

# Characterization of Residual Information for SeaWinds Quality Control<sup>1</sup>

M. Portabella, A. Stoffelen

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

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

e-mail: [portabel@knmi.nl](mailto:portabel@knmi.nl), [stoffelen@knmi.nl](mailto:stoffelen@knmi.nl)

## Abstract

Recent work has shown the important properties of the wind inversion residual or Maximum Likelihood Estimator (MLE) for Quality Control (QC) of QuikSCAT Hierarchical Data Format (HDF) observations. Since March 2000, the QuikSCAT Near-Real Time (NRT) Binary Universal Format Representation (BUFR) product is available. As this product is used for Numerical Weather Prediction (NWP) assimilation purposes, a QC procedure for the BUFR product is needed.

We study the behavior of the MLE in order to determine whether the HDF QC procedure is appropriate for BUFR data. A comparison using real HDF and BUFR data reveals that the MLE distributions of HDF and BUFR differ and are actually poorly correlated. One important difference between BUFR and HDF is the amount of signal averaging prior to wind inversion.

---

<sup>1</sup> *IEEE Trans. Geosci. Rem. Sens.*, Vol. 40, No.12, pp. 2747-2759, 2002. © Institute of Electrical and Electronics Engineers.

The averaging reduces the number of observations used in the wind retrieval for the BUFR product as compared to HDF. We show with a simple example that different MLE distributions are indeed expected due to this averaging. We also run a simulation in order to link theory and reality and better understand the behavior of the MLE.

Despite the different MLE behavior in BUFR and HDF, the quality of the retrieved winds, as compared with the European Centre for Medium-Range Weather Forecasts (ECMWF) winds, is very similar. We develop an MLE-based QC procedure for BUFR similarly to the one in HDF, and we compare both. The skill of the QC in BUFR is again very similar to the one in HDF, showing that despite the different MLE behavior in both formats, the properties of the MLE as a Quality Control indicator remain very similar.

## **1 Introduction**

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 is able to provide accurate winds over the ocean surface and can potentially contribute to improve the situation for tropical and extratropical cyclone prediction ([1], [2], and [3]).

The impact of observations on weather forecast often critically depends on the Quality Control (QC) applied. For example, [4] show a positive impact of cloud motion winds on the

European Centre for Medium-Range Weather Forecasts (ECMWF) model after QC, while the impact is negative without QC. This is also applicable for scatterometer data. Besides its importance for NWP, in applications such as nowcasting and short-range forecasting, the confidence of meteorologists in the scatterometer data is boosted by a better QC. Therefore, in order to successfully use scatterometer data in any of the mentioned applications, a comprehensive QC needs to be done in advance.

Recent and past work on scatterometer QC points out the wind retrieval residual as a good QC indicator. [5], [6], and [7] use a method to detect and reject Wind Vector Cells (WVC) with poor quality wind information using a residual-based parameter for ERS, NSCAT and SeaWinds respectively. Additional work on the relevance of the wind retrieval for QC purposes can be found in [8].

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

The SeaWinds data are first processed at the Jet Propulsion Laboratory (JPL) and then distributed in two different formats. The QuikSCAT Hierarchical Data Format (HDF) product is directly distributed by JPL and the near-real time Binary Universal Format Representation (BUFR) product is distributed by the National Oceanic and Atmospheric Administration (NOAA). Since the first data were distributed only in HDF format, [7] performed a QC procedure for QuikSCAT in this data format. However, since March 2000, the QuikSCAT

near-real time BUFR product is available. As this product is the one used for assimilation purposes, a QC procedure needs to be set for BUFR data.

Figure 1 shows the contour plot of the two-dimensional histogram of the BUFR MLE versus the HDF MLE. The plot shows, as expected, small correlation (around 0.5) between both MLE distributions. Moreover, the mean BUFR MLE value (0.28) is significantly smaller than the mean HDF value (0.57). It may be clear from these results that in order to determine whether to use a similar QC procedure for BUFR than the one used for HDF data or not, a comprehensive characterization of the MLE is needed.

In section 2, the SeaWinds instrument and data are presented. In section 3, we provide an explanation for MLE differences observed by presenting a theoretical example, a simulation study, and a detailed analysis of the real data. Then, the influence of the MLE differences in wind retrieval and QC is tested in section 4. Finally, conclusions and recommendations are presented in section 5.

## **2 Instrument and Data**

The SeaWinds instrument onboard QuikSCAT satellite (launched in June 19, 1999) is a conical-scanning pencil-beam scatterometer. It uses a rotating 1-meter dish antenna with two spot beams, an H-pol beam and a V-pol beam at incidence angles of  $46^\circ$  and  $54^\circ$  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.

The SeaWinds swath is divided into equidistant across-track WVCs or nodes numbered from left to right when looking along the satellite's propagation direction. The nominal WVC size is 25 km x 25 km, and all backscatter measurements centered in a WVC are used to derive the WVC wind solutions. Due to the conical scanning, a WVC is generally viewed when looking forward (fore) and a second time when looking aft. As such, up to four measurement classes (called "beam" here) emerge: 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 46° 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 [9], [10], and [11].

As mentioned in section 1, the QuikSCAT data are distributed in two different formats: HDF and BUFR. In order to characterize the MLE for QC purposes, 3 days of QuikSCAT sweet swath (WVC numbers 9 to 28 and 49 to 68) data in both data formats is used.

The main difference between the HDF and the BUFR products is related to the spatial resolution of  $\sigma^{\circ}$ . In the BUFR product, the  $\sigma^{\circ}$  of a particular beam (fore-inner, fore-outer, aft-inner, aft-outer) is an average of all  $\sigma^{\circ}$ s of that particular beam which fall in the same WVC. In HDF, all separate  $\sigma^{\circ}$ s are provided.

The SeaWinds  $\sigma^{\circ}$ s can be either "eggs" or slices. In a particular WVC, an "egg"  $\sigma^{\circ}$  is the radar backscatter from the whole pulse or footprint whose centre falls in that WVC. The "egg" can be subdivided in individual range-sampling 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

$\sigma^\circ$ . 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 sampling 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 all “egg”  $\sigma^\circ$ s in a WVC is approximately 40 km. Composites enhance the wind resolution mainly in range direction down to 25 km, and have little effect on azimuth resolution.

In order to characterize and validate the QC procedure for BUFR data, we collocate the QuikSCAT data with ECMWF winds and Special Sensor Microwave Imager (SSM/I) rain data. The QuikSCAT data correspond to the preliminary science data products produced by JPL using the NSCAT-2 GMF. [Note that the latest version of QuikSCAT processing uses QSCAT-1 GMF. [12] show that NSCAT-2 better fits the  $\sigma^\circ$  measurements than QSCAT-1, i.e. the mean MLE for NSCAT-2 is lower than for QSCAT-1]

We use the analyses, and 3-hour, 6-hour, and 9-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 0.25° 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.

### 3 MLE Characterization: Effects of $\sigma^o$ Averaging on its Distribution

The wind retrieval minimizes the MLE or residual, which is for SeaWinds defined as (adopted from [10]):

$$MLE = \frac{1}{N} \sum_{i=1}^N \frac{(\sigma_{mi}^o - \sigma_{si}^o)^2}{kp(\sigma_{si}^o)} \quad (1)$$

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

[5] interpret the MLE as a measure of the (squared) distance between a set of  $\sigma_{mi}^o$  and the solution  $\sigma_{si}^o$  set lying on the GMF surface in a transformed measurement space where each axis of the measurement space is scaled by  $Kp(\sigma_{si}^o)$ .

As mentioned in Section 1, the MLE is a good QC indicator in scatterometry. In the case of QuikSCAT, a MLE-based method successfully quality controls the HDF data ([7]). Therefore, if a QC procedure needs to be set for BUFR data, a closer look into the MLE distribution is needed.

As discussed in Section 2, a BUFR  $\sigma^o$  is an average of all the HDF  $\sigma^o$  measurements per beam in a WVC. From a theoretical point of view, the HDF and BUFR MLE distribution characteristics may differ just by using a different number of  $\sigma^o$  in their computation. In this section, we show this with a simple example. In order to validate the theoretical results, a

simulation is performed to bridge the gap between theory and reality. Finally, the remaining differences between real and simulated differences are analyzed in detail.

### 3.1 Theoretical case

This case corresponds to a simplified version of the MLE, which uses the following set of assumptions:

- 1) Typically for SeaWinds, HDF contain  $N > 4 \sigma^0$  per WVC while BUFR contain  $M = 4 \sigma^0$  per WVC in the inner swath. Therefore, the MLE is computed for HDF in a higher dimensional measurement space than for BUFR. In this example we simplify the problem assuming  $N = 2$  and  $M = 1$ .
- 2) The truth or solution lies in the origin of our measurement space for simplicity.
- 3) Since in scatterometry the errors are considered Gaussian, we consider pairs of measurements  $(x, y)$  in the  $N (= 2)$ -dimensional space as Gaussian distributed points around the origin (solution). Therefore, we use the following two-dimensional Gaussian Probability Density Function (PDF):

$$p(x, y) dx dy = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2} \frac{(x^2 + y^2)}{\sigma^2}} dx dy, \quad (2)$$

where the standard deviation in both axis is assumed identical, i.e.  $\sigma = \sigma_x = \sigma_y$ .

