NWP SAF

Satellite Application Facility
for Numerical Weather Prediction

A probabilistic approach for SeaWinds data assimilation: an improvement in the nadir region

Marcos Portabella   *Astronomy and Meteorology Department, University of Barcelona*

Ad Stoffelen   *KNMI*

NWP SAF

Satellite Application Facility for Numerical Weather Prediction

Report of Visiting Scientist Mission

A probabilistic approach for SeaWinds data assimilation: an improvement in the nadir region

Marcos Portabella

Astronomy and Meteorology Department, University of Barcelona

A Stoffelen
KNMI

This documentation was developed within the context of the EUMETSAT Satellite Application Facility on Numerical Weather Prediction (NWP SAF), under the Cooperation Agreement dated 25 November 1998, between EUMETSAT and the Met Office, UK, by one or more partners within the NWP SAF. The partners in the NWP SAF are the Met Office, ECMWF, KNMI and Météo France.

Change record			
Version	Date	Author / changed by	Remarks
1.0	22.10.02	M.Portabella, A.Stoffelen	

Contents

Abstract

Scatterometer sea-surface wind observations are being successfully assimilated into Numerical Weather Prediction (NWP) models. However, the impact of such observations often critically depends on their quality. In this respect, the quality of the winds retrieved from the new SeaWinds scatterometer (onboard QuikSCAT) depends on the subsatellite cross-track location. In particular, the poor azimuth separation or diversity between views (beams) in the nadir region results in poor quality winds.

The standard wind retrieval procedure consists of considering the Maximum Likelihood Estimator (MLE) cost function local minima as the potential (ambiguous) wind solutions that are used by the Ambiguity Removal (AR) procedure to select the observed wind. In the QuikSCAT nadir region, where the cost function minima are broad, the use of the standard procedure results in arbitrary and inaccurate winds. A scheme, which allows more ambiguous wind solutions when the retrieval results in broad cost function minima, i.e., a multiple solution scheme (MSS), is proposed as alternative to the standard procedure. The probability of every ambiguous solution of being the "true" wind is empirically derived and used in the AR procedure to make the scheme flexible enough to accept many wind solutions. The AR scheme uses National Centre for Environmental Prediction (NCEP) 24-hour forecasts as NWP background.

A comparison between the standard wind retrieval and the MSS procedures at 100-km resolution is then performed, using the European Centre for Medium-range Weather Forecast (ECMWF) First Guess at Appropriate Time (FGAT) model winds for validation. The MSS turns out to be more in agreement with ECMWF than the standard procedure, especially at nadir. Moreover, it shows more spatially consistent and realistic winds by more effectively exploiting the information content of the observations. In fact, AR results in winds with generally higher a priori probability and generally good agreement between a priori probability and AR selection. As such, the MSS concept is potentially beneficial for QuikSCAT data assimilation purposes in NWP. Finally, the lack of an effective Quality Control (QC) at 100-km resolution, essential for assimilation purposes, is discussed and several methods are recommended for further investigation.

1 Introduction

The forecast of extreme weather events is not always satisfactory, while their consequences can have large human and economic impact. Since many weather disturbances develop over the oceans, sea surface wind observations can help to improve the prediction of the intensity and position of such disturbances.

Nowcasting, short-range forecasting and numerical weather prediction (NWP) assimilation can benefit from the sea surface wind observations. In this respect, *Stoffelen and Anderson* (1997a) show that spaceborne scatterometers, which are able to provide accurate winds over the ocean surface, have a beneficial impact on analyses and short-range forecast, mainly due to improvements on the sub-synoptic scales. Moreover, the impact of assimilating sea surface winds into NWP models significantly depends on the data coverage. *Stoffelen and Van Beukering* (1997) and *Undén et al.* (1997) show a much more positive impact by duplicating the sea surface wind data coverage.

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 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.

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 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.* (1997), *JPL* (2001), and *Leidner et al.* (2000).

In comparison with previous scatterometers, the SeaWinds system has a much higher coverage and, as such, is potentially very useful for data assimilation in NWP models. However, because of its rotating mechanism, the SeaWinds antenna beam geometry varies across the subsatellite track. As reported by *Portabella and Stoffelen* (2002a) and *Stiles et al.* (2002), the quality of the retrieved winds depends on the azimuth angle separation among beams (or views), i.e. on the azimuth diversity. The poorer the azimuth diversity, the lower the quality of the retrieved winds is. In particular, the nadir region of the QuikSCAT swath has poor azimuth diversity, i.e. inner and outer views are close in azimuth and fore and aft views are close to 180° apart. This region represents a considerable portion of the QuikSCAT inner swath, i.e. about 500 km. Therefore, in order to successfully assimilate QuikSCAT winds into NWP models, additional effort is required to improve the wind retrieval in the nadir region.

Portabella and Stoffelen (2002a) extensively examined the wind retrieval problem in the QuikSCAT nadir region, determining why the standard wind retrieval procedure produces such poor quality winds and proposing an alternative method, i.e. the multiple solution scheme (MSS), to overcome such problem. In this introduction (see following subsections) we extensively reproduce the work performed by *Portabella and Stoffelen* (2002a). In chapter 2, we compare the standard wind retrieval procedure with the MSS at 100-km resolution, using the European Centre for Medium-range Weather Forecast (ECMWF) model winds as reference. In chapter 3 we address the problem of quality control (QC) at 100-km. Finally, in chapter 4, the summary and conclusions are presented.

1.1 Cost function

In remote sensing, the relationship between any observation or set of observations and one or more geophysical state variables is generally represented with the following equation:

$$\mathbf{y} = K_n(\mathbf{x}) \tag{1}$$

where $\mathbf{y}$ is the vector of observations, $\mathbf{x}$ is the vector of state variables that $\mathbf{y}$ depends on, and the operator K_n is the so-called forward model, which relates the state variables to the observations; the subscript n reminds us that it might be non-linear. The process of deriving the best estimate of $\mathbf{x}$ for a given $\mathbf{y}$, allowing for observation errors, is called inversion. There are several approaches for inverting remotely sensed variables, including Bayes' theorem, exact algebraic solutions, relaxation, least squares estimation, truncated eigenvalue expansions, etc (*Rodgers, 2000*). The most general approach to the problem is the Bayesian approach. This approach is also used in scatterometry, where the inversion process is highly non-linear.

Several optimization techniques, which depend on the desired statistical objective, can be applied when using the Bayesian approach, including maximum likelihood, maximum posterior probability, minimum variance, minimum measurement error, etc. The maximum likelihood estimation is the most commonly used technique to invert winds in scatterometry (*Pierson, 1989; Stoffelen, 1998*).

For SeaWinds, the Maximum Likelihood Estimator (MLE) is defined as [adopted from *JPL* (2001)]:

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

where N is the number of measurements, σ_{mi}^o is the backscatter measurement, σ_{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 (noise). Strictly speaking, when assuming Gaussian errors, a term $\ln(Kp(\sigma_{si}^o))$ should be added to the right-hand side of equation 2 but this term is not significant and, as such, is not used. [Note: the Kp is usually taken proportional to either σ_{mi}^o or σ_{si}^o; the latter is chosen to derive winds at 25-km resolution, following the MLE definition for QuikSCAT given by the Jet Propulsion Laboratory

(JPL). On the other hand, recent experiments (see Appendix A) seem to indicate that, for SeaWinds, a *Kp* proportional to $\sigma_{mi}{}^o$ is slightly better than a *Kp* proportional to $\sigma_{si}{}^o$ at 100-km resolution (see equation 2); as such, the former is used in chapter 2.]

According to the Bayes' theorem, the MLE value represents the probability of a trial wind vector (solution) being the "true" wind. The SeaWinds optimization technique consists of looking for the minima of equation 2, which represent the local solutions with maximum probability of being the "true" wind. Since it is computationally expensive to search for minimum MLE in the entire wind domain, the following procedure is usually applied in scatterometry:

- For a particular wind direction, the minimum MLE is searched as a function of wind speed, which, in contrast with wind direction, behaves quasi-linearly and a single well-determined minimum is usually found. The search is generally performed at the speed step size given by a look-up-table (LUT) (0.2 m/s for QuikSCAT).

- The same operation is repeated for every wind direction, at the step size given by the LUT (2.5° for QuikSCAT). The resulting minimum MLE as a function of wind direction is referred to as MLE cost function.

In the standard wind retrieval procedure, the MLE cost function is searched for minima. There are typically up to four minima, which are called ambiguous wind solutions. A spatial filter or ambiguity removal (AR) scheme is then used to select the observed wind field from the ambiguous wind field.

1.1.1 Wind retrieval skill

The MLE (see equation 2) can be interpreted as a measure of the distance between a set of $\sigma_{mi}{}^o$ values 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 $Kp(\sigma_{si}{}^o)$ (*Stoffelen and Anderson, 1997b*). The shape of the MLE cost function is determined by the σ° modulation of any view and the relative geometry among views. By using the MLE cost function minima in the retrieval (standard procedure), the shape of the cost function will determine the skill of the wind retrieval.

Figure 1 shows an example of the MLE cost function for QuikSCAT as a function of wind direction. The diamond symbols indicate the ambiguous wind solutions detected by the inversion procedure. The shape of the minima determines the accuracy of the wind retrieval. The broader the minima, the less accurate the retrieved winds are, since we are ignoring the neighbouring wind solutions to the minima, which are of comparable probability of being the "true" wind, i.e., comparable MLE value. The depths of the minima relative to each other determine in this case the likelihood of each ambiguous solution of being the "true" wind and therefore the ambiguity or uncertainty of the system. The closer the depth of the secondary minima to that of the primary (deepest) minimum and the larger the number of (deep) minima, the more ambiguous the wind retrieval is.

The modulation of the cost function (difference between maximum and minimum in Figure 1) is also important in terms of wind retrieval accuracy. It shows how unlikely the lowest likelihood points of the cost function are compared to the highest likelihood points. For example, the low

GMF modulation at low winds results in a low cost function modulation. In this case, the wind direction solutions coming out of the inversion are not so meaningful anymore, since the standard procedure is ignoring many cost function points of comparable probability to that of the ambiguous solutions. As such, a low cost function modulation corresponds to a low wind direction skill[1].

The MLE cost function is an output of the inversion, and as such is reflecting the inherent inversion problems. Using the minima of the MLE cost function as the only ambiguous wind solutions can lead to poor quality retrievals. As we will see in section 1.3, if we properly use the information on accuracy and ambiguity derived from the MLE cost function (inversion), the wind retrieval may improve significantly.
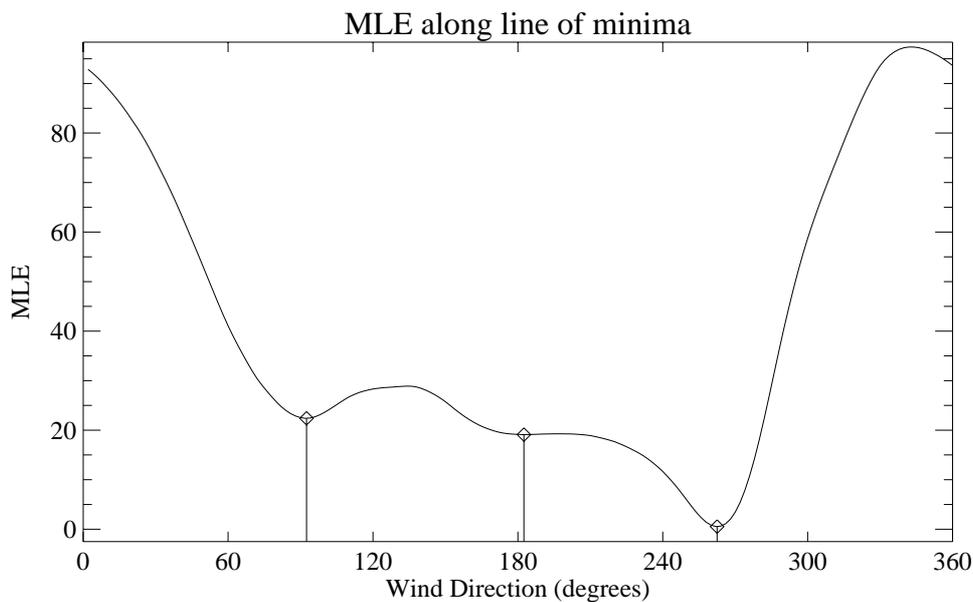

Figure 1 Example of MLE cost function for QuikSCAT node number 33. The diamond symbols indicate the locations of the minima found by the inversion procedure.

1.1.2 QuikSCAT example

As already mentioned, the wind retrieval performance decreases in certain regions of the QuikSCAT swath. This is an inherent problem of the QuikSCAT inversion, which is reflected in the shape of the MLE cost function.

The example shown in Figure 1 corresponds to node number 33. This WVC is inside the nadir region (WVC numbers 29-48), close to the sweet region (WVC numbers 9-28 and 49-68). As we approach the nadir sub-track of the satellite (nodes 38 and 39) and the azimuth diversity decreases, the MLE cost function minima tend to become broader and therefore wind retrieval less accurate. In contrast, when approaching the sweet region and the azimuth diversity increases, the minima become steeper and consequently the wind retrieval more accurate. In the outer region (WVC numbers 1-8 and 69-76), the wind vector is not anymore overdetermined since

---

[1] Wind direction information is meteorologically less meaningful for low winds. We generally find that the wind vector error does not depend on wind speed.

there are only two views. The MLE cost function will have most of the times four minima with nearly equal and low MLE values. The outer region is therefore the most ambiguous of the QuikSCAT swath. The minima in this region will be steep and therefore as accurate as those in the sweet swath, except for the nodes at the edges of the swath, where the two outer views are close to each other (poor azimuth separation) and therefore broad minima in wind direction are again present.

In order to better illustrate the QuikSCAT inversion problem, we have inverted QuikSCAT winds, using inversion software available at the Royal Dutch Meteorological Institute (KNMI), and performed collocations with ECMWF model winds over a period of 12 hours (more than 7 orbits).

Figure 2 shows the two-dimensional histograms of the 1st rank (deepest cost function minimum) KNMI-retrieved wind solution versus the ECMWF wind for wind speed (left plots) and wind direction (right plots), and for different parts of the swath: sweet (top plots), nadir (middle plots) and outer (bottom plots) regions. Note that the right plots are computed for ECMWF winds larger than 4 m/s. This is done to avoid noise in the plots, produced by the typical low wind direction skill at low winds, i.e., for a constant wind vector error the wind direction error is increasing with decreasing wind speed. The ambiguity of the system is reflected in the quality of the 1st rank solution. In other words, the deeper the 1st rank in comparison with the secondary minima, the higher the likelihood of the corresponding rank-1 wind to be the "true" wind (higher 1st rank skill), i.e., the lower the ambiguity. It is clearly discernible from the plots that the sweet swath is the region with the best 1st rank skill. It has the lowest bias and standard deviation (SD) values and the highest correlation values of the entire swath in both speed and direction. As expected, the worst 1st rank skill corresponds to the outer regions. The uncertainty or ambiguity is revealed in the wind direction contour plots as data accumulation away from the main diagonal (see 0° line departure in the plots). In particular, the typical 180° ambiguity of scatterometer data is shown as data accumulation along the 180° diagonals. Again, the sweet region (plot b) shows little data accumulation away from the main diagonal, mainly located along the 180° diagonals. In the nadir swath (plot d), the data accumulation away from the main diagonal is larger and somewhat more spread in comparison with the sweet swath, denoting a slightly worse ambiguity problem. Note the large accumulations of data along the 180° diagonals and elsewhere in the outer swath (plot f), denoting the significant ambiguity of the system in these regions.

