**Lab Report**

Title: *Using RGB Indices to Predict Ground Cover*

Notice: Dr. Bryan Runck, Dr. Patrick Ewing

Author: Lucas Rosen

Date: 5/25/2022

**Project Repository:** *https://github.com/LRosen656/USDA\_GEMS\_RGB\_COVER*

**Abstract**

Cover crops have many benefits to agriculture. A study was done to collect red-green-blue images over 7 dates to show changes in clover cover crop over time and different plant method. Images were cropped to the center and moved to avoid any erroneous data. Four different indices were evaluated: Excess green, Excess Green minus Red, Green Leaf Index, and Visible Atmospherically Resistant Index. Otsu’s threshold both locally and globally as well as a zero threshold were evaluated with each index to classify the image as vegetation or non-vegetation. Ground cover percent was then determined. Results showed that Excess Green minus Red with zero threshold was the most accurate.

# Problem Statement

Cover crops have many benefits to agriculture. They reduce soil erosion, improve weed control, and increase nutrient cycling (Snapp et al., 2005). Ground cover percentage can summarize cover crop growth over a given area. Because ground cover percentage correlates with multiple biophysical factors affecting productivity including light interception, remote sensing can automate our understanding of plant growth and development.

In 2021, a visible imagery (RGB) was collected from the USDA-ARS Research station. RGB Indices such as excess green have been used to classify between vegetation (living plant) and non-vegetation (soil or non-living residue). The goal of this study is to (1) compare the accuracy if different RGB-indices and thresholds to classify vegetation vs non-vegetation and (2) use classified images to find correlations between ground cover and common in field metrics. **Table 1** shows the requirements for this project.

**Table 1: Project Requirements**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **#** | **Requirement** | **Defined As** | **Spatial Data** | **Attribute Data** | **Dataset** | **Preparation** |
| 1 | Plot images | Raw jpeg dataset from USDA | Plot images | NA | USDA-ARS Cart Imagery | Rename images to Plot Map. Crop images to center. Convert from jpeg to tiff. |
| 2 | Normalized Band (RBand(N)) | RBand/(RRed +RGreen + RBule) | Plot images | NA | USDA-ARS Cart Imagery (Prepared) | NA |
| 3 | Excess Green Index (ExG) | 2RGreen(N) -RRed(N) - RBlue(N) | Plot Images | NA | USDA-ARS Cart Imagery (Prepared) | NA |
| 4 | Excess Green minus Excess Red Index (ExGR) | 3RGreen(N) -2.4RRed(N) - RBlue(N) | Plot Images | NA | USDA-ARS Cart Imagery (Prepared) | NA |
| 5 | Green Leaf Index (GLI) | (2RGreen ­­– R­Red  - R­­Blue ) / (2RGreen + RRed ­+ R­Blue) | Plot Images | NA | USDA-ARS Cart Imagery (Prepared) | NA |
| 6 | Visible Atmospherically Resistant Index | (RGreen ­­– R­Red ) / (RGreen + RRed ­- R­Blue) | Plot Images | NA | USDA-ARS Cart Imagery (Prepared) | NA |
| 7 | Thresholds | Local Otsu’s, Global Otsu’s, Zero | Vegetation Indices | NA | Vegetation Indices | NA |
| 8 | Image Classification | Value < Thresh = 0  (Non-vegetation)  Value > Thresh = 1 (Vegetation) | Vegetation Indices | NA | Vegetation Indices | NA |
| 9 | Vegetation Cover | Classified Vegetation/ Total Area | Classified Images | NA | Classified Images | NA |
| 10 | Mean Vegetation Greenness | RGreen(N) Vegetation/Total Vegetation | Plot and Classified Images | NA | USDA-ARS Cart Imagery (Prepared),  Classified Images | Mask plot images to classified vegetation |

# Input Data

Red-Green-Blue (RGB) JPEG imagery of 120 plots over 7 dates were collected from oat-study done by Ewing Laboratory over the 2021 season at a USDA-ARS Research Station north of Brookings, South Dakota. Images were collected by a Canon PowerShot ELPH 190 IS approximately 8’ in the center of each plot using a cart. Images were uploaded to a Google drive and downloaded to an external hard drive. The dates of collection were: July 8th, August 21st, September 9th, September 29th, October 5th, October 15th, and October 25th.

