EUMETSAT QUIKSCAT FELLOWSHIP : PROGRESS REPORT

by

M. Portabella and A. Stoffelen

December 2000

Preface

This report contains an overview of the work performed in the first 19 months of the EUMETSAT QuikSCAT fellowship, carried out at the Royal Netherlands Meteorological Institute (KNMI) in De Bilt (The Netherlands) within the research group of the Satellite Data Division of the Observations and Modeling Department.

Contents

1 Introduction

The SeaWinds on QuikSCAT mission (from NASA) is a “quick recovery” mission to fill the gap created by the loss of data from the NASA Scatterometer (NSCAT), when the ADEOS-1 satellite lost power in June 1997. QuikSCAT was launched from Vandenberg Air Force Base (USA) in June 19, 1999. A 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 conically scanning pencil-beam scatterometer, which in comparison with the NSCAT fan-beam scatterometer has the following advantages: higher signal-to-noise ratio, smaller in size, and superior coverage. On the other hand, the new measurement geometry provides a variable mix of polarizations and azimuth looks that require renewed data characterisation efforts as employed in this fellowship.

The SeaWinds instrument uses a rotating 1-meter dish antenna with two spot beams, an H-pol beam and a V-pol beam at incidence angles of 46º and 54º respectively, that 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 to 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) 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: 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 data we refer to [[*Spencer et al.* (1994), *JPL* (1999), *Leidner et al.* (2000)]].

In this report we describe our work on QuikSCAT product validation in chapter 2, inversion of the backscatter data to winds in chapter 3, and in chapters 4 and 5 a new procedure to quality control (QC) SeaWinds data in HDF and BUFR formats respectively, in particular to screen out rain-contaminated points. Our QC method is based on a methodology that was used to screen ERS (*Stoffelen and Anderson*, 1997) and NSCAT backscatter measurements (*Figa and Stoffelen*, 2000). The methodology checks whether the consistency of the backscatter measurements at a particular WVC is compatible with the consistency as predicted by the Geophysical Model Function (GMF). Incompatible measurement sets are assumed to be anomalous and result in invalid wind retrieval. To validate the QC and rain screening procedure, we collocated

QuiksCAT data with ECMWF winds and with a SSM/I rain poduct. We seeked for a statistical scheme resulting in a proper rejection threshold and verified the settings of the scheme in many cases of which we show a few here.

2 Product Validation

During the pre-launch period, a full validation of the different simulated QuikSCAT sample data products was performed. In particular, the work focused on the Level 2A and 2B HDF (Hierarchical Data Format) Science product and on the near-real time BUFR (Binary Universal Format Representation)product. The Level 2A contains the radar backscatter or sigma0 (σ^o) related information and the level 2B the surface wind related information, while the BUFR product is containing both σ^o and wind information.

The main difference between the HDF and the BUFR product is related to the spatial resolution of σ^o. In each WVC, the σ^o of a particular beam (fore-inner, fore-outer, aft-inner, aft-outer) in the BUFR product is an average of all σ^os of that particular beam in the HDF product which fall in the same WVC.

The SeaWinds σ^os can be either “eggs” or slices. In a particular WVC, an “egg” σ^o is the radar backscatter from the whole pulse or footprint whose centre falls in that WVC. The “egg” can be subdivided in individual range-resolution elements or slices; the slices of a particular “egg” whose centre fall in the same WVC are weight-averaged (the weighting factor is directly dependent on the noise of each slice “measurement”) to become a composite σ^o. The antenna footprint or “egg” is an ellipse approximately 25-km in azimuth by 37-km in the look (or range) direction. The slices are 25-km in azimuth by a variable range resolution of approximately 2 to 10 km (the nominal width is 6 km).

The HDF real data are given (up to now) in “egg” resolution. Therefore, although the size of the WVC is 25 km, the actual resolution of the winds retrieved from the “egg” σ^os is approximately 40 km. Composites enhance the wind resolution mainly in range direction down to 25 km, and have little effect on azimuth resolution.

Real BUFR data have been released to the Science Working Team (SWT) recently. The BUFR real σ^os are finally weighted average of composites.

During the pre-launch period, two different versions of sample data were released by JPL and NOAA in HDF and BUFR format respectively. The “old” version was containing “egg” σ^o and the “new” one was based on slice σ^o. The most significant results of the QuikSCAT product validation on these sample files are explained in the following sections.

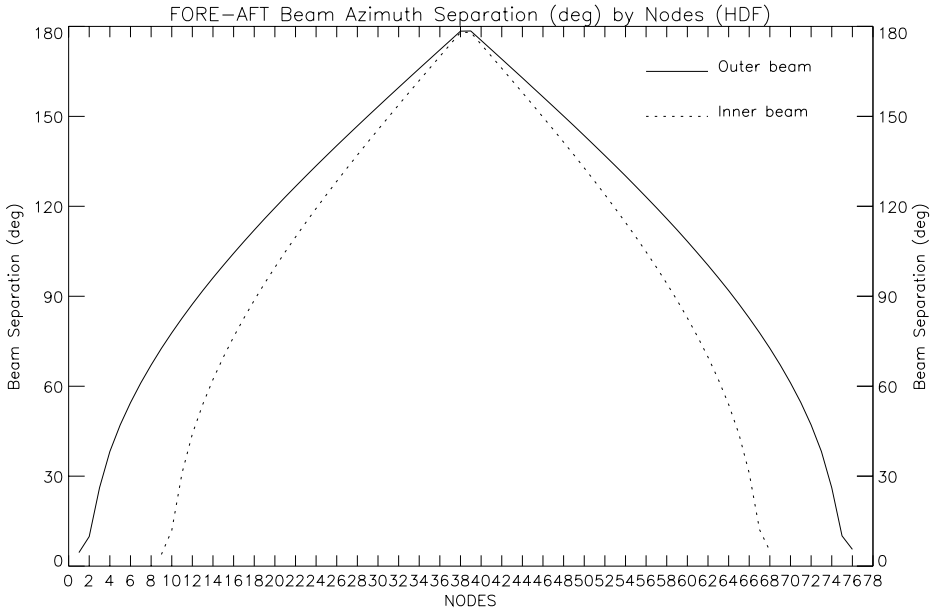
2.1 Azimuth averaging

In order to check inconsistencies in the way the azimuth information is computed across the swath for any beam, we look at the relative azimuth distribution for all node numbers (or WVC numbers). Figure 1 shows the mean azimuth separation between fore and aft beams by node for a few revolutions of HDF data (plot a) and the Standard Deviation (STD) of the azimuth separation for each node (plot b). The plots show a very consistent azimuth separation distribution going from small separation at the edges of both swaths, i.e., the outer (nodes 1-2 and 75-76) and the inner swath (nodes 9-10 and 67-68), and increasing gradually to a separation of almost 180º at nadir (nodes 38 and 39). The STD plot also shows consistent results with small values all across the swath, except at the edges of both the inner and the outer swaths, where the STD increases as the number of measured σ^os significantly decreases. (These nodes are set to accommodate occasional measurements that lie outside of the pre-defined 1800-km swath for the outer beam and 1400-km swath for the inner beam).

If we look at the same STD plot (Figure 2) but for BUFR (the mean azimuth separation plot is not shown because it is very similar to Figure 1a), we see a large peak around node 46 for the inner swath and node 48 for the outer swath. This peak, which does not appear in the HDF product, clearly shows some problem in the way the σ^o azimuths are averaged from HDF to BUFR. After taking a closer look to the azimuths at those particular nodes, we discovered that there was an error while averaging the azimuths close to 0º. In those particular areas, the azimuths can be either close to 360º or close to 0º. The software was not taking into account this periodicity and just averaged azimuths, leading to averaged azimuth values very different from 360º or 0º, the problem depending on the ratio of the number of azimuth values close to 360º and the number of azimuths close to 0º.

The problem was reported to JPL and corrected straightaway.

a)



b)

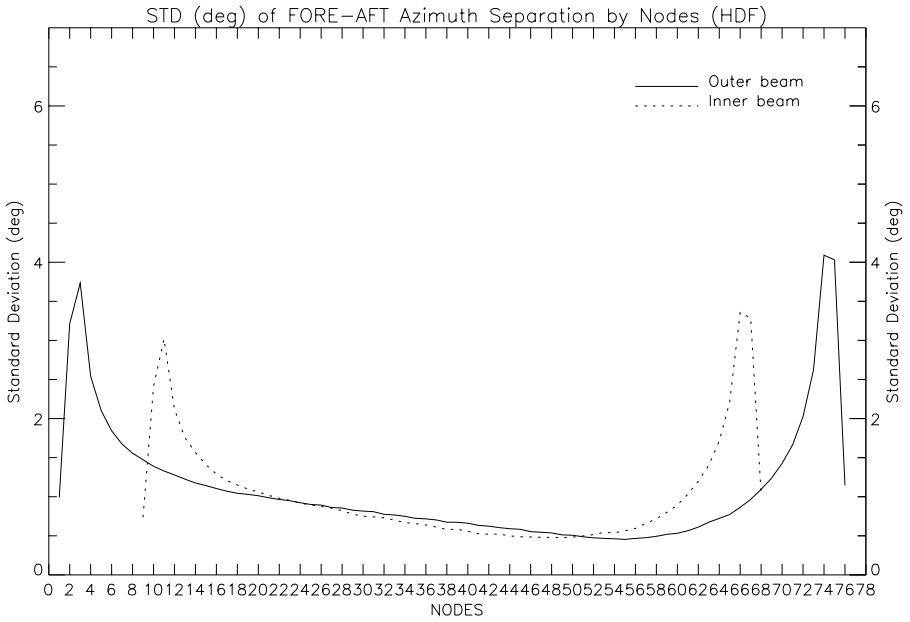


Figure 1 a) Mean fore-aft azimuth separation by node for one revolution of HDF data; the outer beam separation is in solid line and the inner beam separation in dotted line; b) Standard Deviation of the azimuth separation.

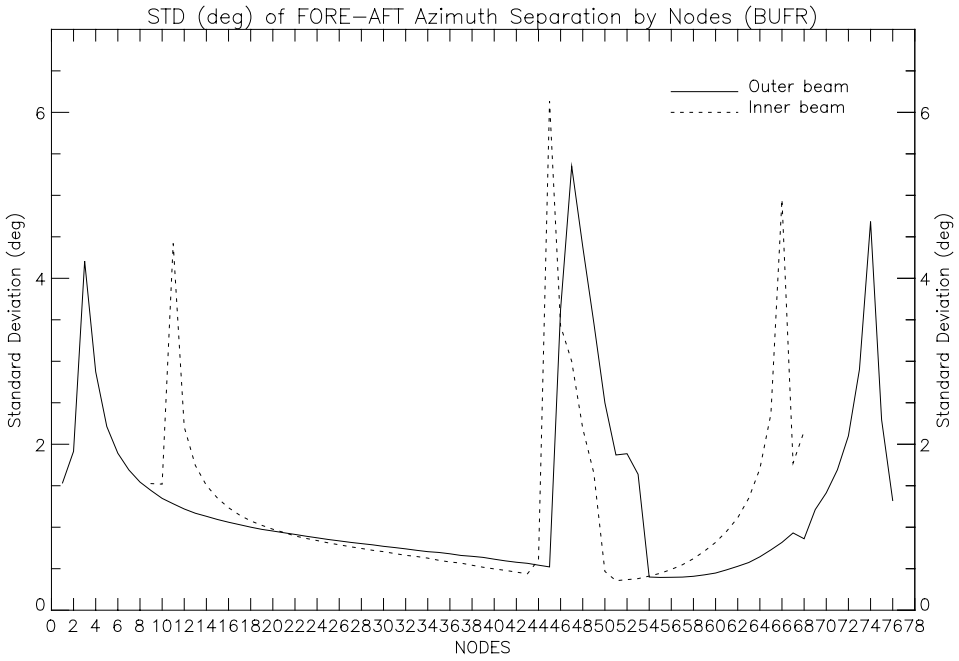


Figure 2 same as Figure 1b but for BUFR data

2.2 Wind Direction Convention

The σ^o sample data were generated applying the NSCAT-2 model function to NWP winds (from NCEP 1º spatial resolution and 12-hour forecast, which are included in the product), including additional realistic measurement noise (*Kp*) generated by a random selection from a noise distribution with width $\sqrt{Kp}$. Then, these simulated measurements of σ^o (σ^o_m) were inverted (using JPL's objective function as described in the "QuikSCAT Science Data Product User's Manual") into winds. In order to check for any inconsistency in the simulation process we simulate σ^o (σ^o_{sim}) from the JPL retrieved winds, applying the NSCAT-2 model function. Figure 3 shows the mean σ^o_m and the mean σ^o_{sim} by node for the fore-inner beam (the results for the other beams are very similar) in one orbit (or revolution) of data. The results are very consistent although there is a systematic bias in the σ^o_{sim}. The reason is that the σ^o_m were corrected for atmospheric attenuation prior to retrieving the JPL winds, while the σ^o_{sim} were computed without "uncorrecting" for atmospheric attenuation. Therefore, σ^o_{sim} is systematically higher than σ^o_m.

Figure 4a is showing the same plot as Figure 3 but for the new version of sample data. The results are very different compared to the old version of sample data, showing no systematic bias at all. Figure 4b shows the same as Figure 4a but changing the JPL wind direction by 180º prior to compute the σ^o_{sim}. The systematic bias of σ^o_{sim} is now clearly discernible again, suggesting that JPL did change the wind direction

conventions in the product but forgot to report it in the User's Guide. This problem was confirmed by JPL and is corrected already.

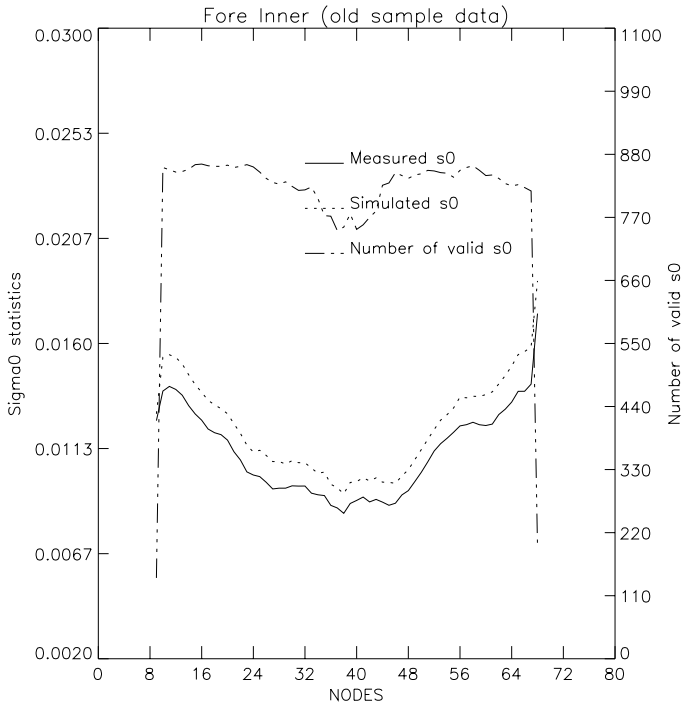


Figure 3 Mean σ°_{m} (solid) and σ°_{sim} (dotted) versus node number for the old sample data. The number of σ° used is in dash-dotted line.

The final wind direction conventions for both products is the following:

- HDF: both NWP and retrieved winds are in oceanographic or flow vector convention (0º meaning wind blowing towards North with positive angles increasing in the clockwise direction).

- BUFR: both NWP and retrieved winds are in meteorological convention or flags (0º meaning wind blowing from North with positive angles increasing in the clockwise direction).

2.3 Flags

All the flags present in the BUFR and HDF product were checked for inconsistencies. We found no meaningful information in the **wvc_quality_flag** (this flag indicates the quality of wind retrieval within a given WVC) of the BUFR product. After reporting the problem to JPL, it turned out that this flag was containing less information and in a different way than the same flag of the HDF product. In order to avoid confusion, they set this flag in BUFR in the same way as in HDF.

a) b)

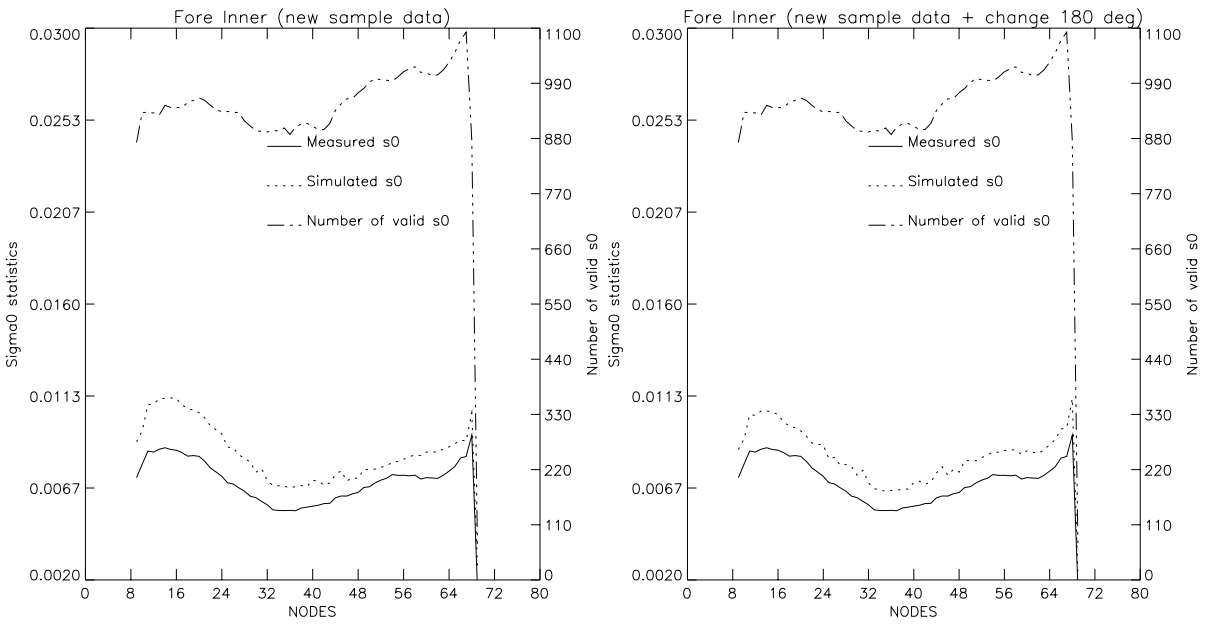


Figure 4 a) same as Figure 3 but for new sample data; b) same as plot a) but changing the wind direction 180 ° prior to computing σ^{o}_{sim}

2.4 Other Problems

Every variable contained in the data product is defined between a range of values. The best way to look for bugs in the processing software is to look for out of range values and distribution of values in the variables. Here there is a summary of the most significant bugs we found:

a) A few null Kp_β values per orbit were found in both HDF and BUFR, as well as some null Kp_γ values in HDF.

b) The Kp_α values of the new sample data were all out of range and the mean Kp_α value was twice as high as the mean Kp_α value of the old sample data.

[Note: Kp_α, Kp_β, and Kp_γ are the coefficients used to compute the measurement noise (Kp)]

c) A few values per orbit of the **wind_speed_err** (estimated error in the retrieved speed) were above 50 m/s; in the same way, some values of **wind_dir_err** (estimated error in the retrieved wind direction) were above 180º.

d) σ^{o}<–70 dB were also found.

All these problems were reported to JPL. We have performed this kind of validation on the real HDF data and none of these problems are present. The σ^{o} range has been extended to [20,–300] dB in the real data. Note that there is no physical lower limit of σ^{o}.

The same validation has been recently performed on the real BUFR data. There still are some out of range values in **wind_speed_err** and **wind_dir_err** variables as well as a few null Kp_β values per orbit. As before, we have already reported these problems to JPL.

3 Inversion

The inversion software (mimicing JPL's inversion and based on ERS scatterometer processing software) that Julia Figa (NSCAT EUMETSAT fellowship 97-99) developed for NSCAT has been adapted for QuikSCAT. In order to test the quality of this inversion software, the σ^o_m (originally generated from the NWP winds) have been processed, using our inversion software, and compared to the JPL winds (retrieved winds in the product).

