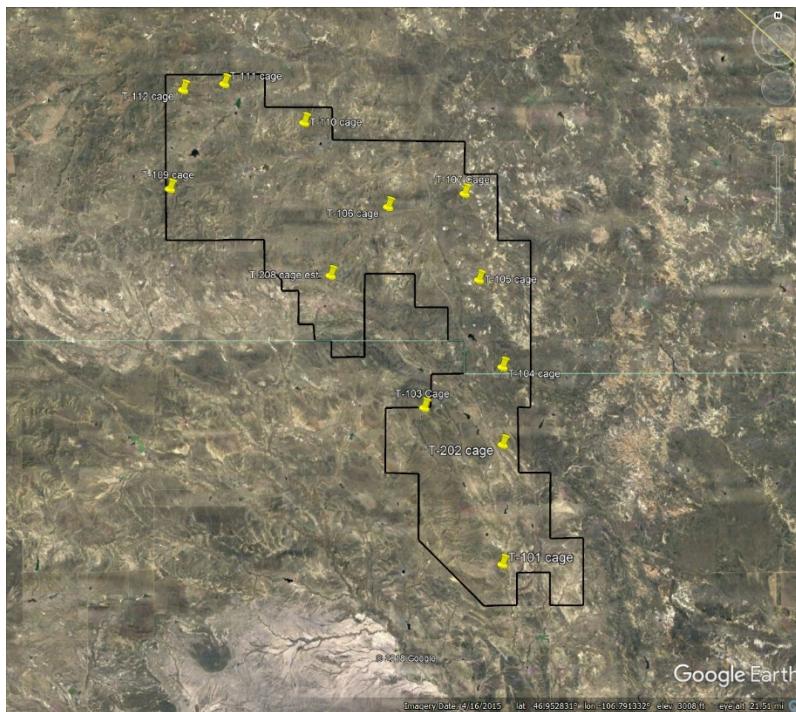
This research seeks to contribute to the above scholarship. It is distinct for a few reasons. First, as far as I am aware, this is the first correlational analysis in the literature that uses Sentinel-2 satellites for remote sensing data acquisition. The satellites in this constellation (Sentinel-2A and Sentinel-2B) are cutting edge; fresh into orbit, they were launched in June 2015 and March 2017 respectively and are endowed with *R*, *G*, *B*, and *NIR* bands at 10m resolution, and *red edge* and *SWIR* bands at 20m resolution (outperforming Landsat’s 30m spatial resolution). Additionally, the return time of the satellites (taken together) is 5-days, surpassing Landsat’s 16-day mark. Together, these attributes allow remote sensing data to be captured closer in space to ground-sampling points and closer in time to ground-sampling dates. Second,

as far as I am aware, this analysis is the first to correlate SATVI with vegetative biomass (in addition to other indices).

Methodology

A .kml file of the ranch (~ 50,000 acres, SE Montana) was given to me by *Grasslands, LLC*, including a polygon of the ranch periphery and precise markers for 12 “livestock-exclusion” cages¹ (Image 1). Additionally, average vegetative biomass ground-measurements were given for each sample site for each year since 2013, when the ground-sampling protocol began. Lastly, sampling dates for each cage site for each year were given (Image 2).

Image 1



Google Earth .kml layer, showing ranch border and cage sites

Image 2

<input checked="" type="checkbox"/> T-101 Stud	7: WPGC
<input checked="" type="checkbox"/> T-103 Ringsveit West	6: WPGC
<input checked="" type="checkbox"/> T-104 Ringsveit East - N...	7: WPGC
<input checked="" type="checkbox"/> T-105 Cottonwood	7: WPGC
<input checked="" type="checkbox"/> 2013-09-19 : W P G C	
<input checked="" type="checkbox"/> 2014-10-16 : W P G C	
<input checked="" type="checkbox"/> 2015-10-15 : W P G C	
<input checked="" type="checkbox"/> 2016-07-10 : W P G C	
<input checked="" type="checkbox"/> 2016-11-08 : W P G C	
<input checked="" type="checkbox"/> 2017-09-20 : W P G C	
<input checked="" type="checkbox"/> 2018-08-01 : W P G C	
<input checked="" type="checkbox"/> T-106 Pumphouse	7: WPGC
<input checked="" type="checkbox"/> T-107 North High Hill	7: WPGC
<input checked="" type="checkbox"/> T-109 Fawn	7: WPGC
<input checked="" type="checkbox"/> T-110 Johnson	7: WPGC
<input checked="" type="checkbox"/> T-111 Horsecreek	7: WPGC
<input checked="" type="checkbox"/> T-112 Jasper	7: WPGC

Sampling dates for each cage site for each year were given

¹ “Livestock exclusion” cages enclosed small experimental areas of the ranch. These areas are segmented off from livestock and are used to assess grazing management decisions against a control.

