Ambiguity Removal and Product Monitoring for SeaWinds

J. de Vries, A. Stoffelen, J. Beysens

Observation Research Division
KNMI
Wilhelminalaan 10, 3732GK
De Bilt, Netherlands

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.

Copyright 2003, EUMETSAT, All Rights Reserved.

Change record			
Version	Date	Author / changed by	Remarks
1.0	21-7-03	J. de Vries	Initial version
1.1	10-01-04	A. Stoffelen	Product Monitoring Flag inclusion
1.2	05-Oct-04	A.H. Verhoef	Addition of Appendix A
1.2	02-Mar-05	A.H. Verhoef	Some additions and clarifications in Ch. 5

Table of Contents

INTRODUCTION

This report is a technical note on the implementation of Ambiguity Removal (AR) and the product monitoring flag in the QuikSCAT Data Processor (QDP). It is written for the NWP community, specifically for users who intend to use the SeaWinds wind product for applications like data assimilation. The purpose of this document is to inform the intended user on the tests and implementation details concerning AR and product monitoring in the development of a 100km surface wind product based on SeaWinds measurements.

In chapter 1 a general introduction is given on the SeaWinds instrument describing the instrument details that are assumed known to the reader in the subsequent chapters. Chapter 2 gives a theoretical description of variational ambiguity removal. Chapter 3 describes the specification of the background error covariance matrix. Chapter 4 deals with Variational Quality Control needed to filter out observations that contain gross errors. Chapter 5 reports on the development of a product monitoring flag. Finally chapter 6 gives a summary and outlook. Chapter 6 summarizes the references used throughout this report.

1. THE SEAWINDS INSTRUMENT

SeaWinds on QuikSCAT constitutes a "quick recovery" mission to fill the gap created by the loss of data from the NASA Scatterometer (NSCAT), when the ADEOS-1 satellite lost power in June 1997. QuikSCAT was launched from VandenBerg Air Force Base (USA) in June 19, 1999. A similar version of the SeaWinds instrument flew on the Japanese ADEOS-II satellite launched in late 2002, which was regretfully lost in October 2003.

1.1 Instrument Characteristics

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

The SeaWinds instrument uses a rotating 1-meter dish antenna with two spot beams, an H-pol beam and a V-pol beam at incidence angles of 46º and 52º 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 sub-track, making approximately 1.1 million 25-km ocean surface wind vector measurements and covering 90% of the Earth's surface every day. These measurements will help to determine atmospheric forcing, ocean response and air-sea interaction mechanisms on various spatial and temporal scales.

The SeaWinds swath is divided into equidistant across-track wind vector cells (WVC) numbered from left to right when looking into the satellite's propagation direction. The nominal WVC size is 25 km x 25 km, and all backscatter measurements centered in a WVC are used to derive the WVC wind solutions. Due to the conical scanning, a WVC is generally viewed when looking forward (fore) and a second time when looking aft. As such, up to four measurement classes (called "beam" here) emerge: H-pol fore, H-pol aft, V-pol fore, and V-pol aft, in each WVC. Due to the smaller swath (1400 km) viewed in H-pol at 46º degrees incidence, the outer swath WVCs have only V-pol fore and aft backscatter measurements. For more detailed information on the QuikSCAT instrument and data we refer to [*Spencer et al* (1994), *JPL* (1999) and *Leidner et al* (2000)].

1.2 Swath Definition

SeaWinds on QuikSCAT has an antenna geometry that is dependent on WVC number or cross-track location due to its circular scan on the ocean. As shown in figure 1.1, the SeaWinds swath is subdivided in several regions classified in three different categories. The nadir region or *nadir swath* (category II) has fore and aft looks of both beams (H-pol and V-pol) that are nearly 180º apart. At the outer regions or *outer swath* (category III) there is no inner beam information. Moreover at the edges of the swath, the outer V-pol beam fore and aft looks are nearly in the same direction.

The skill of the wind retrieval algorithm or *inversion* depends very much on the number of backscatter measurements, the beam polarizations and the 'azimuth diversity', i.e. the spread in azimuth angle among the measurements in a WVC [*Portabella*, 2002]. In swath areas II and III the skill of the wind retrieval algorithm is expected to decrease with respect to the rest of the swath, the so-called sweet regions (category I), where there are nominally four measurements (fore-inner, fore-outer, aft-inner and aft-outer) with sufficient azimuth diversity. In the near real-time BUFR product of NOAA/NESIDS there are 76 WVCs in the across-track direction. The distribution of the WVCs over the three categories is as follows: WVC numbers 1-11 and 66-76 correspond to the outer swath, WVC numbers 12-28 and 49-65 to the sweet swath and WVC numbers 29-48 to the nadir swath. The KNMI 100km SeaWinds wind product is derived from the NOAA/NESDIS product and is created with the QuikSCAT Data Processor (QDP) (*de Vries*, 2003). In the 100km product there are 19 WVCs in the across-track direction. WVC numbers 1-3 and 17-19 correspond to the outer swath; WVC numbers 4-7 and 13-16 represent the sweet swath and WVCs 8-12 are in the nadir swath.

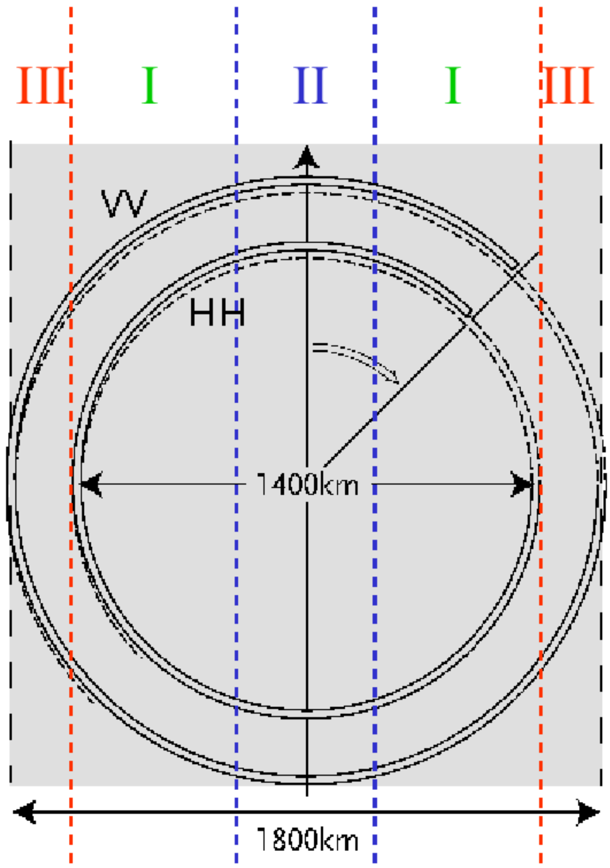

Figure 1.1: Schematic illustration of the SeaWinds swath. The circles represent the two scanning rotating antennae with H-pol (HH) and V-pol (VV) at 46° and 54° incidence angles respectively. Regions I, II and III represent parts of the swath with different wind retrieval skill. Adopted from Stoffelen(1998). In all parts of the swath multiple wind vector solutions will result from the wind retrieval, where the ambiguous properties are different in the different swath regions [Stoffelen, 1998].

2. VARIATIONAL AMBIGUITY REMOVAL

Ambiguity Removal (AR) is the process of selecting the surface wind from an ambiguous set of wind vector solutions at each observation point in a way that is both spatially and meteorologically consistent. At KNMI a *variational* scheme named 2D Variational Ambiguity Removal (2DVAR) was developed for the near real-time ambiguity removal in ERS winds (*De Vries, Stoffelen*, 2000). This scheme was enhanced and generalized within the framework of the NWP-SAF to be suitable for the ambiguity removal applied to other types of scatterometers like SeaWinds.

The 2DVAR scheme uses a short range forecast of the 10m wind from a numerical weather prediction (NWP) model and scatterometer observations consisting of sets of ambiguous wind vector solutions obtained from the inversion of the GMF, ordered on the swath and combines them in a statistically optimal way. The scheme is based on Bayesian probability theorem and uses a maximum probability approach (*Lorenc*, 1986). It assumes that background errors and observation errors are *normally distributed* and uses an *incremental formulation* (*Courtier*, 1994) for computational efficiency. The incremental form is obtained by subtracting the background field from the ambiguous SeaWinds wind vector solutions and from the control variable

$$\delta\underline{x} = \underline{x} - \underline{x}_b = \begin{pmatrix} u' \\ v' \end{pmatrix}$$

where *u* and *v* are the wind vector component increments. The optimal solution or *analysis* is obtained by minimizing the cost function

$$J = \delta\underline{x}^T B^{-1} \delta\underline{x} + \left[\sum_{i=1}^{nsol} \left[\left(\mathrm{H}\delta\underline{x} - \underline{d}_i \right)^T R^{-1} \left(\mathrm{H}\delta\underline{x} - \underline{d}_i \right) - 2\ln \underline{w}_i \right]^{-p} \right]^{-1/p}$$

In the expression for the cost function *B* and *R* are the background error and observation error covariance matrix respectively. H is the forward model, which is simply horizontal bi-linear interpolation from the analysis grid to the observation points; *d* symbolizes the observation-minus-background differences and *w* is the normalized solution probability for each wind vector solution. The value of p in the exponents is currently 4 (*Stoffelen & Anderson*, 1999). The observational term between the small square brackets was developed for SeaWinds on QuikSCAT. For more information on this cost function and especially on the concept of solution probability see (*Stoffelen, Voorrips, De Vries*, 2000).

For the minimization of the cost function 2DVAR applies a quasi-Newton method (*Gilbert and Lemarechal*, 1989). This method requires the gradient of the cost function,

$$\nabla J = B^{-1}\delta\underline{x} + \nabla J_o^{scat}$$

with,

$$\nabla J_o^{scat} = \left[\sum_{i=1}^{nsol} J_i^{-p} \right]^{-(1+1/p)} \cdot \sum_{i=1}^{nsol} J_i^{-(p+1)} \frac{\partial J_i}{\partial \delta\underline{x}}$$

and

$$J_i = \left(\mathrm{H}\delta\underline{x} - \underline{d}_i\right)^T R^{-1}\left(\mathrm{H}\delta\underline{x} - \underline{d}_i\right) - 2\ln \underline{w}_i$$

To enhance the convergence of the minimization preconditioning is applied with the transformation

$$\delta\underline{Z} = \Lambda^{-1/2} WF\delta\underline{x}$$

In this transformation $\delta\underline{Z}$ is the new preconditioned spectral control variable, $\Lambda^{-1/2}$ is the inverse square root of the background error covariance matrix of stream function ψ and velocity potential χ, which is diagonal (see chapter 3). F is the Fourier operator and W the Helmholtz operator that converts wind components into ψ and χ . Through this transformation of variables the minimization takes place in the spectral domain, which allows the minimization routine to adjust every wave number separately. The transformation also simplifies the structure of the background error covariance matrix to the identity matrix, which is beneficial for the computational efficiency of 2DVAR. With the new control variable the cost function becomes

$$J = \delta\underline{Z}^T \delta\underline{Z} + J_o^{scat}\left(\delta\underline{Z}\right)$$

Because the minimization is in the spectral domain special care has to be taken w.r.t. the periodic nature of Fourier transform. If no precautions are taken then solution increments due to the presence of observations near an edge of the grid will affect the solution at the opposite edge in the physical domain. Therefore in 2DVAR an extended grid is used as depicted in figure 2.1. A square bounding box is created around the observations extended with zones that do not contain observations. The scatterometer swath may be situated inside the bounding box in two ways. When making use of the regular grid of the swath (I), it can be aligned with a corner of the bounding box. Another way is to center the bounding box on the center of gravity of the scatterometer swath (II). On the area including the extensions a regular grid is defined so that the bounding box edge and the outer edge coincide with grid lines. The grid has nominally 32x32 nodes, a spacing of 100km and extensions of 5 nodes. Solution increments due to observations at the edge of the bounding box change the solution locally in the extension zone but not inside the bounding box on the opposite side.