- 4) For simplicity, we also assume constant  $K_p$  noise values for both HDF and BUFR. As such, the MLE is equivalent to a squared distance to the origin weighted by a constant



factor. Moreover, this assumption is important since it will allow us to show the significant change between the mentioned MLE distributions just by setting  $N > M$ .

### ***Mathematical demonstration***

In order to show the difference between two distributions, we use the following mathematical definitions:

- The mean or expected value of a function  $f(x,y)$  is defined in terms of the PDF  $p(x,y)$  by

$$E(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \cdot p(x,y) \cdot dx \cdot dy \quad (3)$$

- The standard deviation (SD) of a function  $f(x,y)$  is defined in terms of the PDF  $p(x,y)$  by

$$SD(f) = \sqrt{VAR(f)} = \sqrt{E(f^2) - (E(f))^2} \quad (4)$$

where  $VAR$  is the variance.

- Finally, the correlation between two functions  $f(x,y)$  and  $g(x,y)$  is defined by

$$COR(fg) = \frac{(E(fg) - E(f) \cdot E(g))}{SD(f) \cdot SD(g)} \quad (5)$$

Using the above assumptions, we can rewrite equation 1 for both HDF and BUFR cases:

- a) HDF ( $N=2$ )

Using the above assumed measurement noise, i.e.  $\sigma$  (see assumption 4), the MLE in the 2D case is:

$$MLE_{2D} = \frac{x^2 + y^2}{2\sigma^2} \quad (6)$$

b) BUFR (M=1)

In this case, the measurement is an average of the two measurements,  $z = \left(\frac{x+y}{2}\right)$ , with  $z = 0$  as solution. The measurement noise can be easily computed using equations 2, 3, and 4 and has the following value:  $Kp_z = (SD(z))^2 = \sigma^2 / 2$ . Thus, the MLE in the 1D case is:

$$MLE_{1D} = \frac{z^2}{Kp_z} = \frac{x^2 + y^2}{2\sigma^2} + \frac{x \cdot y}{\sigma^2} \quad (7)$$

In order to show that the distributions of  $MLE_{1D}$  and  $MLE_{2D}$  (analogous to MLE in BUFR and HDF, respectively) differ, we compute their mean values, standard deviations and the correlation using the above mathematical definitions (equations 3, 4, and 5).

The results show that reducing the number of dimensions from two to one in the observational space by averaging the observations, does not affect the mean MLE value ( $E(MLE_{1D})=E(MLE_{2D})=1$ ) but produces an increase in the SD of the MLE distribution ( $SD(MLE_{1D})=1$  while  $SD(MLE_{2D})=\sqrt{2}$ ). Moreover, there is a clear decorrelation between the 2D and the 1D MLE distributions ( $COR(MLE_{1D}MLE_{2D})=1/\sqrt{2} \approx 0.7$ ). Therefore, it is clear that the distributions differ.

In Figure 1, we directly compare the MLE distributions from the HDF and BUFR products for the entire set of 3 days of QuikSCAT HDF data collocated with the QuikSCAT BUFR data. The contour plot of the two-dimensional histogram of the BUFR MLE versus the HDF MLE shows indeed small correlation between both MLE distributions, as presented in the introduction. However, the results differ somewhat from the theoretical example. The

correlation is smaller in the real case and the ratio between the mean values of HDF and BUFR distributions is 1 in the case of the theoretical example while 2 in the real case where the mean BUFR MLE value (0.28) is substantially smaller than the mean HDF value (0.57). This can be expected, since the theoretical example is just a simplification of the problem as discussed above. In order to understand the real results in more detail, a simulation is needed.

### **3.2 MLE Simulation**

In the simple example that we theoretically solve in Section 3.1, we show that the small correlation between HDF and BUFR MLE distributions is due to the  $\sigma^0$  averaging, assuming two measurements for HDF and one for BUFR. However, in the real case, where HDF contains more than four measurements and BUFR typically four, the results, although similar, present some differences with respect to the theoretical case. In particular, the correlation is significantly smaller (0.5) compared to the simple theoretical example (0.7).

In this section, we simulate HDF and BUFR MLEs, assuming a realistic number of measurements for both sets. The simulation is done to show that the theoretical demonstration can be extrapolated to the real case by using a larger number of  $\sigma^0$  in both HDF and BUFR products. We also simulate the effects of varying the number of  $\sigma^0$  on the MLE distributions.

### 3.2.1 Simulation procedure

We use the JPL selected winds of the BUFR files as truth to simulate two sets of measurements. The first set simulates the HDF product, using realistic noise values and number of measurements per beam. Then, similar to the real data, these measurements are averaged per beam to generate the second set which simulates the BUFR product. The more HDF observations per beam in a particular WVC that we simulate, the larger the measurement noise that we assume for each individual measurement, such that the information content is the same in each simulated HDF and BUFR WVC. Once we have simulated both sets of measurements, we invert them, using equation 1, to derive the simulated MLE.

#### *Number of $\sigma$*

In order to adequately simulate both products we have to use a realistic number of  $\sigma$  per WVC. In the case of the HDF simulation, we produce a variable number of measurements depending on the WVC number and beam.

Figure 2 shows the histogram of the number of measurements per WVC and beam for one day of HDF data. [Note that because of symmetry, both the fore-beam and the aft-beam histograms are identical; therefore, only one of them is shown in the Figure]. Plot a corresponds to WVC number 12 and plot b to WVC number 55. 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. However, to simplify the

simulation, we have chosen the number of  $\sigma^\circ$  corresponding to the peak of each distribution as the fixed value which will represent the number of  $\sigma^\circ$  for each particular WVC and beam.

As explained above, the BUFR measurements are produced by averaging the HDF measurements per beam. Therefore, the number of  $\sigma^\circ$  in BUFR will depend on the number of  $\sigma^\circ$  in HDF. Since we perform this simulation in the sweet parts of the swath, we use 4  $\sigma^\circ$  per WVC for BUFR.

### 3.2.2 Simulation results

In order to provide a realistic simulation, we use the Kp and the wind distributions as provided in the JPL product together with the realistic number of measurements for HDF and BUFR computed in section 3.2.1.

Figure 3 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 1), it is clear that we have successfully reproduced the same small correlation (about 0.5 in both cases) by simply assuming a different number of measurements (more  $\sigma^\circ$  in HDF than BUFR). The remaining differences between the simulated and the real distributions, which can be explained by many issues, are analyzed in detail in section 3.3. Nevertheless, it is clear from the results that the simulation is a good reflection of reality. Therefore, since averaging  $\sigma^\circ$  from HDF to BUFR is the main assumption of the realistic simulation, we conclude that this is the main cause of the low correlation of the MLE values (see Figure 1 or Figure 3).

The main difference between the real/simulated distributions and the theoretical case is in the mean MLE values. The ratio between HDF and BUFR mean values is above 1.5 in the former and unity in the latter (see section 3.1). In order to see the effects of extrapolating the theoretical case to a higher dimensional order of the measurement space for HDF and BUFR, we also perform a more constrained simulation<sup>1</sup>. The latter gives similar ratio between HDF and BUFR mean MLEs to that of the realistic simulation, i.e. about 1.5. The reason for this difference between the real/simulated distributions and the theoretical ones is that, in the theoretical case, the solution is a point in the multi-dimensional space while, in the simulation (also for real data), the solution is a multi-dimensional folded surface with a strong non-linear behavior. This non-linearity may contribute to the change in the MLE properties when going from HDF to BUFR.

The general results of the constrained simulation are similar (not shown) to those of the realistic simulation (see Figure 3). This shows that the constraining assumptions have no significant effect on the low correlation of the MLE values. Therefore, this result validates the assumptions used in the theoretical example.

### ***MLE distribution dependence on number of $\sigma^0$***

Figure 4 is similar to Figure 3. We use the same simulation procedure but in this case we fix the number of  $\sigma^0$  used in the simulated HDF instead of using a realistic number. In the case of fixing the number of HDF measurements to 5 (plot a), one beam has two  $\sigma^0$  measurements and the rest of the beams have only one each. In the case of 6  $\sigma^0$  measurements (plot b), two

---

<sup>1</sup> This simulation includes two additional constraints based on the assumptions 2 and 4 (see section 3.1), i.e. we consider only one truth (origin in the theoretical case), which in this case is an eastward wind of 7.8 m/s and we use fixed  $K_p$  values for both HDF and BUFR, and fixes the number of measurements in HDF and BUFR, i.e.  $N=8$  and  $M=4$  respectively.

beams have two measurements each and two beams have only one each. Analogous explanation goes for the cases of 7 (plot c) and 8 (plot d) measurements. [Note that the different combinations of measurements / beams (e.g. in the case of 5  $\sigma^0$ , you may use two measurements for the fore inner, the fore outer, the aft inner or the aft outer beam) do not affect the two-dimensional histograms (not shown)].

The plots in Figure 4 clearly show a decreasing correlation value with increasing number of HDF measurements (from 0.78 in plot a to 0.53 in plot d). Moreover, this effect is also seen in the shapes of the two-dimensional histograms, which are progressively smeared away from the diagonal. The correlation value of the theoretical case (0.7) is in between the correlation values of plots a (0.78) and b (0.66). This suggests that the decorrelation of the MLE when going from N=2 to M=1 is similar to the one from N=5 or 6 to M=4. The two-dimensional histogram in plot d is very similar to that of Figure 3. Both histograms present as well similar correlation values. This is due to the fact that the realistic distribution of the number of  $\sigma^0$  measurements used for HDF in Figure 3 contains 8  $\sigma^0$  measurements per WVC on average for the sweet parts of the swath. We can therefore interpret Figure 4 as a transition from the theoretical case to reality.