Figure 5 shows the two-dimensional histogram of KNMI winds (our inverted winds using the same atmospheric attenuation correction used by JPL prior to the inversion) versus the JPL winds. Plots a) and b) show the comparison of the first rank solution inverted wind speeds and directions, respectively. In these plots, the good agreement between the KNMI and the NWP winds is clearly discernible showing a good performance of the KNMI inversion. The high value of the wind direction STD (49.18º) is due to a small amount of KNMI solutions which are approximately 180º apart from the JPL solution and which increase the STD. This is an expected result as both implementations are not exactly the same and sometimes the wind direction solutions are ranked in a slightly different order (typically 1st rank and 2nd rank solutions are permuted in KNMI with respect to JPL or vice versa). The bottom plots show the same as the upper plots but for the KNMI solution closest to the JPL 1st rank solution. We note that in this case no more solutions are 180º apart, leading to a significant decrease of the STD down to an acceptable value of 5.54º.

In contrast with NSCAT, QuikSCAT has an antenna geometry which is dependent on node number or cross-track location, due to its circular scans on the ocean. The skill of the wind retrieval algorithm depends very much on the number of measurements and their polarization and azimuth diversity, where "azimuth diversity" is defined as the spread of the azimuth looks among the measurements in the WVC. As we see in Figure 1a, the nadir region (area close to and centered in nodes 38-39) has fore and aft looks of both beams (H-pol and V-pol) nearly 180º apart. At the edges of the swath (areas close to node 1 and 76) the outer V-pol beam fore and aft looks are nearly in the same direction and no inner beam information is available. In both areas, the skill of the wind retrieval algorithm is expected to decrease with respect to the rest of the swath (called the sweet or NSCAT-like parts of the swath) where there are four measurements (fore-inner, fore-outer, aft-inner and aft-outer) with enough azimuth diversity.

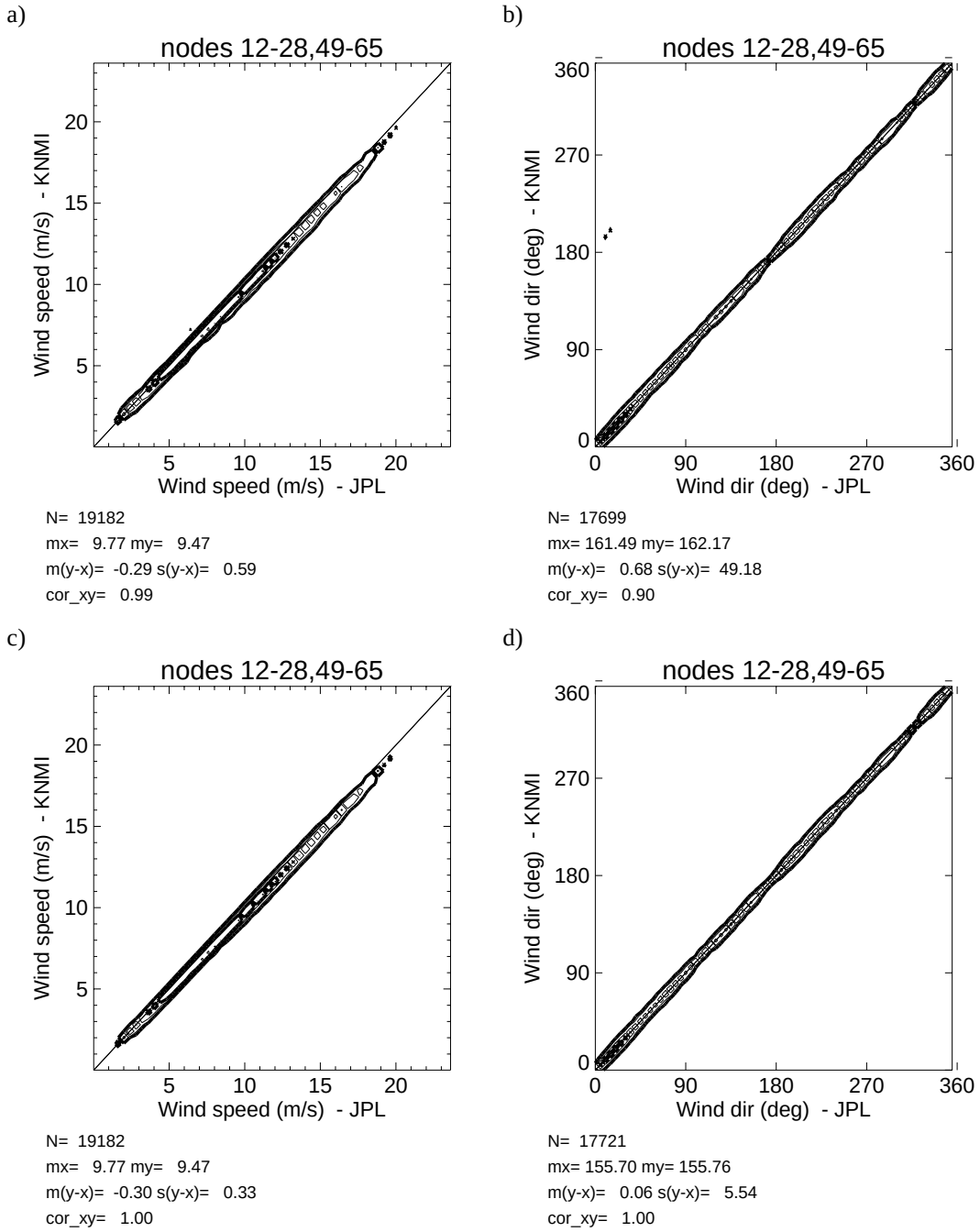


Figure 5 Two-dimensional histograms of KNMI wind solutions versus 1st rank JPL wind solutions in the NSCAT-like parts of the swath. Left plots are for wind speeds and right plots for wind directions. The y-axis for the upper plots represents the 1st rank KNMI solutions while for the bottom plots represents the KNMI solution closest to the 1st rank JPL solution. The width of the bins is 0.5 m/s in speed and 5° in direction. N is the number of points; mx is the mean value of the x-axis; my is the mean value of the y-axis; m(y-x) is the bias; s(y-x) is the standard deviation of the x and y difference; and corr_xy the Pierson correlation of x and y.

In order to illustrate this problem, we have run some simulations for the different parts of the swath and compared them to the "true" winds. The simulation process is the following: we choose the NWP winds as the "true winds" and using the NSCAT-2 model function we simulate the measured backscatter. Then, we add some noise

(using the Kp information available in the data product) to the simulated σ^o_m, and we invert the noisy σ^o_m (with the KNMI inversion scheme) into winds.

Figure 6 shows the two-dimensional histograms of KNMI winds (closest solution to NWP wind) versus the NWP winds for wind speed and wind direction for different parts of the swath: NSCAT-like (plots a and b), nadir swath (plots c and d) and outer swath (plots e and f). The results from this simple simulation show a higher standard deviation of speeds and directions in the outer and nadir parts compared to the NSCAT-like parts of the swath. No significant bias is seen in any part of the swath.

We are planning to work on the QuikSCAT inversion problem in the near future.

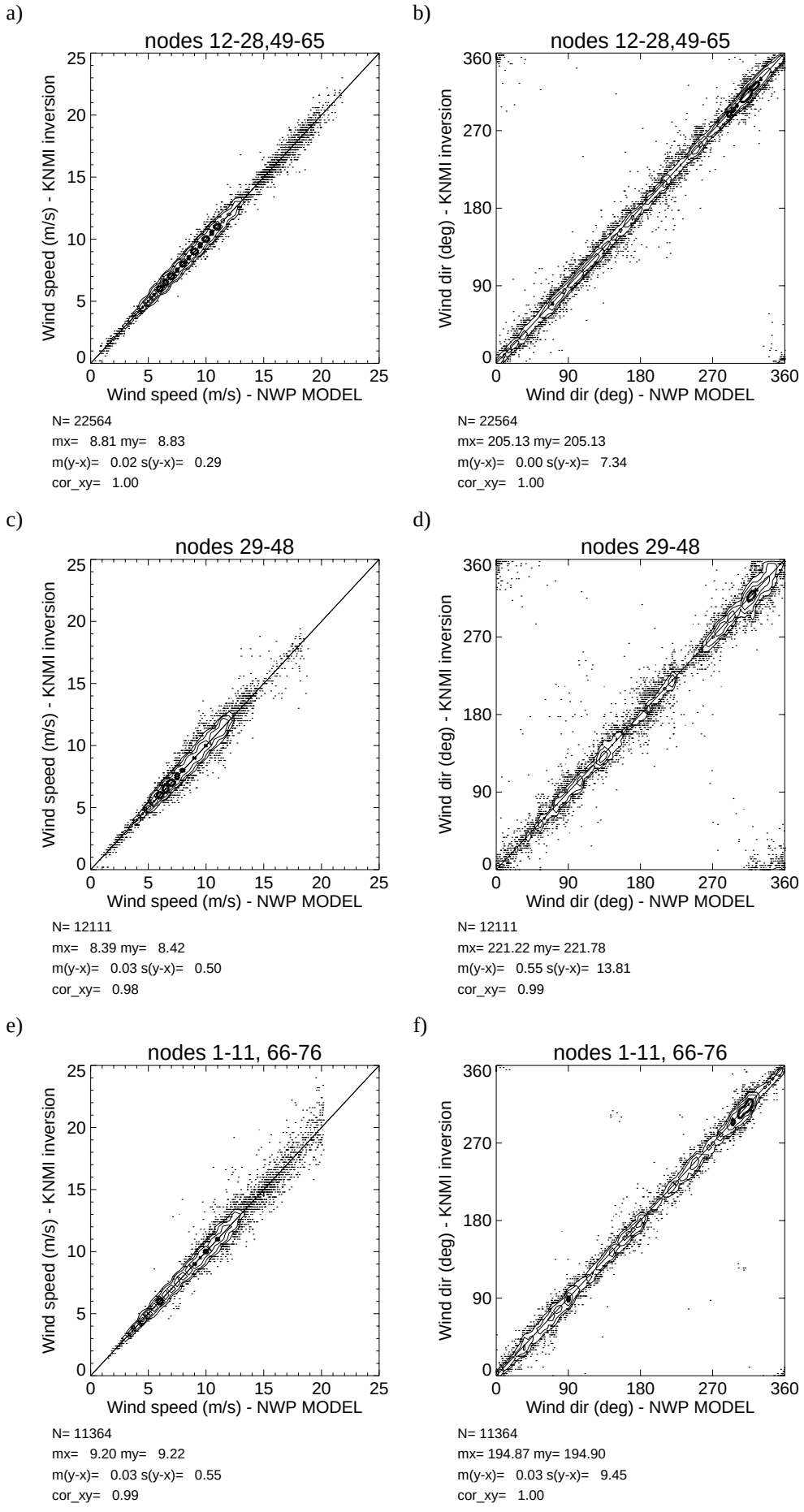


Figure 6 Two-dimensional histogram of KNMI-retrieved wind (closest solution to NCEP) versus NCEP winds in the different parts of the swath: the NSCAT-like (plots a and b), the Nadir (plots c and d) and the outer (plots e and f) regions. The left column is for wind speed the right one for wind direction. The width of the bins is 0.5 m/s in speed and5° in direction. N is the number of points; mx is the mean value of x; my is the mean value of y; m(y-x) is the bias; s(y-x) is the standard deviation of the x and y difference; and corr_xy the Pierson correlation of x and y.

4 Quality Control

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 space-borne 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 Beukering* (1997)]. The impact of observations on weather forecast often critically depends on the Quality Control (QC) applied. For example, *Rohn et al.* (1999) show a positive impact of cloud motion winds on the ECMWF model after QC, while the impact is negative without QC. Besides its importance for NWP, in applications such as nowcasting and short-range forecasting, the confidence of meteorologists in the QuikSCAT data is boosted by a better 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 ERS and NSCAT respectively. Here we adapt this method for QuikSCAT.

The Maximum Likelihood Estimator (MLE) is defined as [adopted from *JPL* (1999)]:

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

where N is the number of measurements, $\sigma_{mi}{}^{o}$ are the backscatter measurements, $\sigma_{s}{}^{o}$ is the backscatter simulated through the Geophysical Model Function (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 $\sigma_{mi}{}^{o}$ and the solution $\sigma_{s}{}^{o}$ lying on the GMF surface in a transformed measurement space where each axis of the measurement space is scaled by $kp(\sigma_{s}{}^{o})$.

The MLE therefore indicates how well the backscatter measurements used in the retrieval of a particular wind vector fit the 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 for example rain, confused sea state, or ice, and as such the MLE provides a good indication for the quality of the retrieved winds.

Rain Effects

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 (inversely proportional to wavelength to the fourth power). Large drops are relatively more important in the scattering and smaller wavelengths more sensitive. For example, *Boukabara et al.* (1999) 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. A dense rain cloud results in a radar cross section corresponding to a 15-20 m/s wind.

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) about 50 times stronger. 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 measured σ^o, 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.

4.1 Quality Control in HDF

In this Section, we define and characterize a parameter based on the MLE to quality control the QuikSCAT data. Since the first QuikSCAT data available are provided by JPL in HDF format (late September 1999), we will first investigate the Quality Control procedure with this data format.

4.1.1 Methodology

The method consists in normalizing the MLE with respect to the wind and the node number (or cross-track location). For a given wind and node number, we compute the expected MLE. 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 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 polarization 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 further 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 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.

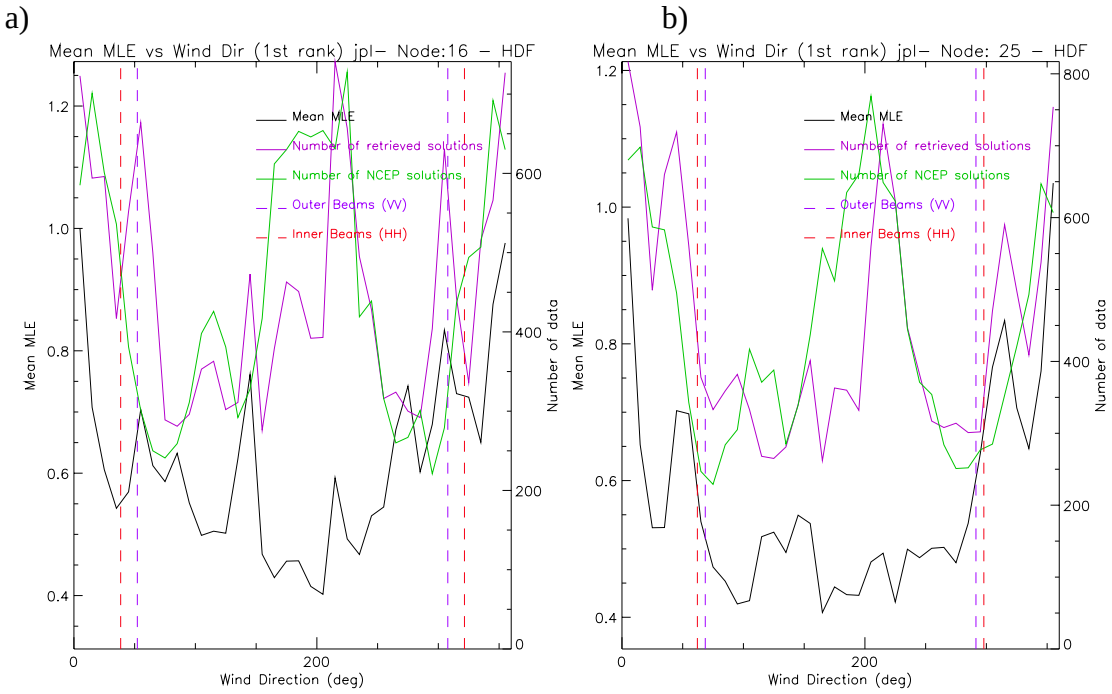


Figure 7 Mean JPL-selected MLE versus JPL-selected wind direction (solid black line) and wind direction distribution of JPL-selected winds (solid purple line) and NCEP winds (solid green line) for nodes 16 (plot a) and 25 (plot b). The direction binning is 10° and the vertical dashed lines correspond to the azimuth of inner (red) and outer (purple) beams with respect to the cross-track flight satellite direction.

Figure 7 shows the mean MLE of the JPL-selected solution versus JPL-selected wind direction together with the wind direction distribution of both the NCEP model and the JPL retrieved solutions (the selected solution after ambiguity removal) for nodes 16 (plot a) and 25 (plot b). The JPL direction distribution shows some significant differences (peaks and troughs) as compared to the NCEP distribution, which may be associated to 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. This is an expected result as measurement sets far away from the GMF solution surface in measurement space (*Stoffelen*, 1998), i.e., with large MLE, are systematically assigned to certain wind directions (the shape of the GMF solution surface makes certain wind directions to be favoured 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>.

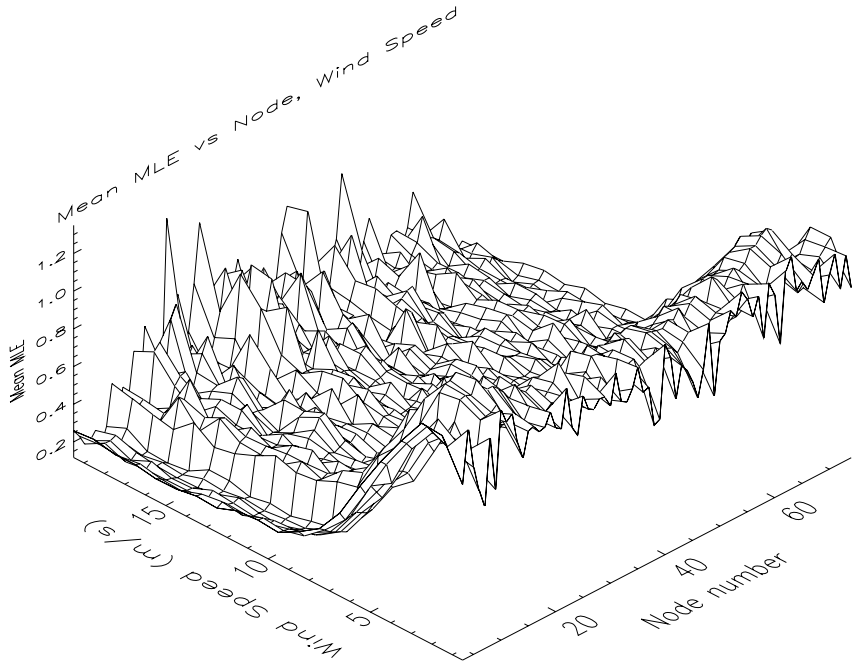
As such, <MLE> is computed as a function of wind speed and node number. The method to compute <MLE> is as follows:

- 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 HDF data. Figure 8a shows a 3D plot of this mean MLE. 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 in rejecting the MLEs which are at least 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. The resulting surface is shown in Figure 8b. The peaks have disappeared in general and at high speeds the surface is much smoothed. 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 9. It is clearly discernible that the shape of the surface remains the same and only the noise 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 (see Appendix A). 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 result of the fitting is shown in Figure 10. The 3D surface is the expected MLE and compares well to the computed mean MLE in the inner part of the swath.

a)



b)

