Coastal resource conditions of backcountry campsites in Kenai Fjords National Park

Shannon Wesstrom

April 17, 2020

Introduction

Kenai Fjords National Park (KEFJ), a 669,983 acre park, is in the unique position of offering experiences with receding tidewater glaciers, endangered wildlife, and backcountry adventures all within a three hour drive of Anchorage, Alaska. Although there have been consistent variances in overnight backcountry usage throughout the years, camping in a remote location in KEFJ is still a popular activity for those visiting the park. While much of the park is inaccessible by foot, many guests seek to find secluded campsites by kayak, boat, or float plane. This study aims to investigate the ecological changes of campsites over time in the backcountry of KEFJ.

In 2017 there were 674 backcountry overnight campers in KEFJ – a slight increase from prior years. However, backcountry camping is following the trend of a significant decrease from decades prior. Just ten years ago, in 2009 there were 1,692 backcountry users and in 1999 there were a record 2,989 backcountry campers (Stats Report Viewer, 2019). Even with lower numbers in backcountry campers for recent years, this group still represents an important visitor type to the park with their individual motivations and expectations. Understanding their influence on the environment is very important to park managers so they can better establish guidelines to reduce the anthropogenic impacts on the environment.

Multiple studies have been conducted that suggest some generalizations of campsite effects on the environment. Namely, Cole and Hall (1992) discovered that even with increases in campsite size, vegetation cover remained consistent over an 11-year study period in Eagle Cap Wilderness, Oregon. Phillips *et al.* (2015) summarized this well by explaining that on established sites, changes in “areal extent” or surface area were more obvious than the changes in the impact intensity. Additionally, increases in site numbers (which imply increases in surface area disturbances) over time may be more of a concern for managers than the degradation at the individual site level. This opens up a popular discussion for managers in terms of planning strategies. More confined “designated camping areas” limit the formation of new sites and allow for more desirable well-maintained sites which work well in locations with high volumes of visitors (Hampton and Cole, 2003; Leung and Marion 2000). Also, Cole *et al.* (2008) found in a study of Grand Canyon National Park that, over 20 years, informal sites were created even under a functioning confinement campsite plan, resulting in an increase in total disturbed area. This study leads managers to believe that a more dispersed campsite strategy plan works in areas where visitor campsite demand is low.

Given the unique dispersal of backcountry users over time in KEFJ, we have examined the ecological changes on campsites in the backcountry of KEFJ. Specifically, we identified if our indicator variables (i.e. vegetation cover loss, tent rocks, mineral soil exposure, etc.) could predict campsite area measurements. Determining the extent of ecological changes due to recreational influences in KEFJ help managers improve backcountry experiences and open up avenues for different types of park visitors. When completed, this study could back the rationale for designated dispersed campsites, promote more non-motorized means of travel to the backcountry, and even provide supplementary evidence for Congress to designate this defined eligible wilderness area as true Wilderness.

Methods

Data Collection

Following the direction of the work conducted by Dr. Chris Monz (Utah State University) and KEFJ staff in 2008 – 2012, an analysis of the 101 sampled sites on 15 beaches within the 669,983 acre park were conducted. To focus the study on areas of more popular use, survey efforts were directed to campsites in Aialik Bay, Northwestern Fjord, and Resurrection Bay. To discover potential campsites, informal trails (trails not created by park management) were followed to find larger plots of vegetation, surface layers, or soil disturbances that may indicate human influences. Once the plot was determined to be a campsite, a metal center point marker with a unique identification number was buried in the center of the campsite. The location and marker number was recorded in a Trimble GPS unit and a campsite assessment and radial transect data collection were conducted.

Table 1

Impact assessment parameters, assessment methods and measurement scale.

Site attribute	Method used	Measurement scale
Area of observable impact	Radial transect measurement	Square area of campsite
Condition class	Ocular estimation	Five level condition class scale
Fire sites	Counts	Total number of fire sites present
Informal trails	Counts	Total number of trails present
Mineral soil exposure on site	Ocular estimation	Six level cover scale, 0–5%; 6–25%; 26–50%; 51–75%; 76–95%; 96–100%)
Stumps/cut shrubs	Counts	Total number of cut stumps present
Vegetative ground cover on site and in control areas	Ocular estimation	Six level cover scale, 0–5%; 6–25%; 26–50%; 51–75%; 76–95%; 96–100%)
Campsite substrate type	Observation	Cobble, organic soil, sand
Human waste sites	Ocular estimation	Three level human waste scale
Litter and trash	Ocular estimation	Four level trash quantity scale
Root exposure on site	Ocular estimation	Three level root exposure scale
Tree damage	Ocular estimation	Three level tree damage scale

Table 1. Indicator variables methods and measurement scales. Adapted from Monz et al. 2010.

Two types of campsite assessments were conducted as part of a long-term monitoring program for KEFJ: Rapid and Complete. Rapid assessments were done by one person and took about five minutes to complete. These quick assessments are designed to discover new campsites not previously recorded and to check in on campsites the monitoring program was already aware of. As this was a long term monitoring project, these assessments are performed in between regular monitoring field seasons. Complete assessments required two to three technicians and about fifteen to twenty minutes to complete. Using a Trimble GPS unit, several observations of the campsite were noted and recorded: tag number, location description, area measure type, distance from high tide line, landing substrate, camping site substrate, tree canopy cover, cover onsite, cover offsite, substrate exposure, tree damage, ghost tree damage, root exposure, number of stumps, number of ghost stumps, number of fire sites, number of trails, number of tent rocks, trash, human waste, defining a condition class, comments, field notes, and campsite photos are taken (Table 1). All data was then uploaded, stored, and analyzed in GPS Pathfinder Office to be reviewed and evaluated later (Phillips *et al.* 2015).

