2002-12-07

An assessment of SeaWinds on QuikSCAT wind retrieval

David Draper
Brigham Young University - Provo

David G. Long
Brigham Young University - Provo, long@ee.byu.edu

Original Publication Citation
Draper, D. W., and D. G. Long, An assessment of SeaWinds on QuikSCAT wind retrieval, J. Geophys. Res., 107(C12), 3212, doi:10.1029/2002JC001330, 2002.

An assessment of SeaWinds on QuikSCAT wind retrieval

David W. Draper and David G. Long
Department of Electrical and Computer Engineering, Brigham Young, University, Provo, Utah, USA

Received 30 January 2002; revised 30 May 2002; accepted 3 June 2002; published 7 December 2002.

[1] The scatterometer ocean wind retrieval process produces several possible solutions or ambiguities at each point, requiring a separate ambiguity selection step to infer a unique wind vector field. An ambiguity selection error occurs when the selected wind vector is not the closest ambiguity to the true wind. The current ambiguity selection routine for SeaWinds is ad hoc, but performs well under most circumstances. Factors such as instrument noise and rain can also cause the estimated wind flow to deviate from the true wind. A quality assurance (QA) analysis is performed to assess the ambiguity selection effectiveness and noise level of the retrieved wind using a low-order wind field model. The wind field model is data-driven and shown to be rather insensitive to the training data set. The QA analysis demonstrates that the SeaWinds ambiguity selection process is at least 95% effective. Ambiguity selection errors are correlated with storms and rain corruption. A subjective analysis on a set of cyclonic storm passes confirms that the wind retrieval is somewhat less effective in storm regions. *INDEX TERMS:* 4275 Oceanography: General: Remote sensing and electromagnetic processes (0689); 4504 Oceanography: Physical: Air/sea interactions (0312); 4594 Oceanography: Physical: Instruments and techniques; *KEYWORDS:* scatterometry, QuikSCAT, wind retrieval, ambiguity selection, dealiasing

Citation: Draper, D. W., and D. G. Long, An assessment of SeaWinds on QuikSCAT wind retrieval, *J. Geophys. Res.*, *107*(C12), 3212, doi:10.1029/2002JC001330, 2002.

1. Introduction

[2] Spaceborne scatterometers afford a valuable means of studying the oceans at high accuracy without the limitations associated with surface-based measurement schemes. SeaWinds on QuikSCAT, the latest spaceborne scatterometer employed by NASA, observes global marine winds in all weather conditions and independent of time of day. The SeaWinds wide-swath design affords near-full coverage of our planet's oceans daily at 25 km resolution [*Spencer et al.*, 1997]. SeaWinds, launched in mid-1999, replaces the NASA Scatterometer (NSCAT) which failed in 1997 after less than 9 months of operation. A second SeaWinds instrument is scheduled for launch in November 2002 aboard ADEOS II.

[3] The scatterometer wind retrieval process has two steps: First, a set of potential solutions or ambiguities are generated from the backscatter (σ°) values by a maximum likelihood estimation (MLE) technique at each wind vector cell (wvc) [*Chi and Li*, 1988; *Long and Mendel*, 1991]. Then, an ad hoc ambiguity selection routine is used to choose a unique wind vector field from the set of ambiguities [*Shaffer et al.*, 1991].

[4] There are two major types of error in scatterometer winds: First, the variability of the ambiguity closest to the true wind is influenced by factors such as instrument noise, measurement geometry, rain contamination, and accuracy of the geophysical model function (GMF) relating σ° to the wind speed and direction [*Freilich and Dunbar*, 1999]. These factors generally cause the selected wind flow to appear "noisy". For a discussion on GMF inaccuracies and the limitations of the comparison data, see *Brown* [2000].

[5] Second, data quality is affected by the accuracy of the ambiguity selection process. An ambiguity selection error occurs when a selected wind vector is not the closest ambiguity to the true wind. Due to the nature of the ambiguity selection process, ambiguity selection errors generally occur in clusters and result in 90° or 180° shifts in the selected wind flow. Because both noise and ambiguity selection errors can cause the retrieved winds to have inconsistent flow, the noise level and ambiguity selection performance can be assessed by evaluating the general consistency of the ambiguity-selected wind [*Gonzales and Long*, 1999].

[6] Traditionally, quality assessment is accomplished by comparing ambiguity-selected winds to numerical weather prediction (NWP) fields or buoy measurements. These methods allow a validation check of the data, but are sensitive to the accuracy of the comparison data and interpolation methods. Validation of NSCAT data against both buoy measurements and NWP fields has yielded good results [*Freilich and Dunbar*, 1999; *Vershell et al.*, 1999]. For our quality assessment, we adapt and extend a self-contained method developed by *Gonzales and Long* [1999] for NSCAT. The NSCAT quality assurance (QA) assessment method compares ambiguity-selected winds to a low-order Karhunen-Loève (KL) wind model fit. Regions exceeding error thresholds are identified and tallied.

0148-0227/02/2002JC001330$09.00

[7] Our implementation of this model-based assessment method uses an empirically determined set of error thresholds and other criteria to locate possible ambiguity selection errors in the presence of noise and estimates the overall self-consistency of the wind. This technique does not evaluate the absolute correctness of the scatterometer data. Rather, it is a method of identifying inconsistent wind flow, which suggests a wind retrieval ambiguity selection error. Natural phenomena such a fronts or fine-scale storms may contain apparently inconsistent wind flows and thus be erroneously identified by the method as a possible ambiguity selection error. An additional analysis of storm cases is provided to better quantify the wind retrieval performance in storm areas. Further, it is possible that an ambiguity selection error can result in a seemingly consistent flow, which is not correct. We assume that such errors are infrequent. Although this model-based approach has limitations, it works well for most wind data has the distinct advantage of being self-contained.

[8] In this paper, the QA analysis method is described along with rationale for the selection of parameters. Section 2 describes the QuikSCAT instrument and data set. The KL wind model size and truncation point are discussed with an analysis of the KL model's data sensitivity in section 3. The QA analysis method is described in detail in section 4. The QA analysis is applied to 2 years of SeaWinds data and the results are presented as a function of cross-track position, RMS wind speed, time, and latitude band in section 5. Higher noise is shown to occur in low wind speed regions and at nadir. Possible ambiguity selection errors tend to be highly correlated with rain and storm occurrences. The ambiguity selection is determined to be at least 95% effective for wind speeds above 3.5 m/s. An additional subjective analysis for cyclonic storm areas is presented in section 6, indicating that scatterometer wind estimation is less effective in cyclonic storm regions.

2. Data

[9] SeaWinds is a Ku-band wide-swath scanning pencil-beam scatterometer. The design employs two antenna beams: an outer (v-pol) beam, and an inner (h-pol) beam, giving an effective swath width of approximately 1800 km [*Spencer et al.*, 1997]. The rotating dish antenna traces a helical pattern on the surface and gives dense sampling of the ocean at constant incidence and varying azimuth angles. The standard QuikSCAT wind data set (Level 2B) is processed on a resolution grid of 25 $\times$ 25 km wvcs, with a swath size of 76 $\times$ 1624 wvcs. Cells not containing valid ocean wind data (wvcs over land or ice) are not processed.

[10] SeaWinds infers near-surface ocean winds by measuring the normalized radar backscatter cross-section (σ°) of the ocean surface at incidence angles of 46° for the inner beam and 54° for the outer beam. The radar signal scatters mainly from wind-induced capillary waves on the order of the electromagnetic wavelength due to the Bragg effect. The returned power is sensitive to the size and orientation of the waves, making σ° a function of the vector wind [*Moore and Fung*, 1979].

[11] Wind speed and direction are related to σ° through the GMF. Because the inverse mapping of σ° to the wind is not unique, several scatterometer observations at different azimuth angles are required to retrieve the wind. SeaWinds achieves this azimuthal diversity with fore and aft observations at each wvc. The azimuth angle difference between fore and aft measurements varies with cross-track position. There is little azimuthal diversity at the far swath, and a near-180° difference at nadir (center swath). Due to symmetry in the GMF, the MLE process results in several (1 to 4 with SeaWinds) possible wind vector choices or ambiguities. The ambiguities are ordered according to likelihood. Since the first (most likely) ambiguity is not always the closest ambiguity to the true wind, a separate ambiguity selection step is required to obtain a unique wind vector field.

[12] Ambiguity selection routines are generally based on ad hoc considerations and are prone to error. The ambiguity selection algorithm currently used by JPL in processing SeaWinds data has two parts: nudging and median filtering. In traditional nudging, each wvc is initially set to the ambiguity that most closely matches NWP fields produced by the National Centers for Environmental Prediction (NCEP). JPL implements a variant of traditional nudging known as thresholded nudging [*Stiles et al.*, 2002] with SeaWinds. In thresholded nudging, the estimated instrument skill (probability of a correct first ambiguity) at a wvc dictates the set of ambiguities used in the initialization. Where the estimated instrument skill is high, only ambiguities with high likelihood values are selected by nudging. Where the instrument skill is low, all ambiguities may be involved in the selection.

[13] After selecting an initial ambiguity field by nudging, a modified point-wise median filter is applied [*Shaffer et al.*, 1991]. For each wvc, the point-wise median filter selects the ambiguity that minimizes the directional error between the ambiguity and the surrounding 7 $\times$ 7 selected ambiguities. The point-wise median filter is iterated until convergence criteria are met.

