Paper 1789-2014

Visualizing Lake Michigan Wind with SAS® Software

Aaron C Clark, Grand Valley State University; David Zeitler,
Grand Valley State University

ABSTRACT

A wind resource assessment buoy, residing in Lake Michigan, uses a pulsing laser wind sensor to measure wind speed and direction offshore up to a wind turbine hub-height of 175m and across the blade span every second. Understanding wind behavior would be tedious and fatiguing with such large data sets. However, SAS/GRAPH® 9.4 helps the user grasp wind characteristics over time and at different altitudes by exploring the data visually. This paper covers graphical approaches to evaluate wind speed validity, seasonal wind speed variation, and storm systems to inform engineers about the energy potential of Lake Michigan offshore wind farms.

KEYWORDS

Big data, time series, visualization, offshore wind farms, renewable energy, Lake Michigan, moving window, PROC SGPLOT, PROC GRADAR

INTRODUCTION

Wind speeds off of large bodies of water have often been hailed for their prime wind energy candidacy, and Lake Michigan is no exception. The U.S. Department of Energy's (DOE) National Renewable Energy Laboratory (NREL) has cited Lake Michigan with an "outstanding" wind resource with the potential of generating and estimated 600- 800 watts/minute^2 at 50 meters above the water surface in 2007. In a continued effort to explore the candidacy of wind energy on Lake Michigan and to compromise with shoreline property owners' claim that turbines contribute to visual pollution, offshore wind farms propose a promising alternative. In 2012 Grand Valley State University deployed a wind resource assessment buoy called the WindSentinel™ in Lake Michigan's mid-lake plateau, 35 miles west of Muskegon, MI, a prime area for development in approximately 250 feet of water. The feature technology: a pulsing laser wind sensor (LWS) is mounted on the buoy to accurately measure wind speed and direction up to a wind turbine hub-height at 175m and across the blade span every second. Predecessors to the WindSentinel™ would aggregate wind speeds to ten minute averages only a few feet above the water, lacking detailed data. The WindSentinel's™ primary objective is wind monitoring using the LWS, but many water, atmospheric, and bird/bat characteristics are also captured using other onboard devices.

One challenge to determining wind farm plausibility on Lake Michigan is confirming the validity of the wind measurements we observe in preparation for data analysis. In some experimental high altitude cases, the LWS struggles to collect consistent and validated wind speed records due to lack of reflective particulates and movement of existing particles in the atmosphere over the open water. Data quality indicators provided by the LWS vendor have proven to flag relatively good data as bad. Therefore, before exploration of seasonal and storm activity on the lake for a turbine-friendly assessment, we need to ensure we are examining all valid data values. Visualizing the state of these data using SAS® 9.4 will prove useful to identify "bad data" and inform an algorithm to sort it from the "good data." For example, at times, reported wind speeds are too constant or too extreme to be real.

The second challenge is previewing seasonal variation with inherent patterns in the wind behavior. Understanding the seasons will be important in forecasting how often a wind turbine is operating with optimal power output limitations. That is, optimal power output cannot be achieved if the wind is too fast or too slow; when it is, the turbine will not collect any wind, and will shut itself off. Since wind is cyclical, we need to understand how often our target wind speed is maintained during these seasonal variations.

The last undertaking is to explore the phenomenon of storms to reveal any possible challenges for turbines in the middle of the lake. During high wind storms, can we expect the turbines to operate? Is the WindSentinel™ collecting valid wind during these storm periods? These concerns will be addressed.

Considering all these challenges, visualizations of these data will help the user gain a clear perspective on activity and key wind characteristics much easier for short periods of time than descriptive statistics would be able to portray. This paper will serve as a guide to how we approached these large time series datasets using visualizations.