Figure 3 shows the same as in Figure 2 but for the KNMI-retrieved wind solution closest to the ECMWF wind. The quality of the closest solution gives an idea of the accuracy of the wind retrieval. Note that the wind speed and wind direction contour lines of both the sweet (top plots) and the outer swaths (bottom plots) are close to the diagonal line, denoting high accuracy of the wind retrieval. However, this is not the case for the nadir swath (middle plots). Moreover, the bias and SD values are significantly larger than in the rest of the swath, denoting relatively poor wind retrieval accuracy at nadir.
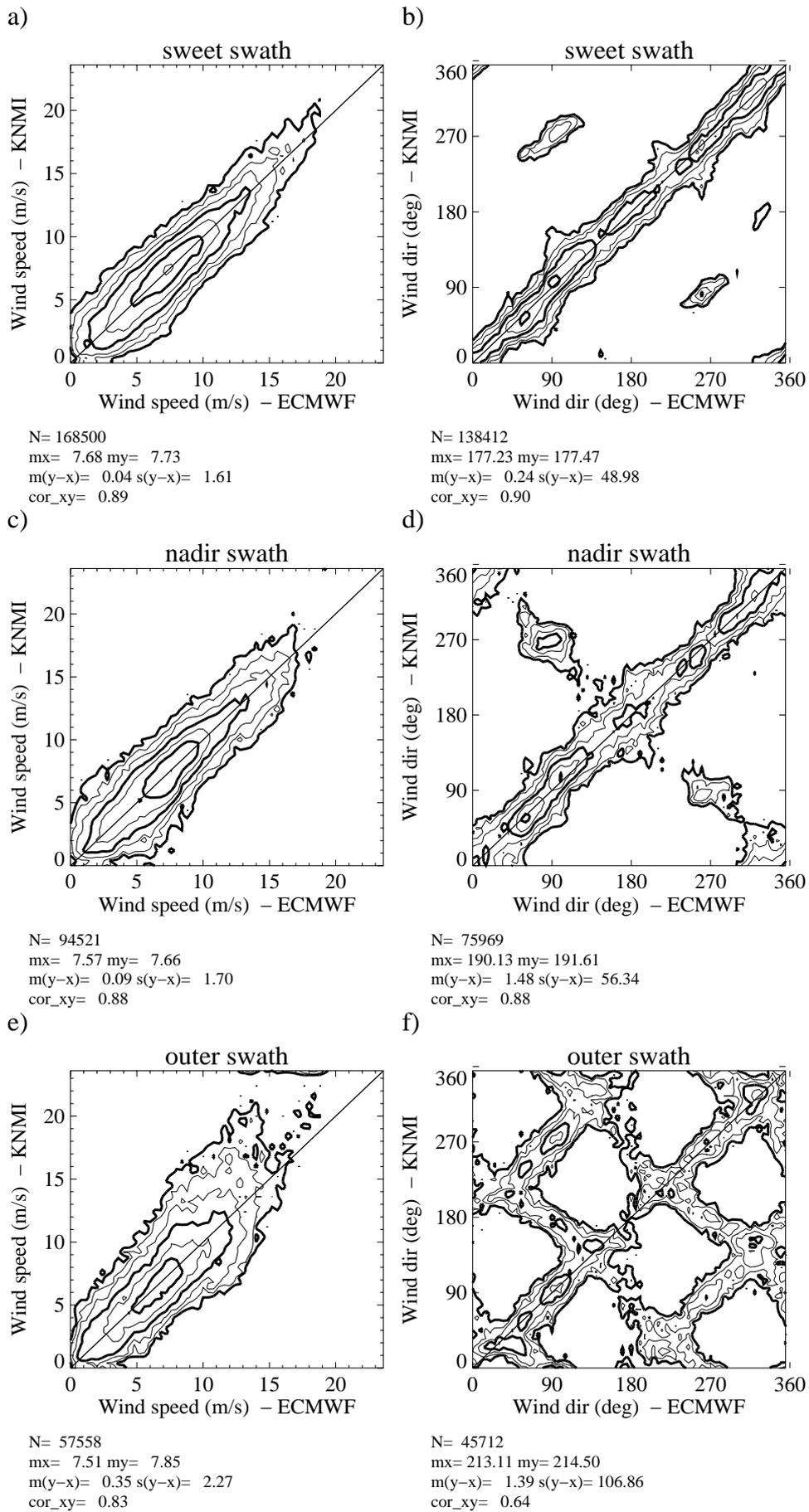
a)

b)

sweet swath

N= 168500
mx= 7.68 my= 7.73
m(y−x)= 0.04 s(y−x)= 1.61
cor_xy= 0.89

N= 138412
mx= 177.23 my= 177.47
m(y−x)= 0.24 s(y−x)= 48.98
cor_xy= 0.90

c)

d)

nadir swath

N= 94521
mx= 7.57 my= 7.66
m(y−x)= 0.09 s(y−x)= 1.70
cor_xy= 0.88

N= 75969
mx= 190.13 my= 191.61
m(y−x)= 1.48 s(y−x)= 56.34
cor_xy= 0.88

e)

f)

outer swath

N= 57558
mx= 7.51 my= 7.85
m(y−x)= 0.35 s(y−x)= 2.27
cor_xy= 0.83

N= 45712
mx= 213.11 my= 214.50
m(y−x)= 1.39 s(y−x)= 106.86
cor_xy= 0.64

Figure 2 Two-dimensional histogram of the 1st rank KNMI-retrieved wind solution versus ECMWF wind in the different parts of the swath: the sweet (top plots), the nadir (middle plots) and the outer (bottom plots) regions. The left plots correspond to wind speed (bins of 0.4 m/s) and the right plots to wind direction (bins of 2.5°). The latter are computed for ECMWF winds larger than 4 m/s. N is the number of data; mx and my are the mean values along the x and y 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 x- and y-axis 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/8000 data points.
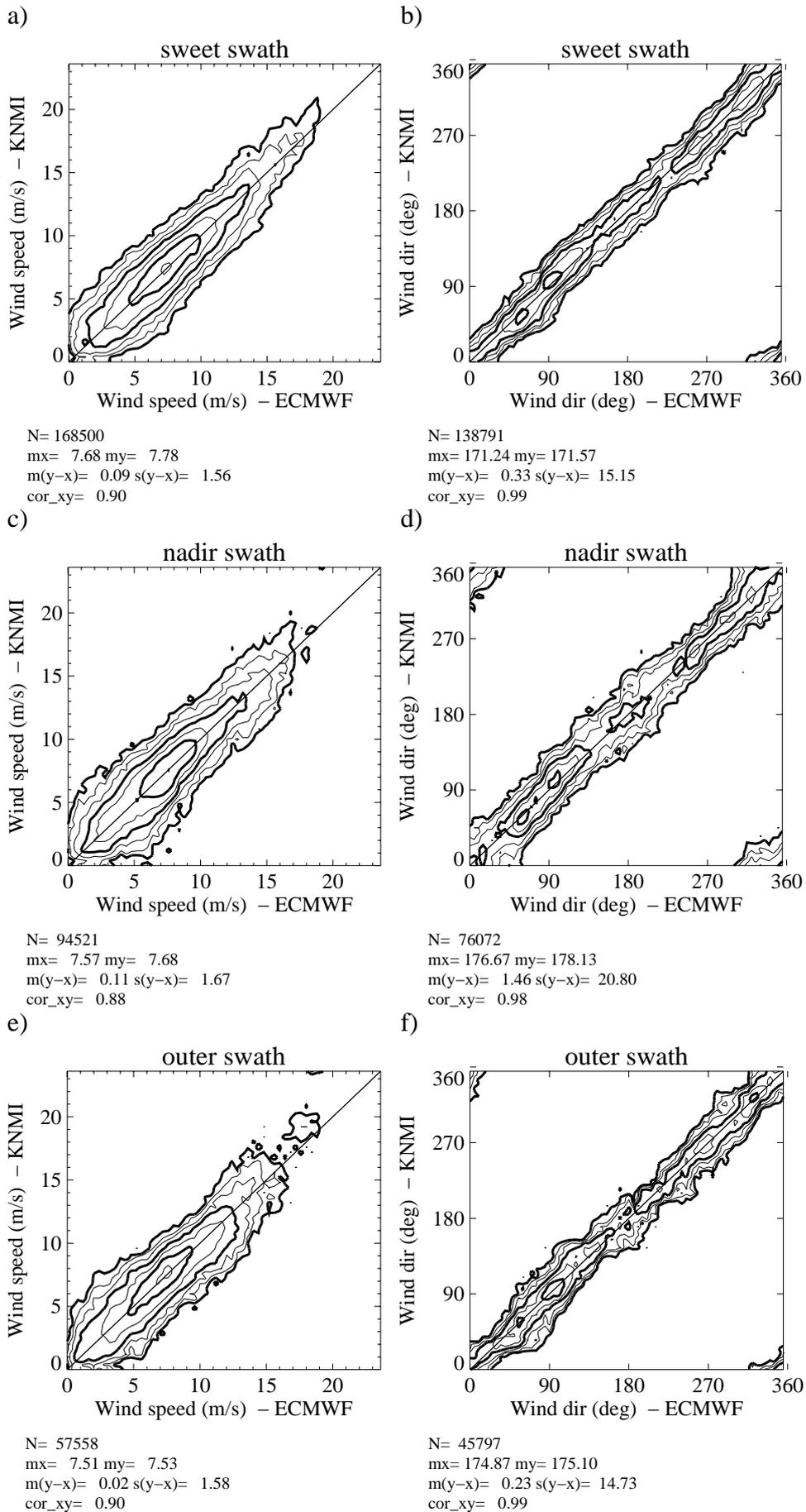
a)

sweet swath

(y-axis: Wind speed (m/s) – KNMI; x-axis: Wind speed (m/s) – ECMWF)

N= 168500
mx= 7.68 my= 7.78
m(y−x)= 0.09 s(y−x)= 1.56
cor_xy= 0.90

b)

sweet swath

(y-axis: Wind dir (deg) – KNMI; x-axis: Wind dir (deg) – ECMWF)

N= 138791
mx= 171.24 my= 171.57
m(y−x)= 0.33 s(y−x)= 15.15
cor_xy= 0.99

c)

nadir swath

(y-axis: Wind speed (m/s) – KNMI; x-axis: Wind speed (m/s) – ECMWF)

N= 94521
mx= 7.57 my= 7.68
m(y−x)= 0.11 s(y−x)= 1.67
cor_xy= 0.88

d)

nadir swath

(y-axis: Wind dir (deg) – KNMI; x-axis: Wind dir (deg) – ECMWF)

N= 76072
mx= 176.67 my= 178.13
m(y−x)= 1.46 s(y−x)= 20.80
cor_xy= 0.98

e)

outer swath

(y-axis: Wind speed (m/s) – KNMI; x-axis: Wind speed (m/s) – ECMWF)

N= 57558
mx= 7.51 my= 7.53
m(y−x)= 0.02 s(y−x)= 1.58
cor_xy= 0.90

f)

outer swath

(y-axis: Wind dir (deg) – KNMI; x-axis: Wind dir (deg) – ECMWF)

N= 45797
mx= 174.87 my= 175.10
m(y−x)= 0.23 s(y−x)= 14.73
cor_xy= 0.99

Figure 3 Same as Figure 2 but for the KNMI-retrieved wind solution closest to ECMWF wind.

In summary, as seen in Figures 2 and 3, the sweet regions show the best wind retrieval skill of the QuikSCAT swath, in terms of ambiguity and accuracy. Although there is a significant ambiguity problem in the outer swath, its accuracy is comparable to that of the sweet swath[1]. The wind retrieval accuracy in the nadir region is significantly poorer compared to the outer and sweet regions of the QuikSCAT swath and has no precedent in scatterometry; as such, special attention should be given to it.

The QuikSCAT azimuth diversity smoothly changes with the node number in the inner swath. In other words, there is no discontinuity between the sweet and the nadir regions. As such, it seems reasonable to consider the sweet swath as well for this study. Therefore, we focus our research on improving wind retrieval in the inner swath (sweet + nadir), giving special attention to the nadir region. However, this does not mean that the methodology applied for the inner swath is not valid for the outer swath. In principle, the results from this study are applicable to the outer swath as well.

1.2 Standard procedure

The scatterometer wind retrieval procedure consists of inversion and AR. In this section, we describe the standard inversion + AR methodology used in scatterometry.

1.2.1 Inversion

The MLE-based inversion has already been discussed in section 1.1. The standard procedure gives up to four ambiguous wind solutions, corresponding to the cost function minima. In the process of deriving such minima, several parameters can be tuned to improve the inversion in terms of ambiguity and quality. An example on how to perform a comprehensive inversion tuning, in this case for QuikSCAT, can be found in *Portabella and Stoffelen* (2002a). The tuning, although improving the overall wind retrieval skill, does not solve any of the already discussed inherent inversion problems.

As an interface between the inversion and the AR, a natural step in scatterometry is to convert the MLE into a solution probability. According to Bayes theorem and the formulation of the MLE explained in section 1.1, the probability of being the "true" wind given a set of scatterometer observations is related by definition to the MLE in the following way:

$$p(v \mid \sigma^o) = \frac{1}{k} e^{-MLE/2},$$
(3)

---

[1] Note that a feature of the closest is that the more solutions are available, the better the apparent quality. However, it is clear that quality degrades with the number of solutions. The outer swath results are therefore too optimistic, since the outer swath represents more solutions (typically four) than the rest of the swath (on average, between two and three).

---

where v represents the "true" wind and σ^o the set of backscatter measurements, and k is a normalization factor. The theoretical relationship is therefore an exponential. In other words, as the MLE, which represents the misfit of the measurements with the solution lying on the GMF surface, increases, the probability of that particular solution being the "true" wind decreases exponentially. In reality, some of the contributions to the observation error are not properly accounted for (see *Portabella, 2002*) and, as such, the shape of the exponential may differ from the theory. A comprehensive characterization of the solution probability for QuikSCAT, based on the empirical methodology described by *Stoffelen et al.* (2000), follows.

Empirical solution probability

- Instead of the MLE, we use a normalized MLE or normalized residual (Rn) used by *Portabella and Stoffelen* (2001) for QuikSCAT quality control (QC) purposes to avoid the already mentioned problem in the measurement noise estimation, such that equation 3 is re-written as:

$$p(v \mid \sigma^o) = \frac{1}{k'} e^{-Rn/l} \tag{4}$$

where k' is again a normalization factor, and l is the parameter that we want to empirically derive. Further details on how the Rn is computed in this case can be found in *Portabella and Stoffelen* (2002a).

- In order to empirically derive equation 4, we can ignore the a priori knowledge on the exponential behavior of the probability, and make the following assumption: There exists a function $p_s(x)$ such that, if we have a set of inversion solutions v_i with normalized residual Rn_i, then the probability that rank j is the one closest to the true wind, denoted by $s=j$, is given by

$$P(s = j \mid Rn_i, i \in \{1, N\}) = \frac{p_s(Rn_j)}{\sum_{i=1}^{N} p_s(Rn_i)} \tag{5}$$

- To determine $p_s(x)$, we concentrate first on only those cases which have exactly two solutions. We process about 2.5 days of QuikSCAT BUFR data and we collocate them with ECMWF winds. The closest solution to the ECMWF wind is used as the "selected" wind. Therefore, we can construct a two-dimensional histogram showing the relative probability of selecting the 1st rank (or the 2nd rank), as a function of Rn_1 and Rn_2. But according to our assumption, by applying equation 5 with $N=2$, we find that the probability of selecting the 1st rank is given by

$$P(s = 1 \mid Rn_1, Rn_2) = \frac{p_s(Rn_1)}{p_s(Rn_1) + p_s(Rn_2)} = \{1 + p_s(Rn_2) / p_s(Rn_1)\}^{-1} \tag{6}$$

- Therefore, by re-arranging equation 6, the two-dimensional histogram gives an estimate of $p_s(Rn_2) / p_s(Rn_1)$ for every combination of Rn_2 and Rn_1. Figure 4a shows such experimentally determined ratios as a function of $Rn_2 - Rn_1$, for several values of Rn_1. Although for $Rn_1 = 2.5$ the ratio is somewhat noisy, it is discernible that the ratio is a fairly invariant function of Rn_2 -

Rn_1. Since Rn_1 is constant and therefore $p_s(Rn_1)$ is also a constant, this plot is actually showing the shape of $p_s(x)$.

- As we know from equation 4, the shape of $p_s(x)$ is exponential and therefore we just have to fit the exponential to the experimental function of Figure 4a by adjusting the l parameter. Figure 4b shows the best fit to Figure 4a, which is represented by the following function:

$$p_s(x) = e^{-x/1.4} \tag{7}$$

where x is representing the Rn.

In order to check whether the assumption is correct and the $p_s(x)$ we found can be generalized for any number of solutions and not only for two, we use the probability function to predict how often a certain solution rank corresponds to the "true" solution for a varying number of solutions and varying distributions of Rn_i (remember that we have used only a few constant Rn_1 values to fit the distributions of Figure 4a).