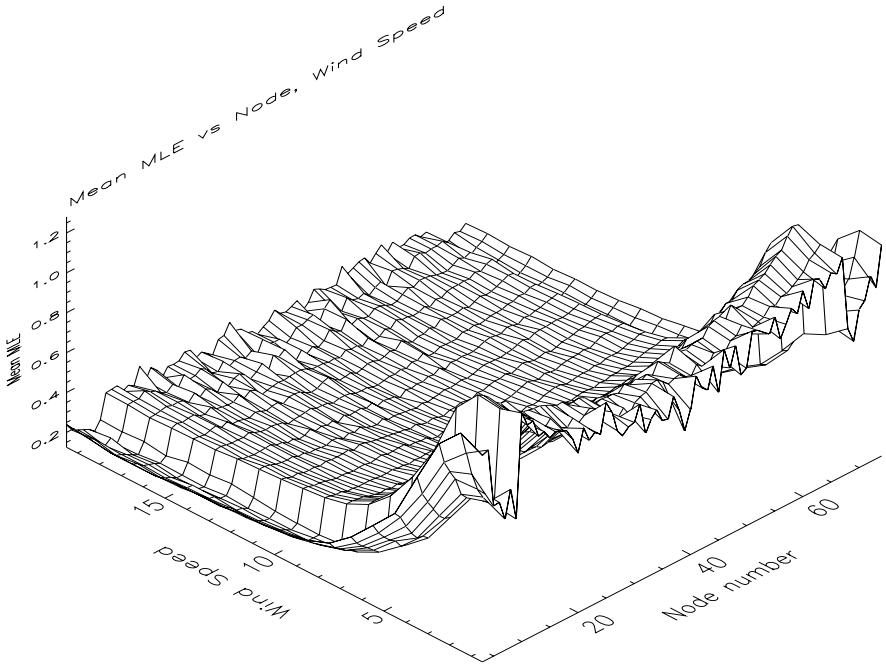


Figure 8 Mean JPL-selected MLE (plot a) and "filtered" meanJPL-selected MLE (plot b) versus JPL-selected wind speed and node number. The speed binning is 1 m/s and the node binning is 1

a) b)

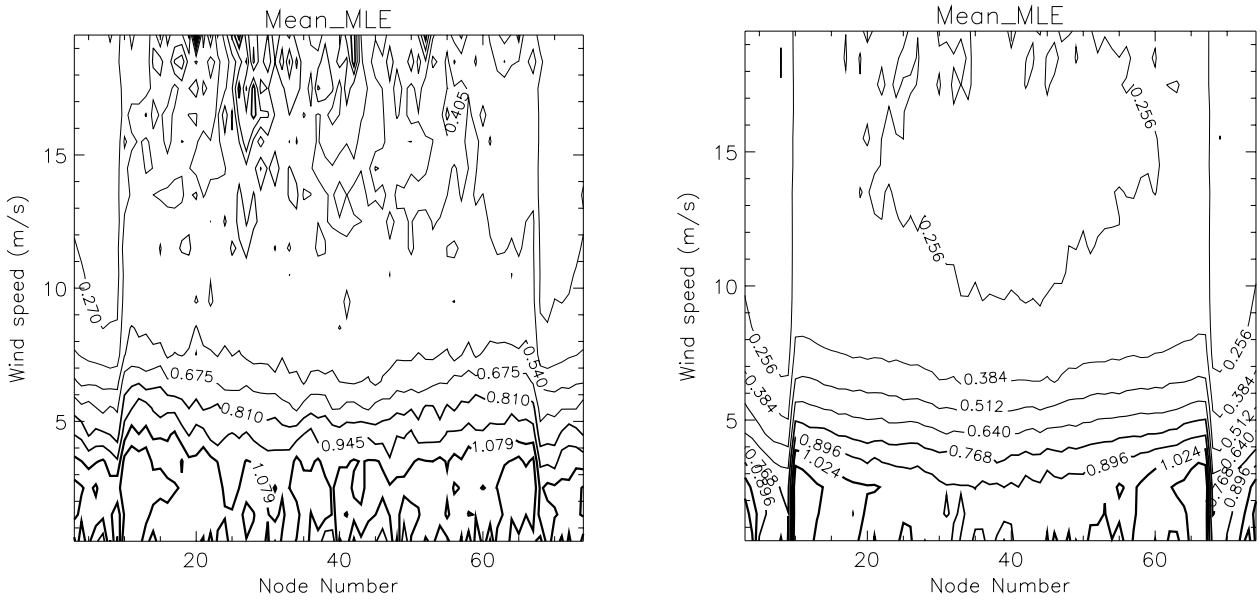


Figure 9 Contour plots of Figure 9a (plot a) and 9b (plot b).

As said before, large departures from fair weather conditions result in a large MLE value. In turn, this MLE, larger than its corresponding expected MLE, results in a large Rn. Therefore, large Rn is indicating bad quality of the retrieved winds for any given WVC.

However, this correlation between the MLE and its expected value is expected to work well when there are more than two measurements and enough azimuth diversity in the σ^o measurements of each WVC, i.e., 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> we can 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, centre 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.

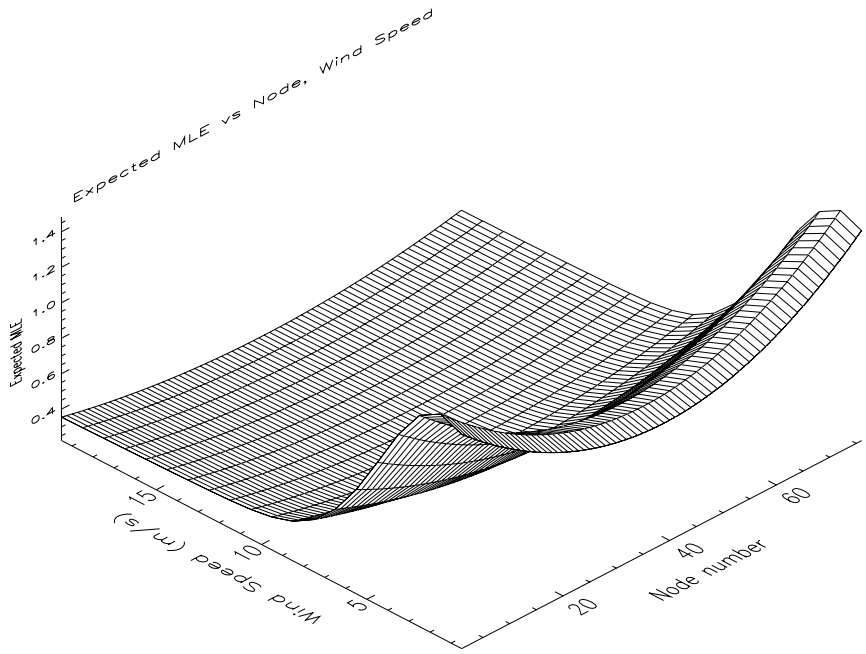


Figure 10 Expected MLE versus wind speed and node number. The speed binning is 1 m/s and the node binning is 1.

4.1.2 Collocations

In order to characterize QC by Rn, we collocate a set of 180 orbits of QuikSCAT HDF data with ECMWF winds and 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 3-hour and 6-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 DMSP (Defense Meteorological Satellite Program) 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.

4.1.3 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 4.1.1, 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. Nodes 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.*, 1999).

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 applicable for the entire inner swath.

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-selected and ECMWF winds (RMS-ECMWF) is used as a quality indicator.

Figure 11 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 data set, 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 plots 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 a different histogram distribution with respect to wind speed. 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.

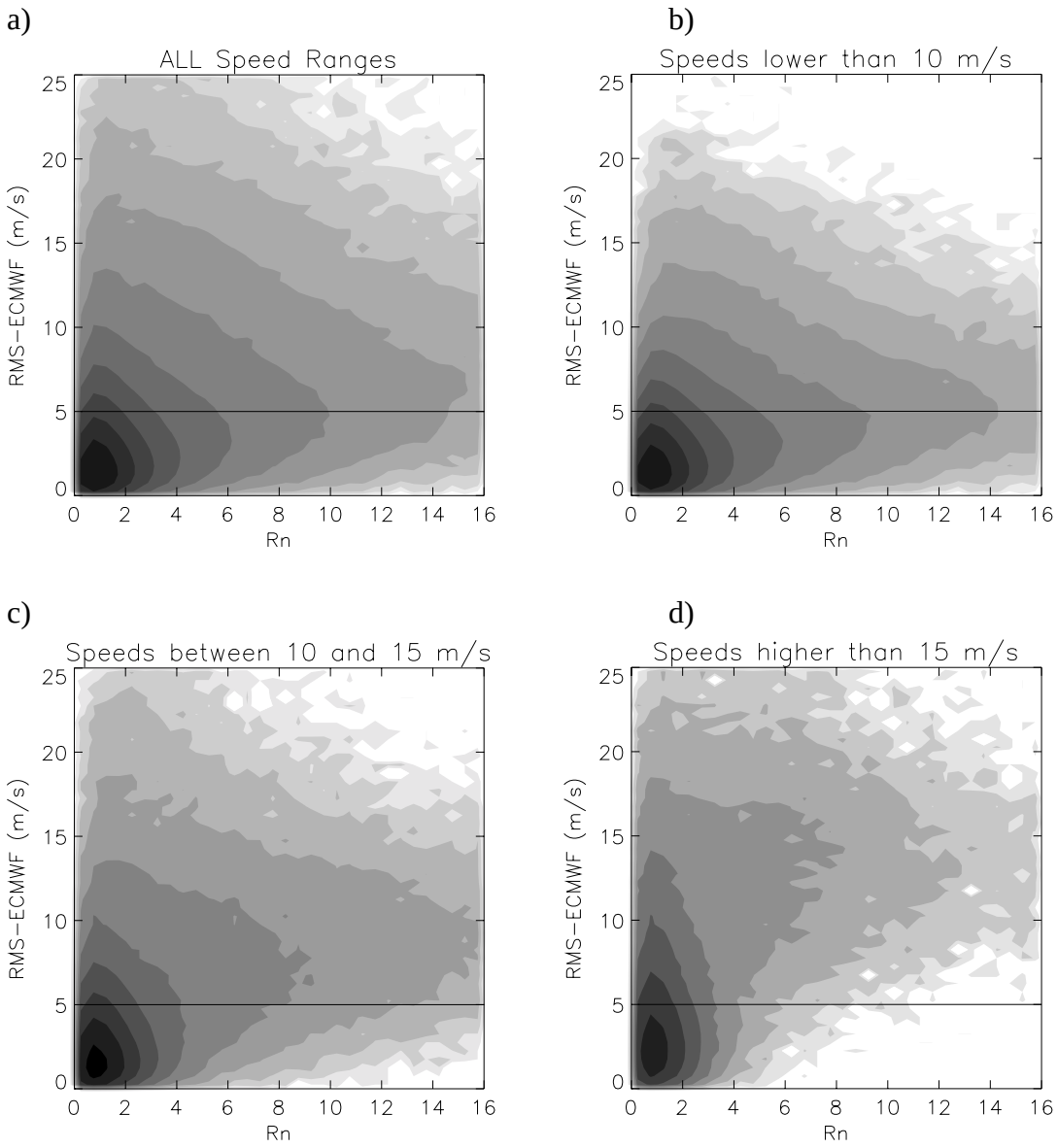


Figure 11 Two-dimensional histograms of RMS-ECMWF versus Rn, for all data (plot a), JPL-selected speeds under 10 m/s (plot b), speeds between 10 and 15 m/s (plot c) and speeds over 15 m/s (plot d). The contouring is in logarithmic scale (two steps corresponding to a factor of 10 in number density) filled from white (unpopulated areas) to black (most populated areas).

Quality Control 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 12 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 increases 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.

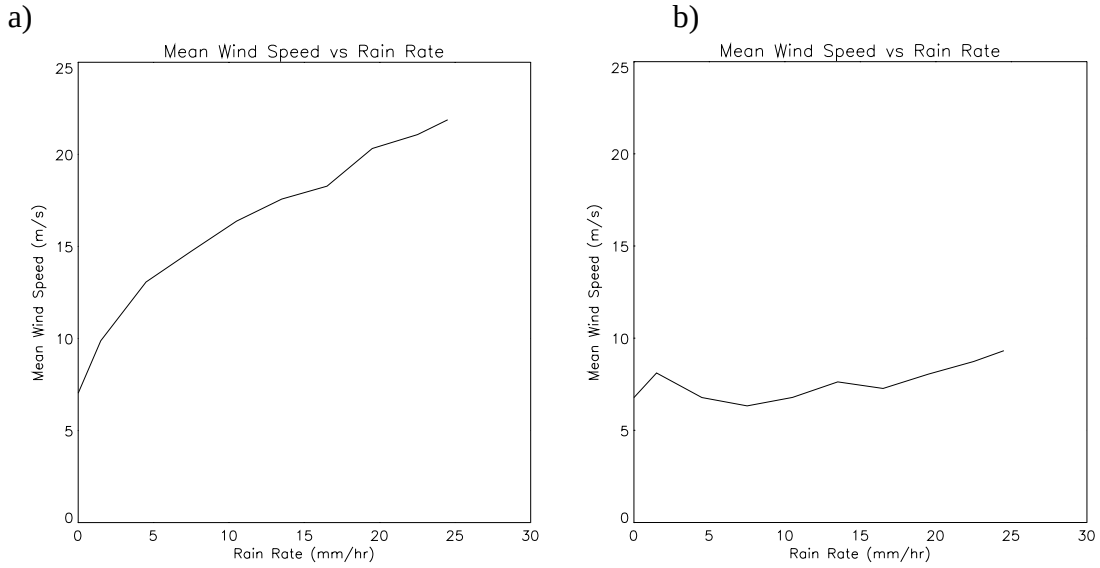


Figure 12 Mean JPL-selected wind speed (plot a) and Mean ECMWF wind speed (plot b) versus rain rate at intervals of 3 mm/hr (except for the rain-free mean speed value, included at 0 mm/hr)

Figure 13 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.

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.

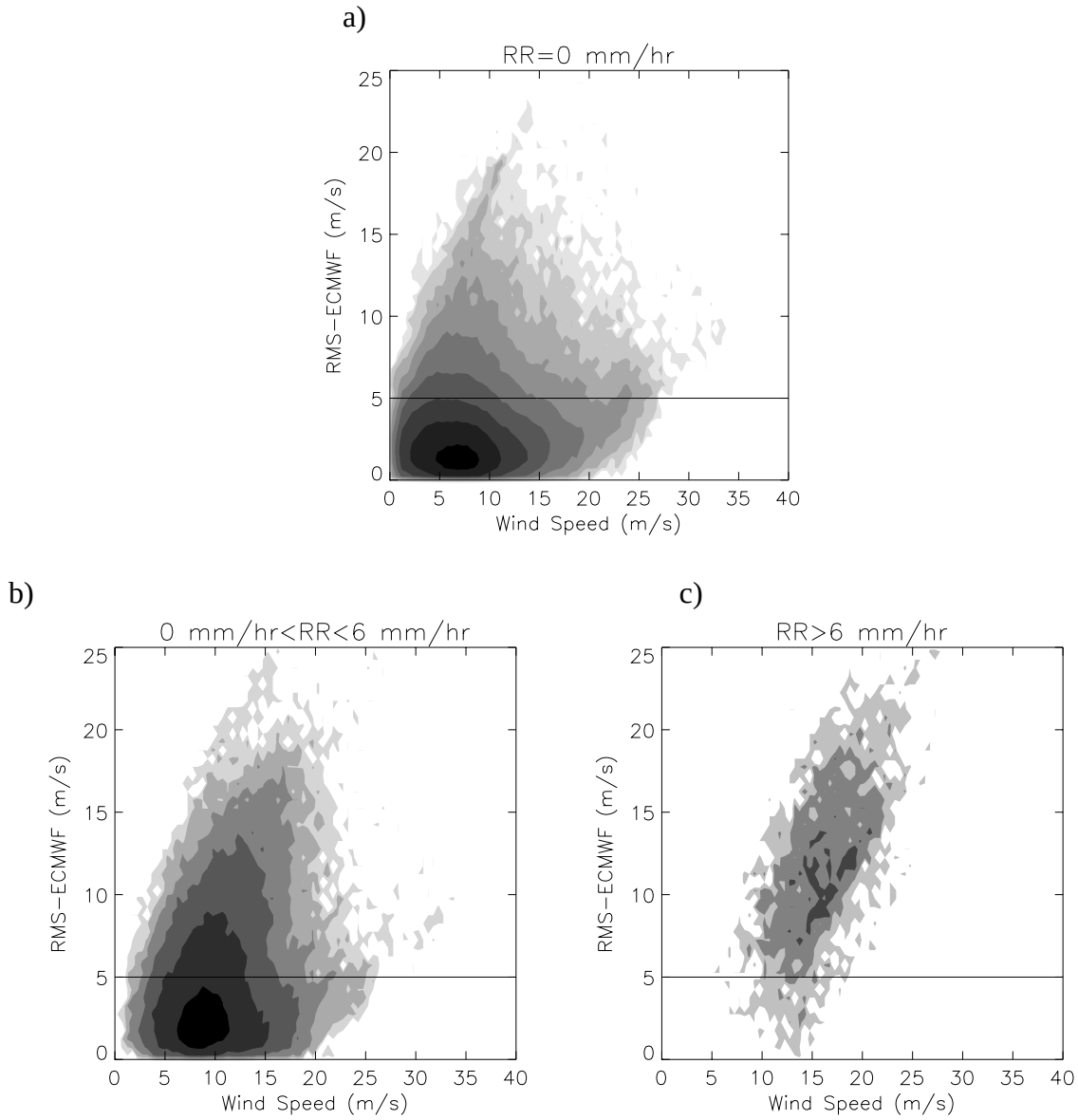


Figure 13 Two-dimensional histograms of RMS-ECMWF versus JPL-selected wind speed for rain-free (plot a), for rain rates from 0 to 6 mm/hr (plot b) and for rain rates above 6 mm/hr (plot c). The contouring is in logarithmic scale (two steps corresponding to a factor of 10 in number density) filled from white (unpopulated areas) to black (most populated areas).

Figure 14 summarizes all these points. The left plots correspond to two-dimensional histograms of Rn versus JPL retrieved wind speeds 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 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 14a and 14b (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 on all higher retrieved speeds would be rejected. Figures 14a and 14b 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 11, 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 14e), we define a minimum threshold value of 2 for speeds higher than 15 m/s. From Figures 14a and 14b, 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 14). Therefore, the defined threshold function is:

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

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.

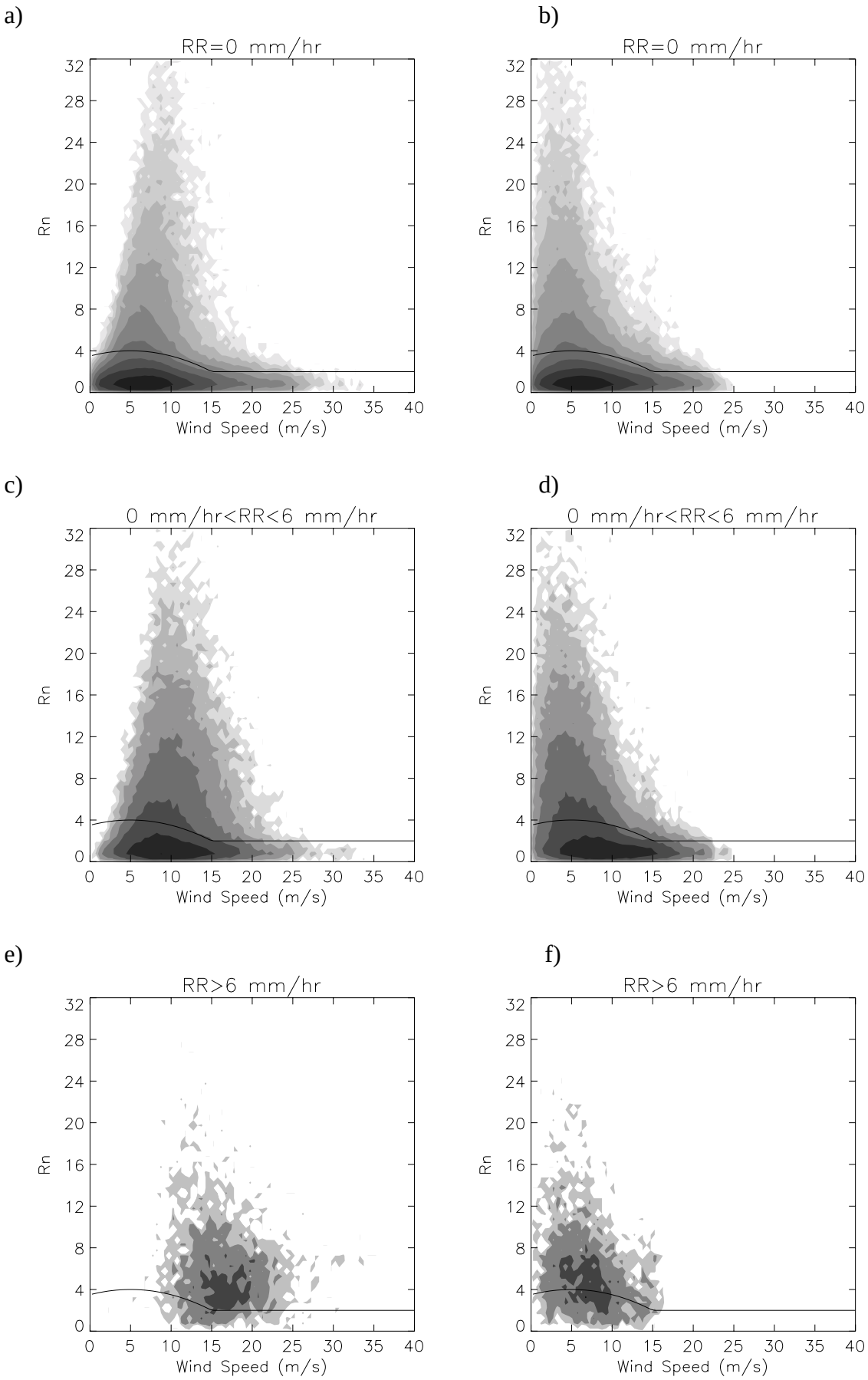


Figure 14 Two-dimensional histograms of Rn versus JPL-retrieved wind speed (left plots) and versus ECMWF speeds (right plots) for rain-free data (plots a and b), for rain rate from 0 to 6 mm/hr (plots c and d), and for rain rate above 6 mm/hr (plots e and f). The contouring is in logarithmic scale (two steps corresponding to a factor of 10 in number density) filled from white (unpopulated areas) to black (most populated areas).

