Energy Efficiency of Buildings Designed in Ecotect

Icy Hot
Nathan Becker | Levi Lamming | Thoms Verzino

Stat 5346/CS 5204
Professor Olbricht
4/28/2018

Background

This data set explores how the efficiency of heating and cooling a home is affected by the building characteristics of that home. To test this, the creators of the dataset designed buildings in a simulator called Ecotect. Finding a model to fit this data could help indicate which features of a home affect the heating and cooling costs most significantly.

Data

768 different building shapes and characteristic combinations were observed, where changes were made to a number of different characteristics. The two response variables are the Heating Load and Cooling Load needed to maintain internal climate. The predictors are Relative Compactness, Surface Area, Wall Area, Roof Area, Height, Orientation, Glazing Area, and Glazing Area Distribution. All predictors are numerically measured.

Statistical Methods

Using SAS Studio, we performed multiple linear regression on our data using the procedure reg. We constructed two models, seperate ones for heating and cooling loads. We used /vif to determine which variables show multicollinearity and could be removed from our model selections. We used /selection=rsquare in the procedure to apply the best subset method, to determine which variable combination creates the best model. We used proc transreg to run a Box-Cox Analysis to determine an appropriate approximation for our response variables.

Results

The Scatterplot Matrix shown in Figure 1 reveals Heating Load and Cooling Load are highly correlated. This is expected because maintaining an internal climate is the essence of both responses. This matrix also shows how relative compactness and surface area are very related which is expected. We will need to test for multicollinearity and likely will remove one of these.