For each year for each cage site, vegetative biomass samples were collected according to the following specifications: samples were to be taken at least 10 meters from cage centroid², samples were to be taken within 30 meters from cage centroid (Image 3), and at least two samples were to be taken for each cage site. Sampling consisted of (1) randomly selecting a point within the sampling zone ($10m < \text{distance_from_centroid} < 30m$), (2) laying down a 4.8ft^2 hoop at that point (Image 4), (3) clipping all biomass inside the hoop, (4) weighing the raw biomass, (5) applying a dry weight multiplier, (6) and applying a conversion factor to get the approximate forage lbs/acre. Samples are then averaged across each cage site (Image 5).

Image 3

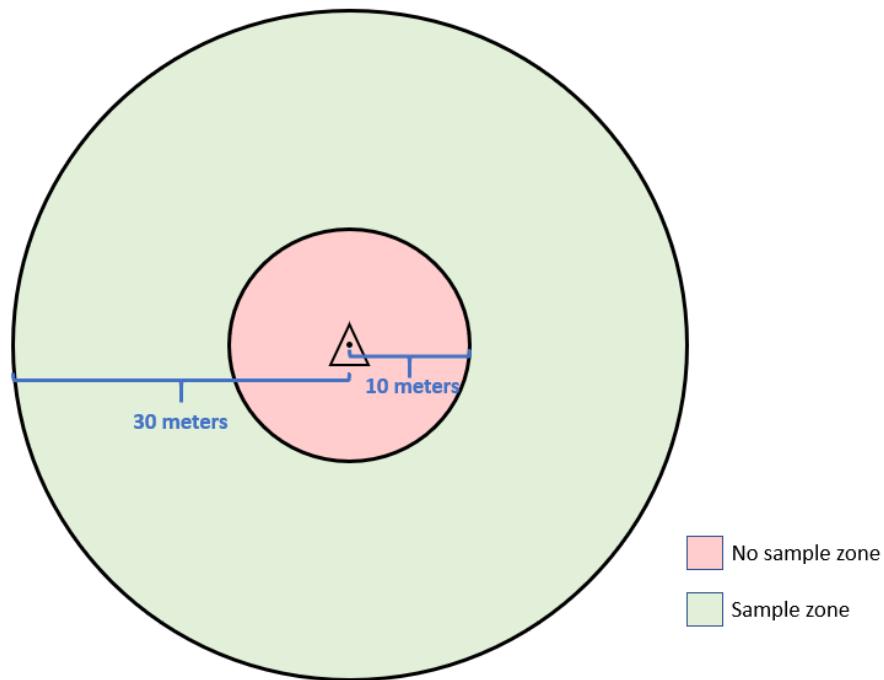


Diagram of sampling zone specified in protocol. For each cage, samples are to be taken at least 10 meters from cage centroid, but not greater than 30 meters from cage centroid

² Cattle tend to cluster immediately around the grazing cages (they like to use the cages to scratch themselves). Cattle clustered in such a manner has a heavy impact on the land. Thus, vegetative biomass samples taken within 10 meters of cage centroid would not be representative of the site as a whole.

Image 4



Examples of hoops laid down in the sampling zone. Vegetative biomass within each hoop is clipped and weighed to estimate vegetative biomass. Notice also, the mix of green and senescent vegetation apparent at these sites.

Image 5

Clip	Pasture Production (all weights in grams)	N/A	Clipped	Estimated:
# 1: (Raw Wt:	63	- Bag Wt: 0) X %Dry: 95 = Production: 1197 Lbs/Acre	% of Pasture:	
# 2: (Raw Wt:	90	- Bag Wt: 0) X %Dry: 95 = Production: 1710 Lbs/Acre	% of Pasture:	
# 3: (Raw Wt:		- Bag Wt:) X %Dry: = Production: Lbs/Acre	% of Pasture:	
# 4: (Raw Wt:		- Bag Wt:) X %Dry: = Production: Lbs/Acre	% of Pasture:	
# 5: (Raw Wt:		- Bag Wt:) X %Dry: = Production: Lbs/Acre	% of Pasture:	
Average Pasture Production: 1454 Lbs/Acre Clip(s): #1, #2 Total % (should be 0 or 100): 0				

Vegetation biomass samples were averaged across each cage site for each year.