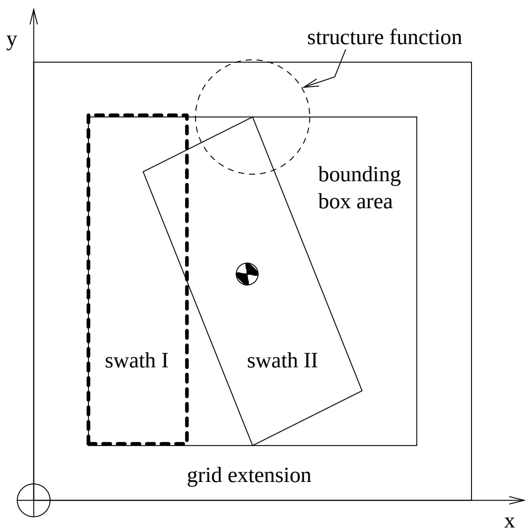

Figure 2.1: The computational grid with extension zone.

After the minimization the analysis increment field is added to the background field to obtain the analysis wind field.

$$\underline{x}_a = \underline{x}_b + \delta \underline{x}$$

The analysis wind field is interpolated to the observation points and used to select the closest of the ambiguous wind vector solution in terms of vector rms difference. The selected ambiguity is considered to be the actually observed surface wind.

For the implementation in QDP of the 2DVAR scheme, several tests and comparisons have been carried out to define and characterize the SeaWinds cost function and validate its correctness. The tests are described in the following chapters.

3. THE BACKGROUND ERROR COVARIANCE MATRIX

Implementation of the cost function of 2DVAR for SeaWinds observations requires specification of the background error covariance matrix B. 2DVAR is a spatial filter and it is the B-matrix that constraints the spatial structure of the solution in the minimization. In 2DVAR, B is constant and modeled with prior knowledge on the spatial structure of the background error of the surface wind field in NWP models. It is necessary to model the structure of B with efficiency in mind so that its repeated inversion will be computationally feasible.

The background errors are assumed spatially homogeneous and isotropic, which translates into two-point error covariances being functions of the separation distance only (Daley, 1985). As a result B becomes a *symmetric* and *positive-definite* matrix, i.e.

$$B_{\delta \underline{x}} = B_{u,v} = \begin{pmatrix} \langle u', u'^T \rangle & \langle u', v'^T \rangle \\ \langle v', u'^T \rangle & \langle v', v'^T \rangle \end{pmatrix} = \begin{pmatrix} B_{uu} & B_{uv} \\ B_{vu} & B_{vv} \end{pmatrix}$$

with the prime denoting error, which we will drop from the notation. The structure of B turns the matrix-vector multiplication with δx into a convolution, which, subject to Fourier transform, reduces into a vector-vector product for scalars, i.e. B becomes diagonal. For vector quantities like wind this implies that the background error covariance matrix becomes block diagonal in Fourier or *spectral* space. Due to its symmetry in the physical domain, B is a real-valued matrix in spectral space.

The cross-covariances in B are generally not equal or close to zero. Cross-covariances of errors in stream function Ψ and velocity potential χ however are very small. If we introduce, as a first step to the transformation for preconditioning mentioned above, a change of the control variable to increments of ψ and χ,

$$\delta \underline{X} = WF\delta \underline{x} = \begin{pmatrix} \psi' \\ \chi' \end{pmatrix}$$

the covariance matrix of background errors of stream function and velocity potential becomes,

$$B_{\psi,\chi} = \begin{pmatrix} \langle \psi', \psi'^T \rangle & \langle \psi', \chi'^T \rangle \\ \langle \chi', \psi'^T \rangle & \langle \chi', \chi'^T \rangle \end{pmatrix} = \begin{pmatrix} B_{\psi\psi} & \approx 0 \\ \approx 0 & B_{\chi\chi} \end{pmatrix} = \Lambda$$

Because this matrix is diagonal a further factorization is now possible into error variances and error correlations, i.e.

$$\Lambda = \Sigma C \Sigma$$

with

$$\Sigma = \begin{pmatrix} \Sigma_\psi & 0 \\ 0 & \Sigma_\chi \end{pmatrix}, \; C = \begin{pmatrix} C_{\psi\psi} & 0 \\ 0 & C_{\chi\chi} \end{pmatrix}$$

Stream function and velocity potential are not observed quantities but their error variances and error correlations can be derived from those of wind by Helmholtz theorem and implemented in W. The derivation gives for the relation between the error-variances of ψ and χ and wind (Daley, 1985) in the physical domain

$$\sigma_\psi^2 = \left(1-\nu^2\right)\sigma_u^2 L_\psi^2,\ \sigma_\chi^2 = \nu^2 \sigma_u^2 L_\chi^2$$

The error correlations in matrix *C* when transformed to the physical domain represent circular symmetric functions of distance also known as *structure functions.* For error correlations many empirical model functions are known from literature that can be used as these structure functions for wind (*Daley*, 1985), which in turn determine the structure functions for ψ and χ. Conversely, appropriate structure functions for ψ and χ define the structure functions for wind and should exhibit realistic wind increment structures. A typical structure function contains parameters that have to be set with values that characterize the behavior of the particular geophysical parameter. Specification of the *B*-matrix comes down to finding suitable values for these structure function coefficients. In 2DVAR the structure function used for stream function and velocity potential are

$$c_{\psi\psi} = \left(1-\nu^2\right)e^{-\left(r/L_\psi\right)^2},\ c_{\chi\chi} = \nu^2 e^{-\left(r/L_\chi\right)^2}$$

The characteristic length scales L_ψ and L_χ, the ratio between the rotational and divergent part in the total kinetic energy of the background wind error ν^2, as well as the background error variances are the parameters that have to be set to specify *B*. Through the structure function of ψ (χ) the amount of rotation (divergence) in the solution for the wind field increments can be specified.

To obtain suitable settings for the *B*-matrix different settings for the structure functions were used in a comparison study with an alternative AR scheme that served as a reference. The settings for the structure functions resulting in realistic wind structures in 2DVAR that are used in the comparison are given in table 3.1.

Setting	1		2		3		4	
	L_Ψ / L_χ	ν^2	L_Ψ / L_χ	ν^2	L_Ψ / L_χ	ν^2	L_Ψ / L_χ	ν^2
NH	400km	0.0	400km	0.2	500km	0.2	300km	0.1
Trop.	500km	0.5	500km	0.4	400km	0.5	600km	0.5
SH	400km	0.2	400km	0.0	300km	0.2	300km	0.1

Table 3.1: Settings for the structure function parameters in 2DVAR.

The background error standard deviation and observation error standard deviation were kept constant. The observation error standard deviation is set to 1.7 m/s. The standard deviation of error of ψ and χ depend on the characteristic length scale and ν. They are assigned values based on a constant value of the background wind vector component error standard deviation of 2.0 m/s and are thus not taken constant.

A comparison with the alternative AR scheme for SeaWinds has not been accomodated. Therefore a comparison was made with the PRESCAT method used for AR in ERS observations. PRESCAT is a statistical filter that uses wind vector continuity as a constraint and relies on the input of a background wind field albeit for initialization only (*Stoffelen*, 1998).

3.1 Method of Comparison

The method of comparison is the same as the statistical analysis described in *Stoffelen* (2002). The difference with 2DVAR used in that comparison is mainly an increase of the error standard deviation, which decreases the weight of the background field in the analysis and, secondly, the application of latitude-dependent structure functions. We expect that these changes enable a better fitting of the observations and so enhance the AR skill of 2DVAR relative to other AR-schemes. Our expectation is based on a detailed comparison and performance analysis of AR scheme in *Stoffelen* (2002).

The method of comparison was carried out with the near real-time processing of global ERS measurements in mind. This means that the available background (BG) is often a 36h forecast of 10m winds from a global NWP model. This BG is of poorer quality than a first guess, FG, (3-9h forecast) from the next model run in a 24h assimilation cycle (see, e.g., *Stoffelen*, 1998, p.V-25). Therefore in an off-line setting the BG is used in the AR schemes and the FG is used for verification. The relative performance of 2DVAR with different settings compared to PRESCAT can be evaluated and a best performing setting can be selected.

3.2 Data

In the comparison about a month of ERS-2 scatterometer retrieved winds are used from January 2000. The retrieved winds are represented on a 500 km wide swath, sampled every 25km resulting in 19 nodes across the swath. Along the swath the sampling is also 25km. The nodes are not independent however, and the effective spatial resolution is 50km. The retrieved winds consists of two ambiguities at each node and are organized in messages of 19 rows in the along direction of the swath containing a total of 361 nodes. The data was quality controlled with a procedure that identifies and rejects 1-2% anomalous triplets at the wind retrieval (*Stoffelen*, 1998).

The data set is not entirely spatially homogeneous, which is relevant because AR schemes are spatial filters and 2DVAR in particular can only constrain spatial consistency over a limited domain. PRESCAT on the other hand is able to process all continues pieces of ocean swath from pole to pole. A procedure was used in the processing with 2DVAR in which the messages were ordered in batches of four times the spatial extend of a message length. Batches contain between 1-4 messages depending on the spatial distance from one message to the next and the first message to the last. Messages were aligned according to swath I in figure 2.1. Inspection of plots of selected surface winds showed no problems with continuity.

The BG wind field used for the two AR schemes is retrieved from ECMWF. For observation times of

- $12h<t\leq 24h$ a BG from forecasts with a lead time from 24 to 36 hours at 6-hourly intervals is used;
- $00h<t\leq 12h$ a BG from forecasts with a lead time of 36 to 48 hours at 6-hourly intervals is used .

The FG at appropriate time is used for verification. This is a cubic time interpolation between a 3-, 6-, and 9-hour forecast to the actual observation time. These ECMWF fields are available for the 0, 6, 12, and 18 UTC analyses of January 2000.

3.3 Results

The activity and performance of both schemes were investigated for all wind speeds. It turned out, as expected, that for winds below 3m/s the activity was high and the performance was low for both schemes. Therefore winds below 3m/s were not investigated any further.

3.3.1 Activity

After initialization of the AR process with the BG the selecting process gives equal selections for the BG and the AR scheme (trivial) and selections would either agree or disagree with the FG, see situations 1 and 2 in figure 3.1. After the AR process the selections of the AR scheme may have changed either from situation 1 into 3 or 2 into 4. The selections may have also remained unchanged. The change in the selections of an AR scheme with respect to the selections of the BG is what we call *activity*. In figure 3.1, activity leading from situation 1 to 3 is a change for the *worse*. Going from situation 2 to 4 is a change for the *better*.

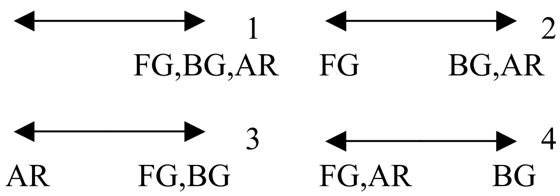

Figure 3.1: Consistency scenarios of possible selections by FG, BG and AR (see text).

Figure 3.2 shows histograms of activity scores for PRESCAT and for 2DVAR with the settings of table 3.1. In each panel the rightmost column gives the difference in selections between the FG and the BG. The total activity (worse and better added) of the AR schemes is almost always less than BG, so there is some room for increasing activity. The scores do show that both schemes are beneficial to the AR process. Changes made in the selections are mostly made for the better, except for 2DVAR with setting 3 in the tropics.

All results are discussed relative to 2DVAR with setting 1, which shows increased activity for the better compared to PRESCAT in the NH for the whole speed range. A relative increase in the divergent kinetic energy of up to 20% in the structure functions of 2DVAR (setting 2) does not affect activity very much over the whole speed range. Having, in addition, a longer characteristic length scale of 500km is slightly detrimental at both high wind speeds and low wind speeds with a decrease in activity in the latter case. A shorter characteristic length scale of 300km and a 10% increase in divergent kinetic energy gives rise to increased activity for the better. Apparantly rotational structure functions with a 300km length scale are most suitable in general to fit wind field structures in the NH.

In the tropics activity is relatively low. The plots indicate that 50% divergence in the structure functions seems right and a length scale below 500km is very detrimental and does more bad than good especially at low wind speeds and results in a significant increase in activity. A longer length scale is slightly detrimental at high wind speeds and neutral at low wind speeds. Divergent structure functions with length scales of 500-600km seem to work good for 2DVAR. PRESCAT performs best in the tropics.

Activity in the SH is somewhat lower than in the NH for all AR schemes. The schemes verify better with the FG than in the NH. Results show that it is better to reduce divergence and length scale at high wind speeds. At low wind speeds a smaller length scale only seems better.

