WIND SCATTEROMETRY WITH IMPROVED AMBIGUITY SELECTION AND RAIN MODELING

by

David Draper

A dissertation submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering

Brigham Young University

December 2003

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a dissertation submitted by

David Draper

This dissertation has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Dr. David G. Long, Chair
Date	Dr. Michael A. Jensen
Date	Dr. Karl F. Warnick
Date	Dr. Brian D. Jeffs
Date	Dr. Travis E. Oliphant

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the dissertation of David Draper in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Dr. David G. Long
Chair, Graduate Committee

Accepted for the Department

Dr. Michael A. Jensen
Graduate Coordinator

Accepted for the College

Dr. Douglas M. Chabries
Dean, College of Engineering and Technology

ABSTRACT

WIND SCATTEROMETRY WITH IMPROVED AMBIGUITY SELECTION AND RAIN MODELING

David Draper

Electrical and Computer Engineering

Doctor of Philosophy

Although generally accurate, the quality of SeaWinds on QuikSCAT scatterometer ocean vector winds is compromised by certain natural phenomena and retrieval algorithm limitations. This dissertation addresses three main contributers to scatterometer estimate error: poor ambiguity selection, estimate uncertainty at low wind speeds, and rain corruption. A quality assurance (QA) analysis performed on SeaWinds data suggests that about 5% of SeaWinds data contain ambiguity selection errors and that scatterometer estimation error is correlated with low wind speeds and rain events.

Ambiguity selection errors are partly due to the "nudging" step (initialization from outside data). A sophisticated new non-nudging ambiguity selection approach produces generally more consistent wind than the nudging method in moderate wind conditions. The non-nudging method selects 93% of the same ambiguities as the nudged data, validating both techniques, and indicating that ambiguity selection can be accomplished without nudging.

Variability at low wind speeds is analyzed using tower-mounted scatterometer data. According to theory, below a threshold wind speed, the wind fails to

generate the surface roughness necessary for wind measurement. A simple analysis suggests the existence of the threshold in much of the tower-mounted scatterometer data. However, the backscatter does not "go to zero" beneath the threshold in an uncontrolled environment as theory suggests, but rather has a mean drop and higher variability below the threshold.

Rain is the largest weather-related contributer to scatterometer error, affecting approximately 4% to 10% of SeaWinds data. A simple model formed via comparison of co-located TRMM PR and SeaWinds measurements characterizes the average effect of rain on SeaWinds backscatter. The model is generally accurate to within 3 dB over the tropics. The rain/wind backscatter model is used to simultaneously retrieve wind and rain from SeaWinds measurements. The simultaneous wind/rain (SWR) estimation procedure can improve wind estimates during rain, while providing a scatterometer-based rain rate estimate. SWR also affords improved rain flagging for low to moderate rain rates. QuikSCAT-retrieved rain rates correlate well with TRMM PR instantaneous measurements and TMI monthly rain averages. SeaWinds rain measurements can be used to supplement data from other rain-measuring instruments, filling spatial and temporal gaps in coverage.

ACKNOWLEDGMENTS

I want to thank my wonderful wife who has supported me in my seemingly endless efforts and hours away from home. I also want to thank Dr. Long for his vast knowledge and insights that he has instilled in me.

Contents

Acknowledgments **vii**

List of Tables **xviii**

List of Figures **xxix**

1 Introduction **1**

1.1 Wind scatterometry background 2

1.1.1 History of scatterometers 3

1.1.2 Wind estimation 3

1.2 The SeaWinds scatterometer 8

1.3 Dissertation objectives 15

2 An Assessment of SeaWinds on QuikSCAT Wind Retrieval **17**

2.1 Error in scatterometer winds 17

2.2 Analysis of the KL wind model 20

2.2.1 KL model size and truncation point 22

2.2.2 KL basis data sensitivity analysis 24

2.3 QA analysis method 26

2.3.1 Overview 27

2.3.2 Performance of the ambiguity selection error detection method on the training data set 33

2.4 QA analysis 33

2.4.1 Overall SeaWinds QA results 34

2.4.2 Cross track/RMS wind speed dependence 36

2.4.3 Temporal QA statistics 38

2.5 Subjective analysis of cyclonic storm regions 43
2.6 Discussion and summary 46

3 An Advanced Ambiguity Selection Algorithm for SeaWinds 49
3.1 Background 49
3.2 Overview of BYU point-wise ambiguity selection 50
3.2.1 Initial estimate 51
3.2.2 Estimating swath edges 56
3.2.3 Repairing inconsistencies 57
3.3 Comparison of BYU and JPL Ambiguity Selection 60
3.3.1 Direct comparison of selected ambiguities 62
3.3.2 Quality assurance assessment 64
3.3.3 Analysis of storm and non-storm regions 66
3.4 Summary and discussion 69

4 Evidence of a Threshold Wind Speed in Tower-mounted Scatterometer Data 71
4.1 Overview 71
4.2 The YSCAT scatterometer 72
4.3 Theoretical threshold wind speed 74
4.3.1 Detection and estimation of YSCAT threshold wind speeds . . 75
4.3.2 Dual threshold model 80
4.4 Summary and discussion 82

5 Evaluating the Effect of Rain on SeaWinds Scatterometer Measurements 85
5.1 Data 86
5.2 Modeling σ° in rain and wind 88
5.2.1 Determining model parameters from TRMM and NCEP . . . 89
5.2.2 Relating PR atmospheric parameters to rain rate 95
5.3 Analysis 98

5.3.1 Estimating the surface perturbation 98
5.3.2 Combined rain effect model 102
5.3.3 Data regimes 106
5.4 Conclusions 107

6 Simultaneous Wind and Rain Retrieval Using SeaWinds Data 113
6.1 Methodology 114
6.1.1 Wind/rain measurement model 115
6.1.2 Simultaneous wind/rain retrieval 118
6.1.3 Rain-corrected wind retrieval 119
6.2 Beam/ WVC filling 119
6.3 Estimation of K_{pe} 122
6.3.1 K_{pe} due to model uncertainty and non-uniform rain 123
6.3.2 K_{pe} yielding lowest RMS wind error 124
6.4 Simulations and validation 126
6.4.1 Simulation results 126
6.4.2 Validation 133
6.5 Synoptic example 137
6.6 Conclusions 140

7 Assessing the Quality of SeaWinds Rain Measurements 149
7.1 Background 150
7.2 Cramér-Rao bound 153
7.2.1 C-R / Monte-Carlo simulation analysis 154
7.3 Improved rain flag 160
7.4 Monthly rain rate validation with TMI 164
7.5 Conclusions 168

8 Summary and Conclusions 171
8.1 Contributions 172
8.1.1 Impact of error sources on vector wind 172

8.1.2 High quality ambiguity selection without nudging 172
8.1.3 The effect of a low wind speed threshold on scatterometer data 173
8.1.4 The effect of rain on the SeaWinds scatterometer 173
8.1.5 Rain measurement from SeaWinds 174
8.1.6 Improvement of current rain flagging 174
8.2 Future research 175
8.2.1 Improvement of Non-nudging Ambiguity Selection in Storms . 175
8.2.2 Application of rain model to SeaWinds on ADEOS II 175
8.2.3 High resolution rain retrieval 175
8.2.4 Combined model-based/rain estimation 176

A Fourier Analysis and Model Size Trade-offs of the QuikSCAT KL Model **179**
A.1 Generating a KL wind field model 179
A.2 Frequency spectrum of the KL wind model 180
A.2.1 Model size comparison 180
A.2.2 Frequency groupings of the QuikSCAT KL model 183
A.2.3 QuikSCAT point-wise wind spectrum 185
A.3 Determining KL model truncation points 187
A.3.1 Trade-offs between smaller and larger models 188
A.4 Conclusion 188

B Selecting Variable WVC Thresholds **191**
B.1 Optimal direction error thresholds 192
B.2 Optimal vector error thresholds 196

C Ambiguity Selection Error Detection Consistency Checks **199**
C.1 Model based consistency check 199
C.2 Directional histogram-based consistency check 200
C.3 Region RMS error threshold 201
C.4 RMS wind speed threshold 202

D QA Bit Flag **203**

D.1 Individual WVC flag 203

D.2 Region WVC flag 204

D.3 QA bit flag results 205

E Automated Storm Rating Procedure **209**

F Dilation and Erosion Steps of the BYU Point-wise Method **213**

G Alternative Method of Determining The Low Wind Speed Threshold **217**

H Derivation of the Combined Rain/Wind Model Mean and Variance **223**

I Gradient of the Rain/Wind Model **225**

I.1 Gradient of the MLE mean 225

I.2 Gradient of the MLE variance 226

J Rain Model Bias Correction **227**

Bibliography **237**

List of Tables

2.1 Comparison of several KL models corrupted with artificially-induced ambiguity selection errors to the original (non-corrupted) KL model. 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 left column ($L_{A,B}$) where A is the corrupted KL model and B is the original model. A and B both have 6 basis vectors. 26

2.2 Constant WVC thresholds determining the flagging of a vector. The term u_{rms} is the region root mean square (RMS) wind speed defined by $u_{rms} = (\mathbf{w}_o^T \mathbf{w}_o / N)^{\frac{1}{2}}$ 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. (†Denotes values used by *Gonzales and Long*, [1999]). 30

2.3 Thresholds determining the classification of a region. (†Denotes values used by *Gonzales and Long [1999]*). 31

2.4 Overall results of the QA analysis for the SeaWinds data set and for the training data set. Also, the percent of ambiguity selection errors subjectively flagged in the training data set. 35

2.5 Latitude bands 41

2.6 Correlation coefficients for each latitude band. C_{sa} is the correlation coefficient between the smoothed ambiguity selection error and storm data of Figure 2.13. C_{ra} is the correlation coefficient between the smoothed ambiguity selection error and rain data of Figure 2.13. . . 42

2.7 Number of storms identified as rating "1", "2", and "3" for two time periods: July 1999, and January 2000. For each time period, the data is further subdivided into northern and southern hemisphere locations. 44

2.8 Average distance (km) between the NCEP storm centers and the L2B storm centers for storms rated 1 or 2. 46
3.1 Percentages of ambiguities chosen by the JPL L2B product and the BYU algorithm. 63
3.2 Percentage of ambiguities chosen the same and changed from 1st to 2nd or from 1st or 2nd to 3rd or 4th for the sample set of BYU and JPL ambiguity-selected winds 64
3.3 Total numbers and percentages of regions subjectively rated "good" or "poor" in both BYU and JPL data sets for 38 × 38 WVC areas without and with cyclonic features. Percentages are calculated for regions with and without cyclones separately. 68
4.1 Values of parameters given in Eq. (4.3) for a water temperature of 10° C and zero salinity. 74
4.2 Theoretical threshold wind speeds (m/s) as a function of frequency and incidence angle for YSCAT 75
5.1 Coefficients of the linear and quadratic fits to the parameters PIA and $\sigma_{r(PR)}$ 98
5.2 Estimated parameters of $\tilde{\sigma}_{\mathbf{sr}}$ for h-pol and v-pol for the optimum γ. 102
5.3 Estimated parameters of σ_e for h-pol, v-pol and both linear and quadratic fits. 104
6.1 K_{pe} value yielding the lowest RMS wind speed error when compared with calibrated NCEP numerical weather prediction winds over the TRMM/QuikSCAT co-located dataset. Also, the corresponding RMS errors for both variance equations of Eq. (6.13) and Eq. (6.16). . . . 125
6.2 Delineations of wind speed, wind direction, rain rate, and cross track position for which the simulations are performed. 127
7.1 Percentage of rain detection for SWR rain flag, MUDH rain flag, and combined rain flag for various PR rain rate (R_{PR}, km mm/hr) bins over the co-located TRMM/QuikSCAT dataset. 162

7.2 Percentage of WVCs flagged as containing rain, and not containing rain for the MUDH and combined rain flags for 4 months of SeaWinds on QuikSCAT data. 163

7.3 Mean bias and correlation coefficients for monthly averages of QuikSCAT-derived rain rates and TMI rain rates. A positive mean bias indicates larger average QuikSCAT rain rates. A negative bias indicates larger TMI rain rates. 165

A.1 Each QuikSCAT model parameter of the 24 × 24 model was modeled by the enlarged NSCAT 12 × 12 WVC model. The model parameters with the greatest energy used by the NSCAT model are enumerated. In addition, the approximate range of NSCAT model parameters used to model the QuikSCAT KL model is shown. 189

A.2 Comparison of frequency bin groupings between different models. The last model parameter that used only the first N bins is listed. 190

A.3 Cutoff model parameter numbers for 8 × 8, 12 × 12, 16 × 16, 20 × 20, and 24 × 24 QuikSCAT models. 190

D.1 Percent of valid individual WVCs flagged according to several QA bit flag classifications. The symbol '×' indicates a "don't care." The classifications have the following meanings: A - No Flag: the cell is not excessively noisy and is not estimated to be an ambiguity selection error, B - Flagged with constant thresholds (noisy vector), C - Flagged with variable thresholds (likely to lie near ambiguity selection errors), D - "Good" region classification, E - "Fair" region classification, F - "Poor" region classification, G - "Ambiguity selection error" region flag, H - Both region and cell flagged as an ambiguity selection error 207

E.1 Sample means and standard deviations of 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) (%), and 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. . 210

E.2 Number of storms subjectively rated "1", "2", or "3" vs. number of storms rated "1", "2", or "3" by the automated method. 210

J.1 Corrected model parameters for the quadratic log-log model of the effective rain backscatter σ_e. 227

List of Figures

1.1 Idealized ocean waves generated by wind with SeaWinds backscatter. 4

1.2 GMF for SeaWinds inner beam: h-pol, 46° incidence, showing σ° as a function of the relative azimuth angle χ for several wind speeds. . . 6

1.3 Sample objective function showing multiple minima, corresponding to ambiguities. The true wind is shown as a black arrow. 8

1.4 The QuikSCAT satellite with the SeaWinds instrument. 9

1.5 SeaWinds measurement geometry 10

1.6 SeaWinds measurement geometry for several cases: a) swath edge, b) sweet spot, and c) nadir swath region. 11

1.7 A sample QuikSCAT ambiguity field. Also shown, the first and second ambiguity fields. 13

2.1 a) Region containing some noisy vectors due to estimation error, poor measurement geometry, rain, etc.. b) Region containing an ambiguity selection error. Notice that the wind vectors in the central region oppose the surrounding flow. 18

2.2 The truncated KL model used in the QA algorithm. 21

2.3 The first 50 eigenvalues of the 8×8 KL model. 22

2.4 The wind fields corresponding to the first 6 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. 23

2.5 The wind fields corresponding to KL basis vectors 7 - 12 and the magnitudes of their corresponding cosine transforms. 24

2.6 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. 27
2.7 8×8 Region of QuikSCAT point-wise selected wind and the KL model fit. Also shown are all possible ambiguities and the difference field between the observed wind and model fit. 29
2.8 The a) directional and b) vector error thresholds per cross track and RMS wind speed that give a constant false alarm rate. 32
2.9 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. 34
2.10 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. 36
2.11 a) Percent of WVCs per cross track position with 1 to 3 ambiguities and b) Percent of WVCs per cross-track position with 4 ambiguities averaged over 600 revs of SeaWinds data. 37
2.12 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. 38
2.13 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. . . . 40
2.14 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. 45
3.1 Flow chart of the BYU ambiguity selection method. 51

3.2 Initial ambiguity selection method example: a) First ambiguity field, b) Inner-beam portion with rain masked out, c) Decimated 20 × 20 field d) First model fit, e) First and second ambiguities with inconsistent flow masked out, f) Second model fit, g) reconstructed 60 × 60 field, h) Closest ambiguity to second model fit. 54
3.3 A flow diagram summarizing the initial estimate. The shaded bubbles show the most important steps. 55
3.4 A flow diagram summarizing the estimation of swath edges. The shaded bubbles show the crucial steps. 58
3.5 An example of the steps involved with flagging ambiguity selection error edges. (a) The ambiguity-selected wind, (b) the median-filtered wind, (c) the average-filtered wind, (d) the difference field. Where the errors are large, the WVCs are flagged as inconsistent (circled). . . . 59
3.6 Example of the repair process: a) A sample wind field containing a large ambiguity selection error in the initial estimate. b) The initial estimate including swath edge estimation. c) Ambiguity selection error region flagged by the inconsistency flag (red vectors). d) Final field after repairs. 61
3.7 A flow diagram summarizing the repair process. The shaded bubbles show the most important steps. 62
3.8 Percentage of individual WVCs selected the same for both JPL and BYU methods per (a) cross-track position and (b) average wind speed of the ambiguities. 63
3.9 Percentage of 8 × 8 regions flagged as containing possible ambiguity selections error by the QA analysis method (a) per cross-track position and (b) RMS wind speed for both JPL and BYU methods (left axes). A histogram of wind speeds is show in (b), indicating the number of regions inspected per RMS wind speed bin (right axis). 65
3.10 Examples of the various classifications of regions *without* cyclonic features when comparing the JPL and BYU ambiguity selection routines. 66

3.11 Examples of the various classifications of regions *with* cyclonic features when comparing the JPL and BYU ambiguity selection routines. . . 67
4.1 Wind wave/electromagnetic wave interaction for Bragg scattering. . 73
4.2 The theoretical threshold wind speeds for the range of Bragg wavelengths observed by YSCAT. Also, the estimated threshold wind speeds for v-pol and h-pol observations obtained from the YSCAT data where a threshold wind speed is detected. The dashed and dot-dashed lines show a second order polynomial fit to the h-pol and v-pol threshold wind speeds respectively. 77
4.3 H-pol YSCAT data plotted as a function of wind speed for each Bragg wavelength. The straight line through the data is the linear least-squares fit. The curved line is the non-parametric fit. The solid vertical line is the theoretical threshold wind speed (Th). The dashed line is the threshold wind speed estimated from the YSCAT data (Es). . . 78
4.4 V-pol YSCAT data plotted as a function of wind speed for each Bragg Wavelength and polarization (See Figure 4.3). 79
4.5 An example of the roll-off, hysteresis and moderate wind speed regions in YSCAT data. 81
4.6 Subjectively chosen low-end and high-end cutoffs for the hysteresis region and a least-squares fit. 82
5.1 Swath geometry of the TRMM PR and SeaWinds on QuikSCAT instruments. 87
5.2 Histogram of rain rates for the co-located data set derived from TRMM PR data 91
5.3 Relationship of $\sigma_{r(PR)}$ to $\sigma_{m(QSCAT)}$. 94
5.4 Ambiguity selected wind in a QuikSCAT/TRMM PR co-location. Also, an enhanced resolution backscatter image for the fore and aft beams (h-pol), and integrated rain backscatter calculated from TRMM PR. 95

5.5 Parametric estimation of the parameters σ_r and PIA as a function of rain rate for a) h-pol and b) v-pol data. 97
5.6 Each parameter of $x_{sr}^{\gamma}(n)$ as a function of γ for both h-pol and v-pol observations and both linear and quadratic fits. 99
5.7 Ratio of the attenuated surface perturbation $\alpha_r\sigma_{sr}(\gamma)$ to the modified atmospheric rain backscatter $\gamma\sigma_r$ for several values of γ plotted as a function of rain rate for a) h-pol and b) v-pol. 100
5.8 Objective function with respect to γ (see text). The γ that minimizes the objective function is displayed with a dashed line. 101
5.9 Non-parametric and least-squares fits to the parameter σ_e. a) and b) show the data and non-parametric fit in normal space for h-pol and v-pol. c) and d) show the non-parametric, and linear and quadratic fits in log-log space. 103
5.10 Non-parametric, and combined rain effect models for a) h-pol and b) v-pol. Also, error between parametric and non-parametric models. . 106
5.11 Backscatter regimes for SeaWinds as a function of rain rate and effective wind backscatter for a) h-pol and b) v-pol models. Also plotted is a contour plot of the combined rain effect model for σ_m. 108
5.12 QuikSCAT measured backscatter σ_m plotted as a function of effective non-raining wind σ°, σ_w for h-pol. Also plotted is a non-parametric estimate of the data, and the combined rain effect model. 110
5.13 QuikSCAT measured backscatter σ_m plotted as a function of effective non-raining wind σ°, σ_w for v-pol. Also plotted is a non-parametric estimate of the data, and the combined rain effect model. 111
6.1 Variance of Eq. 6.7 for K_{pa} very large and of Eq. 6.7 ($K_{pa} = 0$) as a function of rain rate, demonstrating that even at large K_{pa}, the overall variance is not very sensitive to K_{pa}, and thus the variability of the attenuation can be ignored. 117

6.2 Sample likelihood function, displayed as a function of wind direction and rain rate. For each pixel, the wind speed corresponding to the likelihood minimum for that wind direction and rain rate is chosen. . 120
6.3 The geometry of a sample WVC (bold square) and the 6 dB gain contours of the individual σ° measurements (ellipses) comprising the WVC. A PR-derived rain map is shown in the background. 123
6.4 The WVC-average (unweighted) rain rate versus the weighted average rain rate seen by SeaWinds. 124
6.5 Ratio in RMS error between the a) simultaneous wind/rain retrieval RMS_{wr} and the wind-only retrieval RMS_{wo}, and b) the rain-corrected wind retrieval RMS_{rc} and the wind-only retrieval RMS_{wo} as a function of rain fraction. 128
6.6 Directional error statistics as a function of true direction for 4 rain rate cases with wind speed of 7 m/s. In each plot, the rain-corrected wind retrieval is shown on top, simultaneous wind/rain retrieval is in the middle, and wind-only retrieval is on the bottom. 129
6.7 Wind speed error statistics as a function of true direction for 4 rain rate cases with wind speed of 7 m/s. In each plot, the rain-corrected wind retrieval is shown on top, simultaneous wind/rain retrieval is in the middle, and wind-only retrieval is on the bottom. Notice the scale differences of the plots. 131
6.8 Wind speed error statistics as a function of wind speed for various rain rates. The rain-corrected wind retrieval is shown on top, simultaneous wind/rain retrieval is in the middle, and wind-only retrieval is on the bottom. Each point is horizontally offset for clarity. 132
6.9 Rain rate error statistics as a function of rain rate for several different wind speeds. Each point is horizontally offset for clarity. 133
6.10 Scatter plot with density curves of the bias-corrected NCEP wind speeds versus simultaneous wind/rain retrieved wind speeds for various rain fraction bins. 134

6.11 Scatter plot with density curves of the bias-corrected NCEP wind speeds versus wind-only retrieved wind speeds for various rain fraction bins. 135
6.12 Scatter plot with density curves of NCEP wind directions versus simultaneous wind/rain retrieved wind directions for various rain fraction bins. 136
6.13 Scatter plot with density curves of NCEP wind directions versus wind-only retrieved wind directions for various rain fraction bins. 137
6.14 Scatter plot of TRMM PR derived "effective" weighted average rain rates versus QuikSCAT-derived rain rates. Density curves are shown, along with the equality line (solid). The dotted line represents the best quadratic fit to TRMM rain rate (in log space). 138
6.15 Synoptic example of a hurricane: a) QuikSCAT simultaneous wind/rain derived wind vectors. b) QuikSCAT wind-only retrieval. c) QuikSCAT simultaneous wind/rain derived rain rates. d) Raw PR rain rates. The box shows the coverage of the PR data. Black pixels represent WVCs flagged as containing land. 139
6.16 Example of a rain event a) QuikSCAT simultaneous wind/rain derived wind vectors. b) QuikSCAT wind-only retrieval. c) QuikSCAT simultaneous wind/rain derived rain rates. d) Raw PR rain rates. The box shows the coverage of the PR data. Black pixels represent WVCs flagged as containing land. 140
6.17 Mean wind speed bias as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval. 142
6.18 Wind speed standard deviation as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval. 143

6.19 Mean directional bias as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval. 144
6.20 Wind direction standard deviation as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval. 145
6.21 Mean rain rate bias as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval. 146
6.22 Rain rate standard deviation as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval. 147
7.1 Example of wind/rain estimation for cross-swath blowing winds: a) QuikSCAT simultaneous wind/rain derived wind vectors. b) QuikSCAT wind-only retrieval. c) QuikSCAT simultaneous wind/rain derived rain rates. d) Raw PR rain rates. The box shows the coverage of the PR data. Black pixels represent WVCs flagged as containing land. 151
7.2 Standard deviation as a function of cross track position of retrieved wind speed, wind direction, and rain rate from QuikSCAT computed via a) simulation and b) the unbiased C-R bound for three rain rates. Notice the anomalous C-R bound spike in the nadir region for the directions, resulting from a large C-R bound due to a bias in the estimate. The true wind velocity is 7 m/s and 0 degrees. 155
7.3 Standard deviation as a function of wind direction of retrieved wind speed, wind direction, and rain rate from QuikSCAT computed via a) simulation and b) the unbiased C-R bound for three rain rates. Notice the anomalous C-R bound spikes at 60, 120, 240, and 300 degrees for the C-R bound, resulting from a large C-R bound due to a bias in the estimate. The true wind speed is 7 m/s at cross track position 14. . 156

7.4 Wind speed and rain rate bias (lines) and standard deviations (error bars) for all wind directions and several rain rates with a wind speed of 7 m/s for WVC 14. 157
7.5 Monte Carlo Simulations of SeaWinds data at WVC 14 for several wind directions, a true wind speed of 7 m/s, and true rain rate of 0 km mm/hr. Also shown are concentration ellipses representing the unbiased C-R bound 158
7.6 GMF for a true wind speed of 7 m/s plotted for the inner H-pol SeaWinds beam and outer V-pol SeaWinds beam at WVC 14. Also plotted are the expected fore and aft backscatter measurements for each beam given an upwind direction (wind blowing toward the instrument or $d = 90^\circ$ relative to the direction of the spacecraft motion), or a downwind direction (wind blowing away from the instrument or $d = 270^\circ$ relative to the direction of the spacecraft motion). 159
7.7 Rain rate thresholds that yield a constant false alarm rate given wind speed, wind direction, and cross-track position. 161
7.8 Percentage of WVCs flagged by the MUDH rain flag versus the combined rain flag per cross-track position. 164
7.9 Monthly rain rate average for a) QuikSCAT-derived rain rates using the combined rain flag and b) TMI-derived rain rates. 166
7.10 a) Monthly rain height average from TRMM PR, interpolated to $1^\circ \times 1^\circ$ grid; b) Error between QuikSCAT and TMI monthly average rain rate for August 1999. 167
A.1 The x-axis is the one dimensional frequency spectrum of the QuikSCAT 16 $\times$ 16 KL model for the u component. The y-axis is the model parameter number. As the model parameter increases, the frequency spectrum widens. 181
A.2 The average modeling error over the wind field of the QuikSCAT 16 $\times$ 16 model being modeled by successive numbers of the enlarged QuikSCAT 12 $\times$ 12 model. 184

A.3 Average FFT over the cross track of QuikSCAT winds 185
A.4 Running sum of energy in the FFT bins over the cross track 186
B.1 The average directional standard deviation for both RMS wind speed and cross track position. 192
B.2 The direction thresholds per cross track and RMS wind speed that minimizes the probability of false alarm beneath a threshold of 2.5%. 194
B.3 The direction error thresholds per cross track and RMS wind speed that minimize the probability of false alarm beneath a threshold of 2.5%. 195
B.4 The vector error thresholds per cross track and RMS wind speed that minimize the probability of false alarm beneath a threshold of 2.5%. 195
B.5 Probability of missed detection with optimized direction and vector thresholds for different region thresholds. 196
C.1 A wind field containing an ambiguity selection error "edge" and the corresponding directional histogram. Note the bi-modal nature. . . . 200
C.2 Histogram of the RMS errors of all the regions manually flagged as "poor", the false alarms generated by the algorithm, and the missed detections generated. 201
D.1 QA bit flag example containing a possible region of ambiguity selection error. Missing vectors indicate the position of land. Notice that the QA algorithm flags the position of inconsistencies in the estimated wind flow. 205
D.2 A few examples of the QA bit flag in stormy regions or regions of ambiguity selection error. 206
F.1 The "nearest neighbor" structuring element. A value of "1" indicates a nearest neighbor. The center pixel is the origin. 214
F.2 (a) Dilation and (b) erosion steps, demonstrating how an isolated region is "filled in." The initial state matrix is derived from the example in Fig. 3.5. To enhance readability, state 0 cells are left blank. . . . 215
G.1 An threshold wind speed calculation example 221

J.1 Validation of QuikSCAT rain rates with TRMM PR antenna average rain rates using the corrected rain model. 228
J.2 Validation of QuikSCAT rain rates with TRMM PR antenna average rain rates using the corrected rain model. 229

Chapter 1

Introduction

Oceans are the central element of global energy circulation. The interaction of the atmosphere and the ocean is of great interest to scientists and oceanographers involved with understanding the world's climate and predicting future weather patterns. Wind scatterometry is a method of remotely sensing ocean winds by relating radar backscatter from wind-roughened seas to near-surface wind speed and direction. Wind scatterometry affords frequent global coverage of the Earth's oceans, providing a better understanding of the global winds than can be obtained from ground-based techniques.

The most recent spaceborne scatterometers launched by NASA are SeaWinds on QuikSCAT (July 1999) and SeaWinds on ADEOS II (November 2002). The SeaWinds instrument is a conically scanning Ku-band (14 GHz) active radar with a wide-swath design that affords near-full coverage of our planet's oceans on a daily basis at 25 km resolution [1]. Previous scatterometers such as the SEASAT Active Scatterometer System (SASS) flown in 1978 and the NASA Scatterometer (NSCAT) in 1996-1997 by the National Aeronautics and Space Administrations (NASA) utilized fixed fan-beam antennas and did not measure winds in the nadir sub-track region. The SeaWinds design presents both unique advantages and challenges to wind scatterometry.

Wind estimation is possible due to the relationship between the normalized backscattering cross-section of the ocean surface and the vector wind. This relationship is known as the Geophysical Model Function (GMF) and is symmetric with respect to wind direction. Because of the symmetry of the GMF, the wind vector solution obtained via scatterometry is generally non-unique. At each resolution cell,

there are typically 2 to 4 possible wind vector solutions known as ambiguities that could have given rise to the measurements. Numerical weather prediction data is typically used to "nudge" or influence the scatterometer estimate, creating a dependence upon outside information to select a unique wind vector field from the ambiguities. While scatterometer wind estimates are generally accurate, scatterometer wind estimation quality is degraded at low wind speeds due to a low signal-to-noise ratio (SNR). Low return signal power at low wind speeds is theorized to stem from a low wind speed threshold in backscatter below which the small waves formed by wind fail to be generated. Further, rain events corrupt the wind-induced backscatter response, making wind estimates less accurate in the presence of rain.

The objective of this dissertation is to analyze, understand, and mitigate typical errors in current scatterometer wind retrieval. The dissertation addresses three main areas: performing high quality ambiguity selection without nudging, understanding the backscatter response at low wind speeds, and modeling the effects of rain on SeaWinds backscatter in order to improve wind estimation in the presence of rain and to provide a scatterometer-based estimate of the rain rate. These objectives are expanded in Section 1.3.

In this introduction, a brief history of scatterometry and a general overview of wind estimation is presented in Section 1.1. A description of the SeaWinds scatterometer and the SeaWinds wind retrieval method is given in Section 1.2, along with a discussion of error sources in scatterometer data. Section 1.3 provides an overview of the dissertation.

1.1 Wind scatterometry background

Wind scatterometry is a sub-discipline of Earth remote sensing. Many active and passive sensors have been used to study the Earth from outer space including optical, infrared, and microwave-frequency instruments. An advantage of radar instruments is that they are relatively insensitive to cloud conditions and solar illumination. Also, radar makes it possible to observe features of the surface that are not "visible" at higher electromagnetic frequencies.

1.1.1 History of scatterometers

Radar return from the ocean surface was first studied in connection with clutter rejection in military radar systems during World War II. It was later discovered that radar cross section of the ocean surface was related to the wind velocity. From this discovery, the idea of the scatterometer was born. The scatterometer is a radar specifically designed to measure wind over the ocean. In 1973-1974, the first scatterometer was flown as a part of the Skylab mission. Later, in 1978, the National Aeronautics and Space Administration launched the SEASAT-A Scatterometer (SASS). Although the spacecraft failed after only 6 weeks of operation, the short SASS mission proved that accurate ocean wind measurements could be obtained via remote measurements from outer space. As a follow-up mission to SASS, the NASA scatterometer (NSCAT) aboard the Advanced Earth Observing Satellite (ADEOS) was launched in 1996. The ADEOS platform, however, failed after only nine months of operation [2].

During the NSCAT era, the next generation scatterometer called SeaWinds was being designed. The SeaWinds instrument utilizes a rotating pencil beam, rather than fixed fan antenna beams like SASS and NSCAT, to measure the radar cross-section of the ocean. To fill the gap in coverage caused by the NSCAT failure, a SeaWinds instrument was flown on the QuikSCAT satellite as a "quick recovery" mission in June 1999. The next SeaWinds instrument was launched in November 2002 aboard ADEOS II. Both SeaWinds instruments are currently functioning to provide unprecedented earth-wide coverage of ocean winds. Additional scatterometers have been launched by the European Space Agency aboard the European Remote Sensing-1 (ERS-1) and ERS-2 satellites [3].

1.1.2 Wind estimation

The underlying concept behind scatterometer wind estimation is that radar instruments can remotely measure the wind-generated "roughness" of the ocean surface. The profile of the ocean surface follows a power-law dependence on wave number, with saturation occurring at low wave numbers, and a cutoff at high wave numbers.

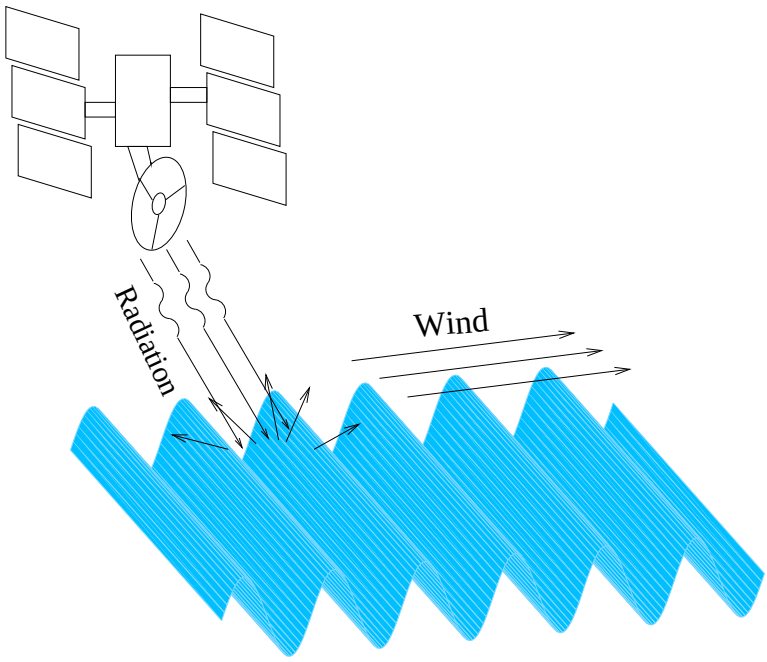


Figure 1.1: Idealized ocean waves generated by wind with SeaWinds backscatter.

Thus, in a rough ocean, nearly all reasonable spatial frequencies of waves exist simultaneously. The ocean waves with the highest wave numbers on the order of a few centimeters are known as capillary waves and are the most sensitive to the wind. The wind transfers momentum into the ocean by forming and sustaining these small capillary waves. Transfer of momentum from the wind to the ocean is modeled as a drag force per unit area, known as wind stress.

Under neutral stability conditions (equal surface air and surface temperature with an adiabatic lapse rate), the wind speed above the surface at a reference height is directly related to the wind stress. Also, a nonlinear monotonic relationship exists between the wind stress and amplitude of capillary waves, relating the capillary wave amplitude to the neutral stability wind [4].

Scatterometers transmit an electromagnetic pulse and measure the power in the return echo (see Figure 1.1). From the return power, the normalized radar

cross-section, σ° of the ocean surface is calculated via the radar equation,

$$\sigma^\circ = \frac{(4\pi)^3 r^4 L}{P_t G^2 \lambda^2 A_{eff}} P_r \tag{1.1}$$

where r is the distance to the earth, L is the loss of the antenna, P_t is the power transmitted, G is the gain of the antenna, λ is the electromagnetic wavelength, A_{eff} is the effective area of the ocean surface illuminated by the antenna, and P_r is the power in the return signal. Scatterometers measure σ° at high incidence angles (20-60 degrees). The main backscattering mechanism at these incidence angles is Bragg resonance from waves on the order of the electromagnetic wavelength [5] (capillary-sized waves for SeaWinds). Bragg resonance occurs when scattering from the troughs and crests of the wave field constructively combine to yield an enhanced backscatter return to the radar. The water wavelength to which the scatterometer signal resonates is known as the Bragg wavelength. The rougher the surface, the higher the return echo. The Bragg-induced backscatter is a function of the amplitude and orientation of the small wind-generated capillary waves, and thus a function of the vector neutral stability wind.

The relationship between wind velocity and σ° is described by the GMF. Figure 1.2 shows the GMF used with SeaWinds on QuikSCAT for several wind speeds. The GMF has a sinusoidal nature in 2χ, and is symmetric about $\chi = 180^\circ$. The highest σ° occurs at 0° or wind blowing directly toward the satellite. The lowest σ° occurs when the wind is perpendicular to the satellite azimuth (90 and 270 degrees). Another lower peak occurs at 180°, when the wind blows away from the radar. The sensitivity of σ° to wind speed and direction allows for scatterometer-based wind retrieval. The derivation for various model functions can be found in [6, 7, 8, 9]

The GMF, $\mathcal{M}$, is a function of the wind speed, relative azimuth angle, incidence angle, and polarization,

$$\sigma^\circ = \mathcal{M}(u, \chi, \theta, \text{pol}, f) \tag{1.2}$$

where u is the wind speed, χ is the relative azimuth angle defined as the difference between the wind direction d and azimuth angle ϕ of the instrument, θ is the surface

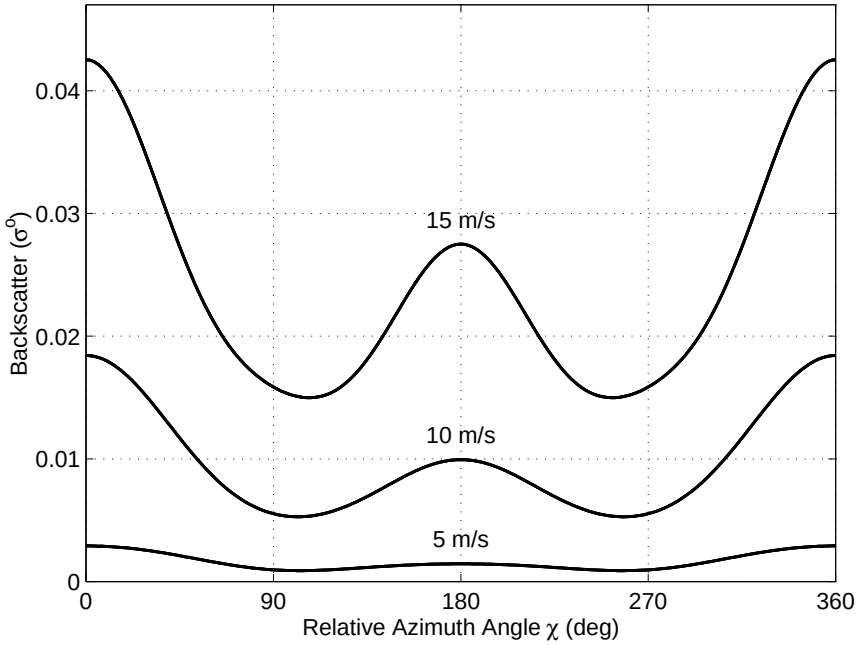


Figure 1.2: GMF for SeaWinds inner beam: h-pol, 46° incidence, showing σ° as a function of the relative azimuth angle χ for several wind speeds.

incidence angle, pol is the electromagnetic polarization, and f is the radar frequency. The azimuth angle, incidence angle, and polarization are determined by the instrument, and are known for each measurement. Thus, the GMF effectively maps two parameters (speed and direction) to one (σ°). Because of this, a single backscatter measurement does not uniquely represent a wind vector, but rather a range of possible wind speeds and directions. Inversion of the GMF from one measurement is the non-linear equivalent of the "one equation; two unknowns" scenario in linear algebra.

In order to provide a unique wind vector solution from scatterometer observations, the scatterometer is designed to provide multiple σ° measurements from various azimuth angles [10]. The satellite σ° data is gridded into square bins called wind vector cells (WVCs), with each WVC containing measurements from multiple azimuth looks. From the gridded σ° measurements, two wind retrieval methods have been introduced. The standard wind estimation procedure, known as point-wise wind retrieval, estimates the wind at each WVC individually using the σ° measurements

assigned to the WVC. An alternative method, known as field-wise wind retrieval, estimates entire N $\times$ N WVC regions by finding the optimal low-order wind field model fit to the data [11].

For both point-wise and field-wise wind retrieval, the wind estimation procedure involves finding the wind solution that minimizes an objective function or error measure between the wind projected through the GMF and the measured σ° values. The typical objective function used in wind estimation has the form [12]

$$J(\mathbf{u}, \sigma^\circ) = \sum_j \sum_k \left\{ \left| \frac{f(\sigma^\circ_{j,k}) - f(\mathcal{M}_k(\mathbf{u}_j))}{\varsigma^2_{j,k}} \right|^p + q \ln \varsigma_{j,k} \right\} \tag{1.3}$$

where j indexes the WVC (note: j only indexes one WVC for point-wise retrieval), k indexes the measurement within the WVC, $\mathbf{u}_j$ is the estimated wind vector at the j^{th} WVC, $\varsigma^2_{j,k}$ is a weighting term which may be a function of either the wind or the measured σ° values, and f is some weighting function (typically the identity or the logarithm). The term $\mathbf{u}$ is the wind field parameterized by the low-order wind model for field-wise wind estimation, and the wind vector estimate for point-wise estimation.

Because of noise and symmetry in the GMF, the objective function usually has several local minima. The wind solutions corresponding to these minima are known as ambiguities or aliases (see Figure 1.3). The ambiguities are ordered by increasing objective function value. The global minimum is termed the "first ambiguity", and the ambiguity with next lowest likelihood value the "second ambiguity" and so forth. The first ambiguity is most likely to be the correct wind vector choice; however, any of the ambiguities may be the closest ambiguity to the true wind and thus the best estimate. Because the estimation process results in multiple wind solutions, an ambiguity selection or dealiasing step is required to yield a unique wind vector field [13].

Ambiguity selection techniques are generally *ad hoc*, and not geophysically based. The typical procedure includes two steps: selection of an initial field, and filtering. The initial field can be "nudged" or "un-nudged." For nudged winds, the initial field is selected as the closest ambiguity at each WVC to an outside estimate of the wind. Traditionally, only a subset of ambiguities are used in nudging to mitigate

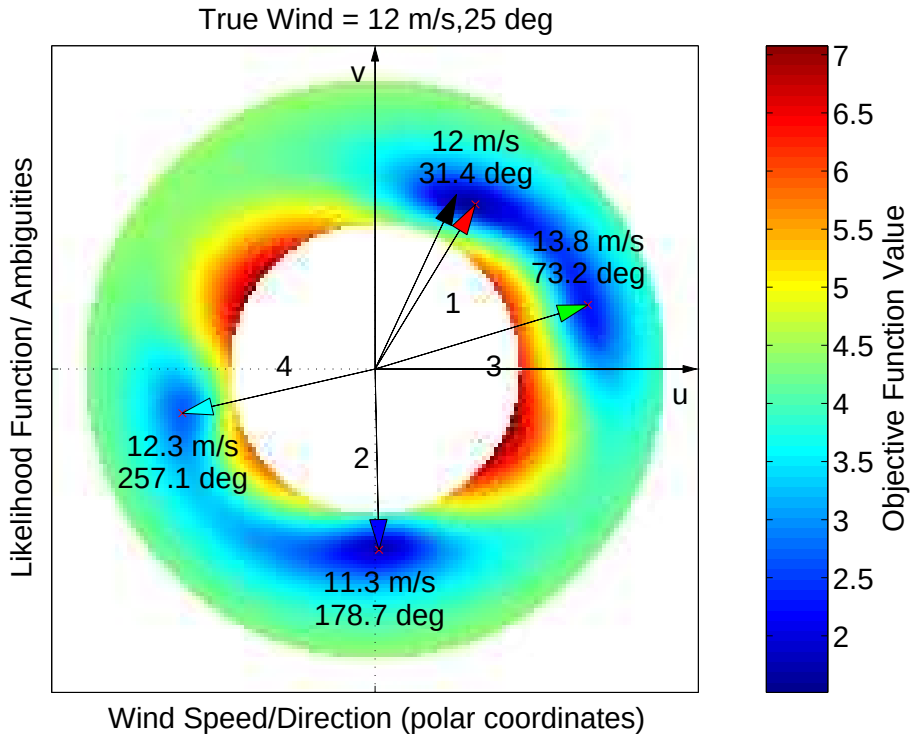


Figure 1.3: Sample objective function showing multiple minima, corresponding to ambiguities. The true wind is shown as a black arrow.

the influence of the outside data [14]. The un-nudged technique uses the field of first ambiguities as an initial wind field estimate. After an initial wind field is selected, a filtering process is used to "correct" isolated patches of "incorrect" initial wind solutions by replacing the initial ambiguity selection with ambiguities that are the most consistent with the surrounding selected ambiguities. A description of the point-wise filtering method used with the standard SeaWinds data product is given in the following section.

1.2 The SeaWinds scatterometer

As previously discussed, wind retrieval requires several measurements at different azimuth angles to reduce ambiguity in the estimated wind. The diversity in azimuth looks determines the accuracy of the retrieved winds. Instruments such as SASS and NSCAT utilized fixed fan beam antennas on either side of the spacecraft

Figure 1.4: The QuikSCAT satellite with the SeaWinds instrument.

oriented at different angles to obtain azimuthal diversity [10]. Such a design affords constant performance for all wind vector cells due to the fixed azimuth angles of the antennas. One draw back to the fan-beam design is that nadir-facing measurements cannot be utilized in wind retrieval because there is little wind directional sensitivity at low incidence angles. Thus, the data from fan beam instruments has a $\sim$ 300 km nadir region gap in coverage. To remedy this problem, the SeaWinds instrument is designed to have a constant incidence angle over the entire swath. The SeaWinds antenna is a rotating dish with two offset-fed antenna beams: a horizontally polarized (h-pol) beam at 46° incidence, and a vertically polarized (v-pol) beam at 54° incidence

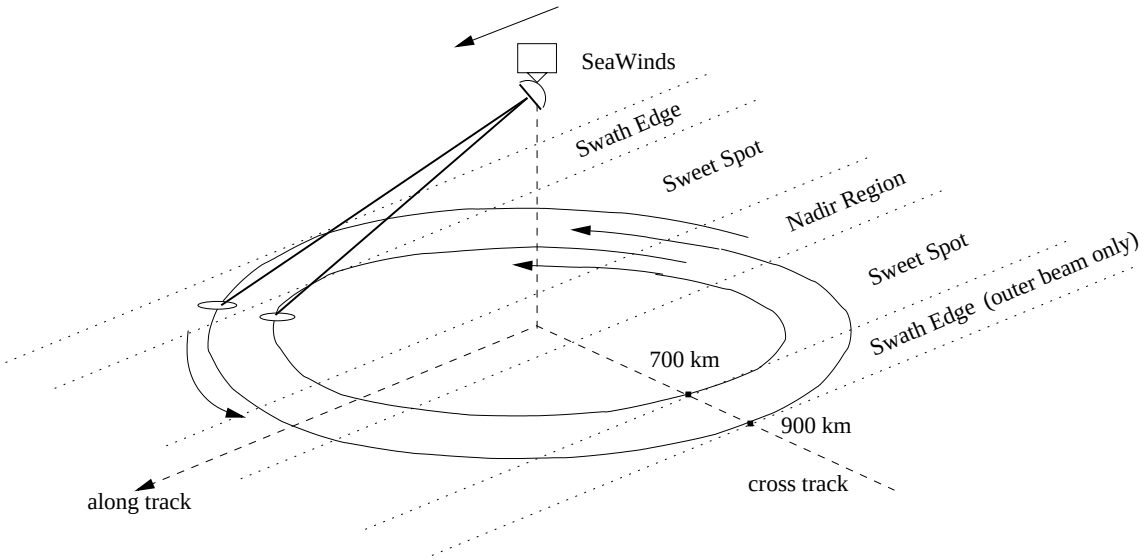


Figure 1.5: SeaWinds measurement geometry

(see Figure 1.4). The antenna traces a helical pattern on the surface and achieves the needed azimuthal diversity by measuring each point on the surface at least four times, twice by each beam fore and aft as the antenna rotates [1]. The constant incidence angles afford wind measurement in the nadir region due to the constant incidence angles, yielding a wider coverage (1800 km swath vs. 500-1200 km) than fan-beam instruments. The SeaWinds instrument orbits the earth in a near-polar orbit and yields 90% coverage of the earth in a 24-hour period.

Unlike fan-beam instruments, the pencil beam measurement geometry varies along the cross track. In the center of the swath (nadir region), the fore and aft beams are nearly 180 degrees apart, while the difference in azimuth between fore and aft beams go to zero on the swath edges. Also, the outer eight 25 km WVCs on either side of the swath only obtain measurements from the outer v-pol beam. Thus, the swath edges and nadir regions have somewhat poor viewing geometry for wind estimation. In the off-nadir inner-beam regions, known as the "sweet spots", the azimuthal diversity is very well suited for wind retrieval [15]. The measurement geometry for various swath positions is illustrated in Figure 1.6

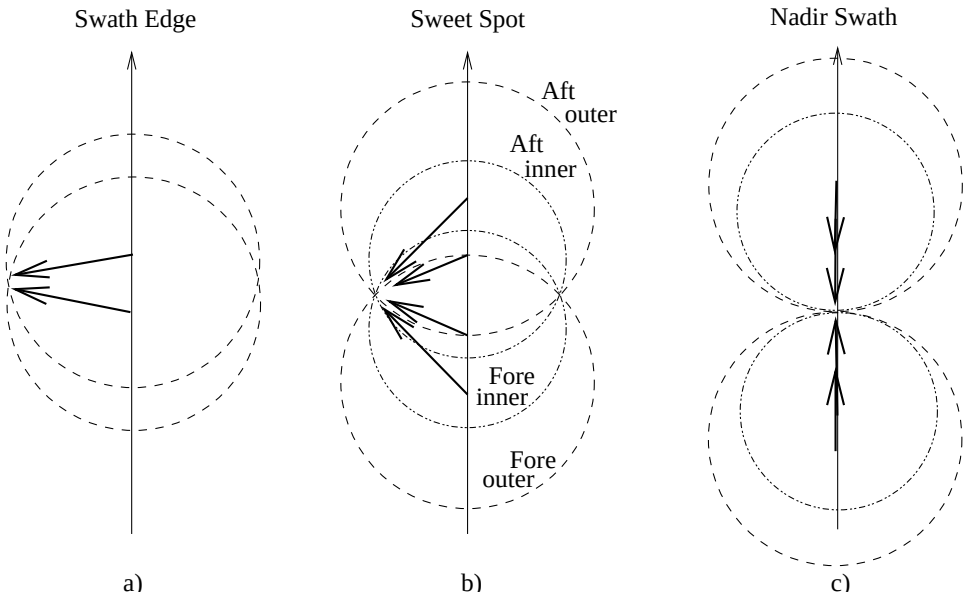


Figure 1.6: SeaWinds measurement geometry for several cases: a) swath edge, b) sweet spot, and c) nadir swath region.

In standard SeaWinds processing, the satellite swath is segmented into approximately 25 km square wind vector cells (WVC). The σ° measurements whose centers fall within the 25 $\times$ 25 km WVC are used to create a wind vector estimate at that WVC. WVCs that contain land or ice are discarded. The overall swath size for a single orbit is 76 $\times$ 1624 WVCs. Wind estimation is performed by finding the local minima to the point-wise objective function. Specifically for SeaWinds, the objective function is derived from a maximum likelihood estimation technique [12].

Assuming an independent Gaussian noise model, the probability of the retrieved measurements given the wind is

$$p(\sigma^\circ|\mathbf{u}) = \prod_k \frac{1}{\sqrt{2\pi}\varsigma_k} \exp\left\{-\frac{1}{2}\frac{(\sigma_k^\circ - \mathcal{M}_k(\mathbf{u}))^{\mathbf{2}}}{\varsigma_k^2}\right\} \tag{1.4}$$

where ς_k^2 is the measurement variance. The mean of the noise model is the GMF. The variance is a combination of uncertainty in the GMF, signal noise due to fading and thermal noise. Typically, the variance is defined in terms of K_{pm}, the normalized standard deviation of the GMF, and K_{pc}, the normalized standard deviation of the

communication or signal noise. The total variance on σ° is given by

$$\varsigma^2(\mathbf{u}) = (K_{pc}^2 + K_{pm}^2 + K_{pc}^2 K_{pm}^2)\mathcal{M}^2(\mathbf{u}) \tag{1.5}$$

where K_{pc} is generally written as

$$K_{pc} = \sqrt{\alpha + \frac{\beta}{\sigma_t} + \frac{\gamma}{\sigma_t^2}}. \tag{1.6}$$

The coefficients α, β and γ depend on fading characteristics of the surface scatterers, the receiver design, and the signal to noise ratio (SNR) at the receiver [15, 16]. Also, σ_t is the true σ° without communication noise.

For SeaWinds processing, K_{pm} is assumed to be constant, and σ_t is approximated by the model function estimate of σ°,

$$\sigma_t \cong \mathcal{M}(\mathbf{u}). \tag{1.7}$$

Using the approximation from Eq. (1.7), and combining Eq. (1.5) and Eq. (1.6) yields

$$\begin{aligned} \varsigma^2(\mathbf{u}) \;\; &\cong \;\; \alpha\mathcal{M}(\mathbf{u})^2 + \beta\mathcal{M}(\mathbf{u}) + \gamma \\ &\quad + K_{pm}^2 \left[(1+\alpha)\mathcal{M}(\mathbf{u})^2 + \beta\mathcal{M}(\mathbf{u}) + \gamma\right]. \end{aligned} \tag{1.8}$$

SeaWinds processing simplifies this expression by the assumption

$$(1+\alpha)\mathcal{M}(\mathbf{u})^2 \gg \beta\mathcal{M}(\mathbf{u}) + \gamma, \tag{1.9}$$

yielding, after some algebra,

$$\varsigma^2(\mathbf{u}) \cong [(1+\alpha)K_{pm}^2 + \alpha]\mathcal{M}^2(\mathbf{u}) + \beta\mathcal{M}(\mathbf{u}) + \gamma. \tag{1.10}$$

The coefficients kp_alpha $= (1+\alpha)$, kp_beta $= \beta$, and kp_gamma $= \gamma$ are computed and stored in the standard SeaWinds L2A product. The SeaWinds data processing estimates the wind via the simplified objective function formed by removing the $\frac{1}{\sqrt{2\pi}\varsigma_k}$ term in Eq. (1.4) and taking the negative logarithm, yielding a reduced form of the objective function $l(\mathbf{u}, \sigma^\circ)$,

$$l(\mathbf{u}, \sigma^\circ) = \sum_k \frac{(\sigma_k^\circ - \mathcal{M}_k(\mathbf{u}))^{\mathbf{2}}}{\varsigma_k(\mathbf{u})^2}. \tag{1.11}$$

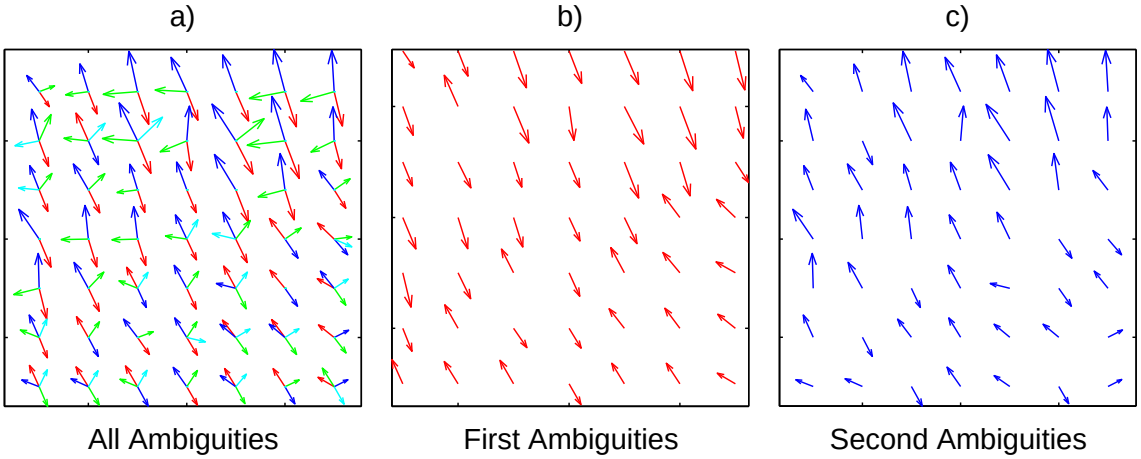


Figure 1.7: A sample QuikSCAT ambiguity field. Also shown, the first and second ambiguity fields.

Thus, the objective function is of the form of Eq. (1.3) with f the identity function, $p = 2$, and $q = 0$.

The objective function has one or more local minima corresponding to possible wind vector solutions or ambiguities. SeaWinds processing keeps the first four ambiguities from which it performs ambiguity selection. As explained earlier, ambiguity selection requires two steps: nudging and filtering. The standard SeaWinds product uses a variant of nudging known as thresholded nudging [14] with numerical weather prediction (NWP) winds from the National Centers for Environmental Prediction (NCEP). In thresholded nudging, the ambiguities whose objective function lies outside of a threshold from the global minimum are not used in the nudging process. Thresholded nudging inherently takes into account the instrument skill at the WVC. The instrument skill is defined as the estimated percentage of correct first ambiguities. Where the instrument skill is high, separation between the objective function values of the ambiguities is generally large, and thus, the least likely ambiguities are discarded in nudging. Where the instrument skill is low, the objective function values of each ambiguity are about the same, and all ambiguities are used. Figure 1.7 shows an example wind field and its ambiguities.

After nudging, a modified median filter iteratively selects the ambiguity at each WVC that most closely matches the flow of the surrounding selected wind. The point-wise median filter has the form

$$n = \underset{i}{\operatorname{argmin}} \sum_{j} d(\mathbf{u}_i, \mathbf{u}_j) \tag{1.12}$$

where i indexes the ambiguities at a given WVC, j indexes the surrounding 7×7 WVC region of selected wind vector, and n is the new selected ambiguity. For SeaWinds processing, $d(\cdot, \cdot)$ represents a directional error measure, defined as the difference in directions between the two vectors on the set 0 to 180 degrees. A vector difference has also been suggested as an error measure [17]. Thus, the median filter selects the ambiguity that minimizes the sum of directional errors between it an the surrounding selected WVCs. The median filter is repeated until convergence is reached [17].

The SeaWinds wind retrieval performance is not constant for all conditions. First, because the measurement geometry varies across the swath, the instrument skill is not uniform for every cross track position. Due to non-uniform measurement geometries, the wind retrieval process is somewhat ill-conditioned on the swath edges and at nadir [1]. Second, due to limitations in the GMF, noise, and a possible backscatter threshold wind speed, wind retrieval is less accurate at low wind speeds ($\lesssim 4$ m/s) [18]. Also at high wind speeds, the GMF saturates, making wind retrieval less accurate at high wind speeds ($\gtrsim 25$ m/s) [19]. Third, the wind retrieval process is somewhat sensitive to the nudging data used in ambiguity selection. Where the nudging data is less accurate, the ambiguity-selected winds retrieved winds are adversely affected [14]. Fourth, unmodeled geophysical phenomena such as sea temperature variations, salinity, foam, breaking waves, modulation by larger waves, and rain add to errors in the scatterometer winds. The effects from most of such phenomena are assumed to be small. However, rain backscatter has been shown to be significant and at times dominating [20].