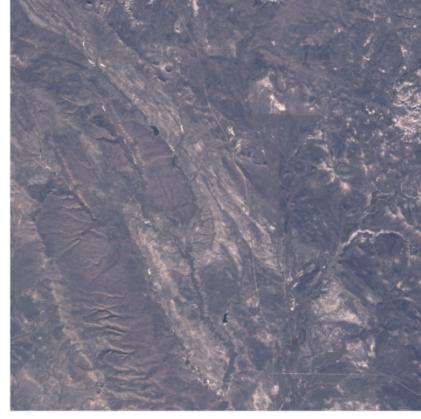
Satellite data processing was performed in Google Earth Engine. All Sentinel-2 images, which (1) encompassed the ranch and (2) fell within 7 days of an on-the-ground sampling date were identified (Appendix 1). Images that met the above criteria were atmospherically corrected³ using a python-based Sentinel-2 atmospheric correction script (in Jupyter Notebooks),

³ While atmospherically corrected data products are available for most satellites directly from the Google Earth Engine catalog, this is not yet the case for Sentinel-2

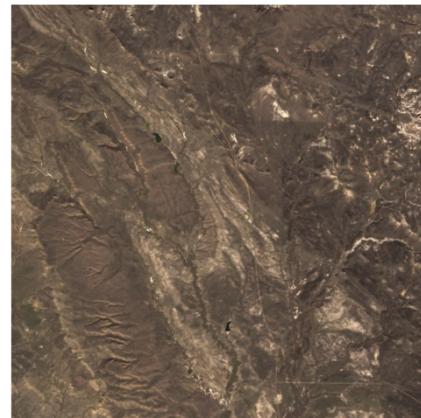
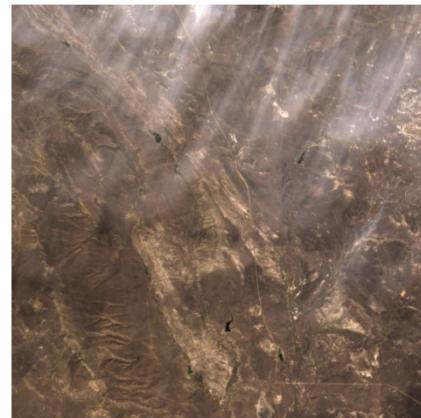
which was initially developed by GitHub user Sam Murphy, and which I further amended to correct for minor errors (Appendix 2). Example images before and after atmospheric correction are shown below (Image 6). All atmospherically corrected images were brought back into the Earth Engine JavaScript console for further processing.

Image 6

Before



After



Sample Sentinel-2 images before and after atmospheric correction

First, (1) all images were clipped to the silhouette of the ranch and masked for clouds and cloud shadows. Second, (2) point markers and a sampling zone buffer ring (like that represented in Image 3) were generated for each of the 12 cage-sites. Third, (3) five vegetation indices (Table 1) were calculated for every pixel in each image.

Table 1

<i>Normalized Difference Vegetation Index:</i>	$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$
<i>Soil Adjusted Vegetation Index:</i>	$SAVI = \frac{NIR - RED}{(NIR + RED + L)} * (1 + L)$
<i>Modified Secondary Soil Adjusted Vegetation Index:</i>	$MSAVI2 = 0.5 * \left[(2NIR + 1) - \sqrt{(2NIR + 1)^2 - 8(NIR - R)} \right]$
<i>Normalized Difference Senescent Vegetation Index:</i>	$NDSVI = \frac{\rho_{band5} - \rho_{band3}}{\rho_{band5} + \rho_{band3}}$
<i>Soil Adjusted Total Vegetation Index:</i>	$SATVI = \frac{\rho_{band5} - \rho_{band3}}{\rho_{band5} + \rho_{band3} + L} (1 + L) - \frac{\rho_{band7}}{2}$

