The Reconstruction of SDHN4 Wind Data Using A Multiple Linear Regression

Joshua Coupe

Introduction

On October 28th, Hurricane Sandy underwent an extratropical transition that not only failed to weaken the storm, but led to an expansion in the storm's wind field (Galarneau et al., 2012) Hurricane Sandy brought major coastal flooding mostly to parts of New Jersey and New York. In addition, the hurricane force winds and storm surge led to the destruction of the meteorological equipment at the Sandy Hook (SDHN4) station. The equipment at Sandy Hook was not replaced until January 2014, leaving an enormous data gap. Fortunately, offshore buoys remained intact and continued to record wind observations. This analysis has the following aims: (1) To determine the veracity of the RU-WRF model in terms of windspeed and direction at Sandy Hook and buoys further offshore, (2) to determine the relationship between windspeed and direction at the land based SDHN4 station and ocean based buoys 44065 and 44025, (3) and to reconstruct the SDHN4 station wind observations, using both buoy observations and RU-WRF model winds.

Methods & Data

The RU-WRF numerical weather prediction model is run twice daily by Rutgers University at 9km, 3km, and 1km resolution at 00Z and 12Z. The 3km model is used in this analysis because it has the least amount of gaps and can reasonably be expected to represent the winds that the observations show. The model has been run nearly continuously since April 2012. To construct a continuous timeseries of RU-WRF data, forecast hours 6 through 18 from each 00z and 12z run are spliced together. The first 5 hours of every forecast are not used to avoid errors due to model spin-up (Weiss et al., 2008). The RU-WRF model incorporates an advanced SST algorithm to resolve the unique wind dynamics relevant to the areas offshore of New Jersey (Glenn and Dunk, 2013). The initial conditions for the RUWRF model are on a 13km resolution grid using the NCEP Rapid Refresh assimilation/modeling system. Boundary conditions for the 3km RU-WRF are based on the NCEP North American Mesoscale (NAM) modeling system at 12km resolution.

The observations which were compared to the model are a combination of buoys and land based meteorological stations, as seen in Figure 1. All buoy/meteorological stations are obtained from the National Data Buoy Center (NDBC). SDHN4 is a land based meteorological station recording wind speed and direction. Data from 2012 through the end of 2016 is used, with the exception of the period between October 28th, 2012 and January 1st, 2014. Following Superstorm Sandy, SDHN4's anemometers were destroyed and not replaced until 2014. Buoys 44065 and 40025 off of Long Island survived and continued to report wind speed and direction. Data for buoys 44065 and 44025 were nearly continuous for the 2012-2016 period. To reconstruct the Sandy Hook windspeed and direction data for October 28th, 2012 through

January 1st, 2014, a number of multivariate linear regressions between the model wind speeds, the two offshore buoys, and SDHN4 will be performed for SDHN4 to determine the sensitivity of SDHN4 to variations in the other datasets. The model that explains the highest variance will be used to reconstruct the period with the missing data.

Model Verification

To verify that the RU-WRF model simulates coastal winds, the methodology of a previous study was adapted on all available RU-WRF 3km data. (Glenn and Dunk, 2013). A combination of verification metrics from Dvorak et al., (2012) and from The National Renewable Energy Laboratory, or NREL, have been used to assess RU-WRF's handling of the buoy and land station network. These metrics are as follows:

(1) The standard deviation of the RU-WRF model winds must be similar to the standard deviation of the observed winds, therefore the percent-difference must be no greater than 10%.

(2) The centered root mean square error between the observations and the model winds must be less than the standard deviation of the observed winds, where

centered root mean square error $= \sqrt{\frac{1}{n}(\Sigma\,[(\widehat{y}_i - \overline{\widehat{y}_i}) - (y_i - \overline{y_i})]^2)}$, $\widehat{y}$ is the model windspeed, y is the observed windspeed, $\overline{y}$ represents the mean of that variable, and n is the number of datapoints.

If a station fails any of the two conditions, it fails the verification process. RU-WRF was verified to adequately represent the 10 meter winds at almost a dozen buoy and weather stations as shown in Figure 2. Buoys that passed include 44009 at Delaware Bay, 44025 at W. Long Island, 44065 at C. Long Island, 44017 at E. Long Island, 44014 off Cape Hatteras, CHLV2 or the Chesapeake Lighthouse, 44066 at the Texas Tower, and BRND1 at Brandywine Shoals. Observation stations that failed to verify include BUZM3 at Buzzards Bay, SJSN4 at Saint John Shoals, and also SDHN4 at Sandy Hook.