1.3 Dissertation objectives

As indicated above, many factors can decrease the accuracy of scatterometer-retrieved winds. Therefore, my goal in this dissertation is to develop methods to detect, better understand, and mitigate errors in scatterometer wind estimation. I first perform a quality assurance (QA) analysis to determine the impact of error sources on the current wind product. I then specifically address the three main issues indicated by the QA analysis: the ambiguity selection procedure, low wind speed uncertainty, and rain contamination.

First, because the quality of ambiguity selection is sensitive to outside data used in the initialization step, I develop an advanced ambiguity selection algorithm using a low-order wind model rather than outside data to enhance self-consistency of retrieved winds. Using the new ambiguity selection method, I demonstrate that high-quality ambiguity selection can be obtained without the use of outside data. The new method independently produces winds that are similar to the nudged winds, thereby also validating the nudging approach.

Second, it is theorized that radar backscatter falls off below a wind speed threshold where capillary waves fail to be generated, creating low-quality wind estimates. Because the threshold wind speed is difficult to analyze directly with SeaWinds data, I use data from a tower-mounted scatterometer YSCAT to evaluate the affect of the threshold wind speed on radar backscatter in an uncontrolled marine environment. With the tower-mounted data, I illustrate a roll-off in σ° that occurs beneath the threshold wind speed for much of the data. Although my research is not designed to improve scatterometry through this analysis, I present valuable evidence of the threshold wind speed which aids to validate threshold wind speed theory and demonstrate its effects in uncontrolled data.

Third, to better understand the effects of rain corruption on radar backscatter, I develop a combined rain/wind backscatter model using co-located measurements from the Tropical Rainfall Measuring Mission precipitation radar to estimate the

impact of rain on SeaWinds scatterometer data. This research expands our understanding of both surface and atmospheric rain effects on Ku-band scatterometers, and suggests that rain can be retrieved using SeaWinds data.

Fourth, I use the wind/rain backscatter model to simultaneously retrieve wind/rain and rain from SeaWinds data. The simultaneous wind/rain retrieval method improves the wind estimate in the presence of rain and allows for SeaWinds scatterometer-based estimation of rain rate. I validate the simultaneous wind/rain procedure with passive and active rain measurement instruments from the Tropical Rainfall Measuring Mission satellite.

Fifth, I analyze the simultaneous wind/rain retrieval via simulation and the Cramér-Rao lower bound on estimate accuracy. This analysis illustrates the regimes where the simultaneous wind/rain retrieval works well and the limitations associated with the method. I show that using simultaneous wind/rain retrieval, the current rain flagging method can be improved.

Overall, this dissertation affords a greater understanding of the current issues associated with improving scatterometer data quality. In addition, the research presents practical methods of improving SeaWinds on QuikSCAT ambiguity selection and estimation quality in the presence of rain.

In Chapter 2, a quality assessment of SeaWinds on QuikSCAT data over a two-year period is presented. Chapter 3 describes and evaluates a new method of ambiguity selection that does not require nudging. In Chapter 4, I present an analysis of the threshold wind speed using tower-mounted scatterometer data. Then, Chapters 5, 6, and 7 develop, evaluate, and validate the simultaneous wind/rain retrieval method. Conclusions are given in Chapter 8.

Chapter 2

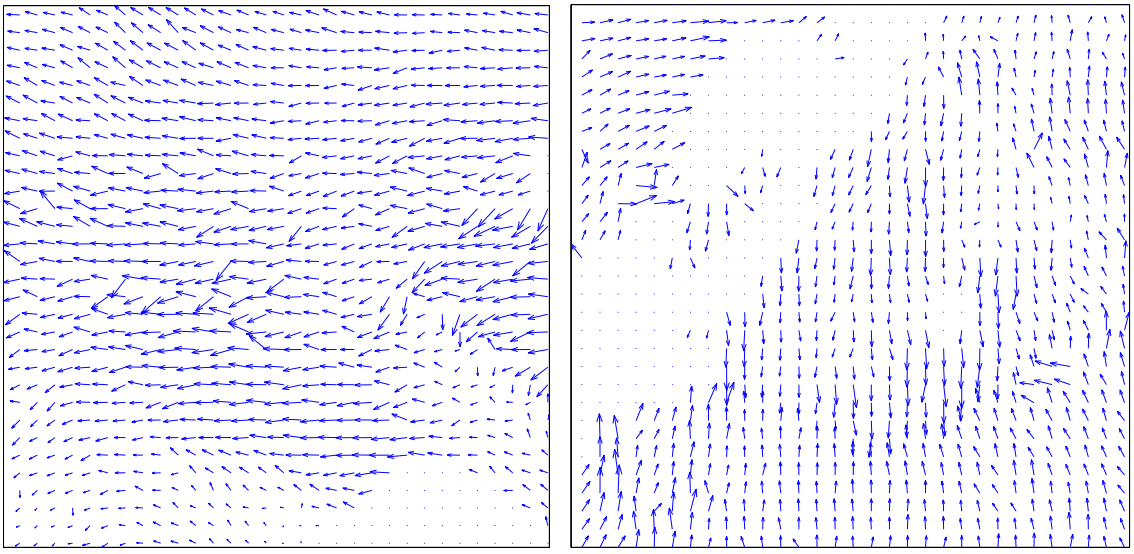
An Assessment of SeaWinds on QuikSCAT Wind Retrieval

This chapter assesses the quality of SeaWinds on QuikSCAT wind retrieval using a self-contained quality assurance (QA) method. The QA method uses a low-order wind model to assess the general quality of the data and examine the ambiguity selection skill. The QA analysis given in this chapter infers that the ambiguity selection at least 95% effective and that scatterometer errors are generally associated with instrument geometry, low wind speeds, fine-scale wind features and rain. Several issues indicated by the QA analysis are addressed in subsequent chapters, including the ambiguity selection procedure, low wind speed uncertainty, and rain contamination.

2.1 Error in scatterometer winds

Error in scatterometer winds can be divided into two major classifications: estimation error, and ambiguity selection error. Estimation error is uncertainty due to the estimation process and is evidenced by variability in the ambiguity closest the true wind. Estimation error is influenced by factors such as instrument noise, measurement geometry, rain contamination, and accuracy of the GMF. These factors generally cause the selected wind flow to appear "noisy."

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 (see Figure 2.1 for a comparison of noisy vectors and ambiguity selection error).



a) b)

Figure 2.1: a) Region containing some noisy vectors due to estimation error, poor measurement geometry, rain, etc.. b) Region containing an ambiguity selection error. Notice that the wind vectors in the central region oppose the surrounding flow.

Because both estimation noise and ambiguity selection errors can cause the retrieved winds to have inconsistent flow, the error due to these factors can be assessed by evaluating the general consistency of the ambiguity-selected wind [21].

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 [22, 23].

For this quality assessment, I adapt and extend a self-contained method developed by *Gonzales and Long* [21] for NSCAT. The NSCAT quality assurance assessment method compares ambiguity-selected winds to a low order Karhunen-Loève (KL) wind model fit. The model fit gives an estimate of the noise-reduced underlying wind flow. Regions exceeding error thresholds are assumed to contain either estimation or ambiguity selection error and are identified and tallied.

Our implementation of this model-based assessment method first addresses the overall consistency of the wind due to both estimation and ambiguity selection error. Then the method uses an empirically-determined set of error thresholds and other criteria to locate possible ambiguity selection errors in the presence of estimation noise (thus attempting to separate the effect of ambiguity selection error from estimation error). This technique does not evaluate the absolute correctness of the scatterometer data. Rather, it is a method of identifying inconsistent wind flow which suggests wind retrieval 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. Such errors are assumed to be infrequent. Although this model-based approach has limitations, it works well for most wind data has the distinct advantage of being self-contained.

In this chapter, the QA analysis method is described along with rationale for the selection of parameters. The KL wind model size and truncation point are discussed with an analysis of the KL model's data sensitivity in Section 2.2. The QA analysis method is described in detail in Section 2.3. The QA analysis is applied to two 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 2.4. Higher noise is shown to occur in low wind speed regions and at nadir. 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 2.5, indicating that scatterometer wind estimation is less effective in cyclonic storm regions.

2.2 Analysis of the KL wind model

The 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.

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 [24] 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 $\times$ 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. This is known as the standard vector form of the wind field. The autocorrelation matrix ($\hat{R}$) is constructed by combining the information from each wind field vector ($\mathbf{w_n}$) by

$$\hat{R} = \frac{1}{M}\sum_{n=1}^{M}\mathbf{w}_n{\mathbf{w}_n}^T \tag{2.1}$$

where M is the number of N $\times$ 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 representing the component of the wind with the most 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 [24]. The KL model basis vectors used in the QA analysis is shown in Figure 2.2.

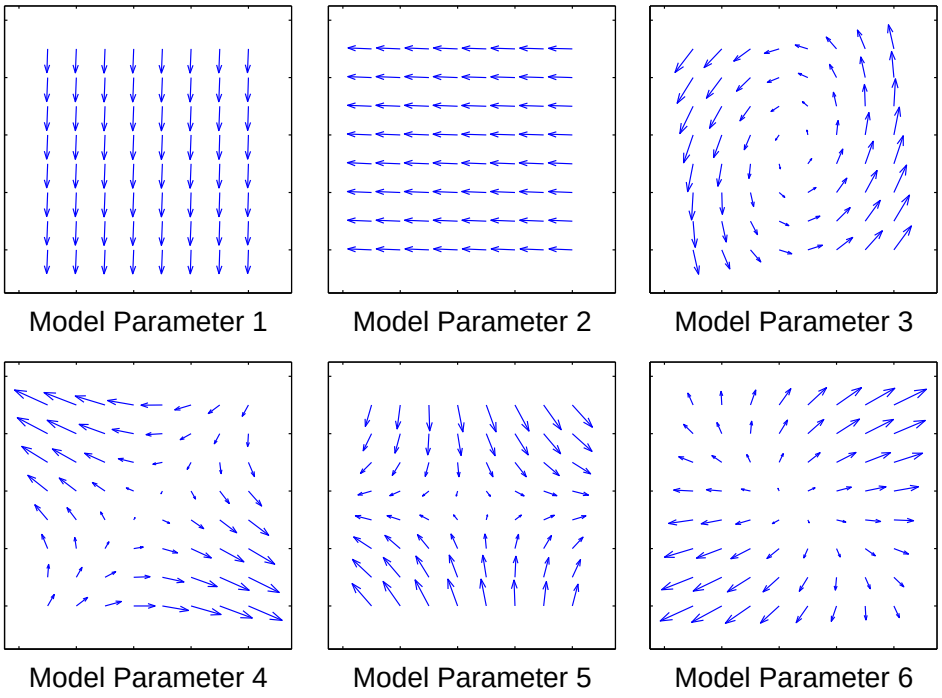


Figure 2.2: The truncated KL model used in the QA algorithm.

Because the wind has a generally red spectrum [25] (a spectrum where the lowest frequency content contains the highest energy), the low-order 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 [26]. 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 error, estimation error, or fine-scale wind features outside of the space spanned by the truncated basis, but are most commonly associated with estimation error and ambiguity selection error.

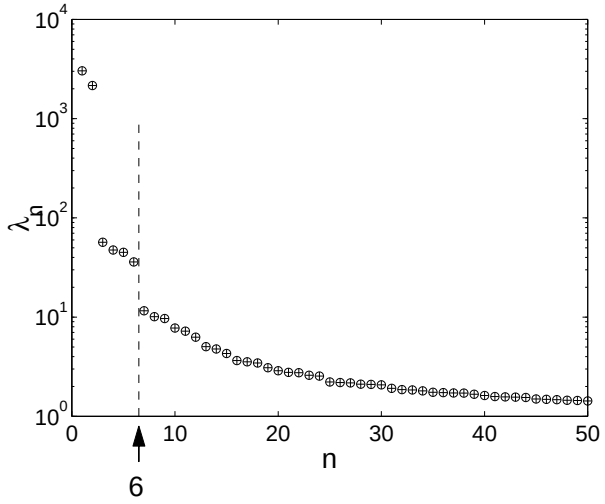


Figure 2.3: The first 50 eigenvalues of the 8 × 8 KL model.

2.2.1 KL model size and truncation point

The N × N size of the KL model used in the QA analysis method is determined to be 8 × 8, and the truncation point is determined to be 6 basis vectors via the following analysis.

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 any sized model as long as the number of basis vectors is chosen appropriately (see Appendix A). The size of the KL model used in the QA analysis is determined by a trade-off between the ability to pin-point 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

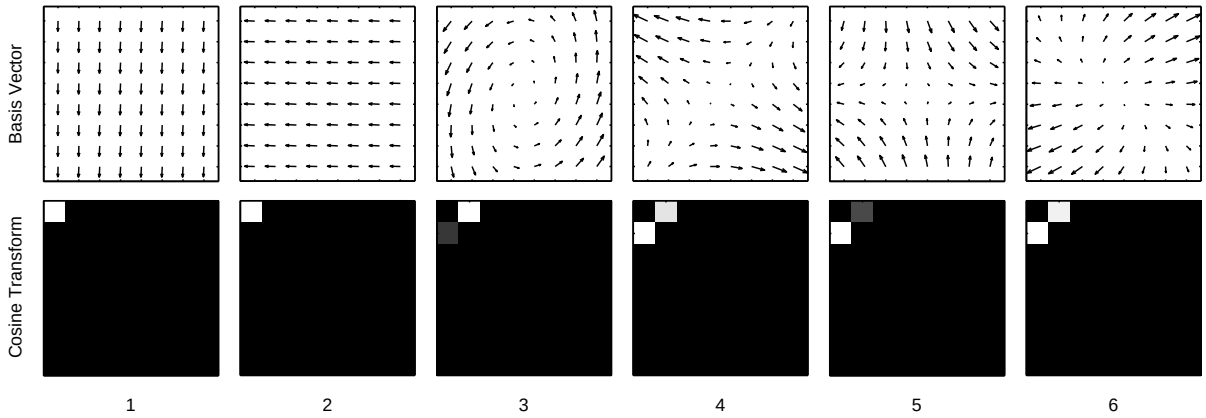


Figure 2.4: The wind fields corresponding to the first 6 KL basis vectors (top) and the magnitudes of their corresponding cosine transforms (bottom). The vertical dimension of the cosine transform corresponds to increasing frequency in the vertical direction. Likewise, the horizontal dimension of the cosine transform corresponds to increasing frequency in the horizontal direction.

not particularly sensitive to the model size as long as the truncation point is chosen such that the spatial frequency resolution is similar.

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 2.3), there are noticeable discontinuities in eigenvalues after the 2nd and 6th eigenvalues. Because the first 6 eigenvectors characterize over 95% of the wind variation, six is a reasonable truncation point for a simple, low-order model.

The discontinuities in the eigenvalues after the 2nd and 6th value can be explained by examining the spatial frequency of the eigenvectors via the 2-D vector cosine transform. Figure 2.4 displays the first 6 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 4 basis vectors represent various common wind features and are linear combinations of the (1,2) and (2,1) cosine basis terms. The succeeding 6 KL basis fields consist of the (1,3), (2,2), and (3,1) cosine bases (Figure 2.5). This

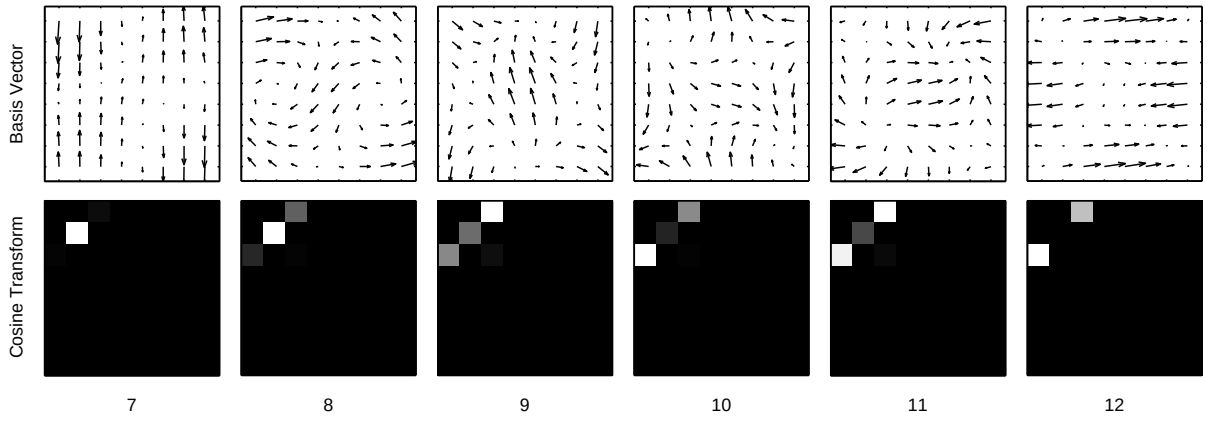


Figure 2.5: The wind fields corresponding to KL basis vectors 7 - 12 and the magnitudes of their corresponding cosine transforms.

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.

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

2.2.2 KL basis data sensitivity analysis

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 injected into the data set 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, and demonstrate that the space spanned by the KL model is insensitive to the artificially-induced ambiguity selection errors. The analysis suggests that the original model is unaffected by ambiguity selection errors inherent in the training data.

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, $A = [\mathbf{a_1}\ \mathbf{a_2}\ \cdots\ \mathbf{a_N}]$ and $B = [\mathbf{b_1}\ \mathbf{b_2}\ \cdots\ \mathbf{b_N}]$ where $\mathbf{a_n}$ and $\mathbf{b_n}$ are the n^{th} basis vectors of A and B respectively. We map each basis vector of A onto the space spanned by B via the inner product,

$$\hat{\mathbf{x}}_n = B^T \mathbf{a}_n. \tag{2.2}$$

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

$$L_{A,B} = \frac{1}{N} \sum_{n=1}^{N} \|B^T \mathbf{a}_n\|_2^2. \tag{2.3}$$

$L_{A,B}$ 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 Eq. (2.3), we obtain

$$\begin{aligned} L_{A,B} &= \frac{1}{N} \sum_{n=1}^{N} (B^T \mathbf{a}_n)^T B^T \mathbf{a}_n \\ &= \frac{1}{N} \mathrm{tr}\{(B^T A)^T B^T A\} \\ &= \frac{1}{N} \|B^T A\|_F^2 \end{aligned} \tag{2.4}$$

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

Table 2.1: Comparison of several KL models corrupted with artificially-induced ambiguity selection errors to the original (non-corrupted) KL model. 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 left column ($L_{A,B}$) where A is the corrupted KL model and B is the original model. A and B both have 6 basis vectors.

Induced ambiguity selection errors	$L_{A,B}$ 6 basis vectors
0%	1.0000
4%	0.9996
8%	0.9992
12%	0.9989
16%	0.9984
20%	0.9981

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 2.1. From this analysis, each 6-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 2.4.1), the low-order model is considered insensitive to the ambiguity selection of the training set.

2.3 QA analysis method

This section describes the ambiguity selection method in detail. 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 2.2, 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

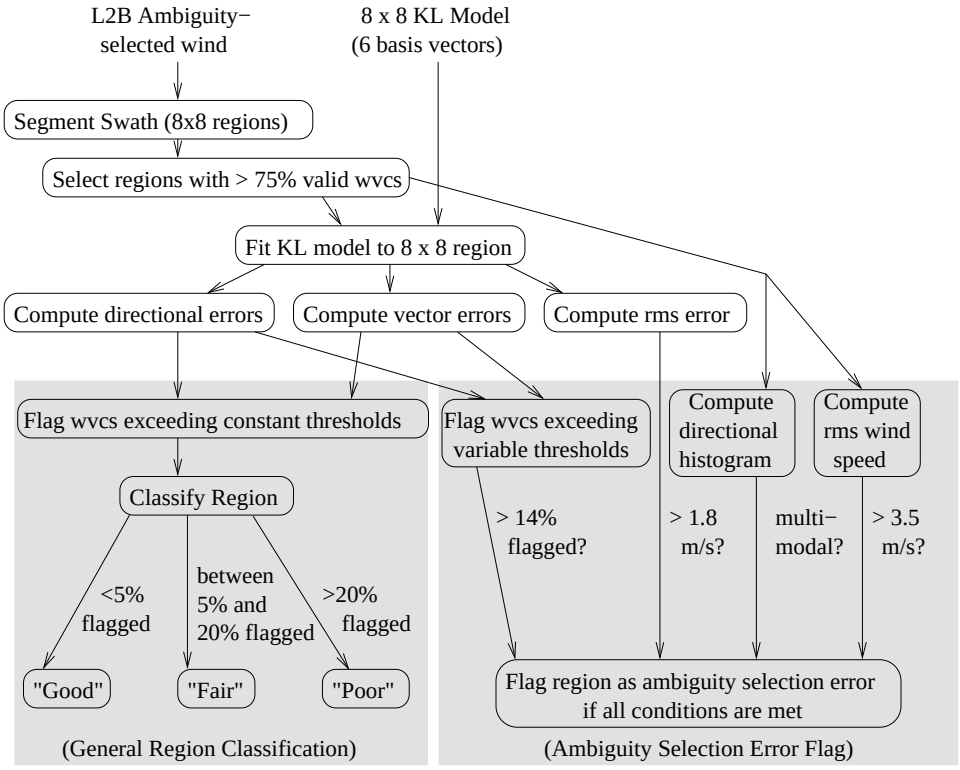


Figure 2.6: A flow diagram describing the QA analysis method. The left shaded box defines the region classification thresholds, and the right shaded box defines the thresholds involved with flagging a region as an ambiguity selection error.

can generally attribute large errors between the model fit and the ambiguity-selected wind to estimation error or ambiguity selection error.

Besides estimation 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 [21].

2.3.1 Overview

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 estimation error, ambiguity selection error, 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 estimation error in order to better locate those regions containing ambiguity selection error. A flow diagram of the QA method is shown in Figure 2.6 with a summary description given in the following sections.

Weighted least-squares model fit

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-squares fit,

$$\hat{\mathbf{w}}_m = F(F^T W F)^{-1} F^T W \mathbf{w}_o \tag{2.5}$$

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 in standard vector form defined in Section 2.2. 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.

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.

$$\phi_{\mathbf{e}} = |\phi_m - \phi_o|_{0 \le \phi_{\mathbf{e}} \le 180} \tag{2.6}$$

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.

$$k_e = \left((u_m - u_o)^2 + (v_m - v_o)^2\right)^{\frac{1}{2}} \tag{2.7}$$

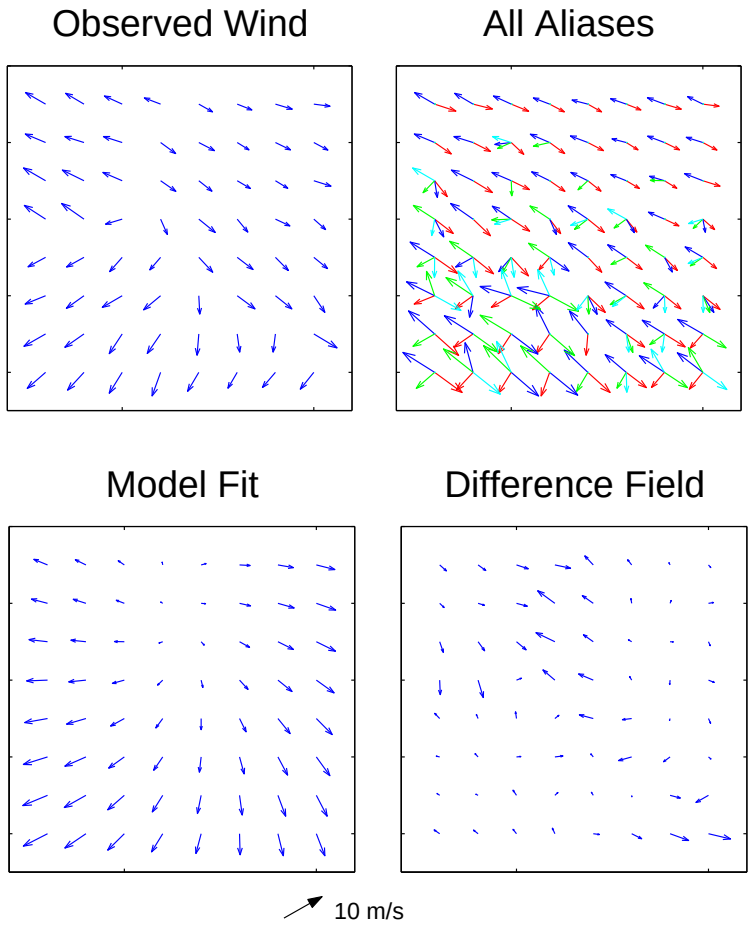


Figure 2.7: 8×8 Region of QuikSCAT point-wise selected wind and the KL model fit. Also shown are all possible ambiguities and the difference field between the observed wind and model fit.

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.

Figure 2.7 shows an 8×8 WVC region of point-wise selected wind and its least squares approximation. This example contains inconsistent flow evidenced by the divergent nature of the central section. The restricted basis cannot accurately represent this feature.

Table 2.2: Constant WVC thresholds determining the flagging of a vector. The term u_{rms} is the region root mean square (RMS) wind speed defined by $u_{rms} = (\mathbf{w}_o^T \mathbf{w}_o / N)^{\frac{1}{2}}$ 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. (†Denotes values used by *Gonzales and Long*, [1999]).

WVC threshold	value
direction error	$23°$ †
vector error	$\max \begin{cases} 2.7 \text{ †} \\ 0.5u_{rms} \end{cases}$ m/s

Region classification (good, fair, poor)

After generating a model fit and thresholding the wind vectors, the QA method first evaluates 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 2.3.1). These WVC thresholds are given in Table 2.2.

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 estimation 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 2.3. Bulk general region classification statistics for the QuikSCAT data set are given in Section 2.4.

Table 2.3: Thresholds determining the classification of a region. ([†]Denotes values used by *Gonzales and Long [1999]*).

Classification	Percentage of cells flagged per region
"Good"	$< 5\%$
"Fair"	5% - 20%
"Poor"	$> 20\%$ [†]

Ambiguity selection error detection

The next stage of the QA analysis is ambiguity selection error detection. Here, we suppress flagging of regions due to estimation noise error 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.

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 2.3.1). 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 2.8). 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. The variable thresholds suppress over-flagging of regions due to estimation noise. A full description of the method used to compute the variable thresholds is given in Appendix B.

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 help 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

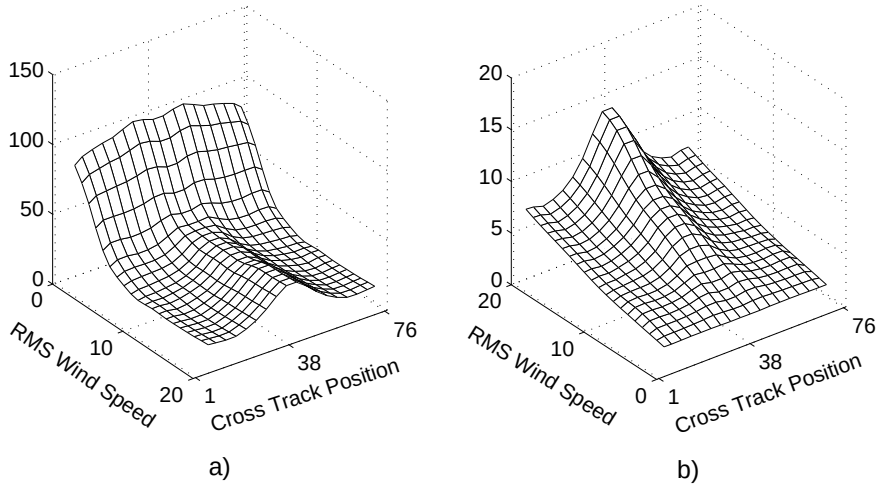


Figure 2.8: The a) directional and b) vector error thresholds per cross track and RMS wind speed that give a constant false alarm rate.

present, a high estimation 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. When a region fails both model-based and directional-histogram-based consistency checks, it is identified as containing possible ambiguity selection errors. These two consistency checks are described in more detail in Appendix C.

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 [21]. Similarly, through subjective examination of SeaWinds data, the estimation 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.

The QA data can be used in conjunction with the current L2B product via a WVC bit flag, created as an overlay product. Appendix D presents a bit flag that indicates the overall quality of retrieved winds in the swath using the QA analysis method.

2.3.2 Performance of the ambiguity selection error detection method on the training data set

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

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 2.9. 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 4 m/s bin because the data at these wind speeds tend to be somewhat noisier than for other wind speed bins.

2.4 QA analysis

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

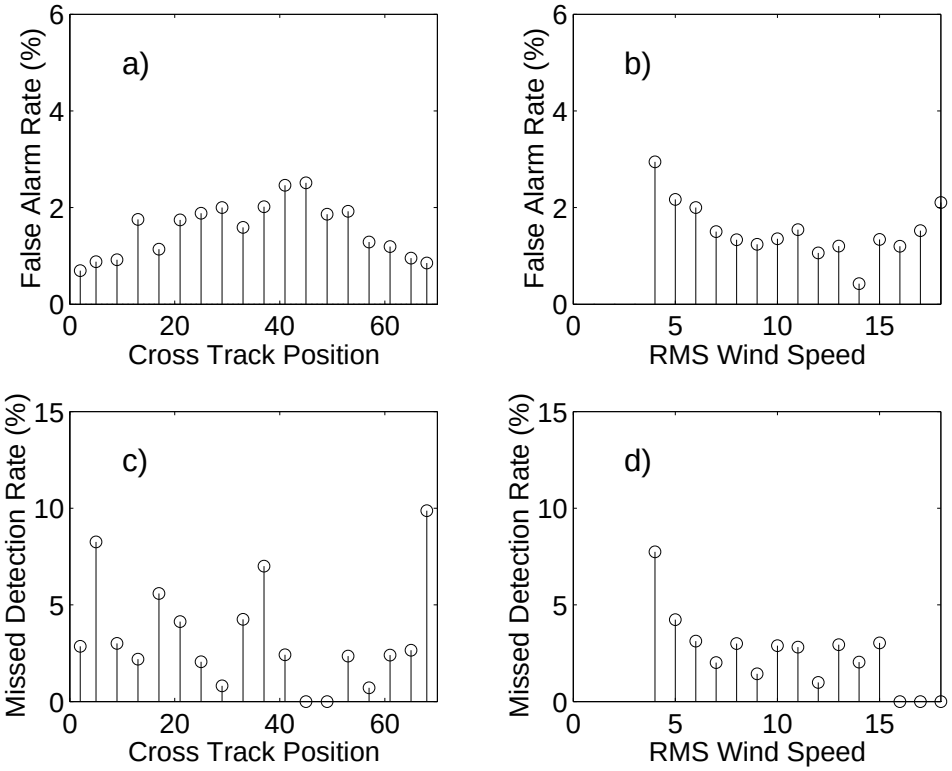


Figure 2.9: False alarm and missed detection rates for the ambiguity selection error detection method per cross track and RMS wind speed. This data is taken from a test set of 15 subjectively analyzed revs.

position and RMS wind speed. Lastly, 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.

2.4.1 Overall SeaWinds QA results

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 2.4 summarizes the percentage of 8 $\times$ 8 regions classified as "good," "fair," or "poor" using the method described in Section 2.3.1. Also given is the percent of regions flagged as possible ambiguity selection errors using the method described in Section 2.3.1. The majority of the regions examined are classified as "good" (> 65 %). This

Table 2.4: Overall results of the QA analysis for the SeaWinds data set and for the training data set. Also, the percent of ambiguity selection errors subjectively flagged in the training data set.

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)

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 estimation noise level or contain ambiguity selection errors. However, using the ambiguity selection error detection method, only five percent 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.

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.

The results given here for SeaWinds are very comparable to the QA results obtained for NSCAT [21]. 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

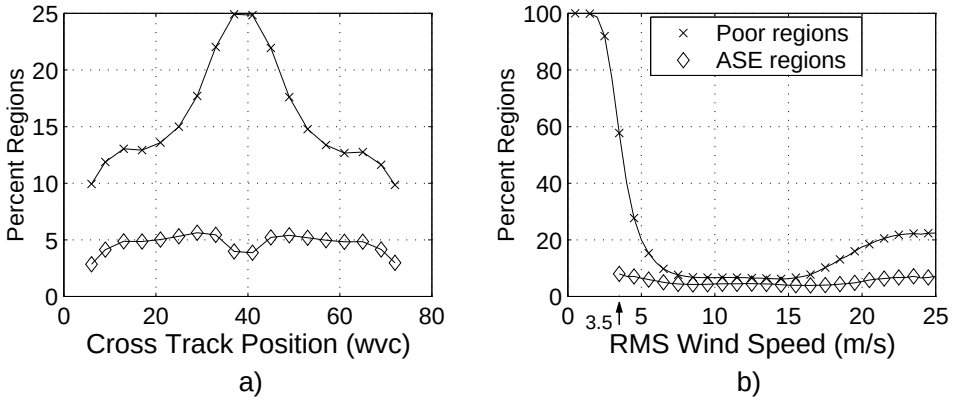


Figure 2.10: Percentage of all regions flagged as "poor" and flagged as containing ambiguity selection errors (ASE regions) per a) cross-track position and b) RMS wind speed.

determined to be at least 95% effective for wind speeds exceeding 3.5 and 4 m/s respectively.

2.4.2 Cross track/RMS wind speed dependence

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 2.10). Poor regions occur in higher percentages at nadir and where there is low wind speed. These areas generally contain higher estimation noise, and are thus more apt to be rated "poor." Poor wind retrieval at low wind speeds are additionally observed by [27].

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 1 to 3 ambiguity cases per cross track is compared to the fraction of 4 ambiguity cases averaged over 600 revolutions of SeaWinds data in Figure 2.11. The general shape of

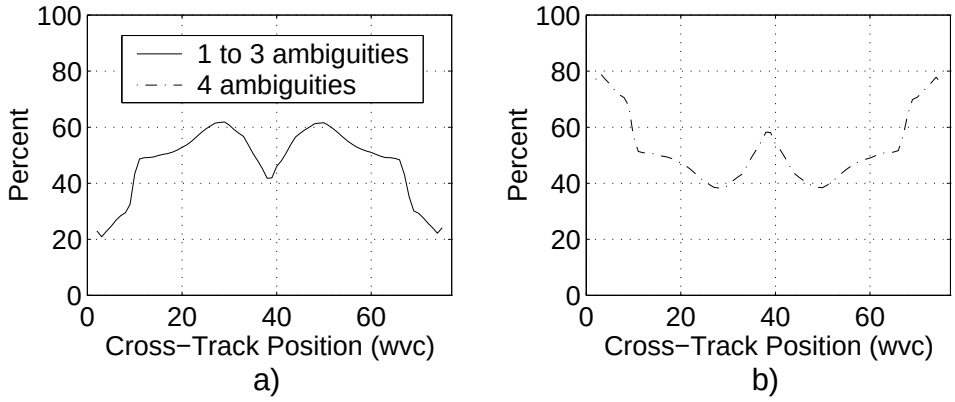


Figure 2.11: a) Percent of WVCs per cross track position with 1 to 3 ambiguities and b) Percent of WVCs per cross-track position with 4 ambiguities averaged over 600 revs of SeaWinds data.

the curve representing the 1 to 3 ambiguity cases of Figure 2.11 closely mirrors the percent of ambiguity selection errors per cross-track position shown in Figure 2.10. At nadir and on swath edges (where there are fewer estimated ambiguity selection errors), there is a higher likelihood of having four ambiguity choices.

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 2.4.3 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 [14], 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.

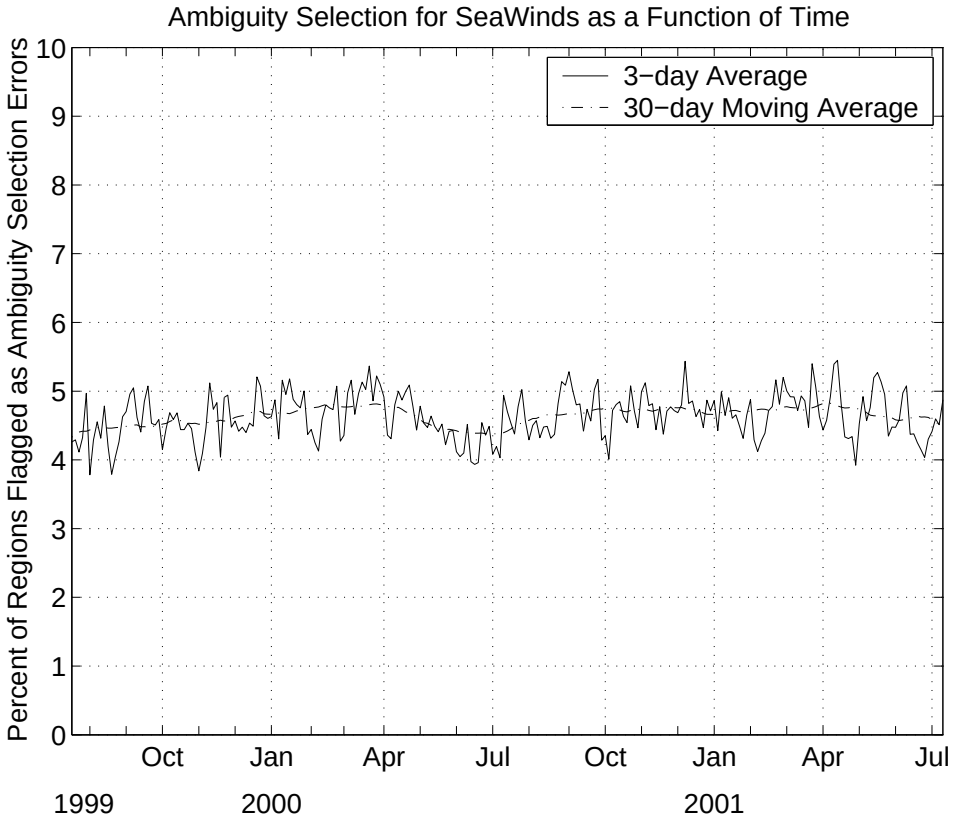


Figure 2.12: Ambiguity selection errors as a function of time. The data line is the fraction of ambiguity selection errors averaged over 3 days. The smooth line is a 30 day moving average.

2.4.3 Temporal QA statistics

Next, we examine SeaWinds ambiguity selection as a function of time. Figure 2.12 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 two 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.

In order to understand weather pattern variations that affect SeaWinds' performance, we divide the QuikSCAT wind data into latitude bands (see Table 2.5). For each band, the average percentage of possible ambiguity selection errors detected

by the QA method, the average number of cyclonic storms passed by SeaWinds per degree latitude (see Section 2.5 for details on cyclonic storm detection), and the number of WVCs flagged by the JPL rain flag [28] per day are computed and given in Figure 2.13.

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 mid-latitude 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.

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.

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 [21]. 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 2.5, the wind retrieval is worse in storm regions.

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 [5]. In addition, rain disrupts the wind-generated capillary waves on the ocean surface, altering the wind-induced backscatter signature [29]. 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

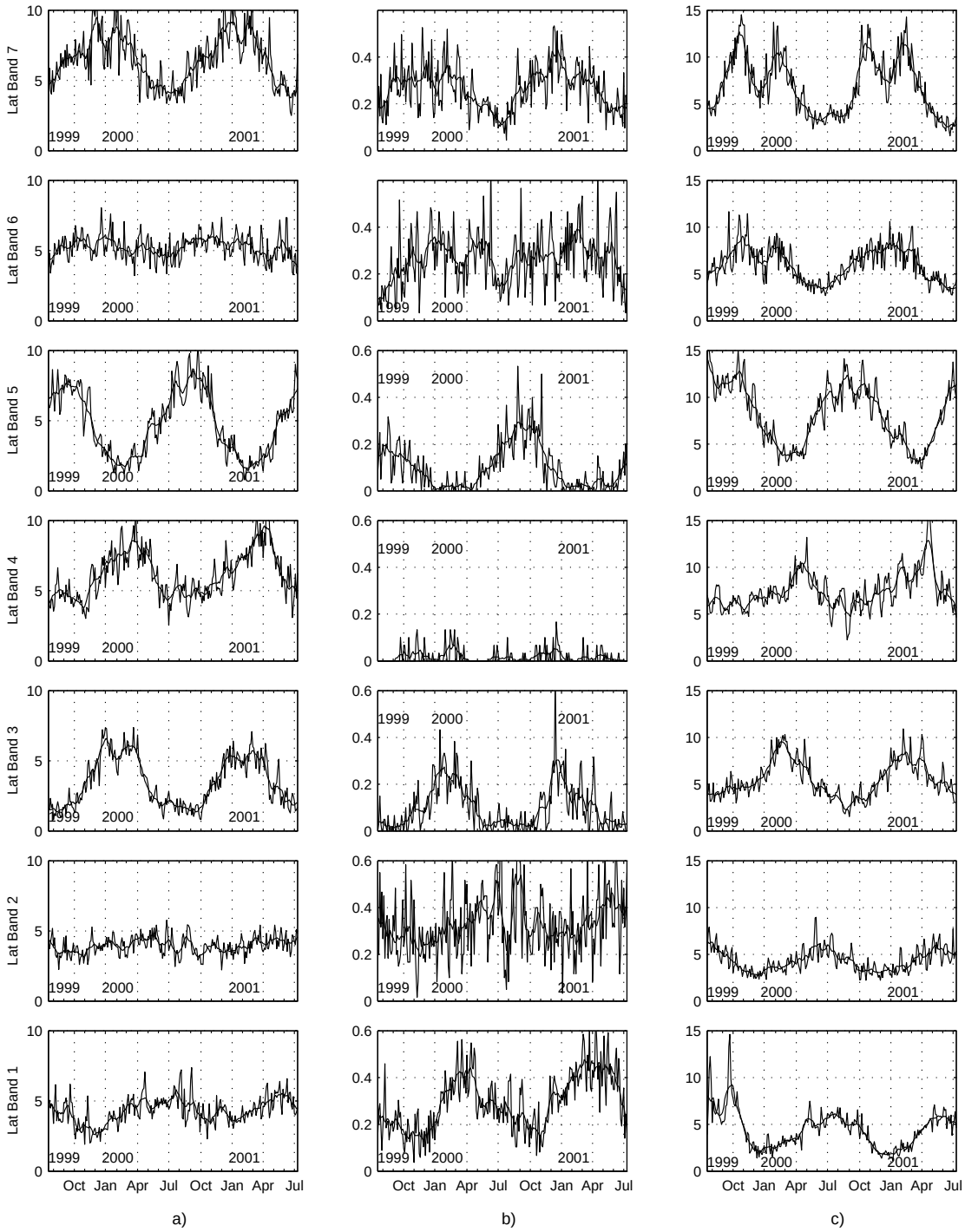


Figure 2.13: a) Percent of ambiguity selection error regions, b) number of cyclonic storms per degree latitude, c) percent of WVCs flagged with L2B rain flag for each latitude band averaged over 3 days per data point.

Table 2.5: 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°

the cross track because rain is a nearly isotropic scatterer, giving an equal response for both fore and aft looking observations. Likewise, winds blowing cross-wise to the satellite flight direction also give a near-equal response from fore and aft observations [28]. Rain effects on wind scatterometer backscatter returns have been found to be significant [29, 30].

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 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.

From visual inspection of Figure 2.13, 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 2.6 for each latitude band. The highest correlation occurs in the 3rd, 5th, and 7th latitude bands. This correlation suggests that rain and storms contribute to poor ambiguity selection and inconsistent wind flow.

Table 2.6: Correlation coefficients for each latitude band. C_{sa} is the correlation coefficient between the smoothed ambiguity selection error and storm data of Figure 2.13. C_{ra} is the correlation coefficient between the smoothed ambiguity selection error and rain data of Figure 2.13.

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

The equatorial band (latitude band 4) has the fewest storms. Storms around the equator are rare because the Coriolis 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.

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 2.13b and 2.13c, 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 5th 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.

Though more total storms occur in bands 2 and 6 (mid-latitude regions) than in the tropics, the mid-latitude 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 3 and 5, resulting in more seasonally uniform performance of scatterometer wind retrieval. Mid-latitude rain averages are higher in the northern hemisphere, causing a slightly worse performance of SeaWinds ambiguity selection in this region.

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 land masses 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 2nd 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.

2.5 Subjective analysis of cyclonic storm regions

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 2.4). Where the mean square of parameters 3 and 6 is greater, the region is flagged as a cyclonic storm.

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

Table 2.7: Number of storms identified as rating "1", "2", and "3" for two time periods: July 1999, and January 2000. For each time period, the data is further subdivided into northern and southern hemisphere locations.

	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

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 2.14.

Two separate time periods are examined: 2 weeks of QuikSCAT data from July 19th to August 1, 1999, and 2 weeks of QuikSCAT data from January 17th to January 30th, 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 2.7.

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.

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 2.8). The northern hemisphere July 1999 case exhibits the statistically best storm retrieval and also demonstrates

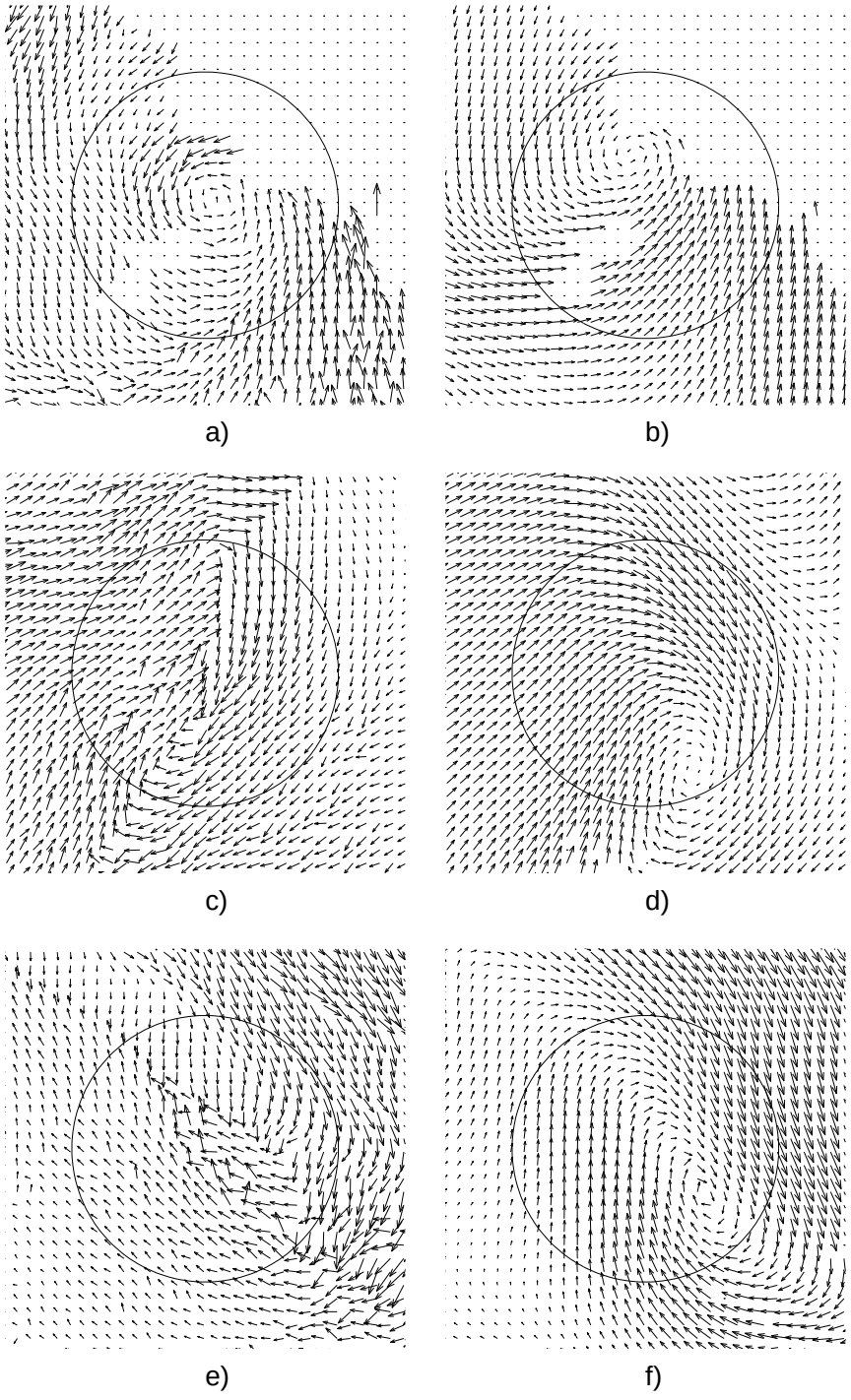


Figure 2.14: a) QuikSCAT storm region rated "1", b) corresponding NCEP data. c) QuikSCAT storm region rated "2", d) corresponding NCEP data. e) QuikSCAT storm region rated "3", f) corresponding NCEP data.

Table 2.8: Average distance (km) between the NCEP storm centers and the L2B storm centers for storms rated 1 or 2.

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

the best collocation between NCEP and QuikSCAT storms. This example suggests that more accurate nudging in this area improves the wind retrieval performance.

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.

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 mis-location 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 E.

2.6 Discussion and summary

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.

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.

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.

Although limitations in scatterometer retrieved winds exist, point-wise estimated scatterometer winds are of very high quality in non-rain 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 [31, 14, 27, 32]. In addition, research is ongoing to improve the GMF at low, *e.g.* [33], and high wind speeds, *e.g.* [34].

In the following chapters, three main issues indicated by the QA analysis are addressed: ambiguity selection, low-wind speed variability, and rain modeling and correction. Through the following analysis, I demonstrate that

1. Quality ambiguity selection can be performed without the use of nudging data.
2. A low wind speed threshold in backscatter may be a significant contributer to scatterometer error at low wind speeds.
3. Using a simple rain backscatter model, the effects of rain contamination in SeaWinds data can be reduced.
4. Rain can be measured using the SeaWinds scatterometer.

Chapter 3

An Advanced Ambiguity Selection Algorithm for SeaWinds

The QA analysis presented in the previous chapter suggests that ambiguity selection errors contribute to inconsistent wind vector estimates in about 5% of SeaWinds data. In order to improve the self-consistency of the ambiguity selected wind flow, this section describes an advanced KL-model-based method of ambiguity selection [35]. The new method avoids using outside data in its initialization and improves the overall self-consistency of the wind. The new ambiguity selection method produces winds similar to the JPL product and demonstrates that ambiguity selection can be accomplished without nudging.

3.1 Background

In conventional SeaWinds point-wise ambiguity selection, NASA's Jet Propulsion Laboratory (JPL) uses a nudging method to initialize the ambiguity selection process and enhance self-consistency [14]. As explained in Chapter 1, nudging selects the ambiguity at each wind vector cell (WVC) that most closely matches the flow of numeric weather prediction (NWP) fields to initialize the point-wise median filter. The main limitation to the nudging method is that it creates a dependence on the quality of outside information to select a unique solution from the ambiguity sets.

As an alternative to the nudged point-wise winds, Long developed a field-wise wind estimation technique [11]. In field-wise wind retrieval, estimates are made on a region-by-region basis using a low-order linear wind field model. The low-order model assumes an inherent correlation between neighboring wind vectors, which restricts the solution to a wind field satisfying the correlation constraints.

In addition to its field-wise utility, the low-order wind model can also be used to make point-wise ambiguity selection. Gonzales and Long demonstrated that some ambiguity selection errors in point-wise retrieved winds can be corrected by selecting the ambiguity that is closest to a least-squares model fit to the JPL winds [21].

This chapter describes a non-nudging point-wise ambiguity selection method for SeaWinds developed at Brigham Young University (BYU) [35]. The BYU method uses the low-order KL wind model to create an initial estimate of the overall wind flow which replaces nudging. A model-based correction technique then reselects ambiguities where the data is not self-consistent. The new method is self-contained and computationally efficient.

In Section 3.2, we describe the BYU technique in detail. Because we lack truth data, the ambiguity-selected JPL winds are used as a reference data set to which the BYU method is compared in Section 3.3. We find that the BYU method generally selects the same ambiguities as traditional JPL point-wise ambiguity selection with more self-consistent performance in winds with low spatial variability. The BYU method, however, is somewhat less able to correctly define fine-scale cyclonic flow than the JPL nudged method. The fact that both the BYU and JPL methods produce approximately the same result simultaneously helps validate both the nudged and non-nudged approaches.

3.2 Overview of BYU point-wise ambiguity selection

In this section, the BYU algorithm is described in detail. To avoid nudging, we utilize the KL model introduced in Chapter 2 to determine an initial ambiguity selection over the high instrument skill inner-beam portion of the swath. The point-wise median filter is initialized from this initial selection. The swath edges, where the instrument skill is lower, are separately estimated by extrapolation of the inner-beam wind flow. A correction routine locates and repairs inconsistencies in the selected wind field. The masking and repairing steps are repeated until the wind field meets convergence criteria. Fig. 3.1 outlines the steps involved in the BYU method. Section

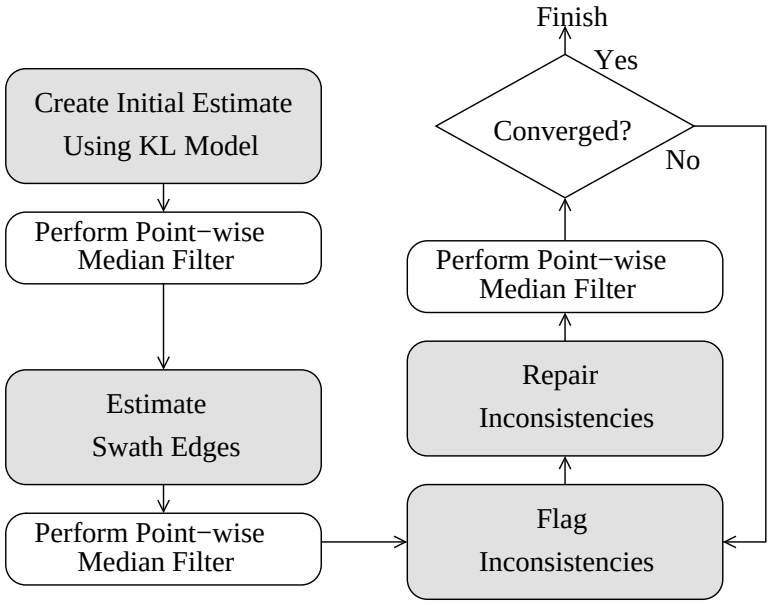


Figure 3.1: Flow chart of the BYU ambiguity selection method.

3.2.1 describes the initial estimate. Section 3.2.2 describes ambiguity selection of the swath edges. Lastly, section 3.2.3 details the repair process.

3.2.1 Initial estimate

Methodology

As described in Chapter 1, the SeaWinds sub-track has three distinct regimes with different wind retrieval characteristics: the swath edge, the sweet spot, and the nadir region. Each regime has a different instrument skill or percentage of correct first ambiguities. For SeaWinds, each WVC in the inner-beam region, including the sweet spot and nadir region, has at least two fore-looking observations (one for each beam) and at least two aft-looking observations, which provide sufficient measurement density and azimuthal diversity to afford a high instrument skill. On the swath edges, poor instrument geometry and observations from the outer-beam only produces a much lower instrument skill. Also, the outer two cross track positions on either side (1,2,75 and 76) are not always estimated in point-wise retrieval due to very poor instrument geometry. Thus, the effective wind swath width is 72 WVCs.

The high instrument skill of the inner-beam portion of the SeaWinds swath affords estimation of the main wind flow using the first ambiguities. Although we cannot calculate the actual instrument skill without truth data, a reasonable estimate is the average percentage of first ambiguities selected by the JPL method. On average, the JPL ambiguity selection method selects over 70% of first ambiguities in the inner-beam portion of the swath. In the outer-beam portion of the swath, only about 35% first ambiguities are selected. Because a vast majority of WVCs in the inner-beam region have "correct" first ambiguities, the inner-beam first ambiguity field can be utilized to initialize the ambiguity selection process.

Incorrect first ambiguities can be either isolated or clustered. In general, isolated incorrect first ambiguities can be corrected by simply applying the point-wise median filter to the field of first ambiguities. Small to large clusters of incorrect first ambiguities on the order of half the filter size or larger remain incorrect after median filtering. These errors can significantly affect the ambiguity-selected wind flow.

Rather than simply using the field of first ambiguities for an initial ambiguity selection, the BYU method replaces first ambiguities that oppose the main flow by second ambiguities. Since the JPL method selects over 90% first and second ambiguities in the inner-beam region, a first/second ambiguity field yields better final solution. Further, where neither the first or second ambiguity produces a self-consistent solution, the ambiguity selection may be replaced by a third or fourth ambiguity to create an even better initial estimate. In this way, ambiguities are inserted into the swath based on priority. This priority-based ambiguity selection is the basic idea behind the BYU initial estimate.

In order to replace clusters of incorrect first ambiguities with second ambiguities, a low-order constraint is placed on the wind flow of the first ambiguity field. To do this, an initial low-order KL model fit is made to the first ambiguity field. High spatial frequency content caused by small clusters of incorrect first ambiguities are smoothed by the model fit. All WVCs are set to the nearest first or second ambiguity. This step replaces isolated vectors and small to medium clusters of incorrect first ambiguities with second ambiguities. The order of the KL model dictates the

"smoothness" of the model fit, and thus influences the cluster size that is replaced by second ambiguities. Wind vector cells exhibiting large errors between the first/second ambiguity field and the model fit are flagged. Next, a second low-order model fit is performed to the first/second ambiguity field, weighting out flagged WVCs. Where neither first or second ambiguities are consistent with this model fit, the closest third or fourth ambiguity may be chosen. Figure 3.2 illustrates the initial estimate.

Although this initial estimate corrects small to medium clusters of incorrect vectors (depending on the KL model order), large regions where the first ambiguity field is incorrect must be repaired separately with a more sophisticated routine (see Section 3.2.3).

Detailed description of the initial estimate: application to SeaWinds swath

The large swath size prohibits applying a KL model to the entire swath due to computational concerns. As a result, the swath is divided into 60 × 60 WVC sections overlapping by 75% in the along track direction. The 60 × 60 region size is chosen because it spans the entire inner-beam portion of the cross-track. A square region size is chosen for historical reasons [11, 21]. Since the model fit requires the inversion of a matrix that has on the order of N^4 elements where N is the width of a region, each 60 × 60 section is additionally decimated into 9 interleaved fields of 20 × 20 WVCs. This is equivalent to increasing the measurement spacing from 25 km to 75 km for each field. Segmentation and decimation allow the use of a reasonably sized model to reduce the computational expense of the problem.