Vegetation indices calculated for analysis

Fourth, (4) all images within a 7-day range of a given ground-based sample date were mosaicked together. The 7-day range is the maximum threshold that I've set for images to be considered in the index statistics around a given sample date. I will refer to this range as the satellite sampling period. The mosaicking procedure was specified such that for all images that fell within a given satellite sampling period, the image nearest in time to the ground-based sampling date was stacked on top, with the second closest image (if there was one) stacked below, and the third (if there was one) still below. Pixels in the second layer of a satellite sampling period stack would only show up in the mosaic if the first image had a “hole” in it (a “NoData” area generated by the cloud or cloud-shadow masks). Likewise, pixels in the third layer of a satellite sampling period stack would only show up in the mosaic the first and second images both had “holes” in the same spot. This mosaicking step also served the purpose of creating a spatially continuous single image for each satellite sampling period (the Sentinel-2

orbit pattern segmented the ranch into different tiles for a given image date). Image 7 displays this workflow. The result was a set of ten image mosaics- one for each ground-sampling date.

Image 7

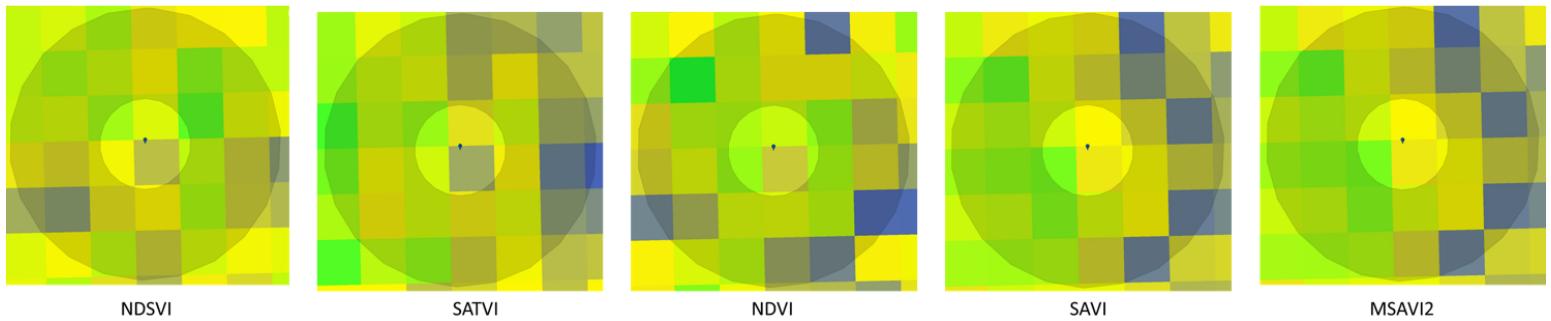


For ground sample date July 29, 2018, Sentinel-2 images for July 27, 2018 and August 02, 2018 fell within the 7-day satellite sampling period. Two images existed for each satellite image date, capturing different parts of the Ranch. The mosaicking procedure stacked the July 27 images on top of the August 02 images because July 27 is closer to the July 29 sampling date. Holes in the July 27 image are filled by the August 02 image in the final mosaic. Note also the imperfections of the cloud-masking procedure, which failed to mask a sizeable portion of clouds from the July 27 image.

mosaics times 12 ground-sample-cage-sites) were exported from Google Earth Engine to a .csv file (Appendix 3). Data was cleaned so that there was a single cage-site observation for each year of the ground-sampling protocol (where Sentinel images were available). Observations missing index values (due to cloud masking) were eliminated. A total of 31 observations were left in the final dataset. Ground-based forage biomass measurements for a give cage-site-year observation were manually added to the dataset. Straight line regressions were conducted in R to assess the correlation between remotely sensed vegetation indices and ground-based data.

Image 8

Ground-Sample Cage Site: T-103; Ground-Sample Date: October 21, 2015



Pixel-by-pixel vegetation indices shown for ground-sample cage site T-103 based on ground-sample date October 21, 2015. Greener colors indicate higher index values. Bluer colors indicate lower index values. Index values for all sample date mosaics were averaged across each ground-sample cage site buffer zone. Note similarities and discrepancies between indices.

Results

Image 9 displays scatterplots of ground-based data against the five remotely sensed indices. Linear trendlines are included along with R^2 values, which range from .007 to .087.