4.1.4 Threshold validation

We test the defined threshold 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. 5.6% of data are rejected and the rejection rate is increasing with wind speed. This is an expected result. As “rainy” cells result in higher retrieved wind speeds (the larger the rain rate the larger the speed bias) and we want to get rid of those cells, the amount of rejections should increase with wind speed. However, in order to reject rain we have defined a threshold which 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 mean 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 mean 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 behaviour of the Rn threshold. When the rain is lower or equal to 6 mm/hr, the percentage of rejections decrease 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 enough wind information) which we may want to keep, but still there is a significant portion of low winds (see discussion of Figure 14) 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).

Figure 15 shows the two-dimensional histograms of RMS-ECMWF versus retrieved wind speed for different rain intervals. The left plots correspond to the accepted solutions and the right plots to the rejected solutions. It is clear when comparing the contour lines of the left with the right plots that the latter show a much more vertical orientation with the maximum (darkest contour) significantly higher than the former (accepted solutions). This is a way to show the mean RMS difference between the accepted and the rejected solutions presented in table 2. For rain rates over 6 mm/hr (see bottom plots) most of the solutions are rejected.

Comparing the distributions of Figures 13a and 13b (prior to QC) with the distributions of Figures 15a and 15c (accepted solutions), it is discernible that either for no rain or for rain rate lower than 6 mm/hr, the distributions have become flatter (less vertically oriented) after QC. This indicates a general decrease of the mean RMS and therefore a good performance of the method.

Tables 4, 5 and 6 are the same as tables 1, 2 and 3 but for the nadir swath. In the nadir swath, there is not always enough azimuth diversity in the σ^o beams. In particular, WVCs very close to the nadir have fore and aft beams 180^o apart, which is almost the same as having only one of the two beams. As said in Section 4.1.1, when there is not enough azimuth diversity this QC procedure may not work well. Therefore, we expect a lower skill of the QC in the nadir swath compared to the sweet swaths. And this is what we see in the results shown in the tables mentioned above.

Comparing table 4 with table 1, we see a larger percentage of rejections in the nadir swath, which increases with wind speed. At speeds higher than 15 m/s, 23.8% of the data are rejected. This represents almost 5% more rejections than in the sweet spots.

Comparing table 5 with table 2, there is a slightly larger percentage of rejections at RMS-ECMWF < 5m/s and a smaller percentage of rejections at RMS-ECMWF > 5m/s in the nadir swath, indicating a slight decrease in the performance of the QC procedure. Although the mean RMS-ECMWF of the accepted solutions is slightly higher in the nadir swath, the mean RMS-ECMWF difference between accepted and rejected solutions is the same (4 m/s), showing a comparable result in both regions.

Comparing table 6 with table 3, when rain is over 6 mm/hr there are slightly less rejections in the nadir swath. This shows again a slightly worse performance in the nadir swath, especially if we consider that overall (see tables 1 and 4) this region suffers more rejections (especially at high winds, where the rain is "located"). However, the percentage of rejections for rain under 6 mm/hr is about 6% higher in the nadir swath. Most of these rejections have an RMS-ECMWF over 5 m/s. This result is unexpectedly positive, as even if the overall portion of rejections with RMS-ECMWF > 5 m/s (see tables 2 and 5) is about 6% smaller in the nadir swath, the portion of rejections when the rain is below 6 mm/hr is around 6% higher for the nadir swath.

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 the wind solutions and their corresponding MLE values. Ambiguity removal then picks the geophysically most consistent solution. Therefore, we recommend to use the QC based on the selected solution.

TABLE 1

	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

	RMS<5	RMS>5	Mean 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

	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

TABLE 4

	Total	V<10	10≤V≤15	V>15
Num. Points (n/a)	2812095	2186477	511131	114487
Accepted (%)	93.7	95.9	88.5	76.2
Rejected (%)	6.3	4.1	11.5	23.8

TABLE 5

	RMS<5	RMS>5	Mean RMS (m/s)
Total (n/a)	2483112	329113	2.81
Accepted (%)	96.8	70.6	2.55
Rejected (%)	3.2	29.4	6.62

TABLE 6

	RR=0	0<RR≤6	RR>6
Num. Points (n/a)	572894	47529	2526
Accepted (%)	96	64.5	14.8
Rejected (%)	4	35.5	85.2

Note : RMS is refered 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.

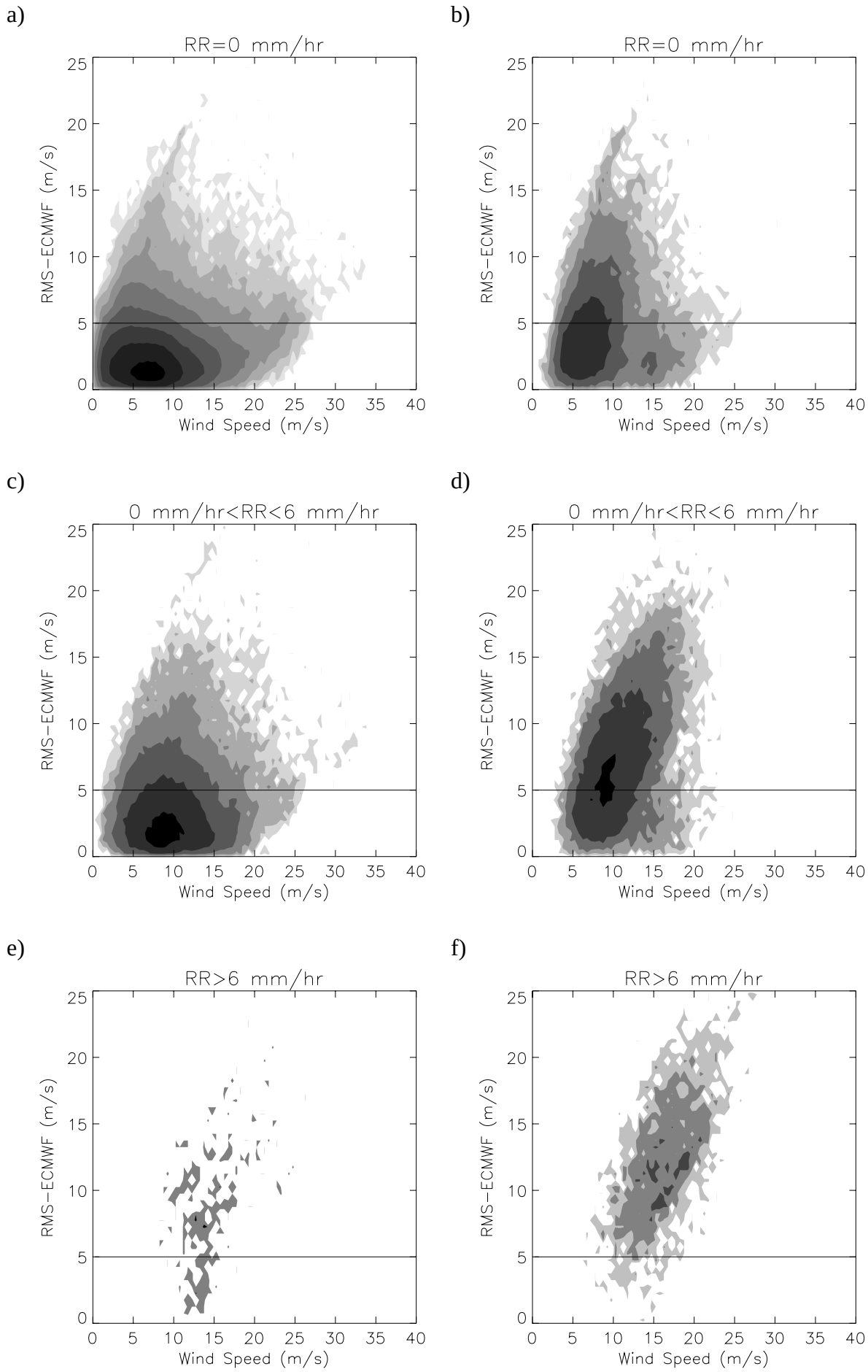


Figure 15 Two-dimensional histograms of RMS-ECMWF versus JPL-retrieved wind speed for the accepted (left plots) and rejected WVCs (right plots). Plots a and b correspond to rain-free data, plots c and d to rain rate from 0 to 6 mm/hr and plots e and f to rain rate above 6 mm/hr. The contour areas are in logarithmic scale (steps of 0.5) filled from white (unpopulated areas) to black (most populated areas).

4.1.5 Cases

In this section, we show a few wind field examples where the QC procedure has been applied. Figures 16, 17 and 18 show triple collocated QuikSCAT-ECMWF-SSM/I data. The arrows in plot a) correspond to the QuikSCAT JPL-selected wind solutions and the colors represent the accepted (green) and the rejected (red) 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 violet solid lines divide the different regions of the swath (outer, sweet and nadir).

In Figure 16, there is a 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 centre to the right side of the plot. This area is dividing a mid and high wind speed area (south part) from a low speed one (north part), suggesting the presence of a wind front. The QC is performing 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, showing 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 less sensitive to wind direction changes. Our QC will not detect these 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.

In Figure 17a, the presence of a wind front is clearly discernible in the middle of the plot, where again a confused sea state and therefore poor quality winds are expected. WVCs along the front line are rejected by QC. This is also the case for the centre 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 wind 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 4.1.1). 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 centre of the low and the associated wind front is not so pronounced as in the QuikSCAT plot. This example illustrates again the potential positive impact of assimilating QuikSCAT winds into ECMWF after using our QC.

Figure 18 is a clear pure wind case. No fronts were predicted by ECMWF (plot b) and almost no rain was observed by SSM/I. Most of the wind solutions have been accepted by the QC. Indeed, QuikSCAT winds show a meteorologically consistent pattern, indicating again a good performance of this QC.

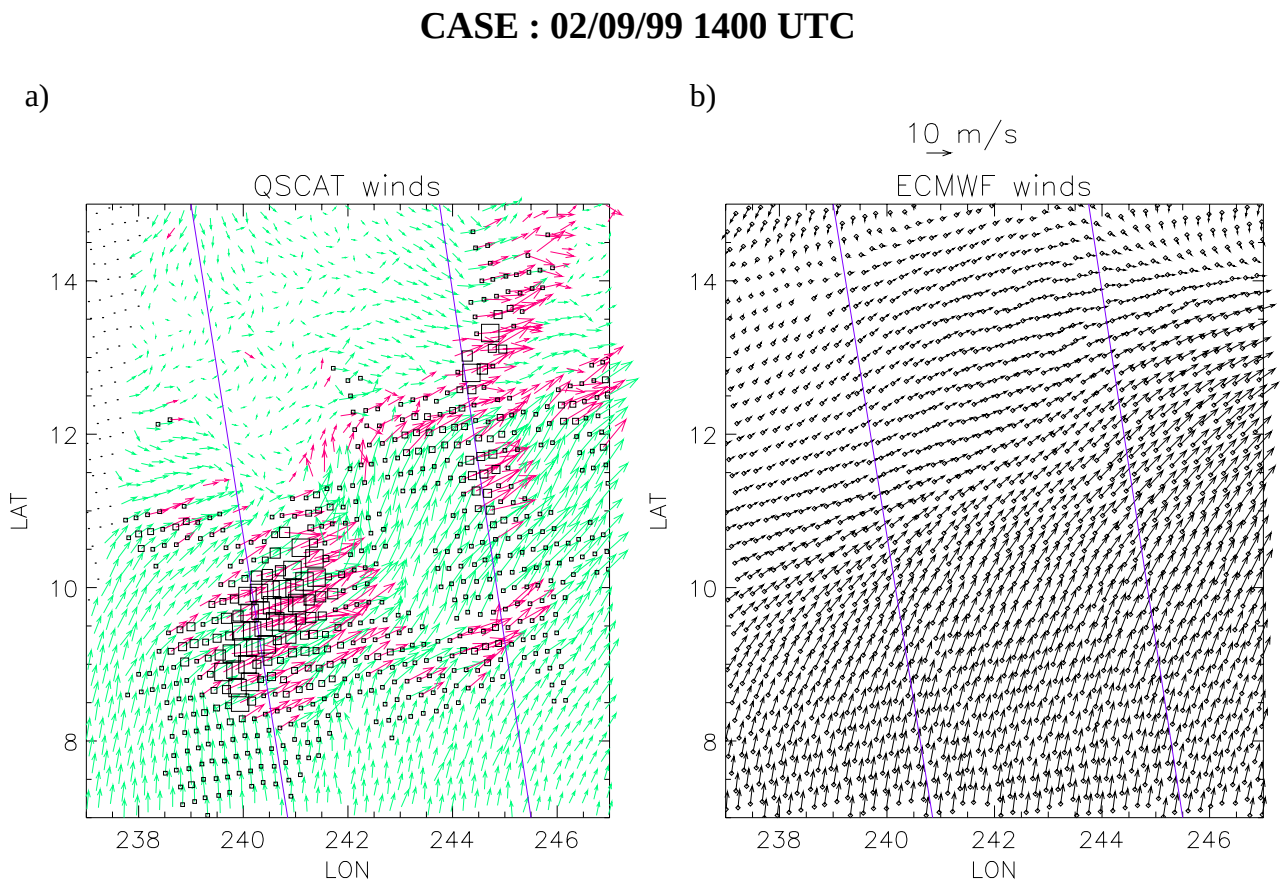


Figure 16 Collocated QuikSCAT-ECMWF-SSM/I data. Plot a shows QuikSCAT wind arrows (JPL-selected winds), where green color correspond to accepted WVCs and red color to rejected WVCs. The size of the squares represent the different rain rates from 0 mm/hr (no square) to 25 mm/hr (the largest ones). Plot b shows the collocated ECMWF winds. The violet lines separate different regions of the swath. In this case, the left side of the plot corresponds to the sweet-left region, the middle to the nadir region and the right side to the sweet-right region. The acquisition date was September 2 1999 at 14 hours UTC

CASE : 28/08/99 0500 UTC

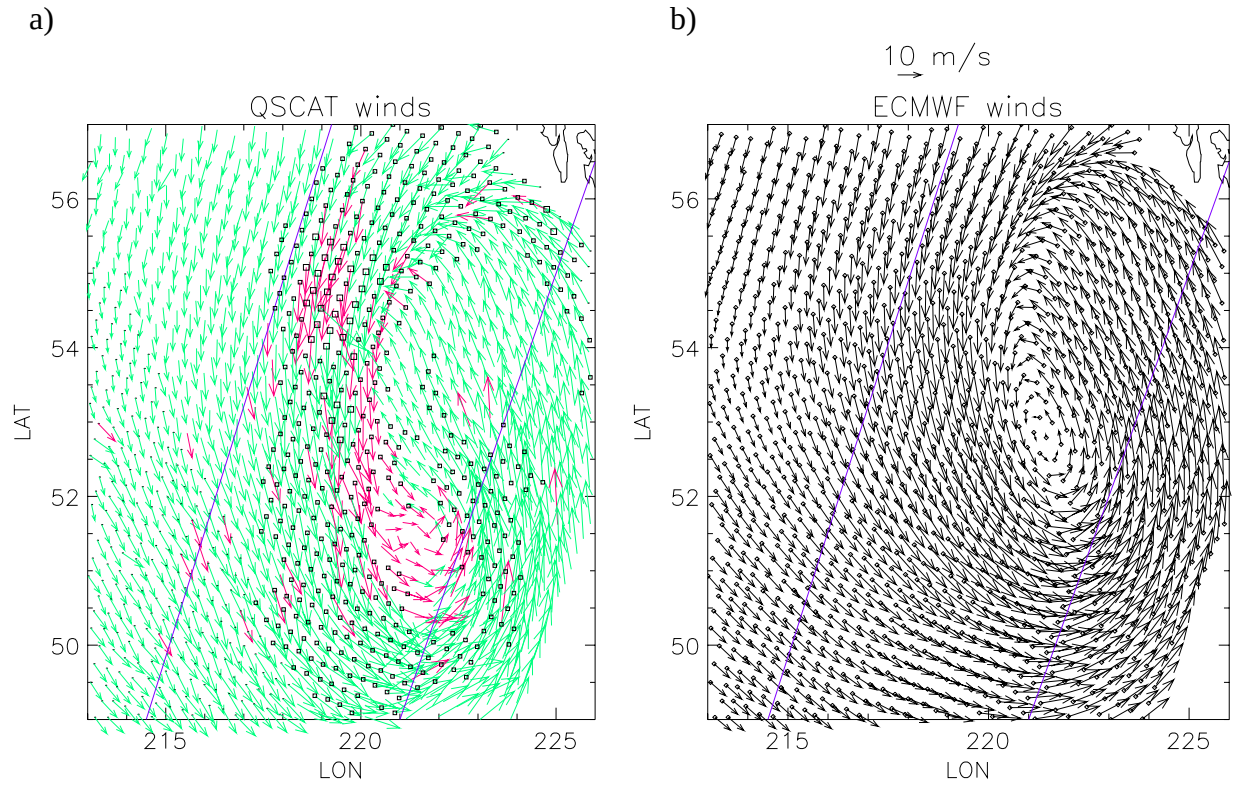


Figure 17 Same as Figure 16 but for different date (August 28 1999 at 5 hours UTC) and location. The violet lines separate the nadir (left side), the sweet-left (middle) and the outer-left (right side) regions.