Stations that did not meet the verification standards for the 2012-2016 time period share a common trait in that they are either land-based meteorological stations or are very close to the land. Additional friction from the land surface is initially hypothesized to be the reason for this inability to verify. Verification results from a comparison of RU-WRF model wind speeds with a number of NJ area buoys can be seen in Table 1.
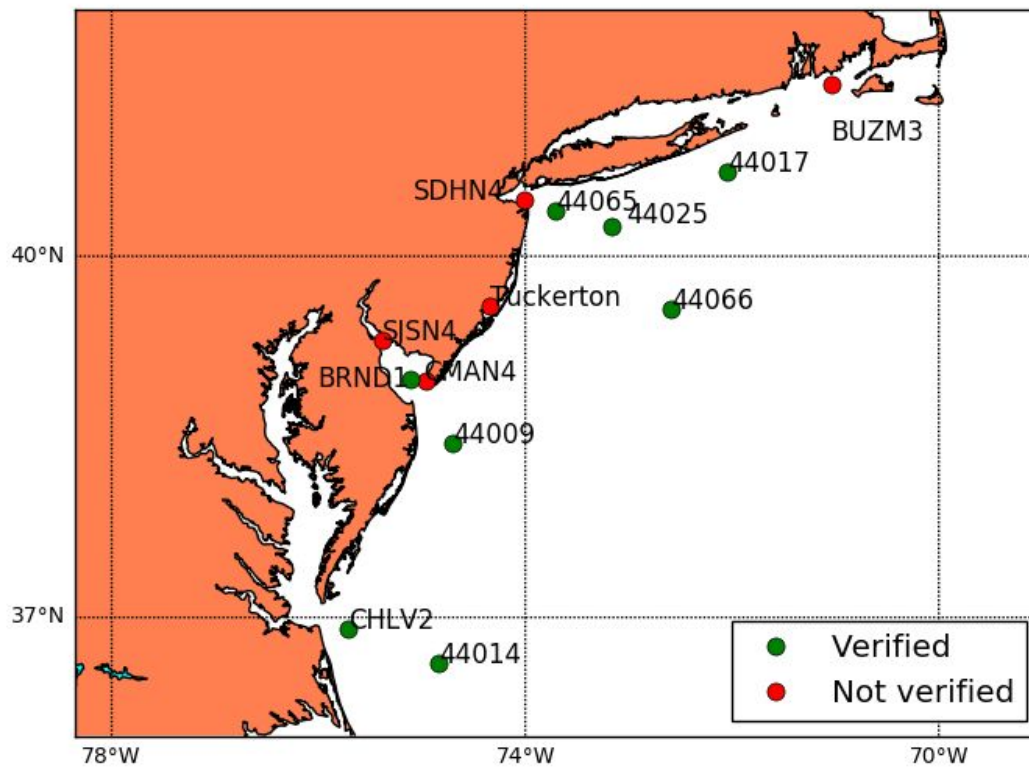
Figure 2. Stations whose observations were compared with RU-WRF model output.

2012-2016	44009	44025	44065	44017	44014	CHLV2	44066	BRND1H	BUZM3H	SJSN4H	SDHN4	CMAN4
WRF SD (U,V)	4.508, 5.152	4.952, 5.077	4.758, 4.878	5.252, 5.111	4.268, 5.637	5.193, 6.741	5.397, 5.204	4.795, 5.245	4.991, 4.975	3.048, 3.491	2.818, 3.059	4.434, 5.064
Obs SD (U,V)	4.406, 5.426	5.045, 5.144	4.979, 5.043	5.479, 4.882	4.294, 5.838	5.059, 7.113	5.586, 5.347	4.822, 5.583	5.529, 5.967	4.497, 4.537	3.819, 3.341	3.164, 3.204
% diff (U,V)	2.3%, 5.1%	1.85%, 1.30%	4.4%, 3.3%	4.2%, 4.5%	0.6%, 3.5%	2.58%, 5.2%	3.4%, 2.7%	5.5%, 6.1%	9.7%, 16.6%	32.2%, 32%	26.2%, 8.4%	28.6%, 36.7%
CRMSE (U,V)	2.287, 2.683	2.135, 2.347	2.328, 2.459	2.223, 2.340	2.285, 2.715	2.986, 3.305	2.699, 2.781	2.859, 2.894	2.661, 2.959	2.985, 2.564	2.201, 2.096	5.402, 5.878
# of points	9598	22578	29172	19816	24153	26598	13418	18277	32616	30429	16063	31187
Height (m)	5	5	5	5	5	43.3	5	19.3	24.8	20.4	5.5	9.7

Table 1. Verification statistics for buoys/towers for 2012-2016.