Data and Assumptions

The 101 campsites were spread across 15 beaches. Figure 1 shows the area (m^2) distributions of campsites on all 15 beaches. Our median campsite size is 19.593m^2 with the largest measuring 290.239m^2 and the smallest being 0m^2. Sites that measured 0m^2 were measured as such, indicating that a site that was once there, had recovered.

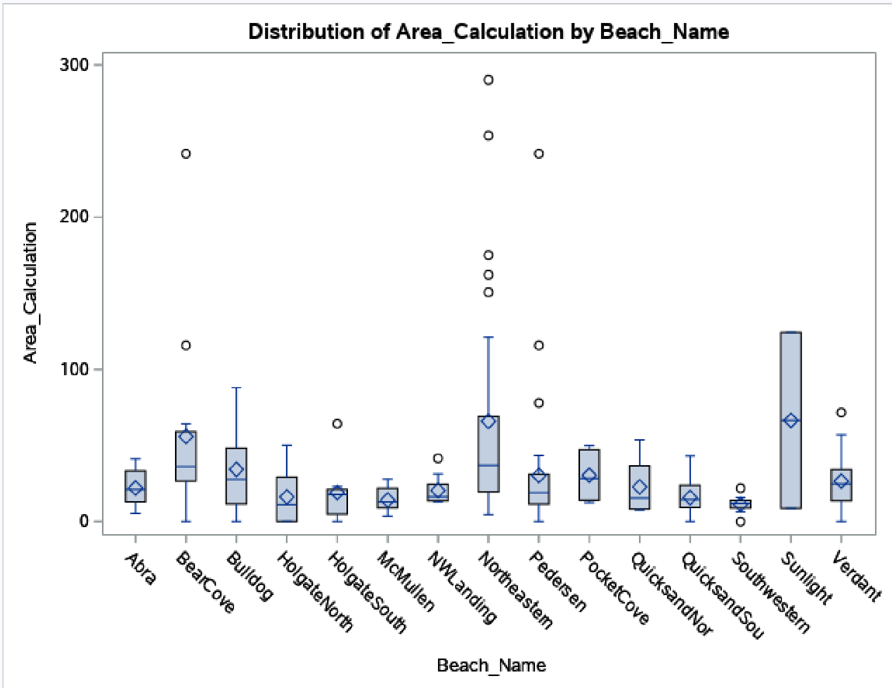


Figure 1. Area measurement distributions across all 15 beaches.

Figure 2 demonstrates the distribution of area measurement data for our sample. The clear right tail indicated the need for a data transformation in plot 2a. Plot 2b shows the distribution of area measurement data after a square root transformation. A square root transformation was chosen due to the $0m_2$ values that need to be retained. Log transformations and any variation that included adding an integer, did not work and eliminated the areas with $0m_2$ recorded. With the transformation completed, a crude initial multiple linear regression was conducted.

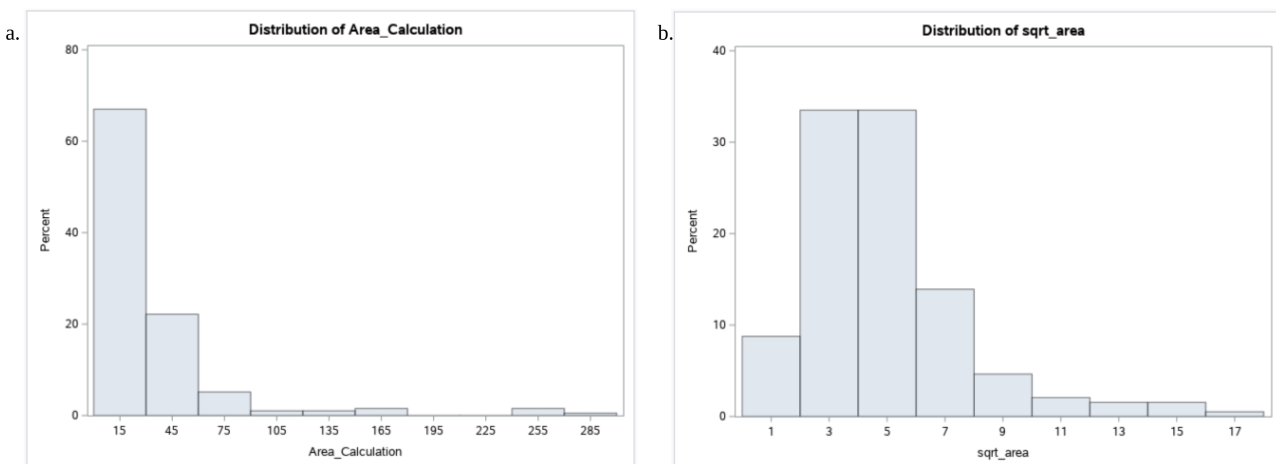


Figure 2. Histogram plots of area measurement data. Right tail indicates the need for a transformation (a). Plot b is the distribution of data after a square root transformation.