Figure 5 shows separately the distributions (one-dimensional histograms) of simulated BUFR and HDF MLEs. The different plots correspond to the different number of measurements simulated in HDF, in the same way as for Figure 4. As explained at the beginning of section 3, the BUFR simulated measurements are a beam averaging of the HDF simulated measurements. Figure 5 clearly shows that the MLE distribution for simulated BUFR is invariant to the number of HDF measurements used prior to the BUFR averaging. This is an expected result since the number of BUFR simulated measurements per WVC is always the same (4, one for each beam). However, the distribution of HDF MLE is significantly

changing with the number of simulated HDF measurements, increasing its peak and mean value with increasing number of simulated measurements (see evolution from plots a to d). Since the MLE value is a measure of the distance from the measurements to the GMF, this distribution change indicates that the more measurements (or the more dimensions in the measurement space) we use, the lower the probability to be close to the solution or GMF. As discussed in the first simulation, the dependence of the mean MLE value on the number of measurements is due to the non-linearities in the GMF.

The decorrelation between HDF and BUFR MLE is explained by the change in the distribution characteristics of the latter. Although the non-linear behavior of the GMF is affecting the MLE distributions, it is clear from the simulation results that the decorrelation is mainly due to a much smaller number of  $\sigma^0$  used in the inversion for BUFR compared to HDF (about half, since typically BUFR contains 4 and HDF 8 measurements).

In section 3.1, we demonstrate the change in the MLE distribution characteristics when averaging from a two-dimensional measurement space to a one-dimensional one. With this simulation we are able to better characterize the evolution of the MLE distributions when encountering higher dimensional measurement spaces.

### ***3.3 Detailed analysis of MLE differences: real versus simulated***

In sections 3.1 and 3.2, we clearly show the change in the MLE distribution characteristics when averaging the  $\sigma^0$  information (from HDF to BUFR), which leads to a small correlation of the HDF and BUFR MLEs. However, some differences are visible in the simulated MLE



distributions compared to the real MLEs. In this section, we perform a deeper analysis of these differences.

### ***Distributions***

Similar to Figure 5, Figure 6 shows the one-dimensional histograms of HDF and BUFR MLEs but for real (plot a) and simulated (realistic) (plot b) data. Note that the shape of the simulated HDF and BUFR distributions is different compared to the real distributions. In particular, the mean value of the BUFR distributions is larger for the simulated MLE than for the real MLE. Moreover, the SD value of the HDF distributions is significantly larger for the real compared with the simulated MLE. We also note that in the HDF real data there is a larger accumulation of values in the vicinity of zero MLE.

### ***Mean values versus node number and wind speed***

In order to better understand these differences in the MLE distributions, we study the behavior of the mean simulated and real MLE as a function of wind speed and cross-track location.

Figure 7 shows the mean MLE surface as a function of wind speed and WVC number for the HDF (plot a) and BUFR (plot b) real data. The noise in the MLE surfaces, caused by geophysical effects (such as rain) and/or small amount of data (at high winds), is filtered out. The MLEs used in this Figure correspond to the MLEs of the selected solutions provided in both QuikSCAT data products (HDF and BUFR). Figures 8a and 8b show the same surfaces as Figures 7a and 7b, respectively, but for simulated data. The MLEs used in this Figure

correspond to the MLEs of the 1<sup>st</sup> rank solutions provided by our inversion software. [Note: Figure 7 looks almost identical if we use the 1<sup>st</sup> rank solution instead of the selected solution. The reason to show the selected solution Figure is because it is used for QC purposes in section 4.]

In HDF, there is a slight increase of the surface with increasing distance to nadir (Figures 7a and 8a). The reason for this increase lies in the inversion. As the inversion is a non-linear process, the scaling (linear correction), that is, the Kp noise (see equation 1), is not sufficient to compensate for the increase in the MLE due to the increase in the number of  $\sigma^\circ$ . Nevertheless, the mean MLE surfaces show that this effect is minor. Note that the increase is stepwise in the simulated data (Figure 8a) and not monotonic as for real data (Figure 7a) because of the approximation in the number of  $\sigma^\circ$  made in the simulation of HDF data, that is, we use a constant number of measurements in each WVC while in reality the number of measurements per WVC varies (see Figure 2).

In BUFR, there is no increase in MLE with increasing distance to nadir (Figures 7b and 8b), as the number of  $\sigma^\circ$  is kept constant for all WVCs. As for the HDF case, the simulated MLE behaviour across track (Figure 8b) compares well with the real case in BUFR (Figure 7b).

If we compare the mean MLE behaviour as a function of wind speed between the real (Figure 7) and the simulated (Figure 8) cases, we see a large discrepancy at low speeds. In reality, the MLE increases with decreasing speed while in the simulated case the MLE decreases with decreasing speed.

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. The Kp noise contains two

terms: the instrument noise and the geophysical noise. [6] 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 (geophysical noise). 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  $K_p$  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 however not the case. The problem lies in the inversion and, among others, in the fact that we assume that the measurement noise is proportional to the true value. The latter leads to a  $K_p$  which is proportional to  $\sigma_{si}^\circ$  (simulated  $\sigma^\circ$  from the GMF) in the denominator of equation 1. [13] explains on 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 9 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 ( $K_{p_t}$ ). Using this  $K_{p_t}$  we simulate the measurement pair (triangle inside the solid circle). The dashed circle represents its corresponding estimated noise ( $K_{p_m}$ ). After inversion, we get a positively biased solution (star) which has its proportional noise ( $K_{p_s}$ ) represented by the dotted circle. As  $K_{p_s}$  increases significantly, the MLE decreases ( $K_{p_s}$  is the denominator of equation 1) and this effect is more acute as we approach the origin corresponding to lower speeds.

Finally, it is clearly discernible that the mean MLE at mid and high speeds is significantly larger for simulated data compared to real data (see the surface plateau level of Figure 8 compared to Figure 7). This means that there is an overestimation of the Kp (or measurement) noise at these speed regimes. Since the largest uncertainty at these speeds lies in the instrument noise and not in the geophysical noise, we conclude that there is probably an overestimation of the instrument noise.

### *Main differences*

According to the analysis, we conclude that the remaining differences between the simulated and the real distributions can be attributed to the following:

- The simulation of the number of  $\sigma^\circ$  per WVC and beam in HDF is just a rough approximation. In the real data a WVC can contain a variable number of  $\sigma^\circ$  (see Figure 2) 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 7 and 8). In the real data, the estimated Kp values, and more specifically the geophysical noise values, are underestimated.
- An overall overestimation in the real data of the estimated Kp values (except at low winds where the opposite occurs), more specifically, the instrument noise values. This is deduced from the higher mean MLE values of both HDF and BUFR simulated distributions (Figure 8) compared to the mean values of the real distributions (Figure 7).

There may be other reasons that could cause minor differences in the distributions, such as processing of eggs or composites, i.e., for real data HDF uses eggs and BUFR composites

(see section 2), whereas for simulated data we treated both HDF and BUFR as composites. Nevertheless and as discussed in the previous section, these differences are not so relevant as the simulation is a good reflection of reality.

## **4 Wind Retrieval and Quality Control skills in BUFR versus HDF**

In the previous section, we have shown how different the MLE distributions are in HDF and BUFR. In the BUFR product,  $\sigma^0$  measurements are combined to result in only 4 independent  $\sigma^0$  observations. In HDF on average 8  $\sigma^0$  measurements are available per WVC in the sweet swath. This data reduction could cause information in the  $\sigma^0$  measurements to be lost. As such, the poor BUFR and HDF MLE comparison should be taken seriously. The MLE is the residual parameter output from wind retrieval and is very important for Quality Control ([5], [6], [7], and [8]). Therefore, in this section, we investigate in some detail the wind retrieval and QC performance properties of SeaWinds BUFR as compared to HDF.

### **4.1 Wind retrieval**

Figure 10 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 1<sup>st</sup> rank solution.

From the upper plots we note that the BUFR and HDF retrieved winds are not identical, although very similar. Plot a) shows almost no bias in speed and a very small STD (0.58 m/s). Plot b shows a typical effect of comparing 1<sup>st</sup> rank solutions, which is the secondary distribution around 180°. This is due to the fact that 1<sup>st</sup> and 2<sup>nd</sup> rank solutions (often with very similar wind speed but wind direction 180° apart) can have very similar MLE values and therefore be switched from one data product to the other. This effect leads to very high directional SD 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 conclude that the difference in the MLE distributions is not affecting the quality of the retrieved winds.

## **4.2 Quality Control**

The mean MLE distributions of both formats are similar (see Figure 7). Since the mean MLE distribution is used for determining the QC of HDF data (see [7]), it seems reasonable to test the same QC procedure for BUFR.