Tables 1 and 2 compare the predicted distributions over the different ranks with the "observed" distributions (using the closest to ECMWF) in the sweet and the nadir swaths respectively, for the set of about 2.5 days of collocated QuikSCAT-ECMWF data. The number of solutions corresponds to the number of minima in the MLE cost function and the solution ranking goes from the deepest to the shallowest cost function minimum in ascending order. The first row corresponds to the number of data stratified by number of solutions. As shown, when comparing the left side to the right side of the columns, the correspondence is remarkable. Therefore, we
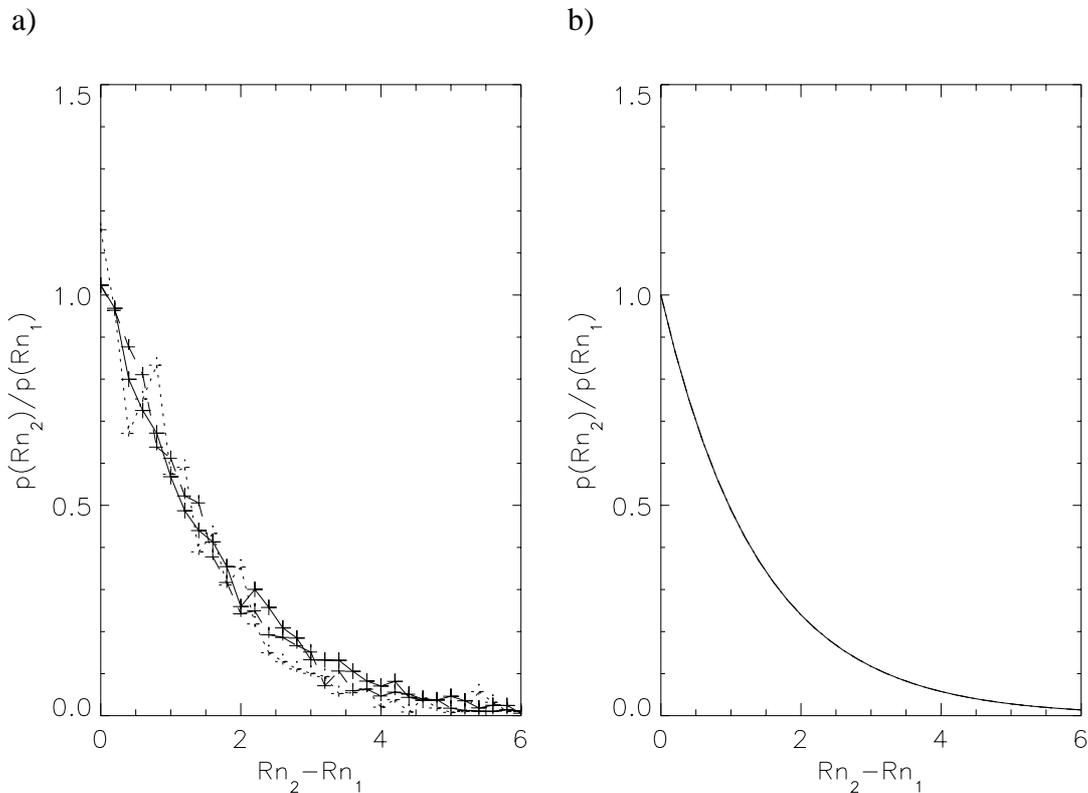
a)   b)

Figure 4 Plot a shows the ratio of the number of realizations of Rn_2 and the number of realizations of Rn1 as a function of $Rn_2 - Rn_1$, and for values of Rn_1=0.1 (solid), Rn_1=1.1 (dashed), and Rn_1=2.5 (dotted). Plot b shows the single exponential fit to the curves of plot a.

conclude that the assumption is correct and that equation 7 can be used to determine the solution probability.

Table 1 Predicted / observed distributions at 25-km (sweet swath).

	2 Solutions	3 Solutions	4 Solutions	All Solutions
Number of Data	331666	233477	317373	882516
Rank 1	91 / 90	82 / 82	77 / 79	84 / 84
Rank 2	9 / 10	15 / 15	18 / 17	14 / 14
Rank 3	-	3 / 3	4 / 3	2 / 2
Rank 4	-	-	1 / 1	0 / 0

Table 2 Predicted / observed distributions at 25-km (nadir swath).

	2 Solutions	3 Solutions	4 Solutions	All Solutions
Number of Data	262753	172506	45638	480897
Rank 1	82 / 80	79 / 79	65 / 66	79 / 79
Rank 2	18 / 20	17 / 17	20 / 19	18 / 18
Rank 3	-	4 / 4	8 / 8	2 / 2
Rank 4	-	-	7 / 7	1 / 1

1.2.2 Ambiguity removal

In order to understand the importance of the solution probability for AR, a brief description of AR follows. The AR is the process of selecting a unique wind vector out of a set of ambiguous wind vectors at each WVC. The AR is not computed in a WVC-by-WVC basis but over many neighbouring WVCs at once. There are two AR techniques, which are commonly used in scatterometry: spatial filters, e.g., median filter for QuikSCAT, and variational analysis.

Median filter

The median of a group of data values is that value for which there are equal numbers of data values of greater and lesser magnitude. This conventional definition of the median can only be applied to non-circular (i.e., linear and scalar) data in which the ordering of the values is obvious. For circular data or vector data such as scatterometer winds, an alternative definition of median is used. The median of a set of data x(1), x(2),…,x(N) is defined as the number x(M) which minimizes:

$$\sum_{i=1}^{N} |x(M) - x(i)| \tag{8}$$

where $1 \le M \le N$.

The medians of circular and vector data calculated using the alternative definition have similar characteristics to the median of non-circular data, i.e., extreme and isolated data are ignored.

The median filter is used by JPL for QuikSCAT AR (*JPL, 2001*) and works as follows:

- The wind field over an entire revolution of scatterometer data is initialised with the help of an NWP model. For each particular WVC, the 1st rank or the 2nd rank wind vector solution, whichever is closer to the NWP field, is selected as first guess wind. The number of ranked solutions used for initialisation does not necessarily need to be two (see section 1.3).

- The wind vectors in a 7 x 7 filter window determine a median vector for the center WVC. The median vector is compared with the ambiguities in that WVC, and the closest ambiguity to the median is selected for use in the next iteration. The entire revolution is filtered in that way. The process continues until it converges, i.e., when no new replacements of vectors have been made.

The MLE (or probability) information is implicitly used in the median filter. The probability can play an important role in the selection of ambiguities used in the initialization and filtering processes (this is further discussed in section 1.3). However, it is never explicitly used in this AR technique.

Variational analysis

The variational analysis is a commonly used technique for data assimilation into NWP models. It consists of combining the background field (NWP) with the observations, assuming that both sources of information contain errors and these are well characterized, to get an analysis field, which is spatially consistent and meteorologically balanced. This analysis field can then be used for scatterometer AR, that is, to select the closest ambiguous wind solution to the analysis field at each WVC. At KNMI, a simple 2D (at surface level only) variational analysis scheme (2D-Var) has been specifically developed for AR (*Stoffelen et al.*, 2000), which attempts to minimize the cost function

$$J(\delta x) = J_b + J_o^{scat}, \tag{9}$$

where J_b is the background term and J_o^{scat} is the observation term. It uses an incremental formulation with the control variable of wind increments, $\delta x = x - x_b$, defined on a rectangular equidistant grid. The control variable x_b is the background field, which in 2D-Var is a NWP model forecast. The forecast is also used as first guess making the control variable equal to the null-vector at the start of the minimization.

The J_b is a quadratic term that contains the inverse of the background error covariance matrix. It penalizes the deviation from the background field. The J_o^{scat} expresses the misfit between the ambiguous wind vector solutions and the control variable at each observation point. The contribution of the wind solutions in each observation point is weighted by the solution probability in the following way (adopted from *Stoffelen and Anderson, 1997a; Stoffelen et al., 2000*):

$$J_o^{scat} = \left(\frac{1}{\sum\limits_{i=1}^{N} K_i^{-p}} \right)^{1/p} \tag{10}$$

where N is the number of solutions and K_i is:

$$K_i = \left(\frac{u - u_i}{\varepsilon_u} \right)^2 + \left(\frac{v - v_i}{\varepsilon_v} \right)^2 - 2\ln P_i \tag{11}$$

where u and v are the wind component control variables; u_i and v_i, the wind solution i in zonal and meridional components, respectively; ε_u and ε_v the corresponding observation errors; and P_i the solution probability.

In order to solve the minimization problem, a conjugate gradients method is used, which also requires the gradient of the cost function. After convergence, the control variable vector of wind increments is added to the background field to obtain the wind analysis. The analyzed wind field is then used for AR, as already discussed.

The solution probability is used explicitly in this AR technique (see *Stoffelen et al., 2000*). It plays a very important role in the minimization and therefore must be characterized in a comprehensive way. In this respect, the empirically derived solution probability, shown in the previous section, is essential for a successful use of a variational AR.

### 1.2.3 Relevance of spatial resolution

KNMI has a NRT 100-km resolution QuikSCAT wind product, which includes inversion, QC and ambiguity removal. *Stoffelen et al.* (2000) show that the 25-km QuikSCAT winds are often too noisy, especially at low winds and in the nadir region. They also show that the averaging of the radar backscatter information, and therefore the reduction of the spatial resolution, significantly reduces the noise of the inverted winds and increases the rank-1 probability (see also *Portabella et al., 2001*). For applications such as mesoscale NWP data assimilation, where the

effective analysis resolution is at least 100-200 km, the use of reduced resolution QuikSCAT winds is effective. In this respect, several High-resolution Limited Area Model (HIRLAM) project countries and ECMWF are now operationally using a reduced resolution QuikSCAT wind processing in data assimilation. As such, a comparison between the 25-km and the 100-km inversions seems appropriate at this stage, and can in turn help to better understand the QuikSCAT inversion problem.

Probability at 100-km

We can perform this comparison in terms of the probability, since it is a closer stage to AR (see section 1.2.1) than the MLE. Therefore, we first compute the probability for the 100-km product, following the same methodology as for the 25-km product (see section 1.2.1):

- The *Rn* is computed at 100-km resolution (see *Portabella and Stoffelen, 2002a*).

- The shape of $p_s(x)$ is found by processing about 10 days of QuikSCAT data and shown in Figure 5 for the same values of Rn_1 as used in Figure 4a. The curves are noisier than in Figure 4a, since the number of data used in the 100-km two-dimensional histogram is about four times smaller than that used in the 25-km histogram (one 100-km WVC corresponds to sixteen 25-km WVCs). Despite this noise, note that the curve of Figure 4b fit also fairly well the curves of Figure 5. Therefore, we also use equation 7 to compute the solution probability at 100-km, where *x* is in this case the *Rn* at 100-km resolution.
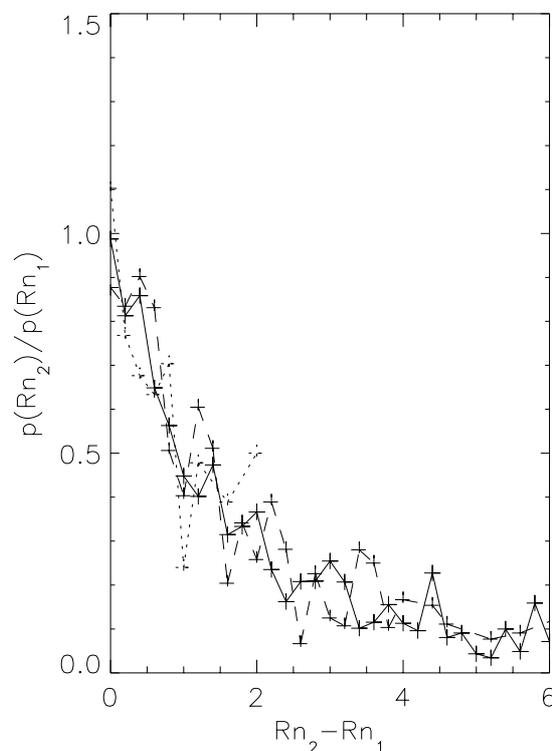
Figure 5 Same as Figure 4a but for the 100-km resolution Rn

- Similar to tables 1 and 2, the results for the verification of the 100-km probability are shown in tables 3 and 4, respectively. The correspondence between the predicted and the observed distributions is also remarkable, confirming the validity of equation 7 for computing 100-km probability.

Comparison

By comparing tables 1 and 2 to tables 3 and 4, respectively, one can clearly see the substantially higher 1st rank skill of the 100-km product, denoting a smaller ambiguity problem (see section 1.1.2), compared to the 25-km product (note the higher percentages of the rank-1 row in the 100-km tables in comparison with the 25-km tables).

In order to compare both products, we have transformed the MLE cost function into a probability cost function by using equation 7. We invert the already mentioned sets of BUFR data (2.5 days for the 25-km and 10 days for the 100-km) and keep the probability cost function information. [Note that discussing about peaks or maxima in the probability cost function is equivalent to the discussion about minima in the MLE cost function]. Figure 6 shows the statistical results of looking at several characteristics of the cost function.

Table 3 Predicted / observed distributions at 100-km (sweet swath).

	2 Solutions	3 Solutions	4 Solutions	All Solutions
Number of Data	53753	67947	73269	194969
Rank 1	97 / 96	94 / 93	92 / 92	94 / 93
Rank 2	3 / 4	5 / 5	7 / 7	5 / 6
Rank 3	-	1 / 2	1 / 1	1 / 1
Rank 4	-	-	0 / 0	0 / 0

Table 4 Predicted / observed distributions at 100-km (nadir swath).

	2 Solutions	3 Solutions	4 Solutions	All Solutions
Number of Data	66618	40478	9344	116440
Rank 1	83 / 83	93 / 93	78 / 74	86 / 86
Rank 2	17 / 17	6 / 6	16 / 19	13 / 13
Rank 3	-	1 / 1	3 / 4	1 / 1
Rank 4	-	-	3 / 3	0 / 0

The top plots of Figure 6 show the histograms of the difference between the maximum (Pmax) and the minimum (Pmin) probabilities for the 25-km (plot a) and the 100-km (plot b) products. The distributions of Figure 6b are much broader and shifted towards higher probability difference values than the distributions of Figure 6a, denoting a better probability modulation and therefore less accuracy (see section 1.1.1) of the 100-km product. Comparing the sweet (solid lines) with the nadir (dotted lines), we see a better probability modulation for the former in both products.

The middle plots of Figure 6 show the histograms of the number of cost function points with probability larger than 10% for the 25-km (plot c) and the 100-km (plot d) products. As discussed in section 1.1, the cost function is computed at the direction step size of the GMF LUT (2.5°) and therefore contains 144 points. The fact of having at least one point above 10% probability is an indication of a good probability modulation since it shows how likely these points are with respect to the remaining cost function points with average likelihood of $\frac{1}{144} = 0.7\%$. In this sense, notice the larger amount of times that the 25-km product cost function does not have any probability value above 10% compared to the 100-km product, showing again a better probability modulation of the latter. In a similar way, if we compare the sweet with the nadir swaths, we notice a larger number of times (the double or more) where no cost function points were above 10% probability in the latter. However, the fact of having more or less points above 10% does not necessarily show a better modulation. For example, if we look at the shape of the distributions in Figure 6d, we see that the nadir swath distribution is shifted towards a larger number of points compared to the sweet swath. Since the nadir swath does not usually have more than 3 solutions (look at the number of data with 4 solutions in comparison with the number of data with 2 or 3 solutions in table 4), the relatively large number of points above 10% probability could be an indication of a flat peak, as expected from this region of the swath (see section 1.1.2).

The bottom plots of Figure 6 show the histograms of the difference between Pmax and the mean probability (Pmean) over an interval of ±12.5° around Pmax for the 25-km (plot e) and the 100-km (plot f) products. This difference gives an indication of the peak modulation. The larger the difference, the steeper the maximum (or main peak) of the cost function and therefore the better the accuracy of retrieved winds is (see sections 1.1.1 and 1.1.2). The larger accumulations of data at low difference values in the nadir swath (dotted) with respect to the sweet (solid) swath confirms the existence of flatter peaks in the former as discussed above. Moreover, this is not only valid for the 100-km product but also for the 25-km product. The reason why we could not infer flat peaks in the 25-km product from the middle plot distributions is that the flat peaks are below the 10% probability level imposed in such plots. However, as we see from the larger accumulation of data at low Pmax-Pmean values in Figure 6e with respect to Figure 6f, the peaks are much flatter (lower peak modulation) at 25-km than at 100-km resolution.

