Rain Detection and Quality Control of SeaWinds[1]

M. Portabella, A. Stoffelen

KNMI, Postbus 201, 3730 AE De Bilt, The Netherlands

Phone: +31 30 2206827, Fax: +31 30 2210843

e-mail: portabel@knmi.nl, stoffelen@knmi.nl

Abstract

A good assessment of the information content of scatterometer winds is particularly important in order to assimilate them in weather analysis. Besides retrieval problems in cases of a confused sea state, a particularly acute problem of Ku-band scatterometry is the sensitivity to rain. Elimination of poor quality data is therefore a prerequisite for the successful use of the National Aeronautics and Space Administration (NASA) Scatterometer (NSCAT) or QuikSCAT winds. Following the Quality Control for the European Remote-Sensing Satellite ERS and NSCAT scatterometers performed at Royal Netherlands Meteorological Institute (KNMI), the authors further develop this methodology for QuikSCAT and define a quality indicator, called the normalized residual (Rn). In order to characterize and validate the

[1] Manuscript reference: Portabella, M., and Stoffelen, A., "Rain Detection and Quality Control of SeaWinds," *J. Atm. and Ocean Techn*., Vol. 18, No. 7, pp. 1171-1183, 2001, © American Meteorological Society.

normalized residual, the authors use collocated Special Sensor Microwave Imager (SSM/I) rain and European Centre for Medium-Range Weather Forecasts (ECMWF) wind data. The results show indeed correlation between Rn and data quality. A wind speed dependent Rn threshold is shown to be adequate in terms of rejecting poor quality data (particularly rain) and keeping fair quality data. This opens the way to a quantitative use of SeaWinds measurements in weather analysis.

1 Introduction

The SeaWinds on QuikSCAT mission is a “quick recovery” mission to fill the gap created by the loss of data from the the National Aeronautics and Space Administration (NASA) Scatterometer (NSCAT), when the ADEOS-1 satellite lost power in June 1997. QuikSCAT was launched from Vandenberg Air Force Base (USA) on 19 June, 1999. A very similar version of the SeaWinds instrument will fly on the Japanese ADEOS-II satellite currently scheduled for launch in late 2001.

The SeaWinds instrument is an active microwave radar designed to measure the electromagnetic backscatter from the wind-roughened ocean surface. The instrument is a conical-scanning pencil-beam scatterometer, which in comparison with the NSCAT fan-beam scatterometer has the following advantages: higher signal-to-noise ratio, smaller size, and superior coverage.

The SeaWinds instrument uses a rotating 1-meter dish antenna with two spot beams, an horizontally polarized (H-pol) beam and a vertically polarized (V-pol) beam at

incidence angles of 46º and 54º respectively, which sweep in a circular pattern. The antenna radiates microwave pulses at a frequency of 13.4 GHz (Ku-Band) across a 1800-km-wide swath centered on the spacecraft's nadir subtrack, making approximately 1.1 million 25-km ocean surface wind vector measurements and covering 90% of the Earth's surface every day. These measurements will help determine atmospheric forcing, ocean response, and air-sea interaction mechanisms on various spatial and temporal scales.

The SeaWinds swath is divided into equidistant across-track wind vector cells (WVC) or nodes numbered from left to right when looking along the satellite's propagation direction. The nominal WVC size is 25 km x 25 km, and all backscatter measurements centered in a WVC are used to derive the WVC wind solutions. Due to the conical scanning, a WVC is generally viewed when looking forward (fore) and a second time when looking aft. As such, up to four measurement classes (called "beam" here) emerge in each WVC: H-pol fore, H-pol aft, V-pol fore, and V-pol aft, in each WVC. Due to the smaller swath (1400 km) viewed in H-pol at 46º degrees incidence, the outer swath WVCs have only V-pol fore and aft backscatter measurements. For more detailed information on the QuikSCAT instrument and Jet Propulsion Laboratory (JPL) data we refer to [*Spencer et al* (1997), *JPL* (1999), *Leidner et al* (2000)].

The forecast of extreme weather events is not always satisfactory, while its consequences can have large human and economic impact. The lack of observations over the oceans, where many weather disturbances develop, is one of the main problems of Numerical Weather Prediction (NWP) for predicting their intensity and position. A spaceborne scatterometer with extended coverage like SeaWinds is able to

provide accurate winds over the ocean surface and can potentially contribute to improve the situation for tropical and extratropical cyclone prediction [*Isaksen and Stoffelen* (2000) and *Stoffelen and Van Beukering* (1997)].

The impact of observations on weather forecast often critically depends on the Quality Control (QC) applied. *Rohn et al.* (1998) show a positive impact of cloud motion winds on the European Centre for Medium-Range Weather Forecasts (ECMWF) model after QC, while the impact is negative without QC. Therefore, in order to successfully assimilate QuikSCAT data into NWP models, a comprehensive QC needs to be done in advance. *Stoffelen and Anderson* (1997) and *Figa and Stoffelen* (2000) use a method to detect and reject WVCs with poor quality wind information using a Maximum-Likelihood-Estimator-based (MLE) parameter for European Remote-Sensing Satellite (ERS) and NSCAT, respectively. Here, we adapt this method for QuikSCAT.

The MLE indicates how well the backscatter measurements used in the retrieval of a particular wind vector fit the Geophysical Model Function (GMF), which is derived for fair weather wind conditions. A large inconsistency with the GMF results in a large MLE, which indicates geophysical conditions other than those modeled by the GMF, such as rain, confused sea state, or ice, and as such, the MLE provides a good indication for the quality of the retrieved winds.

Rain is known to both attenuate and backscatter the microwave signal. *Van de Hulst* (1957) explains these effects. Rain drops are small compared to radar wavelengths and cause Rayleigh scattering (inverse proportional to the fourth power wavelength). Large drops are relatively more important in the scattering, and smaller wavelengths more sensitive. For example, *Boukabara et al.* (2000) show the variation of the signal

measured by a satellite microwave radiometer with the rain rate. As the rain rate increases, the spaceborne instrument sees less and less of the radiation emitted by the surface, and increasingly sees the radiation emitted by the rainy layer that becomes optically thick due to volumetric Rayleigh scattering.

Comparing Ku-band (13.4 GHz) to C-band (5 GHz) scatterometers, the higher frequency of the former makes both effects (rain attenuation and scattering) much more relevant. In particular, as SeaWinds operates at high incidence angles and therefore the radiation must travel a long path through the atmosphere, the problem of rain becomes acute.

In addition to these effects, there is a “splashing” effect. The roughness of the sea surface is increased because of splashing due to rain drops. This increases the radar backscatter (σ^o) measured, which in turn will affect the quality of wind speed (positive bias due to σ^o increase) and direction (loss of anisotropy in the backscatter signal) retrievals.

In section 2 of this paper we present the set of data used; in section 3, the MLE-based QC method; in sections 4 and 5, the characterization and validation of the normalized residual (Rn), respectively; in section 6, a few cases to illustrate the performance of this QC; and finally in section 7, some conclusions and recommendations.

2 Data

In order to characterize and validate the QC, we collocate a set of 180 orbits of QuikSCAT Hierarchical Data Format (HDF) data with ECMWF winds and Special Sensor Microwave Imager (SSM/I) rain data. The HDF data correspond to the preliminary science data product produced by JPL using the NSCAT-2 GMF.

We use the analyses, and 3-hour forecast ECMWF winds on a 62.5-km grid, and we interpolate them both spatially and temporally to the QuikSCAT data acquisition location and time, respectively.

The collocation criteria for SSM/I rain data are less than 30 minutes time and 0.25º spatial distance from the QuikSCAT measurement.

The SSM/I instruments are on board Defense Meteorological Satellite Program (DMSP) satellites. We have used DMSP F-13 and F-14 satellites (the most recent ones). Most of the collocations with F-13 were found at low latitudes (Tropics), while collocations with F-14 were found at mid and high latitudes.

3 Methodology

The method consists in normalizing the MLE with respect to the wind and the node number (or cross-track location). In this paper, we use the following MLE (adopted from *JPL, 1999*) :

$$MLE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(s_{mi}^{o} - s_{s}^{o}\right)^2}{kp\left(s_{s}^{o}\right)}, \qquad (1)$$

where N is the number of measurements, σ_{mi}^{o} is the backscatter measurement, σ_{s}^{o} is the backscatter simulated through the GMF for different wind speed and direction trial values, and $kp(\sigma_{s}^{o})$ is the measurement error variance.

Stoffelen and Anderson (1997) interpret the MLE as a measure of the distance between a set of σ_{mi}^{o} and the solution σ_{s}^{o} lying on the GMF surface in a transformed measurement space where each axis of the measurement space is scaled by $kp(\sigma_{s}^{o})$.

In order to normalize the MLE, we compute the expected MLE for a given wind and node number. Then we define the normalized residual as

$$Rn = MLE / \langle MLE \rangle, \qquad (2)$$

where MLE is the maximum likelihood estimator of a particular wind solution (given by the inversion) and <MLE> is the expected MLE for that particular WVC (node number) and wind solution.

