Midterm

RJ Cass

Abstract

We analyzed data of the total Basal Area of trees in a plot and their corresponding environmental factors. We found that the factors that most influence tree growth are Elevation, Aspect, Longitude, and Latitude. We found that the combination of factors most conducive to tree growth was a Longitude and Latitude of -109.34, 40.99 (the maximum values for each), an elevation of 9755.8, and an aspect of 360 (facing North), with slope not being significant. We created a GAM model using these factors to generate predcitions of total Basal Area for other plots not surveyed by the FIA, which are shown in Table 4 in the Appendix.

1: Introduction

We have a dataset representing the total basal size of all Lodgepole Pines within a plot, along with various environmental factors that may contribute to tree growth. We want to use these data to answer the following questions:

1) How is lodgepole growth is affected by its environment?

2) What environments are conducive to lodgepole growth?

3) Given the environmental data for areas the FIA was not able to survey, what is the predicted basal area of the trees in that area?

Examining the data, the output variable (Basal Area) does not appear to be normally distributed (see Figure 1). As such, the normality assumption required for many modeling methods is not met and, if not addressed, generated models will not properly reflect the relationships between the input and output variables, and will have reduced predictive power. To account for this, we examined a log and a square root transform, and found that the square root transform better fits the normality assumptions (also shown in Figure 1). Thus, in models where normality is assumed, we will use the square root transform of Basal Area.

We also identified that several of the factors appear either to be non-linear, or possibly non-significant in determining Basal Area. Shown in Figure 2 are the scatter-plots of Basal Area plotted against each factor.

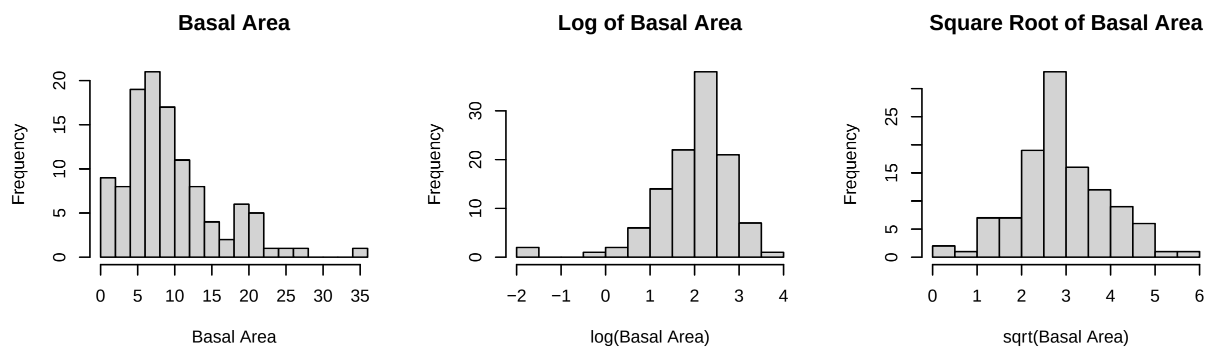

Figure 1: *Histograms of Basal Area (left) and its Log (middle) and Square Root (right) transforms. Note the skewness in the left and middle graphs.*

