Development of a Rain Flag for QuikScat.

I. The Effect of Rain on Wind Vector Retrieval for Scatterometers.

Scatterometers such as QuikScat obtain the surface wind vector over the ocean by measuring the radar signal returned from the sea surface. The sea surface radar cross section σ_0 is measured for several different azimuth angles and for both horizontally and vertically polarized radiation. The wind vector can be retrieved by fitting these measurements to a geophysical model function that describes the expected σ_0 as a function of wind speed, wind direction relative to the look angle, and the incidence angle. The presence of rain in the atmosphere through which the radar signal is traveling affects the measured σ_0. This occurs in three ways

1. The radar signal is attenuated by the rain as it travels to and from the earth's surface. This process reduces the measured σ_0.

2. The radar signal is scattered by the raindrops. Some of this scattered power returns to the instrument. This increases the measured σ_0.

3. The roughness of the sea surface is increased because of splashing due to the raindrops. This increases the measured σ_0

A simple expression for the measured cross-section σ_{meas} that takes these effects into account is shown here.

$$\sigma_{meas} = \sigma_{scat}(R) + \tau(R)^2 \left[\sigma_{wind} + \sigma_{splash} \left(R/h \right) \right].$$

In this equation, R is the rain column rate (rain rate times column height h and sec θ), σ_{scat} is the contribution due to scattering from rain drops, τ is the one-way attenuation due to the rain, σ_{wind} is the contribution due to surface roughness caused by wind, and σ_{splash} is the contribution due to surface roughness caused by raindrops. Although this is equation is somewhat simplified, it describes the important effects of rain on scatterometry. At low wind speeds, where σ_{wind} is small, the additional scattering effects dominate, and σ_{meas} is greater than σ_{wind}. Eventually, as the wind speed increases, σ_{wind} becomes large enough that the attenuation effect becomes important, and begins to cancel the effect of the scattering terms.

II. The Effect of Rain on the QuikScat Instrument.

For the incidence angles and radar frequency used by QuikScat, the scattering effect is dominant up to a wind speed of about 10 m/s. The scattering and attenuation effects cancel completely at a wind speed between 15 m/s and 20 m/s, depending on the relative wind direction and which beam (h-pol or v-pol) is being considered.

At low wind speeds and not too high rain rates, it is relatively easy to detect when a QuikScat measurement is affected by rain. For low winds, the h-pol σ_{wind}, $\sigma_{H,wind}$ is significantly (~3 dB) smaller than the v-pol σ_{wind}, $\sigma_{V,wind}$. The rain-scattered contribution is either polarization independent (for small, spherical rain drops) or has a larger h-pol component (for large, flattened rain drops). When the rain-scattered component is added into the sea-surface component, σ_H is too large compared to σ_V to be consistent with any wind speed and direction. For higher wind speeds or very high rain rates, the situation is not so favorable. As the wind speed increases, $\sigma_{H,wind}$ increases faster than $\sigma_{V,wind}$. For very high winds and for some wind directions, $\sigma_{H,wind}$ even exceeds $\sigma_{V,wind}$. The magnitude of σ_{wind} for both polarizations is also much larger, and the scattering component of σ_{meas} becomes less important. Under these conditions, the presence of rain modifies σ_H and σ_V, but usually not by so much they become inconsistent with any wind vector. Instead, the wind vectors are typically changed in magnitude and point toward a direction slightly more perpendicular to the track of the satellite. The $\sigma_{H,wind}$ to $\sigma_{V,wind}$ ratio in the model function is maximized when the retrieved vector is aligned perpendicular to the satellite track. Note that the same effect can occur even for low wind speeds, if both σ_H and σ_V are increased by the rain so much that they are consistent with a strong, cross-swath wind. These effects make it difficult to impossible to use the σ_H to σ_V ratio to detect rain at high values of σ_0.

Rain is more highly variable over short distances than is the wind. For each 25 km wind vector cell, typically more than 12 egg measurements are used to retrieve the wind vector. It is possible that in rainy instances, these eggs, which may be separated by more the 25 km, will view significantly different rain rates. Eggs that are viewing the sea surface from the same direction and with the same polarization, and therefore would usually measure similar values of σ_0, might show significant variation. This, too, can form a component of a rain detection algorithm.

III. The Rain Detection Algorithm.

Although it may be difficult to detect rain in high wind regions, it is still a worthwhile endeavor to develop a rain detection algorithm based solely on σ_0 measurements for several reasons. First, it may not be so important to detect rain in high wind regions, since the effects of the rain are less pronounced. Second, it appears that even a small amount of rain in regions of very low winds can significantly modify the retrieved winds. The amounts of rain are so small that a rain detection method based on using the QuikScat instrument in radiometer mode would probably insufficiently sensitive. Third, if the algorithm is also sensitive to the spatial variability within a wind vector cell, it may be able to detect instances of rain even at high wind speeds.

We have developed a rain detection algorithm that is simple, and sensitive to both the σ_H to σ_V ratio, and to high variability within a wind vector cell. The algorithm is based on an empirically normalized objective function (ENOF), which is defined for each wind vector cell, and is given by the equation below.

$$ENOF = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(\sigma_{i,MEAS} - \sigma_{i,NSCAT2}\right)^2}{W_i}$$

Here, $\sigma_{i,meas}$ is the measured σ_0 for each observation in the cell, and $\sigma_{i,NSCAT-2}$ is the value of σ_0 calculated for each observation, using the retrieved wind vector and the NSCAT-2 model function. (This is the model function originally used to retrieve the wind). W_i is a weighting function that represents the "expected" or "typical" contribution of the ith measurement to the overall variance, and the sum is over the N observations used to retrieve the wind vector for the cell. W_i is chosen, using the procedure described in the next paragraph, so that if the ENOF is greater than one, the fit of the measured σ_0's to the model function is significantly worse than is typical. Note that this procedure is sensitive both to cases where σ_H is too large to σ_V, and to cases with large variability between σ_{meas} that should be the same (i.e. when measured at the same angle and polarization).