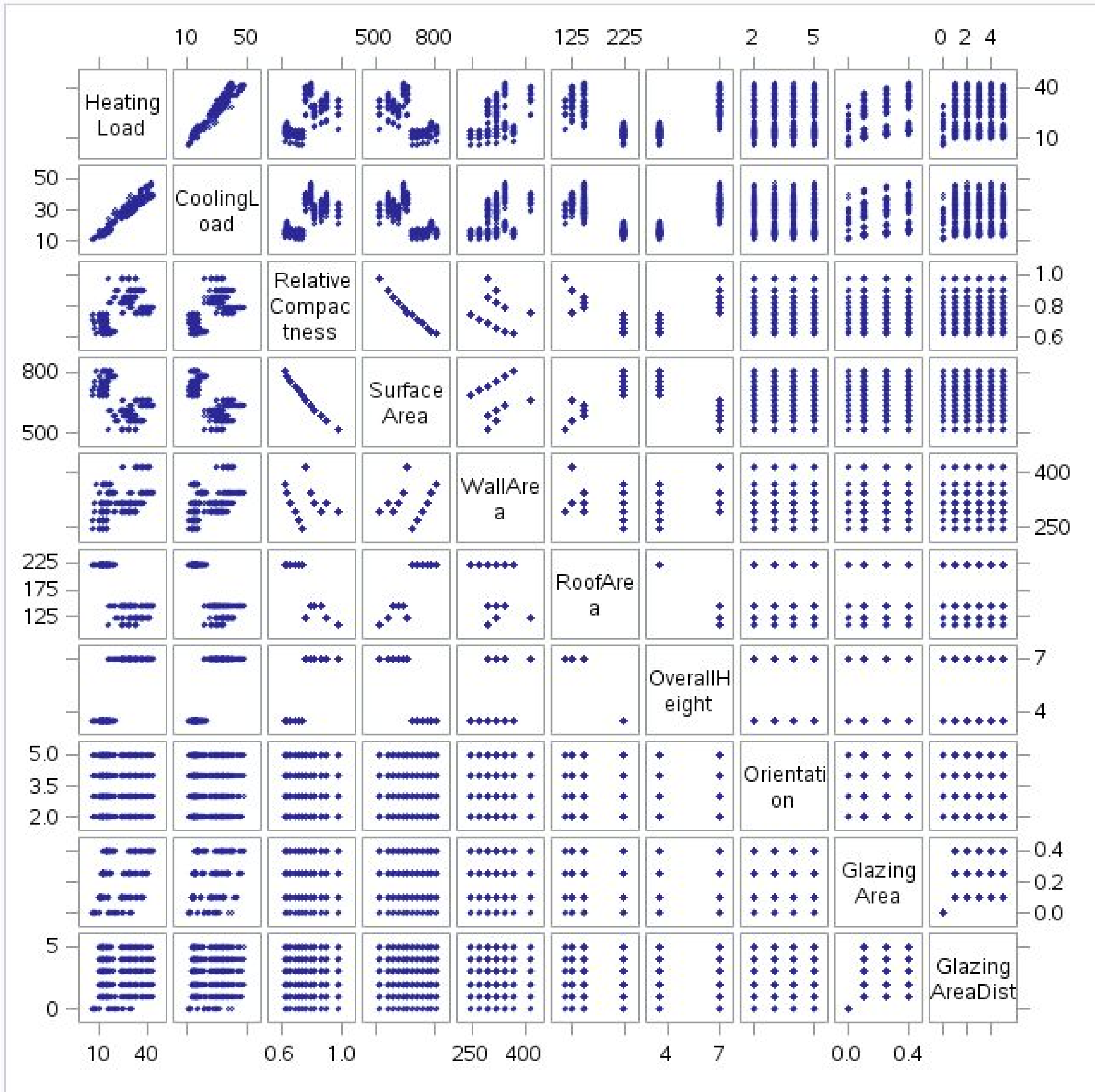

Figure 1: Scatterplot Matrix

Surface Area is extremely correlated with Relative Compactness, which is especially easy to see in Table 1. For this reason, SAS removes Surface Area because this causes the model to be not full rank. The predictors having the highest correlation with the response variables are Relative Compactness, Surface Area, Wall Area, Roof Area, and Overall Height.

Table 1: Correlation Matrix

Pearson Correlation Coefficients, N = 768										
	HeatingLoad	CoolingLoad	RelativeCompactness	SurfaceArea	WallArea	RoofArea	OverallHeight	Orientation	GlazingArea	GlazingAreaDist
HeatingLoad HeatingLoad	1.00000	0.97586	0.62227	-0.65812	0.45567	-0.86183	0.88943	-0.00259	0.26984	0.08737
CoolingLoad CoolingLoad	0.97586	1.00000	0.63434	-0.67300	0.42712	-0.86255	0.89579	0.01429	0.20750	0.05053
RelativeCompactness RelativeCompactness	0.62227	0.63434	1.00000	-0.99190	-0.20378	-0.86882	0.82775	0.00000	0.00000	0.00000
SurfaceArea SurfaceArea	-0.65812	-0.67300	-0.99190	1.00000	0.19550	0.88072	-0.85815	0.00000	0.00000	0.00000
WallArea WallArea	0.45567	0.42712	-0.20378	0.19550	1.00000	-0.29232	0.28098	0.00000	0.00000	0.00000
RoofArea RoofArea	-0.86183	-0.86255	-0.86882	0.88072	-0.29232	1.00000	-0.97251	0.00000	0.00000	0.00000
OverallHeight OverallHeight	0.88943	0.89579	0.82775	-0.85815	0.28098	-0.97251	1.00000	0.00000	0.00000	0.00000
Orientation Orientation	-0.00259	0.01429	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000
GlazingArea GlazingArea	0.26984	0.20750	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000	0.21296
GlazingAreaDist GlazingAreaDist	0.08737	0.05053	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.21296	1.00000

Running the models with VIF indicates that multicollinearity exists, as seen in Table 2. Relative Compactness and Surface area are extremely high, and greater than 10. Overall Height may also be also have multicollinearity with Roof Area. When we decide to remove Roof Area and Surface Area, all Variable Inflation Factors are less than 10, as shown in Table 3.

Table 2: Parameter Estimates with Variance Inflation Factors

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	B	84.01342	19.03361	4.41	<.0001	0
RelativeCompactness	RelativeCompactness	B	-64.77343	10.28945	-6.30	<.0001	105.52405
SurfaceArea	SurfaceArea	B	-0.08729	0.01708	-5.11	<.0001	201.53113
WallArea	WallArea	B	0.06081	0.00665	9.15	<.0001	7.49298
RoofArea	RoofArea	0	0	.	.	.	.
OverallHeight	OverallHeight	B	4.16995	0.33799	12.34	<.0001	31.20547
Orientation	Orientation	1	-0.02333	0.09470	-0.25	0.8055	1.00000
GlazingArea	GlazingArea	1	19.93274	0.81399	24.49	<.0001	1.04751
GlazingAreaDist	GlazingAreaDist	1	0.20378	0.06992	2.91	0.0037	1.04751

Table 3: Parameter Estimates with Variance Inflation Factors - No Roof Area or Surface Area

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	-12.32617	2.71031	-4.55	<.0001	0
RelativeCompactness	RelativeCompactness	1	-14.53240	3.09634	-4.69	<.0001	9.25028
WallArea	WallArea	1	0.03498	0.00439	7.97	<.0001	3.16193
OverallHeight	OverallHeight	1	5.60675	0.19080	29.39	<.0001	9.62610
Orientation	Orientation	1	-0.02333	0.09626	-0.24	0.8086	1.00000
GlazingArea	GlazingArea	1	19.93274	0.82732	24.09	<.0001	1.04751
GlazingAreaDist	GlazingAreaDist	1	0.20378	0.07106	2.87	0.0043	1.04751

We decided to run models with Surface Area and Roof Area removed. However, Figure 2 and Figure 3 display that these models showed non-constant variance. This indicated we should run a Box-Cox Analysis, seen in Figure 3 and Figure 4. The Box-Cox Analysis indicated a power transformation of natural log should be applied to the response variables. The Box-Cox for Heating Load and Cooling Load both indicated 0.