Therefore, we conclude that, for QuikSCAT, the 100-km product is less ambiguous and more accurate than the 25-km product and therefore more suitable for wind retrieval purposes than the 25-km product. In this study, we will therefore use the 100-km product.
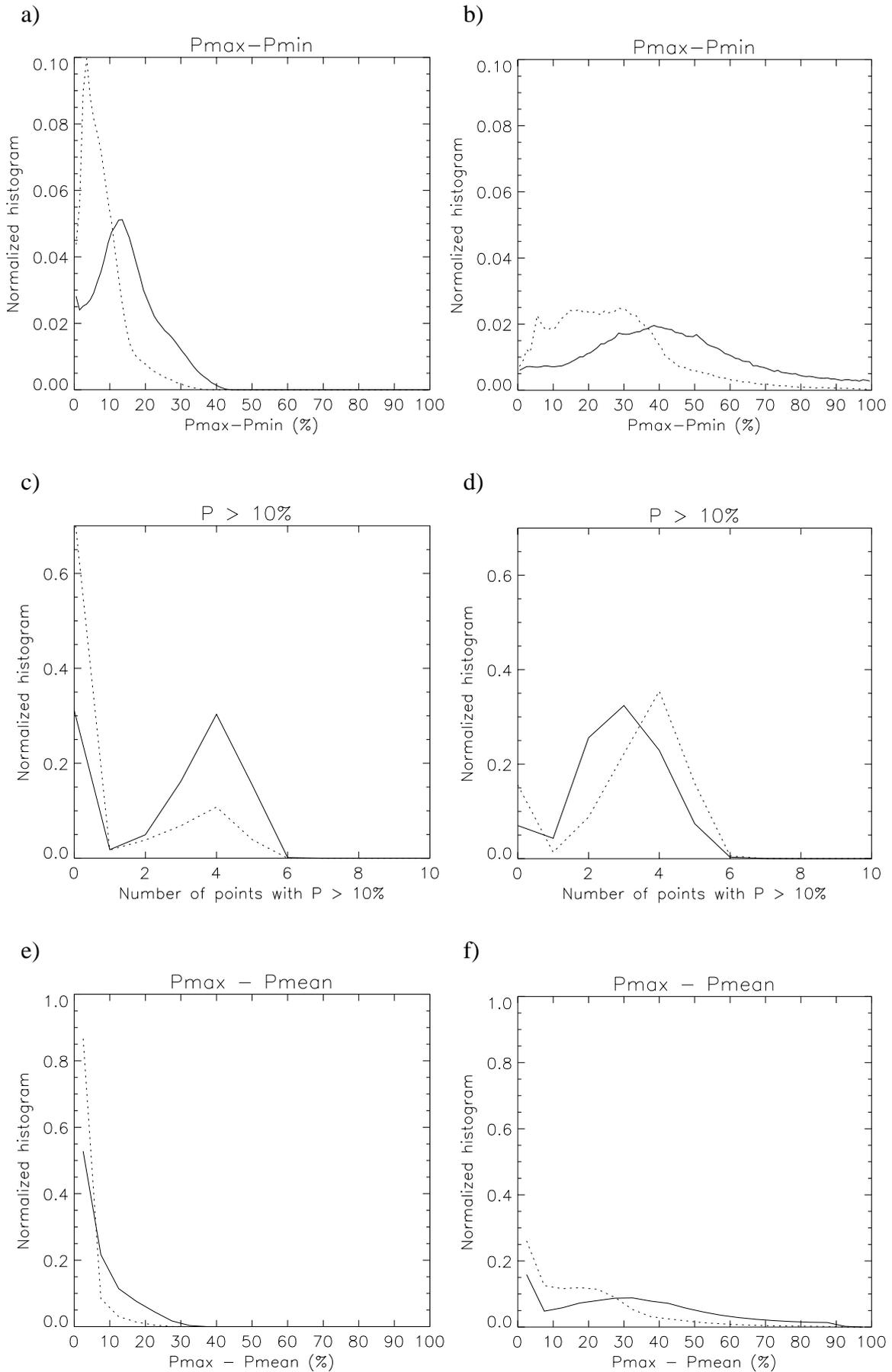
Figure 6 Histograms of the difference between the maximum (Pmax) and the minimum (Pmin) probabilities (top plots), the number of cost function points with probability larger than 10% (middle plots), and the difference between Pmax and the mean probability (Pmean) over an interval of ±12.5° around Pmax (bottom plots), for the sweet (solid lines) and the nadir (dotted lines) regions and for the 25-km (left plots) and the 100-km (right plots) products.

1.3 Multiple solution scheme

So far, we have extensively examined the inversion problem for QuikSCAT and determined the relation between the relative probability of a solution and the MLE in order to prepare QuikSCAT ambiguous solutions for AR. We have learned that in the nadir swath, the accuracy of the inverted winds is low compared to the sweet swath, due to low peak modulation in the probability cost function. For low winds, the accuracy is also low due to the low cost function modulation. The worst scenario therefore occurs for low winds in the nadir swath, where the cost function modulation is rather flat.

The number of solutions in the nadir swath is smaller than in the sweet swath (see the relatively small amount of data with 3 and 4 solutions compared with 2 solutions in table 4, in contrast with table 3). This may be caused by the noise and/or the shape of the cost function, i.e., a cost function that has well defined and steep probability peaks (or MLE minima) may have a larger number of peaks than a cost function that has broad peaks. However, it seems contradictory to provide only few wind solutions to AR when the cost function peaks are less well defined, since these do not represent the full information content of the wind retrieval. Along a broad peak, there are several wind solutions with almost the same relative probability as the peak. However, by selecting only one (as the inversion is doing), we assign zero probability to the rest of the points that belong to the broad peak. On the other hand, by selecting all of the points of the broad peak, we are transferring to AR all retrieved quality information; that is, the inversion could not find a clear candidate for that particular region of the cost function, but rather a few candidates with comparable probability.

Precedent

At JPL a procedure, based on a multiple solution inversion output (not constrained to four solutions) in combination with AR, called DIRTH (*Stiles et al., 2002*) was developed. It includes an initialization technique for the median filter, called the Thresholded Nudging (TN), and a multiple solution selection scheme as input to the median filter, called the Direction Interval Retrieval (DIR).

The TN allows for more than two ambiguities in the initialization (see section 1.2.2) and works as follows. The probability[1] of the cost function is normalized with the probability of rank 1, and the number of ambiguities (up to four) with normalized probability above 0.2 is used in the initialization.

The DIR performs AR in the following way. Given a threshold T (0.8), a set of cost function points around each of the local maxima (resulting in as many segments as local maxima) is selected such that the number of points is minimized and the integral of the cost function over the interval of such points is T. Then, AR is performed in the usual manner (except for using the TN for initialization), and only the segment of points around the selected ambiguity is further used by the median filter (see section 1.2.2).

---

[1] *Stiles et al.* (2002) use the theoretical relation between MLE and probability, i.e., equation 3, to compute the latter.

By examining many wind field cases, we conclude that the DIRTH winds are often very smooth and unrealistic in the nadir swath. Here we identify some possible reasons for this result:

- By applying the median filter only on the segment that was selected in the first place by the "traditional" AR, the scheme is subject to the accuracy of the latter. That is, if the traditional AR fails in an area and produces the wrong solutions, all the segments used in that area will in turn produce a more or less smooth field (probably following some segment extremes, depending on the segment width) but wrong.

- When using a threshold T of 0.8 to define the segments, it may well happen that the remaining cost function points that sum a probability of 0.2 (1-T) contain valuable information indeed. In particular, if we look at the Pmax - Pmin distributions in the nadir swath for 25-km resolution (Figure 6a), we see a relatively poor probability modulation. In such region, many cost function points with substantial probability may be left out of the segment selection. This will in turn decrease the quality of the wind retrieval.

The reason for setting such threshold T is to prevent oversmoothing. That is, if we use T=1, all data in the cost function will be used by the median filter, which in turn will result in a wind field inhibited by the NWP reference and the median filter characteristics. This is due to a very important limitation of the median filter AR, which is not explicitly using the relative probability of each solution, but rather considering all the solutions with identical probability. Despite the mentioned threshold and as already discussed, the resulting wind field is still substantially smooth in areas with large solution segments, i.e., the nadir region. Since the median filter does not ensure meteorologically balanced fields, the retrieved winds are not only oversmoothed but also unrealistic in some (of such) areas.

Alternative

The 2D-Var AR (see section 1.2.2) explicitly uses the probability of all ambiguous solutions. This AR therefore allows the possibility of using as many ambiguous solutions as we desire without a substantial risk of oversmoothing. Moreover, since the variational analysis is always constrained to spatial consistency and meteorological balance, we can ensure realistic retrieved winds by using a scheme based on a multiple solution inversion output in combination with such AR.

Figure 7 shows a QuikSCAT retrieved wind field, using the standard inversion output (up to four ambiguous wind solutions) and the 2D-Var AR. In the nadir region, it is clearly discernible that the retrieved wind field is spatially inconsistent. Since the 2D-Var analysis field (not shown) is spatially consistent, the problem is most likely in the ambiguous solution distribution.

Figure 8a shows the standard ambiguous solution distribution (MLE cost function minima) for the same case of Figure 7. As we can clearly see in the nadir region, the wind solution pattern shows almost no solutions in the direction of the mean flow. Therefore, even if the 2D-Var analysis field were of acceptable quality, there is no way to select a consistent wind field from such solution pattern.
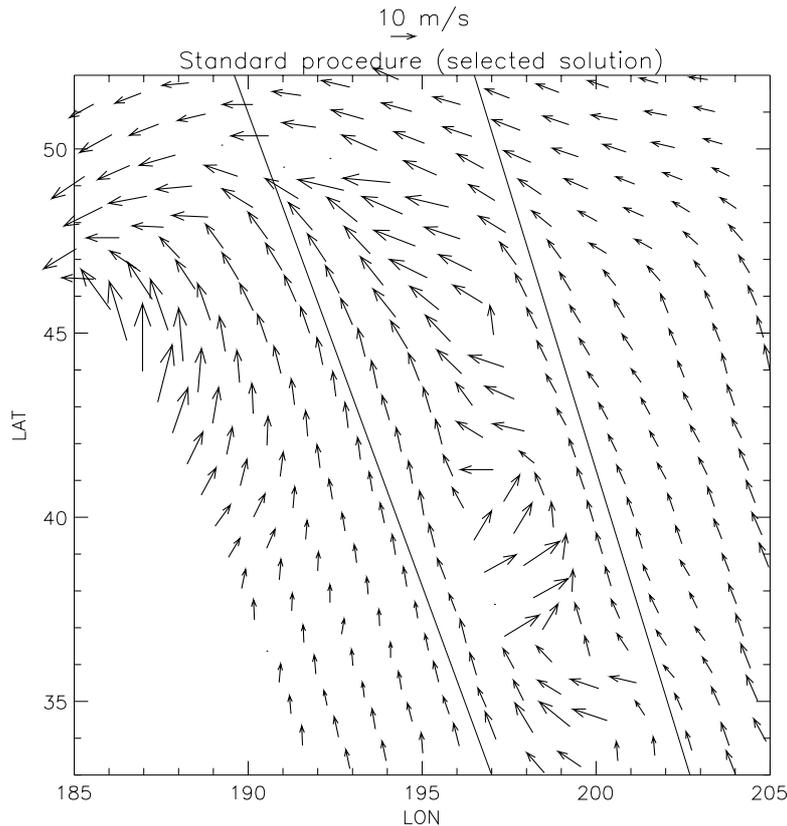
Figure 7 QuikSCAT retrieved wind field using the standard inversion output (cost function minima) and the 2D-Var AR. The acquisition date is January 15 2002 at 16 hours UTC. The solid lines separate the sweet-left (left side), the nadir (middle), and the sweet-right (right side) regions of the QuikSCAT swath.

Figure 8b shows the multiple ambiguous solution (not constrained to four) distribution again for the same meteorological case as Figures 7 and 8a. We show all the cost function solutions with probability above a guessed threshold[1] of $2x10^{-7}$. Notice how often the ambiguous solutions in the sweet swath are around the cost function minimum, which is in the direction of the mean flow, denoting little ambiguity (main cost function minimum much deeper than the remaining minima) in comparison to the nadir swath. Note also that the number of solutions in the nadir region is large, indicating lower accuracy (broader minima) than in the sweet swath. In comparison with Figure 8a, we are providing much more information content to the AR using the multiple solution inversion output. As already discussed, the 2D-Var uses the information in an appropriate way (the ambiguous solutions are weighted by their computed probability) and therefore, from a theoretical point of view, the multiple solution concept may considerably improve the resulting analysis field. Moreover, the AR will now result in a spatially consistent wind field since the multiple solution concept does provide solutions aligned with the mean flow (see solution distribution in the nadir swath of Figure 8b). [Note: the dots in Figures 7 and 8 represent quality-controlled points. This issue is discussed more in depth in chapter 3.]

It seems reasonable to test the multiple solution scheme (MSS) against the standard procedure. Since using all the points of the cost function with non-zero probability (up to 144) as solution ambiguities for the 2D-Var AR is computationally expensive, we use the mentioned probability thresholds, i.e., 10^{-5} for the standard procedure and $2x10^{-7}$ for the MSS, as a first guess.

---

[1] The reason for choosing a different probability threshold in the standard procedure and the MSS is due to the normalization of the probability; the former is normalized with up to 4 solutions and the latter with up to 144.

a)

10 m/s

Standard procedure (solution pattern)
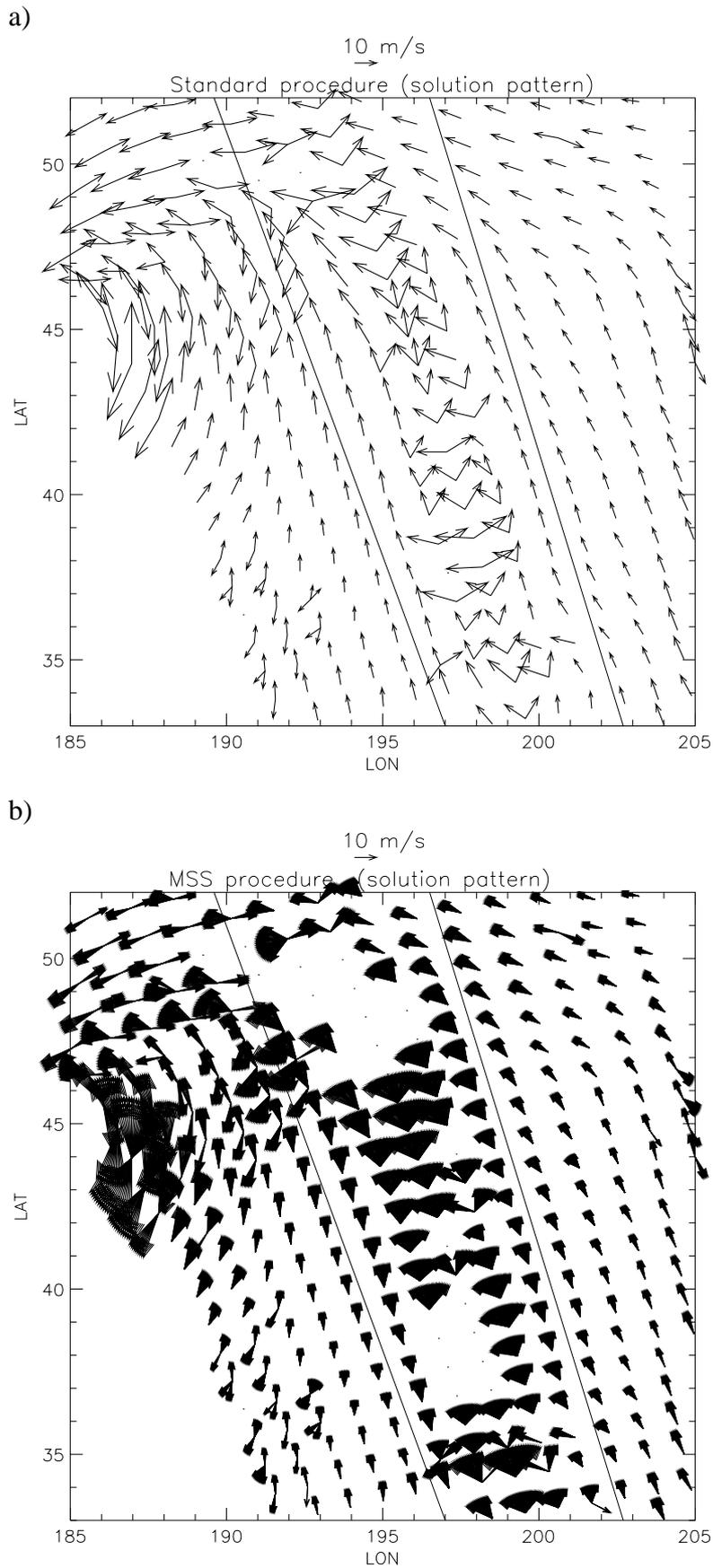

b)

10 m/s

MSS procedure (solution pattern)

Figure 8 Same as Figure 7 but for QuikSCAT ambiguous wind field using (a) the standard inversion output (cost function minima); and (b) the multiple solution scheme. Only solutions with probability above 10^{-5} (a) and $2x10^{-7}$ (b) are shown.

2 Comparison between the standard procedure and the MSS

As discussed in section 1.2.2, the 2D-Var background term is a NWP forecast field. The QuikSCAT data products distributed by JPL and the National Oceanographic and Atmospheric Administration (NOAA) include collocated National Center for Environmental Prediction (NCEP) wind information. The latter is used for AR purposes, i.e., as background term. As such, a different reference should be used to compare the standard wind retrieval and the MSS procedures. In this study, we use ECMWF winds as reference.

