Spatial Analysis of Recreational Impacts in Mount Rainier National Park

Meghan Halabisky
SFR 502 Final Project

Abstract:

The National Park Service (NPS) has a dual mandate to both protect the natural resources within park boundaries as well as promote the recreational use of these natural resources. In order to monitor the balance of this dual mandate the NPS requires each National Park to report on the level of human impacts within the park. A previous study identified four landscape metrics based on social trail inventories that can be used to quantitatively describe the level of human impacts. For this research study I wanted to determine if management units could be categorized into levels of disturbance (ex. high, medium, low) for management purposes. In addition, I wanted to condense the four landscape metrics into one "disturbance" metric to allow for easier comparison between management units and for ranking levels of disturbance throughout the park. It is difficult to rank management units from high to low levels of impact when using multiple landscape metrics as the ranking is different depending on which landscape metric is chosen. In addition, I wanted to determine which explanatory variables; percent meadow, distance from visitor center, elevation, percent bare ground, and density of park trails could predict which management units had the greatest level of human impacts.

To determine if the different management units can be grouped into understandable levels of disturbance I conducted a hierarchical cluster analysis. The results of the cluster analysis showed that the management units could be grouped into four levels of disturbance. I conducted a Principal Component Analysis (PCA) to simplify the landscape metrics into a "disturbance metric" by using the first principal component. Finally, by conducting a PERMANOVA I was able to determine that park trail density was the greatest predictor of disturbance within a management unit, followed by distance from the visitor center. The results of this study provide a valuable technique to simplify multiple landscape metrics for management and reporting purposes. Use of cluster analysis techniques and principal component analysis will be especially useful when applied to the hundreds of management units throughout Mt. Rainier National Park.

1. Introduction

The purpose of the U.S. National Park Service (NPS) is *"...to promote and regulate the use of the...national parks...which purpose is to conserve the scenery and the natural and historic objects and the wild life therein and to provide for the enjoyment of the same in such manner and by such means as will leave them unimpaired for the enjoyment of future generations."* (National Park Service Organic Act. 16 US.C.1.)

This mission outlines a dual mandate to protect and conserve natural resources and to promote and regulate use of national parks for the public that is challenging to balance. In order to monitor how effectively the NPS balances this dual mandate each National Park must report the level of human impacts within the park as part of the "Visitor Experience and Resource Protection" (VERP) framework.

Social trails are informal trails created by erosion due to foot traffic from people and animals. Social trails are not part of the official National Park trail network and are an indication of human disturbance. Inventorying social trails combined with spatial analysis is one way to detect and monitor the level of human impacts to natural areas within the park (Leung & Louie 2008).

As part of a study for Yosemite National Park, Leung & Louie (2008) identified specific landscape metrics based on social trail inventories that could be used as indicators of human impacts. These metrics can be used to compare human impacts between management units and can also be used to monitor a management unit over time. This information is useful to develop management strategies to protect natural areas from future disturbance and for prioritization of restoration activities.

2. Background

The landscape metrics developed for Yosemite by Leung & Louie were recently tested out on two management zones (Paradise Meadows and Spray Park) within Mt. Rainier National Park (Moskal & Halabisky 2010). As part of this study the landscape metrics were also tested on the management sub-units within Paradise Meadows (subsequently referred to as "management units") (Figure 1). The results of this study showed that the landscape metrics developed for Yosemite National Park were successful at explaining the level of recreational impacts for Mt.

Rainier National Park and provided valuable information for tracking human impacts across the entire park and at the sub-unit level. The results of this study were aimed at providing NPS staff with the ability to compare human impacts of management units across Mt. Rainier National Park. The results could be used to develop a system of prioritization for restoration efforts. However, the results from the study are not easily understood for this purpose as the management units rank differently depending on which landscape metric is used for ranking. For this reason it is difficult to determine which management units should be targeted for restoration efforts.

3. Research Goals

Interpreting multiple landscape metrics can be confusing, especially across a broad landscape such as Mt. Rainier National Park (Table 1). Grouping continual datasets into categories can be useful for reporting and monitoring purposes. For example, understanding how many management units are in critical environmental condition can help with budgeting, management, and prioritization of restoration of these areas. In addition, the use of multiple landscape metrics makes it difficult to rank management units from high to low levels of impact. Because landscape metrics are often correlated to one another (Ritters, 1995) and because all four of the landscape metrics were derived from the same social trail inventory I suspected that the data from the landscape metrics could be reduced into one meaningful "disturbance metric". This "disturbance metric" could then be used for ranking management units in regard to their levels of human impacts.

A second objective of this study was to explore the reasons behind why some management units are more heavily disturbed than other management units. It is assumed that land cover is a major factor in determining which areas are more heavily impacted than others. Meadows are easy to walk through, provide unobstructed views, and are filled with highly photogenic wildflowers in the summer months. In addition, meadows are more sensitive to human impacts. Therefore, it is assumed that management units that contain high percentages of meadows would have the greatest human impacts. By understanding which variables influence recreational impacts across the landscape natural resource managers can make better management decisions to minimize impacts on the landscape.

3.1 Research Objectives

For these reasons I chose the following research objectives:

1. Understand if the management units within Paradise Meadows on Mt. Rainier differ in regards to recreational impacts as measured by social trails. Can management units be categorized into levels of disturbance (ex. high, medium, low) for management purposes?

2. Simplify landscape metrics into a single disturbance metric for ranking management units in regards to recreational impacts.

3. Use the above information to test the following hypothesis: Management units containing greater amounts of meadow will have a higher degree of human impacts as described by the landscape metrics. Other variables such as distance of management unit from visitor center, elevation, and density of official park trails are also correlated to the level of disturbance a management unit contains.

4. Methods

4.1 Data Collection Methods

Inventories of social trails were conducted between 1986 and 1988 by mapping trails on aerial photos and topographic maps during field surveys (Rochefort & Swinney 2000). Social trails were then digitized and entered into an ArcGIS database. In addition non-linear disturbances were delineated as polygons either by global positioning systems (GPS) or through manual photo interpretation and entered into an ArcGIS database. Some areas within Mt. Rainier have very detailed information regarding the width and depth of the trail, while other areas lack specific trail information. An earlier study showed there was a not a significant difference in the results of the spatial analysis when a default trail width of ½ meter was used instead of the actual trail widths (Moskal & Halabisky, 2010). Because actual trail widths are not available for the entire study area I used a default trail width of ½ meter.

4.2 Landscape Metrics

I selected four landscape metrics from the 2010 study (Moskal & Halabisky) that were calculated on 21 management units within Paradise Meadows in Mt. Rainier Park (Figure 1); Largest Patch Index (LPI-3), Weighted Mean Patch index (WMPI), density of social trails, and percent of impact area.