The activity gives a good indication of the sensitivity of the scores with the variation of the structure function parameters. It tells about *how* the scores change. It does not say *how much* things change (performance).

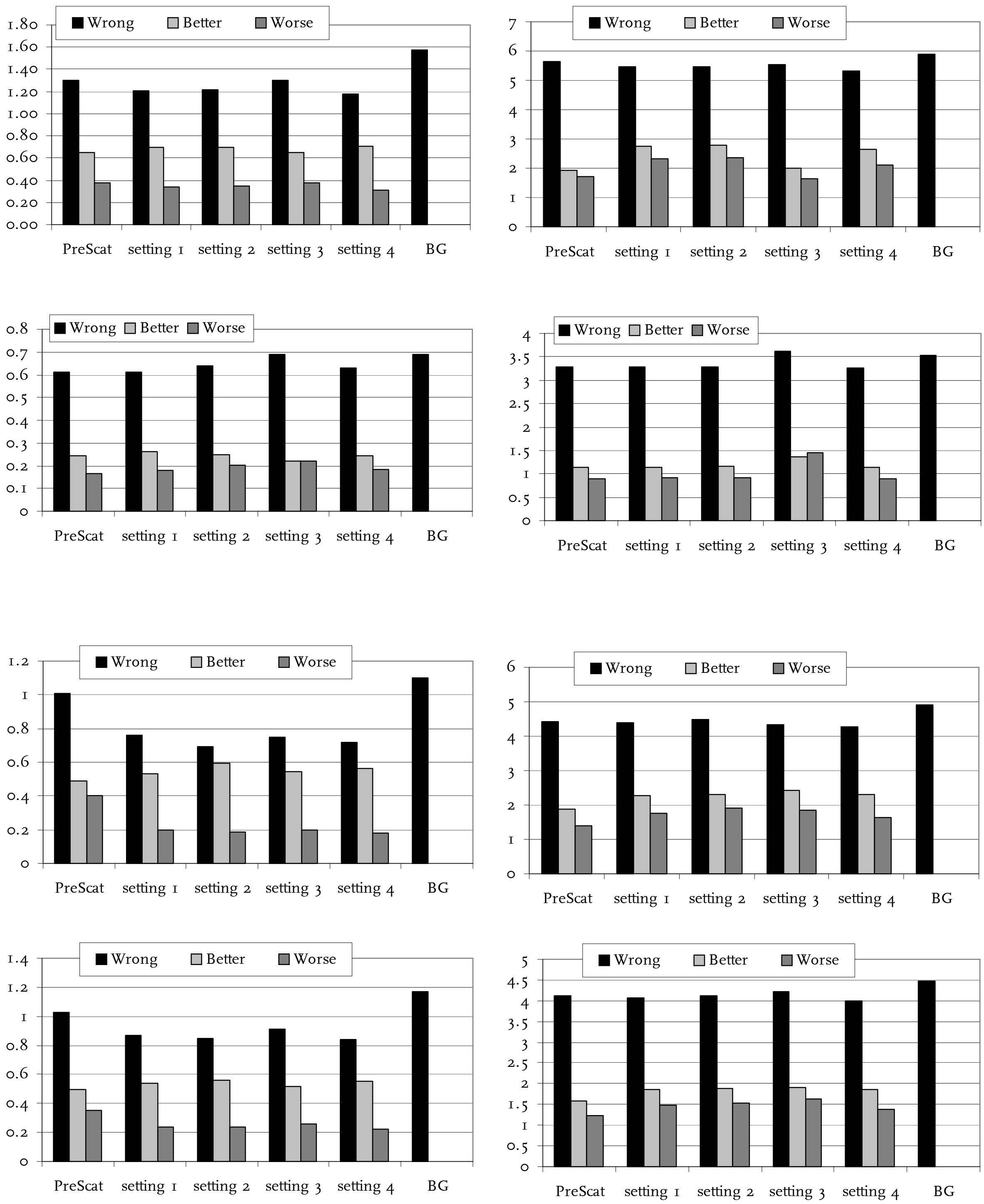

Figure 3.2: Percentage of AR scores. Left panels represent activity for wind speeds above 7m/s. Right panels are for wind speeds 3-7m/s. From top to bottom panels are for NH, tropics, SH and global respectively.

3.3.2 Performance

The *selection* performance is the difference in selection between the AR scheme and the FG. The scores from the comparison are given in figure 3.3.

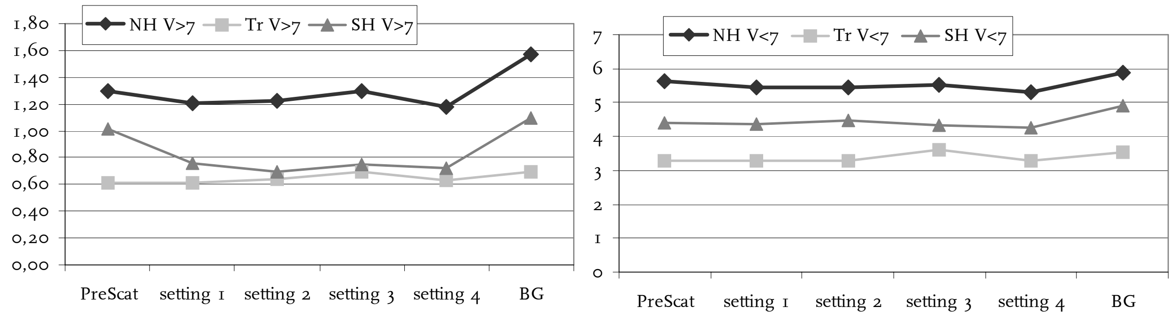

Figure 3.3: Percentage of difference with the FG for wind speeds 3-7m/s (right) and above 7m/s (left)

It is evident that PRESCAT and 2DVAR with all settings verify better with the FG than the BG, but improvements are small. The schemes verify best with the FG in the tropics and better in the SH than in the NH. Scores are better at high wind speeds. The PRESCAT scores are slightly better in the tropics than 2DVAR, except with setting 4 with which it is on par.

The *wind vector* performance translates the selection performance into a wind vector RMS performance. Results are shown in figure 3.4. Too long a characteristic length scale adversely affects the wind vector performance. 2DVAR generally performs better than PRESCAT in the NH

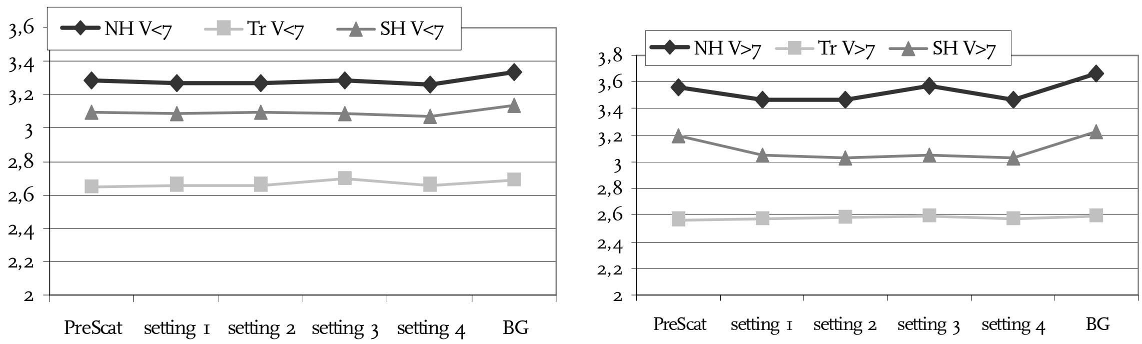

Figure 3.4: RMS of wind vector difference with the FG of the AR schemes for wind speeds of 3-7m/s (left) and above 7m/s (right).

and SH. PRESCAT is better in the tropics. 2DVAR with setting 4 gives the best results overall.

The performance of 2DVAR relative to other ERS schemes has been tested before (*Stoffelen et al*, 2002) and based on this study settings were improved here. Figure 3.5 shows the wind vector performance of 2DVAR in the currently best performing setting 4 in comparison with the schemes from that investigation. The 2DVAR scheme now performs best in the NH and SH and is on par with PRESCAT in the tropics.

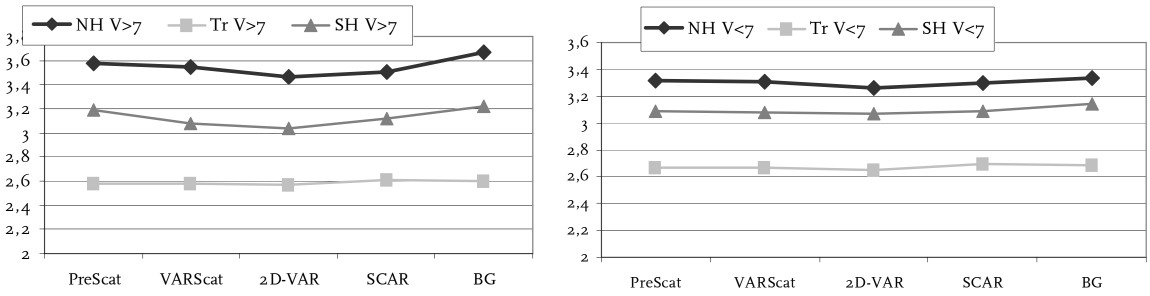

Figure 3.5: RMS of wind vector difference of AR scheme with FG versus AR scheme for wind speeds of 3-7m/s (right) and wind speeds above 7 m/s (left). Adopted from *Stoffelen et al*, 2002.

This is to be expected because the settings for 2DVAR in the NH and SH were improved based on the recommendations of *Stoffelen et al* (2002). In the tropics the recommendations imply that solution increments are divergent (see section 3.4). This favors vector continuity (trade winds), which is an explicit constraint for PRESCAT. All schemes perform best in the tropics, but improvements over BG are small.

3.4 Single Observation Test

As mentioned in the previous sections the shape and extent of the structure functions used to specify the *B*-matrix determine the spatial structure of the background wind increments. A single observation test visualizes the structure functions that define the *B*-matrix. Figure 3.6 shows the wind increment structure as a result of an observed wind vector pointing due north. The left panel shows the rotational structure for the extra tropics. The right panel shows the divergent structure as specified for the tropics for setting 4 in table 4.1.

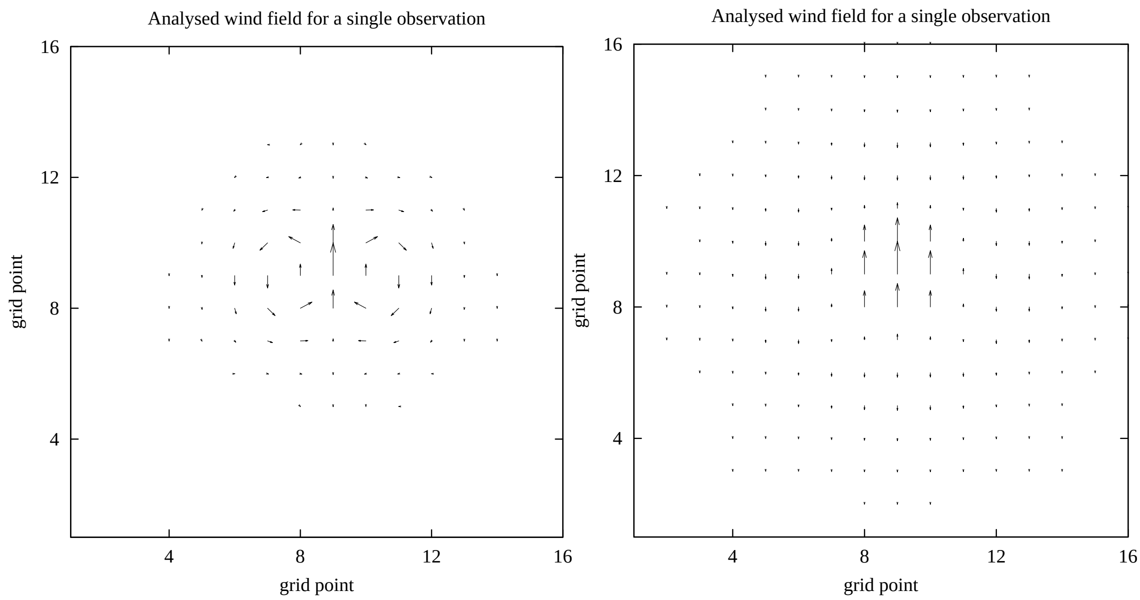

Figure 3.6: rotational wind and divergent wind increment structure

3.5 Discussion and recommendations

We have compared the 2DVAR scheme with PRESCAT using ERS winds. The results have shown that 2DVAR with setting 4 performs best in terms of wind vector RMS but it is necessary to make some comments:

- The density of the ERS data, represented on a 25 km resolution grid is much higher than the resolution of 100 kilometers for the KNMI SeaWinds product. ERS data only resolve 50-km structures, however, resulting in correlated wind solution data. As such, this may be detrimental to the performance of 2DVAR running in ERS mode, a simple up weighing of the observations relative to the BG may suffice in case of SeaWinds.
- The data set that was available only covered a winter period in the NH. During the winter the atmosphere is more dynamic, which indicates that AR schemes perform less in highly dynamic conditions. For a more balanced setting of 2DVAR that takes into account seasonal effects an annual data set would have been more appropriate. It is however encouraging that an identical setting for the SH (summer) and NH (winter) was selected here.
- The ERS scatterometer operates in C-band and is not affected by rain so Variational Quality Control (see chapter 4) was switched off in 2DVAR. SeaWinds is affected by rain so 2DVAR will operate in a different mode, which may affect performance.
- Because no direct and fair comparison with other AR schemes for SeaWinds is possible to date, the only way to assess the quality of the selected solutions is at the product level, which involves the whole processing chain (*De Vries*, 2003).