Non-verifying stations

Because the RU-WRF model fails to adequately represent higher order variability in the land based observations, an analysis of the model errors is performed to eliminate systematic biases. The V component winds at SDHN4 met the threshold for verification, while the U component winds failed. The SDHN4 observations are taken at a location with land to the north and east and open water to the south, leading to the hypothesis that north and easterly winds would be compromised. However, the CRMSE for easterly winds was 2.01 m/s, while westerly winds was 2.5 m/s. Greater error existed for westerly winds, which were also twice as common and likely to include wind gusts. This is demonstrated by the higher standard deviation for westerly winds, 3.49 m/s, compared to easterly winds, 2.51 m/s. Normalized, this results in a CRMSE of 0.8 standard deviations for U winds and 0.7 standard deviations for V winds, a negligible difference. Visualizing differences between the forecast winds and the observed winds is typically done with a quantile-quantile plot, shown in Figure 3. Each point represents a percentage of the distribution of points at that wind speed for both model and observations. At slow wind speeds, the model and observations typically agree, but at much higher windspeeds, the model has a tendency to overestimate. Representing friction on the coastline is incredibly difficult, and if the land area around SDHN4 is not completely flat, winds may be artificially higher or lower than the model's idealized conditions at any given time. Ultimately, a systematic bias was very difficult to locate and correct for.

Figure 3. Quantile-quantile plot showing the distribution of observed windspeeds at SDHN4 compared to RU-WRF windspeeds at SDHN4 over 2012-2016 in meters per second.
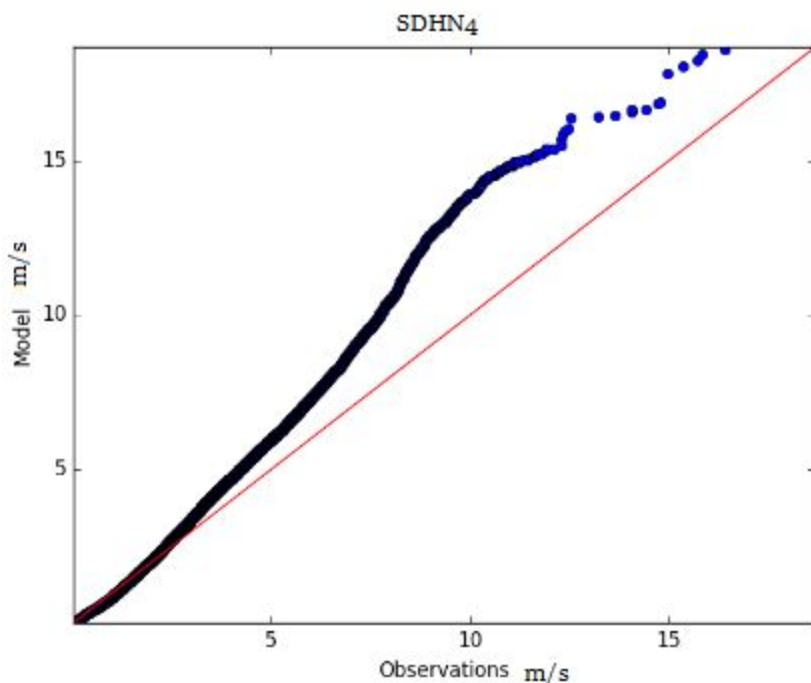
Multivariate linear regressions

To reconstruct the missing SDHN4 data, a multiple linear regression of SDHN4 based on observations at 44065 and 44025, model data at SDHN4, 44065, and 44025 is performed. This model explains 86% of the variance. Although buoys 44065 and 44025 have a nearly complete data record data, inevitably some data is missing during the period of time of missing SDHN4 data. Additionally, there are small gaps in the RU-WRF model data. Therefore, three different regression models are used to reconstruct the missing SDHN4 observations. When buoy 44065 is missing data, the RU-WRF and buoy 44025 data are used to develop the model. Likewise, when buoy 44025 is missing data, RU-WRF and buoy 44065 are used. These models described 79% and 81% of the variance of the SDHN4 observations, respectively. The three different regressions for the three different circumstances are as follows:

