TRIED TPB-04 Report

Carl Robinson - 4th Dec 2017

Part 2: Framing by approximation of the variance

0) Objective, Data and Method

Objective & Method

A multilayered perceptron (PMC1) was trained on the training set data, with the optimum architecture was found using the validation set. This first network is a model that allows us to determine the horizontal polarization of σ0, as a function of the wind (direction and modulus). In order to attach confidence intervals to these results, we will estimate the variance of noise as a function of the data using a second multilayered perceptron.

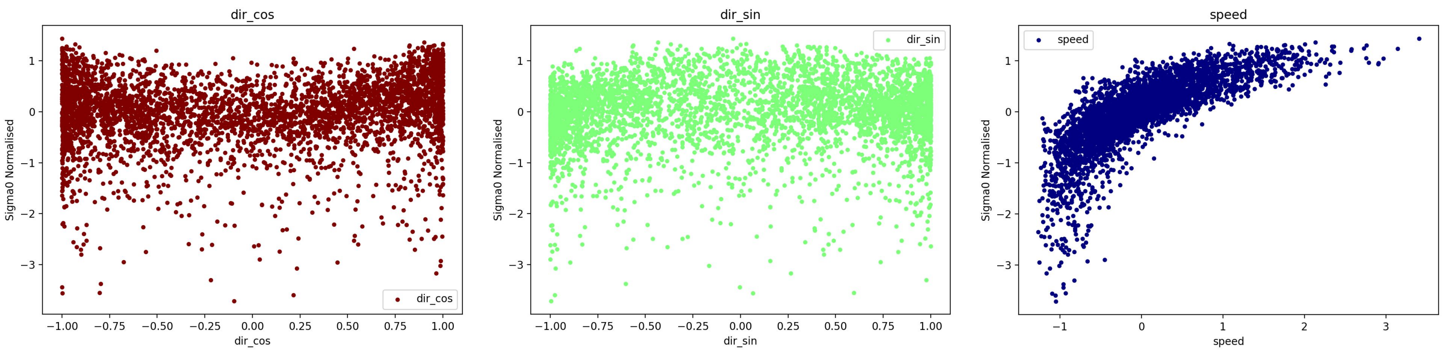
Dataset

The database is a collection of NSCAT radar measurements by the SeaScat satellite. This scatterometer is an active radar for observing the ocean surface.

4 files are included:

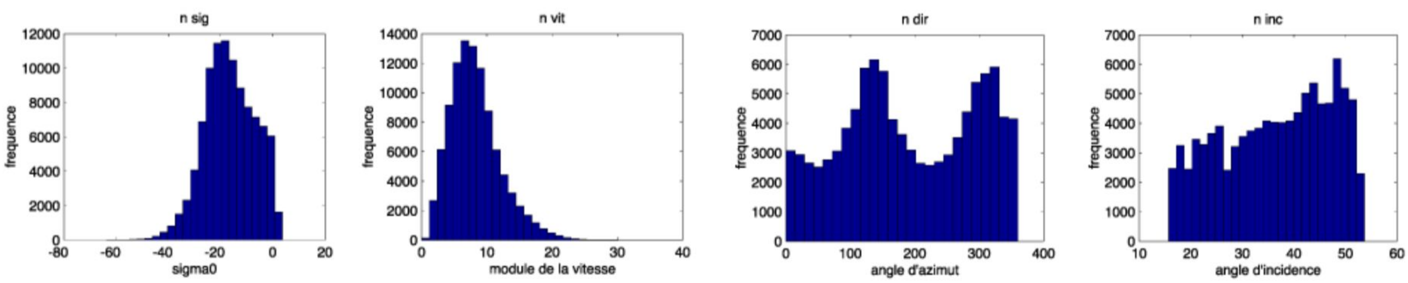
- Diffu_Dir.dat: The azimuth angle (in degrees) of the wind direction (χ) relative to the signal.
- Diffu_Inc.dat: The angle of incidence (in degrees) of the signal (θ)
- Diffu_Vit.dat: The modulus of the wind speed: ν (in m/s)
- Diffu_Sig.dat: The backscattering coefficient Sigma0 (or σ0), HH component (horizontal polarizations) only (in dB).

Relationship of sigma to wind direction (sin and cos components) and speed (modulus)



- This database is not equalized, so there are many more examples for low wind speeds than for high wind speeds, and some azimuth angles are better represented than others.
- The complete NSCAT database contains over 100,000 measurements (see histograms below). We will work with a subset of 4098 measurements, for a fixed angle of incidence of ~35 degrees. As such, incidence angle information is discarded from the input data.