We hypothesize that the MLE is very much altered in the case of rain and therefore is very different from the expected MLE. A set of σ^o_m coming from a "rainy" WVC (or a WVC where some geophysical phenomena other than wind is "hiding" the wind-related information) is expected to be inconsistent with any set of σ^o_m that is simulated from the GMF, since basic properties of the backscatter measurements, such as HH–to-VV polarisation ratio (Wentz et al., 1999) and the isotropy of scattering at the ocean surface are expected to be different. In other words, the set of backscatter measurements coming from a "rainy" WVC is expected to be farther away from the GMF than a set of measurements coming from a "windy" WVC (which should lie very close to the GMF). Therefore, the MLE is much higher than <MLE>, and the normalized residual (Rn) is high. In contrast, the MLE of a "windy" WVC is closer to the <MLE>, and consequently we have Rn values of the order of 1.

In order to compute the expected MLE for a given wind and node number, we study the dependencies of the MLE with respect to the wind speed, wind direction, and node number over 60 revolutions of real data.

[Note: The method to compute Rn can be based on the MLE of the first-rank or the selected (after ambiguity removal) solution of the inversion scheme. In this paper, we present the results of the Rn based on the MLE of the JPL selected solution.]

Figure 1 shows the mean JPL-selected MLE versus the JPL-selected wind direction (dotted line) together with the wind direction distribution of both the JPL-selected solutions (dashed line) and the National Center for Environmental Prediction (NCEP) model (solid line) for WVC number 16. The JPL direction distribution shows some significant differences (peaks and troughs) as compared to the NCEP distribution, which may be associated with some deficiencies in the inversion and the NSCAT-2

model function. We note that the mean MLE is following these relative peaks and troughs of the JPL wind direction distribution with respect to NCEP, not only in this particular WVC but also in the rest of the swath (not shown). This is an expected result, as measurement sets far away from the GMF solution surface in measurement space (*Stoffelen*, 1998), that is, with large MLE, are systematically assigned to certain wind directions (the shape of the GMF solution surface makes certain wind directions to be favored in such cases). However, these peaks are due to an inversion problem and not to a realistic MLE dependency on wind direction. In other words, the mean MLE peaks are not always showing a real MLE dependency on wind direction but just some artificial accumulation of wind direction solutions due to some deficiencies in the inversion. Therefore, we discard the wind direction dependency when computing the <MLE>.

As such, <MLE> is computed as a function of wind speed and node number. The method to compute <MLE> is as follows:

- We compute the mean MLE of the JPL-selected solution with respect to the JPL-selected wind speed and the node number for the 60 revolutions of data. The surface is a bit noisy, which is mainly due to geophysical effects such as rain, which we want to discard. At high wind speeds, additional noise is present due to the small amount of data we get at these speeds and node numbers.

- In order to filter the noise on the surface, we set up an iterative process, which consists of rejecting the MLEs that are two times higher than the mean MLE for that particular wind speed and node number, and we compute again the mean MLE surface. Then, we start the rejection process again in an iterative mode until it converges (no more rejections). The process converges very rapidly after two

iterations, and the number of data rejected in each speed and node bin is very small (up to 7% in some high-speed bins). This gives faith in the noise filtering method, as it shows that only the tail of the MLE distribution is cut in each bin, corresponding to geophysical anomalies. In order to show the consistency of this filtering procedure, we show the contour plots of both mean MLE surfaces (before and after filtering) in Figure 2. It is clearly discernible that the shape of the surface remains the same and only the noise (especially at high speeds) has been removed.

- In order to extrapolate to high wind speeds, we fit in a very simple way a two-dimensional function to the filtered surface. The function is only fit in the inner swath (nodes 12 to 65) and extrapolated for the outer swath (nodes 1 to 11 and 66 to 76) (see discussion below). The filtered mean JPL-selected MLE surface and the result of the fitting are shown in Figure 3. The surface fit (plot b) is the expected MLE and compares well to the computed mean MLE (plot a) in the inner part of the swath.

As said before, anomalous weather conditions are expected to result in a large Rn. However, this correlation between Rn and data quality is expected to work well when there are more than two measurements and enough azimuth diversity in the σ^o measurements of each WVC, that is, when the inversion problem is overdetermined. In the outer parts of the swath, where there are only one or two beams (fore and/or aft VV), the wind vector is not overdetermined, and generally, multiple wind speed and direction combinations exist that exactly fit the measurements. Then the MLE is going to be zero or very close to zero in most of the cases, regardless of the quality of the data. Only for the exceptional case, when the MLE is substantially larger than our extrapolated <MLE>, can we infer that the data are of bad quality in these parts of the

swath. This means that our QC procedure is generally not going to work well in the outer regions. Nevertheless, to provide a gross check and in order to arrive at a simple function fit to the <MLE> surface, we have extrapolated the inner swath function to the outer regions.

Once Rn is defined, we have to characterize it. The way to characterize Rn is to test it against a variety of geophysical conditions such as rain, confused sea state (in wind fronts, center of lows, coastal regions), or just pure wind cases. As the method is expected to work fine in the inner swath, especially in its sweet parts (nodes 12 to 28 and 49 to 65), we are going to focus our research on these regions.

4 Rn characterization

In this section, we study the correlation between Rn and the quality of QuikSCAT winds. Collocated ECMWF winds and SSM/I rain are used as characterization and validation tools. Note that both the ECMWF winds and SSM/I rain data contain uncertainties and obey different space and time representations than the QuikSCAT winds.

Characterizing Rn results in a QC procedure by finding a threshold value of Rn, which separates the good-quality from the low-quality retrieved winds.

As said in section 3, the Rn is defined from the MLE of the JPL-selected solution. Therefore, if we identify a low-quality wind selected solution, we will assume that all wind solutions in that particular WVC are of low quality. This means that the QC is

performed on a node-by-node basis. The WVCs that are accepted may have wind solutions with MLE above the threshold. These solutions are kept but will be down-weighted in the data assimilation procedure (*Stoffelen et al.*, 2001).

We characterize Rn in the sweet parts of the swath, where it is most meaningful. However, as we show in the validation, the threshold is useful in the entire inner swath.

4.1 Rn as a quality indicator

The first step in the characterization of this QC procedure is to confirm the correlation between Rn and the quality of the data. The vector RMS difference between the JPL-retrieved (selected solution) and ECMWF winds (RMS-ECMWF) is used as a quality indicator.

Figure 4 shows a contour plot of a two-dimensional histogram of RMS-ECMWF against Rn. We set an arbitrary threshold at RMS=5 m/s, which is roughly separating the "good" from the "bad" quality cases. Plot a), which represents the whole collocated dataset, shows a clear correlation between RMS-ECMWF and Rn. Most of low Rn cases, represented by the two darkest grey-filled contours (remember that the contours are in logarithmic scale), are of good quality. The RMS-ECMWF increases as Rn increases, which means that as expected, the quality of the data is decreasing while Rn increases (i.e., Rn is a good quality indicator).

Plots b), c), and d) show histogram distributions for different wind speeds. The RMS-ECMWF is increasing more rapidly with Rn at higher wind speeds. The quality of the

data is poor for lower values of Rn as the retrieved wind speed increases. This suggests a Rn threshold dependent on the retrieved wind speed, with a threshold value smaller at high wind speeds than at low speeds.

4.2 Effects of rain

As said before, the Ku-band signal is known to be distorted in the presence of rain. In order to study this distortion effect, SSM/I collocations are used as a rain indicator.

Figure 5 shows both the mean retrieved wind speed (plot a) and the mean ECMWF wind speed (plot b) versus the rain rate. The retrieved wind speed is increasing with the rain rate while the ECMWF wind speed shows obviously no significant dependence on the rain. As the rain rate increases, the density and size of the droplets increase and the probability of having a homogeneous rainy WVC (no patches with absence of rain) increases. Therefore, the wind information contained in a particular WVC is increasingly hidden, and the backscatter signal is becoming more and more “rain-related” instead of “wind-related.” From these plots, one may infer that the “rainy” WVCs produce high winds in the retrieval process.

Figure 6 shows the two-dimensional histogram of RMS-ECMWF versus the retrieved wind speed for rain-free (plot a) and for different rain rate intervals (plots b and c). The upper plot shows a generally horizontal orientation of the contour lines, while the bottom plots show mainly a vertical orientation, suggesting a decline of the data quality (higher mean RMS-ECMWF) in the presence of rain. At rain rates higher than 6 mm/hr, most of the data are above the RMS threshold of 5 m/s, indicating no useful

wind information in them. However, when the rain is lower than 6 mm/hr, there is still a significant portion of the retrieved winds with low RMS and therefore containing significant wind information in their backscatter signal. We want to define a threshold capable of removing those “rainy” WVCs with rain rates over 6 mm/hr and those with lower rain rates but no significant wind information (high RMS-ECMWF values) in them.

4.3 Rn threshold

Up to now, we have achieved three major conclusions:

a) The Rn seems a good quality indicator

b) When it is rainy, the retrieved wind speed is in general too large by an amount which is proportional to rain rate

c) For SSM/I rain rates above 6 mm/hr, the WVCs contain no valuable wind information.