For each 20 × 20 decimated region, a model fit is made to the first ambiguity field. Non-ocean WVCs and WVCs that contain significant rain contamination as determined by the L2B rain flag [28] are weighted out of the model fit. The model fit is computed via a weighted regularized least-squares estimate,

$$\hat{F}^{\dagger}_{opt} = (F^T W F + \Lambda_F^{-1})^{-1} F^T W \tag{3.1}$$

where Λ_F contains the eigenvalues of the basis vectors in F and W is a weighting matrix with diagonal elements of "1" corresponding to valid data cells and "0" corresponding to non-data cells or cells that are to be ignored. The coefficients for the

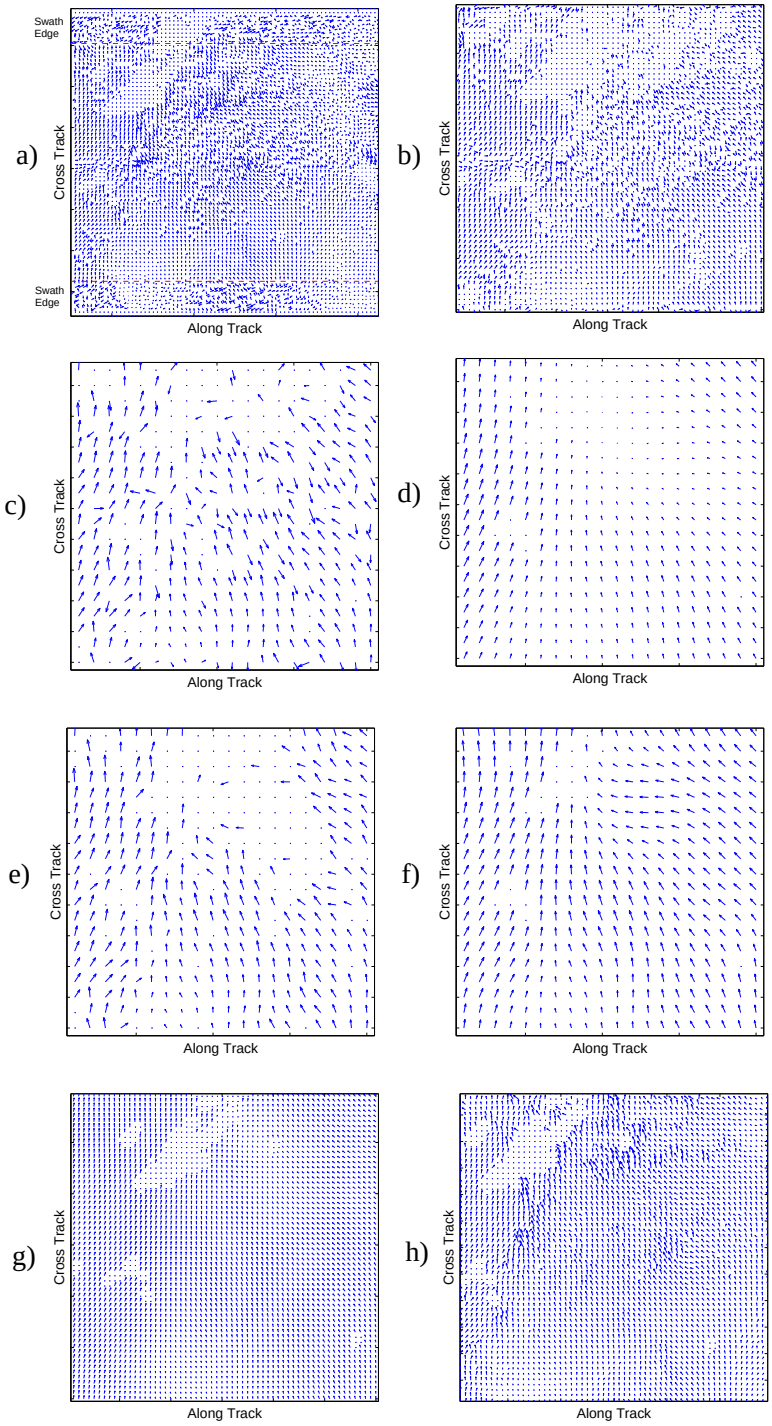


Figure 3.2: Initial ambiguity selection method example: a) First ambiguity field, b) Inner-beam portion with rain masked out, c) Decimated 20 × 20 field d) First model fit, e) First and second ambiguities with inconsistent flow masked out, f) Second model fit, g) reconstructed 60 × 60 field, h) Closest ambiguity to second model fit.

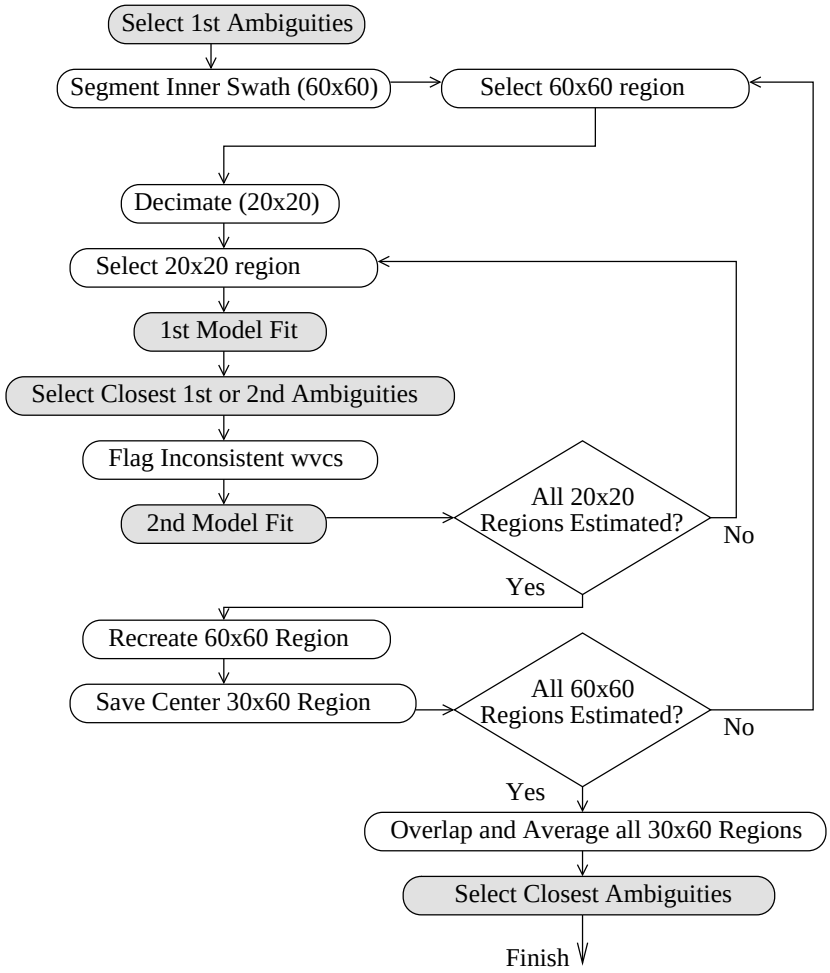


Figure 3.3: A flow diagram summarizing the initial estimate. The shaded bubbles show the most important steps.

basis fields, $\hat{\mathbf{x}}_{opt}$, are

$$\hat{\mathbf{x}}_{opt} = \hat{F}^{\dagger}_{opt}\mathbf{w} \tag{3.2}$$

where $\mathbf{w}$ is the standard vector form of the wind field. The model fit field $\hat{\mathbf{w}}_{opt}$ is constructed by the equation

$$\hat{\mathbf{w}}_{opt} = F\hat{\mathbf{x}}_{opt}. \tag{3.3}$$

This model fit is regularized by the eigenvalues to give a realistic solution to points that have been weighted out.

After a model fit is made to the first ambiguity field, all WVCs in the 20 $\times$ 20 region are set to the nearest first or second ambiguity to the model fit. The directional and vector error between the new ambiguity-selected field and the model fit are computed for each WVC over the region. Cells where the directional error exceeds 45° or the vector error is greater than the average wind speed of the region are flagged as poor. A second higher order model fit is then made in which vectors flagged as poor are weighted out. The second model fit interpolates new values for the flagged cells.

After second model fits are made to all 9 fields corresponding to a 60 $\times$ 60 region, the entire 60 $\times$ 60 region is reconstructed by interleaving the second model fit fields. The u and v components of the 60 $\times$ 60 reconstructed field are then median filtered to ensure consistency among the interleaved fields. To rebuild the entire swath, the center 30 along track rows are saved. All along track sections are likewise estimated and each section is overlapped and window averaged. A simple triangular averaging window is used in our implementation, although the shape of the window is not critical. This creates a low resolution initial wind field close to the wind flow dictated by the first and second ambiguities. Lastly, each WVC is set to the nearest ambiguity to the low-resolution wind field. All ambiguities are included in the selection process. The result is an ambiguity-selected field dictated mainly by the flow of the first and second ambiguities. The steps involved with creating the initial estimate are shown in Fig. 3.3.

The initial ambiguity-selected field replaces the nudging data used in traditional point-wise estimation. Next, the point-wise median filter described in Chapter 1 is employed to insure self-consistency among each selected ambiguity and its neighbors. This is the same point-wise median filter implemented by JPL, although JPL initializes the filter with NWP model fields.

3.2.2 Estimating swath edges

The initial selected ambiguity field only includes the inner-beam portion of the swath. Due to low instrument skill in the outer beam region, ambiguity selection

of swath edges must be performed separately. We make no assumption about the correctness of the first ambiguities on the swath edges. Instead, we use the wind flow of the inner-beam region to infer a solution for the outer-beam region by extrapolating values for the outer cells via the KL model.

The swath edges, including the outer-beam WVCs and the outer edge of the inner-beam, are divided into 16 $\times$ 16 WVC regions overlapping by 50% in the along track direction. These regions contain 9 cross-track rows assigned unique ambiguities by the previous steps and 7 cross-track rows of unselected outer-beam WVCs. The choice of $N = 16$ is arbitrary, but provides a good compromise between computational efficiency and inclusion of inner-beam data. The outer cross-track row is not included because wind retrieval is not performed there. A low-order model fit is made to each 16 $\times$ 16 region using only the inner-beam wind vectors. This yields an estimate of the outer 7 cross track rows via interpolation of the model fit. A new 16 $\times$ 16 field is constructed from the closest ambiguity to the model fit. If the root mean square (RMS) error between the closest alias field and the model fit falls beneath a threshold, the new ambiguity selections are inserted into the swath. This threshold is set at 1.5 m/s and is relaxed (increased by 1.5 m/s) for each pass until all WVCs have a unique vector selected. Thus, the best solutions are inserted first and used to influence subsequent estimates. The swath is again passed through the point-wise median filter after estimating the swath edges. The estimation of the swath edges is summarized in Fig. 3.4.

3.2.3 Repairing inconsistencies

The previous steps yield a unique ambiguity at each WVC. Although this selected ambiguity field could be the final product, due to possible large clusters of incorrect first ambiguities in the initial estimate, there still may be areas of significant ambiguity selection errors. Clusters of ambiguity selection errors are generally characterized by 90 to 180 degree shifts in the selected wind flow. The transition

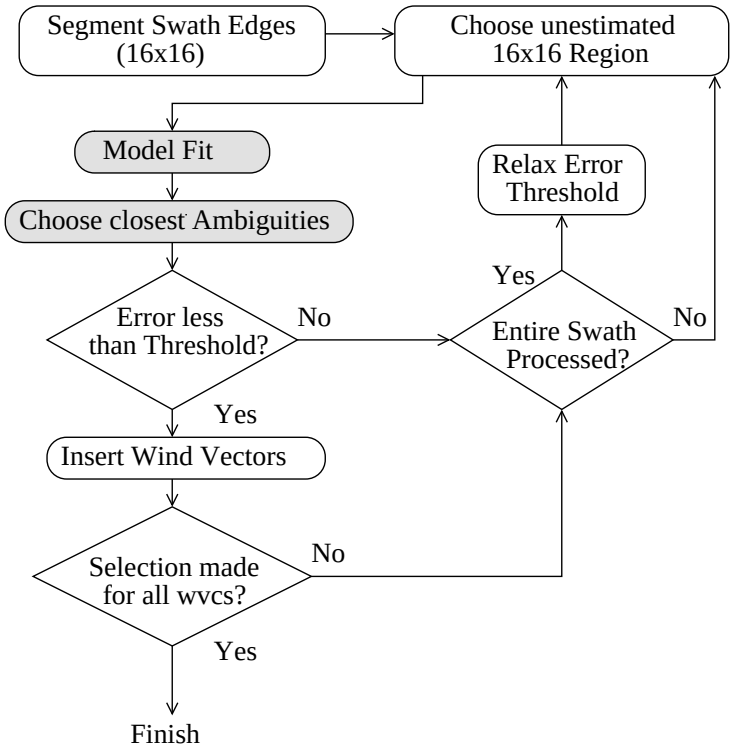


Figure 3.4: A flow diagram summarizing the estimation of swath edges. The shaded bubbles show the crucial steps.

between a cluster of ambiguity selection errors and the correctly selected wind is typically sharp, and results in unnatural, inconsistent wind flow. In order to correct such possible errors, we develop an inconsistency flag followed by a correction algorithm.

Inconsistency flag

Wind vector cells neighboring the sharp transitions from the correctly selected winds to the ambiguity selection errors can be identified by evaluating the consistency of the wind flow [36]. In order to flag inconsistent WVCs, several image processing tools are used.

A true median filter is applied to the u and v components of the ambiguity-selected wind (the value at each WVC is replaced by the median of the values around it within a 3×3 WVC region). This technique reduces the noise in the wind field while preserving edges caused by inconsistent wind flow. The median wind field

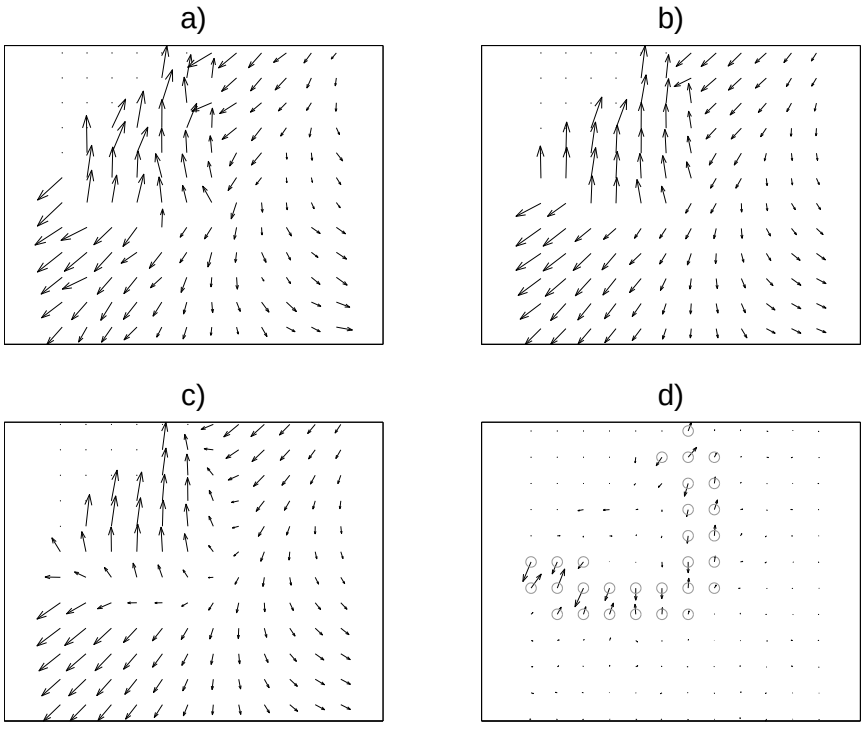


Figure 3.5: An example of the steps involved with flagging ambiguity selection error edges. (a) The ambiguity-selected wind, (b) the median-filtered wind, (c) the average-filtered wind, (d) the difference field. Where the errors are large, the WVCs are flagged as inconsistent (circled).

is then filtered with a 3 × 3 WVC averaging filter (the components of each cell are replaced by the average of those around it). This technique reduces noise, but smoothes edges. The median field and the averaged field are compared. Cells are flagged as *inconsistent* where the normalized vector difference between the averaged and median field is large (greater than 0.25). Fig. 3.5 shows an example of the steps involved with flagging ambiguity selection error edges.

The WVCs flagged as inconsistent indicate the location of ambiguity selection error edges. In order to repair areas of ambiguity selection error, we flag WVCs isolated by the inconsistency flag and other features such as low wind speed areas, the swath edge, or land. The "filling in" of such isolated regions is accomplished through modified dilation and erosion techniques (see Appendix F for details).

Selecting a consistent wind field

All isolated ambiguity selection error regions are re-estimated through interpolation using the KL model. The swath is divided into 72 × 72 WVC sections overlapping by 50% in the along-track direction. This region size is chosen because it includes the entire cross track except the outer 2 cells on either side which do not always contain estimated ambiguities. Each region is decimated into nine 24 × 24 WVC smaller interleaving regions to reduce computational expense. This region size is larger than the 20 × 20 WVC region size used earlier because the outer beam portion of the swath is now included in the estimation process. Each region is model fit using a truncated 24 × 24 KL model, weighting out flagged cells. Flagged cells are re-estimated through interpolation, and the ambiguities closest to the model fit are selected. Each 72 × 72 WVC region is then reconstructed, and the center 36 along track rows are kept. The outer along track rows are discarded because they are more likely to contain modeling errors. The 50% overlap provides an estimate of the entire swath from the individual pieces. The swath is reconstructed from the 72 × 36 pieces and point-wise median filter ambiguity selection is performed on the entire swath. Figure 3.6 shows an example of the repair process.

Iteration of the repair process

For each iteration of the repair process, the output field is compared to the input field. When the number of changing cells falls beneath a threshold (set at 20) or a maximum number of iterations is reached, the algorithm stops. Most swaths converge in about 10 iterations. The repair process is summarized in Fig. 3.7.

3.3 Comparison of BYU and JPL Ambiguity Selection

Without an extensive truth data set, assessing the quality of the BYU ambiguity selection is difficult [21]. As a result, we use the JPL level 2B (L2B) nudged winds as a reference data set. We perform three comparisons of BYU to standard JPL processing on a set of 102 revs. Ambiguity selection is performed on revs 1000-1050 and revs 6000-6050 of QuikSCAT data. In section 3.3.1, we perform a

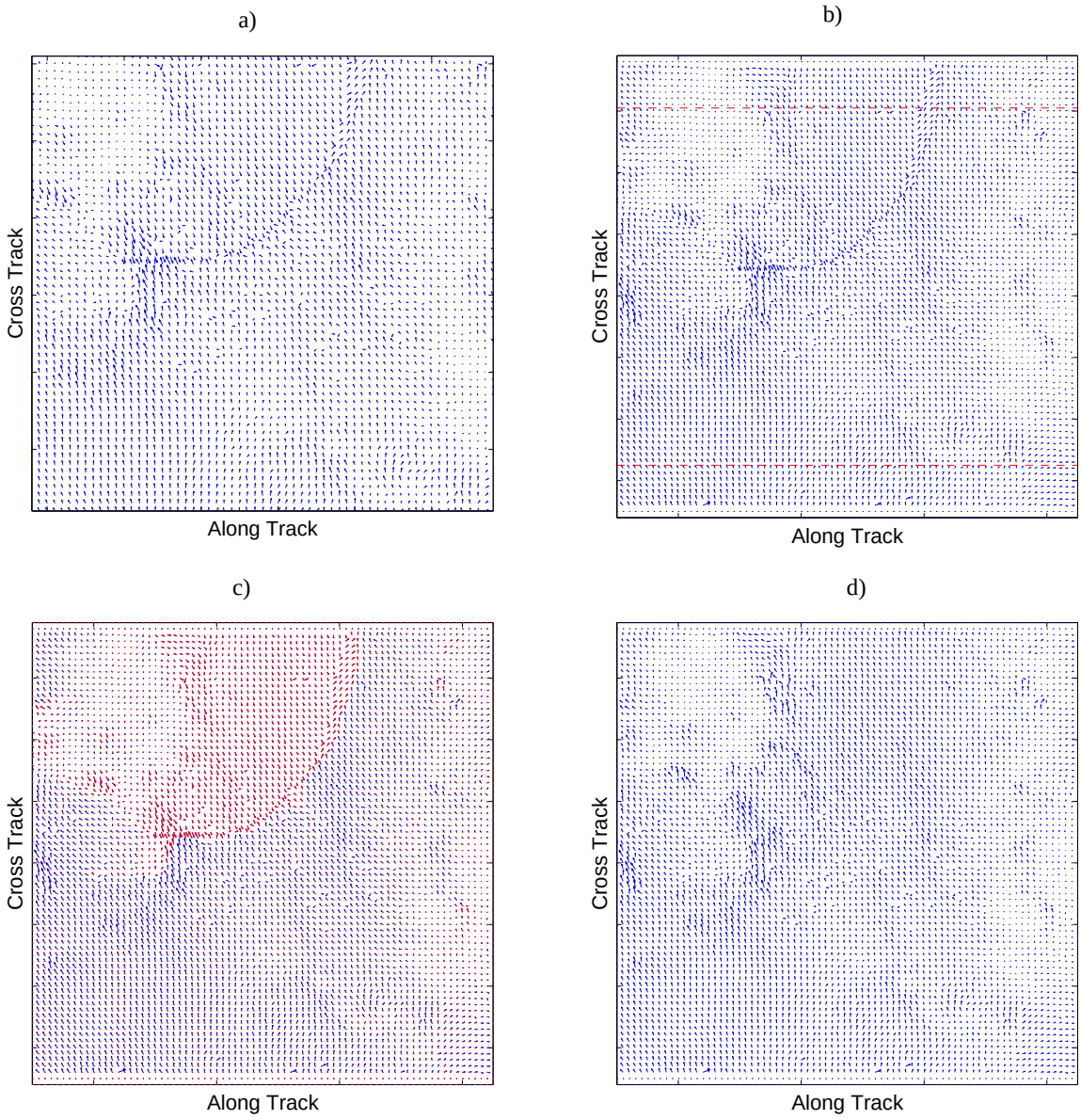


Figure 3.6: Example of the repair process: a) A sample wind field containing a large ambiguity selection error in the initial estimate. b) The initial estimate including swath edge estimation. c) Ambiguity selection error region flagged by the inconsistency flag (red vectors). d) Final field after repairs.

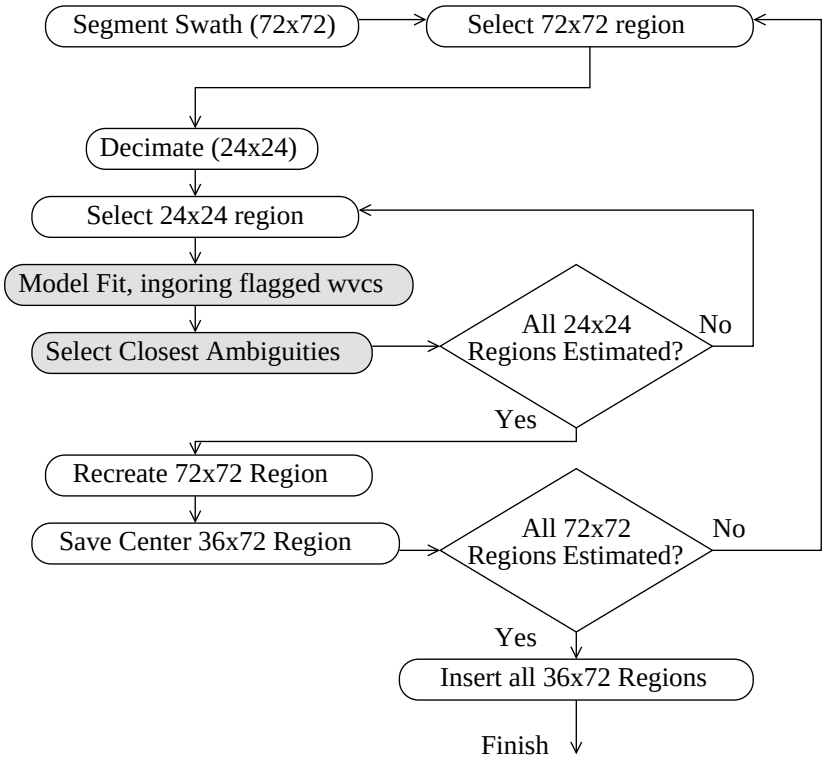


Figure 3.7: A flow diagram summarizing the repair process. The shaded bubbles show the most important steps.

direct comparison of the ambiguities selected by both techniques. In section 3.3.2, we perform a quality assurance analysis with a model-based technique [36]. In section 3.3.3, we present statistics based on manual inspection of the data.

3.3.1 Direct comparison of selected ambiguities

To quantify the similarity of the BYU ambiguity-selected winds to the JPL reference data set, we compare the percentage of each ambiguity selected by both techniques over the test data set in Table 3.1. The BYU algorithm selects slightly fewer first ambiguities than the JPL product and slightly more of the other ambiguities. Like the JPL product, the BYU algorithm selects a majority of first and second ambiguities (about 85%) which is consistent with our assumption that the

Table 3.1: Percentages of ambiguities chosen by the JPL L2B product and the BYU algorithm.

Category	JPL L2B product	BYU method
1st Ambiguity	65.47%	64.80%
2nd Ambiguity	20.58%	20.76%
3rd Ambiguity	8.63%	8.89%
4th Ambiguity	5.32%	5.55%

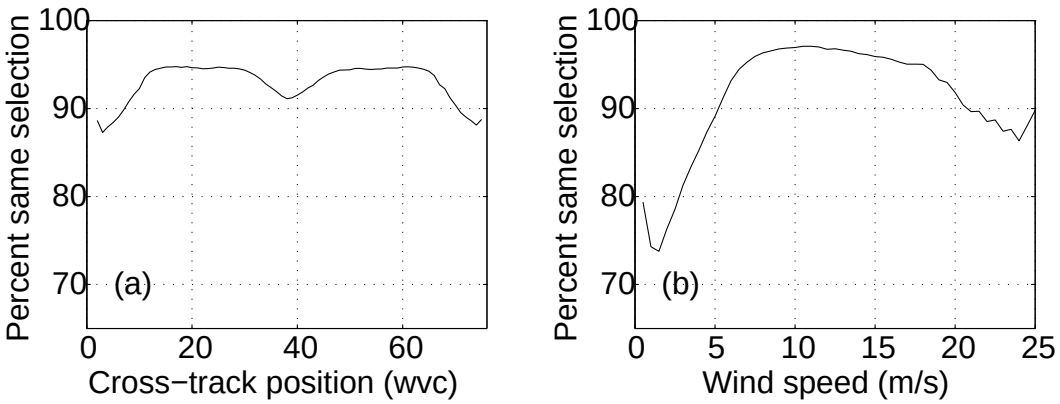


Figure 3.8: Percentage of individual WVCs selected the same for both JPL and BYU methods per (a) cross-track position and (b) average wind speed of the ambiguities.

overall flow of the wind is dictated mainly by the first two ambiguities. The other ambiguities are chosen in approximately the same proportion as the JPL product.

Further, we compare the percentage of ambiguities similarly selected by both the BYU and JPL algorithms and the percent of each ambiguity changed in Table 3.2. From Table 3.2, the ambiguity selection is the same for both BYU and JPL in 93% of WVCs. Thus, the BYU algorithm gives generally the same wind vector solution as the JPL product which simultaneously helps validate both BYU and JPL techniques. This result is significant because the BYU method was independently developed and not tuned against the JPL reference data. It suggests that the SeaWinds noisy ambiguity sets contain a sufficient percentage of correct first ambiguities to allow self-contained ambiguity selection that is very close to a nudged solution.

Table 3.2: Percentage of ambiguities chosen the same and changed from 1st to 2nd or from 1st or 2nd to 3rd or 4th for the sample set of BYU and JPL ambiguity-selected winds

Category	Percentage
Same ambiguity selected for JPL and BYU	93.03%
JPL selected 1st ambiguity and BYU selected 2nd ambiguity	1.41%
JPL selected 2nd ambiguity and BYU selected 1st ambiguity	1.13%
JPL selected 1st or 2nd ambiguity and BYU selected 3rd or 4th ambiguity	2.23%
JPL selected 3rd or 4th ambiguity and BYU selected 1st or 2nd ambiguity	1.74%

Next, we compare the percentage of similar ambiguities selected as a function of cross-track position and wind speed. To assign a unique wind speed to a WVC, we average wind speeds of all ambiguities. Since all ambiguities at a WVC generally have similar wind speeds the averaging does not significantly affect the results. Fig. 3.8 displays the percentage of ambiguities chosen the same per cross-track position and wind speed. Fig. 3.8 indicates that the ambiguity selection differs the most along swath edges and at nadir. This is intuitive, because more ambiguities are generally produced in these regions, allowing for a higher probability of a different selection. Also, low and high wind speed data (which tend to be the noisiest) have a higher percentage of changed WVCs.

3.3.2 Quality assurance assessment

In this section, we present a QA assessment of the ambiguity-selected (see Chapter 2 for a description of the QA method). The QA method is applied to both JPL and BYU ambiguity-selected test sets. Of JPL ambiguity-selected winds, 4.5% of 8×8 WVC regions inspected by the QA method are determined to contain possible ambiguity selection errors. Of the BYU wind vector selections, about 3.4% of all regions are determined to contain possible ambiguity selection errors. These numbers

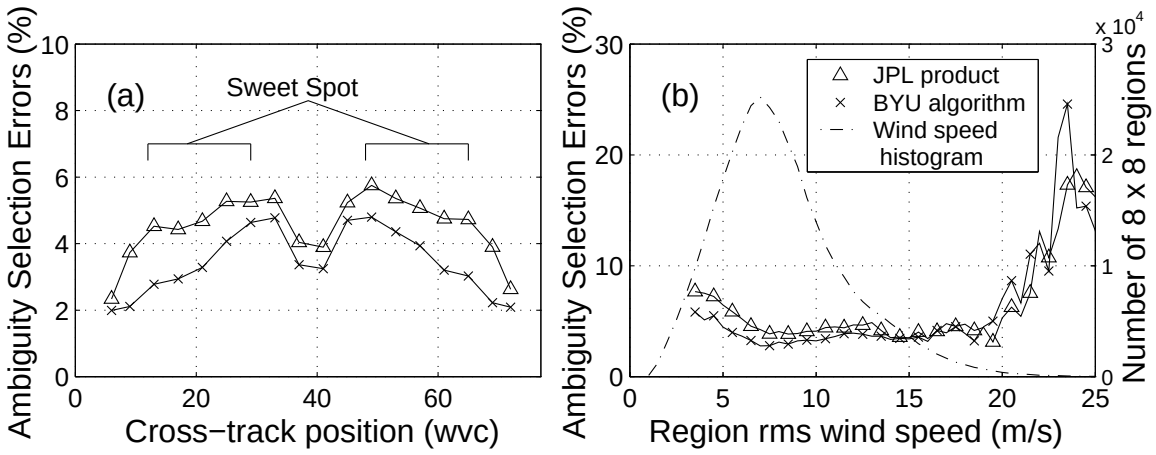


Figure 3.9: Percentage of 8 × 8 regions flagged as containing possible ambiguity selections error by the QA analysis method (a) per cross-track position and (b) RMS wind speed for both JPL and BYU methods (left axes). A histogram of wind speeds is show in (b), indicating the number of regions inspected per RMS wind speed bin (right axis).

suggest that the overall consistency of the BYU method is somewhat better than JPL.

Next, we examine the percentage of QA-determined ambiguity selection errors as a function of cross-track position and region RMS wind speed (see Fig. 3.9). The region RMS wind speed is defined as

$$\left(\frac{1}{n}\sum_{i,j} U_{i,j}^2\right)^{\frac{1}{2}} \tag{3.4}$$

where $U_{i,j}$ is the wind speed at cell {i,j} of the region, and n is the number of valid wind data WVCs in the region (WVCs over ocean). The BYU method performs particularly better in the "sweet spot" (off-nadir region characterized by high instrument skill) and at low to moderate wind speeds (4-10 m/s). Many of the "sweet spot" cases occur in rain-corrupted areas where thresholded nudging fails to select ambiguities consistent with the overall flow of the non-rain-corrupted wind [36]. At moderately high wind speeds (10-18 m/s) both methods perform about the same. At extreme

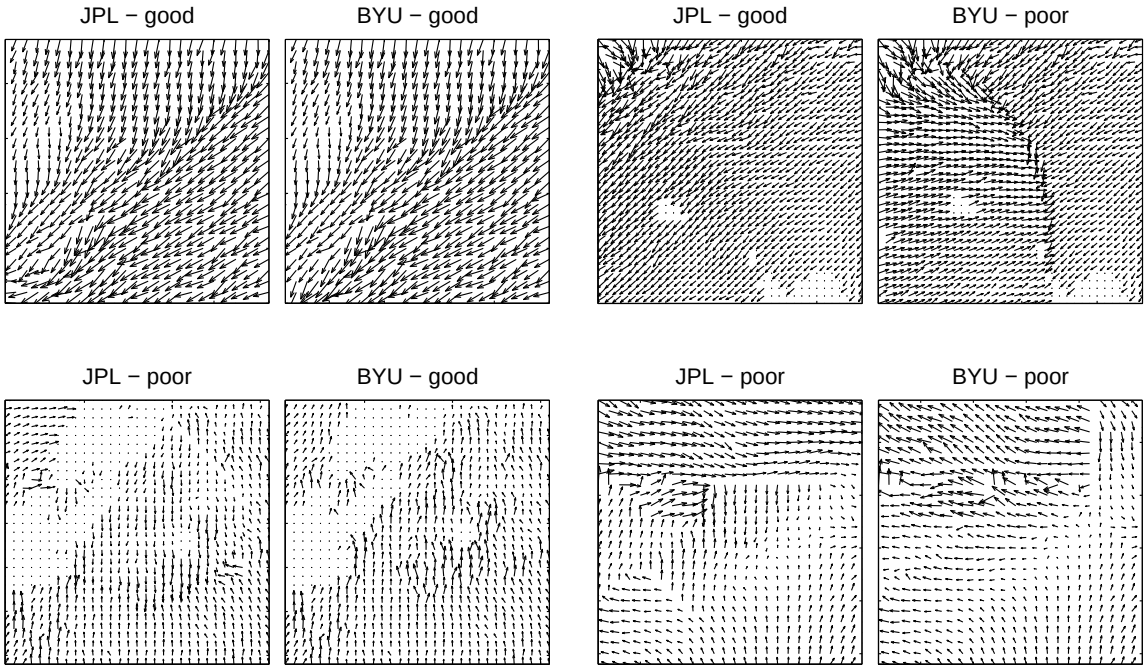


Figure 3.10: Examples of the various classifications of regions *without* cyclonic features when comparing the JPL and BYU ambiguity selection routines.

winds (18+ m/s) the JPL method performs better. Extreme winds, however, only occur a small percentage of the time.

3.3.3 Analysis of storm and non-storm regions

This section provides a subjective comparison of BYU ambiguity selection with the JPL selection in regions with and without cyclonic features. For each revolution of test data, the swath is segmented into 38 × 38 WVC sections. Because cyclones represent areas of high spatial variability, regions containing cyclonic features are analyzed separately. A 38 × 38 region is rated "good" if there is a visually consistent wind flow. If the region contains an area of visually inconsistent flow, it is rated "poor". These ratings are applied to each 38 × 38 WVC region of the 102 swath test data set. Examples of each type of region are found in Figs. 3.10 and 3.11. Table 3.3 summarizes statistics on the regions for the study.

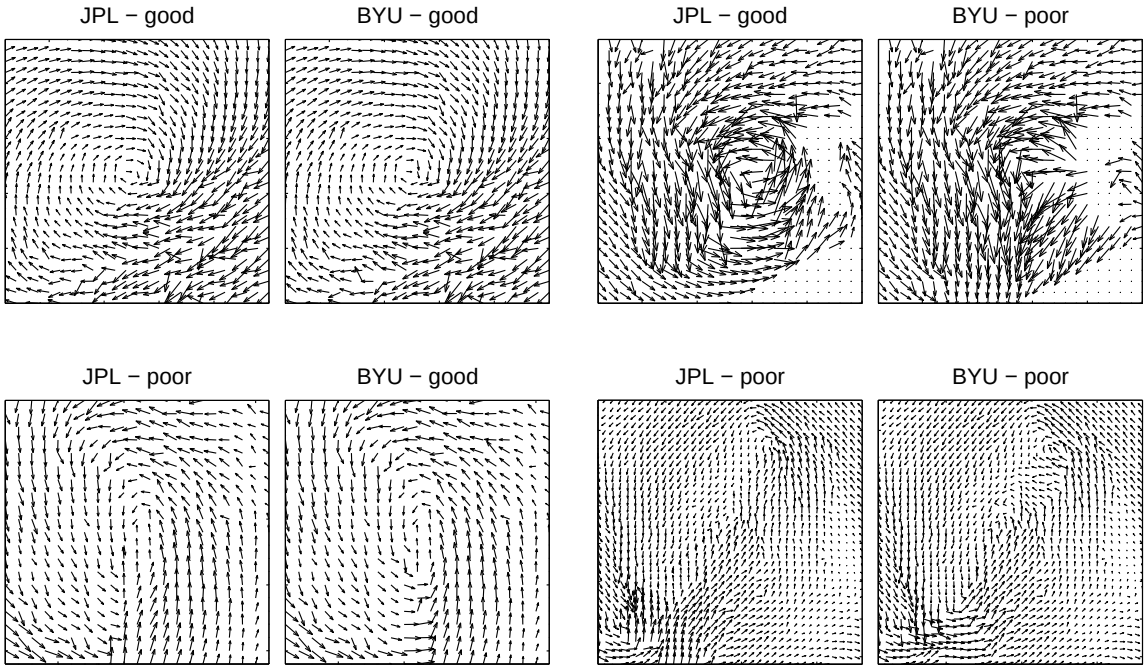


Figure 3.11: Examples of the various classifications of regions *with* cyclonic features when comparing the JPL and BYU ambiguity selection routines.

For non-cyclonic areas, both JPL and BYU methods produce very visually consistent flow in about 95% of the test data set. The BYU algorithm on average produces somewhat more non-cyclonic "good" regions. From visual observation, JPL poor areas often occur where the first ambiguity field is corrupted by rain from which the JPL algorithm cannot recover. This is consistent with results found in [36]. Ambiguity selection errors in the BYU data are often associated with large regions of incorrect first ambiguities or fine-scale wind features. An interesting observation is that in some cases, large areas of incorrect first ambiguities occur near the southern end of the swath at moderately high wind speeds, creating ambiguity selection errors in the BYU data, but not in the nudged JPL winds.

For cyclonic cases, the BYU method performs the same as the JPL winds in about 71% of the cases. Of the remaining cases, BYU winds are subjectively better in about 9% of the cases, and worse in about 20% of the cases. Thus, the BYU

Table 3.3: Total numbers and percentages of regions subjectively rated "good" or "poor" in both BYU and JPL data sets for 38 × 38 WVC areas without and with cyclonic features. Percentages are calculated for regions with and without cyclones separately.

	Without Cyclone		With Cyclone	
	BYU		BYU	
JPL	Good	Poor	Good	Poor
Good	6154 (94.8%)	63 (1.0%)	91 (56.2%)	32 (19.7%)
Poor	147 (2.3%)	125 (1.9%)	15 (9.3%)	24 (14.8%)

method is somewhat less able to produce realistic cyclonic features. Although the JPL nudging technique is not perfect in storm regions, it provides an initial guess of the placement of cyclonic rotational features. These features may not be clearly evident in the noisy first ambiguity field. Additionally, in the BYU method, cyclonic storms are sometimes overly smoothed by the low-pass effect of the KL model, resulting in poor ambiguity selection. Of the poorly retrieved storms for the BYU algorithm, a higher percentage occur on the swath edges than in the inner-swath. Eighteen storms are centered on the swath edge in the test data set. Of these storms, eleven (61%) are subjectively identified as "poor," a much higher percentage than the overall 34.5% of BYU storm cases that are identified as "poor". The higher rate of poorly retrieved storms on the swath edge is probably due to the fact that solutions for the swath edge are selected via extrapolation of the inner swath. Thus, small-scale features located in the other-swath region are more likely to be missed by the algorithm.

The analysis in this section suggests that the BYU method is somewhat less effective than JPL in creating visually consistent cyclonic storm regions, but is generally better in areas lacking cyclonic features. The BYU method could be improved by detecting and separately processing storm regions with a specialized algorithm or by mitigating the effects of rain in storm regions. Further work is planned to accomplish this.

3.4 Summary and discussion

BYU point-wise ambiguity selection uses a data-driven model rather than a nudging field to produce self-consistent wind fields for scatterometer wind ambiguity selection. A correction routine locates and corrects further large inconsistencies in the wind. The algorithm is applied to a set of test revs. The algorithm generates the same basic wind flow as the current JPL product without using the nudging field.

Largest differences between the JPL and BYU selections occur in low and high wind speed regions, at nadir and along swath edges. Our QA analysis demonstrates that low to moderate wind speed data (between 4 and 10 m/s) is generally more self-consistent for the BYU method, while higher wind speed data is somewhat less self-consistent. Also, there is increased performance in the sweet spot for the BYU method, especially in rain-corrupted regions.

From a manual inspection of the JPL and BYU ambiguity selected winds, we conclude that BYU produces fewer possible ambiguity selection errors in regions without cyclonic storms. Ambiguity selection errors in the BYU algorithm are generally associated with storms, extreme winds, and large areas of incorrect first ambiguities. These cases, however are rare. Overall, the BYU method selects 93% of the same ambiguities as the JPL. This result indicates that ambiguity selection can be generally accomplished without the use of a nudging field. The BYU ambiguity selection method is well suited for operational ambiguity selection since the NWP nudging field is not required and the algorithm is computationally efficient.

Chapter 4

Evidence of a Threshold Wind Speed in Tower-mounted Scatterometer Data

The QA assessment classifies a higher percentage of regions as "poor" when the RMS wind speed is less than about 5 m/s (see Figure 2.10b). At very low wind speed, the SNR of SeaWinds may be too low to validate the wind vector estimates. In addition, *Donelan and Pierson* [18] postulated the existence of a threshold wind speed below which there is insufficient friction between the wind and water to generate capillary waves. The threshold wind speed is mainly dependent upon water temperature and Bragg wavelength. Below the threshold wind speed, the Bragg-induced backscatter theoretically goes to zero. Thus, according to the theory, wind retrieval below the threshold is not possible. Because it is difficult to directly observe this phenomenon with spaceborne measurements, this chapter analyzes the threshold wind speed using data from the tower-mounted YSCAT scatterometer. This chapter demonstrates that the possible effects of the low wind speed threshold are detectable in scatterometer data in an uncontrolled marine environment. The implications for SeaWinds are explained in the discussion section at the end of the chapter.

4.1 Overview

The geophysical model function (GMF) described in Chapter 1 is generally accurate for wind speeds above about 3 to 4 m/s. Below the threshold wind speed, the capillary waves theoretically disappear and the Bragg-induced backscatter goes to zero. Although the minimum wind speed for wave generation has been carefully studied in wave tank data [37], and have been documented in airship data [38], the implications of the low wind speed threshold to space-based wind scatterometry are

still uncertain. An approach of correcting the GMF for threshold wind speed effects is given in [33], however no compensation for the threshold wind speed is currently operationally used.

In order to analyze the low wind speed threshold, we examine data from the tower-mounted YSCAT scatterometer. Data from the YSCAT scatterometer provides *in situ* wind speed measurements which are generally unavailable in connection with SeaWinds data. YSCAT is an ultra-wideband (2-14 GHz) tower-mounted scatterometer, which was deployed on Lake Ontario for a period of 6 months in 1994 [39]. Data was collected for a variety of frequencies and incidence angles which span a range of Bragg wavelengths. This chapter examines the YSCAT data for evidence of a threshold wind speed as a function of Bragg wavelength. A simple experiment is conducted in which approximately 1/2 of the Bragg wavelengths observed by YSCAT evidence a threshold wind speed. The YSCAT-estimated threshold corresponds well to the theoretical threshold for both h-pol and v-pol data. Although the effect of the threshold wind speed is detectable, the YSCAT data does not suggest that the backscatter "goes to zero" in an uncontrolled environment as theory indicates; rather, backscatter measurements for winds below the threshold have a somewhat decreased mean.

4.2 The YSCAT scatterometer

YSCAT collected data at 2, 3, 5, 10, and 14 GHz, and at 0° (nadir), 20°, 25°, 30°, 40°, 50°, and 60° incidence angles at both horizontal and vertical polarizations. In addition, a variety of environmental parameters were observed including wind speed and direction, rainfall, and water temperature. The antenna was specially designed to provide a near constant beam width over most of the operating bandwidth and produce a footprint of approximately one meter for mid-range incident angles [39].

Post processing of YSCAT data provides minute averages of σ^o and *in situ* measurements of wind velocity. Rain-contaminated data, data corrupted by instrument malfunctions or other sources of error, and data with fluctuating wind

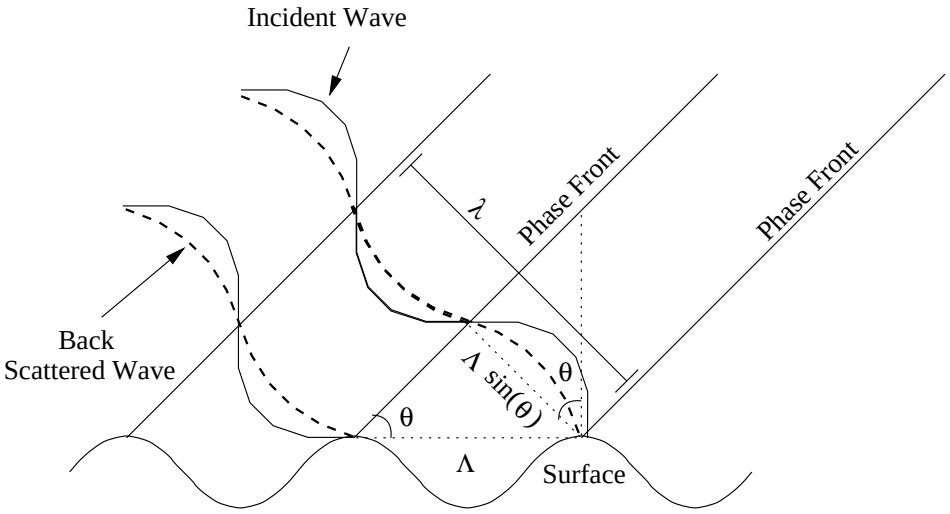


Figure 4.1: Wind wave/electromagnetic wave interaction for Bragg scattering.

speed measurements are discarded. The minute-averaged data is binned according to frequency, polarization, incidence angle, and direction. Data is additionally binned into upwind and downwind directions. The data is normalized by dividing out the mean of each record and then multiplying by the aggregate mean of the corresponding bin [39].

The large number of frequencies and incidence angles provides a wide range of Bragg wavelengths observed by the YSCAT instrument. The Bragg wavelength is the component of the water wave spectrum that is resonant with the incident frequency. Bragg scattering occurs when the phase fronts of electromagnetic backscatter from successive surface waves align, creating an enhanced backscatter response. The geometry associated with Bragg scattering is shown in Figure 4.1. This first order Bragg theory relates the Bragg wavelength Λ to the electromagnetic wavelength λ by

$$\lambda = 2\Lambda \sin(\theta) \tag{4.1}$$

or

$$\Lambda = \frac{\lambda}{2\sin(\theta)}. \tag{4.2}$$

Table 4.1: Values of parameters given in Eq. (4.3) for a water temperature of 10° C and zero salinity.

parameter	value
ρ_a	1.19 g/L
ρ_w	0.997 g/cm^3
τ	67.94 mN/m
ν	0.890 mPas

where θ is the incidence angle. The wide range of Bragg wavelengths observed by YSCAT affords us access to understanding the wind/sea interaction for a large number of wave sizes. The Bragg wavelengths measured by YSCAT range from 1.2 cm to 43 cm.

4.3 Theoretical threshold wind speed

The theoretical threshold wind speed is derived by *Donelan and Pierson* [18]. At low wind speeds, friction from the wind is too light to overcome the viscous effects of water and no capillary/gravity waves are generated. If we accept Bragg scattering as the predominate scattering mechanism, under these conditions, there should be no backscatter at all [33]. The threshold wind speed at $\pi/k = \lambda/2$ height is derived to be [18]

$$\bar{U}(\pi/k) = C(k) + 2\left(\nu k \frac{C(k)}{0.194\rho_a/\rho_w}\right)^{1/2}. \tag{4.3}$$

In Eq. 4.3, $C(k)$, the phase speed of the capillary/gravity Bragg waves, is given by

$$C(k) = \sqrt{g/k + \tau k/\rho_w}. \tag{4.4}$$

Also, ρ_a and ρ_w are the densities of air and water respectively, τ is the surface tension, g is the acceleration of gravity, and ν is the viscosity of water. The viscosity has a dependence on water temperature. Thus, the threshold wind speed is a function of Bragg wavelength and temperature. It is also a weak function of salinity, which is ignored here. The mean water temperature for the YSCAT experiment is assumed

Table 4.2: Theoretical threshold wind speeds (m/s) as a function of frequency and incidence angle for YSCAT

Frequency	Incidence Angle						
(GHz)	10°	20°	25°	30°	40°	50°	60°
2.00	1.98	1.94	1.95	1.98	2.03	2.09	2.14
3.05	1.93	1.99	2.03	2.09	2.19	2.28	2.36
5.30	1.96	2.15	2.26	2.36	2.55	2.72	2.87
10.02	2.14	2.56	2.77	2.98	3.40	3.79	4.12
14.00	2.30	2.92	3.24	3.57	4.23	4.85	5.40

to be 10° C. Using 10° C temperature and zero salinity, the parameters of Eq. (4.3) are calculated and given in Table 4.1.

The wind speed at π/k height is related to the wind speed at 10 m height by the equation

$$\bar{U}(\pi/k) = \bar{U}(10)\left(1 - \frac{[A + B\bar{U}(10)]^{1/2}}{\kappa}[\ln k - \ln(\pi/10)]\right) \tag{4.5}$$

where A = $0.96e^{-3}$, B = $0.041e^{-3}$, and κ is the von Karman constant assumed to be 0.4. The threshold wind speed at 10m ($\bar{U}(10)$) must be numerically solved. The theoretical threshold wind speeds for YSCAT at 10° C water temperature are given in Table 4.2.

4.3.1 Detection and estimation of YSCAT threshold wind speeds

A simple experiment is performed to detect and estimate the threshold wind speed versus Bragg wavelength using YSCAT data. Assuming that downwind and upwind observations exhibit the same threshold characteristics, downwind and upwind measurements are combined in the analysis. The backscatter cross-section above the threshold wind speed (which we initially approximate as 4 m/s) has a power-law relationship with wind speed [39],

$$\sigma^{\circ}(U > \bar{U}(10)) = aU^{b}. \tag{4.6}$$

Using YSCAT minute average wind speeds and σ° values, the parameters a and b for each Bragg wavelength are computed for wind speeds above 4 m/s by converting

both sides of Eq. 4.6 to decibels,

$$\sigma^{\circ}_{\mathrm{dB}} = a_{\mathrm{dB}} + bU_{\mathrm{dB}} \tag{4.7}$$

and applying linear least-squares estimation to the YSCAT data [39]. After estimating a_{dB} and b, all σ° values further than 2 standard deviations from the model fit are discarded. The estimation is performed again, giving the final values for a_{dB} and b.

Next, non-parametric estimation is performed on each Bragg wavelength data set using a Gaussian kernel. This method gives a smoothed line through the data relating U_{dB} and $\sigma^{\circ}_{\mathrm{dB}}$. The non-parametric curve affords a higher resolution estimate of the data than the linear least-squares fit.

In order to select a threshold wind speed for each Bragg wavelength data set, we compare the non-parametric curve to the linear least-squares fit. Since the data is obtained from an uncontrolled marine environment and is minute-averaged for each observation, we do not expect the σ° values to go to zero below the threshold wind speed, but we do expect σ° to decrease in mean value and have a higher variability. We estimate the threshold wind speed to be the point where the non-parametric curve falls 1.5 dB (25%) beneath the linear least-squares fit. The threshold of 1.5 dB is subjectively chosen, but the results are not particularly sensitive to this value. A higher threshold has the effect of slightly lowering the estimated low wind speed thresholds. If the curve *never* falls 1.5 dB beneath the linear fit, we label the data a "non-detection" (N/D). An alternative method for estimating the low wind speed threshold is given in Appendix G. The alternative method, although more theoretical, is less robust. Thus, the method described above is used in the further analysis.

Using this method, 20 of the 35 data Bragg wavelength bins observed by YSCAT have a detectable threshold wind speed for h-pol observations. Of the fifteen non-detections, 4 sets exhibit a noticeable noise floor that prevents observation of the threshold wind speed. These sets occur at the highest frequencies and incidence angles. In addition, one set has no usable data (5 GHz, 25°, not shown in Figures 4.3 and 4.4).

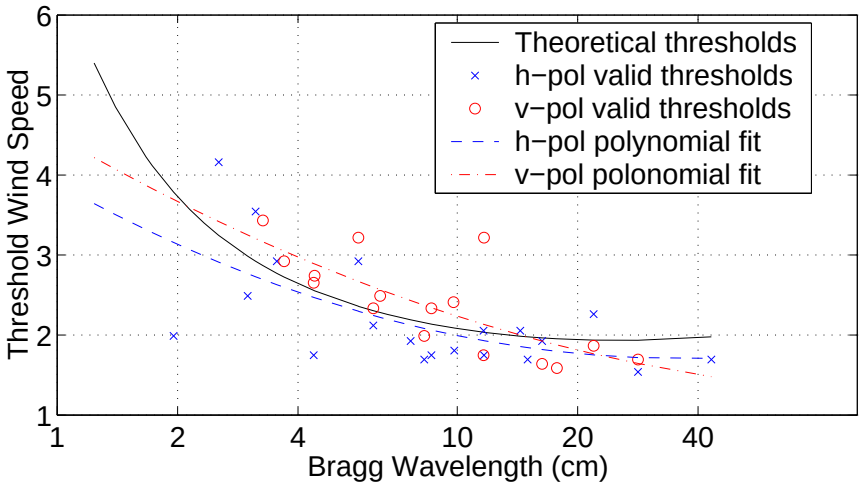


Figure 4.2: The theoretical threshold wind speeds for the range of Bragg wavelengths observed by YSCAT. Also, the estimated threshold wind speeds for v-pol and h-pol observations obtained from the YSCAT data where a threshold wind speed is detected. The dashed and dot-dashed lines show a second order polynomial fit to the h-pol and v-pol threshold wind speeds respectively.

Of the v-pol observations, 16 sets have a detectable threshold wind speed. Of the nineteen non-detections, seven sets exhibit a noticeable noise floor and one set contains no usable data. The v-pol data is much noisier than the h-pol data, thus giving a higher variance in the estimates of the wind speed threshold and possibly prohibiting detection in some cases.

The theoretical threshold wind speeds are plotted along with the results obtained from this experiment in Figure 4.2. Also in Figure 4.2, a quadratic fit is made to the h-pol and v-pol thresholds. Both quadratic fits give a fairly close approximation of the thresholds in the range of valid detections. The v-pol data is slightly more variable than the h-pol data, giving a fit that deviates more from the theoretical thresholds. These results help verify the existence of the threshold wind speed, and give support to the accuracy of the model.

The YSCAT data along with the theoretical and estimated threshold wind speeds are shown in Figure 4.3 and 4.4. These figures allow for visualization of the threshold wind speed and resulting decrease in σ° below the threshold. As illustrated

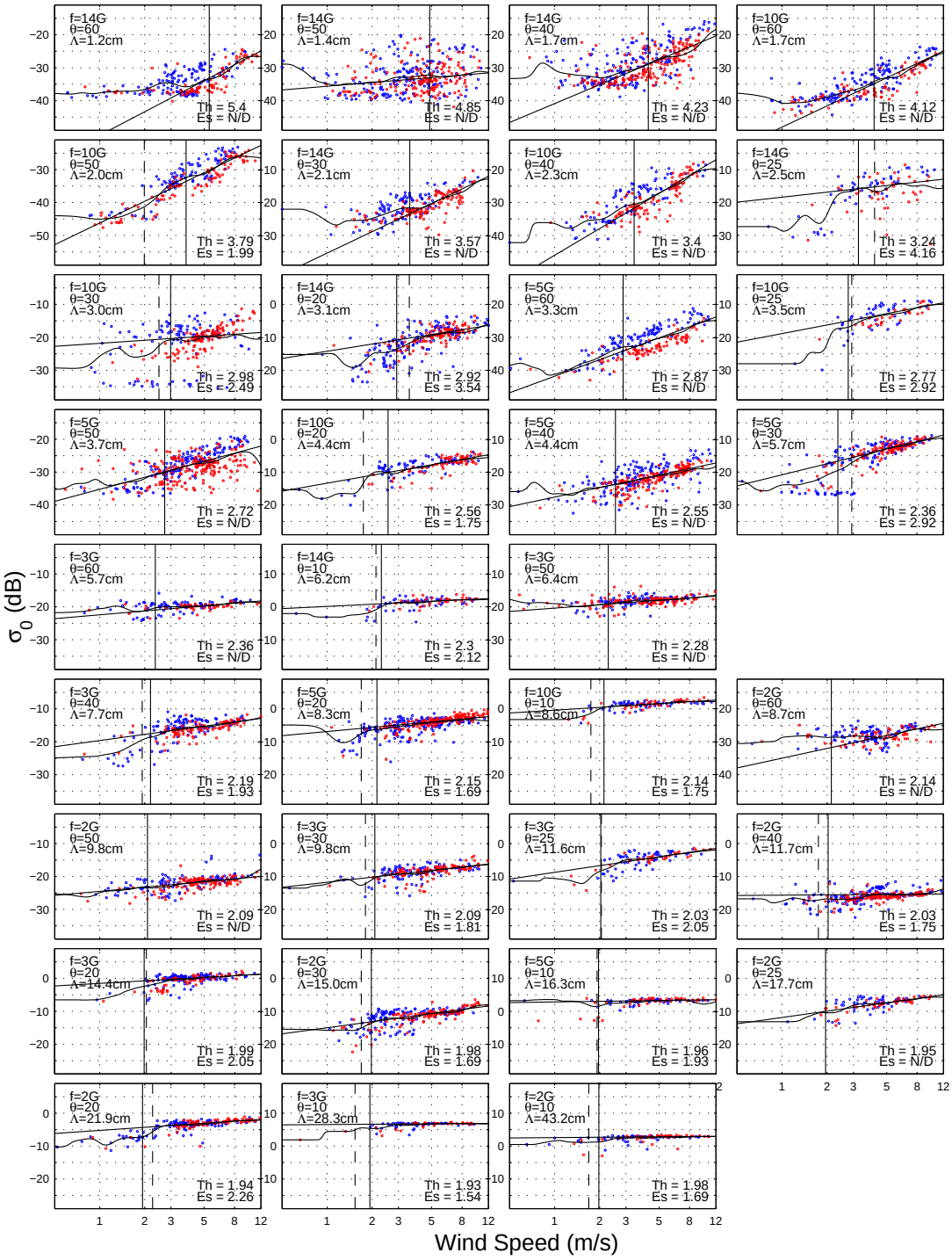


Figure 4.3: H-pol YSCAT data plotted as a function of wind speed for each Bragg wavelength. The straight line through the data is the linear least-squares fit. The curved line is the non-parametric fit. The solid vertical line is the theoretical threshold wind speed (Th). The dashed line is the threshold wind speed estimated from the YSCAT data (Es).