We conclude

- The 2DVAR scheme is recommended for the removal of wind direction ambiguity in ERS data; PRESCAT and 2DVAR performance is on par in the tropics, but 2DVAR performs better in the NH and SH.
- 2DVAR is also recommended for SeaWinds, but the variational quality procedure, VarQC, should then be used, as discussed in next chapter.

4. VARIATIONAL QUALITY CONTROL

The use of variational quality control (VarQC) is necessary to flag observations from the SeaWinds instrument affected by geophysical conditions other than wind, such as rain, and that are not rejected by Quality Control prior to AR.

The SeaWinds scatterometer is sensitive to rain and measurements affected by rain generally result in a stronger backscatter signal (*Portabella*, 2002). The selected wind vector solution from these measurements will exhibit noticeable differences with surrounding wind observations in terms of wind speed and/or wind direction. It may be said that the affected observations contain gross errors and VarQC is used to filter out these observations.

To incorporate VarQC in 2DVAR we follow *Anderson and Järvinen* (1998). For a given prior probability of gross error *Pg* for both wind components, the probability density function (pdf) for the components of the scatterometer wind ambiguities of a single observation is

$$P_{u,v}^{QC} = \left[1-\left(1-P_g\right)^2\right]\frac{\pi}{2d^2} + \left(1-P_g\right)^2 e^{-\frac{J_o^{scat}}{2}}$$

The first term on the RHS of this expression is the uniform probability distribution of gross error (white noise) with a finite range defined as a multiple *d* of the observation error standard deviation. It expresses the probability that the observation does not contain meteorological information on surface wind. The probability density function for VarQC implies that an observation is considered rejected if one or both wind components are affected by gross error. The observational term in the cost function follows from

$$J_o^{scat,QC} = -2\ln P_{u,v}^{QC}$$

Figure 4.1 gives a graphical depiction of a cross section of the observational term of the cost function and its gradient with VarQC. It shows the gradient going to zero very quickly away from the ambiguous solutions. In an iterative minimization process this behavior of the gradient has the effect of limiting contributions of inconsistent observations to the cost function gradient leaving the solution to be determined by surrounding consistent observations.

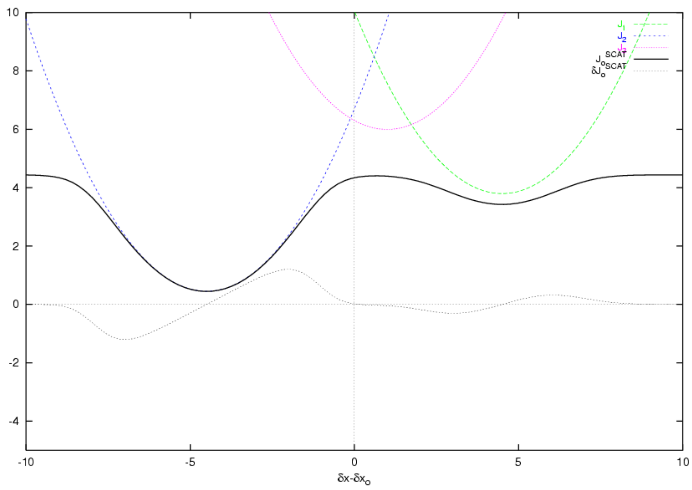

Figure 4.1: The SeaWinds cost function and gradient with VarQC. Notice the gradient

The observations however are not rejected permanently and may regain full importance at a later stage in the minimization if they become more consistent with the updated analysis.

4.1 Tuning of VarQC parameters

In order to use VarQC in 2D-VAR we need to specify parameters *Pg* and *d.* To obtain suitable settings for the prior probability of gross error and the range parameter we studied normalized departures of selected ambiguities from 2DVAR analysis through $\sqrt{J_o^{scat}}$.

The departures were obtained by running 2DVAR with retrieved ambiguous winds at 100km resolution as input and VarQC switched off. The SeaWinds data set that was used was from January 27th to February 3rd of 2002. The ambiguous winds were quality controlled at the inversion stage (*Portabella*, 2002), which can be considered as a form of consistency check.

Figure 4.2 shows the probability density distribution of $\sqrt{J_o^{scat}}$ for Wind Vector Cell (WVC) 3 to 17 out of 19 that cover the SeaWinds swath in the across track direction. We also fitted the mode of the distributions to a normal distribution, thereby correcting for outliers.

Evident from figure 4.2 is that the shape of the distributions varies from one WVC to the next and pairs of WVCs symmetrically situated around the nadir node (WVC 10) have distributions of similar shape. At nadir WVC 10 the peak in the distribution occurs at a higher cost function value, due mainly to the fact that WVC 10 has more cases with two ambiguities and where the probability of the selected solution is low compared to the other WVCs (not shown). All distributions exhibit a long tail compared to the normal distributions. For moderate values of $\sqrt{J_o^{scat}}$ (between 3.0 and 5.0) the actual distribution is larger than the normal distribution indicating a fitting problem, which may be related to non-optimality of the structure functions and not to gross error. As a result the mode in the actual distribution is somewhat under-sampled compared to the normal distribution, but generally appears about in the right location. From a theoretical piont of view a Rayleigh distribution will provide a better fit, but this is not tested here.

The variability in the distribution shape can be explained by the fact that for SeaWinds the scanning geometry is a function of across-track swath position. The scanning geometry is near optimal for the sweet swath (WVCs 4-7 and 13-16), resulting in a retrieval cost function that has deep well defined minima which can be accurately determined (small residuals). The scanning geometry in the nadir and outer part of the swath is less optimal, i.e. broad minima and the retrieved winds are less accurate.

We determined the values of *Pg* and *d* from figure 4.2 such that rejections occur for significantly higher values of $\widehat{f}$. We assume that data is rejected when the gradient is reduced to 0.25 of that of the cost function without active VarQC, i.e. with a *posterior probability of gross error* (*Pp*) of 0.75. The rejection limit can be expressed as (Anderson and Järvinen, 1998)

$$J_o^{scat} > \ln \frac{P_p}{\left(1-P_p\right)\gamma}$$

with

$$\gamma = \frac{\left\lfloor 1-\left(1-P_g\right)^2 \right\rfloor}{\left(1-P_g\right)^2} \frac{\pi}{2d^2}$$

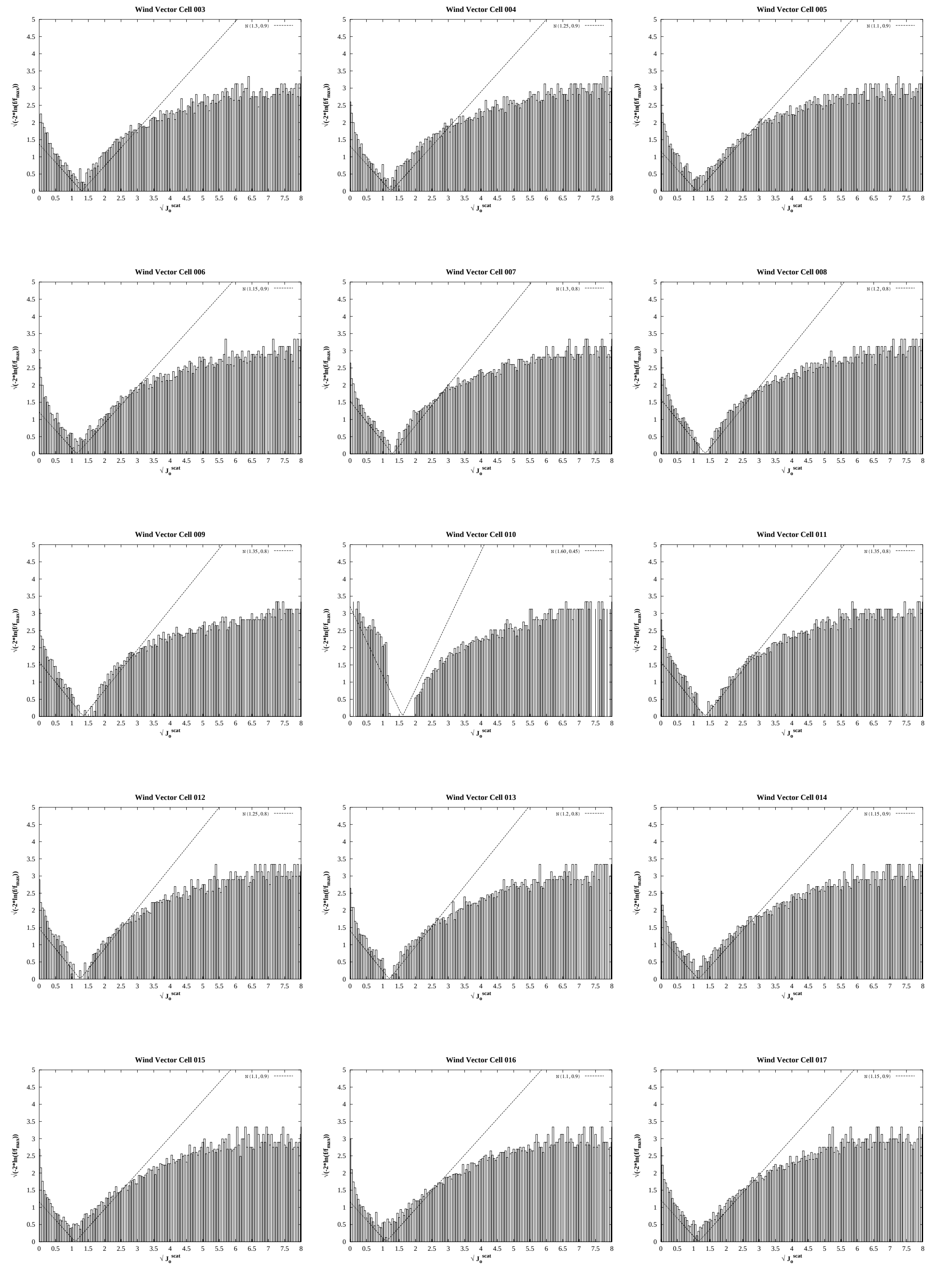

Figure 4.3: Transformed histogram of $\sqrt{J_o^{scat}}$. The straight lines represent the normal distribution fitted to the frequency distribition in figure 4.2

WVC	Mean (m/s)	SD (m/s)	Rejection limit	*d*	*Pg*
3	1.3	0.95	5.10	4	0.818E-5
4	1.25	0.95	5.05	4	0.818E-5
5	1.1	0.95	4.90	4	0.818E-5
6	1.15	0.95	4.95	4	0.818E-5
7	1.3	0.85	4.70	4	0.146E-3
8	1.25	0.85	4.65	4	0.146E-3
9	1.35	0.85	4.75	4	0.146E-3
10	1.6	0.45	4.00	4	0.199E+0
11	1.35	0.85	4.75	4	0.146E-3
12	1.25	0.85	4.65	4	0.146E-3
13	1.2	0.85	4.70	4	0.146E-3
14	1.15	0.95	4.95	4	0.818E-5
15	1.1	0.95	4.90	4	0.818E-5
16	1.1	0.95	4.90	4	0.818E-5
17	1.15	0.95	4.95	4	0.818E-5

Table 4.1: The mean and standard deviation of the normal distribution, the rejection limit, the range parameter, and the prior probability of gross error for all WVCs.