2.1 Statistical results

Three days of QuikSCAT and ECMWF collocated winds at 100-km resolution are used in the comparison. Table 5 shows the mean root-mean-square (RMS) of wind vector differences between ECMWF and three different wind sources: standard wind retrieval, MSS and NCEP. Comparing the standard procedure and the MSS, the latter shows better performance, i.e., agreement with ECMWF. As expected, the major difference between the two procedures is in the nadir region, where the RMS difference is more than 0.5 m/s lower for the MSS. In the sweet swath, the MSS also works better. This is due in part to an improvement at low winds, where low cost function modulation is expected, and in part to the improvement of the analysis field, i.e., a better 2D-Var analysis in nadir is expected to positively impact the analysis in the sweet regions. Indeed, the results (see table 6) indicate better agreement of MSS analysis (compared to standard analysis) with ECMWF in both the sweet and the nadir swath.

Table 5 Mean vector RMS[1] (m/s)

Swath region	Standard procedure	MSS	NCEP
Sweet	2.48	2.23	2.85
Nadir	2.98	2.45	2.96

[1] The vector RMS is referred to as the RMS of the wind vector difference between ECMWF and the different wind sources shown in the table.

Both the standard procedure and the MSS show generally better scores (against ECMWF) than NCEP (see table 5). This suggests that 2D-Var is successfully exploiting the observations rather than to follow the background (i.e., NCEP). As such, the quality of the background does not significantly affect the quality of the retrieved winds. This is also true in the nadir region. As discussed in section 1.3, the MSS provides a larger number of equally likely ambiguous solutions in the nadir swath, compared to the sweet regions, thus resulting in a larger influence of the background term in 2D-Var. However, the impact of NCEP in the nadir is also minor, as seen from the substantial difference in vector RMS between the MSS (2.45 m/s) and NCEP (2.96 m/s). The observations and the constraints on meteorological balance and spatial consistency are therefore the most dominant factors in the retrieval.

Swath region	Standard procedure (analysis)	MSS (analysis)
Sweet	2.14	2.04
Nadir	2.39	2.24

[1] The vector RMS is referred to as the RMS of the wind vector difference between ECMWF and the 2D-Var analysis of the different wind sources shown in the table.

Figure 9 shows the two-dimensional histograms of the selected solutions by the standard procedure (top plots) and by the MSS (bottom plots) against ECMWF winds, for wind speed (left plots) and wind direction (right plots), in the nadir swath. The MSS shows a slight improvement in the wind speed accuracy compared to the standard procedure, as denoted by their corresponding SD values (see left plots). The main improvement is in wind direction. It is clear that the contour lines in Figure 9d are closer to the diagonal than those of Figure 9b. The better wind direction accuracy of the MSS is confirmed by the SD scores, where the standard procedure is more than 4° higher than the MSS. The fact that the main improvement is in wind direction is an expected result since the MSS leaves essentially a larger wind direction choice to the AR procedure (i.e., 2D-Var) than the standard procedure. The fact that the MSS choice of wind direction also improves the wind speed scores indicates a more consistent selection for MSS.

The overall results (table 5) show that the difference in wind vector accuracy between the nadir and the sweet regions is 20% for the standard procedure, while only 10% for the MSS. This is mainly due to the substantial improvement of the MSS in wind direction accuracy at nadir. The MSS clearly reduces noise as compared to the standard procedure, due to the spatial smoothing constraints, i.e., flow rotation and little divergence, and the improved J_o^{scat} (equation 10). We now further investigate the effect of J_o^{scat}.

MSS probabilistic behavior

A way to test the consistency of the MSS is to verify the a priori probabilities of the solutions. Figure 10a shows how often a solution with a particular probability value is selected (diamond symbols) or is closest to NCEP (star symbols) as a function of probability. Both the x-axis and the y-axis are in logarithmic scale. As such, the diagonal denotes a consistent probabilistic behavior, i.e., a solution with probability value 10^{-2} (for example) is expected to be "selected" 1% of the time. The closest solution turns out to be probabilistically rather inconsistent as shown by the large discrepancy with the diagonal. This essentially means that if the MSS systematically selects the closest solution, it would be doing a poor job since it would not correct the differences between QuikSCAT and NCEP (background) observing systems, where they exist. The selected solution shows a more consistent probability pattern than the closest, especially in the most populated region, i.e., probabilities between 10^{-2} and $10^{-0.4}$ (see solid line in Figure 10b), where the diamonds clearly follow the diagonal. The reason for this is that many closest-to-NCEP low-probability solutions are not selected and high-probability solutions are selected instead. This indicates that in general 2D-Var is successfully resolving the large number of solutions provided by the MSS, thus confirming the small dependency (of the MSS) on the background discussed at the beginning of this section.
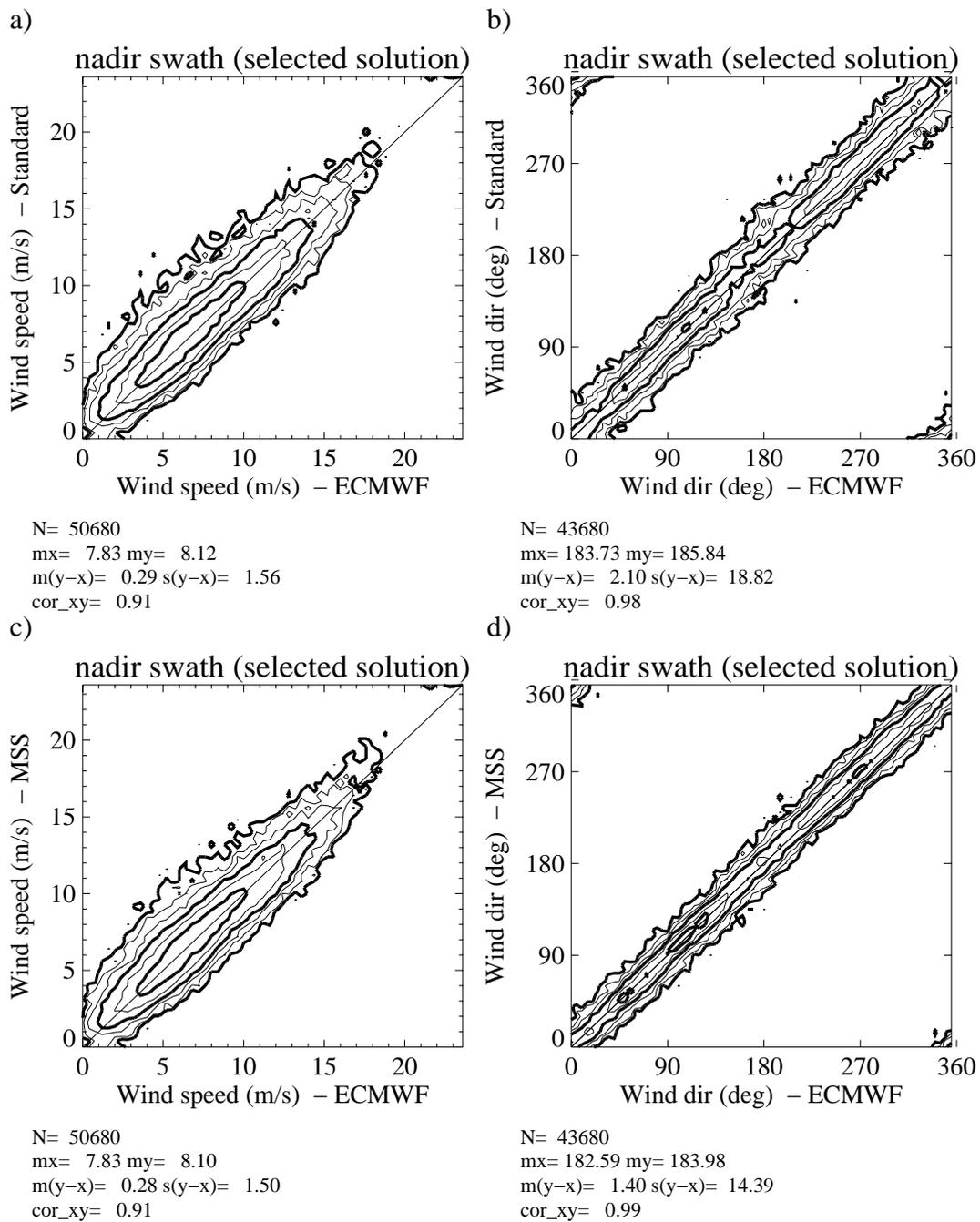
a)

nadir swath (selected solution)

N= 50680
mx= 7.83 my= 8.12
m(y−x)= 0.29 s(y−x)= 1.56
cor_xy= 0.91

b)

nadir swath (selected solution)

N= 43680
mx= 183.73 my= 185.84
m(y−x)= 2.10 s(y−x)= 18.82
cor_xy= 0.98

c)

nadir swath (selected solution)

N= 50680
mx= 7.83 my= 8.10
m(y−x)= 0.28 s(y−x)= 1.50
cor_xy= 0.91

d)

nadir swath (selected solution)

N= 43680
mx= 182.59 my= 183.98
m(y−x)= 1.40 s(y−x)= 14.39
cor_xy= 0.99

Figure 9 Two-dimensional histogram of the selected wind solution by the standard (top plots) and by the MSS (bottom plots) versus ECMWF wind in the nadir region. The left plots correspond to wind speed (bins of 0.4 m/s) and the right plots to wind direction (bins of 2.5°). The latter are computed for ECMWF winds above 4 m/s. N is the number of data; mx and my are the mean values along the x and y 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 x- and y-axis 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/8000 data points.

A remaining question is what to do with both tails of the distribution, i.e., probabilities below 10^{-3} and above $10^{-0.2}$ (see solid line in Figure 10b), where the probabilistic behavior is far from being consistent. Figure 10b shows the quality of the data (star symbols) as a function of probability. Note that the quality is decreasing (i.e., increasing RMS) as we approach the extremes of the

distribution[1]. In particular, below 10^{-4}, the data are of poor quality (close to 4 m/s RMS), indicating that the probability threshold of 2×10^{-7} initially used by MSS (see section 1.3) may be increased to improve the quality of the retrievals. This is a QC issue, which will be further discussed in chapter 3.
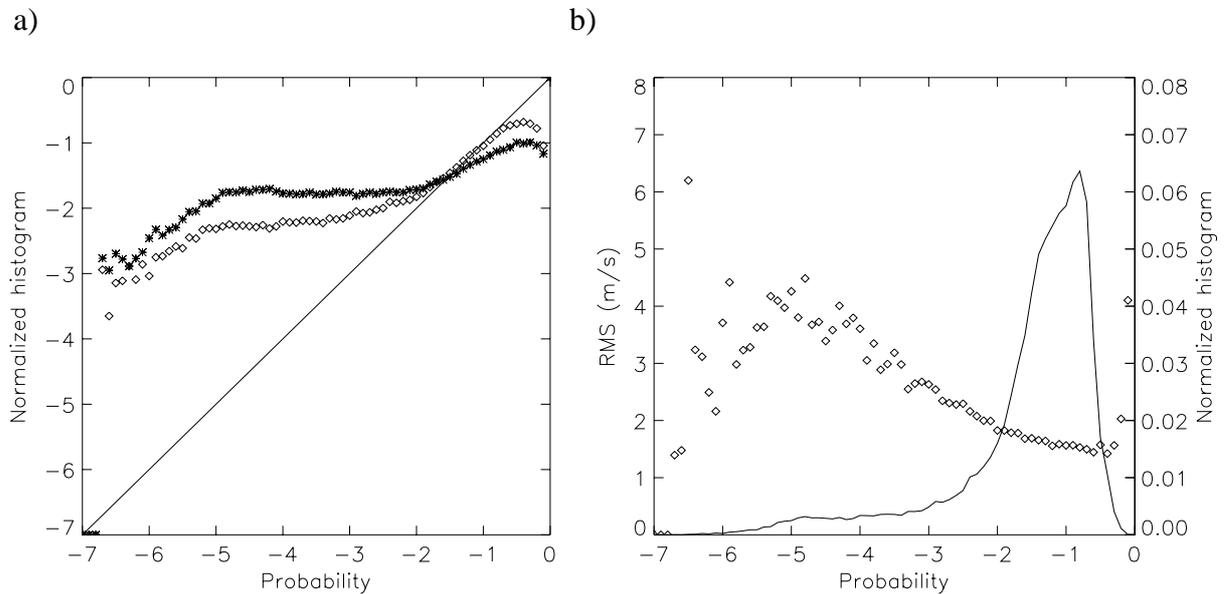
a)   b)

Figure 10 (a) Number of times (normalized and in logarithmic scale) that a solution with a particular probability value is selected (diamond) or closest to NCEP (star) versus probability (logarithmic scale). (b) Normalized histogram of selected solutions (solid line) and mean RMS of vector difference between the selected solutions and ECMWF winds (star) versus probability (logarithmic scale).

2.2 Cases

Many meteorological cases were examined in this comparison. In order to better illustrate the statistical results of the previous section we show some of these cases here [Note: some additional cases are shown in Appendix B].

Figure 11 shows the MSS selected wind field for the same poor-quality case as Figures 7 and 8. As discussed in section 1.3, in contrast with the standard procedure, the MSS provides solutions in the direction of the mean flow in the nadir swath (see Figure 8). As such, a spatially more consistent and realistic wind field is found when using the MSS. This is shown in Figures 7 and 11, especially in the middle of the plot. A few inconsistent wind arrows (probably rain contaminated), which should be quality controlled (see discussion on QC at 100-km resolution in chapter 3), are still present though.

---

[1] Below probability of 8×10^{-6} the number of data is very small (see solid line in Figure 10b) and therefore not statistically significant, as denoted by the noisy RMS values in the left part of Figure 10. This is also true for probability above $10^{-0.2}$.
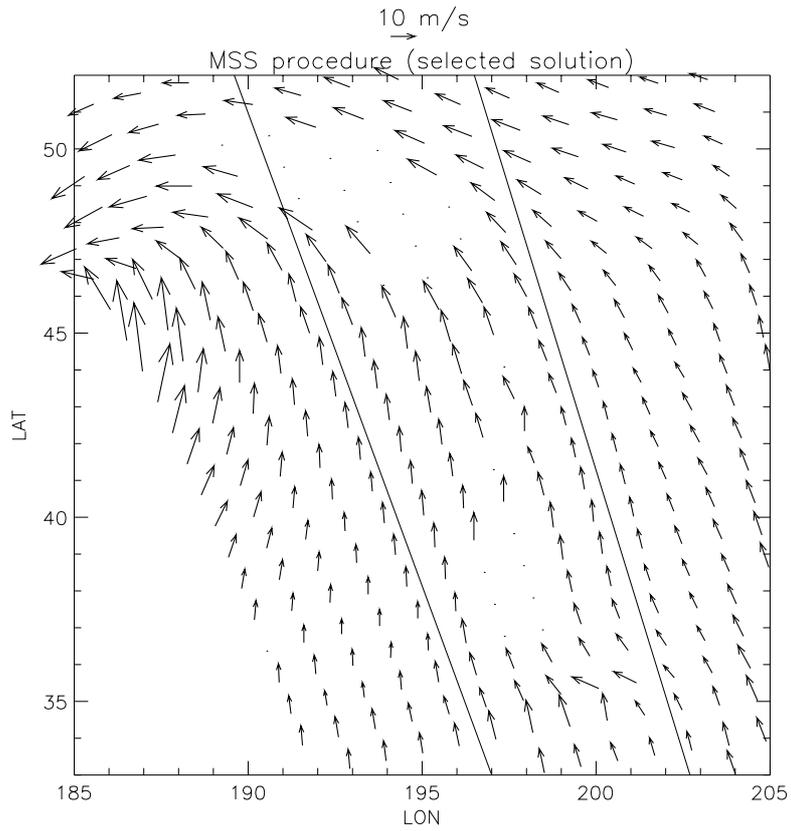
Figure 11 Same as Figure 7 but for MSS retrieved wind field.