CASE : 28/08/99 1000 UTC

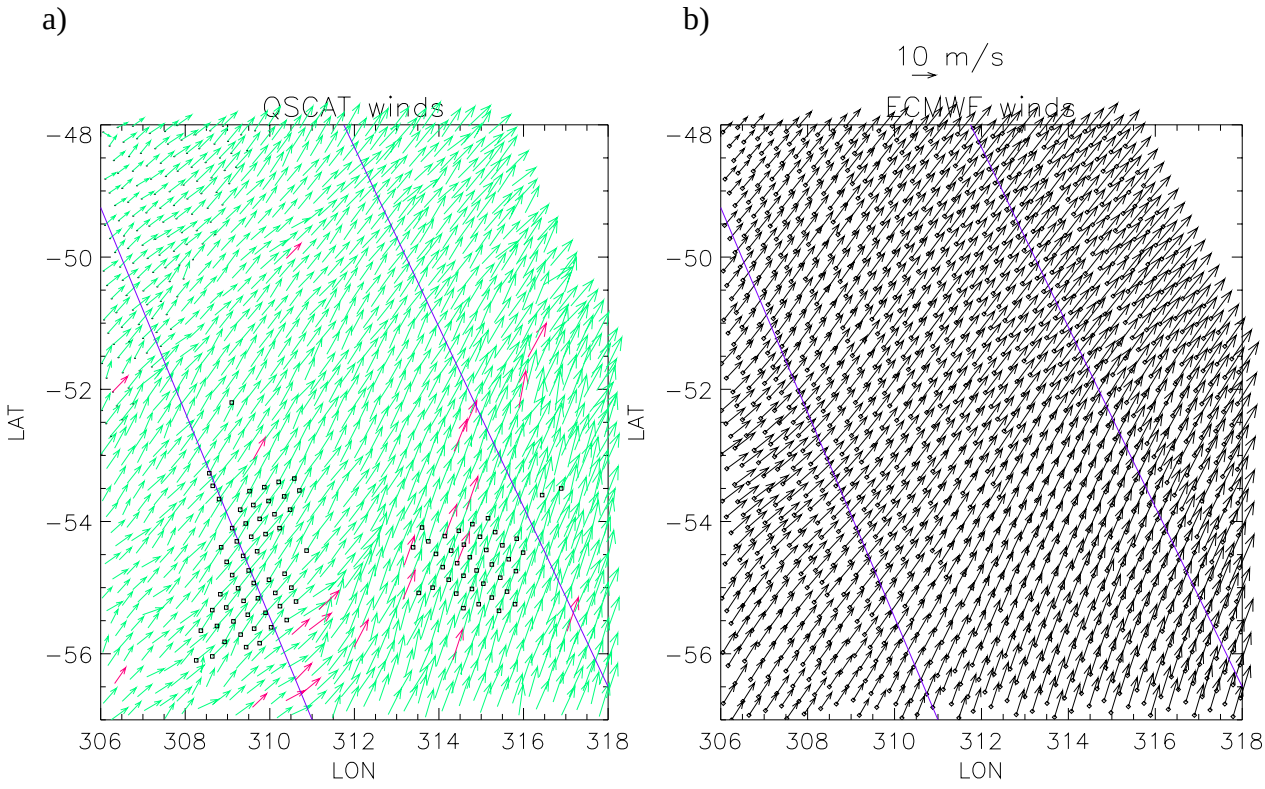


Figure 18 Same as Figure 16 but for different date (August 28 1999 at 10 hours UTC) and location. The violet lines separate the nadir (left side), the sweet-right (middle) and the outer-right (right side) regions.

4.2 Quality Control in BUFR

As shown in Section 4.1, there is a potential positive impact of assimilating QuikSCAT winds into NWP models, such as ECMWF, after using our QC. Since March 2000, the QuikSCAT Near Real Time (NRT) BUFR product is available. As this product is the one used for assimilation purposes, we also want to set a QC for the BUFR product.

4.2.1 Data Repetition

First of all, we would like to make a few comments about a common problem in the QuikSCAT BUFR data. It appears that certain rows of data are repeated in the data stream. For a given row that repeats, the data values may be identical or different, depending on whether additional σ^o values are present in one of the row records or not. The row of lower quality is the one with fewer σ^o values. In all cases, the higher quality data is expected to be furthest from the edge of the data pass. A parameter called **time_to_edge**, which indicates the time difference to the nearest edge of the data pass, has been set in the product. Therefore, in order to avoid repetitions and keep the higher quality data, we first have to detect the repeated rows and then keep the one with the largest value of **time_to_edge** (*Leidner et al., 2000*).

The above procedure always requires reading the data an extra time and this may be computationally very inefficient when used in operational assimilation. In order to avoid this inefficiency problem, we have made an exhaustive analysis over a set of 6-hour QuikSCAT BUFR data files available at the ECMWF MARS archive. Here are the key points of this analysis :

a) there is no more than one repetition;

b) when a row is repeated, the one of highest quality can be either before or after the one of lower quality;

c) there is 10% (approx) of repetitions per file;

d) if we only look at WVC with 4 σ^o, there is only 1% of repetitions;

e) the retrieved winds of repeated WVCs with 4 σ^o are almost identical.

Therefore, if we always keep WVCs with 4 σ^o and we reject repetitions when they occur (no matter if the repeated or second one is of higher quality as in this case the winds are almost identical), the data files do not need to be read in advance as with the other procedure.

This procedure can only be used in the inner swath as in the outer swath we have a maximum of 2 σ^o per WVC. Since for now we are only interested in the inner swath (see Section 4.1.1) and the amount of WVCs which contain less than 4 σ^o and wind-

derived information is only 1% of the total in this part of the swath, the procedure is not expected to distort the results in the QC.

Furthermore, we can apply the same concept in the outer swath but keeping WVC with 2 σ^o and rejecting repetitions and extend this method to any operational use of the BUFR data.

4.2.2 Differences with the QC procedure in HDF

As for HDF, the QC in BUFR is based on Rn. In order to compute and characterize Rn in BUFR, we have applied the same procedure than for HDF. However, a few differences were found which we think are important to report.

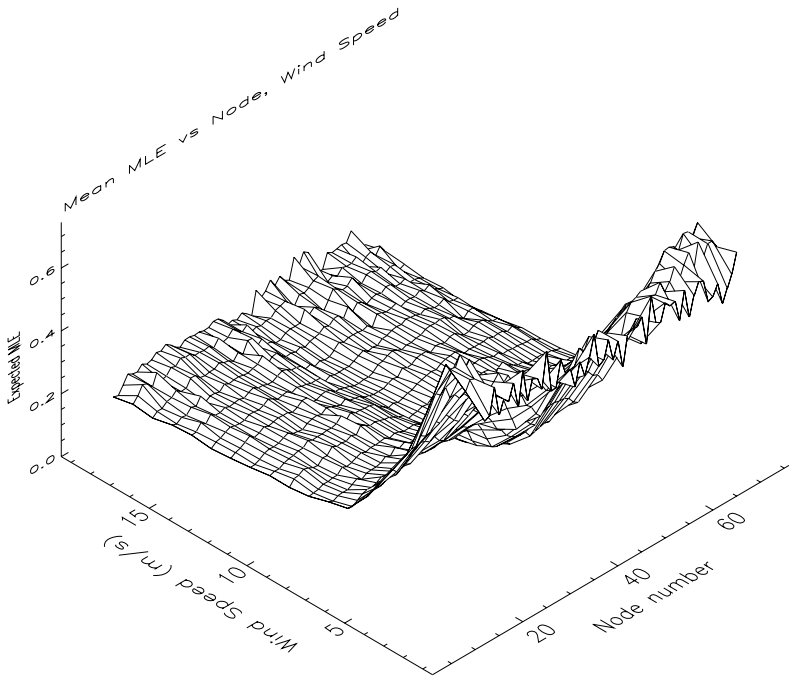
The expected MLE is also computed from 60 orbits of real data (BUFR in this case). From the mean MLE surface versus wind speed and node number, the noise is filtered using the same iterative process as for HDF. However, the MLEs rejected are three (or more) times higher than the mean MLE instead of two times as for HDF. This is done to keep consistency in the filtering procedure in terms of rejecting a small amount of data and conserving the shape of the original function (see discussion in section 4.1.1).

Figure 19a shows the filtered mean MLE versus wind speed and node number for BUFR. Comparing this surface with the one for HDF (Figure 8b), both are very similar although the BUFR surface looks more irregular for speeds higher than 7 m/s. These irregularities make the two dimensional function fit (see Appendix B) to the BUFR surface less accurate. As said in section 4.1.1, the function fit is required for extrapolation purposes. Figure 19b shows the function fit (or expected MLE surface). It is clearly discernible that the irregularities seen in Figure 19a are filtered out in the fit, but the main shape of both surfaces remains the same and therefore the accuracy of the resulting Rn is not expected to decrease significantly.

Both surfaces in Figure 19 are for speeds lower than 20 m/s and for the inner nodes. The two dimensional function fit is used in the same way as in HDF to extrapolate the expected MLE surface for winds higher than 20 m/s and the outer nodes.

Figure 20 shows the contour plot of a two-dimensional histogram of RMS-ECMWF against Rn for two weeks of BUFR data. As in Figure 11 (same plot but for HDF), the RMS-ECMWF increases as Rn increases, or in other words, the quality of data is decreasing with increasing Rn. From plots b), c) and d) we can also say that the RMS-ECMWF is increasing more rapidly with Rn at higher wind speeds. However, when comparing both Figures, it is clearly discernible that the RMS-ECMWF in HDF is increasing more rapidly with respect to Rn for all wind speeds, suggesting a better behaviour of the HDF Rn as a quality control indicator.

a)



b)

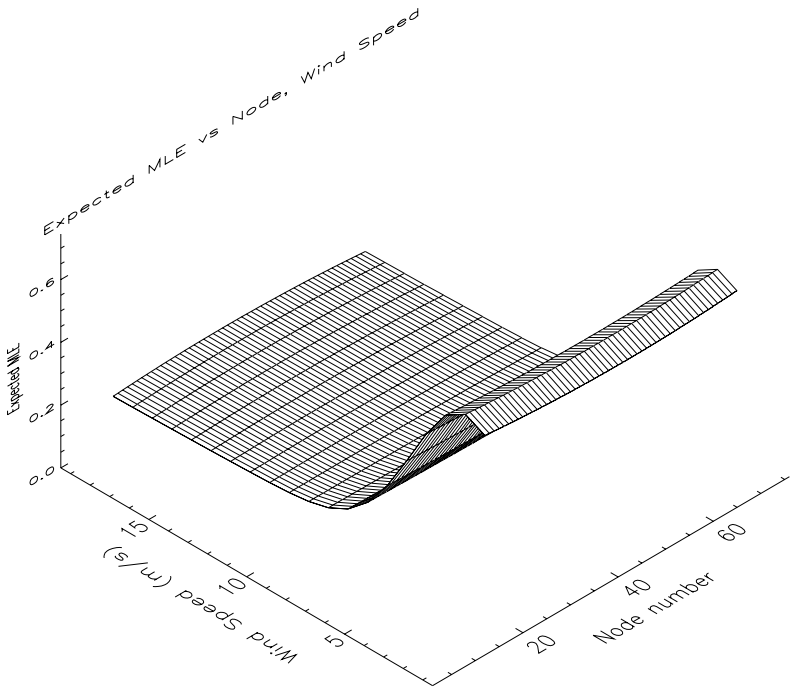


Figure 19 "Filtered" mean MLE (plot a) and Expected MLE versus wind speed and node number for the inner swath nodes. The speed binning is 1 m/s and the node binning is 1.

Nevertheless, we have looked at the same plots as in Figures 12 and 13 but for BUFR (not shown) and they are very similar to HDF. Therefore, the rain rate is also

proportionally increasing the retrieved wind speed and above 6 mm/hr produces undesirable “rainy” WVCs. Moreover, the same plot as in Figure 14 but for BUFR (not shown) is also very similar to HDF, suggesting that the optimal BUFR Rn threshold may be the same as that used for HDF.

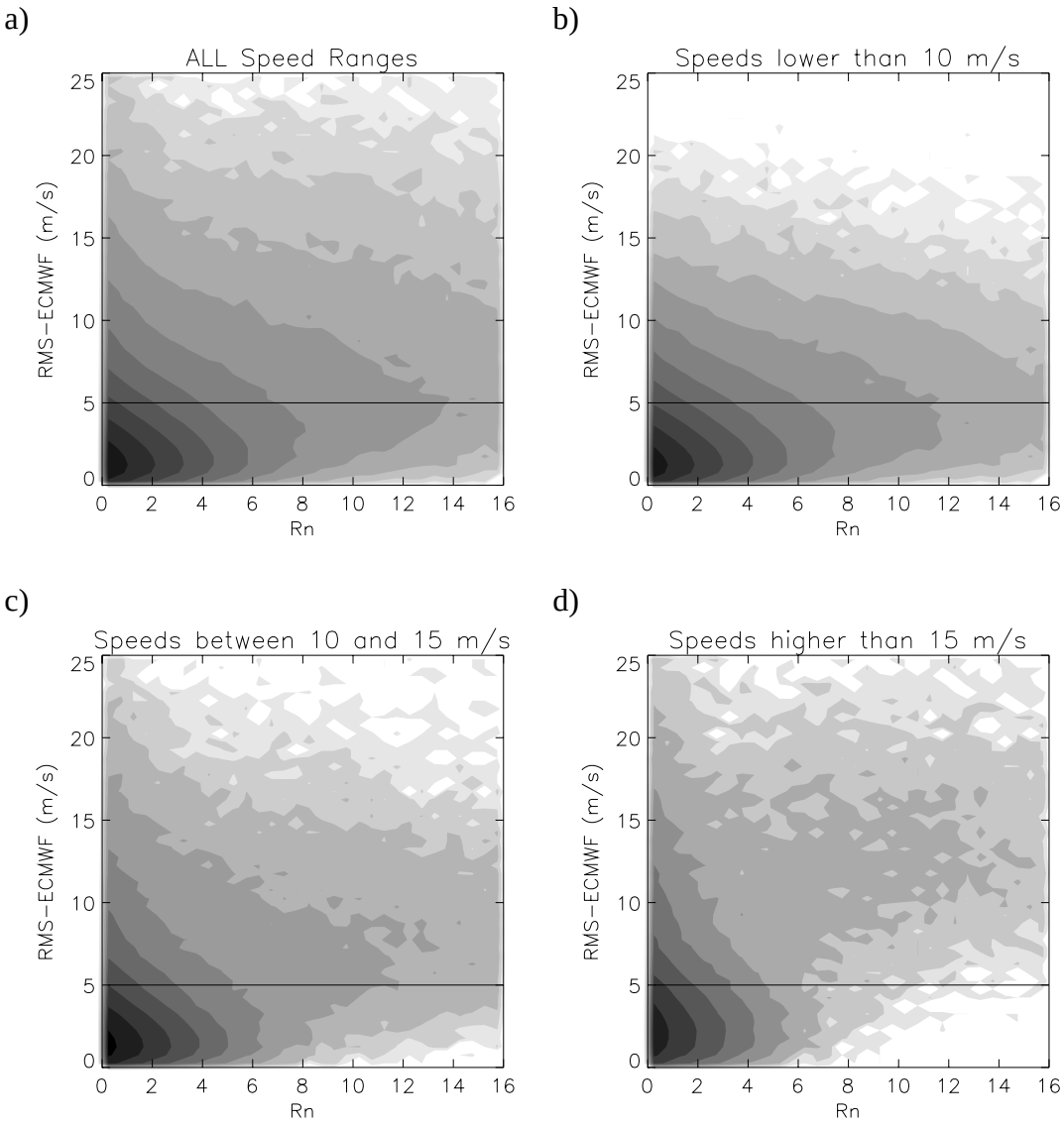


Figure 20 Same as Figure 11 but for BUFR data.

Note that, as for HDF, we have tested different thresholds. The one which has given better results is the one used in HDF.

4.2.3 Threshold Validation

In the same way as in HDF (see section 4.1.4), we test the defined threshold against ECMWF and SSM/I collocations. The results for the sweet parts of the swath are shown in tables 7,8 and 9, and the results for the nadir parts are shown in tables 10, 11 and 12. Tables 7-12 are in the same format as tables 1-6 (see section 4.1.4) respectively.

In general, the results show that the BUFR QC works slightly worse than the HDF (see discussion on the Rn behaviour in the previous section).

Comparing table 7 to table 1, we appreciate a larger percentage of rejections in BUFR at all speeds. From tables 8 and 2, we see that this excess of rejections is mostly concentrated below the RMS value of 5 m/s, which in turn makes the mean RMS-ECMWF smaller. The mean RMS-ECMWF difference between accepted and rejected solutions is about 3 m/s in BUFR while in HDF it is 4 m/s. This indicates a better performance of the HDF QC. Moreover, from tables 9 and 3, the amount of “rainy” WVCs rejected is slightly lower for BUFR, perhaps as a result of the slightly poorer performance.

Comparing the BUFR QC in the nadir (tables 10-12) with the HDF QC in the same region (tables 4-6), we can draw similar conclusions than for the sweet parts of the swath, except that in this case the total number of rejections in BUFR is comparable to HDF. The reason why the number of rejections in the nadir swath is not higher for BUFR is the fact that the <MLE> in BUFR is misfit towards higher values (compared to the filtered mean MLE values) in the nadir parts (see Figures 19a and 19b). This in turn decreases the value of Rn and therefore decreases the number of rejections.

In general, the BUFR QC works fine, although its performance is slightly worse than the HDF QC. This is due to the fact that the properties of the Rn as a quality control indicator are somehow smeared in BUFR compared to HDF, as it is further explained in the next section.

TABLE 7

	Total	V<10	10≤V≤15	V>15
Num. Points (n/a)	3005557	2261475	617140	126942
Accepted (%)	93.3	94.9	90.5	79.2
Rejected (%)	6.7	5.1	9.5	20.8

TABLE 8

	RMS<5	RMS>5	Mean RMS (m/s)
Total (n/a)	2805852	203084	2.26
Accepted (%)	95.5	63.8	2.07
Rejected (%)	4.5	36.2	4.92

TABLE 9

	RR=0	0<RR≤6	RR>6
Num. Points (n/a)	647292	56939	2904
Accepted (%)	95.3	72.7	16.1
Rejected (%)	4.7	27.3	83.9

TABLE 10

	Total	V<10	10≤V≤15	V>15
Num. Points (n/a)	1744647	1290254	372353	82040
Accepted (%)	93.9	95.9	90.3	79.1
Rejected (%)	6.1	4.1	9.7	20.9

TABLE 11

	RMS<5	RMS>5	Mean RMS (m/s)
Total (n/a)	1585453	160703	2.48
Accepted (%)	96.1	72.3	2.29
Rejected (%)	3.9	27.7	5.49

TABLE 12

	RR=0	0<RR≤6	RR>6
Num. Points (n/a)	360953	28150	1536
Accepted (%)	95.9	70.7	19.4
Rejected (%)	4.1	29.3	80.6

Note : RMS is refered 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.

4.2.4 Collocations HDF/BUFR

In order to study this smearing effect in the BUFR Rn, we perform some collocations between the HDF and the BUFR product.

Figure 21 shows the contour plot of the two-dimensional histogram of the BUFR MLE versus the HDF MLE. As the Rn is just a normalized MLE, the smearing effect in the BUFR Rn is expected to be seen in the BUFR MLE as well. The plot confirms this. The distribution shows only small apparent correlation between both MLEs.

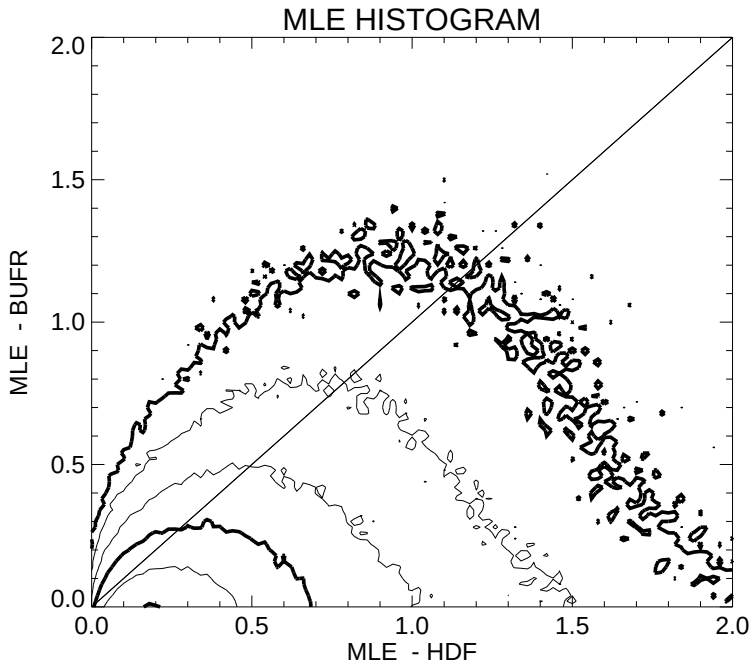


Figure 21 Contour plot of the two-dimensional histogram of the BUFR MLE versus the HDF MLE. The contour lines are in logarithmic scale (each step is half an order of magnitude).