1. All available data:
Regression for u = 0.3198*WRF_sdnh4 + 0.1029*buoy_40025 + -0.0492*WRF_44025 + 0.0322*WRF_44065 + 0.5034*buoy_44065
Regression for v = 0.3003*WRF_sdhn4 + -0.0771*buoy_44025 + -0.0412*WRF_44025 + 0.0757*WRF_44065 + 0.4902*buoy_44065

2. Only model data available:
U regression = 0.3775*WRF_44025 + 0.0064*WRF_44065 + 0.5537*WRF_sdhn4
V regression = 0.2334*WRF_44025 + 0.0249*WRF_44065 + 0.5698*WRF_sdhn4

3. Only observations available:
U regression = 0.1746*buoy_44025 + 0.5646*buoy_44065
V regression = -0.0631*buoy_44025 + 0.6508*buoy_44065

Figure 4 shows the timeseries of U and V winds for both the regression model and the observations and model for a period of time where SDHN4 observations are available. Figure 5 shows the timeseries of U and V during Superstorm Sandy, showing when SDHN4 observations went offline. Figure 6 shows the entire reconstruction of data for SDHN4.

Figure 4. U and V velocity winds from 09/14/2012 at 16 UTC through 09/23/2012 at 00 UTC.
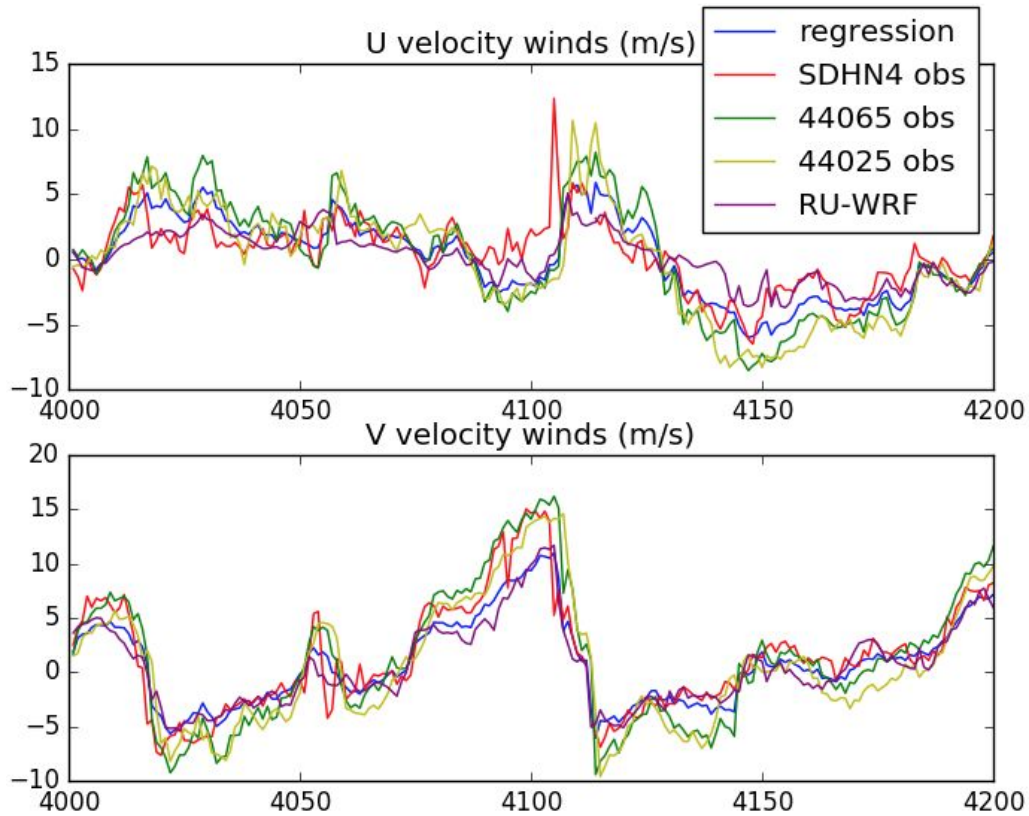