### 4.2.1 Methodology

The method, as described by [7], consists of normalizing the MLE with respect to the wind speed and the node number (or cross-track location). For a given wind speed and node number, we compute the expected MLE. Then we define the normalized residual as:

$$R_n = \text{MLE} / \langle \text{MLE} \rangle \quad (8)$$

where MLE is the maximum likelihood estimator of a particular wind solution (given by the inversion) and  $\langle \text{MLE} \rangle$  is the expected MLE for that particular WVC (node number) and wind solution. The  $\langle \text{MLE} \rangle$  is a surface fit to the mean MLE surface (Figure 7). This fit is made in order to be able to extrapolate to higher wind speeds. The computation of the  $\langle \text{MLE} \rangle$  for BUFR data (see Appendix) is slightly different than for HDF data since the surfaces are somewhat different (compare Figure 7a and 7b).

[7] found that the QC procedure works slightly better when using the MLE information of the selected solution rather than the 1<sup>st</sup> rank solution. Therefore, similar to the HDF case, the  $R_n$  is computed with the MLE of the selected solution given in the BUFR product.

Following the definition of  $R_n$  (see equation 8), an  $R_n$  threshold, which separates the good quality winds from the poor quality winds, needs to be defined. The same wind-dependent  $R_n$  threshold as defined for HDF (see [7]) is used for BUFR data. It is a parabolic threshold with a maximum  $R_n$  value of 4 at 5 m/s, which reaches a value of 2 at 15 m/s and then remains constant for higher wind speed values. The reference wind speed is the selected wind solution.

The QC by  $R_n$  works as follows. The  $R_n$  (equation 19) of the selected solution of any WVC is computed. If the  $R_n$  is lower or equal to the threshold, the WVC is accepted; otherwise, the WVC is rejected. We test the QC procedure against ECMWF and SSM/I collocations.

#### 4.2.2 Results

The QC procedure should be optimized in order to achieve the following goals: maximum low-quality data rejection (including rain contaminated WVCs) and minimum good-quality data rejection. Table 1 summarizes the results of the BUFR QC compared with the HDF QC (the latter adopted from [7]). As in [7], the results correspond to the sweet parts of the swath (nodes 9 to 28 and 49 to 68). The total amount of data used is over 3 million for both BUFR and HDF.

The first column shows the percentage of accepted and rejected data. We note a larger percentage of rejections in BUFR (6.7%) than in HDF (5.6%).

The second column shows the average Root Mean Square (RMS) of wind vector difference between the QuikSCAT selected winds and the ECMWF winds (RMS-ECMWF). The RMS-ECMWF is providing information about the quality of the QuikSCAT data, i.e. the larger the RMS-ECMWF value the lower the quality of the retrieved wind is according to ECMWF. The difference in RMS-ECMWF between the accepted and rejected solutions is 4 m/s in HDF, while in BUFR is slightly lower than 3 m/s. Therefore, the excess of rejections in BUFR with respect to HDF (see column 1) is mostly concentrated at low RMS-ECMWF values, which in



turn makes the mean RMS-ECMWF smaller (4.92 m/s in BUFR, while 6.24 m/s in HDF). This indicates a better performance of the HDF QC with respect to the BUFR QC.

The third column shows the percentage of accepted and rejected “rainy” data. We consider “rainy” data, when SSM/I rain rate is above 6 mm/hr. [7] investigated the quality of QuikSCAT winds in the presence of rain. They found that for SSM/I rain rate above 6 mm/hr, the radar backscatter is mainly coming from the rain and therefore the quality of the retrieved winds is generally unacceptable. Therefore, the QC is expected to maximize rain rejection. This is indeed what occurs, since most of the “rainy” WVCs are rejected. However, the percentage of rejections is slightly higher for HDF (87.3%) than for BUFR (83.9%), denoting again a slightly better performance of the HDF QC.

As for HDF, in the region close to nadir (nodes 29 to 48), the results (not shown) show a slightly worse skill of the QC compared with the sweet parts of the swath. This is due to the poorer azimuth diversity in the  $\sigma^0$  measurements of each WVC at the nadir region.

Note that, as for HDF, 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.

In general, the BUFR QC works fine, although its performance is slightly worse than the HDF QC. Therefore, the properties of the MLE as a QC indicator are almost independent of the data format used, despite the important differences between the MLE distributions discussed in section 3. In other words, anomalies like rain cause similar effects in both HDF and BUFR  $\sigma^0$  products, that is, HDF and BUFR are strongly correlated in such cases.

## 5 Conclusions

A comparison between HDF and BUFR MLE data shows that both MLE distributions are poorly correlated. It is assumed that the different level of averaging of the data in HDF and BUFR is the main cause.

In order to set a QC procedure for QuikSCAT BUFR data using the experience of the recent QC procedure developed by [7] for QuikSCAT HDF data, a comprehensive characterization of the MLE is performed.

A very simple example is solved theoretically to predict the MLE behavior with respect to the data format used. In this theoretical case, for simplicity, HDF is assumed (among other assumptions) a measurement system providing two observations and BUFR a system providing one (combined) observation. We show how the MLE distribution changes, due to the  $\sigma^0$  averaging that is performed when going from HDF format to BUFR format. As a consequence of this change, the correlation between both MLE distributions is 0.7.

The results of the comparison between the HDF and BUFR real data are close to the theoretical case. However the MLE distribution characteristics and the correlation values are somewhat different. In particular the correlation value for real data is 0.5. We postulate that the main difference between theory and reality is the different number of measurements used and we perform a simulation to check it.

First, a simulation, which includes a realistic number of  $\sigma^0$  observations and realistic noise values for both HDF and BUFR formats, is performed. The MLE distributions of this realistic

simulation compare well with the real distributions. Indeed, the simulated correlation value between both formats is the same as in reality (about 0.5). Since the main assumption in the simulation is the change in number of measurements when simulating BUFR  $\sigma^o$  from HDF  $\sigma^o$ , we conclude that the low correlation of the MLE distributions of both formats is due to the  $\sigma^o$  averaging.

In order to understand the differences between the mean MLE distributions of the simulated/real data and the theoretical case, we performed a more constrained simulation. This simulation is simply an extrapolation of the theoretical case to a higher dimensional measurement space. The results of this simulation are similar to those of the realistic simulation. The differences between the simulated (and real) and the theoretical distributions are explained by the fact that, in contrast with the theory where a single point is used as solution, in the simulation (and reality) the solution is a multi-dimensional surface with highly non-linear behavior. The good agreement between the realistic simulated MLE distributions and the “constrained” simulated distributions validates the assumptions used in the theoretical case.

Further simulations show how the MLE distributions change as a function of the number of observations taken. The higher the difference in the number of HDF and BUFR observations, the lower the correlation and higher the mean MLE value difference between the two products.

The remaining differences between the realistic simulation and reality are analyzed in detail. Misestimations of the real measurement noise and simplification in the computation of the number of measurements for both formats in the simulation are pointed out as the main cause for these differences.

Collocations of HDF and BUFR winds with ECMWF winds show that the quality of the retrieved winds of both formats is very similar. As such, the different MLE behavior is not affecting the quality of the wind retrievals.

Finally, we use the same QC procedure for BUFR and for HDF. The skill of the BUFR QC is very similar to the HDF QC. Therefore, and despite the different MLE distribution of the BUFR product compared to HDF, the MLE turns out to be a good QC indicator regardless of the data format. This in turn shows that in the presence of rain (or other anomalies that produce low quality winds) HDF and BUFR are strongly correlated indeed.

As such, despite the smaller information content of the BUFR product due to the averaging procedure, its usefulness for meteorological applications is very similar to the HDF product.

## **Appendix: <MLE> surface fit for JPL-retrieved winds in BUFR format**

For HDF, a simple fit to the filtered mean MLE surface (see Figure 7a) is performed, by finding the MLE dependence on wind speed for a certain node and assuming that the shape of this function is nearly constant over all nodes of the inner swath (see [7]). Looking at the filtered mean MLE surface for BUFR (see Figure 7b), we note that assuming a constant shape of the MLE dependence on wind speed over all nodes of the inner swath is not valid anymore.

In this case, we fit a Gaussian + 2<sup>nd</sup> order polynomial function to the filtered mean MLE for each node of the inner swath separately. Then, we fit a 2<sup>nd</sup> order polynomial function to the

evolution of each coefficient of the previous function with respect to the node number.

Therefore, the 2D function which well fits the filtered mean MLE surface is the following:

$$\langle MLE \rangle_{fit} = A_0 \cdot e^{-\frac{1}{2} \left( \frac{v-A_1}{A_2} \right)^2} + A_3 + A_4 \cdot v + A_5 \cdot v^2 \quad (A1)$$

$$A_i = A_{i0} + A_{i1} \cdot n + A_{i2} \cdot n^2 \quad ; \quad \forall i \in [0,5], \forall n \in [1,76] \quad (A2)$$

where  $\langle MLE \rangle$  is the expected MLE,  $v$  is the wind speed and  $n$  the node number.

The coefficient values are the following:

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

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

$$A_{20} = 2.75000; A_{21} = 0.00000; A_{22} = 0.00000$$

$$A_{30} = 0.21210; A_{31} = -2.49E-3; A_{32} = 3.02E-5$$

$$A_{40} = -7.41E-3; A_{41} = 3.13E-4; A_{42} = -4.08E-6$$

$$A_{50} = 1.18E-4; A_{51} = -4.76E-6; A_{52} = 6.24E-8$$

## Acknowledgements

As a member of the KNMI QuikSCAT team, John de Vries has contributed to the work described in this paper. We acknowledge the help and collaboration of our colleagues working at ECMWF and KNMI. The QuikSCAT data were obtained from the NASA Physical

Oceanography Distributed Active Archive Center, at the Jet Propulsion Laboratory / California Institute of Technology, and the National Oceanic and Atmospheric Administration. Furthermore, this work is only possible due to the EUMETSAT grant for a research fellowship post at KNMI. We greatly appreciate the two reviewers who helped to improve the paper.