Different oat/soy and clover cover crop treatments were used in the plots. Because only one image date (7/8/2021) was collected before the oat crop `harvest (8/6/2021), this study will focus on the different red clover (*Trifolium pratense???*) planting treatments. “Under-Seeded” was clover planted (more details needed) during the growing season (date needed), “Post-Harvest” was clover planted after the oat harvest (8/12/2021). Finally, “Fallow” had no cover crop and was not weeded. **Figure 1** shows the layout of the plots and treatments.



Figure : Layout of Plot Treatments



**Table 2: Input Data**

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Title** | **Purpose in Analysis** | **Link to Source** |
| 1 | Plot Images | Classifying Vegetation vs Ground over time | [Google Drive](https://drive.google.com/drive/folders/1_XWsaLWfEwpgPoJYNq7ora43J4t4vV2b?usp=sharing) |

# Methods

**Figure 2** shows a flowchart of the process. The plot images were downloaded from a shared google drive and stored on a local hard drive. Then, the images were renamed to their corresponding plot id using Pandas. Once renamed, the images were cropped to the center and converted from a jpeg to a tiff using OpenCV. Vegetation indices were done (using Scikit Image) on each image and included: excess green (ExG), excess green minus excess red (ExGR), green leaf index (GLI) and visible atmospherically resistant index (VARI).

Diagram

Description automatically generated

Figure : Data Flow Diagram

Thresholds were used to distinguish vegetation from non-vegetation. In this context, non-vegetation is defined as anything that is not living vegetation (e.g., soil and plant residue). Otsu’s Threshold is an unsupervised method that minimizes variance with a tested threshold (**Equation 1**) (Otsu, 1979).

**(1)**

Where within class variance () at a given threshold level (k) is determined by the sum of within class variance below the threshold () and the sum of within class variance above the threshold (). ω­0 is a weight determined by the distance from the threshold to the beginning of the histogram and ω­1 is a weight determined by the distance from the threshold to the end of the histogram. The ideal threshold is where within class variance is the smallest.

ExG often uses Otsu’s method to determine a threshold ((Li et al., 2019; Meyer & Neto, 2008). A local Otsu’s threshold was done on each image and a global threshold was done by using Otsu’s method on all images per index. ExGR and VARI often have a threshold of zero (Meyer & Neto, 2008; Wang et al., 2022). Therefore, the three thresholds that were calculated for each plot index were local Otsu’s-threshold, global Otsu’s-threshold, and a zero-threshold. Finally, the images were reclassified using each threshold and a ground-truth was done for an accuracy assessment.

## Preprocessing: Renaming and Cropping images

Except for one date, the images on the google drive were given arbitrary names assigned by the camera. For meaningful data, the first step was to copy and rename the images. The order of the images recorded on each date was collected in a csv spreadsheet. So, the images were copied and relabeled using the “shutil” function. The output was to a new directory with the correct name and jpeg as the file format. July 8th was just copied to that directory as the images already had the correct names.

Now that the images are correctly labeled, the next step was to crop them to the center. This was done for two reasons: (1) to avoid errors near the edge of the image and (2) to remove any erroneous objects including the cart, labels, and flags. The center ninth of the image was used for the data processing. To do this, OpenCV was used to square and divide the image height and width into thirds and resize the image to its center. Once done, we noticed that there was an object from the cart in the center of each image. So, the center was offset slightly to avoid the object. For most of the images, shifting the center 500 pixels up and 500 pixels to the right resulted with an image closest to the center without any objects. The output was saved to a new directory and converted to a tiff. The output also had the plot number and date associated with the image (i.e., plot\_yyyymmdd.tif). One final note is that some of the original images had labels to orient the picture order. These images were skipped. **Figure 3** shows an image crop example.

Figure Crop example.



## Vegetation Indices

Vegetation indices are used to enhance vegetation against non-vegetation. It also summarizes the color bands into a single band which is necessary for thresholding. Our dataset was limited to the visible bands (red, green, and blue) and therefore could only use indices that were restricted to those colors. The selected indices were excess green (ExG), excess green minus excess red (ExGR), green leaf index (GLI) and visible atmospherically resistant index (VARI).

