Validation of the Radar Gust Equation Used at the Kennedy Space Center and Cape Canaveral Air Force Station

Jared Rennie

Department of Atmospheric Sciences and Chemistry, Plymouth State University, Plymouth, New Hampshire

ABSTRACT

Forecasting for convective winds can be a rigorous task due to their small spatial resolution. Nonetheless the 45th Weather Squadron is responsible for predicting wind gusts that exceed 35 kts for the Kennedy Space Center and Cape Canaveral Air Force Station. While many resources are available, research has been performed in attempt to use WSR-88D technology. An equation developed by Loconto (2006) implements information from the storm structure alphanumeric product to predict the speed of a gust from a convective storm. This research will attempt to validate this equation outside the Cape Canaveral area. Scatter plots of reported wind gusts versus predicted wind gusts for three different areas (Continental Interior, Florida, Northeast Seacoast) show that Loconto’s equation may not be useful in all areas. In attempt to create a threshold for each area, multiple linear regression models are run in order to generate a new location specific equation to predict convectively driven gusts.

1) Introduction and Objectives

Convective winds are mesoscale in nature and are hard to accurately predict. Nonetheless these winds create a major aviation hazard, due to the strong low level wind shear that can be generated. Because of this, convective wind events are one of the greatest concerns at both the Kennedy Space Center (KSC) and Cape Canaveral Air Force Station (CCAFS) in Florida. Accurate prediction and lead time of these wind gusts are critical, in order to avoid unnecessary equipment costs, and to ensure human safety.

The 45th Weather Squadron is responsible for issuing convective wind advisories for thunderstorms in the area with the potential to produce wind gusts of 35 knots or greater (Loconto 2006). Many resources are available to them in generating these products, including a network of thirty six wind towers, radiosondes launched twice a day, radar data from the local WSR-88D in Melbourne, Florida (KMLB) as well as a new polarization radar on site, and equations developed by Loconto (2006) to help predict the value of these peak wind gusts.

Convectively driven winds are primarily generated from microbursts in a storm cell. Therefore the use of WSR-88D can be highly effective as a precursor in determining wind events. In fact, Loconto introduced an equation in attempt to determine the peak wind gusts based on data from the WSR-88D's storm structure product (Loconto 2006). However only forty-four cases were selected for this equation and radar data from only two different sites were used. How well does this equation hold up in other areas around the country and is there a need for a new equation or equations?

This research will attempt to test the validity of this radar gust equation, as well as provide new equations that could be used for different areas across the United States.

2) Background

a) Storm Structure Alphanumeric Product

When determining the wind gust from a downburst, it is important to figure out information from that specific storm cell, and not the entire area surrounding the radar data acquisition unit (RDA). Fortunately one of the derived products the WSR-88D can provide is storm structure data. This product, which is archived by the National Climatic Data Center (NCDC), provides information about the storm cell on each successive scan. The storm structure data is derived from the storm cell investigation and tracking algorithm (SCIT).

The SCIT algorithm will input volumetric reflectivity from WSR-88D data and attempt to plot a three dimensional field of the storm cell once all of the elevation scans are completed. The cells are then given an ID number and are ranked by their cell based vertically integrated liquid. Once the next volumetric scan is completed, the algorithm will attempt to locate similarities with the previous cell and define a storm track. If a correlation is made, the cell is given the same ID as before and all attributes are updated. If no correlation is made, it is given a new ID (Johnson *et al* 1998).

Once a storm cell is identified, the algorithm will calculate many different parameters to determine the information about that cell. The relevant information that can be obtained from storm structure data for this research includes the following:

- Cell-based vertically integrated liquid (VIL) (kg m^{-2})
- Echo top (in kilofeet)
- Maximum reflectivity (in dBZ)
- Height of the maximum reflectivity (in kilofeet)

If the cell that produced the maximum wind gust can be located, then one (if not all) of these values could be used to not only forecast convective wind events, but also determine how high the wind gust will be.

Additionally, another value can be derived from the storm structure data to help determine the convectively driven wind gust: VIL Density. VIL Density is defined as

$$\mathrm{VILD} = \frac{\mathrm{VIL}}{\mathrm{ET}} \tag{1}$$

Where VIL is the cell based vertically integrated liquid (kg m^{-2}) and ET is the echo top (kft). The units of VIL Density are kg m^{-3}.

b) Past Research

The idea of using WSR-88D data was first introduced by Stewart (1996), who created an equation to help determine the maximum downdraft speed of a microburst event based on derived WSR-88D products. The equation is the following:

$$w = \sqrt{(20.628571 \cdot VIL) + (-3.125 \times 10^{-6} \cdot ET^{2})} \tag{2}$$

where VIL is the vertically integrated liquid (kg m^{-2}), ET is the echo top (kft) and w is the maximum predicted downdraft speed (ms^{-1}). Because downdrafts generally diverge near the surface, if the downdraft speed is known, then one can derive the peak gust for an area near the storm cell.

A study performed by Sullivan (1999) evaluated three different predictive wind gust equations to see which one was the most effective at the Kennedy Space Center. Using the KSC wind tower network to locate the value of the maximum peak gust, and storm structure radar data from the KMLB radar, he computed the root mean square error (RMSE) and the mean absolute error (MAE) for each of the three equations against fifteen microburst events in and around the Cape Canaveral area.

He concluded that the equation created by Stewart (1996) best predicted the observed gust. However accuracy values were noticeably large, with a RMSE of 11.9 kts and a MAE of 9.7 kts (Sullivan 1999). Additionally, the question of statistical significance is introduced because of the low number events used in the study. Perhaps if more cases were used, the results would have been more conclusive.

For his radar meteorology project, Loconto (2006) updated the results of Sullivan (1999), who increased the number of events from 15 to 30 that met the 45th Weather Squadron criteria of 35 knots or greater. Additionally these events were divided into two sections: 11 strong events (gusts greater than 50 kts) and 19 weak events (gusts between 35 and 49 kts). For consistency, he used the same VIL/ET relationship that was favored by Sullivan (1999).

His results were a little more promising than Sullivan (1999). The RMSE and MAE for all 30 events were 5.93 and 5.23 knots, respectively. The weak wind events had lower values of RMSE and MAE (5.45 and 4.95 kts, respectively) and strong wind events had higher values (6.67 and 5.72 kts, respectively). While the results show more promise, Loconto argued that more variables could be added and a new equation could be constructed (Loconto 2006).

Using that motivation as a part of his thesis, Loconto (2006) set out to create a more effective equation to determine the maximum peak wind gust based on radar variables from KMLB storm structure data. Using a new data set of 44 cases, 30 of which met warning criteria (gusts greater than 35 kts) and 14 that were below criteria (gusts less than 35 kts), he recalculated the correlation coefficient between the peak wind gust and all variables of the storm structure data. He concluded that the cell based VIL and maximum reflectivity were the best correlated values. Using forward stepwise regression as a variable selector and a multiple linear regression model, the following equation was developed by Loconto (2006):

$$GU = (.4138 \times VIL) + (.9194 \times MaxZ) + (.6253 \times height) - 28.7719 \quad (3)$$

where *GU* is the maximum peak wind gust (kts), *VIL* is the cell-based vertically integrated liquid (kg m^{-2}), *MaxZ* is the maximum reflectivity (dBZ), and *height* is the height of the maximum reflectivity (kft).

To test this equation, an independent set of twenty-two cases were randomly selected and the wind gusts reported were checked against the wind gusts generated by the equation. The results show that the values were fairly correlated, with a correlation coefficient of approximately 0.7 (Loconto 2006). This suggests that the new radar gust equation can accurately predict the maximum wind gust much better than the first equation developed by Stewart (1996).

Today, the equation generated by Loconto (2006) is a highly effective tool for nowcasting convective winds. However this equation has only been tested at or near the Kennedy Space Center Complex. While convective wind events are frequent in the state of Florida, how well does this equation work across the United States? Other examples that can generate convective winds include a dry line producing supercells in west Texas, or an overnight Mesoscale Convective Complex (MCC) in Iowa. Additionally, squall lines in the New England area can generate destructive winds. How well will radar sites in those areas be able to predict wind gusts generated by these events? For this paper we would like to answer the following scientific question: How well does Loconto's radar gust equation accurately predict wind gusts from convectively driven events across the United States?

3) Data and Methodology

To solve this question, three individual groups must first be defined: Florida, the Continental Interior, and the Northeast Seacoast. Florida experiences convection frequently, due to its tropical like air mass and small scale forcing (for example, sea breeze fronts). For the purposes of this research, the continental interior is defined by five mid-western states: Oklahoma, Kansas, Nebraska, South Dakota, and North Dakota. These five states are chosen because they lie in generally the same area of longitude. This area is located on the lee side of the Rocky Mountains, where in the presence of a strong low-level jet, mesoscale convective systems can form. The northeast seacoast is subject to complex terrain, and experiences convective events frequently during the summer months.

Two datasets will be extracted over the summer months of 2007 and 2008 (June, July and August). Summer months are chosen because convection generally occurs in all three defined areas. The first dataset is the wind gust report from the Storm Prediction Center's storm reports page. These values, which are in miles per hour, will be converted to knots and will be considered the "true" value for this research. If there are multiple wind reports over one day, then the largest wind gust will be used.

Overall twenty to thirty cases will be selected for each defined area. In order to ensure a wide range of wind gusts, some of the cases will have peak wind gusts below the KSC criteria of 35 kts. This poses a problem, since a lot of SPC reports will have unknown values. In order to determine a proper wind value, the nearest METAR report with a valid wind gust will be used. The METAR must be within a few miles of the report, or else it will be disregarded.

It is also important to note that Loconto's (2006) equation is only valid for convectively driven winds. A wind gust induced by synoptic flow or a frontal system cannot be used. To account for this, surface maps produced by the Hydrometeorological Prediction Center will be subjectively looked at to confirm that the area of the reported wind gust is associated with a weak pressure gradient.

Once the cases are identified for each group, the nearest WSR-88D is located and storm structure data are analyzed. The data files are archived by the National Climatic Data Center (NCDC) and the volume scan at, or just before the time of the SPC report will be used. Using NCDC's Java NEXRAD Data Viewer, the cell closest to the location of the SPC report will be chosen and the alphanumeric text product will be generated.