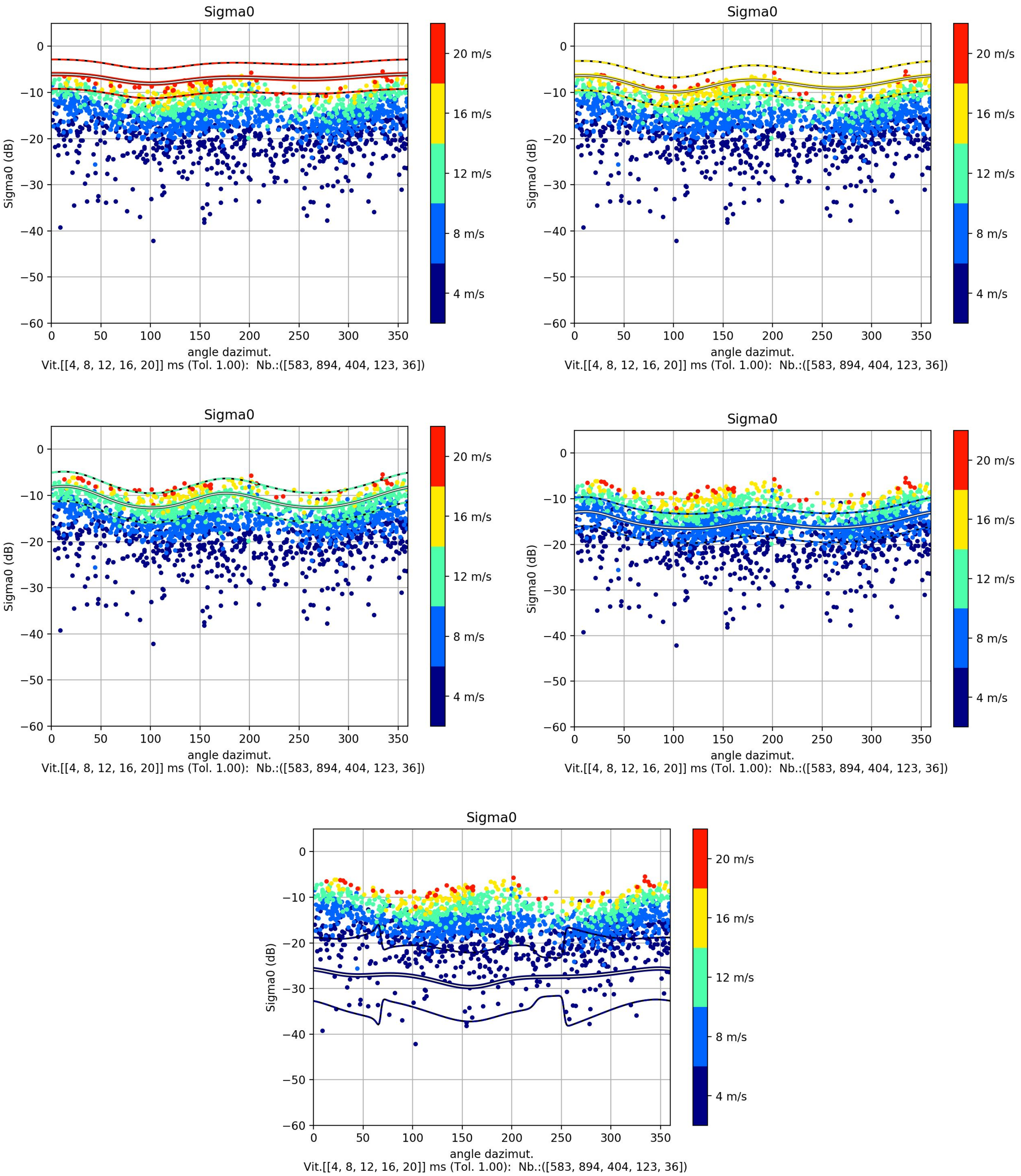
Frequency histograms of sigma, wind speed, wind direction and angle of incidence