THE WIND TECHNOLOGY AT A GLANCE

The Grand Valley State University-owned WindSentinel™ is a product of AXYS Technologies Inc., and the first wind assessment tool of its kind in the world. Its monitoring systems include wave, current, water quality, water temperature, basic wind (anemometers), atmospheric pressure, solar radiation, laser wind, sonar, and audio recordings. The vessel is approximately 15 feet in length and capable of powering all its systems with an onboard turbine and solar panels and storing the energy in forty batteries located in the hull. In case no wind or sun energy is available, a backup diesel generator lies within the hull as well. As you can see in diagram 1 (right), the LWS monitors wind at six distinct altitudes called range gates (RG). Their altitudes are 75m, 90m, 105m, 125m, 150m, and 175m for range gates 1 through 6, respectively. The laser sensor pulses more than 600 times per second in three angled directions, auto-correcting for buoy movement.

The buoy’s location during the 2012 season (May 7 to December 19) was approximately W43^{o} 20’ 31.20’’, N -87^{o} 7’ 12.00’’. The offshore wind assessment project is three years in total, spending time on Muskegon Lake, the Mid-Lake Plateau, and various other locations on Lake Michigan. To date, over 65 individuals from various institutions have collaborated on the project, including the U.S. Department of Energy, Federal and State agencies, Grand Valley State University, Michigan State University, University of Michigan, University of Delaware, and Michigan Tech University.

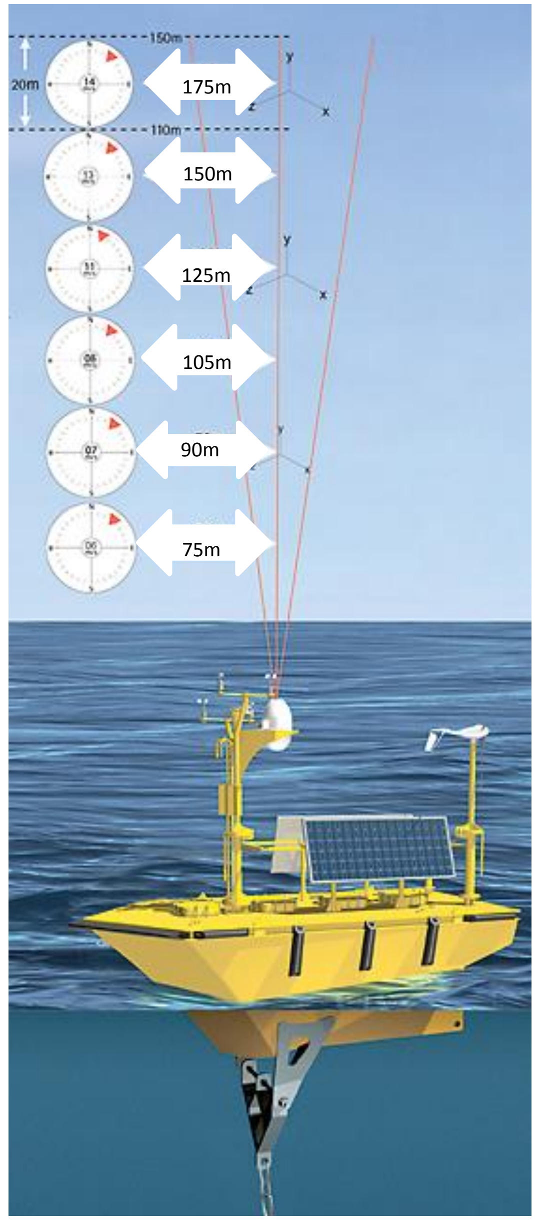

Diagram 1: Conceptual view of WindSentinel™ configuration and the Laser Wind Sensor’s 6 range gates

VISUALIZING WIND VALIDITY

We have a number of SAS/GRAPH® tools at our disposal to visualize invalid data. Range gate 6 (175m) is a test range gate to observe the performance of the LWS at an extreme operating height limit for this configuration. Thus, performance degradation was expected. To analyze only valid data at each range gate, we will build a quantitative and indicator measure to identify wind speeds that are extremely high or too constant to be real.