Four main components of the storm structure product will be extracted (cell based VIL, echo top, maximum reflectivity, and height of the max reflectivity) and placed in a spreadsheet. Using the echo top and cell based VIL, the VIL Density will also be calculated. Next, the cell based VIL, max reflectivity, and height of the max reflectivity will be inserted into Loconto's gust equation, and the resulting gust value will be considered the derived estimates of the wind gust.

Once the true and derived values of the wind gust are computed, statistical tests will be applied to test its correlation. First, scatter plots will be generated for each of the three areas with the predicted estimates on the x-axis, and actual estimates on the y-axis.

From these scatter plots, a simple linear regression equation will be created with a correlation coefficient (R^2) value. The correlation coefficient is a number between 0 and 1, where 0 gives us no correlation, and a value of 1 gives us a perfect correlation. The higher the value, the better the data are correlated with each other. For our purposes, we would like to see high R^2 values, which indicate a strong correlation between the derived and actual wind gusts. Additionally the average bias will be calculated to see if the radar gust equation over predicts or under predicts the actual wind gust.

Because we are comparing radar data to the real or “true” estimate, the root mean square error (RMSE) will also be calculated. The RMSE will show an averaged difference between the derived and true values of the wind gust. A small value indicates the derived values are relatively close to the true values. For our purposes, we would like to see small values of RMSE.

If poor correlations and high RMSE values are noted (which is expected to be true), new equations will be generated to predict convectively driven wind gusts in each of the three defined areas. To do this, each of the individual variables from the storm structure data (cell based VIL, echo top, max reflectivity, height of max reflectivity), as well as VIL density, will act as predictors and the true wind gust will be the response.

From here, scatter plots of the individual predictors will be plotted against the true wind gust estimates, in order to determine how well each variable correlates with the actual wind gust. Additionally, using the Minitab statistical software package, the best subsets function will be run in order to choose the best predictors for each model. The function attempts to determine which variables are best fitted for a multiple linear regression model using different values. These values include the correlation coefficient, the Mallows Cp value, and the standard deviation.

To prevent collinearity (in other words, over fitting or under fitting the model with too many or too little predictors), the Mallows Cp variable is used in determining model selection. The best model output is chosen when the Mallows Cp value is approximately equal to the number of predictors in the model (Wilks 2005). To sum up, the best model that should be chosen has a high correlation coefficient, a low standard deviation, and a Mallows Cp value that is approximately equal to the number of predictors. Because the function does not choose which model to use, the model is chosen subjectively.

Once the predictors are chosen a multiple linear regression model will be run and an equation will be generated. To test the equations significance, an analysis of variance (ANOVA) table will be shown, which displays the degrees of freedom, mean squared error, F-Statistic, and p-value.

The F-Statistic is defined as the following: a ratio of the mean-squared regression value to the mean-squared residual indicating the relative appropriateness of the model. (Wilks 2005). For this test, we define the null hypothesis as the following: the predictors in the model cannot explain the variance of the wind gust. The alternative hypothesis is just the opposite, in that the predictors can explain the variance of the wind gust.

Depending upon the degrees of freedom for the regression (predictors) and residuals, an F-critical value can be computed. If the F-value from the ANOVA table is higher than this critical value, along with a low p-value, we can reject the null in favor of the alternative. This in other words means that the model is statistically significant, and the resulting equation could be used for its respective area.

4) Results

Table 1 below shows the number of events that were found for each area. Overall there were 88 convective events, 60 of which met the warning criteria of 35 knots, and 28 which were below 35 knots.

	Criteria	Non-Criteria	Total
Continental Interior	23	8	31
Florida	21	12	33
Northeast Seacoast	16	8	24
ALL	60	28	88

Table 1: Number of events ranked by area, and criteria vs. non-criteria

Part a will go over the validity of Loconto's equation, using scatter plots, RMSE, average biases, and R^2 values. Part b will attempt to generate new equations via multiple linear regression. All figures are located in the appendix for reference.

a) Predicted Wind Gust vs. Actual Wind Gust

Figure 1 shows the scatter plot, linear regression, average bias, and RMSE for the Continental Interior area. The predicted wind gust from Loconto's equation is on the x-axis, and the actual gust from the SPC/METAR report is on the y-axis. Overall the data does not appear to be well correlated with each other, as the correlation coefficient is 0.2497 and a very large RMSE of 15.88 kts is shown. The bias is -2.05, which indicates that the radar gust equation will underestimate the actual wind gust.

Figure 2 shows the scatter plot, linear regression, average bias, and RMSE for the state of Florida. Their results are not much better, with an RMSE of 14.86 kts and an R^2 value of 0.3015. However the bias is somewhat low at 0.86 kts, but that still indicates that the predictive equation will tend to overestimate the actual wind gust.

Out of the three defined areas, it appears that the Northeast Seacoast saw the best results. Their scatter plot is shown in Figure 3, with an R^2 value of 0.4407, and an RMSE of 12.27 kts. The bias however is the largest out of the three at 3.56 kts. This indicates that the equation will tend to overestimate the actual wind gust for the Northeast Seacoast

Figure 4 shows the scatter plot, linear regression, average bias, and RMSE for all of the three defined areas. The overall R^2 value is 0.3233, with an overall bias and RMSE error of 0.57 and 14.58 knots, respectively. While the bias is low, the correlation coefficient and RMSE are still not values we would like them to be.

Table 2 below shows the R^2 value, RMSE, and average bias for all four cases. Overall the results do not appear to match up with the results Loconto (2006) suggested in his thesis. Using an independent study of twenty-two cases, he was able to generate a statistically significant outcome with an R^2 value of approximately 0.7. All of our cases, including all three cases combined, could not meet that threshold. This suggests that the equation derived by Loconto, may not be able to accurately predict wind gusts outside of the Cape Canaveral area, and more work needs to be done.

	R^2	RMSE (kts)	Bias (kts)
Continental Interior	0.2497	15.88	-2.05
Florida	0.3015	14.86	0.86
Northeast Seacoast	0.4407	12.27	3.56
All Three Cases	0.3233	14.58	0.57

Table 2: Correlation Coefficient, Root Mean Square Error, and Bias between the predicted wind gust and actual wind gust for each area

b) Multiple Linear Regression Model:

In attempt to create some sort of threshold for accurately predicting wind gusts, a multiple linear regression model was created for all three cases. Using the actual wind gust as the response, and the five variables from the storm structure data (top, VIL, maxZ, height, and VILD) as the predictors, scatter plots are generated and the correlation coefficient is calculated. These values can be seen in Table 3 below. In addition, the best subsets function was run to determine which of the five predictors were best fit for each area, and then a multiple linear regression model, along with an ANOVA table were generated.

	Top	VIL	VILD	MaxZ	Height
Continental Interior	0.0956	0.3253	0.3764	0.2396	0.0111
Florida	0.2116	0.3029	0.2656	0.2912	0.0435
Northeast Seacoast	0.3616	0.4726	0.2369	0.2760	0.0586

Table 3: Correlation coefficient between the reported wind gust and Echo Top, Vertically Integrated Liquid, VIL Density, Maximum Reflectivity and height of the maximum reflectivity for each area

I) Continental Interior

Figures 5-9 show the scatter plots for each of the five predictors against the actual wind gust for the Continental Interior area. It appears that VIL and VILD have the best results, with correlation coefficient values of 0.3253 and 0.3764, respectively. On the other hand, both top and height have the worst results, with correlation coefficient values of 0.3253 and 0.3764, respectively.

Using the best subsets function (seen in Table 7 in the appendix), four out of the five variables were deemed as the best fit: top, maxZ, height, and VILD. Using these four variables gives us the highest R^2 value of 0.447 and a Mallows Cp of 4.1, which is approximately equal to the number of predictors. However height had the worst correlation of 0.0111. Because of this the next best model was used. Interestingly enough, it was the same model as above, just without the height predictor. This gives us an R^2 value of 0.428, a mallows Cp of 3, and the lowest standard deviation out of all the models (13.711).

Inserting these three predictors into the multiple linear regression model, we get the following equation:

$$\textbf{Gust} = 60.0 + 0.523\ \textbf{Top} - 1.02\ \textbf{MaxZ} + 9.89\ \textbf{VILD} \tag{4}$$

Where Gust is the wind gust in kts, Top is the echo top in kilofeet, MaxZ is the maximum reflectivity in dBZ, and VILD is the VIL Density in kg m^{-3}. The ANOVA table for this model is below:

Source	DF	SS	MS	F	P
Regression	3	3799.6	1266.5	6.74	0.002
Residual Error	27	5078	188		
Total	30	8875.6			

Table 4: ANOVA table for the Continental Interior multiple linear regression model.

Table 4 indicates that with 3 degrees of freedom for regression and 27 degrees of freedom for residual error, the F-Critical value is approximately 2.960, because our F-statistic is higher (6.74) with a p-value of 0.002, we can reject the null in favor of the alternative, meaning that the four predictors in this regression can explain the total variance of the wind gust.

II) Florida

Figures 10-14 show the scatter plots of each of the five predictors against the actual wind gust for the Florida area. Height had the worst correlation, with a correlation coefficient value of 0.0435. The other four predictors also have low correlation values, with the best being VIL at 0.3029.

The best subsets function (Table 8 in the appendix) suggests that the best model can be generated with top, VIL, and VILD. The R^2 value would end up being 0.414 and the Mallows Cp would be 2.1. The standard deviation is also the lowest out of all the other models, at 13.870.

Inserting these three predictors into the multiple linear regression model, we get the following equation for the Florida area:

$$\mathbf{Gust = -\,2.8 + 1.13\ Top - 0.611\ VIL + 10.4\ VILD} \tag{5}$$

Where Gust is the wind gust in kts, Top is the echo top in kilofeet, VIL is the cell based vertically integrated liquid in kg m^{-3}, and VILD is the VIL Density in kg m^{-3}. The ANOVA table for this model is below:

Source	DF	SS	MS	F	P
Regression	3	3934.3	1311.4	6.82	0.001
Residual Error	29	5578.6	192.4		
Total	32	9512.9			

Table 5: ANOVA table for the Florida multiple linear regression model.

