# A Quality Assurance Algorithm for NASA Scatterometer Ambiguity Removal

**Amy E. Gonzales and David G. Long**
**Brigham Young University, MERS Laboratory**
**459 CB, Provo, UT 84602**
**801-378-4884, FAX: 801-378-6586 gonzalae@sage.ee.byu.edu**

*Abstract*— The recently launched NASA Scatterometer (NSCAT) estimates the wind speed and direction of near-surface ocean wind. This is done by directing microwaves toward the earth's surface and measuring the backscattered radiation. From this, several possible wind vectors are identified for each point over the swath. The correct wind must be distinguished from these in a step called ambiguity removal.

Unfortunately, ambiguity removal algorithms are subject to error. Because the true wind is not known, where these errors occur is difficult to determine, and there is little information in the measurements alone to detect the errors in this removal step. We have developed a method to assess the accuracy of the ambiguity removal algorithm by comparing the point-wise retrieved wind to winds inferred with a wind field model.

The performance of the algorithm achieves its goal to identify at least 95% of regions containing ambiguity removal errors. The algorithm provides a very simple tool to indicate regions of possible ambiguity removal errors in the point-wise retrieved winds for NSCAT data. This paper describes this algorithm and its performance for real NSCAT data.

## INTRODUCTION

The NASA Scatterometer (NSCAT) is a microwave instrument capable of accurately measuring vector winds over the ocean during all weather conditions [1]. Scatterometers do not directly measure the wind; rather the speed and direction of the wind are inferred from the normalized radar cross section ($\sigma^o$) measurements of the ocean surface. The wind is related to $\sigma^o$ via a geophysical model function. However, there are several possible wind vectors for any particular $\sigma^o$. Although the speeds are typically the same, the directions exhibit a 180 degree ambiguity. 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 $\sigma^o$ 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.

A second method to determine wind measurements is model based wind retrieval [2]. 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. 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 [2]. The models are either data-driven or physics-based and have been shown to provide more accurate wind measurements than point-wise wind retrieval [3].

The wind field models can also be used to improve the point-wise wind product by identifying ambiguity removal errors. One way to do this is to fit the estimated point-wise wind to a simple wind field model over a small area. Large errors in the fit suggest possible ambiguity removal errors while small errors suggest a realistic wind field. This is exploited in the ambiguity quality assurance algorithm which follows.

## WIND FIELD MODELS

As discussed in [2, 3], a simple wind field model can be developed which is expressed as

$$\mathbf{W} = \mathbf{F}\mathbf{X}$$

where $\mathbf{X}$ is an L-element vector containing the model parameters and $\mathbf{F}$ is a constant model matrix where the rows of $\mathbf{F}$ form a basis set for possible wind fields. There are several different models for which this model matrix changes. Two particular twenty-two order models were examined for this algorithm: the Parameterized Boundary Conditions (PBC) model [2] and the Karhunen-Loeve (K-L) model [4].

To use the model as a quality assurance for the point-wise wind retrieval, the model is fit in a least-squares sense to the observed point-wise wind field. The error in the fit provides some information about how realistic the observed wind actually is.

A least squares estimate of the model parameter vector $\mathbf{X}$, $\underline{\mathbf{X}}$, can be obtained from the observed wind field $\mathbf{W}_0$ using the pseudo-inverse of $\mathbf{F}$, $\mathbf{F}^\dagger$, i.e., $\underline{\mathbf{X}}=\mathbf{F}^\dagger\mathbf{W}_0$. The reconstructed wind field $\mathbf{W}$, also known as the model-fit field, is $\mathbf{W} = \mathbf{F}\underline{\mathbf{X}}$ with the reconstruction error field $\mathbf{W}_E$ given by

$$\mathbf{W}_E=\mathbf{W}-\mathbf{W}_0=(\mathbf{F}\mathbf{F}^\dagger-\mathbf{I})\mathbf{W}_0.$$

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.