Figure 3, is a representation of the crude initial model. There is a reasonable amount of spread in the residuals, with some clustering. The data overall appears quite normal given the histogram and normal QQ plot. There is still a bit of trail straying from the projected regression line as area increases, but is an improvement from the untransformed data. The Cook's D and Out vs. Leverage plots however, single out points 15 and 133 as potential outliers or influential points. By deleting those points, we expected an improvement in how the data follows the regression line in the QQ plot. After deleting those points, we ran the criteria assumption diagnostics again. As seen in Figure 4, the diagnostic plots show little to no improvement. The Cook's D plot only decreased to a 0.25 level, only a 0.05 difference and the QQ plot still shows trailing points as area increases. With this information, points 15 and 133 will be left in the model. After several attempts of trying to transform the data and deleting potential outliers, this is the best way the data met the model assumptions, as it is real ecological data.

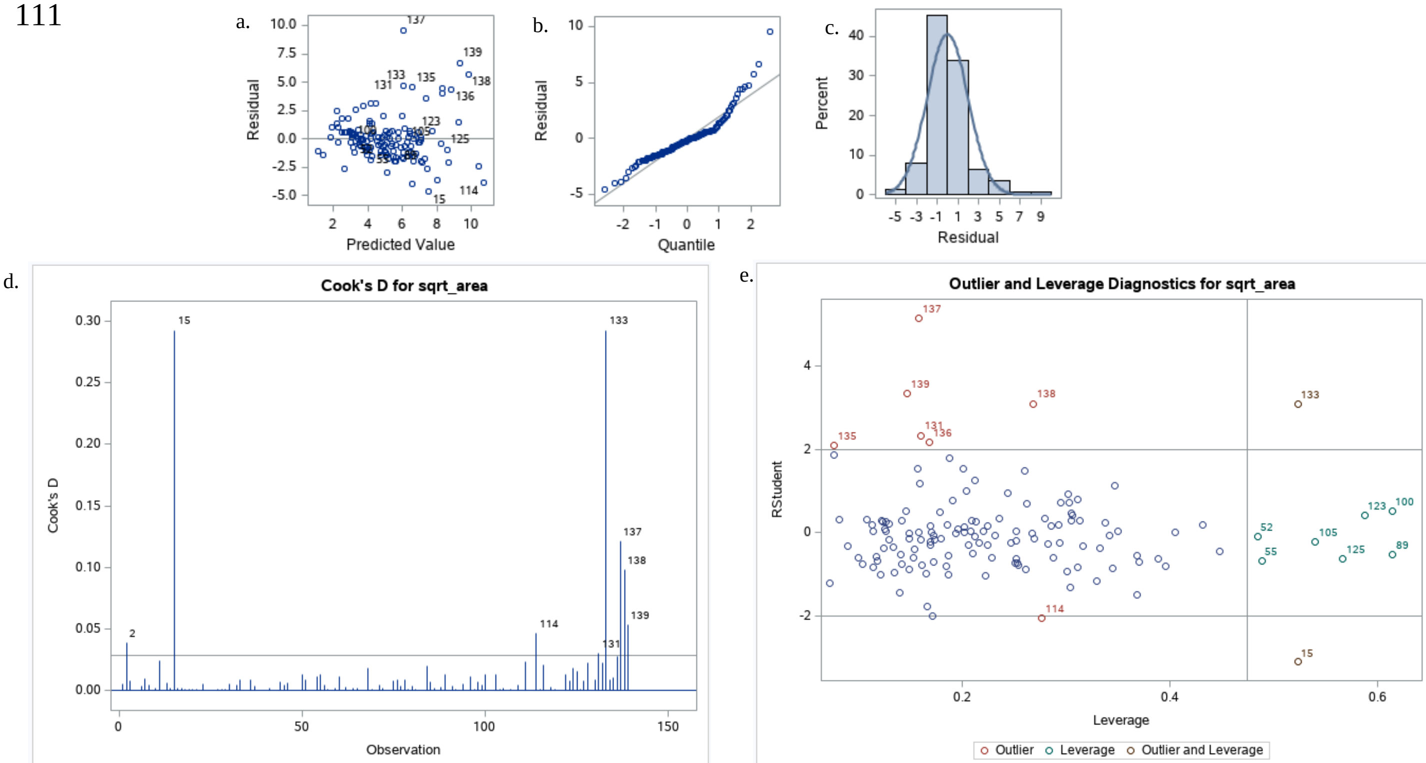


Figure 3. Model Assumption Graphical Diagnostics. A. Residuals vs. Predicted Values. B. Normal QQ. C. Histogram. D. Cook's D. E. Outlier vs. Leverage.

