AN ASSESSMENT OF THE NASA SCATTEROMETER AMBIGUITY REMOVAL TECHNIQUE

Amy E. Gonzales

Department of Electrical and Computer Engineering

Master's Degree, May 1998

ABSTRACT

The NASA Scatterometer (NSCAT) estimates the wind speed and direction of near-surface ocean wind. Several possible wind vectors are estimated for each resolution element known as a wind vector cell (wvc). Typically, the speeds of the possible wind vectors are nearly the same, but the directions are very different. The correct wind must be distinguished in a step called ambiguity removal. Unfortunately, ambiguity removal algorithms are subject to error. In an attempt to evaluate the accuracy of the Jet Propulsion Laboratory (JPL) NSCAT product, I developed a new model-based quality assurance (QA) algorithm which uses only NSCAT data. The algorithm segments the swath into overlapping 12x12 wvc regions and classifies each region according to estimated quality. The nine month NSCAT mission dataset is analyzed. In 82% of the regions, the ambiguity removal is over 99% effective with the ambiguity rmoval errors correctable using a model-based correction technique. In 5% of the regions, areas of significant ambiguity removal errors are found. For remaining regions, all of which have root mean square (rms) wind speeds less than 4 m/s, there is too much uncertainty in the wind field model or too much noise in the measurements to uniquely evaluate ambiguity removal accuracy with sufficient confidence. I thus conservatively conclude that for the set of regions with rms wind speed greater than 4 m/s, NSCAT ambiguity removal is at least 95% effective, i.e. NSCAT may not be effective for the 5% of regions with significant errors.

COMMITTEE APPROVAL:

_____
David G. Long, Committee Chairman

_____
Michael   Jensen, Committee Member

_____
Linton   Salmon, Committee Member

AN ASSESSMENT OF THE NASA SCATTEROMETER AMBIGUITY REMOVAL TECHNIQUE

A Thesis

Submitted to the

Department of Electrical and Computer Engineering

Brigham Young University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Amy E. Gonzales

May 1998

This thesis by Amy E. Gonzales is accepted in its present form by the Department of Electrical and Computer Engineering of Brigham Young University as satisfying the thesis requirement for the degree of Master of Science.

_____

David G. Long, Committee Chairman

_____

Michael Jensen, Committee Member

_____

Linton Salmon, Committee Member

_____   _____

Date   Michael Rice, Graduate Coordinator

DEDICATION

To my husband
Michael

ACKNOWLEDGMENTS

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

The NASA Scatterometer (NSCAT) is the latest generation in a series of spaceborne wind scatterometers that make measurements over the ocean under all weather conditions. From these measurements, the ocean wind is estimated. There are typically several possible solutions for each wind vector estimate (termed "ambiguities"). The actual wind vector is then distinguished in a step called ambiguity removal. However, ambiguity removal is not always correct. A quality assurance (QA) algorithm is needed to determine the accuracy of NSCAT ambiguity removal.

In this thesis, a wind field model is developed and used as the basis for an assessment of the accuracy of NSCAT ambiguity removal. A technique is developed to detect and correct ambiguity removal errors using only NSCAT data. The results of using this technique on the nine month NSCAT mission data set are presented. The results suggest that NSCAT ambiguity removal is at least 95% effective for the entire set of regions with rms wind speeds greater that 4 m/s. This is the first quality assurance algorithm that requires only NSCAT data.

1.1 Background

Ocean wind measurements are crucial to an improved understanding of the ocean, weather, and climate. Unfortunately, observational data (ships, island stations, buoys, etc.) are severely limited in coverage and accuracy. Additionally, observational data are rarely recorded during storms or over the large expanses of the ocean, leaving much of the wind over the ocean unknown.

This has lead to the development of microwave remote sensing devices that can measure the wind from space. In 1978, the first scatterometer (Ku-band) flew on Seasat (SASS) and demonstrated that scatterometers could accurately measure vector winds over the ocean. SASS was followed by two scatterometers from the European Space Agency, ERS-1 and ERS-2 (both C-band), in 1991 and 1995 respectively. NASA launched its own scatterometer (the Ku-band NSCAT) in 1996 (Naderi et al., 1991).

The NASA Scatterometer (NSCAT) is a microwave instrument capable of measuring vector winds over the ocean during all weather conditions. NSCAT has the ability to map the winds over the oceans of the entire earth every three days. An incredible amount of data is now available to all scientists studying air/sea interactions. The winds over the ocean modulate all air/sea fluxes. Accurate knowledge of the winds over the oceans will aid numerous disciplines in their study of the earth and its processes (Naderi et al., 1991).

One of the most important applications for measuring ocean winds is weather forecasting. Current methods of weather forecasting use mathematical models that take the current weather as input and predict future weather. NSCAT data has recently been shown to have a significant positive input on numerical weather prediction (Atlas, 1997). This is critical for forecasting severe events such as monsoons, hurricanes, and typhoons that take thousands of lives each year.

1.2 Wind Retrieval

Scatterometers do not directly measure the wind; rather the speed and direction of the near-surface wind are inferred from the normalized radar cross section (σ°) measurements of the ocean surface. The wind is related to σ° via a geophysical model function. However, given the scatterometer measurements at an observation point or wind vector cell (wvc), there are several possible wind vectors for any particular set of σ° measurements (Long and Mendel, 1991). Although the speeds are very similar, the directions vary with two to four possible directions for each wvc.

Traditional point-wise wind retrieval consists of two steps and uses only the σ° measurements for a single wvc to retrieve the wind for that cell. The first step is to find the multiple wind vectors (aliases) for each cell of the scatterometer swath. The second step, ambiguity removal, selects one unique wind vector estimate for each of these cells. Various ambiguity removal schemes have been developed (Schroeder et al., 1985; Hoffman, 1982; Atlas et al., 1987). For NSCAT, JPL has chosen to use a modified median filter technique is used (Shaffer et al., 1991; Shultz, 1990). Correct ambiguity removal results in the identification of the point-wise ambiguity that is closest to the actual wind vector. Unfortunately, ambiguity removal algorithms are prone to error.

Another method to determine wind measurements is model-based wind

retrieval (Long, 1993). The wind field model provides a description of the near-surface wind field over the scatterometer measurement swath and is optimized for scatterometer wind retrieval. The swath is sectioned into rectangular regions and the wind is extracted over the entire region instead of by individual resolution elements. The model relates the components of the wind vector field over this region to a set of model parameters (Oliphant, 1996). The models may be data-driven or physics-based and have been shown to provide more accurate wind measurements than point-wise wind retrieval (Long, 1993).

The wind field models used in model-based wind retrieval can also be used to improve the point-wise wind product by identifying and correcting ambiguity removal errors. One way to do this is to fit the estimated point-wise wind to a simple wind field model over a small area. Since ambiguity removal errors typically cause 90 degree or 180 degree shifts in wind direction, large errors in the fit suggest possible ambiguity removal errors while small errors suggest a realistic wind field. Areas of errors can be corrected by choosing the alias closest in direction to the model-fit.

1.3   Contribution

Current quality assessments of the NSCAT ambiguity removal technique rely on comparisons with buoy/ship data and numerical weather prediction winds. At present, quality assurance algorithms that use only NSCAT data are primitive.

Wind field models can be used as a quality assurance for the NSCAT point-wise wind product and to correct ambiguity removal errors. The model of interest for this baseline algorithm is the data-driven Karhunen-Loeve model (Gunther and Long, 1994). The Microwave Earth Remote Sensing Laboratory (MERS) received the NSCAT point-wise wind product from the Jet Propulsion Laboratory(JPL) for the nine month NSCAT mission (September 15, 1996–June 29, 1997), and I have successfully demonstrated that wind models can be effectively used to evaluate the quality of NSCAT winds and validate the ambiguity removal algorithms (Gonzales and Long, 1997). I have also developed an operational algorithm for use by JPL and the National Oceanic Atmospheric Administration (NOAA) to evaluate ambiguity removal errors and correct them (Gonzales and Long, 1998a).

The wind characteristics of the NSCAT data for the production of this

operational algorithm have been examined. A number of wind fields have been manually inspected to identify ambiguity removal errors. After segmenting the data into small regions, the models were fit in the least-squares sense to the wind fields over each region. Several statistics were calculated for each of these regions. The set of model parameters was recorded for all the data available as well as RMS fit errors and direction errors.

The statistics of the wind field have been used to establish thresholds for the algorithm. The rms, normalized rms, component, direction fit errors and the model parameters give some indication of the accuracy of NSCAT ambiguity removal. When any of these values are large, the wind field is not realistic. The above method locates the boundaries of the regions that have possible ambiguity removal errors. It is designed as a check of the consistency of the unique wind field. The algorithm is then extended to make corrections in the ambiguity removal and used to assess the accuracy of NSCAT ambiguity removal. This QA algorithm requires only NSCAT data. The nine month NSCAT mission is analyzed with this algorithm. For the set of regions with rms wind speed greater than 4 m/s, the results suggest that NSCAT ambiguity removal is at least 95 % effective. This work has resulted in a significant publication (Gonzales and Long, 1998b).

Specifically, the contributions of this thesis are:

- An operational algorithm to detect errors in the NSCAT point-wise wind product is developed.

- An operational algorithm for correcting ambiguity removal errors is developed.

- The characteristics of different wind fields are examined for the algorithm.

- An assessment of the NSCAT ambiguity removal process is made.

- The algorithms have been provided to operational users at JPL and NOAA.

Because of the importance of accurate wind measurements over the ocean, development of an accuracy assessment and correction algorithm for this data is essential to establish the integrity of the NSCAT data product. Accurate observational data of the winds over the ocean will greatly benefit both meteorologists and oceanographers. The improvement of numerical weather prediction, with the aid of

scatterometer data, will be invaluable to the lives and property of those who use the ocean.

1.4   Overview

Chapter 2 establishes background for this thesis. Some introductory explanation of both point-wise and model-based wind retrieval is provided. The chapter describes the JPL NSCAT point-wise wind product that is analyzed using the QA algorithm. The explanation of model-based wind retrieval provides background information that is useful in understanding the choice of a model-based quality assurance technique.

In Chapter 3, the wind field model used in the QA algorithm is derived from NSCAT data. The choice of this particular wind field model is explained, and a description of how the model is used in the QA algorithm is also provided.

In Chapter 4, a technique is developed to detect and correct ambiguity removal errors using the model derived in Chapter 3. The threshold values used in this technique are described. The technique is applied to the nine month NSCAT mission, and the results are presented.

Finally, Chapter 5 provides the summary and contribution of this thesis. Possible extensions to this work are also presented.

Chapter 2

WIND SCATTEROMETRY BACKGROUND

Since the quality assurance algorithm is used to assess the accuracy of the NSCAT product, a description of that product as well as the technique used to produce it is necessary for an understanding of this thesis. This chapter will provide some of the fundamentals of wind scatterometry and the steps involved in producing a wind field product. First, multiple $\sigma°$ measurements are recorded for each wvc. The Geophysical Model Function which provides a relationship between $\sigma°$ and the ocean wind vector is used to obtain each of the multiple vector wind aliases. The pointwise objective function then ranks each of the aliases in order of probability. Finally, a modified median filter technique is used to choose the alias that is closest to the true wind. Each of these steps will be described in this chapter.

2.1 Principles of Scatterometry

There are a number of principles of wind scatterometry that play a part in generating wind estimates over the ocean. Reviewing these principles is helpful in understanding how JPL produces its NSCAT wind product. These include the interplay between ocean winds and ocean waves, the relationship between ocean waves and radar backscatter and how this backscatter is measured.

2.1.1 Ocean Winds Generate Ocean Waves

When the wind blows across the oceans, small ripples form on the surface of the water as energy is transferred from the wind to the water. These small ripples are called capillary waves. They have wavelengths on the order of a centimeter. Capillary waves can be found superimposed on other types of waves on the oceans as well. Capillary waves are of interest to scatterometry because they are directly related to the winds that cause them. Knowledge about capillary waves from scatterometer measurements provides an indirect method of estimating the winds over the ocean (Naderi et al., 1991; Ulaby et al., 1981a).

6

2.1.2   Ocean Waves Modulate Radar Backscatter

The microwave pulses generated by the scatterometer interact with the ocean surface, and the scatterometer measures the radiation that is scattered back to the scatterometer. Scatterometers take advantage of Bragg scattering or Bragg resonance for capillary waves. Using the basic physics of Bragg resonance, the frequency of NSCAT was chosen to be sensitive to capillary waves. NSCAT was designed at 14 GHz which will be optimal for capillary waves with wavelengths of about 3-8 cm (Naderi et al., 1991).

2.1.3   Radar Equation

When the scatterometer transmits microwave pulses to the oceans' surface, some of this energy will be scattered off the surface. Ocean winds generate ocean waves that modify this backscattered energy. Rougher surfaces scatter more energy back to the scatterometer. The normalized radar backscatter, σ°,, is a measure of this returned energy and is a function of the transmitted and backscattered power. This is quantified in the radar equation (Ulaby et al., 1981b):

$$P_s = \frac{P_t G^2 \lambda^2 A}{(4\pi)^3 R^4} \sigma^{\circ} , \qquad (2.1)$$

where P_t is the power transmitted, P_t is the power backscattered, G is the gain of the transmitting antenna, λ is the wavelength of the electromagnetic wave, A is the effective illuminated area, and R is the distance from the scatterometer to the target.

The backscattered power cannot be measured directly. A measure of returned power, P_r, is taken by the scatterometer which is the sum of backscattered power, P_s, and the noise power, P_N. Thus, a noise only measurement must taken and is subtracted from P_r to estimate P_s.

2.2   NSCAT

The basic principles of scatterometry are combined with the specific design of NSCAT to produce wind estimates. An important design feature of NSCAT is the antenna pattern. NSCAT uses a fan-beam design with an illumination pattern as shown in Figure 2.1. As can be seen from the figure, NSCAT has three antennas on

SUBSATELLITE TRACK

BEAM 6V

BEAMS 5H,5V

45°

65°

135°

BEAM 4V

600 KM

LEFT SWATH

BEAM 1V

45°

115°

BEAMS 2H,2V

135°

BEAM 3V

600 KM

RIGHT SWATH
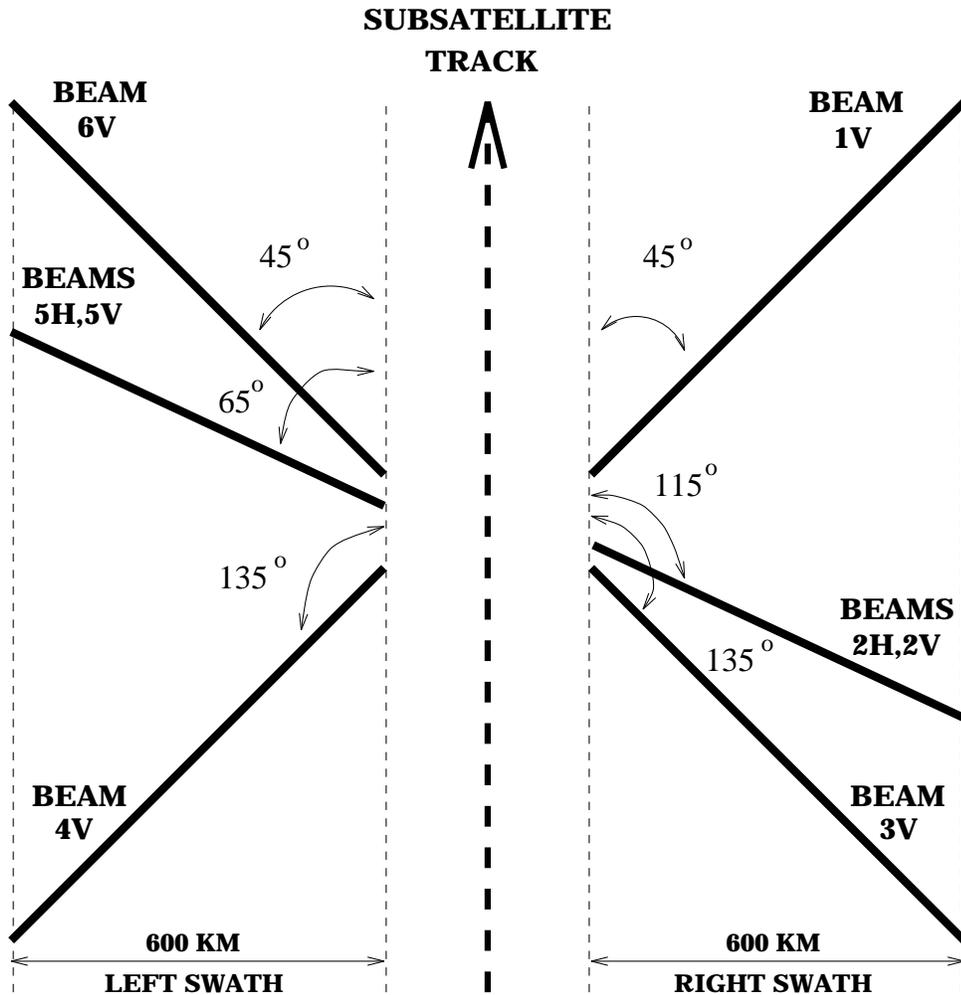
NSCAT

Figure 2.1: *Measurement geometry for the NSCAT satellite.*

each side, one of which is dual polarization. Thus, four measurements are taken on each side of the scatterometer. This will be important in the next section.