[14] Unlike previous fan-beam instruments, SeaWinds retrieves wind in the nadir region, increasing the total swath coverage. Due to SeaWinds' constant incidence angles, observations of the nadir region are at oblique incidence. (For a further description of SeaWinds geometries, see *Spencer et al.* [1997].) A near-180° azimuth difference between fore and aft measurements at nadir causes the MLE technique to be ill-conditioned, giving rise to a noisy nadir wind field [*Oliphant and Long*, 1999]. Also, low wind speed data is generally noisy due to a low signal-to-noise ratio (SNR) and possible effects of a low-wind speed cut-off in σ° [*Donelan and Pierson*, 1987]. Model function inaccuracies also exist at high wind speeds due to high-wind saturation in relating winds to backscatter [*Donelan and Pierson*, 1987; *Donnelly et al.*, 1999; *Zeng and Brown*, 1998]. Thus, the performance of the SeaWinds wind estimation is variable with both cross-track position and wind speed [*Oliphant and Long*, 1999].

3. Analysis of the KL Wind Model

[15] Our QA analysis method is based on comparing QuikSCAT retrieved winds to a low-order KL wind model fit and identifying regions where the model fit significantly deviates from the retrieved winds. The KL model fit effectively smooths the wind data. Other smoothing methods (truncated Fourier spectrum, wavelets, averaging) may be used; however, the KL model optimally compresses the

low-order wind information into a few basis vectors that resemble true phenomenological features. These basis vectors span most of the low-order wind flow, and give a very good estimate of the underlying wind field.

[16] In general, the KL transform (also known as the method of principle components and commonly used in image processing) maps a signal (an image or wind field) on to a set of orthogonal basis vectors [*Jain*, 1989] formed from the eigenvectors of an autocorrelation matrix. In wind applications, the autocorrelation matrix is empirically generated from a training set of QuikSCAT retrieved winds. Each "square" N × N wind sample used to form the autocorrelation matrix is extracted from a swath of QuikSCAT winds and reshaped into a $2N^2$ vector by column scanning and then stacking the cross-track (u) and along-track (v) orthogonal components of the wind field. The autocorrelation matrix ($\hat{R}$) is constructed by combining the information from each wind field vector ($\mathbf{w_n}$) by

$$\hat{R} = \frac{1}{M}\sum_{n=1}^{M} \mathbf{w}_n \mathbf{w}_n^T \tag{1}$$

where M is the number of N × N wind fields in the training data set. The eigenvalues, or principle components, of the autocorrelation matrix indicate the amount of energy in the training set that is in the direction of the corresponding eigenvectors. The eigenvectors are ordered by decreasing eigenvalue, with the first eigenvector having the largest energy. The KL approach optimally compresses the greater part of the signal variability into the lowest order KL basis vectors and minimizes the basis restriction error [*Jain*, 1989].

[17] The KL wind model is formed from the lowest order KL basis vectors, i.e., a truncated KL basis. An N × N wind field can be written as a linear combination, using a least squares fit, of the KL model basis vectors. Because the wind has a generally red spectrum [*Freilich and Chelton*, 1986], the KL model basis vectors represent the low-frequency component of the wind. Thus, the low-order model fit retains the general flow of the wind and rejects high frequency content [*Moon and Stirling*, 2000]. Where the difference between the model fit and the ambiguity-selected (observed) wind is large, the truncated basis is not sufficient to characterize the selected wind flow. Large errors between the model fit and the observed wind may be due to ambiguity selection errors, noise, or fine-scale wind features outside of the space spanned by the truncated basis, but are most commonly associated with noise and ambiguity selection errors.

3.1. KL Model Size and Truncation Point

[18] The spatial frequency resolution of the KL model is determined by the model size (N) and the truncation point (number of basis vectors retained). A desired frequency resolution can be attained with most model sizes as long as the number of basis vectors is chosen appropriately. The size of the KL model used in the QA analysis is determined by a tradeoff between the ability to pinpoint error regions and minimize modeling error. The use of a smaller model affords better localization of error regions and is less computationally expensive. A larger model requires more basis vectors to achieve the same resolution, but can more accurately represent the overall flow of the wind and is thus less prone to modeling error. Test versions of the QA analysis method were evaluated for different model sizes. An 8 × 8 KL model was subjectively determined to be the best compromise between error region localization and modeling error minimization. The QA analysis, however, is not particularly sensitive to the model size as long as the truncation point is chosen such that the spatial frequency resolution is similar.

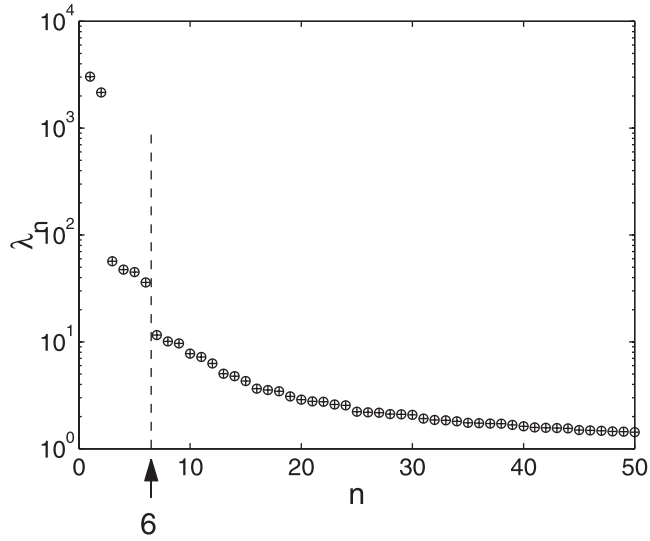

Figure 1. The first 50 eigenvalues of the 8 × 8 KL model.

[19] The general approach in choosing the KL model truncation point is to remove higher-order eigenvectors containing mostly noise. Examining the eigenvalues of the wind autocorrelation matrix for an 8 × 8 region size (see Figure 1), there are noticeable discontinuities in eigenvalues after the second and sixth eigenvalues. Because the first six eigenvectors characterize over 95% of the wind variation, six is a reasonable truncation point for a simple, low-order model.

[20] The discontinuities in the eigenvalues after the second and sixth value can be explained by examining the spatial frequency of the eigenvectors via the 2-D vector cosine transform. Figure 2 displays the first six basis vectors and the magnitude of their corresponding vector cosine transforms. The first two basis vector wind fields are the mean wind, and consist solely of the constant (1,1) cosine basis term (upper left corner). The next four basis vectors represent various common wind features and are linear combinations of the (1,2) and (2,1) cosine basis terms. The succeeding six KL basis fields consist of the (1,3), (2,2), and (3,1) cosine bases (Figure 3). This analysis suggests that the eigenvalue discontinuities correspond to edges of discrete frequency groupings. With six basis vectors, the model can represent a wide range of wind features, including cyclonic storms, col points, and divergent fields.

[21] The effective spatial resolution achieved by the QuikSCAT KL model is similar to that of the NSCAT KL model. *Gonzales and Long* [1999] used a larger (12 × 12) region size with an arbitrary truncation point of 22 parameters in analyzing NSCAT data. Due to QuikSCAT's finer

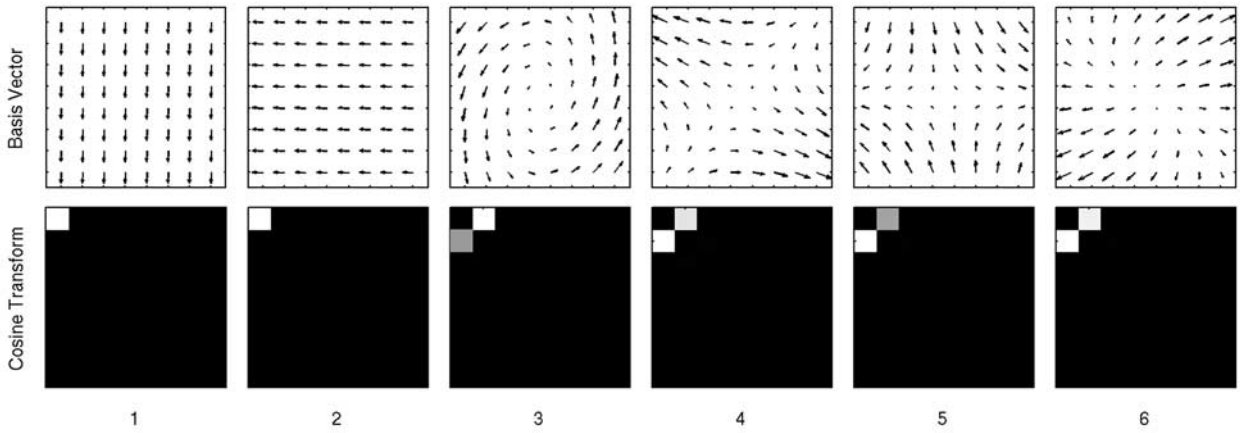

Figure 2. The wind fields corresponding to the first six KL basis vectors (top) and the magnitudes of their corresponding cosine transforms (bottom). The vertical dimension of the cosine transform corresponds to increasing frequency in the vertical direction. Likewise, the horizontal dimension of the cosine transform corresponds to increasing frequency in the horizontal direction.

grid, the actual spatial resolution achieved by the 6-parameter QuikSCAT model is on the order of the resolution achieved by the larger NSCAT model.

3.2. KL Basis Data Sensitivity Analysis

[22] Because the KL model is derived from SeaWinds ambiguity-selected wind fields, it is important to evaluate the sensitivity of the model to ambiguity selection errors in the training data. To do this, we employ Monte Carlo simulation to the ambiguity-selected winds to form several KL models whose training data sets are "corrupted" with artificially induced ambiguity selection errors. A cluster of artificial ambiguity selection errors is introduced into SeaWinds data by adding an arbitrary angle between 60° and 300° to a randomly shaped patch of ambiguity-selected winds and choosing the nearest ambiguities. The ambiguity selection errors are added until a certain percentage of wind vector cells have been corrupted. A KL model is then generated from the corrupted data. We then evaluate the vector space spanned in common between the corrupted and uncorrupted models.