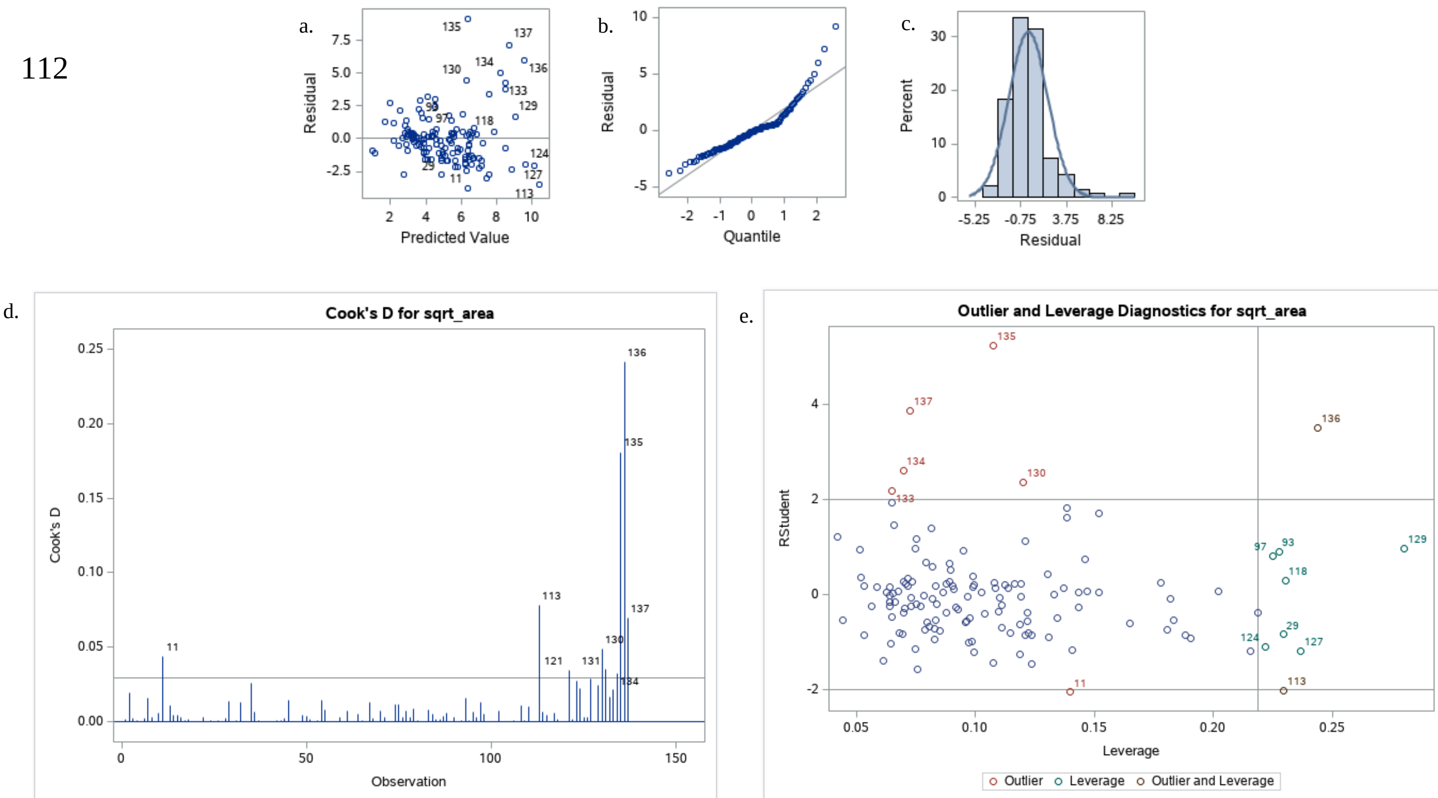


Figure 4. Model Assumption Graphical Diagnostics. A. Residuals vs. Predicted Values. B. Normal QQ. C. Histogram. D. Cook's D. E. Outlier vs. Leverage with points 15 and 133 removed.

Variable Selection and Interaction Terms

With a combination of quantitative and qualitative data, there were 40 variables to consider for the regression model. It was originally thought there may have been some correlation between beaches and area measurements, as some have generally smaller campsites as seen in Figure 1. In conducting a scatterplot and correlation matrix there did not appear to be any correlation between the beach and square root area measurements, although, there was some correlation between the number of trails and moderate root exposure. The Pearson Correlation Coefficient was above 0.50 and the p-value significant ($p < 0.0001$). However, after conducting tests for multicollinearity, Variance Inflation Factors (VIF) and Condition Indices (CI), there appears to be no collinearity occurring within the variables. None of the VIFs were above ten and the average of all of them combined were just over 1. The largest C.I. was 21.582 and tried to connect trash levels of none to a handful and the intercept. While it explained 96% of the variance in the intercept, that was the only variable it explained and does not count in this analysis. Since it did not explain two variables at 50%, there was no collinearity.

Variables selected for the model were chosen using a variety of tests. First, by comparing Adj. R_2, Mallow's C(p), AIC, and SBC values. These values are found in Table 1.

Table 2

Adj. R_2	Mallow's C(p)	AIC	SBC
0.4033	2.738	224.535	271.486

Table 2. Selection Criteria Values.

The 15 variables chosen for this model include: mineral soil exposure, number of fire rings, number of trails, number of tent rocks, condition class rating, Holgate North, Northeastern, Quicksand North, Southwestern, and Sunlight Beaches, no tree canopy, tree damage is not applicable, ghost trees are not applicable, moderate root exposure, and none to a handful of trash found on the campsite.

Variables using a stepwise selection method include: number of trails, number of tent rocks, condition class rating, Northeastern, Southwestern, and Sunlight beaches, none to slight ghost tree damage, and none to a handful of trash found on the campsite. This method produced an R^2 value of 0.425 and a Mallow's C(p) of -2.43 with 8 variables.

The final variable selection method was the LASSO method. This produced the selection criteria values found in Table 2. The 11 variables included in this method were: number of trails, number of tent rocks, condition class rating, Northeastern, Southwestern, and Sunlight beaches, tree damage is not applicable, ghost trees are not applicable, moderate root exposure, and none to a handful or greater than a handful of trash was found on the campsite.