Choosing the weights W_i so that the ENOF will be large in the presence of rain is somewhat complicated. Formally, W_i is a function of the true wind speed, the wind direction relative to the observation azimuth angle, the incidence angle, and the polarization. Due to the two-beam conical scanning geometry of QuikScat, we can replace the dependence on the incidence angle and polarization with a dependence on beam index. Previous work has indicated that the expected variance is not a strong function of relative wind direction. We do expect some dependence of W_i on the azimuth and polarization diversity in the wind vector cell, since it will be easier for the retrieval process to return a wind that fits the model function when fewer polarizations are measured, or when there is not much variation in measurement azimuth angle. We absorb this dependence into a dependence on cross-swath wind-vector cell number.

The largest problem occurs because we do not know the true wind speed. Using the retrieved wind speed is not acceptable because the typical effect of rain is to increase all σ_0 measurements, yielding a retrieved wind speed that is much higher than the true wind speed. This in turn would give us a much too large value for W_I, and we would be less sensitive to rain-induced increases in variance. Instead of using the retrieved wind, we use a rain-desensitized proxy for wind speed to describe the wind speed dependence. We have constructed a simple wind-speed retrieval regression based on a rain-desensitized average σ_0 given by

$$\overline{\sigma}_{0,rain-desensitized} = \overline{\sigma}_V - 0.5\overline{\sigma}_H(1.0 - 0.05\overline{\sigma}_H),$$

where $\overline{\sigma}_V$ and $\overline{\sigma}_H$ are the averaged V-pol and H-pol σ_0's for all observations in the cell. The idea is that since the presence of rain often increases σ_H more that σ_V, by subtracting away part of σ_H we can construct an average σ_0 that is less sensitive to increases due to rain. At high winds, and therefore large values of σ_0, the value of σ_H approaches that of σ_V, so we reduce the degree to which σ_H is subtracted. This average σ_0 is then used in a polynomial regression of the following form to retrieve a pseudo wind speed.

$$W_{PSEUDO} = a_{1/2}\sigma_0^{1/2} + a_1\sigma_0 + a_2\sigma_0^{2} + a_3\sigma_0^{3} + a_4\sigma_0^{4}.$$

The coefficients $a_{1/2} - a_4$ are chosen by fitting to the full NSCAT-2 model function.

We then use this pseudo wind speed to divide the measurements into bins by polarization, cross-swath cell number, and wind speed. Only QuikScat measurements for which a 3 x 3 pixel area of SSM/I measurements centered on the QuikScat measurement both were measured within 60 minutes and were rain free were used. All rain-free collocated measurements for QuikScat revs 900-999 were used. For each bin,

a histogram of squared-difference deviations was calculated, and the value of the 95th percentile was used for a first-guess value for W_i. The idea is that any measurement for which the squared difference exceeds this value is of suspect quality. The 95th percentile was used, rather than the mean, to help account for variations in the width of the distribution of squared-difference deviations for different cells, wind speeds, and polarizations. A surface plot of the Logarithm of the 95th percentile as a function of pseudo wind speed and cell number is shown in Figures 1a (H-Pol) and 2a (V-Pol). Since the unsmoothed values were noisy, especially at high wind speed where there was not much data in the 100 orbits used to determine Wi, we performed a series of smoothing steps to obtain the surfaces shown in Figures 1d and 2d.

The values of Wi shown in Figures 1 and 2 are used to compute the values of the ENOF for each wind vector cell. A threshold is then selected which translates the ENOF value to a binary rain flag and results in a significant chane that the wind vector cell is contaminated by rain. We chose three representative thresholds that flag ~5%, ~7.5%, and ~10% of the wind vector cells when applied globally. These thresholds are summarized in the table below.

ENOF Threshold Value	Percent of Cells Flagged, revs 1550-1850
2.16	5.0%
1.51	7.5%
1.24	10.0%

IV. Rain Detection Performance.

To estimate the performance of the rain flag, we compare the results with the SSM/I rain-rate product produced by RSS. Only SSM/I measurements that occur within 30 minutes of QuikScat measurements were used. Before comparison, the SSM/I rain rates were converted to columnar rain densities by multiplying the rain rate by and estimate of the column height. The estimate we used matched the estimate originally used in the SSM/I algorithm to retrieve rain rate, since SSM/I more directly measures rain column than rain rate. We convert back to columnar rain, since the scattering and attenuation due to the rain is directly related to the path length through the rain multiplied by the density of the rain. This product is directly proportional to rain column.

We begin by showing two examples, one example where the rain flag works well, and another where it fails. The first Example is shown in Figure 3. This figure consists of four plots. In Fig. 3a, we show the wind vectors from the ECMWF global analysis, color-coded by wind speed. In the Fig. 3b, we plot the QuikScat wind vectors, color-coded by SSM/I rain column. Green indicates light rain, with a rain column between 0.0 and 3.0 km mm/hr. Red indicates heavier rain, with rain above 3.0 km mm/hr, black indicates no rain, and blue indicates no collocated SSM/I data. In Fig 3c, we again show the QuikScat wind vectors, color-coded using the ENOF rain flag. Vectors with ENOF values above 1.51 are shown in red, those below are shown in blue. Note that we have effectively flagged most of the vectors that SSM/I has measured to be in regions of moderate to heavy rain, except in the outer swath region, where only V-Pol observations are available. In the Southern part of the region, we have flagged many vectors for which SSM/I reports no rain, but which are adjacent to rainy regions. We believe that many of these vectors are in fact influenced by rain and are correctly flagged. The vectors are in many cases obviously wrong (wind speed much too high, and pointing cross-swath), and are in regions where SSM/I reports significant cloud

cover, as shown in Figure 3d. The figure also shows a significant number of isolated vectors that are flagged as rain. Since they often occur in regions of low clod cover, these are probably truly false alarms, caused by random noise in the σ_0 measurements. These could easily be removed using a spatial filtering technique that requires that one or more near neighbors also be flagged, or the flag is turned off. We have demonstrated the effectiveness of such a flag, but it is not implemented at this time.