### Getting the Band Arrays

Each plot image has the radiance (R) of each band between 0 and 255 (8-bit). To get to the band arrays, Scikit Image was used to extract each band matrix. When using this package, red is the first band, green is the second band and blue is the third band (it varies with different packages). Once each band was extracted, the datatype was converted from bytes to floats to be compatible with base 10 math. For ExG and ExGR, the bands also had to be normalized. This was done by dividing each band by the sum of all bands (**Equation 2**)(Meyer & Neto, 2008). The result shows each band value between 0 and 1. A note that any pixel where all values were 0 (i.e., 0/0) the output was redefined as zero for all indices.

**(2)**

### ExG and ExGR

Perhaps the most direct indices are ExG and ExGR. Once the bands are normalized, ExG takes the 2 times green band and subtracts the red and blue band. ExGR takes three times the green band, subtracts 2.4 of the red band followed by the blue band. **Equation 3 and 4** show ExG and ExGR respectively(Li et al., 2019; Meyer & Neto, 2008).

**(3)**

**(4)**

### GLI

GLI is similar to ExG, but it also normalizes so the bands to have the index values between -1 and 1. **Equation 5** Shows GLI (Louhaichi et al., 2001).

**(5)**

### VARI

The final index used to analyze the images was VARI. VARI assumes that blue band is atmospheric noise and reduces its effect in the output index (Gitelson, Kaufman, et al., 2002) It takes the green band and subtracts the red band and divides that by green plus red minus blue (**Equation 6**).

**(6)**

## Finding Thresholds

Scikit Image was used to find all three threshold methods for each image index. Ostu’s threshold on each image resulted in a unique (local) threshold. To find a global threshold for each index, Numpy was used to append each image per index and flatten it to a single array. Otsu’s threshold was then taken for that array. The zero threshold was manually put in for each index. Anything greater than the threshold was reclassified as 1 (vegetation) and anything less than the threshold was reclassified as 0 (non-vegetation)

## Ground Truthing

Accuracy assessment was done by using Scikit Image and Matplotlib to ground truth each image. Eight images were randomly selected from each date. Once Scikit uploaded the image, Numpy and Matplotlib were used to plot 30 random pixels. The reference value (0 or 1) was found visually by the user input. **Figure 4** shows an example of ground truthing an image. A note that subjectivity between vegetation and non-vegetation may have led to errors.

A picture containing text

Description automatically generatedA close-up of some plants

Description automatically generated with low confidence

Figure : Ground truth example. The red x shows the sampled pixel (non-vegetation).

## Finding Percent Ground Cover

Percent ground cover was defined as the total area of plants divided by the total area. To determine this, the most accurate image classification index and threshold (ExGR-Zero). Because vegetation was classified as 1, Scikit Image was used to take the sum of each classified image (i.e., total number of pixels classified as vegetation) and divide that by the total number of pixels. The decimal result is the vegetation fraction, and the percent vegetation multiplies that by 100.

## Correlating Clover Harvest Metrics

ExGR-Zero classified image was also used to determine the mean greenness of vegetation. Scikit Image was used to take the normalized green of each image. The plots were then masked to the vegetation by multiplying the image by the classified vegetation. This turned all non-vegetation values to zero. Finally, the sum of the masked matrix was divided by the area of vegetation to get a mean value.

Clover metrics from post-seeded and under-seeded treatments were measured at the end of the season and compared to the final date of image collection (10/25/2021). The metrics chosen for correlation were biomass (kg/ha) to vegetation fraction, nitrogen (percent by mass) to average vegetation greenness, and carbon-nitrogen ratio to average vegetation greenness.