The former two scenarios are displayed in figure 1 (bottom right) which previews wind at RG6 on July 1st, 2012. A trivial PROC SGPLOT with a SCATTER statement produce figure 1:

```
PROC SGPLOT DATA=LIB2.JULY1;
  SCATTER Y= WindSpeedHorRG6 X=TS;
RUN;
```

Occasionally, the LWS will record an observation not characteristic of the wind speeds surrounding it. For example, figure 1 displays an extreme point at around 4am nearing 48 m/s (~107 mph) which is clearly unrealistic, especially considering the wind activity during the rest of the day. Therefore, we must train our validity indicator to recognize such occurrences. This is easily achieved by scanning the data for high wind speeds, taking into account wind before and after each occurrence. Visualizations such as figure 1 are all that are needed to show these instances.

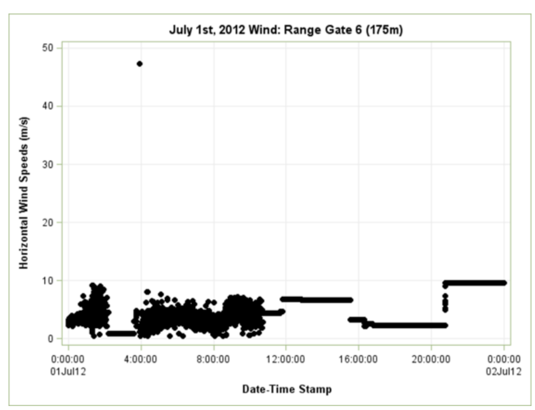

Figure 1: Range Gate 6 (175m) wind speeds on July 1st, 2012

Next, identifying wind speeds that are too constant to be real is slightly more complex and requires a computational solution since short time intervals of "dead wind" are difficult to observe with the naked eye. Figure 1 has some obvious instances of this, especially after noon. These winds cannot be real for three reasons:

1. Winds come and pass in gusts. That is, it is unlikely for wind to maintain a constant speed for more than a couple seconds.
2. Our measurements extend out to the tenth of an m/s, so exact readings for consecutive time stamps are less probable to this degree.
3. The LWS measurements are inherently variable. As a result, it will likely never measure the exact same wind speed for consecutive time stamps.

Our solution utilizes a 5-second moving window standard deviation (5-sec stddev). This statistic satisfies our need for a quantitative measure of variability because the window is short enough to measure delicate spikes in wind and long enough to be conservative about how long wind can remain constant. Using this information, we produce figure 2, a display of the relationship between wind speeds and their 5-sec stddev by adding another SCATTER statement to the code that produced figure 1. Focusing on figure 2, whenever the wind speeds remain constant, the 5-sec stddev equals zero. Thus, we can train our validity indicator to classify "bad data" as having a zero value for this statistic. Grouping by the validity indicator using a GROUP= option on the SCATTER statement produces figure 3 below.

```
PROC SGPLOT DATA= JULY1;
 SCATTER Y= WindSpeedHorRG6 X=TS;
 SCATTER Y=_5sSdRG6 X=TS;
 YAXIS MIN=0 MAX=11 LABEL="Horizontal Wind Speeds (m/s)";
RUN;
DATA JULY1;
 SET JULY1;
 IF _5sSdRG6 = 0 THEN ValidRG6 = 0;
                ELSE ValidRG6 = 1;
RUN;
PROC SGPLOT DATA= JULY1;
 SCATTER Y= WindSpeedHorRG6 X=TS / GROUP=ValidRG6;
 YAXIS MIN=0 MAX=12.5 LABEL="Horizontal Wind Speeds (m/s)";
RUN;
```

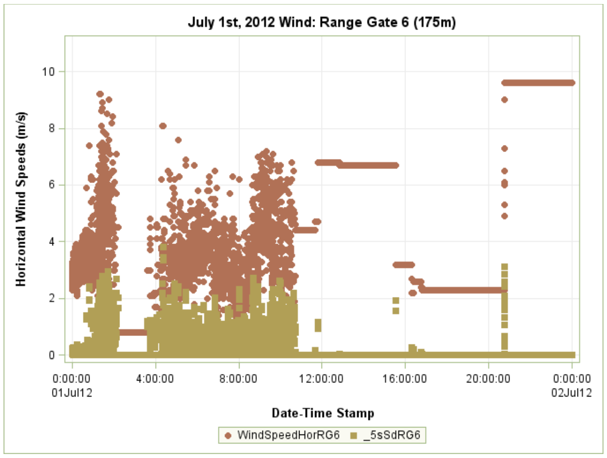

Figure 2. Display of 5-second moving window standard deviation against wind speeds

