AN ALGORITHM TO ASSESS THE ACCURACY OF NSCAT AMBIGUITY REMOVAL

Amy Elizabeth Gonzales
Brigham Young University
459 CB, Provo, UT 84602
801-378-4884, FAX: 801-378-6586, e-mail: gonzalae@ee.byu.edu

Abstract–A wind field model can be used to evaluate the accuracy of pointwise ambiguity removal for NASA Scatterometer (NSCAT) data. Errors in pointwise ambiguity removal result in large model-fit errors when the pointwise wind estimates are assimilated into the model. By thresholding the error, regions containing ambiguity removal error can be identified. For these regions, the ambiguity selection can be improved using the model-fit field. I have developed a new automated algorithm for evaluating the quality of the pointwise ambiguity selection and for correcting the ambiguity selection. This paper presents this correction algorithm, which is generally applicable to other scatterometers, and the results for NSCAT data.

INTRODUCTION

The NASA Scatterometer (NSCAT) is a microwave instrument capable of accurately measuring vector winds over the ocean during all weather conditions [Naderi et al., 1991]. Scatterometers do not directly measure the wind; rather the speed and direction of the near-surface wind are inferred from the normalized radar cross section (σ^o) measurements at an observation point or wind vector cell (wvc). The wind is related to σ^o via a geophysical model function. However, there are several possible wind vectors for any set of σ^o measurements. Although the speeds are typically the same, the directions vary with two to four possible directions for each wvc. An ambiguity removal algorithm must be employed to determine the correct direction.

Point-wise wind retrieval is the traditional method for estimation of the winds over the ocean. It consists of two steps and uses only the σ^o measurements for a single wind vector. The first step is to find the multiple wind vectors for each cell of the scatterometer swath. The second step, ambiguity removal, selects one unique wind vector estimate for each of these cells, though this algorithm is prone to error. A quality assessment of these algorithms is essential to maintain the integrity of the data.

Another method to determine wind measurements is model based wind retrieval [Long, 1993]. The wind field model provides a description of the near-surface wind field over the scatterometer measurement swath and is optimized for scatterometer wind retrieval. Wind field models are based on the spatial correlation between wind vectors. The swath is sectioned into rectangular regions and the wind is extracted over the entire region instead of by individual resolution elements. The model relates the components of the wind vector field over this region to a set of model parameters [Long, 1993] [Oliphant, 1996]. The models are either data-driven or dynamics-based.

The wind field model can be used to improve the point-wise wind product by identifying and correcting ambiguity removal errors. The quality of the fit of the estimated point-wise wind to a simple wind field model over a small area provides a measure of possible ambiguity removal errors. Large errors in the fit suggest possible ambiguity removal errors while small errors suggest a realistic wind field. Areas with errors can be corrected by choosing the alias closest in direction to the model-fit.

WIND FIELD MODELS

As discussed in [Long and Mendel, 1990, Long, 1993], a simple wind field model can be developed which is expressed as

$$\mathrm{W} = \mathrm{FX}$$

where X is an L-element vector containing the model parameters and F is a constant model matrix where the rows of F form a basis set for possible wind fields. There are several different models for which this model matrix changes.

While [Long, 1993] used a simple dynamics-driven model, in this paper we adopt a data driven model. We use the Karhunen-Loeve (KL) model since it is known to minimize the basis restriction error. The KL model is derived from the eigenvectors of the autocorrelation matrix of the sampled wind [Jain, 1989]. Using standard eigenvalue/eigenvector decomposition methods, the KL model is formed as the lower subset of the sorted eigenvectors of the sample autocorrelation matrix. In this paper, the model matrix was subjectively chosen as the first 22 eigenvectors

for the tradeoff between modeling error and the ability to locate regions with ambiguity removal errors. We note, however, that there is little performance difference in the algorithm when using between 20 and 30 eigenvectors.

A least squares estimate of the model parameter vector X, $\underline{X}$, can be obtained from the observed wind field W_0 using the pseudo-inverse of F, $F^\dagger$, i.e., $\underline{X}=F^\dagger W_0$. The reconstructed wind field W_R, also known as the model-fit field, is $W_R = F\underline{X}$ with the reconstruction error field W_E given by

$$W_E=W_R-W_0=(FF^\dagger-I)W_0.$$

If the reconstruction error is small, the model-fit is good and the observed wind field is "realistic" for the specific model. Large reconstruction errors suggest that the observed wind field is not realistic due to either ambiguity selection errors or poor modeling. Thresholds for the reconstructed error field detect regions with possible ambiguity removal errors.