Figure 1 is an example of the observed JPL wind field and the model-fit field for a particular 12x12 region. Notice the smoothing. As can be seen from the all alias plot, the model-fit wind field corrects ambiguity removal errors in the observed wind field. Thus, the model-fit is a reasonable basis for determining realistic wind fields and locating regions with possible errors.

However, there are several considerations when implementing this simple algorithm. First, the error in the model-fit might be high in regions where the wind estimates are very noisy even if ambiguity removal is correct. Second, the wind field model inherently smoothes the wind field over the entire region, and cyclones and sharp fronts are not modeled well. The error in these regions will be very high due to the limitations of the model. Third, at low wind speeds, the wind is highly variable which is complicated by the low signal to noise ratio in these regions. As a result, the error in low wind speed regions is larger than in high wind speed regions–even for perfect ambiguity removal. Finally, 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, regions with significant amounts of land or missing measurements are not processed. 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.

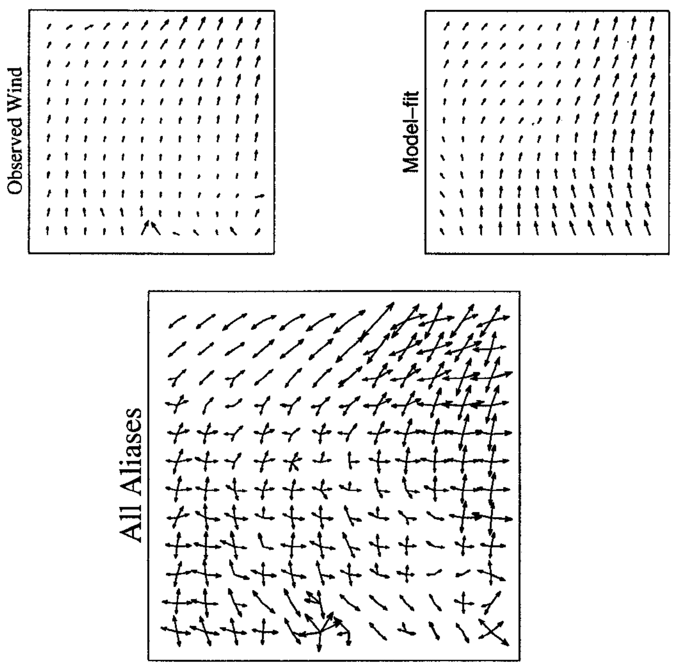


**Figure 1**: *The region on the upper left is the JPL product while the region on the upper right is the model-fit to this field. The lower figure plots all the alliases for this region.*

The model-fit error locates the boundaries of the regions that have possible ambiguity removal errors. The algorithm is very successful at identifying these regions, but so far does not correct the errors. It is designed only as a check of the consistency of the unique wind field.

## 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 $\mathbf{W}$, the reconstruction error field $\mathbf{W}_E$, the model parameter vector $\underline{\mathbf{X}}$, and the statistics of $\mathbf{W}_E$. 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 the statistics, including those for the model parameter vector $\underline{\mathbf{X}}$, are larger than the thresholds. If so, the region is identified as containing possible ambiguity selection errors.

## IDENTIFYING 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.

### Model Parameter Statistics

To determine the thresholds for the model parameters, a histogram of the these parameters is examined for each model. Figure 2 shows the histograms of four of the parameters for the K-L model using 5488 regions of valid NSCAT data. As can be seen, while the X parameters are not Gaussian, they exhibit a Gaussian-like shape.

It has been found that large values for any of the parameters correspond to regions with possible errors. After some examination of the values for the parameters, the thresholds are subjectively set at twice the standard deviation for each of them. This provides an initial starting place for altering these numbers as needed to correctly identify error-prone regions.