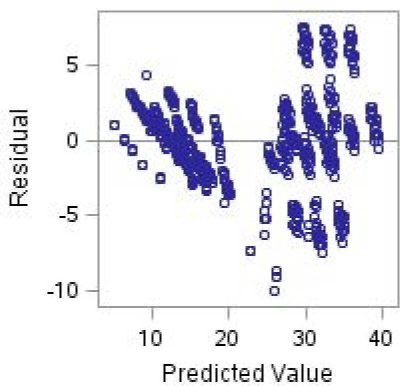

Figure 2: Heating Residuals vs. Predicted Values

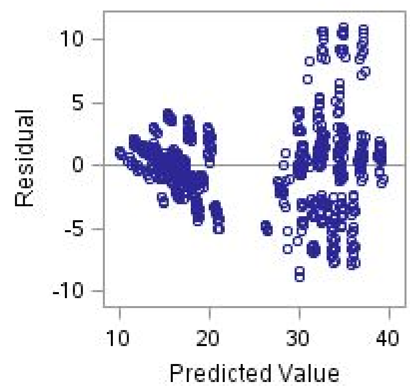

Figure 3: Cooling Residuals vs. Predicted Values

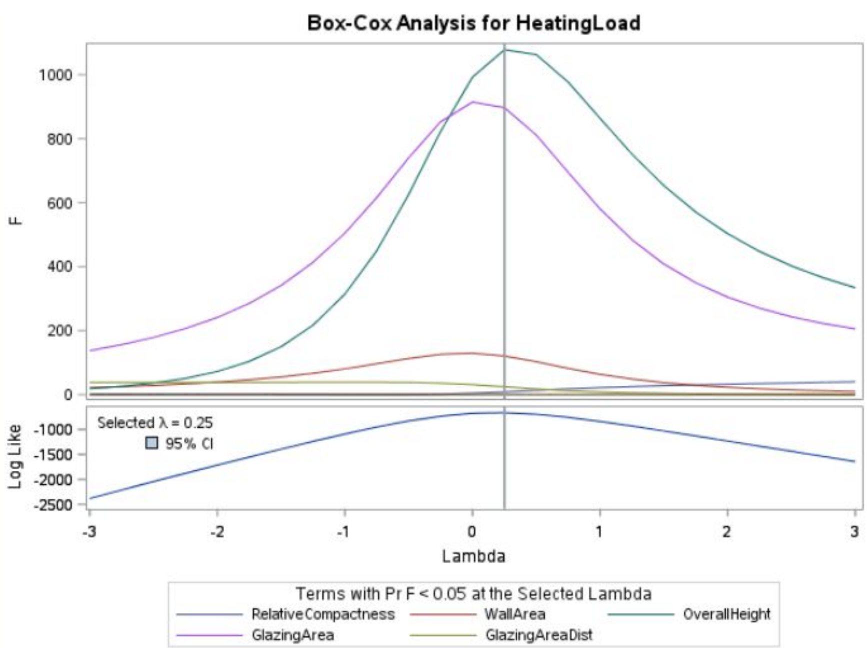

Figure 3: Box-Cox Analysis for Heating Load

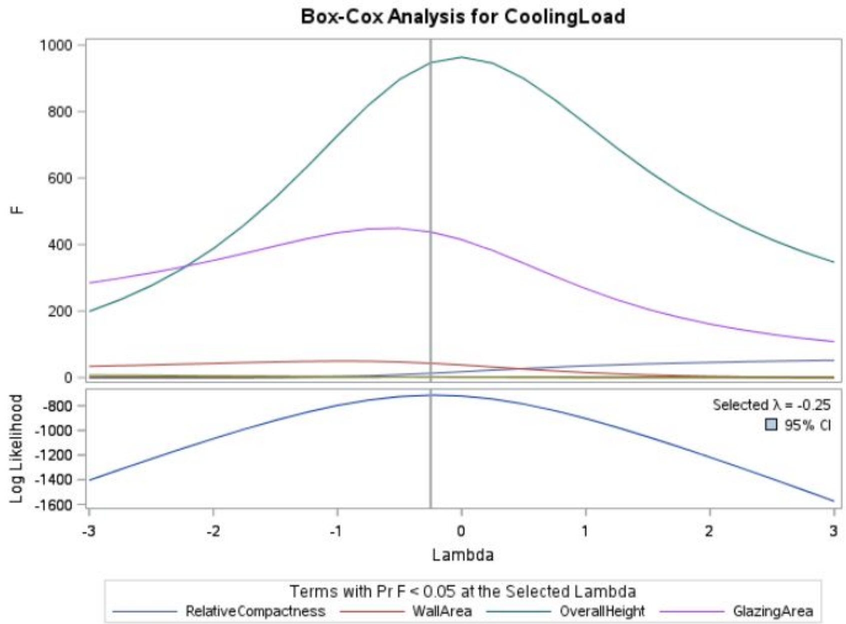

Figure 4: Box-Cox Analysis for Cooling Load

After applying a natural log transformation to both Heating Load and Cooling Load, the variance became more constant. Normality also improved after applying the transformation because the Residual vs. Quantile Plot had less curvature and became more linear. Error independence is not