In order to investigate whether this anomalous behaviour of the BUFR Rn and MLE is affecting the quality of the retrieved winds, we perform some triple collocations of HDF, BUFR, and ECMWF winds.

Figure 22 shows the two-dimensional histograms of BUFR winds versus HDF winds (upper plots), BUFR versus ECMWF (mid plots) and HDF versus ECMWF (bottom plots). The left plots correspond to the histograms of wind speeds and the right plots to the histograms of wind directions. Both BUFR and HDF retrieved winds correspond to the 1st rank solution.

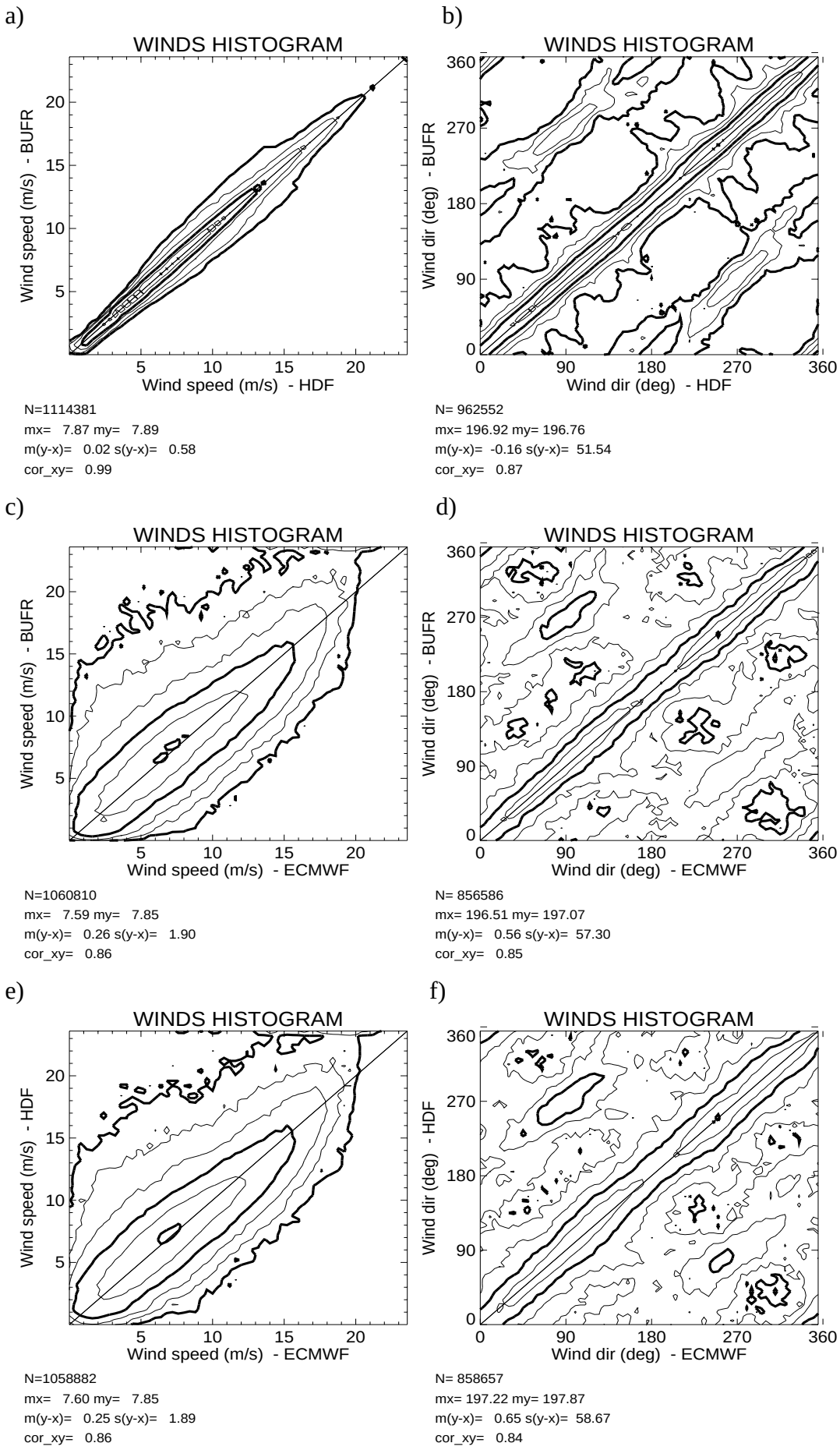


Figure 22 Two-dimensional histogram of BUFR winds versus HDF winds (plots a and b),BUFR winds versus ECMWF winds (plots c and d) and HDF winds versus ECMWF winds (plots e and f). The left plots correspond to wind speeds (bins of 0.4 m/s) and the right plots to wind directions (bins of 2.5 °).

From the upper plots it is discernible that the BUFR and HDF retrieved winds are not identical, although very similar. Plot a) shows almost no bias in speeds and a very small STD (0.58 m/s). Plot b shows a typical effect of comparing 1st rank solutions, which is the secondary distribution around 180°. This is due to the fact that 1st and 2nd rank solutions (usually with very similar wind speeds but wind directions 180°apart) can have very similar MLE values and therefore be switched from one data product to the other. This effect is leading to very high STD values. Still, we can see from the correlation factor (0.87) that the retrieved directions are similar.

Looking at the mid and the bottom plots, we can see almost no difference between HDF and BUFR when compared to ECMWF winds. Plots c) and e) show almost identical wind speed distributions with almost the same bias and STD. Plots d) and f) show very similar wind direction distributions with almost the same correlation factor.

Therefore, we can conclude that the smearing effect in the BUFR Rn is not affecting the quality of the retrieved winds.

4.2.5 MLE Simulation

The σ^o averaging prior to the wind inversion reduces the number of σ^o observations used in the wind retrieval for the BUFR product as compared to HDF. We believe that the smearing of the BUFR MLE compared to the HDF MLE is mainly due to this lower number of observations that are used in the inversion. When the number of observations N is low, the distribution of the MLE will vary considerably if we vary N.

In this section, we use the JPL selected winds of the BUFR files as truth to simulate two sets of measurements. The first set simulates the BUFR product and the second one the HDF product.

Number of σ^o

In order to adequately simulate both products we have to use a realistic number of σ^o per WVC. From section 4.2.1, it is obvious that using 4 σ^o per WVC, one for each beam, to simulate the BUFR set is realistic. In the case of the HDF simulation, we produce a variable number of measurements depending on the WVC number and beam. The more observations of a particular WVC that we simulate, the larger the measurement noise that we assume, such that the information content is the same in each simulated HDF and BUFR WVC.

Figure 23 shows the histogram of the number of measurements per WVC and beam for one day of HDF data. Plot a corresponds to WVC number 12 and plot b to WVC number 38. It is clearly discernible from the different distributions of plots a and b that the number of measurements in HDF varies from one WVC to another. Moreover, these distributions are broad, indicating that the number of measurements

is considerably varying in each WVC as well. To simplify the simulation, we have chosen the number of σ^o corresponding to the peak of each distribution as the fixed value which will represent the number of σ^o for each particular WVC and beam.

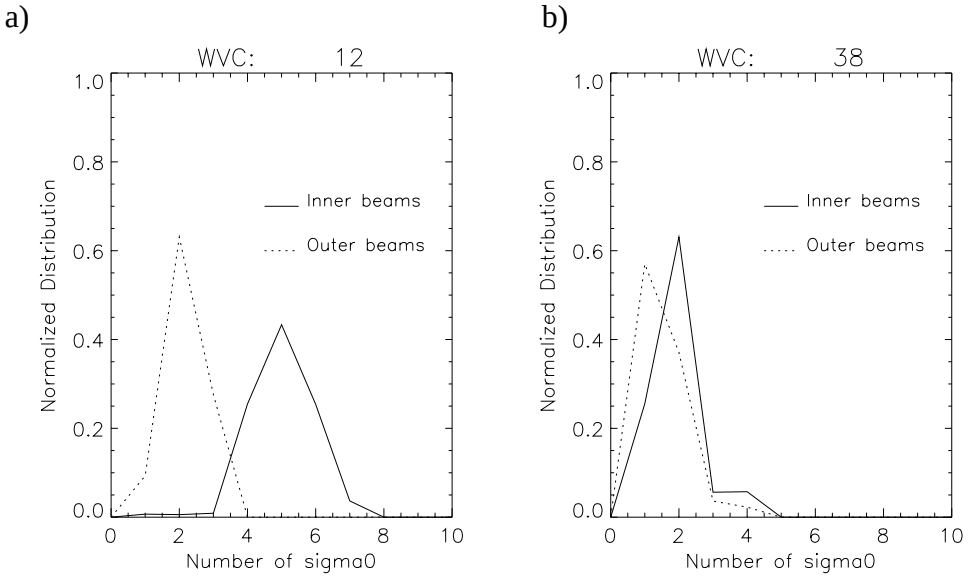


Figure 23 Normalized histogram of the number of sigma0 for WVC number 12 (plot a) and 38 (plot b). The solid line correspond to the inner swath beams (fore and aft) and the dotted line to the outer swath beams.

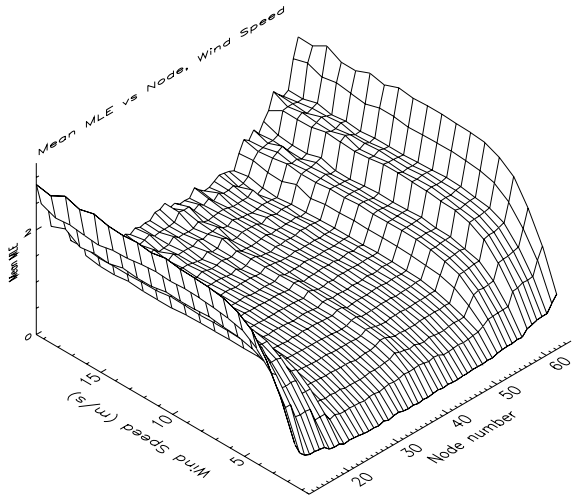
MLE scaling

Once we have simulated both sets of measurements, we invert them to derive the MLE.

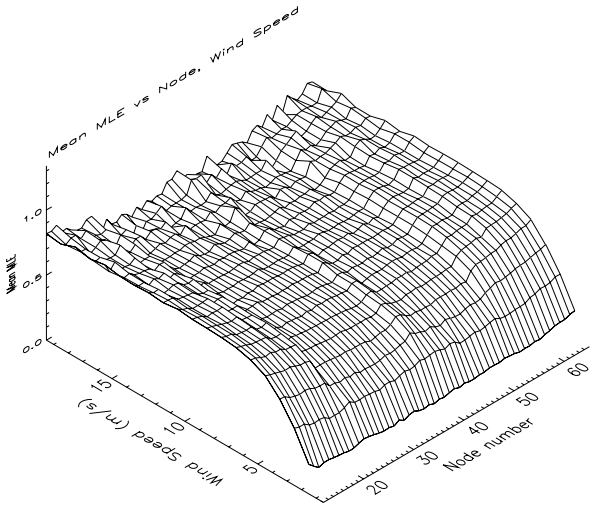
Figure 24a shows the mean MLE surface as a function of wind speed and WVC number for the HDF simulation. The surface show some steps in the cross-track axis which correspond to the different number of σ^o used in the simulation (the number increases with the distance to nadir). These steps are not seen in Figure 8b (same surface but for HDF real data). This is due to the fact that the MLE has to be scaled with the number of σ^o used in the inversion in order to get a normalized MLE along the cross-track axis. Figure 24b shows the same surface as Figure 24a but scaled with the number of σ^o. The steps, although very much reduced, are still visible in the surface. This slight increase of the surface with increasing distance to nadir is also seen in Figure 8b. The increase in Figure 24b is stepwise and not monotonic as for real data (Figure 8b) because of the approximation in the number of σ^o made in the simulation of HDF data. The reason for this increase lies in the inversion. As the inversion is a non-linear process, the scaling (linear correction) is not sufficient to compensate the increase in the MLE due to the increase in the number of σ^o. Nevertheless, the mean MLE surfaces show that this effect is minor.

Figure 24c shows the mean MLE surface versus wind speed and WVC number for the BUFR simulation. As the number of σ^o is kept constant for all WVC, no increase in MLE is produced with increasing distance to nadir. As for the HDF case, the MLE behaviour across track compares well with the real case in BUFR (Figure 19a).

a)



b)



c)

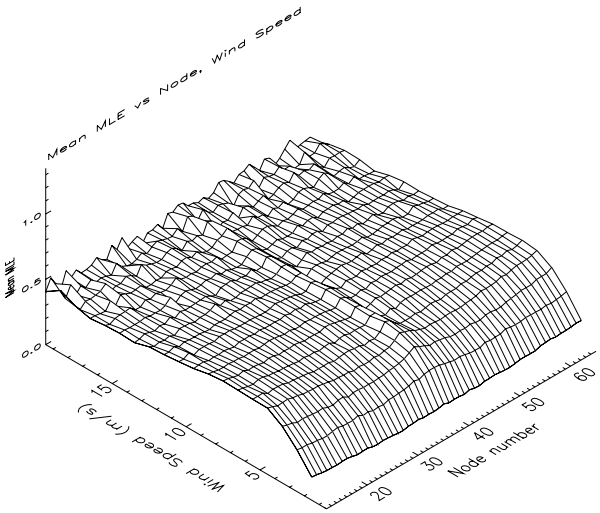


Figure 24 Mean MLE versus wind speed and node number. Plots a and b are for HDF simulated data (without and with scaling respectively) and plot c is for BUFR simulated data. . The speed binning is 1 m/s and the node binning is 1

If we compare the mean MLE behaviour as a function of wind speed between the real cases (Figures 8b and 19a) and the simulated cases (Figures 24b and 24c respectively), we see a large discrepancy at low speeds. In the real cases the MLE increases with decreasing speeds while in the simulated cases the MLE decreases with decreasing speeds.

The reason for this MLE increase in the real case is that the observation error (Kp) is underestimated for low wind speeds. From equation 1, an underestimation in the Kp (denominator term) will in turn produce an increase in the MLE. *Figa and Stoffelen* (2000) provide a physically-based model for the NSCAT backscatter observation error. They find that for low wind speed, the largest uncertainty lies in the spatial variability of the geophysical target. Since the different beam and polarization measurements in a WVC do not sample exactly the same area, the geophysical collocation error variability becomes substantial at low backscatter levels.

However, in the simulated case, the Kp is considered as a true value and therefore we would expect no increase or decrease in the MLE value at low wind speeds. This is not the case. The problem lies in the inversion and the fact that we assume that the measurement noise is proportional to the true value. The latter leads to a Kp which is proportional to σ_s^o (simulated σ^o from the GMF) in the denominator of equation 1. *Stoffelen (1998)* explains in page III-29 how proportional errors cause positive bias in the solution (after inversion). This positive bias will in turn produce a decrease in the MLE. Figure 25 illustrates the problem in the case of a two beam measurement system (QuikSCAT has four beams, but for simplicity we draw a 2D case). The solid curves represent the solution space. The diamond represents the pair of "true" measurements which are the starting point in the simulation process. The solid circle around the diamond represents the "true" measurement noise (Kp_t). Using this Kp_t we

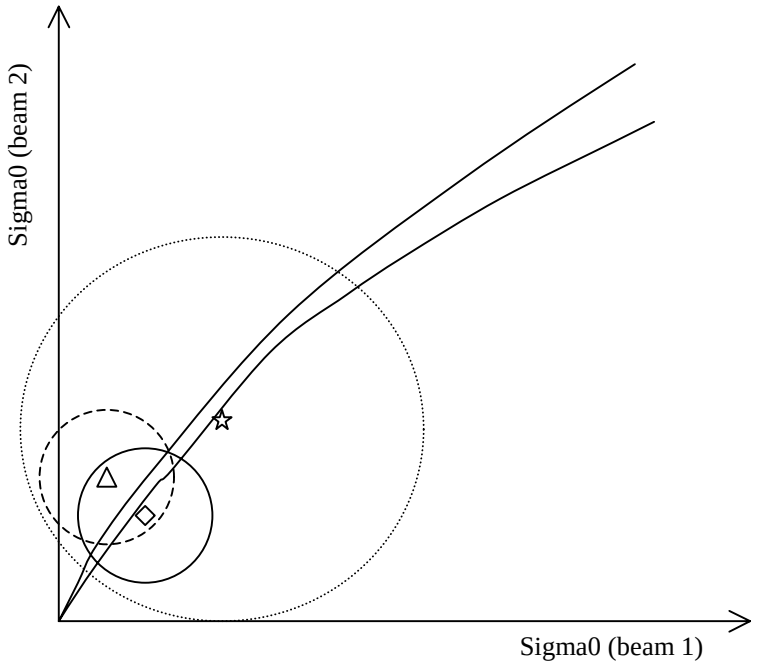


Figure 25 schematic illustration of the problem in a 2D measurement space

simulate the measurement pair (triangle inside the solid circle). The dashed circle represents its corresponding estimated noise (Kp_m). After inversion, we get a positively biased solution (star) which has its proportional noise (Kp_s) represented by the dotted circle. As Kp_s increases significantly, the MLE decreases (Kp_s is the denominator of equation 1) and this effect is more acute as we approach the origin corresponding to lower speeds.

Direct MLE comparison

Figure 26 shows the contour plot of the two-dimensional histogram of the simulated BUFR MLE versus the simulated HDF MLE. Although the distribution differs somewhat from the real case (Figure 21), it is clearly discernible that we have successfully reproduced the smearing effect by simply assuming a different number of measurements (more σ^o in HDF than BUFR).

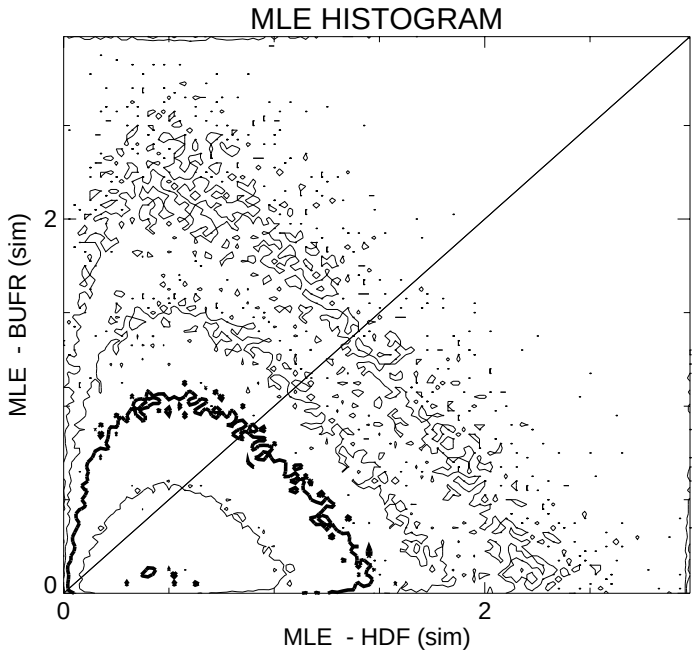


Figure 26 Contour plot of the two-dimensional histogram of the simulated BUFR MLE versus the simulated HDF MLE. The contour lines are in logarithmic scale (each step is half an order of magnitude).

The remaining difference between the simulated and the real distributions may be produced for many reasons. Here we address some of them :

- The simulation of the number of σ^o per WVC and beam in HDF is just a rough approximation. In the real data a WVC can contain a variable number of σ^o (see Figure 23) and in the simulation we have fixed this number.

- The different behaviour of the real and simulated MLEs at low speeds as discussed above (see Figures 8b, 19a, 24b and 24c), and due to ignoring wind variability, i.e. all σ^o in the WVC were produced from the WVC mean σ^o by assuming only Kp noise.

- In the real data, the HDF σ^o is an "egg" and the BUFR σ^o a weighted average of a composite (see chapter 2); in the simulation we have treated HDF as a weighted average of composites and therefore some differences in the simulated HDF measurement noise compared with the real HDF noise are expected.