Figure 5. U and V velocity winds from 10/28/2012 at 10 UTC through 11/05/2012 at 18 UTC.
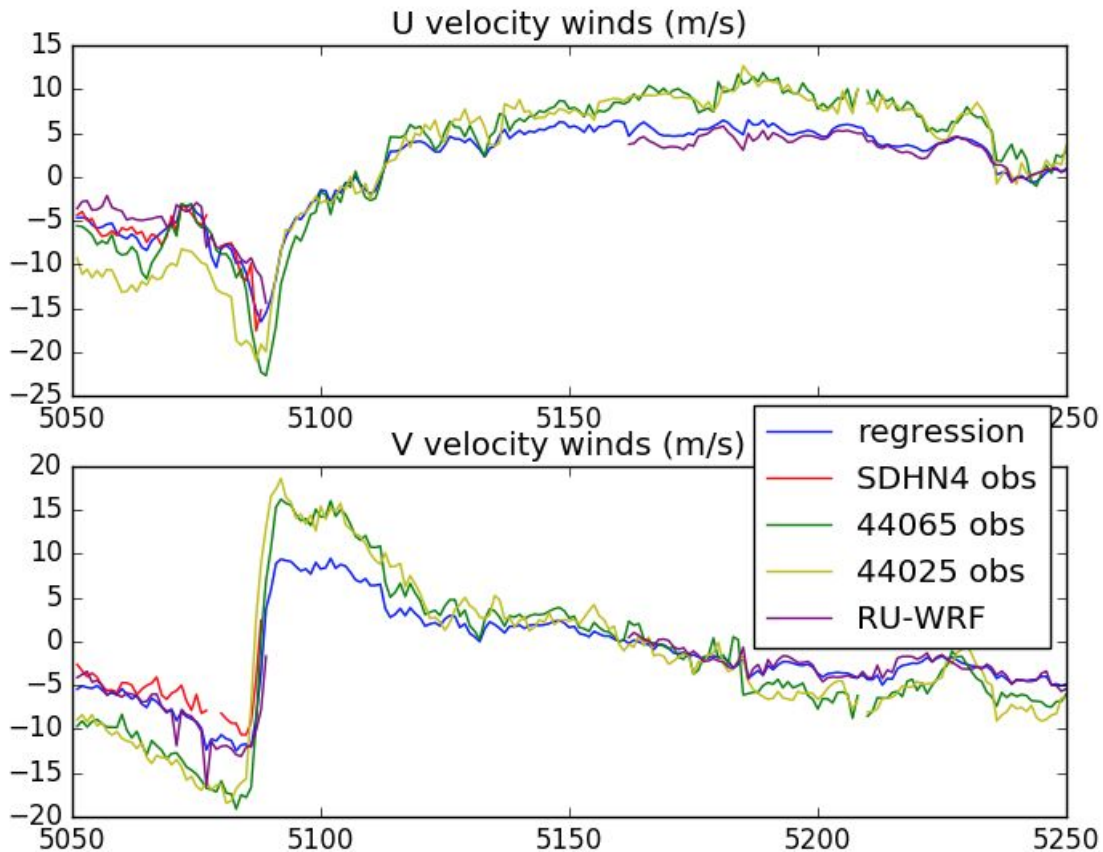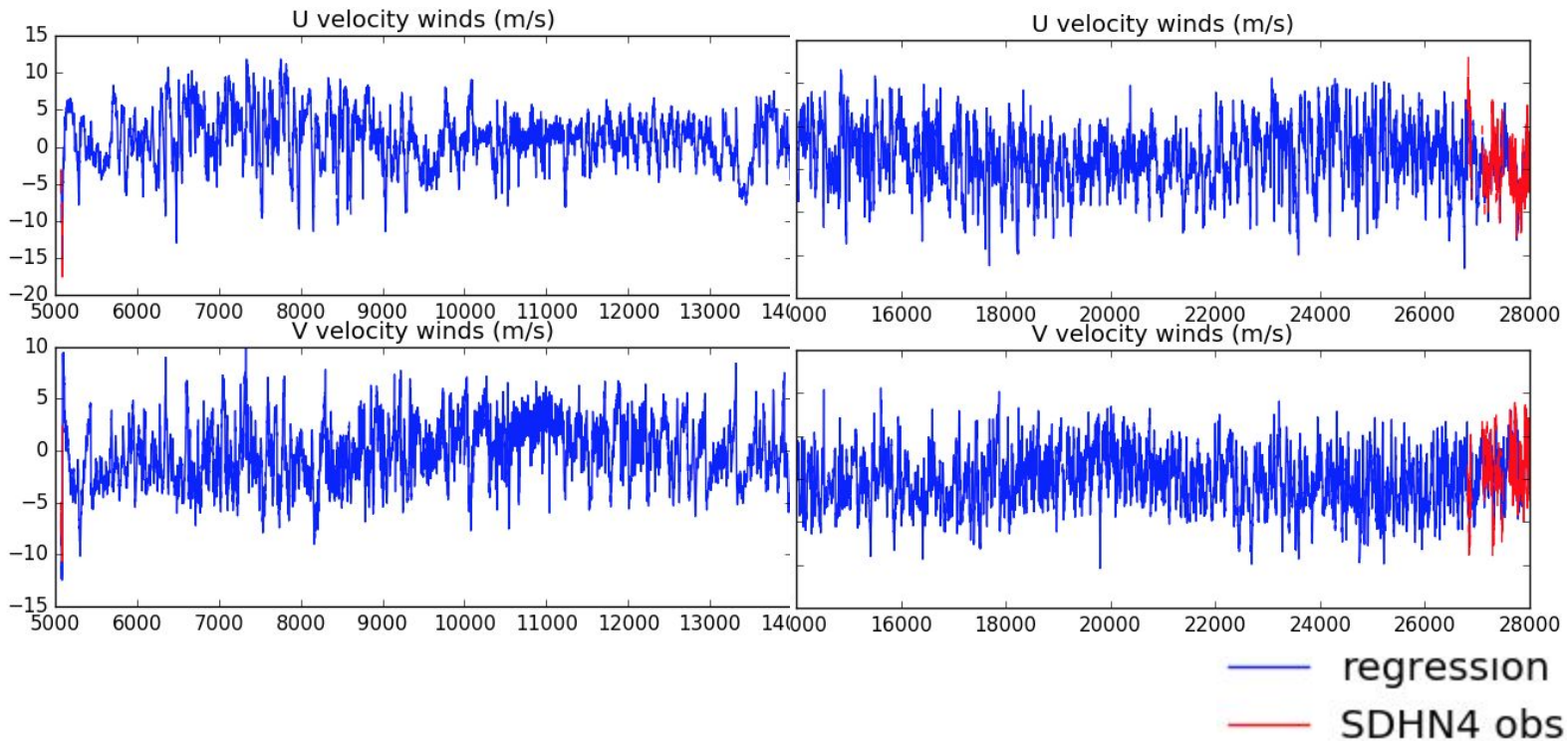
Figure 6. Entire reconstruction from April 2012 through December 31st, 2017.