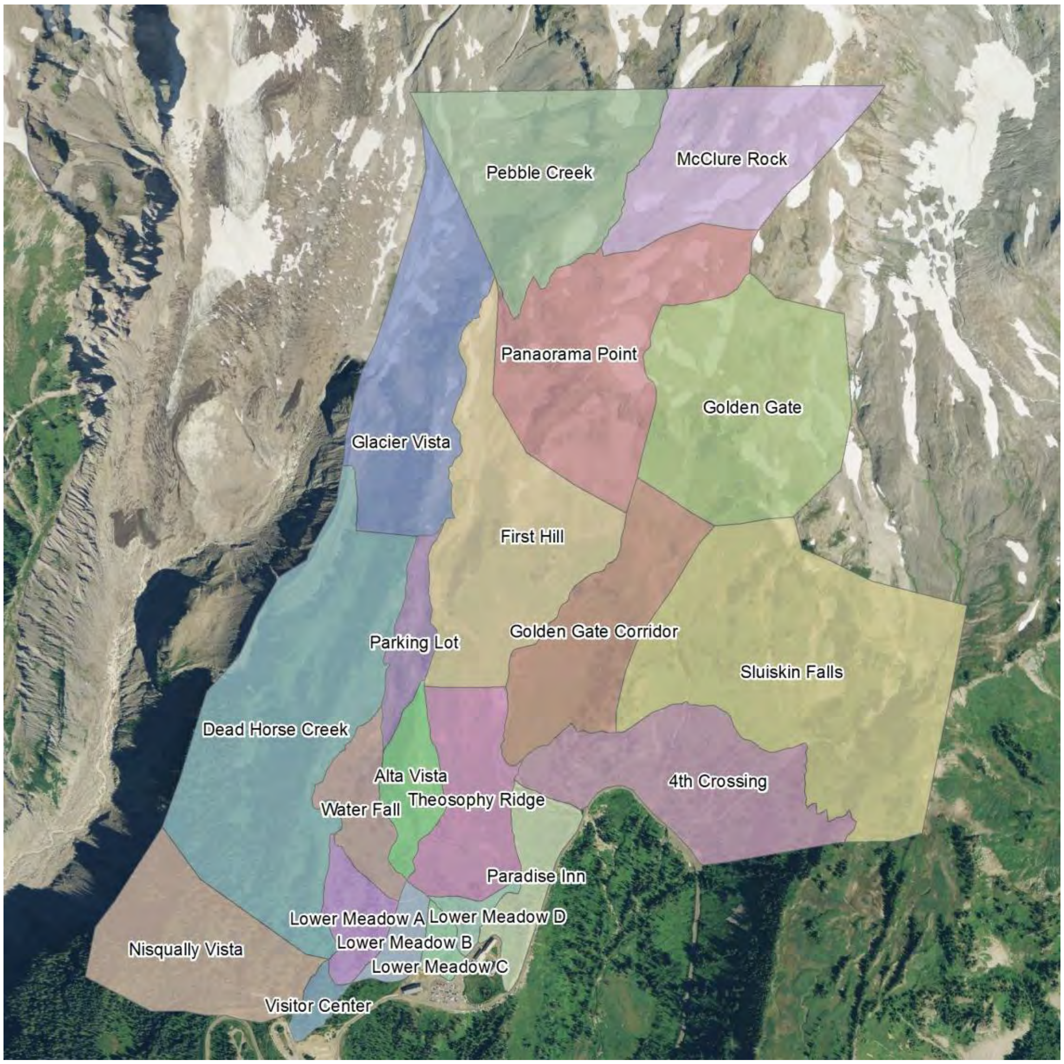

Figure 1: Study Area
Paradise Meadows, Mt. Rainier National Park, Washington, USA

Landscape metrics used as indicators of human impact (Leung & Louie 2008):

Patch density, patch size and variability metrics:

- **Weighted Mean Patch Index (WMPI) (Leung & Louie 2008)**

Metric: $WMPI = wf * (\sum a_{ij} / n) * (1/10000)$

where $wf = (\sum a_{ij} / A)$

Notations: a_{ij} = area (m^2) of patch ij, n = total # of patches, wf = weight factor, A = landscape/meadow area
Unit: hectare

Decreasing values indicate increasing degrees of fragmentation. This index is weighted by the size of the management unit, to allow for comparison between management units of varying sizes. Increasing spatial extent of informal trails would result in reduced index values even if the average patch size does not change.

- **Largest Patch Index – 3 (LPI -3) (Leung & Louie 2008)**

Metric: $LPI\text{-}5 = \sum max_5 (a_{ij}) / A * 100\%$

Notations: max_i = the largest *i* patches; a_{ij} = area (m^2) of patch ij, A = area (m^2) of the landscape (meadow)

This metric is based on the three largest patches within a management unit.
Decreasing values would suggest increasing degrees of fragmentation.

.

Area metrics:

- **Density of Social Trails**

Metric: Length (m) / Area (ha.)

Increasing values would suggest that more of the area is being impacted

- **Percent of Impact Area**

Metric: Total Impact Area m2 / Total Area m2

This metric is calculated on the total area of social trails.

Increasing values would suggest that a greater percentage of the overall area is affected.

Table 1: Landscape Metrics for Each Management Unit in Paradise Meadows. The top four management units with the greatest human impacts are highlighted for each spatial metric from high (red) to low (green). Although a pattern is evident the ranking of each management unit changes depending on which spatial metric is chosen.

Management Area	WMPI	LPI -3	Percent of Impact Area	Density of Social Trails (m/ha)
4th Crossing	0.47	81.37%	1.15%	124.15
Alta Vista	0.16	81.67%	5.15%	241.19
Dead Horse Creek	1.15	91.97%	1.77%	110.51
First Hill	0.77	98.18%	1.13%	107.78
Glacier Vista	0.59	93.62%	2.15%	116.25
Golden Gate	1.35	92.51%	1.80%	76.47
Golden Gate Corridor	2.06	96.45%	1.24%	75.47
Lower Meadow A	0.46	93.10%	4.71%	136.07
Lower Meadow B	0.74	96.25%	3.75%	119.68
Lower Meadow C	0.16	90.57%	4.01%	260.24
Lower Meadow D	0.12	73.16%	4.91%	263.87
McClure Rock	0.59	98.87%	0.60%	66.85
Nisqually Vista	1.34	93.71%	1.70%	64.57
Panaorama Point	0.51	76.73%	1.86%	123.00
Paradise Inn	0.32	96.82%	1.98%	233.23
Parking Lot	0.13	83.85%	7.68%	401.82
Pebble Creek	0.94	96.89%	0.66%	57.11
Sluiskin Falls	0.90	80.00%	1.46%	86.45
Theosophy Ridge	0.61	95.94%	1.22%	102.57
Visitor Center	0.11	68.18%	20.92%	357.65
Water Fall	0.23	68.38%	8.41%	314.78

4.3 Explanatory Variables

Explanatory variables were calculated in ArcGIS using data supplied by the U.S. National Park Service or created through remote sensing analysis. The explanatory variables are:

- Elevation - Elevation at the centerpoint of the management unit.
- Distance - Distance from Visitor Center to centerpoint of management unit.
- Percent meadow – Percent landcover of meadow within a management unit.
- Percent bare ground – Percent landcover of bare ground within a management unit.
- Park trail density – Density of official park trails within a management unit.