There may be other reasons that could cause minor differences in the distributions. Nevertheless, it is clear from the results that having a different number of measurements (HDF versus BUFR) will produce uncorrelated MLE values (see Figure 21 or 26).

This smearing is however not causing a significant difference in the skill of BUFR QC with respect to HDF.

4.3 Quality Control with new GMF

Since May 2000, JPL wind retrievals are produced using a new GMF called QSCAT-1. This is the first empirically derived GMF from QuikSCAT measurements, as the one used up to now, NSCAT-2, was derived from NSCAT data.

If we invert winds using a different GMF, we will get different MLE values. Although these differences are not expected to be significant, it may well be that Quality Control is affected and therefore it needs to be revised.

Assuming no major changes, we first test Rn using the <MLE> surface computed with the NSCAT-2 GMF. The test is performed for two weeks of QuikSCAT HDF data.

4.3.1 Rn Characterization

As in the previous sections, we have collocated QuikSCAT data with ECMWF winds and SSM/I rain data and looked at the same kind of plots as in Figures 11 to 14 in order to characterize and validate Rn. The plots show very similar features as in section 4.1.3. The only difference is a slight increase in the Rn values.

Figure 27 shows the contour plot of a two-dimensional histogram of RMS-ECMWF against Rn. As in Figure 11 (same plot but for the NSCAT-2 GMF), the RMS-ECMWF increases (quality of data decreases) as Rn increases and the RMS-ECMWF

is increasing more rapidly with Rn at higher wind speeds (see plots b, c and d). However, when comparing both Figures, it is discernible that the distributions (see contour lines) in Figure 27 are slightly shifted towards higher Rn values compared to Figure 11. The shift is more significant at mid and high wind speeds (see plots c and d in both Figures).

As we have used the same <MLE> surface to compute the Rn, this shift means that in general MLE values coming out from the inversion are slightly higher using the QSCAT-1 GMF than using the NSCAT-2 GMF. This is an indication that the NSCAT-2 GMF better fits the backscatter measurements than the QSCAT-1 GMF.

Therefore, if we keep the same Rn threshold as in Section 4.1, we would expect to have more rejections in this case.

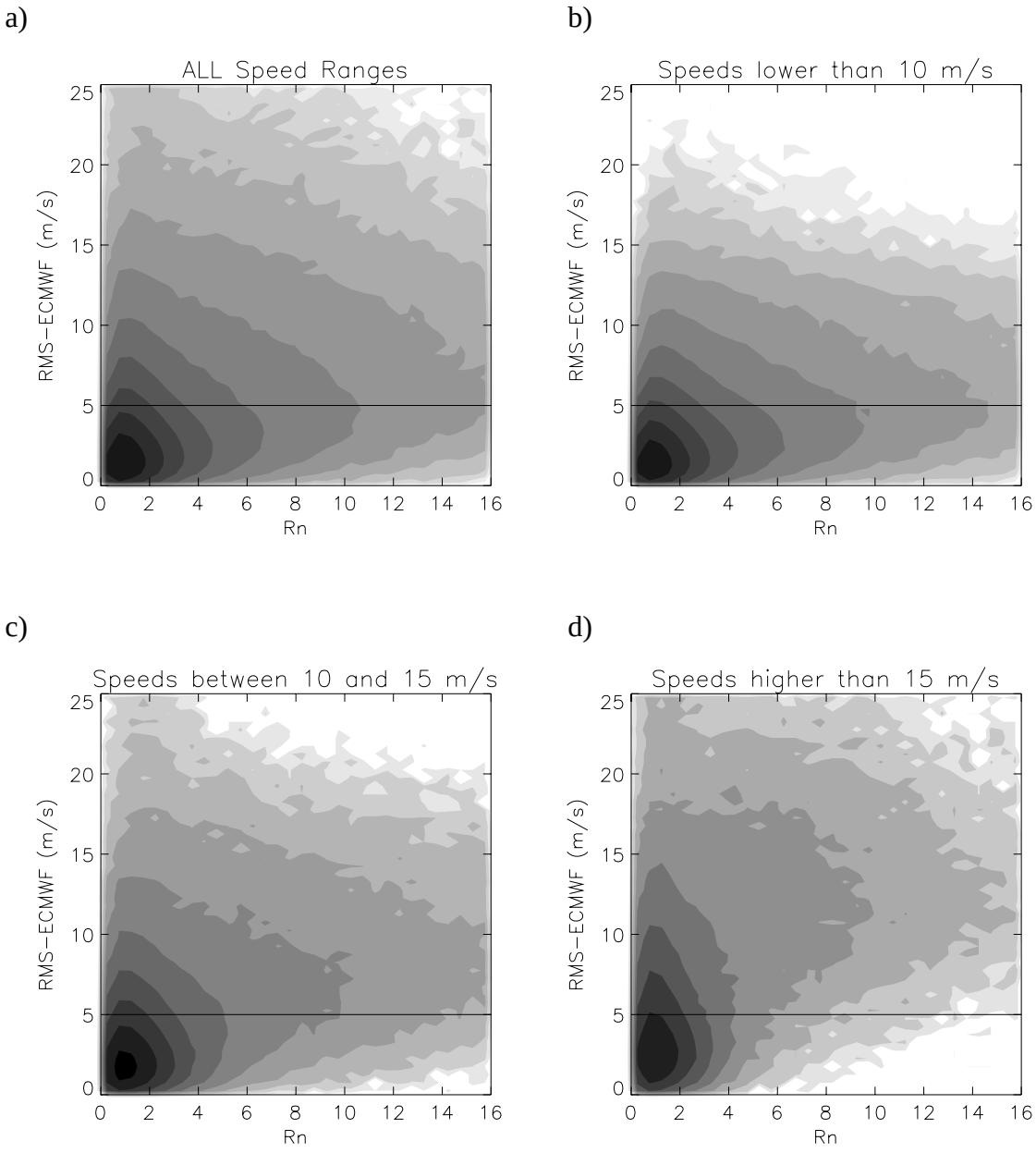


Figure 27 Same as Figure 11 but for the QSCAT-1 GMF.

4.3.2 Threshold Validation

We test the same Rn threshold as defined in section 4.1.4 against ECMWF and SSM/I collocations. The results for the sweet parts of the swath are shown in tables 13, 14 and 15, and the results for the nadir part are shown in tables 16, 17 and 18. Tables 13-18 are in the same format as tables 1-6 (see section 4.1.4) respectively.

Comparing table 13 to table 1, we appreciate a larger percentage of rejections in the former at all speeds, and more in particular at mid and high winds. From tables 14 and 2, we see a considerable amount of this excess of rejections concentrated below the RMS value of 5 m/s, which in turn makes the mean RMS-ECMWF slightly smaller. The mean RMS-ECMWF difference between accepted and rejected solutions is slightly smaller in table 14 (3.6 m/s) compared to table 2 (4 m/s). This indicates a somewhat better performance of the HDF QC using the NSCAT-2 GMF. However, this excess in rejections is positively contributing to rain detection. From tables 15 and 3, the amount of "rainy" WVCs rejected is higher in the former, especially at rain rates below 6 mm/hr where there is 9.7% more of rejections.

Comparing the HDF QC using the QSCAT-1 GMF in the nadir (tables 16-18) with the HDF QC using the NSCAT-2 GMF (tables 4-6), we can draw the same conclusions than for the sweet parts of the swath. In terms of quantitative results, the only significant differences with respect to the sweet parts are the following: the excess of rejections is 1.6 % in total while in the sweet parts is 1.3% (see tables 1, 13 and 16); and the excess of rejections for rain rates below 6 mm/hr is 6.6% compared to the 9.7% in the sweet parts (see tables 3, 15 and 18).

In general, the results show that with the QSCAT-1 GMF the QC rejects more data than with the NSCAT-2 GMF. However, this gives a positive impact on rain detection, especially for rain rates below 6 mm/hr. Therefore, we believe that using the same QC procedure (i.e. same <MLE> surface and Rn threshold) as in Section 4.1 for the new GMF (QSCAT-1) is appropriate and there is no need to tune the QC procedure to the new data.

TABLE 13

	Total	V<10	10≤V≤15	V>15
Num. Points (n/a)	5170647	3922963	982175	265509
Accepted (%)	93.0	95.2	88.9	77.1
Rejected (%)	7.0	4.8	11.1	22.9

TABLE 14

	RMS<5	RMS>5	Mean RMS (m/s)
Total (n/a)	4726341	444378	2.49
Accepted (%)	95.9	62.1	2.24
Rejected (%)	4.1	37.9	5.83

TABLE 15

	RR=0	0<RR≤6	RR>6
Num. Points (n/a)	1188320	89416	4742
Accepted (%)	95.7	60.9	9.0
Rejected (%)	4.3	39.1	91.0

TABLE 16

	Total	V<10	10≤V≤15	V>15
Num. Points (n/a)	3006927	2295287	555318	156322
Accepted (%)	92.1	95.0	85.9	71.1
Rejected (%)	7.9	5.0	14.1	28.9

TABLE 17

	RMS<5	RMS>5	Mean RMS (m/s)
Total (n/a)	2634399	372757	2.85
Accepted (%)	95.5	68.0	2.56
Rejected (%)	4.5	32.0	6.23

TABLE 18

	RR=0	0<RR≤6	RR>6
Num. Points (n/a)	670388	48638	3370
Accepted (%)	95.0	57.9	12.9
Rejected (%)	5.0	42.1	87.1

Note : RMS is referred 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.

4.4 KNMI QC versus JPL rain flag

Since May 2000, the QuikSCAT data products have a unique rain flag based on the combination of two parameters, the mp_rain_probability and the nof_rain_index developed by *Huddleston and Stiles* (2000) and *Mears et al.* (2000) respectively.

In this section, we test the JPL rain flag against our QC.

4.4.1 JPL Rain Flag Description

In January 2000, JPL incorporated in the QuikSCAT products two different rain flags based the mp_rain_probability and the nof_rain_index respectively. However, since May 2000, JPL merged both techniques into a single rain flag. This rain flag is actually based on the mp_rain_probability and the nof_rain_index is incorporated as an additional parameter in the mp_rain_probability procedure, called MUDH (Multidimensional Histogram) rain algorithm.

Briefly, mp_rain_probability is the probability of encountering a columnar rain rate that is greater than 2km*mm/hr. This probability value is read directly from a table based on eight input parameters including nof_rain_index and the average brightness temperature. The space spanned by these parameters can detect whether the set of σ^o used in wind retrieval contain a noteworthy component created by some physical phenomenon other than wind over the ocean's surface, assuming that the most likely phenomenon is rain.

The nof_rain_index is based upon a simplified version of the standard GMF to determine a MLE and a wind speed for each WVC. The MLE is based upon the sum of the squared differences between the set of σ^o that were used to retrieve winds and the corresponding model function σ^o that would generate the ambiguity with the greatest MLE. The wind speed is based upon a modified σ^o which is specifically calculated to be less sensitive to rain. The simplified MLE is normalized by a tabular empirical estimate for the 95^{th} percentile of the squared difference distribution. These tabular values are indexed by beam polarization, cross-track location in the measurement swath and wind speed. The normalized MLE is then divided by the number of σ^o in the WVC, multiplied by thirty, and rounded to the nearest integer value.

The nof_rain_index is most effective for wind speeds under 10 m/s and not very effective for wind speeds greater than 15 m/s.

The final rain flag deduced from the MUDH rain algorithm is also incorporated in the QuikSCAT products and can be found in the wvc_quality_flag variable.

4.4.2 Comparison

The JPL rain flag separates “rain” (rain rate above 2mm/hr) from “no rain” (rain rate below 2 mm/hr) cases and the KNMI QC separates cases of low quality to be rejected (above Rn threshold) from those of good quality to be accepted (below Rn threshold).

Both the JPL rain flag and the KNMI QC are meant to separate the usable data from the non-usable data. Therefore, the user should use only “no rain” data according to JPL rain flag and reject the “rain” data. In the same way, the user should accept or reject data according to KNMI QC, and therefore a study of the difference in behaviour of both procedures is of interest.

In order to make a consistent comparison we have processed two weeks of HDF data and classified the results in four different categories: A) “JPL Rain Flag - No Rain” and “KNMI QC - Accepted”; B) “JPL Rain Flag - Rain” and “KNMI QC - Accepted”; C) “JPL Rain Flag - No Rain” and “KNMI QC - Rejected”; and D) “JPL Rain Flag - Rain” and “KNMI QC - Rejected”. In line with the previous paragraph, categories A and D show similarities and categories B and C show discrepancies between the two procedures.

In tables 19-21 we present the results of collocating 2 weeks of QuikSCAT HDF data (sweet parts of the swath only) with ECMWF winds and SSM/I rain data. In total, there are about 5.2 million collocations with ECMWF and 1.1 million collocations with SSM/I. Following JPL’s definition (see previous section), we refer to rain data when SSM/I rain rate (RR) value is above 2 mm/hr, and to rain-free data when SSM/I rain rate value is below 2 mm/hr.

Table 19 shows by category the percentage of total data, the QuikSCAT mean speed, the ECMWF mean speed, the mean bias (QuikSCAT-ECMWF), the mean RMS-ECMWF, the percentage of data with rain (RR>2 mm/hr), and the percentage of all rain points (RR>2 mm/hr).

Results in table 19 show a very good agreement between both procedures as 94% of the data corresponds to categories A and D (91.1% in A and 2.9% in D). Moreover, category A shows good quality (0.5 m/s bias and 2.2 m/s RMS) rain-free (only 0.1% of data a rain contaminated) data while category D shows very low quality (5.1 m/s bias and 8.2 m/s RMS) and rain-contaminated (31.9% of data are rain contaminated) data.

Categories B and C contain 6% of the data and correspond to the differences in behaviour of both procedures.

Comparing both categories in terms of SSM/I rain detection, category B contains 13.9% of all the rain data while category C is only containing 2.4%. Therefore, JPL rain flag is more efficient since only 7.6% (5.2% in A and 2.4% in C) of all rain data is not rejected, while the KNMI QC is accepting 19.1% (5.2% in A and 13.9% in B).

In terms of quality of the data, both categories contain data of low quality, with similar bias (2.4 m/s in B and 1.7 m/s in C) and RMS (4.8 m/s in B and 4.1 m/s in C) values. The KNMI QC is more efficient in rejecting more low quality data than the JPL rain flag as there is a 4% of data in category C while only a 2% in category B. However, the JPL rain flag seems to work reasonably well as a Quality Control flag as categories B and D show that only 27% of that data (13.6% in B and 31.9% in D) are rain contaminated data and therefore the rest are rain-free but still low quality data.

TABLE 19

	JPL Rain Flag No Rain	JPL Rain Flag Rain
KNMI QC Accepted	Number of data (%): 91.1 QuikSCAT Mean Speed (m/s): 7.6 ECMWF Mean Speed (m/s): 7.1 Bias-ECMWF (m/s): 0.5 RMS-ECMWF (m/s): 2.2 Rain > 2 mm/hr (%[1]): 0.1 Rain > 2 mm/hr (%[2]): 5.2	Number of data (%): 2.0 QuikSCAT Mean Speed (m/s): 14.2 ECMWF Mean Speed (m/s): 11.8 Bias-ECMWF (m/s): 2.4 RMS-ECMWF (m/s): 4.8 Rain > 2 mm/hr (%[1]): 13.6 Rain > 2 mm/hr (%[2]): 13.9
KNMI QC Rejected	Number of data (%): 4.0 QuikSCAT Mean Speed (m/s): 9.1 ECMWF Mean Speed (m/s): 7.4 Bias-ECMWF (m/s): 1.7 RMS-ECMWF (m/s): 4.1 Rain > 2 mm/hr (%[1]): 1.0 Rain > 2 mm/hr (%[2]): 2.4	Number of data (%): 2.9 QuikSCAT Mean Speed (m/s): 12.3 ECMWF Mean Speed (m/s): 7.2 Bias-ECMWF (m/s): 5.1 RMS-ECMWF (m/s): 8.2 Rain > 2 mm/hr (%[1]): 31.9 Rain > 2 mm/hr (%[2]): 78.5

[1] : % of data in this category with rain (RR> 2 mm/hr)

[2] : % of all rain points (RR>2mm/hr)

Tables 20 and 21 are similar to table 19 but only for rain-free data and rain data respectively. Table 20 contains about 1.1 million data and table 21 about 17000 data.

Table 20 shows very similar results to table 19. The most significant result is that for rain-free data, categories B and D contain low quality data, as seen from the high bias (2.2 m/s in B and 4.4 m/s in D) and RMS (4.4 m/s in B and 7 m/s in D). This confirms the JPL rain flag as a Quality Control flag as well.

Table 21 shows clearly the effect of rain in the quality of the data. All categories have larger bias and RMS values compared to tables 19 and 20. In particular, category A contains 5.2% of rainy data, which are clearly of low quality (2.4 m/s bias and 5.5 m/s RMS). These data are detected neither by the JPL rain flag nor by the KNMI QC.

The results clearly show that category B contains low quality data, including a significant amount of rainy data. Therefore, it seems a good idea to incorporate the JPL rain flag to the KNMI QC in order to improve the Quality Control of QuikSCAT

data. However, according to the results in these three tables, ECMWF winds in category B are in general 3-5 m/s higher than in the rest of the categories. This means that category B corresponds to dynamically active situations. Therefore, it could well be that this category systematically corresponds to frontal or low-pressure system areas where the discrepancy between ECMWF and QuikSCAT is indeed of valuable interest and therefore we want this data to be kept and not rejected.

In order to determine the convenience of incorporating the JPL rain flag in the KNMI QC, some meteorological cases need to be examined.

TABLE 20

	JPL Rain Flag No Rain	JPL Rain Flag Rain
KNMI QC Accepted	Number of data (%): 92.7 QuikSCAT Mean Speed (m/s): 7.0 ECMWF Mean Speed (m/s): 6.5 Bias-ECMWF (m/s): 0.5 RMS-ECMWF (m/s): 2.0	Number of data (%): 1.3 QuikSCAT Mean Speed (m/s): 12.2 ECMWF Mean Speed (m/s): 10.0 Bias-ECMWF (m/s): 2.2 RMS-ECMWF (m/s): 4.4
KNMI QC Rejected	Number of data (%): 3.5 QuikSCAT Mean Speed (m/s): 7.8 ECMWF Mean Speed (m/s): 5.9 Bias-ECMWF (m/s): 1.9 RMS-ECMWF (m/s): 4.2	Number of data (%): 2.5 QuikSCAT Mean Speed (m/s): 10.3 ECMWF Mean Speed (m/s): 5.9 Bias-ECMWF (m/s): 4.4 RMS-ECMWF (m/s): 7.0

TABLE 21

	JPL Rain Flag No Rain	JPL Rain Flag Rain
KNMI QC Accepted	Number of data (%): 5.2 QuikSCAT Mean Speed (m/s): 10.8 ECMWF Mean Speed (m/s): 8.4 Bias-ECMWF (m/s): 2.4 RMS-ECMWF (m/s): 5.5	Number of data (%): 13.9 QuikSCAT Mean Speed (m/s): 13.7 ECMWF Mean Speed (m/s): 9.0 Bias-ECMWF (m/s): 4.7 RMS-ECMWF (m/s): 8.2
KNMI QC Rejected	Number of data (%): 2.4 QuikSCAT Mean Speed (m/s): 9.9 ECMWF Mean Speed (m/s): 6.6 Bias-ECMWF (m/s): 3.3 RMS-ECMWF (m/s): 6.1	Number of data (%): 78.5 QuikSCAT Mean Speed (m/s): 14.4 ECMWF Mean Speed (m/s): 6.4 Bias-ECMWF (m/s): 8.0 RMS-ECMWF (m/s): 11.2

4.4.3 Cases

In this section, we show a few wind field examples where both the KNMI QC and the JPL rain flag have been applied. Figures 28 and 29 show QuikSCAT winds collocated with ECMWF winds. The arrows in plot a) correspond to the QuikSCAT JPL-selected wind solutions and the colors represent categories A (green), B (yellow), C (purple) and D (red). The arrows in plot b) correspond to the collocated ECMWF winds. Plot c) is the same as plot a), but arrows corresponding to categories C and D are substituted by dots.