[23] The following derivation introduces a basis comparison metric that calculates the amount of energy in common between the spans of two truncated KL basis models. We begin the derivation by defining the two truncated bases, $A = [\mathbf{a}_1\ \mathbf{a}_2 \cdots \mathbf{a}_\mathbf{N}]$ and $B = [\mathbf{b}_1\ \mathbf{b}_2 \cdots \mathbf{b}_\mathbf{N}]$ where $\mathbf{a_n}$ and $\mathbf{b_n}$ are the nth basis vectors of A and B, respectively. We map each basis vector of A onto the space spanned by B via the inner product,

$$\hat{\mathbf{x}}_n = B^T \mathbf{a}_n. \tag{2}$$

Because the columns of B are orthonormal, each element in $\hat{\mathbf{x}}_n$ gives the magnitude of the projection of $\mathbf{a}_n$ in the direction of the B basis vectors. Also, because each $\mathbf{a}_n$ has unity length, the magnitude of $\hat{\mathbf{x}}_n$ is between zero and one. A value of "0" indicates that $\mathbf{a}_n$ is orthogonal to B while "1" indicates that $\mathbf{a}_n$ can be completely represented by B.

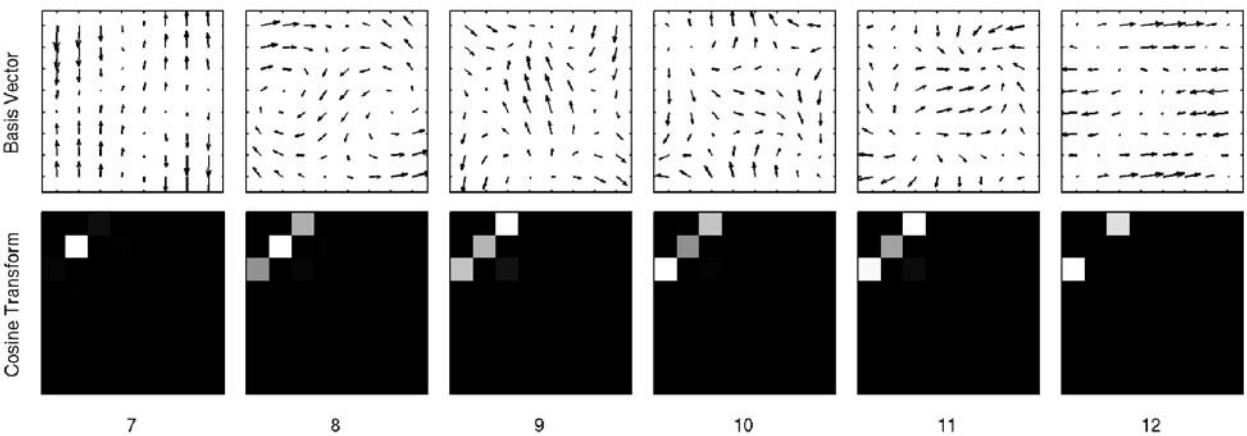

Figure 3. The wind fields corresponding to KL basis vectors 7–12 and the magnitudes of their corresponding cosine transforms.

Table 1. Comparison of Several KL Models Corrupted With Artificially Induced Ambiguity-Selection Errors to the Original (Noncorrupted) KL Model[a]

Induced Ambiguity-Selection Errors, %	$L_{A,B}$, Six Basis Vectors
0	1.0000
4	0.9996
8	0.9992
12	0.9989
16	0.9984
20	0.9981

[a] The percentage of wvcs corrupted by induced ambiguity-selection errors in the training data is given in the left column. The basis comparison metric is given in the right column ($L_{A,B}$) where A is the corrupted KL model and B is the original model. A and B both have six basis vectors.

The total mean energy of the first N projected basis vectors is

$$L_{A,B} = \frac{1}{N}\sum_{n=1}^{N} \|B^T \mathbf{a}_n\|_2^2. \tag{3}$$

$L_{A,B}$ is known as the basis comparison metric. The value of $L_{A,B}$ is also between 0 and 1, indicating the fraction of the energy in the span of A that is also spanned by B. Simplifying equation (3), we obtain

$$\begin{aligned} L_{A,B} &= \frac{1}{N}\sum_{n=1}^{N} (B^T \mathbf{a}_n)^T B^T \mathbf{a}_n \\ &= \frac{1}{N}\mathrm{tr}\left\{(B^T A)^T B^T A\right\} \\ &= \frac{1}{N}\| B^T A \|_F^2 \end{aligned} \tag{4}$$

where $\| \cdot \|_F$ is the Frobenius norm.

[24] The basis comparison metric ($L_{A,B}$) is applied to the uncorrupted KL model against each of the corrupted models and the results are given in Table 1. From this analysis, each six-parameter model derived from corrupted data is shown to span over 99% of the same vector space as the original KL model. This indicates that the KL model is not sensitive to ambiguity selection errors (at least as high as 20%) in the training data set. Because the ambiguity selection errors inherent in QuikSCAT data are estimated to only be about 5% (see section 5.1), the low-order model is considered insensitive to the ambiguity selection of the training set.

4. QA Analysis Method

[25] Because we do not know the true wind, an absolute assessment of the performance of QuikSCAT wind retrieval is unachievable. It is possible, however, to evaluate the self-consistency of the ambiguity-selected winds. From the KL analysis in section 3, we see that the wind is dominated by low spatial frequency content. By mapping the observed winds onto the low-order KL eigenvectors and comparing the observed wind to the model fit, we identify the wind variability outside of the typical wind spectrum. We can generally attribute large errors between the model fit and the ambiguity-selected wind to noise or ambiguity selection errors.

[26] Besides noise and ambiguity selection errors, certain fine-scale wind features cannot be accurately modeled by the KL model's limited basis set. Such areas may differ from the model fit, albeit the ambiguity selection is the best possible. Also, it is possible for ambiguity selection errors to result in a consistent wind flow. Due to these considerations, the QA analysis method is considered effective only in identifying "possible" ambiguity selection errors [*Gonzales and Long*, 1999].

4.1. Overview

[27] Our QA analysis method has two parts: First, each 8 × 8 region is classified as "good," "fair," or "poor" by the absolute level to which it deviates from the model. This region classification rates the level of noise, ambiguity selection errors, and poorly modeled fine-scale wind features. Second, possible ambiguity selection errors regions are identified using a more sophisticated approach. This ambiguity selection error detection method is optimized to suppress certain known effects of noise in order to better locate those regions containing ambiguity selection errors. A flow diagram of the QA method is shown in Figure 4 with a summary description given in the following sections.

4.1.1. Weighted Least Squares Model Fit

[28] The swath is segmented into 8 × 8 regions overlapping by 50% in the cross-track and along-track directions. Regions containing more than 25% invalid cells are ignored. An invalid cell is a wvc over land or ice where wind retrieval is not performed. The KL model is applied to each 8 × 8 region using a weighted least squaress fit,

$$\hat{\mathbf{w}}_m = F(F^T W F)^{-1} F^T W \mathbf{w}_o \tag{5}$$

where F is the truncated basis model, W is a weighting matrix, $\mathbf{w}_o$ is the observed wind, and $\hat{\mathbf{w}}_m$ is the model-fit estimate. The vectors $\mathbf{w}_o$ and $\hat{\mathbf{w}}_m$ and the columns of F are

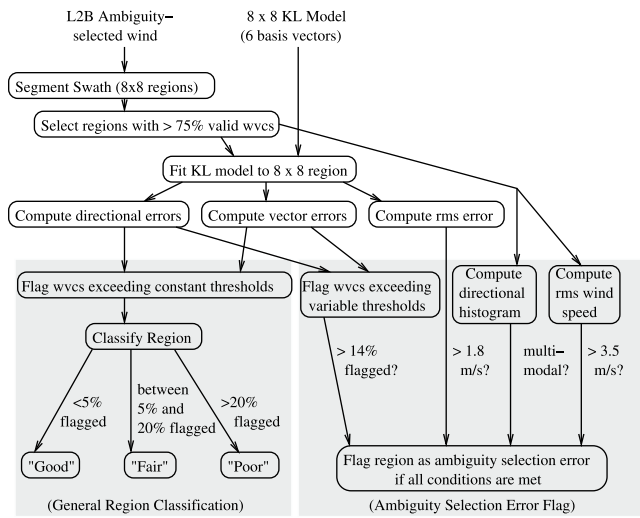

Figure 4. A flow diagram describing the QA analysis method. The left shaded box defines the region classification thresholds, and the right shaded box defines the thresholds involved with flagging a region as an ambiguity-selection error.

Table 2. Constant wvc Thresholds Determining the Flagging of a Vector[a]

wvc Threshold	Value
Direction error	23°[b]
Vector error	max{2.7[b], 0.5u_{rms}} m/s

[a] The term u_{rms} is the region root mean square (rms) wind speed defined by $u_{rms} = (\mathbf{w}_o^T \mathbf{w}_o / N)^{1/2}$, where $\mathbf{w}_o$ is the standard vector form of the observed wind field and N is the number of valid data cells in the region.
[b] Denotes values used by *Gonzales and Long* [1999].

in standard vector form defined in section 3. The diagonal weighting matrix W places a weighting of "1" on valid wvcs and a weighting of "0" on invalid wvcs. Thus, wvcs containing invalid data do not contribute to the modeled wind field.