The rain flag is quite successful in Figure 3, removing most if not all of the vectors that are clearly influenced by rain. This was possible because in this instance, the rain was accompanied by low wind speed. As discussed above in Section II, it is much easier to see the effects of rain in regions of low wind speed, due to the low sea-surface σ_0's, and to the low σ_H to σ_V ratio. Performance similar that illustrated here is exhibited in the low wind, high rain tropical convergence regions in the Pacific, Atlantic and Indian Oceans. This is very encouraging, since it is in these regions that the wind vectors disagree most strongly with modeled wind, and wind speeds retrieved by SSM/I.

Figure 4 is a second example of our rain flag. In this example, the rain-flagging algorithm is much less successful at identifying regions of rain. Here, the rain is associated with a mid-latitude storm with high wind speed, located southeast of the southern tip of Africa. As can be seen in Figs. 4a and 4b, there are large regions of the storm with high wind speed and high rain rate. None of these regions are flagged as having rain by the ENOF rain flag, even though there are some QuikScat vectors that appear to be affected by rain, such as the vectors pointing cross-swath near 65 E, 50S. As discussed in Section II, we expected that rain would be much harder to detect in regions of high wind speed. We also see similar failure of the rain flag in the high wind areas of tropical cyclones.

Performance metrics are used to further evaluate the rain detection algorithm. An SSM/I rain column threshold $R_{SSM/I}$ of was required to determine the cells contain rain. We then evaluated the performance of the algorithm as a function of this threshold. The results we discuss here are for the case when the ENOF threshold has been set to 1.51, and about 7.5% of the wind vector cells are flagged as rain. We also only consider data in wind vector cells 10-66, where both H-Pol and V-Pol measurements are available. We primarily use two metrics to determine the success of our rain flag algorithm. These are summarized below.

Misclassification Rate	The percent of cells containing rain as determined by SSM/I, i.e. cells with an SSM/I rain column above $R_{SSM/I}$, but that are not flagged by the rain algorithm. In other words, the percent of raining wind vector cells that are missed.
False Alarm Rate	The percent of cells for which SSM/I indicates **no** rain, but that are flagged by the rain algorithm. The idea is that if SSM/I indicates any rain at all, it is OK for the algorithm to flag it as rain. Note that the False Alarm Rate does not depend on the $R_{SSM/I}$.

Ideally, we would like both the misclassification rate and the false alarm rate to be as small as possible. In Figure 5, we plot these rates as a function of $R_{SSM/I}$ for QuikScat revs 1550-1850. As the rain-column threshold increases, the misclassification rate falls from 33% at 3 km mm/hr to about 20% at 10 km mm/hr. The false alarm rate is just over 5%. Use of the spatial filtering technique described above

improves the misclassification rate as shown by the dashed line. When we separate the data into three wind speed bins, we find the lowest misclassification rates occure for the lowest wind speed bin, from 3 m/s to 8 m/s. The dotted-dashed line shows that at these wind speeds, we are misclassifying less than 10% of the wind vectors for columnar rain rates above 5 km mm/hr. For winds from 8 m/s to 15 m/s, the misclassification rate increases to about 30%, and to 65-75% for winds above 15 m/s. This confirms what we saw in the examples above – the rain flag works well for low wind speeds, but misses many instances of rain for high wind speeds.

The use of the ENOF rain flag to exclude rain-contaminated wind vector cells significantly improves the agreement of QuikScat wind vectors with those produced by general circulation models such as those used by NCEP and ECMWF for weather forecasting and data assimilation. In the table below, we report the RMS wind speed difference and the RMS wind direction difference between QuikScat and ECMWF when different methods of rain flagging are used. These statistics were obtained by averaging collocated QuikScat-SSM/I observation over QuikScat revs 1550-1850.

	No Rain Flag	Rain Flagged with ENOF (7.5%)	Rain Flagged with SSM/I
RMS Wind Speed Difference	2.22	1.72	1.63
RMS Wind Direction Difference	25.5	23.4	23.3

The SSM/I rain flag used to produce these statistics is very conservative. If the closest SSM/I cell to the QuikScat observation, or any of its eight nearest neighbors shows any rain, the observation is discarded. Use of the ENOF flag improves the agreement with the ECMWF model nearly as much as using this conservative SSM/I rain flag. This result is very encouraging, indicating that the ENOF rain flag removes erroneous wind vectors almost as effectively as SSM/I collocated rain measurements, but does not require an SSM/I collocated measurement.

