

# A probabilistic approach for SeaWinds data assimilation<sup>1</sup>

By M. PORTABELLA and A. STOFFELEN

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

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

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

## SUMMARY

Scatterometer sea-surface wind observations are being successfully assimilated into Numerical Weather Prediction (NWP) models. 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 in the nadir region results in poor quality winds. In the QuikSCAT nadir region, where the local cost function minima are broad, the use of the standard procedure results in arbitrary and inaccurate winds. A new scheme, which accounts for broad cost function minima by allowing more ambiguous wind solutions, 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 ambiguity removal procedure to make the scheme flexible enough to accept many wind solutions.

A comparison between the standard wind retrieval and the MSS procedures at 100-km resolution is then performed, using independent model winds for validation. The MSS turns out to be more in agreement with the model reference 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.

**KEYWORDS:** Scatterometer, wind inversion, variational analysis, quality control.

---

<sup>1</sup> Quart. J. R. Met. Soc., vol. 130, no. 596, pp. 127-152, 2004. © Royal Meteorological Society

## 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 doubling 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  $46^\circ$  and  $54^\circ$  respectively, that sweep in a circular pattern. The antenna radiates microwave pulses at a frequency of 13.4 GHz (Ku-Band) across a 1800-km-wide swath centered on the spacecraft's nadir subtrack, making approximately 1.1 million 25-km resolution ocean surface wind vector measurements and covering 90% of the Earth's surface every day.

The SeaWinds swath is divided into 76 equidistant across-track wind vector cells (WVCs) or nodes numbered from left to right when looking along the satellite's propagation direction. The nominal WVC size is 25 km x 25 km, and all backscatter measurements centered in a WVC are used to derive the WVC wind solutions. Due to the conical scanning, a WVC is generally viewed when looking forward (fore) and a second time when looking aft. As such, up to four measurement classes (called "beam" here) emerge: H-pol fore, H-pol aft, V-pol fore, and V-pol aft, in each WVC. Due to the smaller swath (1400 km) viewed in H-pol at  $46^\circ$  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 pattern of azimuth angles in a single SeaWinds WVC 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  $180^\circ$  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.

In section 2 of this paper we present the wind retrieval problem of scatterometers, in particular of QuikSCAT. The standard wind retrieval and the Multiple Solution Scheme (MSS) procedures are presented in sections 3 and 4, respectively. In section 5, a comparison between the two

procedures is performed at 100-km resolution. Then, the need for an effective Quality Control (QC) procedure (at low resolution) is discussed in section 6. Finally, the conclusions and recommendations are presented in section 7.

## 2. COST FUNCTION

In remote sensing, the process of deriving the best estimate of one or more geophysical state variables for a given set of observations, allowing for observation errors, is called inversion. Among the several approaches for inverting remotely sensed variables (see Rodgers 2000), the most general one 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)]:

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

where  $N$  is the number of measurements,  $\sigma_{mi}^o$  is the backscatter measurement,  $\sigma_{si}^o$  is the backscatter simulated through the Geophysical Model Function (GMF) for different wind speed and direction trial values, and  $K_p(\sigma_{si}^o)$  is the measurement error variance (noise). Strictly speaking, when assuming Gaussian errors, a term  $\ln(K_p(\sigma_{si}^o))$  should be added to the right-hand side of (1) but this term is not significant and, as such, is not used. [Note: the  $K_p$  is usually taken as function of either  $\sigma_{mi}^o$  or  $\sigma_{si}^o$ , 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 Portabella and Stoffelen 2002b) seem to indicate that, for SeaWinds, a  $K_p$  dependent on  $\sigma_{mi}^o$  is slightly better than a  $K_p$  dependent on  $\sigma_{si}^o$  at 100-km resolution; as such, the former is used in section 5.]

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 (1), 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 since a single well-determined minimum is usually found; the same operation is repeated for every wind direction, at a step size of typically  $2.5^\circ$ . 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.

### (a) *Wind retrieval skill*

The MLE (1) can be interpreted as a measure of the distance between a set of  $\sigma_{mi}^\circ$  values and the solution  $\sigma_{si}^\circ$  set lying on the GMF surface in a transformed measurement space where each axis of the measurement space is scaled by  $Kp(\sigma_{si}^\circ)$  (Stoffelen and Anderson 1997b). The shape of the MLE cost function is determined by the  $\sigma^\circ$  modulation of any view and the relative geometry among views. By only using the MLE cost function local minima in the retrieval (as in the standard procedure), the skill of the wind retrieval is compromised.

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 precision 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 Fig. 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<sup>1</sup>.

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 4, if we properly use the information on accuracy and ambiguity derived from the MLE cost function (inversion), the wind retrieval may improve significantly.

---

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

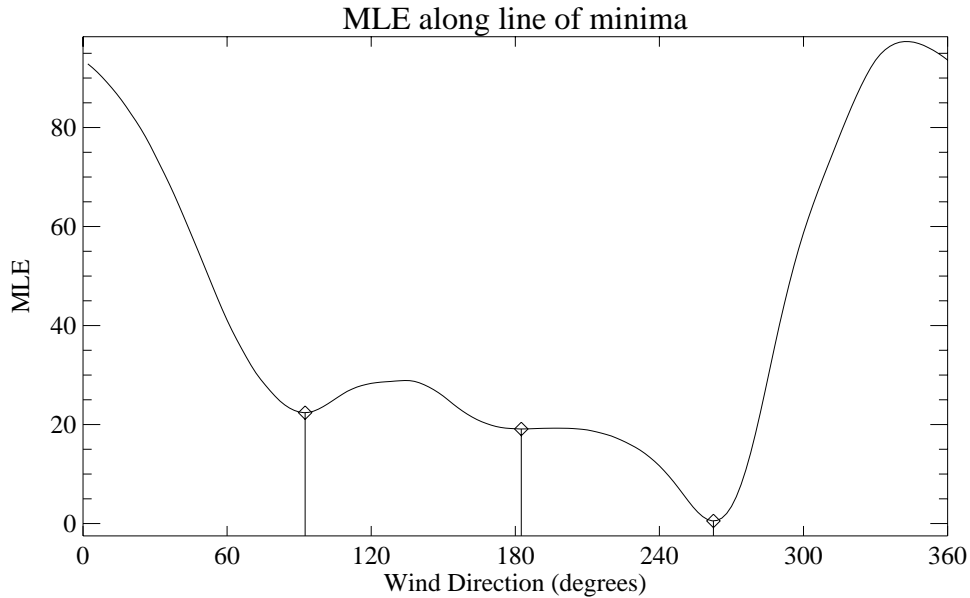


Figure 1 Example of MLE cost function for QuikSCAT node number 33. The diamond symbols indicate the locations of the minima found by the inversion procedure.