Figure 7 summarizes all these points. The left plots correspond to two-dimensional histograms of Rn versus JPL-retrieved wind speeds (selected solution) for different rain rate intervals. The right plots are the same histograms of Rn but versus ECMWF wind speed. In the absence of rain (upper plots), we clearly discern the significant difference between the retrieved and ECMWF wind speeds at Rn values larger than 4 (see speed shift in the contour line), denoting a poor quality of the retrieved solutions. Thus in case of no rain, high Rn is seemingly associated with systematically wrong

winds. This wind speed difference at Rn values larger than 4 becomes significantly larger (2-3 m/s) in the mid plots, while for low Rn (darkest contour), there is no significant difference. This is denoting that although at mid and high winds the wind retrieval is not very much affected by low rain rates, at low winds, the sensitivity to rain is so important that even at low rain rates the quality of the retrievals is poor. This is an expected result, as for low winds, you get lower backscatter than for high winds, and therefore the backscattering from the rain droplets (see section 1) becomes more significant. Comparing the contours from the left and the right plots, there is a positive shift of the left ones with respect to the right ones (indicating a positive bias of the retrieved speeds with respect to ECMWF speeds) as the rain rate increases. This shift is becoming excessively large and unacceptable (more than 10 m/s) for rain rates over 6 mm/hr (bottom plots), denoting again the poor quality of the retrieved solutions.

In the definition of a Rn threshold, we would like to achieve the following goals:

- Maximum low-quality data rejection, including rain;
- Minimum good-quality data rejection.

As said before, the Rn threshold may be dependent on the retrieved wind speed. Figures 7a and 7b (no rain) suggest that the threshold should include and follow the contour lines that are very similar in both plots (showing good quality data). Obviously, this threshold should become constant at a certain wind speed. Otherwise, we would start rejecting more and more data for increasing wind speed until the threshold reaches zero at a certain wind speed from where all higher retrieved speeds

would be rejected. Figures 7a and 7b do not suggest poor quality of all high wind speeds. The constant threshold value has to be a compromise between the amount of high-wind data we want to keep and the amount of "rainy" data we want to reject.

From Figure 4, it is obvious that for higher winds, we should be more critical with the Rn threshold. Therefore, and in order to reject most of the "rainy" data (see Figure 7e), we define a minimum threshold value of 2 for speeds higher than 15 m/s. From Figures 7a and 7b, we define a parabolic threshold with a maximum value of 4 at 5 m/s, which reaches a value of 2 at 15 m/s (see Rn threshold in black solid lines in Figure 7). Therefore, the defined threshold function is

$$
\begin{aligned}
&v \le 15 m/s \Rightarrow y = y_0 + A \cdot (v - v_0)^2 \\
&v > 15 m/s \Rightarrow y = 2 \\
&where, \\
&y_0 = 4, A = -\frac{2}{100}, v_0 = 5,
\end{aligned}
\qquad (3)
$$

v is the retrieved wind speed and *y* the Rn threshold value.

Note that we have tested different thresholds, including: 1) different parabolas with maxima and minima at different Rn/Speed locations; 2) a constant value for all wind speeds; and 3) a constant value for all speeds but with a step (change in value) at different wind speed locations. None of them have given better results than the one defined above, according to our statistics and the two previously mentioned goals.

5 Threshold validation

We test the threshold defined against the ECMWF and SSM/I collocations. The test consists of looking at the Rn of the selected solution of any WVC. If the Rn is lower or equal to the threshold, the WVC is accepted; otherwise, the WVC is rejected. The results for the sweet parts of the swath are shown in Tables 1, 2 and 3.

Table 1 shows the percentage of accepted and rejected WVCs from all the WVCs, segregated by wind speed intervals. Of the data, 5.6% are rejected, and the rejection rate is increasing with wind speed, as expected. As “rainy” cells result in higher retrieved wind speeds (the larger the rain rate the larger the speed bias) and we want to screen out those cells, the amount of rejections should increase with wind speed. However, in order to reject rain, we have defined a threshold that is decreasing with wind speed (up to 15 m/s where it remains constant), and therefore, we might reject an increasing amount of “good” solutions as well.

Table 2 shows the total and the percentage of the accepted and rejected solutions for above and below a RMS-ECMWF threshold of 5 m/s. For the total, accepted and rejected classes, the different average RMS-ECMWF value is also shown. On the one hand, there is a very small portion of rejected data (2.9 %) with RMS values below 5 m/s, indicating that most of the “good” quality solutions have been accepted. On the other hand, there is a significant percentage of rejected data (35.2%) with RMS values over 5 m/s, showing that the Rn threshold is effective in rejecting poor quality data.

The difference between the average RMS of rejected and accepted data is 4 m/s, showing again the effectiveness of the Rn threshold.

Table 3 shows the percentage of the accepted and rejected solutions divided by rain rate intervals. When there is no rain, the percentage of rejections is 3.4%. If we compare this result with the total portion of rejections given in Table 1 (5.6%) we can conclude that in more than 2% of the cases, we are rejecting "rainy" cells. When the rain is over 6 mm/hr, most of the "rainy cells" are rejected (87.3%), denoting a very good behavior of the Rn threshold. When the rain is lower than (or equal to) 6 mm/hr, the percentage of rejections decreases significantly (29.4%) compared to higher rain rates. As said in the previous section, at these rain rates, we are still getting "fair" quality winds (with sufficient wind information), which we may want to keep, but still there is a significant portion of low winds (see discussion of Figure 7) of low quality, which are rejected. In this sense, we achieve a good compromise in the screening of cases in the absence of rain (3.4% of rejections) and in cases of SSM/I rain over 6 mm/hr (87.3% of rejections).

In the WVCs close to nadir, there is not always enough azimuth diversity in the σ^o beams. As said in Section 3, when there is not enough azimuth diversity our QC procedure may not work well. Therefore, we expect a lower skill of the QC in the nadir parts of the swath compared to the sweet swath parts. Nevertheless, the results in the nadir swath (not shown) are reasonably good.

In general, the skill of the QC procedure is good in both regions of the swath, although it is slightly better in the sweet region.

We have also tested a QC based on the MLE of the first rank instead of the selected solution. It shows similar results although the QC based on the selected solution is marginally better. A possible explanation for this small difference is that there is more correlation between a geophysical disturbance and the MLE of the selected solution rather than with the MLE of the first-rank solution. In other words, there is some correlation between the data quality and the number of wind solutions and their corresponding MLE values. Ambiguity removal then picks the most geophysically consistent solution. Therefore, it seems more suitable to use the QC based on the selected solution.

6 Cases

In this section, we show two wind field examples where the QC procedure has been applied. Figures 8 and 9 show triple collocated QuikSCAT-ECMWF-SSM/I data. The arrows in plot a) correspond to the QuikSCAT JPL-selected wind solutions and the grey scales represent the accepted (grey) and the rejected (black) solutions by the Rn threshold (QC). The squares correspond to the collocated SSM/I rain data, where the size of the squares annotates rain rate. The arrows in plot b) correspond to the collocated ECMWF winds. The solid lines divide the different regions of the swath (outer, sweet and nadir).

In Figure 8, there is a tropical case of significant rain (up to 25 mm/hr) over the entire plot, especially in the mid-left and upper-right parts. It is clearly discernible that most of the areas with rain rate above 6 mm/hr (mid-large squares) are rejected by the QC.

At about 12º latitude, there is a “band” of rejections going from the center to the right side of the plot. This area is dividing a modest and high wind speed area (south part) from a low speed one (north part), suggesting the presence of a wind front. The QC performs well, as in the frontal area confused sea state is expected (due to high temporal wind variability), and therefore poor quality wind solutions exist. The wind field in plot b (ECMWF wind field) does not at all reflect the spatial detail seen in plot a, indicating a potential positive impact of assimilation of QuikSCAT winds into the ECMWF model.

Although the low wind speed region shows some erratic flow patterns, most of the wind solutions have been accepted by the QC. This region is mostly located in the nadir part of the swath. As said before, in the nadir regions, there is a lack of azimuth diversity in the σ^o beams. This is going to affect the skill of the wind retrieval, in particular at low wind speeds where the GMF is not so sensitive to wind direction. Here, our QC may not detect anomalous points, since they do not exhibit large Rn. However, we think that improved inversion schemes could produce solution patterns that are more consistent. This will be investigated in the future.

Figure 9 shows a cyclone case. In plot a, the presence of a front is clearly discernible in the middle of the plot, where again a confused sea state and therefore poor quality winds are expected. The WVCs along the front line are rejected by the QC. This is also the case for the center of the low at the bottom of the plot, where there is probably extreme temporal and spatial sea state variability or rain. At the left side of the front, we see a region of significant rain (above 6 mm/hr) which has been successfully detected by the QC. In the outer swath region (right side of the plot), there are very few rejections as expected (see section 4.1). In general, the QC does not

detect much of the poor quality data in the outer regions. However, in this case, the flow looks consistent, and therefore the QC apparently seems to work.

The ECMWF forecast (plot b) does not accurately place the center of the low, and the associated wind front is not so pronounced as in the QuikSCAT plot. This example illustrates again the potential value of assimilating QuikSCAT winds into ECMWF after using our QC.

7 Conclusions and Future Work

Collocations of QuikSCAT data with ECMWF winds and SSM/I rain were used to characterize and validate the QC by a normalized residual (Rn).

The results show a good correlation between the RMS-ECMWF (vector RMS difference between ECMWF and QuikSCAT winds) and the Rn. The data quality, as measured by the inverse of RMS-ECMWF, decreases with increasing Rn, and the decrease rate is increasing with retrieved wind speed; data quality is relatively poor at low Rn values when retrieved speeds are high.