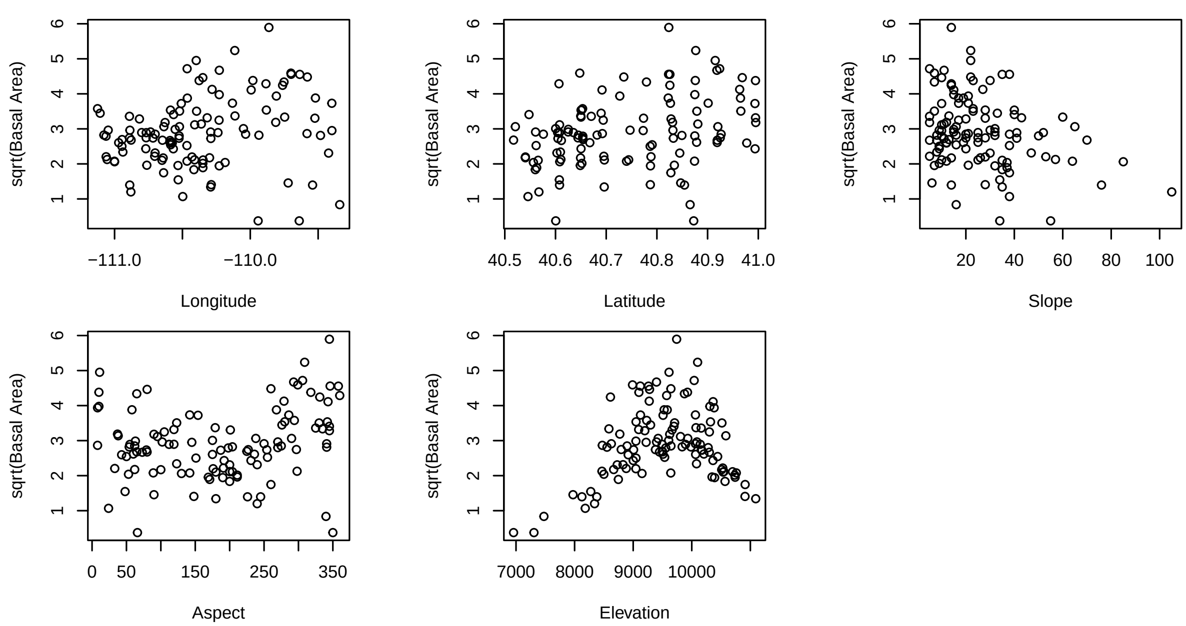

Figure 2: *Scatterplots of the square root of Basal Area vs. each explanatory factor. Note that Aspect and Elevation appear to be non-linear relationships.*

2: Methodology

2.1 Proposed Models

Model 1

The first model considered was a Generalized Additive Model (GAM). To determine which variables to include we fit an inital GAM model using all factors to the square root of Basal Area. Reviewing the plots in Figure 2, we chose Longitude, Latitude, and Slope to be linear relationships. We chose to fit Elevation using a natural spline because it forces linearity outside the range of given data which will help with better predictive power for values outside our range. We chose to fit Aspect using a basis spline because Aspect is bounded between 0 and 360 so we do not need to worry about values outside that range. Reviewing the p-values, we found that with an $\alpha = .05$ Slope was not significant so we removed it from the model.

GAM models require the LINE assumptions of Linearity (in the case each variate is linear compared to the rest of the model), Independence, Normality, and Equal Variance. If these conditions are not met, the model will not accurately predict, nor represent the correct explanatory relationships between values. This model will allow us to answer the research questions by identifying what combination of factors yields the highest Basal Area, as well as provide predictions for land plots outside of the measured values.

Model 2

For our second model we created a LOESS (locally estimated scatterplot smoothing) model. Given a LOESS model is normally limited to 4 factors due to increased complexity with higher dimensionality, and we previously found that Slope was not significant, we chose to exclude Slope from this model as well. LOESS models do not have any assumptions that need to be met. This model can be used to answer the research questions by identifying which combinations give the largest Basal Area, and by predciting for land plots outside the measured values.

2.2 Model Evaluation and Selection

To compare the models we performed Leave-One-Out Cross Validation (LOOCV) and calculated the RMSE of each model, as shown in Table 1. The GAM model yielded an In Sample R^2 value of 0.814. The LOESS model, being non-parametric, does not give a standard R^2 value so we will compare the models using the LOOCV RMSE. The RMSE of the two models are very close and their confidence intervals overlap. However, the GAM model is much more interpretable due to being a direct transform of the explaining factors. Thus, we will use the GAM model to answer our research questions.