Table 3

Adj. R^2	AIC	SBC
0.359	371.964	266.178

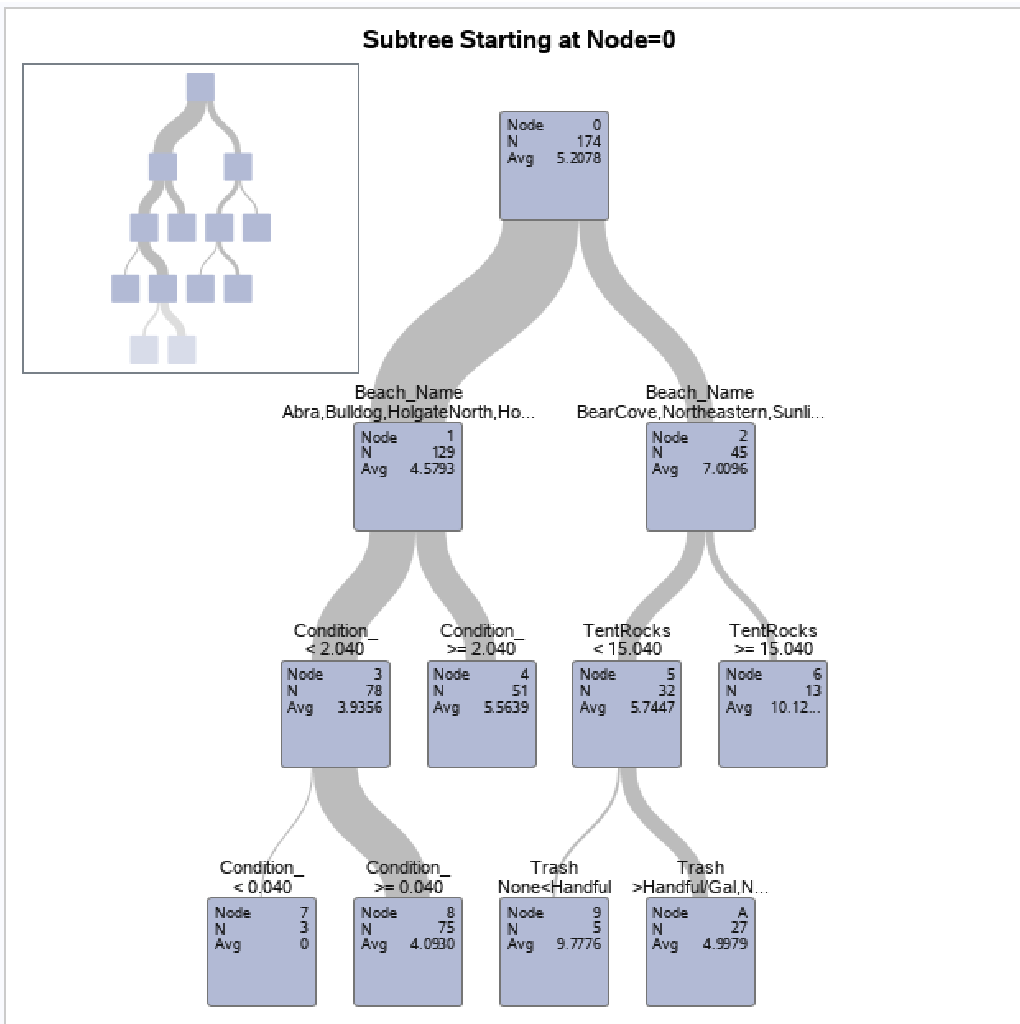
Table 3. Selection Criteria Values for Lasso Method.

Given this information, we believe the LASSO selected variables were the best to use. This is because this method involved a penalized approach and there are fewer variables here than in the model selected based on the 4 selection criteria values. Also, in the stepwise model, a negative Mallow's C(p) value indicates overfitting and is not the best model to use. The fact that majority of the selected variables appear in all three variable selection techniques, is of some comfort. It was suspected that there may be an interaction between the variables trash none to a handful and trash greater than a handful. An additional interaction term was thought to be between the number of trails and a moderate rate of root exposure. After combining the two trash variables, that interaction term did not prove significant and did not improve the model. However, the number of trails and a moderate rate of root exposure did prove significant and improved the model, bumping up the Adj. R^2 to 0.404. Final graphical assumption diagnostics were run for the finalized model with the interaction term. All of the graphs looked identical to those found in Figure 3, with the exception of the Cook's D plot. Suddenly, sample points 15 and 133 shoot up to 0.6, above the appropriate threshold and 0.3 units above the set 0.3 that was the limit in the crude model. In running the model again, excluding points 15 and 133, all of the graphs were identical to those found in Figure 4. The Cook's D plot had a lower limit of 0.3 but points 15 and 133 will remain in the model as they did not influence the models regression fit.