Since the satellite moves relative to the surface of the earth, NSCAT uses Doppler processing to achieve the cross-track resolution making it possible to partition the 600 km swath into cross-track cells each having a resolution of 25 km. To achieve along-track resolution, NSCAT must take measurements at intervals corresponding to

a 25 km difference in distance traveled. In this way, the swath is segmented into 25x25 km along-track cells. To improve wind estimation accuracy, these measurements are then grouped into 50x50 km cells over which the wind is estimated. Thus, each side of the swath is 12 cross-track cells wide (Naderi et al., 1991).

2.3   Geophysical Model Function

The relationship between σ° and wind speed and direction is quantified in the Geophysical Model Function. Much research has been done in this area, but a theoretical equation for this relationship has not been developed. Such an equation would require complete knowledge of air/sea interactions and the relationship between electromagnetic radiation and the sea surface.

The flight of Seasat in 1978 made available the millions of σ° measurements for use in the study of an empirical model. These measurements were combined with in situ data such as buoy and ship measurements to establish a Ku-band Geophysical Model Function (Wentz et al., 1984) which can be written as

$$\sigma^\circ = \mathcal{M}(\theta, U, \chi, f, p)$$

where θ is the incidence angle, U is the wind speed, χ is wind direction, f is the frequency, and p is the polarization.

A new model function was developed for the Ku-band NSCAT. This model function was optimized for NSCAT data and is referred to as the NSCAT-1 model function (Freilich and Dunbar, 1998; Wentz and Smith, 1998). This is the model function used for the data analyzed in this thesis.

Figure 2.2 plots σ° versus wind direction, χ for several incidence angles and speeds. From this plot, it can be seen that σ° increases with higher wind speeds, decreases with larger incidence angles, and has a $\cos(2\chi)$ dependence on χ. Since σ° increases with wind speed, the estimation of speed for wind vectors is relatively accurate. However, the double-cosine nature of σ° with wind direction makes it difficult to estimate the direction of the wind vector. Notice the multiple wind directions that could give rise to a σ° measurement of .6 in the upper right hand plot. This suggests a need for multiple σ° measurements for each wind vector cell.

Another way to justify multiple measurements is to plot the wind speeds

9
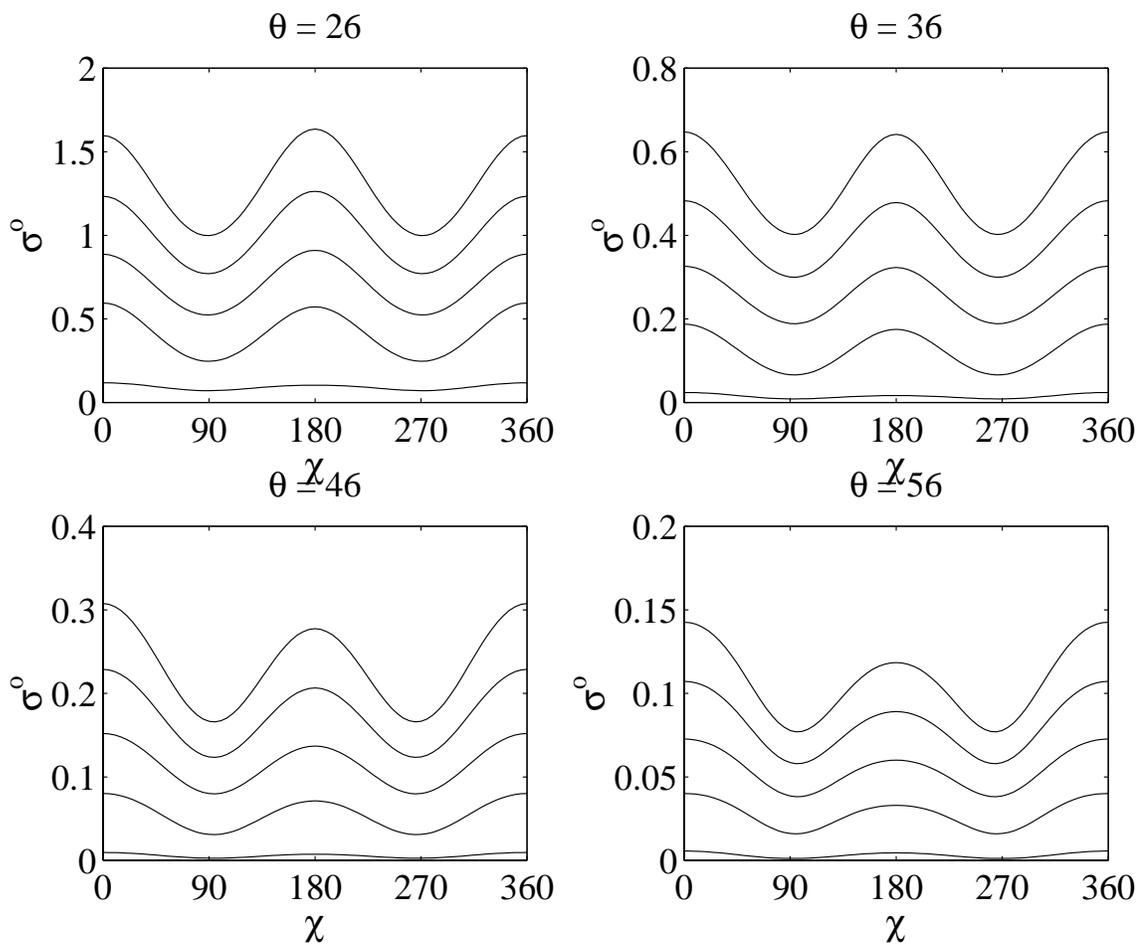
Figure 2.2: *Plots of σ° versus relative wind direction, χ for various incidence angles and speeds. The speeds chosen for are 5, 15, 25, 35, and 45 m/s where σ° is larger for higher wind speeds.*
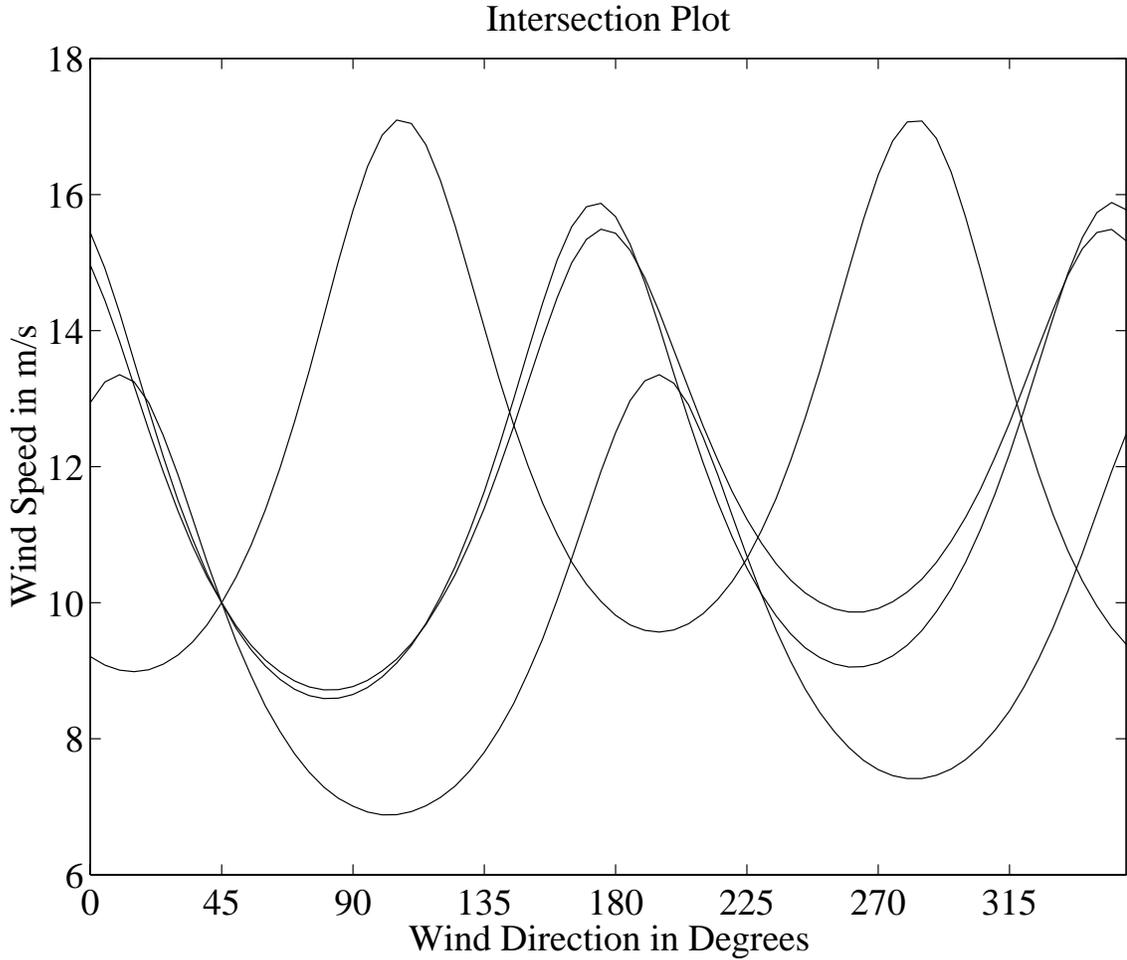
Figure 2.3: *Plots of wind vectors that would support the measurements taken for a particular wind vector cell for NSCAT.*

and directions that give rise to a single σ° measurement according to the Geophysical Model Function. Unfortunately, there are an infinite number of wind speeds and directions that would account the σ° measurement. To combat this problem, NSCAT has effectively four antennas at different azimuth angles from which to make σ° measurements over one wind vector cell. Figure 2.3 plots the four curves that would result from the σ° measurements of a particular wind vector cell for NSCAT. There is a clear intersection at a relative direction of 45°. However, as can be seen from the plot, there are still several near intersections. With the added noise from the system, these near intersections become possible solutions. There are typically between two and four possible wind vectors for every cell.

2.4 Pointwise Objective Function

As mentioned in the previous section, four scatterometer measurements are taken at varying azimuth angles for each wind vector cell. The measurements are denoted z_k, and the vector of measurements, $\mathbf{z} = [z_1, \ldots , z_K]$. These measurements provide between two and four ambiguities following the inversion of the Geophysical Model Function. Each of these ambiguities can be ranked according to probability using classic maximum likelihood techniques.

The joint density of $\mathbf{z}$ given the wind, $\mathbf{w}$, can be written as

$$p_{\mathbf{z}}(\mathbf{z}|\mathbf{w}) = \prod_{k=1}^{K} \frac{1}{\varsigma_{z_k}\sqrt{2\pi}} \exp\left[\frac{-(z_k - \sigma_k^\circ)^2}{2\varsigma_{z_k}^2}\right], \tag{2.2}$$

$$\varsigma_{z_k}^2 = var(z_k) \tag{2.3}$$

$$\sigma_k^\circ = \mathcal{M}(\mathbf{w}, k). \tag{2.4}$$

Thus, given the observations, $\mathbf{z}$, the wind can be estimated by selecting the vector $\hat{\mathbf{w}}$ that maximizes the above density function. This is classic maximum likelihood:

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}} p_{\mathbf{z}}(\mathbf{z}|\mathbf{w}). \tag{2.5}$$

For NSCAT, we maximize the log-likelihood which is equivalent:

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}} L(\mathbf{w}, \mathbf{z}), \tag{2.6}$$

$$L(\mathbf{w}, \mathbf{z}) = \log p_{\mathbf{z}}(\mathbf{z}|\mathbf{w})$$
$$= -\sum_{k=1}^{K}\left\{\frac{[z_k - \sigma_k^\circ]^2}{2\varsigma_{z_k}^2} + \frac{1}{2}\log[\varsigma_{z_k}^2] + \frac{1}{2}\log(2\pi)\right\}. \tag{2.7}$$

Since σ_k° is found empirically, this optimization is done numerically.

$L(\mathbf{w}, \mathbf{z})$ typically has several local maxima because of the symmetry in the Geophysical Model Function discussed earlier. In fact, there are two prominent local maxima and several smaller local maxima. These correspond to the near intersections visible in Fig. 2.3. The values at each of these maxima are used rank the likelihood of each ambiguity.

2.5 Median Filter Technique

The correct ambiguity must be distinguished with the use of an ambiguity removal technique. A number of ambiguity removal schemes have been developed

including field-wise estimation (Long, 1993; Hoffman, 1982), and various ad hoc techniques (Wurtele et al., 1982). NSCAT uses a modified median filter technique to remove ambiguous wind vectors (Naderi et al., 1991).

Median filters in image processing define the median of N nearby values as $(N + 1)/2$ for N odd to replace the center value. They are advantageous over linear filters in many applications, because extremely large or small values do not affect the median and edges are preserved. For wind images, this becomes important. However, traditional median filters for image processing applications cannot be used for ambiguity removal because each resolution element contains two values, speed and direction. A modified median filter technique has been developed for the purposes of ambiguity removal (Shaffer et al., 1991; Shultz, 1990).

As described in (Shaffer et al., 1991), the median for vector data $V(i)$ is the vector $A(m)$ which minimizes $E(m)$ where $E(m)$ is defined as

$$E(m) = \sum_{i=1}^{N} \| A(m) - V(i) \| \quad and \quad 1 \leq m \leq N. \tag{2.8}$$

This equation is further modified to accommodate two-dimensional data, the likelihood values of each of the ambiguities, and the filter window.

The median filter is implemented by passing an NxN window (N is odd) over the data. Each of the elements of the filter can have different weights which affect the contribution of each element. For NSCAT, the size of the filter is 7x7 with equal weights on each of the elements. For each location (i, j) of the swath, there exists a set of k ambiguities. Each ambiguity has an associated likelihood value, L_{ij}^k, as described in the previous section.