Table 1

Model	LOOCV RMSE
GAM	8.523 ± 1.875
LOESS	8.438 ± 1.861

The GAM model follows this format with the calculated β values shown in Table 3 in the Appendix:

$$g(\mathbb{E}(\sqrt{BasalArea})) = \beta_0 + \beta_1 Longitude + \beta_2 Latitude + \sum_{k=1}^{3} \beta_k b_k(Aspect) + \sum_{k=1}^{3} \beta_k n_k(Elevation)$$

As shown in Figure 1, we performed a square root transform on Basal Area to meet the normality assumptions. Independence (that one plot of trees does not impact another plot of trees) is assumed . Figure 3 outlines the Linearity (left) and Equal Variance (right) assumptions. The left plot appears to have some shape/pattern to it, indicating that Linearity may not be a clear assumption. However, we believe it to be sufficient for this analysis. The variance of points in the line on the right plot appear equal, indicating the Equal Variance assumption is met.

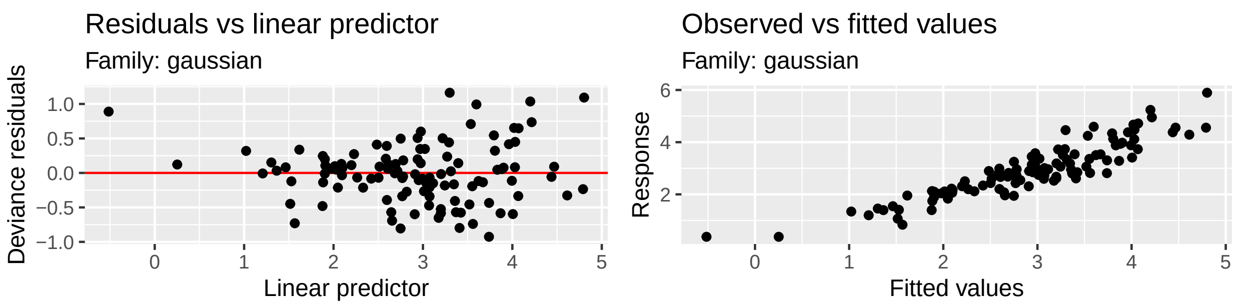

Figure 3: *Left - residual plot, note the change in variance across the x-axis. Right - fitted vs. actual values, note the variance appears equal over the fit*

3: Results

We can use the GAM model identified previously to answer our research questions:

1. The significant factors in determining Basal Area using the data provided are Longitude, Latitude, Aspect, and Elevation. Provided in Figure 4 are the plotted model fits of each significant explanatory value. Because Aspect is a measure of degrees from North, it is a little easier to understand when represented in terms of cardinal directions, as shown in the same plot.

2. Using our constructed model we evaluated points over the full range of provided data and identified the conditions most conducive to tree growth. These values are shown in Table 2.

Table 2

Longitude	Latitude	Slope	Aspect	Eelevation
-109.34	40.99	1	360	9755.76

3. We also used the model to generate predicted values for plots of land not surveyed by the FIA. These values are provided in Table 4 in the Appendix.