Regression Tree

The final portion of this study was the creation of a regression tree to understand the importance of the explanatory variables in predicting the square root area. The regression tree and its corresponding Table 4, inform us that the average square root value of the mean area for all campsites is 5.208m (27.123m^2). From there, the most important variable that influences area is the beach that the campsite is located on. Beaches Abra, Bulldog, Holgate North, and Holgate

South have a smaller mean square root area of 4.580m than beaches Bear Cove , Northeastern, and Sunlight where the mean square root area of a campsite is 7.010m. Next, the presence and amount of tent rocks is the most significant variable. For the larger beaches only, area increases to a mean square root value of 10.12m when there are more than 15 tent rocks present. The mean square root area decreases to 5.745m when there are less than 15 tent rocks. This makes intuitive sense as campers traveling in larger groups will likely have larger sites with more tents and the need for more tent rocks. Sites with less than 15 tent rocks were split again by the variable, trash. Interestingly, sites with the amount of trash exceeding a handful decreased the mean square root area to 4.998m. Sites with none to a handful of trash present increased their mean square root area to 9.778m.



Model-Based Fit Statistics for Selected Tree

N Leaves	ASE	RSS
7	3.9424	686.0

Variable Importance

	Training		
Variable	*Relative*	*Importance*	*Count*
Beach Name	1.000	14.037	1
Tent Rocks	0.948	13.312	1
Condition	0.813	11.406	2
Trash	0.670	9.817	1
Ghost Tree	0.263	3.689	1

Table 4. Regression tree Variable Importance Table.

For beaches with a smaller area to begin with (Abra, Bulldog, Holgate North, and Holgate South) the condition class rating is then of most importance. Campsites with a rating greater than 2 have an increased mean square root area of 5.564m while sites with a rating less than 2 have a decreased mean square root area of 3.936m. This is logical because higher condition class rating (indicating more obvious use) were given to sites with larger areas. Condition class is then used as a leaf again for those smaller sites that have a condition class rating of less than 2. Sites that were given a rating of 0 had an area of 0m indicating that they recovered and were given the best condition class rating possible. Sites that had a condition class rating between 0 and 2 had an increased mean square root area of 4.093m.

Model Inference and Validation

The full model, with all of the 40 variables had an MSPR of 6.112. The LASSO model had an MSPR of 5.320, and the null model with only the intercept had an MSPR of 7.343. Because it had the lowest MSPR value, the LASSO model was the best model to use. The estimated equation for this model is:

$$\hat{Y} = 6.625 + 0.059X_1 + 0.127X_2 + 0.563X_3 + 2.637X_4 - 1.601X_5 + 3.615X_6 - 0.766X_7 - 0.932X_8 - 2.091X_9 - 3.963X_{10} - 3.774X_{11} + 0.687X_{1,9}$$

This means, with no information about the campsites provided, the square root area measurement would be 6.625m or 43.891m^2. Since the median campsite area of the sample is 19.593 m^2, less than half of this estimate, this implies that the indicator variables have a heavy influence on the size of the area. With all other variables held constant, with each increase in trail numbers, the square root area increases by 0.059m. Again, with all other variables held constant, if the campsite is located on Southwestern beach, it's square root area decreases by 1.601m. Like the other previously explained variables, with all other variables held constant, the square root area of a campsite increases by 0.127m with each additional tent rock, 0.563m with a one point increase in the condition class rating, 2.637m if the campsite is located on Northeastern beach, and 3.615m if the campsite is located on Sunlight beach. A campsite decreases in square root area by 0.766m if there are no trees to receive damage, 0.932m when there are no ghost trees to be damaged, 2.091m with moderate root exposure, 3.963m if the trash found on the site exceeds a handful, and 3.77m if there is less than a handful of trash found on the site. The number of trails and moderate root exposure interaction term increases campsite square root area by 0.687m. First the number of trails needs to be multiplied by 0.059 and then if there is moderate root exposure multiply 0.059 by -2.091. That value is then multiplied by 0.687 to determine the square root area estimate.

Conclusions

This study examined indicator variables that influence campsite size in remote areas of Kenai Fjords National Park. The increased number of trails, number of tent rocks, condition class ratings, sites located on Northeastern and Sunlight beach, and the combination of root exposure and number of trails increase campsite size. Campsites decrease when there are no trees or ghost trees, moderate root exposure, trash, and on Southwestern beach. These coefficients and their positive or negative effect on campsite area made logical sense. Beaches Northeastern and Sunlight had the highest average campsite size, thus; it made sense they had the largest positive coefficients adding to the square root area estimates. As previously discussed, an increase in the number of tent rocks is indicative of an increased number of tents located on the site, increasing the overall area of the site.

Conversely, campsites on Southwestern beach decreasing the area site also makes sense as Southwestern beach has the smallest average campsite size. Additionally, there being no damage to surrounding trees or ghost trees, perhaps even no trees at all, could decline campsite area if the site is located on cobble stone making evidence of human influence harder to identify. The variable trash decreasing the size of a campsite is puzzling, and we believe this may be due to the measurement scale used to survey trash. The options a surveyor chose from included “none less than a handful”, “none to a handful”, or “greater than a handful to a gallon”. With “none” being listed in two options makes it more difficult to select the most accurate description of how much trash was actually on the site. For future studies using this method, we recommend changing the scale to “none”, “less than a handful”, or “greater than a handful” to avoid confusion.

Results from research such as this, provide park managers the information they need to determine how best to mitigate recreation ecological damage. Management suggestions might include; designating campsites, moving to a reservation system for use, promoting dispersed camping use. This analysis has garnered evidence to improve existing campsite monitoring protocols. Furthermore, we believe the ecological conclusions brought forth in this paper can be extended to public land campsites, off of a road system, in temperate coastal rainforests. Campsites in British Columbia, Canada and the Pacific Northwest region of the United States of America come to mind.