Each swath is considered separately and must be initialized by one of the ambiguities for each location. This initial field is an array of vectors U_{ij} selected from the A_{ij}^k ambiguities, typically with A_{ij}^1 which is the most likely ambiguity. Without following all of the details of the derivation in (Shaffer et al., 1991), the ambiguity A_{ij}^k at location (i, j) which minimizes E_{ij} is substituted for the center vector. E_{ij}^k is defined as

$$E_{ij}^k = (L_{ij}^k)^{-2} \sum_{m=i-3}^{i+3} \sum_{n=j-3}^{j+3} \| A_{ij}^k - U_{mn} \| \tag{2.9}$$

where the U_{mn} are each of the vectors surrounding the center vector. The filter is

moved over the swath and the entire process is repeated until convergence is reached (Shaffer et al., 1991; Naderi et al., 1991).

There are two techniques for initializing the swath. The first technique is to initialize by the most probable ambiguity. Figure 2.4 is an example of the median filtering technique on a region initialized by the most probable ambiguity. The most probable ambiguity is the closest to the actual wind approximately 60% of the time. However, local skill (the percentage of the most probable wind vectors that are closest to the actual wind in a local area) can be lower than this value. This sometimes causes the median filter to propagate ambiguity removal errors (Naderi et al., 1991; Shaffer et al., 1991). Consequently, a second technique for initializing the swath was developed. Since one of the two most probable ambiguities is closest to the actual wind approximately 90% of the time (Naderi et al., 1991), data from global surface analysis fields from the National Center for Environmental Prediction (NCEP) is used to determine which of the two most probable ambiguities with which to initialize the swath (Freilich and Dunbar, 1998). The wind product produced using these two techniques are referred to as "unnudged" and "nudged" data respectively.

Median filters are ideal for removing random selection errors while keeping the dominant features of the wind. Consequently, the initial field for the median filter must contain a majority of the true features of the wind. Thus, the nudging technique is preferred over the unnudged technique.

In summary, pointwise wind retrieval is the traditional method of estimating the wind from scatterometer measurements. The first step is to obtain the four scatterometer measurements for each cell location. The second step is to invert the Geophysical Model Function for each of these measurements and find the two to four intersections. The third step is to rank each of these four solutions using maximum likelihood techniques. Finally, the modified median filter is applied to produce the final wind product.

2.6   Model-based wind retrieval

Model-based wind retrieval is another method for estimating the winds over the oceans using scatterometer data. Model-based wind retrieval assumes that the wind is correlated, i.e., that it is spatially consistent. By imposing this correlation on the structure of the wind, the wind can be retrieved over an area instead of by

All Aliases
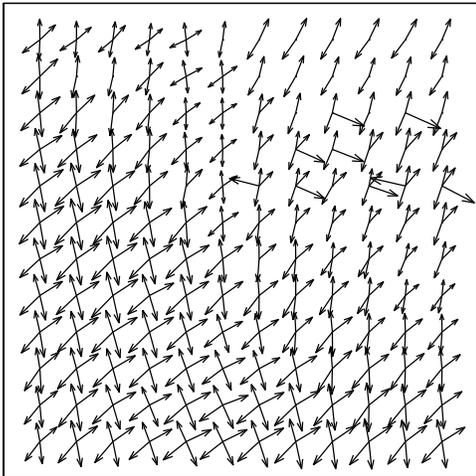
Selected Wind
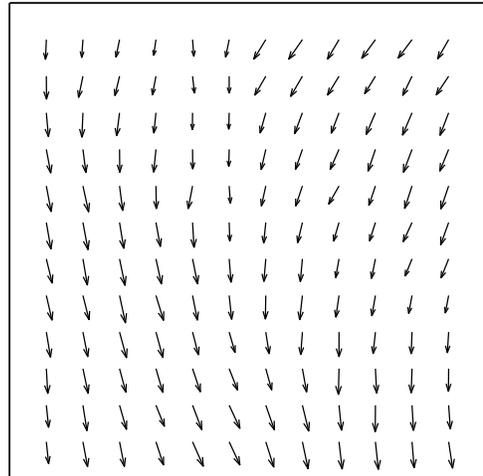
Figure 2.4: *An example of the median filter technique on a region initialized by the most probable ambiguity for ascending revolution number 2454.*

15

individual elements.

The model-based technique is a recent development that improves on some of the problems associated with point-wise ambiguity removal (Long, 1993). The model is linear and can be expressed as

$$W = FX$$

where X is a vector containing the model parameters and F is a constant model matrix described in more detail in the next chapter. Instead of estimating the wind, W, the must smaller vector of parameters, X, is estimated and the wind obtained (Long, 1993).

Model-based wind retrieval can yield more accurate results than point-wise wind estimation. In addition, as demonstrated in this thesis, the wind field model can be used as a quality assurance test on point-wise ambiguity removal. This is the focus of this work and is described in detail in Chapters 3 and 4.

Chapter 3

WIND FIELD MODELS

In this chapter, a wind field model is developed which is used to assess the accuracy of NASA Scatterometer (NSCAT) ambiguity removal. A technique is then developed to use this model for the quality assurance of NSCAT ambiguity removal.

3.1   The Wind Field Model

The data used for the NSCAT ambiguity removal assessment is the NASA Jet Propulsion Laboratory (JPL) Level 2.0 product for the NSCAT mission (Naderi et al., 1991). The wind resolution is 50 kilometers. The scatterometer makes wind observations over a dual-sided swath that is 600 kilometers or 12 wind vector cells (wvc's) on each side. Figure 3.1 is an example section of the observed wind field produced by JPL for the ascending revolution number 847. Two data sets were examined, each processed with the NSCAT-1 geophysical model function (Freilich and Dunbar, 1998; Wentz and Smith, 1998) which was tuned to NSCAT data. The same maximum likelihood wind retrieval technique is used for both data sets. The data contains up to four ambiguities per cell ranked by maximum likelihood with a flag to indicate the ambiguity selected by JPL. Point-wise ambiguity removal has been performed on the data using a median filter technique (Shaffer et al., 1991). In median filtering, each swath is initialized separately by the most probable alias for "unnudged" data and by global surface analysis fields from the National Center for Environmental Prediction (NCEP) for "nudged" data (Freilich and Dunbar, 1998). In the "nudged" processing, data from NCEP is used to select which of the two most probable solutions are used to initialize the swath for implementing the median filter (Freilich and Dunbar, 1998). The unnudged data exhibits more obvious ambiguity removal errors than nudged data.

3.1.1   Determination of the model

As mentioned, wind field models can be used to assess the accuracy of ambiguity removal algorithms. In this work, a linear model similar to Long (Long,

17
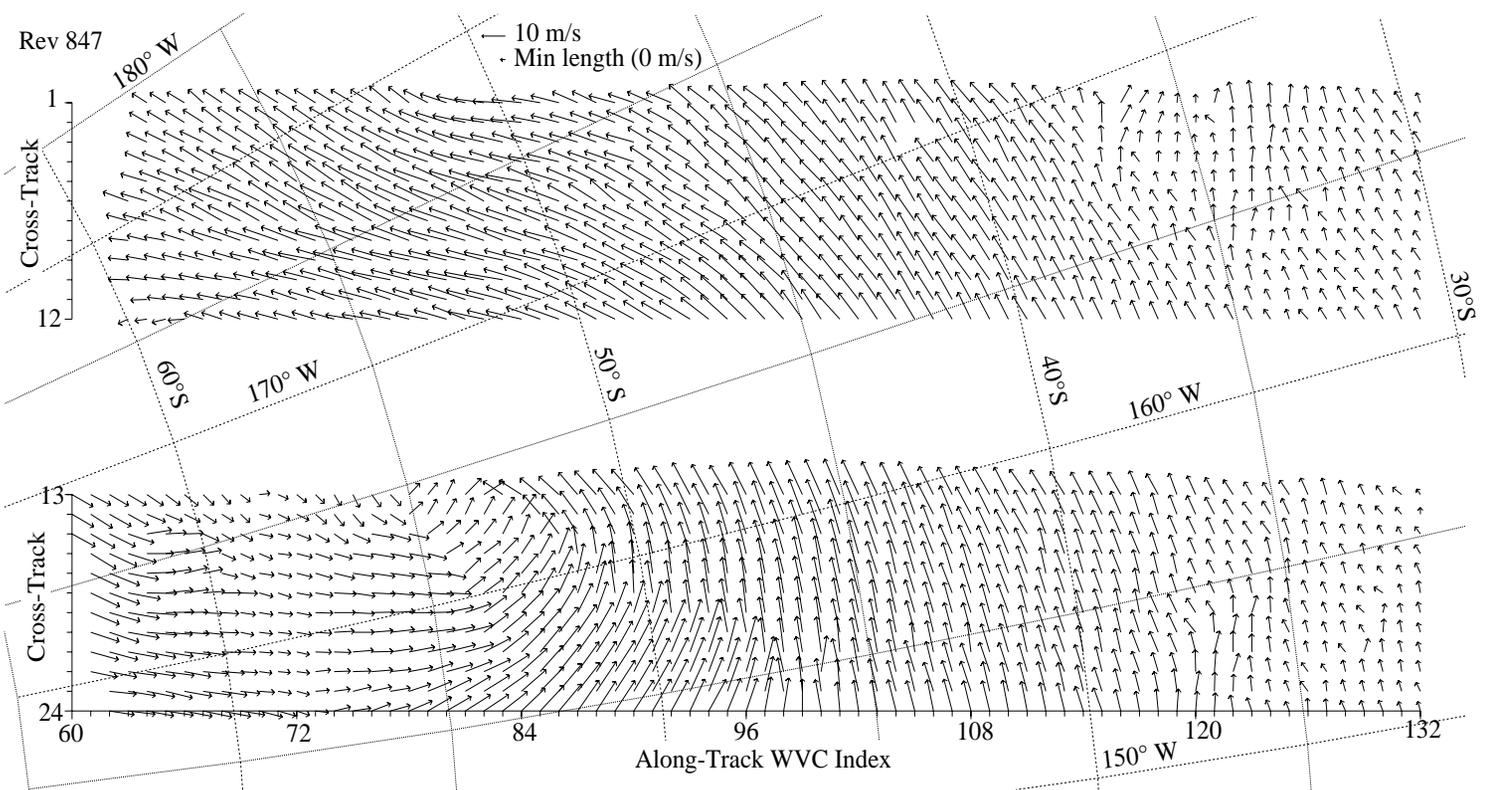
Figure 3.1: A sample wind field over the Pacific Ocean from the nudged JPL product for the ascending rev number **847**. A minimum vector length has been used in this figure to clarify the presentation of very low wind speed vectors.

1993) is used. It can be expressed as

$$W = FX$$

where X is an L-element vector containing the model parameters and F is a constant model matrix where the rows of F form a basis set for possible wind fields. W is a row-scanned vector of winds sampled at the scatterometer observations over a small (12wvc x 12wvc) section of the swath. For each 12x12 region, W is defined in terms of the components of the wind:

$$\mathbf{W} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}, \tag{3.1}$$

where U is a row-scanned version of the 12x12 matrix of east components of the wind and V is a row-scanned version of the 12x12 matrix of north components of the wind. For both U and V, the rows vary with cross-track and the columns vary with along-track.

While (Long, 1993) used a simple dynamics-driven model for F, in this paper we adopt a data driven model matrix with a minimum number of basis vectors. We use the Karhunen-Loeve (KL) model since it is known to minimize the basis restriction error (Jain, 1989).

The KL model matrix F is derived from the eigenvectors of the autocorrelation matrix R of the sampled wind field (Jain, 1989; Long, 1993). R is defined as $E[\mathbf{WW^T}]$. Since R is not known, it must be estimated from the sample autocorrelation. While a sample correlation could be computed from global circulation models (e.g., ECMWF or NCEP), these models are low resolution in comparison to the 50 kilometer NSCAT resolution. Instead, the point-wise wind estimates for NSCAT data are used to compute an estimate of R.

A portion of three weeks (128 revs) is used to estimate the sample auto-correlation. Each swath is segmented into 12x12 overlapping regions (approximately 53,000 regions) and W is determined for each of the regions. The estimate of R is then the sample average of the autocorrelation matrix:

$$\mathbf{R} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{WW^T}$$

where N is the number of regions.

19

Using standard eigenvalue/eigenvector decomposition methods, the model matrix F is formed as the lower subset of the sorted eigenvectors of the sample autocorrelation matrix. The eigenvectors corresponding to the largest eigenvalues are the most important and are used as the columns of F. Eigenvectors with very low eigenvalues describe wind field components that are relatively rare or less important. Plots of the eigenvalues and the model-fit difference are useful for determining where to truncate the eigenvector series (see Fig. 3.2 and Fig. 3.3). Visible in the eigenvalue plot, Fig. 3.2, are some natural breakpoints, and the similarity between the unnudged and nudged data sets is apparent. Figure 3.3 shows the model-fit difference versus the number of basis vectors in the model. It was generated by fitting the model to nudged NSCAT data and calculating the vector rms difference. In this paper, the model matrix was subjectively chosen as the first 22 basis vectors of F for the tradeoff between modeling error and the ability to locate regions with ambiguity removal errors. We note, however, that there is little performance difference in the QA algorithm when truncating the model between basis vectors 20 through 30.

3.1.2 Model Basis Vectors

The truncated KL model is only minimally dependent on which data set is used to generate it even though the unnudged winds contain many more ambiguity removal errors than the nudged data set. Separate KL models were computed for left and right swath and nudged and unnudged JPL products. The first few basis vectors are essentially identical for all cases. The basis vectors beyond the truncation point are the least important and have little effect on the truncated model. The truncated model admits basis vectors that describe the common wind fields which are essentially the same for nudged and unnudged data.