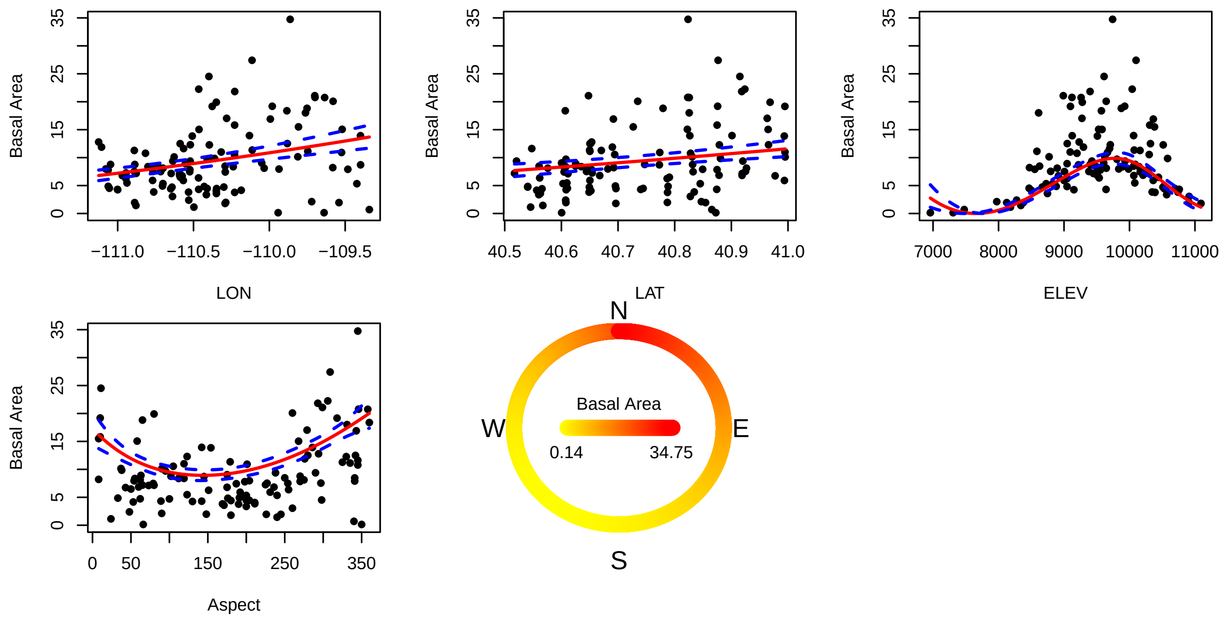

Figure 4: *Plots showing the model fit of each variable. Each plot holds all other variables constant at their mean value. Because Slope was excluded from the model it is just flat so we excluded it here. The last plot shows the fitted Aspect in terms of cardinal direction.*

4: Conclusions

We used the Lodgepole data to understand the relationship between various environmental factors and the Basal Area of plots of Lodegpole pines. Of the explanatory factors, we identified that Slope was not significant in determining Basal Area so we created a GAM model consisting of Longitude, Latitude, Aspect, and Elevation as related to the square root of Basal Area. We used this model to identify that the conditions most conducive to Lodgepole growth are a Longitude and Latitude of (-109.34, 40.99), an Aspect of 360 (facing North), and an Elevation of 9755.76 (with Slope not being significant). We also used the model to generate predicted values for land plots not visited by the FIA, as shown in Table 4.

We believe that the significance of Latitude and Longitude is likely driven due to other factors such as an interaction with elevation, or changing temperatures as you move in a certain direction, etc. A future step for this analysis would be to identify any interaction among explanatory factors. We also suggest further study into other environmental factors that would impact tree growth, such as distance to the nearest water source.

Appendix

Table 3: *Table showing the coefficients of the chosen model. It is important to remember that these coefficients come from a model of the square root of Basal Area.*

Parameter	Coefficient	CI_low	CI_high
(Intercept)	13.5620137	-22.39933758	49.5233650
LON	0.6116082	0.41394341	0.8092730
LAT	1.3064051	0.69683842	1.9159718
bs(Aspect)1	-1.9333634	-2.85167880	-1.0150479
bs(Aspect)2	-0.7828835	-1.37540043	-0.1903667
bs(Aspect)3	0.4552546	-0.02012635	0.9306354
ns(ELEV, df = 3)1	3.7474787	3.32810833	4.1668491
ns(ELEV, df = 3)2	7.3697091	6.05789836	8.6815199
ns(ELEV, df = 3)3	0.3409309	-0.11168763	0.7935493

Table 4

LON	LAT	Slope	Aspect	ELEV	Lodgepole
-109.5198	40.87226	6	306	7430	1.028137202
-110.9515	40.55974	6	90	9874	6.735208355
-110.9954	40.56292	6	156	9278	4.719442174
-110.6357	40.87199	6	70	10880	3.153959852
-110.7610	40.69594	7	327	10824	6.068041911
-109.5815	40.86886	8	43	7366	0.184081470
-110.4061	40.83268	8	3	11151	4.187557170
-110.4822	40.51814	9	111	8474	1.734507356
-109.5704	40.55577	10	250	6128	9.449841891
-109.9319	40.82875	10	352	10601	15.178192935
-110.9980	40.51779	11	260	9821	8.406509605
-111.0560	40.56657	12	170	9125	3.988007665
-111.1230	40.83439	12	303	9309	11.414640090
-110.2258	40.65166	13	151	10627	4.117587314
-110.7646	40.51741	14	254	10105	8.059824926
-109.4653	40.68576	15	212	8184	3.032943452
-109.7064	40.59679	17	205	8798	6.476771407
-109.9927	40.83059	17	280	11036	4.628703858
-110.1713	40.73923	17	300	11007	4.743006491
-109.7558	40.55485	20	175	6982	2.296745594
-110.9438	40.92702	20	335	8909	11.907304616
-110.5285	40.78882	20	87	10816	3.111740565
-110.8764	40.68762	20	100	10081	6.948048284
-109.9360	40.69535	21	75	10605	6.504532065
-109.6455	40.55103	22	192	6546	6.024550684