Future research would include continued longitudinal monitoring of these campsites. To have data that spans decades and has consistent sampling periods would provide a more robust analysis. Research such as this should also expand to other locations. By conducting this study in areas with similar biomes could corroborate and support these results or provide new insight into variables that impact campsite size. All future research would be beneficial in understanding recreational ecological impacts.

References

Cole, D.N. and T. E. Hall. 1992. Trends in campsite condition: Eagle Cap Wilderness, Bob Marshall Wilderness, and Grand Canyon National Park. Res. Pap. INT-453 Ogden: UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Hampton, B. and Cole, D. 2003. Soft paths: how to enjoy the wilderness without harming it (3rd ed.). Harrisburg, PA: Stackpole Books.

Leung, Y. and J. L. Marion. 1999. Characterizing backcountry camping impacts in Great Smoky Mountains National Park, USA. Journal of Environmental Management. 57(3): 193-203.

Phillips, L. M., F. Klasner, J. Cusick, and C. Monz. 2015. Kenai Fjords National Park Coastal Campsite Monitoring: 2008-2012. NPS/KEFJ/NRR—2015. National Park Service, Fort Collins, Colorado.

National Parks Service. 2019 "Stats Report Viewer." U.S. Department of the Interior. https://irma.nps.gov/Stats/SSRSReports/Park%20Specific%20Reports/Overnight%20Stays%20(1979%20-%20Last%20Calendar%20Year)?Park=KEFJ