To illustrate, Fig. 1 is a region with clear ambiguity removal errors in the upper left corner of the region. The model-fit field exhibits large errors at some locations which correspond to the boundary of the ambiguity removal error region. These are easily seen in the difference field. By finding these areas of significant wind error in the model-fit, ambiguity removal errors are identified.

There are a number of considerations when implementing this simple technique. First, the model must be fit to the wind field over a region. To produce an adequate fit, the input wind must be defined over the full region. Thus, for this simple algorithm, only those regions with fewer than eight cells of land or missing measurements are used. The missing measurements are replaced with the average of the cells surrounding it and then processed. Second, the wind field model inherently smoothes the wind field over the entire region due to modeling error; the model matches the general flow of the wind, but may not adequately model the center of a cyclone or the boundary of a front. Such regions can be flagged as containing errors, because the modeling error is large. Third, the error in the model-fit can be high in regions where the wind estimates are very noisy even if ambiguity removal is correct. Thus, the region may be flagged as having possible ambiguity removal errors even if the ambiguity removal is correct. Finally, at low wind speeds, the wind is highly variable, resulting in significant modeling error which is further complicated by the low signal to noise ratio in these regions. Manual ambiguity removal is also very difficult in such regions. As a result, we are unable to verify the ambiguity removal accuracy for

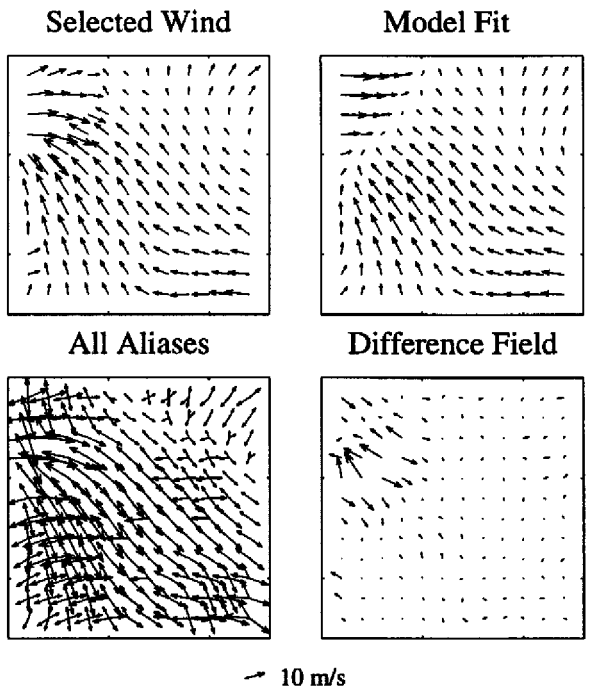

Figure 1: A wind field that exhibits a significant area of ambiguity removal errors in the upper left corner.

low wind speed regions. Figure 2 illustrates one such region.

This method locates the boundaries of the regions that have possible ambiguity removal errors. Once the regions with possible ambiguity removal errors are identified, it is natural to try to find a means of correcting these errors. To this end, a technique for correcting the errors has been developed.

An important consideration in making the corrections is that some regions are poorly modeled by the wind field model resulting in a poor model-fit for the reasons given above. So, regions with con-

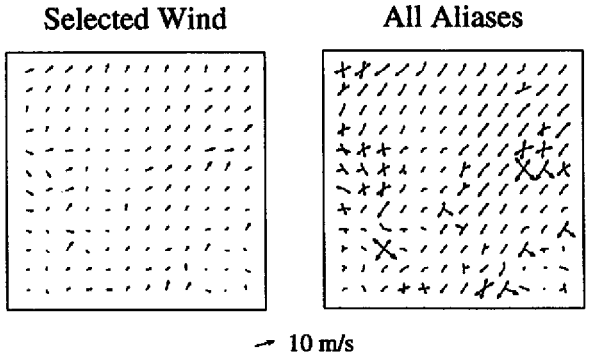

Figure 2: A region of low wind speed.