(continued)

LON	LAT	Slope	Aspect	ELEV	Lodgepole
-110.7659	40.56251	22	17	9609	11.803366747
-109.9384	40.91927	22	174	9225	10.651474995
-111.0192	40.68235	22	151	10240	5.143457661
-109.4018	40.86244	23	338	7755	4.620680957
-110.8768	40.74240	23	8	10304	11.371865979
-109.3462	40.64482	24	160	6761	2.943244377
-109.9273	40.55387	25	80	7379	0.249763493
-111.0590	40.69536	25	290	9749	11.265867211
-111.0023	40.70136	25	320	9944	13.328234834
-110.4913	40.61982	25	96	10943	1.481387352
-110.0470	40.73850	25	220	10739	5.045371890
-110.1708	40.87685	25	123	11147	1.349052676
-109.3440	40.60056	28	192	6280	8.130401967
-111.1161	40.55838	28	187	7530	0.980619635
-110.2923	40.65113	28	22	10755	6.358859068
-109.6829	40.90886	30	2	6422	1.632258256
-109.8083	40.96409	30	16	7831	3.431203258
-110.0642	40.68469	30	137	10715	3.921821157
-109.7430	40.93585	31	359	6778	0.005091188
-109.6443	40.68962	31	246	9222	12.347845708
-110.2293	40.56296	32	165	8359	1.647648423
-109.9427	40.96765	33	347	9336	22.514149227
-109.8191	40.68913	33	138	10217	9.243685393
-109.4612	40.64508	34	256	7524	0.281112608
-111.0617	40.88088	35	230	8769	4.779396011
-109.5109	40.96192	37	241	6402	3.639945427
-110.9359	40.52199	37	344	9805	14.000623640
-109.4547	40.82271	38	140	7902	1.290106877
-109.6761	40.93287	40	142	6246	7.860386483
-111.0643	40.92369	40	158	7673	0.049375731
-110.1114	40.60659	40	218	7844	0.304951803
-110.3880	40.54058	40	55	9153	7.781217387
-109.8784	40.91685	42	72	8457	5.709489227
-109.5247	40.73060	42	244	8439	6.066635029
-109.9880	40.91924	45	210	9538	13.231604788
-110.0552	40.92061	45	165	10000	11.773868358
-110.8227	40.56215	49	311	10202	10.642406409
-109.4527	40.87241	50	50	6980	0.246257696
-110.1679	40.59570	50	331	10010	16.778760304
-109.9209	40.93394	51	255	8068	3.366622923
-109.9954	40.60632	51	355	8928	14.801476990

(continued)

LON	LAT	Slope	Aspect	ELEV	Lodgepole
-110.3487	40.60104	53	223	8225	1.529135666
-110.4628	40.69858	55	355	10674	10.743801766
-109.4559	40.91507	60	320	6069	4.196680645
-109.8199	40.65146	60	73	8896	7.791459132
-110.8226	40.78491	60	147	10381	5.317253303
-109.6055	40.93588	65	160	6738	2.401893411
-110.1774	40.64587	70	65	8071	1.318880762
-110.7088	40.83536	74	336	10595	10.725607625
-109.6448	40.64241	75	118	7809	0.371635433
-110.9982	40.78764	75	64	8241	1.412278916
-110.7017	40.56653	75	114	8685	2.565330352
-110.9391	40.74638	80	220	9544	8.347995522