Table 5 above shows that with 3 degrees of freedom for the regression predictors and 29 degrees of freedom for the residual error, the F-Critical value is approximately 2.934. Because the F-Statistic (6.82) is higher than this critical value with a p-value of 0.001, we can say that this model is also statistically significant. We can reject the null in favor of the alternative, meaning that for Florida, the VIL, VIL Density, and Echo Tops explain for most of the wind gust variance.

III) Northeast Seacoast

Figures 15-19 show the scatter plots of each of the five predictors against the actual wind gust for the Northeast Seacoast area. Both the echo top and VIL had the best correlation with the wind gust, as correlation coefficient values of 0.3616 and 0.4726 respectively are shown. Once again the height of the maximum reflectivity had the worst correlation, with an R^2 value of 0.0586. Height is the only predictor not used in any of the three models.

The best subsets function (Table 9 in the appendix) indicates that the predictors with the highest correlation coefficients(Top and VIL) will also serve as the best model for the Northeast Seacoast. The correlation coefficient of the model is high (0.479), the Mallows Cp is 0.3, and it's standard deviation is relatively small at 12.115.

Inserting these predictors into the multiple linear regression model, we get the following equation for the Northeast Seacoast:

$$\mathbf{Gust = 17.6 + 0.240\ Top + 0.620\ VIL} \tag{6}$$

Where Gust is the wind gust in kts, Top is the echo top in kilofeet, and VIL is the cell based vertically integrated liquid in kg m^{-3}. The ANOVA table for this model is below:

Source	DF	SS	MS	F	P
Regression	2	2832.8	1416.4	9.65	0.001
Residual Error	21	3082.2	146.8		
Total	23	5915			

Table 6: ANOVA table for the Northeast Seacoast multiple linear regression model.

Table 6 above indicates that with 2 degrees of freedom for the regression predictors and 21 degrees of freedom for the residual error, the F-Critical value is approximately 3.467. Because the F-Statistic is much higher (9.85) than this critical value with a p-value of 0.001, we can say that this model, just like the models for the Continental Interior and Florida, is statistically significant. We can reject the null in favor of the alternative, meaning that for the Northeast Seacoast, the cell based VIL and echo top can explain for most of the variance in the wind gust.

5) Conclusions

Using Loconto's radar gust equation, it appears that it does not accurately predict convectively driven winds across the three defined areas. The root mean square errors were very high (between approximately 12 and 16 kts), and the correlation coefficient values were low (between approximately 0.24 and 0.44). This is possibly due to the fact that all 44 events used in his thesis were all located at, or near the Kennedy Space Center. There could be different parameters in the storm structure data that are more favorable in different areas.

That idea was tested using multiple linear regressions for each of the three individual areas. We can see from Table 3 above that different parameters were better correlated in some areas and not others. It is also important to note that the height parameter had the worst correlation for all three areas, and was never used in any of the regression models.

Echo top was deemed important in all three areas. This makes some sense, because when the storm cell is higher in the atmosphere, there is a better chance for more precipitation in a vertical column. If the conditions are dry enough near the surface, more precipitation falling into this area could generate a much larger downburst or microburst.

All of the models above gave us different equations with different parameters. The ANOVA table suggests that each of the models appear to be statistically significant and that their respective predictors can describe most of the variance. Unfortunately due to time constraints, verification of these equations could not be assessed during the time of research.

It is also important to note some flaws during data acquisition. First, the SCIT algorithm is dependent on the volume control pattern the RDA is using. Different VCP's have different number of elevation angles. The more angles there are, the better the approximation of cell based VIL will be. This may cause a misrepresentation of the predicted wind gusts.

Additionally, we assumed that the wind gust from either the SPC report or METAR was assumed to be the true value. This is not necessarily the case. First, how was the report taken? If it was by a human, what did they use to record the wind? If it was automated, did the instrumentation have any contamination issues at the time? Additionally, depending on how far away the report was from the actual storm cell, the downburst could have decreased over time due to surface friction, or loss of energy. This once again provides a misrepresentation of the results.

Also, the storm structure data were extracted either at or near the time of the reported wind gust. Perhaps if these cases were tested over longer periods of time, we would not only see better results, but also have a chance to predict these convectively driven wind events a few minutes (if not more) before event onset.

Finally, the question of statistical significance is still an issue. Only cases in the summer months of 2008 and 2007 were used. Months of May and September could also be added, as well as many more years. Perhaps for future research, more data can be added and automated to provide better results in a less amount of time.

6) Summary

This research attempted to validate Loconto's radar gust equation used for predicting convectively driven winds at Cape Canaveral Air Force Station and the Kennedy Space Center. Using SPC and METAR reports as the actual wind gust, and storm structure text products from NCDC to determine the derived wind gust, statistical regression was calculated for Florida, the Continental Interior, and Northeast Seacoast. All three areas had relatively high values of RMSE and low correlation coefficients.

In order to provide an estimate of convectively driven winds at other areas, four predictors from the storm structure product, as well as VIL density, were inserted into multiple linear regression models and new equations were generated. While the results were not as great as Loconto's equation, it can be noted that more research needs to be done.

Many assumptions and issues have not been accounted for in this research, and the question of statistical significance still remains. However the results from this project is a motivation for thesis research this summer to provide better nowcasting tools for predicting convectively driven wind gusts with WSR-88D products.

7) References and Acknowledgements

Johnson, J.T., MacKeen, P.L., Witt, A., Mitchell, E.D., Stumpf, G.J., Eilts, M.D., and Thomas, K.W., 1998. The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm. *Weather and Forecasting*, **13**, 263-276.

Loconto, A.N., 2006. Improvements of Warm-Season Convective Wind Forecasts at the Kennedy Space Center and Cape Canaveral Air Force Station. M.S. Thesis, Dept. of Chemical, Earth, Atmospheric and Physical Sciences, Plymouth State University, Plymouth, NH.

---, 2006 Prediction of Warm-Season Convective Wind Gusts at the Kennedy Space Center Using the Echo Top / VIL Wind Gust Potential Equation *Radar Meteorology*.

Stewart, S.R., 1996. Wet Microbursts—Predicting Peak Wind Gusts Associated with Summertime Pulse-Type Thunderstorms. Preprints, 15th Conf. on Weather and Forecasting. Amer. Meteor. Soc., 324-327.

Sullivan, G.D., 1999. Using WSR-88D to Forecast Convective Winds at the Kennedy Space Center and Cape Canaveral Air Station. M.S. thesis, Air Force Institute of Technology. 97 pp.

Wilks, D. S., 2005: *Statistical Methods in the Atmospheric Sciences, Volume 91, Second Edition*. Academic Press, 648.

8) Appendix

a) Predicted Wind Gust vs. Actual Wind Gust

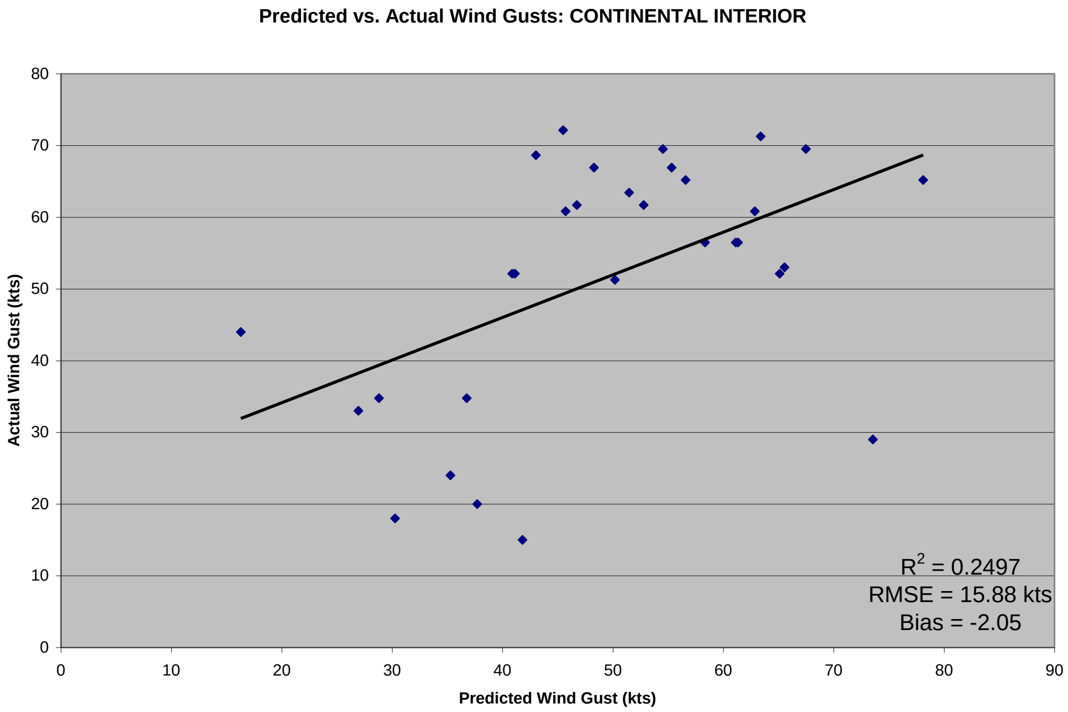

Figure 1: Scatter plot, correlation coefficient, root mean square error, and average bias for the Continental Interior.