# Results

**Figure 5** shows an example of each classified threshold.

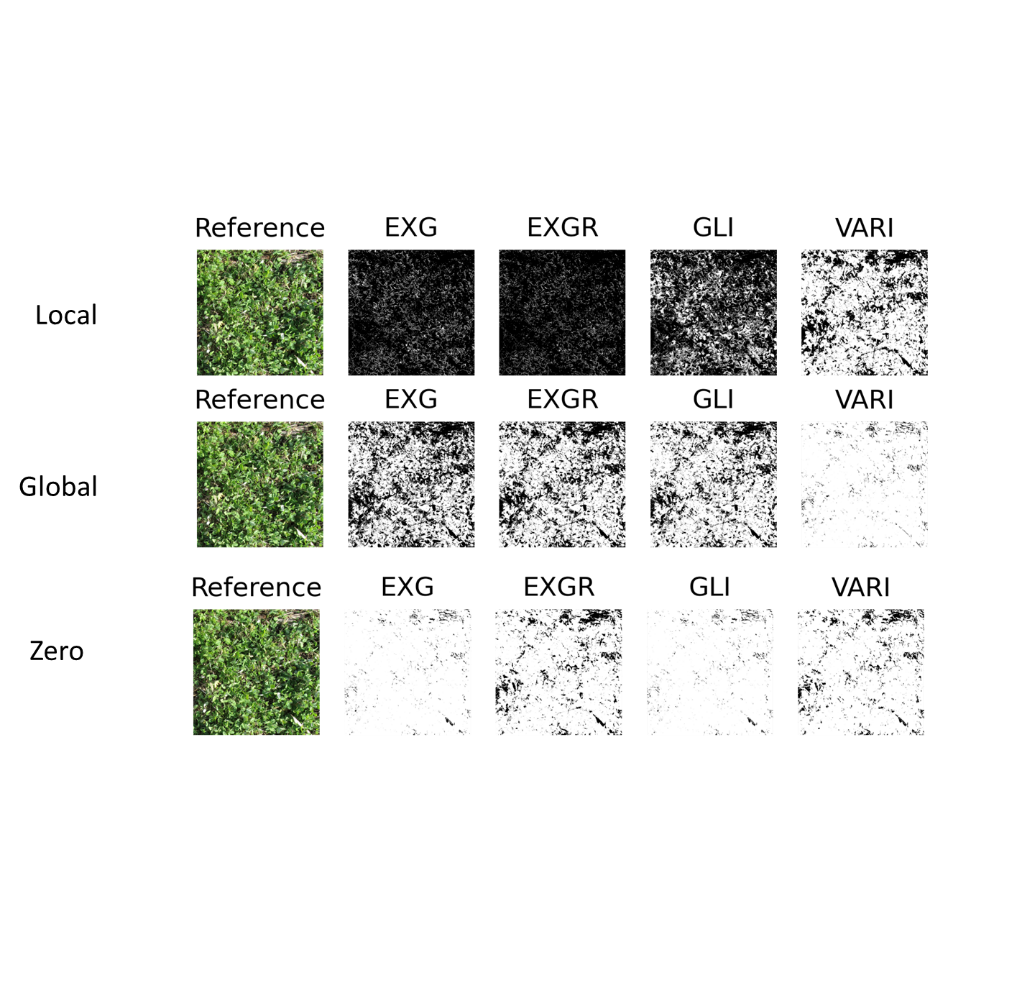


Figure : Reference and classified image example.

## Threshold Values

Using Otsu’s Threshold locally across all indices, values tended to be negative areas where vegetation cover was low (early dates/ fallow) and positive when vegetation cover was high (later dates/ under). Otsu’s Threshold globally varied by the vegetation index: ExG had a threshold of ~0.30, ExGR had a threshold of ~0.20, GLI had a threshold of ~0.16, and VARI had a threshold of ~-0.26.

## Threshold Accuracy

**Table 3** shows the overall accuracy of each threshold method. Using Otsu’s Method locally for each image had the lowest classification for each vegetation index with an overall accuracy of 51% or less. This was likely due to over classification of vegetation in the early dates, and under classification of vegetation in later dates. Otsu’s Method globally did significantly better with an overall with an average overall accuracy over 78%. The most common error with global threshold were under classifying vegetation. The best overall accuracy was using a threshold of zero on the ExGR with an accuracy nearly 85%. VARI was also acceptable with an accuracy around 79%. Both ExG and GLI had identical low accuracies at a zero-threshold due to overclassifying vegetation.