The presence of rain artificially increases the retrieved winds, proportionally to the rain rate. For rain rates over 6 mm/hr, the backscatter measurements contain insufficient wind information but rather rain information, leading to poor quality retrieved winds.

We define a Rn threshold dependent on wind speed, which is optimized to separate the good quality cases from the poor quality ones (including rain) in the inner swath

(WVC number 12 to 65). The results show a RMS-ECMWF difference between accepted and rejected data of 4 m/s, most of the “rainy” cells rejected, and more than 97% of good quality data (low RMS-ECMWF) accepted.

The QC procedure works well in the whole inner swath, although the skill is slightly better in the sweet regions than in the nadir region.

The effectiveness of this QC procedure is furthermore illustrated with a few examples. Cases with meteorologically inconsistent spatial wind patterns are generally removed, while important information on the dynamical structures is kept. Cases that are meteorologically consistent are kept in general.

The QC by Rn is in line with the QC investigated for NSCAT (*Figa and Stoffelen*, 2000) and the QC applied in case of the ERS scatterometer (*Stoffelen*, 1998), which is, in contradiction to NSCAT and SeaWinds, not sensitive to rain. As such, QC rejection for ERS is only activated in case of confused sea state, ice occurrence, etc.

Noise measurements of QuikSCAT, which are ocean view measurements without a signal return, can be used as a radiometer signal with an accuracy of about 10-15 K. Particularly in the Tropics, this signal may be used for rain detection. A problem here is the large footprint of the radiometer of about 75 km (*Jones et al.*, 1999). This concept, together with an MLE-based method (similar to our QC by Rn), which is computing the probability of rain according to the MLE, wind speed and direction, and normalized beam difference (between the s^0 measurements and the GMF) of the first-rank solution, has been proposed by JPL (*Huddleston and Stiles*, 2000). Another recent rain flag based on an Empirical Normalized Objective Function (*Mears et al.*,

2000) has been included in the JPL product. We plan to test in the near future JPL's rain detection flags against our QC.

The wind retrieval skill decreases in the outer and nadir regions of the swath compared to the sweet regions (*Portabella and Stoffelen*, 2000). We plan to work on the inversion problem in these parts of the swath to improve the current wind retrieval skill.

The further idea is to use the Rn as a probability factor for the solutions at a particular WVC, that is, a larger probability will be assigned to the low-Rn wind solutions than to the high-Rn solutions. This probability factor should help to accept or to reject a particular wind solution in the assimilation process, and therefore improve the ambiguity removal.

Acknowledgements

As members of the KNMI QuikSCAT team, Julia Figa and Aart Voorrips have extensively contributed to the work described in this report. We acknowledge the help and collaboration of our colleagues working at ECMWF and KNMI, in particular Mark Leidner (now at AER). The QuikSCAT project team is greatly appreciated for their efforts to provide the QuikSCAT data to us. Furthermore, this work is only possible due to the EUMETSAT grant for a research fellowship post at KNMI. We greatly appreciate the three reviewers who helped to improve the paper.

Appendix : <MLE> Surface Fit

In order to fit a 2D function to the <MLE> surface, we first fit a function for the <MLE> dependence on wind speed at a certain node. Then we verify that the shape of this function is nearly constant over all nodes of the inner swath, and we compute the variation of its mean value over the node number domain.

Figure 10a shows the fit of the <MLE> versus wind speed for node number 25 with a Gaussian + 2^{nd} order polynomial function. The dotted line represents the extrapolated values for wind speeds higher than 20 m/s. It is clearly discernible that the fit is very accurate for that particular node.

Figure 10b shows the averaged <MLE> over all wind speeds and normalized, with the speed dependent function (fit on Figure 10a) versus the node number in the inner swath. The fit is a 2^{nd} order polynomial function (node dependent function). The dotted line shows the extrapolation over the outer swath.

The fact that we have found a two-dimensional function, which fits reasonably well to the computed mean MLE, makes our assumption of considering the shape of the speed-dependent function constant over the node domain valid.

The 2D function, which fits the <MLE> surface, is simply the product of the speed and the node dependent functions. The expression is the following:

$$< MLE >_{fit} = f(v) \cdot f'(n) \qquad \text{(A1)}$$

where *f* is the wind speed dependent function, *f'* is the node dependent function, *v* is the wind speed, and *n* the node number.

$$f(v) = A_0 \cdot e^{-\frac{1}{2}\cdot\left(\frac{v-A_1}{A_2}\right)^2} + A_3 + A_4 \cdot v + A_5 \cdot v^2 \qquad \text{(A2)}$$

$$f'(n) = B_0 + B_1 \cdot n + B_2 \cdot n^2 \quad ; \forall n \in [1,76] \qquad \text{(A3)}$$

where,

$A_0 = 0.78519$; $A_1 = 1.47396$; $A_2 = 2.91577$

$A_3 = 0.31881$; $A_4 = -4.2426 \times 10^{-3}$; $A_5 = 6.9633 \times 10^{-5}$

$B_0 = 1.37840$; $B_1 = -0.02713$; $B_2 = 3.4853 \times 10^{-4}$

References

Boukabara, S.A., Hoffman, R.N., and Grassotti, C., "Atmospheric Compensation and Heavy Rain Detection for SeaWinds Using AMSR," *Atmospheric Environmental Research Inc.*, 131 Hartwell Ave., Lexington, Massachussets (USA), 2000.

Figa, J., and Stoffelen, A., "On the Assimilation of Ku-band Scatterometer Winds for Weather Analysis and Forecasting," *IEEE Trans. on Geoscience and Rem. Sens.* 38 (4) pp. 1893-1902, 2000.

Huddleston, J.N., and Stiles, B.W., "A Multi-dimensional Histogram Technique for Flagging Rain Contamination on QuikSCAT," *Proc. of IEEE International*

Geoscience and Remote Sensing Symposium, Vol. 3, Honolulu (USA), IEEE, pp. 1232-1234, 2000.

Isaksen, L., and Stoffelen, A., "ERS Scatterometer Wind Data Impact on ECMWF's Tropical Cyclone Forecasts," *IEEE Trans. on Geoscience and Rem. Sens.* 38 (4) pp. 1885-1892, 2000.

Jones, L., Mladen, S., Park, J., and Mehershadi, R., "A Quality Control Rain Flag using QuikSCAT Radiometric Observations," *Proc. of QuikSCAT Cal/Val Workshop*, Pasadena/Arcadia (USA), JPL, 1999.

JPL, "QuikSCAT Science Data Product User's Manual," version 1.1, *JPL* D-18053, pp. 84, 1999.

Leidner, M., Hoffman, R., and Augenbaum, J., "SeaWinds Scatterometer Real-Time BUFR Geophysical Data Product," version 2.2.0, *NOAA/NESDIS*, pp. 45, 2000.

Mears, C., Smith, D., and Wentz F., "Detecting Rain with QuikSCAT," *Proc. of International Geoscience and Remote Sensing Symposium*, Vol. 3, Honolulu (USA), IEEE, pp. 1235-1237, 2000.

Portabella, M., and Stoffelen, A., "EUMETSAT QuikSCAT Fellowship : Progress Report," *KNMI* [available online at http://www.knmi.nl/scatterometer/quikscat/], 2000.

Rohn, M., Kelly, G., and Saunders, R., "Experiments with Atmospheric Motion Vectors at ECMWF," *Proc. of Fourth International Winds Workshop*, EUM P24, ISSN 1023-0416, Saanenmoser (Switzerland), EUMETSAT, pp. 139-146, 1998.

Spencer, M.W., Wu, C., and Long, D.G., "Tradeoffs in the Design of a Spaceborn Scanning Pencil Beam Scatterometer: Apllication to SeaWinds," *IEEE Trans. Geosci. Remote Sensing*, Vol. 35, No. 1, pp. 115-126, 1997.

Stoffelen, A., “Scatterometry,” *PhD thesis* ISBN 90-393-1708-9, University of Utrecht, pp. 195, 1998.

Stoffelen, A., and Anderson D., “Scatterometer Data Interpretation: Measurement Space and Inversion,” *J. Atm. And Ocean Techn.*, 14(6), pp. 1298-1313, 1997.

Stoffelen, A., Van Beukering, P., “Implementation of Improved ERS Scatterometer Data Processing and its Impact on HIRLAM Short Range Weather Forecasts,” *Beleidscomissie Remote Sensing Report*, contract NRSP-2/97-06, pp. 75, 1997.

Stoffelen, A., de Vries, J., Voorrips, A., “Towards the Real-Time Use of QuikSCAT Winds,” *Beleidscomissie Remote Sensing Report*, contract USP-2/00-26, pp. 77, 2001.

Van De Hulst, H.C., “Light Scattering by small particles,” *John Wiley and Sons*, New York, pp. 428, 1957.

Wentz, F., Smith, D., and Mears, C., “Rain and the QuikSCAT Winds,” *Proc. of QuikSCAT Cal/Val Workshop*, Pasadena/Arcadia (USA), JPL, 1999.

Figure Captions

Figure 1 Mean JPL-selected MLE with respect to wind direction (dotted line) and wind direction distribution of JPL-selected winds (dashed line) and NCEP winds (solid line) for WVC number 16. The direction binning is 10°.