[29] After creating the model fit, the directional and vector error between the observed wind and model-fit wind for each wvc is computed. The directional error is the difference in direction between the model-fit cell and the selected ambiguity, i.e.

$$\phi_{\mathbf{e}} = |\phi_m - \phi_o|_{0 \le \phi_{\mathbf{e}} \le 180} \tag{6}$$

where ϕ_m and ϕ_o are the directions of the modeled and observed vectors. The direction error is always between 0° and 180°. The vector error is the magnitude of the vector difference between the model fit and the selected ambiguity, i.e.

$$k_e = ((u_m - u_o)^2 + (v_m - v_o)^2)^{\frac{1}{2}} \tag{7}$$

where (u_m, v_m) and (v_o, v_o) are the cross-track and along-track components of the model-fit and observed wvc, respectively.

4.1.2. Region Classification (Good, Fair, Poor)

[30] In order to evaluate the overall consistency of the wind, each 8 × 8 region is classified according to the number of wvcs exceeding "constant" directional or vector thresholds. The term "constant" denotes that the thresholds are independent of cross-track position (as opposed to "variable" thresholds explained in section 4.1.3). These wvc thresholds are given in Table 2.

[31] When the region exceeds 20% flagged cells (those that exceed the "constant" thresholds), it is classified as "poor." A poor rating indicates that the region is not spatially consistent due to a high noise level, significant ambiguity selection errors, or fine-scale wind features such as fronts or storms. If the region contains 5% to 20% flagged wvcs, it is classified as "fair." A fair rating indicates that the region may have several noisy vectors or some fine-scale wind variations. Small isolated ambiguity selection errors may also receive a fair rating. A "good" rating (less than 5% flagged cells) indicates a spatially consistent wind field with possibly only a few noisy vectors. These region thresholds are summarized in Table 3.

4.1.3. Ambiguity Selection Error Detection

[32] The next stage of the QA analysis is ambiguity selection error detection. Here, we suppress flagging of regions due to noise in order to better locate ambiguity selection errors. To detect 8 × 8 regions containing possible ambiguity selection errors, two types of consistency checks are performed: a model-based consistency check, and a directional-histogram-based consistency check.

[33] First, we perform a model-based consistency check. The region is compared to the KL model fit and wvcs are flagged according to variable directional and vector error thresholds (similar to the region classification explained in section 4.1.2). The variable wvc thresholds are tuned to a manually inspected training data set. The training set consists of 15 swaths of SeaWinds data in which ambiguity selection errors are manually flagged. The variable wvc thresholds are set to equalize the false alarm rate for all cross-track/rms wind speed bins, giving constant performance across the swath (see Figure 5). Where the region exceeds a limit in the number of cells flagged (set at 14%) and an RMS error threshold, it is flagged as inconsistent. This model-based consistency check is explained further in Appendix A.

[34] Second, we perform a directional-histogram-based consistency check by inspecting the 8 × 8 wvc region for multiple directional flows. The purpose of this non-model-based consistency check is to ensure that the region is not being flagged solely due to noise. Generally, the directions of noisy wind vectors have some random distribution about a mean flow. When no ambiguity selection errors are present, a high noise level may cause the region to deviate significantly from the model, albeit there is only one mean flow. Ambiguity selection errors, however, generally cause neighboring patches of wvcs to point in contradictory directions, creating multiple main wind directions in a single 8 × 8 region. This can be detected by inspecting the histogram of directions for multiple modes. This supplementary consistency check is described further in Appendix B. When a region fails both model-based and directional-histogram-based consistency checks, it is identified as containing possible ambiguity selection errors.

[35] In addition to these consistency checks, all regions under 3.5 m/s RMS wind speed are not examined for possible ambiguity selection errors because the SNR is too low to validate the wind direction estimates. From experience with the NSCAT QA analysis, most NSCAT regions with RMS wind speed less than 4.0 m/s are flagged primarily because of noise [*Gonzales and Long*, 1999]. Similarly, through subjective examination of SeaWinds data, the noise level for regions below 3.5 m/s RMS is determined to be too high to subjectively assess the ambiguity selection. Approximately 7% of the total number of regions fall beneath this threshold.

4.2. Performance of the Ambiguity Selection Error Detection Method on the Training Data Set

[36] Ambiguity selection error detection is performed on the training data set, and the number of false alarms and

Table 3. Thresholds Determining the Classification of a Region

Classification	Percentage of Cells Flagged per Region, %
"Good"	<5
"Fair"	5–20
"Poor"	>20[a]

[a] Denotes values used by *Gonzales and Long* [1999].

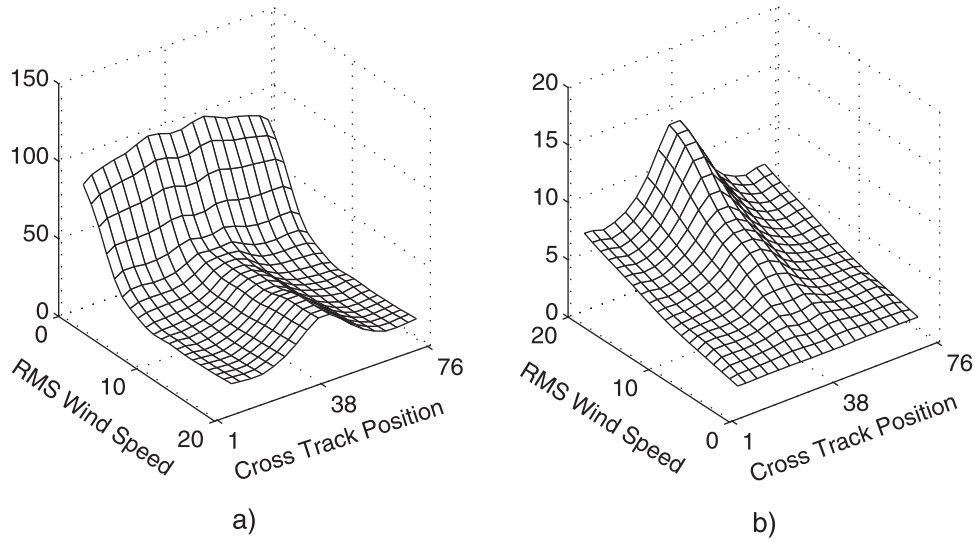

Figure 5. The (a) directional and (b) vector error thresholds per cross track and RMS wind speed that give a constant false alarm rate.

missed detections are tabulated. A region is considered a false alarm if it was not subjectively identified as an ambiguity selection error, but is flagged by the detection method. A region is considered a missed detection if the region was subjectively identified as an ambiguity selection error, but neither it nor an overlapping region is flagged by the detection method. The false alarm rate (number of false alarms per number of regions not subjectively flagged as ambiguity selection errors) is determined to be approximately 1.5%. The missed detection rate (number of missed detections per number of regions subjectively flagged as ambiguity selection errors) is found to be about 3%. This means that the ambiguity selection errors are correctly identified 97% of the time.

[37] The false alarm and missed detection rates are also computed as a function of cross-track position and RMS wind speed. These results are summarized in Figure 6. From this, we see that the far swath produces less false alarms than the inner- and mid-swath regions. Also, the false alarm rate and missed detection are slightly higher in the 3.5–4.5 m/s bin because the data at these wind speeds tend to be somewhat noisier than for other wind speed bins.

5. QA Analysis

[38] In this section, the QA analysis method is applied to 2 years of QuikSCAT data and a statistical account is presented. First, the overall quality of the data set is evaluated. Next, the quality assessment is presented as a function of cross-track position and RMS wind speed. Last, we present the QA results as a function of time and latitude band. We also compare the ambiguity selection errors to the number of cyclonic storms and the percentage of wvcs corrupted by rain.

5.1. Overall SeaWinds QA Results

[39] An aggregate assessment of the ambiguity selection and self-consistency of retrieved winds for 2 years of QuikSCAT data is provided in this section. Table 4 summarizes the percentage of 8×8 regions classified as "good," "fair," or "poor" and the percent flagged as possible ambiguity selection errors. The majority of the regions examined are classified as "good" (>65%). This indicates that in general, most QuikSCAT point-wise derived winds have a relatively low noise level. A substantial portion (15.5%) of the regions are classified as poor. These regions have a high noise level or contain ambiguity selection errors. However, using the ambiguity selection error detection method, only 5% of regions are flagged as possible ambiguity selection errors, suggesting that only about one third of the poor regions are a result of ambiguity selection errors. Thus, we conclude that the SeaWinds ambiguity selection is at least 95% effective for wind speeds exceeding 3.5 m/s.

[40] We also note that the automated ambiguity selection error detection method classifies regions in the 2-year data set in approximately the same proportion as the training data set, suggesting that the training data is representative of the 2-year SeaWinds data collection.

[41] The results given here for SeaWinds are very comparable to the QA results obtained for NSCAT [*Gonzales and Long*, 1999]. With NSCAT, 65% of regions are classified as "perfect" or "good," which is similar to a "good" classification for the SeaWinds QA analysis. In addition, 18% of NSCAT regions are classified as "poor" which is only slightly more than for SeaWinds. Also, both SeaWinds and NSCAT ambiguity selection is determined to be at least 95% effective for wind speeds exceeding 3.5 and 4 m/s, respectively.