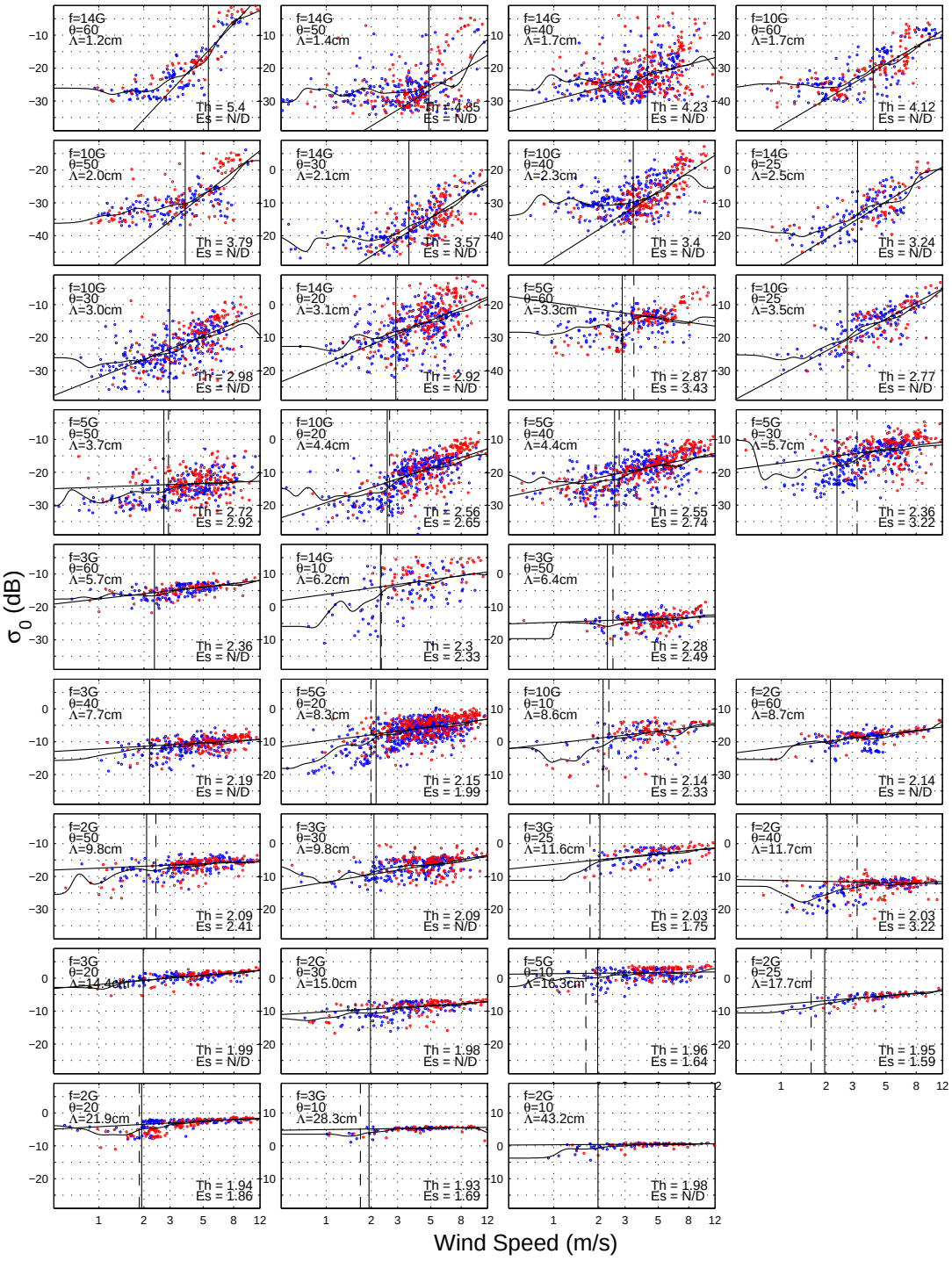


Figure 4.4: V-pol YSCAT data plotted as a function of wind speed for each Bragg Wavelength and polarization (See Figure 4.3).

in Figure 4.3 and 4.4, the roll-off in σ° below the threshold is not as pronounced as in controlled wave tank data, and does not drop to zero as theory suggest. Nevertheless, a mean decrease in σ° is detectable in much of the data, suggesting that for small-footprint instruments such as YSCAT, effects of the threshold wind speed are detectable in an uncontrolled environment.

The fact that the backscatter does not drop to zero beneath the threshold wind speed may be due to several effects. We discuss two reasons: First, the backscatter measurements are affected by the sea state induced by previous winds. When the wind decreases below the threshold, there is a lagging or hysteresis effect that occurs, i.e. it takes time for waves previously generated by the wind to subside. The radar then receives additional backscatter due to the residual waves. This hysteresis effect is a function of Bragg wavelength. The longer the waves, the longer they take to subside, and the less noticeable the low wind speed threshold.

Second, the backscatter data is minute-averaged. Even though the average wind may be below the threshold, the wind may somewhat vary during the minute interval, generating at times capillary waves which give enhanced backscatter to the radar. Thus, the averaging suppresses the overall effect of the low wind speed threshold.

4.3.2 Dual threshold model

In addition to the single-threshold model, we introduce a dual-threshold model. An dual-threshold model allows us to observe the hysteresis effect of the wind-water interaction on different Bragg wavelengths. This is subjectively determined by manually dividing each Bragg wavelength bin of data into three regions:

- moderate wind speed region.
- hysteresis region
- roll-off region

The moderate wind speed region is where there is no evidence of σ^o roll-off. It is separated from the hysteresis region by the "high-end" cutoff. The "high-end" cutoff

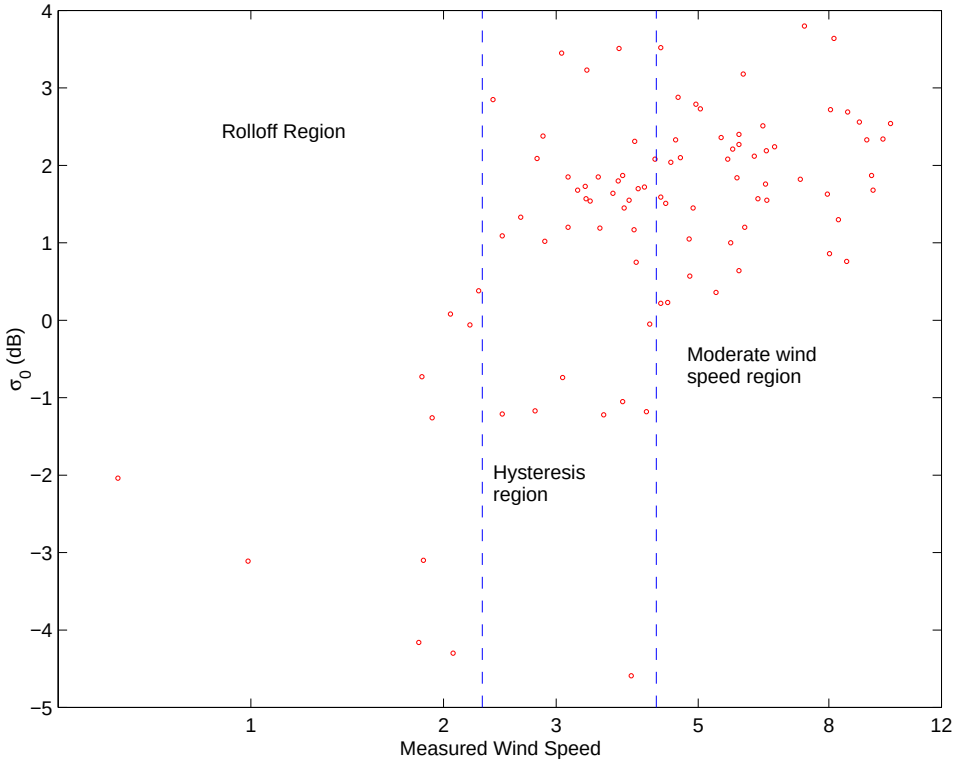


Figure 4.5: An example of the roll-off, hysteresis and moderate wind speed regions in YSCAT data.

(χ_h) is the high end of the hysteresis region where the σ^o roll-off starts. This cutoff indicates the threshold at which wind starting up begins to create capillary waves. The "low-end" cutoff (χ_l) is the low end of the hysteresis region. This cutoff is where slowing winds fail to give enough energy to the water to create the waves. Below that is the roll-off region where the roll-off is evident in nearly all σ^o measurements. Figure 4.5 shows these regions.

Figure 4.6 shows the subjective estimates for the low-end and high-end cutoffs of the hysteresis region as a function of Bragg wavelength. The low-end cutoff decreases and the high-end cutoff increases with increasing Bragg wavelength. This suggests that the hysteresis region widens with Bragg wavelength. This phenomenon has physical meaning. Short waves are very sensitive to the wind. Thus, when the wind is starting up, the shorter waves are created quickly. However, longer waves take

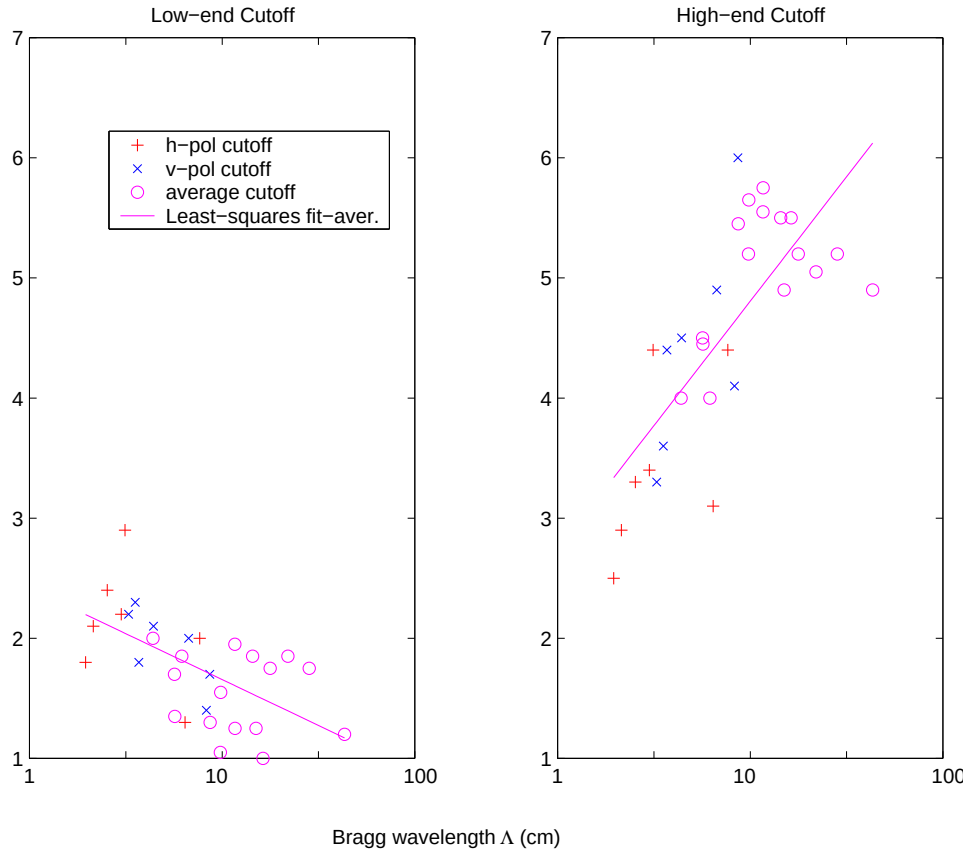


Figure 4.6: Subjectively chosen low-end and high-end cutoffs for the hysteresis region and a least-squares fit.

more energy to get started. Thus, the high-end cutoff is higher for longer wavelengths. On the other hand, once the larger waves are in motion, it takes longer for them do die out because they are less sensitive to the wind. Therefore, the low-end cutoff decreases for larger waves and the hysteresis widens.

4.4 Summary and discussion

YSCAT data demonstrates strong evidence of a threshold wind speed in an uncontrolled marine environment. The experiment in this paper demonstrates that in approximately 1/2 of the YSCAT Bragg wavelength observations, a detectable roll off in σ° is evident beneath a threshold wind speed. Although the backscatter

"rolls off," it does not decrease to zero as theory suggests. Nevertheless, the YSCAT-estimated threshold wind speeds correspond well to the theoretical threshold wind speeds derived by [18], and help support the theory of a threshold wind speed.

We do not use the YSCAT analysis to directly correct SeaWinds scatterometer data. Rather, we give a brief discussion of implications of the YSCAT analysis to SeaWinds. Unfortunately, a noise floor prohibits YSCAT observation of the threshold wind speed at SeaWinds incidence angles and frequency ($\sim$ 14 GHz, 46° and 54° incidence.) Thus, directly relating the YSCAT threshold effect to SeaWinds is difficult. We can only make general observations.

The main observation from the YSCAT experiment is that backscatter from winds beneath the low wind speed threshold are often lower than expected, with increased variability. The variability is due in part to the hysteresis effect, and is likely a main factor in the decrease in accuracy of SeaWinds vector wind data at low wind speeds. The wind speed does not go to zero as theory predicts in an uncontrolled environment, probably due to scattering effects from phenomena such as residual waves from subsiding winds. In order to relate the effect of the low wind speed threshold to SeaWinds requires an understanding of the variability of the wind within the large footprint, i.e. the beamfilling effect. Over the footprint, there may be areas with wind speeds above the threshold, and other areas with wind speeds below the threshold. Depending on the spatial variability of the wind, this beamfilling effect may significantly alter the backscatter from the mean. However, directly correcting for the variability of the wind within the measurement via the scatterometer is not possible, since knowledge of sub-footprint phenomena is not generally accessible. Such effects can only be treated statistically.

A statistical method for improving wind estimation at low wind speeds is given by *Shankaranarayanan and Donelan* [33]. The method adjusts the GMF for a mean bias created by the threshold effect. The corrected GMF from [33] assumes that for low wind speeds, the backscatter falls to zero in a statistically-determined area of the 25 km footprint where the wind speed is locally beneath the threshold. Thus, the GMF is lowered at low wind speeds to account for the threshold effect. The YSCAT

analysis, however, suggests that the backscatter does not drop to zero in uncontrolled environments. Thus, the new GMF presented in [33] may be over-compensated.

Chapter 5

Evaluating the Effect of Rain on SeaWinds Scatterometer Measurements

The quality assessment presented in Chapter 2 and other validation analyses have found SeaWinds on QuikSCAT data to be highly accurate in non-raining, moderate wind speed areas [36, 40]. However, the quality of QuikSCAT retrieved winds is degraded by rain contamination, especially in storm regions [41, 42]. Globally, rain is estimated to affect between 4% and 10% of SeaWinds data. Although rain flagging techniques have been developed [28, 43], the current wind retrieval method does not incorporate rain effects. As a result, rain attenuation and backscatter are interpreted as wind-induced features [44].

Rain corrupts the wind retrieval process by altering the wind-induced radar backscatter signature. Rain striking the water creates rings, stalks, and crowns from which the signal scatters [45]. Rain also alters the wind-induced capillary wave field, suppressing the wind/backscatter directional dependence [46], which may limit wind estimation in cases of wide-spread rain [47]. The scatterometer signal is additionally scattered and attenuated by falling hydrometeors.

In this chapter, we use a simple phenomenological backscatter model to characterize the effects of rain on SeaWinds on QuikSCAT backscatter. We assume that all rain effects are isotropic, allowing us to model the backscatter as a function of integrated rain rate and effective non-raining σ° in closed form. For comparison to rain-free backscatter conditions, the effective non-raining σ° from wind is estimated using numerical weather prediction (NWP) winds from the National Centers for Environmental Prediction (NCEP) projected through the GMF. In order to estimate the rain-induced parameters of the model, we use 100 co-located regions of Precipitation

Radar (PR) data from the Tropical Rainfall Measuring Mission (TRMM) satellite and SeaWinds on QuikSCAT data. Each co-located region contains the portion of the overlapping swaths in which the TRMM PR time tags are within ±10 minutes of the corresponding QuikSCAT time tags. The co-location set spans a 3-month period from August to October 1999. We use only the co-located regions where the overlapping TRMM PR swath contains more than 2.5% measurements flagged as "rain certain" in the TRMM 2A25 files, eliminating rain free areas from the analysis. The co-located set is limited to the tropics within ±37° of the equator, where rain contamination is the most frequent. While we use data from TRMM PR, previous analyses of the effects of rain on Ku-band backscatter have used data collected from ship [47], buoy [44], and radiometer [20] observations.

In Section 5.1, the SeaWinds and TRMM PR instruments are described. In Section 5.2, the rain/wind backscatter model is developed and the parameters of the model are estimated. In Section 5.3, an analysis of the surface perturbation due to rain is presented along with validation of the backscatter model as a function of rain rate and surface σ° from wind. Conclusions are given in Section 5.4. We demonstrate that the surface backscatter from rain dominates the rain-induced signal for low to moderate rain rates while atmospheric scattering dominates at high rain rates. Overall, the simple backscatter model estimates 94% of the co-located SeaWinds backscatter values to within 3 dB and corresponds well with a non-parametric estimate of the measured σ° as a function of integrated rain rate and effective non-raining wind σ°. Using the simple backscatter model, the conditions are determined for which estimation of rain from SeaWinds measurements is possible, and for which wind retrieval is not possible.

5.1 Data

Here, we describe each data source used in our rain model analysis, including the SeaWinds on QuikSCAT scatterometer, NCEP NWP winds, and the TRMM PR.

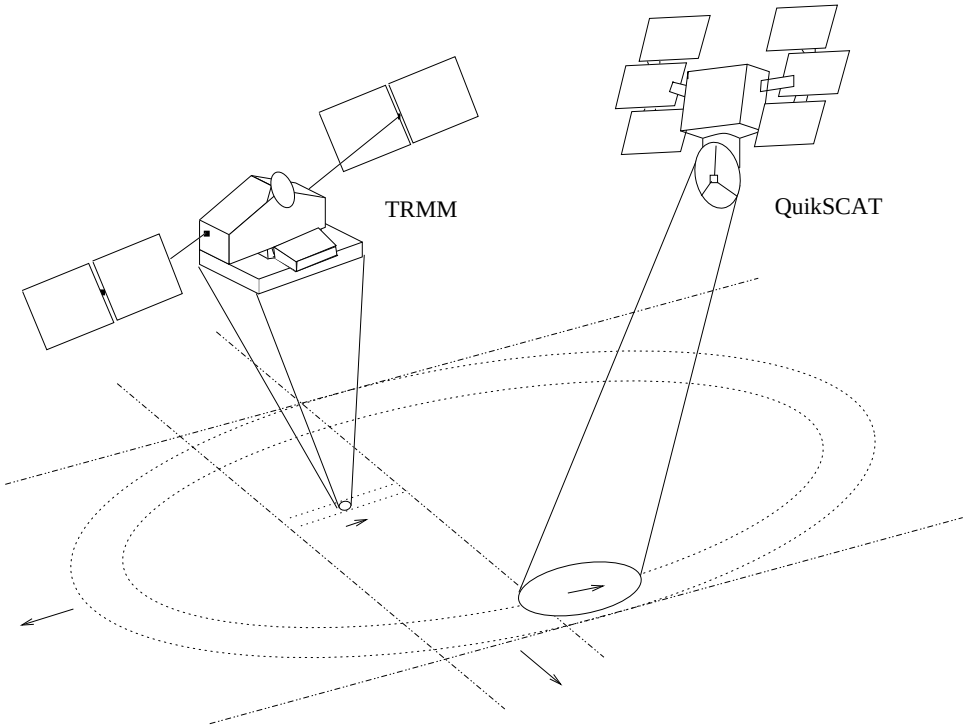


Figure 5.1: Swath geometry of the TRMM PR and SeaWinds on QuikSCAT instruments.

As explained in the Chapter 1, SeaWinds on QuikSCAT scatterometer is the first scanning pencil-beam instrument designed to measure marine winds. QuikSCAT revolves in a near-polar orbit and obtains measurements from over 90% of the earth daily. The outer (v-pol) beam operates at 54° incidence and has a 3 dB footprint of 37 × 52 km. The inner (h-pol) beam operates at 46° incidence and has a 3dB footprint of 34 × 44 km [1]. Because the h- and v-pol measurements have different characteristics, it is necessary to analyze them separately.

The wind estimation process produces several possible solutions at each wind vector cell, requiring "nudging" from an outside estimate of the wind to produce a unique vector field [14]. These nudging winds are obtained from NCEP 1000 millibar (mb) winds, included in the QuikSCAT level 2B (L2B) data files. They provide an estimate of the non-raining wind used in our analysis. The NCEP winds are trilinearly interpolated (in space and time) from a 2.5° × 2.5° latitude-longitude grid with a

temporal resolution of 6 hours [7] to the QuikSCAT wind vector cell locations. It is commonly known that the 1000 mb NCEP wind speeds are generally 10%-20% higher than the 10 m neutral stability winds measured by SeaWinds. This, along with prediction and interpolation errors, introduces a region-dependent bias in the SeaWinds-NCEP comparison data. To ameliorate these effects, a bias correction for NCEP is presented in Section 5.2.1. Overall, the RMS differences between SeaWinds on QuikSCAT and NWP analysis fields are $\sim$ 1.4 m/s for speeds from 3–20 m/s and 14 degrees for speeds from 5–20 m/s [48].

TRMM PR data is used to evaluate the effect of rain on the SeaWinds h- and v-pol backscatter. The TRMM satellite has been in operation since 1997 and orbits at a low inclination angle of 35°, providing coverage of the tropics. The PR instrument has a much higher resolution ($\sim 4 \times 4$ km) than SeaWinds and is electronically scanned within 17° of nadir [49]. Since the radar is near-nadir looking, the 200 km PR swath width is much narrower than SeaWinds 1800 km swath. Figure 5.1 compares the swath geometry of the TRMM PR and SeaWinds instruments. Although the viewing geometry of TRMM and SeaWinds is quite different, both instruments operate at approximately the same Ku-band frequency (13.8 Ghz for TRMM PR versus 13.4 GHz for SeaWinds). Thus, the atmospheric effects of rain on the SeaWinds scatterometer and TRMM PR are similar. Standard PR data products used here include unadjusted reflectivity in 3 spatial dimensions (Z_m) from the level 1C21 data files, level 2A25 integrated rain rate (R), and level 2A25 path integrated attenuation (PIA) [50].

5.2 Modeling σ° in rain and wind

This section presents a simple backscatter model and estimates the model parameters using the co-located data sources. As previously noted, rain striking the surface perturbs the wind-wave field and causes additional scattering from rain-induced surface features. Along with surface effects, rain in the atmosphere also attenuates and scatters the scatterometer signal. The scattering and attenuation

parameters can be combined into a simple phenomenological model,

$$\sigma_m = (\sigma_w + \sigma_{sr})\alpha_r + \sigma_r \tag{5.1}$$

where σ_m is the measured σ°, σ_w is the wind-induced radar backscatter, σ_{sr} is the surface backscatter perturbation due to rain striking the water, α_r is the two-way atmospheric rain attenuation, and σ_r is volume scattering due to falling rain droplets.

Equation (5.1) combines all surface rain effects into one additive perturbation parameter σ_{sr}. Although this relationship for the surface backscatter does not fully represent the complicated nature of surface scattering from wind- and rain-induced ocean waves, we are interested only in the average change in SeaWinds surface σ°. Thus, an additive relationship is sufficient for our first-order analysis.

The model of Eq. (5.1) is used to compare the surface perturbation to the atmospheric backscatter in Section 5.3.1. However, in the final analysis, we are only interested in the bulk augmentation of the scatterometer signal due to rain. Thus, the model is further simplified in Section 5.3.2 by summing the attenuated surface perturbation and the atmospheric scattering terms, creating a single effective rain backscatter parameter σ_e. The combined rain effect model is

$$\sigma_m = \sigma_w \alpha_r + \sigma_e \tag{5.2}$$

where $\sigma_e = \sigma_{sr}\alpha_r + \sigma_r$. This simplification is also used in [20].

5.2.1 Determining model parameters from TRMM and NCEP

We calculate the parameters from Eq. (6.1) by synergistically combining co-located data from the TRMM PR and NWP fields to estimate the SeaWinds-observed backscatter and attenuation from rain, and the effective non-raining backscatter from wind. Because TRMM PR is higher resolution than SeaWinds, comparison of the two datasets requires averaging of the TRMM parameters over the SeaWinds footprint to account for the aperture of the SeaWinds antenna. In order to make the notation tractable, when needed, we prime ($'$) parameters that have the resolution of the PR and remove the prime after averaging over the SeaWinds footprint. Also, parentheses

in the subscripts of symbols are used to indicate the source of the data used to estimate that parameter.

The SeaWinds on QuikSCAT measured backscatter $\sigma_{m(QSCAT)}$ is taken from the QuikSCAT L1B data set and corrected for atmospheric (non-rain) attenuation using the QuikSCAT L2A attenuation map. These data were obtained from the NASA Physical Oceanography Distributed Active Archive Center at the Jet Propulsion Laboratory/California Institute of Technology.

In evaluating the effect of rain on SeaWinds, the effective vertically-integrated rain rate observed by SeaWinds is estimated from the TRMM 2A25 integrated rain rate. The PR rain rate $R'_{(PR)}$ is averaged over the SeaWinds footprint, weighting each data point by the SeaWinds two-way antenna gain pattern,

$$R_{(PR)} = \frac{\sum_{i=1}^{N} G_i R'_{i(PR)}}{\sum_{i=1}^{N} G_i}, \tag{5.3}$$

where G_i is the SeaWinds antenna gain pattern at the i^{th} PR measurement, N is the number of PR data points within a 6-dB antenna pattern contour, and $R'_{i(PR)}$ is the PR-observed rain rate. We note that this rain rate estimate is integrated (essentially) normal to the surface, and thus does not take into account the slant range of the SeaWinds instrument. The antenna-weighted rain rate differs somewhat from an unweighted average due to the spatial variability of the rain within the antenna footprint. This "beam-filling" variability is examined in Chapter 6 by evaluating the error between the average rain rate weighted by the antenna beam pattern and the unweighted average rain for individual measurements. A histogram of the PR-derived rain rates for SeaWinds over the co-located data set is shown in Figure 5.2. The co-located data set allows evaluation of rain effects for rain rates up to about 100 km mm/hr.

The effective non-raining wind backscatter σ_w is computed from co-located NWP estimates of the wind from NCEP included in the QuikSCAT level 2B data files. As previously noted, these wind fields are interpolated to the same grid as the SeaWinds 25 $\times$ 25 km wind product. For each SeaWinds observation, we further interpolate the NCEP wind data to the center of each SeaWinds σ° measurement

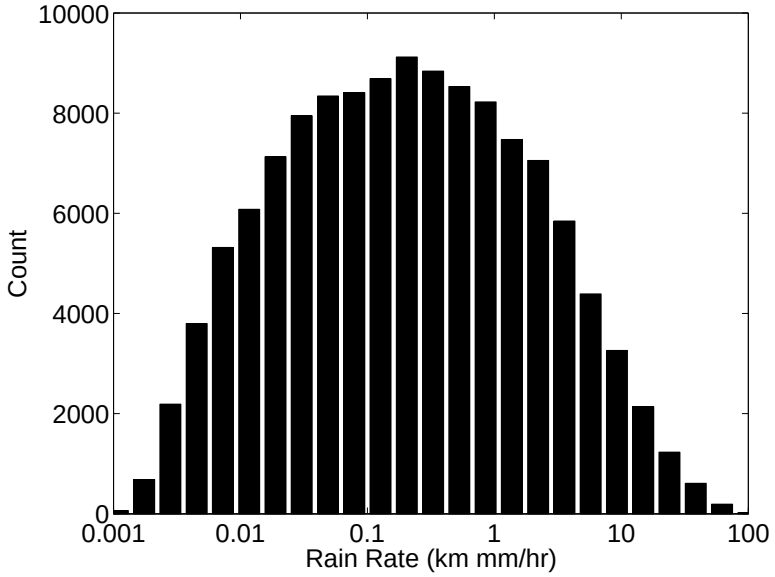


Figure 5.2: Histogram of rain rates for the co-located data set derived from TRMM PR data

using cubic spline interpolation of the zonal and meridional components of the wind. The process is not particularly sensitive to the interpolation method since the NCEP winds are low resolution. Each NCEP wind vector is then projected through the GMF to yield an estimate of the non-raining wind backscatter,

$$\sigma_{w(NCEP)} = \mathcal{M}(u_{(NCEP)}, \chi_{(NCEP)}, \theta, \text{pol}) \tag{5.4}$$

where $u_{(NCEP)}$ is the NCEP model wind speed, $\chi_{(NCEP)}$ is the relative azimuth angle of the NCEP wind direction, θ is the incidence angle and pol is the polarization of the QuikSCAT measurement.

As mentioned earlier, the wind backscatter estimate from NCEP has some bias due to prediction errors, errors in the GMF, and differences in reference height between the 10 m SeaWinds wind product and 1000 mb NCEP winds. Because NCEP is low resolution, we assume that the bias error is spatially correlated. For each SeaWinds backscatter measurement, we estimate the bias error ϵ as a Gaussian-weighted average of the error between SeaWinds and NCEP σ° values for surrounding measurement cells in the same look direction (i.e. fore/aft) with very little or no rain.

For the j^{th} measurement,

$$\epsilon^j = \sum_i W^{ij}(\sigma^i_{m(QSCAT)} - \sigma^i_{w(NCEP)}) / \sum_i W^{ij} \tag{5.5}$$

where i sums over all measurements in the same look direction with rain rate less than 0.05 km mm/hr as determined by TRMM PR. The threshold of 0.05 km mm/hr is chosen because it is approximately the rain rate achieved for a SeaWinds footprint when only one co-located PR cell contains a rain rate of 0.7 km mm/hr. The value of 0.7 km mm/hr is the lower bound of the sensitivity of the TRMM PR for a rain height of 1 km [49]. The weighting W^{ij} is calculated by

$$W^{ij} = \exp\left(\frac{1}{2}\frac{d(i,j)^2}{20^2}\right) \tag{5.6}$$

where $d(i,j)$ is the euclidean distance between the i^{th} and j^{th} measurements in km. The distance scale of 20 km (yielding an effective 3-σ diameter of 120 km) is subjectively chosen; however, because the NCEP wind are low resolution, the bias estimate is not particularly sensitive to this value. Our estimate of σ_w is now written as the NCEP-determined value plus the bias error,

$$\sigma_w = \sigma_{w(NCEP)} + \epsilon. \tag{5.7}$$

The mean ϵ over the QuikSCAT/TRMM PR data set is $\sim$ -0.0025 while the standard deviation is $\sim$ 0.0064, suggesting that NCEP predicted wind backscatter is biased slightly high, as expected.

The estimate of the NCEP-QuikSCAT σ° bias error is most accurate when there are no fine-scale wind features in the near vicinity of the QuikSCAT measurement. Because fine-scale features may be misplaced or absent in the low-resolution NCEP data, the estimate of σ_w error is expected to be less reliable in such areas. Also, in regions of wide-spread rain, the estimate is less reliable because there are fewer no-rain measurements in the vicinity from which to make our estimate.

The two-way atmospheric attenuation factor α_r of Eq. (5.1) is calculated from the PIA estimates in the TRMM 2A25 data sets. First, the PIA measurements are adjusted for the PR slant range and converted to normal space,

$$\alpha'_{r(PR)} = 10^{-\cos\theta_{(PR)}\mathrm{PIA(2A25)}/10}. \tag{5.8}$$

Values for $\alpha'_{r(PR)}$ where there is no rain detected are set to one. For each SeaWinds observation, the co-located values of $\alpha'_{r(PR)}$ are weighted-averaged over the 6dB SeaWinds footprint to yield the effective SeaWinds on QuikSCAT columnar path integrated attenuation, and then scaled by the secant of the SeaWinds incidence angle $\theta_{(QSCAT)}$ to adjust for the SeaWinds slant range,

$$\mathrm{PIA} = -\sec\theta_{(QSCAT)} 10\log_{10}\left(\frac{\sum_{i=1}^{N} G_i \alpha'_{ri(PR)}}{\sum_{i=1}^{N} G_i}\right) \tag{5.9}$$

where $\alpha'_{ri(PR)}$ is the i^{th} PIA value. The attenuation parameter is then converted to normal space,

$$\alpha_{r(PR)} = 10^{-\mathrm{PIA}/10}. \tag{5.10}$$

Estimates for the volume-scattering cross-section (σ_r) are calculated from measured reflectivities (Z_m) obtained from the TRMM 1C21 data set. The actual (attenuation-adjusted) reflectivity of the atmospheric rain (Z_e) is related to the measured reflectivity through the equation,

$$Z_m(r) = Z_e(r)\alpha_r(r) \ \ \mathrm{mm^6 m^{-3}} \tag{5.11}$$

where r is the range, and $\alpha_r(r)$ is the path integrated two-way attenuation at range r. The volume backscattering coefficient can be found from [5],

$$\sigma_{vr}(r) = 10^{-10}\frac{\pi^5}{\lambda_0^4}|K_w|^2 Z_e(r) \ \ \mathrm{m^2/m^3} \tag{5.12}$$

where λ (cm) is the electromagnetic wavelength of SeaWinds, and $|K_w|^2$ is a coefficient related to the absorption properties of water (assumed to be 0.9). The quantity σ_{vr} is radar backscatter cross-section per unit volume.

The volume backscattering cross-section observed by the satellite (σ_{vro}) is attenuated by the two-way attenuation factor, $\alpha_r(r)$ and is equal to,

$$\begin{aligned}
\sigma_{vro}(r) &= \sigma_{vr}(r)\alpha_r(r) \\
&= 10^{-10}\frac{\pi^5}{\lambda_0^4}|K_w|^2 Z_e(r)\alpha_r(r) \\
&= 10^{-10}\frac{\pi^5}{\lambda_0^4}|K_w|^2 Z_m(r).
\end{aligned} \tag{5.13}$$

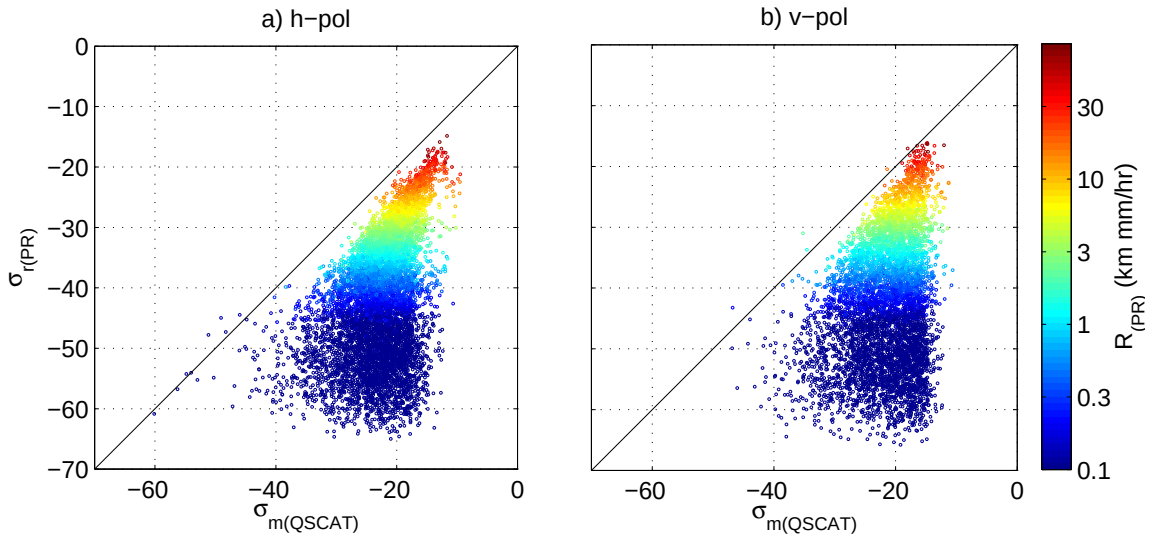


Figure 5.3: Relationship of $\sigma_{r(PR)}$ to $\sigma_{m(QSCAT)}$.

The total atmospheric rain backscatter as seen by the PR ($\sigma'_{r(PR)}$) is σ_{vro} integrated through the PR antenna beam to the lowest no-surface-clutter range (r_{nc}),

$$\begin{aligned} \sigma'_{r(PR)} &= \int_0^{r_{nc}} \sigma_{vro}(s)ds \quad \mathrm{m}^2/\mathrm{m}^2 \\ &\approx \sum_{s=1}^{N_{nc}} \sigma_{vro}(s)\Delta r \quad \mathrm{m}^2/\mathrm{m}^2 \end{aligned} \tag{5.14}$$

where Δr is the vertical range resolution of the PR and N_{nc} is the lowest no-surface-clutter range bin. All valid $\sigma'_{r(PR)}$ values are weighted-averaged over the SeaWinds footprint,

$$\sigma_{r(PR)} = \frac{\sum_{i=1}^{N} G_i \sigma'_{ri(PR)}}{\sum_{i=1}^{N} G_i} \tag{5.15}$$

where $\sigma'_{ri(PR)}$ is the i^{th} PR rain backscatter value. Values for $\sigma'_{ri(PR)}$ where there is no rain detected are set to zero. Figure 5.3 is a scatter plot illustrating the relationship between the QuikSCAT measured backscatter and the atmospheric rain backscatter observed by the PR. In addition, the effect of the rain on SeaWinds backscatter and retrieved winds is shown in Figure 5.4, depicting a high correlation between the integrated backscatter from rain seen by the PR, and the SeaWinds backscatter.

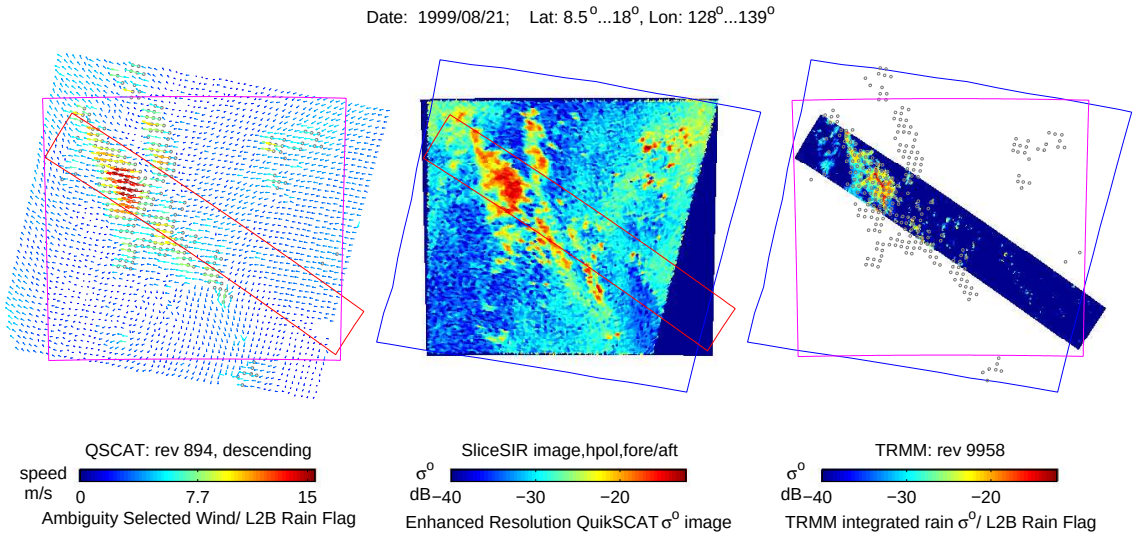


Figure 5.4: Ambiguity selected wind in a QuikSCAT/TRMM PR co-location. Also, an enhanced resolution backscatter image for the fore and aft beams (h-pol), and integrated rain backscatter calculated from TRMM PR.

Thus far we have derived estimates for the atmospheric scattering and attenuation from rain as seen by SeaWinds. We have assumed that the PR-derived atmospheric backscatter σ_r can be directly applied to SeaWinds. However, because SeaWinds has a longer slant range and a different incidence angle, we expect some error in our estimate of σ_r. The TRMM PR and SeaWinds measurements are only co-located at the ocean surface. For increasing height, the TRMM PR and SeaWinds beams become misaligned. Temporal changes in rain profiles between the TRMM and SeaWinds observation times introduce additional errors. Also, we expect a lower sensitivity at v-pol than h-pol [20]. We do not attempt to model the all the sources of error in σ_r. Rather, we evaluate the sensitivity of the overall model to the PR-derived estimate of σ_r in Section 5.3.1.

5.2.2 Relating PR atmospheric parameters to rain rate

Here, we relate the PR-derived atmospheric parameters $\alpha_{r(PR)}$ and $\sigma_{r(PR)}$ to the integrated rain rate. Power-law (linear log-log) models are widely used to relate rain rate R to the rain extinction coefficient κ_{er} and reflectivity Z_e, from which α_r

and σ_r are derived [5, 50]. The energy transfer equations relate κ_{er} and Z_e to α_r and σ_r by

$$\text{PIA} = -10\log_{10}(\alpha_r) = 2\int_0^{r_{nc}} \kappa_{er}(r)dr \tag{5.16}$$

and

$$\sigma_r = C\int_0^{r_{nc}} Z_e(r)10^{-\left(2\int_0^r \kappa_{er}(r')dr'\right)/10}dr \tag{5.17}$$

where $C = 10^{-10}\pi^5\lambda_0^{-4}|K_w|^2$. Although we are not directly interested in κ_{er} and Z_e, the linear nature of Eq. (5.16) suggests that a linear log-log model may suffice for estimating PIA. However, the non-linear attenuation term of Eq. (5.17) may cause σ_r to somewhat deviate from a power-law model. Due to this non-linearity in Eq. (5.17) and the fact that we do not include information about rain profile or type, we model the PR-derived atmospheric parameters with both linear and quadratic log-log fits, i.e.

$$10\log_{10}(\text{PIA}) = \quad f_a(R_{dB}) \quad = \sum_{n=0}^{N} x_a(n)R_{dB}^n \tag{5.18}$$

$$10\log_{10}(\sigma_r) = \quad f_r(R_{dB}) \quad = \sum_{n=0}^{N} x_r(n)R_{dB}^n \tag{5.19}$$

where $R_{dB} = 10\log_{10}(R)$, and N is either one or two.

The values for α_r and σ_r can be cast into matrix equations of the form

$$\mathbf{PIA} \quad = \quad \mathbf{R}\mathbf{x_a} \tag{5.20}$$

$$\tilde{\sigma}_\mathbf{r} \quad = \quad \mathbf{R}\mathbf{x_r} \tag{5.21}$$

where

$$\mathbf{R} = [\mathbf{1} \quad \mathbf{r}] \tag{5.22}$$

for the linear log-log model and

$$\mathbf{R} = [\mathbf{1} \quad \mathbf{r} \quad \mathbf{r}^2] \tag{5.23}$$

for the quadratic log-log model. In Eq. (5.20), the elements of **PIA** are equal to $10\log_{10}$ PIA where PIA is related to α_r by Eq. (5.16). In Eq. (5.21), $\tilde{\sigma}_\mathbf{r}$ is a vector containing σ_r values in decibels. In Eqs. (5.22) and (5.23), $\mathbf{r}$ is a vector containing the

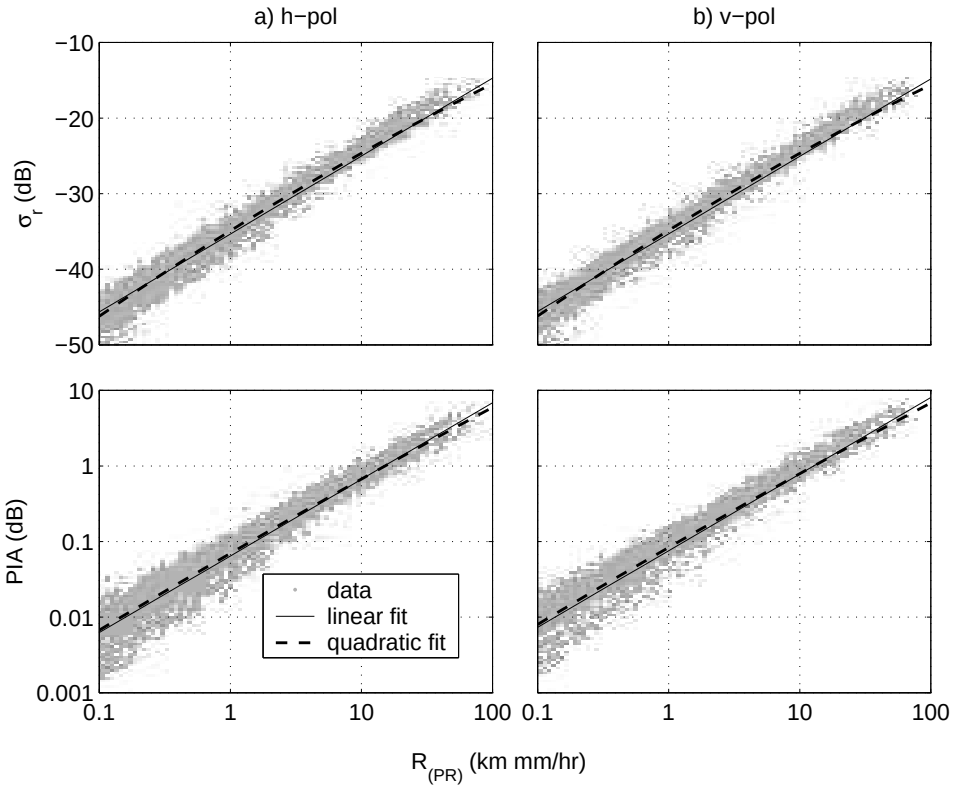


Figure 5.5: Parametric estimation of the parameters σ_r and PIA as a function of rain rate for a) h-pol and b) v-pol data.

rain rate in decibels (R_{dB}). The estimation of $\mathbf{x_a}$ and $\mathbf{x_r}$ is performed by first making a non-parametric (moving average) estimate of $10\log_{10}\mathrm{PIA}$ and $10\log_{10}\sigma_{r(PR)}$ as a function of rain rate at regular logarithmically spaced rain rate bins. In doing this, we use an Epanechnikov kernel [51] with a support of 3 dB in rain rate. The parameters $\mathbf{x_a}$ and $\mathbf{x_r}$ are solved for using linear least-squares,

$$\mathbf{x_a} = \mathbf{R}^{\dagger}\mathbf{PIA} \tag{5.24}$$

$$\mathbf{x_r} = \mathbf{R}^{\dagger}\tilde{\sigma}_{\mathbf{r}} \tag{5.25}$$

where $\mathbf{PIA}$ and $\tilde{\sigma}_{\mathbf{r}}$ contain the non-parametric estimated values and $\mathbf{R}^{\dagger}$ is the least-squares pseudo-inverse of $\mathbf{R}$ which is formed using the rain rate bins. The PR data and model fits are shown in Figure 5.5. The estimated parameters are given in Table 5.1.

Table 5.1: Coefficients of the linear and quadratic fits to the parameters PIA and $\sigma_{r(PR)}$

	h-pol		v-pol	
	linear	quadratic	linear	quadratic
$x_a(1)$	-11.90	-11.55	-11.21	-10.78
$x_a(2)$	1.01	1.00	1.01	1.00
$x_a(3)$	–	-0.0017	–	-0.0021
$x_r(1)$	-35.34	-34.90	-35.32	-34.85
$x_r(2)$	1.03	1.07	1.03	1.08
$x_r(3)$	–	-0.0053	–	-00057

5.3 Analysis

Thus far, all rain-induced parameters of Eq. (6.1) except the surface perturbation σ_{sr} have been estimated. The surface rain perturbation is often ignored by theoretical models due to a lack of understanding of its effects. However, it is important to include the surface perturbation to afford the best modeling of rain. This can be done directly using the wind/rain backscatter model from Eq. (6.1), or indirectly using the combined rain effect model of Eq. (5.2). In Section 5.3.1, we estimate the surface perturbation directly using the co-located QuikSCAT/TRMM PR data and compare the effect of the surface perturbation to that of the atmospheric backscatter. In Section 5.3.2, we apply the combined rain effect model to SeaWinds on QuikSCAT data. Using the combined rain effect model, the conditions for which wind and rain retrieval are possible from scatterometer data is given in Section 5.3.3.

5.3.1 Estimating the surface perturbation

This section estimates the surface perturbation and presents a qualitative comparison of the effect of the surface perturbation to the atmospheric rain backscatter. Surface backscatter from rain striking the water has been studied at low incidence angles (30°) in wave-tank data [29, 52], and at higher incidence angles ($31°, 41°, 51°, 76°$) from ship observations [47], but not at SeaWinds' specific geometries. The QuikSCAT/TRMM PR co-located data affords us the ability to evaluate

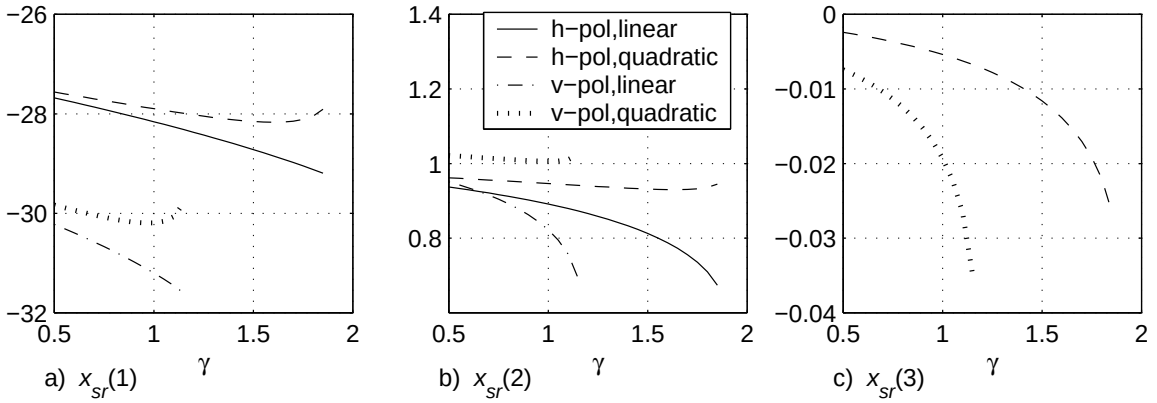


Figure 5.6: Each parameter of $x_{sr}^{\gamma}(n)$ as a function of γ for both h-pol and v-pol observations and both linear and quadratic fits.

the contribution of the surface rain perturbation to the total rain backscatter. An estimate of the surface perturbation is obtained by solving Eq. (6.1) for σ_{sr} and adding appropriate subscripts,

$$\sigma_{sr} = \alpha_{r(PR)}^{-1}(\sigma_{m(QSCAT)} - \sigma_{r(PR)}) - (\sigma_{w(NCEP)} + \epsilon). \tag{5.26}$$

In comparing the estimated surface perturbation to the atmospheric backscatter, we recognize that error in the PR-derived estimates of the atmospheric rain backscatter $\sigma_{r(PR)}$ arises from differences in slant range, polarization, and calibration between the PR and SeaWinds and introduces a bias into the estimate of σ_{sr}. To evaluate the sensitivity of σ_{sr} to the error, we adopt a variable calibration parameter γ that is multiplied by $\sigma_{r(PR)}$. The modified estimate for σ_{sr} is

$$\sigma_{sr}(\gamma) = \alpha_{r(PR)}^{-1}(\sigma_{m(QSCAT)} - \gamma\sigma_{r(PR)}) - (\sigma_{w(NCEP)} + \epsilon). \tag{5.27}$$

As with the PR-derived atmospheric parameters, we write the "calibrated" surface perturbation $\sigma_{sr}(\gamma)$ as a linear and quadratic function of rain rate in log-log space, and solve for it using a similar method as described in Section 5.2.2 for varying γ. The model for σ_{sr} is written as

$$10\log_{10}(\sigma_{sr}(\gamma)) \approx f_{sr}^{\gamma}(R_{dB}) = \sum_{n=0}^{N} x_{sr}^{\gamma}(n) R_{dB}^{n}. \tag{5.28}$$

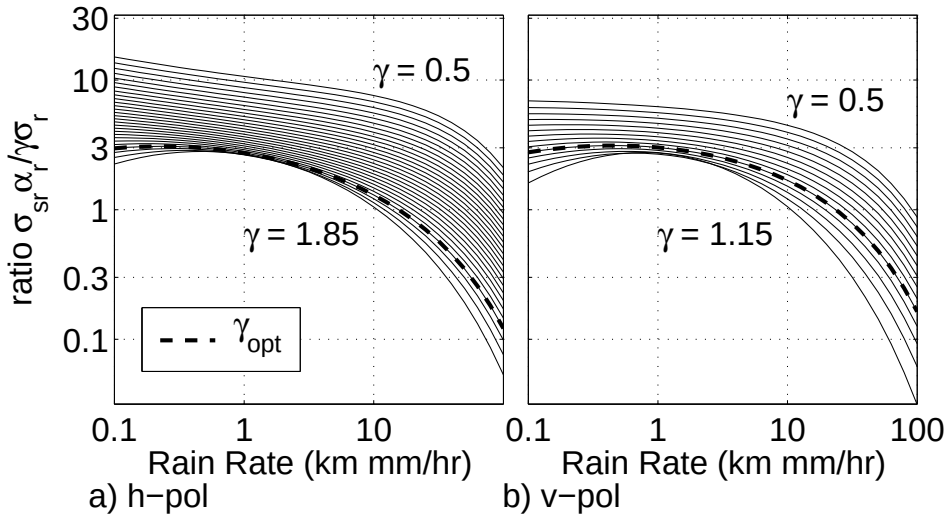


Figure 5.7: Ratio of the attenuated surface perturbation $\alpha_r\sigma_{sr}(\gamma)$ to the modified atmospheric rain backscatter $\gamma\sigma_r$ for several values of γ plotted as a function of rain rate for a) h-pol and b) v-pol.

Because the estimate of the surface perturbation $\sigma_{sr}(\gamma)$ can be negative, the non-parametric estimate described in Section 5.2.2 is performed in normal-space for $\sigma_{sr}(\gamma)$. The non-parametric estimate is then converted to decibels and a least-squares estimate of $\sigma_{sr}(\gamma)$ is formed for a range of γ values.

The estimated parameters of $\sigma_{sr}(\gamma)$ are plotted as a function of the calibration parameter γ in Figure 5.6 for both linear and quadratic fits. As a note, the estimates are only valid for v-pol observations up to $\gamma = 1.15$ and h-pol observations up to $\gamma = 1.85$. At a larger γ, the non-parametric estimate of σ_{sr} becomes negative. From Figure 5.6a, the constant term $x_{sr}^{\gamma}(1)$ is not particularly sensitive to the choice of γ. The highest order terms, $x_{sr}^{\gamma}(2)$ for the linear model and $x_{sr}^{\gamma}(3)$ for the quadratic model, are the most sensitive, especially as γ increases.

The observed backscatter due to rain is the sum of the attenuated surface perturbation $\alpha_r\sigma_{sr}$ and the atmospheric rain backscatter σ_r. A direct comparison of the estimated surface perturbation parameters (Figure 5.6) to the atmospheric scattering parameters (Table 5.1) shows that for all γ, the constant term of the

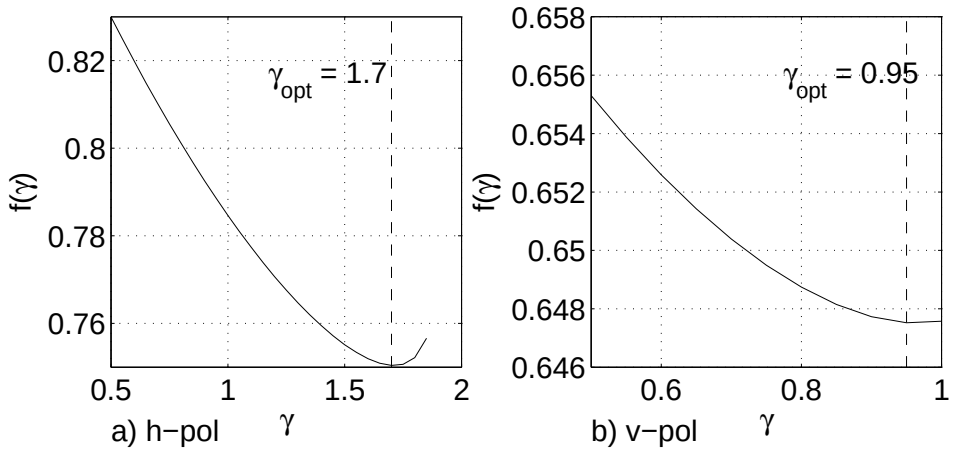


Figure 5.8: Objective function with respect to γ (see text). The γ that minimizes the objective function is displayed with a dashed line.

surface perturbation is significantly higher than the constant term of the atmospheric scattering (compare Figure 5.6a to Table 5.1 row 4). Also, the linear and quadratic terms of $x_{sr}^{\gamma}(n)$ are on the same order of magnitude as the the linear and quadratic terms of $x_r(n)$. This comparison demonstrates that the surface perturbation is a significant contributer to the observed rain backscatter.

We further compare the contribution of the surface perturbation to the atmospheric backscatter by computing the ratio of the attenuated surface perturbation $\alpha_r \sigma_{sr}(\gamma)$ for each γ to the calibrated atmospheric rain backscatter $\gamma\sigma_r$ as a function of rain rate (see Figure 5.7). The ratio $\alpha_r \sigma_{sr}(\gamma)/\gamma\sigma_r$ is greater than one for rain rates less than 10 km mm/hr, suggesting that the surface rain backscatter dominates at low to moderate rain rates. At higher rain rates, the surface signal is suppressed by attenuation and atmospheric scattering dominates.

Although we have oversimplified the error in σ_r by modeling it as a simple multiplicative constant γ, it is of value to consider the question "What γ gives the best fit to the data and is the overall model sensitive to our choice of γ?" We approach this question by finding an optimum γ that minimizes a least-squares objective function given varying values of γ. The least-squares objective function is a measure of the

Table 5.2: Estimated parameters of $\tilde{\sigma}_{\mathbf{sr}}$ for h-pol and v-pol for the optimum γ.

	h-pol ($\gamma = 1.70$)	v-pol ($\gamma = 0.95$)
$x_{sr}^{\gamma}(1)$	-28.1	-30.2
$x_{sr}^{\gamma}(2)$	0.93	1.01
$x_{sr}^{\gamma}(3)$	-0.017	-0.017

least-squares error between the QuikSCAT-measured σ° values and the model values given a certain γ,

$$f(\gamma) = \sum_{i} \left(\sigma_{m}^{i} - \sigma_{m(CALC)}^{i}(\gamma)\right)^{2} \tag{5.29}$$

where i is the data index and $\sigma_{m(CALC)}(\gamma)$ is the model-calculated σ° value given γ and the corresponding model for σ_{sr}. The quadratic model for σ_{sr} is used rather than the linear model because it it accounts for more of the variance than the linear model. Now, we compute $f(\gamma)$ at each valid value of γ. The objective function for both h-pol and v-pol is plotted against γ in Figure 5.8. For h-pol, the objective function has a minimum where $\gamma = 1.70$. For v-pol, the objective function has a minimum where $\gamma = 0.95$. The values of $x_{sr}^{\gamma}(n)$ corresponding to the optimum γ are given in Table 5.2. The optimum values for γ confirm that the SeaWinds h-pol response to atmospheric rain is greater than the v-pol response. The shape of the objective function, however, suggests a high variance on our estimate, and that the overall model is not very sensitive to the exact choice of γ.