Another method that is useful for evaluating the performance of the rain flag is to examine histograms of what is and is not flagged by the ENOF and SSM/I methods. In Figure 6, we show wind speed histograms of four subsets of ECMWF and QuikScat wind vectors. In Figure 6a, we show histograms of the ECMWF wind speeds (black) and QuikScat wind speeds (red) for wind vector cells that were flagged as having rain by both SSM/I and the ENOF rain flag. Note that the distribution of QuikScat winds is shifted to much higher wind speed that the ECMWF winds, indicating that these winds were significantly affected by the presence of rain, and **should** be eliminated. In Figure 6b, we show a similar set of histograms, except that these are cells that SSM/I indicated contained rain, but were not flagged by the ENOF algorithm. These cells contribute to the misclassified percentage discussed above. Once again, the QuikScat winds are much higher, indicating that these winds are in fact affect by the presence of rain, i.e. they **should** have been flagged. The ECMWF winds for this subset are much higher than in Figure 6a. This, once again, is indicative of the relative insensitivity of the ENOF method at high true wind speed. In Figure 6c, we plot histograms for the false alarm cases, where the ENOF indicates that rain is present, but no rain is measured by SSM/I. Here too, the QuikScat winds are shifted to higher wind speed, indicating that rain, or something, is influencing the retrieved winds – i.e. these winds too, are erroneous and **should** be eliminated. In fig 6d, we plot histograms for the case when neither SSM/I nor the ENOF algorithm

indicated rain. In this case, the wind speeds agree quite well, except for a small high bias of the quikscat wind speeds relative to ECMWF. If a similar comparison to NCEP wind speeds is done, the QuikScat winds show much less bias, indicating that this bias may be due to a bias in the ECMWF model. The low overall bias for this case lends weight to the significance of the large biases observed for the other three cases. It also indicates that we were "right" about these winds – they **should not** be eliminated.

V. Summary.

We have developed a rain detection algorithm based on an empirically normalized objective function. The algorithm flags wind vector cell where the measured σ_0's are in poor agreement with those calculated using the retrieved wind vector and the model function. We judge that the method is only effective in the central part of the QuikScat swath, where both H-Pol and V-Pol observations are present, and is not currently implemented far-swath regions. The algorithm does an excellent job of flagging rain at low wind speeds, where most of the significant rain is detected. At higher wind speeds, the efficiency of the algorithm is not as good. It may be that the flagging of the rain is not as important at higher winds, since the measured σ_0's are less affected by the presence of rain. This is supported by the nearly identical level of agreement between QuikScat and ECMWF winds obtained when two different methods of rain flagging, ENOF and SSM/I, were used to choose a rain-free subset of the QuikScat observations.

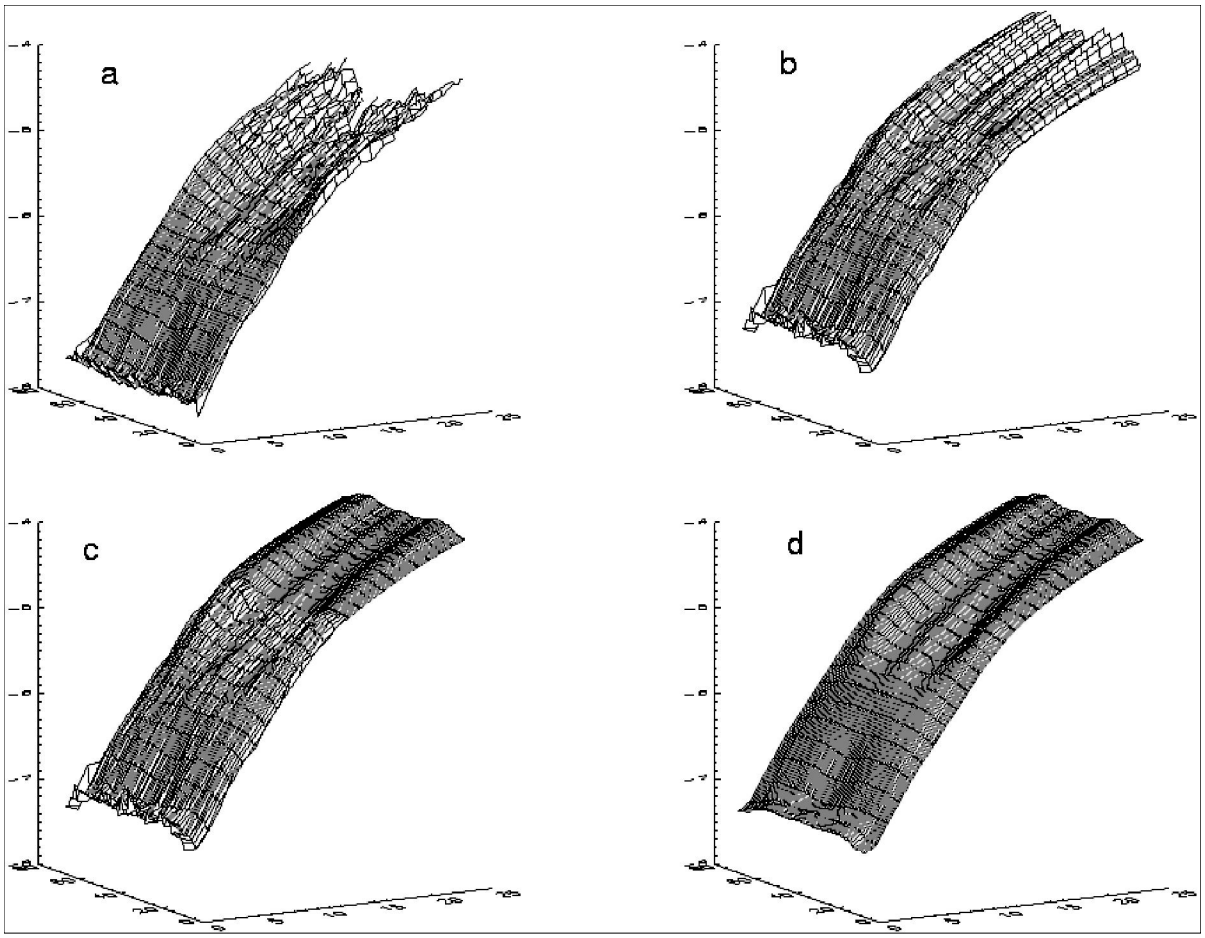

Figure 1. Surface plot of the Logarithm of W_i as a function of pseudo wind speed and cross-swath cell index. (a) The unsmoothed statistic. In (b), we have fit a power law for the data from 13 m/s to 18 m/s, and used it to replace the data above 18 m/s. In (c) we have smoothed the fit parameters in the cross-swath direction. In (d), we have smoothed the data below 18 m/s. In all cases, the smoothing procedure also enforces symmetry with respect to the center of the swath.