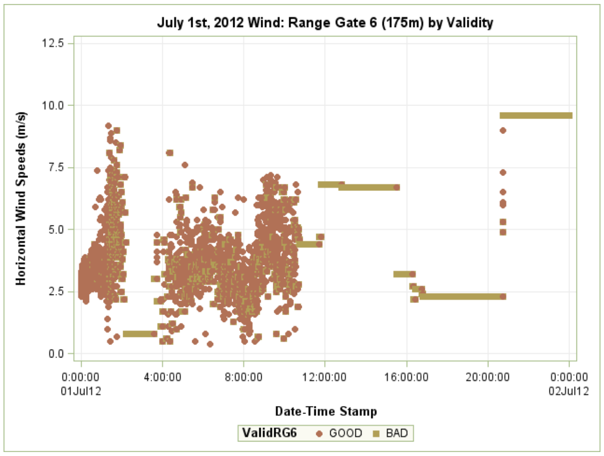

Figure 3. Display of wind speeds grouped by manually built validity indicator

Note that we not only identify obvious constant wind that occurs after noon on July 1st, but we also captured short time intervals (1 to 3 minutes, for example) where the wind remained constant that were not visible at this day-level visualization. Now that the data has been classified as "good" and "bad", it is appropriate to explore the data further. The tools identified in this section will prove useful in dissecting key wind characteristics in the remainder of the report.

VISUALIZING SEASONAL VARIATION

Your average 850kW wind turbine, the prospective model appropriate for offshore wind farms in Lake Michigan, generates the most energy when capturing wind between 11 and 14 m/s. However, the turbine will remain ON as long as the wind is between 4 and 14 m/s. Any wind outside of this range is not sufficient on the low end and potentially damaging to the turbine on the high end, so the turbine will shut off. Seventy-three percent of the 2012 data from the mid-lake plateau location was between these values. From now on, our visualizations will include a horizontal reference lines at 4 and 14 to offer this perspective. To accomplish this, we only need the following in any SGPLOT procedure:

```
REF 4 14;
```

To attain the “big picture” perspective that visualizing seasonal wind activity seeks, detail of second-level data in our plots would be quite excessive. That is, the 2012 season is composed of roughly 23 million data observations. To plot such detail on in one pass would prove useless as the points would be too packed, plotted over one another, and take the form of a big blob. Therefore, we have opted to utilize ten minute averages of the data, reducing our total observation count to approximately 38 thousand and smoothing out possible outliers. To reduce the detail even more, we ask SAS® for only two of these observations be plotted for each day: the max and min. The result (figures 4- 6) will provide a “channel” of possible wind values measured by day when using two SERIES statements (one for day-minimums and one for day-maximums) in PROC SGPLOT, viewed by season. An alternative would be to use PROC SGPANEL and a statement to PANELBY season. We will choose the former so that we can append more graphics that assess seasonal data, and group them accordingly.

Namely, we will create a windrose plot for each season to evaluate the frequency of several wind speed magnitudes and wind directions in one plot. To do so, we create discrete categories for wind speeds and directions and produce a cross tabulation of their frequencies for input into PROC GRADAR.

The SAS® syntax below illustrates the entire seasonal visualization process:

```
PROC SGPLOT DATA= Seasons;
  SERIES Y=dayMIN X=DateStamp;
  SERIES Y=dayMAX X=DateStamp;
  BY Season;
  REFLINE 4 14;
RUN;
PROC FREQ DATA = Midlake;
  TABLES DiscreteAvgDir*DiscreteAvgSpd/ NOROW NOCOL OUT= Freqs;
BY Season;
RUN;
DATA FALLFREQS; SET Freqs; IF Season="Fall"; RUN;
DATA SPRINGFREQS; SET Freqs; IF Season ="Spring"; RUN;
DATA SUMMERFREQS; SET Freqs; IF Season ="Summer"; RUN;
PROC GRADAR DATA= SPRINGFREQS;
  CHART DiscreteAvgDir / sumvar=percent windrose noframe speed=DiscreteAvgSpd;
RUN;
```