5.3.2 Combined rain effect model

The wind/rain backscatter model affords a closed-form relationship between ocean backscatter due to wind and rain rate. However, because we do not precisely know the surface perturbation or error in the atmospheric backscatter estimate, we adopt the combined rain effect model of Eq. (5.2) to further analyze the effect of rain on SeaWinds scatterometer measurements. Using the co-located data, we calculate estimates for the effective rain backscatter,

$$\sigma_{e} = \sigma_{m(QSCAT)} - (\sigma_{w(NCEP)} + \epsilon)\alpha_{r(PR)}, \tag{5.30}$$

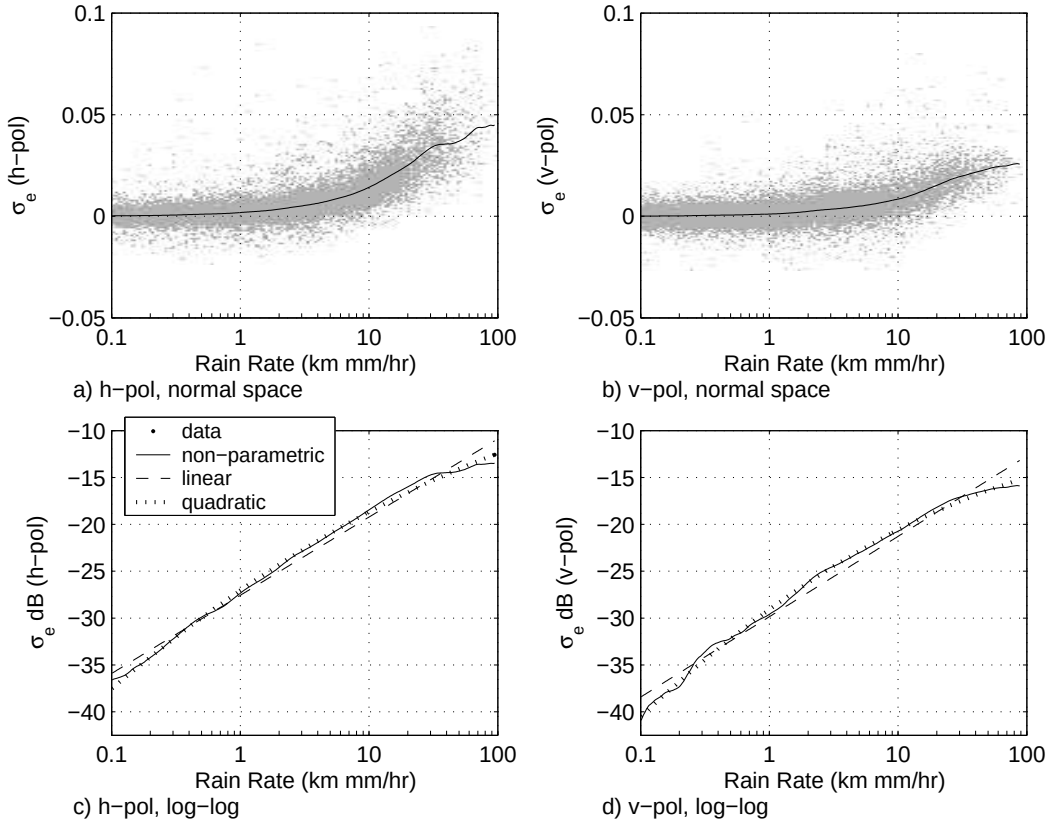


Figure 5.9: Non-parametric and least-squares fits to the parameter σ_e. a) and b) show the data and non-parametric fit in normal space for h-pol and v-pol. c) and d) show the non-parametric, and linear and quadratic fits in log-log space.

and use the same estimation technique to calculate σ_e as a function of rain rate as performed with σ_{sr} (see Section 5.3.1). The effective rain backscatter can be written as a function of rain rate,

$$10\log_{10}(\sigma_e) \approx f_e(R_{dB}) = \sum_{n=0}^{N} x_e(n) R_{dB}^n. \tag{5.31}$$

The optimization is performed and the estimated values for $x_e(n)$ for both linear and quadratic fits are given in Table 5.3. Graphics demonstrating the model fits are shown in Figure 5.9. The estimates σ_e range from -40 to -15 dB for the observed rain rates, demonstrating a significant sensitivity of SeaWinds on QuikSCAT data to rain.

Table 5.3: Estimated parameters of σ_e for h-pol, v-pol and both linear and quadratic fits.

	h-pol		v-pol	
	linear	quadratic	linear	quadratic
$x_e(1)$	-27.57	-27.04	-29.86	-29.09
$x_e(2)$	0.83	0.94	0.86	1.00
$x_e(3)$	-	-0.011	-	-0.015

Now that the parameters of the wind/rain backscatter models have been estimated, the measured SeaWinds backscatter can be parameterized by the surface backscatter due to wind and rain rate as

$$\sigma_m(\sigma_w, R) = \sigma_w \; \alpha_r(R) + \sigma_e(R) \tag{5.32}$$

for the combined rain effect model and

$$\sigma_m(\sigma_w, R) = (\sigma_w + \sigma_{sr}(R)) \; \alpha_r(R) + \sigma_r(R) \tag{5.33}$$

for the full model. Each rain-induced parameter has a linear or quadratic log-log relationship to rain rate of the form

$$\sigma_e(R) = 10^{f_e(R_{dB})/10} \tag{5.34}$$

$$\sigma_{sr}(R) = 10^{f_{sr}^{\gamma}(R_{dB})/10} \tag{5.35}$$

$$\alpha_r(R) = 10^{-10^{f_a(R_{dB})/10}/10} \tag{5.36}$$

$$\sigma_r(R) = \gamma_{opt} 10^{f_r(R_{dB})/10}. \tag{5.37}$$

Comparing the two models, we find that the full model and the combined rain effect model are very similar. Thus, for simplicity, the remaining analysis is applied only to the quadratic combined rain effect model. Similar results follow for the full model.

Beamfilling: Worst Case Scenario

Chapter 6 addresses the effect of beamfilling with respect to the averaging procedure. Here, we examine the beamfilling effect with respect to the spatial

distribution of the rain. In doing so, we consider a worst case scenario of two measurements where the integrated rain rate, averaged over the SeaWinds footprint, is 0.5 km mm/hr. In the first measurement, the rain rate is 0.5 km mm/hr, uniformly distributed over the measurement area. In the second measurement, the rain rate is 50 km mm/hr, isolated to the area of a single PR measurement cell (with approximately 100 PR cells co-located within the SeaWinds measurement). Using the quadratic combined rain effect model, the effective rain backscatter for the uniform 0.5 km mm/hr rain rate is $\sigma_e = 0.001$ (-30 dB). The effective rain backscatter for the isolated rain event of $R = 50$ km mm/hr, is $\sigma_e = 0.039$. Averaging over the SeaWinds footprint yields an effective backscatter of $\sigma_e = 0.00039$ (-34 dB). Thus, the effect of the uniform rain differs from the worst-case non-uniform rain by about 4 dB. However, the effect of non-uniform beam filling is generally much less substantial. We treat the effect of non-uniform rain as a source of unknown variability in the measurement model, and evaluate the overall accuracy of the model in the following section.

Model Validation

Here we compare the combined rain effect model to actual SeaWinds backscatter measurements as a function of $R_{(PR)}$ and $\sigma_{w(NCEP)} + \epsilon$. Figures 5.12 and 5.13 show the quadratic combined rain effect model plotted against SeaWinds on QuikSCAT backscatter for h- and v-pol respectively. Visually, the model follows the data very well. In fact, the model is within 3 dB of the measurements more than 94% of the time. The standard deviation of the log error is about 1.6 dB for h-pol and 1.5 dB for v-pol. The relatively low variance of the model indicates that the simple closed-form model is sufficient to describe the effects of rain on the SeaWinds-measured σ° to a high level of accuracy.

For further comparison of the model to the data, we create a non-parametric estimate of σ_m on a regular grid with axes of $\sigma_{w(NCEP)} + \epsilon$ and $R_{(PR)}$. The non-parametric estimate is formed using a two-dimensional Epanechnikov kernel [51]. The non-parametric estimate, along with the quadratic combined rain effect model

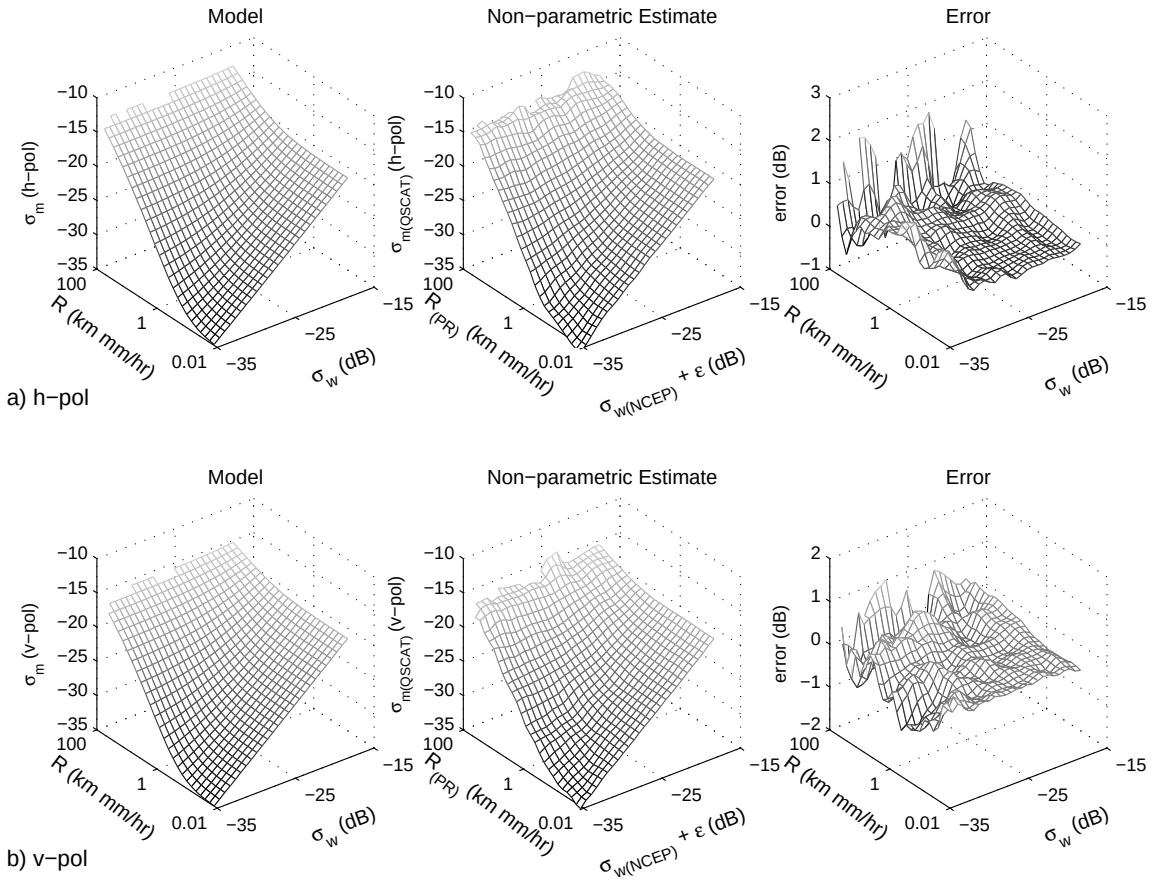


Figure 5.10: Non-parametric, and combined rain effect models for a) h-pol and b) v-pol. Also, error between parametric and non-parametric models.

are computed on the regular grid and shown in Figure 5.10 for h-pol and v-pol. Both h-pol and v-pol models are very close to the non-parametric estimate. The largest error occurs at high rain rates. However, there is less data at high rain rates, increasing the uncertainty in the estimate.

5.3.3 Data regimes

By examining the wind-rain model, we notice three distinct backscatter regimes. Regime one is where the rain backscatter σ_e is large compared to the attenuated wind-induced surface backscatter $\alpha_r \sigma_w$ and rain dominates the backscatter return. Regime two is where the backscatter from rain and wind effects are on the

same order of magnitude. Regime three occurs where the wind-induced backscatter dominates rain effects. We can identify these regions by thresholding the ratio of σ_e to the model estimate of σ_m. We define regime 1 by $\sigma_e/\sigma_m > .75$ and regime 3 as $\sigma_e/\sigma_m < .25$. Between these regimes is regime two. The regimes are plotted along with the combined rain effect model in Figure 5.11.

The three regimes help us understand under what conditions rain and/or wind information can be extracted from SeaWinds data. Using the model, rain rates can readily be computed from data in regime 1, but not from data in regime 3. Wind information from data in regime 1 may be unobtainable, while current wind retrieval methods may be sufficient for accurate estimates from data in regime 3. It is also likely that rain and wind information may be simultaneously retrieved from data in regime 2 by incorporating a rain model into the wind retrieval method (and improving the current scatterometer wind estimate).

Of all h-pol QuikSCAT measurements in the co-location data set with significant rain ($> .1$ km mm/hr), only 14% fall in regime 1. About 41% fall in regime 2, and 44% fall in regime 3. For v-pol, about 7% of measurements fall in regime 1, 29% in regime 2, and 64% in regime 3. If we accept the 3-month co-location data set as typical, these percentages suggest that by including the rain model in the wind retrieval method, winds can be improved for a significant fraction of rain-corrupted measurements (data from regime 2). Also, because 36% to 56% of the measurements with significant rain fall into either regime 1 or 2, we may be able to retrieve the rain from nearly half of all QuikSCAT rain-contaminated measurements. We again note that only 4-10% of the measurements are effected by rain.

5.4 Conclusions

Rain is one of the main sources of error in QuikSCAT scatterometer winds. Current wind retrieval methods are limited by our knowledge of the effects of rain on scatterometer data. Although backscattering and attenuation from rain are successfully modeled, rain surface effects are not generally well understood.

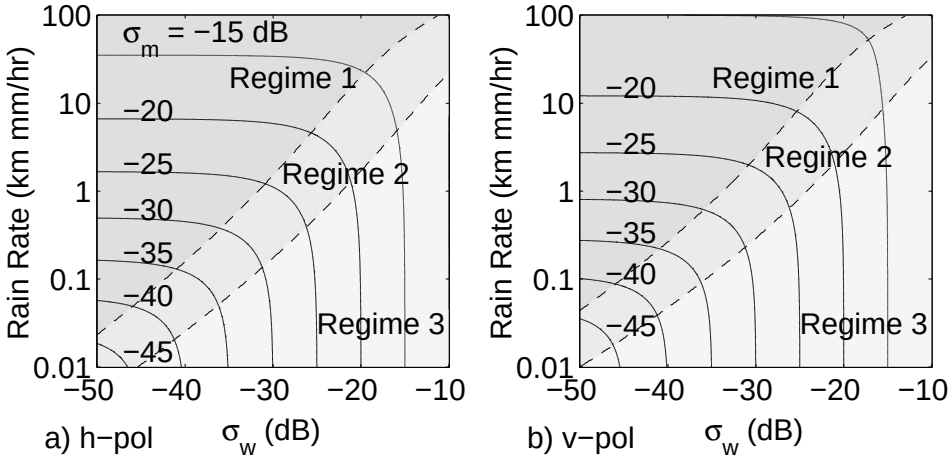


Figure 5.11: Backscatter regimes for SeaWinds as a function of rain rate and effective wind backscatter for a) h-pol and b) v-pol models. Also plotted is a contour plot of the combined rain effect model for σ_m.

By synergistically combining TRMM PR and QuikSCAT data, we have demonstrated that a simple low-order model is adequate for describing the wind/rain interaction as a function of rain rate and non-raining effective wind σ°. Also, our analysis suggests that the surface perturbation due to rain striking the water dominates the total scattering from rain for low to moderate rain rates (< 10 km mm/hr). Since backscatter from rain-induced waves significantly contribute to the total backscatter, surface effects must be included in the rain/wind model.

In the rain/wind model, there are three distinct regimes. In the first regime, the rain dominates. It may be possible to directly calculate the rain rate from data in this regime. In the second regime, the signal from the rain and the wind are on the same order. Data from the second regime may be used to simultaneously retrieve wind and rain. In the third regime, the signal from the wind dominates and current wind retrieval techniques are adequate. These regimes help us understand where rain can be retrieved and where rain-contaminated wind can be corrected.

Chapter 6 introduces a simultaneous wind/rain retrieval method using the combined rain effect model. The new method demonstrates that the use of the rain

model in wind estimation can reduce the adverse effects of rain on the retrieved winds and that rain measurement can be performed from SeaWinds on QuikSCAT data. The wind/rain model may be especially useful in connection with the SeaWinds on ADEOS II instrument launched in November 2002. SeaWinds on ADEOS II includes a dual-pol radiometer that measures rain rate at a high accuracy. Synergistic use of the radiometer and scatterometer via a rain model may significantly improve the wind estimation process.

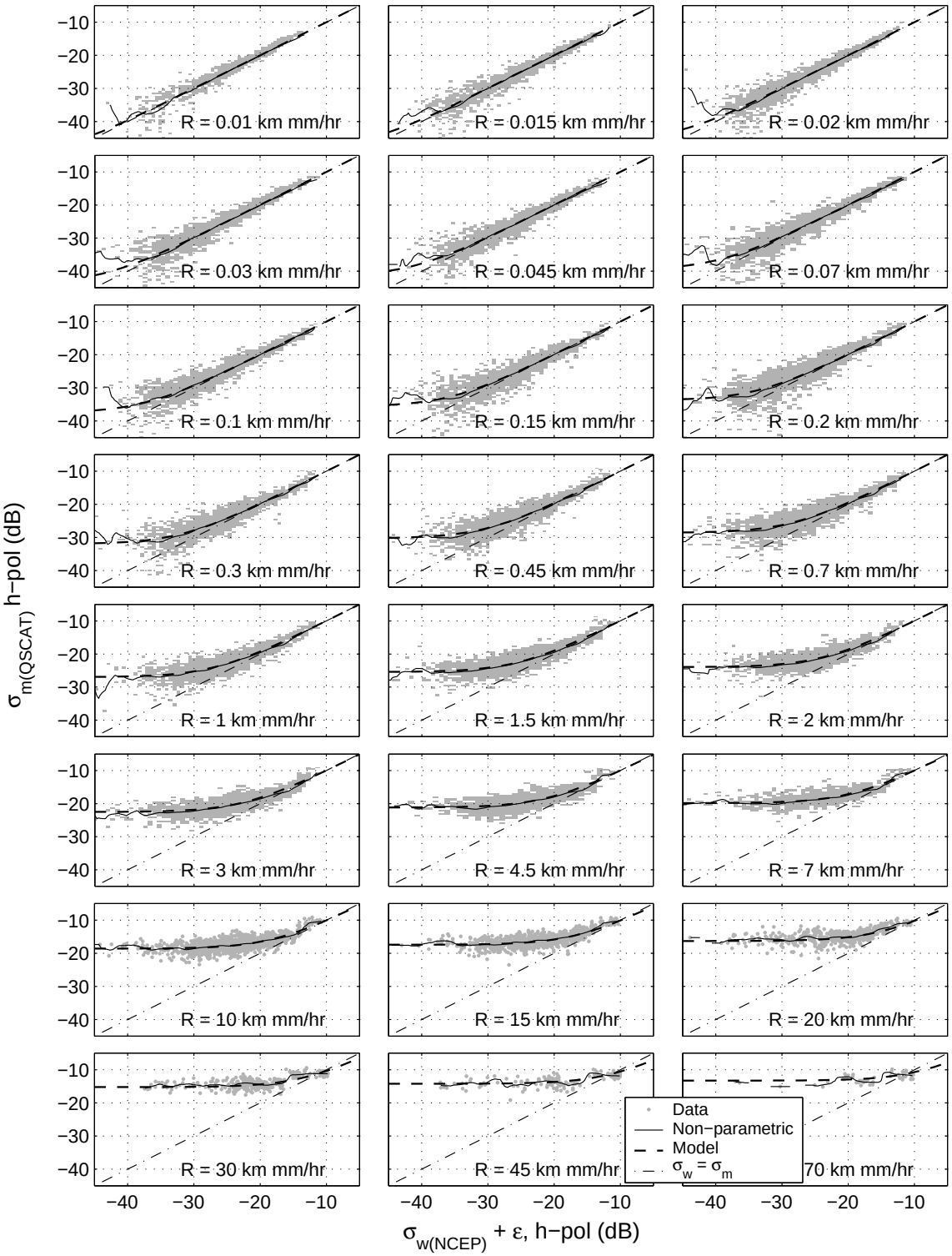


Figure 5.12: QuikSCAT measured backscatter σ_m plotted as a function of effective non-raining wind σ°, σ_w for h-pol. Also plotted is a non-parametric estimate of the data, and the combined rain effect model.

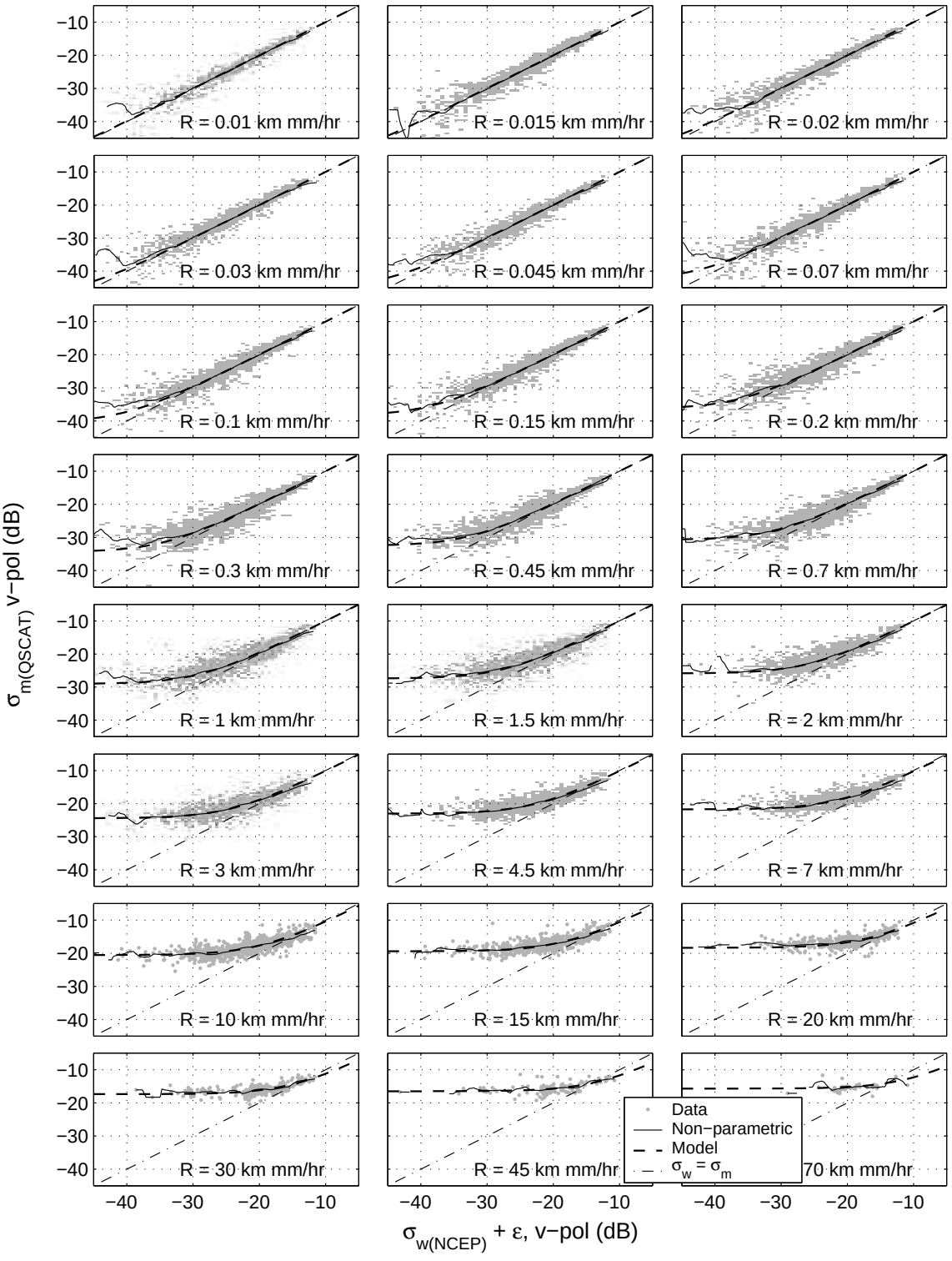


Figure 5.13: QuikSCAT measured backscatter σ_m plotted as a function of effective non-raining wind σ°, σ_w for v-pol. Also plotted is a non-parametric estimate of the data, and the combined rain effect model.

Chapter 6

Simultaneous Wind and Rain Retrieval Using SeaWinds Data

This chapter discusses a new method for improving wind estimates in the presence of rain using the simple rain-backscatter model developed in Chapter 5 [53]. The chapter focuses on two wind retrieval methods: simultaneous wind/rain estimation and rain-corrected wind retrieval.

First, in the absence of an independent estimate of the rain rate, the wind velocity and integrated rain rate is simultaneously retrieved using maximum likelihood estimation (MLE). This method is particularly useful for post-correction of SeaWinds on QuikSCAT data where co-located rain measurements are not available, and to provide an auxiliary earth-wide database of rain rates.

Second, when knowledge of the rain rate is available, such as from the Advanced Microwave Scanning Radiometer (AMSR) aboard ADEOS II, the GMF is directly corrected for the known rain rate. The wind is then retrieved using the adjusted GMF. The rain-corrected retrieval technique can be used to provide a more accurate wind estimate than simultaneous wind/rain retrieval by eliminating rain rate ambiguity.

The combined wind/rain MLE approach is described in Section 6.1. In conjunction with SeaWinds rain retrieval, beamfilling and wind vector cell filling issues are discussed in Section 6.2, along with a calculation of the normalized standard deviation of the rain backscatter model in Section 6.3. This chapter evaluates the wind retrieval skill for both simultaneous wind/rain and rain-corrected methods via simulation, comparing each method to the conventional wind-only retrieval in Section 6.4. In addition, validation studies are performed for the MLE approach with data from the Tropical Rainfall Measuring Mission (TRMM) precipitation radar (PR), and

winds from the National Centers for Environmental Prediction (NCEP). A synoptic example is given in Section 6.5, demonstrating the utility of simultaneous wind/rain retrieval. Simulation and validation demonstrate that the simultaneous wind/rain retrieval method works best in the swath sweet spots, and is not usable on the swath edges. As expected, the nadir region tends to be noisier than the sweet spots. The method works well for most wind and rain conditions, although some wind/rain states cause ambiguity between wind speed and rain rate, resulting in spurious rain rate estimates or abnormally low wind speeds. Such anomalies usually occur in extreme wind or rain conditions, and in connection with cross-swath blowing winds. In zero rain conditions, simultaneous wind/rain retrieval produces somewhat noisier results than the conventional method, especially at nadir. Further work is intended for quality control in such areas of degraded performance.

6.1 Methodology

The "wind-only" estimation process is described in Section 1.1.2 and assumes that the effects from unmodeled factors such as salinity, sea and air temperature, sea foam, and rain are small. The GMF variance term K_{pm} helps account for small perturbations due to these unknowns in the estimation process. Rain effects, however have been shown to be appreciable, and at times dominating [20, 53].

The effect of rain on σ° can be parameterized by the additional scattering and attenuation of the signal,

$$\sigma_m = \sigma_w \alpha_r + \sigma_e \tag{6.1}$$

where σ_m is the measured backscatter, σ_w is the component of the backscatter due to wind, α_r is the two-way atmospheric attenuation from falling rain, and σ_e is the effective rain backscatter due to surface perturbations and atmospheric scattering [20, 53, 54]. The parameters α_r and σ_e are assumed to be independent of azimuth angle and wind velocity, and thus solely a function of the vertically integrated rain rate in km mm/hr. Here, we adopt the empirically-derived quadratic log-log model for α_r and σ_e given in [53] for use with simultaneous wind/rain retrieval.

The simple rain-σ° model of Eq. (6.1) can be used in conjunction with the GMF to create a combined rain/wind model function of the form

$$\mathcal{M}_r(\mathbf{u}, R, ...) = \mathcal{M}(\mathbf{u}, ...)\alpha_r(R) + \sigma_e(R) \tag{6.2}$$

where R is the integrated rain rate and $\mathcal{M}_r$ is the combined wind/rain model function. When R is unknown, Eq. (6.2) can be used to simultaneously retrieve the wind and rain. If R is known for each measurement, Eq. (6.2) can be used to directly correct the model values in the wind estimation process. The measurement model for the combined wind/rain case is derived in Section 6.1.1, and the simultaneous wind/rain and rain-corrected wind retrieval methods are described in Sections 6.1.2 and 6.1.3 respectively.

6.1.1 Wind/rain measurement model

Applying an MLE technique to simultaneous wind/rain retrieval requires a measurement model for the signal and noise in the combined wind and rain signal. In the measurement model, the noise in the measurement is assumed to be white Gaussian like the non-raining case. Also, the communication noise coefficients α, β and γ used in the conventional wind-only case are assumed to not change under raining conditions.

We assume that the wind model uncertainty and uncertainty in the rain model are independent. Given the true wind and rain, the true backscatter σ_t can be written as

$$\begin{aligned}\sigma_t = \mathcal{M}(1 + \eta_1 K_{pm}) 10^{-\alpha_{dB}(1+\eta_2 K_{pa})/10} \\ +\sigma_e(1 + \eta_3 K_{pe})\end{aligned} \tag{6.3}$$

where η_1, η_2, and η_3 are zero-mean Gaussian random variables, K_{pa} is the normalized standard deviation of α_{dB} (the model α_r in decibels), and K_{pe} is the normalized standard deviation of model effective rain backscatter. The non-linear nature of the attenuation term introduces difficulty in calculating the overall variance of the model. Assuming that the attenuation is not very large, the attenuation term can be

simplified by truncating the Taylor's series expansion of 10^x about $x = 0$, yielding

$$\begin{aligned} 10^{-\alpha_{dB}(1+\eta_2 K_{pa})/10} &= \alpha_r 10^{(-\eta_2 K_{pa}/10)} \\ &\cong \alpha_r(1 - \ln(\alpha_r) K_{pa}\eta_2). \end{aligned} \tag{6.4}$$

Further, by truncating the Taylor's series expansion of $\ln(x)$ about $x = 1$, we arrive at

$$10^{-\alpha_{dB}(1+\eta_2 K_{pa})/10} \cong \alpha_r(1 + (1 - \alpha_r) K_{pa}\eta_2). \tag{6.5}$$

Using this approximation, the expected value of σ_t is easily calculated, giving the original model equation,

$$\mathrm{E}\{\sigma_t\} = \mathcal{M}\alpha_r + \sigma_e = \mathcal{M}_r. \tag{6.6}$$

The variance of σ_t is also calculated using the approximation of Eq. (6.5), yielding

$$\begin{aligned} \mathrm{Var}\{\sigma_t\} &\cong \mathcal{M}^2\alpha_r^2[K_{pm}^2 + K_{pa}^2(1 - \alpha_r)^2(1 + K_{pm}^2)] \\ &\quad +\sigma_e^2 K_{pe}^2. \end{aligned} \tag{6.7}$$

For low to moderate rain rates, α_r is nearly unity, making the K_{pa} term of Eq. (6.7) negligible. At high rain rates α_r is small, and the K_{pe} term dominates. Thus, the variability due to attenuation can be ignored, reducing the variance to

$$\mathrm{Var}\{\sigma_t\} \cong \mathcal{M}^2\alpha_r^2 K_{pm}^2 + \sigma_e^2 K_{pe}^2. \tag{6.8}$$

Figure 6.1 supports the argument that the variability of the attenuation is negligible in wind retrieval.

Adding communication noise, the scatterometer measurement z is modeled as

$$\begin{aligned} z &= \sigma_t(1 + K_{pc}\eta_4) \\ &\cong [\mathcal{M}\alpha_r(1 + \eta_1 K_{pm}) \\ &\quad +\sigma_e(1 + \eta_3 K_{pe})](1 + K_{pc}\eta_4). \end{aligned} \tag{6.9}$$

Using this model, the expected value of z is

$$\mathrm{E}\{z\} = \mathcal{M}\alpha_r + \sigma_e = \mathcal{M}_r \tag{6.10}$$

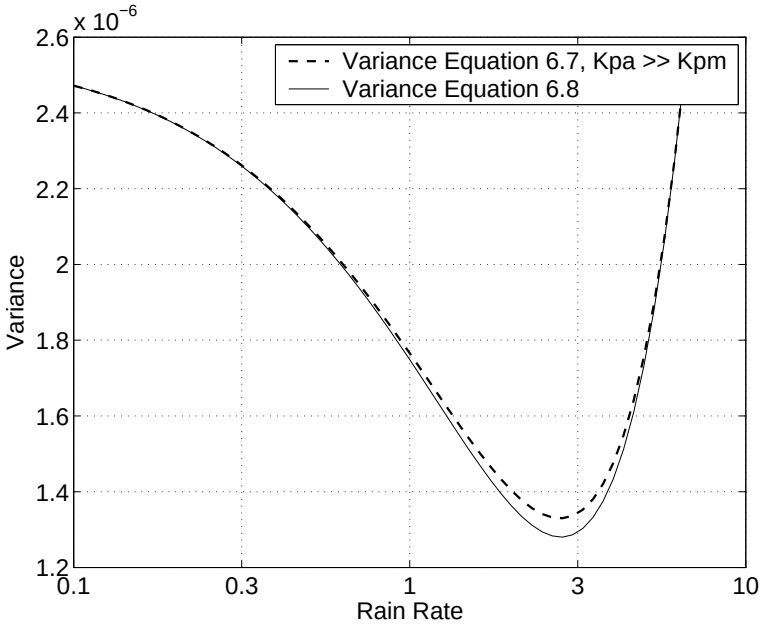


Figure 6.1: Variance of Eq. 6.7 for K_{pa} very large and of Eq. 6.7 ($K_{pa} = 0$) as a function of rain rate, demonstrating that even at large K_{pa}, the overall variance is not very sensitive to K_{pa}, and thus the variability of the attenuation can be ignored.

and the variance is

$$\begin{aligned} \mathrm{Var}\{z\} &= (\mathcal{M}^2\alpha_r^2 K_{pm}^2 + \sigma_e^2 K_{pe}^2)(1 + K_{pc}^2) \\ &\quad + K_{pc}^2(\mathcal{M}\alpha_r + \sigma_e)^2 \end{aligned} \tag{6.11}$$

(see Appendix H). As in the non-raining derivation of the variance, we replace σ_t with its mean from Eq. (6.6). Using K_{pc} from Eq. (1.6), and making the assumption (similar to Eq. (1.9))

$$1 + \alpha \gg \frac{\beta}{\sigma_t} + \frac{\gamma}{\sigma_t^2}, \tag{6.12}$$

Eq. (6.11) becomes

$$\begin{aligned} \mathrm{Var}\{z\} &= (\mathcal{M}^2\alpha_r^2 K_{pm}^2 + \sigma_e^2 K_{pe}^2)(1 + \alpha) \\ &\quad + \alpha\mathcal{M}_r^2 + \beta\mathcal{M}_r + \gamma. \end{aligned} \tag{6.13}$$

For zero rain conditions, $\alpha_r \to 1$, $\sigma_e \to 0$, and $\mathcal{M}_r \to \mathcal{M}$, reducing Eq. (6.13) to the non-raining variance of Eq. (1.10).

To further simplify the variance, the first term on the right hand side of Eq. (6.13) is manipulated by completing the square to yield,

$$\begin{aligned} \mathrm{Var}\{z\} &= [(\mathcal{M}\alpha_r K_{pm} + \sigma_e K_{pe})^2 \\ &\quad -2K_{pm}K_{pe}\mathcal{M}\alpha_r\sigma_e](1+\alpha) \\ &\quad +\alpha\mathcal{M}_r^2 + \beta\mathcal{M}_r + \gamma. \end{aligned} \tag{6.14}$$

Given that the total model backscatter $\mathcal{M}_r$ is a constant, the negative cross term in Eq. (6.14) is parabolic in σ_e, of the form

$$-2K_{pm}K_{pe}(\mathcal{M}_r - \sigma_e)\sigma_e \tag{6.15}$$

which has a minimum of $-.5K_{pm}K_{pe}$ at $\sigma_e = \mathcal{M}_r/2$. Although this negative cross term is not negligible, the parabolic nature has an adverse effect on the MLE by lowering the variance of potential estimates where the raining backscatter is on the order of half the total backscatter. The parabolic nature also raises the likelihood function, artificially discouraging estimation of rain rates in this regime. In order to eliminate this problem, we present a reduced version of the variance which eliminates the negative cross term,

$$\begin{aligned} \mathrm{Var}\{z\} &\cong (\mathcal{M}\alpha_r K_{pm} + \sigma_e K_{pe})^2(1+\alpha) \\ &\quad +\alpha\mathcal{M}_r^2 + \beta\mathcal{M}_r + \gamma. \end{aligned} \tag{6.16}$$

To validate this simplification, wind retrieval skill using both Eq. (6.13) and Eq. (6.16) is compared in Section 6.3, demonstrating that Eq. (6.16) yields better wind retrieval. It is interesting to note that if $K_{pe} = K_{pm}$, the form of Eq. (6.16) reduces to the form of the non-raining variance as intuition suggests,

$$\mathrm{Var}\{z\} \cong [(1+\alpha)K_{pm}^2 + \alpha]\mathcal{M}_r^2 + \beta\mathcal{M}_r + \gamma. \tag{6.17}$$

6.1.2 Simultaneous wind/rain retrieval

The approximate MLE likelihood function for simultaneous wind/rain retrieval is written as

$$l_r(z|\mathbf{u}, R) = \sum_k \frac{(z_k - \mathcal{M}_{rk}(\mathbf{u}, R, ...))^2}{\varsigma_{rk}(\mathbf{u}, R)^2} \tag{6.18}$$

where $\varsigma_{rk}(\mathbf{u}, R)$ is the variance derived in the previous section. Simultaneous wind and rain estimates are found by minimizing the likelihood function for $\mathbf{u}$ and R given the backscatter measurements.

As in the non-raining case, the likelihood function has several local minima corresponding to possible ambiguities (see Figure 6.2). Each ambiguity has a corresponding wind speed, wind direction and rain rate. To yield a unique wind vector field, ambiguity selection is performed. In order to follow the conventional wind-only retrieval method as much as possible, a nudging/median filtering ambiguity selection scheme is implemented. For simplicity, nudging for the simultaneous rain/wind retrieval is not thresholded as with the JPL product [14]; all ambiguities are used. Median filtering is performed using the modified vector-median filter described in [17].

6.1.3 Rain-corrected wind retrieval

In the case of known rain rate, such as from the AMSR on ADEOS II, simultaneous wind/rain retrieval is simplified by evaluating the MLE likelihood function of Eq. (6.18) at the "true" rain rate only. This technique is known as rain-corrected wind retrieval and is equivalent to a Bayesian estimation technique using a delta-distributed rain rate prior. In the case of synergistic use of AMSR radiometer rain rates from ADEOS II, a realistic Bayesian prior given the AMSR rain rates may also be developed. Simultaneous wind/rain retrieval and rain-corrected retrieval are evaluated in Section 6.4.

6.2 Beam/ WVC filling

Because of the relatively high spatial variability of rain, beamfilling and WVC-filling effects are significant factors in interpreting SeaWinds-retrieved rain rates. The retrieved rain rate corresponding to the selected local minimum of the likelihood function is an irregularly-weighted average over an area larger than the 25 $\times$ 25 km WVC. For each measurement, the antenna beam response function weights the backscatter values received from the atmospheric and surface rain scatterers. Thus, depending on the spatial distribution of the rain within the footprint, the backscatter response is altered from the true unweighted average. This is known as the beamfilling

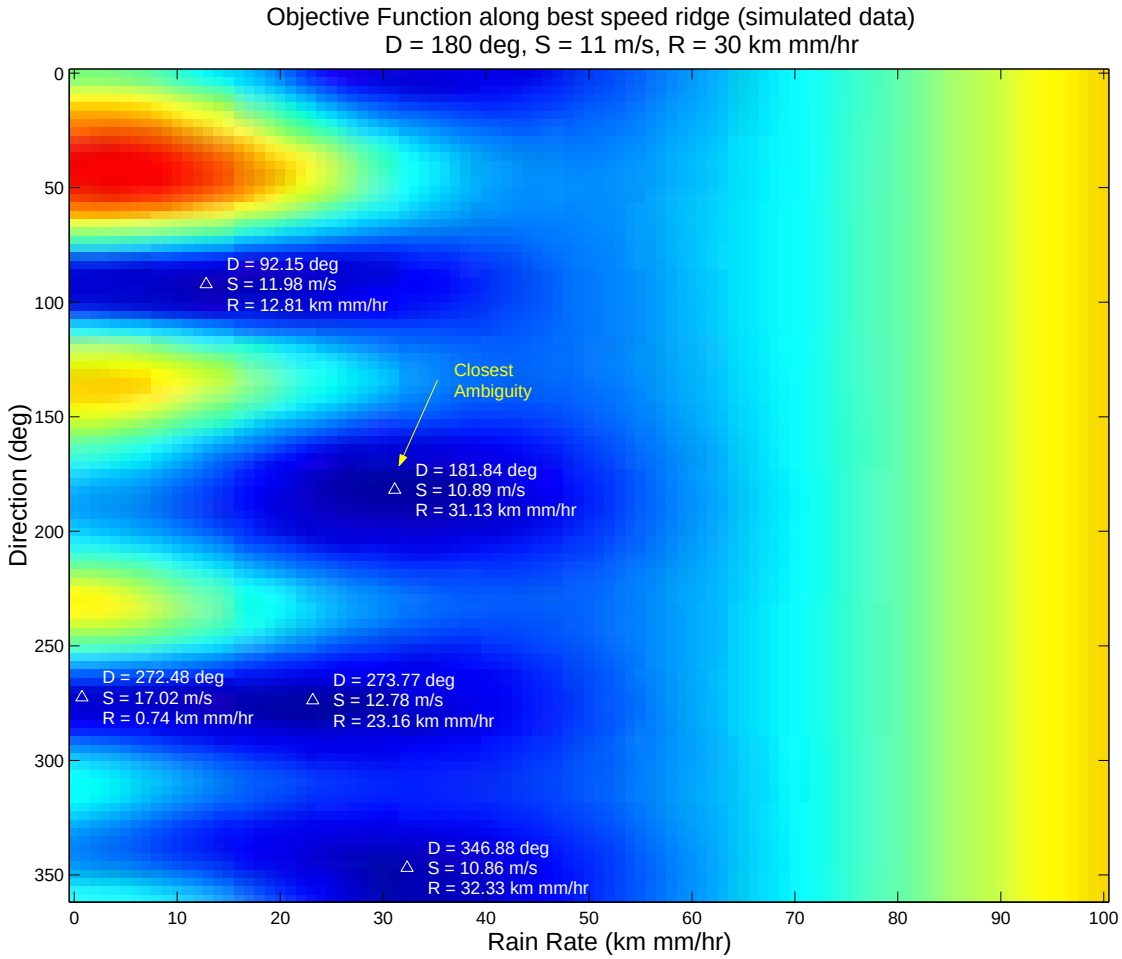


Figure 6.2: Sample likelihood function, displayed as a function of wind direction and rain rate. For each pixel, the wind speed corresponding to the likelihood minimum for that wind direction and rain rate is chosen.

effect. Additionally, the spatial layout of the σ° measurements within a WVC, with the associated overlap and possible lack of coverage, yields an irregularly-weighted rain rate estimate. This is the WVC-filling effect.

To evaluate the beamfilling and WVC-filling effects, we use co-located data from the TRMM PR. The TRMM/SeaWinds data include 100 regions co-registered temporally within 10 minutes. The co-located data span three months from September through November 1999 and are limited to the tropics. The TRMM PR obtains measurements at a much higher resolution than the SeaWinds scatterometer (about 4 km), affording an excellent data set for analysis of beamfilling and WVC-filling effects.

First, we examine the beamfilling effect by evaluating the error between the average rain rate weighted by the antenna beam pattern and the non-weighted average rain for individual measurements. Using the TRMM/SeaWinds co-located data set, both antenna-weighted (R_w) and non-weighted (R_{nw}) rain rate estimates for each SeaWinds measurement are computed from the PR-derived integrated rain rates via the equations

$$R_w = \frac{\sum_{i=1}^{N} G_i R'_{i(PR)}}{\sum_{i=1}^{N} G_i} \tag{6.19}$$

and

$$R_{nw} = \frac{1}{N} \sum_{i=1}^{N} R'_{i(PR)} \tag{6.20}$$

where G_i is the gain of the SeaWinds antenna pattern at each co-located TRMM PR measurement within the 6dB SeaWinds footprint, $R'_{i(PR)}$ is the PR integrated rain rate at each 4 km PR resolution cell, and N is the number of PR measurements within the SeaWinds footprint. The statistics of the normalized error between the antenna-weighted and non-weighted measurements with significant non-weighted rain rates ($>$ 2 km mm/hr) are calculated for the entire co-located data set. The normalized error is defined as the error divided by the non-weighted measurements, i.e. $(R_w - R_{nw})/R_{nw}$. The mean normalized error is approximately zero, while the standard deviation is 0.21, suggesting that while beamfilling introduces variance into the rain estimate, it does not shift the mean.

Next, we examine the WVC-filling effect by evaluating the error between the "effective" weighted average rain rate seen by SeaWinds and the non-weighted rain rate averaged over each 25 × 25 km WVC. The weighted average rain rate R_w^{wvc} is calculated by first computing the PR-derived antenna-weighted rain rate for each measurement in the co-located TRMM/SeaWinds data set (see Eq. (6.19)), and then averaging the antenna-weighted rain rates corresponding to each WVC. An example 25 × 25 km WVC, the 6 dB contours of individual measurements comprising the WVC, and co-located 4-km PR rain rates are displayed in Figure 6.3. To facilitate a comparison, WVC-average non-weighted rain rate estimates R_{nw}^{wvc} are obtained by averaging all raw 4 km PR rain rates lying within each 25 × 25 km square WVC. Figure 6.3 illustrates the large area covered by the measurements and the high variability of rain within the view of the large SeaWinds footprints.

A scatter plot comparing the WVC-average rain rates to the weighted average rain rates for the co-located TRMM/SeaWinds dataset is displayed in Figure 6.4. The statistics of the normalized error are calculated for all measurements with WVC-average rain rates greater than 2 km mm/hr. The mean of the normalized error for the co-located dataset is about -.04, suggesting a slight negative bias of the weighted-average estimates. The standard deviation is 0.39, indicating that the variability due to the WVC-filling effect is significant. Thus, it is important to interpret SeaWinds-derived rain rates as an irregularly-weighted average of the rain, rather than an unweighted average of rain over the 25 × 25 km WVC.

6.3 Estimation of K_{pe}

In order to apply the MLE method to simultaneous wind/rain estimation, the normalized standard deviation of the effective rain backscatter K_{pe} must be determined. Several factors may contribute to K_{pe} including uncertainty in the model and inter-measurement variability due to non-uniform rain. We present estimated K_{pe} values due to both of these factors in Section 6.3.1. Also, we present an empirical approach to estimating K_{pe} by finding the value of K_{pe} that yields the most accurate

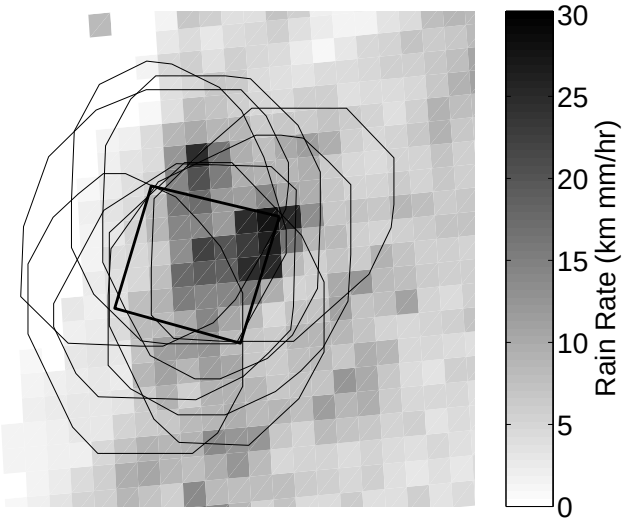


Figure 6.3: The geometry of a sample WVC (bold square) and the 6 dB gain contours of the individual σ° measurements (ellipses) comprising the WVC. A PR-derived rain map is shown in the background.

wind speeds in Section 6.3.2. The empirical K_{pe} from the rain detection analysis is used in the remainder of the paper.

6.3.1 K_{pe} due to model uncertainty and non-uniform rain

We first examine the rain backscatter model uncertainty. Validation of the wind/rain backscatter model (see Eq. (6.1)) is given in [53], where the normalized standard deviation due to uncertainty in the model is determined to be 1.6 dB, corresponding to a K_{pe} value of 0.45. This, however, may be an overestimate, augmented by the inherent variability of the NCEP winds and TRMM PR rain rates used to validate the model. Thus, we expect the actual variability to be somewhat lower than this estimated value.

Second, we examine the inter-measurement variability of the rain backscatter within a WVC due to rain non-uniformity. The contribution to K_{pe} from inter-measurement variability from rain non-uniformity is estimated using the co-located QuikSCAT/TRMM data by comparing the model backscatter computed from the

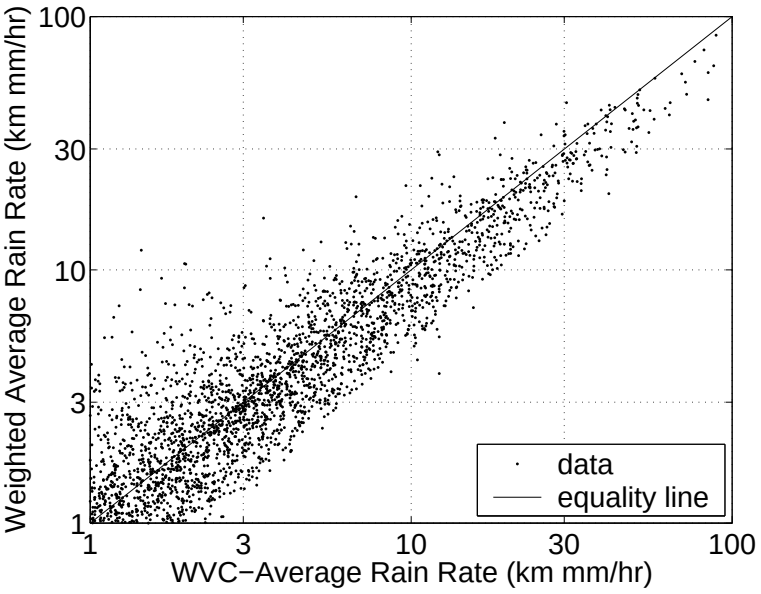


Figure 6.4: The WVC-average (unweighted) rain rate versus the weighted average rain rate seen by SeaWinds.

weighted-average rain rate of the entire WVC to the model backscatter of the rain corresponding to individual measurements. The model estimate of the backscatter for the entire WVC, σ_e^{wvc}, is computed by projecting the weighted WVC-average rain rate R_w^{wvc} through the rain model. Likewise, for all QuikSCAT measurements in the WVC, the model backscatter corresponding each measurement σ_e is computed by forward projection of the antenna-weighted rain rate R_w through the rain model. An estimate of K_{pe} is obtained by taking the standard deviation of the normalized error,

$$K_{pe} = \text{std}\{(\sigma_e - \sigma_e^{wvc})/\sigma_e^{wvc}\} \tag{6.21}$$

for all rain rate observations where the weighted average rain rate is greater than 2 km mm/hr. The resulting value is $K_{pe} \cong 0.39$, indicating that inter-measurement variability is important and on the order of the model uncertainty.

6.3.2 K_{pe} yielding lowest RMS wind error

Since the value of K_{pe} affects the likelihood function, which in turn influences simultaneous wind/rain retrieval, an alternate approach to estimating K_{pe} is to

Table 6.1: K_{pe} value yielding the lowest RMS wind speed error when compared with calibrated NCEP numerical weather prediction winds over the TRMM/QuikSCAT co-located dataset. Also, the corresponding RMS errors for both variance equations of Eq. (6.13) and Eq. (6.16).

Variance form	K_{pe}	RMS error
Eq. (6.13)	0.18	2.27 m/s
Eq. (6.16)	0.16	2.14 m/s

find the value of K_{pe} that yields the best wind speed retrieval in real SeaWinds data. Here, we compare the retrieved wind speed from simultaneous wind/rain retrieval to NCEP wind speeds over the QuikSCAT/TRMM/NCEP co-located dataset.

In comparing QuikSCAT to NCEP, we note that a bias exists between non-raining QuikSCAT data and NCEP winds [8, 53]. We adjust for the bias with a multiplicative constant determined using least-squares linear estimation over the non-raining QuikSCAT winds (as determined from the TRMM PR rain rate). The bias constant is determined to be $u_{qscat} = 0.83 \times u_{ncep}$. The remaining discussion uses NCEP wind speeds adjusted for the bias.

To find the value of K_{pe} that gives the best speed wind retrieval, we perform simultaneous wind and rain retrieval over the co-located TRMM/SeaWinds data set for varying values of K_{pe} and for both likelihood function variance equations (Eq. (6.13) and Eq. (6.16)). Then, we choose the value of K_{pe} and corresponding variance equation that yields the lowest RMS wind speed error overall. We note that we use the JPL default value of $K_{pm} = 0.16$ for the model function variability in the retrievals. The optimal values of K_{pe} for each variance equation and the RMS error are given in Table 6.1.

The optimal value of K_{pe} is 0.16 corresponding to the reduced variance form of Eq. (6.16). Since this value of K_{pe} is the same as the K_{pm} value used, the variance reduces to the non-raining variance equation. Thus, for the remaining analysis, we utilize the non-raining variance equation (see Eq. 6.17) in the retrievals with $K_{pm} = 0.16$.

6.4 Simulations and validation

In order to evaluate the quality of the new wind/rain estimation procedure, we present a simulation and validation study for simultaneous wind/rain retrieval and rain-corrected wind retrieval. For SeaWinds baseline wind-only retrieval, it is known that for some wind speeds and cross track positions, the wind retrieval performance of SeaWinds is somewhat degraded [15, 36]. This degradation often occurs at low and extremely high winds, at nadir, and on the swath edges. At low wind speeds, low SNR often causes wind estimates to be noisy. At high wind speeds, a saturation in σ° occurs, decreasing the accuracy of the winds [19, 55]. On the swath edges and at nadir, poor viewing geometry causes the MLE to be ill-conditioned. However, at moderate wind speeds, and especially in the "sweet spots" of the swath, wind retrieval performance is very good.

In the absence of rain, the inclusion of a rain rate parameter into the estimation process inherently makes the MLE more ill-conditioned than wind-only retrieval. However, when rain is present, simultaneous wind/rain retrieval can significantly improve the wind estimate. It is thus important to evaluate the performance of the wind/rain MLE procedure with and without rain. Also, because the wind retrieval accuracy varies with cross-track position and wind velocity, we evaluate the performance given a variety of cross track and wind conditions. The MLE is evaluated via simulation in Section 6.4.1 for both simultaneous wind/rain and rain corrected retrievals. Validation of simultaneous wind/rain retrieval with NCEP winds and PR rain rates is given in Section 6.4.2.

6.4.1 Simulation results

To analyze the performance of the MLE, we perform simulations of the backscatter return for various conditions and evaluate the statistics of the retrieved wind and rain. Simulations are conducted for varying wind speeds, rain rates, wind directions, and cross track positions, spanning a wide range of the parameter space (see Table 6.2). Nominal values of the K_{pc} coefficients α, β, and γ are used with typical measurement geometries at each WVC.

Table 6.2: Delineations of wind speed, wind direction, rain rate, and cross track position for which the simulations are performed.

Speed	3, 7, 11, 15, 20, 25 m/s
Direction	0°, 15°, 30°, ..., 345°
Rain Rate	0, 0.3, 1, 3, 10, 30 km mm/hr
Cross Track Position	WVC 2, 6, ... 38; 39, 43, ... 75

For each combination of conditions, we project the speed, direction, and rain rate through the backscatter model (Eq. (6.1)) for all measurements corresponding to that WVC. Next, zero-mean Gaussian random noise with the variance given in Eq. (6.17) is added. Retrieval is then performed for 500 noise realizations for each set of conditions. The wind vector ambiguity realization that is closest to the true wind vector is selected.

For each simulation, three retrievals are performed: conventional wind-only retrieval, simultaneous wind/rain retrieval, and rain-corrected wind retrieval. In wind-only retrieval, the baseline wind-only likelihood function of Eq. (1.11) is used. In simultaneous wind/rain retrieval, ambiguities are determined as the local minima of the wind/rain likelihood function of Eq. (6.18). In rain-corrected retrieval, the simultaneous rain/wind likelihood function (Eq. 6.18) is evaluated at the true rain rate only, requiring knowledge of the true rain rate available from simulation. Full simulation statistics for simultaneous wind/rain retrieval are given in Figures 6.17–6.22 at the end of the chapter.

In presenting the simulation results, we first examine the RMS error of the ambiguity closest to the true wind as a function of cross track position and "rain fraction." The rain fraction F is defined as the effective rain backscatter divided by the total model backscatter given the ambiguity selected rain rate and vector wind averaged over the measurements,

$$F = \sum_k \sigma_{ek}(R)/\mathcal{M}_{rk}(R, \mathbf{u}, ...). \tag{6.22}$$

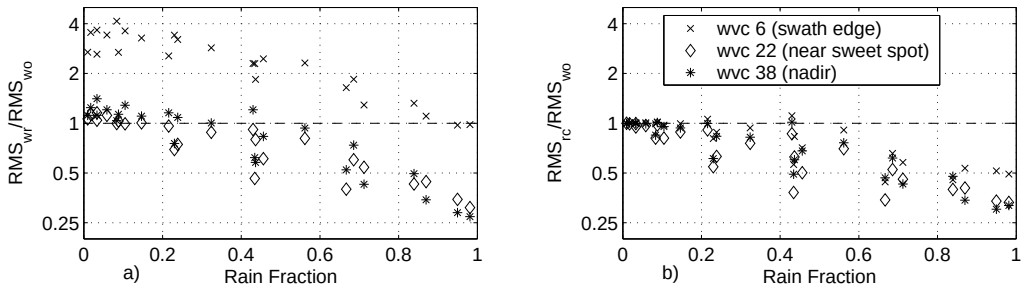


Figure 6.5: Ratio in RMS error between the a) simultaneous wind/rain retrieval RMS_{wr} and the wind-only retrieval RMS_{wo}, and b) the rain-corrected wind retrieval RMS_{rc} and the wind-only retrieval RMS_{wo} as a function of rain fraction.

The rain fraction F indicates the level to which rain affects the backscatter measurements, with zero meaning rain has no significant effects and one meaning rain dominates the observed backscatter.

The wide scope of wind speeds and rain rates provides a large range of rain fraction bins, each corresponding to a different wind/rain combination. To allow compact comparison of wind-only retrieval to simultaneous wind/rain and rain-corrected retrievals, the ratio of RMS error between each is shown in Figure 6.5 for each cross-track position.

Figure 6.5a demonstrates that simultaneous wind/rain retrieval on the swath edge performs poorly as indicated by a high error ratio. Also from Figure 6.5a, simultaneous wind/rain retrieval is less accurate than wind-only retrieval for zero to low rain fraction data (corresponding to relatively low rain rates). However, for most rain fractions above 0.2, simultaneous rain/wind retrieval has a lower RMS error, especially for sweet-spot observations. These simulations suggest that simultaneous wind/rain retrieval works well for most rain corrupted cases, while it slightly degrades wind retrieval performance in zero-rain conditions.