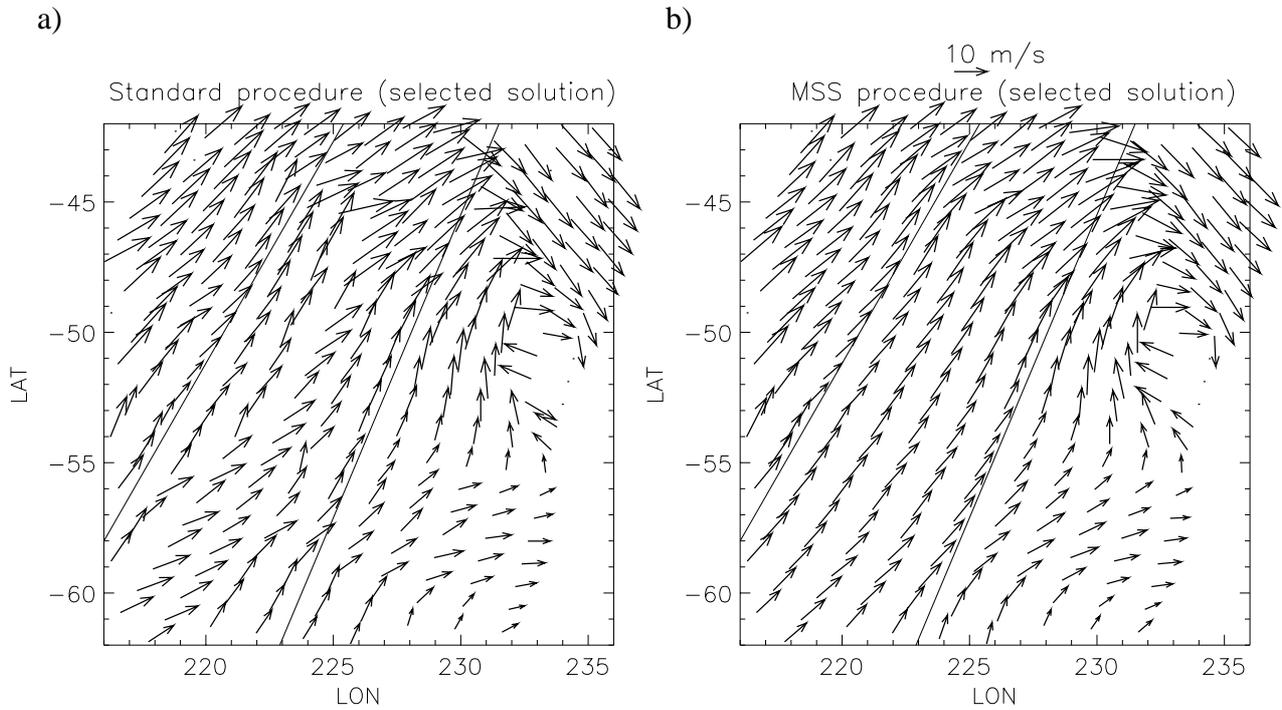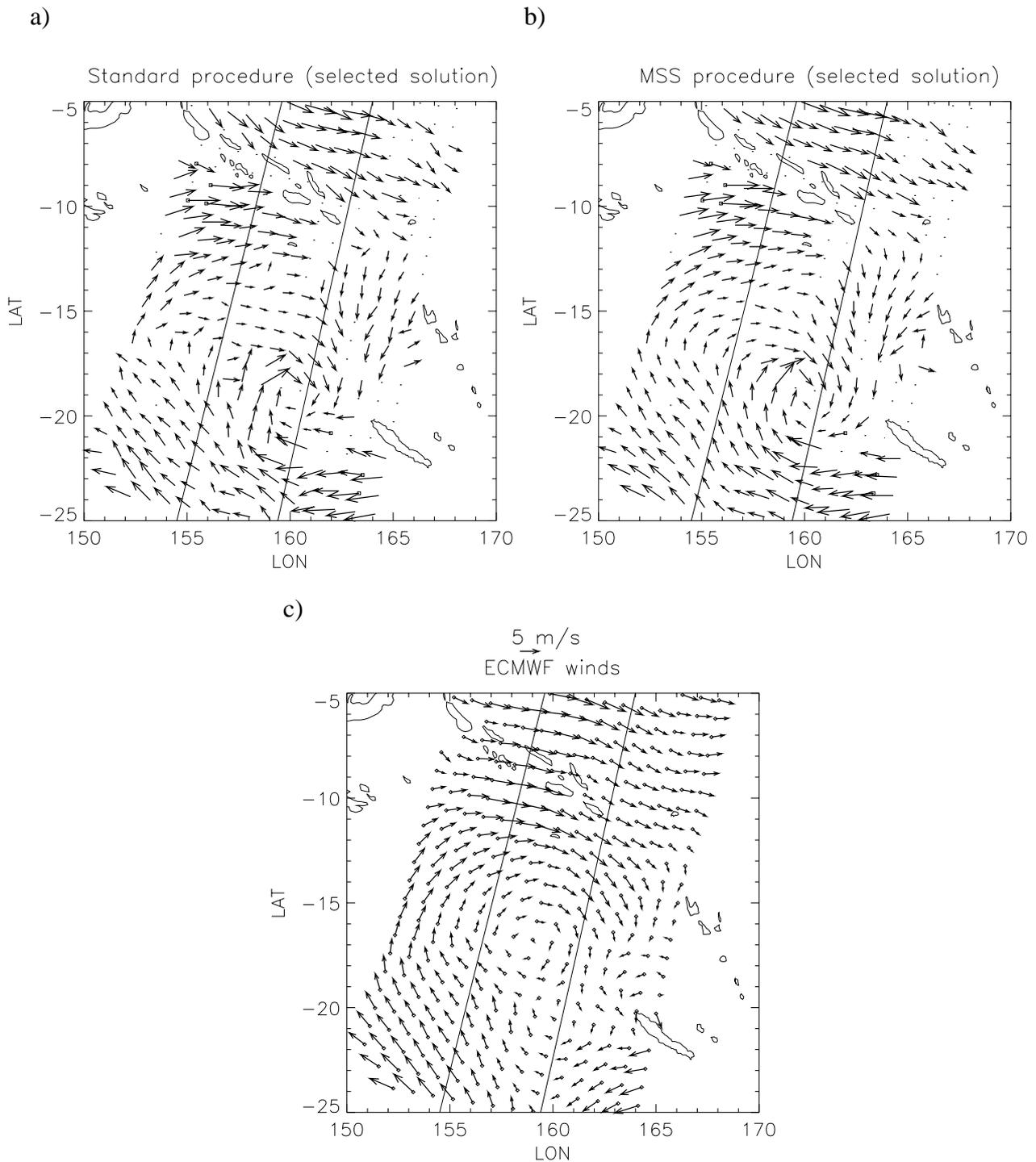a)   b)

Figure 12 QuikSCAT retrieved wind field using the standard procedure (a) and the MSS (b). The acquisition date is February 3 2002 at 2 hours UTC. The solid lines separate the sweet-right (left side), the nadir (middle), and the sweet-left (right side) regions of the QuikSCAT swath.

Figure 12 shows another interesting case of how the MSS is improving the quality of the retrieved wind field in the nadir with respect to the standard procedure. Note the noisy and granular wind field over the entire nadir swath in Figure 12a. The MSS (Figure 12b) is successfully filtering this noise, keeping at the same time the dynamical information of this case (intensity and location of the low-pressure system are the same in both plots).

Figure 13 shows a low wind speed case. Again, the standard wind field (Figure 13a) shows a noisy pattern in the nadir swath, which is successfully filtered by the MSS (Figure 13b). The

a)

b)

c)

Figure 13 QuikSCAT retrieved wind field using the standard procedure (a) and the MSS (b), and ECMWF wind field (c). The acquisition date is February 3 2002 at 7 hours UTC. The solid lines separate the sweet-right (left side), the nadir (middle), and the sweet-left (right side) regions of the QuikSCAT swath.

presence of a low-pressure system is better depicted by the MSS. Moreover, the standard wind field is also somewhat noisy in the sweet swath, as may be expected from the low cost function modulation at low winds (see section 1.3). As shown in Figure 13, the MSS is successfully filtering the noise in the sweet swath as well.

Figure 13c shows the ECMWF wind field. Both the intensity and location of the low-pressure system are in disagreement with the observations. The assimilation of a well-defined and spatially consistent wind field such as the MSS could help very much to improve ECMWF forecast.

3 Need for a Quality control at 100 km resolution

An important aspect of the 100-km product used in this study, which needs to be examined, is the QC. Up to now, the 100-km product is using the MLE-based QC at 25-km resolution (KNMI QC) developed by *Portabella and Stoffelen* (2001) in the following way: if there is sufficient information on the 100-km WVC after QC (at least half of the 25-km WVCs within the 100-km WVC), the wind retrieval is performed.

Problems using 25-km QC in 100-km WVC

The problem of using a 25-km QC procedure in 100-km WVCs is illustrated in Figure 8 (see section 1.3). Figure 8b shows the effects of using the 25-km QC recommended by *Portabella and Stoffelen* (2002b), i.e., KNMI QC[1] + JPL rain flag[2] in the nadir and only KNMI QC in sweet regions, in comparison with Figure 8a, where only the KNMI QC has been applied. On the one hand, as reported by *Portabella and Stoffelen* (2002b), the JPL rain flag is rejecting a considerable amount of consistent winds, as seen in the Northern part (nadir region) of the wind flow (see WVCs with consistent wind solutions in Figure 8a removed in Figure 8b). On the other hand, the 25-km QC (using JPL rain flag) is able to reject several WVCs of poor quality, probably rain contaminated (see the nadir region WVCs with inconsistent solution pattern, both in speed and direction, in the lower half of Figure 8a, removed in Figure 8b). These poor-quality WVCs show zero probability in the direction of the flow (not shown) and therefore it is of great importance to identify these cases and reject them, regardless of the solution scheme, i.e., the standard procedure or the MSS, we use. However, even if the 25-km QC is able to remove most of the poor-quality WVCs, a few of them still remain in Figure 8b (notice the absence of solutions aligned with the mean flow in a few nadir WVCs).

Alternatives

Using the background error spatial structure functions, large discrepancies between the wind solutions provided by the MLE inversion and the analysis (i.e., output from variational AR) can be interpreted as poor-quality retrieved solutions. After a comprehensive validation, a threshold, which relates these discrepancies to the quality of the observations, can be set. This gross error check is the so-called variational QC. The inconsistent nadir winds could therefore be rejected using this QC. Moreover, in contrast with the JPL rain flag, it would generally keep the consistent wind flow. However, the rejection of too many discrepancies with the analysis could lead to a retrieved field too close to the background and, as such, not useful in data assimilation, i.e., the

---

[1] The KNMI QC uses the normalized MLE (Rn) information at 25-km resolution to filter poor quality data, i.e. a Rn threshold, which maximizes the good quality acceptance and the poor quality rejection, is set.

[2] The rain flag developed by JPL (see *Huddleston and Stiles, 2000*) looks for 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 several input parameters including average brightness temperature (both H-pol and V-pol), normalized inter-view σ° difference, wind speed, wind direction relative to along track, and a normalized MLE. The space spanned by these parameters can detect whether the set of σ° values used in wind retrieval is affected by rain.

impact of assimilating observations that are well in agreement with the NWP background is expected to be negligible. Consequently, an extensive testing is required prior to using such QC.

As discussed in section 2.1, the MSS selected solutions with low probability values, i.e., below 10^{-4}, are of poor quality (see Figure 10b). As such, a more straightforward QC (prior to variational analysis) can be set by using a higher probability threshold than the one used by the MSS (i.e., 2×10^{-7}). However, by increasing the probability threshold, we will also decrease the number of MSS ambiguous solutions (see section 1.3). This may lead to some additional noise in the nadir swath, i.e., the lower the range of solutions the larger the number of cases with no solution aligned with the "true" direction. Nevertheless, large discrepancies with the mean flow will most generally occur when the observation is of poor quality. Therefore, a variational QC could then be used to remove such poor quality cases.

Another possibility is to set up a QC procedure for 100-km resolution in a similar way as it was done for 25 km, i.e., computing Rn (at 100km) and setting an optimal threshold in terms of maximum good quality acceptance and poor quality rejection. The 100-km QC would be able to reject the 100-km WVCs that despite they contain good-quality 25-km information (after 25-km QC), they result in poor-quality 100-km winds; for example, a 100-km WVC crossed by a front line, which still contains enough quality controlled 25-km WVCs for wind retrieval.

A way to avoid a decrease in the number of MSS ambiguous solutions and still remove the WVCs that contain low probability selected solutions is to use an appropriate Rn threshold at 100-km resolution. As it is clear from Figure 14, the Rn increases with decreasing probabilities of the selected solution. Since the quality of the data is decreasing with decreasing probabilities (Figure 10b), a Rn threshold would not only remove poor quality data (see above discussion on
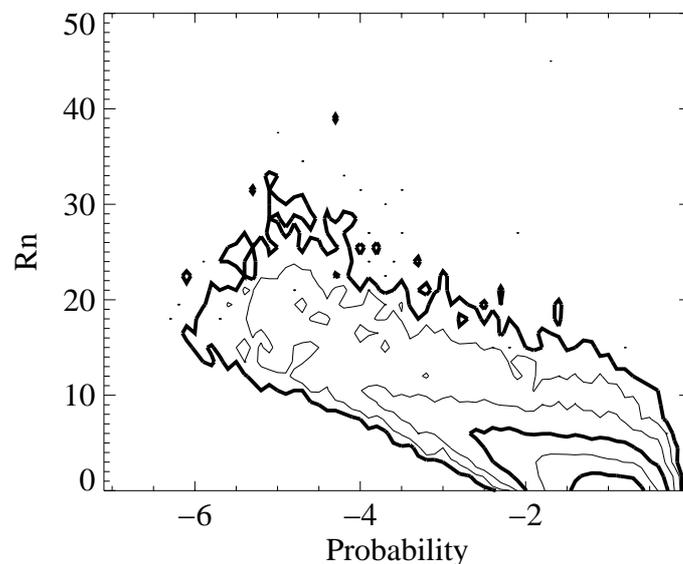

Figure 14 Two-dimensional histogram of the Rn versus probability of the selected solution. The total number of data is 50642. The contour lines are in logarithmic scale (two steps corresponding to a factor of 10 in number density); the lowest level (outer-most contour line) is at 3 data points.

100-km Rn) but also the cases with low probability selected solutions, e.g., a threshold of (let's say) 10 would remove almost all cases with (selected) probability below 10^{-4} and keep most of the cases with probability above 10^{-4}.

In order to define the best strategy for 100-km QC further investigation of the procedures discussed in this chapter is required. A combination of some of these procedures may be more appropriate.

4 Summary and Conclusions

In this study, a probabilistic approach is used to improve the QuikSCAT retrievals, especially in the nadir region, for assimilation purposes into NWP. After the extensive examination of the scatterometer and, in particular, the QuikSCAT inversion problems performed by *Portabella and Stoffelen* (2002a), the standard wind retrieval procedure is compared to a new procedure, the so-called multiple solution scheme, at 100-km resolution. Prior to the comparison, we summarize the most relevant issues investigated by *Portabella and Stoffelen* (2002a):

> In scatterometry, the standard wind retrieval procedure works as follows: the minima of the MLE cost function, considered as the ambiguous wind solutions, are in turn used by the AR procedure to select the observed wind. In these circumstances, it is shown how the shape of the MLE cost function determines the skill of the wind retrieval procedure in terms of ambiguity and accuracy. In particular, for QuikSCAT, the shape gradually changes with the cross-track location (WVC), thus affecting the retrieval skill of the different regions of the swath. In the poor-azimuth-diversity nadir region, where the cost function minima are broad, the accuracy of the retrieved winds is substantially lower compared to the rest of the swath. The standard wind retrieval procedure is therefore further investigated.

> First, and in order to get a more suitable interface between the inversion and the AR schemes, the MLE cost function is transformed into a probability cost function, by experimentally finding the relation between the MLE and the probability of the "true" wind. We use the determined probability function to predict how often a certain solution rank corresponds to the "true" solution, using ECMWF winds as reference. The correspondence is remarkable, indicating that the solution probability function we found is adequate.

> Then, to optimise wind retrieval, the spatial resolution of the retrieved winds is investigated. The QuikSCAT 25-km inverted winds are compared to the 100-km winds. It turns out that the probability function derived for 25-km is also valid for 100-km resolution. The 100-km product, which is less noisy by definition, shows both less ambiguity and more accuracy than the 25-km product and, as such, the former is recommended for QuikSCAT use in NWP data assimilation.

> The results of the extensive study on the QuikSCAT inversion problem show that in order to improve the wind retrieval, notably in the nadir region, more ambiguous wind solutions need to be provided to the AR. In order to be successful with a multiple solution concept, it is very important to characterize each of the ambiguous wind solutions with its corresponding probability of being the "true" wind. Therefore, a median filter AR, in which the probability of each solution is not explicitly used in the final selection, is inappropriate. We propose to use the multiple solution inversion output in combination with a variational analysis AR (i.e., 2D-Var), the so-called MSS. The variational analysis AR is not only capable of explicitly using probability for the multiple solutions but also ensures spatial consistency and meteorological balance of the retrieved winds.