### (b) 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 Fig. 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 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. A better illustration of the QuikSCAT inversion problem can be found in Portabella (2002), where QuikSCAT retrieved winds are compared to ECMWF forecast winds over a period of 12 hours.

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.

### 3. 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.

#### (a) *Inversion*

The MLE-based inversion has already been discussed in section 2. 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 2, the probability of being the “true” wind given a set of scatterometer observations is related by definition to the MLE in the following way:

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

where  $v$  represents the “true” wind and  $\sigma^o$  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; Portabella and Stoffelen 2002c) 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 (2) is re-written as:

$$p(v | \sigma^o) = \frac{1}{k'} e^{-Rn/l} \quad (3)$$

where  $k'$  is again a normalization factor, and  $l$  is the parameter that we want to empirically derive. The Rn computation is described in the Appendix.

- In order to empirically derive (3), 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  $\mathbf{v}_i$  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

$$P(s = j | Rn_i, i \in \{1, N\}) = \frac{p_s(Rn_j)}{\sum_{i=1}^N p_s(Rn_i)} \quad (4)$$

- 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 1<sup>st</sup> rank (or the 2<sup>nd</sup> rank), as a function of  $Rn_1$  and  $Rn_2$ . But according to our assumption, by applying (4) with  $N=2$ , we find that the probability of selecting the 1<sup>st</sup> rank is given by

$$P(s = 1 | Rn_1, Rn_2) = \frac{p_s(Rn_1)}{p_s(Rn_1) + p_s(Rn_2)} = \{1 + p_s(Rn_2) / p_s(Rn_1)\}^{-1} \quad (5)$$

- Therefore, by re-arranging (5), 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 2a 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 (3), the shape of  $p_s(x)$  is exponential and therefore we just have to fit the exponential to the experimental function of Fig. 2a by adjusting the  $l$  parameter. Figure 2b shows the best fit to Fig. 2a, which is represented by the following function:

$$p_s(x) = e^{-x/1.4} \quad (6)$$

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 Fig. 2a).

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 conclude that the assumption is correct and that (6) can be used to determine the solution probability.

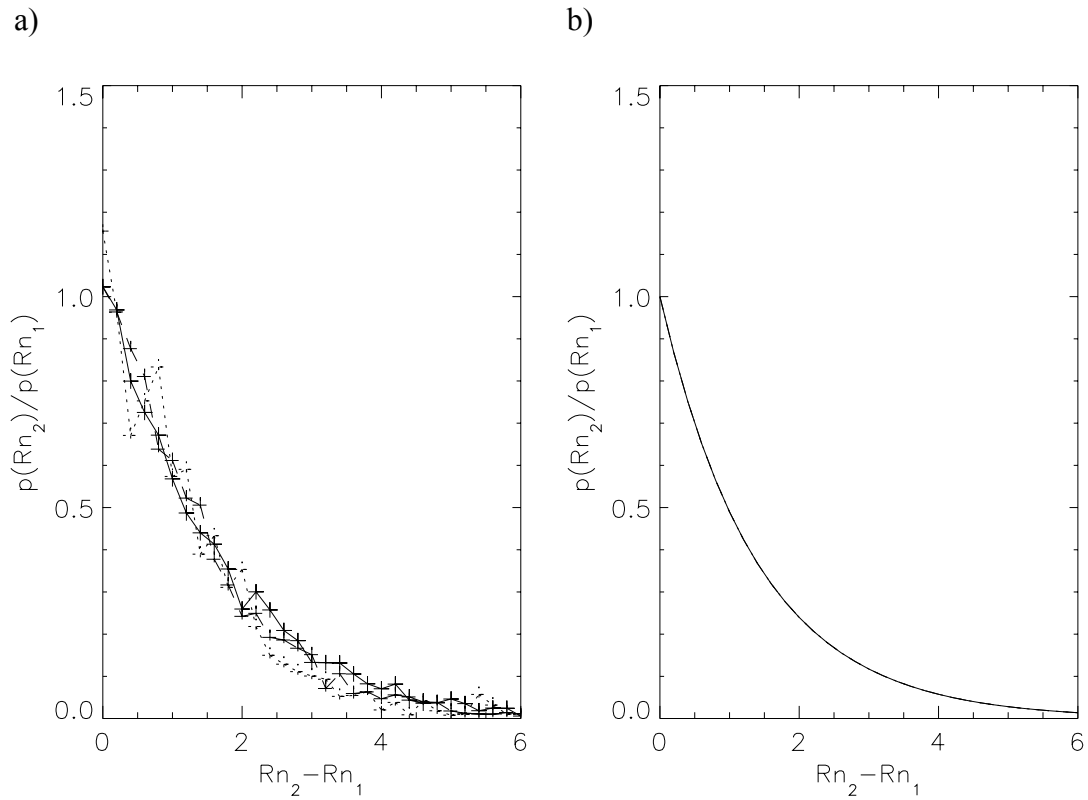


Figure 2 Plot a shows the ratio of the number of realizations of  $Rn_2$  and the number of realizations of  $Rn_1$  as a function of  $Rn_2 - Rn_1$ , and for values of  $Rn_1=0.1$  (solid),  $Rn_1=1.1$  (dashed), and  $Rn_1=2.5$  (dotted). Plot b shows the single exponential fit to the curves of plot a.

**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

(b) *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 filter used by JPL for QuikSCAT AR (JPL 2001) 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 1<sup>st</sup> rank or the 2<sup>nd</sup> 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 4).
- The wind vectors in a 7 x 7 filter window determine a median vector for the center WVC (see definition of median of a group of data in JPL 2001). 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 when used in the initialization and filtering processes (this is further discussed in section 4). However, it is not 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

$$J(\delta x) = J_b + J_o^{scat} , \quad (7)$$

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 reference 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 is 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):

$$J_o^{scat} = \left( \frac{1}{\sum_{i=1}^N K_i^{-p}} \right)^{1/p} \quad (8)$$

where  $N$  is the number of solutions and  $K_i$  is:

$$K_i = \left( \frac{u - u_i}{\varepsilon_u} \right)^2 + \left( \frac{v - v_i}{\varepsilon_v} \right)^2 - 2 \ln P_i \quad (9)$$

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;  $\varepsilon_u$  and  $\varepsilon_v$  the corresponding observation errors; and  $P_i$  the solution probability. Stoffelen and Anderson (1997a) motivate the specification of these equations in wind components rather than in backscatter (as in (1)).

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.

(c) *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. [Note that the solution probability formula (6) is also applicable to 100-km resolution winds; more detailed description on the empirical derivation for 100-km can be found in Portabella and Stoffelen 2002a].