The mean and standard deviation of the normal distribution, the rejection limit, the range parameter, which was kept constant, and the prior probability of gross error are given in table 4.1 for all WVCs.

Having specified *Pg* and *d* the data set was reprocessed with VarQC switched on. With active VarQC the minimization process in 2DVAR first carries out 20 iterations or less without VarQC and another 10 iterations or less with VarQC. Figure 4.4 shows the rejection rate with the settings of table 4.1. The rejection rates are larger than expected based on gross error probability. This is due to the long tails of the distributions in figure 4.2. In fact, a Rayleigh distribution fit is more suitable from a theoretical point of view, but has not been tried here. The rejection limit of WVC 10 was widened to keep it within realistic bounds.

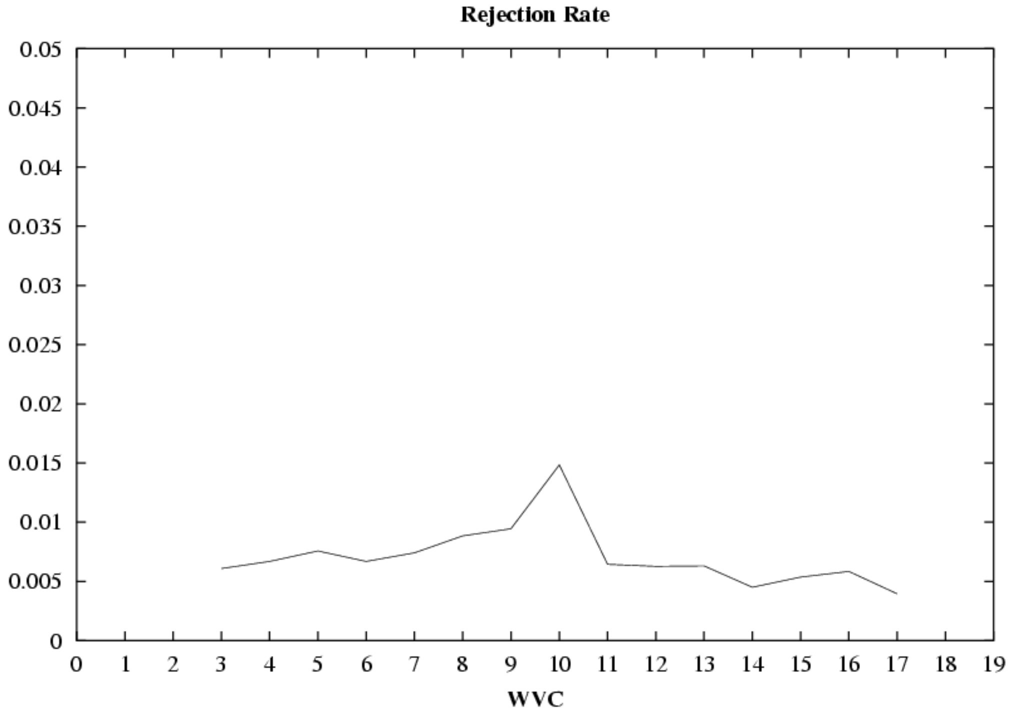

Figure 4.4: The realised rejection rate as function of the across track swath position.

4.1 Statistical comparison

To check if current settings for VarQC are actually taking out bad solutions, a statistical comparison was made in which both the non-rejected and rejected SeaWinds observations were verified with collocated ECMWF analysis winds. Figures 4.5 and 4.6 show the results of this comparison: the wind speed, direction and component statistics of the 100km SeaWinds selected winds are verified against collocated and independent ECMWF 10m winds. The data for these plots are from consecutive orbits of January 27th 2002 until February 3rd 2002. The wind direction statistics are based on winds with wind speeds larger than 4 m/s.

The joint distributions and accompanying statistics in all plots of both figures show that compared to ECMWF the rejected winds are biased higher than the non-rejected winds in terms of wind speed and still show signs of directional ambiguity. The distribution density in the components of the rejected winds is higher away from the diagonal contrary to those distributions for the selected winds. The asymmetric peak in the U-component distribution in figure 4.5 is due to the systematic sampling of the trade winds. The jaggedness of the wind speed distributions in figure 4.5 is due to the fact that the retrieved scatterometer wind speeds stem from the GMF lookup table. Overall it is evident that the distributions of the selected and the rejected winds are quite different and the non-rejected winds compare much better with ECMWF winds. This gives a strong indication that the rejected winds are indeed of low quality.

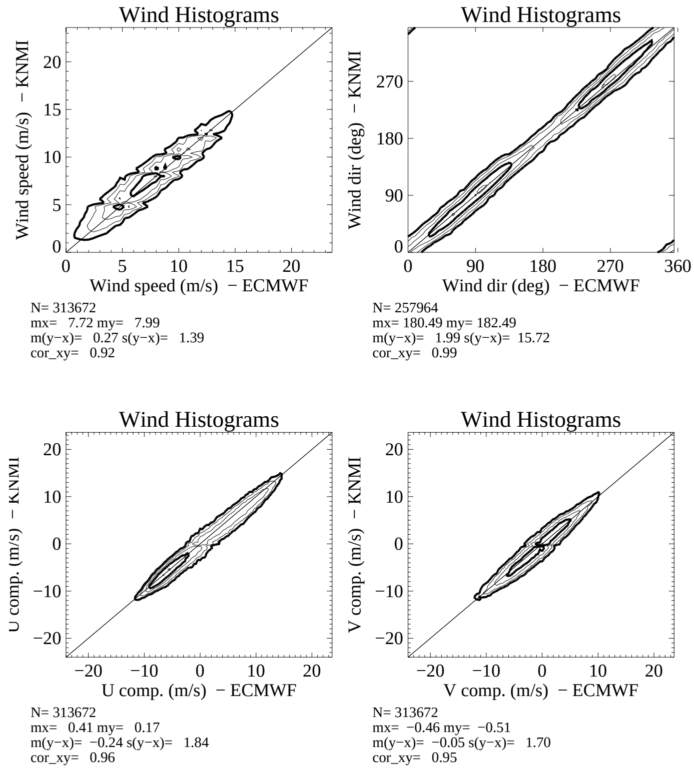

Figure 4.5: Contoured histograms of 100km resolution scatterometer wind observations with VarQC versus ECMWF 10m wind analysis. The value of the top-level contour is equal to the maximum value in the distribution scaled by a factor of 0.9. Every consecutive contour has a value equal to the previous one scaled by a factor of 2.

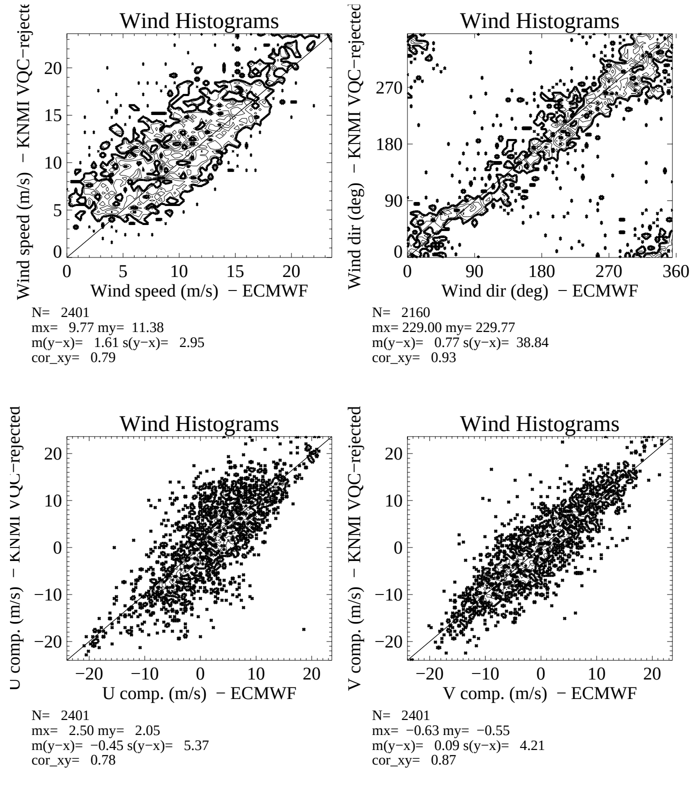

Figure 4.6: As figure 4.5 but for the SeaWinds wind observations rejected by VarQC.

4.2 Case studies

In addition to the statistical analysis a study was made on the behavior of 2DVAR with VarQC under different meteorological conditions. In figures 4.7a-d cases are depicted that give a good indication of the current skill of the variational quality control. In these figures selected SeaWinds wind vector solutions in red overlay a METEOSAT infrared image that correspond best with the observation time of the SeaWinds observations. The blue arrows in figure 4.7c represent a HIRLAM 3h forecast that corresponds closest to the observation time of the SeaWinds observations. The magenta dots are either land or ice contaminated nodes. SeaWinds wind vector solutions rejected by VarQC are colored yellow. Figure 4.7a (upper left) shows a case where a single wind vector is rejected for being very different from neighboring wind solutions in terms wind speed, probably due to sea ice. Figure 4.7b (upper left) is a low-pressure system in the SH where several wind vector solutions are rejected close to the center of the depression. They differ from surrounding wind vectors mainly in terms of wind direction. The rejected solutions are in a clouded area so they may be affected by rain. Figure 4.7c (lower left) is a very complex case. It shows a frontal area with a larger cloud band oriented from north to south at the 0˚ meridian. Large differences in wind speed and direction occur over short distances. The swath is only partly covered with wind observations. All along the front wind solutions are rejected, particularly solutions with high wind speed. At 69N,0E several rejected solutions differ very much from the background direction-wise, which may be due to rain. Rejections further to the north are more extensive. On the one hand this is unfortunate because many solutions along the front seem probable and may have been rejected unjustly. On the other hand, a number of solutions are clearly wrong and have been removed. Figure 4.7d (lower right) finally is case near the ice edge of the South Pole. Rejected solutions are inconsistent in wind direction. These may be due to backscatter measurements over areas containing sea ice.

4.3 Discussion and recommendations

Variational quality control is often a trade-off in individual cases that are complex. Studied cases suggest that the frequency of rejections occurring in complex situations is above average. In situations where the flow is well-organized VarQC is capable to weed out isolated bad solutions. Overall VarQC rejects about 0.75% of the data. The statistical comparison shows that rejected solutions verify less well with ECMWF analyses than non-rejected selected solutions.

We may conclude that VarQC is beneficial to the quality of the wind observations. Therefore it is recommended to use VarQC in the AR process for SeaWinds.