## References

- [1] L. Isaksen and A. Stoffelen, "ERS Scatterometer Wind Data Impact on ECMWF's Tropical Cyclone Forecasts," *IEEE Trans. on Geoscience and Rem. Sens.* 38 (4) pp. 1885-1892, 2000.
- [2] A. Stoffelen and P. Van Beukering, "Implementation of Improved ERS Scatterometer Data Processing and its Impact on HIRLAM Short Range Weather Forecasts," *Beleidscommissie Remote Sensing Report*, contract NRSP-2/97-06, pp. 75, 1997.
- [3] R. Atlas, R.N. Hoffman, S.M. Leidner, J. Sienkiewicz, T.-W. Yu, S.C. Bloom, E. Brin, J. Ardizzone, J. Terry, D. Bungato, J.C. Jusem, "The Effects of Marine Winds from Scatterometer Data on Weather Analysis and Forecasting," *Bull. Amer. Meteorol. Soc.*, Vol. 82, No. 9, pp. 1965-1990, 2001.
- [4] M. Rohn, G. Kelly, and R. Saunders, "Experiments with Atmospheric Motion Vectors at ECMWF," *Proc. of Fourth International Winds Workshop*, EUM P24, ISSN 1023-0416, Saanenmoser (Switzerland), EUMETSAT, pp. 139-146, 1998.
- [5] A. Stoffelen and D. Anderson "Scatterometer Data Interpretation: Measurement Space and Inversion," *J. Atm. And Ocean Techn.*, 14(6), pp. 1298-1313, 1997.
- [6] J. Figa and A. Stoffelen, "On the Assimilation of Ku-band Scatterometer Winds for Weather Analysis and Forecasting," *IEEE Trans. on Geoscience and Rem. Sens.* 38 (4) pp. 1893-1902, 2000.

- [7] M. Portabella and A. Stoffelen, "Rain Detection and Quality Control of SeaWinds," *J. Atm. And Ocean Techn.*, Vol. 18, No. 7, pp. 1171-1183, 2001.
- [8] M. Portabella and A. Stoffelen, "A Comparison of KNMI Quality Control and JPL Rain Flag for SeaWinds," *Can. Jour. of Rem. Sens.* (special issue on Remote Sensing of Marine Winds), Vol. 28, No. 3, 2002.
- [9] M. W. Spencer, C. Wu, and D. G. Long, "Tradeoffs in the Design of a Spaceborn Scanning Pencil Beam Scatterometer: Application to SeaWinds," *IEEE Trans. Geosci. Remote Sensing*, Vol. 35, No. 1, pp. 115-126, 1997.
- [10] JPL, "QuikSCAT Science Data Product User's Manual," version 1.1, *JPL D-18053*, pp. 84, 1999.
- [11] M. Leidner, R. Hoffman, and J. Augenbaum, "SeaWinds Scatterometer Real-Time BUFR Geophysical Data Product," version 2.2.0, *NOAA/NESDIS*, pp. 45, 2000.
- [12] M. Portabella and A. Stoffelen, "Quality Control and Wind Retrieval for SeaWinds," Scientific report WR-2002-01, *Koninklijk Nederlands Meteorologisch Instituut, De Bilt, The Netherlands*, 2002.
- [13] A. Stoffelen, "Scatterometry," *PhD thesis* ISBN 90-393-1708-9, University of Utrecht, pp. 195, 1998.



## List of Figures

Figure 1 Contour plot of the two-dimensional histogram of the BUFR MLE versus the HDF MLE.  $N$  is the number of data;  $m_x$  and  $m_y$  are the mean values along the  $x$  (HDF) and  $y$  (BUFR) axis, respectively;  $m(y-x)$  and  $s(y-x)$  are the bias and the standard deviation with respect to the diagonal, respectively; and  $cor_{xy}$  is the correlation value between the HDF and the BUFR MLE distributions. The contour lines are in logarithmic scale: each step is a factor of 2 and the lowest level (outer-most contour line) is at  $N/4000$  data points.

Figure 2 Normalized histogram of the number of  $\sigma_0$  for WVC number 12 (plot a) and 55 (plot b). The solid line corresponds to the inner swath beams (fore or aft) and the dotted line to the outer swath beams.

Figure 3 Same as Figure 1 but for the realistic simulation.

Figure 4 Same as Figure 3 but using a varying number of measurements in the HDF simulation: a) 5; b) 6; c) 7; and d) 8.

Figure 5 One-dimensional histogram plots of the MLE distributions of Figure 4. The number of measurements used in the HDF simulation is: a) 5; b) 6; c) 7; and d) 8.  $m_h$  and  $m_b$  are the mean values of the HDF and BUFR distributions, respectively;  $s_h$  and  $s_b$  are the standard deviation values of the HDF and BUFR distributions, respectively.

Figure 6 One-dimensional histogram plots of the HDF and BUFR MLE distributions for real data (a) and the realistic simulation (b). The legend is the same as in Figure 5.

Figure 7 Mean MLE versus wind speed and node number for real data: (a) HDF (adopted from Figure 3a of [7]) and (b) BUFR. The speed binning is 1 m/s and the node binning is 1.

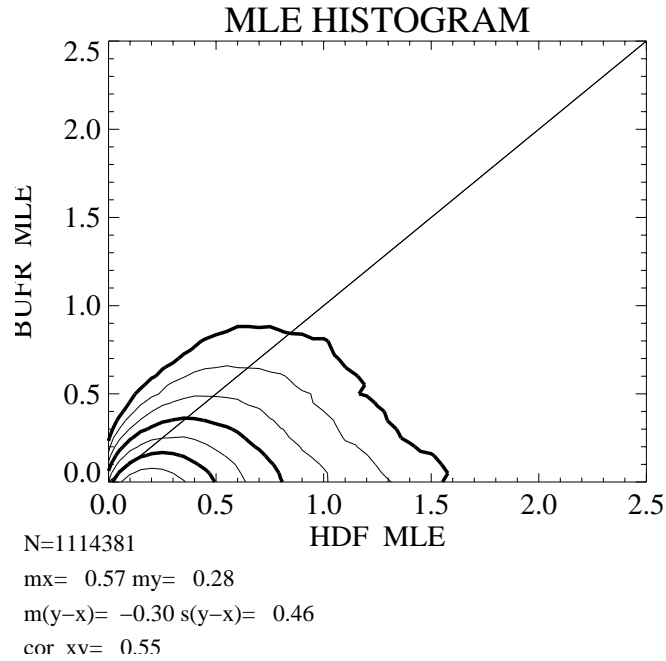
Figure 8 Mean MLE versus wind speed and node number for simulated data: (a) HDF and (b) BUFR. The speed binning is 1 m/s and the node binning is 1.

Figure 9 schematic illustration of the problem in a 2D measurement space

Figure 10 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  $2.5^\circ$ ). The legend is the same as in Figure 1.

Table 1 QC results BUFR / HDF.

## Illustrations



*Figure 1 Contour plot of the two-dimensional histogram of the BUFR MLE versus the HDF MLE.  $N$  is the number of data;  $m_x$  and  $m_y$  are the mean values along the  $x$  (HDF) and  $y$  (BUFR) axis, respectively;  $m(y-x)$  and  $s(y-x)$  are the bias and the standard deviation with respect to the diagonal, respectively; and  $cor_{xy}$  is the correlation value between the HDF and the BUFR MLE distributions. The contour lines are in logarithmic scale: each step is a factor of 2 and the lowest level (outer-most contour line) is at  $N/4000$  data points.*