Only a few of the model parameters are necessary for this algorithm. In the case of the K-L model, choosing

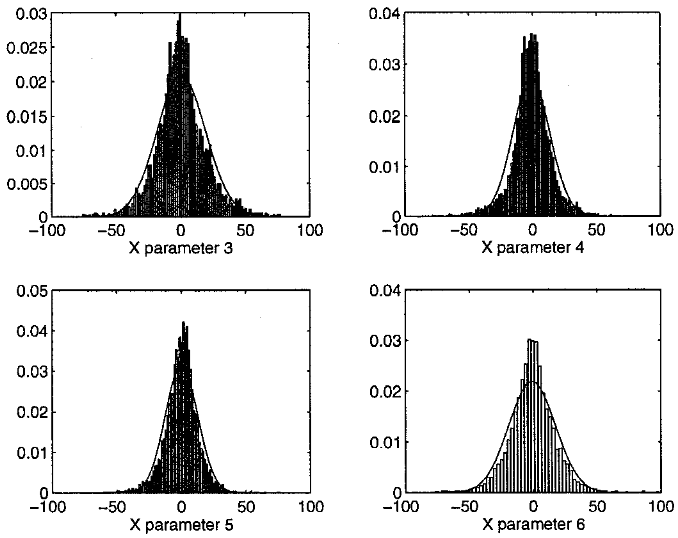


**Figure 2**: *The histograms for parameters three through six for the K-L model. Overlaid is a Gaussian distribution with the same mean and variance.*

them is an easy task. The columns of **F** are basis vectors in decreasing order, and only the first few parameters need be used. However, the PBC model is ordered quite differently, and the fewest number of model parameter thresholds are subjectively chosen that will accurately identify as many regions with ambiguity removal errors as possible.

Error Matrix Statistics

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 squaring the reconstruction error field, dividing by the number of terms, and taking the square root. The normalized rms error is found by squaring the reconstruction error field, dividing by the observed wind field, and taking the square root. The first two are useful for locating regions of obvious error. Both of these error measurements are calculated for the entire region and thus provide information about the region as a whole. The second two error measurements are useful for locating regions in which only a couple of the wind vectors are incorrect. These measurements are calculated by individual components. They locate regions with at least one individual wind vector error. The individual errors are easily identified by finding those that exceed the thresholds.

Selecting Thresholds

To select the threshold values for this algorithm, 527 regions (10 revs) of NSCAT data were inspected by hand, a small subset of the original 5488 regions. The regions were subjectively grouped into three categories: perfect, moderate (those with only a few isolated ambiguity removal error) and poor regions. The statistics of each of the regions were then calculated and compared to the initial two sigma thresholds. The thresholds are adjusted such that the maximum number of poor and moderate regions are identified with a minimum number of false alarms.

For this small set, the algorithm correctly identifies 100% of the poor regions and over 95% of the moderate 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.

## CONCLUSIONS

The thresholds chosen above were tested on a withheld data set of 274 regions (5 revs) and achieved a similar level of performance. The algorithm correctly identified 100% of the poor regions and over 98% of the moderate regions with a false alarm rate of less than 5%.

The algorithm works very well in identifying regions with possible selection errors. Although the algorithm only detects ambiguity removal errors, ongoing research is being done to correct those regions with only a few individual vector errors. In this case, choosing the point-wise ambiguity closest to the model-based wind will correct the problem. How to correct the other regions with possible ambiguity selection errors is the focus of upcoming research.

## REFERENCES

[1] F. Naderi, M.H. Freilich, and D.G. Long, "Spaceborne Radar Measurement of Wind Velocity Over the Ocean--An Overview of the NSCAT Scatterometer System," *Proceedings of the IEEE*, pp. 850-866, Vol. 79, No. 6, June 1991.

[2] D.G. Long and Jerry M. 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*, pp. 349-360, Vol. 28, No. 3, May 1990.

[3] D.G. Long, "Wind Field Model-Based Estimation of SEASAT Scatterometer Winds," *Journal of Geophysical Research*, pp. 14,651-14688, Vol. 98, No. C8, 1993.

[4] Jacob Gunther and D.G. Long, "Models for the Near-Surface Oceanic Vorticity and Divergence," in *Proc. IGARSS '94*, (Pasadena, CA), pp. 951-953,1994.