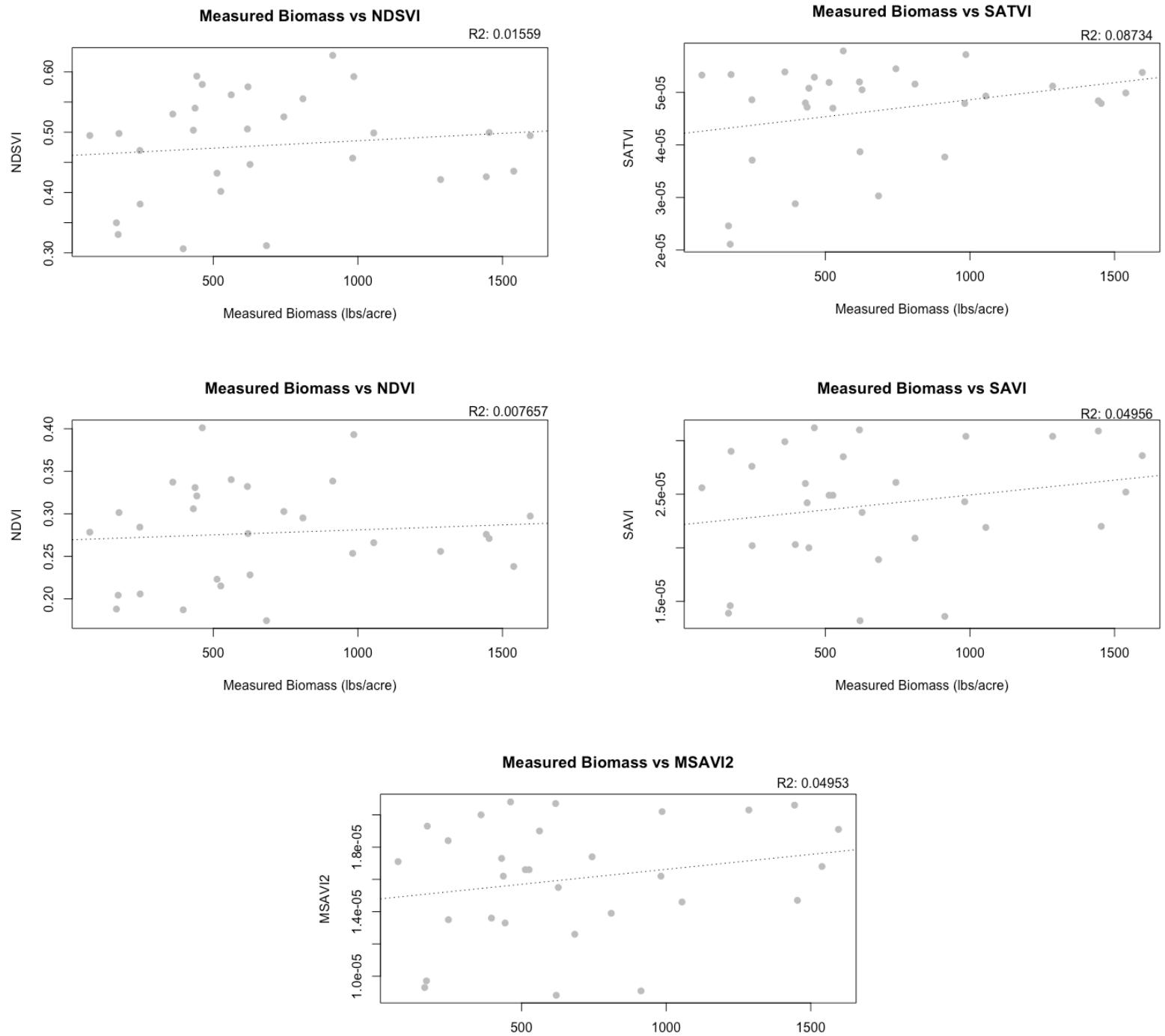
Discussion

R^2 values of .007 to .087 indicate non-existent (or at best extremely weak) correlations between ground-based measures of biomass and remotely sensed vegetative indices. This result was unexpected. A few factors are likely to blame.

First, revisiting Image 7, we see that the Sentinel-2 cloud-masking band was largely insufficient. In such a small dataset, if vegetation indices are computed over cloudy pixels for even a few observations, this would significantly skew the correlation results. Future work should visually inspect all mosaicked images for the presence of clouds over cage-sites and exclude any invalid samples.