The same verification process used on RU-WRF was performed on the regression model for April 2012 through December 31st, 2016. Overall, large improvements in the centered root mean square error (CRMSE) were observed. The CRMSE for the RU-WRF was 2.201 m/s and 2.096 m/s for U and V, while the regression model had a CRMSE of 1.611 m/s and 1.770 m/s for U and V. The percent difference in standard deviation improved substantially for U winds at the expense of V winds. The percent difference for the U winds in the regression model was 7.3% and 9.9%, within the validation criteria used previously.

Untrained data

Finally, an analysis on data not used to train the regression model was performed. RU-WRF model runs, buoy 44065 observations, and buoy 44025 observations from January 1st, 2017 00UTC through January 25th, 2017 18 UTC were used for the regression model. Figure 7 shows the regression model with the different buoy observations. The regression model performed significantly better than the RU-WRF alone at simulating SDHN4 observed windspeeds. The percent difference between the regression model and the SDHN4 observations was 1.3% and 3.6% for U and V winds. The percent difference between the RU-WRF model and the SDHN4 observations was 33% and 12.7% for U and V winds. The CRMSE between the RU-WRF and the SDHN4 observations was 2.42 and 2.02 m/s for U and V winds, while the CRMSE between the regression model and observations was 1.53 m/s and 1.58 m/s for U and V winds. The regression model passes all verification methods from NREL and Glenn and Dunk (2013).