The comparison is performed, using NCEP winds as background term for 2D-Var and ECMWF winds as validation reference. The MSS turns out to be more in agreement with ECMWF than the standard procedure, especially at nadir. As expected, the MSS wind direction is substantially better in nadir, thus validating the procedure proposed. Moreover, the MSS selected solution is, in general, probabilistically consistent, whereas the closest-to-NCEP solution is rather inconsistent with the a priori set probabilities. In other words, the influence of the background in the retrieved field is relatively small. As such, 2D-Var is successfully exploiting the information content of the observations.

Since the ECMWF field used for validation is spatially smooth, it is at this point difficult to assess the effect of the background error structure functions in obtaining a smooth analysis. In particular for applications other than NWP, it may still be worthwhile to evaluate the effect of the spatial filtering by validating different versions of MSS with in-situ data.

The meteorological cases examined clearly show more spatially consistent and realistic wind fields for the MSS than for the standard procedure, especially at nadir. Moreover, the MSS is not only acting as a spatial filter, but is also keeping the wind information (e.g., lows, fronts, etc.) present in the observations. As such, the multiple solution scheme seems to be more appropriate for QuikSCAT data assimilation purposes than the standard scheme.

The MLE-based QC procedure at 25-km resolution (see *Portabella and Stoffelen, 2001*; *Portabella and Stoffelen, 2002b*) is not always satisfactory at 100-km resolution. Alternatives for such QC are discussed. Similar to the QC at 25-km, a threshold of a 100-km-resolution Rn could be set to QC 100-km winds. WVCs with low-probability selected solutions, which are shown to be of poor quality, could also be rejected with such Rn threshold. A variational QC is also pointed out as an effective way of removing large inconsistencies with the analysis field.

In this study, the wind retrieval over the QuikSCAT outer regions is not examined. In such regions, the azimuth separation (diversity) monotonically decreases as we approach the edges of the outer swath. As discussed in section 1.3, the MSS allows a variable number of solutions from inversion, according to the level of determination or azimuth diversity, to be used for AR purposes. It seems reasonable to apply the same methodology (MSS) to the QuikSCAT outer regions. In such regions, there is a substantial ambiguity problem since only two views are available. However, as discussed in section 1.1, the accuracy of a two-view system is comparable to a three-view (or more) system provided that we use an effective AR procedure. Moreover, the variational analysis AR used by the MSS should work significantly better for QuikSCAT than for two-view systems such as the SASS on Seasat, since, in the case of QuikSCAT, the large (i.e., 1400-km wide) and almost unique wind information (i.e., low ambiguity) of the inner swath will be extrapolated to the few nodes of the outer regions, while for SASS, the ambiguity problem is the same over the entire swath.

Portabella and Stoffelen (2002c) show that the characteristics of the MLE change with the dimension of the measurement space. That is, the MLE distributions of two-view (e.g. QuikSCAT outer swath) and four-view (e.g. QuikSCAT inner swath) measurement systems differ. This means that the MSS is applicable to the QuikSCAT outer regions, provided that the solution probability is re-computed using the outer-swath MLE information and the observation term of the AR is tuned to the outer regions. It is also important to say that a comprehensive QC is needed to successfully derive winds in the outer regions. *Portabella* (2002) shows that this is not trivial and therefore further investigation is needed to achieve an effective QC in the outer region prior to operationally assimilate the QuikSCAT outer-region winds into NWP.

Appendix A: MLE norm at 100-km resolution

The MLE, as defined in equation 2 (see section 1.1), can be computed with different norms; for example, a measurement error variance (Kp) proportional to the GMF simulated backscatter ($\sigma_s°$) or a Kp proportional to the backscatter measurement ($\sigma_m°$). Using a Kp proportional to $\sigma_s°$ may cause systematic effects in the wind direction solutions (see *Stoffelen and Anderson, 1997c*). This may be less true for a Kp proportional to $\sigma_m°$ since the MLE norm remains fixed during the inversion process. However, the occurrence of directional biases after inversion depends on the measurement configuration (*Stoffelen and Anderson, 1997c*) and there is no easy way to determine what is the best choice; for such purpose, tests are usually conducted.

A Kp proportional to $\sigma_s°$ is used to compute the MLE at 25-km resolution to mimic the JPL processing (see equation 2). In this study, however, we compare the standard wind retrieval procedure with the MSS at 100-km resolution. Since the MLE norm has not yet been tested at low resolution, it is worthwhile to check which Kp, i.e., proportional to $\sigma_s°$ or proportional to $\sigma_m°$, is best in terms of wind retrieval quality. A set of three days of collocated ECMWF winds is used here for reference.

Figure A.1 shows the wind direction distributions with respect to the satellite flight direction of ECMWF winds (solid lines) and QuikSCAT retrieved solutions closest to ECMWF using Kp($\sigma_s°$) (dotted lines) and Kp($\sigma_m°$) (dashed lines). The left and right plots show the wind direction distributions of the sweet and nadir swath, respectively. It is clear from the plots that the QuikSCAT retrieved distributions present some unrealistic accumulations (see peaks and troughs of both dotted and dashed lines) as compared to ECMWF. Both the Kp($\sigma_s°$) and Kp($\sigma_m°$) distributions are however very similar, showing that none of them is able to avoid these unrealistic wind direction accumulations.

This result is in line with the RMS difference values between the QuikSCAT (closest to ECMWF) and ECMWF wind directions. The RMS difference in wind direction is similar for the Kp($\sigma_s°$) and Kp($\sigma_m°$) distributions, although slightly lower for the former (see table A.1). On the other hand, the RMS difference in speed is slightly lower for Kp($\sigma_m°$) than for Kp($\sigma_s°$), leading to an overall comparable accuracy.

TABLE A.1

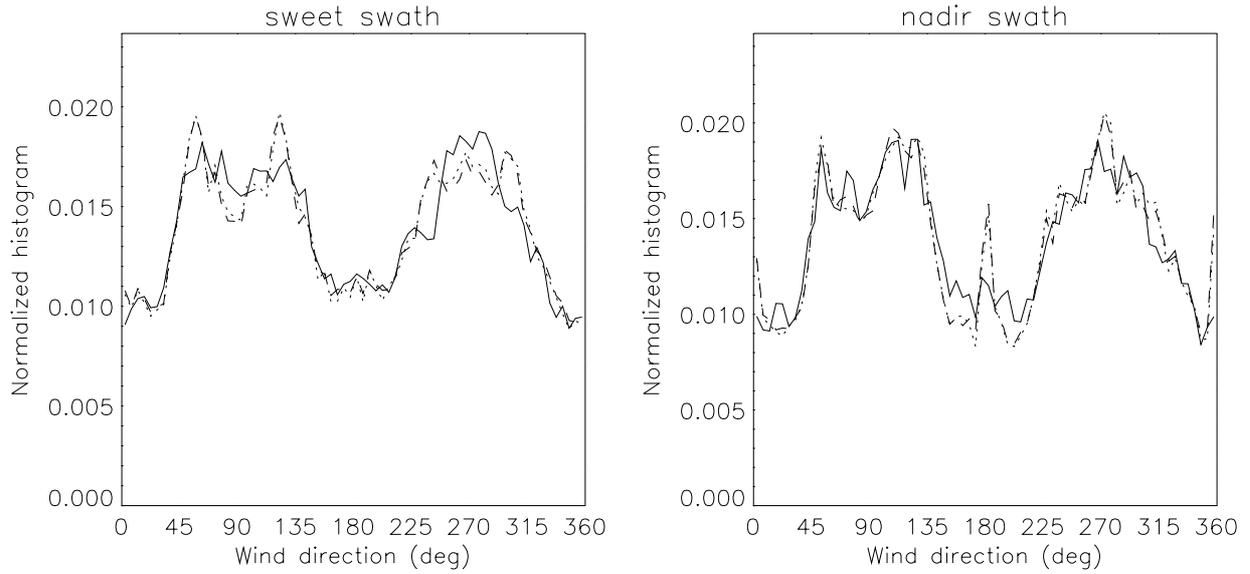|	RMS in Speed (m/s) Kp($\sigma_s°$) / Kp($\sigma_m°$)	RMS in Direction(°) Kp($\sigma_s°$) / Kp($\sigma_m°$)	NRMS Kp($\sigma_s°$) / Kp($\sigma_m°$)
Sweet swath	1.43 / 1.39	18.88 / 19.08	0.3612 / 0.3478
Nadir swath	1.57 / 1.56	22.04 / 22.39	0.4209 / 0.3278

Figure A.1 Wind direction (with respect to the satellite flight direction) histograms of ECMWF winds (solid lines) and QuikSCAT-retrieved solutions closest to ECMWF using $Kp(\sigma_s°)$ (dotted lines) and $Kp(\sigma_m°)$ (dashed lines) for the sweet swath (a) and the nadir swath (b).

Tables A.2 to A.5 show the percentage of "selected" solutions (closest to ECMWF) stratified by number of solutions and rank (same stratification as tables in section 1.2). Tables A.2 and A.3 correspond to the $Kp(\sigma_s°)$ selected solution distributions of the sweet and nadir swath, respectively. Tables A.4 and A.5 correspond to the $Kp(\sigma_m°)$ selected solution distributions of the sweet and nadir swath, respectively.

As discussed in section 1.1.2, the 1st rank skill shows the ambiguity or uncertainty of the inversion. In these tables, the 1st rank skill is shown by the percentage of selections of rank 1 solution. As we see in the tables, the overall 1st rank skill (last column of the tables) is the same for both $Kp(\sigma_s°)$ and $Kp(\sigma_m°)$ in the entire inner (sweet + nadir) swath . However, the number of solutions given by the $Kp(\sigma_m°)$ is significantly smaller than the number given by $Kp(\sigma_s°)$ in both the sweet and nadir swath (see the relatively smaller accumulation of data for 2, 3 and 4 solutions of $Kp(\sigma_m°)$ tables compared to $Kp(\sigma_s°)$ tables). As such, the $Kp(\sigma_m°)$ produces a less ambiguous wind product than the $Kp(\sigma_s°)$.

Stoffelen et al. (2000) computed a more realistic RMS difference in wind direction, called the normalized RMS (NRMS). In using the usual wind direction difference RMS definition, the more ambiguous solutions are provided by the inversion, the smaller the RMS will be, because the chance that one of the solutions will be close to the wind reference will increase. In the limit of an infinite amount of observations, the RMS will even be zero, while the information content of the set of solutions in reality decreases with an increasing number of solutions, because there is no a priori way to say which of the solutions is the correct one. In order to solve this problem, they normalize the RMS with an expected value, which is dependent on the angle separation of the neighboring solutions of the closest solution to reference (ECMWF in this case). For more details, see *Stoffelen et al.* (2000). If we compute the NRMS, we get substantially lower values for the less ambiguous $Kp(\sigma_m°)$ product than for $Kp(\sigma_s°)$ (see table A.1).

Table A.2 Solution distribution for $Kp(\sigma_s^\circ)$ (sweet swath).

	1 Solution	2 Solutions	3 Solutions	4 Solutions	All Solutions
Number of Data	39041	41167	16413	5847	102468
Rank 1	100	89	82	73	91
Rank 2	-	11	12	17	8
Rank 3	-	-	5	6	1
Rank 4	-	-	-	5	0

Table A.3 Solution distribution for $Kp(\sigma_s^\circ)$ (nadir swath).

	1 Solution	2 Solutions	3 Solutions	4 Solutions	All Solutions
Number of Data	24269	22032	3665	738	50704
Rank 1	100	71	75	50	85
Rank 2	-	29	15	18	14
Rank 3	-	-	9	15	1
Rank 4	-	-	-	16	0

Table A.4 Solution distribution for $Kp(\sigma_m^\circ)$ (sweet swath).

	1 Solution	2 Solutions	3 Solutions	4 Solutions	All Solutions
Number of Data	43650	38950	14957	4915	102472
Rank 1	100	88	82	69	91
Rank 2	-	12	13	18	8
Rank 3	-	-	5	7	1
Rank 4	-	-	-	6	0

Table A.5 Solution distribution for $Kp(\sigma_m^\circ)$ (nadir swath).

	1 Solution	2 Solutions	3 Solutions	4 Solutions	All Solutions
Number of Data	26348	20946	2545	865	50704
Rank 1	100	71	67	46	85
Rank 2	-	29	21	24	14
Rank 3	-	-	12	15	1
Rank 4	-	-	-	14	0

In summary, the $Kp(\sigma_m°)$ provides a less ambiguous product than $Kp(\sigma_s°)$ without decreasing the quality of the wind retrieval. In other words, in comparison with $Kp(\sigma_s°)$, $Kp(\sigma_m°)$ is capable of removing a significant amount of unrealistic ambiguous wind solutions. Consequently, the $Kp(\sigma_m°)$ norm will be used for deriving QuikSCAT winds at 100-km resolution.

Finally, it is worthwhile to mention that a fixed (constant) norm has been successfully used to invert ERS winds (*Stoffelen and Anderson, 1997c*). The use of such MLE norm has not yet been tested for SeaWinds but is recommended to be tested.

Appendix B: Meteorological cases

Following the discussion on several meteorological cases presented in section 2.2, we include some additional cases in this appendix.

Figure B.1 shows a high-pressure system in the middle of the plot. Note in Figure B.1a that the standard procedure is successfully retrieving the wind field, showing spatially consistent and realistic winds. Figure B.1b is very similar to Figure B.1a, indicating that whenever the standard procedure is successful, the MSS wind field does not change much. Moreover, the wind front line present at the top left part of Figure B.1a is also visible in Figure B.1b, indicating that the MSS successfully keeps the dynamical information of the observations without significantly oversmoothing the retrieved field.
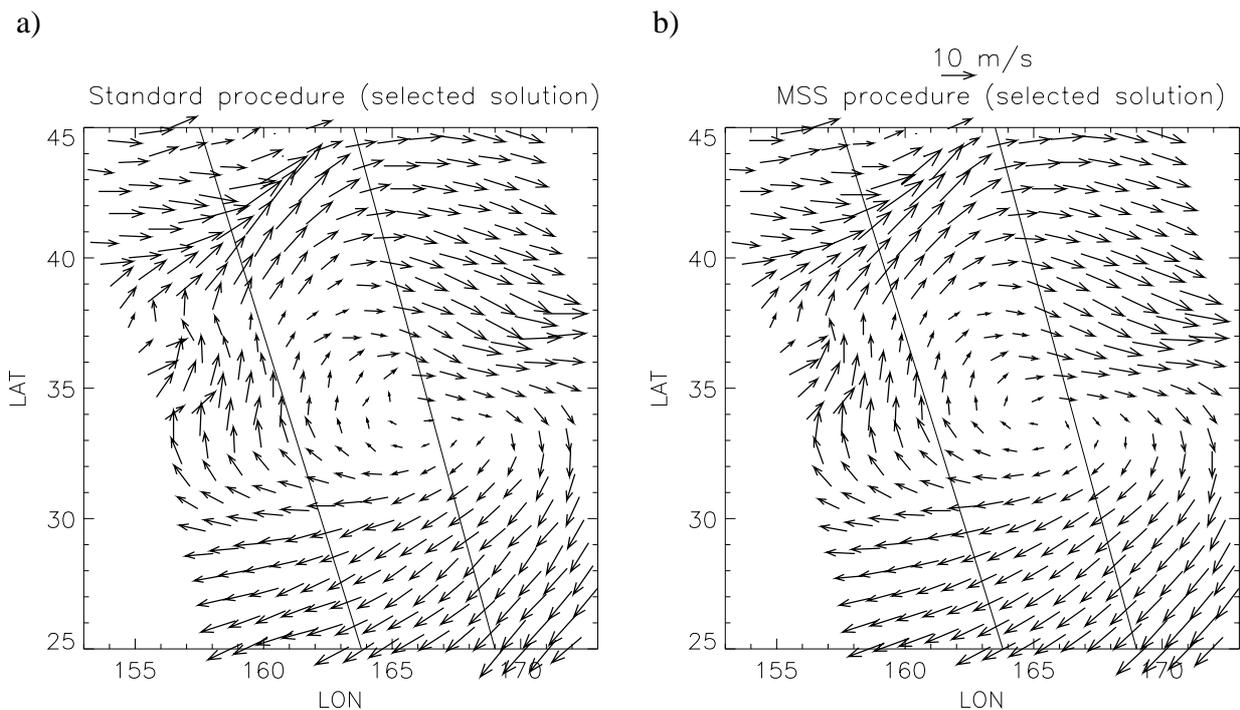
Figure B.1 QuikSCAT retrieved wind field using the standard procedure (a) and the MSS (b). The acquisition date is February 2 2002 at 19 hours UTC. The solid lines separate the sweet-left (left side), the nadir (middle), and the sweet-right (right side) regions of the QuikSCAT swath.
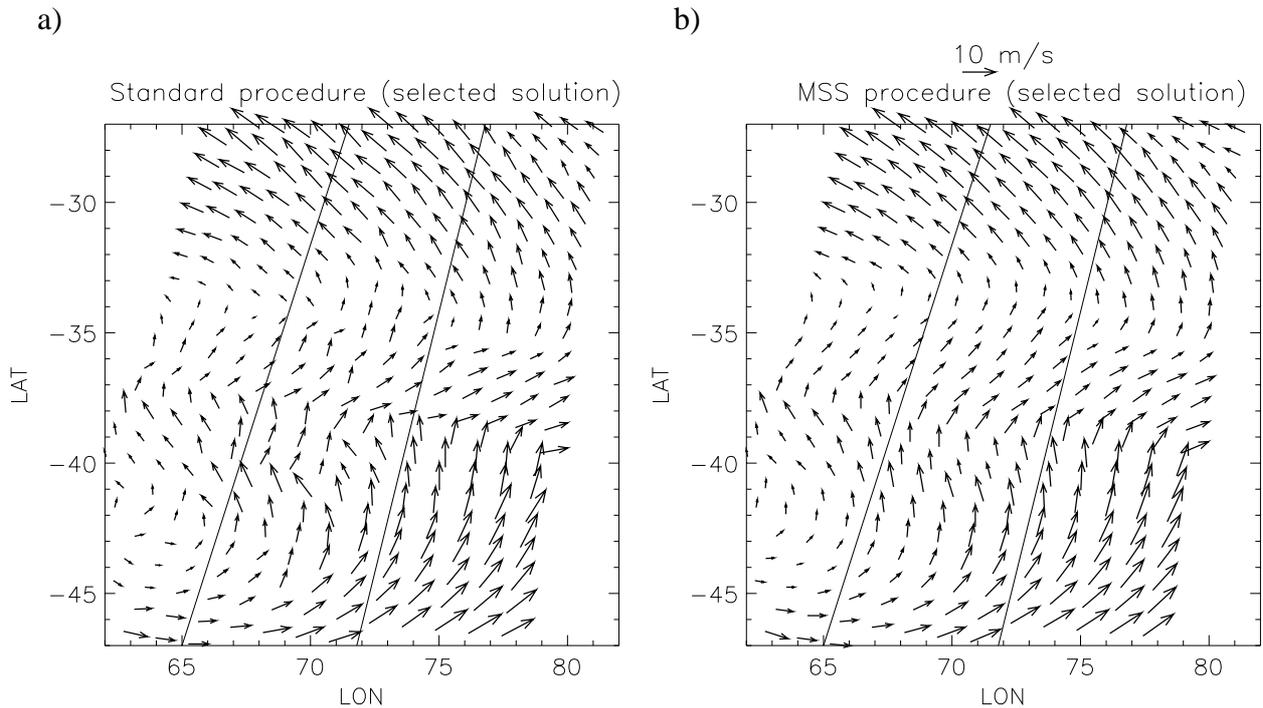
a)   b)