**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 ambiguity of the system is reflected in the quality of the 1<sup>st</sup> rank solution: the deeper the 1<sup>st</sup> rank (deepest MLE cost function minimum) in comparison with the secondary minima, the higher the likelihood of the corresponding rank-1 wind to be the “true” wind (higher 1<sup>st</sup> rank skill), i.e., the lower the ambiguity. By comparing Tables 1 and 2 to Tables 3 and 4, respectively, one can clearly see the substantially higher 1<sup>st</sup> rank skill of the 100-km product, denoting a smaller ambiguity problem, 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 (6). 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 3 shows the statistical results of looking at several characteristics of the cost function.

The top plots of Fig. 3 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 Fig. 3b are much broader and shifted towards higher probability difference values than the distributions of Fig. 3a, denoting a better probability modulation and therefore less accuracy (see section 2(a)) 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 bottom plots of Fig. 3 show the histograms of the difference between Pmax and the mean probability (Pmean) over an interval of  $\pm 12.5^\circ$  around the rank 1 direction for the 25-km (plot c) and the 100-km (plot d) 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 2(a) and 2(b)). 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. This is not only valid for the 100-km product but also for the 25-km product. As we see from the larger accumulation of data at low Pmax-Pmean values in Fig. 3c with respect to Fig. 3d, 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.

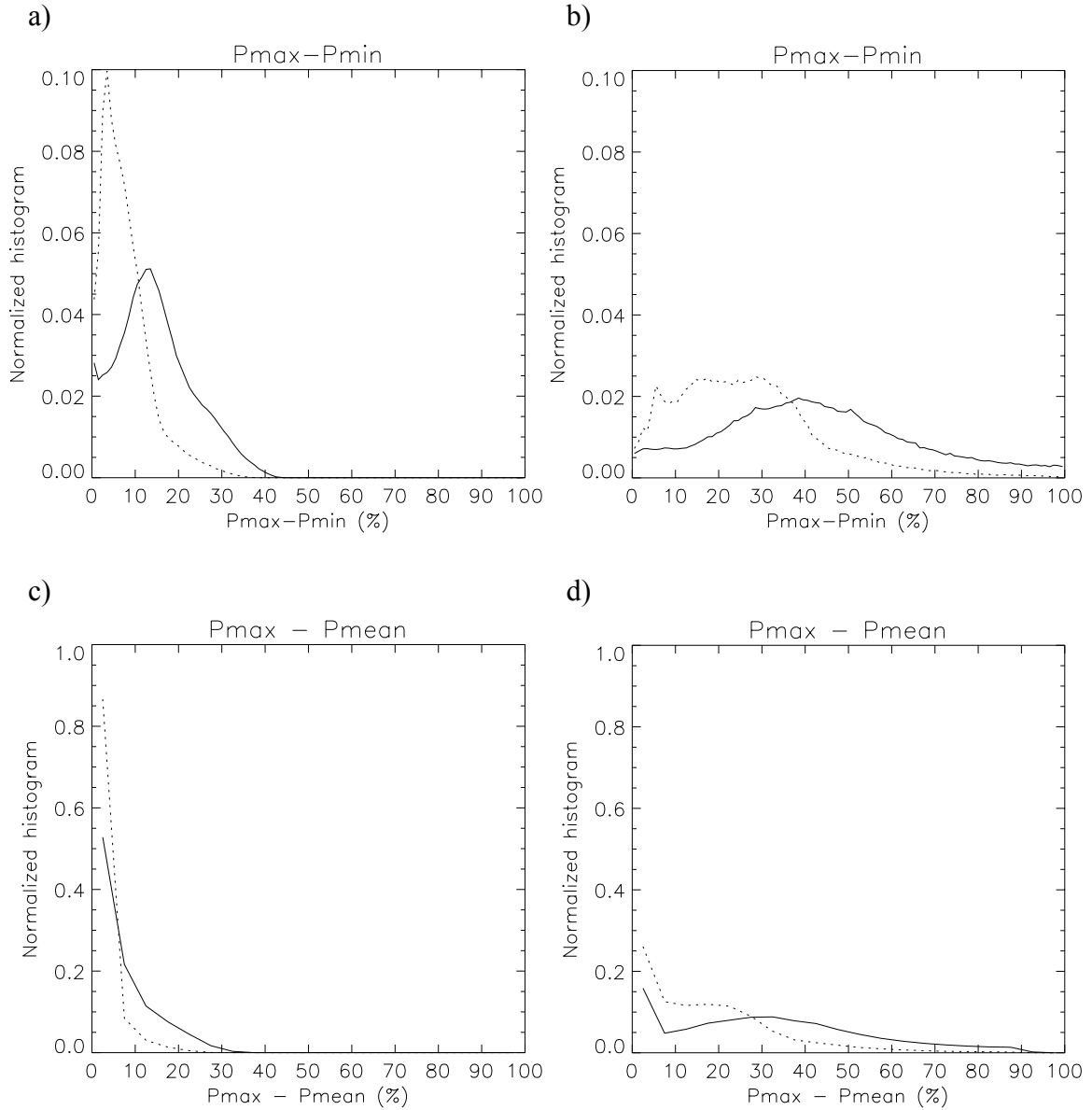


Figure 3 Histograms of the difference between the maximum ( $P_{max}$ ) and the minimum ( $P_{min}$ ) probabilities (top plots), and the difference between  $P_{max}$  and the mean probability ( $P_{mean}$ ) over an interval of  $\pm 12.5^\circ$  around  $P_{max}$  (bottom plots), for the sweet (solid lines) and the nadir (dotted lines) regions and for the 25-km (left plots) and the 100-km (right plots) products.

#### 4. 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 ignore 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.

#### (a) *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 3(b)) and works as follows. The probability<sup>1</sup> 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 3(b)).

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.

---

<sup>1</sup> Stiles *et al.* (2002) use the theoretical relation between MLE and probability (2) to compute the latter.

- 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  $P_{\max} - P_{\min}$  distributions in the nadir swath for 25-km resolution (Fig. 3a), 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.

#### (b) *Alternative*

The 2D-Var AR (see section 3(b)) 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 4 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 5a shows the standard ambiguous solution distribution (MLE cost function minima) for the same case of Fig. 4. 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.