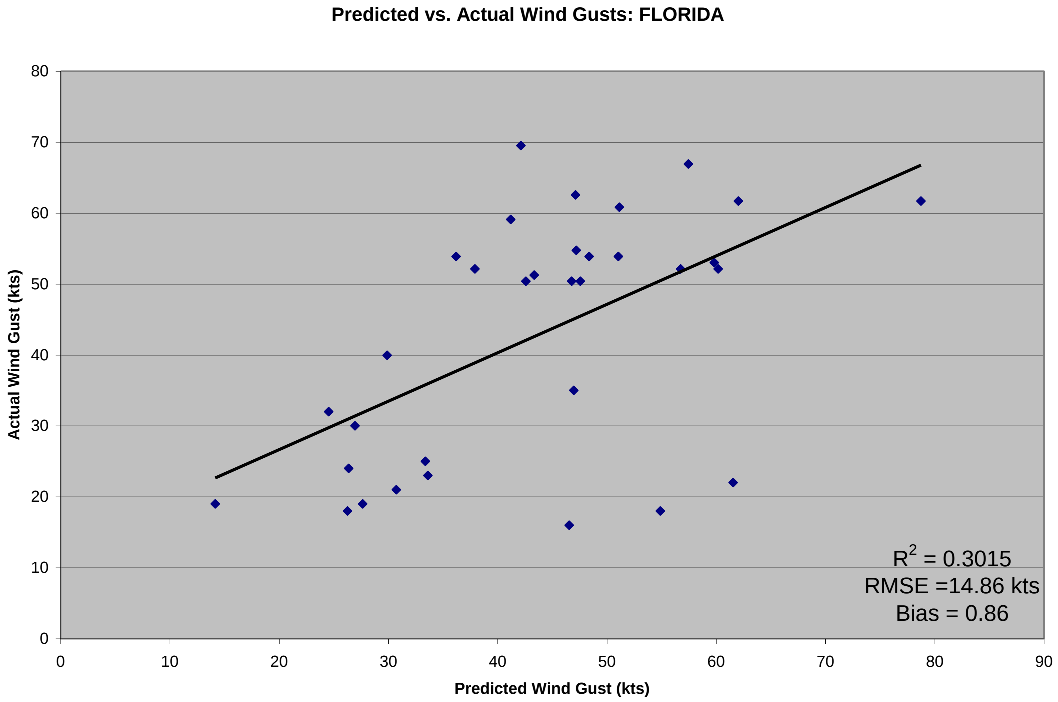

Figure 2: Same as Figure 1, but for Florida

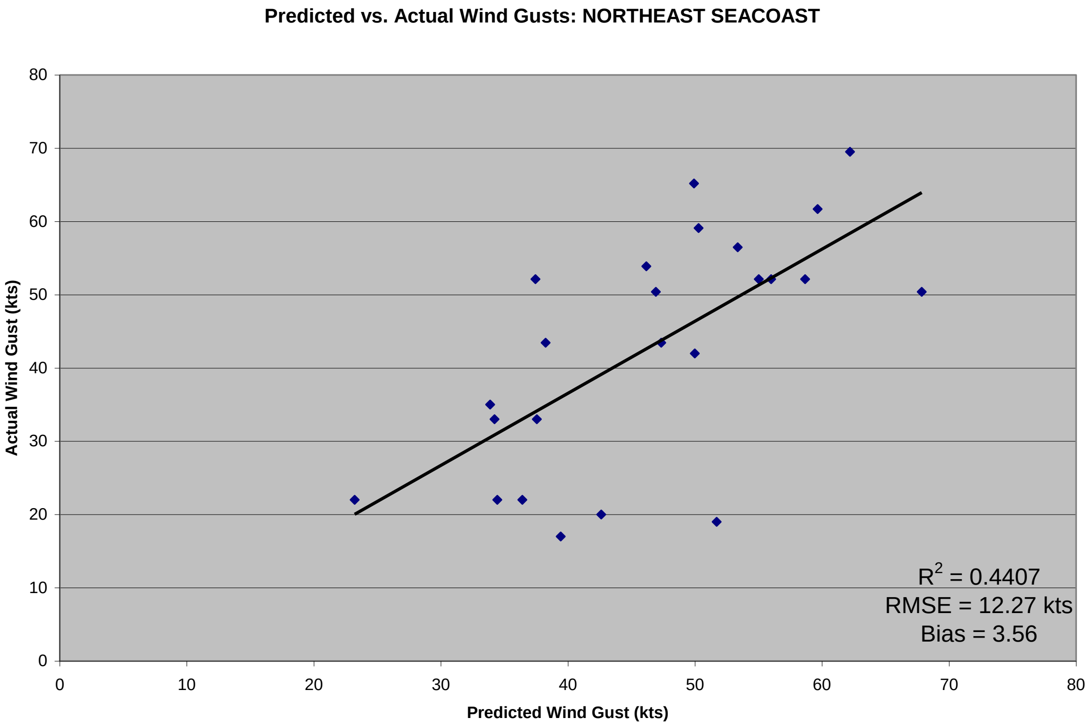

Figure 3: Same as Figure 1, but for the Northeast Seacoast

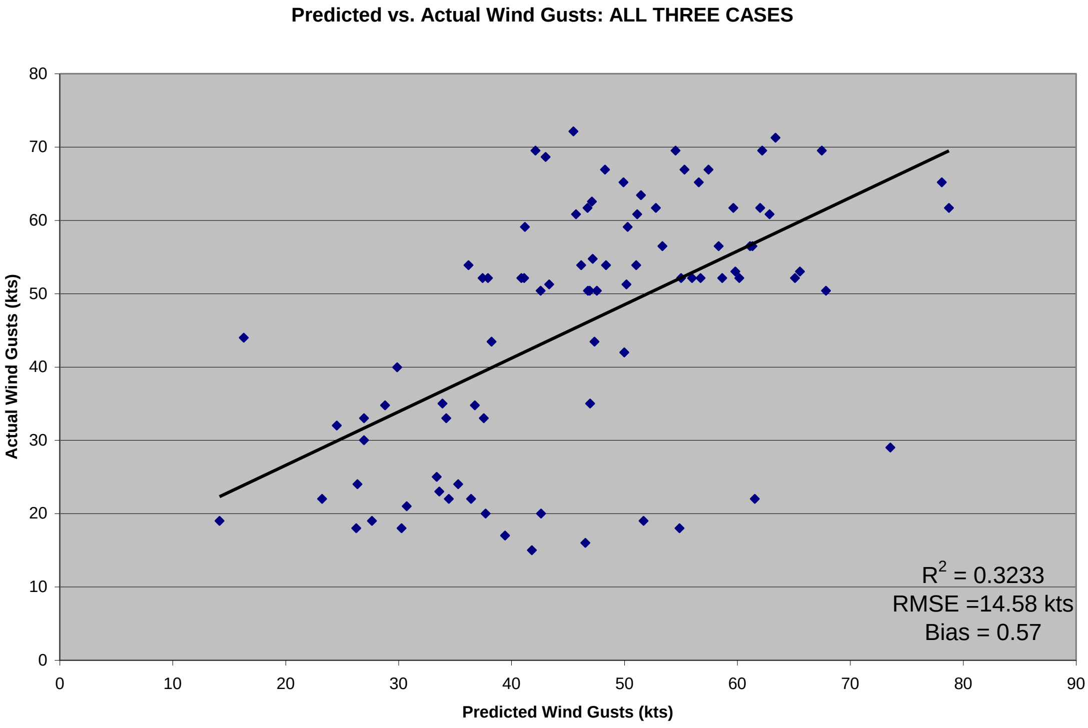

Figure 4: Same as Figure 1, but for all three areas combined

b) Multiple Linear Regressions: Continental Interior

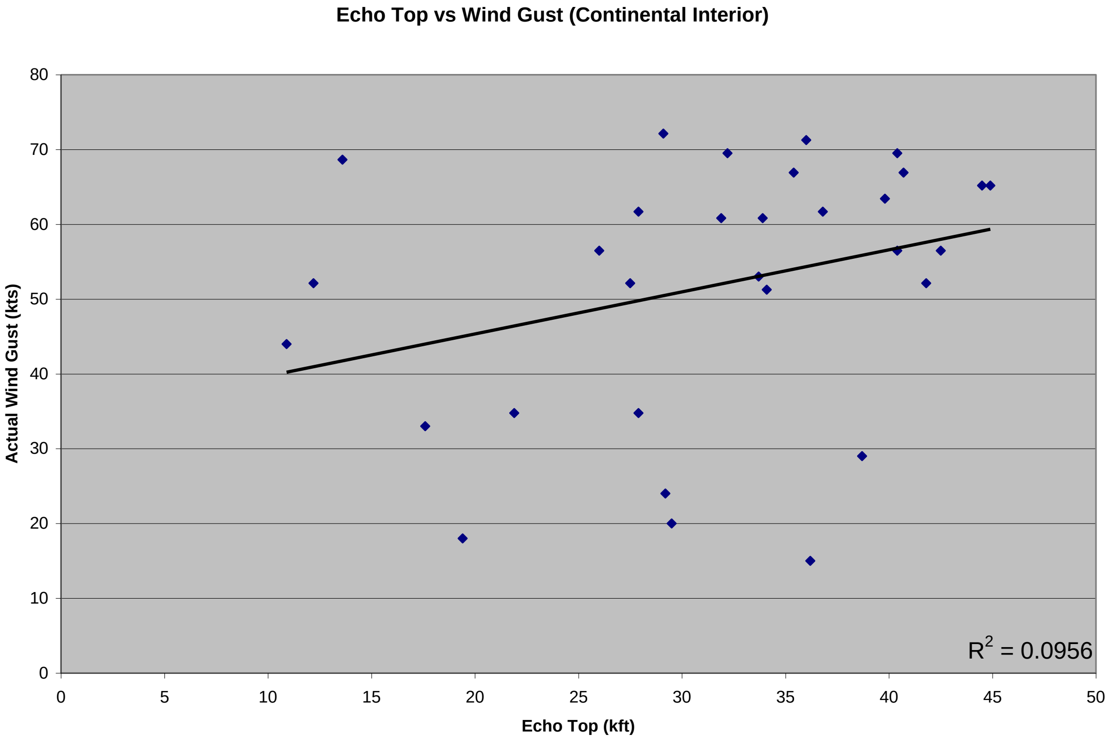

Figure 5: Scatter plot and correlation coefficient of reported wind gust vs. Echo Top for the Continental Interior.