Note that PROC GRADAR does not operate with a BY statement, so three separate procedures are executed, one for each season. The spring season was slightly shorter than the summer and fall seasons as the buoy was deployed May 7, well into this time period. As a result, figure 4 may appear less cluttered than the others. Generally, the averaged wind speeds are mostly within the 4 and 14 m/s turbine constraints, with a tendency to drop below the lower bound more frequently than rise about the upper bound. The percentage of usable wind during the spring season is about 71.18%, not far from the overall percentage of usable wind in 2012 (73%). The wind is primarily out of the south with wind speeds frequent in the 5-10 m/s range.

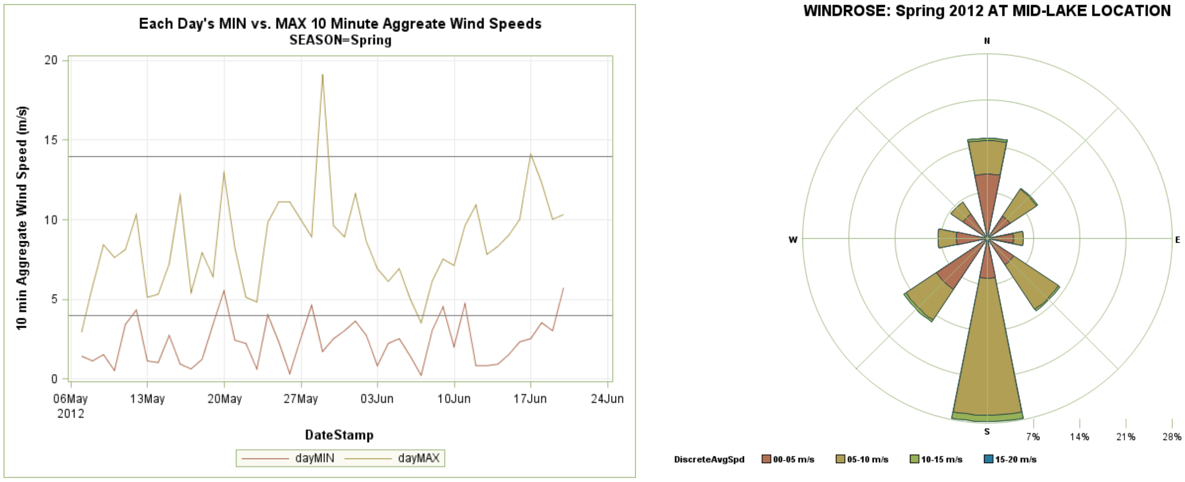

Figure 4. Min and max 10-minute wind speed averages per day, for spring 2012 season with windrose

Next, figure 5 (below) displays a similar trend observed in figure 4. That is, there is more unusable wind below the lower constraint than above the upper constraint. Sixty five percent of the wind speeds during this season were between 4 and 14 m/s, much lower than in the spring. Last, wind from the south and southwest at 5-10 m/s is most frequent, though there seems to be some stronger winds (10-15 m/s) that came from the north briefly.