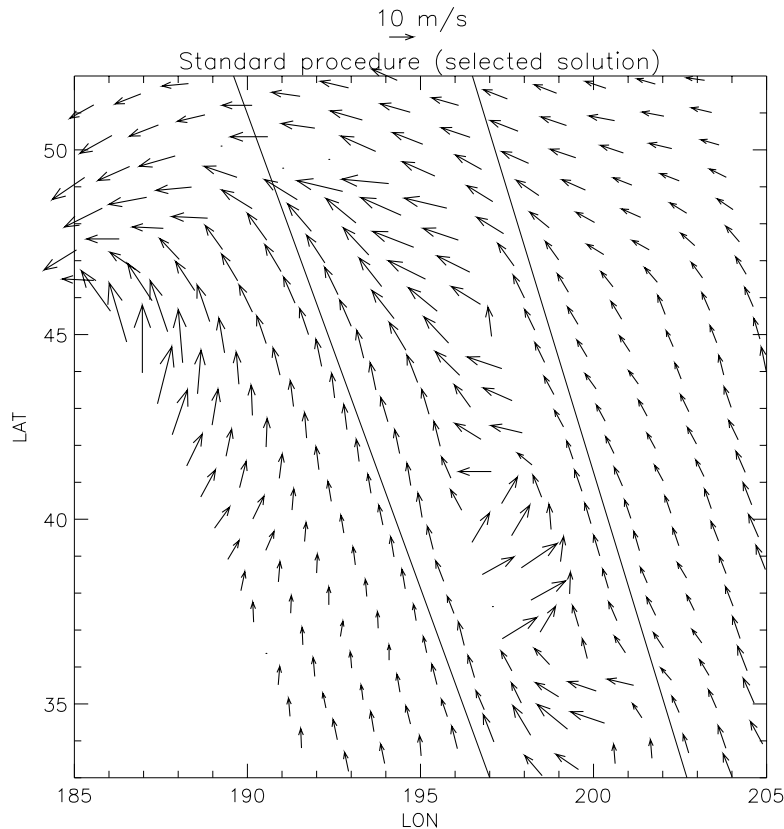


Figure 4 QuikSCAT retrieved wind field using the standard inversion output (cost function minima) and the 2D-Var AR. The acquisition date is January 15 2002 at 16 UTC. The solid lines separate the sweet-left (left side), the nadir (middle), and the sweet-right (right side) regions of the QuikSCAT swath.

Figure 5b shows the multiple ambiguous solution (not constrained to four) distribution again for the same meteorological case as Figs. 4 and 5a. We show all the cost function solutions with probability above a guessed threshold<sup>1</sup> of  $2 \times 10^{-7}$ . 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 Fig. 5a, 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 Fig. 5b). [Note: the dots in Figs. 4 and 5 represent quality-controlled points. This issue is discussed more in depth in section 6.]

<sup>1</sup> 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 (maximum of cost function points at the direction step size of the GMF LUT, i.e.,  $2.5^\circ$ ).



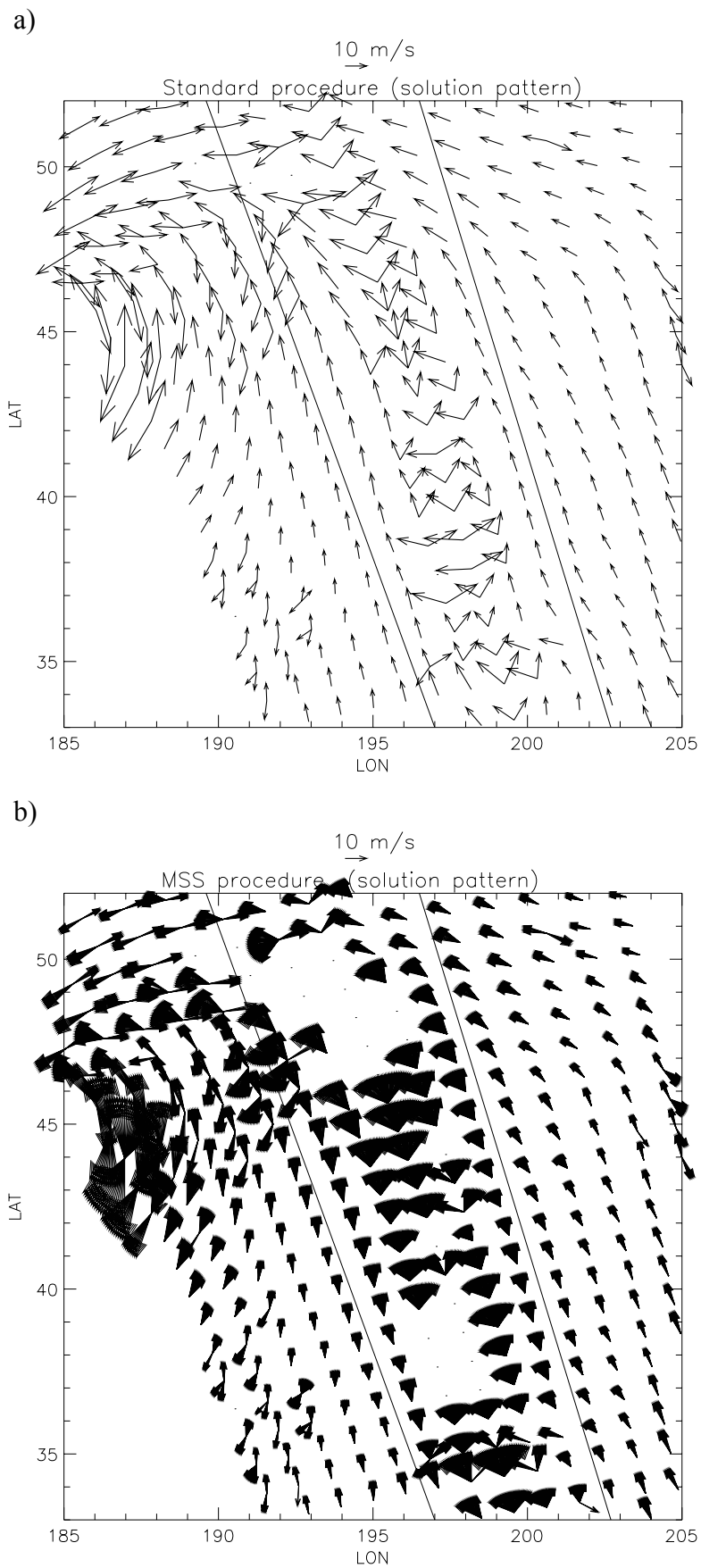


Figure 5 Same as Fig. 4 but for QuikSCAT ambiguous wind field using (a) the standard inversion output (cost function minima); and (b) the multiple solution scheme. Only solutions with probability above  $10^{-5}$  (a) and  $2 \times 10^{-7}$  (b) are shown.