The basis vectors corresponding to the first few eigenvalues are of interest as they mirror common natural wind fields. Figure 3.4 is a plot of the first six basis vectors for the KL model. The two most important basis vectors (i.e. those with the largest eigenvalues) correspond to the mean wind. The importance of these two basis vectors is evident from the large break in the eigenvalue plot between these and the subsequent eigenvalues. The next four also are representative of common wind patterns. The fourth and sixth are representative of cyclonic flows. The third and fifth are both examples of col points. As the eigenvalues for these wind fields suggest
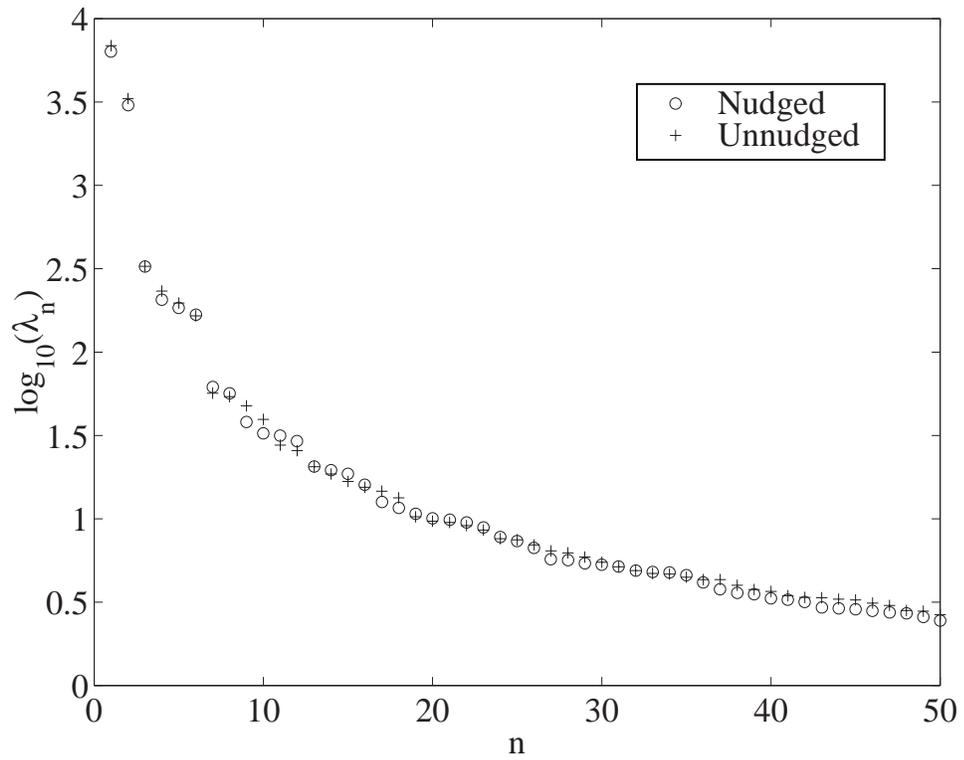
Figure 3.2: Eigenvalues of the sample autocorrelation matrix computed from 128 revs of nudged and unnudged data.
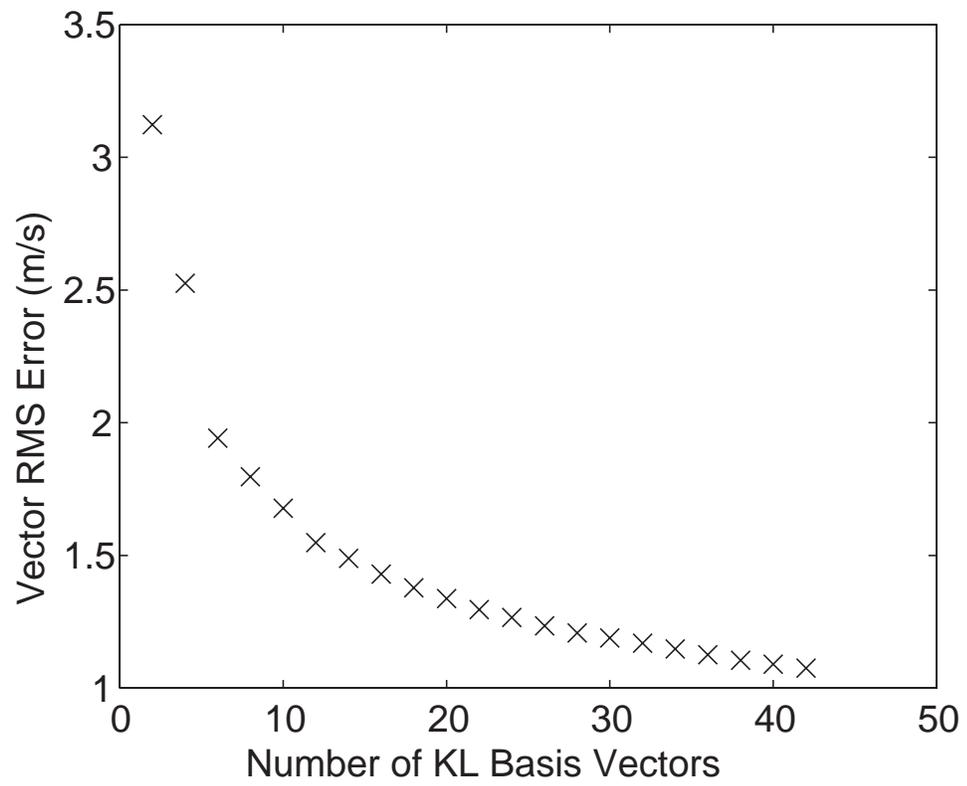
Figure 3.3: Vector RMS error versus the number of KL basis vectors (only even numbers shown) for the 128 rev nudged test data set.
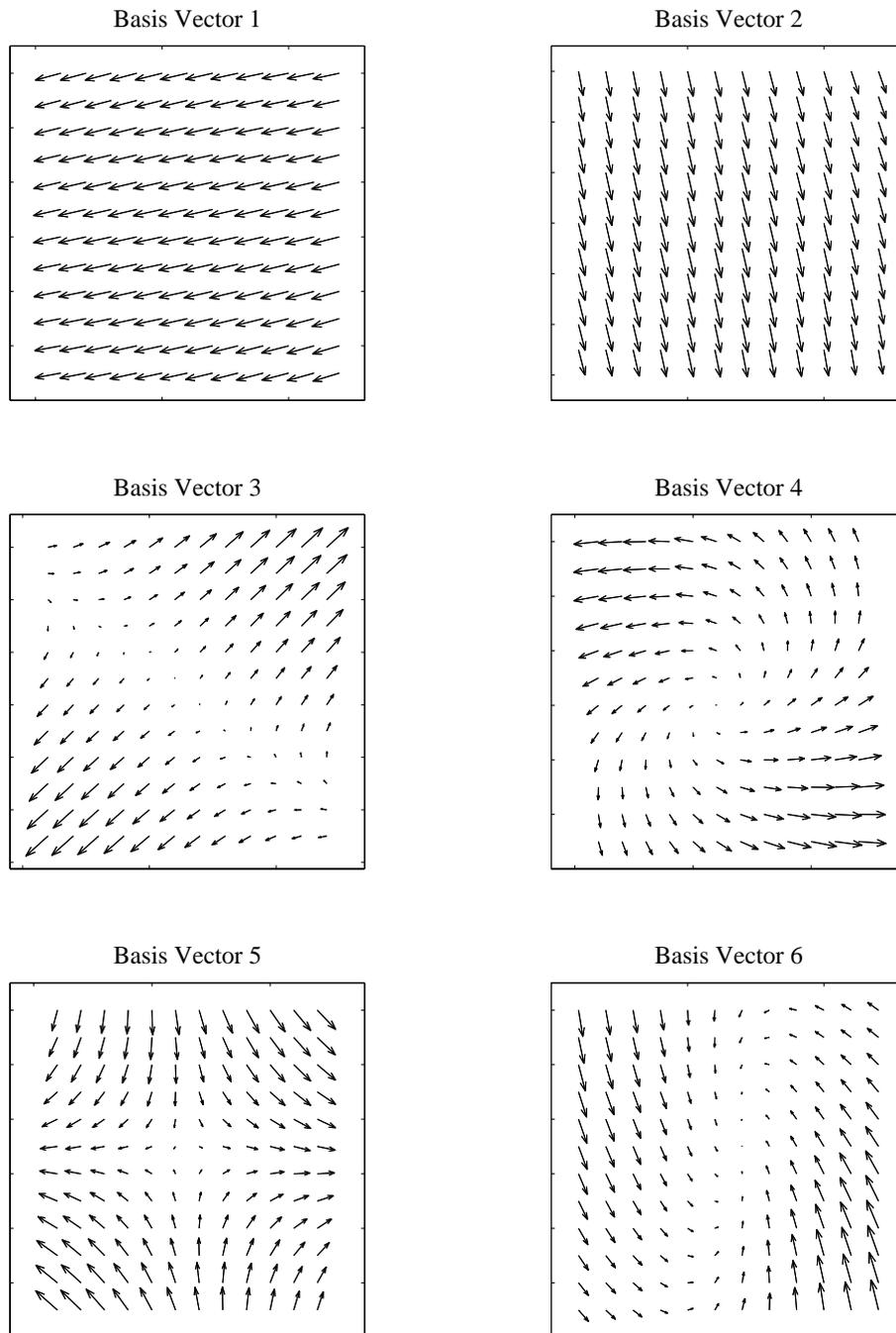
Figure 3.4: The first six basis vectors of the truncated KL model.

(Fig. 3.2), these basis vectors are fundamental and are the bases for most wind fields. The remaining basis vectors are provided in Appendix A.

The truncated model is effective in spanning the majority of common wind fields since wind fields have a red power spectrum, i.e., the low frequency components have the most energy. If the entire model matrix were used, any sampled wind field could be fit exactly to the model. However, by truncating the model, it can be used to identify regions of ambiguity removal errors since, while most realistic wind fields are spanned by the truncated model, fields with ambiguity removal errors are not.

Unfortunately, truncating the model does make some realistic wind fields inadmissible since not all real wind fields are adequately described by only the first 22 basis vectors. This "modeling error" can be significant for some wind fields. As discussed later, modeling error can be confused with ambiguity removal errors; this is a key limiting factor in our approach.

3.2   Methodology

To use the model as a quality assurance for the point-wise wind retrieval, the model is fit in a least-squares sense to the observed point-wise wind field as described in this section. The swath is segmented into overlapping sections and the model-fit is tested for each section. The difference in the fit provides information about the "realism" of the observed wind.

3.2.1   Using the Model-fit

A least-squares estimate of the model parameter vector X, $\underline{X}$, can be obtained from the observed wind field W_0 using the pseudo-inverse of F, $F^{\dagger}$, i.e., $\underline{X}=F^{\dagger}W_0$. The reconstructed wind field W_R, also known as the model-fit field, is $W_R = F\underline{X}$ with the reconstruction difference field W_E given by

$$W_E=W_R\text{-}W_0=(FF^{\dagger}\text{-}I)W_0.$$

If the reconstruction difference is small then the model-fit is good and the observed wind field is considered "realistic" according to the model. Large differences are attributed to possible ambiguity removal errors and flagged. However, the difference can also be affected by noise in the wvc estimate or modelling error.

Selected Wind   Model Fit

All Aliases   Difference Field

↗ 10 m/s
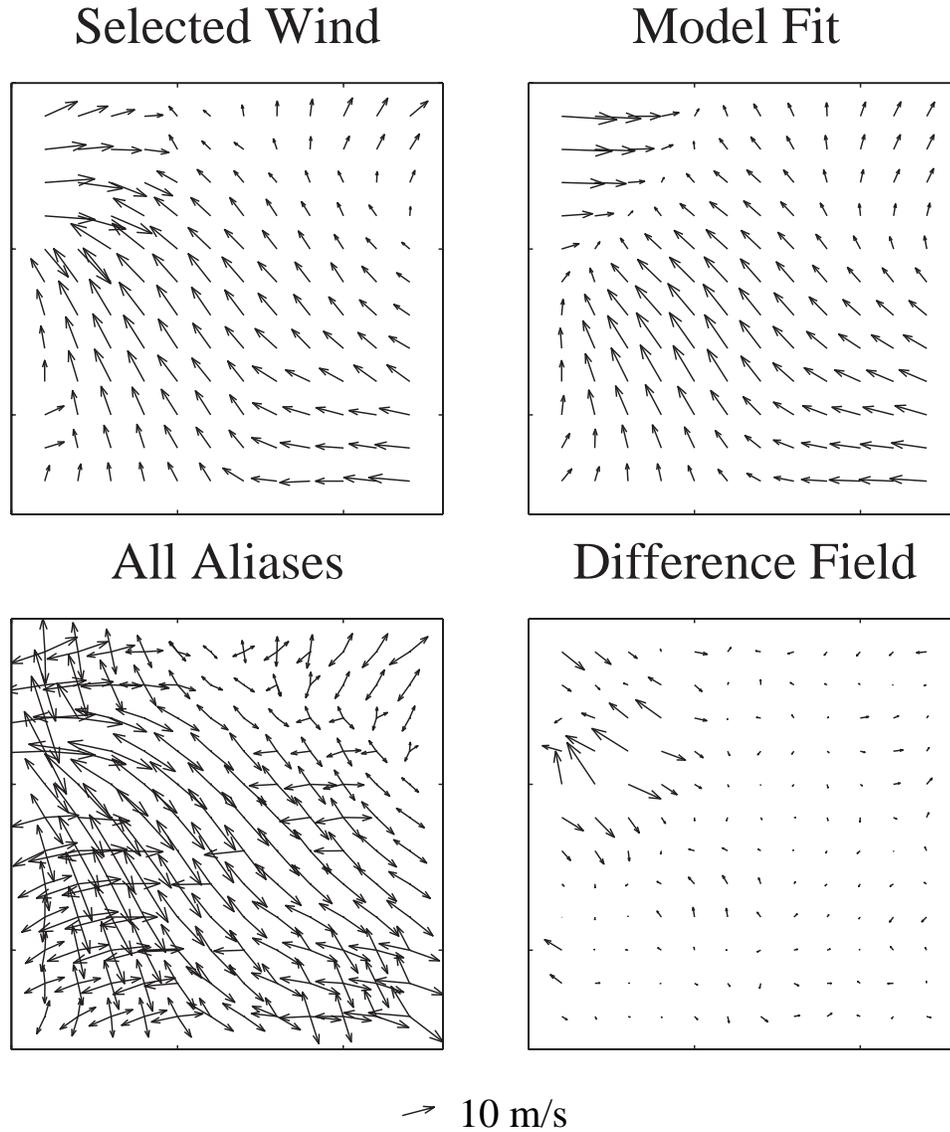
Figure 3.5: A wind field that exhibits a significant area of ambiguity removal errors in the upper left corner. The wind is spatially inconsistent in the upper left corner of the region. This is evident in the difference field where large differences between the selected wind field and the model-fit field are observed. Because of the number of large values in the difference field, this region is classified as poor by the QA algorithm.

To illustrate, Fig. 3.5 is a region with clear ambiguity removal errors in the upper left corner. The model-fit field exhibits large differences at some locations which correspond to the boundary of the ambiguity removal error region. By finding these areas of significant wind error in the model-fit, ambiguity removal errors are identified.

There are a number of considerations when implementing this simple technique. First, the model must be fit to the wind field over a region. To produce an adequate fit, the input wind must be defined over the full region. Thus, for this simple algorithm, only those regions with fewer than eight cells of land or missing measurements are used. Since the reconstruction difference field becomes larger with increasing numbers of missing measurements, the threshold of eight cells was chosen as a conservative estimate. The missing measurements are replaced with the average of the cells surrounding it and then processed. Second, the wind field model inherently smoothes the wind field over the entire region due to modeling error; the model matches the general flow of the wind, but may not adequately model the center of a cyclone or the boundary of a front. Such regions can be flagged as containing errors, because the modeling error is large. Third, the difference in the model-fit can be high in regions where the wind estimates are very noisy even if ambiguity removal is correct. Thus, the region may be flagged as having possible ambiguity removal errors even if the ambiguity removal is correct. Fourth, it is possible for both the JPL field and the model-fit field to be incorrect for a given region though it is impossible to detect this sort of occurance with only NSCAT data. Finally, at low wind speeds, the wind is highly variable, resulting in significant modeling error which is further complicated by the low signal to noise ratio in these regions. Manual ambiguity removal is also very difficult in such regions. As a result, we are unable to verify the ambiguity removal accuracy for low wind speed regions.

Figure 3.6 illustrates one such low wind speed region. Figure 3.7 demonstrates a region which is not represented well by the model-fit. Neither of these regions is spatially consistent and results in large reconstruction differences. While the model-fit seems to show the flow of the wind for this region, it is not clear that it is representative of the actual wind for this region. In such regions, it can be difficult to verify the ambiguity removal accuracy because of the inherent uncertainty between modeling error and ambiguity removal error.