**Table 3: Index and Threshold Accuracies.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ExG** | Overall Accuracy | Precision | Recall | Issue |
| Otsu's Method Global | 78.93% | 83.21% | 46.64% | Underestimated Vegetation |
| Zero | 55.54% | 42.83% | 98.40% | Overestimated Vegetation |
| Otsu's Method Local | 52.26% | 34.61% | 48.04% | Inconsistent |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ExGR** | Overall Accuracy | Precision | Recall | Issue |
| Zero | 84.94% | 74.87% | 82.74% | Overestimated Vegetation |
| Otsu's Method Global | 81.25% | 83.11% | 46.44% | Underestimated Vegetation |
| Otsu's Method Local | 43.99% | 31.41% | 56.93% | Inconsistent |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **GLI** | Overall Accuracy | Precision | Recall | Issue |
| Otsu's Method Global | 81.96% | 80.19% | 61.21% | Underestimated Vegetation |
| Zero | 55.54% | 42.83% | 98.40% | Overestimated Vegetation |
| Otsu's Method Local | 45.77% | 33.74% | 64.41% | Inconsistent |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **VARI** | Overall Accuracy | Precision | Recall | Issue |
| Zero | 79.11% | 63.54% | 88.08% | Overestimated Vegetation |
| Otsu's Method Global | 57.92% | 44.05% | 95.52% | Overestimated Vegetation |
| Otsu's Method Local | 50.42% | 39.22% | 87.72% | Overestimated Vegetation |

## Percent Vegetation Cover

**Figure 6** shows the average percent of vegetation cover per date found through ExGR-Zero thresh method. Percent vegetation dropped from July 8th to August 21st due to crop harvest on August 6th. Clover planted under seed went up sharply until it reached 80% and then stabilized. Post planted clover went up steadily to reach a maximum average cover close to 50%. Areas left fallow consistently stayed under 20% and fluctuated.

Chart, line chart

Description automatically generated

Figure : Percent Vegetation Cover

## Clover Harvest Metrics

Measured results showed that under-seeded clover overall had a higher biomass than post-harvest clover, post-harvest clover had higher percent nitrogen than under-seeded clover, and under-seeded clover had a higher carbon-nitrogen ratio post-harvested clover. Predicted data from 10/25/2021 showed that under-seeded clover had higher vegetation fraction and mean greenness value compared to post-harvested clover.

Correlating results of the harvested metrics showed weak but not significant trends to the predicted data; the R-Squared value was less than 0.60 across all metrics. **Figure 7** shows measured biomass versus the calculated vegetation fraction from the classified ExGR-Zero Threshold. Post-harvest vegetation showed little variation in biomass regardless of vegetation cover. Under-seeded vegetation showed a weakly positive relationship between biomass and vegetation fraction. **Figure 8** shows percent nitrogen by mass versus the calculated vegetation greenness. Both under-seeded and post-harvest clover showed a weakly negative relationship. Finally, **figure 9** shows that there is a weakly positive relationship between carbon-nitrogen ratio and vegetation mean greenness.

Figure : Biomass vs Vegetation Fraction.

Figure : Nitrogen vs Average Greenness.

Figure : Carbon-Nitrogen Ratio vs Average Greenness.

## Results Verification

Looking at the vegetation cover graph (**Figure 6)**, the results appear reasonable: under seeded clover has higher ground cover than post seeded cover, which has a higher ground cover than fallow. We also viewed the plot images of each cover type by date, and they appear similar to the predicted graph. However, **Figure 10** shows that ExGR-Zero had the strongest accuracy in the middle dates and weaker accuracy at the beginning and end dates. This suggests that vegetation may have been omitted especially in dates with high vegetation cover.