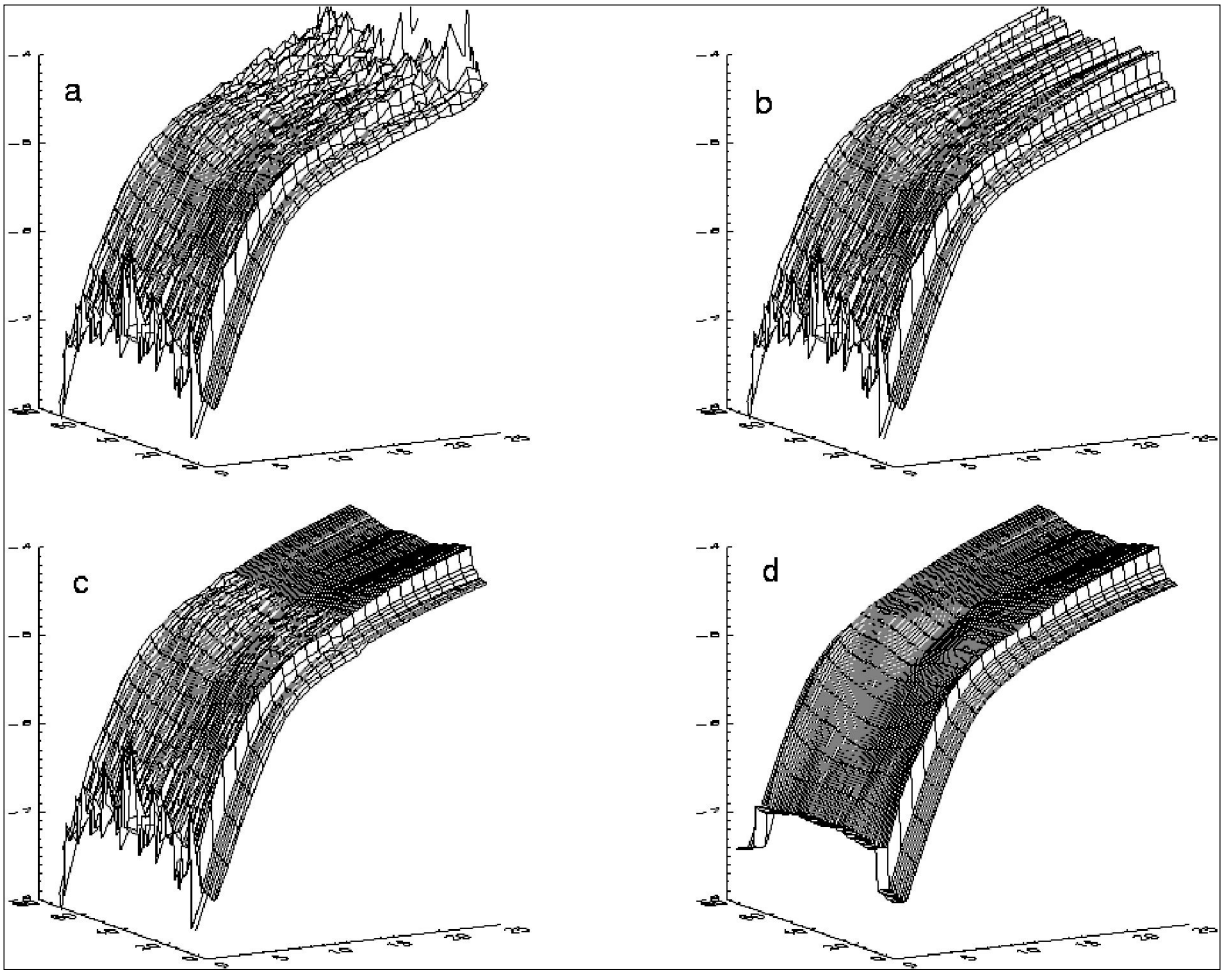

Figure 2. Same as Figure 1, except for V-Pol measurements. Note the marked decrease in expected variance as we move from the mid-swath region, where both V-Pol and H-Pol measurements are present, to the far-swath region, where only V-Pol measurements are present. This strongly suggests that the measurements with the same polarization contain correlated errors that are not present for comparisons between measurements with different polarizations.