In Figure 28, the presence of a low-pressure system to the Northwest of Ireland is clearly discernible in the mid-left part of the plot. A wind front is also visible going from East to Southwest of the low. The KNMI QC has rejected many data in the vicinity of the low and along the front line where confused sea state is expected (see red and blue arrows). There are also some rejections at the right side of the front probably due to rain (see inconsistent flow of red arrows in the mid-right part of the plot) and in low wind areas (blue arrows at the bottom-right area of the plot). However, category-B winds (yellow arrows) are mainly focused in the most dynamical area as anticipated in the previous section.

If we zoom into this area (plot c) and show only categories A and B (accepted winds after KNMI QC), we see that most of the yellow arrows show a spatially consistent flow which we would like to keep. However, there are still a few undesirable yellow arrows in the right side of the plot (most likely rain contaminated) and along the front line (most likely low quality winds).

In Figure 29, we see a low-pressure system Northeast of Florida. The red arrows in the centre and bottom of plot a) clearly show the presence of a wind front with associated rain bands. If we look at the same plot but only for the accepted data after KNMI QC (plot c), we still see some inconsistent yellow arrows, which are most likely rain contaminated and therefore undesirable.

From these meteorological cases, we can conclude that category B winds are primarily located in dynamically active areas and in many cases they show very consistent wind flows. However, there are also several rain-contaminated cases and some confused sea state cases which lead to low quality winds in category B.

From plot c of both Figures, it is clearly discernible that rejecting category B winds is not reducing much the synoptic-scale information content on the meteorological situations and therefore can still be of very valuable use to meteorologists or NWP models.

In conclusion, the JPL rain flag and the KNMI QC are complementary and their combined use is recommended.

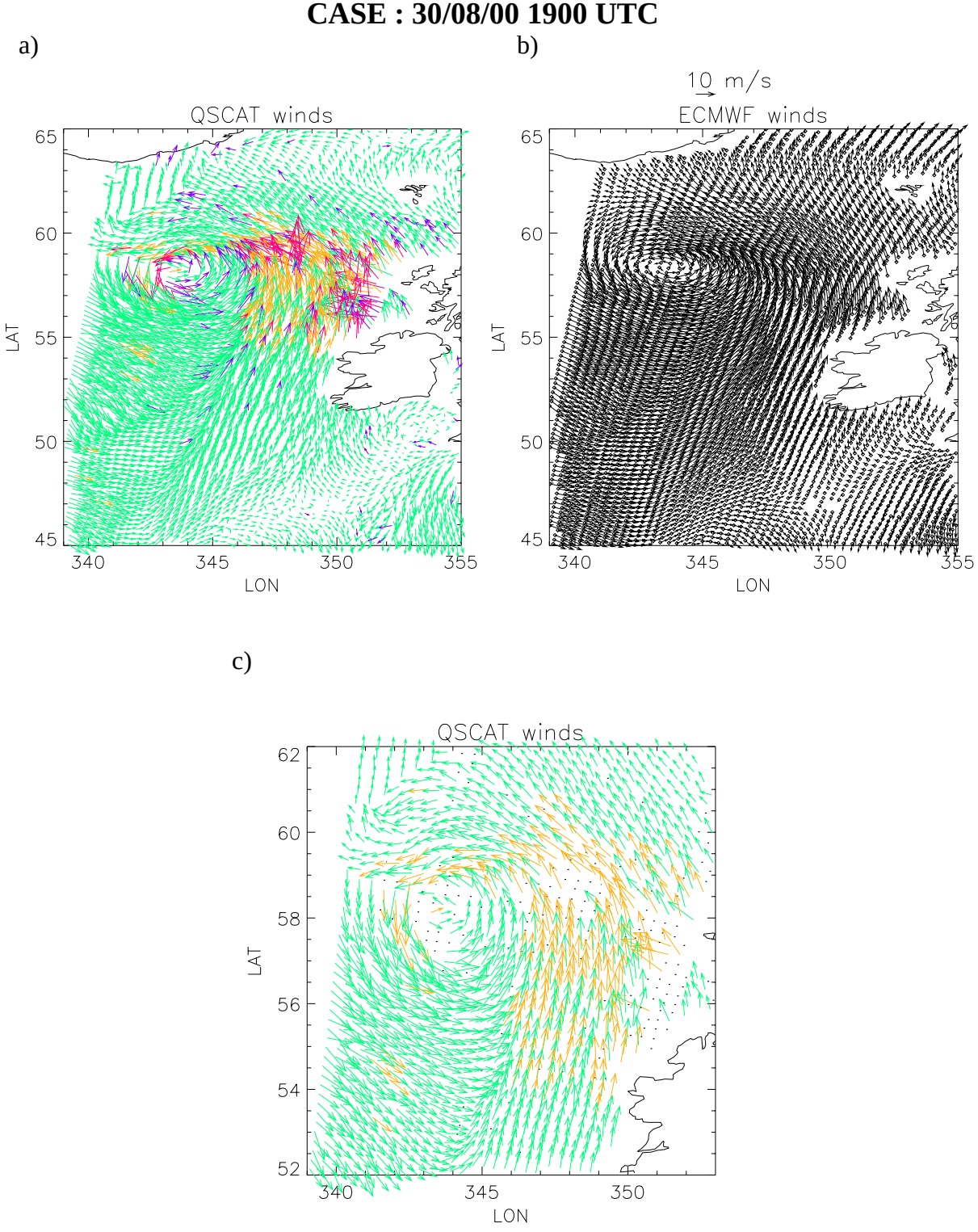


Figure 28 QuikSCAT wind fields (plots a and c) and collocated ECMWF winds (plot b). The colors represent the different categories: green is category A, yellow is B, purple is C, and red is D. The acquisition date was August 30 2000 at 19 hours UTC.

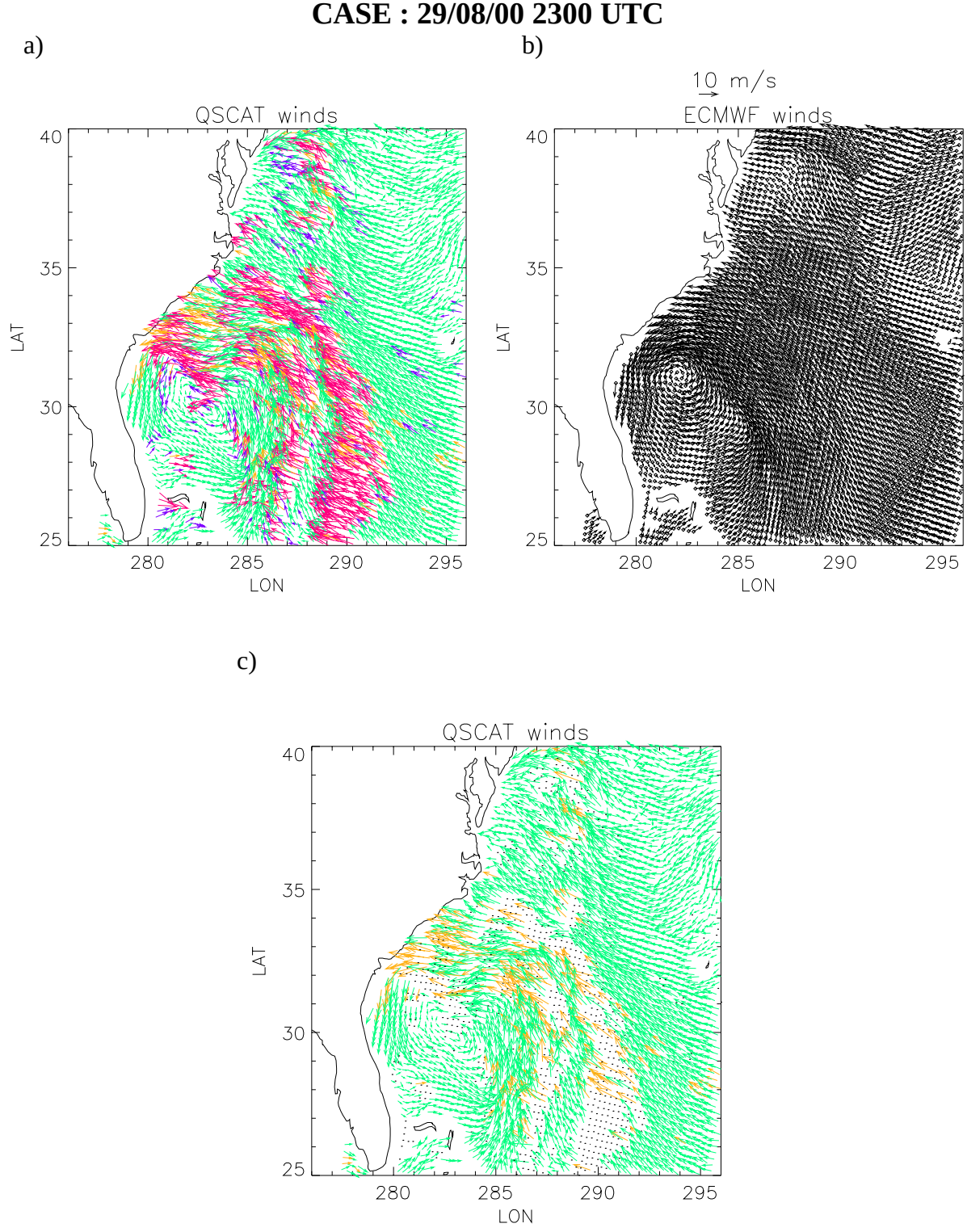


Figure 29 Same as Figure 28 but for different date (August 29 2000 at 23 hours UTC) and location.

5 Conclusions and Future Work

The KNMI inversion scheme for QuikSCAT adapted from ERS and NSCAT processing software with help from ECMWF seems to provide good results. However, further refinement of the scheme is anticipated. The software was used to validate the simulated sample HDF and BUFR products and the real HDF and BUFR QuikSCAT data. Some unwanted features in the data have been reported to JPL and corrected by them.

The QC by normalized residual (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 volumetric rain absorption. As such, QC rejection for ERS is only activated in case of confused sea state, ice occurrence, etc, whereas in the case of QuikSCAT and NSCAT also rain is eliminated.

Collocations of QuikSCAT data with ECMWF winds and SSM/I rain were used to characterize and validate the QC by (Rn) both for HDF and BUFR formats.

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 the retrieved wind speed (data quality becomes poor at lower Rn values when retrieved speeds are higher).

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 of typically 15-20 m/s.

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). For HDF, 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 entire inner swath although the skill is slightly better in the sweet regions than in the nadir region.

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

The BUFR and HDF MLEs appear to be nearly uncorrelated. The consequent different behaviour of the BUFR Rn compared to the HDF Rn is not significantly affecting the skill of the QC procedure nor the quality of the retrieved winds.

As a result from simulating both the HDF and the BUFR products, we show that the MLE decorrelation between the two products is mainly caused by the fact that the BUFR product contains a fewer number of observations per WVC than HDF.

The QC procedure derived for the NSCAT-2 GMF is tested with the new GMF (QSCAT-1). The results show a slight increase of Rn at all speeds, indicating that the former GMF better fits the observed QuikSCAT backscatter measurements. The larger Rn leads to more rejections in general compared to the NSCAT-2 GMF. This excess in rejections however produces a positive impact in rain detection, especially at rain rates below 6 mm/hr. The same QC procedure (i.e. same <MLE> surface and Rn threshold) can therefore be used with the new QuikSCAT data (produced with the QSCAT-1 GMF).

The JPL rain flag is tested against our (KNMI) QC. It turns out that both procedures are complementary. The KNMI QC detects a 4% of low-quality and almost rain-free data, which are not detected by the JPL rain flag. On the other hand, the JPL rain flag detects a 2% of low-quality and partially rain-contaminated data, which are not detected by the KNMI QC. Therefore, the JPL rain flag will be incorporated in the KNMI QC procedure.

As seen in chapter 3, the wind retrieval skill decreases in the outer and nadir regions of the swath in comparison to the NSCAT-like or sweet swath. We plan to work on the inversion problem in these parts of the swath to improve the current wind retrieval and assimilation skills.

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 decide whether to accept or to reject a particular wind solution and therefore improve the ambiguity removal. Quantitative tests are being developed in the context of the NWP Science Application Facility project to exploit this Rn-to-probability relationship. In particular, *Stoffelen et al.* (2000) have worked on the definition of a cost function based on Rn for the assimilation of QuikSCAT data into NWP models.

We plan to extend our QC and cost function to work at lower resolutions (50 and 100 km) to increase data quality for a reliable use in NWP data assimilation and for presentation to meteorologists for nowcasting and short-range forecasting purposes.

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.

Appendix A

In order to fit a 2D function to the filtered mean MLE surface, we first fit a function for the MLE dependence on wind speed at a certain node. Then we assume 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 30a shows the fit of the filtered mean 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 30b shows the averaged MLE over all wind speeds and normalized with the speed dependent function (fit on Figure 30a) 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 2D 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.

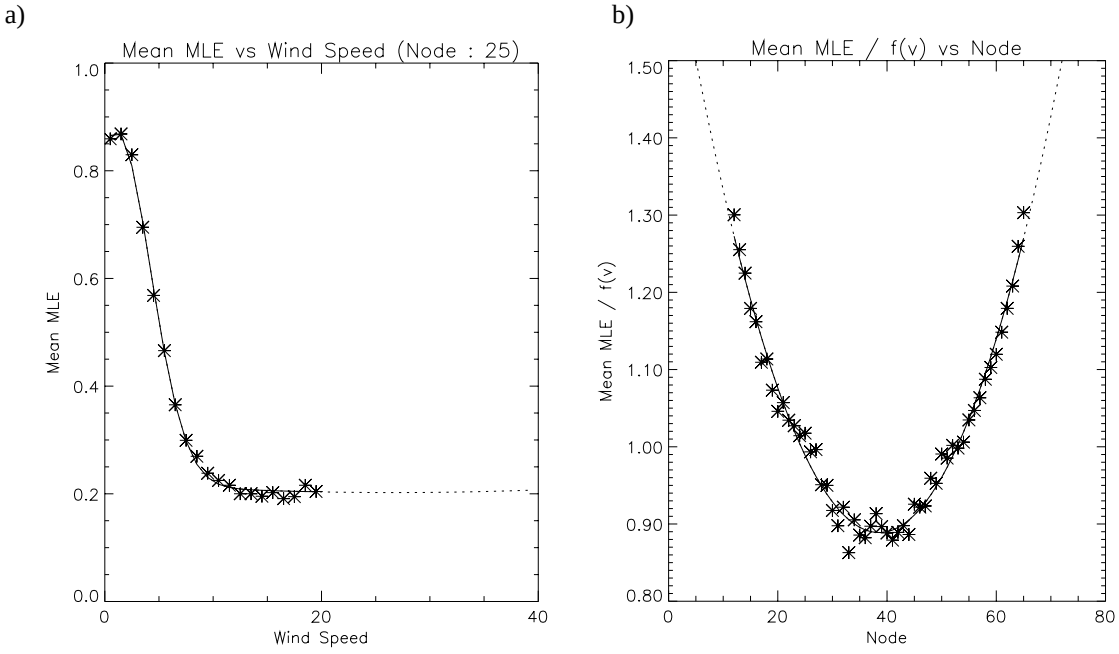


Figure 30 a) Filtered mean MLE versus wind speed for node number 25 (stars), where the solid line shows the function fit and the dotted line the extrapolation for wind speeds higher than 20 m/s; b) averaged MLE over all wind speeds and normalized with the speed dependent function (fit on Figure 18a) 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.

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

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

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

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

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

The coefficient values are the following :

A_0 = 0.78519; A_1 = 1.47396; A_2 = 2.91577

A_3 = 0.31881; A_4 = -4.2426E-3; A_5 = 6.9633E-5

B_0 = 1.37840; B_1 = -0.02713; B_2 = 3.4853E-4

Appendix B

Looking at the filtered mean MLE surface (see Figure 19a), it is clearly discernible that assuming a constant shape of the MLE dependence on wind speed over all nodes of the inner swath (as in Appendix A) is not valid anymore.

In this case, we fit a Gaussian + 2nd order polynomial function to the filtered mean MLE for each node of the inner swath separately. Then, we fit a 2nd order polynomial function to the evolution of each coefficient of the previous function with respect to the node number. Therefore, the 2D function which fits the filtered mean MLE surface is the following:

$$< MLE >_{fit} = A_0 \cdot e^{-\frac{1}{2}\cdot(\frac{v-A_1}{A_2})^2} + A_3 + A_4 \cdot v + A_5 \cdot v^2 \qquad \text{(B1)}$$

$$A_i = A_{i0} + A_{i1} \cdot n + A_{i2} \cdot n^2 \quad ; \ \forall i \in [0,5], \forall n \in [1,76] \qquad \text{(B2)}$$

where <*MLE*> is the expected MLE, *v* is the wind speed and *n* the node number.

The coefficient values are the following:

A_{00} = 0.55000; A_{01} = 0.00000; A_{02} = 0.00000

A_{10} = 1.50000; A_{11} = 0.00000; A_{12} = 0.00000

A_{20} = 2.75000; A_{11} = 0.00000; A_{12} = 0.00000

A_{30} = 0.21210; A_{21} =-0.00249; A_{22} = 3.02E-5

A_{40} =-0.00741; A_{31} = 0.00031; A_{32} =-4.08E-6

A_{50} = 0.00012; A_{41} =-4.75E-6; A_{42} = 6.24E-8

6 Acronyms

ADEOS	Advanced Earth Observation Satellite
AER	Atmospheric Environmental Research Inc.
BUFR	Binary Universal Format Representation
DMSP	Defense Meteorological Satellite Program
ECMWF	European Centre for Medium-range Weather Forecasts
ERS	European Remote Sensing Satellite
EUMETSAT	European Meteorological Satellite Organization
GMF	Geophysical Model Function
HDF	Hierarchical Data Format
H-pol	Horizontal polarization measurement
JPL	Jet Propulsion Laboratories
KNMI	Royal Netherlands Meteorological Institute
MLE	Maximum Likelihood Estimator
NASA	National Air and Space Administration (USA)
NCEP	National Centre for Atmospheric Prediction (USA)
NOAA	National Oceanographic and Atmospheric Administration (USA)
NSCAT	NASA Scatterometer
NWP	Numerical Weather Prediction
RMS	Root-Mean-Squared
SAF	Satellite Application Facility
SSM/I	Special Sensor Microwave Imager
STD	Standard Deviation
SWT	Science Working Team
V-pol	Vertical polarization measurement
WVC	Wind Vector Cell

7 References

Boukabara, S.A., Hoffman, R.N., and Grassotti, C., "*Atmospheric Compensation and Heavy Rain Detection for SeaWinds Using AMSR*", Atmospheric Environmental Research Inc., 1999.

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, July 2000.

Huddleston, J.N., and Stiles, B.W., "*Multidimensional Histogram (MUDH) Rain Flag*", Product Description Version 2.1, available at http://podaac-www.jpl.nasa.gov/quikscat/, September 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, July 2000.

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

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

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

Mears, C., Wentz, F., and Smith, D., "*SeaWinds on QuikSCAT Normalized Objective Function Rain Flag*", Product Description Version 1.2, available at http://podaac-www.jpl.nasa.gov/quikscat/, September 2000.

Rohn, M., Kelly, G., and Saunders, R., "*Experiments with Atmospheric Motion Vectors at ECMWF*", in Proc. of Fourth International Winds Workshop, EUM P24, ISSN 1023-0416, pp. 139-146, October 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, January 1997.

Stoffelen, A., "*Scatterometry*", PhD thesis ISBN 90-393-1708-9, October 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*", Report for the Beleidscomissie Remote Sensing, contract NRSP-2/97-06, 1997.

Stoffelen, A., de Vries, J., and Voorrips, A., "*Towards the Real-Time Use of QuikSCAT Winds*", BCRS Final Report, September 2000.

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*", in Proc. of QuikSCAT Cal/Val Workshop, Pasadena/Arcadia (USA), November 1999.