5.2. Cross-Track/RMS Wind Speed Dependence

[42] Here, we analyze the results of the QA analysis as a function of cross track position and region RMS wind speed. We first compare the percentage of "poor" regions to possible ambiguity selection error regions as a function of cross track position and RMS wind speed (see Figure 7). Poor regions occur in higher percentages at nadir and where there is low wind speed. As discussed in section 2,

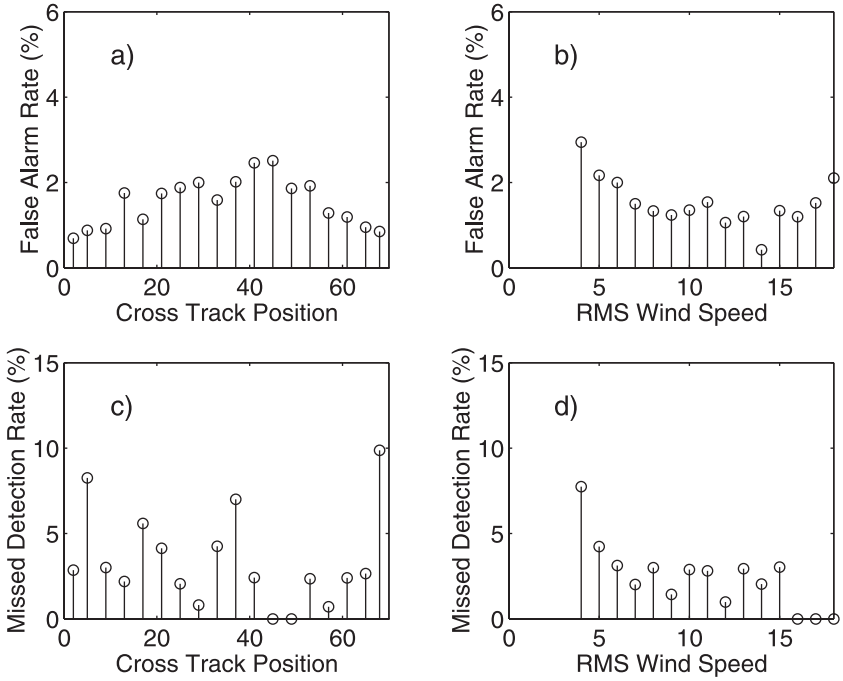

Figure 6. False alarm and missed detection rates for the ambiguity-selection error detection method per cross track and RMS wind speed. This data is taken from a test set of 15 subjectively analyzed revs.

these areas generally contain higher noise, and are thus more apt to be rated "poor". Poor wind retrieval at low wind speeds are additionally observed by *Patoux and Brown* [2001].

[43] An important observation is that fewer possible ambiguity selection errors are inferred at nadir and on the edges of the swath than in the "sweet spot" (the off-nadir region, usually characterized by a high percentage of correct first ambiguities). In order to explain this, we examine the average number of ambiguities produced by the JPL wind estimation algorithm per cross-track position. The fraction of one to three ambiguity cases per cross track is compared to the fraction of four ambiguity cases averaged over 600 revolutions of SeaWinds data in Figure 8. The general shape of the curve representing the one to three ambiguity cases of Figure 8 closely mirrors the percent of ambiguity selection errors per cross-track position shown in Figure 7. At nadir and on swath edges (where there are fewer estimated ambiguity selection errors), there is a higher likelihood of having four ambiguity choices.

[44] These results suggest that a higher number of ambiguities enables the creation of a more self-consistent wind field by the ambiguity selection algorithm. A manual inspection of ambiguity selection errors shows that higher errors in the sweet spot often occur in connection with rain occurrences (see section 5.3 for an objective analysis of the rain effect). In regions of data corruption such as rain contamination where an entire region of first ambiguities may be incorrect, thresholded nudging in the sweet spot can result in blocks of incorrect initial selections. The point-wise median filter alone is insufficient in correcting such errors. Since more ambiguities are used in the nudging process at nadir and along swath edges [*Stiles et al.*, 2002], the result is a more self-consistent initial estimate in the presence of rain. Also, where there are more ambiguities, the point-wise median filter has a wider selection of possible vector directions to match the flow of the surrounding wvcs in regions of rain corruption.

5.3. Temporal QA Statistics

[45] Next, we examine SeaWinds ambiguity selection as a function of time. Figure 9 shows the flagged ambiguity selection errors averaged over 3 days for each point. The percent of flagged ambiguity selection errors stays nominally between 4% and 5% for the 2 years of SeaWinds data analyzed. Although the overall ambiguity selection appears constant, seasonal weather variations in various oceanic regions locally affect the SeaWinds wind retrieval performance.

[46] In order to understand weather pattern variations that affect SeaWinds' performance, we divide the QuikSCAT wind data into latitude bands (see Table 5). For each band, the average percentage of possible ambiguity selection

Table 4. Overall Results of the QA Analysis for the SeaWinds Data Set and for the Training Data Set[a]

Region Classification	Entire Data Set, %	Training Data Set (15 Revs), %
Good	65.2	63.6
Fair	19.3	19.6
Poor	15.5	16.8
Containing ambiguity-selection errors	4.6	4.9 (4.0 subjectively flagged)

[a] Also, the percent of ambiguity-selection errors subjectively flagged in the training data set.

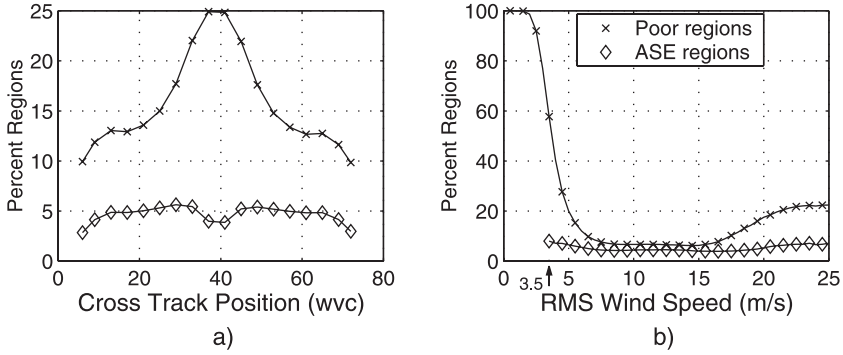

Figure 7. Percentage of all regions flagged as "poor" and flagged as containing ambiguity-selection errors (ASE regions) per (a) cross-track position and (b) RMS wind speed.

errors detected by the QA method, the average number of cyclonic storms passed by SeaWinds per degree latitude (see section 6 for details on cyclonic storm detection), and the number of wvcs flagged by the JPL rain flag [*Huddleston and Stiles*, 2000] per day are computed and given in Figure 10.

[47] We define our count of the number of storms per latitude band as the number of cyclonic features passed by SeaWinds with RMS wind speed >6 m/s divided by the latitude band size in degrees. Due to SeaWinds' polar orbit, storms in the arctic and midlatitude regions may have multiple observations per day, while some tropical storms may only have one or possibly zero observations in a single day. Thus, this scatterometer-observed storm count is somewhat skewed from a true count of the number of storms in each latitude band.

[48] Storms are problematic in wind retrieval for several reasons. First, since storms exhibit fine-scale wind variations, the nudging field must be sufficiently accurate to correctly position such features. The numerical weather prediction models used to nudge SeaWinds data are interpolated from low-resolution estimates. Storm centers and other fine scale features may be misplaced or smoothed due to the interpolation process, resulting in a poor initial estimate of the wind flow.

[49] It is also important to note that small-scale storm features may be erroneously identified as possible ambiguity selection errors due to the limited basis set of the KL model [*Gonzales and Long*, 1999]. Thus, the correlation between storms and estimated ambiguity selection errors may be somewhat artificially induced. However, as is shown by a subjective analysis in section 6, the wind retrieval is worse in storm regions.

[50] Second, rain often exists in stormy areas. Rain affects the ambiguity selection of SeaWinds data by disrupting the scatterometer signal. Falling rain interacts with the signal, causing attenuation and backscatter from the atmospheric hydrometeors [*Ulaby et al.*, 1981]. In addition, rain disrupts the wind-generated capillary waves on the ocean surface, altering the wind-induced backscatter signature [*Bliven and Giovanangeli*, 1993]. Rain-corrupted wvcs are often augmented in speed and contain incorrect direction information. Generally, the first and second ambiguities point in a direction almost parallel with the cross track, independent of the wind's true direction. The ambiguities point parallel with the cross track because rain is a nearly isotropic scatterer, giving an equal response for both fore and aft looking observations. Likewise, winds blowing crosswise to the satellite flight direction also give a near-equal response from fore and aft observations [*Huddleston and Stiles*, 2000]. Rain effects on wind scatterometer backscatter returns have been found to be significant [*Bliven and Giovanangeli*, 1993; *Moore et al.*, 1983].

[51] Nearby ambiguity selection is influenced by rain-corrupted wvcs. When rain occurs in regions of high instrument skill (where thresholded nudging chooses only the first or second ambiguities), incorrect wind vectors may

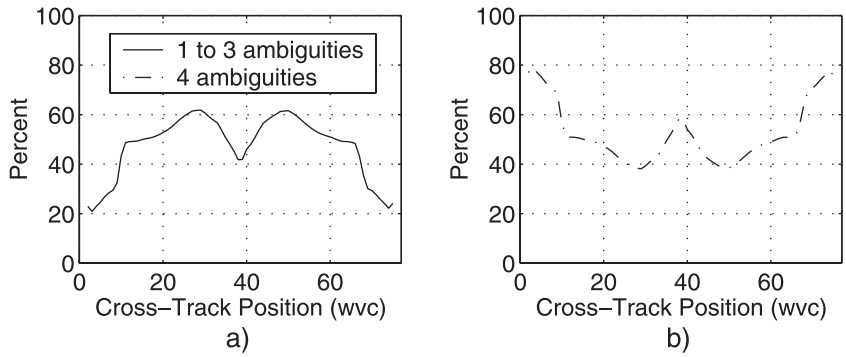

Figure 8. (a) Percent of wvcs per cross track position with one to three ambiguities and (b) percent of wvcs per cross-track position with four ambiguities averaged over 600 revs of SeaWinds data.