The first three explanatory variables were created using ESRI ArcGIS software. I chose elevation based on the assumption that management units at higher elevations were covered in snow for a greater amount of time, but also had less of a growing season to repair injuries to vegetation. I chose distance from Visitor Center to centerpoint of management unit under the assumption that fewer visitors would travel far from the Visitor Center and therefore, impacts would be less. I chose density of official park trails under the assumption that most social trails would originate from park trails. The last two environmental variables were created through an object based remote sensing classification I performed using imagery provided through the National Agriculture Inventory Program (NAIP). The NAIP imagery was flown by the USDA in 2009, has 1-meter pixel resolution, and includes an infrared band. I used Definiens Developed 8.0, a remote sensing software to create the classification for Mt. Rainier. I developed an algorithm to recognize and classify canopy, meadow, and bare ground (Figure 2). These variables were chosen under the previously mentioned assumption that land cover influences movement of people.

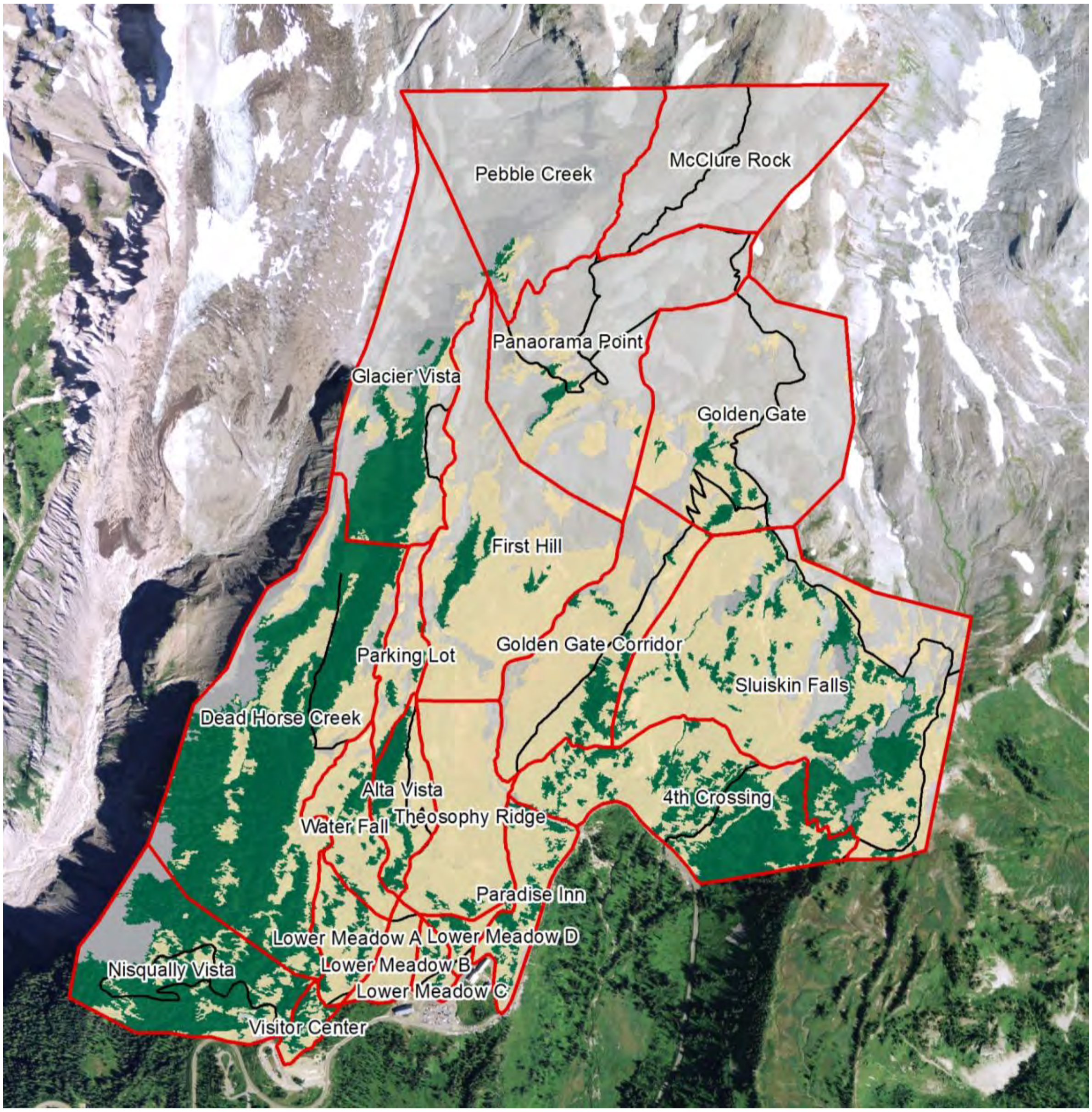

Figure 2: Classification of land cover for Paradise Meadows. Green depicts forest, yellow depicts meadow, and grey depicts bare ground.

4.4 Analysis

All statistical analysis was conducted using "R" statistical software package.

To determine if the different management units can be grouped into understandable levels of disturbance I conducted a hierarchical cluster analysis. To perform my analysis I used the hclust function in the "R" library package called "vegan". Before performing the cluster analysis I relativized the data by the column range. This reduced the variation amongst each variable, especially in the instance of the variable estimating the density of social trails. Next, I converted the raw table to a distance matrix. I chose Euclidean distance as it is appropriate for spatial data (McCune and Grace, 2002). Next, I performed a cluster analysis using Ward's method. Ward's method is an effective linkage method for cluster analysis and works well with Euclidean distances (McCune and Grace, 2002). I conducted a NMDS to examine the results of my cluster analysis in two dimensions using the "ordicluster" function. The results of the NMDS were overlaid on the results of the cluster analysis to double-check the number of appropriate clusters.

I conducted a Principal Component Analysis (PCA) to understand the relationship between the different variables and to simplify the landscape metrics into fewer variables. PCA is an eigenanalysis-based approach, which seeks to reduce the dimensionality of a dataset by creating fewer variables, called principal components. The principal components are uncorrelated to each other and are calculated by eigenvalue measures. Because the data is based on the same social trail dataset I assumed that the data are highly correlated and that PCA could simplify the data for easier comparison between management units. I then ranked management units by the first principal component and compared the results to the cluster analysis.