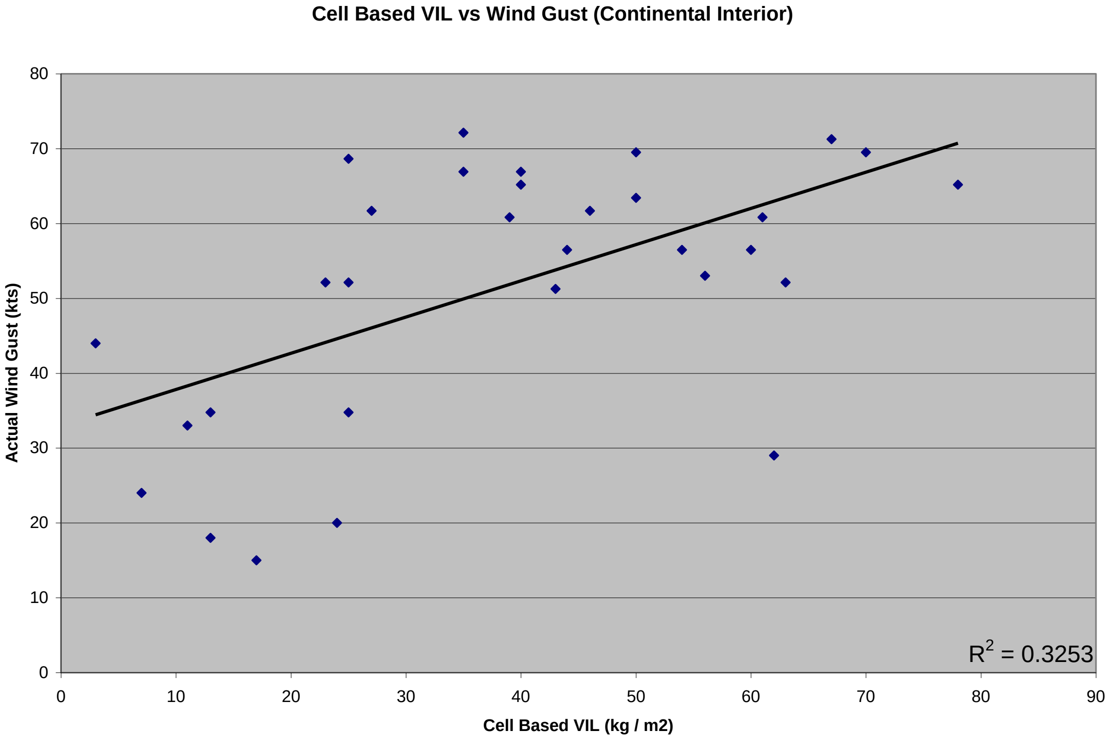

Figure 6: Same as Figure 5, but for cell based vertically integrated liquid

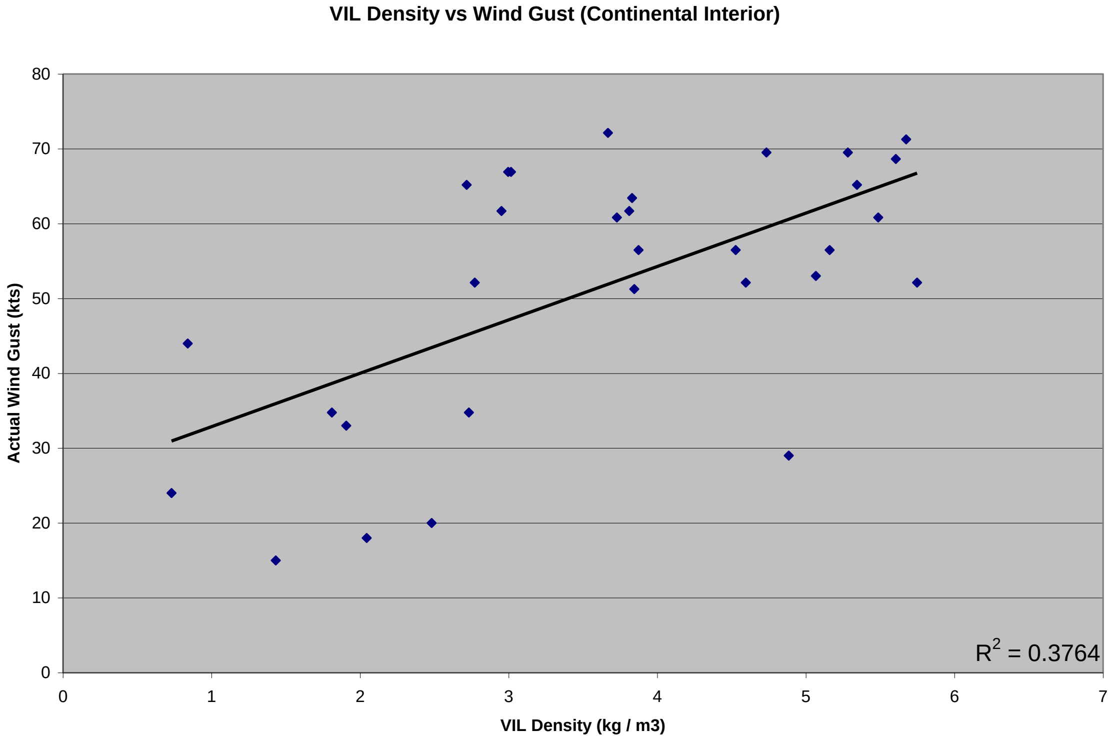

Figure 7: Same as Figure 5, but for VIL Density

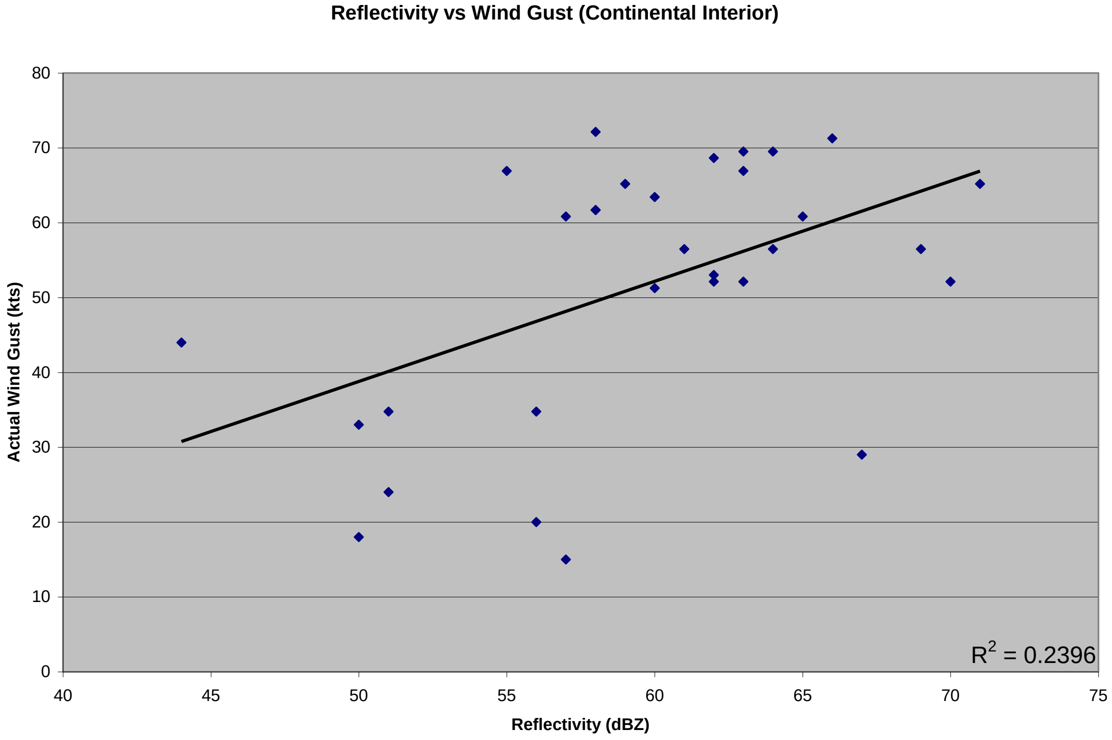

Figure 8: Same as Figure 5, but for maximum reflectivity