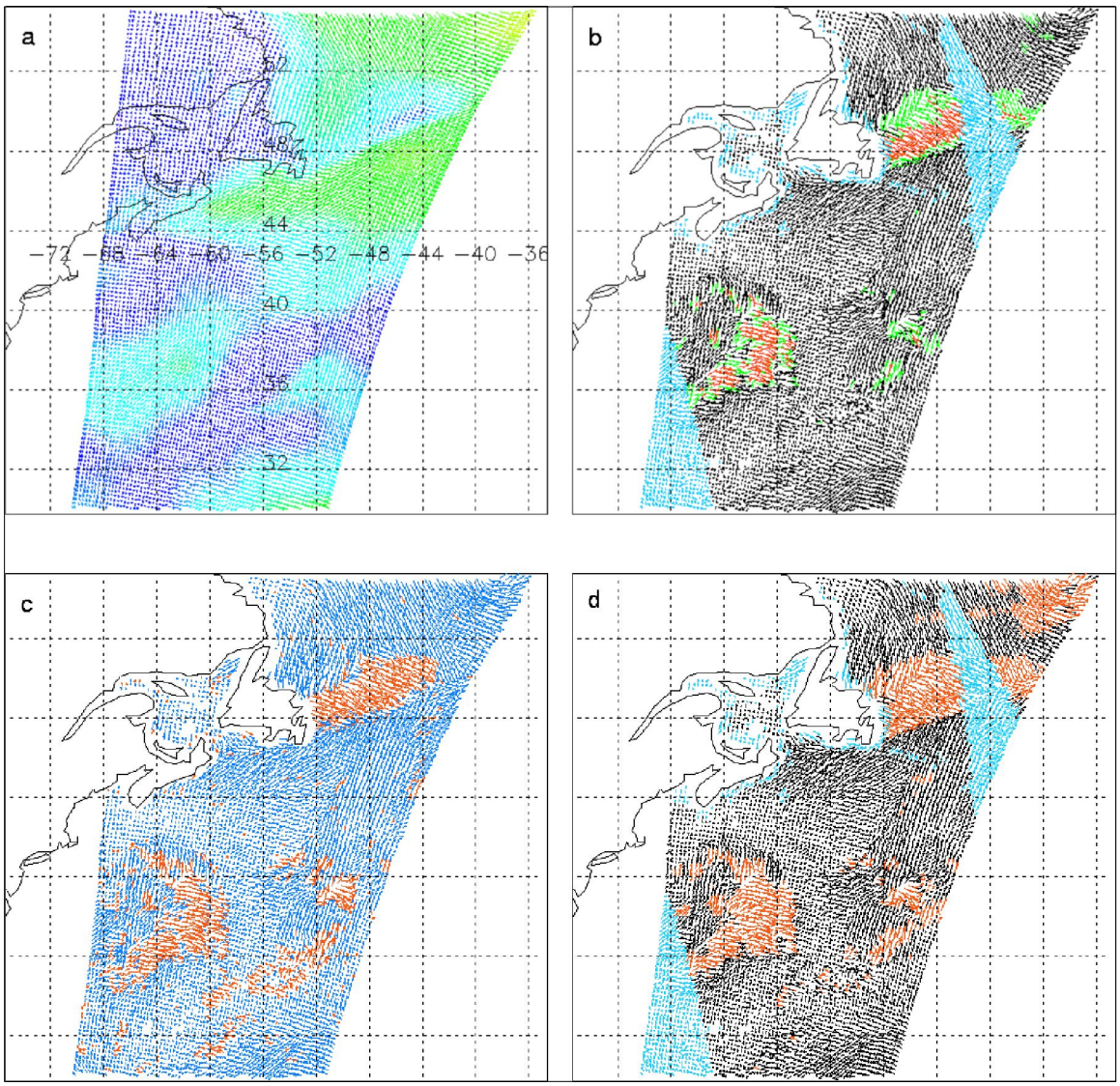

Figure 3. Example of the ENOF rain flag working well. In (a) we show ECMWF global analysis wind vectors for comparison, color-coded by wind speed. In (b), we show QuikScat wind vectors color-coded by SSM/I rain column. Black vectors indicate no SSM/I rain, green vectors indicate rain between 0.0 and 3.0 km mm/hr, and red vectors indicate rain column above 3.0 km mm/hr. Blue vectors indicate no SSM/I data. In (c), we plot QuikScat vectors color-coded by ENOF rain flag with a threshold of 1.51. Red vectors indicate rain and blue vectors indicate no rain. Note the excellent correspondence with (b) except in the far swath region. In (d) we plot QuikScat vectors color-coded by SSM/I cloud water. Red vectors indicate cloud water above 0.15 mm. Note that there is significant cloud water in many of the regions flagged by the ENOF flag as rain, but not flagged by SSM/I as rain.