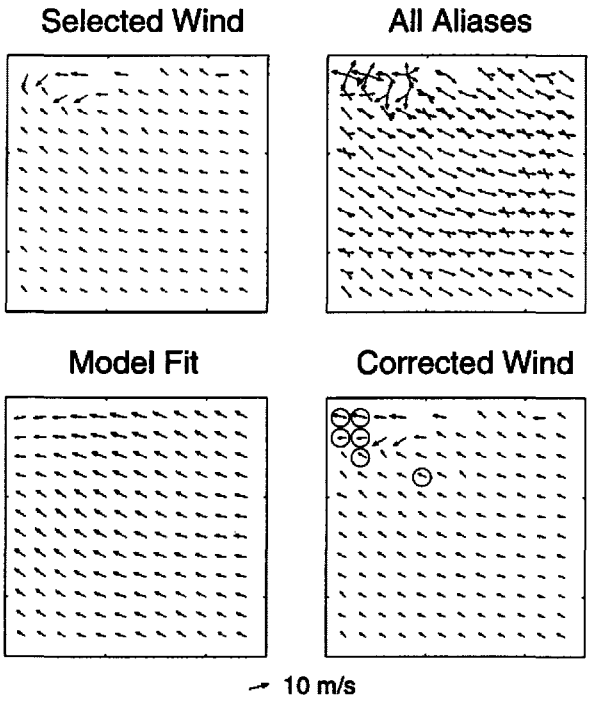

Figure 3: The plots for a corrected wind field. The circled vectors are those that were changed according to the algorithm.

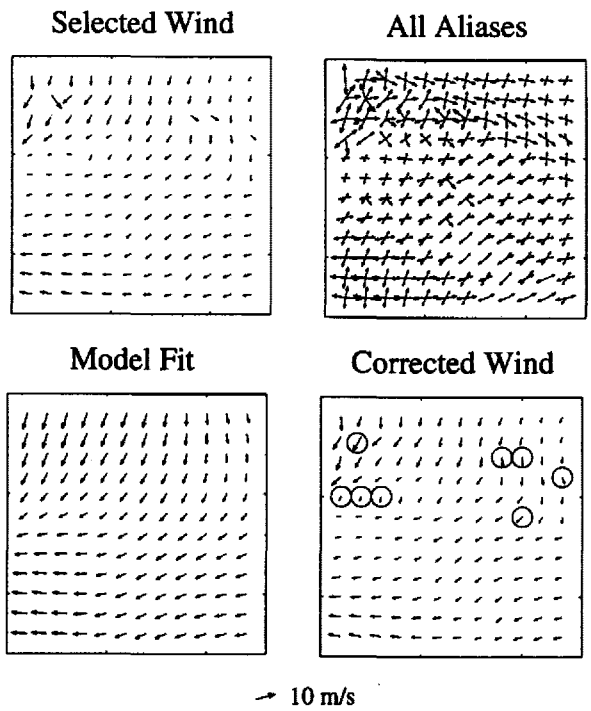

Figure 4: The plots for a corrected wind field. The circled vectors are those that were changed according to the algorithm.

siderable numbers of possible ambiguity removal errors are not considered candidates for the correction algorithm. Large numbers of possible ambiguity removal errors are typically low wind speed regions or regions with significant areas of ambiguity removal errors in which the model-fit should not be used as a means of correction.

Thus, for regions identified as having ambiguity removal errors, the correction technique proceeds as follows: determine the number of possible ambiguity removal errors by identifying those that exceed the thresholds; if this number is greater than a given threshold, do not correct any of the vectors for this region; otherwise, choose the alias closest in direction to the model-fit as the corrected wind.

Figures 3 and 4 demonstrates the use of the correction algorithm. As can be seen, the observed wind product contains several ambiguity removal errors. The algorithm chooses the alias that is closest in direction to the model-fit field, producing a subjectively better corrected wind field.

SELECTING THRESHOLDS

The reconstruction error field provides much information about the difference between the unique wind field and the reconstructed wind field. The value of the model parameter vector is also useful for identifying regions with ambiguity removal errors.

To select the thresholds for the model parameters, a histogram of each parameter is examined. Figure 5 shows the histograms of four of the parameters for the K-L model using 5488 regions of NSCAT data. Experimental testing has shown that large values for any of the model parameters correspond to regions with possible errors. After some examination of the values for the parameters, the thresholds are set at twice the standard deviation for each of them. This provides an initial starting place for subjectively altering these numbers as needed to correctly identify error-prone regions. Only a few of the model parameters are necessary to identify regions of possible ambiguity removal errors. Since the columns of F for the KL model are basis vectors in decreasing order, only the first few parameters are used as thresholds for the QA algorithm.