Standard procedure (selected solution)   MSS procedure (selected solution)

Figure B.2 QuikSCAT retrieved wind field using the standard procedure (a) and the MSS (b). The acquisition date is February 2 2002 at 12 hours UTC. The solid lines separate the sweet-right (left side), the nadir (middle), and the sweet-left (right side) regions of the QuikSCAT swath.

Figure B.2 shows another situation where the standard procedure shows some noisy winds in the nadir swath (see Figure B.2a). As expected, the MSS (Figure B.2b) is successfully filtering the remaining noise (without oversmoothing), improving the quality of the retrieved field compared with the standard procedure (Figure B.2a).

Figure B.3 presents a high wind speed case. Both the standard procedure and the MSS present several spatially inconsistent wind arrows in the vicinity of the low-pressure system (see bottom of Figures B.3a and B.3b). However, it is clear that the low is better resolved by the MSS (Figure B.3b) than by the standard procedure (Figure B.3a). Note also that this is an interesting case for data assimilation into NWP since ECMWF (Figure B.3c) does not accurately predict the intensity and position of the low (see for example, the difference of wind speed between the top plots and the bottom plot).
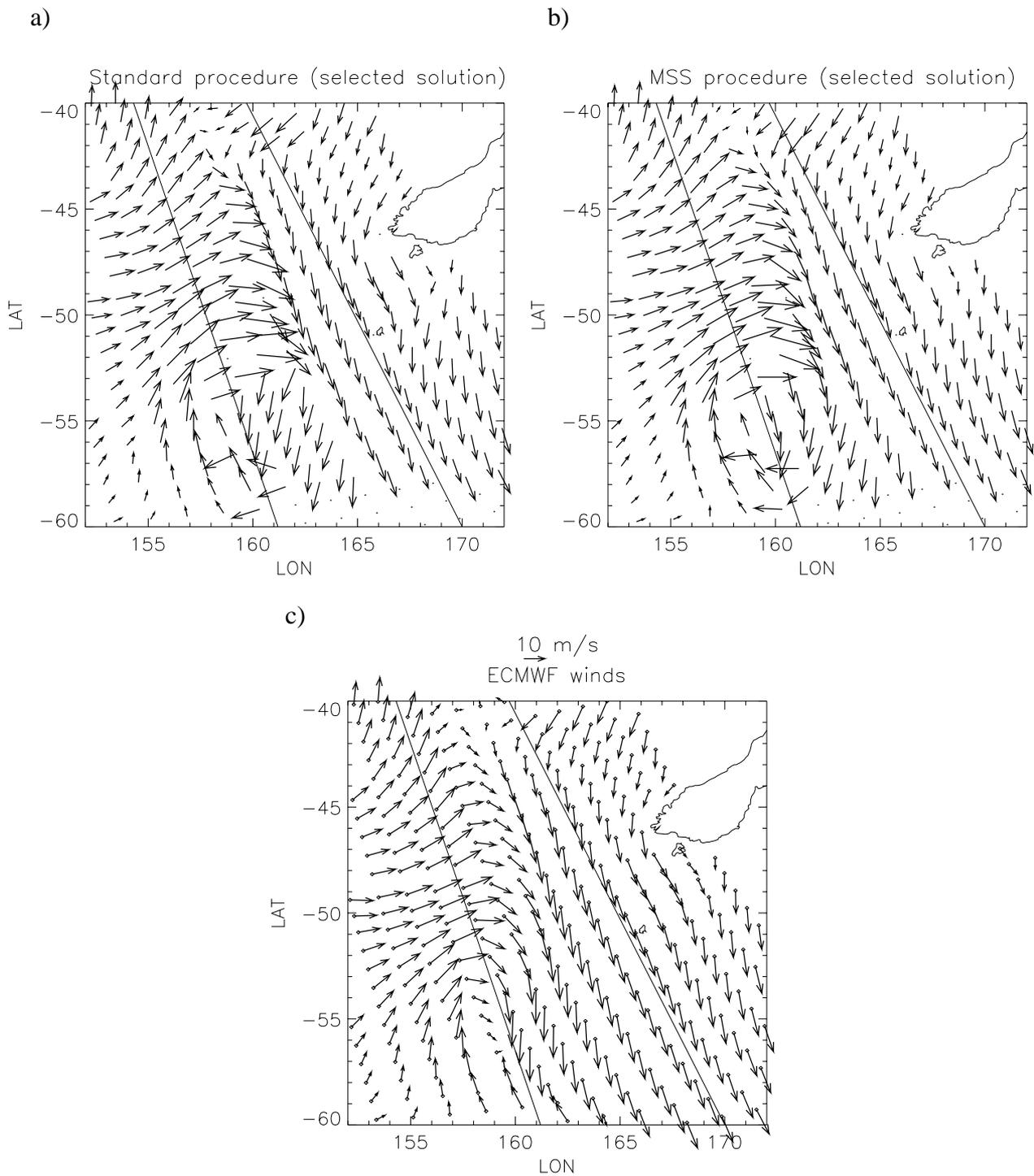
Figure B.3 QuikSCAT retrieved wind field using the standard procedure (a) and the MSS (b), and ECMWF wind field (c). The acquisition date is February 2 2002 at 20 hours UTC. The solid lines separate the sweet-left (left side), the nadir (middle), and the sweet-right (right side) regions of the QuikSCAT swath. The dots represent QC WVCs.
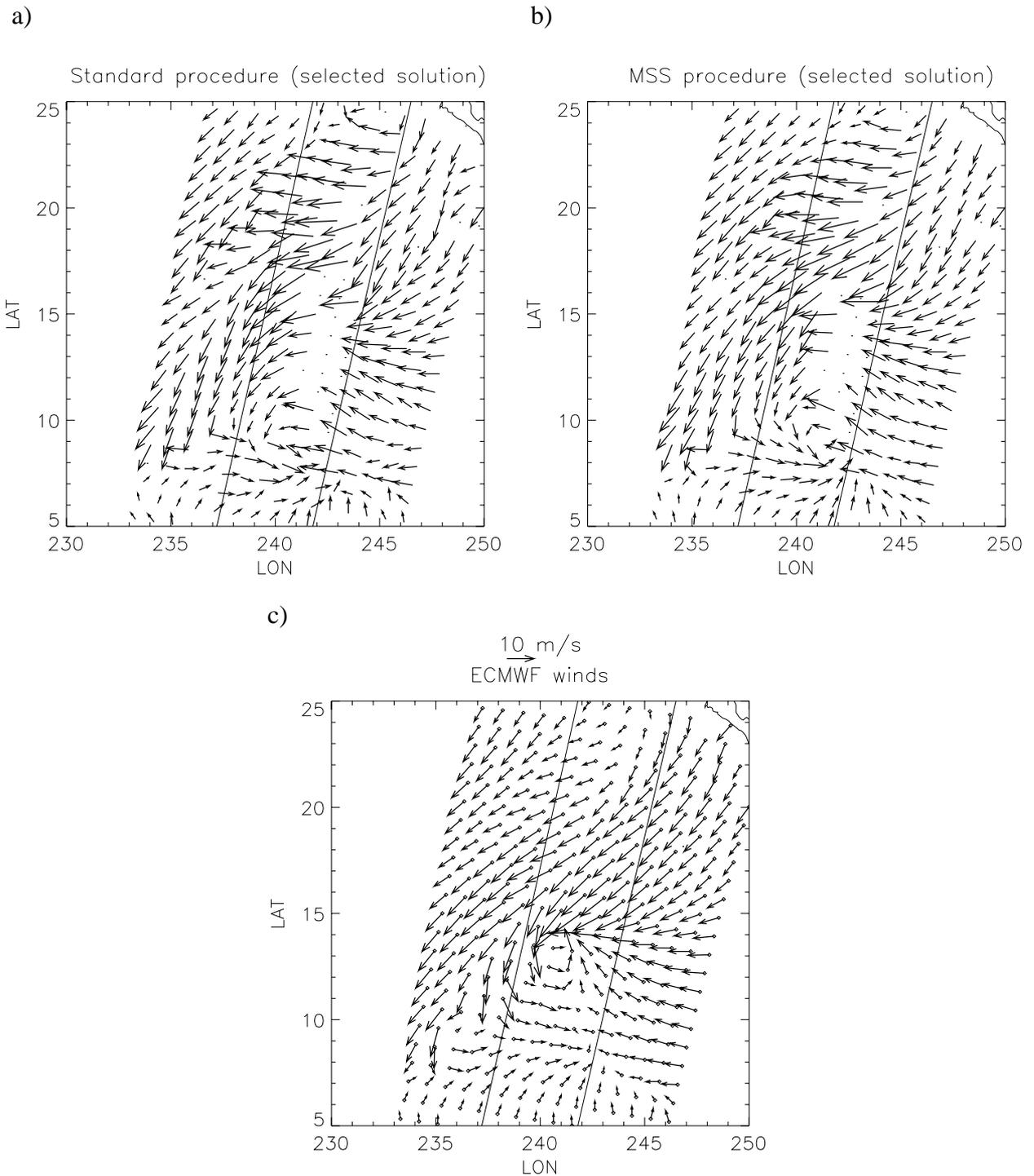
Figure B.4 QuikSCAT retrieved wind field using the standard procedure (a) and the MSS (b), and ECMWF wind field (c). The acquisition date is February 3 2002 at 2 hours UTC. The solid lines separate the sweet-right (left side), the nadir (middle), and the sweet-left (right side) regions of the QuikSCAT swath. The dots represent QC WVCs.

Figure B.4 shows a dynamically active case in the Tropics. Note that the low wind speed area (bottom part of the plots) is better resolved by the MSS (Figure B.4b) than by the standard procedure (Figure B.4a). The wind flow in that area is not only more spatially consistent in the former but also more realistic. Note again, the wind flow difference between ECMWF (Figure B.4c) and QuikSCAT (Figures B.4a and B.4b) in the high wind speed region (center-top part of the plots); both the speed and direction are substantially different, showing once more the potential positive impact of assimilating QuikSCAT data in mesoscale NWP models.

References

Huddleston, J.N., and Stiles, B.W., "Multidimensional histogram (MUDH) rain flag", version 2.1, *Jet Propulsion Laboratory*, available at http://podaac-www.jpl.nasa.gov/quikscat/, September 2000.

JPL, "QuikSCAT science data product user's manual", version 2.2, *Jet Propulsion Laboratory D-12985*, pp. 89, December 2001.

Leidner, M., Hoffman, R., and Augenbaum, J., "SeaWinds scatterometer real-time BUFR geophysical data product", version 2.2.0, *NOAA/NESDIS*, February 2000.

Pierson, W.J., "Probabilities and statistics for backscatter estimates obtained by a scatterometer," *J. Geophys. Res.*, vol. 94, no. C7, pp. 9743-9759, 1989.

Portabella, M., and Stoffelen, A., "Rain detection and quality control of SeaWinds," *J. Atm. and Ocean Techn.*, vol. 18, no. 7, pp. 1171-1183, 2001.

Portabella, M., Stoffelen, A., and De Vries, J., "Development of a SeaWinds wind product for weather forecasting," *Proc. of International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. III, pp. 1076-1078, 2001.

Portabella, M., "Wind field retrieval from satellite radar systems," *PhD thesis at the University of Barcelona,* ISBN 90-6464-499-3, September 2002.

Portabella, M., and Stoffelen, A., "Quality control and wind retrieval for SeaWinds," *Scientific report WR-2002-01*, Koninklijk Nederlands Meteorologisch Instituut, The Netherlands, 2002a.

Portabella, M., and Stoffelen, A., "A comparison of KNMI quality control and JPL rain flag for SeaWinds," *Can. Jour. of Rem. Sens.*, vol. 28, no. 3, pp. 424-430, 2002b.

Portabella, M., and Stoffelen, A., "Characterization of residual information for SeaWinds quality control," *IEEE Trans. Geosci. Rem. Sens.*, in press, 2002c.

Rodgers, C. D., "Inverse methods for atmospheric sounding: theory and practice," *World Scientific Publishing Co.*, 2000.

Spencer, M.W., Wu, C., and Long, D.G., "Tradeoffs in the design of a spaceborn scanning pencil beam scatterometer: application to SeaWinds," *IEEE Trans. Geosci. Rem. Sens.*, vol. 35, no. 1, pp. 115-126, 1997.

Stiles, B.W., Pollard, B.D., Dunbar, R.S., "Direction interval retrieval with thresholded nudging," *IEEE Trans. Geosci. Rem. Sens.*, vol. 40, no. 1, pp. 79-89, 2002.

Stoffelen, A., and Anderson, D., "Ambiguity removal and assimilation of scatterometer data," *Quart. J. R. Met. Soc.,* vol. 123, pp. 491-518, 1997a.

Stoffelen, A., and Anderson D., "Scatterometer data interpretation: measurement space and inversion," *J. Atm. and Ocean. Techn.*, vol. 14(6), pp. 1298-1313, 1997b.

Stoffelen, A., and Anderson, D., "Scatterometer data interpretation: derivation of the transfer function CMOD-4," *J. Geophys. Res.*, vol. 102, no. C3, pp. 5767-5780, 1997c.

Stoffelen, A., Van Beukering, P., "Implementation of improved ERS scatterometer data processing and its impact on HIRLAM short range weather forecasts," *Report NRSP-2/97-06,* Beleidscomissie Remote Sensing, The Netherlands, 1997.

Stoffelen, A., "Scatterometry," *PhD thesis at the University of Utrecht,* ISBN 90-393-1708-9, October 1998.

Stoffelen, A., de Vries, J., and Voorrips, A., "Towards the real-time use of QuikSCAT winds," *Final Report USP-2/00-26*, Beleidscomissie Remote Sensing, The Netherlands, September 2000.

Undén, P., Kelly, G., Le Meur, D., and Isaksen, L., "Observing system experiments with the 3D-Var assimilation system," *Technical Memorandum No. 244*, European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, United Kingdom, 1997.

Acknowledgements