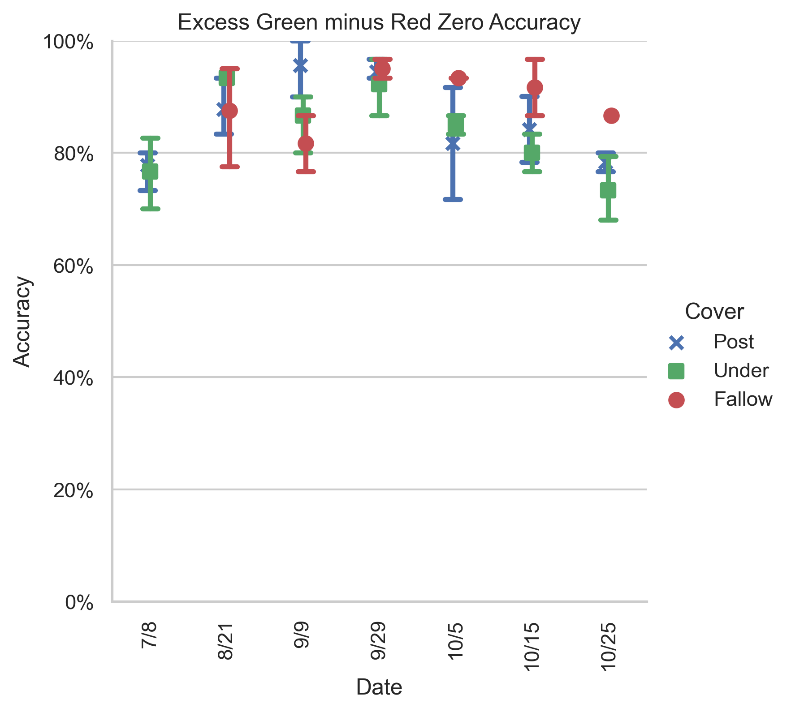


Figure : ExGR-Zero Accuracy.

# Discussion

## RGB Indices

This study showed that Using ExGR with a zero threshold did better than ExG using Otsu’s method. This result is similar to Meyer and Neto’s study that compared ExGR with a zero threshold and ExG with Otsu’s method on various vegetation (Meyer & Neto, 2008). They also found that both ExGR and ExG work better in areas of bare soil rather than areas of residue.

Louhaichi et. al (2001) used GLI and found that areas with high vegetation worked better with a threshold of zero and areas with low vegetation work better with a slightly positive threshold between 0.1 and 0.25. Our global threshold of 0.16 is consistent with their findings.

Finally, VARI is usually done on low resolution imagery from either aircraft or satellites (Gitelson, Kaufman, et al., 2002; Gitelson, Stark, et al., 2002). This reasoning is to remove blue noise from the atmosphere. Because our images were only 8’ from the ground, atmospheric effects were unlikely to be an issue causing the index to result in errors. Wang et al. used VARI with a zero threshold from a UAV and found low accuracy (~66%) compared to other RGB indices.

# Conclusion

Our study showed that RGB indices were able to classify vegetation from non-vegetation. While each index did well in at least one threshold, ExGR with a zero threshold was the most accurate. Furthermore, using classified images, we were able to get ground cover that appears reasonable. This may mean that simple RGB imagery can be used to make correlations and predictions to in field parameters.

### References

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, *80*(1), 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9

Gitelson, A. A., Stark, R., Grits, U., Rundquist, D., Kaufman, Y., & Derry, D. (2002). Vegetation and soil lines in visible spectral space: A concept and technique for remote estimation of vegetation fraction. *International Journal of Remote Sensing*, *23*(13), 2537–2562. https://doi.org/10.1080/01431160110107806

Li, B., Xu, X., Han, J., Zhang, L., Bian, C., Jin, L., & Liu, J. (2019). The estimation of crop emergence in potatoes by UAV RGB imagery. *Plant Methods*, *15*(1), 1–13. https://doi.org/10.1186/s13007-019-0399-7

Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, *16*(1), 65–70. https://doi.org/10.1080/10106040108542184

Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, *63*(2), 282–293. https://doi.org/10.1016/j.compag.2008.03.009

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, *20*(1), 62–66.

Snapp, S. S., Swinton, S. M., Labarta, R., Mutch, D., Black, J. R., Leep, R., … O’Neil, K. (2005). Evaluating cover crops for benefits, costs and performance within cropping system niches. *Agronomy Journal*, *97*(1), 322–332. https://doi.org/10.2134/agronj2005.0322a

Wang, T., Chandra, A., Jung, J., & Chang, A. (2022). UAV remote sensing based estimation of green cover during turfgrass establishment. *Computers and Electronics in Agriculture*, *194*, 106721. https://doi.org/10.1016/j.compag.2022.106721

**Self-score**

*Fill out this rubric for yourself and include it in your lab report. The same rubric will be used to generate a grade in proportion to the points assigned in the syllabus to the assignment.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Description** | **Points Possible** | **Score** |
| **Structural Elements** | All elements of a lab report are included **(2 points each)**:  Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score | 28 |  |
| **Clarity of Content** | Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**. | 24 |  |
| **Reproducibility** | Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified. | 28 |  |
| **Verification** | Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**. | 20 |  |
|  |  | 100 |  |