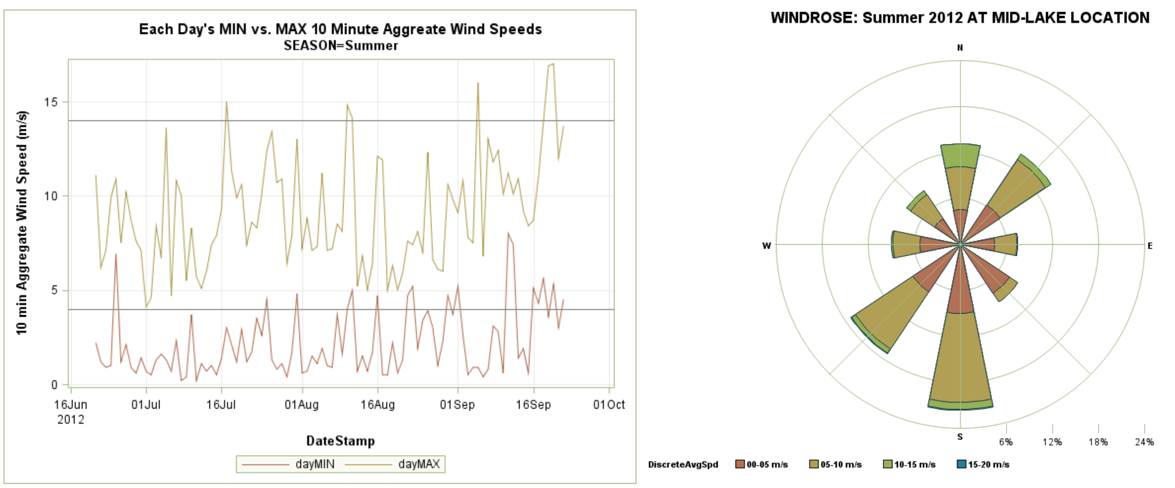

Figure 5. Min and max 10-minute wind speed averages per day, for summer 2012 season with windrose

Finally, figure 6 (below) displays the results for the fall season:

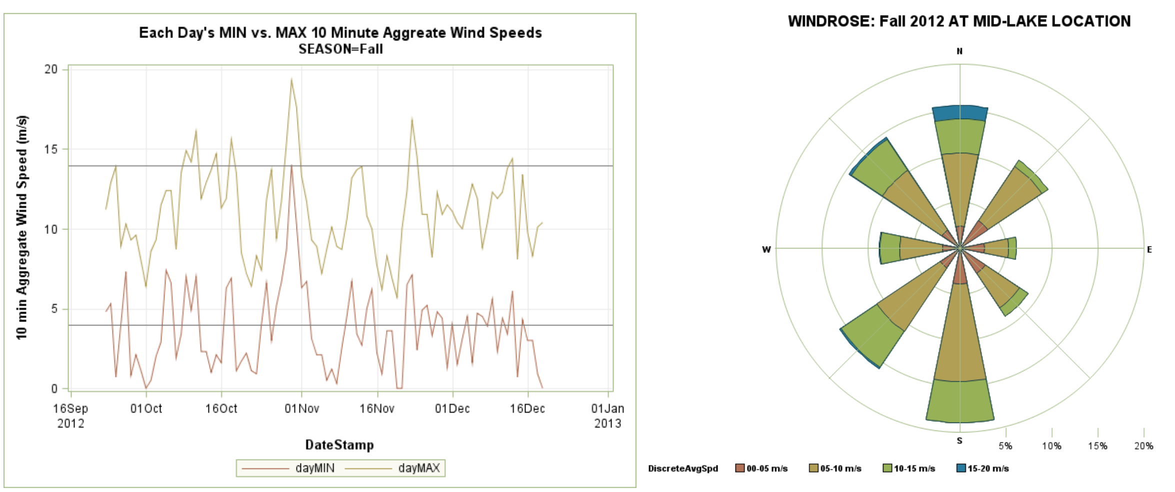

Figure 6. Min and max 10-minute wind speed averages per day, for fall 2012 season with windrose

During the fall season, 82.8% of the wind fell within the usable constraints defined by an 850kW turbine. That makes this season's usable wind the highest out of all other season. However, around the end of October, the min and max 10-minute averaged wind speeds rise above 14 m/s. This phenomenon is due to residual effects of the devastating Hurricane Sandy that hit the east coast impacting the U.S. all the way up into Lake Michigan. Actual (not averaged) wind speeds reached gale force levels of 26 m/s (58 mph). One gust reached 29.9 m/s (or nearly 67 mph). As such, the windrose also displays significantly higher winder wind speeds in accordance with this finding, but with wind coming from the S, SW, NW, and N a fairly uniform amount of time. The may be due to the circular motion of hurricane storm systems.