Appendix

```
FILENAME REFFILE '/home/u45031692/my_courses/Final_Project/Campsite_Sampling.csv';

PROC IMPORT DATAFILE=REFFILE replace
	DBMS=CSV
	OUT=WORK.campsites;
	GETNAMES=YES;
RUN;

/* Check hard upper bound on area */
proc univariate data=campsites;
	var Area_Calculation;
	histogram Area_Calculation;
run;

/* Add ID variable where we sort by the Campsite Area */
proc sort data=campsites; by Area_Calculation;
data campsites; set campsites; ID = Tag_Number;

/* Plot the Campsite Areas. */
proc sgplot data=campsites;
	scatter x=ID y=Area_Calculation;
run;

/*Boxplot of Area by Beach*/
proc sort data=campsites out=sort_beach; by Beach_Name;
proc boxplot data=sort_beach;
	plot Area_Calculation*Beach_Name/
	boxstyle=schematic boxwidth=10 haxis=axis1
	cboxfill=lavender cboxes=bgr;
run;

/*Log transform Area to make it look more normal*/
data campsites2; set campsites;
	log_area = log(Area_Calculation);
run;
/*Not chosen because of zeros in data*/

/*Log + 1 transform Area to make it look more normal and include recovered areas*/
data campsites3; set campsites2;
	log_area1 = log(Area_Calculation) + 1;
run;
/*Did not fix problem with zeros*/

/*Square Root Transformation of Area*/
```

```
data campsites4; set campsites3;
	sqrt_area = sqrt(Area_Calculation);
run;

/* Check hard upper bound on area */
proc univariate data=campsites4;
	var sqrt_area;
	histogram sqrt_area;
run;

/*Model with sqrt Area data*/
proc glmmod data=campsites4 outdesign=GLMDesign outparm=GLMParm NOPRINT;
	class Beach_Name Tree_Canop Tree_Damag Ghost_Tree Root_Expos Trash
HumanWaste;
	model sqrt_area = Vegetation_Loss Mineral_So Num_Stumps Num_Ghost_
Num_Fire_S
	Num_Trails TentRocks Condition_ Beach_Name Tree_Canop Tree_Damag Ghost_Tree
Root_Expos
	Trash HumanWaste;
run;

/* Separate Into Training and Test Sets.
Only Fit Models to the Training Set. The variable
"Selected" separates training (0) from test (1) */

proc surveyselect data=GLMDesign seed=12345 out=campsites6
	rate=0.2 outall; /* Withold 20% for validation */
run;

data train2; set campsites6;
if Selected = 0;
run;

data test2; set campsites6;
if Selected = 1;
run;

/*Adding ID column*/
data train3; set train2;
	ID = _n_;
run;

/*Square Root Crude Model*/
proc reg data = train3 plots(label)=(CooksD RStudentByLeverage DFFITS DFBETAS);
	ID;
	model sqrt_area = COL1-COL40;
```

```
   title1 'Crude Model Square Root Area';
run;

/*Square Root Variable Selection*/
proc reg data = train3 plots(label)=(CooksD RStudentByLeverage DFFITS DFBETAS);
        ID;
        model sqrt_area = COL1-COL40
                / selection = AdjRSq Cp AIC SBC;
   title1 'Compare Selection Criteria';
run;

/*Stepwise Selection*/
proc reg data = train3 plots(label)=(CooksD RStudentByLeverage DFFITS DFBETAS);
        ID;
        model sqrt_area = COL1-COL40
            / selection = stepwise slentry = .10 slstay = .10;
   title1 'Stepwise Selection';
run;

/* lasso variable selection*/
proc glmselect data=train3 plots=(criterion ase);
model sqrt_area = COL1-COL40
        / selection=lasso(adaptive choose=sbc stop=none);
run;

/*Removing point 15 and 133*/
data train4; set train3;
        if ID=15 then delete;
        if ID=133 then delete;
run;
/*Square Root Lasso Model without 15 and 133*/
proc reg data = train4 plots(label)=(CooksD RStudentByLeverage DFFITS DFBETAS);
        ID;
        model sqrt_area = COL7-COL9 COL17 COL22 COL23 COL28
        COL31 COL33 COL37 COL38;
   title1 'Square Root Area Lasso Model';
run;

/*Test for Multicollinearity*/
proc reg data=train5 plots=(CooksD RStudentByLeverage DFFITS DFBETAS);
        model sqrt_area = COL7-COL9 COL17 COL22 COL23 COL28
        COL31 COL33 COL37 COL38 / vif collin;
store regModel;
run;

/* Look at scatterplot matrix */
```

```
/*Only Trash Beaches and Tree Damage and Ghost Trees*/
proc sgscatter data=train5;
matrix sqrt_area COL17 COL22 COL23 COL28
        COL31 COL37 COL38/
        markerattrs=(symbol=CIRCLEFILLED size=6pt);
title1 'Campsite Data';
run;

/* Look at correlation matrix */
proc corr data=train5;
var sqrt_area COL7-COL9 COL17 COL22 COL23 COL28
        COL31 COL33 COL37 COL38;
title1 'Correlation Matrix: Campsite Data';
run;

/*Creating Interaction terms*/
/*Trash none-handful and >handful to gallon*/
/*Trails and Root Exposure*/
data train5; set train3;
        trash= Col37 * Col38;
        trail_root = COL7 * COL33;
run;

/*Trash Interaction*/
proc reg data=train5 plots=(CooksD RStudentByLeverage DFFITS DFBETAS);
        model sqrt_area = COL7-COL9 COL17 COL22 COL23 COL28
        COL31 COL33 COL37 COL38 trash;
        title1 'Is Trash None to Handful and Trash Greater than Hnadful Interacting Terms?';
run;
/*Interaction terms are not significant*/

/*Trails and Root Exposure*/
proc reg data=train5 plots=(CooksD RStudentByLeverage DFFITS DFBETAS);
        model sqrt_area = COL7-COL9 COL17 COL22 COL23 COL28
        COL31 COL33 COL37 COL38 trail_root;
        title1 'Are Trails and Root Exposure Interacting Terms?';
run;

/*BEST MODEL TO USE*/
proc reg data=train5 plots(label)=(CooksD RStudentByLeverage DFFITS DFBETAS);
        ID;
        model sqrt_area = COL7-COL9 COL17 COL22 COL23 COL28
        COL31 COL33 COL37 COL38 trail_root;
        title1 'Square Root Area Lasso Model with Interaction Term';
run;
```

```
/* regression tree */
proc hpsplit data=campsites4 seed=123 maxdepth=15 maxbranch=2;
	class Beach_Name Tree_Canop Tree_Damag Ghost_Tree Root_Expos Trash
HumanWaste;
	model sqrt_area = Vegetation_Loss Mineral_So Num_Stumps Num_Ghost_
Num_Fire_S
	Num_Trails TentRocks Condition_ Beach_Name Tree_Canop Tree_Damag Ghost_Tree
Root_Expos
	Trash HumanWaste;
output out=out2;
run;

/* regression tree no beach name */
/*This one does look a bit better*/
proc hpsplit data=campsites4 seed=123 maxdepth=15 maxbranch=2;
	class Tree_Canop Tree_Damag Ghost_Tree Root_Expos Trash HumanWaste;
	model sqrt_area = Vegetation_Loss Mineral_So Num_Stumps Num_Ghost_
Num_Fire_S
	Num_Trails TentRocks Condition_ Tree_Canop Tree_Damag Ghost_Tree Root_Expos
	Trash HumanWaste;
output out=out3;
run;


/* Calculate MSPR (ASE) */
/*Full Model*/
proc reg data = train3 plots(label)=(CooksD RStudentByLeverage DFFITS DFBETAS);
	ID;
	model sqrt_area = COL1-COL40;
store regModel;
run;

data test; set test;
	sqrt_area = sqrt(Area_Calculation);
run;

proc plm restore=regModel;
score data=test out=newTest predicted;
run;

data newTest; set newTest;
ASE = (sqrt_area - Predicted)**2;
run;

proc means data = newTest;
```

```
var ASE;
run;

/*Lasso Model*/
proc reg data=train5 plots(label)=(CooksD RStudentByLeverage DFFITS DFBETAS);
	ID;
	model sqrt_area = COL7-COL9 COL17 COL22 COL23 COL28
	COL31 COL33 COL37 COL38 trail_root;
	title1 'Square Root Area Lasso Model with Interaction Term';
store regModel2;
run;

data test2; set test;
	trail_root = COL7 * COL33;
run;

proc plm restore=regModel2;
 score data=test2 out=newTest2 predicted;
 run;

data newTest2; set newTest2;
ASE = (sqrt_area - Predicted)**2;
run;

proc means data = newTest2;
var ASE;
run;

/*Null Model*/
proc reg data=train3 plots=(CooksD RStudentByLeverage DFFITS DFBETAS);
 model sqrt_area = ;
store regModel3;
run;

proc plm restore=regModel3;
 score data=test out=newTest3 predicted;
 run;

data newTest3; set newTest3;
ASE = (sqrt_area - Predicted)**2;
run;

proc means data = newTest3;
var ASE;
run;
```