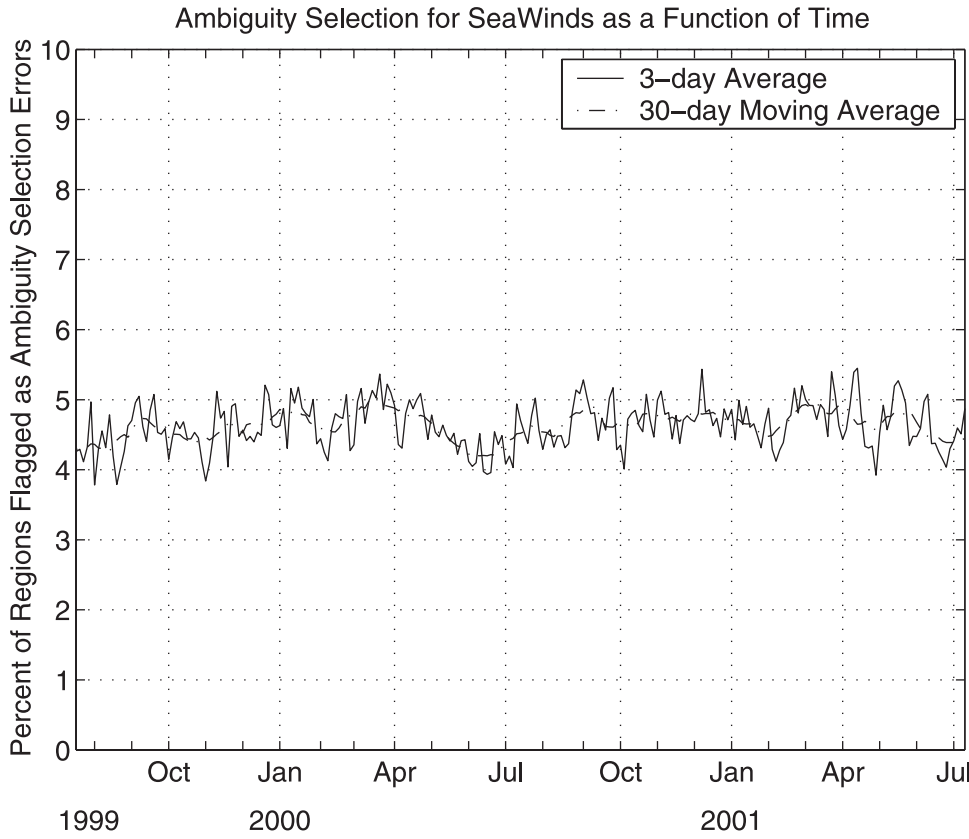

Figure 9. Ambiguity-selection errors as a function of time. The data line is the fraction of ambiguity-selection errors averaged over 3 days. The smooth line is a 30-day moving average.

be chosen to initialize the median filter. Because these corrupted wind vectors are given the same weight as other wind vectors, they influence neighboring cells causing an entire area of wvcs to point in an incorrect direction. Also, because the direction of isolated cells that have been rain contaminated may be incorrect, they may be flagged as possible ambiguity selection errors, since there is no better choice.

[52] From visual inspection of Figure 10, a correlation exists between the possible ambiguity selection errors, number of storms, and rain percentages. We quantify this correlation by computing the correlation coefficients between storms and ambiguity selection errors and between wvcs flagged as rain and ambiguity selection errors. These coefficients are listed in Table 6 for each latitude band. The highest correlation occurs in the third, fifth, and seventh latitude bands. This correlation suggests that rain and storms contribute to poor ambiguity selection and inconsistent wind flow.

[53] The equatorial band (latitude band 4) has the fewest storms. Large cyclonic storms around the equator are rare because the Coreolis effect that drives cyclonic circulation disappears at the equator. Thus, of storms and rain, rain has the larger influence on the self-consistency of the ambiguity-selected wind in that area. From visual inspection, both rain and possible ambiguity selection errors for the equatorial band peak around April. This time period in the equatorial band has some of the highest wind retrieval error rates.

[54] Latitude bands 3 and 5 (north and south tropical regions) demonstrate the most noticeable seasonal trend in storms, rain, and possible ambiguity selection errors. During the Austral summer months (November to May), rain, storms, and ambiguity selection errors increase in the Southern Hemisphere tropical band (see Figures 10b and 10c, latitude band 3). In the Austral winter months (May to November), there are decreased rain, storms and ambiguity selection errors. A similar, shifted seasonal trend occurs in the Northern Hemisphere in the fifth latitude band. Peak rain averages in band 5 (Northern Hemisphere tropics) are higher than band 3 for these years, resulting in higher peak ambiguity selection errors.

[55] Though more total storms occur in bands 2 and 6 (midlatitude regions) than in the tropics, the midlatitude bands are the most stable with respect to ambiguity selection errors. The seasonal variation of rain and ambiguity selection errors in these bands are not as distinct as in bands

Table 5. Latitude Bands

Latitude Band	Range
7	45° to 90°
6	25° to 45°
5	5° to 25°
4	−5° to 5°
3	−25° to −5°
2	−45° to −25°
1	−90° to −45°

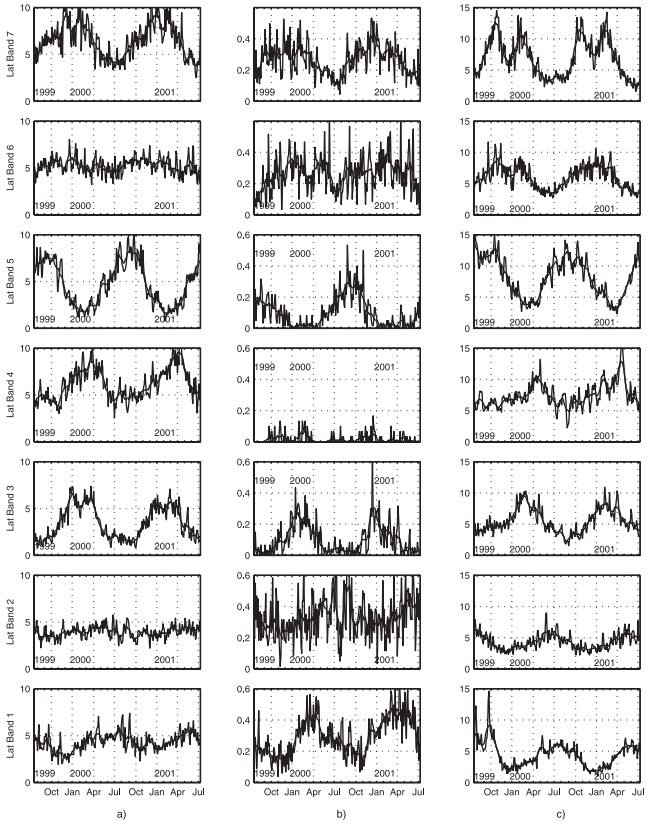

Figure 10. (a) Percent of ambiguity-selection error regions, (b) number of cyclonic storms per degree latitude, (c) percent of wvcs flagged with L2B rain flag for each latitude band averaged over 3 days per data point.

3 and 5, resulting in more seasonally uniform performance of scatterometer wind retrieval. Midlatitude rain averages are higher in the Northern Hemisphere, causing a slightly worse performance of SeaWinds ambiguity selection in this region.

[56] The two polar regions (bands 1 and 7) have distinctly different characteristics. The key difference between the polar regions is the position of the Earth's landmasses at each of the poles. In the Southern hemisphere (band 1), wind retrieval is not performed over the pole due to the position of Antarctic land and ice. The estimated winds are only from the upper part of the band, and are therefore very similar in their characteristics to the winds from the second latitude band. In the Arctic region (band 7), SeaWinds retrieves winds in the ice-free areas of the Arctic Ocean. Since the weather in the Arctic region widely varies with the seasons, the performance of SeaWinds also varies. The peak in ambiguity selection errors for the arctic region occurs from October to April during the stormy winter months.

6. Subjective Analysis of Cyclonic Storm Regions

[57] To better understand the scatterometer wind retrieval performance in storms, we perform a subjective analysis of the winds in regions containing cyclonic storm features. Cyclonic storm features are located by fitting the KL model to the NCEP fields and comparing the mean square of parameters 3 and 6 to the mean square of model parameters 1 and 2. Parameters 3 and 6 have cyclonic features, while parameters 1 and 2 represent uniform wind flow (see Figure 2). Where the mean square of parameters 3 and 6 is greater, the region is flagged as a cyclonic storm.

[58] The circular region surrounding the storm center with radius 10 wvcs (250 km) is manually examined in both the NCEP and ambiguity-selected data. The region is subjectively given a rating of "1," "2," or "3." A "1" rating indicates that the ambiguity-selected storm's cyclonic flow is very well defined and realistic. A "2" rating occurs when the storm has a mostly cyclonic flow, but there are some noticeable ambiguity selection errors or rain-corrupted wind vectors. A rating "3" indicates that the cyclonic flow is not well defined and the region may contain significant ambiguity selection errors or rain corruption. Examples of each rating are shown in Figure 11.

[59] Two separate time periods are examined: 2 weeks of QuikSCAT data from 19 July to 1 August 1999, and 2 weeks of QuikSCAT data from 17 January to 30 January 2000. Cyclonic features bordering land or ice, and those with excessively low wind speed (less than 6 m/s rms) are ignored. Additionally, the data is divided into Northern and Southern Hemisphere. The number of storms given each rating for each time period and hemisphere is shown in Table 7.

[60] Overall, less than 40% of the manually inspected storm cases are identified as "1." The remaining 60% of storm cases have some level of inconsistent wind flow which can often be attributed to ambiguity selection errors. In addition, 26% are given a "3" rating, indicating very poor wind retrieval.