Selected Wind

All Aliases

→ 10 m/s
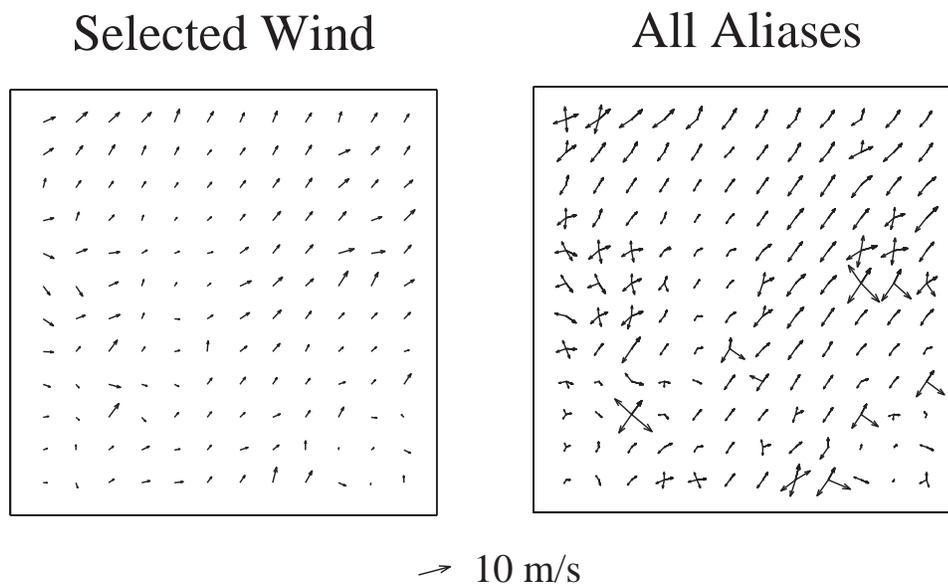
Figure 3.6: A region of low wind speed that is classified as poor by the QA algorithm.

Selected Wind   Model Fit

All Aliases   Difference Field

⇀ 10 m/s
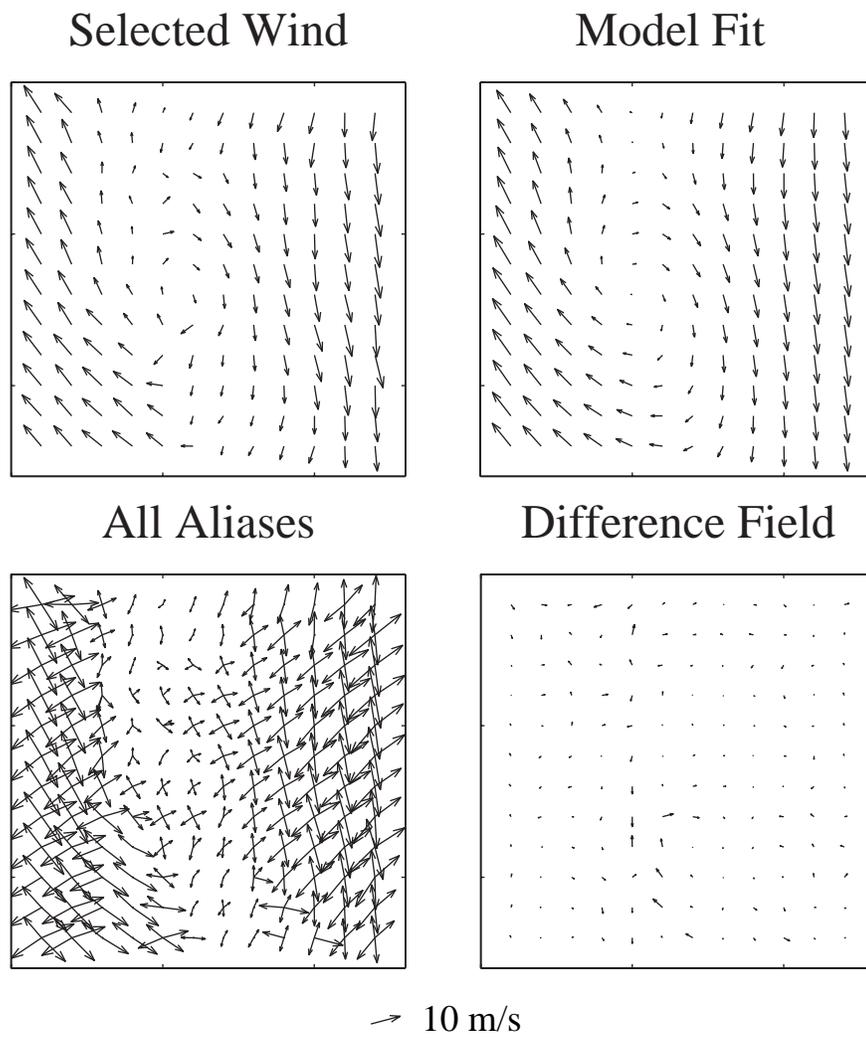
Figure 3.7: A region which is not represented well by the model-fit and is flagged as poor.

Chapter 4

QUALITY ASSESSMENT

Using the results from the previous chapter, a technique is developed to detect and correct ambiguity removal errors. As mentioned in the previous chapter, to use the model as a quality assurance for the point-wise wind product, the model is fit in a least-squares sense (Section 3.2.1) to the observed point-wise wind field. The difference in the fit provides information about the "realism" of the observed wind. Thresholds are found for the model-fit, and regions with statistics exceeding these thresholds are flagged as containing possible ambiguity removal errors. Corrections are then made when possible. The results of using this technique on the data of the NSCAT mission is then presented.

4.1   Selecting Thresholds

To use the model-fit to locate regions with possible ambiguity removal errors, a set of thresholds on the model parameters and the reconstruction difference field are determined in the following. These thresholds are used to classify the quality of the ambiguity removal for each region. A technique for correcting the identified errors is presented and the description of the detection and correction algorithm is given.

To select the thresholds for the model parameters, a histogram of each parameter is examined. Figure 4.1 shows the histograms of four of the parameters for the K-L model using 5488 regions of NSCAT data (6 days from 3 weeks). The rev numbers used to produce these histograms is provided in Appendix B. Manual testing has shown that large values for any of the model parameters correspond to regions with possible errors. After some examination of the values for the parameters, the thresholds are set at twice the standard deviation for each of them. This provides an initial starting place for subjectively altering these numbers as needed to correctly identify error-prone regions. Only a few of the model parameters are necessary to identify regions of possible ambiguity removal errors. Since the columns of F for the KL model are basis vectors in decreasing order, only the first few parameters are
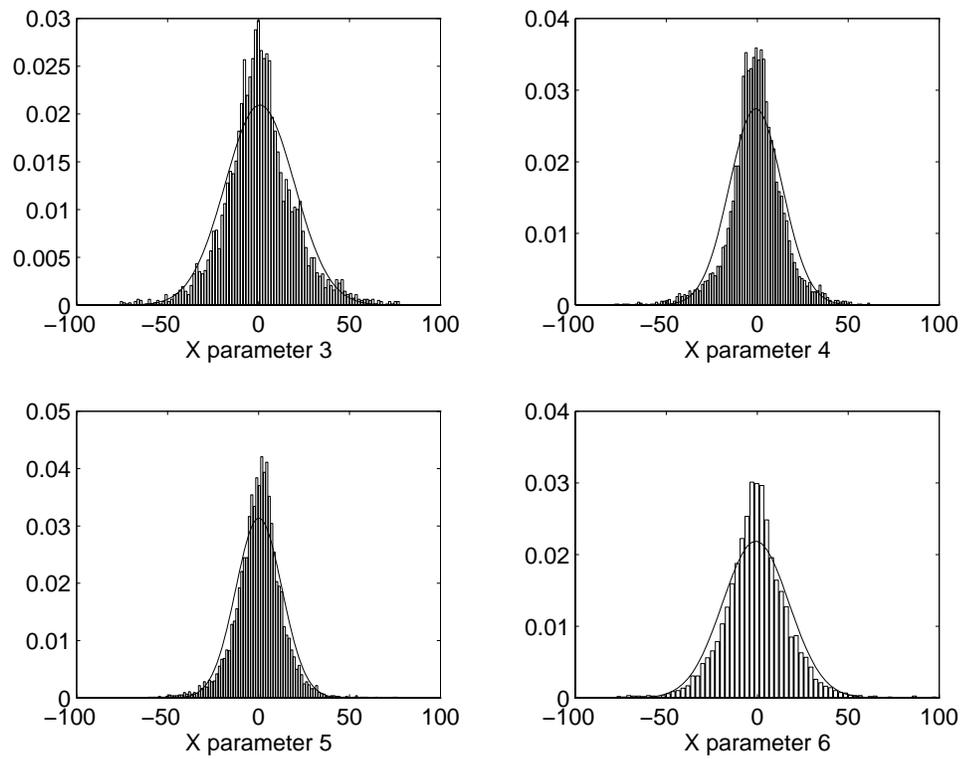
Figure 4.1: Histograms for parameters three through six for the K-L model. Overlaid is a Gaussian distribution with the same mean and variance.

used as thresholds for the QA algorithm. Table 4.1 lists the parameters used for this algorithm and their values. Appendix C and D provide some of the statistics for the parameters and describe how to choose their thresholds.

K-L model (22 parameters)
X(3) = 31.3
X(4) = 31.7
X(5) = 37.9
X(6) = 34.9
X(9) = 16.7

Table 4.1: The X parameter thresholds for the KL model.

The other thresholds for locating ambiguity removal errors are determined from the reconstruction difference field. These thresholds include the rms error, the normalized rms error, the maximum component error, and the maximum direction error for each region. The rms error is found by summing the squared components of the reconstruction difference field, dividing by the number of terms, and taking the square root. The normalized rms error is found by squaring the components of the reconstruction difference field, dividing by the sum of the squared components of the observed wind field, and taking the square root, i.e., nrms $= \sqrt{\frac{\mathbf{W_E}^T\mathbf{W_E}}{\mathbf{W}^T\mathbf{W}}}$. The rms and normalized rms errors aid in locating regions of large error. Both of these values are calculated for the entire region and thus provide information about the region as a whole. The maximum component and maximum direction error values are useful for locating regions in which only a few of the wind vectors are incorrect. The individual errors are identified by finding those that exceed either of these thresholds. These wind vectors are flagged as possible ambiguity selection errors, though as discussed before, the error may exceed the thresholds due to noise, modeling error, or ambiguity removal error.

To select the threshold values for this algorithm, 3309 regions (32 randomly

selected revs) of NSCAT data were manually inspected. The rev numbers are listed in Appendix E. The regions were subjectively grouped into four categories: "perfect" (no errors), "good" (those with only a few isolated ambiguity removal errors), "moderate" (as much as 10% but less than 20% of the wvc's identified as possible ambiguity selection errors), and "poor" (more than 20% of wvc's identified as possible errors). Figures 4.2 and 4.3 are examples of each of these categories. All of the poor regions either have low rms wind speeds making the region difficult to model or have subjectively identified areas of significant ambiguity removal errors. While the possibility of a poor region classification due to modeling error exists, it was not observed in this data set. For this data set, 77% of the poor regions were low wind speed regions (rms speed less than a subjectively chosen threshold of 4 m/s). All of the remaining (with rms speed greater than 4 m/s) were regions with subjectively identified areas of significant ambiguity removal errors. The statistics of each region were calculated and compared to the initial two sigma thresholds. The thresholds were adjusted such that the maximum number of poor, moderate, and good regions are correctly identified as containing ambiguity removal errors with a minimum number of false alarms. Table 4.2 shows each of these thresholds and their means and standard deviations.

	RMS Error (m/s)	Normalized RMS Error	Maximum Component Error (m/s)	Maximum Direction Error
mean	0.88	0.17	0.65	7.38°
σ	0.43	0.10	0.27	6.11°
threshold	0.96	0.26	2.7	23°

Table 4.2: The error thresholds for the KL model and their means and standard deviations.

After this tuning, the algorithm correctly identified 100% of the poor and moderate regions and over 99% of the good regions with a false alarm rate of less than 3% on the tuning data set. Note that the thresholds can be altered to adjust the

Selected Wind   All Aliases

Selected Wind   All Aliases
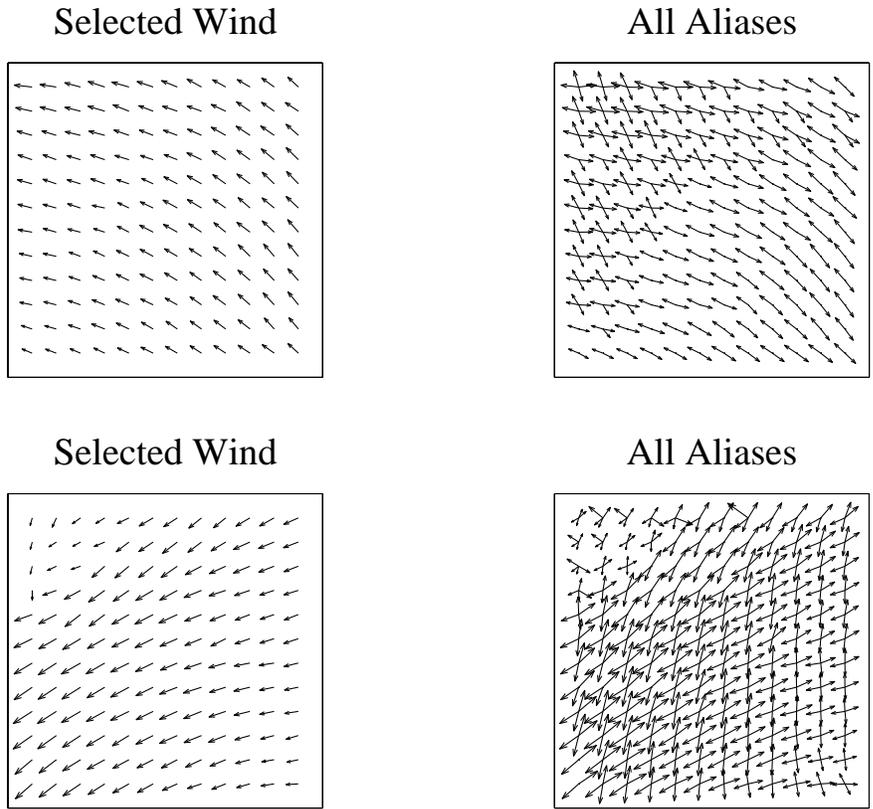
Figure 4.2: The top two wind fields demonstrate the selected and all ambiguity plots for regions classified as perfect. The bottom two wind fields demonstrate the selected and all ambiguity plots for regions classified as good.