Figure 7. Untrained data from January 1st, 2017 00 UTC through January 25th, 2017 18 UTC.
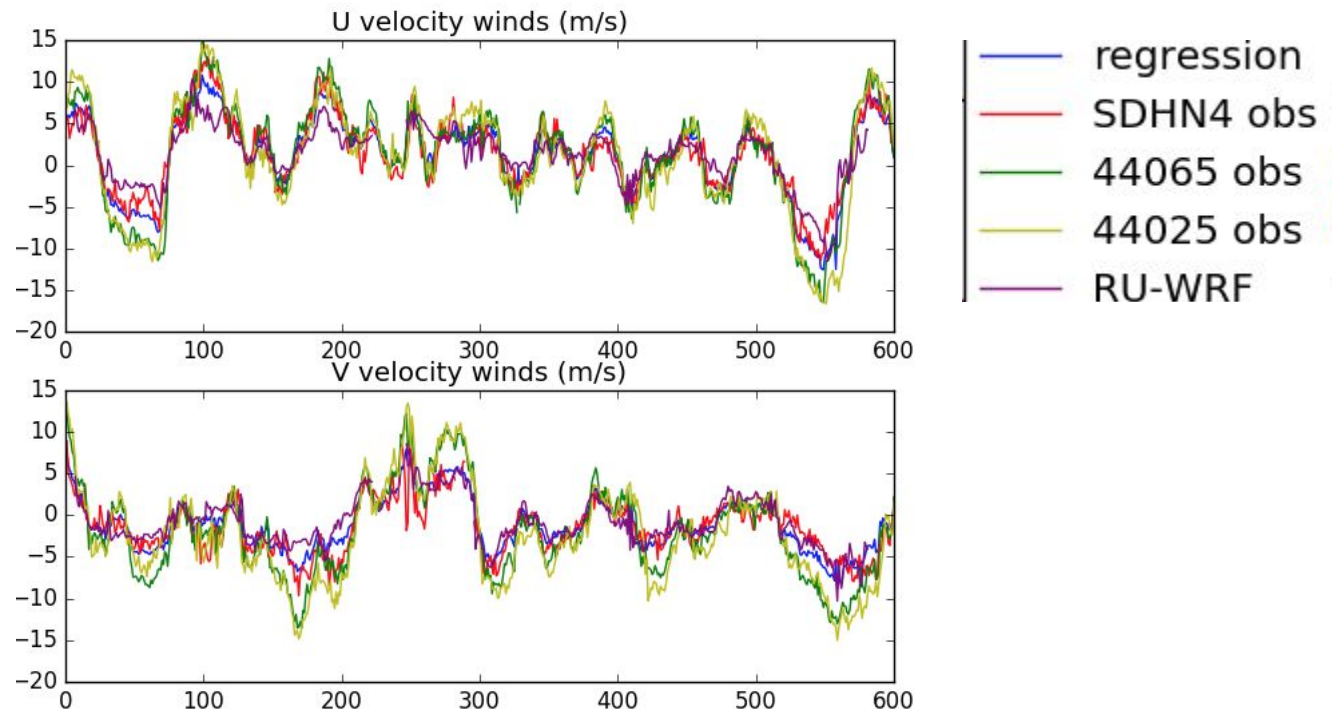
Conclusion

The regression model derived from both RU-WRF and buoy wind speeds has been validated using the same standards used to validate numerical weather prediction models for operational use for managing wind turbines. The reconstructed data using this regression model is the best approximation possible from all available data and outperforms the RU-WRF 3km model at simulating wind speeds at SDHN4. Due to the regression model's exceptional performance on untrained data, the reconstruction is considered to be valid.

References

Galarrneau, T. J., Davis, C. A., Shapiro, M. A., 2012. Intensification of Hurricane Sandy (2012) through Extratropical Warm Core Seclusion. Monthly Weather Review. Vol. 141, 4296-4321.

Glenn, S. and Dunk, R., 2013. An Advanced Atmospheric/Ocean Assessment Program Designed to Reduce the Risks Associated with Offshore Wind Energy Development Defined by the NJ Energy Master Plan and the NJ Offshore Wind Energy Economic Development Act.

Weiss, S.J., Pyle M. E., Janjic,. Z., Bright, D., Kain, J., DiMego, G., 2008: The Operational High Resolution Window WRF Model Runs at NCEP: Advantages of multiple model runs for severe convective weather forecasting. Preprints, *24th Conf. on Severe Local Storms,* Savannah, GA, Amer. Meteor. Soc., P10.8.