Finally, I performed a PERMANOVA to test whether the explanatory variables correlate with the first principle component derived from the four landscape metrics. I used the first principal component as the response variable for the PERMANOVA analysis. I relativized the variables by the column range as the data were at very different scales. I tested each variable in the PERMANOVA model separately before building a final PERMANOVA model. I realized that many of the variables were highly correlated so I chose to build my PERMANOVA model by placing variables I was most interested in, such as percent meadow, first. I ran a stepwise regression using the "Adonis" function in the "vegan" library package. I used euclidean distance and 9,999 permutations for the PERMANOVA model.

5. Results

5.1 Hierarchical Cluster Analysis

The hierarchical cluster analysis demonstrated the presence of two large clusters, which could then be broken down into four smaller clusters (Figure 3). I chose to categorize the management units by four categories as it is more meaningful for natural resource managers than limiting the categories to two classes. The results from the cluster analysis were compared to the raw data to designate the classification level to each cluster. The results from the cluster analysis relate to patterns discerned from Table 1. From the cluster analysis four management units are categorized as heavily impacted; Visitor Center, Parking Lot, Lower Meadow D, and Waterfall.

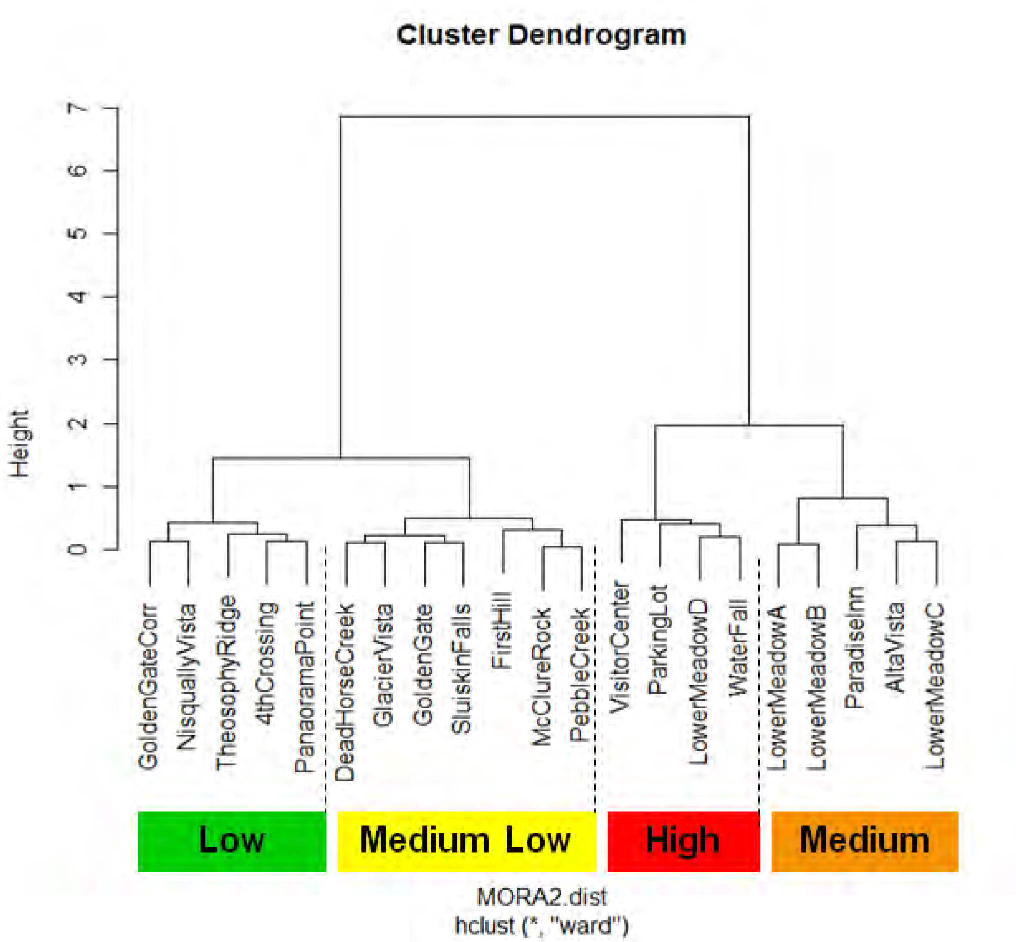

Figure 3: Hierarchical Cluster Analysis. Categories are grouped into high, medium, medium low, and low levels of human impacts.

Conducting a NMDS ordination agrees with the results of the hierarchical cluster analysis. Four distinct groups can be identified when overlaying the results of the NMDS on top of the cluster analysis (Figure 4).