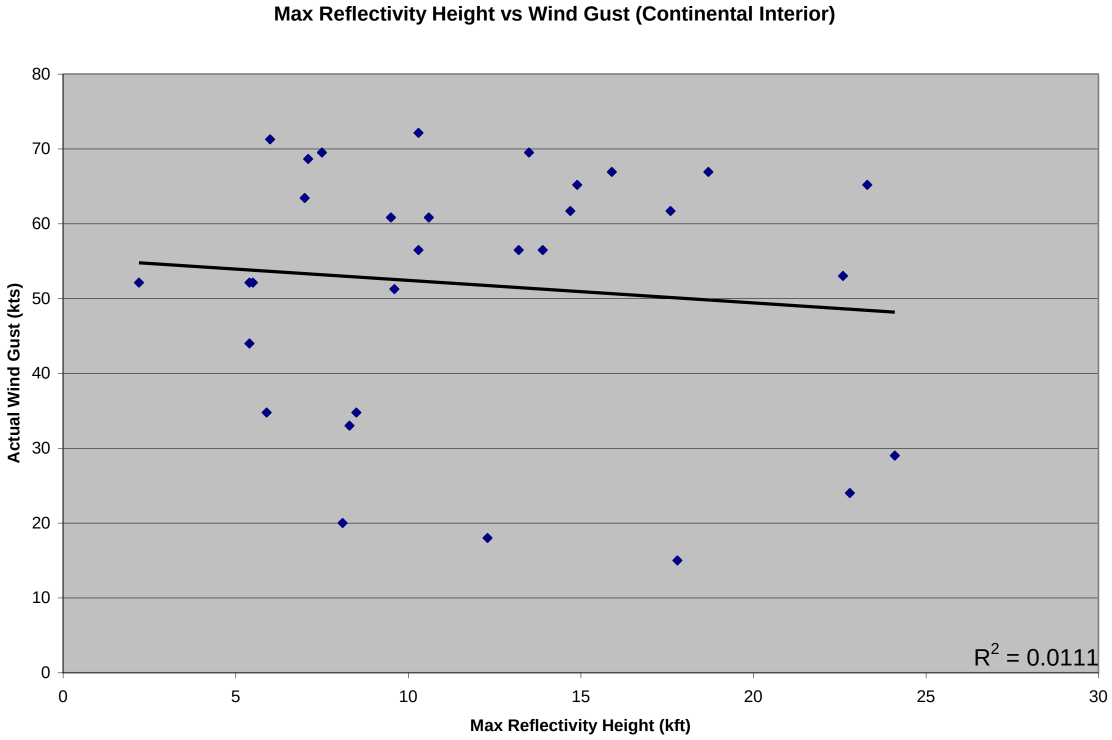

Figure 9: Same as Figure 5, but for height of the maximum reflectivity

c) Multiple Linear Regressions: Florida

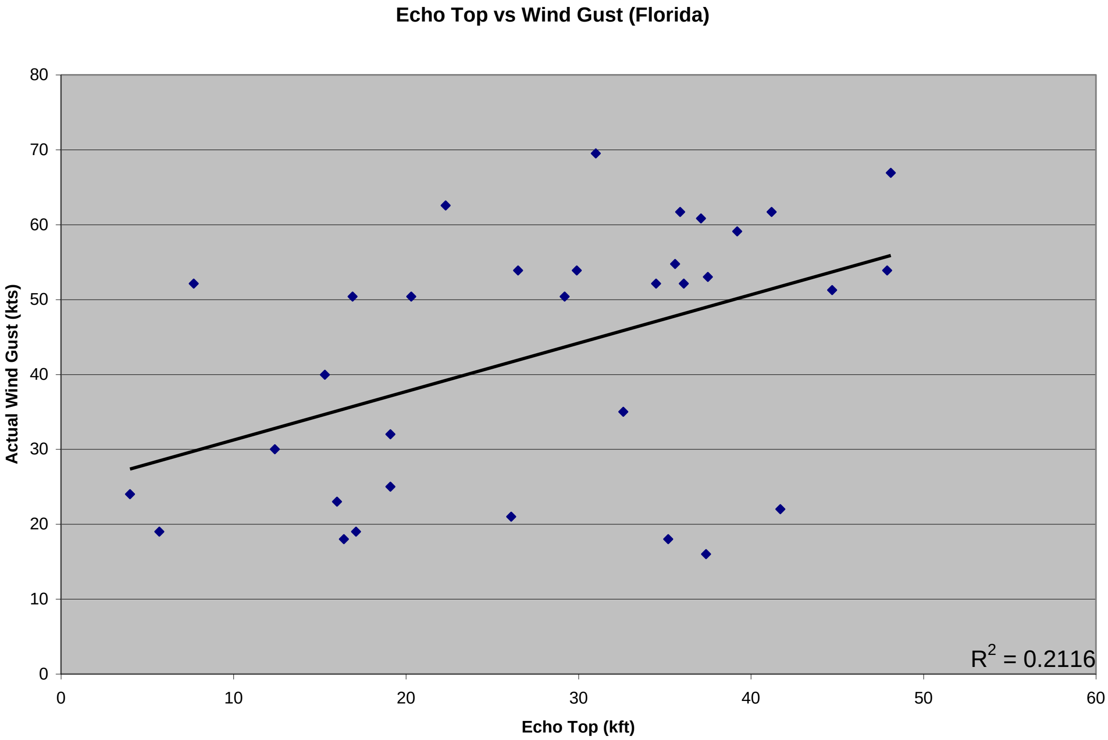

Figure 10: Scatter plot and correlation coefficient of reported wind gust vs. Echo Top for the state of Florida