Comparing Figure 6.5b to Figure 6.5a, the rain-corrected wind retrieval performs better than the wind-only retrieval for almost all cases except at high rain fractions, where the wind is almost totally dominated by rain. The most noticeable improvement over simultaneous wind/rain retrieval is on the swath edges, where

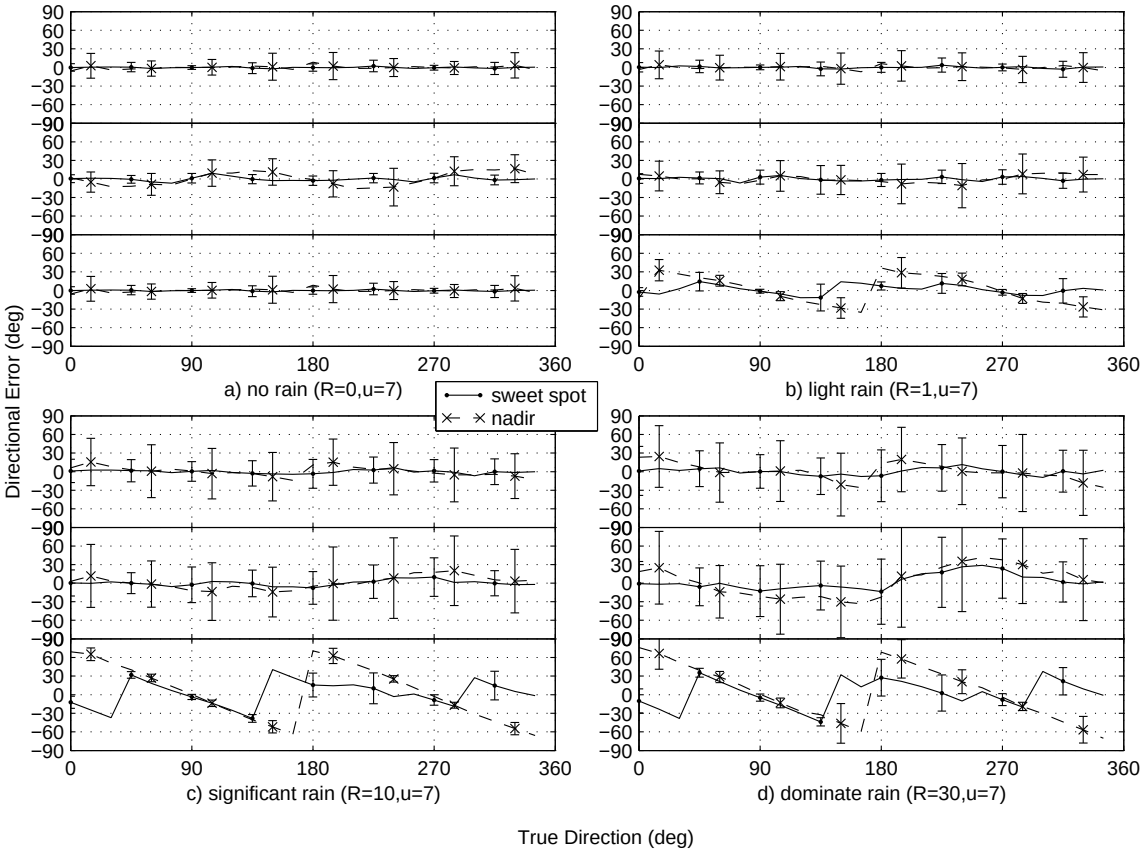


Figure 6.6: Directional error statistics as a function of true direction for 4 rain rate cases with wind speed of 7 m/s. In each plot, the rain-corrected wind retrieval is shown on top, simultaneous wind/rain retrieval is in the middle, and wind-only retrieval is on the bottom.

simultaneous wind/rain retrieval fails. Also, as expected, the rain-corrected wind retrieval performs much the same as wind-only retrieval in zero and low rain cases.

Next, we examine the directional and speed error of the two retrieval methods, comparing them to the wind-only result. The high number of wind/rain/direction/cross-track combinations prohibits displaying all the cases in this paper. Thus, we only show several representative examples. Since 7 m/s is the mean wind speed over the oceans, we show four typical cases with wind speed of 7 m/s and varying rain rates. For clarity, we show nadir and sweet spot simulations only. Figures 6.6 and 6.7 show the directional and speed error statistics for

rain rates of 0, 1, 10 and 30 km mm/hr for rain-corrected retrieval, simultaneous wind/rain retrieval, and wind-only retrieval. For the zero-rain case, the simultaneous wind/rain retrieval of directions is somewhat less accurate than the other two retrieval methods. However, for light rain, simultaneous wind/rain retrieval is more accurate, while wind-only retrieval becomes somewhat biased. For significant rain, the rain-corrected and simultaneous wind/rain retrieval are very close to zero-mean, while the wind-only retrieval is extremely biased in certain directions. The bias in the wind-only data exists because the wind-only ambiguities tend to point in a direction parallel or perpendicular to the swath regardless of the true direction. For the dominate rain case, all three retrievals perform poorly with respect to wind direction as expected; however, the rain-corrected and simultaneous wind/rain retrievals are less biased. Figures 6.6 and 6.7 also demonstrate that simultaneous wind/rain directional retrieval has a lower variance in the sweet spot than in the nadir region. Thus, directional retrieval in the sweet spot is somewhat better than at nadir as expected.

Examining the speed error statistics of Figure 6.7, the simultaneous wind/rain retrieved wind speeds are nearly zero mean for all rain rates with increasing variability at higher rain rates, while the wind speeds for the wind-only retrieval are extremely biased at high rain rates. At higher rain rates, the simultaneous wind/rain retrieved speed is biased slightly low for directions nearing 90° and 270°, especially at nadir. The bias is likely due to cross-track pointing winds appearing to the MLE as rain, and thus decreasing the wind speed while increasing the rain rate. This bias suggests identifiability problems between the wind and rain for cross-track blowing winds.

The speed error for the three retrieval methods is further demonstrated in Figure 6.8. Here, all simulation wind speed and rain rate combinations are shown. As expected, the wind-only retrieval shows considerable biases at low wind speeds for moderate to high rain rates. These biases are almost completely corrected in both rain-corrected and simultaneous wind/rain retrieval which both exhibit a near-zero mean for almost all cases. The rain-corrected retrieval is slightly biased high for very

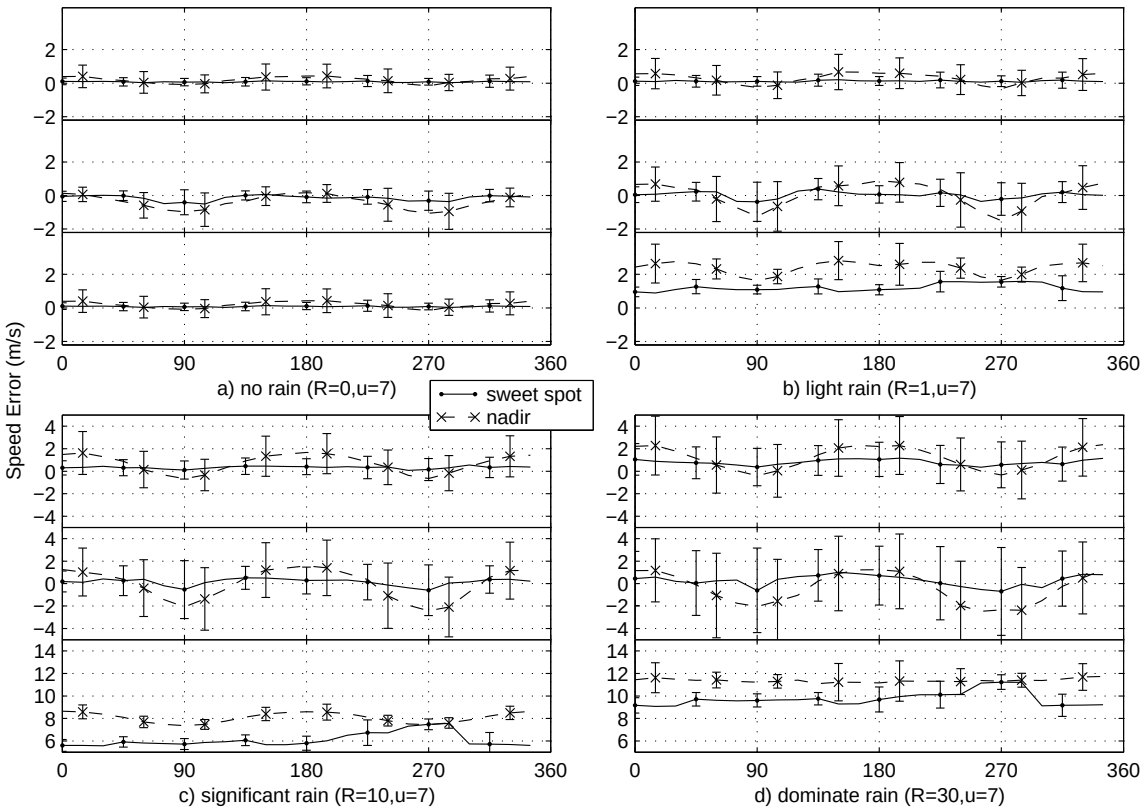


Figure 6.7: Wind speed error statistics as a function of true direction for 4 rain rate cases with wind speed of 7 m/s. In each plot, the rain-corrected wind retrieval is shown on top, simultaneous wind/rain retrieval is in the middle, and wind-only retrieval is on the bottom. Notice the scale differences of the plots.

low rain rates and high wind speeds. Likewise, the simultaneous wind/rain retrieved speeds are slightly biased high for most rain rate/wind speed combinations.

Next, we demonstrate the rain retrieval performance of the simultaneous wind/rain MLE. Figure 6.9 shows the rain rate error statistics of the retrieved rain rates as a function of true rain rate for varying wind speeds. As wind speed increases, the retrieved rain rate becomes increasingly biased. The rain rate bias at high wind speeds is quite high, even at zero rain rate. However, in low to moderate wind speeds (3-11 m/s), the retrieval performs quite well. These simulations demonstrate

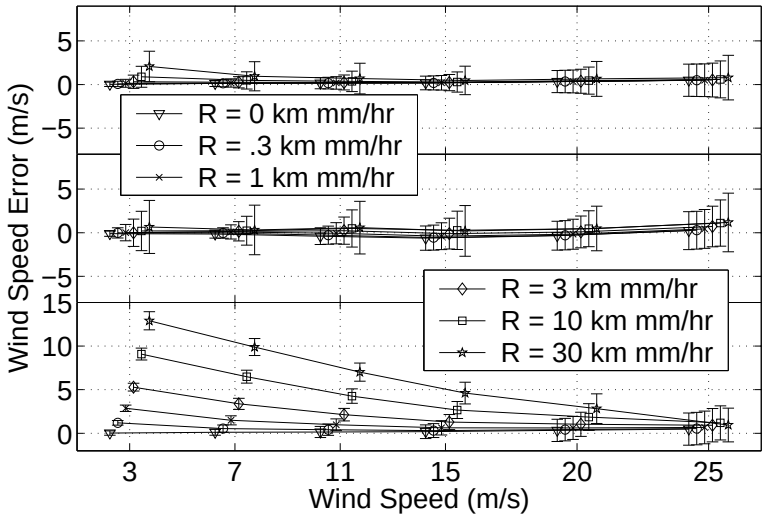


Figure 6.8: Wind speed error statistics as a function of wind speed for various rain rates. The rain-corrected wind retrieval is shown on top, simultaneous wind/rain retrieval is in the middle, and wind-only retrieval is on the bottom. Each point is horizontally offset for clarity.

the limitation of accurately retrieving rain in high wind speed regions where the wind dominates the rain signal. Rain flagging algorithms can be developed to detect erroneous rain rates in high wind speed regions and discard them.

In summary, simultaneous wind and rain retrieval works well for many conditions, especially in the sweet spot and for moderate wind/rain conditions. Nadir-retrieved winds from the simultaneous method tend to be especially noisier than winds retrieved in the sweet spots, and simultaneous wind/rain retrieval is not usable on the swath edges. In the absence of rain, simultaneous wind/rain retrieval is less accurate than wind-only retrieval, and thus it is beneficial to use wind-only retrieval in non-raining regions. The wind speeds of simultaneous wind/rain retrieval are mostly unbiased; however, cross-track blowing winds may cause lower than expected retrieved wind speeds due to increased ambiguity between the wind and rain effects in such conditions. Also, in extreme wind, the retrieved the rain rate tends to be biased high, while in extreme rain, the retrieved wind vectors tend to be highly variable.

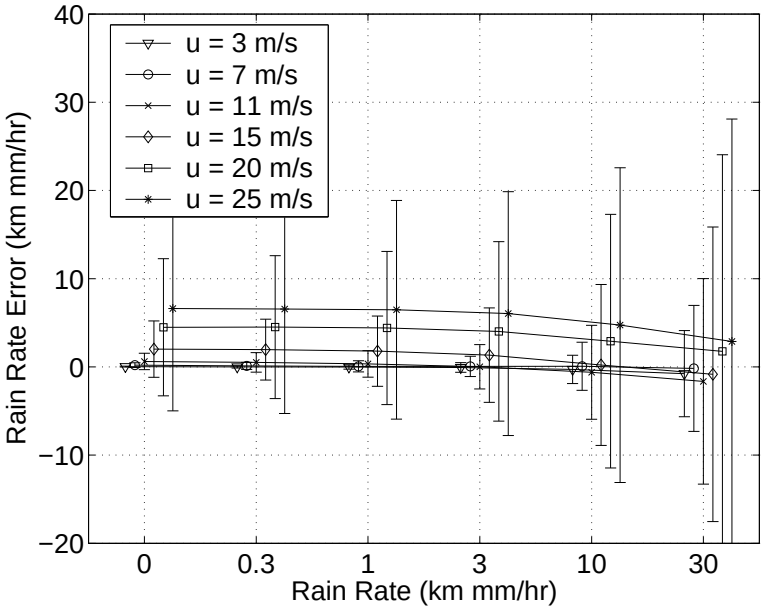


Figure 6.9: Rain rate error statistics as a function of rain rate for several different wind speeds. Each point is horizontally offset for clarity.

6.4.2 Validation

In this section, we present a validation of QuikSCAT simultaneous wind/rain retrieval with NCEP wind speeds/directions in Section 6.4.2, and with TRMM PR rain rates in Section 6.4.2. The analysis of rain-corrected wind retrieval is omitted since it requires some tuning to use in connection with PR rain rates and is thus beyond the scope of this paper. Also, as of yet, AMSR rain rate estimates are not available to use in conjunction with rain-corrected wind retrieval.

Wind vector validation

First, we perform validation of simultaneous wind/rain retrieval with NCEP wind speeds and directions. The validation is performed over the co-located QuikSCAT/TRMM dataset since an estimate of the rain rate is available. Again, we correct for the NCEP bias with a multiplicative constant of 0.83.

A scatter plot displaying the simultaneous wind/rain retrieved and wind-only retrieved wind speed as a function of NCEP wind speed for several rain fraction

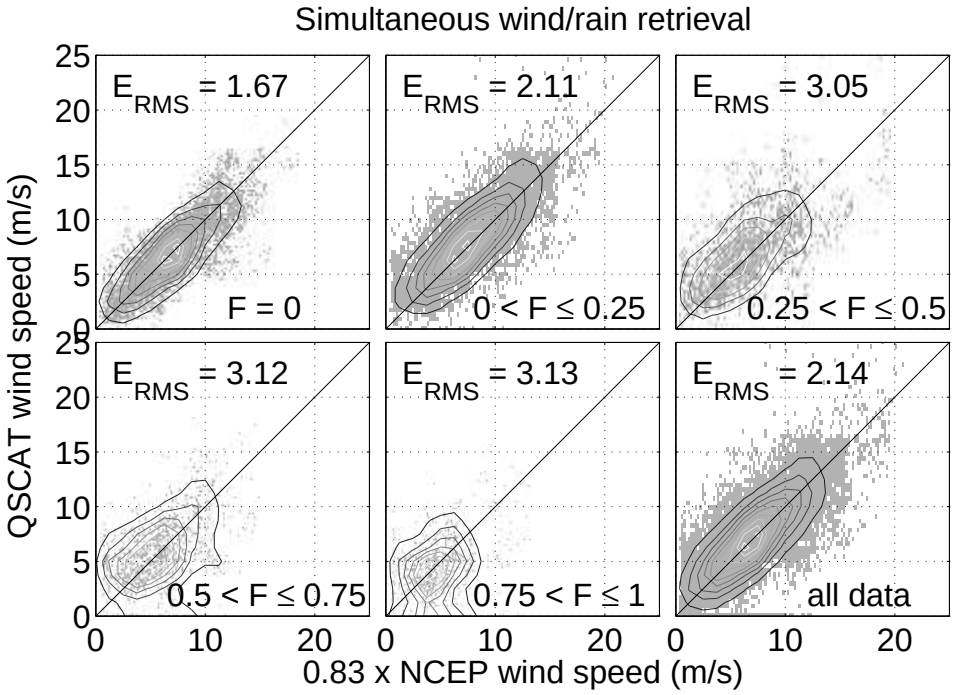


Figure 6.10: Scatter plot with density curves of the bias-corrected NCEP wind speeds versus simultaneous wind/rain retrieved wind speeds for various rain fraction bins.

bins is shown in Figures 6.10 and 6.11. Simultaneous wind/rain retrieval has a slightly higher error than the wind-only retrieval for low rain fractions. However, for increasing rain fractions, the simultaneous wind/rain retrieval remains essentially unbiased, while the wind-only retrieval becomes increasingly biased.

At higher rain fractions, many of the wind speed estimates from simultaneous wind/rain retrieval are driven to zero. A "zero" wind speed indicates that either the rain is sufficiently strong so that it totally dominates the signal, or that the backscatter from the wind appears to the MLE as rain. Thus, in some wind/rain regimes, especially when the rain dominates, the wind may not be identifiable. However, for most cases, the MLE can wind from rain because the co-polarization ratios of wind- and rain-induced backscatter are different.

Validation of directional retrieval is demonstrated in Figures 6.12 and 6.13. We note that we have excluded simultaneous wind/rain wind that have a "0" wind

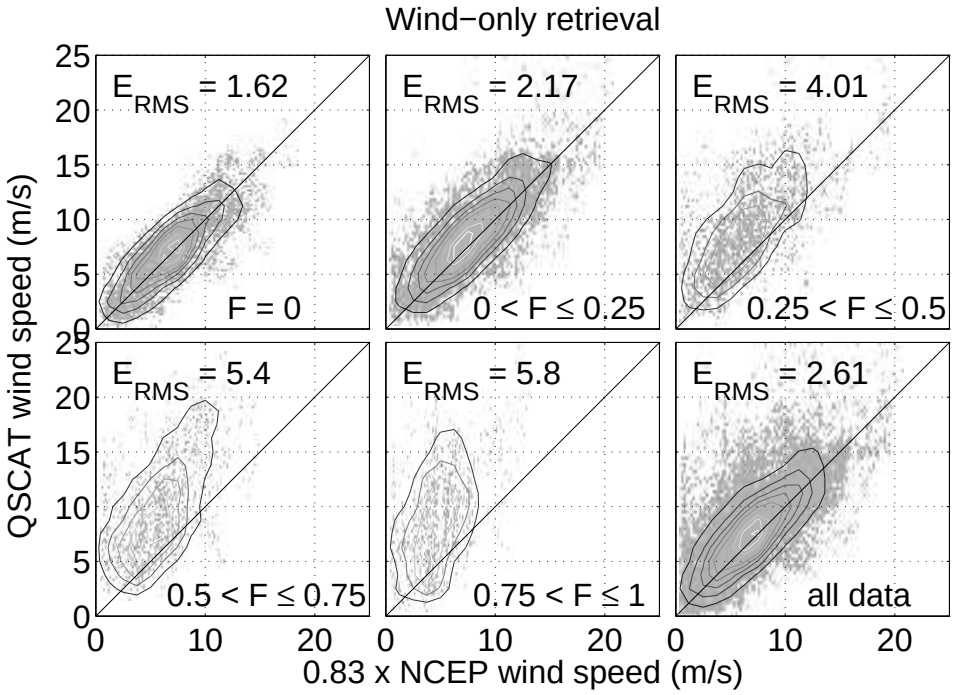


Figure 6.11: Scatter plot with density curves of the bias-corrected NCEP wind speeds versus wind-only retrieved wind speeds for various rain fraction bins.

speed, because the direction is undefined in such situations. At higher rain fractions, the simultaneous wind/rain directions are much closer to the true directions, while at rain fractions above 0.25, the wind-only retrieval begins to have retrieved directions at only 90 and 270 degrees (cross-swath). Thus, where rain is significant, simultaneous retrieval aids in correcting wind directions corrupted by rain. The exception is at very high rain fractions > 0.75, where the rain dominates and neither wind-only or simultaneous wind/rain retrieval are valid.

Rain rate validation

Next, we compare the QuikSCAT-retrieved rain rates to the PR-derived weighted average rain rates over the co-located dataset. A scatter plot of PR rain rates against QuikSCAT rain rates are shown in Figure 6.14. Since the plot is on a

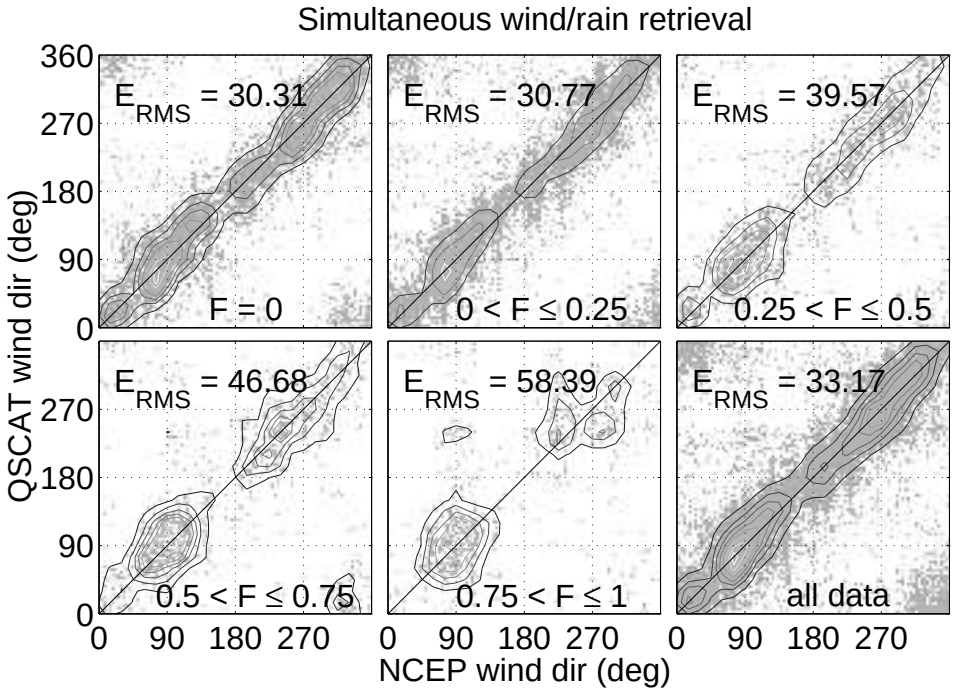


Figure 6.12: Scatter plot with density curves of NCEP wind directions versus simultaneous wind/rain retrieved wind directions for various rain fraction bins.

log-log scale, "zero" rain rate values are not shown. The zero values correspond to false alarms and missed detections.

Although the QuikSCAT derived rain rates have considerable scatter in comparison to the PR rain rates, Figure 6.14 demonstrates a strong correlation between QuikSCAT and PR-derived rain rates. The QuikSCAT rain rates are biased somewhat high, which is expected from the simulation. However, the bias can be corrected. For example, Figure 6.14 shows a quadratic fit of the QuikSCAT rain rates to the TRMM rain rates that can be used to correct for the bias. The bias correction reduces the RMS error by about a fourth. Using the bias correction, the rain backscatter model could be tuned to yield an unbiased estimate (see Appendix I).

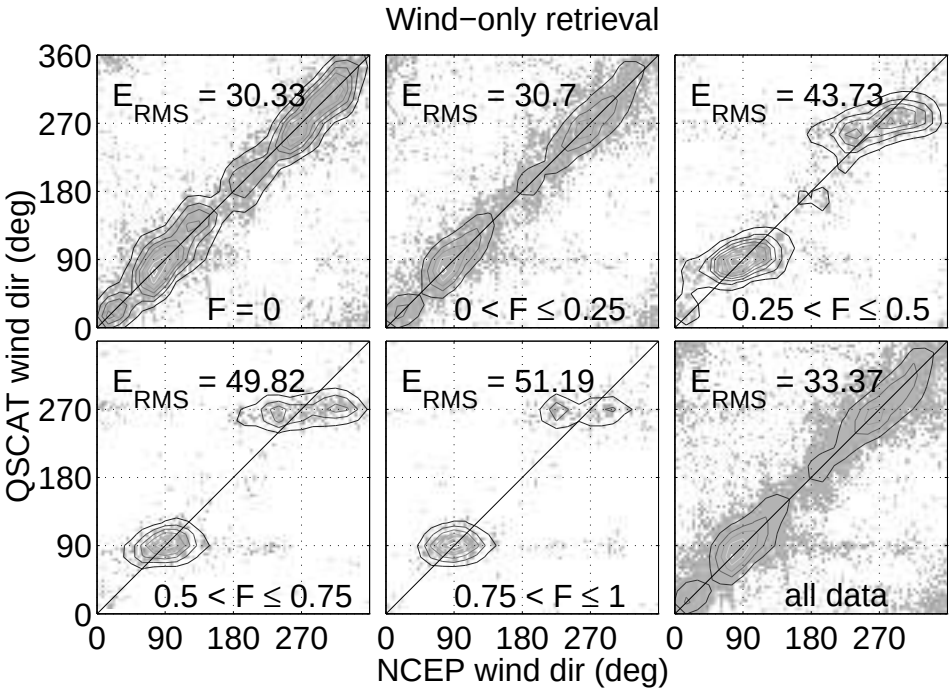


Figure 6.13: Scatter plot with density curves of NCEP wind directions versus wind-only retrieved wind directions for various rain fraction bins.

6.5 Synoptic example

In this section, we examine a co-located PR/QuikSCAT example over a hurricane. The location of the storm is over the Ryukyu Islands off the southern tip of Japan on September 22, 1999. Figure 6.15 shows the QuikSCAT-derived wind vectors for both simultaneous wind/rain and wind-only retrieval, along with the QuikSCAT-derived rain rates and the co-located TRMM PR derived rain rates.

On the far left, the coverage of the storm is limited to the outer-beam region of the QuikSCAT swath. Retrieval of rain in this area is not usable. Thus, the wind-only retrieved wind vectors are shown in Figure 6.15a in the outer swath region.

The wind-only retrieval exhibits many rain-induced features that are corrected by the simultaneous wind/rain retrieval. The most obvious of these features are the rain bands located up to about 24 wind vector cells (600 km) from the center of the storm. The rain band is also visible in the TRMM PR data. The wind-only

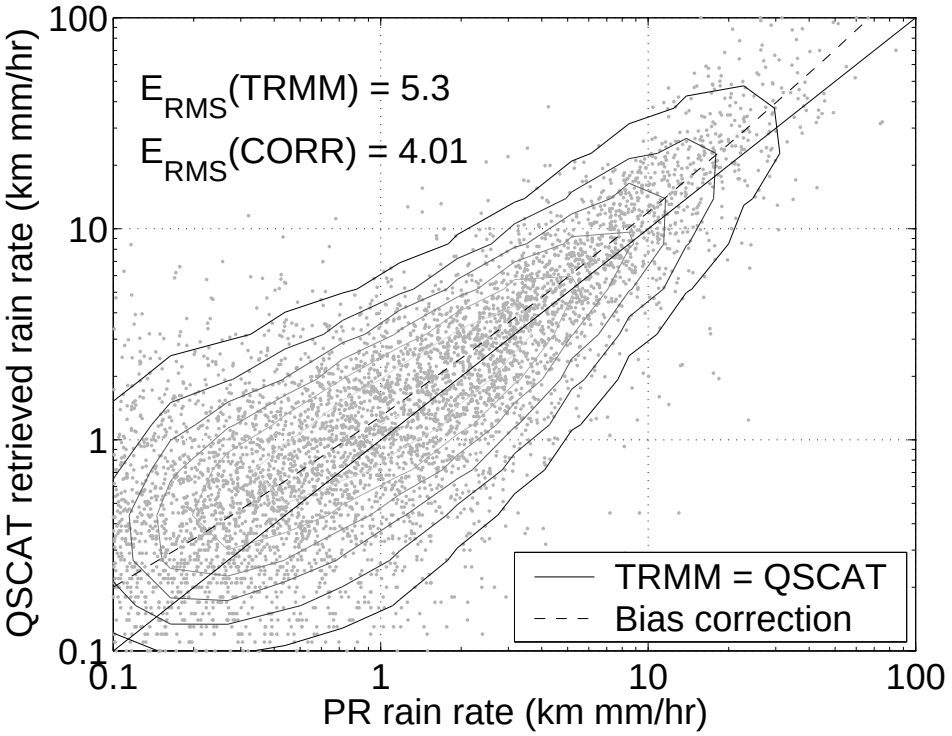


Figure 6.14: Scatter plot of TRMM PR derived "effective" weighted average rain rates versus QuikSCAT-derived rain rates. Density curves are shown, along with the equality line (solid). The dotted line represents the best quadratic fit to TRMM rain rate (in log space).

retrieval shows dramatic "apparent" wind speed increases and corrupted directions in the rain band due to the electromagnetic scattering from rain. The simultaneous wind/rain retrieval yields wind speeds in the rain bands that are more consistent with the wind speeds of neighboring WVCs, along with generally more self-consistent directions.

Another rain feature corrected by simultaneous wind/rain retrieval is the corruption due to the large area of rain just south of the storm center. In this case, the wind-only retrieval shows wind vectors all pointing nearly east (in the cross-track direction), an indicator of rain. The QuikSCAT-retrieved rain rates are very high in this region, which is consistent with the TRMM retrieved rains over the portion covered by the PR. The simultaneous wind/rain retrieval shows a much more

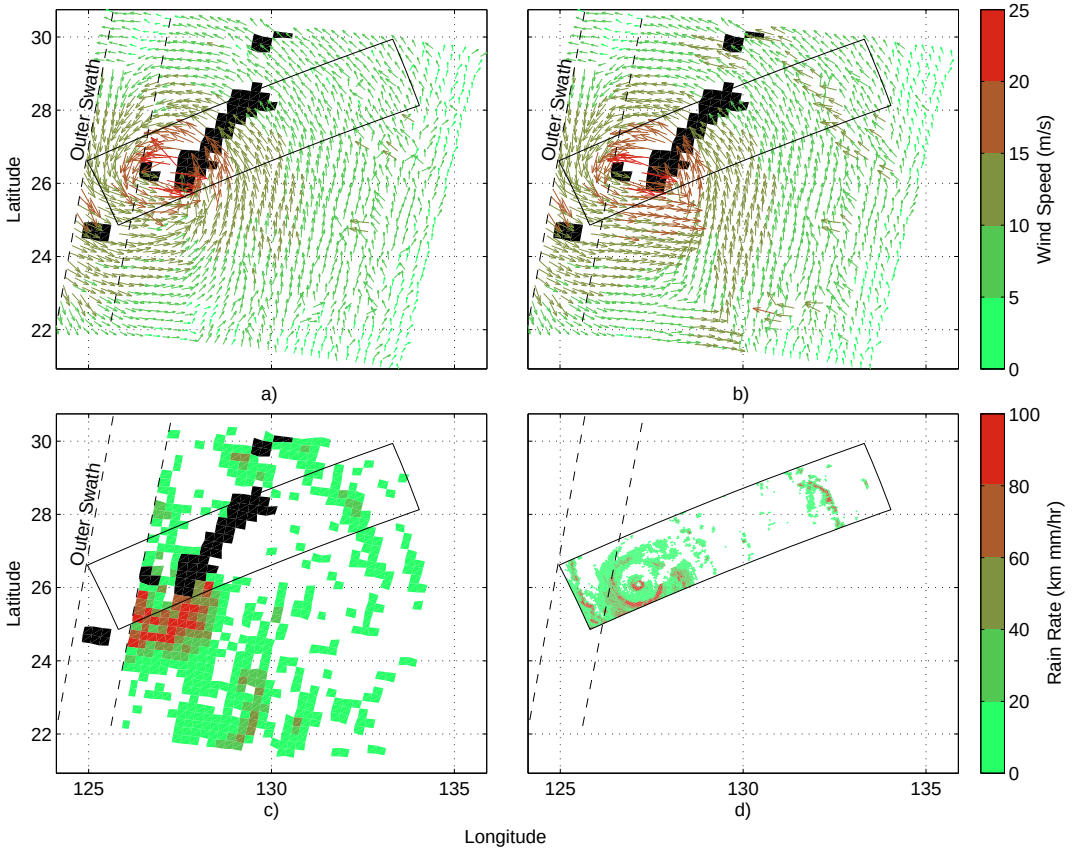


Figure 6.15: Synoptic example of a hurricane: a) QuikSCAT simultaneous wind/rain derived wind vectors. b) QuikSCAT wind-only retrieval. c) QuikSCAT simultaneous wind/rain derived rain rates. d) Raw PR rain rates. The box shows the coverage of the PR data. Black pixels represent WVCs flagged as containing land.

consistent circular flow in this region, suggesting better wind retrieval over the wind-only method. These corrected features demonstrate that simultaneous wind/rain retrieval has the capability of correcting rain-corrupted winds.

Although simultaneous wind/rain retrieval corrects many of the rain-induced features, the rain retrieval process also produces some possible spurious rain rates as seen in the center of Figure 6.15c, where the wind-only retrieval shows little evidence of rain contamination.

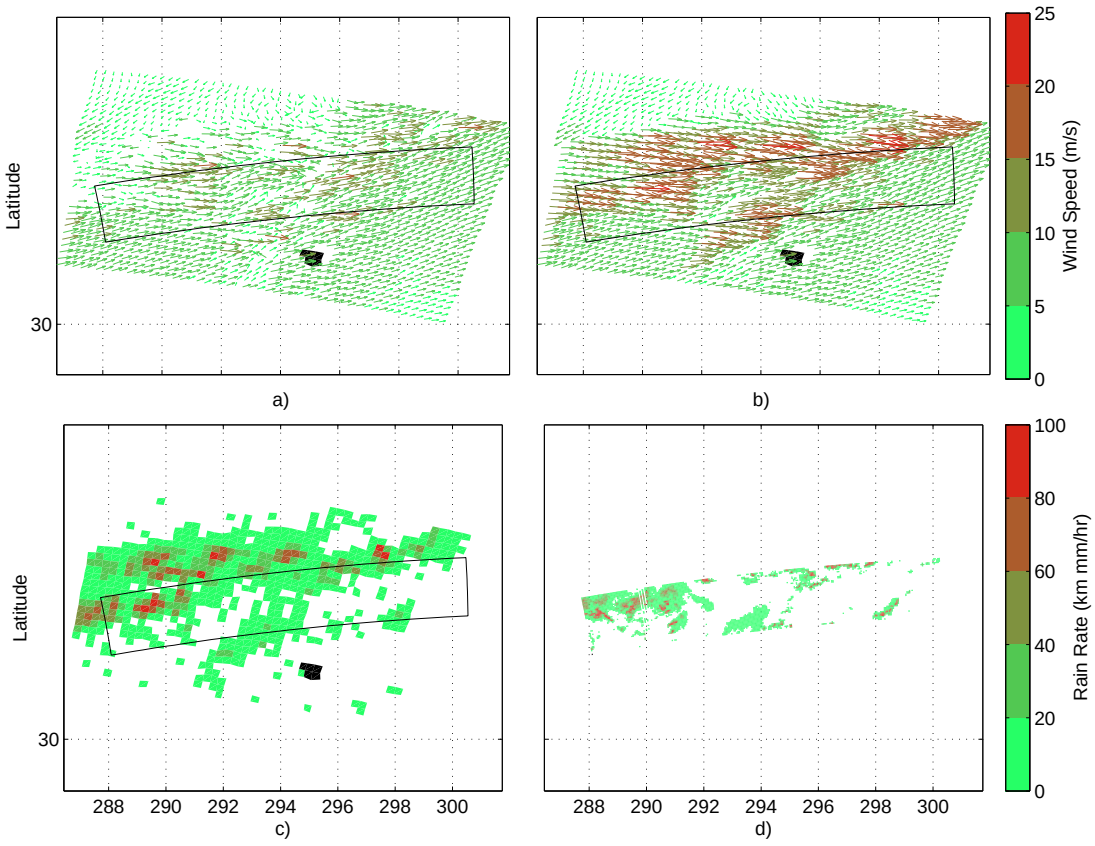


Figure 6.16: Example of a rain event a) QuikSCAT simultaneous wind/rain derived wind vectors. b) QuikSCAT wind-only retrieval. c) QuikSCAT simultaneous wind/rain derived rain rates. d) Raw PR rain rates. The box shows the coverage of the PR data. Black pixels represent WVCs flagged as containing land.

Another example of simultaneous wind/rain retrieval is shown in Figure 6.16. This example further illustrates the ability of QuikSCAT to measure rain. In addition, the simultaneously retrieved wind flow is much more realistic and consistent than the wind-only data. Further consideration of the validation of QuikSCAT rain retrieval is given in Chapter 7.

6.6 Conclusions

Although rain has been shown to be one of the most significant factors that corrupts SeaWinds scatterometer data, the new technique of simultaneously retrieving

ocean winds and rain significantly improves the wind speed estimate for many rain-corrupted areas. As a side benefit, simultaneous wind/rain retrieval provides an estimate of the rain rate, which, while somewhat noisy and biased, has much broader coverage than instruments such as the TRMM PR. When an outside estimate of the rain rate is available, rain-corrected wind retrieval can be used to further improve the rain/wind estimates.

SeaWinds on QuikSCAT rain retrieval has been shown by simulation to give the best results in the sweet spot, and to not work well on the swath edges. Simulation also demonstrates that wind speeds from simultaneous wind/rain retrieval are nearly unbiased, while the wind-only retrieval produces increasingly biased estimates as rain increases. However, in zero rain conditions, simultaneous wind/rain retrieval is generally more ill conditioned than wind-only retrieval, and thus does not perform as well. It is thus beneficial to develop a rain flag, and only perform simultaneous retrieval in raining areas.

The rain-corrected wind retrieval method presented here enables synergistic use of the SeaWinds scatterometer and AMSR radiometer aboard ADEOS II to correct rain-contaminated wind vectors. In addition, the simultaneous wind/rain method can be applied for re-analysis of SeaWinds on QuikSCAT data. In the next chapter, this method is further explored with a method of reducing spurious rain rate estimates to reduce the overall bias of the estimates. Monthly rain rate validation with the TRMM microwave imager is also presented in which QuikSCAT is shown to be consistent (although somewhat biased) with passive data.

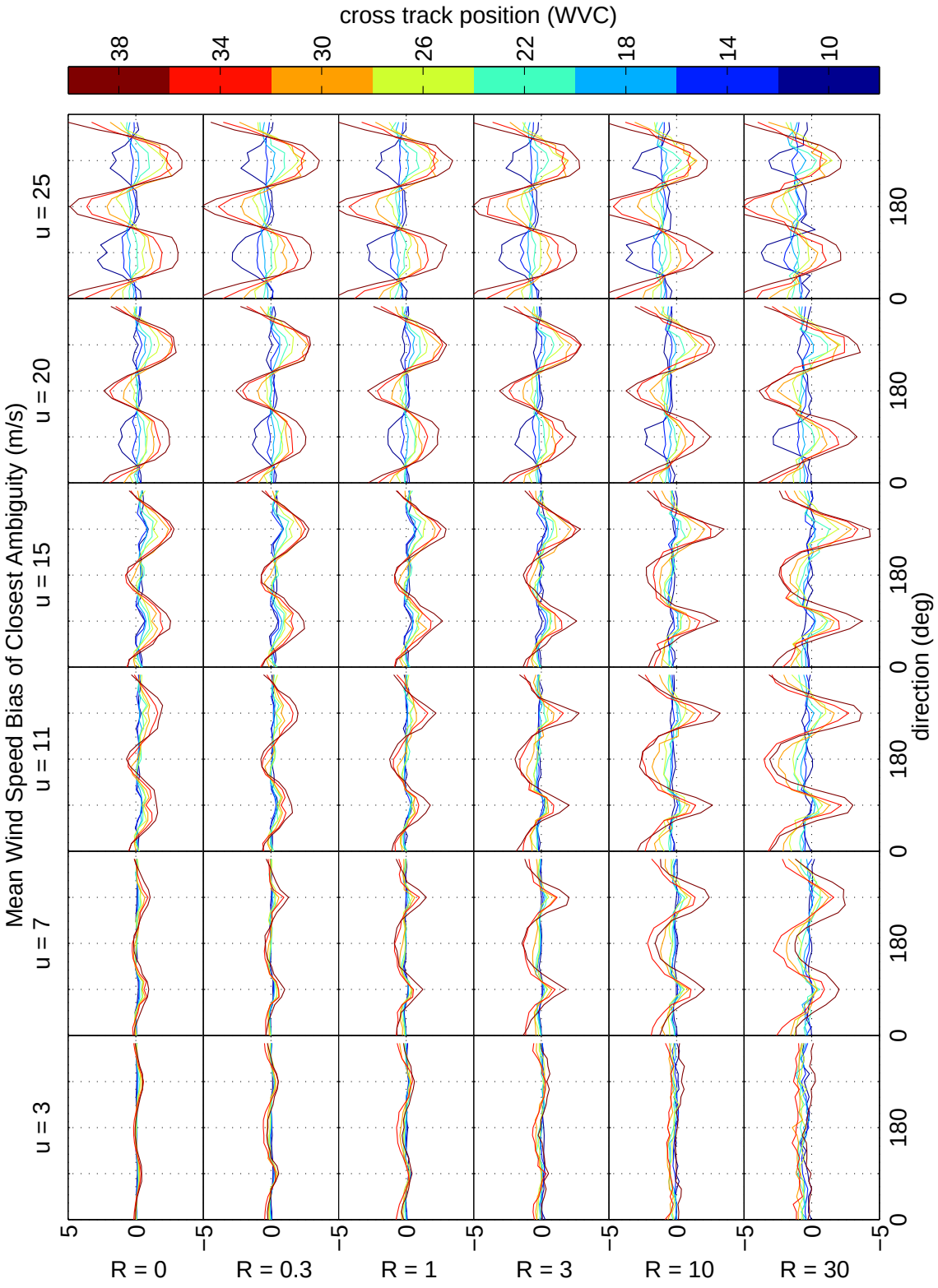


Figure 6.17: Mean wind speed bias as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval.

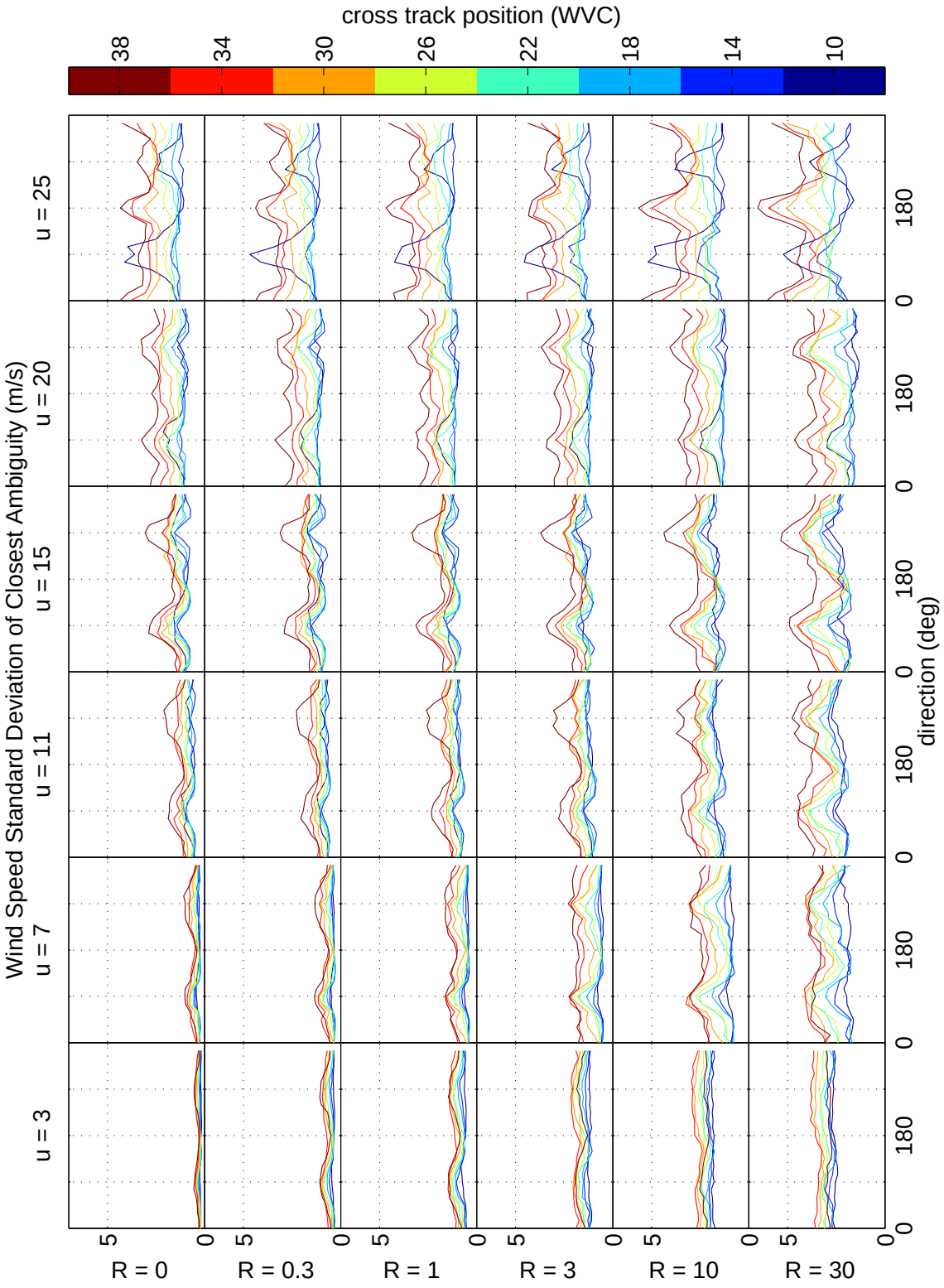


Figure 6.18: Wind speed standard deviation as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval.

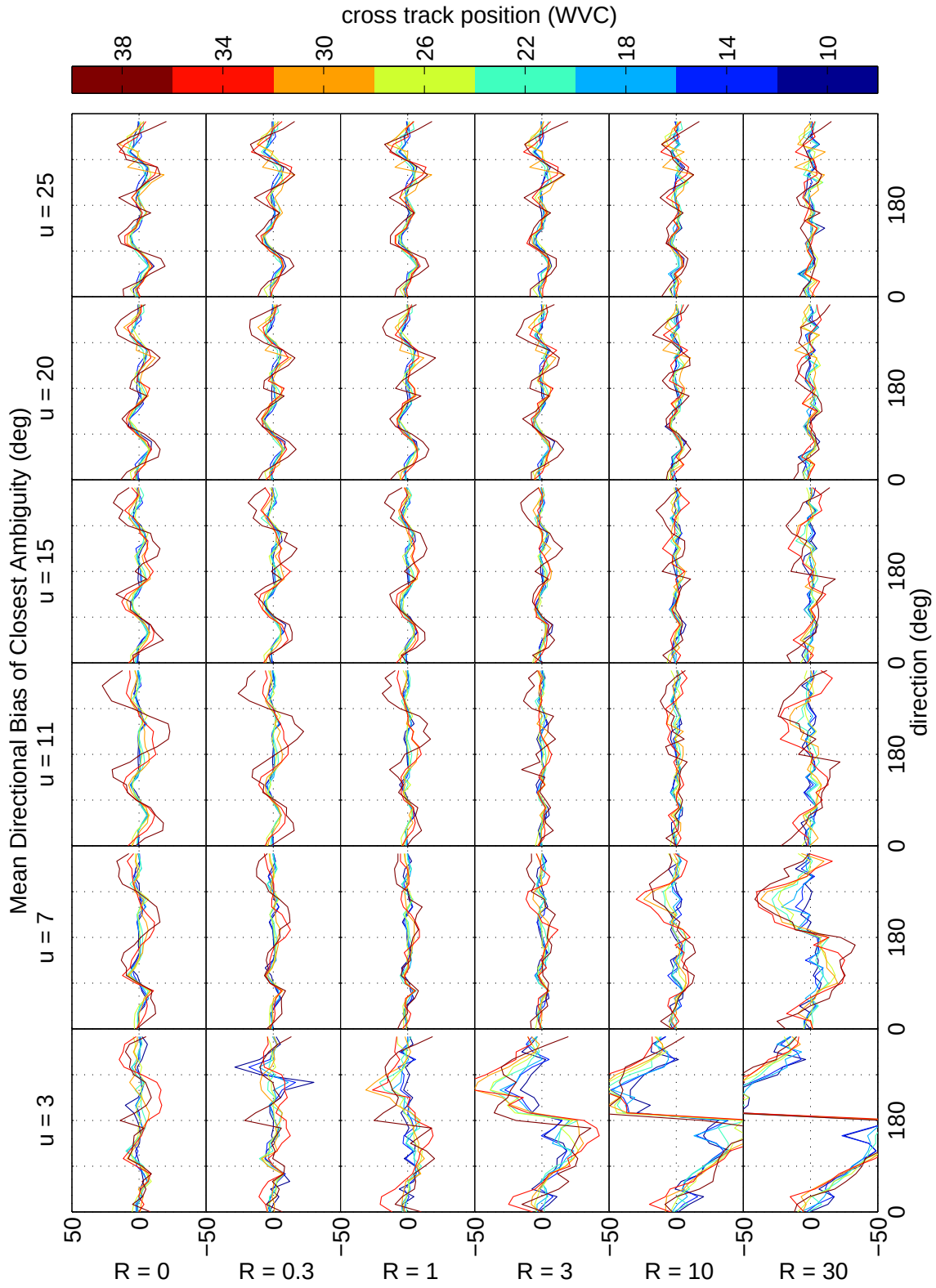


Figure 6.19: Mean directional bias as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval.

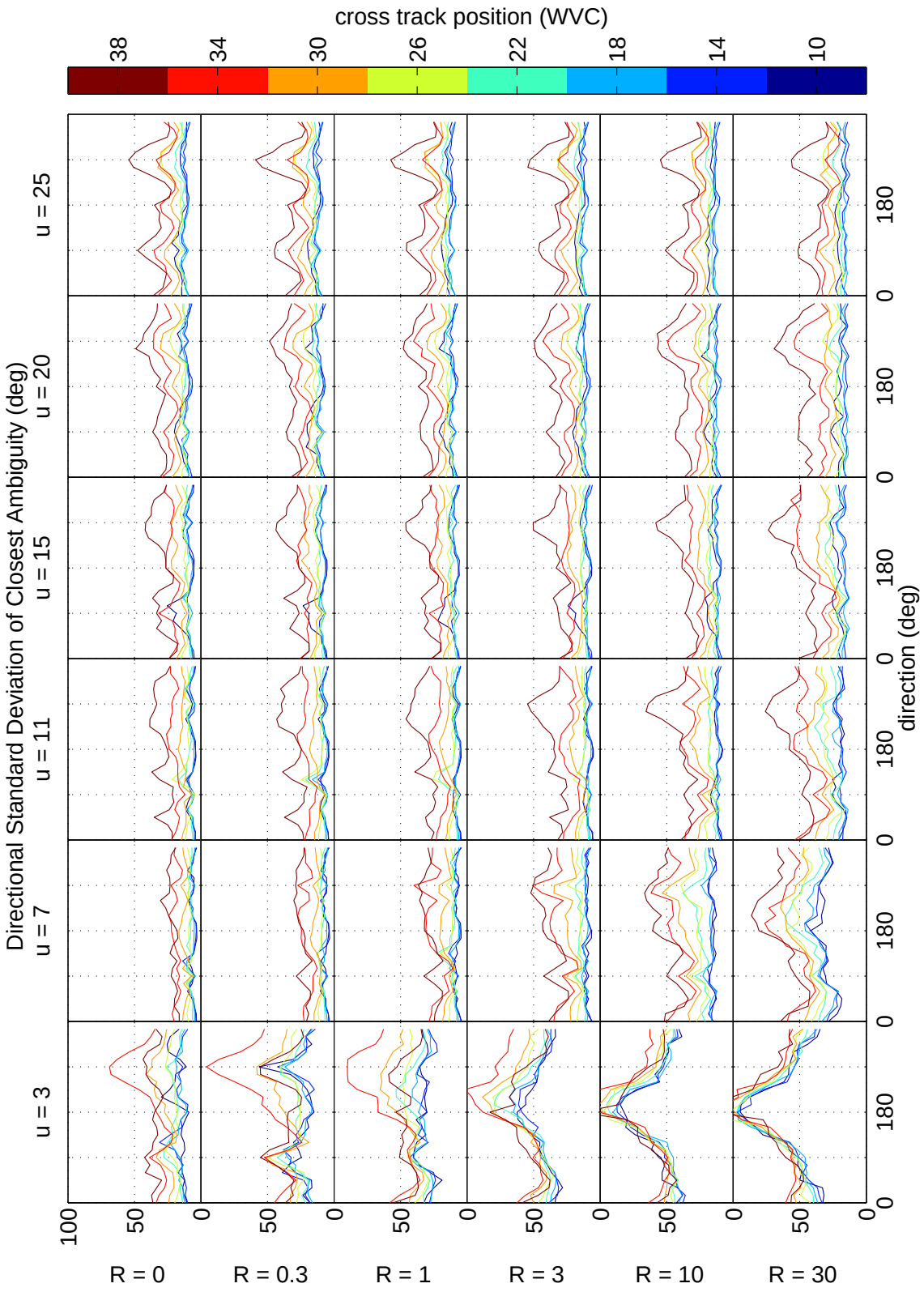


Figure 6.20: Wind direction standard deviation as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval.

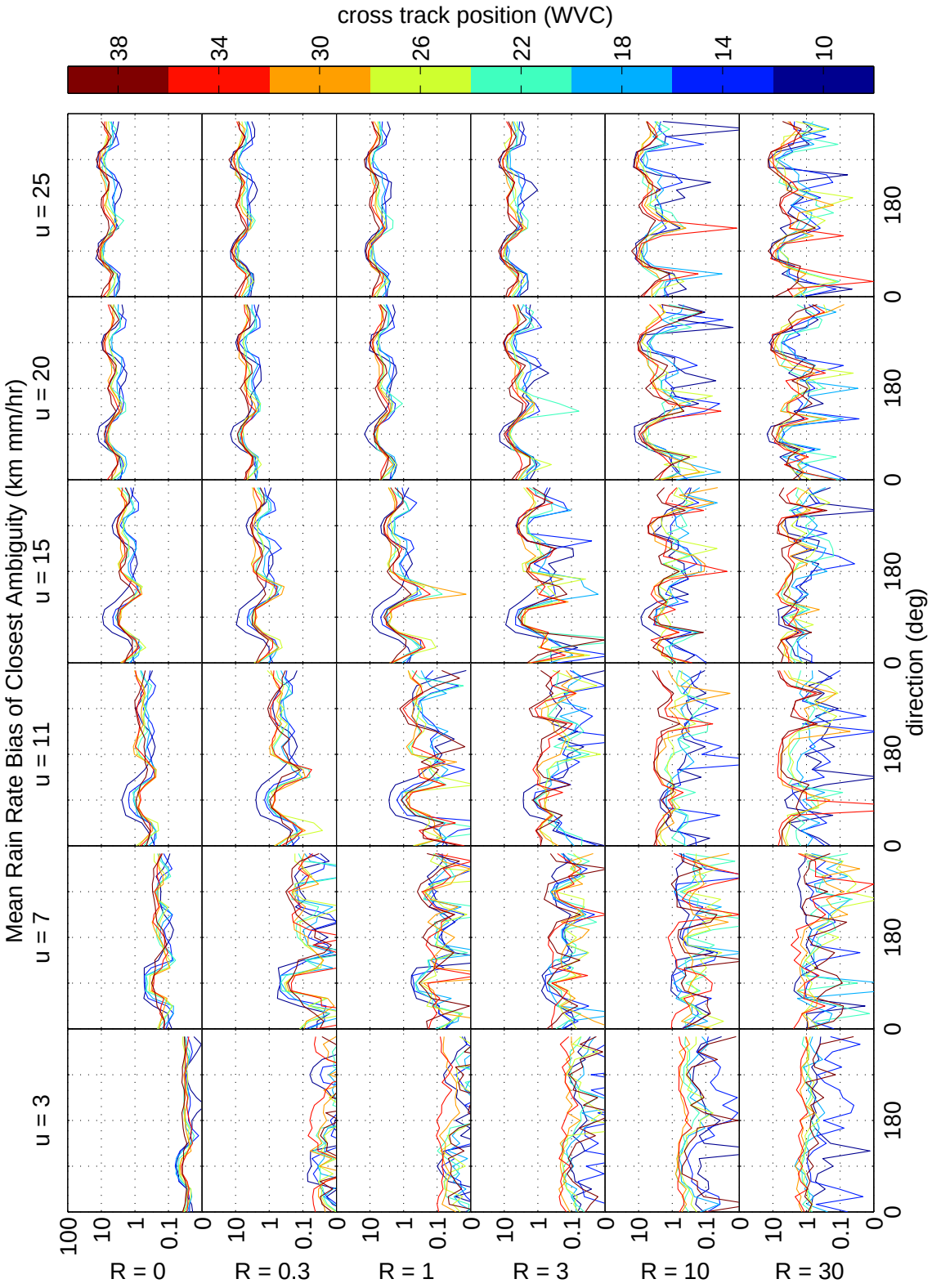


Figure 6.21: Mean rain rate bias as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval.