Selected Wind

All Aliases

Selected Wind

All Aliases
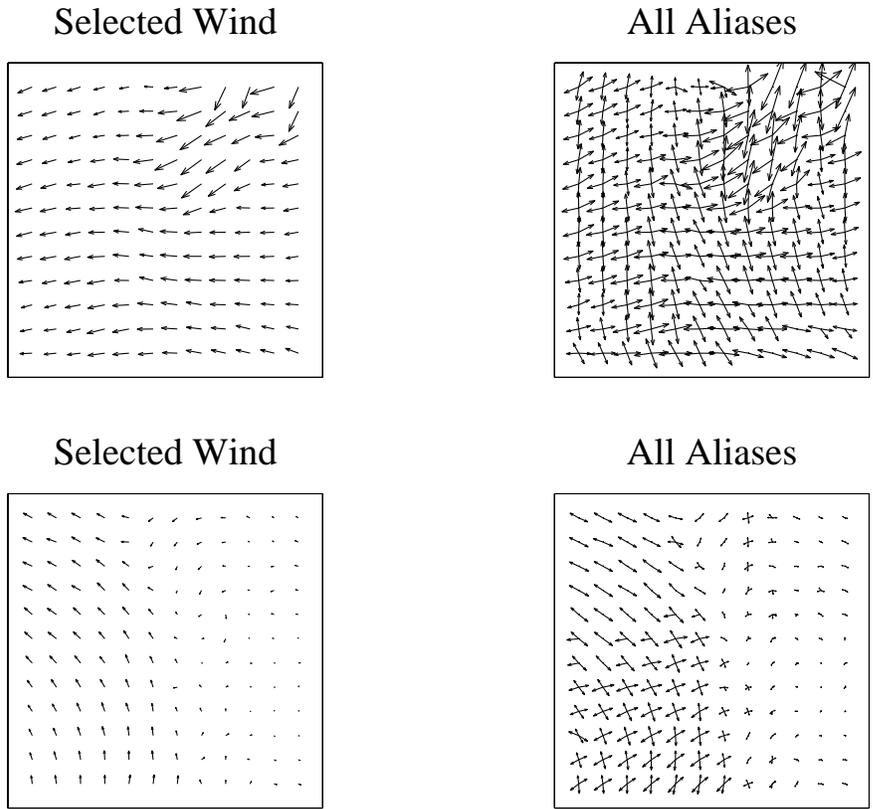
Figure 4.3: The top two wind fields demonstrate the selected and all ambiguity plots for regions classified as moderate. The bottom two wind fields demonstrate the selected and all ambiguity plots for regions classified as poor.

detection and false alarm probabilities since the thresholds are a trade-off between detection and false alarms.

The thresholds chosen for the detection algorithm were tested on a manually classified withheld data set of 1561 regions (16 revs) and achieved a similar level of performance. The rev numbers used for the withheld data set are provided in Appendix F. The algorithm correctly identified 100% of the poor and moderate regions and over 98% of the good regions with a false alarm rate of less than 4%. Combining the statistics for these two data sets results in total detection rate of more than 98% for all regions subjectively identified as containing ambiguity removal errors with less than 4% of the perfect regions misidentified. Thus, though modeling error or noise will sometimes result in an incorrect evaluation of a region as containing possible errors, the vast majority of regions with possible ambiguity removal errors are located using this technique. The classification performance of low wind speed regions was also consistent with the previous results. Regions with low (< 4 m/s) rms wind speeds accounted for 76% of the poor regions with the remaining regions (with rms wind speeds greater than 4 m/s) all containing significant areas of ambiguity removal errors.

Regions with possible errors are then tested for consideration in the correction algorithm in which wind vectors are examined individually. For vectors identified as possible ambiguity removal errors, the point-wise alias closest in direction to the model-fit is chosen as the corrected wind. Since the aliases typically have similar speeds but different directions, the speed field remains similar, but the direction field is more consistent with the model-fit for corrected wind fields. Figures 4.4 demonstrates the use of the correction algorithm. As can be seen, the observed wind product contains several ambiguity removal errors. The algorithm chooses the alias that is closest in direction to the model-fit field, producing a subjectively more realistic corrected wind field. Thus, the model-fit is a reasonable basis for both detecting ambiguity removal errors and correcting at least some ambiguity removal errors.

However, as mentioned, the model does not adequately fit some wind fields, or as a result of significant ambiguity errors, the original wind field cannot be determined with confidence using the model. Thus the model-fit field cannot be used to attempt to correct these wind fields. The number of possible ambiguity removal errors is used as a criterion for determining when a region is a candidate for the correction

Selected Wind

All Aliases

Model Fit

Corrected Wind

10 m/s
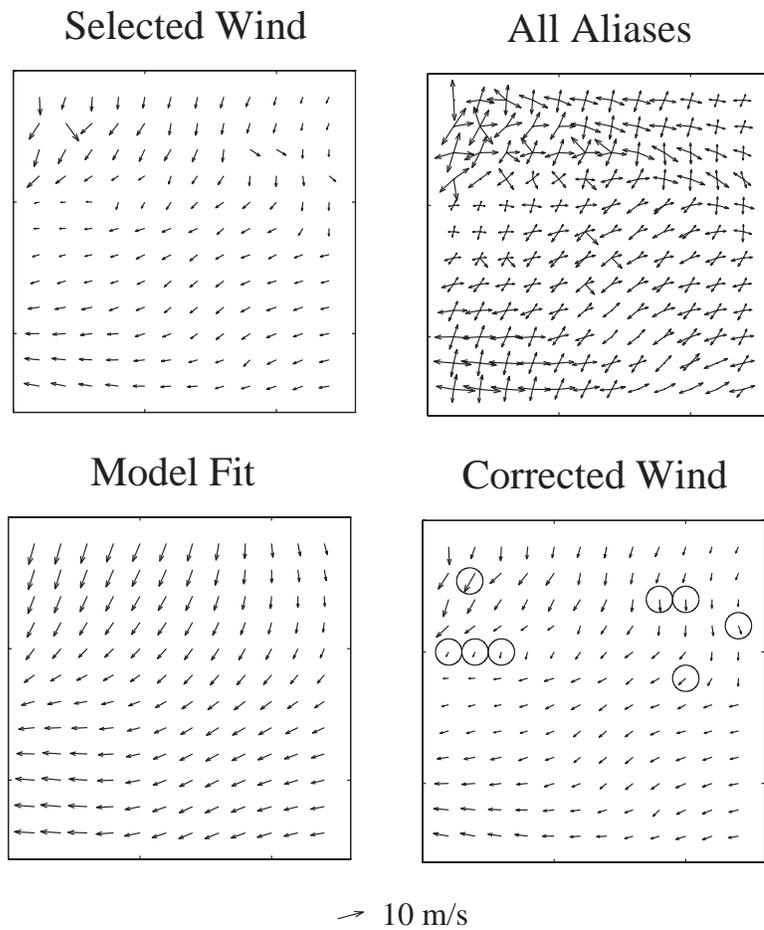
Figure 4.4: A sample corrected wind field. The circled vectors are those that were changed according to the method described in the text.

algorithm. The number of possible ambiguity removal errors, i.e., the number of wvc's flagged by the QA algorithm as having large differences with the model, is determined by those wind vectors that exceed either the maximum angle or maximum component error thresholds. If the total number of errors for a region exceeds this threshold, the region is not considered a candidate for the correction algorithm. The selection of the threshold was determined by trial and error. For this implementation, only regions classified as good or moderate (i.e., with 20% or fewer possible errors) are considered candidates for ambiguity selection correction.

The criteria for an ambiguity removal correction of a wvc is thus extremely conservative. Further, though the vector may be identified as being potentially in error (due, perhaps, to a noisy wind vector estimate), many times the alias closest in direction to the model-fit is, in fact, the original wind vector and thus no change is made. For example, in Fig. 4.4, even after attempted correction, a few of the wind vectors still appear quite noisy and as a result, are still flagged by the algorithm as possibly incorrect even though no better directional ambiguity can be found.

4.2   Analysis

After the algorithm was tuned with ten revs of NSCAT data, the entire nine month nudged NSCAT mission data set was processed to assess the accuracy of NSCAT ambiguity removal. The results were consistent with the results already presented herein for the observation subset used to develop the model. Of 408,069 regions examined, 24% of the regions were classified as "perfect", 41% as "good", 17% as "moderate", and 18% as "poor" where the categories are described in Section 3.2.

For regions classified as perfect, good, or moderate (82% of the total), only 4% of the individual vectors were identified as possible ambiguity removal errors; however, only approximately 10% of these vectors were changed using the model-based correction technique. For the remaining, the ambiguity closest in direction to the model-fit was the original wind vector. Thus, only 0.4% of the individual vectors were corrected using this approach. This result suggests that NSCAT ambiguity removal is thus over 99% effective for these regions.

Figure 4.5 summarizes key statistics for regions (18% of the total) classified as "poor". Of these poor regions, 74% of them have rms speed values of 4 m/s or

37
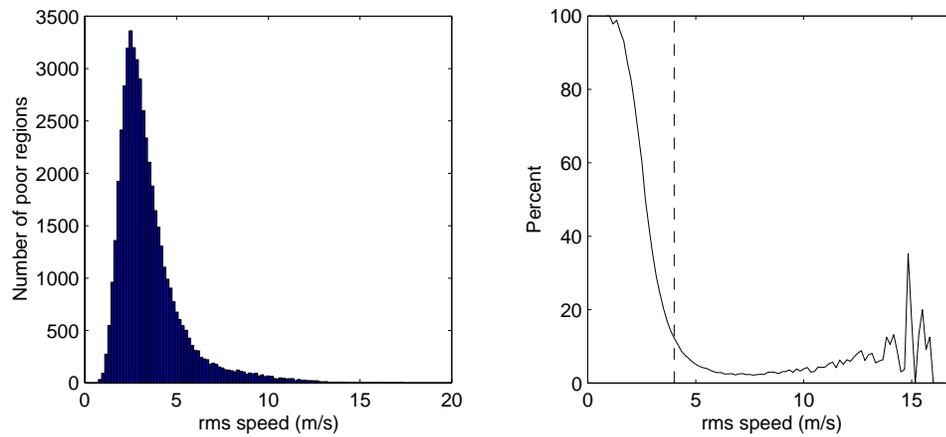
Figure 4.5: (left) Histogram of the rms speed for all regions classified as "poor" in the nine month NSCAT mission. (right) The percent of the total regions which are classified as "poor" at each rms wind speed bin. The vertical dashed line is at 4 m/s.

less, and we are unable to verify the ambiguity removal accuracy due to the difficulty of modeling low wind speed regions and the noise level at low wind speeds. The poor regions with rms wind speeds greater than 4 m/s contain significant ambiguity removal errors. Such regions represent less than 5% of the total number of regions. We note from Fig. 4.5 that not all regions with rms wind speeds less than 4 m/s are rated poor: for rms winds speeds increasing from 2 to 4 m/s, the percentage of regions with a given rms wind speed which are rated poor decreases from 100% to 10%, and all regions with an rms wind speed less than approximately 2 m/s are poor. This is consistent with a low wind speed cutoff in the geophysical model function such as that proposed by (Donelan and Pierson, 1987) who suggested that below a temperature-dependent wind speed threshold of 3 to 5 m/s at Ku-band, depending on incidence angle, the normalized radar cross section falls off rapidly. Such a roll off would decrease the signal-to-noise ratio and reduce the accuracy of the wind estimates.

Although 5% of the total number of regions have large ambiguity removal errors, portions of these regions contain no errors. Since we cannot uniquely resolve corrections using only NSCAT data and this simple technique for these high wind speed (> 4 m/s) poor regions, a conservative approach is to treat each wind vector in the region as a possible ambiguity removal error. Combining this with the previous result of almost complete effectiveness for non-poor regions, we conservatively conclude that, based only on NSCAT data, the effectiveness of NSCAT ambiguity removal is 95% or better for the entire set of regions with rms wind speeds of 4 m/s or greater. This result is consistent with the comparisons with European Centre for Medium-Range Weather Forecasts (ECMWF) winds and the buoy collocation statistics presented by (Freilich and Dunbar, 1998) and (Wentz and Smith, 1998).

The accuracy of NSCAT ambiguity removal is evaluated as a function of time during the mission in Fig. 4.6. This figure shows the percent of non-poor regions as a function of time. There is an apparent slight decrease in the accuracy of NSCAT ambiguity removal over the mission. To understand this effect, the ambiguity removal is evaluated over several Pacific Ocean latitude bands as defined in Fig. 4.7. Figure 4.8 summarizes some of the statistics over the five latitude bands. The expected variation of wind speed with latitude is clearly evident. There is a strong correlation between the ambiguity removal performance and the rms wind speed, with reduced overall ambiguity removal performance (i.e., more poor regions) at lower wind speeds
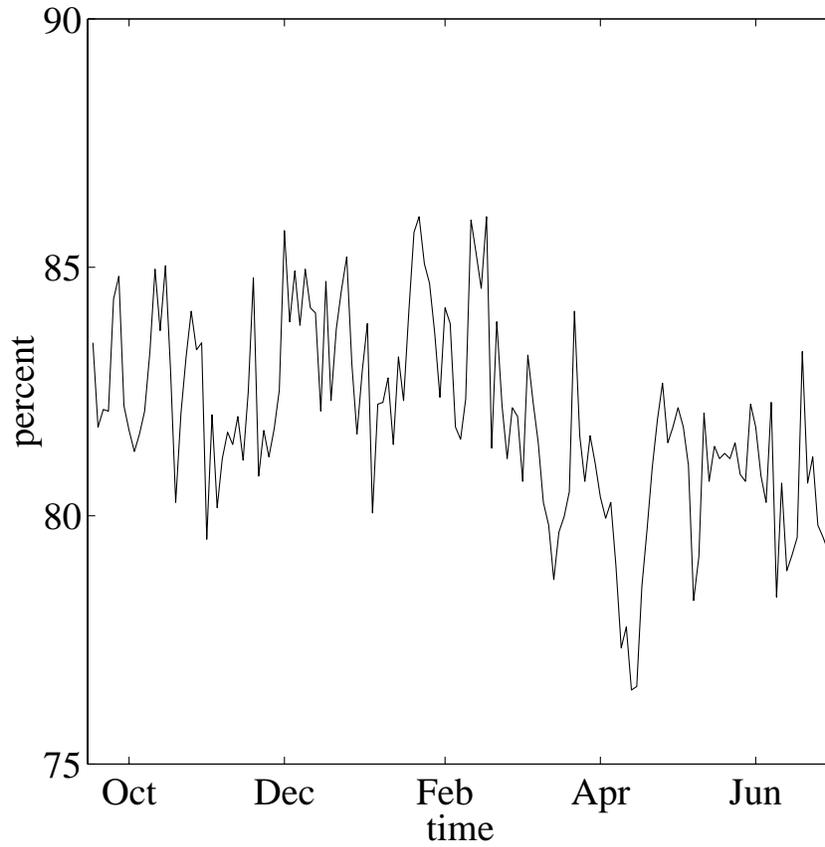
Figure 4.6: The percent of non-poor regions versus time over the nine month NSCAT mission. Each point represents the average computed over approximately two days.
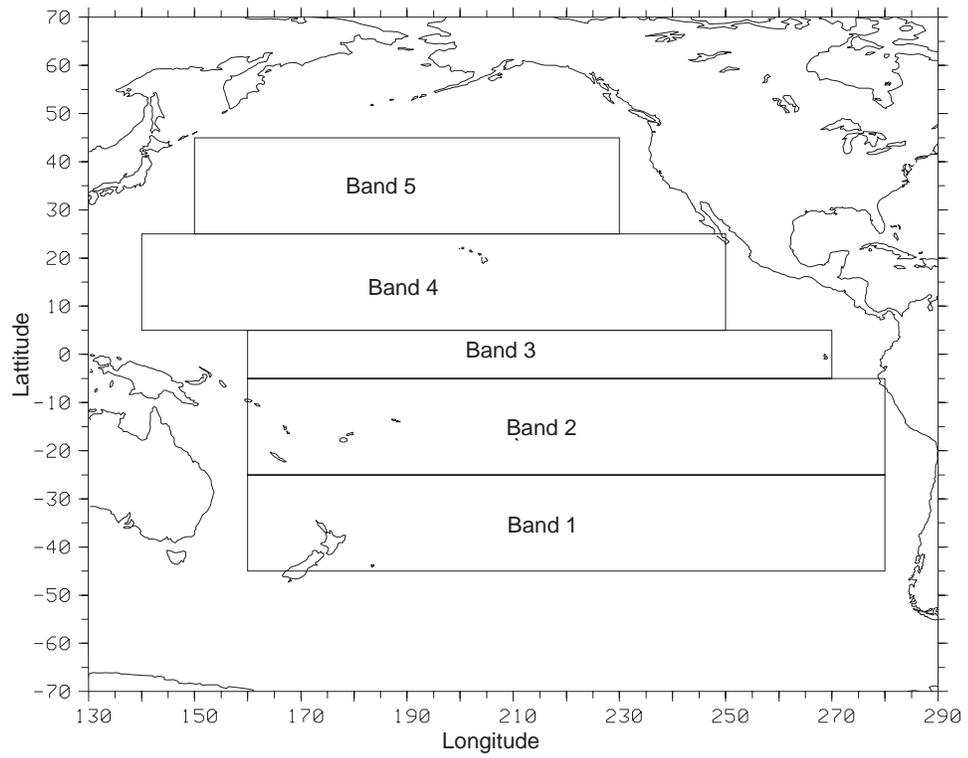
Figure 4.7: Geographical lattitude bands in the Pacific.