The other thresholds for locating ambiguity removal errors are determined from the reconstruction error field. These include the rms error, the normalized rms error, the maximum component error, and the maximum direction error for each region. The rms error is found by summing the squared components of the reconstruction error field, dividing by

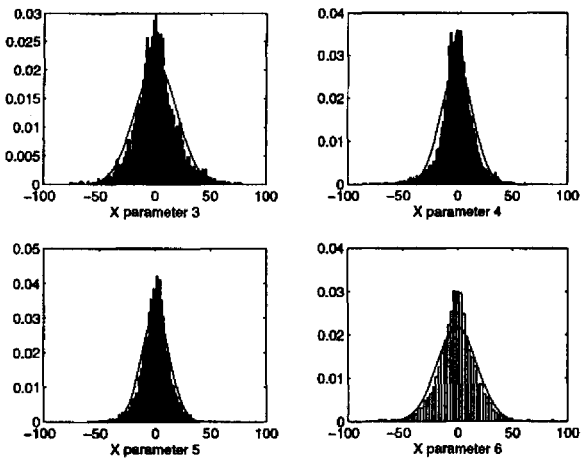

Figure 5: The histograms for parameters three through six for the K-L model. Overlaid is a Gaussian distribution with the same mean and variance.

the number of terms, and taking the square root. The normalized rms error is found by squaring the components of the reconstruction error field, dividing by the observed wind field, and taking the square root. The rms and normalized rms errors aid in locating regions of large error. Both of these error values are calculated for the entire region and thus provide information about the region as a whole. The maximum component and maximum direction error values are useful for locating regions in which only a few of the wind vectors are incorrect. Thresholds on these error values locate regions with at least one individual wind vector error. The individual errors are identified by finding those that exceed the thresholds. These are flagged as possible ambiguity selection errors, though, as discussed before the error may exceed the thresholds due to noise, modeling error, or ambiguity removal error.

To select the threshold values for this algorithm, 527 regions (10 revs) of NSCAT data were inspected by hand. The regions were subjectively grouped into four categories: "perfect" (no errors), "good" (those with only a few isolated ambiguity removal errors), "moderate" (as much as 10% but less than 20% of the wvc's identified as possible ambiguity selection errors), and "poor" (more than 20% of wvc's identified as possible errors). All of the poor regions either have low rms wind speeds making the region difficult to model or have significant areas of ambiguity removal errors. For this data set, 66% of the poor regions were low wind speed regions (rms speed less than a subjectively chosen threshold of 4 m/s). All of the remaining (with rms speed greater than 4 m/s) were regions with significant ambiguity removal errors. The statistics of each region were calculated and compared to the initial two sigma thresholds. The thresholds were adjusted such that the maximum number of poor, moderate, and good regions are correctly identified as containing ambiguity removal errors with a minimum number of false alarms.

For this small set, the algorithm correctly identifies 100% of the poor and moderate regions and over 95% of the good regions with a false alarm rate of less than 5%. It should be understood that the thresholds can be altered to adjust the detection and false alarm probabilities. For example, if all regions with possible errors are to be detected, the number of false alarms will increase. The thresholds are determined for a specific trade-off between detection and false alarms.

The thresholds chosen for the detection algorithm were tested on a manually classified withheld data set of 274 regions (5 revs) and achieved a similar level of performance. The algorithm correctly identified 100% of the poor and moderate regions and over 98% of the good regions with a false alarm rate of less than 5%. Combining the statistics for these two data sets results in total detection rate of more than 97% for all regions subjectively identified as containing ambiguity removal errors with less than 5% of the perfect regions misidentified. Thus, though modeling error or noise will sometimes result in an incorrect evaluation of a region as containing possible errors, the vast majority of regions with possible ambiguity removal errors are located using this technique. The classification performance of low wind speed regions was also consistent with the previous results. Low (< 4 m/s) rms wind speeds accounted for 75% of the poor regions with the remaining regions (with rms wind speed of greater than 4 m/s) all containing significant areas of ambiguity removal errors.

ALGORITHM DESCRIPTION

A general procedural description of the algorithm follows:

1. Segment the swath into 12x12 overlapping regions (50% along track overlap).

2. For each valid region (regions with fewer than eight cells of land or missing measurements), compute the model-fit field W, the reconstruction error field W_E, the model parameter vector $\underline{X}$, and the statistics of W_E. These statistics include the rms error, the normalized rms

error, the maximum component error, and the maximum angle error for each region.