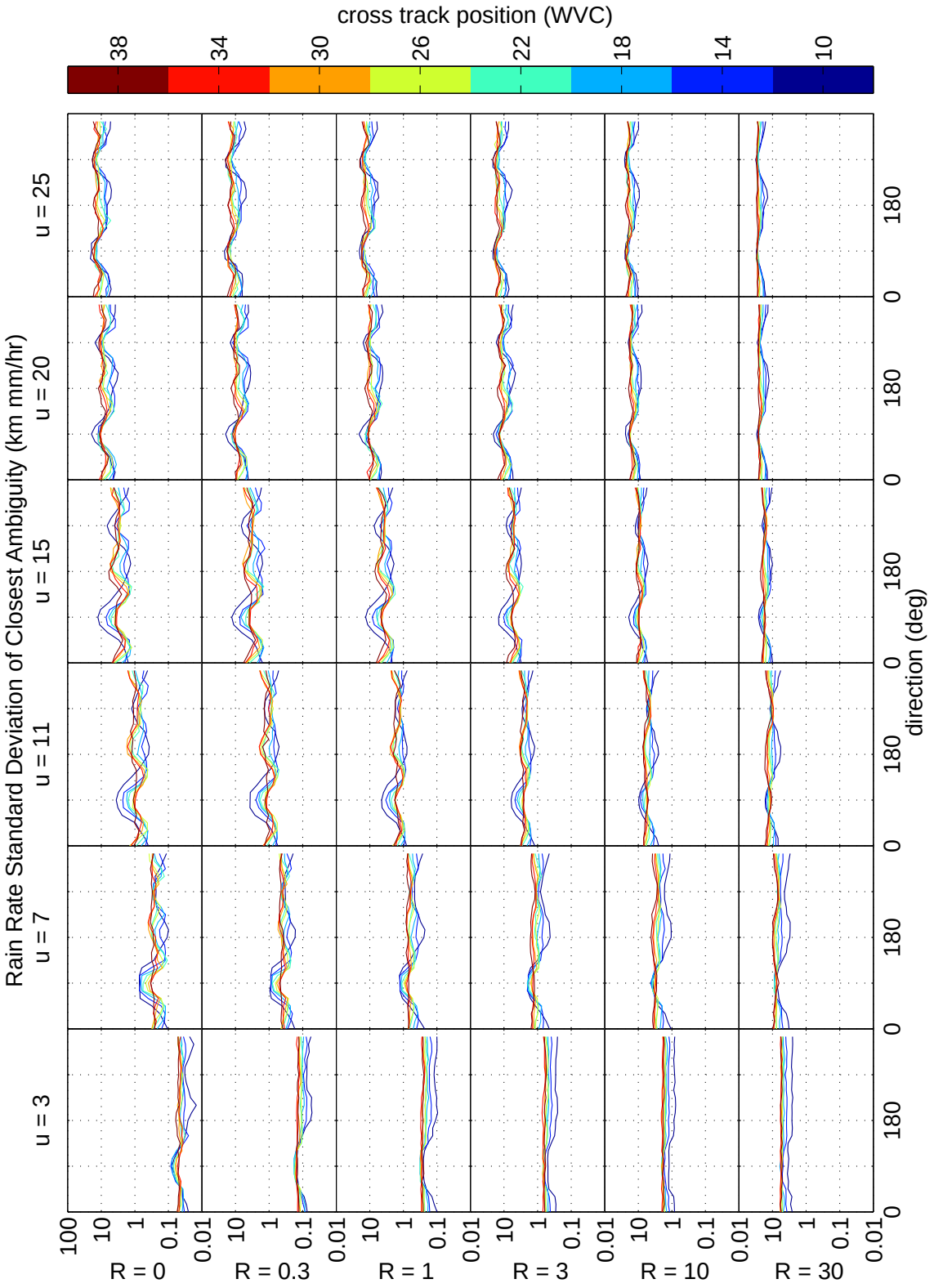


Figure 6.22: Rain rate standard deviation as a function of wind direction for several simulation wind speeds and rain rates for a range of inner-beam cross track positions for simultaneous wind/rain retrieval.

Chapter 7

Assessing the Quality of SeaWinds Rain Measurements

Considering that SeaWinds was not intended to measure rain, simultaneous wind/rain retrieval described in the previous chapter is surprisingly accurate, especially in the sweet spots of SeaWinds swath, and when the true wind direction is not oriented in the cross-swath direction. The main insufficiency in the method is related to the fact that wind speed and rain rate do not have orthogonal effects on backscatter. Thus, depending on the wind direction and measurement geometry, backscatter due to wind can be confused with the response from rain. The effects of this identifiability problem is most noticeable when the wind is oriented cross-swath (see Figure 7.1). With wind in a cross-swath orientation, the backscatter response is similar between the fore and aft beams, appearing isotropic, a signature similar to rain. However, when the wind is not oriented cross-swath, the wind and rain backscatter signatures are sufficiently different to afford good rain retrieval. In the inner-beam region of the swath, where both v-pol and h-pol measurements are available, the effects of this wind/rain identifiability problem is lessened by the fact that the co-polarization ratio of wind-induced backscatter is different than that of rain, i.e. the HH-pol response from rain is higher than VV-pol, while the VV-pol response due to wind-roughened seas is greater than the HH-pol. Thus, the use of measurements at different polarizations affords better rain/wind discrimination than is available with measurements from only one polarization (such as from the v-pol-only outer beam region).

Wind/rain identifiability analysis for cross-swath blowing winds raises two issues in retrieving accurate rain rates: First, for cross-swath winds, the simultaneous wind retrieval method can infer spurious rain rates [56]. Second, the retrieved rain

rates are biased somewhat high. This chapter investigates these issues by further investigation of simulation statistics and via the Cramér-Rao (C-R) theoretical lower bound on estimate accuracy. From the simulations, a rain rate threshold is developed that indicates the level of rain required as a function of wind speed, direction, and cross track position to confidently identify rain in the wind vector cell. This rain flag threshold is set higher where spurious rain rates are more likely to occur. Using simultaneous wind/rain retrieval and the thresholds, a method of improving the current SeaWinds Multi-dimensional Histogram (MUDH) [28] rain flag is introduced. The combined rain flag is shown to perform better than the MUDH flag for low to moderate rain rates when comparing with co-located Tropical Rainfall Measuring Mission (TRMM) precipitation radar (PR) measurements. Using the rain flag, monthly average rainfall statistics are compared between the QuikSCAT data and the TRMM microwave imager (TMI) [57]. Monthly-averaged QuikSCAT rain rates (although noisy) visually correspond well to TMI-derived monthly average rain rates, and are only slightly biased.

After presenting background to the SeaWinds instrument and simultaneous wind/rain retrieval in Section 7.1, we analyze the quality of SeaWinds active rain measurements via the Cramér-Rao bound and simulation statistics in Section 7.2. In Section 7.3, the simultaneous wind/rain retrieval-based rain flag is presented, and validation of QuikSCAT monthly rainfall estimates with TMI data are given in Section 7.4.

7.1 Background

As described in Chapter 5, during rain the backscatter response of the ocean is attenuated by falling hydrometeors. In addition, the nominal backscatter response is augmented by both scattering from falling rain, and surface ripples formed by rain striking the water. These effects are modeled as a net attenuation from the nominal surface backscatter, and a net effective backscatter response,

$$\sigma_m = \sigma_w \alpha_r + \sigma_e \tag{7.1}$$

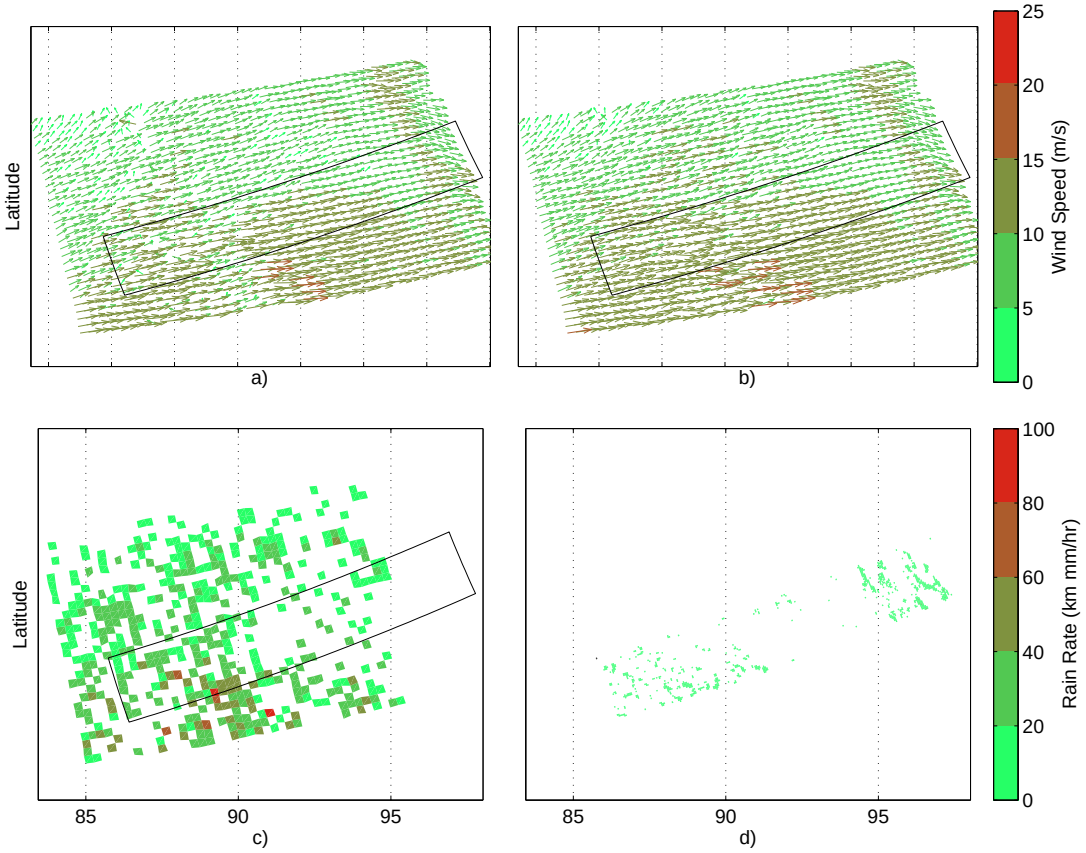


Figure 7.1: Example of wind/rain estimation for cross-swath blowing winds: a) QuikSCAT simultaneous wind/rain derived wind vectors. b) QuikSCAT wind-only retrieval. c) QuikSCAT simultaneous wind/rain derived rain rates. d) Raw PR rain rates. The box shows the coverage of the PR data. Black pixels represent WVCs flagged as containing land.

A parametric model for the attenuation α_r and effective rain backscatter σ_e as a function of integrated rain rate R is computed in Chapter 5 [53]. The parametric model has the form,

$$\alpha_r(R) = 10^{-10^{f_a(R_{dB})/10}/10} \tag{7.2}$$

$$\sigma_e(R) = 10^{f_e(R_{dB})/10} \tag{7.3}$$

where $R_{dB} = 10\log_{10} R$, and $f_a(R_{dB})$ and $f_e(R_{dB})$ are defined as 2^{nd}-order polynomials in R_{dB},

$$f_a(R_{dB}) = \sum_{n=0}^{2} x_a(n) R_{dB}^n \tag{7.4}$$

$$f_e(R_{dB}) = \sum_{n=0}^{2} x_e(n) R_{dB}^n. \tag{7.5}$$

As discussed earlier, the measurement probability density of the combined wind/rain signal has the form,

$$p(\mathbf{z}|u,d,R) = \prod_k \frac{1}{\sqrt{2\pi}\varsigma_{rk}} \exp\left\{-\frac{1}{2}\frac{(z_k - \mathcal{M}_{rk})^2}{\varsigma_{rk}^2}\right\} \tag{7.6}$$

where k indexes the backscatter measurements $\mathbf{z}$ in the WVC, u is the wind speed, and d is the wind direction. The probability distribution can be re-written as a log-likelihood function of the form

$$L(u,d,R) = -\frac{1}{2}\sum_k \left\{\log(2\pi) + \log\varsigma_k^2 + \frac{(z_k - \mathcal{M}_{rk})^2}{\varsigma_k^2}\right\}. \tag{7.7}$$

In Eq. (7.6), the mean and variance are

$$\mathcal{M}_{rk} = \mathcal{M}_k \alpha_{rk} + \sigma_{ek} \tag{7.8}$$

$$\varsigma_{rk} = (1+\alpha_k)\epsilon_k + \alpha_k \mathcal{M}_{rk}^2 + \beta_k \mathcal{M}_{rk} + \gamma_k \tag{7.9}$$

where α_k, β_k, and γ_k are the communication noise coefficients of the k^{th} measurement [16]. Also, Eqs. (7.8) and (7.9) contain the following simplified notation where the dependence on wind and rain is implied,

$$\mathcal{M}_k = \mathcal{M}(u, d-\phi_k, \theta_k, \mathrm{pol}_k) \tag{7.10}$$

$$\alpha_{rk} = \alpha_r(R, \mathrm{pol}_k) \tag{7.11}$$

$$\sigma_{ek} = \sigma_e(R, \mathrm{pol}_k) \tag{7.12}$$

where $\mathcal{M}$ is the GMF, ϕ_k is the antenna azimuth angle, θ_k is the incidence angle, pol_k is the polarization of the k^{th} measurement, and ϵ_k is defined as

$$\epsilon_k = (K_{pm}\alpha_{rk}\mathcal{M}_k + K_{pe}\sigma_{ek})^2 \tag{7.13}$$

where K_{pm} and K_{pe} represent the normalized standard deviation of the GMF and effective rain backscatter, respectively [15, 56].

7.2 Cramér-Rao bound

In this section, we present simultaneous wind/rain estimation statistics obtained from Monte-Carlo simulations and the C-R bound. The C-R bound gives a lower bound on the variance of an unbiased estimator [26] and has also been generalized to include biased estimators. Let $\mathbf{p} = (u, d, R)^T$ be the estimation parameters for simultaneous wind/rain retrieval with u the wind speed in m/s, d the wind direction in degrees, and R the rain rate in km mm/hr. Also, let $\hat{\mathbf{p}}$ be the MLE estimate of the parameters. Using the notation of [15], the covariance of a biased estimate given the measurements $\mathbf{z}$ is approximated by [58]

$$\begin{aligned} C(\mathbf{p}) &= E\{[\hat{\mathbf{p}} - E\hat{\mathbf{p}}][\hat{\mathbf{p}} - E\hat{\mathbf{p}}]^T\} \\ &\gtrsim \frac{\partial E\hat{\mathbf{p}}}{\partial \hat{\mathbf{p}}} J^{-1}(\mathbf{p}) \frac{\partial E\hat{\mathbf{p}}}{\partial \hat{\mathbf{p}}}^T \end{aligned} \tag{7.14}$$

where $J(\mathbf{p})$ is the Fisher-information matrix given by

$$J(\mathbf{p}) = E\left\{ \left[\frac{\partial L(\mathbf{p}, \mathbf{z})}{\partial \mathbf{p}}\right]^T \frac{\partial L(\mathbf{p}, \mathbf{z})}{\partial \mathbf{p}} \right\}. \tag{7.15}$$

If the estimator is unbiased (or biased by a constant value), the derivative of the estimate given in Eq. (7.14) reduces to the identity, and the bound becomes

$$C(\mathbf{p}) \geq J^{-1}(\mathbf{p}). \tag{7.16}$$

The solution of the Fisher-information matrix for the unbiased estimate corresponding to the scatterometer likelihood function of Eq. (7.7) is given in [15] and is equal to

$$J_{ij} = \sum_k \left[\frac{\partial \mathcal{M}_{rk}}{\partial p_i} \frac{1}{\varsigma_k^2} \frac{\partial \mathcal{M}_{rk}}{\partial p_j} + \frac{\partial \varsigma_{rk}^2}{\partial p_i} \frac{1}{2\varsigma_k^4} \frac{\partial \varsigma_{rk}^2}{\partial p_j} \right] \tag{7.17}$$

for the $\{i,j\}^{th}$ element of the matrix. Derivations of the rain model and variance gradients are given in Appendix I.

Since the simultaneous wind/rain retrieval likelihood function has several maxima, each corresponding to a possible wind vector estimate, ambiguity selection must be taken into account in order to discuss statistics. In simulation, ambiguity selection is performed by selecting the ambiguity closest to the true wind. With the C-R bound, explicitly taking into account ambiguity selection is impossible. However, because the C-R bound is computed from derivatives evaluated at the true wind and rain, the computed variance is directly associated with the shape of the likelihood function surrounding the true wind and rain, and not around the other ambiguities. Thus, the C-R bound inherently computes the variability of the ambiguity closest to the true wind. We can thus ignore ambiguity selection in the C-R bound analysis.

7.2.1 C-R / Monte-Carlo simulation analysis

As simulations and validation suggest [56], the wind/rain MLE is somewhat biased. Thus, the most accurate representation of the MLE variance is the biased C-R bound [59]. The biased C-R bound, however requires knowledge of the derivative of the biased estimate, which is difficult to accurately compute. An approach to computing the bias derivative is given in [15].

Rather than compute the biased C-R bound, for simplicity we compute the unbiased bound, which we compare to simulation results. Monte Carlo simulations are performed in which wind and rain are simultaneously retrieved for a variety of wind/rain conditions and cross-track positions [56]. The simulation statistics are computed over 500 noise realizations for each set of conditions. Likewise, the unbiased C-R bound is computed. Standard deviation as a function of wind direction (relative to the direction of the satellite), and cross track position is plotted in Figures 7.2 and 7.3 for a wind speed of 7 m/s . For most sets of conditions in these examples, the C-R bound matches well with simulation results. We note that in some cases the simulation results are somewhat lower than the C-R bound because of biases in simultaneous wind/rain retrieval, and the fact that the model function derivatives

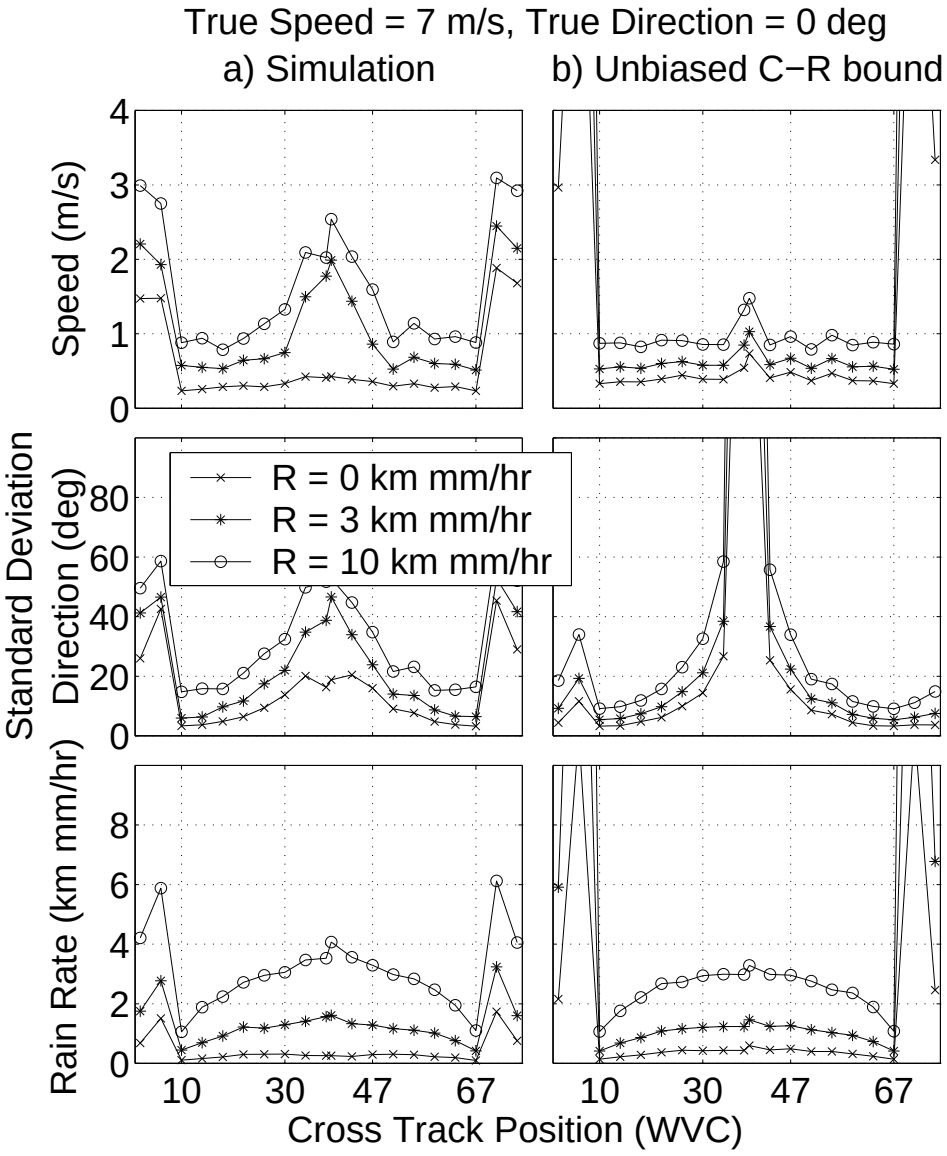


Figure 7.2: Standard deviation as a function of cross track position of retrieved wind speed, wind direction, and rain rate from QuikSCAT computed via a) simulation and b) the unbiased C-R bound for three rain rates. Notice the anomalous C-R bound spike in the nadir region for the directions, resulting from a large C-R bound due to a bias in the estimate. The true wind velocity is 7 m/s and 0 degrees.

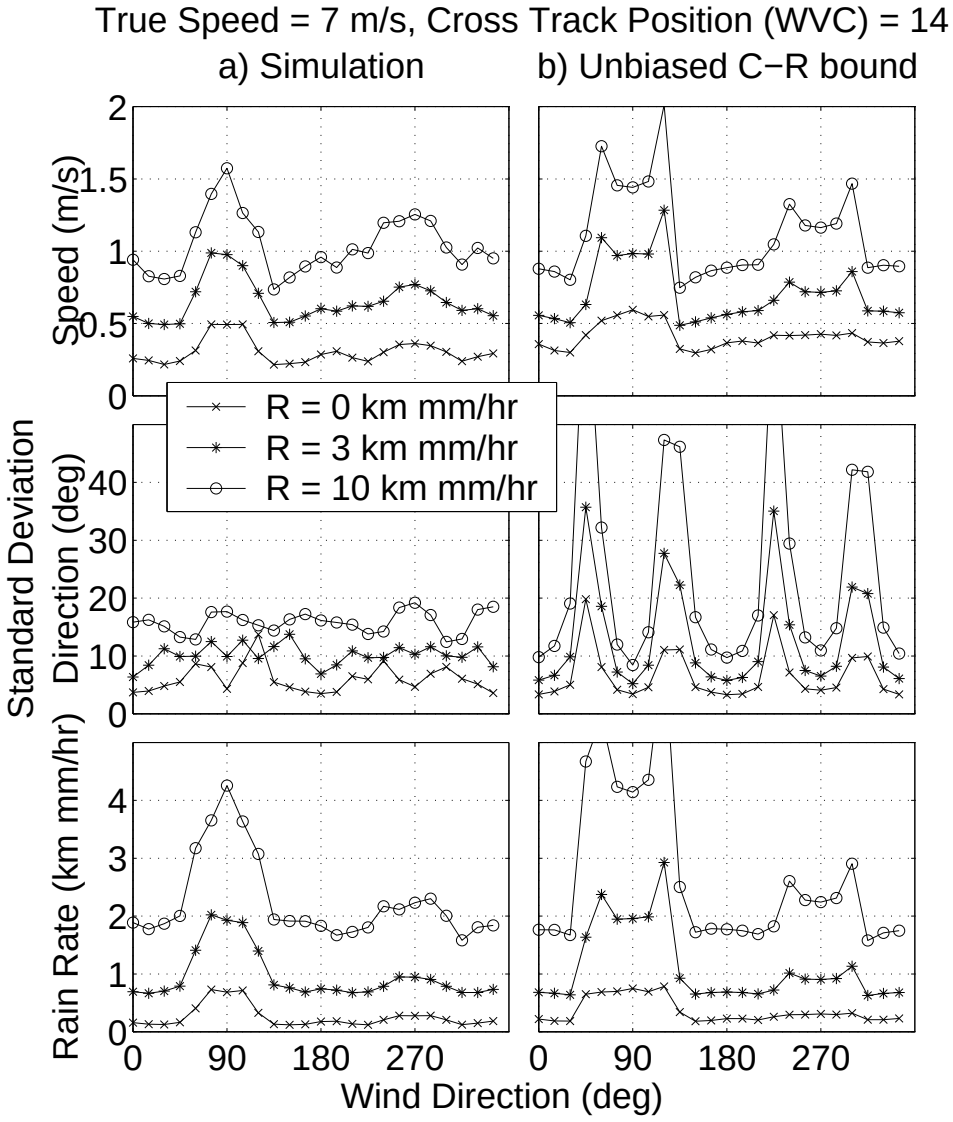


Figure 7.3: Standard deviation as a function of wind direction of retrieved wind speed, wind direction, and rain rate from QuikSCAT computed via a) simulation and b) the unbiased C-R bound for three rain rates. Notice the anomalous C-R bound spikes at 60, 120, 240, and 300 degrees for the C-R bound, resulting from a large C-R bound due to a bias in the estimate. The true wind speed is 7 m/s at cross track position 14.

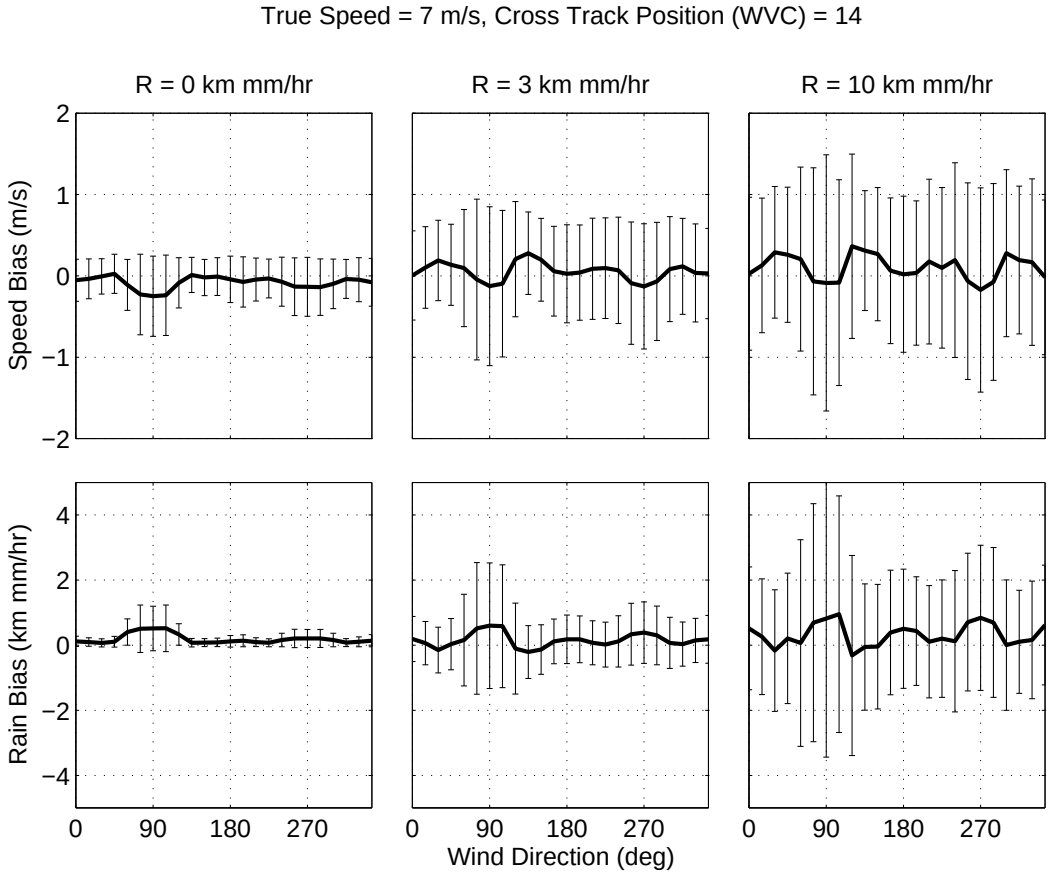


Figure 7.4: Wind speed and rain rate bias (lines) and standard deviations (error bars) for all wind directions and several rain rates with a wind speed of 7 m/s for WVC 14.

are numerically obtained. In addition, there are extreme artifacts in the C-R bound at 60, 120, 240, and 300 degrees, and at nadir (evidenced by larges spikes, especially in the directional standard deviation). These artifacts are most likely due significant biases in such regions. We recognize that the unbiased C-R bound is not useful in such areas.

We note that the SeaWinds geometry was originally designed from wind observation, and thus may be suboptimal for rain measurement. Nevertheless, rain can be accurately retrieved in most conditions. In Figure 7.2, the standard deviation is reasonable for most cross track positions (excluding the swath edge). The standard

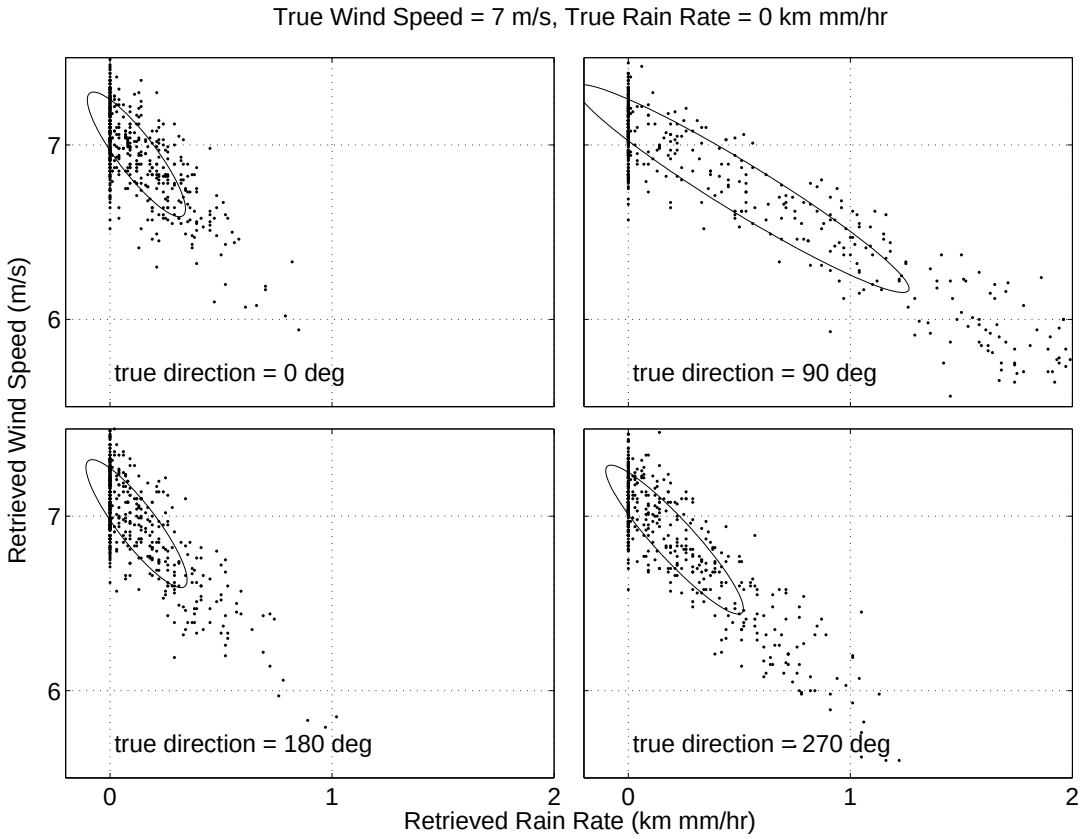


Figure 7.5: Monte Carlo Simulations of SeaWinds data at WVC 14 for several wind directions, a true wind speed of 7 m/s, and true rain rate of 0 km mm/hr. Also shown are concentration ellipses representing the unbiased C-R bound

deviation peaks around nadir for all rain rates. This phenomenon is expected, as SeaWinds has degraded performance in the nadir region. In addition, the rain rate performance is dependent upon cross track position, with highest errors in the nadir region and on the swath edges. Data from the swath edges, however, is not used in simultaneous wind/rain retrieval.

Reasonable standard deviations are shown in Figure 7.3 for most wind directions. Noticeable peaks occur at 90 and 270 degrees relative to the motion of the satellite. These wind directions represent cross-swath blowing winds, where the response from the fore and aft beams equalize, similar to rain. Since this example is located in the left swath, a wind direction of 90° indicates wind blowing toward the

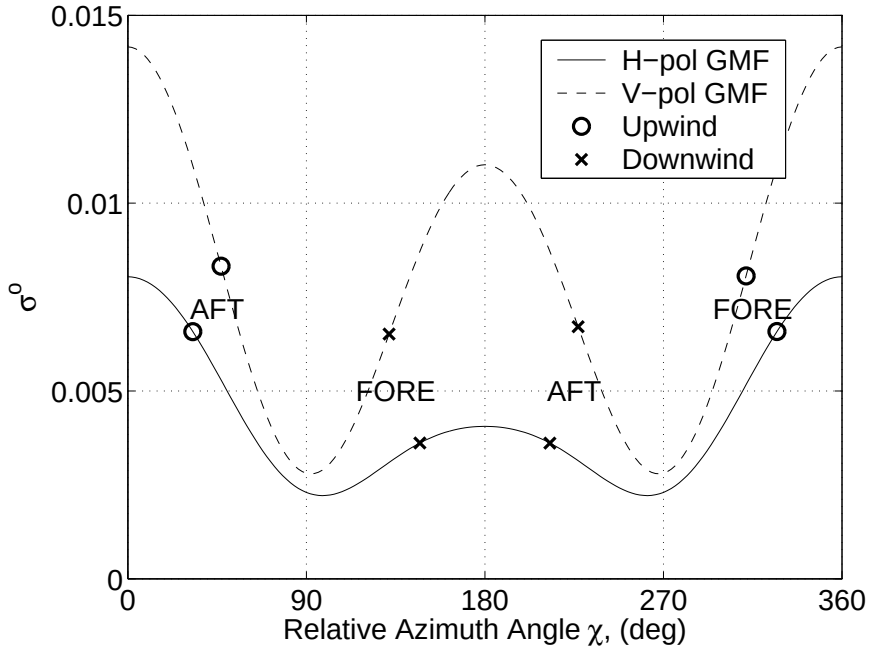


Figure 7.6: GMF for a true wind speed of 7 m/s plotted for the inner H-pol SeaWinds beam and outer V-pol SeaWinds beam at WVC 14. Also plotted are the expected fore and aft backscatter measurements for each beam given an upwind direction (wind blowing toward the instrument or $d = 90°$ relative to the direction of the spacecraft motion), or a downwind direction (wind blowing away from the instrument or $d = 270°$ relative to the direction of the spacecraft motion).

satellite (upwind), while 270° is downwind. For 90°, the fore and aft measurement azimuth angles are centered around a relative (to the wind) azimuth angle of $\chi = 0°$, while the fore and aft measurements at 270° are centered around $\chi = 180°$ (see Figure 7.6). The peak around 90 degrees is somewhat higher than the peak around 270°. This phenomenon is related to the anisotropic nature of the GMF. At a wind direction of 90° (measurements centered around $\chi = 0°$), the GMF reaches the global maximum for a given wind speed, while at 270° (measurements centered around $\chi = 180°$), a lower local maximum in the GMF occurs. Thus, at 90°, the returned $\sigma°$ values have a larger magnitude, which results in greater ambiguity between the wind and the rain than at 270°.

Figure 7.6 also shows that the V-pol backscatter response for wind-roughened seas is higher than the H-pol response. For rain backscatter, the opposite is true [20, 53], i.e. the H-pol response is higher than V-pol. Since noise perturbs the backscatter from the ideal GMF value, the observed V-pol/H-pol co-polarization ratio is often smaller than expected. When the co-polarization ratio is low and winds are pointing cross-swath (resulting in an apparent "isotropic" backscatter between fore and aft beams), simultaneous wind/rain retrieval compensates by "adding rain" to the estimate in order to lower the co-polarization ratio, and thus giving a better fit to the data. The end effect can be a spurious rain rate estimate.

Another result of the wind/rain identifiability problem is a net bias in rain rate. The wind speed and rain rate bias is illustrated in Figure 7.4 as a function of wind direction for various rain rates. The bias is dependent on rain rate, wind speed, wind direction, and cross track position. Even for zero true rain rate, the MLE estimates are biased positive around 90° and 270°. For rain rates of 3 and 10 km mm/hr, the bias is more significant at these directions. One method to remove the bias is to apply a rain rate dependent correction to the wind/rain backscatter model. Although the bias correction is not used in the remainder of the paper, Appendix J presents a general methodology of correcting the rain backscatter model to improve the overall bias using reference data from the TRMM precipitation radar.

7.3 Improved rain flag

The analysis from the C-R Bound and simulated data in Section 7.2 predict a bias and higher variability in the retrieved wind and rain for cross-swath winds. This phenomenon is additionally illustrated in Figure 7.5 for WVC 14, at a true rain rate of 0 km mm/hr. Because the true rain rate in 0 km mm/hr, these spurious rain rates are QuikSCAT false rain detections. As a result of possible spurious rain detections, we present an rain rate thresholding scheme that indicates when simultaneously retrieved rain rates are valid, and thus suggesting corruption in the wind-only speeds.

The simultaneous wind/rain-derived (SWR) rain thresholding scheme is designed to produces a constant false alarm (false rain detection) rate for all wind

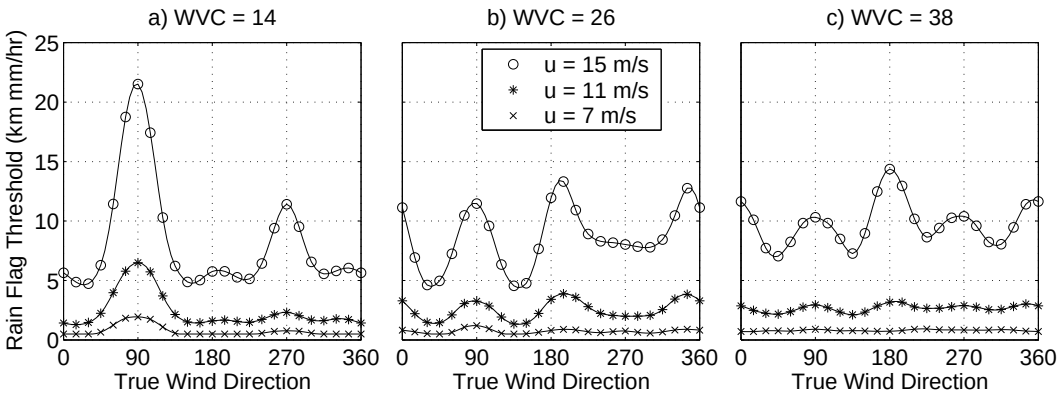


Figure 7.7: Rain rate thresholds that yield a constant false alarm rate given wind speed, wind direction, and cross-track position.

conditions and cross-track positions, at the risk of missed detections in some wind/rain regimes. The SWR method can be used to improve the current Multi-dimensional Histogram (MUDH) QuikSCAT rain flag by reducing the false alarms and expanding the region of detection.

Using the simulation dataset, a rain rate threshold above which we can identify true rain events is chosen for each wind speed, wind direction and cross track position. The threshold is chosen as the rain rate above which the simultaneous wind/rain retrieval estimates rain 5% of the time given zero-rain conditions. This is equivalent to setting a constant false alarm rate of 5% for all wind/cross track conditions. The thresholds are smoothed and interpolated onto a fine grid. In addition, a minimum threshold of 0.5 km mm/hr is set to aid in robustness. We note that 5% is a rather high false alarm rate; however, approximately 30% of false alarms occur beneath the 0.5 km mm/hr threshold, and are thus eliminated.

Ideally, rain rate thresholding uses the "true" wind velocity. Since using the retrieved wind speeds and directions is problematic during rain events, winds from the National Centers for Environmental Prediction (NCEP), available in the QuikSCAT Level 2B data, are used to index the thresholds. The use of NCEP winds to index the thresholds introduces some variability into the rain flagging process,

Table 7.1: Percentage of rain detection for SWR rain flag, MUDH rain flag, and combined rain flag for various PR rain rate (R_{PR}, km mm/hr) bins over the co-located TRMM/QuikSCAT dataset.

R_{PR} bin	0 to 0.1	0.1 to 0.5	0.5 to 2	2 to 5	5 to 10	10 ↑
tot. num.	16007	3226	2266	1137	523	433
SWR	1.7%	9.6%	39.7%	67.5%	85.3%	95.2%
MUDH	2.1%	9.2%	31.7%	70.9%	95.0%	99.5%
comb.	1.8%	10.2%	41.7%	73.7%	92.2%	97.2%

since NCEP winds at times may not accurately represent the true wind. However, in most conditions, especially when there are no fine scale wind features, NCEP winds are adequate.

The SWR rain rate thresholds for various wind speeds, wind directions and cross-track positions are given in Figure 7.7. The low threshold values for $u = 7$ m/s and $u = 11$ m/s illustrate that the SWR retrieval does not significantly confuse rain and wind for low to moderate wind speeds. For higher winds, the rain flag thresholds are set high in the cross-swath directions because high false rain rates may occur there.

In order to better flag rain in areas of high SWR thresholds where the retrieved rain rate is questionable, we describe a simple scheme to combine the SWR rain flag with the current MUDH flag to improve overall performance of both methods. If the retrieved rain falls above the SWR threshold, then it is flagged. However, if the SWR rain flag greater than 2 km mm/hr (the threshold value to which the MUDH rain flag is tuned), and the retrieved rain rate is between 2 km mm/hr and the SWR threshold, then the MUDH flag is used. This scheme allows the use of the MUDH rain flag when the SWR rain rate is questionable.

To validate the SWR and combined rain flags, we use 3 months of co-located TRMM PR data. Table 7.1 shows the percentages of rain detections using the SWR, MUDH, and combined rain flags for various TRMM PR effective average rain rate ranges. The effective average rate is formed by antenna-weighted averaging

Table 7.2: Percentage of WVCs flagged as containing rain, and not containing rain for the MUDH and combined rain flags for 4 months of SeaWinds on QuikSCAT data.

		MUDH		
		No Rain	Rain	Total
Comb.	No Rain	88.0%	2.0%	90%
	Rain	5.3%	4.7%	10%
	Total	93.3%	6.7 %	

of all PR rain rates within each SeaWinds 6 dB contour for each measurement, and then averaging the resulting rain rates corresponding to each WVC.

The combined and SWR and combined rain flags infer less rain than the MUDH rain flag in the 0 to 0.1 bin, indicating that the use of the SWR rain flag yields less false alarms than the MUDH flag. For PR rain rates from 0.1 to 5 km mm/hr, the combined rain flag infers a higher percentage of rain than the MUDH flag, illustrating an improvement in rain flagging in low to moderate rain rates. The MUDH flag, however is slightly more effective in higher rain areas (above 5 km mm/hr).

Further, we compare the combined rain flag to the MUDH rain flag over 4 months of QuikSCAT data from August to September 1999. Table 7.2 gives the total numbers of WVCs flagged by rain for both MUDH and combined rain flags. The combined rain flag identifies a considerably higher percentage of rain cases than the MUDH flag. This is partly due to the fact that the combined rain flag is more sensitive to lower rain rates than the MUDH flag in many areas. However, the SWR rain flag thresholds used in the combined rain flag can be adjusted to give a similar detection rate as the MUDH flag if desired. A comparison of the combined rain flag to the MUDH flag is shown in Figure 7.8 as a function of cross track position. The combined rain flag has rather constant performance in the sweet spots, while flags more WVCS at nadir and on the edges of the inner-beam region. The MUDH rain flag has a triangle-shaped flagging rate which peaks in the center of the swath. The MUDH flag also has increased flagging on the outer-beam swath edges where the QuikSCAT flag is not usable.

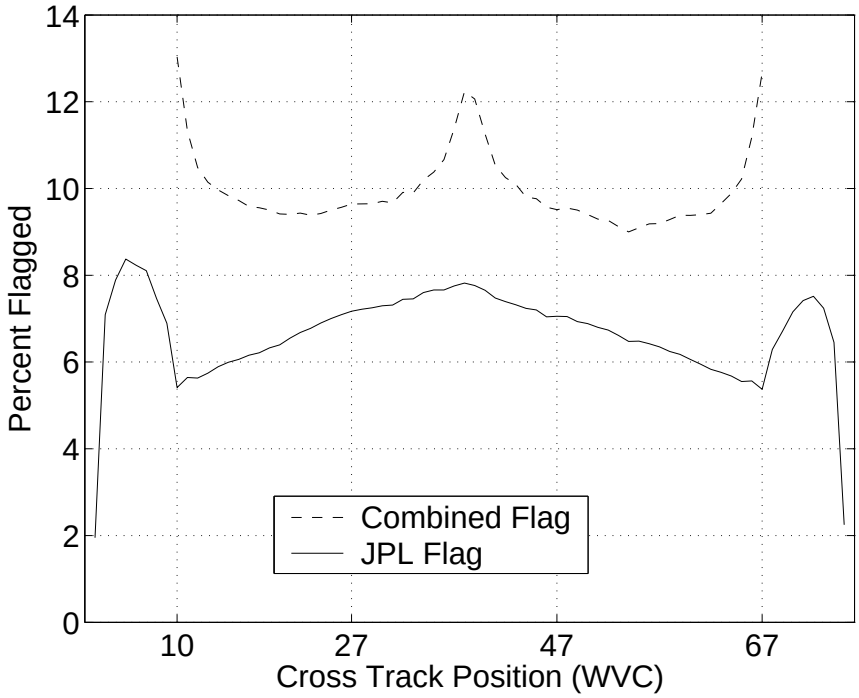


Figure 7.8: Percentage of WVCs flagged by the MUDH rain flag versus the combined rain flag per cross-track position.

7.4 Monthly rain rate validation with TMI

Simultaneous wind/rain retrieval from SeaWinds data, along with the combined rain flag, afford reasonably accurate global observations of ocean rain. Here, we present a comparison of monthly rainfall averages between the TMI and QuikSCAT. In performing the comparison, simultaneous wind/rain retrieval is performed for 4 months of QuikSCAT data from August 99 to November 99. The combined rain flag is applied to the retrieved winds, and the rain rates corresponding to non-flagged WVCs are set to zero. The QuikSCAT retrieved rain rates and TMI surface rain rates are then gridded and averaged separately on a $1^\circ \times 1^\circ$ latitude-longitude grid.

The TMI surface rain rates give an estimate of the instantaneous rain rate in mm/hr. To convert to similar units as QuikSCAT, the surface rain rate is multiplied by a monthly average storm height (obtained from the $5^\circ \times 5^\circ$ PR 3A25

Table 7.3: Mean bias and correlation coefficients for monthly averages of QuikSCAT-derived rain rates and TMI rain rates. A positive mean bias indicates larger average QuikSCAT rain rates. A negative bias indicates larger TMI rain rates.

	08-99	09-99	10-99	11-99
mean bias (km mm/hr)	0.0835	0.0819	0.02	-0.008
corr. coef.	0.5181	0.4768	0.5641	0.4963

monthly average data product, and interpolated to the $1^\circ \times 1^\circ$ grid) to yield an effective integrated average rain rate with units km mm/hr.

An example comparison of the QuikSCAT/TMI monthly average rain rates for August 1999 is given in Figures 7.9 and 7.10. In this example, several global rain features are evident in both the TMI and QuikSCAT rain data. The major noticeable feature is the increased rain activity in the Inter-Tropical Convergence Zone that stretches across the equator. Also, there is a noticeable correlation in the Southern Pacific Convergence Zone located between -40° and -20° latitude and 150° and 270° longitude.

Table 7.3 gives the mean bias statistics (QuikSCAT-TMI) and correlation coefficients for four months of data. Overall, the QuikSCAT data is biased slightly high for three of the months. The bias for August and September is about 15% of the mean TMI rain rate. This bias is comparable to the bias first reported in monthly comparisons between Special Sensor Microwave / Imager (SSM/I) and TMI [60]. The bias is expected from the previous analysis. Overall, the correlation coefficients for the QuikSCAT and TMI data are around 0.5 for the monthly averages. The correlation is likely low due to the noisy nature of the monthly averages.

Examining the monthly error map (see Figure 7.10b), the TMI data is somewhat higher than QuikSCAT (white pixels) in areas of high monthly storm height, indicating that the PR storm height estimate gives somewhat augmented TMI integrated rain averages in some areas. Noticeable augmented storm heights occur around the northern end of South America, in the Southern Pacific, and just south of Japan.

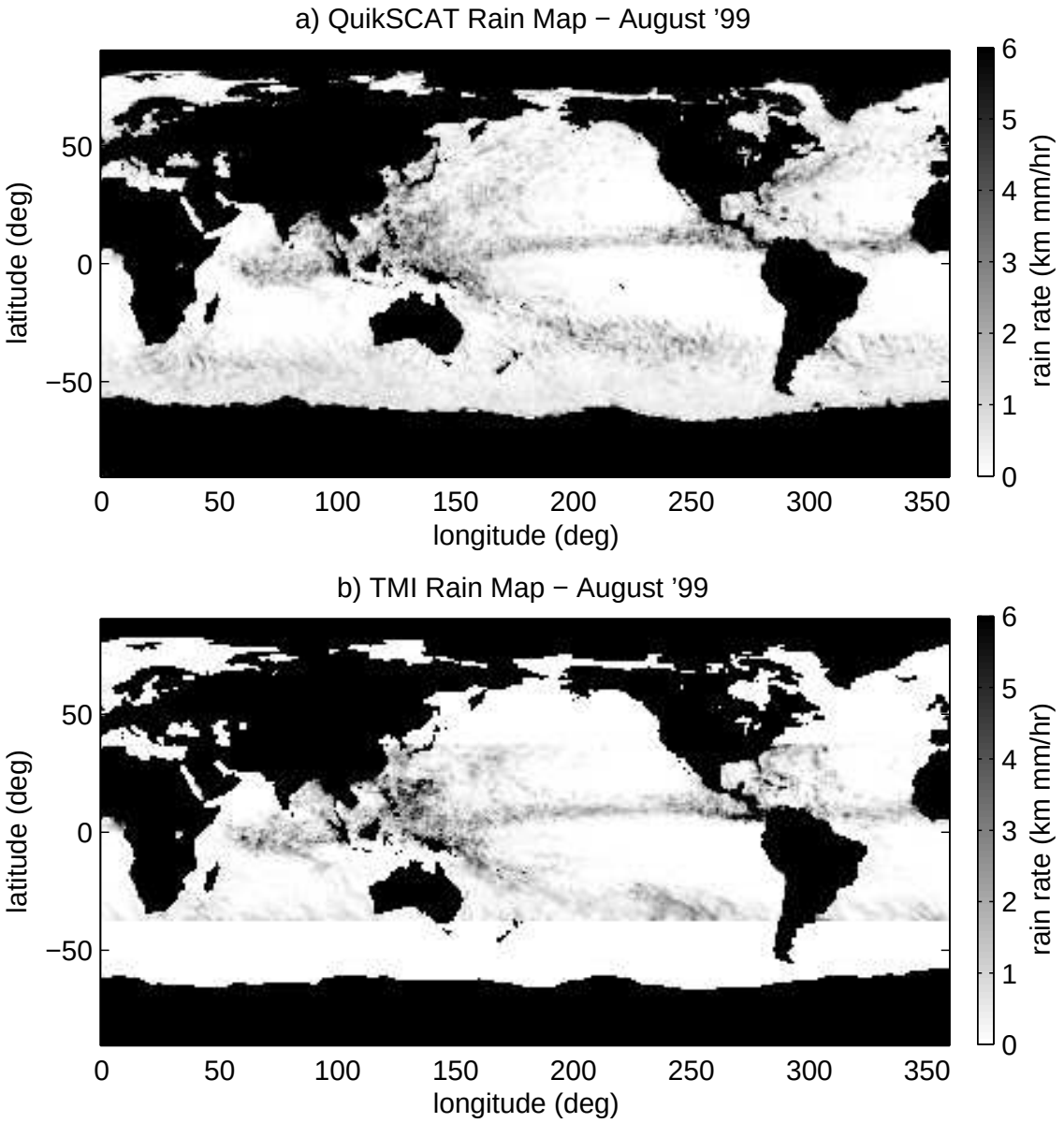


Figure 7.9: Monthly rain rate average for a) QuikSCAT-derived rain rates using the combined rain flag and b) TMI-derived rain rates.

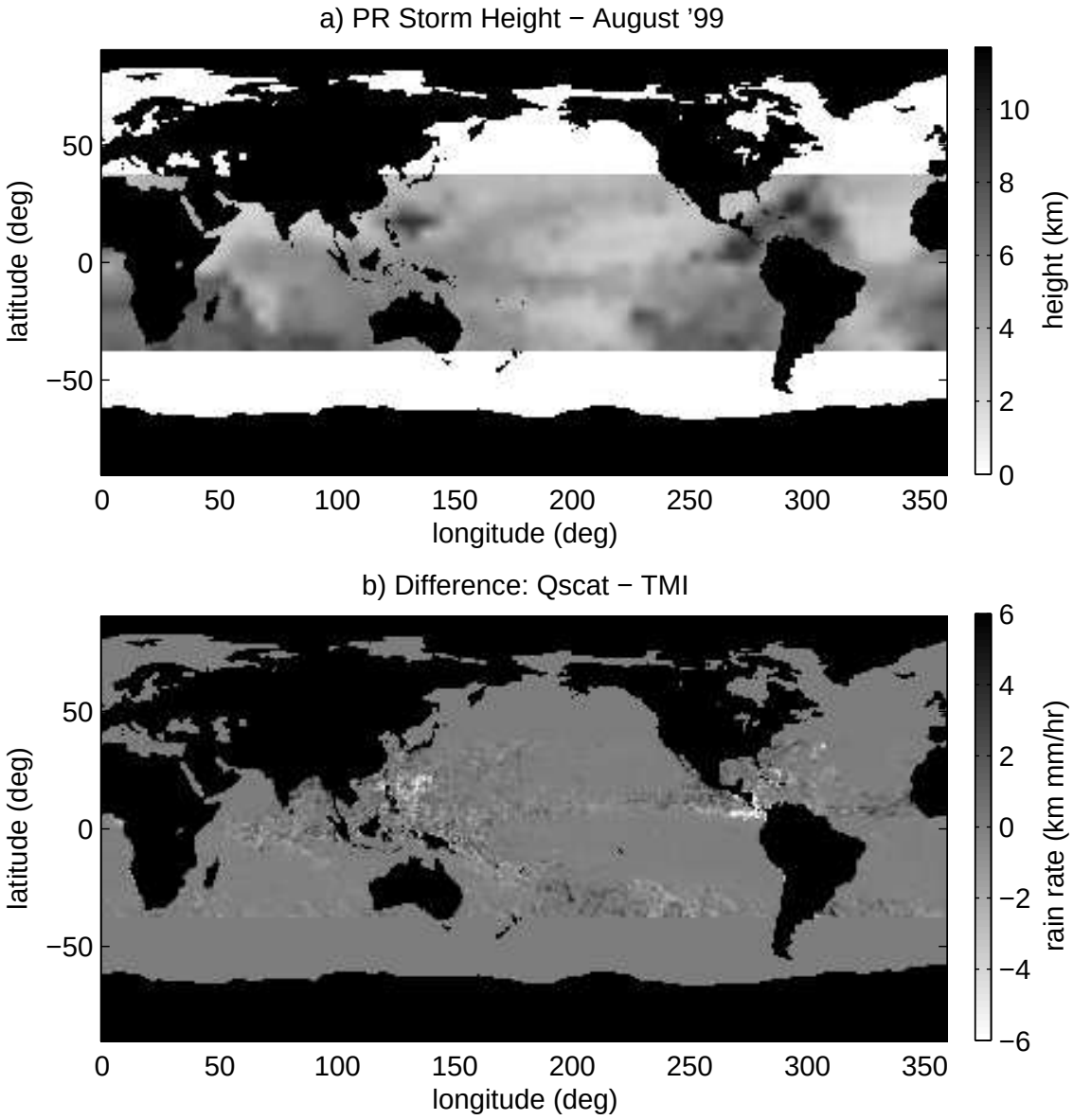


Figure 7.10: a) Monthly rain height average from TRMM PR, interpolated to $1^\circ \times 1^\circ$ grid; b) Error between QuikSCAT and TMI monthly average rain rate for August 1999.

Another noticeable difference between the TMI and QuikSCAT monthly rain map is the noisy nature of the QuikSCAT rain rates in the Southern Pacific Convergence Zone, whereas the TMI data is much smoother. Part of the difference is due to the fact that the TMI data has much higher coverage of this area. For the one month dataset, the TMI data has over 2000 data points per pixel in the southern tropics, while the QuikSCAT data has less than 1000 data points per pixel. Thus, the higher coverage allows for a more consistent time average of the region. In addition, the TMI has a smaller footprint, and thus beam-filling variability is less of an issue.

Validation of QuikSCAT rain rates with TMI illustrates the utility of using Ku-band scatterometer data from instruments such as SeaWinds to measure rain. In addition, SeaWinds affords further active rain measurement to supplement the data obtained by other active instruments such as the PR.

7.5 Conclusions

Although designed to measure near-surface winds, the SeaWinds instrument is also sensitive to rain and thus can be used for global rain measurement. Using the simultaneous wind/rain retrieval method described in [56], rain rate estimates consistent with passive and active instruments such as the PR and TMI can be obtained.

The major issue addressed in this chapter is rain/wind identifiability problems in the measurements, which is especially evident when the wind is oriented cross-swath. In such orientations, spurious rain rates often occur and the retrieved rain rate is biased somewhat high. The problem is evidenced in the Cramér-Rao lower bound and in Monte-Carlo simulations. Because the wind and rain signals may not be completely separable in such orientations, a rain rate thresholding scheme is developed in which the thresholds are set higher in regions of poor wind/rain identifiability to reduce the false alarms. The simultaneous wind/rain retrieval and thresholding scheme are combined to improve the current MUDH rain flag. Using co-located TRMM precipitation radar data, the combined rain flag demonstrates improved results for low to moderate rain rates. Using the simultaneous wind/rain retrieval technique and

combined rain flag, the QuikSCAT-derived rain rates are shown to be consistent with monthly averages derived from TMI.

Chapter 8

Summary and Conclusions

This dissertation addresses several error sources in Ku-band scatterometry, and presents methods for improving ambiguity selection performance and data quality in the presence of rain without the use of outside information. Chapter 1 provides background in scatterometry and describes the current SeaWinds scatterometer design. A quality assurance (QA) analysis technique is described in Chapter 2, evaluating the effects of ambiguity selection, low winds, and rain contamination on SeaWinds data. Chapter 3 describes an advanced ambiguity selection algorithm that does not require nudging. Chapter 4 gives an analysis of the effect of the low wind speed threshold in scatterometer data in an uncontrolled marine environment.

Chapters 5, 6, and 7, present, evaluate and validate a method of simultaneously retrieving wind and rain from scatterometer data. In Chapter 5, a forward model of the rain-induced backscatter and attenuation as seen by the SeaWinds instrument is computed from comparison with co-located TRMM precipitation radar data. The model is found to be accurate to within 3 dB. Using the rain model, the simultaneous wind/rain retrieval method is developed in Chapter 6, demonstrating that rain can be retrieved from SeaWinds data alone. The accuracy of the retrieved winds is evaluated via the Cramer-Rao bound in Chapter 7, and a rain rate flag is presented that indicates when simultaneous wind/rain retrieved rain rates are valid.

The main conclusion of the dissertation is that although ambiguity selection and rain contamination contribute error to the wind retrieval process, the error due to both factors can be reduced without employing outside data. These results are a significant contribution to scatterometry. The following sections provide further contributions made by the dissertation.

8.1 Contributions

8.1.1 Impact of error sources on vector wind

The quality assurance analysis presented in Chapter 2 and published in [36] gives a self-contained method of determining the overall consistency of the wind and identifying ambiguity selection errors. The method expands previous work applied to NSCAT [21]. Although no truth data is available, the chapter describes how to use a low-order model fit to estimate the underlying wind flow, and thus determine the noise level and ambiguity selection. Scatterometer error is shown to be correlated with the ambiguity selection process, low velocity winds, measurement geometry, fine-scale wind features (such as cyclonic storms), and rain. Nadir winds tend to be much noisier than winds in the other parts of the swath, resulting in "poor" wind flow in about 25% of the data. Also, low-velocity winds are shown to be extremely noisy, suggesting that retrieving self-consistent direction flow on the scale of scatterometer winds is not attainable with point-wise retrieval methods. SeaWinds ambiguity selection is demonstrated to be about 95% effective over a two-year data set. The ambiguity selection accuracy varies with latitude and time of year. Comparisons with rain-flagged data and number of cyclonic storms indicates a high correlation between ambiguity selection errors, storms, and rain.

8.1.2 High quality ambiguity selection without nudging

One of the inherent problems with the scatterometer wind retrieval is the fact that a unique wind vector solution is generally not identifiable from the noisy measurements at a single point. Thus, there are typically several possible wind vector choices, known as ambiguities, at each point that could have given rise to the measurements. Numerical weather prediction winds are typically used to initiate ambiguity selection to yield a self-consistent vector field. This technique, known as nudging, creates a dependence on the accuracy of outside information. In Chapter 3, I present a non-nudging approach that provides improved ambiguity selection in non-extreme winds. These results are published in [35], and indicate that high quality

ambiguity selection can generally be performed without nudging by taking into account the low-order variability of the wind via the Karhunen-Loeve wind model. The new ambiguity selection method selects 93% of the same ambiguities as the nudging approach, simultaneously validating both methods.