1) Results

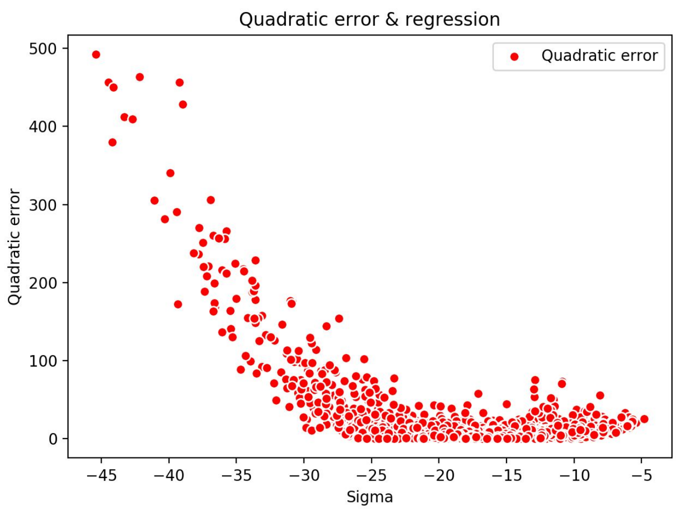
The following graphs plot the raw data points in colours indicating the wind speed. The output of PMC1 gives us an estimate of the expectation E(d/x) of the denormalized sigma0 values, which we use to plot regression curves for each wind speed interval. PMC2 gives us an estimate of the variance of d/x for each value of d, which is displayed as a frame around the regression line.

<u>Sigma0 raw data, PMC1 regression and PMC2 framing, for each of five wind speed intervals</u>

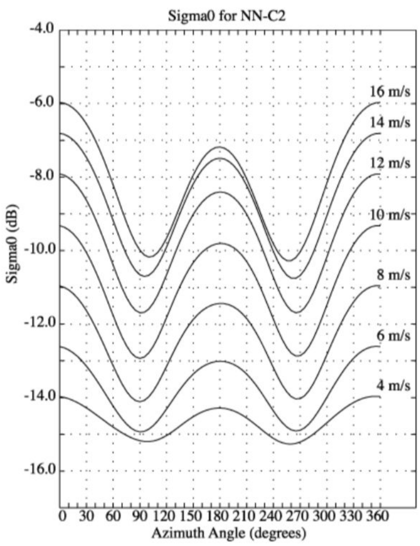


- As variance is a second-order estimator, it is important to have the maximum of data possible in order to estimate the parameters of PMC2. We used all the training data available, which meant we worked without a validation set or test set.
- We can clearly observe in the regression of all five wind speed intervals, the undulating pattern found in the figure for theoretical sigma value against wind direction (azimuth) below.
- The variance of noise for low wind speeds is much greater than that for high wind speeds.
- We see a very uneven/misshapen regression line for wind speed interval 4m/s, suggesting that PMC1 suffers from overfitting to the more sparse data interval.
- For the high wind speed intervals 16m/s and 20m/s, the regression line of PMC2 is unusually high, with most of the points falling beneath the line. The regression line for wind speed intervals 12m/s and 8m/s however appears well placed over the data, while line for wind speed interval 4m/s appear low. This indicates a problem with the ability of PMC1 to accurately estimate of the expectation E(d/x) of the denormalized sigma0 values both ends of the wind speed scale. This is likely due to overfitting, as there is much more data for the low to medium wind speeds than there is for very low or very high wind speeds.
- Unlike the theoretical figure, the regression line for 20m/s is much flatter than the regression line for 4m/s. This is likely due to the fact that strong wind causes greater diffusion in all directions, and results in a stronger measured signal, leading to more consistent signal measurements, while the low wind speed causes more noise in the measurements.
- The code provided did not include a function to count the examples that fell within the framing of the regression, so precise figures for how the variance differs between speed intervals could not be obtained. It is apparent from the graphs that points in the low wind speed interval are more likely to fall outside the framing than points in other intervals. This again is due to the higher error of these measurements due to the weaker signal, and a poorer fitting regression due to lack of data.

PMC1 quadratic error for values of sigma0



Theoretical sigma0 for wind direction



- Additionally, we can plot the quadratic error for sigma0 predictions on the full dataset. We can see that the error is low for high values of sigma0, and high for low values. When referencing the plot of theoretical sigma value against wind direction (azimuth), we see that sigma is lowest when the wind direction is perpendicular to the direction of the signal. Therefore we can conclude that wind direction and quadratic error are highly correlated.
- As the variance for each wind speed bracket regression is now available, for any fixed wind direction (x^0) value it is possible to calculate the confidence interval for the resulting sigma (y^0) value.

$$y_c^0 - t_{\alpha,n-2}\sqrt{VAR\left(y_c^0 - y^0\right)} < y^0 < y_c^0 + t_{\alpha,n-2}\sqrt{VAR\left(y_c^0 - y^0\right)}$$