Figure 2 Contour plots of the (a) mean JPL-selected MLE and (b) the filtered mean JPL-selected MLE versus JPL-selected wind speed and node number. The speed binning is 1 m/s and the node binning is 1.

Figure 3 (a) Filtered mean JPL-selected MLE and (b) expected MLE versus the JPL-selected wind speed and node number. The speed binning is 1 m/s and the node binning is 1.

Figure 4 Two-dimensional histograms of RMS-ECMWF versus Rn, (a) for all data, (b) JPL-selected speeds under 10 m/s, (c) speeds between 10 and 15 m/s, and (d) speeds over 15 m/s (plot d). The contouring is in logarithmic scale (two steps corresponding to a factor of 10 in number density) filled from white (unpopulated areas) to black (most populated areas).

Figure 5 (a) Mean JPL-selected wind speed, and (b) Mean ECMWF wind speed versus rain rate at intervals of 3 mm/hr (except for the rain-free mean speed value, included at 0 mm/hr)

Figure 6 Two-dimensional histograms of (a) RMS-ECMWF versus JPL-selected wind speed for rain-free, (b) for rain rates from 0 to 6 mm/hr, and (c) for rain rates above 6 mm/hr (plot c). The contouring is in logarithmic scale (two steps corresponding to a

factor of 10 in number density) filled from white (unpopulated areas) to black (most populated areas).

Figure 7 Two-dimensional histograms of Rn versus JPL-selected wind speed (left plots) and versus ECMWF speeds (right plots), (a) and (b) for rain-free data, (c) and (d) for rain rate from 0 to 6 mm/hr, and (e) and (f) for rain rate above 6 mm/hr. The contouring is in logarithmic scale (two steps corresponding to a factor of 10 in number density) filled from white (unpopulated areas) to black (most populated areas).

Figure 8 Collocated QuikSCAT-ECMWF-SSM/I data. (a) QuikSCAT wind arrows (JPL-selected winds), where grey corresponds to accepted WVCs and black to rejected WVCs. The size of the squares represents the different rain rates from 0 mm/hr (no square) to 25 mm/hr (the largest ones). (b) The collocated ECMWF winds. The solid lines separate different regions of the swath. In this case, the left side of the plot corresponds to the left sweet region, the middle to the nadir region and the right side to the right sweet region. The acquisition date was September 2 1999 at 14 hours UTC.

Figure 9 Same as Figure 8 but for August 28 1999 at 5 hours UTC and different location. The solid lines separate the nadir (left side), the left sweet (middle) and the outer-left (right side) regions.

Figure 10 (a) <MLE> versus JPL-selected wind speed for node number 25 (stars), where the solid line shows the function fit and the dotted line the extrapolation for wind speeds higher than 20 m/s; (b) averaged <MLE> over all wind speeds and normalized with the speed dependent function (fit on plot a) versus the node number

in the inner swath (stars), where the solid line represents the function fit and the dashed line the extrapolation for the outer swath.

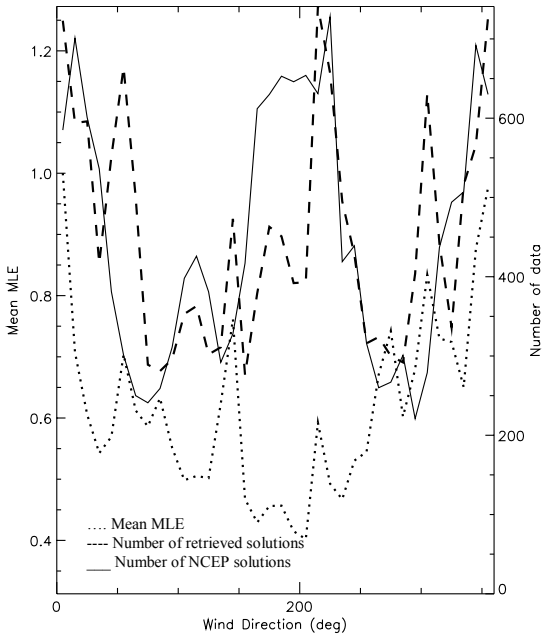
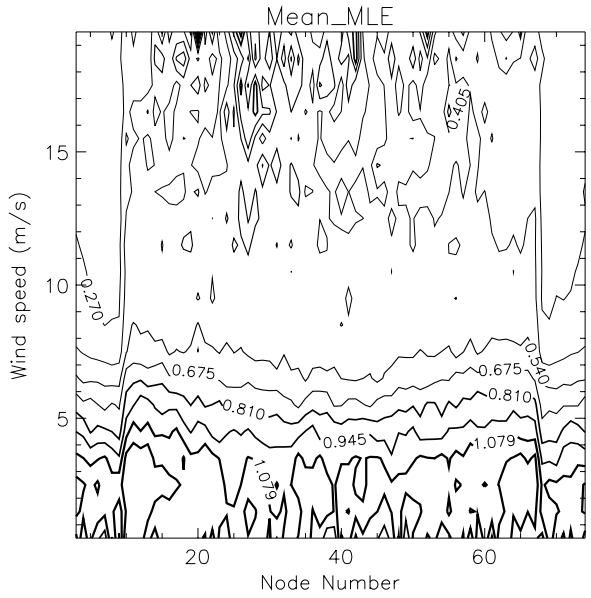


Figure 1 Mean JPL-selected MLE versus JPL-selected wind direction (dotted line) and wind direction distribution of JPL-selected winds (dashed line) and NCEP winds (solid line) for WVC number 16. The direction binning is 10°.

a)



b)

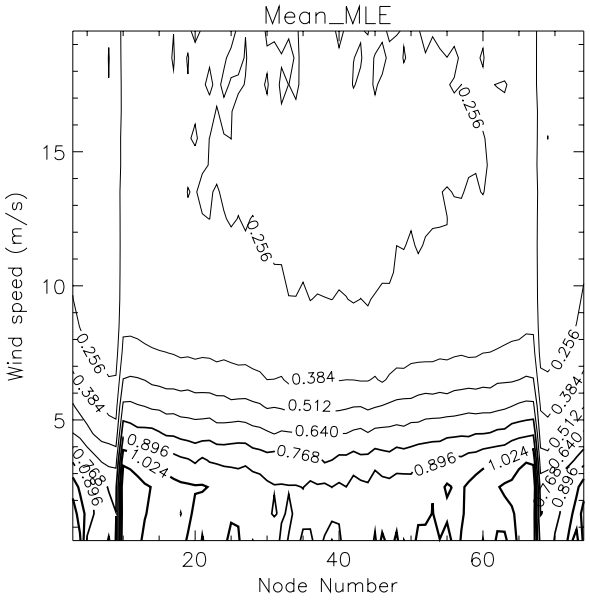
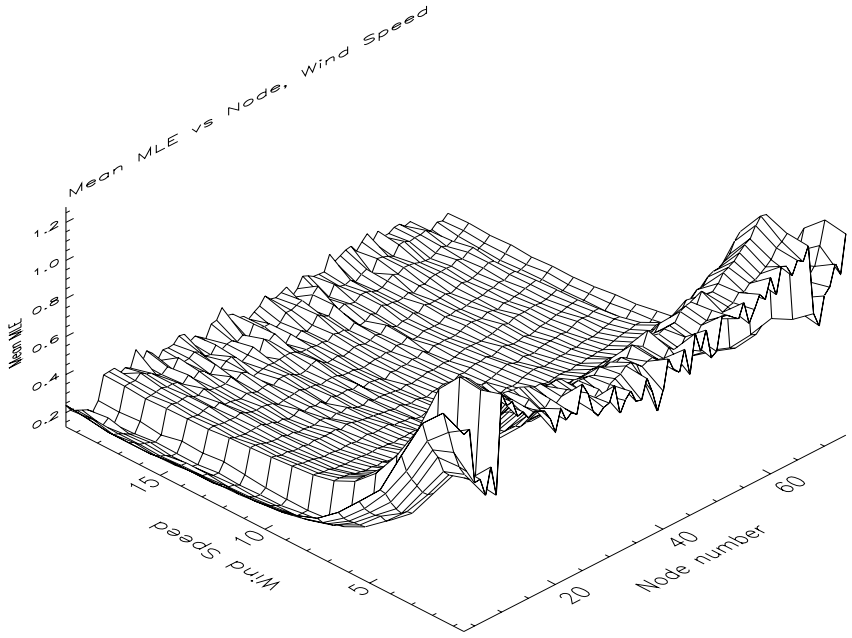


Figure 2 Contour plots of (a) the mean JPL-selected MLE and (b) the filtered mean JPL-selected MLE versus JPL-selected wind speed and node number. The speed binning is 1 m/s and the node binning is 1.

a)



b)