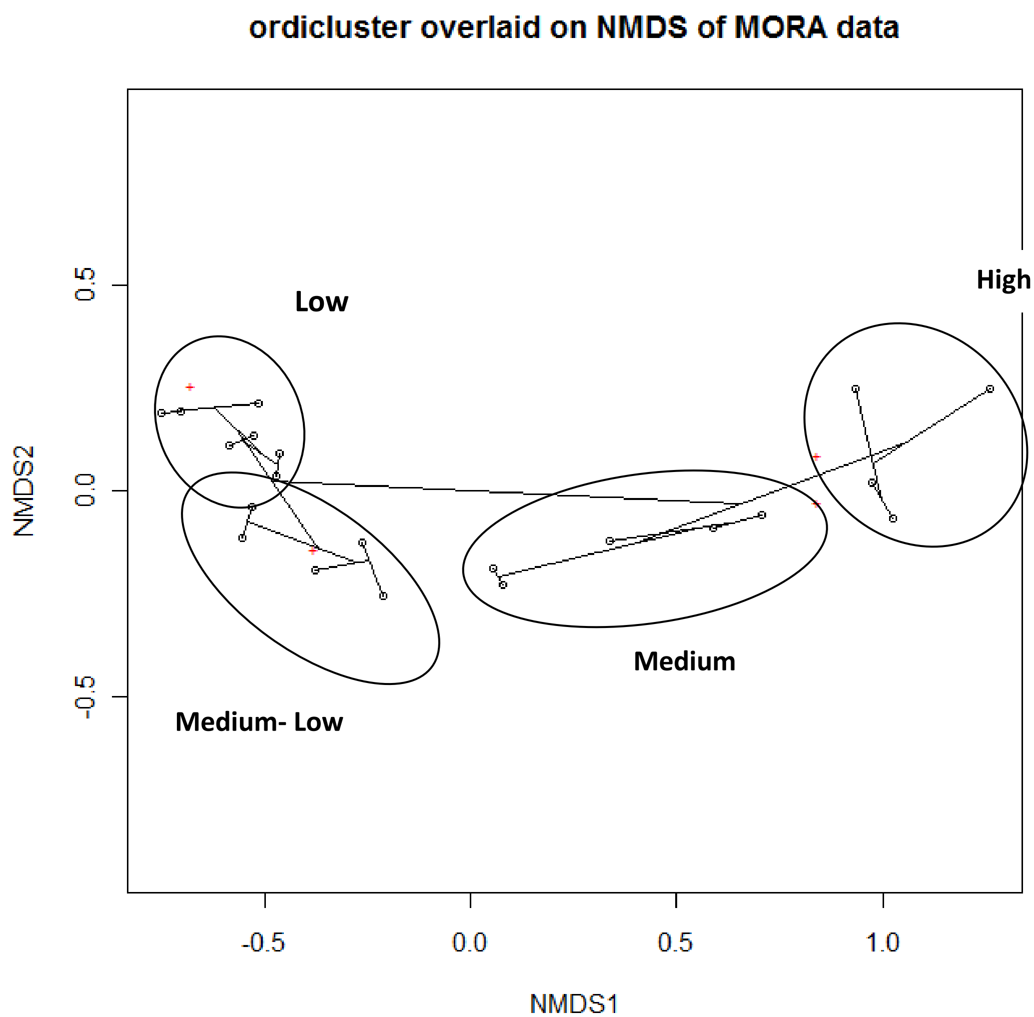

Figure 4: Comparison of cluster analysis to NMDS ordination. Four clusters can easily be identified as shown by the circles.

5.2 Principal Component Analysis

The results from the PCA indicate that the variables are indeed highly correlated to one another. WMPI and LPI-3 are positively correlated to one another and density of social trails and percent impact area are positively correlated to one another. However, the two groups are negatively correlated to each other. This makes sense as WMPI and LPI-3 are both metrics looking at patch size, while percent impact area and density of social trails directly represent the impacts caused by social trails.

The variables can easily be simplified into the first prinicipal component as demonstrated by Figure 5. The first principal component explains 90.7% of the variance.

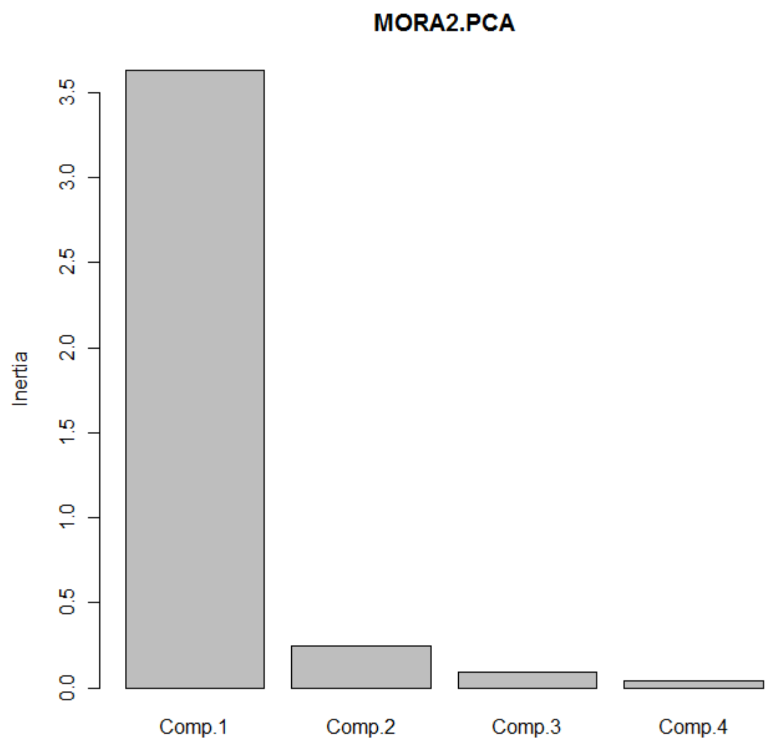

Figure 5. Scree plot of principal component analysis.

After running the PCA, I then ranked the management units by their first principal component (Figure 6). The ranking is in line with the results from the cluster analysis with the exception of Dead Horse Creek and Nisqually Vista. These two management units swapped poisitions in the ranking of the first principal component. These two management units are not very different in regards to the first principal component. This indicates that the line between the medium – low and low category may not be as well defined as the other categories.

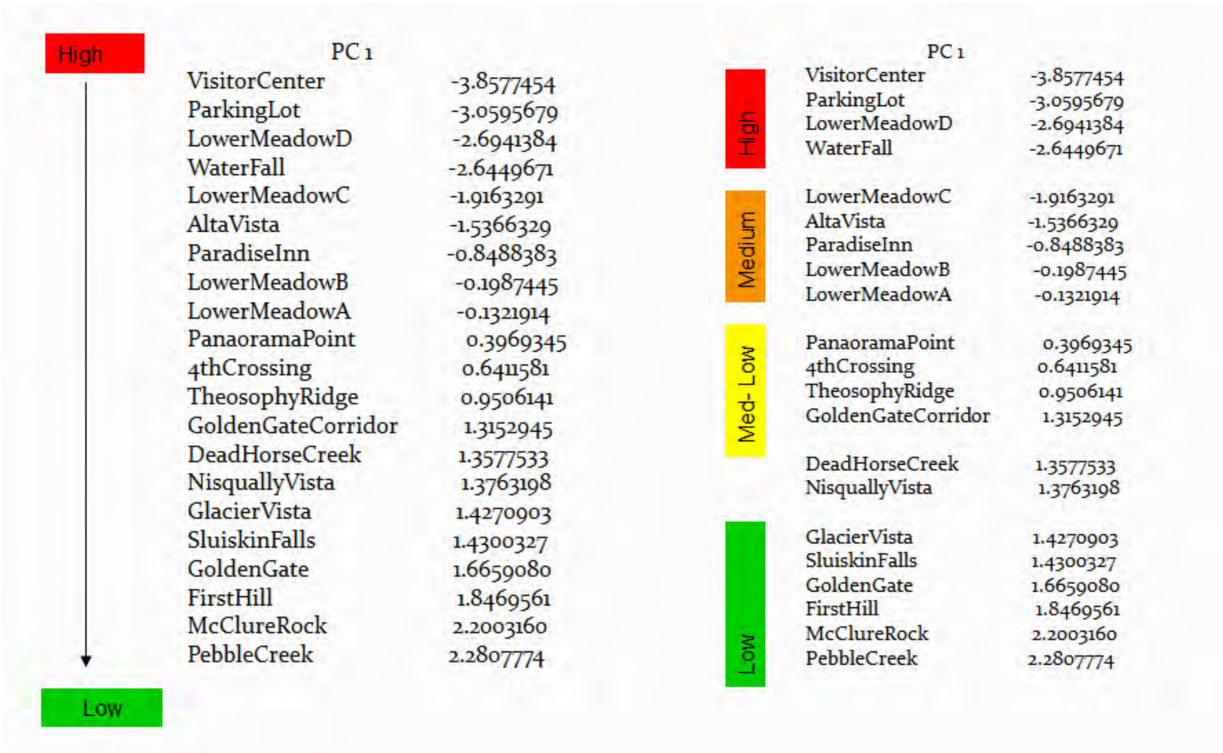

Figure 6. Ranking of management units based on the first prinicipal component. The ranking matches the categories assigned by the cluster analysis with the exception of Dead Horse Creek and Nisqually Vista, which were assigned to different categories in the cluster analysis.