3. For each region, determine if any statistic, including those for the model parameter vector $\underline{X}$, is larger than the threshold. If so, the region is identified as containing possible ambiguity selection errors. Based on the number of possible errors identified for each region, segregate the regions into 4 classes ("perfect", "good", "moderate", and "poor") .

4. For those regions not classified as "poor", correct the ambiguity removal error by choosing the ambiguity closest in direction to the model-fit for those wvc's identified as possible errors.

RESULTS FOR NSCAT

This algorithm was tested on the data for the nine month NSCAT mission. To be considered candidates for the correction technique in this implementation, 20% or fewer of the vectors in the region can be identified as possible errors . Regions not considered candidates are classified as "poor". Of the 408,069 regions examined, approximately 82% were considered candidates for the correction algorithm. Only 4% of the individual vectors in these regions were identified as possible ambiguity removal errors; however, only approximately 10% of these were changed using the model-based correction technique. For the remaining, the alias closest in direction to the model-fit was the original wind vector. Thus, only 0.4% of the individual vectors were corrected using this approach. This suggests that NSCAT ambiguity removal is thus over 99% effective for these regions.

Much can be said about the remaining 18% of the data for the NSCAT mission. Fig. 6 summarizes key statistics for this portion of the data. The majority of these regions (approximately 74%) have rms speeds lower than 4 m/s. The scatterometer does not perform well at such low wind speeds and ambiguity removal algorithms have difficulty distinguishing the correct wind vector at these speeds. Regions with such low wind speeds are thus not included in the assessment of NSCAT ambiguity removal.

"Poor" regions with rms wind speeds in excess of 4 m/s contain significant ambiguity removal errors. This represents only 5% of the total data for the NSCAT mission. Not every wind vector in these regions is in error, a fact verified by a subjective analysis of these regions. However, a conservative approach is to treat each wind vector in these regions as a possible ambiguity removal error. Combining

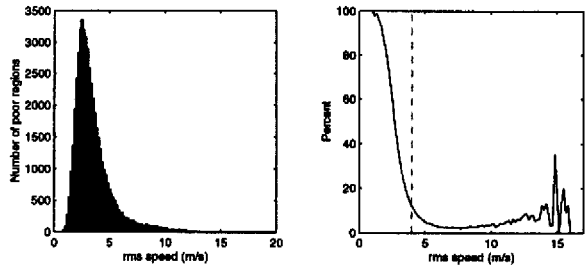

Figure 6: (left) Histogram of the rms speed for all regions classified as "poor" in the nine month NSCAT mission. (right) the percent of the total regions which are classified as "poor" at each rms wind speed bin. The vertical dashed line is at 4 m/s.

this result with the near complete effectiveness of NSCAT for non-poor regions, we conclude that the skill of NSCAT is 95% or better for regions with rms winds speeds greater than 4 m/s.

The performance of NSCAT was also evaluated as a function of time. From Fig. 7, it is clear that the accuarcy of NSCAT declines towards the end of the mission. This is most likely a seasonal effect. To see this more clearly, the performance of NSCAT was evaluated over several lattitude bands in the Pacific Ocean as described in Fig. 8. The statistics are these bands are described in Fig. 9. The expected variation of wind speed with latitude is clearly evident. There is a strong correlation between the ambiguity removal performance and the rms wind speed, with reduced overall ambiguity removal performance (i.e., more poor regions) at lower wind speeds. Thus, the

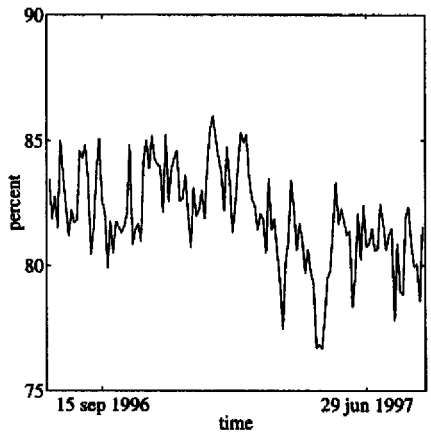

Figure 7: The percent of non-poor regions versus time over the nine month NSCAT mission. Each point represents the average computed over approximately two days.