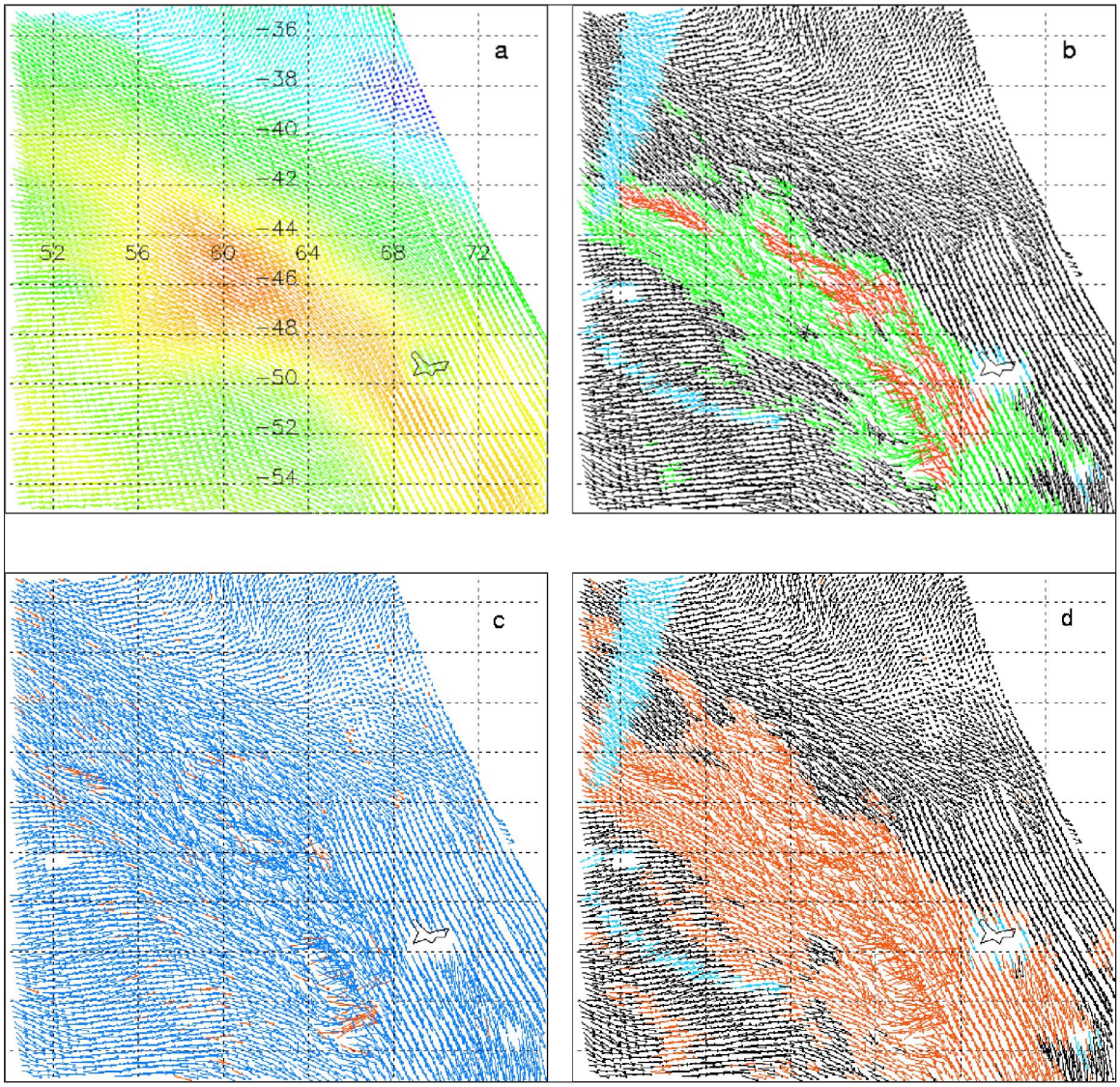

Figure 4. Example of the ENOF rain flag not working well. In (a), we show ECMWF global analysis wind vectors for comparison, color-coded by wind speed. In (b), we show QuikScat wind vectors color-coded by SSM/I rain column. Black vectors indicate no SSM/I rain, green vectors indicate rain between 0.0 and 3.0 km mm/hr, and red vectors indicate rain column above 3.0 km mm/hr. Blue vectors indicate no SSM/I data. In Fig (c), we plot QuikScat vectors color-coded by ENOF rain flag with a threshold of 1.51. Red vectors indicate rain and blue vectors indicate no rain. Note that the rain flag misses nearly all the rain measured by SSM/I, except in a relatively low-wind region near 65E, 52S. In Figure (d), we plot QuikScat vectors, color-coded by SSM/I cloud water. Red vectors indicate cloud water above 0.15 mm.