5.3 PERMANOVA

The results of the PERMANOVA analysis show that all explanatory variables are significant (Table 3). Park trail density had the highest R squared value of 0.727. The next highest R squared value was 0.446 for distance from visitor center. Percent bare ground had an R squared value of 0.293, elevation had an R squared of 0.217, and percent meadow had an R squared value of 0.222.

Table 3. Results of the PERMANOVA analysis when tested out on each individual variable.

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Distance_m	1	32.341	32.341	15.278	0.44571	0.0015
Residuals	19	40.219	2.117	0.55429		
Total	20	72.56	1			

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Elev_m	1	15.707	15.7074	5.2494	0.21647	0.0317
Residuals	19	56.853	2.9922	0.78353		
Total	20	72.56	1			

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Prct_Meadow	1	16.091	16.0906	5.414	0.22176	0.0336
Residuals	19	56.469	2.9721	0.77824		
Total	20	72.56	1			

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
PT_density	1	52.718	52.718	50.481	0.72655	1.00E-04
Residuals	19	19.842	1.044	0.27345		
Total	20	72.56	1			

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Prct_BG	1	21.263	21.2632	7.8758	0.29304	0.0133
Residuals	19	51.297	2.6998		0.70696	
Total	20	72.56			1	

Because I was more interested in building a model from explanatory variables that were not created by management decisions I removed park trail density from my PERMANOVA model. I put percent meadow first in my model as that was the primary hypothesis test. I then included variables based on their significance values. None of variables tested had any significant interactions. As demonstrated in Table 4 both percent meadow and distance from visitor center each explained 16 of the sum of squared difference and showed significance below the 0.05 level.

Table 4: Results from Full PERMANOVA Model

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Prct_Meadow	1	16.091	16.0906	7.7195	0.22176	0.0145
Distance_m	1	16.611	16.6108	7.9691	0.22893	0.0126
Elev_m	1	6.508	6.5079	3.1222	0.08969	0.0987
Prct_BG	1	0	0.0001	0	0	0.9941
Residuals	16	33.351	2.0844	0.45963		
Total	20	72.56	1			

Elevation explained 6.5 of the sum of squares distance, but had significance only at the 0.1 level. Distance from visitor center and elevation are likely highly correlated and the variance is explained by the first variable entered in the model. Percent bare ground did not show any significance in the model. Percent bare ground is likely highly correlated to the percent meadow and elevation as most of the bare ground is at the higher elevations, which are covered in snow for longer period of times. Therefore, most of the variance that could be explained through bare ground is already explained with the first three variables.

6. Discussion and Conclusions

The results of this study provide a valuable method to simplify multiple landscape metrics for management and reporting purposes. Use of cluster analysis techniques and principal component analysis could be especially useful when applied to the hundreds of management units throughout Mt. Rainier National Park. Cluster analysis is a useful technique in determining natural breaks in the data that can be used to groups management units into different levels of disturbance. Principal component analysis is useful in condensing the four landscape metrics into a single "disturbance" metric that can be used for ranking different management units in regards to human disturbance. It can also be used for monitoring an individual management unit or management zone over time. It replaces the need to select one of the four landscape metrics for ranking and monitoring purposes.

All of the explanatory variables selected for this research are related to the level of human impacts within a management unit. The results from the PERMANOVA analysis quantitatively prove the assumption that management units containing greater percent of meadow will have a higher degree of human impacts. This is likely due to the attraction of meadows and because meadows may be more sensitive to human impacts than other land cover classes. Surprisingly, although having a large amount of meadow within a management unit does create a higher level of disturbance it is not as big of a factor as the density of park trails, the distance away from the visitor center, and the percent of bare ground within a management unit.

Density of park trails had the highest sum of squares value when explaining the variance between the different levels of human impacts for each management unit as described by the first principal component. This makes sense as most social trails begin from the official park trail. This is important when considering where to locate official park trails. Because the presence of official park trails is such a significant predictor of human disturbances care should be taken when placing trails nears especially sensitive areas.

7. Recommendations

The results of the cluster analysis and the principal component analysis ranking can be used for annual reporting purposes such as the "Visitor Experience and Resource Protection" (VERP) report to track recreational impacts over time and between management units. The techniques developed here could also be run on the management zone level (e.g. Spray Park, Paradise meadows) to make comparisons between management zones. This analysis only focused on social trails as indicators of human impacts. Additional factors that could be included are; degree of trampling (Holmquist 2007), presence of litter (Rochefort & Swinney, 2000), and evidence of campsites (Rochefort & Swinney, 2000, Moskal & Halabisky 2010).

The PERMANOVA analysis using the selected explanatory variables should be run on the entire park, which would increase the sample size and confirm or deny the results from this study. In addition to the explanatory variables selected for this analysis locations of viewpoints or additional points of interest should be added to the analysis as social trails may increase around these locations.

Finally, it would be interesting to combine the results of this research with field surveys to determine how recreational impacts may influence plant and animal communities. For example, do social trails influence the vegetation communities within each management unit? In order to properly balance the dual mandate outlined by the National Park Service mission it is necessary to understand how recreational impacts influence the natural systems of Mt Rainier National Park. The PERMANOVA could be re-run using the field data as the response variable and the disturbance metrics as the explanatory variables.

In the report titled, "Analysis of Social Trails in Mt. Rainier National Park, Pilot Study" (Moskal and Halabisky, 2010), which looked at applying the spatial analysis techniques developed for Yosemite to Mt. Rainier National Park, the authors recommended that the management units be divided into multiple landcover types, such as meadow, forest, and bare ground. They suggest that if a management area is not split up into habitat types such as meadow and forest one may erroneously conclude that a meadow is not heavily impacted by social trails and campsites. A management area with a large forest area will inflate the landscape metrics for the meadow if they are grouped together in a management area (Figure 7). Analysis run only on meadows would help locate the most heavily disturbed meadows within the park, regardless if the management unit they are located in has large amounts of undisturbed forests.