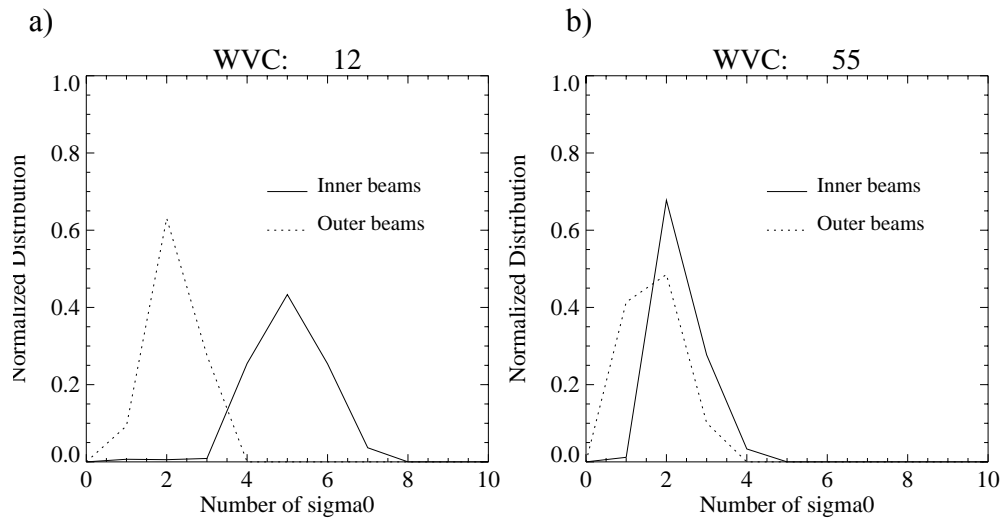
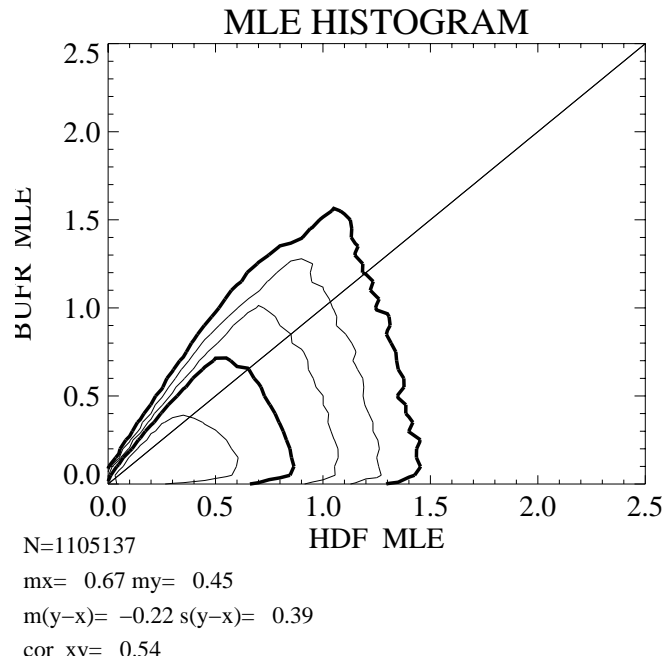


Figure 2 Normalized histogram of the number of sigma0 for WVC number 12 (plot a) and 55 (plot b). The solid line corresponds to the inner swath beams (fore or aft) and the dotted line to the outer swath beams.



*Figure 3 Same as Figure 1 but for the realistic simulation.*

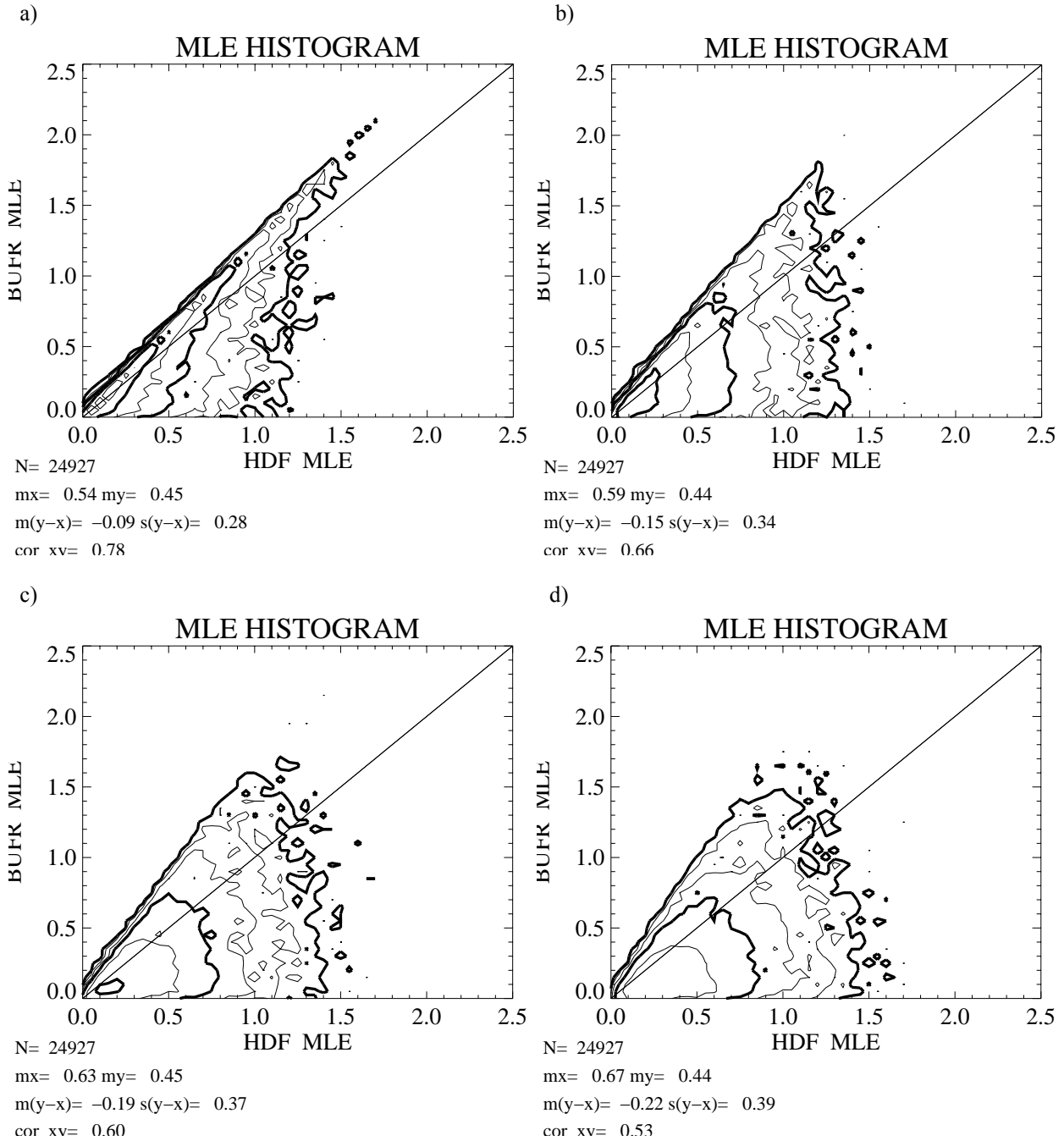
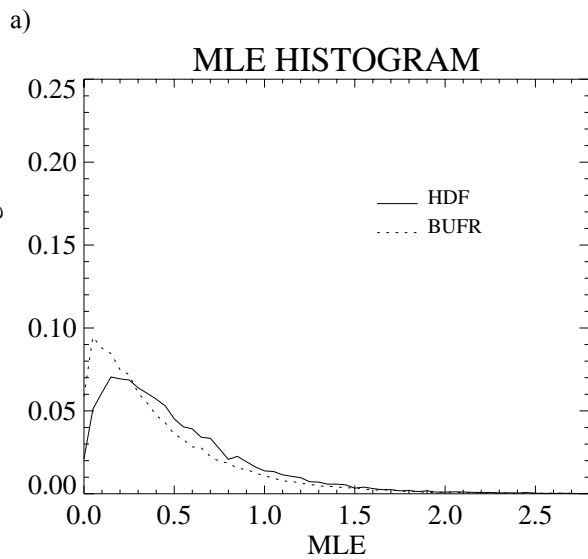
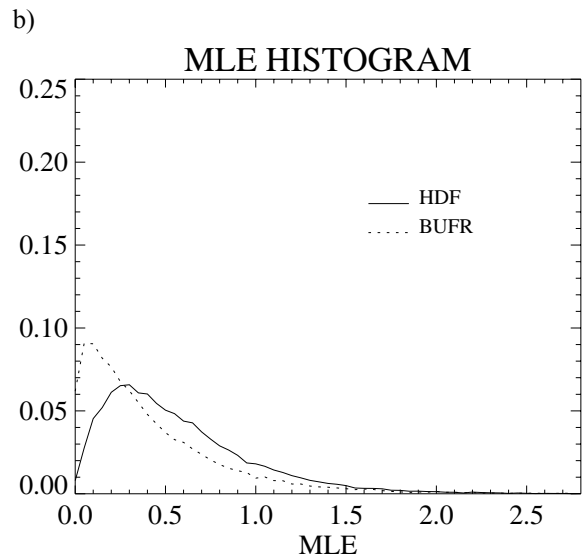


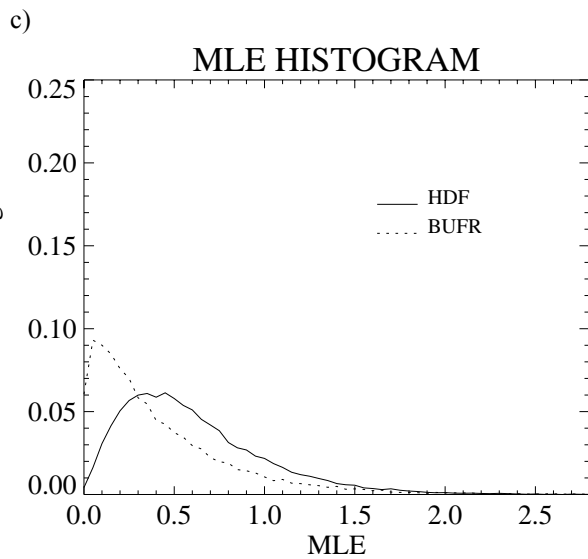
Figure 4 Same as Figure 3 but using a varying number of measurements in the HDF simulation: a) 5; b) 6; c) 7; and d) 8.