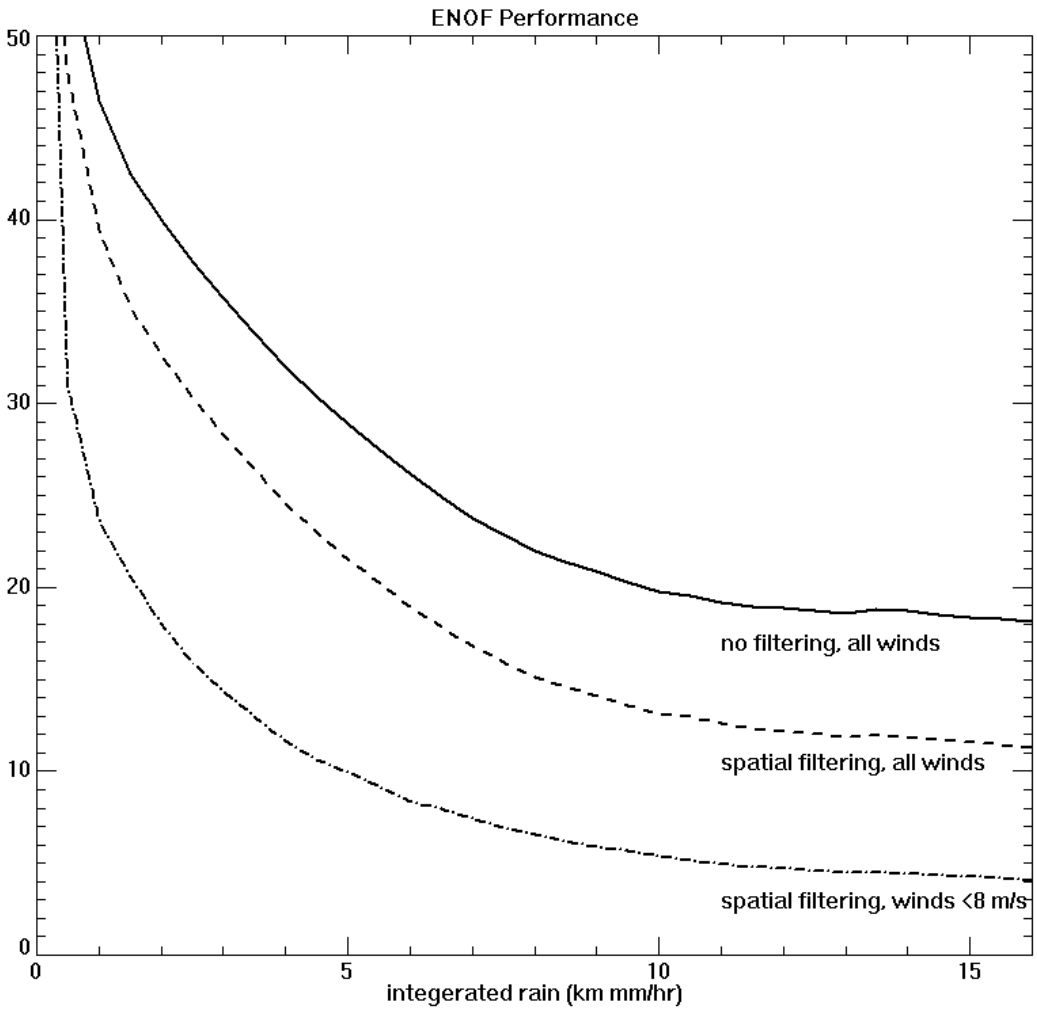

Figure 5. This plot shows the misclassification rate as a function of SSM/I rain column for an ENOF threshold value of 1.51. The solid line represents the performance of the data provided within this dataset. The dashed line represents the performance of the data using a nearest neighbor spatial filtering technique to remove isolated flagged wind vectors. This algorithm performs best at low wind speeds as shown by the dotted-dashed line representing the misclassification rate of spatially filtered wind speeds below 8 m/s. Low winds are most affected by rain within the signal path. Using this rain flag, we can remove greater than 90% of these erroneous wind vectors when the rain column is greater than 5 km mm/hr. The false alarm rates for each curve are 5.5, 6.1 and 6.6 respectively.

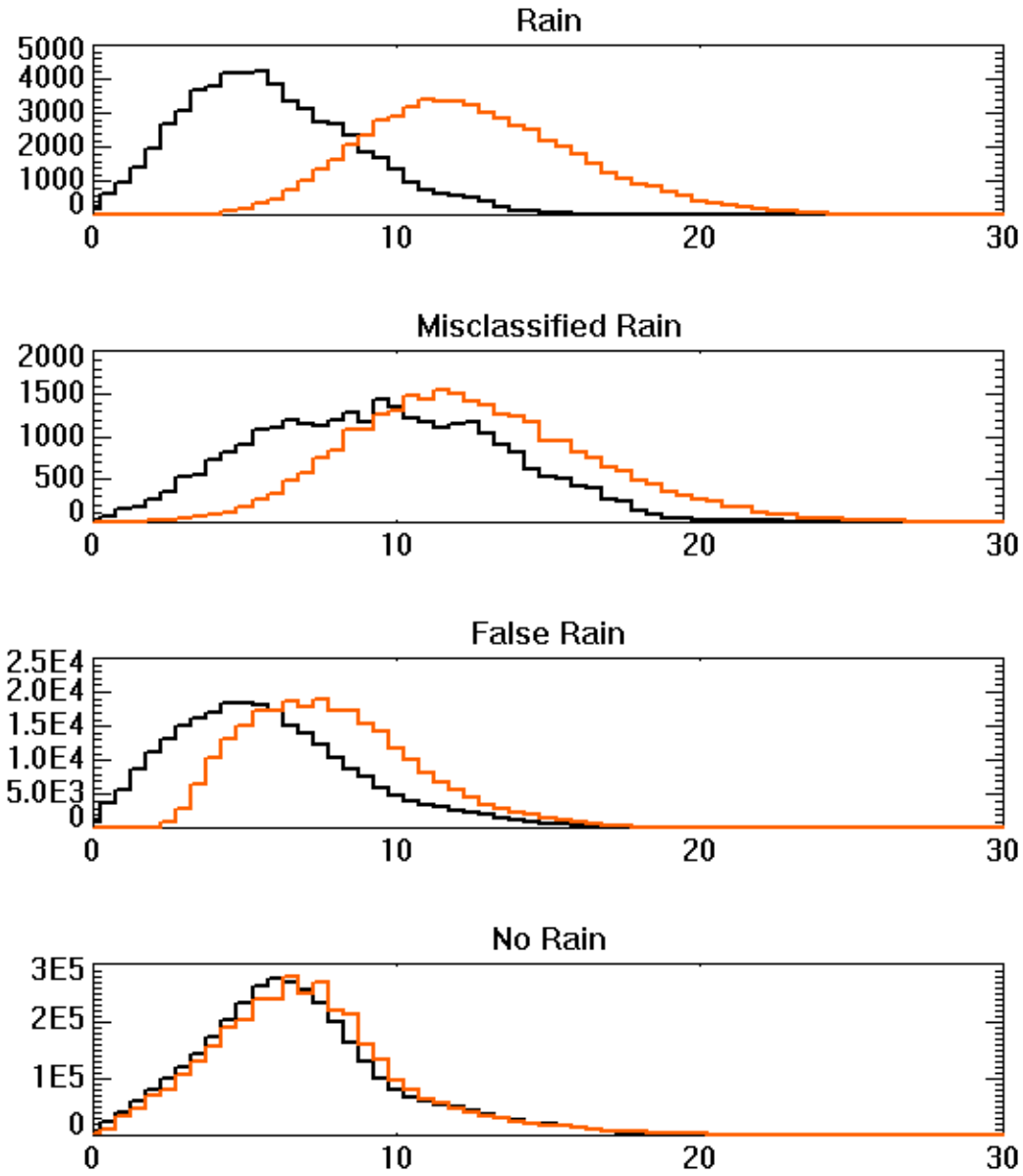

Figure 6. These wind speed histogram plots from collocated QuikScat data for four conditions demonstrate the ability of the rain flag algorithm to remove erroneous QuikScat wind data. For wind vector cells in which both SSM/I and QuikScat rain flag identify rain, we see the increased QuikScat wind speeds (red) due to rain effect on the scatterometer signal (a). The peak of the wind speed histograms for the misclassified rain (b) and the false rain (c) is shifted for the QuikScat data as compared to the ECMWF data (black). The wind speed histograms are in good agreement when no rain is present despite a slight bias between the datasets (d).