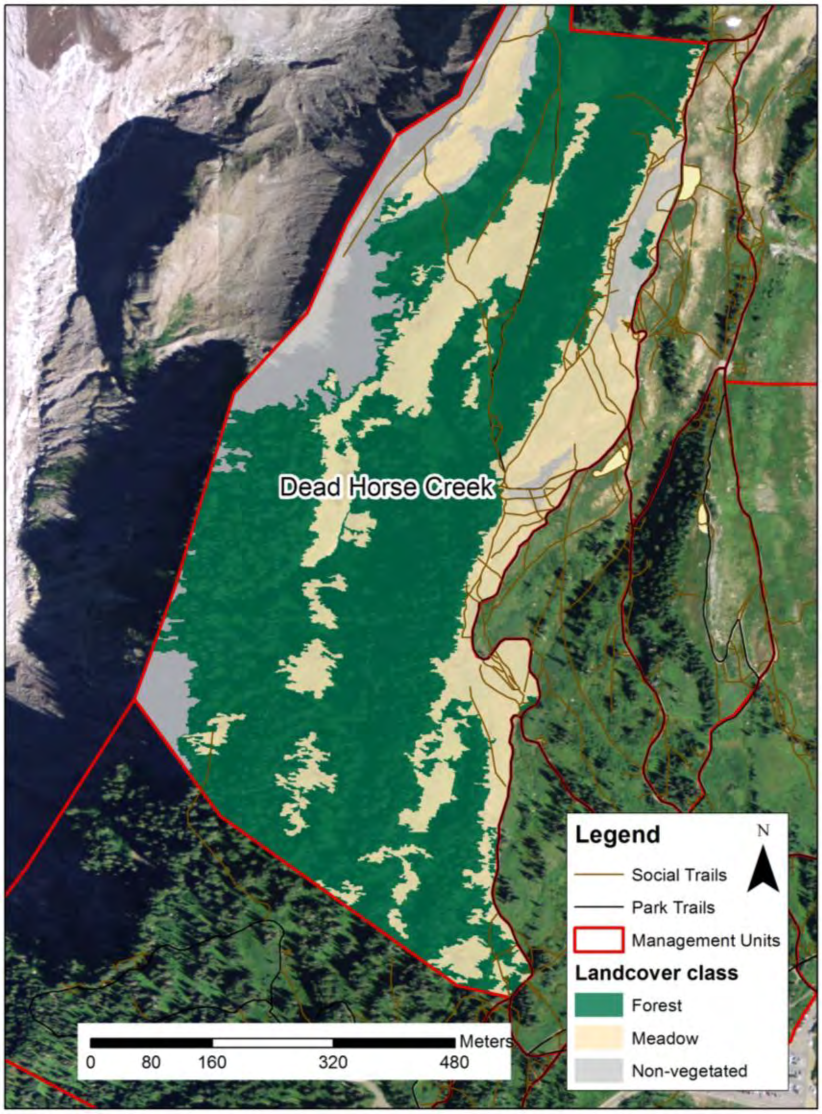

Figure 7. Example of how the locations of mapped social trails are primarily in meadows. A landcover classification could be used to examine impacts only within meadows using the techniques outlined in this analysis.

A basic land cover classification algorithm was created for this analysis that was previously unavailable for earlier work and appears to provide an accurate classification useful for future spatial analysis in Mt. Rainier National Park. An accuracy assessment should be run on the classification to test the accuracy of this dataset. It could then be used for park-wide analysis. I recommend a separate analysis only examining disturbance on meadows within management units. To do this the data could be clipped to meadows and the landscape metrics could be run only on these areas. This would remove the problem of erroneously concluding that a meadow is not heavily impacted by social trails and campsites because the management unit was drawn to include a large portion of forested areas.

References:

Forman, R. T. T. (1995). Some General-Principles of Landscape and Regional Ecology. Landscape Ecology **10**(3): 133-142.

Fraver, S. (1994). Vegetation Responses along Edge-to-Interior Gradients in the Mixed Hardwood Forests of the Roanoke River Basin, North Carolina. Conservation Biology 8(3): 822-832.

Harper, Karen et al. (2005). Edge Influence on Forest Structure and Composition in Fragmented Landscapes. Conservation Biology 19 (3): 768-782.

Holmquist, J.G. Trails and Meadow Fragmentation in Yosemite National Park: Effects on Invertebrate Fauna and Patterns of Abundance and Biodiversity. El Portal, CA: USDI National Park Service, Yosemite National Park

Holmquist, J.G. & Schmidt-Gengenbach, J. (2007). Effect of Experimental Trampling Addition and Reduction on Vegetation, Soils, and Invertebrates in Tuolumne Meadows: Preliminary Report. El Portal, CA: USDI National Park Service, Yosemite National Park

Leung, Y.-F. & Louie, J.. 2008. Visitor Experience and Resource Protection; Data Analysis Protocol: Social Trails – 2007 Field Season. USDI National Park Service, Yosemite National Park.

McCune, B., and J.B. Grace. 2002. Analysis of Ecological Communities. MjM Software Design, Gleneden Beach. ISBN: 0-9721290-0-6

McGarigal, K., & Marks, B. J. (1995). FRAGSTATS: Spatial Pattern Analysis Program for

Moskal, L.M. & Halabisky, M. (2010). Analysis of Social Trails in Mt. Rainier National Park, Pilot Study. Performed under Pacific Northwest Cooperative Ecosystems Studies Unit Task Agreement # J8W07090020

Quantifying Landscape Structure (General Technical Report PNW-GTR-351). Portland, OR: USDA Forest Service, Pacific Northwest Research Station.

Ritters KH, O'Neill RV, Hunsaker CT, et al. (1995) A factor analysis of landscape pattern and structure metrics. Landscape Ecology 10(1):23-39. Available at http://www.springerlink.com/index/10.1007/BF00158551.

Rochefort, Regina M.; Swinney, Darin D. (2000). Human impact surveys in Mount Rainier National Park : past, present future. In: Cole, David N.; McCool, Stephen F.; Borrie, William T.; O Loughlin, Jennifer, comps.2000. Wilderness science in a time of change conference   Volume 5: Wilderness ecosystems, threats, and management; 1999 May 23 27; Missoula, MT. Processings RMRS-P-15-VOL-5. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. P. 165-171.