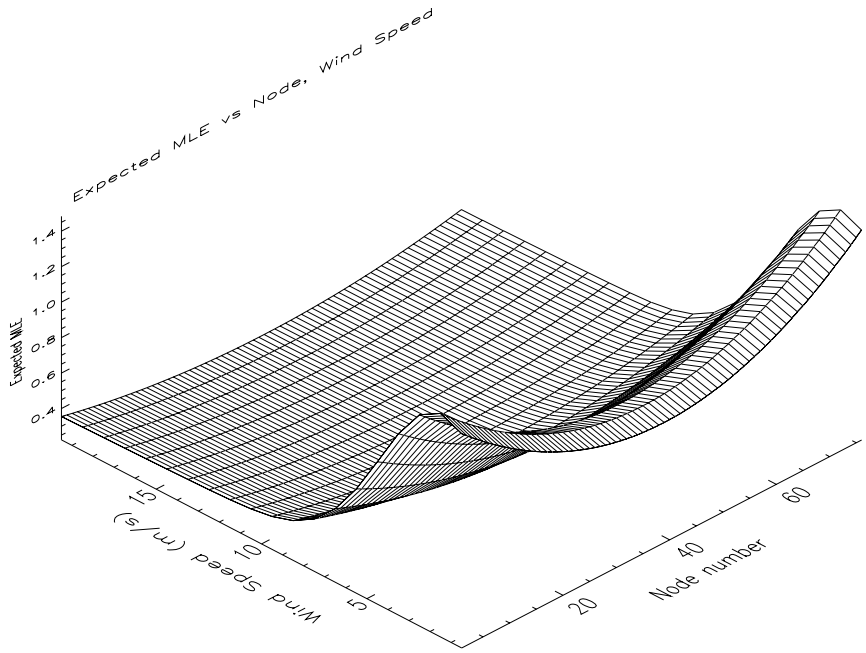


Figure 3 (a) Filtered mean JPL-selected MLE and (b) expected MLE (plot b) versus JPL-selected wind speed and node number. The speed binning is 1 m/s and the node binning is 1.

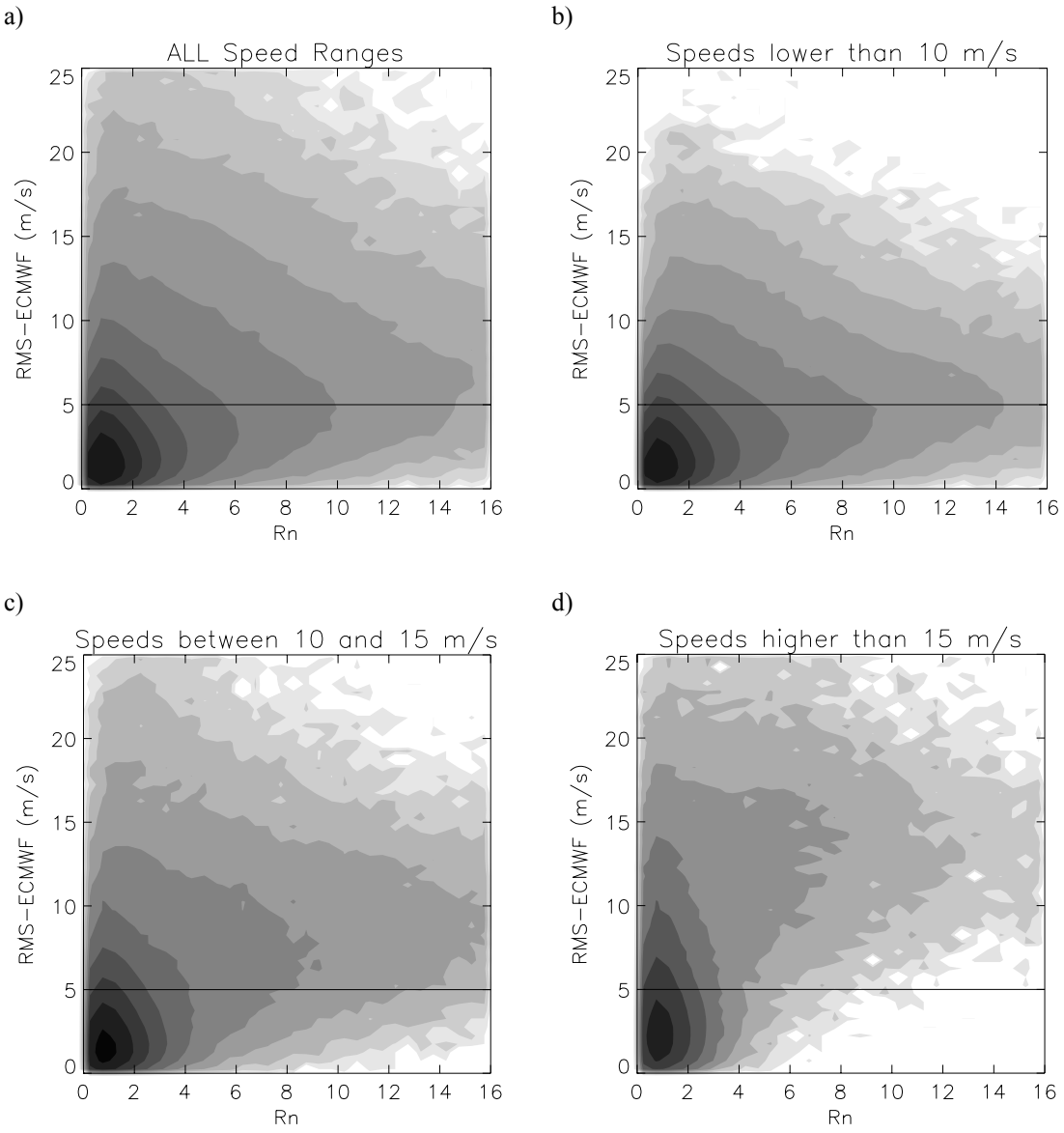


Figure 4 Two-dimensional histograms of RMS-ECMWF versus Rn, (a) for all data, (b) JPL-selected speeds under 10 m/s, (c) speeds between 10 and 15 m/s, and (d) speeds over 15 m/s. The contouring is in logarithmic scale (two steps corresponding to a factor of 10 in number density) filled from white (unpopulated areas) to black (most populated areas).

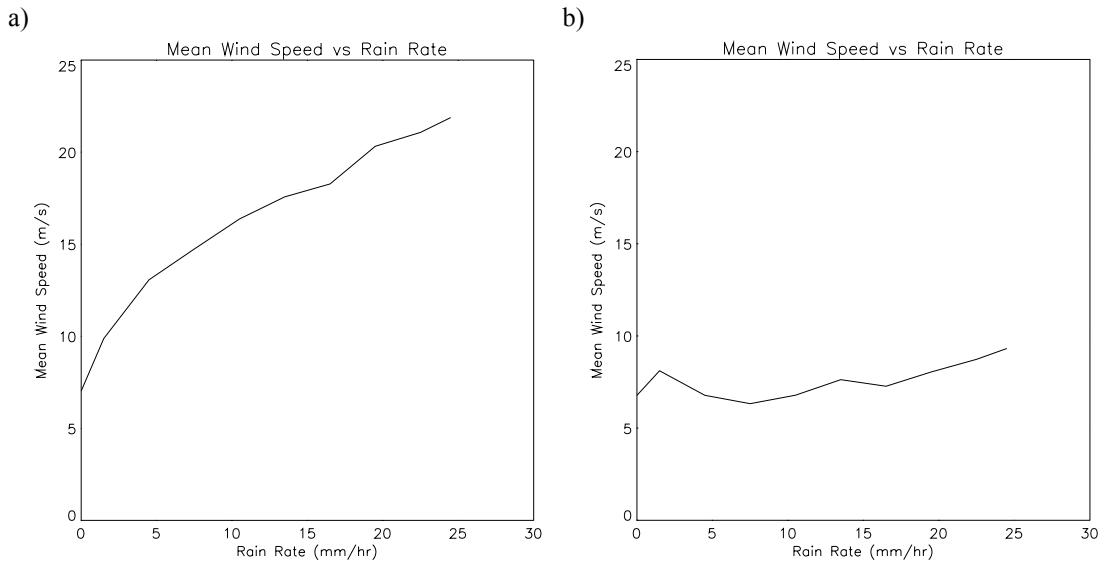


Figure 5 (a) Mean JPL-selected wind speed and (b) Mean ECMWF wind speed versus rain rate at intervals of 3 mm/hr (except for the rain-free mean speed value, included at 0 mm/hr).