It seems reasonable to test the multiple solution scheme (MSS) against the standard procedure. The former has up to 144 solutions. To avoid using the meaningless points of the cost function as solution ambiguities for the 2D-Var AR, we use the mentioned probability thresholds, i.e.,  $10^{-5}$  for the standard procedure and  $2 \times 10^{-7}$  for the MSS. To further reduce the number of 2D-Var gradient computations, these probability thresholds may be tuned (increased), for example, but this was not necessary here.

## 5. COMPARISON BETWEEN THE STANDARD PROCEDURE AND THE MSS

As discussed in section 3(b), 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) forecast 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 First Guess at Appropriate Time (FGAT) ECMWF winds as reference. These have an accuracy of about  $1 \text{ ms}^{-1}$  RMS in the wind components (e.g. Stoffelen, 1998).

### (a) Statistical results

Three days of QuikSCAT and ECMWF FGAT collocated winds at 100-km resolution are used in the comparison. Table 5 shows the 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 is more than  $0.5 \text{ m s}^{-1}$  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 (not shown) indicate better agreement of MSS analysis (compared to standard analysis) with ECMWF in both the sweet and the nadir swath.

**Table 5** RMS of vector differences<sup>1</sup> ( $\text{m s}^{-1}$ )

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

<sup>1</sup> The difference of the two wind sources with ECMWF is taken.

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 4, 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 the RMS between the MSS ( $2.45 \text{ m s}^{-1}$ ) and NCEP ( $2.96 \text{ m s}^{-1}$ ). The observations and the constraints on meteorological balance and spatial consistency are therefore the most dominant factors in the retrieval.

**Table 6** RMS of vector differences<sup>1</sup> at Nadir

	Standard procedure	MSS
Wind Speed ( $\text{m s}^{-1}$ )	1.56	1.50
Wind Direction ( $^{\circ}$ )	18.82	14.39

<sup>1</sup> The difference of the two wind sources with ECMWF is taken.

Similar to Table 5, Table 6 shows the RMS of vector differences between ECMWF and the different QuikSCAT wind sources (i.e., standard procedure and MSS) but separated into wind speed and direction and only for the nadir region. The NCEP scores are similar to the standard procedure (not shown). The MSS shows a slight improvement (better agreement with ECMWF) in the wind speed accuracy compared to the standard procedure. The main improvement is in wind direction, where the MSS is more than  $4^{\circ}$  lower in RMS 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  $J_0^{\text{scat}}$  (8). We now further investigate the effect of  $J_0^{\text{scat}}$ .

### ***MSS probabilistic behavior***

A way to test the consistency of the MSS is to verify the a priori probabilities of the solutions. Figure 6a 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 deviation from 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 Fig. 6b), 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 exploiting probability in resolving the large number of solutions provided by the MSS, thus reducing dependency (of the MSS) on the background, as discussed at the beginning of this section.

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 Fig. 6b), where the probabilistic behavior is far from being consistent. Figure 6b 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<sup>1</sup>. In particular, below  $10^{-4}$ , the data are of poor quality (close to  $4 \text{ m s}^{-1}$  RMS), indicating that the probability threshold of  $2 \times 10^{-7}$  initially used by MSS (see section 4) may be increased to improve the quality of the retrievals. This is a QC issue, which will be further discussed in section 6.

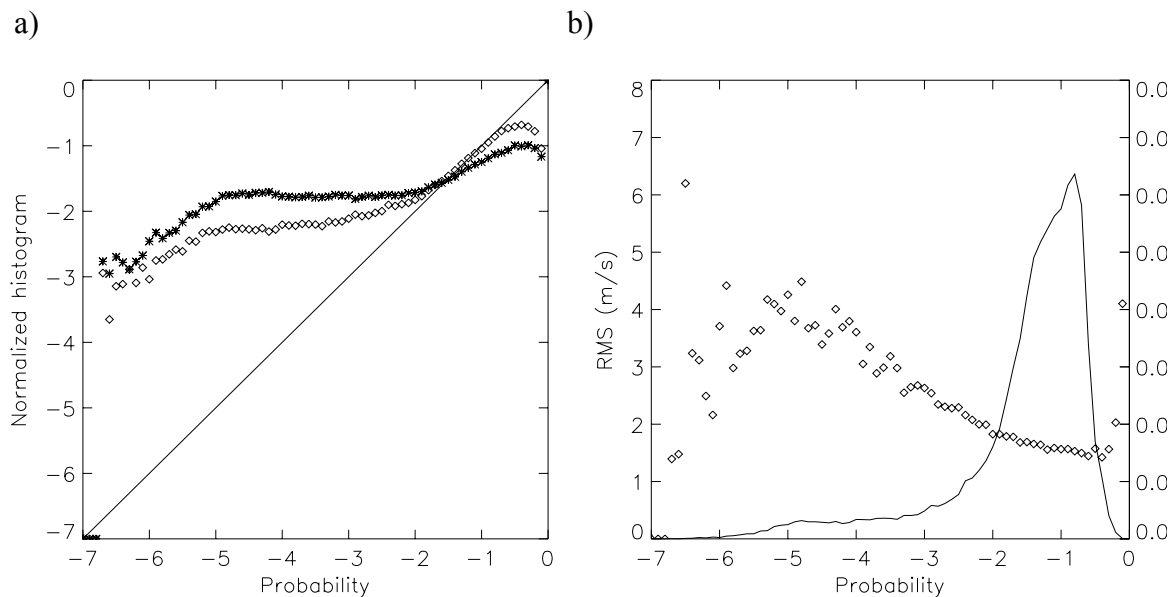


Figure 6 (a) Number of times (normalized and in logarithmic scale) that a solution with a particular probability value is selected (diamond) or closest to NCEP (star) versus probability (logarithmic scale). (b) Normalized histogram of selected solutions (solid line) and mean RMS of vector difference between the selected solutions and ECMWF winds (star) versus probability (logarithmic scale).