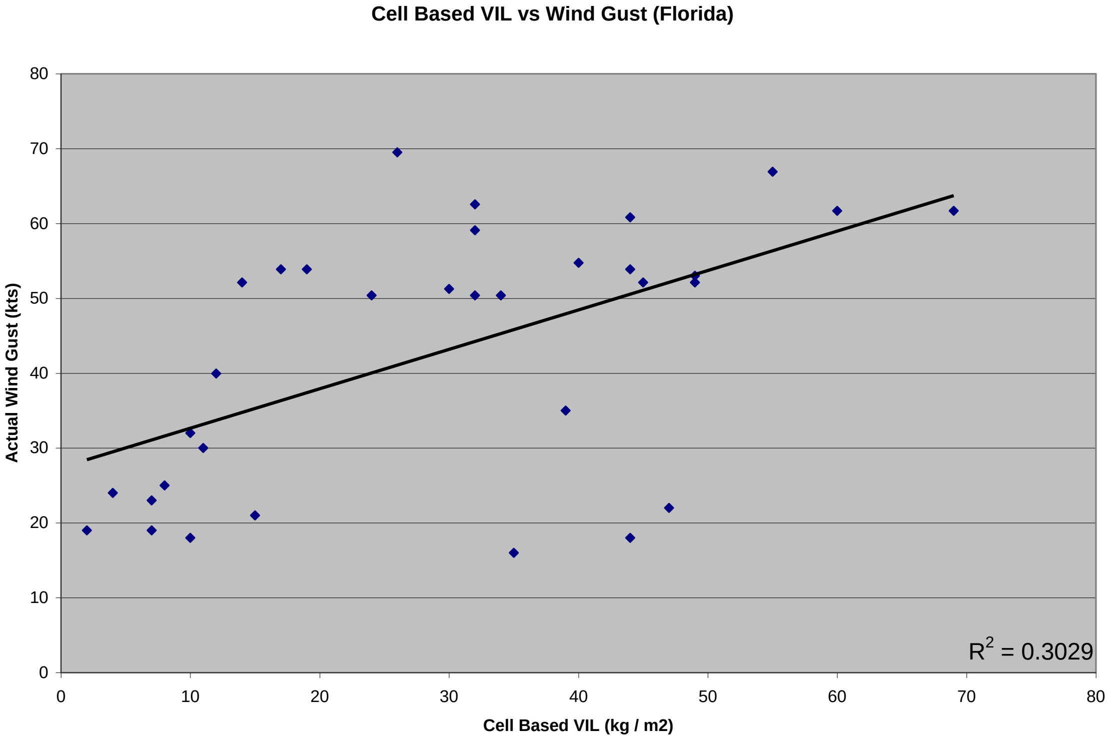

Figure 11: Same as Figure 10, but for cell based vertically integrated liquid

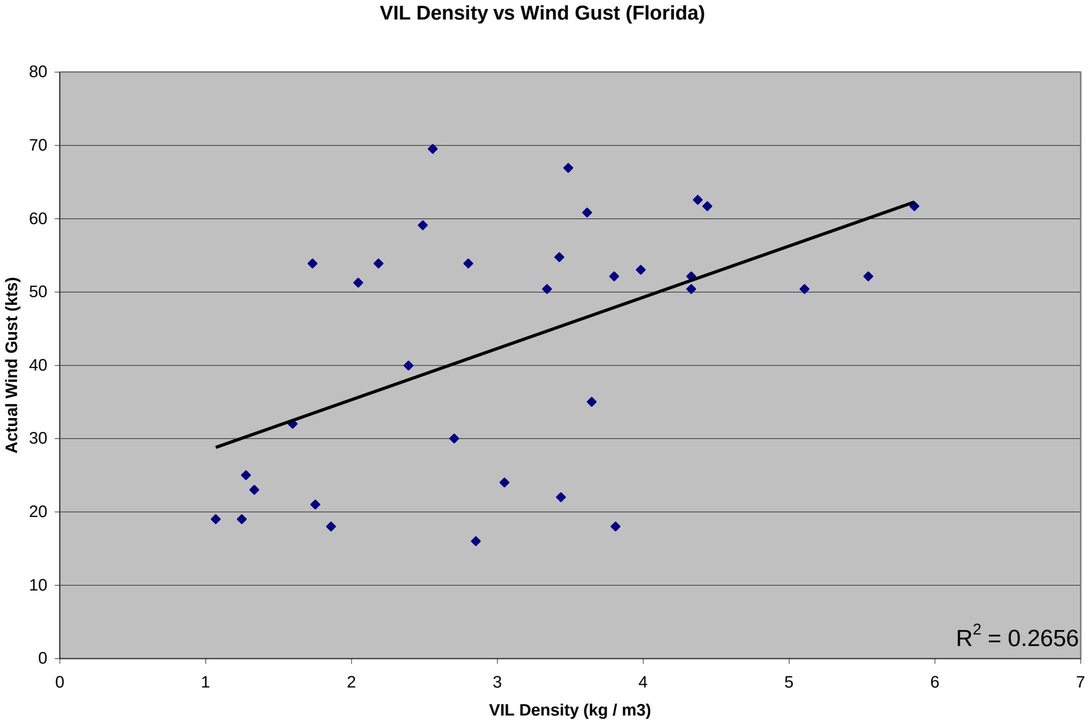

Figure 12: Same as Figure 10, but for VIL Density

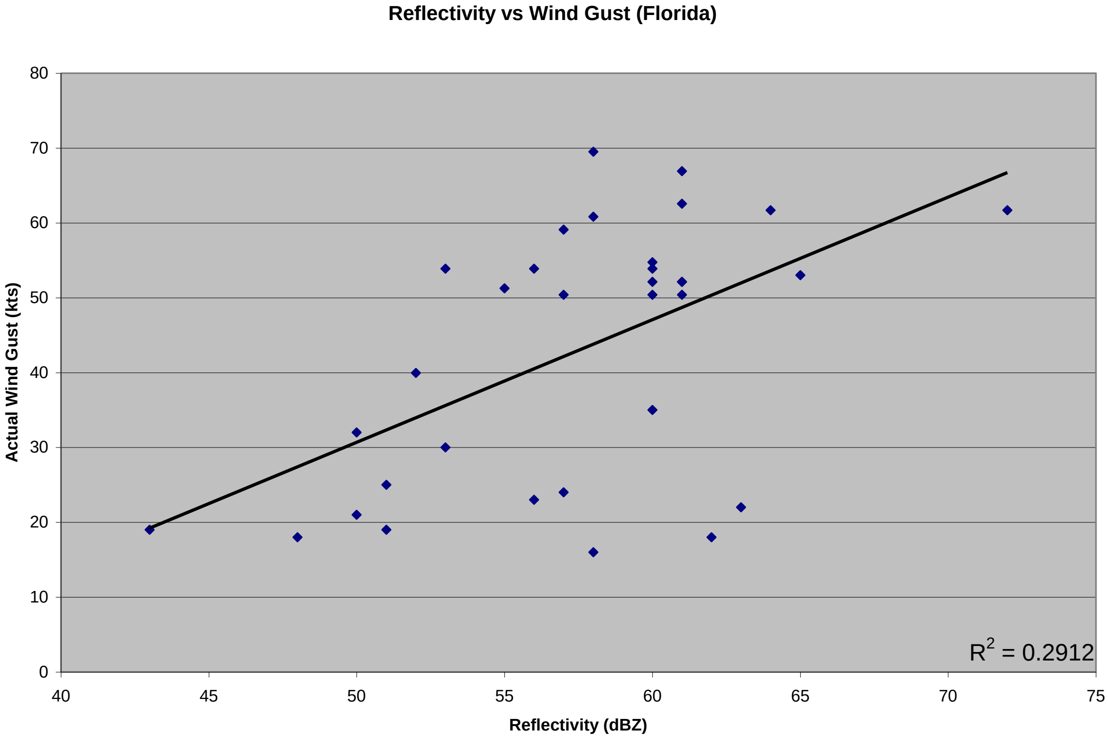

Figure 13: Same as Figure 10, but for maximum reflectivity

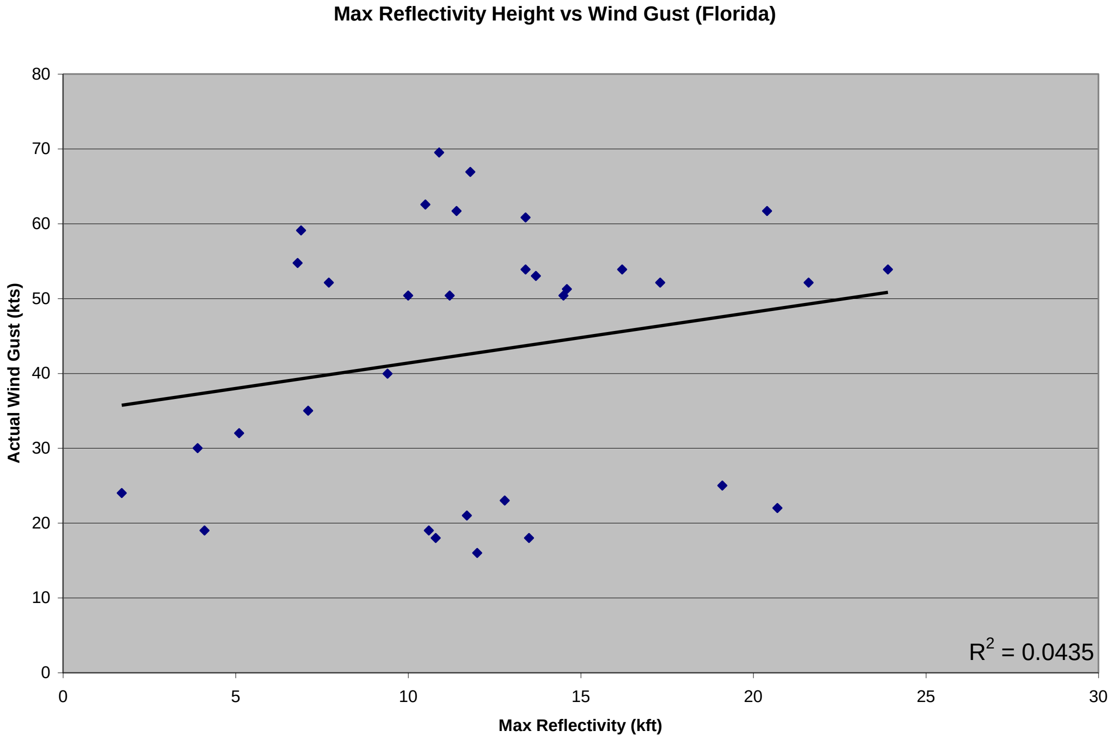

Figure 14: Same as Figure 10, but for height of the maximum reflectivity

d) Multiple Linear Regressions: Northeast Seacoast

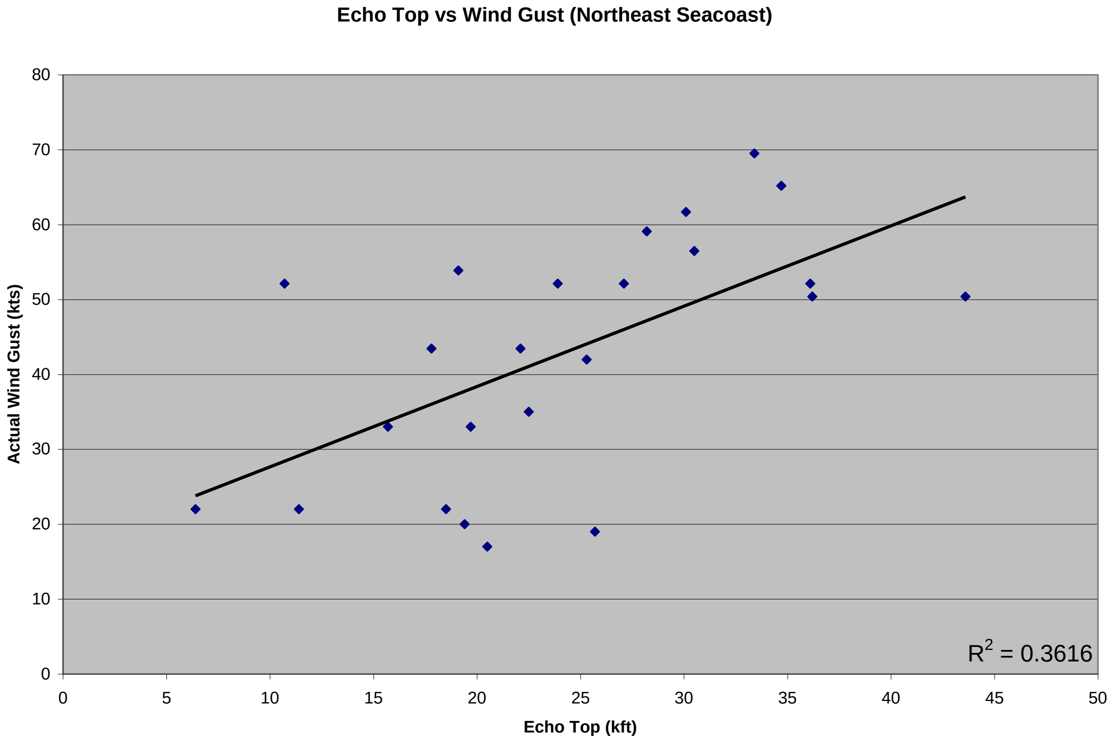

Figure 15: Scatter plot and correlation coefficient of reported wind gust vs. Echo Top for the Northeast Seacoast