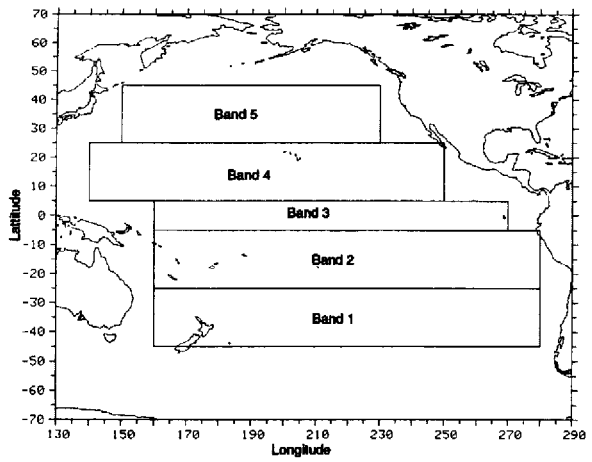

Figure 8: Geographical latitude bands in the Pacific.

wind speed distribution in each band affects the ambiguity removal performance and seasonal changes in the wind speed distribution results in temporal variations in the ambiguity removal performance. In particular, increased storm activity in the Northern Hemisphere results in increased wind speed with improved ambiguity removal during the winter months in Bands 4 and 5. Similarly, the number of poor regions increases during the Southern Hemisphere summer due to a decrease in the rms wind speed. Because of its low rms wind speed, Equatorial Band 3 is the most sensitive to changes in the mean rms wind speed with a significant drop in the percent of non-poor regions corresponding to a small drop in the rms wind speed at the start of 1997.

CONCLUSIONS

The detection algorithm works very well in identifying regions with possible selection errors. Once the errors are detected, they can be corrected by choosing the point-wise alias closest the the model-fit. The correction algorithm consistently produces a subjectively more realistic wind field. This technique provides a quick way in which to measure the accuracy of NSCAT ambiguity removal using only NSCAT data. Further research to optimize this technique, such as finding a theoretical basis for determining when the regions are modeled well, is in progress.

REFERENCES

[Jain, 1989] Jain, A. K., Fundamentals of Digital Image Processing. Prentice Hall, 1989.

[Long, 1993] Long, D. G., "Wind field model-based estimation of Seasat scatterometer winds." Journal of Geophysical Research, vol. 98, no. C8, pp. 14651–14688, 1993.

[Long and Mendel, 1990] Long, D. G. and J. Mendel, "Model-Based estimation of wind fields over the ocean from wind scatterometer measurements, Part1: Development of the wind field model." IEEE Transactions of Geoscience and Remote Sensing, vol. 23, no. 3, pp. 349–360, 1990.

[Naderi et al., 1991] Naderi, F., M. H. Freilich, and D. G. Long, "Spaceborne radar measurement of wind velocity over the ocean–An overview of the NSCAT Scatterometer system." Proc. of the IEEE, vol. 79, no. 6, pp. 850–866, 1991.

[Oliphant, 1996] Oliphant, T. E., New techniques for wind scatterometry. Master's thesis, Brigham Young University, 1996.

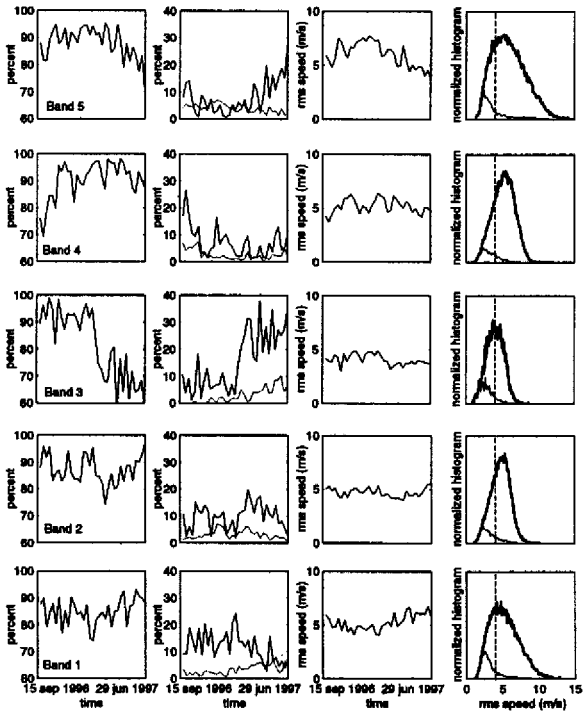

Figure 9: (left) Percentage of non-poor regions as a function of time over the NSCAT mission. (left, middle) Percentage of poor regions with an rms wind speed greater than (solid) and less than (dotted) 4 m/s. (right, middle) Average regional rms wind speed as a function of time. (right) Normalized histograms of (bold) all regions and (light) those classified as poor by the QA algorithm. The vertical dashed line is at 4 m/s.