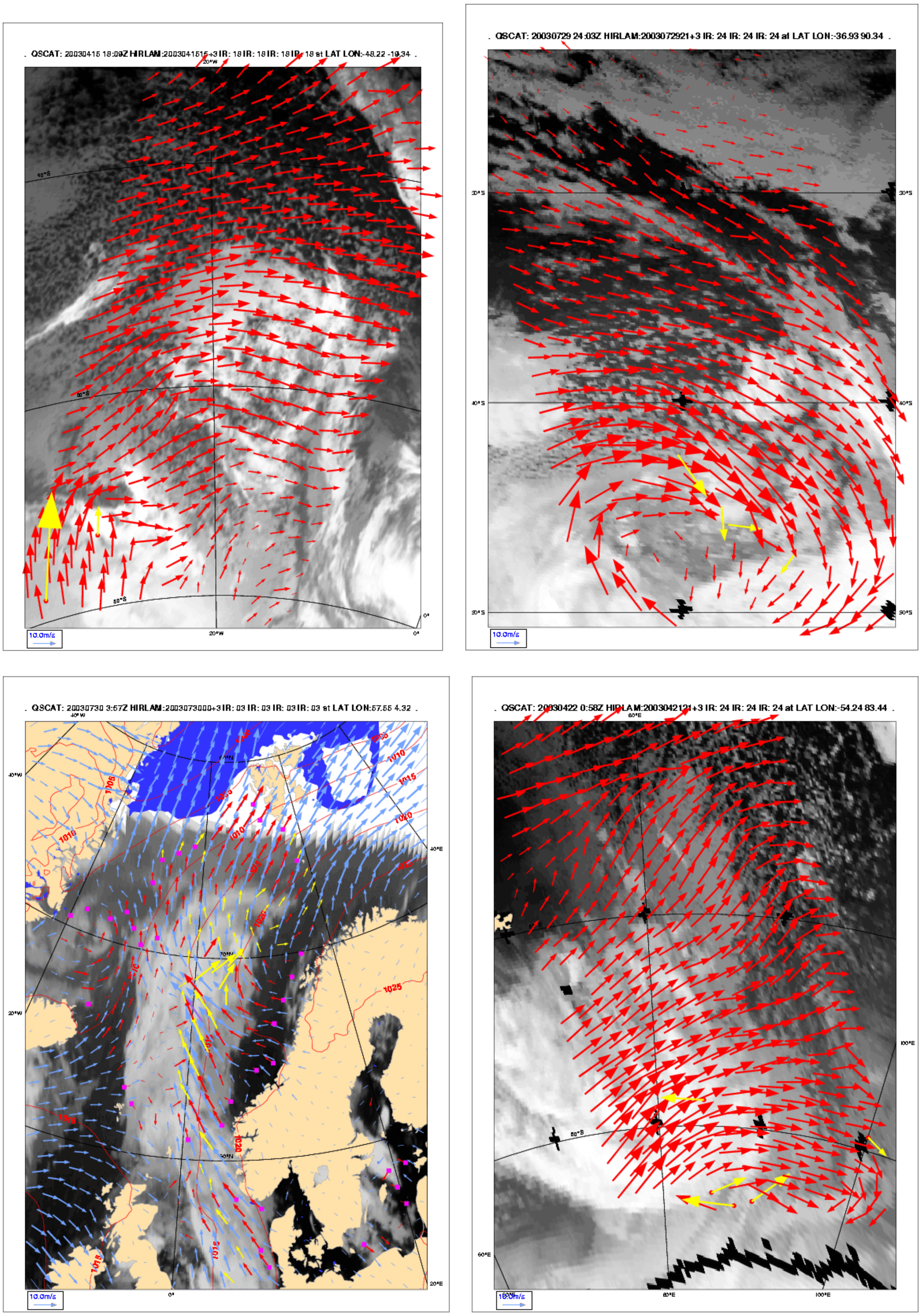

Figure 4.7a-d: Cases with active VarQC.

5. PRODUCT MONITORING FLAG

Meteorological offices more and more rely on the automatic ingestion of observations into their operational procedures for weather forecasting, in particular Numerical Weather Prediction, NWP. Therefore, near-real time quality assurance procedures are needed to prevent detrimental impacts on weather forecasts of occasional sets of anomalous observations. Most forecast systems employ procedures for the control of individual observations, but generally no automatic procedures exist for the control of sets of observations.

Sets of anomalous observations occur in case of failing instrumentation or platforms, but could also be due to artefacts in the processing or data transmission. Due to the nature of the underlying systems, for satellite data streams such failures likely occur on a large scale and, therefore, potentially lead to large-scale forecast failures. On the other hand, satellite data are usually received in large quantities and the statistical distributions of some particular quality indicators may be checked against the expected characteristics of these distributions. For example, one could check the percentage of rejected observations in a data file against an upper threshold in order to judge the complete data file to be suspect.

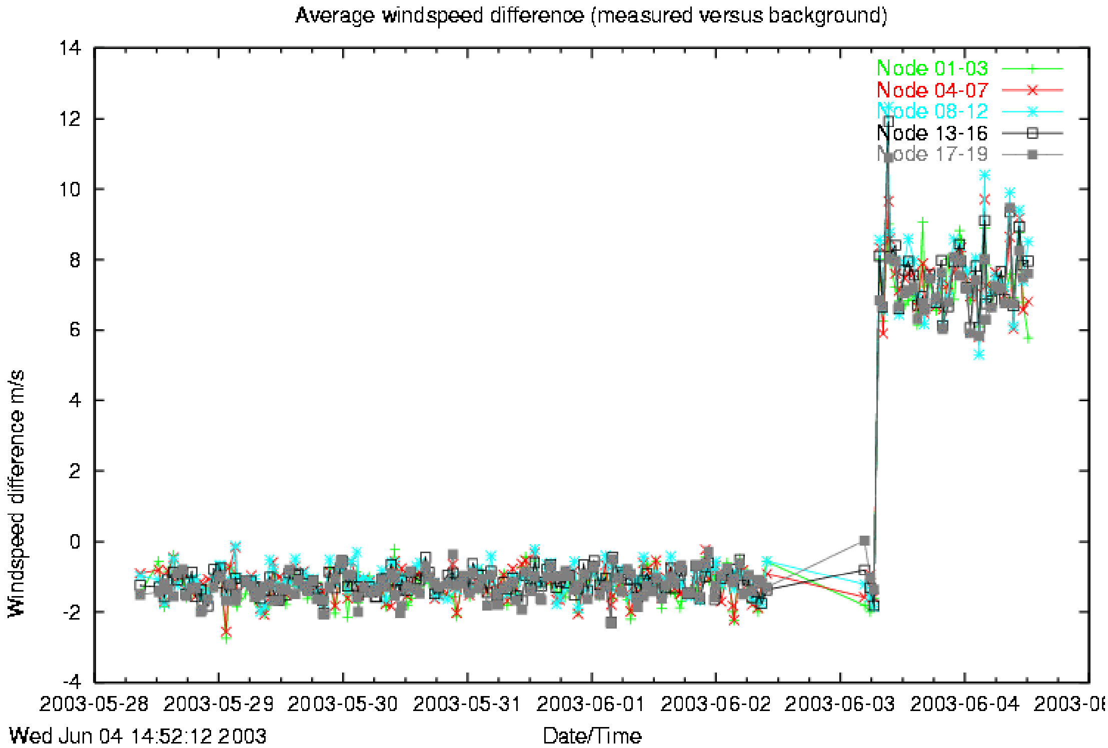

Figure 5.1: SeaWinds minus NCEP wind speed difference versus time for a series of QuikScat product files. After some missing data on the 3rd of June 2003, the NOAA BUFR product did not contain the NCEP background wind information, but zero winds instead, leading to a major increase in wind speed difference. The absence of this information results in badly performing ambiguity removal in the NOAA BUFR product.

Several examples of satellite data anomalies, without prior notification by the data provider, exist where the data have been used in NWP with, as a result, some detrimental impact. In figure 5.1 such an example is shown. In this case the NOAA BUFR product did not contain the NCEP background wind information, but zero winds instead. This information is as such not required by most users, but it turns out that in the absence of this information the ambiguity removal in the NOAA BUFR product is performing badly. The figure shows that this problem can be easily detected by monitoring SeaWinds minus NCEP wind speed differences.

If an observation set in a file can be judged as suspicious on statistical grounds by the user, then such cases may be prevented. On the one hand, assessment of data quality by the user is not desirable in principle, but on the other hand reasons may exist to implement this:

- The data provider has no resources to implement statistical quality checks;
- The user has access to superior reference data, e.g., a NWP first guess;

Here we describe the development of a method to assess the overall quality of files of SeaWinds data. The objective is that files of detrimental quality should be flagged and, at the same time, files of nominal quality not rejected, i.e., high probability of detection, POD, and low false alarm rate, FAR. The KNMI SeaWinds product quality control flag uses a set of predictors and statistical rules based on past data. The flag development and tests are based on our 100-km QuikScat wind product, but the methodology may be easily adopted for other resolutions, or even other instruments.

5.1 Product Quality Flag Definition

The product quality flag consists in selecting a set of predictors, analysing their 2002 statistical distributions and moments for defining and testing the statistical rules constituting the product quality flag. These are subsequently described here.

5.2 Predictors

Standard consistency checks of SeaWinds data are in place at KNMI checking for instance the noise and signal level for each backscatter value. If insufficient backscatter information is available in a particular WVC to allow proper wind retrieval, then the WVC is discarded. Next, each set of WVC backscatter values should be consistent with the wind GMF. This is measured by the residual after wind retrieval, MLE, which has been normalised and is checked against a threshold value for quality control, QC (Portabella). Systematic bias effects in the forward look of the inner beam, for example, will result in a more active QC and an increased MLE. However, a general drop in signal level will not immediately lead to a substantial increase in MLE or associated QC, but does directly affect the magnitude of the retrieved wind speed. These sensitivities lead to the following set of predictors:

1. Fraction of WVCs with valid and sufficient (4 beams) backscatter information for wind retrieval;
2. Ratio of number of WVCs with the QC flag raised (KNMI flag incl. rain) with respect to number of WVCs with sufficient backscatter information;
3. Mean normalised MLE of WVCs containing wind information;
4. Mean wind speed bias of WVCs containing both wind information and valid model wind information;

Note that in predictor 2. also WVCs with insufficient backscatter information are included in the numerator. This makes an important difference since many WVCs with the QC flag raised do not contain sufficient backscatter information (they are rejected by the quality control during processing). If we would omit these WVCs, the predictor becomes very insensitive.

The distribution of points on which the predictors are computed is quite relevant. The first test is done on all WVC, including those over land. Since the fraction of land can vary strongly between one product file and the next, this first predictor is rather coarse. The QC flag, second predictor, is only set for the WVC with valid and sufficient backscatter information. The MLE and wind speed bias are computed for all WVC that passed QC. In nominal conditions the predictors are determined by subsequently:

1. Land and ice distribution; incomplete backscatter data in WVCs at product start and ending;
2. Number of geophysical anomalies (sea state, rain, sea ice, ..);
3. Random noise level;
4. Signal level or systematic error of backscatter measurements.

The predictors are computed in five sets for different parts of the swath; these are WVC ranges 1-3, 4-7, 8-12, 13-16, and 17-19 of the KNMI 100-km product. The symmetry of the SeaWinds swath would nominally provide identical statistical distributions of the predictors in WVC groups 1-3 and 17-19, and 4-7 and 13-16. However, in anomalous conditions this may not be true and the segregation of these WVC ranges indeed provides sensitivity to certain types of problems, for example, loss of attitude control. As such in total 20 predictors are used.

5.3 Expectations

The nominal statistical distributions of the predictors are required in order to determine the non-nominal ranges of these, needed for developing the product quality flag. The KNMI QuikScat archive over year 2002 was used to determine the distributions, but with the following exclusions:

- Anomalous files on days 62-64, and 78, due to improper ice flagging during large-scale sea ice melting events in both the Southern and Northern Hemisphere respectively;
- For ease of computation, the product observation count range was set as depicted in figure 5.2, resulting in a loss of about one-third of data;

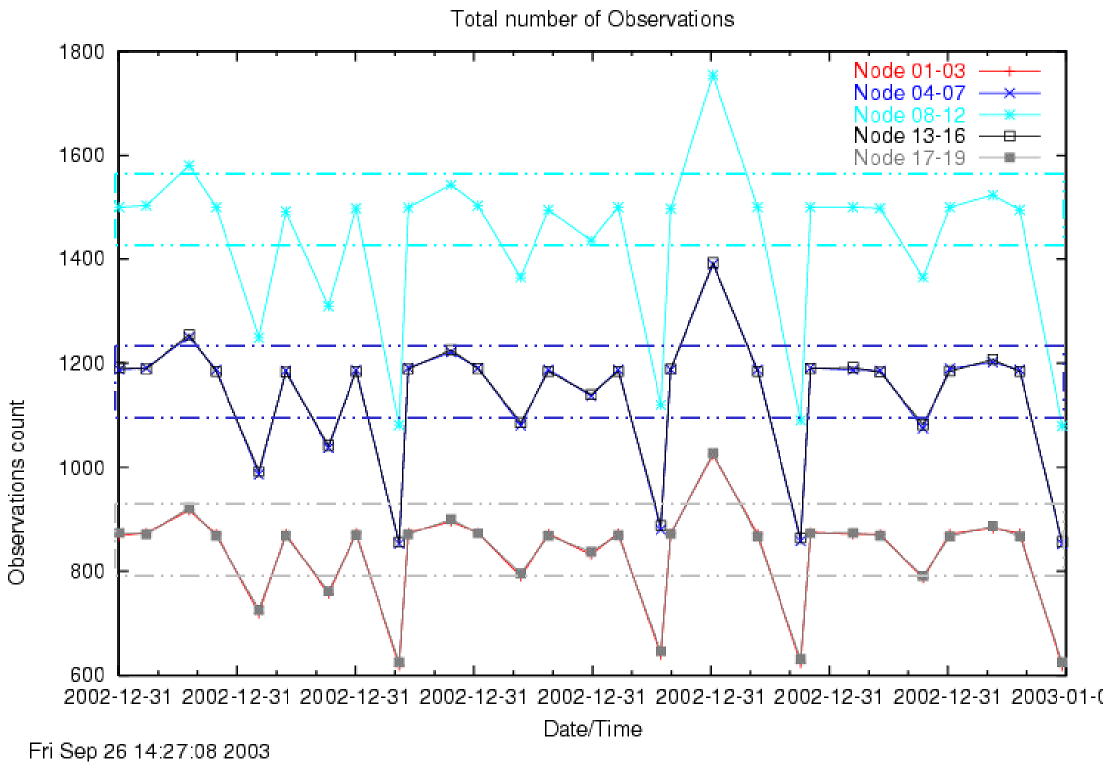

Figure 5.2: Total number of observations in a WVC group (see legend) versus time for a series of QuikScat products. The horizontal lines denote the observation count ranges that were used for computing the average statistics for each WVC group. Note that some lines do not appear since they are overlapping due to the swath symmetry.