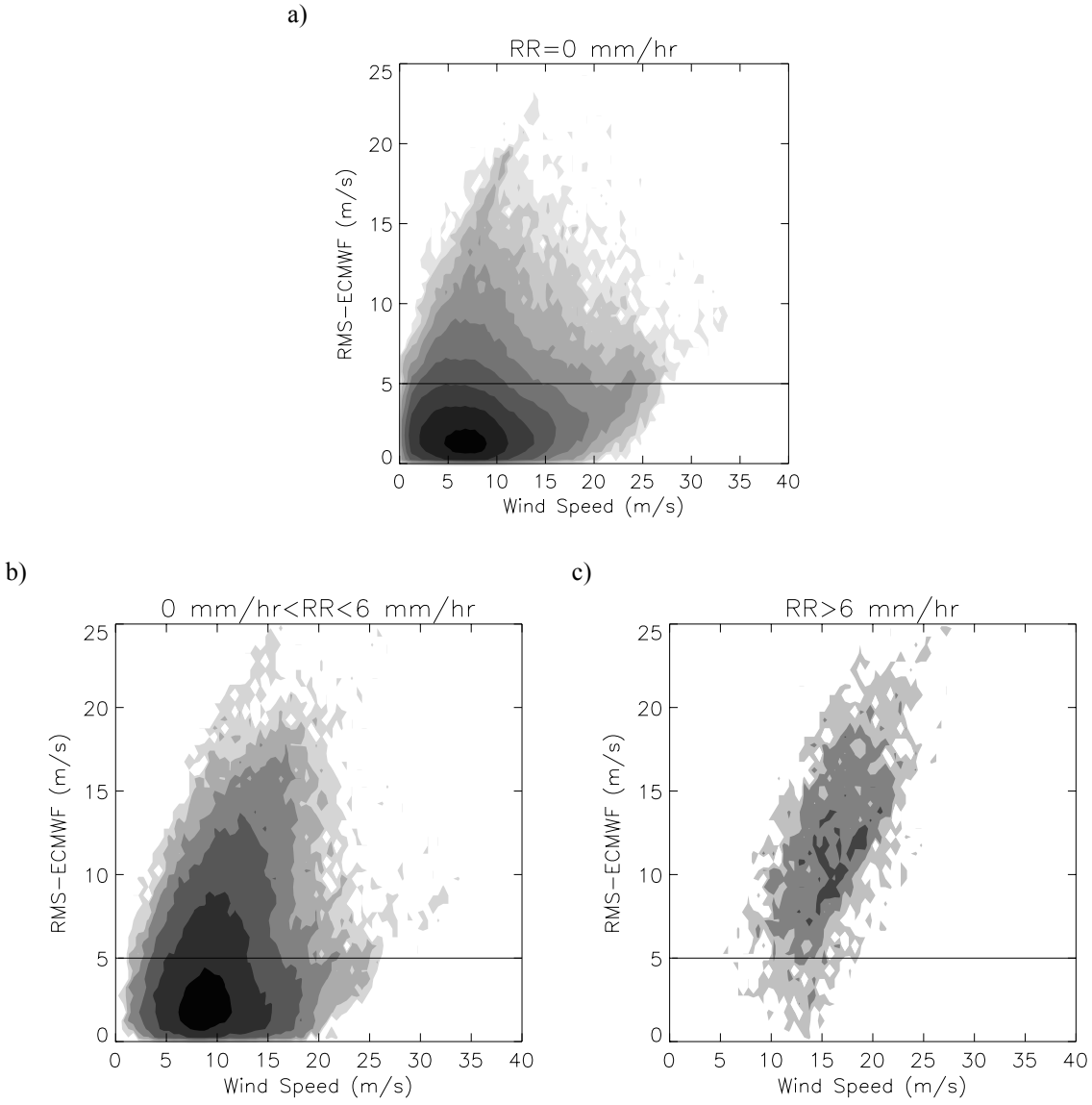


Figure 6 Two-dimensional histograms of RMS-ECMWF versus JPL-selected wind speed (a) for rain-free, (b) for rain rates from 0 to 6 mm/hr, and (c) for rain rates above 6 mm/hr. The contouring is in logarithmic scale (two steps corresponding to a factor of 10 in number density) filled from white (unpopulated areas) to black (most populated areas).

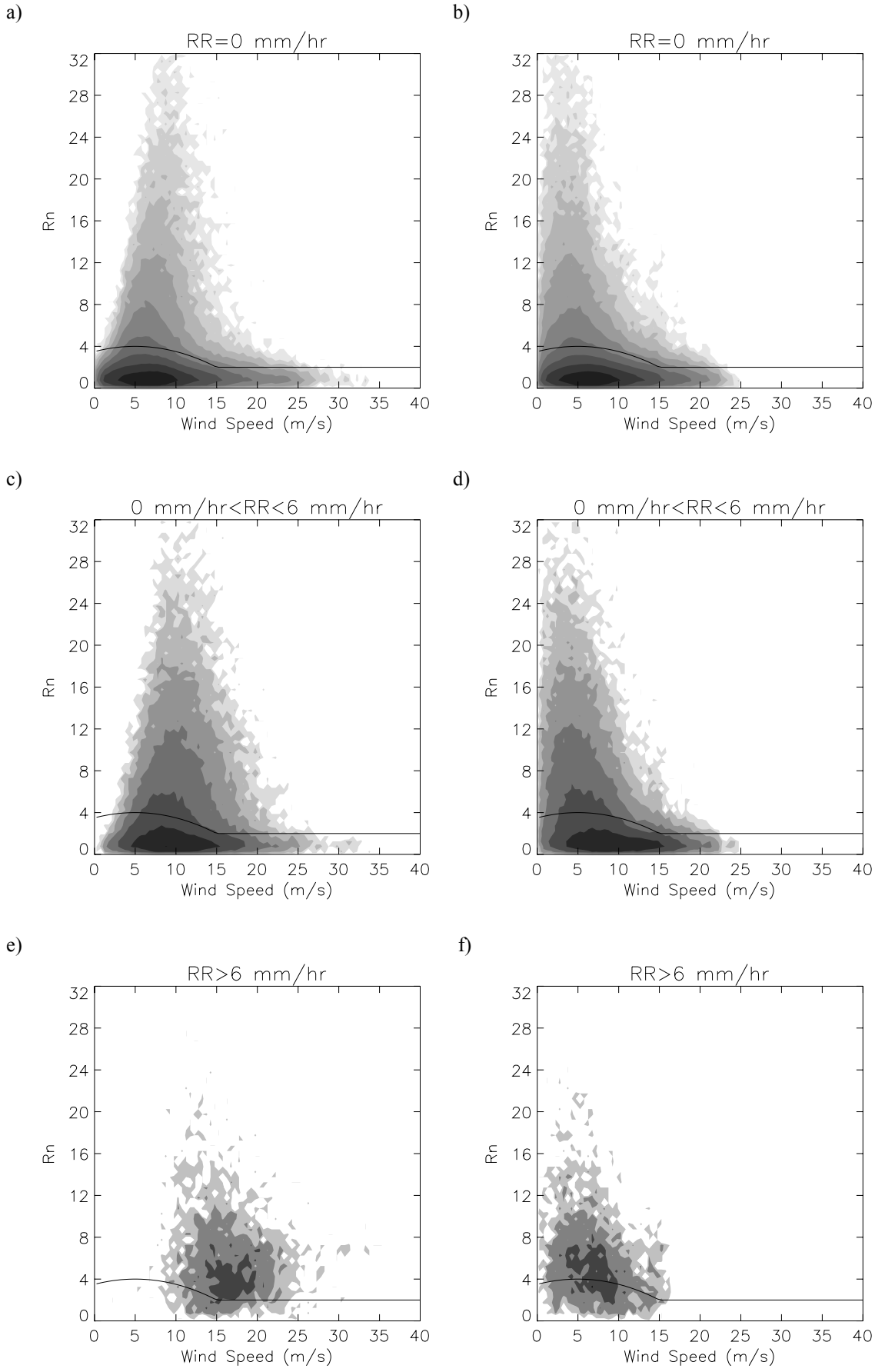


Figure 7 Two-dimensional histograms of Rn versus JPL-selected wind speed (left plots) and versus ECMWF speeds (right plots), (a) and (b) for rain-free data, (c) and (d) for rain rate from 0 to 6 mm/hr, and (e) and (f) for rain rate above 6 mm/hr. The contouring is in logarithmic scale (two steps corresponding to a factor of 10 in number density) filled from white (unpopulated areas) to black (most populated areas).