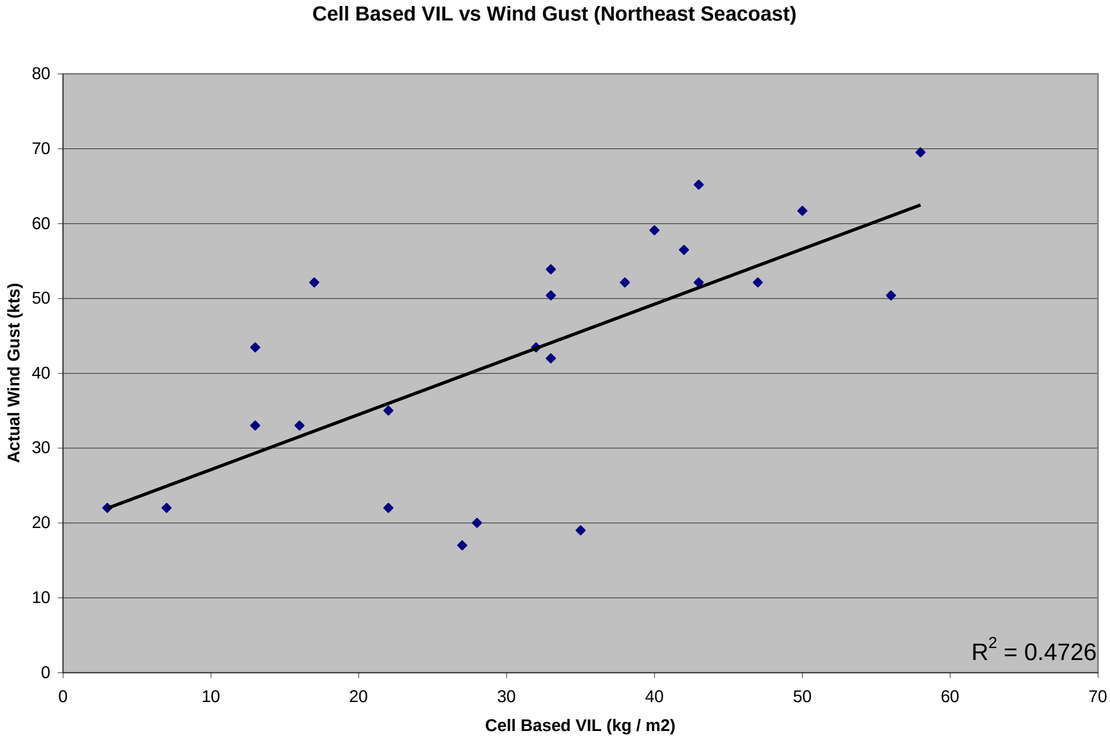

Figure 16: Same as Figure 15, but for cell based vertically integrated liquid

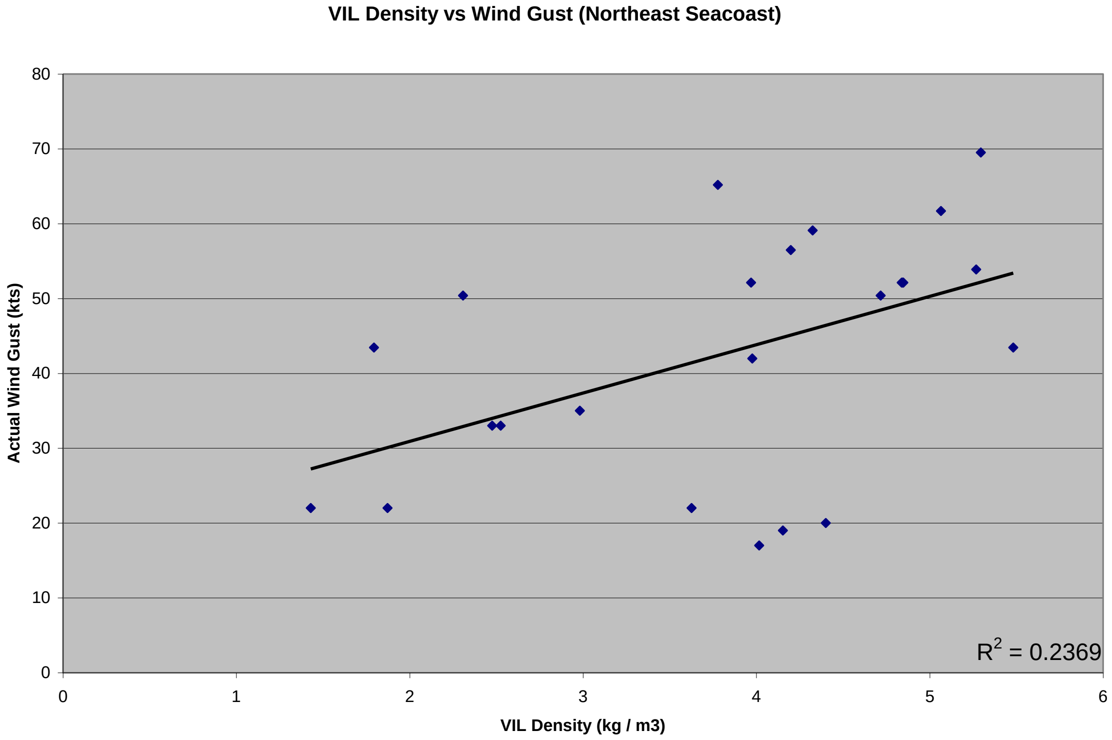

Figure 17: Same as Figure 15, but for VIL Density

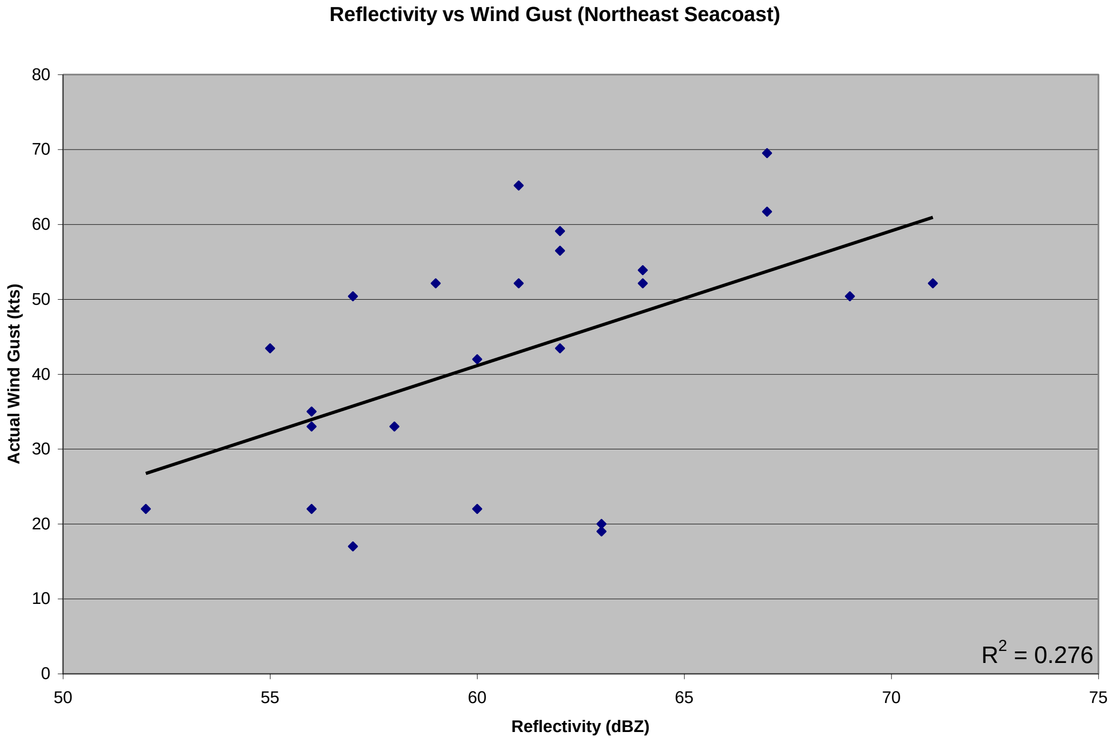

Figure 18: Same as Figure 15, but for maximum reflectivity

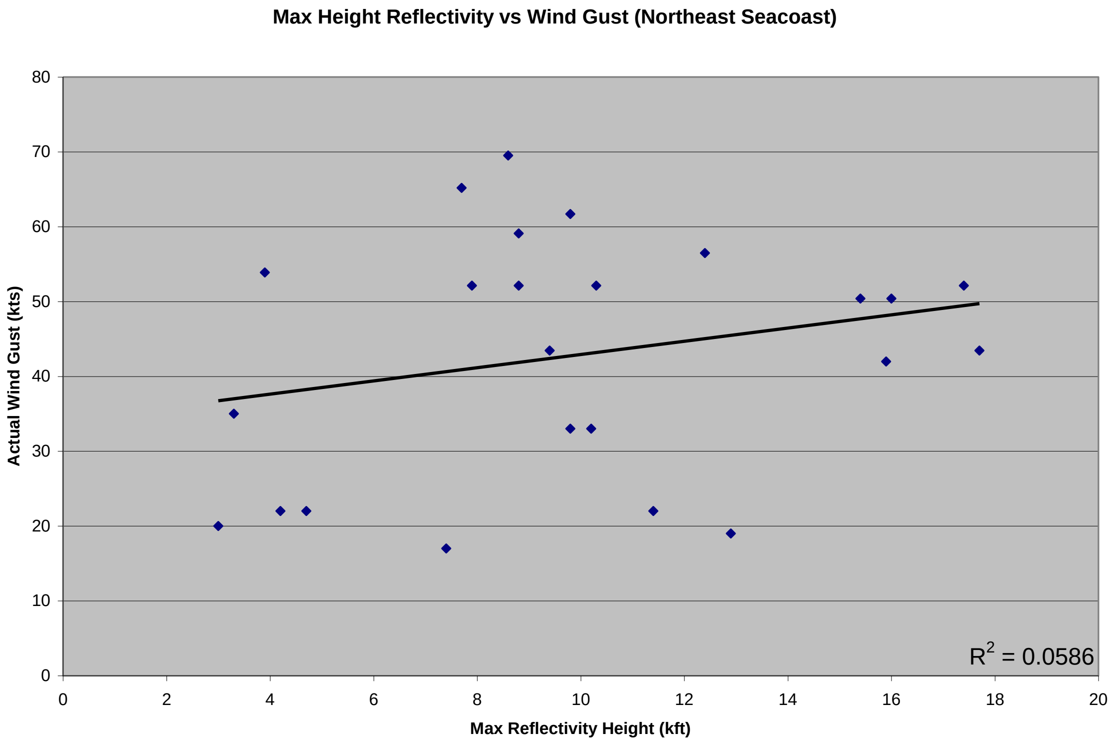

Figure 19: Same as Figure 15, but for height of the maximum reflectivity

e) Best Subsets Function Outputs

```
                                              H
                                              e
                                            M i V
                                        T V a g I
                       Mallows          o I x h L
Vars  R-Sq  R-Sq(adj)       Cp       S  p L Z t D
   1  37.6       35.5      1.3  13.815          X
   1  32.5       30.2      3.6  14.370    X
   2  40.1       35.8      2.2  13.782  X       X
   2  39.7       35.4      2.4  13.830    X     X
   3  42.8       36.5      3.0  13.711  X   X   X
   3  41.8       35.3      3.5  13.836  X     X X
   4  44.7       36.2      4.1  13.743  X   X X X
   4  42.9       34.1      5.0  13.966  X X X   X
   5  45.0       34.0      6.0  13.978  X X X X X
```

Table 7: Output from the best subsets function for the Continental Interior. The correlation coefficient, mallows Cp, and standard deviation values are computed for all the possible variables in the model

```
                                              H
                                              e
                                            M i V
                                        T V a g I
                       Mallows          o I x h L
Vars  R-Sq  R-Sq(adj)       Cp       S  p L Z t D
   1  30.3       28.0      3.3  14.626    X
   1  29.1       26.8      3.8  14.748      X
   2  38.7       34.6      1.4  13.939  X       X
   2  33.9       29.5      3.6  14.482    X     X
   3  41.4       35.3      2.1  13.870  X X     X
   3  38.9       32.6      3.3  14.152  X     X X
   4  41.5       33.1      4.1  14.101  X X   X X
   4  41.5       33.1      4.1  14.102  X X X   X
   5  41.7       30.9      6.0  14.337  X X X X X
```

Table 8: Same as Table 7, but for the state of Florida

```
                                              H
                                              e
                                            M i V
                                        T V a g I
                       Mallows          o I x h L
Vars  R-Sq  R-Sq(adj)       Cp       S  p L Z t D
   1  47.3       44.9     -1.5  11.908    X
   1  36.2       33.3      2.4  13.102  X
   2  48.1       43.1      0.2  12.095    X   X
   2  47.9       42.9      0.3  12.115  X X
   3  48.3       40.5      2.2  12.366  X X     X
   3  48.3       40.5      2.2  12.367  X X   X
   4  48.8       38.0      4.0  12.629  X X   X X
   4  48.3       37.4      4.2  12.686  X X X   X
   5  48.8       34.5      6.0  12.975  X X X X X
```

Table 9: Same as Table 7, but for the Northeast Seacoast