[61] The Northern Hemisphere summer case exhibits the highest percentage of "1" ratings, while the proportion of Northern Hemisphere "3" cases remains approximately the same as the other cases. This increase in accuracy for the Northern Hemisphere summer may be related to the accuracy of the nudging NCEP fields for this case. In order to examine the accuracy of the NCEP winds, we calculate the distance between NCEP and QuikSCAT storm centers for each storm rated "1" or "2." After manually flagging the storm centers, we average the distances between NCEP and QuikSCAT centers for each period (see Table 8). The Northern Hemisphere July 1999 case exhibits the statistically best storm retrieval and also demonstrates the best collocation between NCEP and QuikSCAT storms. This example suggests that more accurate nudging in this area improves the wind retrieval performance.

Table 6. Correlation Coefficients for Each Latitude Band[a]

Latitude Band	C_{sa}	C_{ra}
7	0.9193	0.7626
6	0.5314	0.6285
5	0.8737	0.9152
4	0.1047	0.8399
3	0.8794	0.8456
2	0.6126	0.4388
1	0.5700	0.4072

[a] C_{sa} is the correlation coefficient between the smoothed ambiguity-selection error and storm data of Figure 10. C_{ra} is the correlation coefficient between the smoothed ambiguity-selection error and rain data of Figure 10.

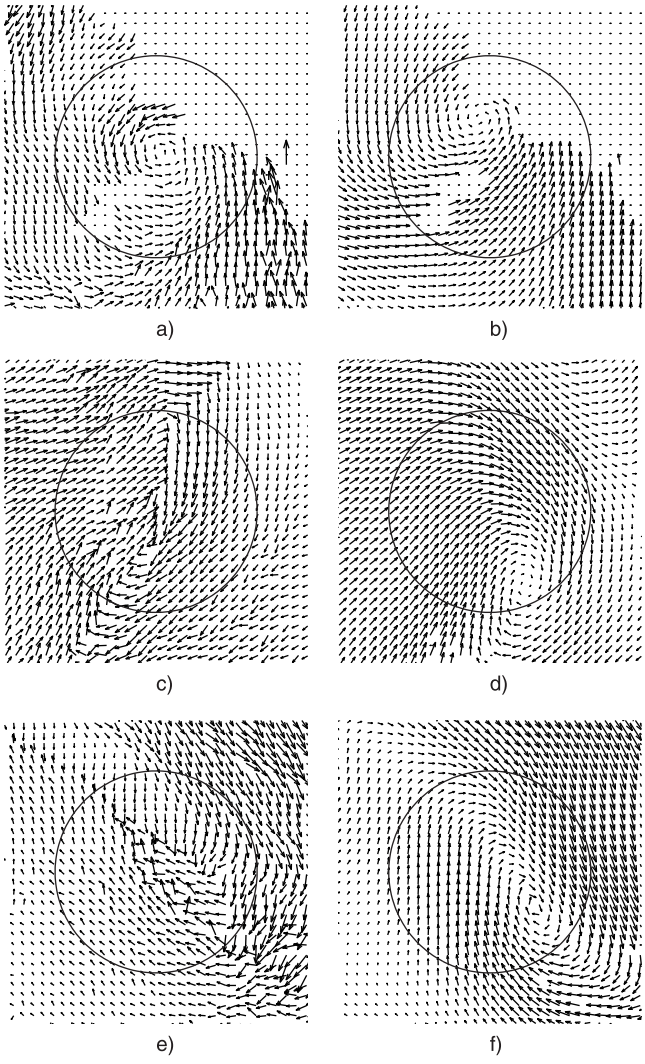

Figure 11. (a) QuikSCAT storm region rated "1," (b) corresponding NCEP data. (c) QuikSCAT storm region rated "2," (d) corresponding NCEP data. (e) QuikSCAT storm region rated "3," (f) corresponding NCEP data.

[62] In addition to the sensitivity to misplaced storms, the scatterometer wind retrieval of storms is also seriously affected by rain. For storms rated "3," on average, approximately 30% of vectors per region are flagged by the QuikSCAT L2B rain flag. For regions ranked "1" or "2," only 20% of vectors are flagged as containing rain. The higher rain averages in storms rated "3" suggests that rain is related to poorly retrieved storms. From manual inspection, where the rain corruption is severe, especially in lower wind speed regions, the backscatter is sufficiently affected as to make retrieval of fine scale features nearly impossible with current methods.

[63] This subjective analysis of QuikSCAT cyclonic storm cases suggests that scatterometer wind retrieval is more error prone in regions of cyclonic storms and lends support to the argument that storms, rain and ambiguity selection errors are correlated. Because rain and storm mislocation along with the QA flag and other factors can be indicators of poorly retrieved storms, automated storm rating procedures can be created using these storm sensitive parameters as inputs. An attempt at a simple maximum likelihood storm-rating technique is given in Appendix C.

Table 7. Number of Storms Identified as Rating "1," "2," and "3" for Two Time Periods: July 1999 and January 2000[a]

	July 1999				January 2000				Total			
Rating	1	2	3		1	2	3		1	2	3	
North	55	17	33	/105	39	54	35	/128	94	71	68	/233
South	42	62	35	/139	74	76	44	/194	116	138	79	/333
Total	97	79	68	/244	113	130	79	/322	210	209	147	/566

[a] For each time period, the data is further subdivided into Northern and Southern Hemisphere locations.

7. Discussion and Summary

[64] Scatterometer wind retrieval offers the opportunity for advanced study of the oceans. Although the estimation process results in ambiguous solutions, we estimate the current point-wise ambiguity selection technique to be 95% effective in creating a self-consistent wind flow. Problems associated with the current wind retrieval process generally are correlated with natural phenomena (i.e., wind speed, storminess, and rain) and instrument geometry. Wind speed and instrument geometry affect the overall noise level of the retrieved winds. Scatterometer winds are especially noisy at low wind speeds and at nadir for SeaWinds. Nearly 100% of regions with RMS wind speeds less than 2.5 m/s are sufficiently corrupted by noise to receive a "poor" rating. In addition, about 25% of regions at nadir are given a "poor" rating.

[65] Ambiguity selection errors are correlated with rain corruption in the scatterometer signal and cyclonic storm features. Rain corruption generally creates significant changes in the σ° values. The large error in the vector estimates for rain contaminated wvcs not only affects the corrupted wvc, but can significantly change the flow of the surrounding wvcs from the point-wise filtering process, thus creating patches of ambiguity selection errors.

[66] Storms present a dual problem. First, the nudging data used to initialize the ambiguity selection process is often in error near fine-scale wind features. Second, rain associated with cyclonic storms often significantly affects the retrieved winds. From the subjective analysis presented in this paper, 37% of examined storm cases have very well defined and realistic flow in the scatterometer-derived wind fields, while 26% of storm cases have very poorly retrieved flow. The effects of rain and nudging-data storm misplacement are significant.

[67] Although limitations in scatterometer retrieved winds exist, point-wise estimated scatterometer winds are of very

Table 8. Average Distance Between the NCEP Storm Centers and the L2B Storm Centers for Storms Rated 1 or 2[a]

	July 1999	January 2000	Combined
North	74	112	95
South	118	106	111
Combined	100	108	105

[a] Values are in km.

Table 9. Sample Means and Standard Deviations of Automated Storm Rating Parameters[a]

Rating	d		R		Q_1		Q_2		U_q		U_n	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
1	82.5	55	18.5	21.2	22.5	9.5	19.2	10.5	11.6	3.3	11.6	2.5
2	127.5	97.5	22.4	20.2	31.3	10.3	29.2	10.3	12.2	3.1	12.0	2.9
3	145	85	29.7	25.6	33.8	11.5	31.2	11.4	11.2	3.2	10.5	2.18

[a] Here, d, distance between NCEP and QuikSCAT storm centers (km); R, wvcs flagged with the L2B rain flag per region (%); Q_1, wvcs flagged per region by the variable thresholds (QA individual cell flag) (%); Q_2, wvcs flagged per region by both the QA region and individual cell flag (%); U_q, average RMS wind speed of storm regions (m/s) from QuikSCAT data; U_n, average rms wind speed of storm regions (m/s) from NCEP data.

high quality in nonrain and moderate wind speed areas. Further research in wind estimation in the presence of rain or in storm regions may aid in higher accuracy of scatterometer retrieved winds. These improved methods may include the addition of a rain rate parameter into the MLE technique and specialized storm retrieval methods using the KL or other storm-specific models. Some current methods designed to improve wind retrieval accuracy include those of *Draper and Long* [2001], *Stiles et al.* [2002], *Patoux and Brown* [2001], and *Long and Mendel* [1990]. In addition, research is ongoing to improve the GMF at low [e.g., *Shankaranarayanan and Donelan*, 2001] and high wind speeds [e.g., *Yueh, et al.*, 2001].

Appendix A: Model-Based Consistency Check

[68] The model-based consistency check is based on comparing the directional and vector errors against a set of "variable" thresholds. The term "variable" indicates that the thresholds are raised in areas of known high noise to suppress flagging of vectors due to noise. As discussed previously, the noise level for SeaWinds is variable with cross-track position and wind speed. In order to reduce flagging of regions due to noise only, the wvc thresholds are individually adjusted for each cross-track position and RMS wind speed.