(See Appendix G). Thus, the wind speed distribution in each band affects the ambiguity removal performance and seasonal changes in the wind speed distribution results in temporal variations in the ambiguity removal performance. In particular, increased storm activity in the Northern Hemisphere results in increased wind speed with improved ambiguity removal during the winter months in Bands 4 and 5. Similarly, the number of poor regions increases during the Southern Hemisphere summer due to a decrease in the rms wind speed. The peak in the percentage of high wind speed poor regions in Band 1 corresponds to early winter in the Southern Hemisphere, a time of large storms in this region. Because of its low rms wind speed, Equatorial Band 3 is the most sensitive to changes in the mean rms wind speed with a significant drop in the percent of non-poor regions corresponding to a small drop in the rms wind speed at the start of 1997.

4.3   Summary and Conclusions

In summary, the steps of the algorithm to detect and correct ambiguity removal errors are:

1. Segment the swath into 12x12 overlapping regions with 50% along track (6 wvc's) overlap.

2. For each valid region (regions with fewer than eight missing measurements or land cells), compute the model-fit field W, the reconstruction error field W_E, the model parameter vector $\underline{X}$, and the statistics of W_E. These statistics include the rms error, the normalized rms error, the maximum component error, and the maximum direction error for each region.

3. For each region, determine if the statistics, including those for the model parameter vector $\underline{X}$, are larger than the thresholds. If so, the region is identified as containing possible ambiguity selection errors. Based on the number of possible errors identified for each region, segregate the regions into 4 classes ("perfect", "good", "moderate", and "poor") .

4. For those regions not classified as "poor", correct the ambiguity removal error by choosing the ambiguity closest in direction to the model-fit for those wvc's identified as possible errors.
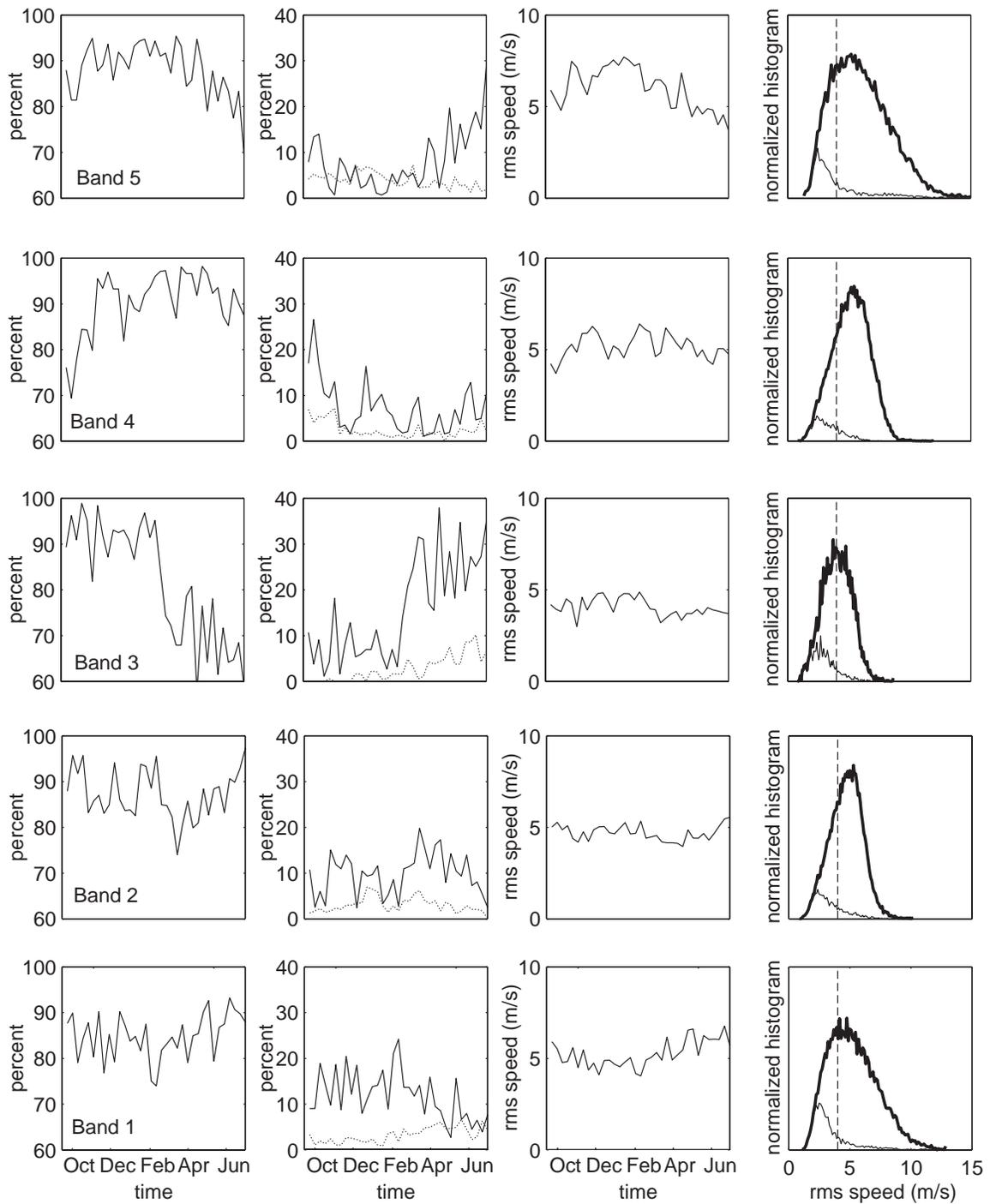
Figure 4.8: (left) Percentage of non-poor regions as a function of time over the NSCAT mission. (left, middle) Percentage of poor regions with an rms wind speed greater than (solid) and less than (dotted) 4 m/s. (right, middle) Average regional rms wind speed as a function of time. (right) Normalized histograms of (bold) all regions and (light) those classified as poor by the QA algorithm. The vertical dashed line is at 4 m/s.

In conclusion, using only NSCAT data the QA algorithm works very well in identifying regions with possible selection errors. The technique allows rapid processing of the data set. Ambiguity removal errors in good or moderate regions can be corrected with a high degree of confidence. Using this technique over the ten month NSCAT mission, the NSCAT ambiguity removal is found to be better than 95% effective for entire set of regions with rms wind speeds greater than 4 m/s.

Chapter 5

CONCLUSIONS

NSCAT estimates the wind speed and direction of near-surface ocean wind. Several possible wind vectors are estimated for each wind vector cell. Typically, the speeds of the possible wind vectors are nearly the same, but the directions are very different. The correct wind must be distinguished in a step called ambiguity removal. Unfortunately, ambiguity removal algorithms are subject to error. As a step toward evaluating the accuracy of the JPL NSCAT product, I have used a new model-based quality assurance algorithm which uses only NSCAT data.

5.1   Summary

The quality assurance algorithm presented in this thesis provides a technique for detecting and correcting ambiguity removal errors. The QA algorithm was used to evaluate the accuracy of NSCAT ambiguity removal.

In summary, the QA algorithm segments the swath into overlapping 12x12 wvc regions and classifies each region according to estimated quality. The nine month NSCAT mission dataset is analyzed. In 82% of the regions, the ambiguity removal is over 99% effective with the ambiguity errors correctable using a model-based correction technique. In 5% of the regions, areas of significant ambiguity error are found. For remaining regions, all of which have root mean square (rms) wind speeds less than 4 m/s, there is too much uncertainty in the wind field model or too much noise in the measurements to uniquely evaluate ambiguity selection with sufficient confidence. I thus conservatively conclude that for the set of regions with rms wind speed greater than 4 m/s, NSCAT ambiguity removal is at least 95% effective.

The accuracy of NSCAT appears to decline over the life of the mission. This is most likely a seasonal effect. Since the NSCAT mission was ended prematurely by the failure of ADEOS, the seasonal effect hypothesis could not be completely validated. However, the latitude analysis supports this view. It is clear that low wind speeds result in lower accuracy levels for NSCAT ambiguity removal. Since lower wind speeds are dominant in the summer months, the seasonal changes in the wind

distribution correspond to the temporal changes in the skill of NSCAT ambiguity removal.

5.2   Contributions

For this thesis, the KL model was extensively researched for nudged and unnudged NSCAT data. Many swaths of wind data were analyzed subjectively and the characteristics of the wind in relation to this algorithm were thoroughly studied. This study suggested a method for detecting regions with possible ambiguity removal errors by establishing a set of thresholds as described in Chapter 3. The technique was perfected and applied to NSCAT data. A technique to correct ambiguity removal errors was developed based on the spatial correlation between wind vectors. The two methods together form the QA algorithm. This algorithm was then applied to the entire NSCAT mission and the accuracy of NSCAT ambiguity removal was evaluated.

Thus, a key contribution of this thesis is the algorithm provided to operational users at JPL and NOAA to detect and correct ambiguity removal errors. The second major contribution is that this algorithm is used to provide an assessment of the NSCAT ambiguity removal process. A significant publication has resulted from this research. This algorithm uses only NSCAT data and is quickly applied to the data. The entire NSCAT dataset can be processed in approximately half a day. Scatterometer data is typically provided at near real-time speeds. Thus, this algorithm would be easily added to current schemes to improve the quality of wind estimation.

5.3   Future Research

The QA algorithm would benefit from future research and can be used with other scatterometers. Logical extensions for this algorithm include application to the QuikScat and Seawinds scatterometers, the addition of phenomenological models, more testing for lower wind speed regions, an objective method of choosing the model order, and additional work regarding a low wind speed cutoff for scatterometry.

As mentioned in chapter 3, the primary limitation to this technique is modeling error. Unfortunately, modeling error frequently occurs in regions with interesting features such as fronts and cyclones. The performance of this algorithm can be improved with the addition of phenomenological models that can accurately model the winds of these important features.

The reported accuracy for all regions of rms wind speeds greater than 4 m/s is 95%. Some reports suggest that the majority of ambiguity removal errors occur for winds between 4 m/s and 6 m/s. The impact of the correction technique is most likely greatest for this range of wind speeds. Since the mean wind speed of winds over the ocean is 7 m/s, many wind estimates fall into this category. Further testing to report these statistics would be helpful to the scientific community.

The choice of model order for this algorithm was based on subjective analysis of the tradeoff between modeling error and the ability to locate regions with ambiguity removal errors. Studies of model sensitivity or an optimizations technique might be helpful in establishing an objective method of choosing model order.

Finally, the low wind speed cutoff proposed by (Donelan and Pierson, 1987) is very interesting when combined with the results of this thesis. Though this work could not validate their conclusions, more research in this area would be a logical extension to this work.

BIBLIOGRAPHY

Atlas, R., "The impact of preliminary NSCAT data on ocean surface analysis and numerical weather prediction." Tech. rep., NASA Scatterometer Science Working Team Meeting, 1997.

Atlas, R., A. J. Busalaci, M. Ghil, S. Bloom, and E. Kalnay, "Global surface wind flux fields from model assimilation of Seasat data." Journal of Geophysical Research, vol. 92, no. C6, pp. 6477–6487, 1987.

Donelan, M. and W. Pierson, "Radar scattering and equilibrium ranges in wind-generated waves with application to scatterometry." Journal of Geophysical Research, vol. 92, pp. 4971–5029, 1987.

Freilich, M. and R. Dunbar, "The accuracy of NSCAT-1 vector winds: Comparisons with NDBc buoys." Journal of Geophysical Research, 1998, submitted.

Gonzales, A. E. and D. G. Long, "A Quality Assurance Algorithm for NASA Scatterometer Wind Retrieval." In Proceedings of SPIE, vol. 3117, pp. 107–114, 1997.

Gonzales, A. E. and D. G. Long, "An Algorithm to Assess the Accuracy of NASA Scatterometer Data." In International Geoscience and Remote Sensing Symposium (IGARSS), 1998a, submitted.

Gonzales, A. E. and D. G. Long, "An assessment of NSCAT ambiguity removal." Journal of Geophysical Research, 1998b, to appear.

Gunther, J. and D. Long, "Models of the near-surface oceanic vorticity and divergence." In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), vol. 2, pp. 951–953, 1994.

Hoffman, R. N., "SASS wind ambiguity removal by direct minimization." Mon. Weather Rev., vol. 110, pp. 434–445, 1982.

Jain, A. K., Fundamentals of Digital Image Processing. Prentice-Hall, 1989.

Long, D., "Wind field model-based estimation of Seasat scatterometer winds." Journal of Geophysical Research, vol. 98, no. C8, pp. 14651–14688, 1993.

Long, D. G. and J. M. Mendel, "Identifiability in wind estimation from scatterometer measurements." IEEE Trans. on Geoscience and Remote Sensing, vol. 29, pp. 268–276, 1991.

Naderi, F., M. Freilich, and D. Long, "Spaceborne radar measurement of wind velocity over the ocean–An overview of the NSCAT Scatterometer system." Proc. of the IEEE, vol. 79, no. 6, pp. 850–866, 1991.

Oliphant, T. E., New techniques for wind scatterometry. Master's thesis, Brigham Young University, 1996.

Schroeder, L. C., W. L. Grantham, E. M. Bracalente, C. L. Britt, K. S. Shanmugam, F. J. Wentz, D. P. Wyle, and B. B. Hinton, "Removal of ambiguous wind directions for a Ku-band wind scatterometer using measurements at three different azimuth angles." IEEE Transactions on Geoscience and Remote Sensing, vol. GE-23, no. 2, pp. 91–100, 1985.

Shaffer, S., R. Dunbar, S. Hsiao, and D. Long, "A median-filter-based ambiguity algorithm for NSCAT." IEEE Transactions on Geoscience and Remote Sensing, vol. 29, no. 1, pp. 167–174, 1991.

Shultz, H., "A circular median filter for resolving directional ambiguities retrieved from spaceborne scatterometer data." Journal of Geophysical Research, vol. 95, pp. 5291–5303, 1990.

Ulaby, F. T., R. K. Moore, and A. K. Fung, Microwave Remote Sensing, vol. 2. Artech House Inc., Norwood, MA, 1981a.

Ulaby, F. T., R. K. Moore, and A. K. Fung, Microwave Remote Sensing, vol. 1. Artech House Inc., Norwood, MA, 1981b.

Wentz, F., S. Peteherych, and L. Thomas, "A model function for ocean radar cross sections at 14.6 GHz." Journal of Geophysical Research, vol. 89, no. C3, pp. 3689–3704, 1984.

Wentz, F. and D. Smith, "A model function for ocean normalized radar cross section at 14.6 GHz derived from NSCAT observations." Journal of Geophysical Research, 1998, submitted.

Wurtele, M. G., P. M. Woiceshyn, S. Peteherych, M. Borowsky, and W. S. Appleby, "Wind direction alias removal studies of Seasat scatterometer-derived wind fields." Journal of Geophysical Research, vol. 87, pp. 3365–3377, 1982.