Second, it is probable that for a given cage-site there were too few on-the-ground measurements to adequately reflect the variability of forage biomass across the entire sample zone. For instance, in their study, Marsett et al. calculated that for their 13,500 m^2 study site, using sampling quadrats of area 1 m^2 , they would have needed to take 77 biomass clippings to achieve a confidence interval of 90% for their estimate of average site biomass (plus or minus 10% of the true mean) (Marsett et al. 2006). This equates to about 1 sample every 175 m^2 . Contrast this with sampling rates at the *Grasslands, LLC* ranch, where the sampling zone for each cage-site was approximately 2,513 m^2 ($\pi \times 30^2 - \pi \times 10^2$), and only 2-3 samples were taken, resulting in a sampling rate of about 1 sample every 1005 m^2 .

Image 9



Scatterplots depicting the relationship between on-the-ground measurements of forage biomass and remotely sensed vegetation indices. R² values are also given.

Of course, the calculated confidence interval is site-specific. Maret et al.'s study looked at rangeland in the arid southwest, which may or may not be less homogenous than rangeland in southwestern Montana. Still these calculations suggest that sampling rates for each cage site at the *Grasslands, LLC* ranch would not accurately reflect the site as a whole. Future work might attempt to circumvent this limitation. For example, consider an analytic technique that relates the slopes in the trendline over time for on-the-ground measures of biomass against the slopes in the trendlines over time for remotely sensed vegetative indices. Assuming that both on-the-ground measurements and remotely sensed indices are equally valid assessments of land-change over time, their relationship can be evaluated in this manner.

Conclusion

This study attempted to correlate remote sensing indices from Sentinel-2 data to on-the-ground forage biomass measurements for a ranch in southeastern Montana. Relationships were generally non-existent or weak, though several factors suggest that the results are (as of yet) invalid. Improvements to the methodology were suggested for future work.

Work Cited

Al-Bukhari, A., Hallett, S., & Brewer, T. (2018). A review of potential methods for monitoring rangeland degradation in Libya. *Pastoralism*, 8(1). doi:10.1186/s13570-018-0118-4

Grasslands LLC. (n.d.). Retrieved from <https://www.grasslands-lcc.com/>

Jinru Xue and Baofeng Su, "Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications," *Journal of Sensors*, vol. 2017, Article ID 1353691, 17 pages, 2017. <https://doi.org/10.1155/2017/1353691>.

Kawamura, K., T. Akiyama, H. Yokota, and M. Tsutsumi. 2005. Monitoring of forage conditions with MODIS imagery in the Xilingol steppe, Inner Mongolia. *International Journal of Remote Sensing*, 26, 1423–1436

Li, Z., & Guo, X. (2015). Remote sensing of terrestrial non-photosynthetic vegetation using hyperspectral, multispectral, SAR, and LiDAR data. *Progress in Physical Geography*, 40(2), 276-304. doi:10.1177/0309133315582005

Marsett, R., Qi, J., Heilman, P., Biedenbender, S., Watson, C., Amer, S., . . . Marsett, R. (2006). Remote Sensing for Grassland Management in the Arid Southwest. *Rangelands*, 59(5). doi:10.2458/azu_jrm_v59i5_marsett

Murphy, S. (n.d.). Sentinel2_atmospheric_correction.ipynb. Retrieved from https://github.com/samsammurphy/gee-atmcorr-S2/blob/master/jupyter_notebooks/sentinel2_atmospheric_correction.ipynb

Reeves, M., R.A. Washington-Allen, J. Angerer, E.R. Hunt, R.W. Kulawardhana, L. Kumar, T. Loboda, T. Loveland, G. Metternicht, and R.D. Ramsey. 2015. Global view of remote sensing of rangelands: Evolution, applications, future pathways [chapter 10]. In *Land resources monitoring, modeling, and mapping with remote sensing*, 237–276. Boca Raton, FL: CRC Press/Taylor and Francis Group

Rouse Jr JW, Haas RH, Deering DW, Schell JA, Harlan JC. 1974. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. In: Type III, final report. Greenbelt: NASA/GSFC.

Todd, S. W., Hoffer, R. M., & Milchunas, D. G. (1998). Biomass estimation on grazed and ungrazed rangelands using spectral indices. *International Journal of Remote Sensing*, 19(3), 427-438. doi:10.1080/014311698216071

Yu, L., L. Zhou, W. Liu, and H. K. Zhou. 2011. Using remote sensing and GIS technologies to estimate grass yield and livestock carrying capacity of alpine grasslands in Golog Prefecture China. *Pedosphere*, 17, 419–424