applicable to this study because the homes aren't dependent on each other. The results of this can be seen in Figure 5 and Figure 6.

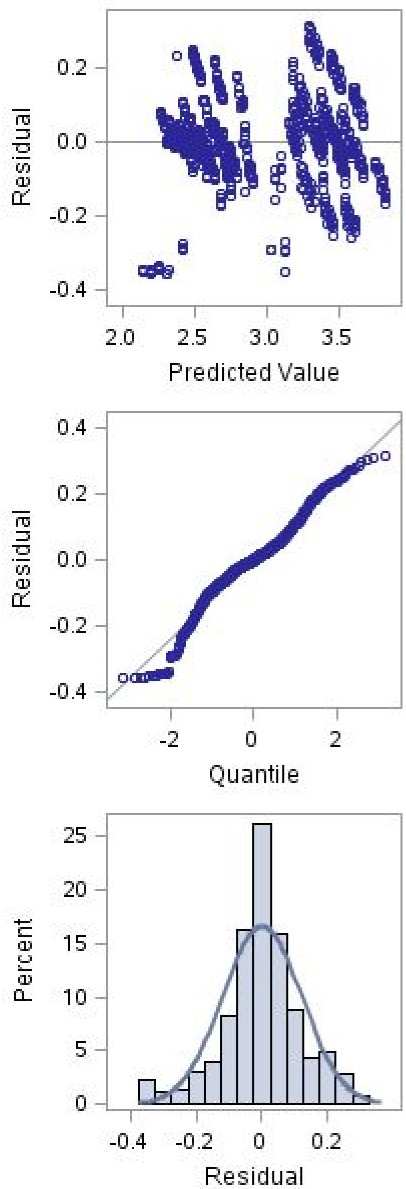

Figure 5: Heating Load Residual Plots

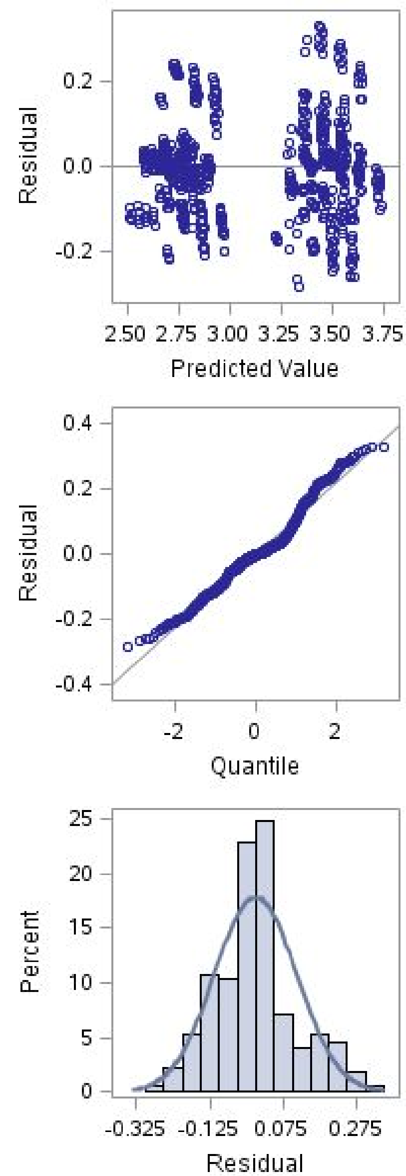

Figure 6: Cooling Load Residual Plots

The best model for heating load is the one with 5 variables (Relative Compactness, Wall Area, Overall Height, Glazing Area,and Glazing Area Distribution). This model has the highest Adjusted R-square value. C(p) is less than p. (5.0465 < 6). This model also has the lowest AIC and second lowest SBC values. This conclusion was made from Table 4.