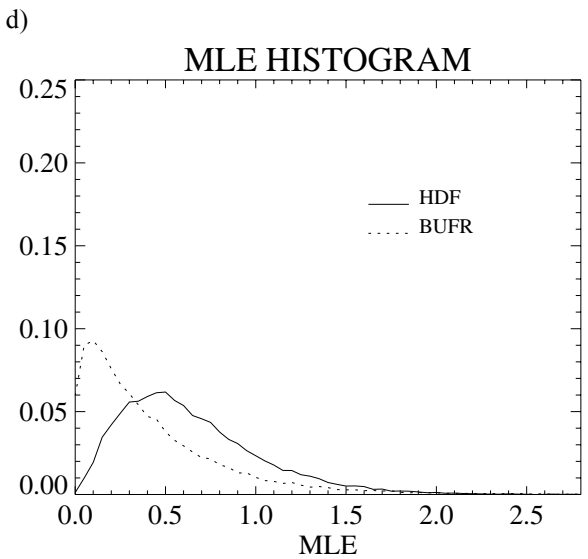
N= 24927  
 mh= 0.54 mb= 0.45  
 sh= 0.43 sb= 0.41



N= 24927  
 mh= 0.59 mb= 0.44  
 sh= 0.42 sb= 0.41



N= 24927  
 mh= 0.63 mb= 0.45  
 sh= 0.40 sb= 0.41



N= 24927  
 mh= 0.67 mb= 0.44  
 sh= 0.39 sb= 0.41

Figure 5 One-dimensional histogram plots of the MLE distributions of Figure 4. The number of measurements used in the HDF simulation is: a) 5; b) 6; c) 7; and d) 8. mh and mb are the mean values of the HDF and BUFR distributions, respectively; sh and sb are the standard deviation values of the HDF and BUFR distributions, respectively.

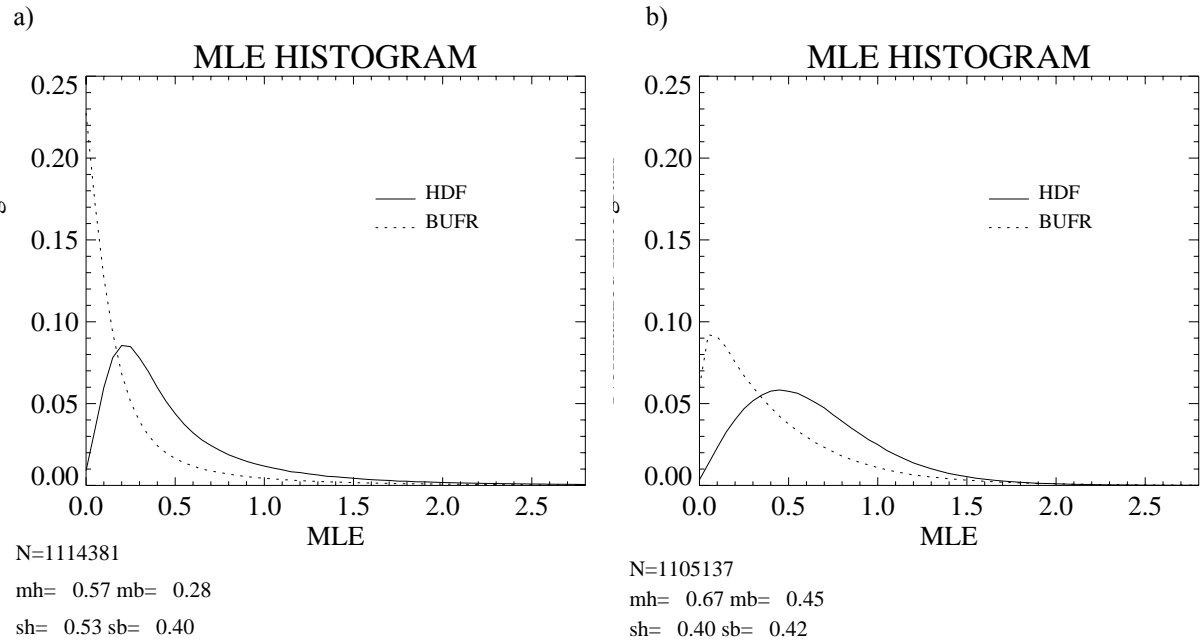
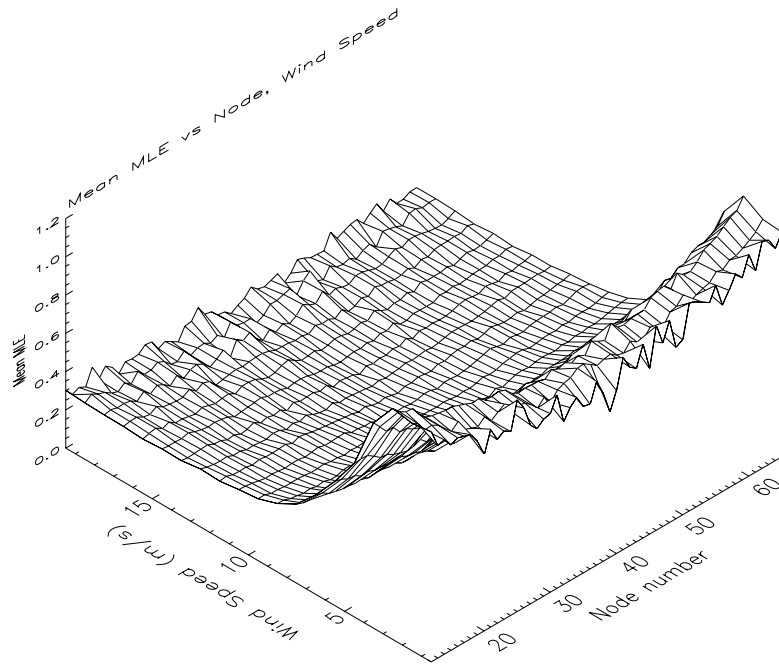


Figure 6 One-dimensional histogram plots of the HDF and BUFR MLE distributions for real data (a) and the realistic simulation (b). The legend is the same as in Figure 5.



a)



b)

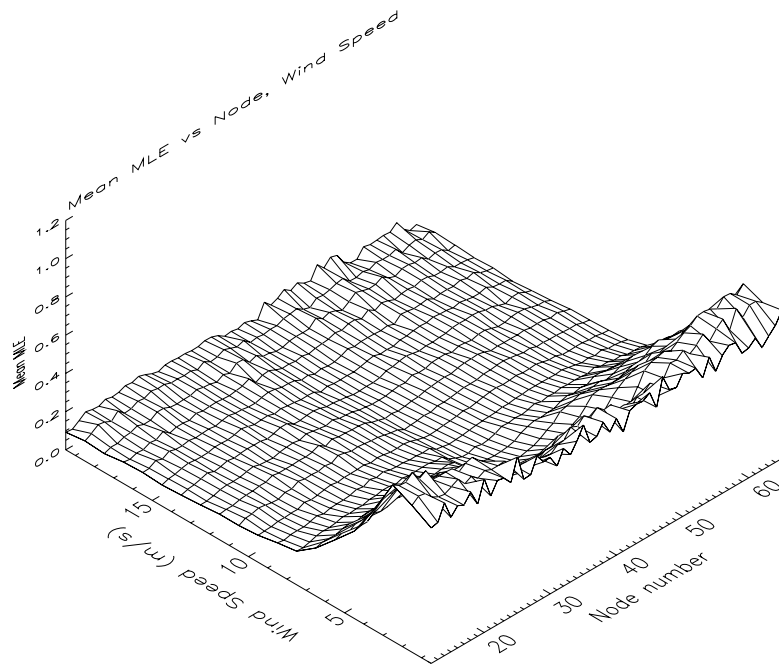
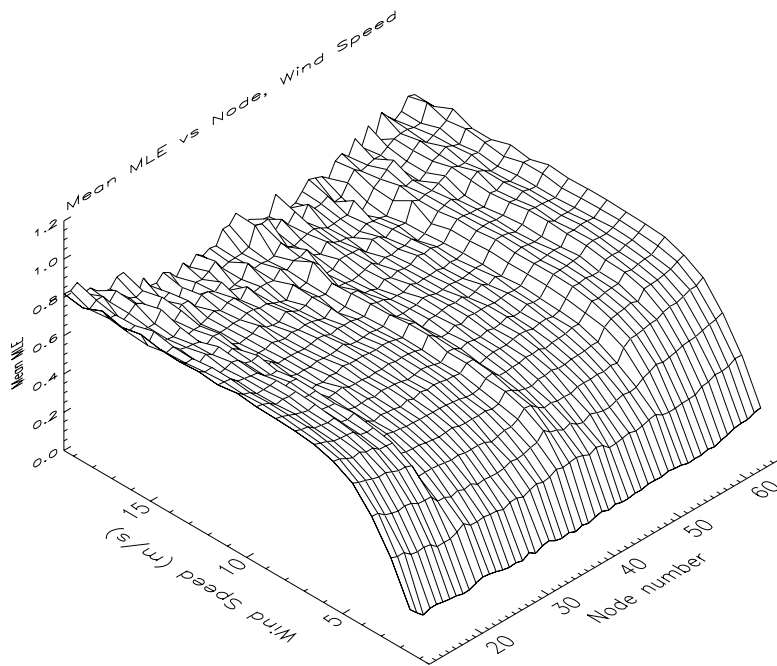


Figure 7 Mean MLE versus wind speed and node number for real data: (a) HDF (adopted from Figure 3a of [7]) and (b) BUFR. The speed binning is 1 m/s and the node binning is 1.

a)



b)