Over 2002 many average and RMS statistics were computed for each product and WVC group:

- For each of the predictors;
- In addition, for the SeaWinds minus NCEP background wind, the wind vector and wind direction statistics; the lowest rank and selected wind solution rank statistics;
- For the VarQC flag occurrence;

These additional statistics were used, on the one hand, to check that anomalous conditions in the five predictors correlate with anomalous conditions in the other statistical quantities mentioned above, but, on the other hand, to check that anomalous conditions do not appear in the additional statistics without anomalies in the predictor set. The latter would indicate that the chosen predictor set is not complete and that other statistical quantities are more sensitive to certain problems. It is verified that the predictor set is adequate indeed.

Moreover, RMS and mean values were used to perform a so-called 3–sigma test to check the Gaussian properties of the predictor distributions. The test consists in assuming a Gaussian distribution with mean μ and standard deviation, SD, σ. If for such a distribution one discards all data x for which $|x - \mu|$ is larger than 3σ, then the expected SD of the remaining points only decreases by about one percent. Moreover, if one substituted σ for the expected SD in a second pass, then the expected SD does not change much, such that the procedure converges very fast. Distributions with long tails converge much slower in such a symmetric test. We indeed found that two iterations are sufficient to estimate a nominal mean and SD for the data in year 2002. These means and SDs were subsequently used to define the product QC tests.

5.4 5-σ predictor test

The mean number of observations, sum, and sum of squares of the predictors for each WVC group for 2002 are subsequently used to test the overall quality of the product files by the following equation:

$$\left|x - \langle x \rangle\right| \leq A + 5 \cdot SD(x) \cdot \left(\langle N \rangle / N\right)^{0.5} \quad (1)$$

where,

x Mean predictor value over *N*;

<*x*> 2002 mean value of *x* (for <*N*> WVC);

N Number of WVCs used to calculate the fraction, ratio or average in a file WVC group;

<*N*> 2002 mean value of *N*;

SD 2002 standard deviation of *x*;

A Uncertainty in ancillary data; *A* = 0 , but 0.5 m/s for wind speed difference with NCEP.

So for predictor 2, N is the number of observations with sufficient backscatter information; for predictor 3, N is the number of WVCs with a wind solution et cetera.

5.5 2002 Product Results

The predictor set is run over 2002, and files failing one of the predictor tests are noted, visually inspected, and the raised flag judged for appropriateness. The following anomalies occur (number of products is noted between brackets):

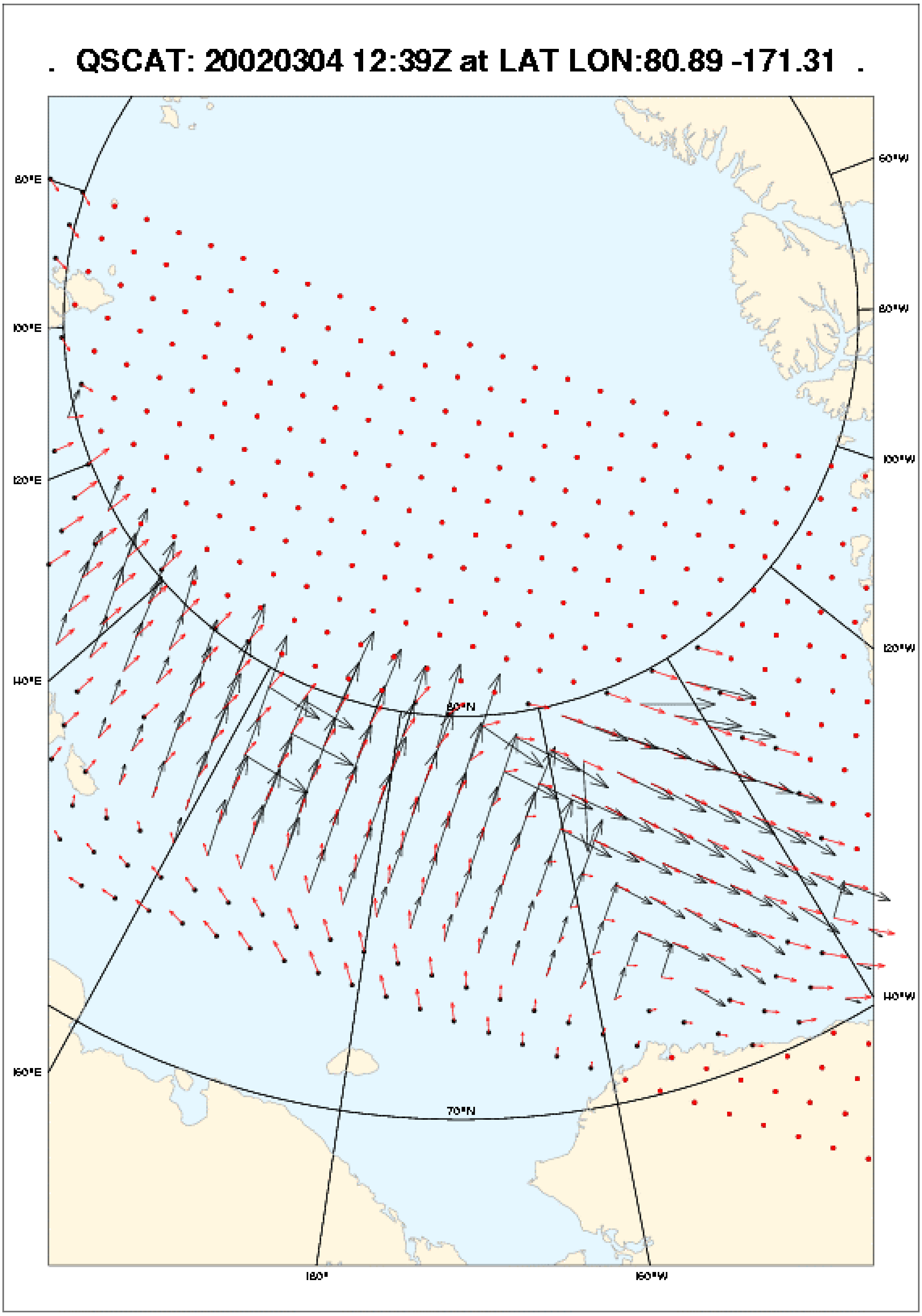

Figure 5.3: Along and across track winds (black) due to NH sea ice on 4 March 2002 with a maximum of about 30 m/s. The red arrows, at the same scale, show the NCEP winds for reference. Bad QuikScat wind data exist over a large area.

1. Days 62-64-: Wind retrieval over SH ice (22);
2. Day 78 : Instrument anomaly (1);
3. A few days around day 240 : Wind retrieval over NH ice (15);
4. Scattered products: No NCEP wind (8);
5. Scattered products : false alarms (20);
6. Day 290: No detection of missing NCEP wind (1).
7. Other “no detections”, e.g., small ice patches (1);
8. No data (= no problem!) (1).

Figure 5.3 shows an example of case 3., but in March, which is due to failure of the JPL ice flag (which is based on polarisation ratio). Besides large inconsistencies with the GMF, resulting in a large fraction of WVC with the QC flag raised and a high mean normalised MLE, retrieved wind directions are typically along or across track. This makes the problem visually tractable as in Figure 5.3. For case 1., often large-scale northerly winds bring warm air over the SH sea ice sheat, resulting probably in widespread water formation over the ice areas. However, due to the presence of the remaining ice, the surface characteristics are different from open sea, and wind retrieval is not possible. The same tests fail in category 3.

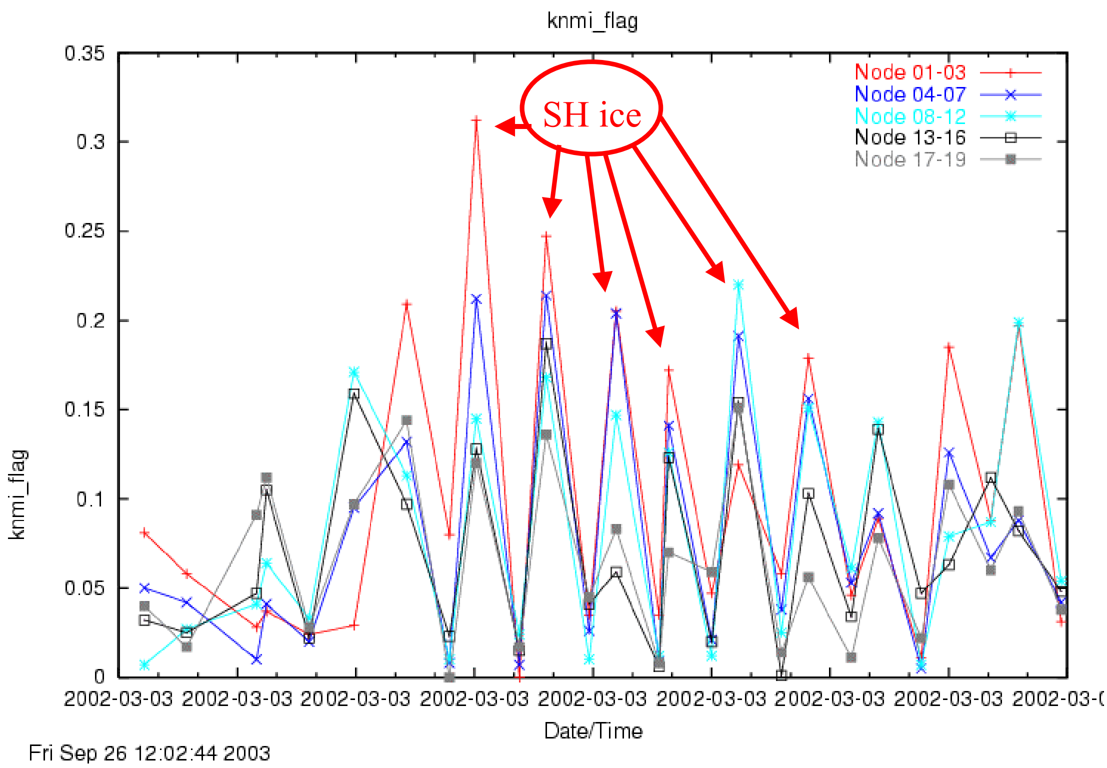

Figure 5.4: KNMI QC flag in a WVC group (see legend) versus time for a series of QuikScat products. The intermittent peaks occur since about one complete orbit is split into two products, only one of which includes the problematic sea ice points over Antarctica (as indicated by the peaks).

Case 2. really represents the kind of instrumental problem that the product flag should detect. Only one such anomaly occurred in 2002, and all predictors went out of range. At this time the data were assimilated at KNMI, ECMWF, and some other weather centres, and wrong winds were assimilated in some cases with detrimental analysis impact.