Table 4: Model Selection - Heating Load

Number in Model	R-Square	Adjusted R-Square	C(p)	AIC	SBC
1	0.8058	0.8055	1563.305	-2393.6561	-2384.36856
2	0.8921	0.8918	530.6588	-2843.2476	-2829.31620
3	0.9335	0.9333	36.2692	-3213.3583	-3194.78310
4	0.9361	0.9358	7.2877	-3241.8367	-3218.61775
5	0.9365	0.9361	5.0465	-3244.1048	-3216.24204
6	0.9365	0.9360	7.0000	-3242.1517	-3209.64517

Number in Model	Parameter Estimates						
	Intercept	RelativeCompactness	WallArea	OverallHeight	Orientation	GlazingArea	GlazingAreaDist
1	1.71327	.	.	0.24432	.	.	.
2	1.46685	.	.	0.24432	.	1.05138	.
3	0.81399	.	0.00232	0.22810	.	1.05138	.
4	0.77830	.	0.00232	0.22810	.	1.01172	0.01599
5	0.99111	-0.25785	0.00202	0.24307	.	1.01172	0.01599
6	0.99405	-0.25785	0.00202	0.24307	-0.00083928	1.01172	0.01599

The best model for cooling load is the one with 4 variables (Relative Compactness, Wall Area, Overall Height, and Glazing Area). This model has the second highest Adjusted R-square value. This model has the lowest AIC and lowest SBC. This conclusion was made from Table 5.

Table 5: Model Selection - Cooling Load

Number in Model	R-Square	Adjusted R-Square	C(p)	AIC	SBC
1	0.8412	0.8410	726.9798	-2845.6551	-2836.36754
2	0.8886	0.8883	284.2468	-3115.6993	-3101.76796
3	0.9168	0.9164	21.3099	-3337.9497	-3319.37457
4	0.9186	0.9182	5.8373	-3353.3196	-3330.10062
5	0.9188	0.9183	6.2700	-3352.8971	-3325.03433
6	0.9189	0.9183	7.0000	-3352.1777	-3319.67115

Number in Model	Parameter Estimates						
	Intercept	RelativeCompactness	WallArea	OverallHeight	Orientation	GlazingArea	GlazingAreaDist
1	2.04618	.	.	0.20570	.	.	.
2	1.89579	.	.	0.20570	.	0.64166	.
3	1.45181	.	0.00158	0.19466	.	0.64166	.
4	1.85388	-0.48718	0.00102	0.22294	.	0.64166	.
5	1.83800	-0.48718	0.00102	0.22294	0.00454	0.64166	.
6	1.83128	-0.48718	0.00102	0.22294	0.00454	0.63418	0.00301

Next we ran proc reg with our chosen predictors for each corresponding model. It confirmed that all the predictors in each model are significant. This is shown in Table 6 and Table 7.

Table 6: Cooling Model with 4 Predictors

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.85388	0.10111	18.34	<.0001
RelativeCompactness	RelativeCompactness	1	-0.48718	0.11661	-4.18	<.0001
WallArea	WallArea	1	0.00102	0.00016531	6.15	<.0001
OverallHeight	OverallHeight	1	0.22294	0.00719	31.03	<.0001
GlazingArea	GlazingArea	1	0.64166	0.03044	21.08	<.0001

Table 7: Heating Load Model with 5 Predictors

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.99111	0.10868	9.12	<.0001
RelativeCompactness	RelativeCompactness	1	-0.25785	0.12513	-2.06	0.0397
WallArea	WallArea	1	0.00202	0.00017737	11.39	<.0001
OverallHeight	OverallHeight	1	0.24307	0.00771	31.53	<.0001
GlazingArea	GlazingArea	1	1.01172	0.03343	30.26	<.0001
GlazingAreaDist	GlazingAreaDist	1	0.01599	0.00287	5.57	<.0001

Conclusions

The model chosen for estimating the Heating Load contains all the predictors of the Cooling Load Model: Relative Compactness, Wall Area, Overall Height, and Glazing Area. The Heating Load Model additional includes the Glazing Area Distribution along with the Cooling Load Model predictors.

References

https://archive.ics.uci.edu/ml/datasets/Energy+efficiency