TROPICAL CASE : 02/09/99 1400 UTC

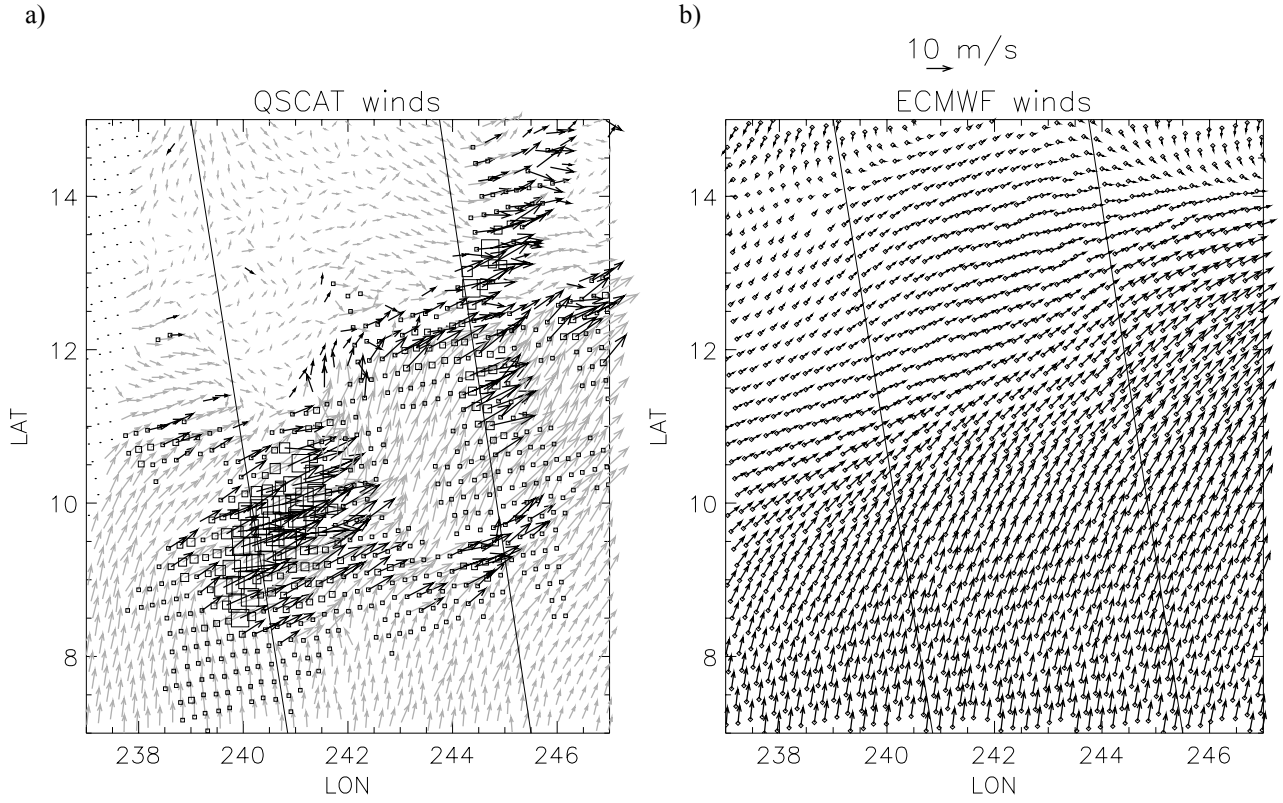


Figure 8 Collocated QuikSCAT-ECMWF-SSM/I data. (a) QuikSCAT wind arrows (JPL-selected winds), where grey corresponds to accepted WVCs and black to rejected WVCs. The size of the squares represents the different rain rates from 0 mm/hr (no square) to 25 mm/hr (the largest ones). (b) the collocated ECMWF winds. The solid lines separate different regions of the swath. In this case, the left side of the plot corresponds to the left sweet region, the middle to the nadir region and the right side to the right sweet region. The acquisition date was September 2 1999 at 14 hours UTC.

CYCLONE CASE : 28/08/99 0500 UTC

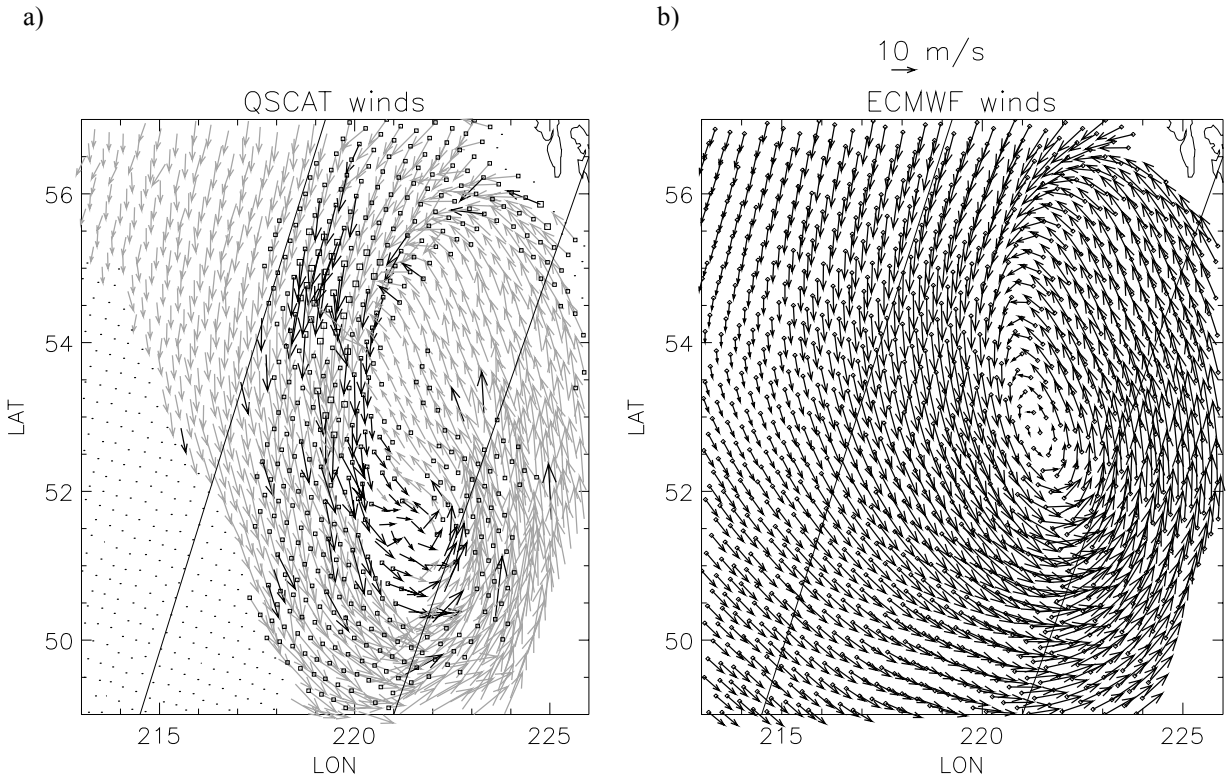


Figure 9 Same as Figure 7 but for August 28 1999 at 5 hours UTC and different location. The solid lines separate the nadir (left side), the left sweet (middle) and the outer-left (right side) regions.

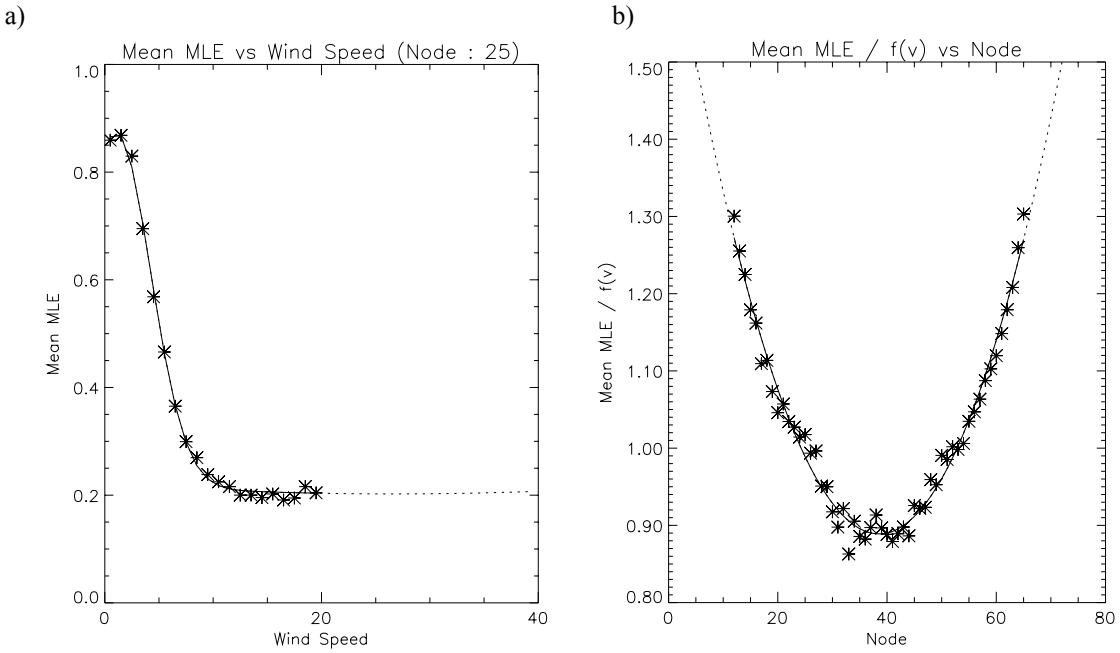


Figure 10 (a) <MLE> versus JPL-selected wind speed for node number 25 (stars), where the solid line shows the function fit and the dotted line the extrapolation for wind speeds higher than 20 m/s; (b) averaged <MLE> over all wind speeds and normalized with the speed dependent function (fit on Figure 18a) versus the node number in the inner swath (stars), where the solid line represents the function fit and the dashed line the extrapolation for the outer swath.

TABLE 1. Accepted and rejected WVCs from all the WVCs.

	Total	V < 10	10≤V≤15	V > 15
Num. Points (n/a)	4826841	3796408	859747	170686
Accepted (%)	94.4	95.8	91	81
Rejected (%)	5.6	4.2	9	19

TABLE 2. Relative quality of accepted and rejected solutions.

	RMS < 5	RMS > 5	Avg. RMS (m/s)
Total (n/a)	4429905	396970	2.46
Accepted (%)	97.1	64.8	2.24
Rejected (%)	2.9	35.2	6.24

TABLE 3. Accepted and rejected solutions divided by rain rate intervals.

	RR = 0	0 < RR ≤ 6	RR > 6
Num. Points (n/a)	1027124	88311	3664
Accepted (%)	96.6	70.6	12.7
Rejected (%)	3.4	29.4	87.3

Note : RMS is referred to as the vector RMS difference between JPL-retrieved winds and ECMWF winds in m/s; V is the JPL-retrieved wind speed in m/s; and RR is the SSM/I rain rate in mm/hr.