In conclusion, the summer months produced the least amount of usable wind while the fall produced the most. However, it is unknown whether the fall season yielded more wind due to high winds from Hurricane Sandy, or for some other reason. A SERIES statement for the mean day 10-minute averaged wind is an alternative to using the min and max. We chose the latter route solely for its "channel" like properties described earlier. A secondary method of visualizing this "channel" is to use a HILOW statement in SGPLOT procedure. This method will plot vertical lines between any two values specified for each day, such as max/min values or even upper/lower bounds to a 95% confidence interval.

VISUALIZING STORM SYSTEMS

As discussed for figure 6's display of hurricane Sandy in *Visualizing Seasonal Variation*, storms play an important role in the functionality of turbines and in the data collected on Lake Michigan. From figure 6 (above), we know that the turbine would shut off because the winds were much too extreme during Hurricane Sandy, a truly powerful storm. However, those winds were not typical of most storms on Lake Michigan. In other words, some storm's wind speeds may be in the "usable range" (between 4 and 14 m/s), but if they are too variable, then the turbine has potential to cease function until winds are more optimal. Additionally, we will explore if storms are affecting the validity of the data collected by the LWS at different altitudes. To achieve both goals, we provide the following example that summarizes what we found to be typical among storms during the 2012 season. Figure 7 displays a brief storm that took place on July 31st, 2012 from approximately 4am to 7am that uses three SCATTER statements (one for each range gate) in a PROC SGPLOT.

```
PROC SGPLOT DATA= Midlake_JULY;
  WHERE DATATIMESTAMP>="31JUL12:00:00:00"DT AND DATATIMESTAMP <="31JUL12:23:59:59"DT;
  SCATTER X=DATATIMESTAMP Y=WindSpeedHorRG1;
  SCATTER X=DATATIMESTAMP Y=WindSpeedHorRG3;
  SCATTER X=DATATIMESTAMP Y=WindSpeedHorRG5;
RUN;
```

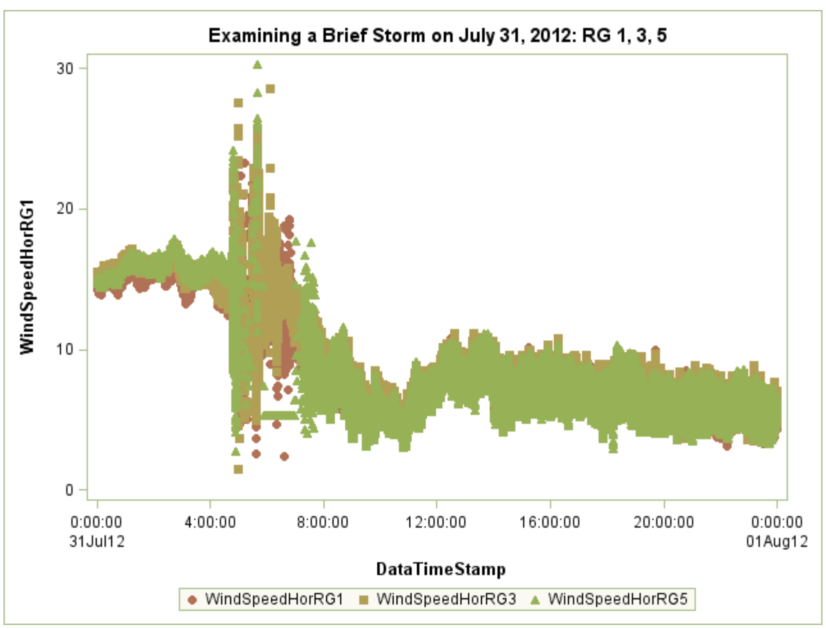

Figure 7. Display brief storm at RGs 1, 3, and 5 on July 31st, 2012

To summarize figure 7: there is data degradation in the higher range gates while the lower range gates records remain intact. There also appears to be a great deal of variation in wind speeds inside the storm vs. the wind before and after the storm. Therefore, we will consult our 5-sec stddev statistic discussed in *Visualizing Wind Validity* and zoom-in on the storm to understand what is happening here. Figure 8 is the result; below is the SAS® syntax required:

```
PROC SGPLOT DATA=JUL31;
  WHERE TS >="31JUL12:04:00:00"DT AND TS <="31JUL12:07:00:00"DT;
  SCATTER X=TS Y=WindSpeedHorRG2 /GROUP=StatusRG2DataGood;
  SCATTER X=TS Y=_5sSdRG2 / GROUP=StatusRG2DataGood;
  REFLINE 4 14;
  XAXIS GRID;
  YAXIS GRID LABEL="WIND SPEED (M/S)";
  TITLE "Examine Wind Variability using 5-sec Stddev by Validity";
RUN;
```