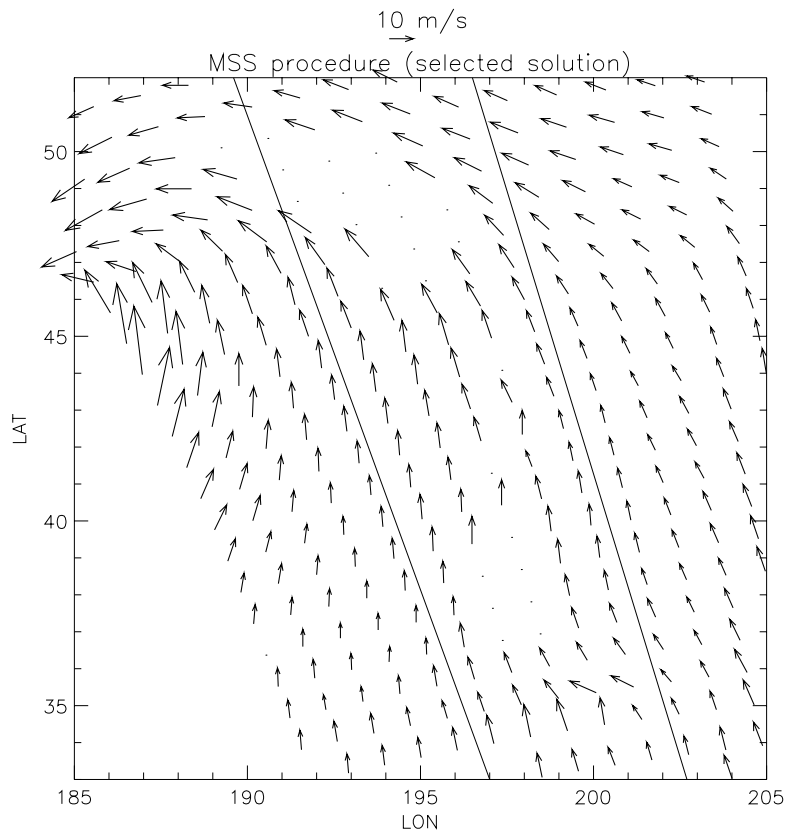
### (b) Cases

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 can be found in Portabella and Stoffelen 2002b].

Figure 7 shows the MSS selected wind field for the same poor-quality case as Figs. 4 and 5. As discussed in section 4, in contrast with the standard procedure, the MSS provides solutions in the direction of the mean flow in the nadir swath (see Fig. 5). As such, a spatially more consistent

<sup>1</sup> Below probability of  $8 \times 10^{-6}$  the number of data is very small (see solid line in Fig. 6b) and therefore not statistically significant, as denoted by the noisy RMS values in the left part of Fig. 6. This is also true for probability above  $10^{-0.2}$ .

and realistic wind field is found when using the MSS. This is shown in Figs. 4 and 7, 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 section 6), are still present though.



*Figure 7 Same as Fig. 4 but for MSS retrieved wind field.*

Figure 8 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 Fig. 8a. The MSS (Fig. 8b) 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).

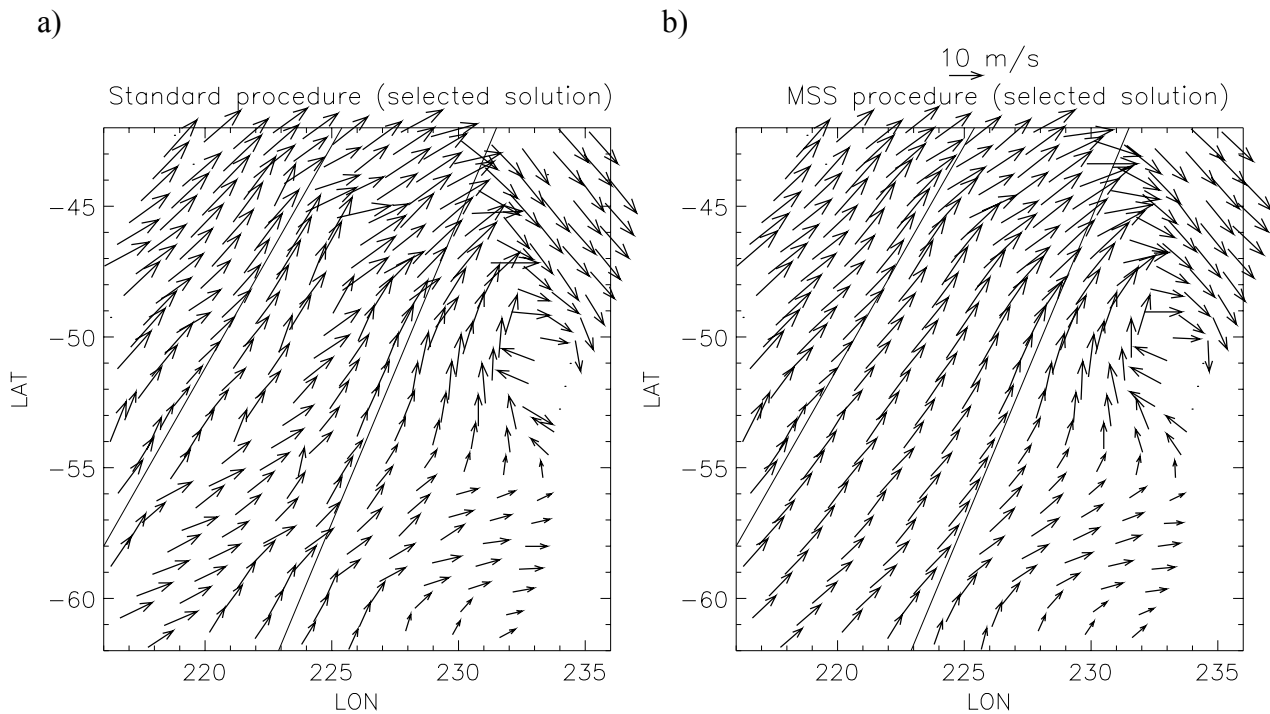


Figure 8 QuikSCAT retrieved wind field using the standard procedure (a) and the MSS (b). The acquisition date is February 3 2002 at 02 UTC. The solid lines separate the sweet-right (left side), the nadir (middle), and the sweet-left (right side) regions of the QuikSCAT swath.

Figure 9 shows a low wind speed case. Again, the standard wind field (Fig. 9a) shows a noisy pattern in the nadir swath, which is successfully filtered by the MSS (Fig. 9b). The 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 4). As shown in Fig. 9, the MSS is successfully filtering the noise in the sweet swath as well.

Figure 9c 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.