Case 4. was rather annoying, but is now being detected in the KNMI SeaWinds processor by a NCEP background wind test, instead of the wind speed bias predictor being used here. On day 290 (case 6.) this predictor did not result in an alarm, while the NCEP winds were missing indeed. The problem is that the ancillary NCEP winds are used in the ambiguity removal, DIRTH at NOAA and 2D-VAR at KNMI, and the NCEP zero winds cause wind direction errors in these procedures. Moreover, the wind solution in the NOAA DIRTH product is affected by DIRTH and thus dependent on NCEP winds. This means that a new wind inversion is needed to get the proper basic wind-solution information. On the other side, the KNMI product contains the complete inversion information, while only the AR fails.

Case 5. is most relevant for the optimisation of equation (1). From the summary of all cases we conclude

- The number of false alarms of 20 in 10,000 is quite acceptable; and
- The detection category of interest is related to non-screened ice patches in the NOAA product.

Concerning the latter, we note that the MLE (and related QC flag) test is quite effective in screening affected products (see Figure 5.4). This seems to indicate that the MLE may be exploited to screen sea ice. This would clean up the SeaWinds data set considerably, resulting in a smaller SD of the predictors then those obtained here. Using equation (1), it may be clear that the predictor tests then become stricter. It is difficult to say at this point whether the false alarms are influenced by some small ice problem here, and whether the false alarm rate would increase or not. This requires some further study once an improved ice screening is in place.

5.6 Predictor test results

Besides the product flag results it is of interest to check the individual predictor tests. Table 5.1 shows the numbers obtained over 2002.

Predictor	Number of occurrence
Wind retrieval fraction	0
KNMI QC flag	42
Mean MLE	26
Wind speed bias	15

Table 5.1: Number of occurrence of the predictor flags over 2002.

The wind retrieval fraction predictor is rather insensitive due to the varying fraction of land and ice points in a product, resulting in a large SD. In this case equation (1) results in a rather large range of acceptable wind retrieval fractions. Moreover, no anomalies appeared resulting in reported invalid σ^0

Naturally, the occurrence of a large KNMI QC flag fraction and a large mean MLE are correlated, since the former is based on a MLE threshold at each WVC. Both predictor tests result in a fail in about 50% of cases. These alarms are most frequent in the sea ice melting season.

A too large wind speed bias with respect to NCEP appears random, and is sometimes due to sea ice melting as well. However, most often it is due to missing NCEP winds coded as zero. As mentioned before, KNMI put in place a simple test for zero NCEP winds in a product and does not use equation (1) for this. Also, these products were avoided in the statistical predictor computations (mean and SD).

On day 78 all flags, but the wind retrieval count, were raised since both the product noise and signal properties were in error.

5.7 Conclusions

A near-real time quality assurance procedure is developed to aid the automatic ingestion of satellite data in weather offices (e.g., NWP). Data files of detrimental quality are successfully flagged (5 in 1000) and, at the same time, only few files of nominal quality are rejected (2 in 1000). Moreover, almost all problems appear detected (99.99%).

In the melting season, both Arctic and Antarctic, patches of melting sea ice are not indicated as such by the JPL ice flag. The QuikScat near-real time users, particularly those near the poles, appear not alerted by wrong NOAA winds over ice, since the data providers were not alerted. This could either be because these users use their own ice masking, or because the SeaWinds data are not so much of interest in the polar melting season. At this stage, this problem causes the product QC flag to be set, and thus make the entire product, ice or no ice, suspicious.

5.8 Recommendations

Improved ice screening procedures would improve the situation. Given the effectiveness of the MLE in indicating such sea ice contamination, the MLE may be used to improve the ice screening. A similar development to ERS, taking into account a wind MLE and a sea ice MLE may be appropriate. After such development and sea ice algorithm improvement, the mean and SD of the predictors need to be re-examined and the product QC flag rates be re-assessed.

Missing NCEP winds may be identified by JPL/NOAA and coded as such in the BUFR product (the code for missing data is unique in BUFR, and thus not zero).

A NRT product QC flag is feasible and recommended.

The product flag should automatically alert the provider, such that immediate action can be taken to cure the eventual anomalies in the data processing or instrument settings.

6. SUMMARY AND OUTLOOK

The specification of the background error covariance matrix *B* and the implementation of variational quality control have shown to give satisfactory results. The settings that have been found are suitable. An indication of the quality of the processing chain including 2D-VAR can be given if we compare the 100km SeaWinds wind product with the existing near-real time NOAA product. *De Vries* (2003) provides such a comparison for a thinned NOAA product at 100km. It shows that the comparison is favorable for the KNMI 100km SeaWinds product.

A product monitoring flag for SeaWinds is described. Sea ice occurrence in the NOAA BUFR product causes some product rejections in a few days a year. A better sea ice model seems feasible that can cure these product rejections.

Future developments are focused on the use of 2DVAR in combination with the multiple solution scheme, MSS, (*Portabella*, 2002) to further enhance the quality of the 100km SeaWinds wind product and to provide a scope for the application of the meteorologically balanced spatial filter in 2D-VAR and the maximum exploitation of backscatter information after wind retrieval, in order to process SeaWinds observations at higher resolution.

7. REFERENCES

Anderson, E., Järvinen, H., 1999, "Variational Quality Control", *Q.J.R. Meteorological Soc*., 125, 697-722.

Gilbert J. Ch., Lemarechal C., 1989, Some numerical experiments with variable storage quasi-Newton algorithms, Mathematical Programming, 45, 407-435.

Hollingsworth, A., 1989: The role of real-time four-dimensional data assimilation in the quality control, interpretation and synthesis of climate data. Oceanic Circulation Models: Combining data and dynamics. Ed. D.L.T. Anderson and J. Willebrand. Kluwer Academic Publishers.

Leidner, M.S., Hoffmann, R.N., Augenbaum, J., 2000, SeaWinds Scatterometer Real-Time BUFR Geophysical Product, *User Guide Version 2.3.0*, NOAA/NESDIS.

Lorenc, Andrew C., 1986, Analysis Methods for Numerical Weather Prediction, Q.J.R. Meteorological Society, 112, 1177-1194.

Portabella, Marcos, 2002, Wind field retrieval from satellite radar systems, PhD thesis, Un. of Barcelona, Spain.

Spencer, M.W., Wu, C., and Long, D.G., "Trade-offs in the Design of a Spaceborn Scanning Pencil Beam Scatterometer: Application to SeaWinds", IEE Trans. Geosci. Remote Sensing, Vol. 3, No 1, pp. 115-126, January 1997.

Stiles, Bryan, Brian D. Pollard, R. Scott Dunbar, 2000, Direction Interval Retrieval with Threshold Nudging: A Method for Improving the Accuracy of QuikSCAT Winds, JPL MS 300-319, Pasadena, CA.

Stoffelen, A., "Scatterometry"', PhD thesis ISBN 90-393-1708-9, October 1998.

Stoffelen, Ad, Aart Voorrips and John de Vries, 2001, Towards the Real-Time Use of QuikSCAT Winds, BCRS project report, USP-2 00-26.

De Vries, J., A. Stoffelen, Variational Ambiguity Removal, 2003, In preparation.

De Vries, J., QuikSCAT Data Processor – User Manual, Version 1.1, NWPSAF-KN-UD001

8. ACRONYMS

2D-VAR	2-dimensional Variational AR
AR	Ambiguity Removal
BUFR	Binary Universal Format Representation
DIRTH	Direction Interval Threshold Nudging
ECMWF	European Centre for Medium-range Weather Forecasts
ERS	European Remote Sensing Satellite
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
GMF	Geophysical Model Function
HH	Horizontal polarisation emitted-Horizontal received
HIRLAM	High-Resolution Limited-Area Model
JPL	Jet Propulsion Laboratory
KNMI	Royal Netherlands Meteorological Institute
MLE	Maximum Likelihood Estimator
NCEP	National Centre for Atmospheric Prediction (USA)
NOAA	National Oceanographic and Atmospheric Administration (USA)
NRT	Near-Real Time
NWP	Numerical Weather Prediction
PRESCAT	Processor of ERS scatterometer data at KNMI
QC	Quality Control
QuikSCAT	NASA scatterometer mission with SeaWinds
RMS	Root-Mean-Squared
SAF	EUMETSAT Satellite Application Facility
SeaWinds	NASA rotating pencil-beam scatterometer
SWVC	Super WVC
VV	Vertical polarisation emitted-Vertical received
WVC	Wind Vector Cell

APPENDIX A: TUNING OF INVERSION AND AR PARAMETERS

In order to optimise the overall performance of the QDP software package, tests were done with different parameter settings for the Ambiguity Removal. The grid spacing for the 2DVAR grid (parameter delta in the software, see Chapter 2) and the length scales of the structure functions (see Chapter 3) were varied.
Moreover, inversion using non-smoothing and 3D interpolation was compared to inversion using smoothing and 1D interpolation (see appendix C in *Portabella*, 2002). Three versions of QDP were compared, all based on version qdp1_0q, which is present in the CVS version control system at KNMI. See table A.1 for the different settings of each version. Other settings were kept constant, e.g., observation weights.

Setting	1		2		3	
	L_Ψ / L_χ	ν^2	L_Ψ / L_χ	ν^2	L_Ψ / L_χ	ν^2
NH	300km	0.1	300km	0.1	300km	0.1
Trop.	500km	0.5	500km	0.5	600km	0.5
SH	300km	0.1	300km	0.1	300km	0.1
delta	220km		220km		100km	
interpolation	1D		3D		3D	

Table A.1: Settings for the parameters in 2DVAR and the interpolation mode.

Setting 1 corresponds to the mode of operation of the current (October 2004) operational QDP version (qdp1_0n). With setting 2, the influence of the inversion mode can be investigated. *Portabella (2002)* showed that 3D interpolation yields better values for the normalised RMS. Setting 3 combines 3D interpolation with the recommended AR settings of Chapter 3 in this report (which are derived using ERS data and not QuikSCAT data!).

Four days of QuikSCAT data, from 3/4 October 2000 (2000-277, 2000-278) and from 7/8 April 2001 (2001-097, 2001-098), were taken from the mass storage device at KNMI. In this data, the NCEP model winds are replaced by ECMWF forecasts of 6-15hrs from the so-called ERA-40 data set. Moreover, the original NOAA input data files are rearranged such that one data file contains all data from one orbit number. In this way, QDP processes each Wind Vector Cell only once and overlap in the original data files is eliminated. This data can be found in the mass storage device in the directories /fa/ao/sat/quikscat/seawinds/YYYY_ecmwf.

After reprocessing the data with QDP, the RMS difference in wind direction between the retrieved QuikSCAT solutions and the ECMWF winds were calculated. Also the normalised RMS values were calculated. These calculations were done using the extractmisc program which is available in version qdp1_0q of the QDP software package. Table A.2 shows the results.

Setting	1	2	3
RMS in direction (°) (all WVCs)	31.53	32.12	33.40
NRMS (all WVCs)	1.5513	0.8907	0.9451
RMS in direction (°) (178031 WVCs)	31.51	32.10	33.38
NRMS (178031 WVCs)	1.5510	0.8900	0.9444

RMS in direction (°) (model speed >= 4m/s)	21.69	22.08	23.34
NRMS (model speed >= 4m/s)	1.0893	0.6113	0.6785

Table A.2: RMS and NRMS results.

First, the RMS and NRMS were calculated for all WVCs where a wind solution could be obtained. This number of WVCs varies somewhat dependent on the setting. To make the results fully comparable, the (N)RMS values were also calculated for a common set of 178031 WVCs that was present in the results of all three settings. The maximum number of valid WVCs was never more than 50 above the number in this common set. It is clear from the table that the values do not change significantly if the common set is used compared to all available WVCs. Finally, the (N)RMS values were calculated only for those WVCs where the model wind speed is 4m/s or higher.

Comparing the RMS values, setting 1 yields the best performance in all cases, although the differences are quite small. The differences between the three settings are clearer if we look at the normalised RMS. Now setting 2 has the best performance in all cases.

Although setting 2 proves to be the best for QDP in this test, there may be other settings yielding even better performance. This was not investigated. Moreover, in the future the SeaWinds processing will be performed using the new SDP software which is developed now. It will be necessary to repeat these tests in order to find the optimal settings for SDP.