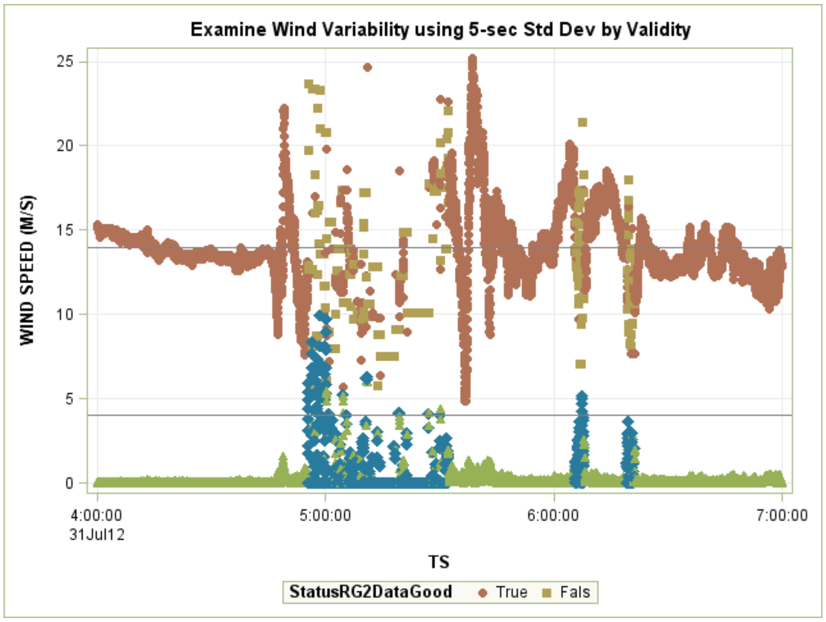

Figure 8. Zoomed-in display of July 31st, 2012 storm wind speeds and 5-sec stddev at Range Gate 2, by validity indicator

By default, SAS® displays a legend for the first grouped data supplied via SCATTER statement with a GROUP= option (the validity indicator for wind speeds), but it is easy enough to understand that the blue diamond symbols correspond to “bad data” and the green triangle symbols correspond to “good data” for the 5-sec stddev. From figure 8, we learn that the variability we observed in figure 7 is not real wind, but is actually invalid data. From the valid data, we can conclude that storm winds vary slightly more than non-storms, but not by as much as originally anticipated. The highest valid 5-sec stddev appears around “5” at 5am, which is half the invalid value observed only few minutes early at “10”. In conclusion, most storms we examined displayed this characteristic: anticipated variable winds were not actually variable, but just invalid. Therefore, further exploration is needed to determine if turbines are subject to winds that are “too variable” and, thus, require shutting down due to inconsistency.

CONCLUSION

Now that we know how to watch out for invalid wind data, wind analysts can properly address how seasonal wind variation and storms will relate with turbine function for an offshore wind farm in Lake Michigan. Undoubtedly, SAS/GRAPH® 9.4 is an exceptionally capable visualization tool to dissect and interpret key wind attributes. The SAS® software will surely aid wind analysts in informing engineers, investors, and developers on the plausibility of Wind Farms on Lake Michigan. However, the wind candidacy is only one small component of wind farm development. Other assessments need to be made on social, economic, political, technologic, and environmental concerns before justifying renewable wind energy on the great lakes. As far as this assessments goes, the wind is there.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Aaron Clark
Organization: Grand Valley State University, Padnos College of Engineering and Computing, Grand Rapids, MI
Email: clarkaa@mail.gvsu.edu
Web: www.linkedin.com/in/dataAaronClark