Appendix A: "R" Code

```
# Bring in table and set row names. Remove variables.
MORA <- read.table(file.choose(), header=TRUE, sep = "\t", na.strings ="NA", dec=".", strip.white=TRUE)
row.names(MORA)=MORA$SiteName
MORA
MORA2 <- data.frame(MORA[,c(3:6)])
MORA2

## Standardize by range.
library(vegan)
MORA2.range<- decostand(MORA2, "range")
MORA2.range

## Convert to distance matrix.
MORA2.dist <- dist(MORA2.range , method = "euclidean")
MORA2.dist

## Next I will use hierarchical clustering. I chose ward's method
MORA.clust<- hclust(MORA2.dist, method= "ward")
MORA.clust
plot(MORA.clust) # cannot figure out how to label my leaves with the site names.
?plot.hclust
identify(MORA.clust)

## Scree plot
scree <- cbind(MORA.clust$merge,MORA.clust$height)
scree
plot(x=11:1, y=scree[10:20,2], xlab = "Number of groups",ylab = "Height")


## It looks like I have 2 major groups or 4 smaller groups with a pretty even distribution.
##  I will also look at NMDS. I think it defaults to bray-curtis distance measures. Changed to distance
measure of NMDS to ?values
   euclidean.
?metaMDS
```

```
NMDS.MORA<- metaMDS(MORA2.range, dist ="euc")
NMDS.MORA
NMDS.MORA2 <- NMDS.MORA
plot(NMDS.MORA, main = "ordicluster overlaid on NMDS of MORA data")
ordicluster(NMDS.MORA, MORA.clust, label= c(row.names))
?ordicluster

## Now I will look at how to choose the number of clusters.
scree<-cbind(MORA.clust$merge, MORA.clust$height)
?plot
plot (x=4:1, y=scree[35:46,3], xlab="Number of groups", ylabd="Height") ## Didn't find this very useful. Basing # of clusters as an efficient management tool is a better idea.

## Now I will try PCA
## This will remove the effects of each variables variance.
scaled.data <- scale(MORA2)
scaled.data
S <- cov(scaled.data)
round (S,3)

##Calculate eigen values of S

S.eigen <- eigen(S)
S.eigen$values  # Eigenvalue
sum(S.eigen$values)# Sums to the number of variable. 4 in this case
S.eigen.prop <- S.eigen$values/sum(S.eigen$values)
S.eigen.prop
round(S.eigen.prop, 3) # Easier to read
## Almost all of the variance is described in the first PCA.
S.eigen$vectors # 1 vector per eigenvalue

## Find PCA for each management unit. # Low PCA represents high impacts and high PCA represent low impacts.
PC.scores <-   scaled.data %*% S.eigen$vectors[,1:3]
PC.scores
PCA.1<- data.frame(sort(PC.scores[,1]))
PCA.1
```

```
MORA.response <- data.frame(PC.scores[,1])
MORA.response

# Dsiplay PCA in a scree plot. Retrieve summary.
MORA2.PCA <- princomp(MORA2.range, cor=TRUE)
MORA2.PCA
summary(MORA2.PCA)
plot(MORA2.PCA)
screeplot(MORA2.PCA)
MORA2.PCA

## PERMANOVA
MORA2
MORA.exp <- read.table(file.choose(), header=TRUE, sep = "\t", na.strings ="NA", dec=".",
strip.white=TRUE)
row.names(MORA.exp)=MORA.exp$SiteName
MORA.exp

MORA.exp2 <- data.frame(MORA.exp[,c(4,5,7,8,10)])
MORA.exp2

#Relativize explanatory variables
library(vegan)
MORA.exp2.range<- decostand(MORA.exp2, "range")
MORA.exp2.range

## PCA of explanatory variables
MORA.exp.PCA<- princomp(MORA.exp2.range, cor=TRUE)
MORA.exp.PCA
summary(MORA.exp.PCA)
plot(MORA.exp.PCA)
screeplot(MORA.exp.PCA)
MORA.exp.PCA

## Eigen values for explanatory variables
scaled.data <- scale(MORA.exp2)
```

```
scaled.data
S <- cov(scaled.data)
round (S,3)

##Calculate eigen values of S

S.eigen <- eigen(S)
S.eigen$values  # Eigenvalue
sum(S.eigen$values)# Sums to the number of variables. 5 in this case
S.eigen.prop <- S.eigen$values/sum(S.eigen$values)
S.eigen.prop
round(S.eigen.prop, 3) # Easier to read
S.eigen$vectors # 1 vector per eigenvalue

## Not really helpful. I will build my model based on  common sense instead.

attach(MORA.exp2.range)
 # if necessary
   detach(MORA.exp2.range)

#Do I need to convert variable to a distance matrix? NO, I can set it in ADONIS.
# But, didn't I do this for PCA anyway. Am I doing a distance matrix twice now?
# Must change the method from the default bray curtis to euclidean.

# Build model - stepwise regression
# Test each of the env. variables
# Build model - stepwise regression

result1 <-adonis(MORA.response ~ Distance_m, method ="euc", perm = 9999)
result1

result2<-adonis (MORA.response ~ Elev_m, method="euc", perm =9999)
result2

result3 <- adonis(MORA.response ~ Prct_Meadow, method = "euc", perm = 9999)
result3
```

```
result4 <-adonis(MORA.response ~ PT_density, method="euc", perm = 9999)
result4

result5<- adonis(MORA.response ~ Prct_BG, method="euc", perm = 9999)
result5

result6<- adonis(MORA2 ~ PT_density + Prct_Meadow + Prct_BG + Elev_m, method="euc", perm =
9999)
result6
## Chose to remove PT_density

final.result <- adonis(MORA.response ~ Prct_Meadow  + Distance_m + Elev_m + Prct_BG ,
method="euc", perm = 9999)
final.result


## Build model
model.1 <- adonis(MORA.response ~  Prct_Meadow * Prct_BG, method="euc", perm=9999)
model.1

model.2 <- adonis(MORA.response ~ Prct_Meadow * Distance_m , method="euc", perm =9999)
model.2

model.2 <- adonis(MORA.response ~ Prct_Meadow *  Elev_m , method="euc", perm =9999)
model.2

model.3 <- adonis(MORA.response ~ Elev_m * Distance_m , method="euc", perm =9999)
model.3

model.4 <- adonis(MORA.response ~ Elev_m * Prct_BG , method="euc", perm =9999)
model.4

model.5 <- adonis(MORA.response ~ Distance_m * Prct_BG, method="euc", perm =9999)
model.5


## CART
```

```
MORA.CART <- data.frame(MORA.response, MORA.exp2.range)
MORA.CART
MORA.CART [,1]
MORA.exp2.range

library(mvpart)
MORA.response

MORA.CART2 <- mvpart(MORA.CART[,1] ~ Distance_m + Elev_m + Prct_Meadow + Prct_BG +
PT_density, data = MORA.CART, xv ="pick", all.leaves=TRUE)
MORA.CART2


MORA.CART2 <- mvpart(MORA.CART[,1] ~ Distance_m, data = MORA.CART, xv =1se",
all.leaves=TRUE)
?mvpart

## This doesn't really help determine which variables to put first. I think I will use commonsense and stick
with my final model that I came up with.
```