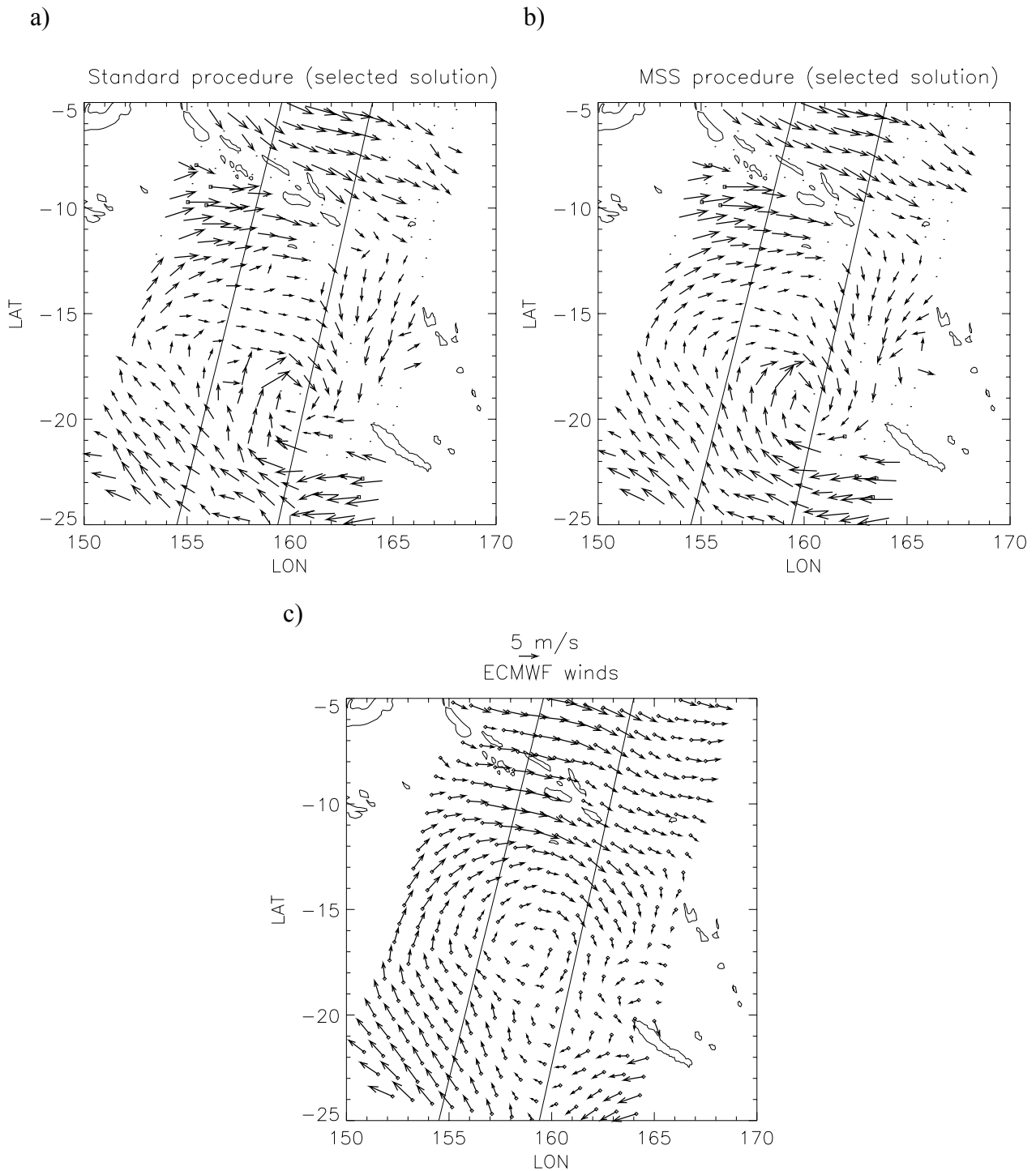


Figure 9 QuikSCAT retrieved wind field using the standard procedure (a) and the MSS (b), and ECMWF wind field (c). The acquisition date is February 3 2002 at 07 UTC. The solid lines separate the sweet-right (left side), the nadir (middle), and the sweet-left (right side) regions of the QuikSCAT swath.

## 6. 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.

### *(a) 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 Fig. 5 (see section 4). Figure 5b shows the effects of using the 25-km QC recommended by Portabella and Stoffelen (2002d), i.e., KNMI QC<sup>1</sup> + JPL rain flag<sup>2</sup> in the nadir and only KNMI QC in sweet regions, in comparison with Fig. 5a, where only the KNMI QC has been applied. On the one hand, as reported by Portabella and Stoffelen (2002d), 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 Fig. 5a removed in Fig. 5b). 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 Fig. 5a, removed in Fig. 5b). 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 Fig. 5b (notice the absence of solutions aligned with the mean flow in a few nadir WVCs).

### *(b) 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 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.

---

<sup>1</sup> 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.

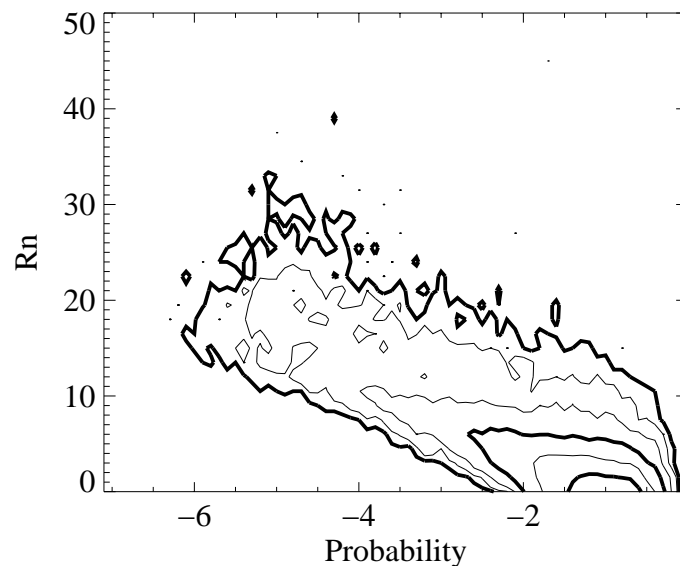
<sup>2</sup> The rain flag developed by JPL (see Huddleston and Stiles 2000) looks for the probability of encountering a columnar rain rate that is greater than 2km\*mm/hr. This probability value is read directly from a table based on several input parameters including average brightness temperature (both H-pol and V-pol), normalized inter-view  $\sigma^\circ$  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  $\sigma^\circ$  values used in wind retrieval is affected by rain.



As discussed in section 5(a), the MSS selected solutions with low probability values, i.e., below  $10^{-4}$ , are of poor quality (see Fig. 6b). 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.,  $2 \times 10^{-7}$ ). However, by increasing the probability threshold, we will also decrease the number of MSS ambiguous solutions (see section 4). 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 Fig. 10, the Rn increases with decreasing probabilities of the selected solution. Since the quality of the data is decreasing with decreasing probabilities (Fig. 6b), a Rn threshold would not only remove poor quality data (see above discussion on 100-



*Figure 10 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.*

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 section is required. A combination of some of these procedures may be more appropriate.

## 7. CONCLUSIONS

In this study, a probabilistic approach is used to improve the QuikSCAT retrievals used for assimilation into NWP. The standard wind retrieval procedure, which only considers as ambiguous solutions the local MLE cost function minima, results in inaccurate winds in the poor-azimuth-diversity nadir region of QuikSCAT.

The MLE cost function is transformed into an adequate probability cost function valid for application to 25 and 100-km retrievals. 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.