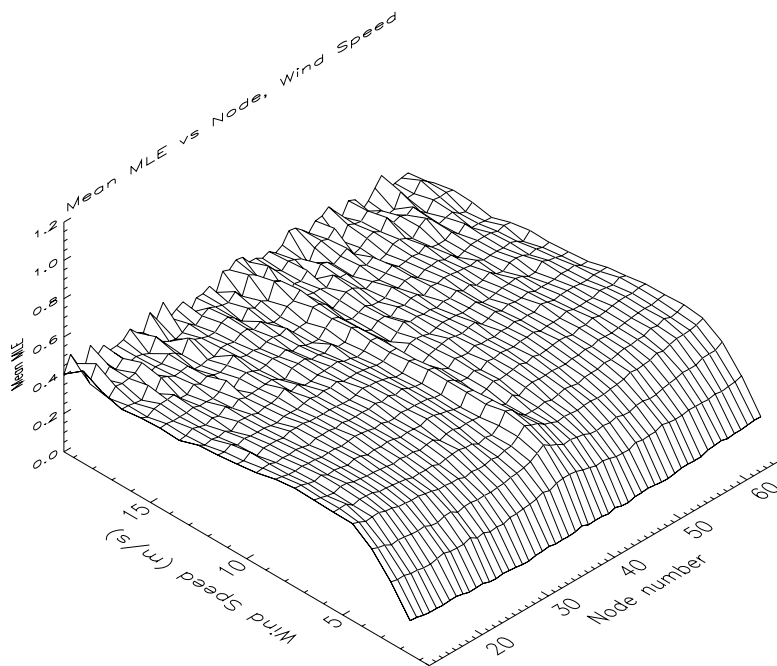


Figure 8 Mean MLE versus wind speed and node number for simulated data: (a) HDF and (b) BUFR. The speed binning is 1 m/s and the node binning is 1.

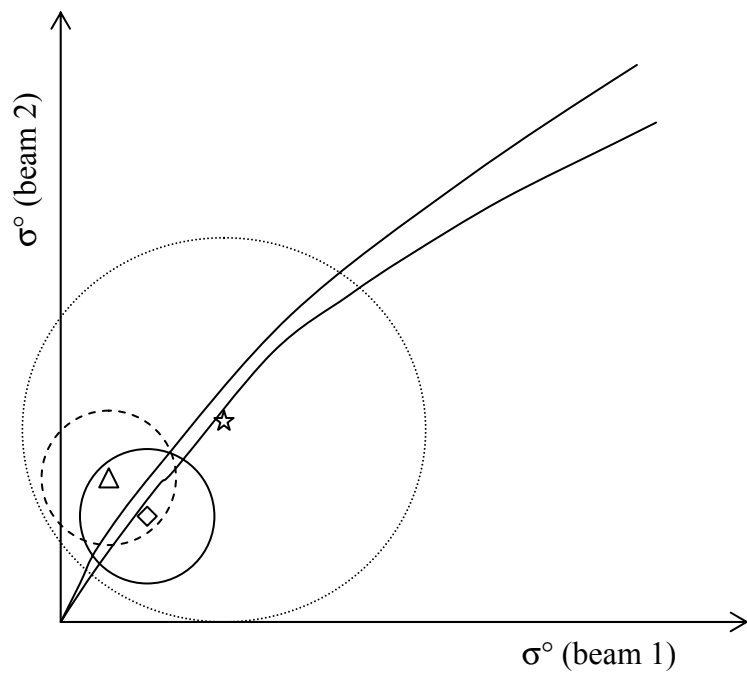


Figure 9 Schematic illustration of the problem in a 2D measurement space

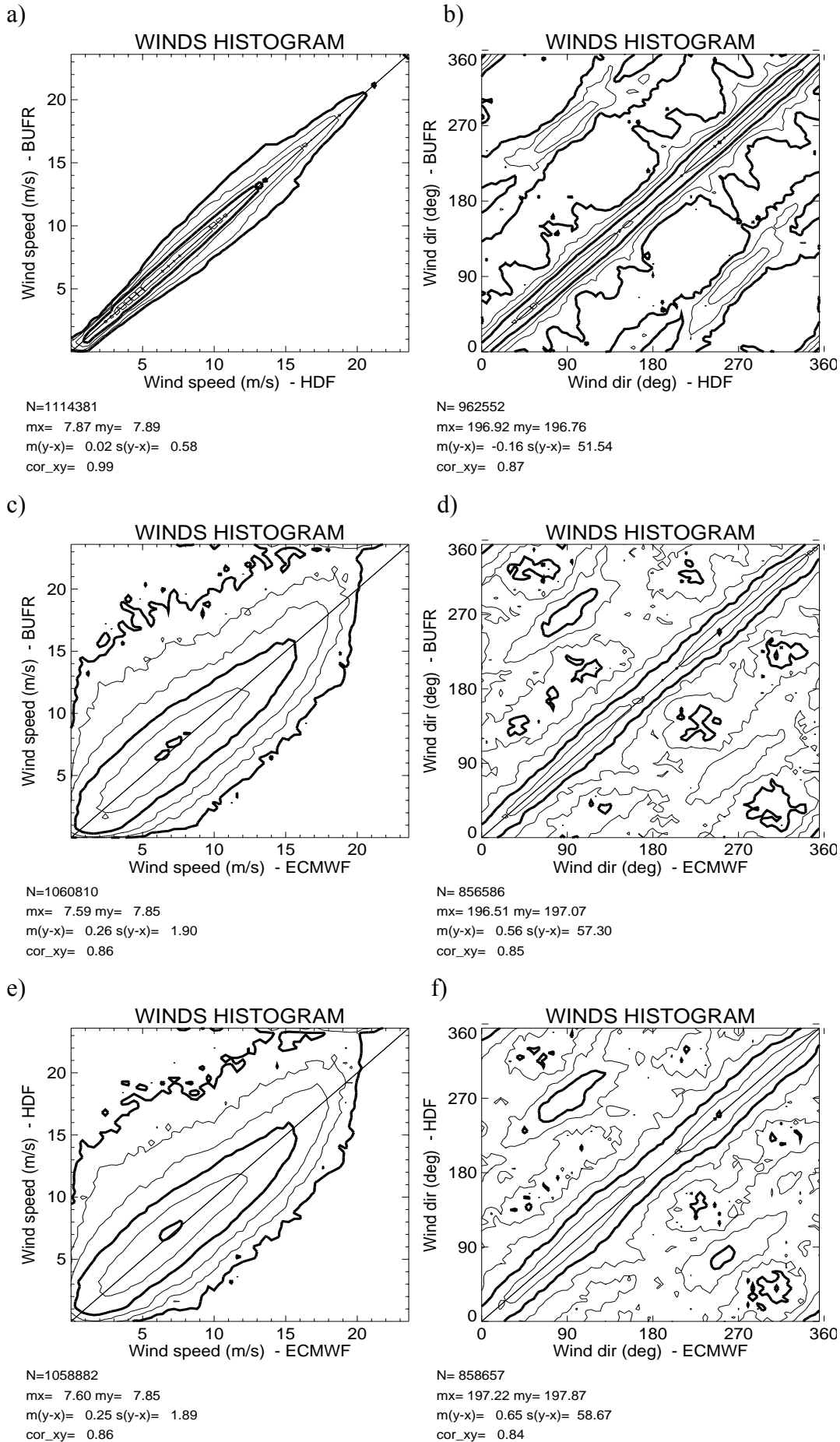


Figure 10 Two-dimensional histogram of BUFR winds versus HDF winds (plots a and b), BUFR winds versus ECMWF winds (plots c and d) and HDF winds versus ECMWF winds (plots e and f). The left plots correspond to wind speeds (bins of 0.4 m/s) and the right plots to wind directions (bins of 2.5 °). The legend is the same as in Figure 1.

## Tables

**Table 1 QC results BUFR / HDF**

	% of Total	Mean RMS (m/s)	RR > 6 (%)
Accepted	93.3 / 94.4	2.07 / 2.24	16.1 / 12.7
Rejected	6.7 / 5.6	4.92 / 6.24	83.9 / 87.3

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

## Biographies



**Marcos Portabella** received his B.Sc. degree in Physics in 1994, with a specialization in atmospheric physics, from the University of Barcelona, Barcelona, Spain, and the M.Sc. in Space Studies in 1995, with a specialization in Remote Sensing, from the Institute of Space Studies of Catalonia, Barcelona, Spain.

He later worked at the European Space Agency (ESA) on wind retrievals from satellite radar systems, such as NASA Scatterometer (NSCAT), ESA scatterometers (ERS-1 and ERS-2) and synthetic aperture radars (SAR). He is currently with the Royal Netherlands Meteorological Institute (KNMI), de Bilt, The Netherlands, where he is working on data interpretation, inversion and quality control of SeaWinds scatterometer, flown aboard the NASA satellite QuikSCAT.



**Ad Stoffelen** was born on February 25, 1962, in the Netherlands. He received the M.Sc. degree in physics in 1988 from the Technical University of Eindhoven, The Netherlands, and the Ph.D. degree from University of Utrecht, Utrecht, The Netherlands.

He is currently with the Royal Netherlands Meteorological Institute (KNMI), de Bilt, The Netherlands, and has worked on scatterometer data interpretation, inversion, calibration, validation, quality monitoring, and assimilation in global and regional weather forecast models. Other involvements include the European Doppler wind lidar in space program, and ozone data assimilation.