Appendix A

KL BASIS VECTORS 7 THROUGH 22
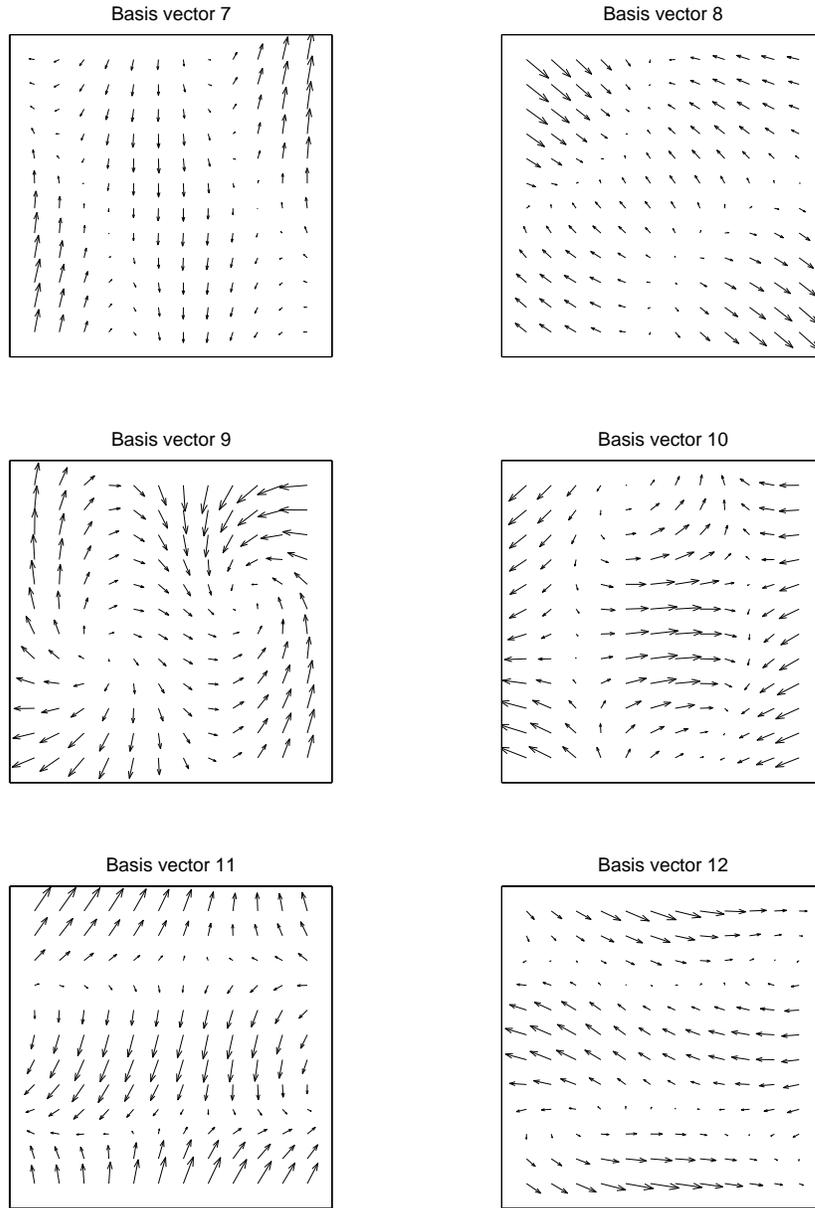
Figure A.1: Basis vectors 7 through 12 of the truncated KL model.

Basis vector 13

Basis vector 14

Basis vector 15

Basis vector 16

Basis vector 17
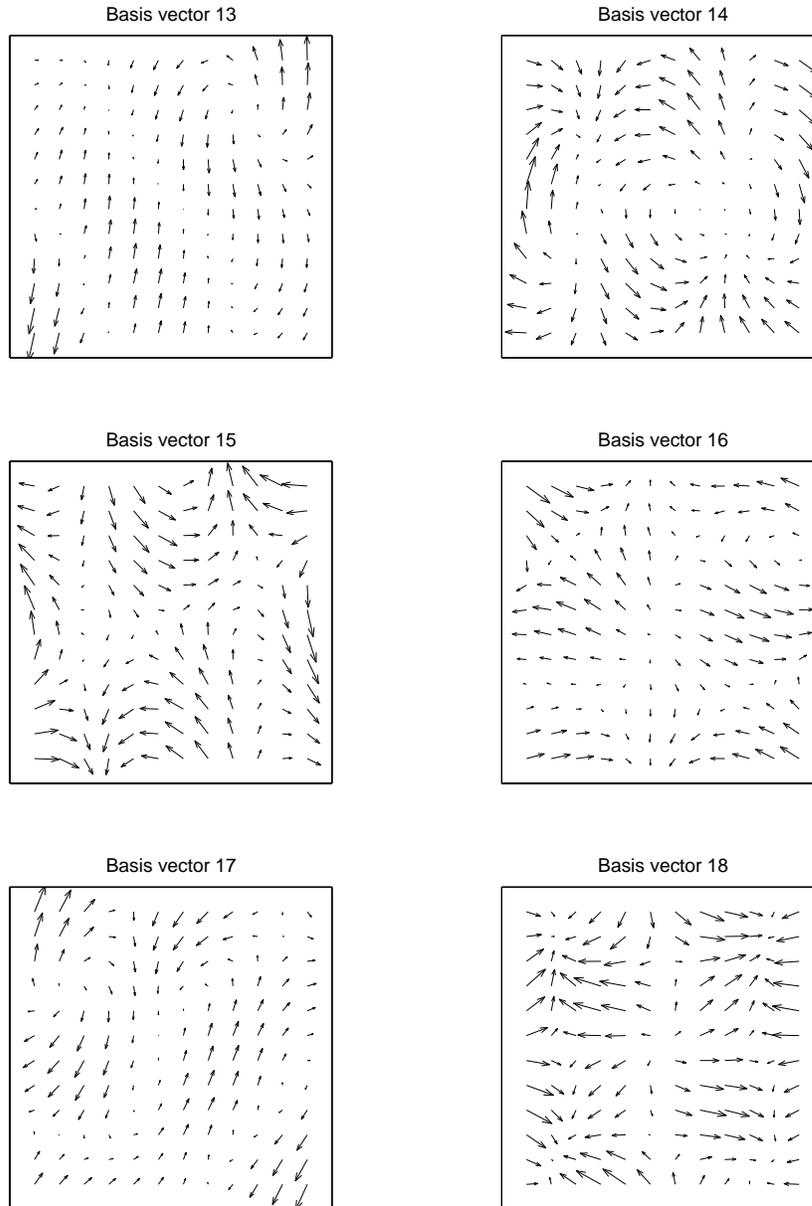
Basis vector 18

Figure A.2: Basis vectors 13 through 18 of the truncated KL model.
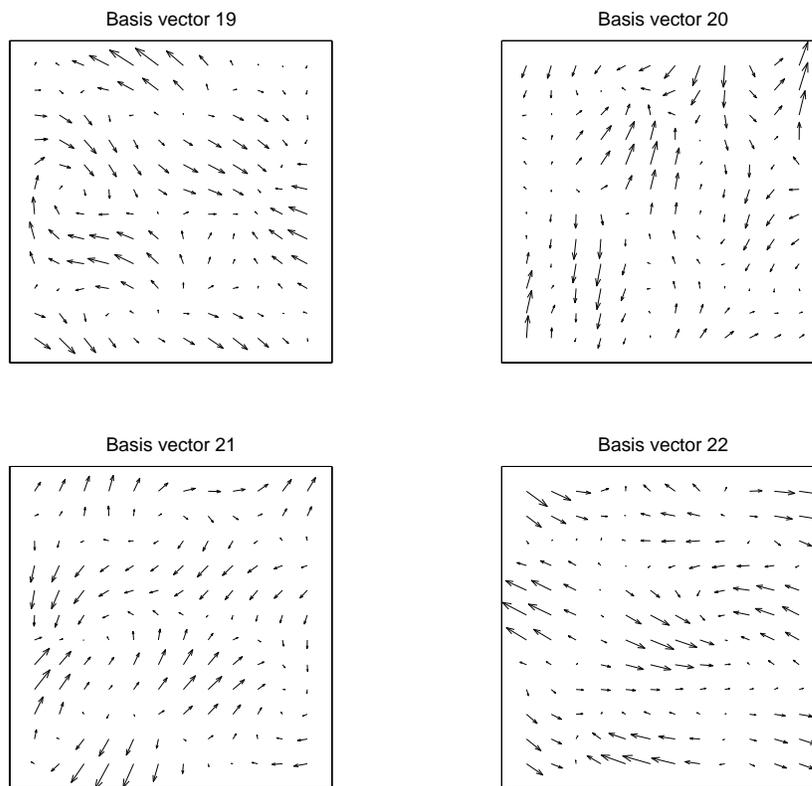
Figure A.3: Basis vectors 19 through 22 of the truncated KL model.

Appendix B

THE REV NUMBERS PROCESSED FOR THE WIND FIELD STATISTICS

830A	835A	840A	845A	850A	855A	860A	865D
830D	835D	840D	845D	850D	855D	860D	866A
831A	836A	841A	846A	851A	856A	861D	866D
831D	836D	841D	846D	851D	856D	862A	867A
832A	837A	842A	847A	852A	857A	862D	867D
832D	837D	842D	847D	852D	857D	863A	868A
833A	838A	843A	848A	853A	858A	863D	868D
833D	838D	843D	848D	853D	858D	864A	869A
834A	839A	844A	849A	854A	859A	864D	869D
834D	839D	844D	849D	854D	859D	865A	

X PARAMETER STATISTICS

2σ		μ	
	X(1) = 122.02		X(1) = 3.05
	X(2) = 154.99		X(2) = -5.18
	X(3) = 41.42		X(3) = 0.51
	X(4) = 34.72		X(4) = 0.80
	X(5) = 38.34		X(5) = 0.66
	X(6) = 32.92		X(6) = 0.14
	X(7) = 16.51		X(7) = 0.22
	X(8) = 16.64		X(8) = 0.21
	X(9) = 16.25		X(9) = 0.14
	X(10) = 18.63		X(10) = -0.2304
	X(11) = 12.07		X(11) = -1.02
	X(12) = 11.82		X(12) = 0.17
	X(13) = 9.92		X(13) = 0.12
	X(14) = 9.95		X(14) = -0.08
	X(15) = 9.81		X(15) = -0.01
	X(16) = 8.71		X(16) = -0.02
	X(17) = 9.00		X(17) = 0.01
	X(18) = 10.13		X(18) = 0.07
	X(19) = 7.47		X(19) = -0.03
	X(20) = 6.98		X(19) = -0.03
	X(21) = 6.83		X(21) = 0.12
	X(22) = 6.16		X(22) = 0.10

Table C.1: The X parameter starting points (twice the standard deviation) and their corresponding means.

Appendix D

HOW TO DETERMINE WHICH X PARAMETERS TO USE AND THEIR VALUES

The X parameter thresholds are originally chosen as 2σ. The thresholds are then tested against the sample dataset. If these thresholds do not locate the regions already subjectively identified as containing possible errors, other thresholds must be used. The thresholds are chosen for the trade-off between false alarms and detection. This is done by making a list of regions that are correctly identified only by the X parameter thresholds and a list of regions that are falsely detected only by the X parameters, i.e., the error statistics do not flag these regions. The following is such a list for the K-L model.

Correctly Identified	False Alarms
$X(4) = 39.91$	$X(4) = 31.66$
$X(4) = 39.43$	$X(4) = 29.89$
$X(3) = 31.36$	$X(5) = 37.81$
$X(6) = 36.17$	$X(5) = 36.70$
$X(6) = 34.57$	$X(5) = 37.65$
$X(9) = 16.84$	$X(6) = 36.87$
$X(3) = 34.94\#$	
$X(5) = 39.20\#$	
$X(9) = 21.22\#$	
$X(4) = 35.91\bullet$	
$X(8) = 16.29\bullet$	
$X(5) = 49.92\circ$	
$X(3) = 32.57\circ$	
$X(5) = 48.38\dagger$	
$X(10) = 20.71\dagger$	

Those values which are marked by the same symbol correspond to a region that was located by more than one of the X parameters. It is clear that parameters five and four can be raised without any adverse effects. However, raising parameter six to the value required to lower the number of false alarms means that at least two regions will go undetected. Thus, this threshold is selected to detect the two regions. Finally, this list also provides a means of deciding which parameters to use for the thresholds. Since only one parameter is needed to detect a region, not all the parameters on the list of correctly identified have to be used. Parameters four, six, three and nine must be used as they are the only parameters to identify a particular region. Once these are chosen, parameters eight and ten can be eliminated. Thus, these thresholds were chosen according to Table 4.1.

Appendix E

THE REV NUMBERS PROCESSED FOR THE RESULTS AND FOR DETERMINING THE THRESHOLDS

1655A	1663A	2454A	2462A
1655D	1663D	2454D	2462D
1656A	1664A	2455A	2463A
1656D	1664D	2455D	2463D
1657A	1665A	2456A	2464A
1657D	1665D	2456D	2464D
1658A	1666A	2457A	2465A
1658D	1666D	2457D	2465D
1659A	1667A	2458A	2466A
1659D	1667D	2458D	2466D
1660A	1668A	2459A	2467A
1660D	1668D	2459D	2467D
1661A	1669A	2460A	2468A
1661D	1669D	2460D	2468D
1662A	1670A	2461A	2469A
1662D	1670D	2461D	2469D

Appendix F

THE REV NUMBERS PROCESSED FOR THE WITHHELD DATA SET

2055A	2059A	2064A	2069A
2055D	2059D	2064D	2069D
2056A	2061A	2065A	2070A
2056D	2061D	2065D	2070D
2057A	2062A	2066A	2071A
2057D	2062D	2066D	2071D
2058A	2063A	2067A	2072A
2058D	2063D	2067D	2072D

Appendix G

THE CORRELATION BETWEEN WIND SPEED AND
AMBIGUITY REMOVAL PERFORMANCE

The correlation between wind speed and ambiguity removal performance is easily seen by plotting the rms speed versus the percent of non-poor regions for each of the lattitude bands. In each of these plots, the correlation is evident. The correlation coefficient for each of the bands is 0.81, 0.80, 0.71, 0.80, and 0.80 beginning with band 5.
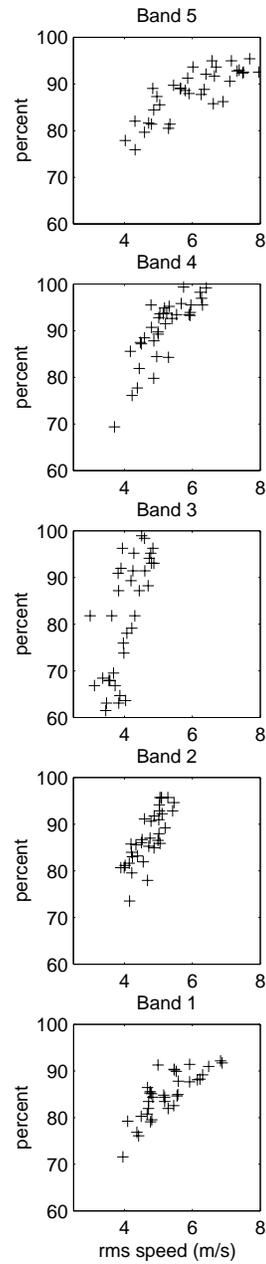
Figure G.1: The average rms wind speed versus the percent of non-poor regions for each of the lattitude bands.