A median filter AR, in which the probability of each solution is not explicitly used in the final selection, is inappropriate since solution probabilities turn out to be very informative. 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 between the standard procedure and the MSS is performed using forecast NCEP winds as background term for 2D-Var and FGAT 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, consistent with the a priori set probability, whereas the closest-to-NCEP solution is rather inconsistent. 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.

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.

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 spectral analysis of the retrieved fields or by validating MSS with in-situ data. This could open the way for testing MSS at higher resolutions.

In addition to the MLE-based QC procedure at 25-km resolution (see Portabella and Stoffelen 2001; Portabella and Stoffelen 2002d) we suggest: 1) a similar procedure, but at 100-km, in order to reject spatially variable and unrepresentative WVCs; 2) a variational QC to remove large inconsistencies with the analysis field.

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 2, 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 since the almost unique wind information (i.e., low ambiguity) of the inner swath will be extrapolated to the few nodes of the outer regions. However, 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 variational QC in the outer region prior to operationally assimilate the QuikSCAT outer-region winds into NWP.

## APPENDIX: NORMALIZED RESIDUAL

The MLE represents a squared distance, which is “normalized” by the measurement error variance or noise (see (1)). The MLE formulation is derived from a Gaussian probability distribution with unit variance of this normalized distance. As such, the MLE should behave uniformly across the swath (WVC number) and over any wind condition. However, in practise the measurement noise is misspecified (a more detailed analysis on this can be found in Portabella and Stoffelen 2002c). In order to cure this problem, for a given wind and node number, an expected MLE value can be estimated, and used to normalize the MLE in the following way:

$$R_n = \text{MLE} / \langle \text{MLE} \rangle \quad (10)$$

where the MLE value represents any point of the cost function for a particular WVC, and  $\langle \text{MLE} \rangle$  is the expected MLE for that WVC (node number) and wind condition.

Note that we are not trying to change the relative weights of (1) but rather provide a uniform MLE. In other words, we do not intend to optimise the inversion, i.e., change the MLE formula, since we believe that (1) is already working reasonably well for such purpose; we rather define a more uniform parameter which may be useful for QC and for assessing solution probability.

### *$\langle \text{MLE} \rangle$ for QuikSCAT*

The purpose of the  $\langle \text{MLE} \rangle$  is to compensate the misestimation of the measurement noise in order to correct the MLE dependencies. Therefore, an accurate knowledge of the instrument and geophysical noise is needed. In this respect, Figa and Stoffelen (2000) used an instrument error model derived by Cavanié (1997) to compute the  $\langle \text{MLE} \rangle$  of the NSCAT  $R_n$  parameter. For QuikSCAT however, there is no instrument error model available and an alternative method has to be sought.

Portabella and Stoffelen (2001) used about 4 days of QuikSCAT data to compute the mean MLE as a function of WVC number and wind condition. After filtering the noise on such surface, they fit a 2D function, which is used to compute the  $\langle \text{MLE} \rangle$  at any observation point. Further details on how to compute  $R_n$  in this case can be found in Portabella and Stoffelen (2002a).

## REFERENCES

Cavanié, A., October 1997. Evaluation of Kp on central and lateral antennas of NSCAT over arctic sea ice, *Report presented at the 12<sup>th</sup> ASCAT Science Advisory Group Meeting*, Darmstadt, Germany.

Figa, J., and Stoffelen, A., July 2000. On the assimilation of Ku-band scatterometer winds for weather analysis and forecasting, *IEEE Trans. on Geoscience and Rem. Sens.*, **38**(4), 1893-1902.

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

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

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

Pierson, W.J., 1989. Probabilities and statistics for backscatter estimates obtained by a scatterometer. *J. Geophys. Res.*, **94**(C7), 9743-9759.

Portabella, M., and Stoffelen, A., 2001. Rain detection and quality control of SeaWinds. *J. Atm. and Ocean Techn.*, **18**(7), 1171-1183.

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

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

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

Portabella, M., and Stoffelen, A., 2002b. A probabilistic approach for SeaWinds data assimilation: an improvement in the nadir region. *Visiting Scientific report for the EUMETSAT NWP SAF*, available at <http://www.eumetsat.de/en/area4/saf/internet/>.

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

- Portabella, M., and Stoffelen, A., 2002d. A comparison of KNMI quality control and JPL rain flag for SeaWinds. *Can. Jour. of Rem. Sens.*, **28**(3), 424-430.
- Rodgers, C. D., 2000. Inverse methods for atmospheric sounding: theory and practice. *World Scientific Publishing Co.*
- Spencer, M.W., Wu, C., and Long, D.G., 1997. Tradeoffs in the design of a spaceborn scanning pencil beam scatterometer: application to SeaWinds. *IEEE Trans. Geosci. Rem. Sens.*, **35**(1), 115-126.
- Stiles, B.W., Pollard, B.D., Dunbar, R.S., 2002. Direction interval retrieval with thresholded nudging. *IEEE Trans. Geosci. Rem. Sens.*, **40**(1), 79-89.
- Stoffelen, A., and Anderson, D., 1997a. Ambiguity removal and assimilation of scatterometer data. *Quart. J. R. Met. Soc.*, **123**, 491-518.
- Stoffelen, A., and Anderson D., 1997b, Scatterometer data interpretation: measurement space and inversion. *J. Atm. and Ocean. Techn.*, **14**(6), 1298-1313.
- Stoffelen, A., Van Beukering, P., 1997. 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.
- Stoffelen, A., 1998. Scatterometry. *PhD thesis at the University of Utrecht*, ISBN 90-393-1708-9.
- Stoffelen, A., de Vries, J., and Voorrips, A., September 2000. Towards the real-time use of QuikSCAT winds. *Final Report USP-2/00-26*, Beleidscomissie Remote Sensing, The Netherlands.
- Undén, P., Kelly, G., Le Meur, D., and Isaksen, L., 1997. Observing system experiments with the 3D-Var assimilation system. *Technical Memorandum No. 244*, European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, United Kingdom.

## ACKNOWLEDGEMENTS

As member of the KNMI QuikSCAT team, John de Vries, has extensively contributed to the work described in this report. We acknowledge the help and collaboration of our colleagues working at KNMI, in particular Jos de Kloe. The QuikSCAT data were obtained from the NASA Physical Oceanography Distributed Active Archive Center, at the Jet Propulsion Laboratory / California Institute of Technology, and the National Oceanic and Atmospheric Administration. Furthermore, this work is only possible due to the European Meteorological Satellite Organization (EUMETSAT) following grants: QuikSCAT fellowship and Visiting Scientist of the Numerical Weather Prediction Satellite Application Facility program.