[69] The variable wvc thresholds were empirically determined through an analysis of false alarms versus missed detections on a training data set consisting of 15 subjectively analyzed swaths (L2B revs 3000-3014). All 8 × 8 wvc regions that subjectively exhibited ambiguity selection errors were identified and binned according to cross-track position and RMS wind speed. Then, the vector and angle thresholds for each bin were separately iteratively applied and the number of flagged wvcs were tallied. The regions that exceeded a limit in wvcs flagged were identified as ambiguity selection errors. The variable wvc thresholds were iteratively adjusted until the region false alarm rate was equalized for all cross-track/rms wind-speed bins. In doing this, region threshold of 14% performed the best and was chosen as the region threshold for flagging a region as an ambiguity selection error. These variable wvc thresholds were then smoothed and further manually adjusted to give subjectively better performance. The final wvc thresholds are given in Figure 5.

[70] The variable thresholds are indexed by the cross-track position and RMS wind speed of the 8 × 8 region and applied to all valid wvcs in the region. When a wvc exceeds either vector or angle thresholds, it is flagged as a possible ambiguity selection error. Where greater than 14% of cells in a region are flagged with the variable thresholds and the RMS region error is greater than 1.8 m/s, the entire 8 × 8 region is considered as possibly containing ambiguity selection errors. The RMS region threshold of 1.8 m/s was subjectively determined. The method, however, is not particularly sensitive to this value.

Appendix B: Directional Histogram-Based Consistency Check

[71] In addition to the model-based consistency check, each 8 × 8 region is inspected for multiple directional flows. A histogram of the vector directions in the region is assembled with a bin spacing of 24°. Then, the histogram is reordered with the lowest bin value first (this eliminates peaks straddling 0° and 360°). The histogram is then numerically differentiated. Multiple modes in the histogram are identified where the derivative crosses the zero line.

[72] The consistency check is relatively insensitive to the bin size. Similar performance on the training data set was achieved for bin size of 20° and 30°. However, a bin size of 24° yielded the least false alarms.

[73] Examining the histogram of wind directions for multiple modes supplements the model-based detection scheme by providing an additional view of the consistency of a region without the issues associated with the restricted basis model.

Appendix C: Automated Storm-Rating Procedure

[74] The data set of subjective storm ratings is used to train an automated storm detection and rating method. With this method, storms are located in the nudging data as previously described, and a finer search is performed in both QuikSCAT and nudging fields in the surrounding area of a detection. The storm center in both QuikSCAT and the nudging data is estimated to be positioned where the ratio of the mean square of model basis coefficients 3 and 6 to the mean square of coefficients 1 and 2 is a maximum. The following parameters are then calculated for the surrounding

Table 10. Number of Storms Subjectively Rated "1," "2," or "3" Versus Number of Storms Rated "1," "2," or "3" by the Automated Method

Subjective Rating	Automated Rating		
	"1"	"2"	"3"
"1"	117	44	24
"2"	51	68	50
"3"	12	17	49

circular region of radius 10 wvcs (250 km): (1) distance between QuikSCAT storm center and nudging storm center, (2) percentage of wvcs flagged by the L2B rain flag, (3) percentage of individual wvcs flagged by the QA variable thresholds, (4) percentage of individual wvcs flagged by the QA variable thresholds where the 8×8 region was additionally flagged, (5) RMS wind speed of the QuikSCAT region, and (6) RMS wind speed of the nudging field region. The means and standard deviations of each of these parameters given the region is subjectively identified as a "1," "2," or "3" are given in Table 9.

[75] Now, assuming a Gaussian distribution for each parameter, a maximum likelihood estimator is used to automatically rate each region,

$$\text{Rating} = \arg\min_{n} \left\{ \sum_{i} \frac{(X_i - \mu_{i,n})^2}{\sigma_{i,n}^2} \right\} \tag{C1}$$

where $(\mu_{i,n}, \sigma_{i,n})$ are the mean and standard deviation of the ith parameter given a rating of n. The quality of this method is demonstrated in Table 10.

[76] The automated method correctly detects a "1" rating with about 65% accuracy. Also, the majority of storms automatically rated "2" or "3" are generally subjectively rated either "2" or "3," although the automated method is not able to distinguish between a "2" and "3" with high precision. Although this simple method has limitations and is not tuned for optimal performance, it suggests that ambiguity selection of a cyclonic storm in scatterometer data can be evaluated by automated methods. The ability to automatically locate and rate storms may aid scientists and those using the data to indicate if the data around a storm is usable. Also, where storms are identified as a "2" or "3," specialized ambiguity selection schemes may be used to increase the quality of the data in those regions.

References

Bliven, L. F., and J. P. Giovanangeli, Experimental study of microwave scattering from rain- and wind-roughened seas, *Int. J. Remote Sens.*, *14*, 855–869, 1993.

Brown, R. A., On satellite scatterometer model function, *J. Geophys. Res.*, *105*, 29,195–29,205, 2000.

Chi, C., and F. K. Li, A comparative study of several wind estimation algorithms for spaceborne scatterometers, *IEEE Trans. Geosci. Remote Sens.*, *26*, 115–121, 1988.

Donelan, M. A., and W. J. Pierson, Radar scattering and equilibrium ranges in wind-generated waves with application to scatterometry, *J. Geophys. Res.*, *92*, 4971–5029, 1987.

Donnelly, W. J., J. R. Carswell, R. E. McIntosh, P. S. Chang, J. Wilkerson, F. Marks, and P. G. Black, Revised ocean backscatter models at C and Ku band under high-wind conditions, *J. Geophys. Res.*, *104*, 11,485–11,497, 1999.

Draper, D. W., and D. G. Long, An advanced point-wise ambiguity selection algorithm: Application to SeaWinds, *Proc. Int. Geosci. Remote Sens. Symp.*, *5*, 2190–2192, 2001.

Gonzales, A. E., and D. G. Long, An assessment of NSCAT ambiguity removal, *J. Geophys. Res.*, *104*, 11,449–11,457, 1999.

Huddleston, J. N., and B. W. Stiles, A multi-dimensional histogram rain flagging technique for SeaWinds on QuikSCAT, *Proc. Int. Geosci. Remote Sens. Symp.*, *3*, 1232–1234, 2000.

Freilich, M. H., and D. B. Chelton, Wavenumber spectra of pacific winds measured by the Seasat scatterometer, *J. Phys. Oceanogr.*, *16*, 741–757, 1986.

Freilich, M. H., and R. S. Dunbar, The accuracy of the NSCAT 1 vector winds: Comparisons with National Data Buoy Center buoys, *J. Geophys. Res.*, *104*, 11,231–11,246, 1999.

Jain, A. K., *Fundamentals of Digital Image Processing*, 163 pp., Prentice-Hall, Old Tappan, N. J., 1989.

Patoux, J., and R. A. Brown, A scheme for improving scatterometer surface wind fields, *J. Geophys. Res.*, *106*, 23,985–23,994, 2001.

Long, D. G., and J. M. Mendel, Model-based estimation of wind fields over the ocean from wind scatterometer measurements, II, Model parameter estimation, *IEEE Trans. Geosci. Remote Sens.*, *28*, 361–373, 1990.

Long, D. G., and J. M. Mendel, Identifiability in wind estimation from scatterometer measurements, *IEEE Trans. Geosci. Remote Sens.*, *29*, 268–276, 1991.

Moon, T. K., and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, 328 pp., Prentice-Hall, Old Tappan, N. J., 2000.

Moore, R. K., and A. D. Fung, Radar determination of winds at sea, *Proc. IEEE*, *67*, 1504–1521, 1979.

Moore, R. K., A. H. Chaudhry, and I. J. Birrer, Errors in scatterometer-radiometer wind measurement due to rain, *IEEE J. Oceanic Eng.*, *8*, 37–48, 1983.

Oliphant, T. E., and D. G. Long, Accuracy of scatterometer derived winds using the Cramér-Rao bound, *IEEE Trans. Geosci. Remote Sens.*, *37*, 2642–2652, 1999.

Shaffer, S. J., R. S. Dunbar, S. V. Hsiao, and D. G. Long, A Median-filter-based Ambiguity Removal Algorithm for NSCAT, *IEEE Trans. Geosci. Remote Sens.*, *29*, 167–174, 1991.

Shankaranarayanan, K., and M. A. Donelan, A probabilistic approach to scatterometer model function verification, *J. Geophys. Res.*, *106*, 19,969–19,990, 2001.

Spencer, M. W., C. Wu, and D. G. Long, Tradeoffs in the design of a spaceborne scanning pencil beam scatterometer: Application to SeaWinds, *IEEE Trans. Geosci. Remote Sens.*, *35*, 115–126, 1997.

Stiles, B. W., B. D. Pollard, and R. S. Dunbar, Direction interval retrieval with thresholded nudging: A method for improving the accuracy of QuikSCAT winds, *IEEE Trans. Geosci. Remote Sens.*, *40*, 79–89, 2002.

Vershell, M. A., M. A. Bourassa, D. E. Weissman, and J. J. Obrien, Ocean model validation of the NASA scatterometer, *J. Geophys. Res.*, *104*, 11,359–11,373, 1999.

Ulaby, F. T., R. K. Moore, A. K. Fung, Extinction and backscattering by rain, in *Microwave Remote Sensing Active and Passive*, vol. 1, pp. 316–326, Artech House, Norwood, Mass., 1981.

Yueh, S. H., B. W. Stiles, W. Tsai, H. Hu, and W. T. Liu, QuikSCAT geophysical model function for tropical cyclones and application to hurricane Floyd, *IEEE Trans. Geosci. Remote Sens.*, *39*, 2601–2611, 2001.

Zeng, L., and R. A. Brown, Scatterometer observations at high wind speeds, *J. Appl. Meteorol.*, *37*, 1412–1420, 1998.

D. W. Draper and D. G. Long, Microwave Earth Remote Sensing (MERS) Laboratory, Brigham Young University, 459 CB, Provo, UT 84602, USA. (draperd@et.byu.edu; long@ee.byu.edu)