8.1.3 The effect of a low wind speed threshold on scatterometer data

In low wind speed conditions, scatterometer winds are much noisier due to a low SNR and the possible effects of a low wind speed threshold in backscatter. The impact of the low wind speed threshold on scatterometer data in an uncontrolled environment has been uncertain. To afford better understanding, I perform a simple low wind speed threshold analysis with tower-mounted scatterometer data in Chapter 4. The instrument measured the backscatter at a range of frequencies and incidence angles, providing observation of the ocean backscatter from a number of Bragg wavelengths. With the tower-mounted scatterometer data, I show that although a low-wind speed threshold is detectable in small-footprint scatterometer data, the backscatter does not "go to zero" as theory suggests. Below the threshold, there is generally a mean drop in backscatter with increased variability. The measured threshold wind speeds correlate well with theory, but the mean drop is usually not more than a few decibels. The increased variability below the threshold may be related to a hysteresis effect and is consistent with space-borne scatterometer observations, suggesting that increasing wind retrieval accuracy below the threshold may not be possible. Space-borne scatterometers, however, observe the ocean with much larger footprints, and are thus not as sensitive to the threshold wind speed as small-foot print instruments. This work has been published as a conference paper (see [61]).

8.1.4 The effect of rain on the SeaWinds scatterometer

Rain is the one of the largest weather-related contributers to scatterometer error in SeaWinds data. Understanding the effect of rain on SeaWinds measurements is imperative to correct rain-contaminated data. Using co-located TRMM precipitation radar, I demonstrate that the effects of rain can be characterized using a very

simple model (to within 3 dB) in Chapter 5. The analysis suggests that scattering from surface perturbations dominates the rain-induced signal for low rain rates. At high rain rates, atmospheric volume rain scattering dominates. Thus, theoretical models must take into account not only volume scattering from the atmosphere, but also surface backscatter to give accurate estimates of the bulk backscatter due to rain. Using the simple backscatter model, I illustrate the regimes in which rain and wind may be estimated from backscatter data in rain. This work has been accepted for publication in the *Journal of Geophysical Research* [53].

8.1.5 Rain measurement from SeaWinds

Although SeaWinds is designed to measure wind, I demonstrate that SeaWinds' sensitivity to rain affords scatterometer-based rain retrieval. Simultaneous wind/rain retrieval is possible using the simple backscatter model presented in Chapter 5 in Chapter 6. The wind/rain model is also useful for possible correction of rain contaminated data from SeaWinds using rain rates obtained from the AMSR aboard ADEOS II. The simultaneous method gives consistent results when compared with the TRMM precipitation radar data, although the wind retrieval quality is somewhat compromised in no-rain conditions, especially for cross-swath blowing winds. Simultaneous wind/rain retrieval reduces scatterometer speed errors in rain events, and provides an auxiliary source for global rain rate estimates. This research is soon to be submitted to the *Journal of Geophysical Research*.

8.1.6 Improvement of current rain flagging

Because SeaWinds was not designed to measure rain, the simultaneous wind/rain retrieval procedure exhibits some limitations, explored further in Chapter 7. The limitations are associated with accurate retrieval of cross-swath winds. Because rain is an isotropic scatterer, and cross-swath blowing winds "appear" isotropic to the scatterometer, the retrieval process may confuse wind and rain, yielding false rain detections, and somewhat degraded wind estimates. This phenomenon is studied via simulation and the Cramér-Rao lower bound on measurement accuracy. Using the simulation data, a set of thresholds are determined which indicates if the estimated

rain rate is believable based on the wind speed, direction and cross track position. Using this rain thresholding scheme, the current multi-dimensional histogram flag can be improved by flagging wind vector estimates whose estimated rain rates fall above the thresholds. The thresholded rain rates are shown to correlate well with monthly averages from the TRMM microwave radiometer. This research is soon to be submitted to the *Journal of Geophysical Research*.

8.2 Future research

8.2.1 Improvement of Non-nudging Ambiguity Selection in Storms

The ambiguity selection method introduced in Chapter 3 is slightly more error-prone in cyclonic storm areas than the current nudging approach. Further research is needed to tune the algorithm for optimal performance in cyclonic storm areas.

8.2.2 Application of rain model to SeaWinds on ADEOS II

The ADEOS II satellite, which houses the second SeaWinds instrument, has a high-precision radiometer known as the Advanced Microwave Scanning Radiometer (AMSR) which will give co-located rain measurements with the scatterometer. These rain measurements can be used to provide rain-corrected wind vector estimates. The method suggested in Chapter 6 can be used to incorporate the AMSR data. The rain model, however, must be tuned to the AMSR rain data, and validation must be performed when AMSR data becomes available.

8.2.3 High resolution rain retrieval

In addition to the 25-km rain product, rain retrieval may also be performed at a higher resolution using the method introduced by *Long* [62]. Because beam-filling is less of an issue at higher resolutions, SeaWinds rain retrieval may operate at approximately the same accuracy as at 25 km resolution. The high-resolution processing can give rain estimates with resolution on the order of the TRMM precipitation radar. Such data would be ideal for hurricane observations of the interplay between rain and wind.

8.2.4 Combined model-based/rain estimation

As illustrated in Chapter 6, simultaneous-wind/rain-retrieved wind speeds and directions during rain events are much more highly variable than in retrieved wind speeds and directions in non-raining conditions. In order to provide a more consistent estimate of the wind, model-based wind retrieval with a rain correction may be employed in rain-corrupted areas. The result would be improved wind retrieval in raining areas at a low computational cost compared with full model-based wind retrieval.

APPENDICES

Appendix A

Fourier Analysis and Model Size Trade-offs of the QuikSCAT KL Model

Because the truncated KL model minimizes the basis restriction error, it is well suited for reduced order modeling of wind fields. This appendix explores the issues involved with adapting the Karhunen-Loeve (KL) wind model to QuikSCAT vector winds for use in model-based retrieval, point-wise ambiguity selection, and quality assurance analysis. The KL model originally designed for NSCAT was a 12 $\times$ 12 KL model truncated at 22 model parameters. Differences in swath size and resolution require a reevaluation of the KL model for use with QuikSCAT and future sensors. Using Fourier analysis techniques, I discusses the issues associated with model truncation point and region size. To explore the advantages of different sized models, I analyze six model sizes including 8 $\times$ 8, 12 $\times$ 12, 16 $\times$ 16, 20 $\times$ 20, 24 $\times$ 24, and 38 $\times$ 38 WVC models.

A.1 Generating a KL wind field model

Generically, a KL model is formed by taking the singular value decomposition of an autocorrelation matrix. To form the model, a subset of QuikSCAT swaths are divided into $N \times N$ regions. No regions are used that contain missing data points. The rectangular (u and v) components for each $N \times N$ region are column scanned to create a $2 \times N^2$ column vector, $\mathbf{w}_n$. The first N^2 vectors contain the column scanned u components, and the last N^2 vectors contain the column scanned v components. The autocorrelation matrix R is estimated:

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

where M is the number of $N \times N$ wind fields examined. The basis set S is extracted by taking the singular value (Σ) decomposition of $\hat{R}$ where

$$\hat{R} = S^T \Sigma S. \tag{A.2}$$

Note that the singular value decomposition is equivalent to the eigenvalue decomposition because $\hat{R}$ is symmetric. The columns of S become the basis fields, or model parameters of the KL model. The basis set is truncated to the appropriate number of basis vectors, represented by F.

An arbitrary wind field can be represented exactly by a linear combination the basis vectors. When restricted, the KL basis set identifies the underlying flow of the wind, reducing noise at the risk of smoothing some fine scale wind features.

A.2 Frequency spectrum of the KL wind model

Since wind has a "red" spectrum, the low frequency wind has the most energy. As a result, the KL model places low frequency parameters first. This phenomenon is demonstrated in Figure A.1. The frequency spectrum of a 16 $\times$ 16 KL model is estimated by taking the Fourier Transform (FFT) of the u component of each model parameter. For the first 150 model parameters, energy is concentrated at lower frequencies. As the number of model parameters increase, the frequency spectrum widens monotonically. When truncated at a low number of parameters, the model acts as a low-pass filter. Because noise is generally a "white" spectrum, most of the noise is contained in the higher order model parameters.

A.2.1 Model size comparison

While the NSCAT KL model was fixed at 12 $\times$ 12 WVCs, the wide QuikSCAT swath affords the flexibility of choosing a more optimally sized model depending on the application. In order to directly compare different sized models, I evaluate them on the same grid, requiring resizing of the KL model to a common region width. Resizing the KL model involves interpolation of the basis wind fields via zero padding in the frequency domain. In doing so, edge discontinuities are avoided

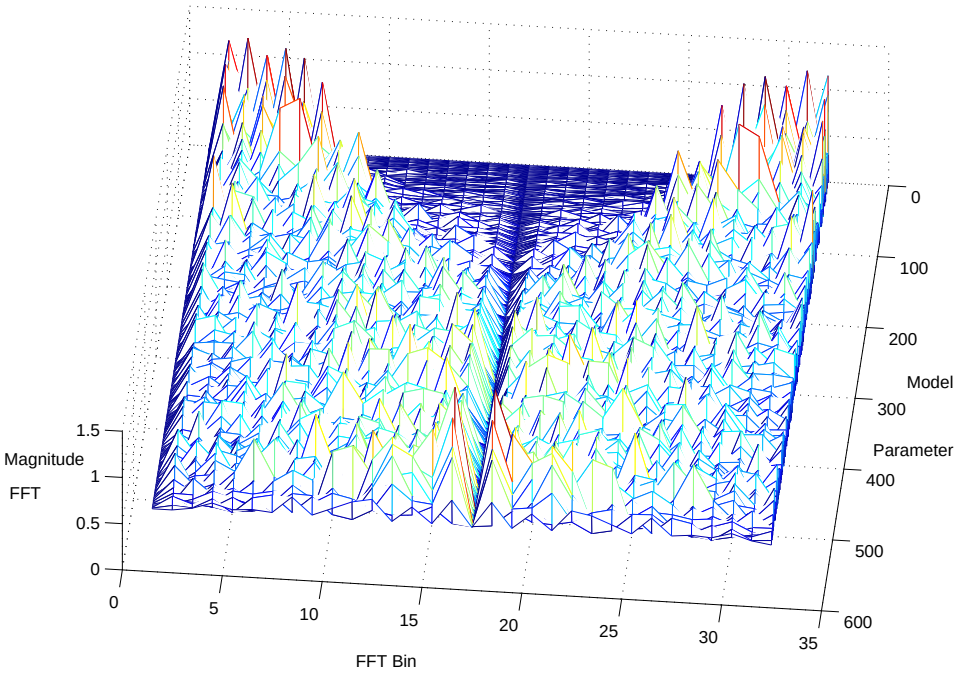


Figure A.1: The x-axis is the one dimensional frequency spectrum of the QuikSCAT 16×16 KL model for the u component. The y-axis is the model parameter number. As the model parameter increases, the frequency spectrum widens.

using conjugate mirroring of the wind field before applying the Fast Fourier Transform (FFT).

Although different sized models are not identical, they contain similar frequency information in the low-order vectors. To demonstrate this, two models A and B are compared by creating a least-squares fit of each parameter in model A using the truncated model B. This analysis is similar to the basis comparison metric given in Chapter 2.

A least squares estimate of the n^{th} model parameter of truncated basis A, $\mathbf{a}_n$, is obtained using the pseudo-inverse of the truncated basis B, $B^{\dagger}$, i.e.,

$$\hat{\mathbf{x}}_n = B^{\dagger}\mathbf{a}_n \tag{A.3}$$

$$\hat{\mathbf{a}}_n = B\hat{\mathbf{x}}_n. \tag{A.4}$$

The coefficients, $\hat{\mathbf{x}}_n$, are examined to see which parameters of model B are used to model the parameters of model A.

Table A.1 shows the groupings of the NSCAT 12 $\times$ 12 WVC model that are significantly used when applied via the least-squares fit to each of th first 22 basis vectors of the QuikSCAT 24 $\times$ 24 model. The 24 $\times$ 24 QuikSCAT model was chosen because it represents the same total area as the NSCAT 12 $\times$ 12 WVC model. Note that the resolution for NSCAT was 50 km and the resolution for QuikSCAT is 25 km. Thus, resizing the NSCAT model to twice the size makes it the same resolution as the QuikSCAT 24 $\times$ 24 model. The models were truncated at 22 because there were no vectors of the NSCAT model available above 22 model parameters.

Table A.1 demonstrates that up until the last few model parameters (20th in this case), the model parameters for NSCAT and QuikSCAT are basically linear combinations of each other. Another interesting point is that there are similar groupings of alike model parameters between the two models. For example, the first 2 QuikSCAT model parameters are modeled exclusively by the first 2 NSCAT model parameters. This is because the first two model parameters are orthogonal DC components of the wind. There is also another grouping of model parameters 3-6. Each of the first 6 QuikSCAT model parameters can be modeled almost perfectly by the

first six model parameters of the NSCAT model. Other noticeable groupings occur at model parameters 7-12, and 13-22. As discussed in Chapter 2, each grouping of model parameters represents a specific range of frequencies. As the number of model parameter increases, the range of frequencies increase.

In analyzing the 8×8, 12×12, 16×16, 20×20, and 38×38 QuikSCAT models, similar groupings are noticed. Each of these cases reveal that there are groupings of model parameters at 1-2, 3-6, 7-12, and 13- $\sim$22. Also, the NSCAT KL model for each case modeled the QuikSCAT model very well until the last 2-4 model parameters. These results were achieved independent of the original model size. This analysis suggests that for meso-scale winds, dynamic features inherent in the wind are independent of size or resolution. Thus, although an 8×8 WVC region covers much less area than a 24×24 WVC region, it contains similar wind structures, but on a smaller scale.

A.2.2 Frequency groupings of the QuikSCAT KL model

The following analysis of the QuikSCAT KL model provides validation to the observation that KL vectors are grouped into separate ranges of frequencies. The frequency content of the QuikSCAT and NSCAT models is examined in the following way: The wind fields are symmetrically mirrored creating periodicity. The FFT for the u component of the rows is taken and averaged. The FFT is computed and averaged for the v component of the rows, u component of the columns, and v components of the columns. The maximum number of FFT bins used for each of the model parameters is recorded. This entails examining each of the 4 FFTs performed, 2 vertical and 2 horizontal, and taking the highest bin used out of the 4 possibilities. A subjectively chosen threshold is set to locate the highest bin. The threshold used in these calculations ranged from .07 – .15 of the maximum bin magnitude depending on model being used. The highest model parameters that used only the first N frequency bins are listen in Table A.2 for each model examined.

Table A.2 indicates that there exist definite groupings of parameters that use the same frequency bins common to all models. In general, the higher frequency

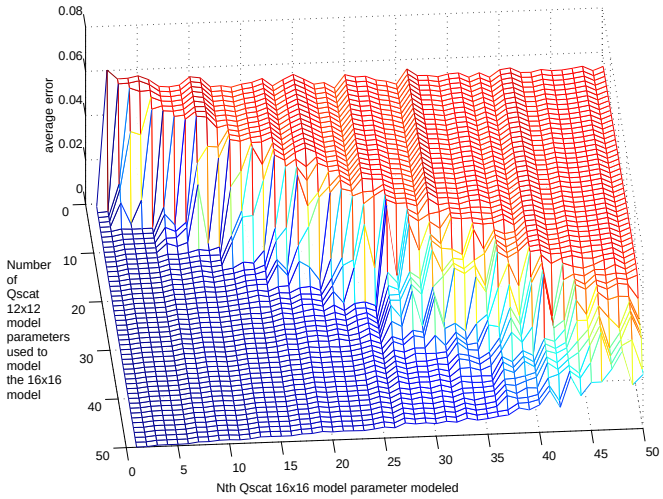


Figure A.2: The average modeling error over the wind field of the QuikSCAT 16 × 16 model being modeled by successive numbers of the enlarged QuikSCAT 12 × 12 model.

groupings become wider as the number of model parameters increase. The results of Table A.2 aid in finding discrete points to best truncate the KL model. If a model is truncated at a point defined in Table A.2, there exist the maximum number of model parameters for a given wind frequency.

An alternate way of showing the results in Table A.2 is to do a model fit of one model with another. Given that truncated model B contains a limited number of spatial frequencies, it should not be able to represent model parameters of model A outside of the frequency grouping due to the orthogonality of the Fourier transform basis. Figure A.2 shows the average component error between the first 50 QuikSCAT 16 × 16 model parameters and model estimates made by the enlarged QuikSCAT 12 × 12 model for increasing numbers of 12x12 model parameters.

As seen in Figure A.2, 12 model parameters from the QuikSCAT 12 × 12 model are sufficient to model anything under 12 of the 16 × 16 model. However, 11 model parameters of the 12 × 12 model do not adequately model the subgrouping from 7 – 12 of the 16 × 16 model. This phenomenon occurs because each of the

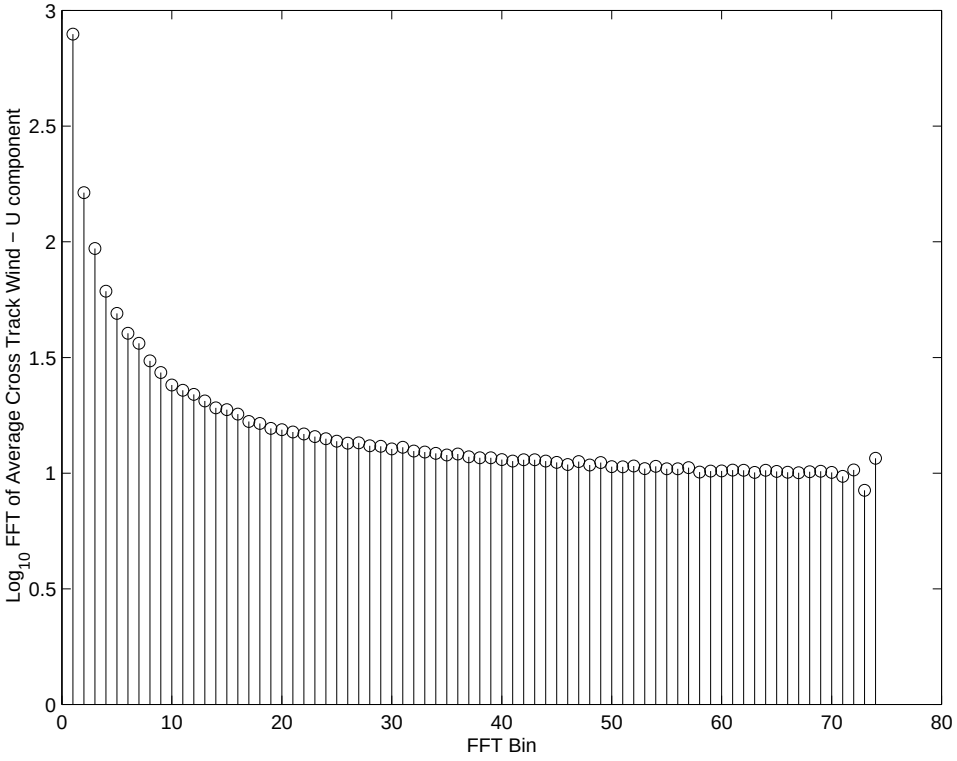


Figure A.3: Average FFT over the cross track of QuikSCAT winds

16×16 model parameters from 7 – 12 contain portions of the frequency content in parameter 12 of the 12×12 model, but do not contain higher components. Additional discontinuities in the modeling error occur at 22, 26, and 36. When an entire frequency grouping is used (e.g. at 12, 26, and 36) the same grouping of another model can be adequately modeled. It is interesting to note that an apparent edge of a grouping in Figure A.2 appears to occur at 16. This may be "sub-grouping" with common frequency components between the 16×16 and 12×12 models.

A.2.3 QuikSCAT point-wise wind spectrum

In deciding reasonable truncation points for the QuikSCAT model, the frequency spectrum of QuikSCAT winds is examined. Using QuikSCAT L2B data, the frequency spectrum is determined by averaging the u- and v-component FFTs for each cross track row containing no data gaps over a set of QuikSCAT winds. In

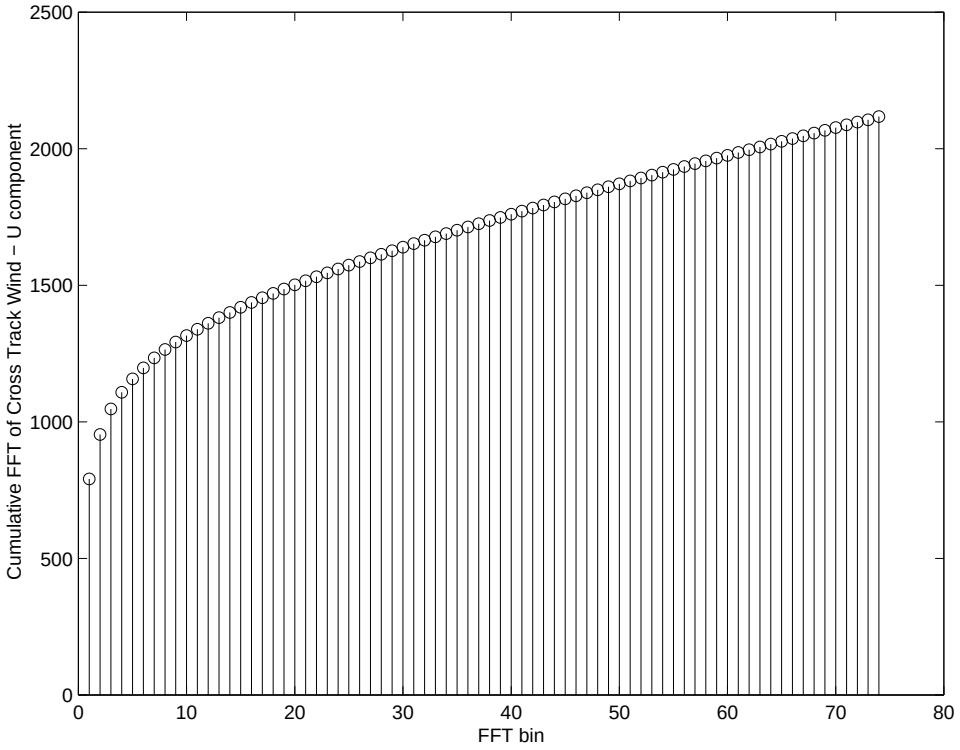


Figure A.4: Running sum of energy in the FFT bins over the cross track

doing this, the data is mirrored for symmetry. Figure A.3 shows the average FFT of the wind over the cross track for 10 revs. As shown, most of the energy in the wind exists at low frequencies. Because of the "red" spectrum of the wind, an ideal wind spectrum should drop to nearly zero at high frequencies. However, the high frequency components saturate due to a "noise floor". Because noise generally has a "white" spectrum, the high frequency component of the wind is likely to be dominated by noise. Thus, appropriately truncating the higher frequencies can reduce the effects noise and not significantly affect the available frequency information in the retrieved winds. Figure A.4 shows a running sum of the total energy in the FFT bins. Notice that the slope becomes constant in the higher frequency areas. This indicates that mostly "white" noise is being summed.

A.3 Determining KL model truncation points

Figures A.3 and A.4 suggest that most of the energy in the wind is contained in the first 1/4 - 1/3 of the frequency content. This section discusses at which model parameters the QuikSCAT models must be truncated in order to keep the lowest 1/4 or 1/3 of the frequency content. In the following discussion, N is the number of bins in the 1/2 the FFT. Note that the wind has been mirrored, creating a vector 2 $\times$ 74 WVCs in length. Thus, N is 74 WVCs. Because the FFT is symmetric, the right-hand side of the FFT has been neglected. Each bin represents a discrete frequency of $\frac{\pi M}{N}$ where M is the bin number. I define T as the sampling period. The real spatial frequency that a bin represents is

$$\frac{M}{N}\frac{\pi}{T}. \tag{A.5}$$

All of the QuikSCAT model parameters have the same sampling period (T = 25 km), making T constant for any size KL model. Given that there are N bins in the FFT and M bins are desired, M/N is the fraction of the total number of bins that will be kept. Equation (A.5) is simply the fraction of bins desired multiplied by a constant. Thus, if the fraction M/N is held constant for any model, it represents the same amount of spatial frequency. For example, if 1/4 of the frequency spectrum was desired for the 16 $\times$ 16 model, N = 16 and M = 4. This is the same spatial frequency as M = 3 for the 12 $\times$ 12 model. Both of these correspond to the same cutoff frequency.

Using this information and the data obtained in Table A.2, QuikSCAT cutoff model parameters can be obtained that correspond to a certain cutoff spatial frequency. These numbers are calculated for each of the models of size 8 $\times$ 8 to 24 $\times$ 24 WVC and tabulated in Table A.3. From Table A.3, it is shown that as the size of the model increases, the number of model parameters needed to obtain the same cutoff frequency also increases. For example, to model 1/3 of the frequency content of a swath, it takes only 12 8 $\times$ 8 model parameters. However, it takes 92 24 $\times$ 24 model parameters. This is such because the 24 $\times$ 24 model covers a larger region. In a larger region, what may look like a uniform wind field for a 8 $\times$ 8 model, actually

has low frequency content that can be modeled by the 24 × 24 model. Thus, the low order wind vectors in larger models model low frequency winds that smaller models cannot see. Table A.3 also shows the smallest wind feature size that can be modeled for each fraction of the FFT. This information was gathered from direct observation of the cutoff model parameters. Notice that using more FFT bins enables the model to fit to smaller wind feature sizes.

A.3.1 Trade-offs between smaller and larger models

The previous discussion affords an understanding of the number of model parameters that must be retained to obtain a certain spacial frequency. The number of parameters to retain is related to the model region size. A trade-off analysis must be made when selecting a model size. Although a large model must be truncated at a high number to obtain the same spatial cutoff frequency as a smaller model, a larger model can more accurately represent large-scale wind flow because the larger model has less error-producing "edges." Processing time, however is drastically increased when a larger model is used. The small model is more computationally efficient and can more easily model ocean regions next to land or non-data points, but introduces higher modeling error. Thus, deciding on a model size and corresponding truncation point involves a trade-off between modeling error and computational efficiency.

A.4 Conclusion

This appendix has shown that there is a direct relationship between the KL model parameter number and the frequency content in the model. The model parameters are grouped according to the amount of frequency content in the wind. It has been shown that for the low order model parameters, all models contain scaled versions of the same wind features. In addition, larger sizes of models contain lower frequency content than smaller models.

For larger models, more model parameters are needed in order to have the same cutoff frequency as the lower models. This creates a trade-off between computational efficiency (with a small model), and low modeling error (with a large model).

Table A.1: Each QuikSCAT model parameter of the 24 × 24 model was modeled by the enlarged NSCAT 12 × 12 WVC model. The model parameters with the greatest energy used by the NSCAT model are enumerated. In addition, the approximate range of NSCAT model parameters used to model the QuikSCAT KL model is shown.

QuikSCAT model par.	NSCAT model parameters used.	Approximate range
1	1-2	1-2
2	1-2	
3	3-6	3-6
4	3-6	
5	3-6	
6	3-6	
7	7-9	7-12
8	7-12	
9	7-13	
10	7-11	
11	7-11	
12	7-12	
13	8-16	13-(22)
14	13-20	
15	13-17	
16	8-17	
17	19-22	
18	14-21	
19	14-21	
20	13-21	
21	Does not model well	
22	Does not model well	

Table A.2: Comparison of frequency bin groupings between different models. The last model parameter that used only the first N bins is listed.

Nth Bin	NSCAT 12 × 12	QSCAT 8 × 8	QSCAT 12 × 12	QSCAT 16 × 16	QSCAT 20 × 20	QSCAT 24 × 24	QSCAT 38 × 38
1	2	2	2	2	2	2	2
2	6	6	6	6	6	6	6
3	12	12	12	12	12	12	12
4		24	26	26	26	26	25
5		32	36	36	40	38	41
6		45	50	53	56	57	52
7		58	65	70	76	78	69
8			75	90	96	92	90

Table A.3: Cutoff model parameter numbers for 8 × 8, 12 × 12, 16 × 16, 20 × 20, and 24 × 24 QuikSCAT models.

QuikSCAT model	Fraction of FFT	Cut off Bin Number	QuikSCAT Cutoff Model Parameter	Smallest Wind Feature Size
8 × 8	1/4	2	6	8 × 8
12 × 12		3	12	
16 × 16		4	26	
20 × 20		5	40	
24 × 24		6	57	
8 × 8	1/3	~ 3	12	6 × 6
12 × 12		4	26	
16 × 16		~ 5	36	
20 × 20		~ 7	76	
24 × 24		8	92	

Appendix B

Selecting Variable WVC Thresholds

This appendix develops the method of determining variable WVC thresholds to detect ambiguity selection errors in the presence of cross track and wind speed dependent noise. In addition to the WVC and region thresholds, we describe several other criteria which must be met to flag a region as a possible ambiguity error. These results are tuned to SeaWinds data, but similar methods may be adapted for use with future instruments.

A selection of 15 revs was manually inspected for ambiguity selection errors in every 8×8 WVC processable region. All regions that exhibited clear ambiguity selection errors were identified. This serves as a *training set* to tune the algorithm.

The following discussion uses the terms *missed detections* and *false alarms*. A missed detection is defined as a region that is manually flagged as an ambiguity selection error, but is not flagged by the algorithm. A false alarm is defined as a region that is flagged by algorithm as an ambiguity selection error, but not manually flagged. The *false alarm rate* or *probability of false alarm* is defined as the total number of false alarms divided by the total number of regions not manually flagged as ambiguity selection errors. The *missed detection rate* or *probability of missed detection* is defined as the total number of missed detections divided by the total number of regions manually flagged as ambiguity selection errors. Because the number of regions not manually flagged as ambiguity selection errors is about 95% of the data, it is valuable to insure a low final false alarm rate. Both direction and vector error thresholds are separately optimized to give a constant false alarm rate of 2.5% for each type of threshold. At this false alarm rate, a region threshold of 14% is found to be optimal.

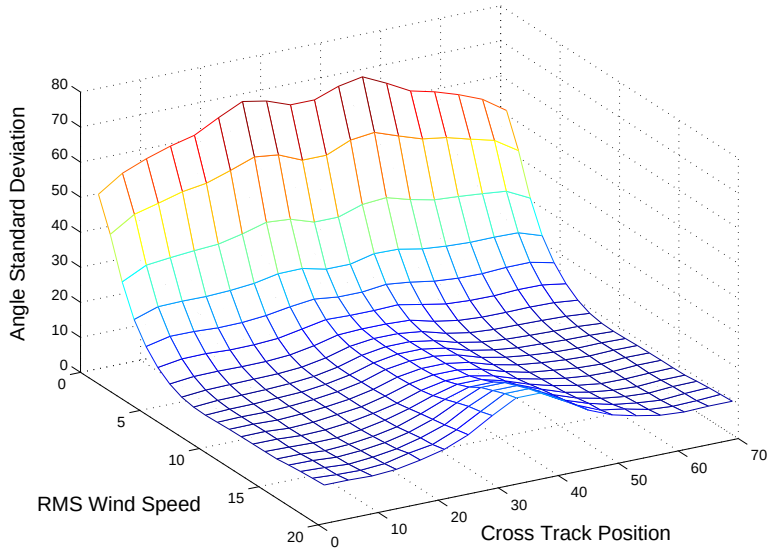


Figure B.1: The average directional standard deviation for both RMS wind speed and cross track position.

B.1 Optimal direction error thresholds

Because it is important to minimize the probability of false alarm, the noise level at each cross track position and for each RMS wind speed must be taken into account. The variability due to noise can be suppressed in the flagging process by setting higher thresholds for statistically noisier regions, thereby equalizing the false alarm rate for all wind speeds and cross track positions.

The standard deviation for each cross track position/RMS wind speed bin averaged over several hundred revs is plotted on a three dimensional grid in Figure B.1. A higher noise variance at nadir and in low wind speed regions make it difficult to set appropriate constant WVC thresholds for locating ambiguity selection problems. Because the quality assurance algorithm is based on flagging individual vectors exceeding a threshold in direction or vector error, using constant thresholds promotes more frequent flagging of high noise regions. This warrants using variable thresholds for different cross track positions and RMS wind speeds.

Assuming that the higher variability in Figure B.1 is caused mostly from noise, setting thresholds higher for regions of a higher estimated directional standard deviation suppresses false alarms due to noise. Thus, an optimally adjusted version of the shape in Figure B.1 should yield a constant false alarm rate for all wind speeds and cross track positions.

To find a set of direction thresholds that gives a constant false alarm rate for all RMS wind speeds and cross track positions, the following approach is implemented:

1. Using the training data set in which ambiguity selection errors have been manually identified, all 8×8 WVC regions are binned according to RMS wind speed and cross track position.

2. The initial WVC threshold is assigned to be the lowest expected threshold value.

3. The observed wind is compared to the model fit and regions are flagged as possible ambiguity selection errors according to the number of poor cells per region (in this case, 14%).

4. The number of false alarms and missed detections is calculated for each bin.

5. If the false alarm rate for a bin is above a certain limit, the WVC threshold for that bin is raised.

6. If the missed detection rate for a bin is greater than zero, and the false alarm rate is significantly smaller than the desired false alarm rate, the WVC threshold is lowered.

7. The number of false alarms and missed detections is recomputed for each cross track and RMS wind speed bin.

8. This process is iterated until either the false alarm rate for all the bins or the average false alarm rate falls beneath a desired threshold.

The threshold determination method was applied to a set of 15 test revs. Figure B.2 shows the resulting set of direction error thresholds. The general shape

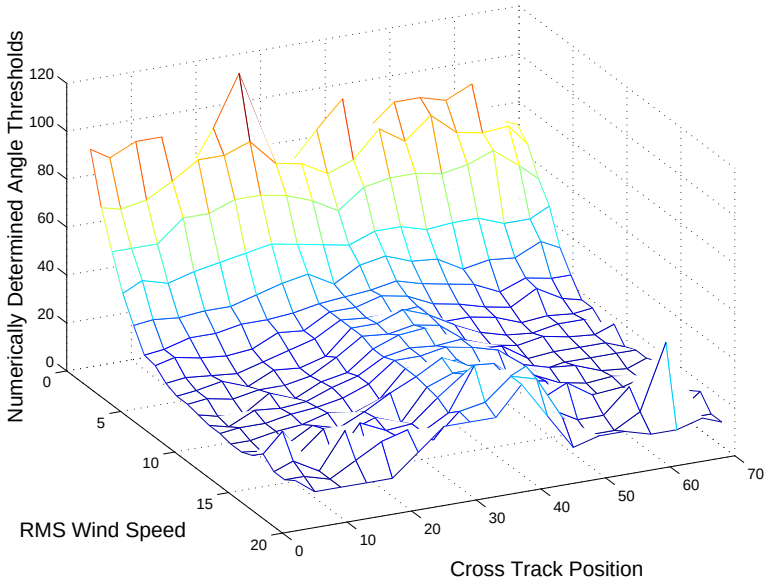


Figure B.2: The direction thresholds per cross track and RMS wind speed that minimizes the probability of false alarm beneath a threshold of 2.5%.

mirrors the directional standard deviations shown in Figure B.1, verifying our assumption.

The thresholds in Figure B.2 are very noisy because of the limited data set used to compute them. However, a good approximation to "smoothed" thresholds is formed by fitting the curve from Figure B.1 to match the values of Figure B.2. The following algorithm accomplishes this in a least-squares sense:

1. For each row of constant RMS wind speed, the values from Figure B.1 and a uniform vector the same length are selected to be "model parameters" for a cross track row from Figure B.2. These two vectors form the columns of matrix T.

2. A least squares fit using these two model parameters is made to the corresponding row of numerically determined threshold values, $\mathbf{r}_0$ by taking the pseudo-inverse of the matrix T, i.e. $\mathbf{x} = T^{\dagger}\mathbf{r_0}$.

3. The modeled row is then $\mathbf{r} = T\mathbf{x}$.

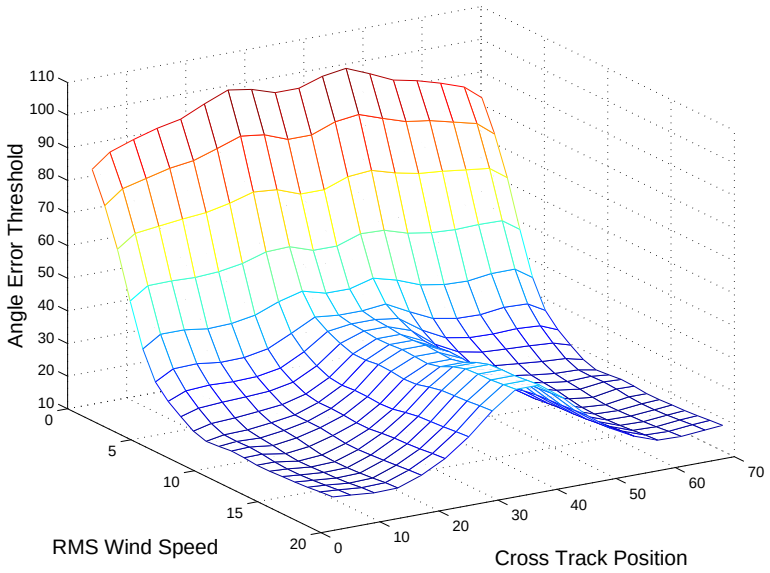


Figure B.3: The direction error thresholds per cross track and RMS wind speed that minimize the probability of false alarm beneath a threshold of 2.5%.

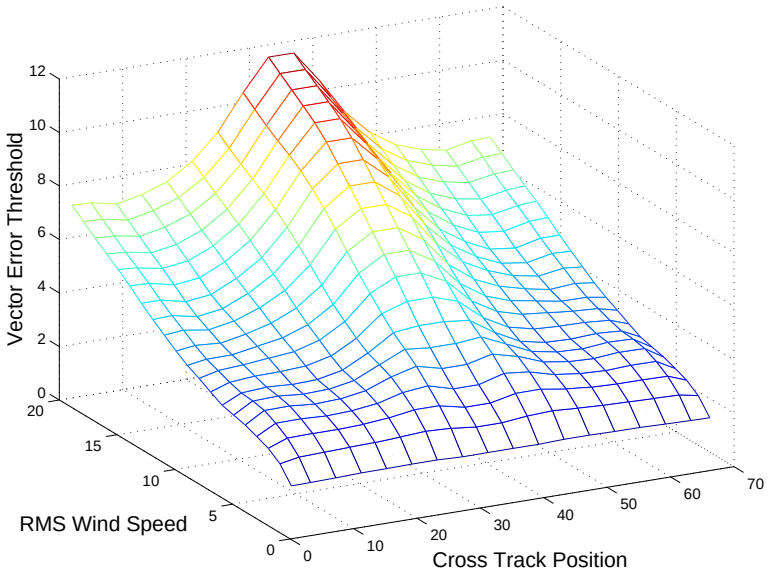


Figure B.4: The vector error thresholds per cross track and RMS wind speed that minimize the probability of false alarm beneath a threshold of 2.5%.

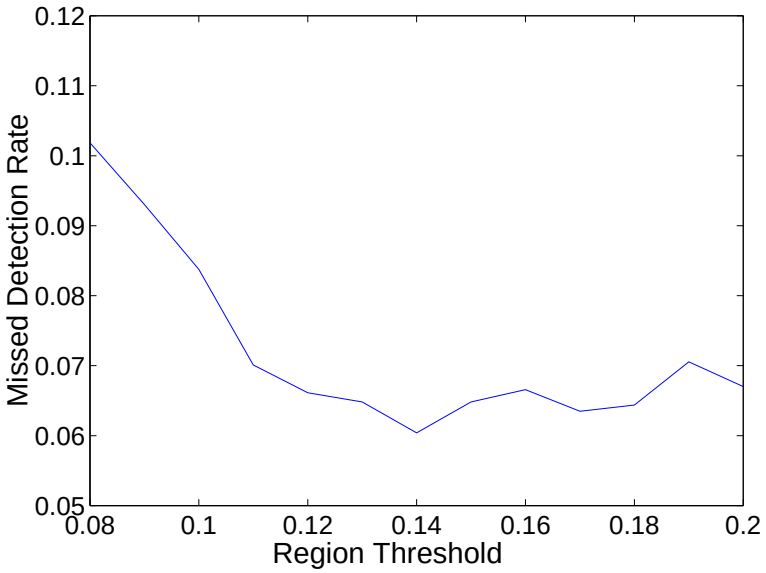


Figure B.5: Probability of missed detection with optimized direction and vector thresholds for different region thresholds.

4. Each of the columns corresponding to constant cross track position is low passed filtered to smooth out any obvious anomalies.

5. Extra adjustments are made manually to improve performance. These include adjusting the thresholds for low (< 4 m/s) and high (> 15 m/s) RMS wind speeds to subjectively give better performance. Insufficient data for these regions warrant manual adjustment.

This method creates "smoothed" WVC thresholds which approximate the WVC thresholds that give a constant false alarm rate for all cross track position and RMS wind speeds on the training data set. The final direction error thresholds are shown in Figure B.3.

B.2 Optimal vector error thresholds

A set of vector error thresholds is also determined which equalizes the false alarm rate for both RMS wind speed and cross track position. These thresholds are

determined in the same way as the direction error thresholds. Figure B.4 shows the imperially determined and "smoothed" vector error thresholds.

In order to classify a region as an ambiguity selection error, the total number of WVCs flagged must exceed a threshold of 14%. To select this threshold, we analyzed the missed detection rate generated by performing the WVC-threshold optimization algorithm with different region thresholds (Figure B.5). As shown, the threshold that results in the minimum missed detections is 14%. Thus, this threshold is used in the algorithm.

Appendix C

Ambiguity Selection Error Detection Consistency Checks

C.1 Model based consistency check

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.

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 B.3.

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

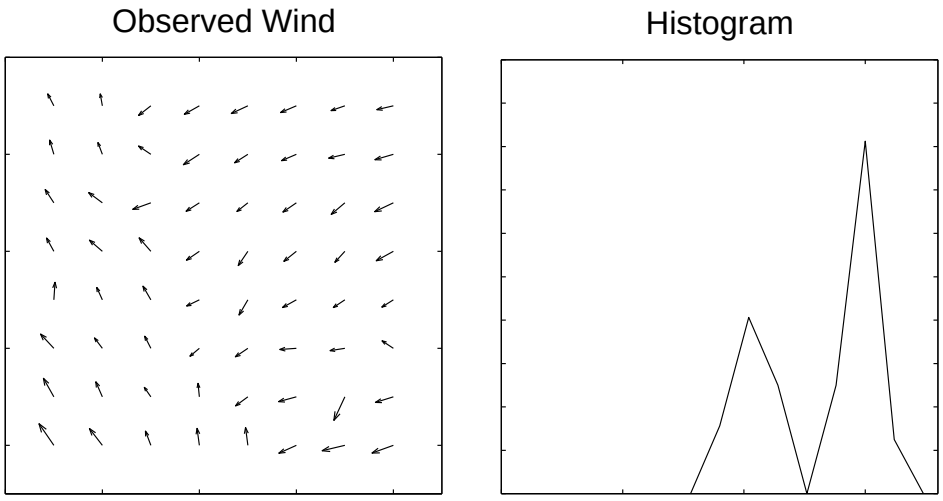


Figure C.1: A wind field containing an ambiguity selection error "edge" and the corresponding directional histogram. Note the bi-modal nature.

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 $\times$ 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.

C.2 Directional histogram-based consistency check

In addition to the model-based consistency check, each 8 $\times$ 8 region is inspected for multiple directional flows (See Figure C.1). 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.

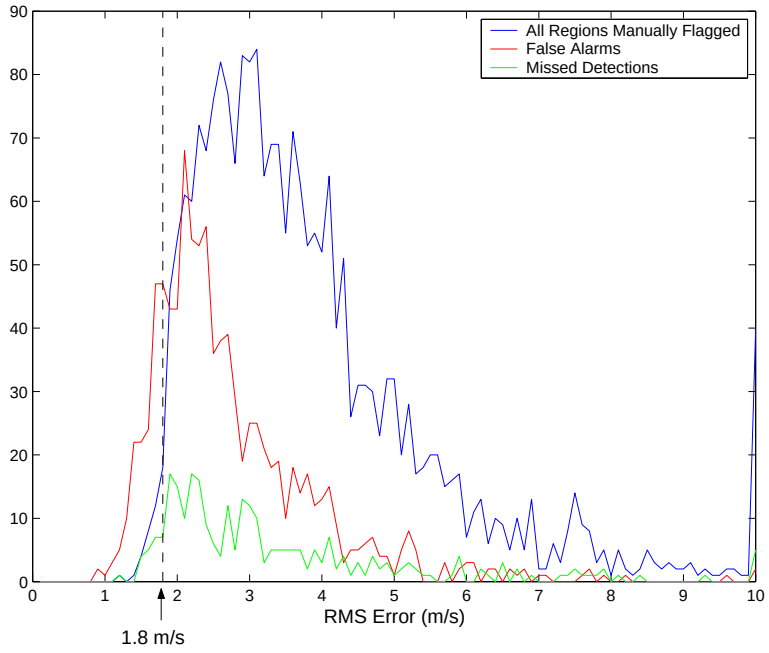


Figure C.2: Histogram of the RMS errors of all the regions manually flagged as "poor", the false alarms generated by the algorithm, and the missed detections generated.

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.

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.

C.3 Region RMS error threshold

In addition to the region threshold for the number of WVCs flagged, an RMS error threshold is applied to each region inspected. The RMS error is defined as the following:

$$\left(\frac{(\mathbf{w}_o - \mathbf{w}_m)^T(\mathbf{w}_o - \mathbf{w}_m)}{N}\right)^{\frac{1}{2}} \tag{C.1}$$

where $\mathbf{w}_o$ and $\mathbf{w}_m$ are the standard vector forms of the observed and modeled winds respectively and N is the number of valid data cells in the region. The number of

false alarms and missed detections per RMS error bin from applying the algorithm to the training data set is computed in Figure C.2. The false alarms for regions whose RMS error is less than 1.8 m/s account for nearly 20% of the total number of false alarms. The missed detection rate is also very high for those regions. This suggests that the algorithm does not perform well for regions where the RMS error is low ($<$ 1.8 m/s). Thus, an *RMS error threshold* is applied which excludes all regions with RMS error less than 1.8 m/s from being flagged as possible ambiguity selection errors. The use of this threshold decreases the false alarm rate by nearly 20% while it increases the missed detection rate by less than 1%. For a region to be flagged as a possible ambiguity selection error, the RMS region threshold must be met.

C.4 RMS wind speed threshold

An RMS wind speed threshold is additionally applied. Regions with RMS wind speeds less than 3.5 m/s are considered not processable because the SNR may be too low to create valid wind direction estimates. From experience with NSCAT, most regions with RMS wind speeds less than 4.0 m/s were flagged primarily because of noise. Through examination of SeaWinds data, we concluded that it is very difficult to subjectively locate ambiguity selection errors when the region RMS wind speed is less than 3.5 m/s. Approximately 7% of the total number of regions fall beneath this threshold.

Appendix D

QA Bit Flag

In order to make the QA data more accessible, we introduce a QA bit flag that can be used in conjunction with the standard wind product. In this section, a level bit flag is introduced that indicates the inferred quality of each cell in a rev. Wind data is generally stored as a swath-shaped array with each element representing a WVC. A corresponding QA flag array is produced by the quality assurance algorithm. The flag is a 4 bit integer at each WVC. The following sections describe the meaning of each bit.

D.1 Individual WVC flag

The two least significant bits of the QA flag, $[q_1, q_0]$ indicate individually flagged vectors. Bit q_0 is set when a vector is flagged using the constant thresholds in any of the overlapping regions (noisy vector flag). Bit q_1 is set when a vector is flagged using the variable thresholds (possible ambiguity selection error or ASE cell flag). Recall the variable thresholds are set higher in nadir and low wind speed regions to produce a constant false alarm rate over the swath, while the constant thresholds have variable performance over the swath. The constant thresholds serve as a noisy vector flag while the variable thresholds indicate vectors more likely to be ambiguity selection errors. In summary, the first two bits have the following meanings:

q_1, q_0

×, 1 - *Noisy vector*: The cell is flagged using constant thresholds in at least one of the overlapping regions, indicating a generally noisy wind vector.

1, × - *ASE cell flag*: The cell is flagged using variable thresholds in at least one of the overlapping regions. Multiple neighboring flagged cells may indicate ambiguity selection errors in the vicinity.

D.2 Region WVC flag

The two most significant bits, $[q_3, q_2]$ correspond to the region flag. Because the 8 × 8 regions overlap, the WVC is classified with the highest classification for any of the overlapping regions. The flagging scheme is:

q_3, q_2

0, 0 - *Good:* All overlapping regions containing the WVC are flagged "good." Each region contains less than 5% of the individual WVCs flagged by the constant thresholds. For these regions the wind flow fits the model estimate very well and is spatially consistent having a low noise level.

0, 1 - *Fair:* At least one overlapping region containing the WVC is flagged "fair." The region contains 5-20% of its WVCs flagged by the constant thresholds. The wind flow is consistent, but some vectors may contain moderate amounts of noise and/or possible ambiguity selection errors. Wind fields with fine scale spatial variations (e.g. fronts) may also be flagged.

1, 0 - *Poor:* At least one overlapping region containing the WVC is flagged "poor." The region contains more than 20% of WVCs flagged by the constant thresholds. The wind flow is not consistent due to ambiguity selection errors high levels of noise. Nadir regions and low wind speed regions (both of which are noisier) are more likely to flagged "poor."

1, 1 - *ASE region flag:* At least one overlapping region is flagged as containing ambiguity selection errors. The region exceeds 14% cells flagged by the variable WVC thresholds, has an RMS wind speed of above 3.5 m/s, the RMS error between the observed and the modeled wind is more than 1.8 m/s, and the histogram of the wind directions in the region is multi-modal. Given all of

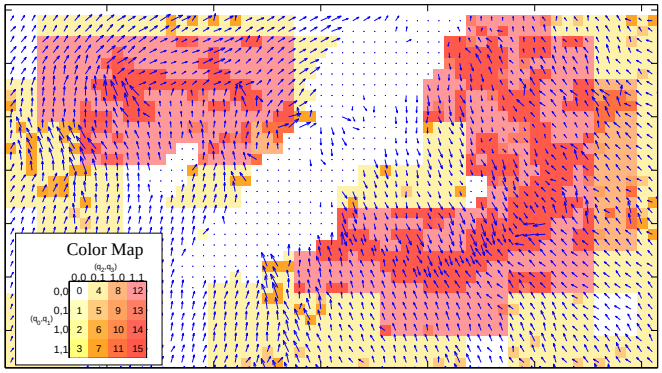


Figure D.1: QA bit flag example containing a possible region of ambiguity selection error. Missing vectors indicate the position of land. Notice that the QA algorithm flags the position of inconsistencies in the estimated wind flow.

these considerations, the region is estimated to contain substantial ambiguity selection errors. Very high noise corruption and rain contamination may also cause the region to be flagged.

The algorithm is performed for each rev and provides a flag value for each WVC. Cells without wind estimates are flagged with zeros. Note that isolated WVCs or isolated small WVC groups cannot be effectively flagged and are, by default, flagged zero. The QA flag can be viewed as an integer number from 0 to 15 indicating the increasing uncertainty in the correctness of the wind flow at that cell. Figure D.1 illustrates a region containing an ambiguity selection error flagged by the algorithm. Additional examples of the QA bit flag are shown in Figure D.2. Notice that the algorithm clearly identifies inconsistencies in the estimated wind flow. This information can be used to locate isolated regions of ambiguity selection errors in order to correct them.

D.3 QA bit flag results

The results presented thus far have given the statistics for the QA algorithm on a region-by-region basis. However, because the regions overlap, this does not fully reflect the actual percentage of WVCs flagged by the QA algorithm. Using the QA bit

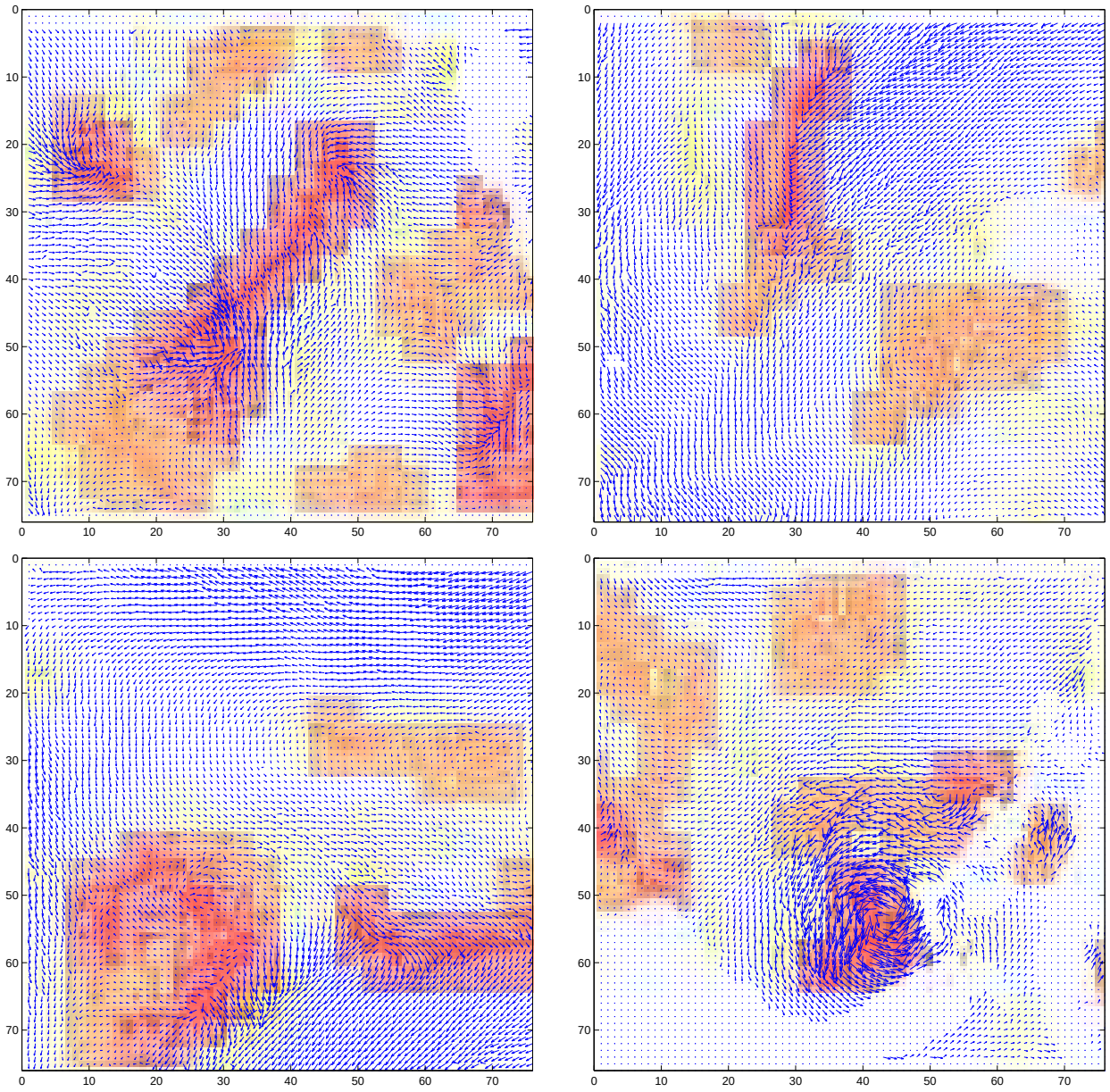


Figure D.2: A few examples of the QA bit flag in stormy regions or regions of ambiguity selection error.

Table D.1: Percent of valid individual WVCs flagged according to several QA bit flag classifications. The symbol '×' indicates a "don't care." The classifications have the following meanings: A - No Flag: the cell is not excessively noisy and is not estimated to be an ambiguity selection error, B - Flagged with constant thresholds (noisy vector), C - Flagged with variable thresholds (likely to lie near ambiguity selection errors), D - "Good" region classification, E - "Fair" region classification, F - "Poor" region classification, G - "Ambiguity selection error" region flag, H - Both region and cell flagged as an ambiguity selection error

	Classification	Bit Flag $[q_3, q_2, q_1, q_0]$	Percent WVCs
A	no flag	$[0, 0, 0, 0]$	52%
B	noisy cell	$[\times, \times, \times, 1]$	12%
C	ASE cell	$[\times, \times, 1, \times]$	6%
D	"good" region	$[0, 0, \times, \times]$	53%
E	"fair" region	$[0, 1, \times, \times]$	22%
F	"poor" region	$[1, 0, \times, \times]$	16.5%
G	ASE region	$[1, 1, \times, \times]$	8.5%
H	ASE region & ASE cell	$[1, 1, 1, \times]$	4.4%

flag, we generate statistics indicating the percent of cells flagged individually by the algorithm and the percentage of cells classified into each region classification. These are shown in Table D.1

From Table D.1, we see that the due to the overlap of regions, actual percentage of cells in regions of ambiguity selection errors is higher than the percentage of regions flagged (compare Table D.1 case G to Table 2.4). However, the percentage of individual cells flagged in the regions of ambiguity selection error are approximately equal to the percent of regions flagged by the algorithm ($\sim 5\%$: Table 2.4, case H).

Appendix E

Automated Storm Rating Procedure

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 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 E.1.

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

$$\text{Rating} = \arg\min_{n} \left\{ \sum_{i} \frac{(X_i - \mu_{i,n})^2}{\sigma_{i,n}^2} \right\} \tag{E.1}$$

where $(\mu_{i,n}, \sigma_{i,n})$ are the mean and standard deviation of the i^{th} parameter given a rating of n. The quality of this method is demonstrated in Table E.2. Based on a simple missed detection/ false alarm analysis, storm sensitive parameters 1-4 performs the best in correctly rating the storms.

Table E.1: Sample means and standard deviations of 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) (%), and 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.

	d		R		Q_1		Q_2		U_q		U_n	
Rating	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

Table E.2: Number of storms subjectively rated "1", "2", or "3" vs. number of storms rated "1", "2", or "3" by the automated method.

	Automated Rating		
Subj. Rat.	"1"	"2"	"3"
"1"	117	44	24
"2"	51	68	50
"3"	12	17	49

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.

Appendix F

Dilation and Erosion Steps of the BYU Point-wise Method

Dilation and erosion are morphological operations that expand or contract features of a binary image [24]. Morphological operations involve a structuring element, similar to a convolution kernel. One of the simplest structuring elements is a "nearest neighbor" element. The "nearest neighbor" element is shown in Fig. F.1. If the "nearest neighbor" element is used, a dilation step turns on a pixel where any neighboring pixel is turned on. Erosion turns off a pixel where any neighboring pixel is off. A general technique of filling bounded regions is to dilate for several iterations, and then erode for several iterations.

Morphological operations are a subclass of cellular automata. A cellular automaton is an array of identically programmed cells which interact with each other. For each cell, there is a state (in the binary case, on or off), a neighborhood, and a set of rules on how the state changes. Morphological operations are binary, but a multi-valued operation is needed in the BYU method. Thus, we modify the dilation and erosion techniques to better suit our application by defining a cellular automaton over the swath.

For each WVC, we define 4 states. State 1 is assigned to all WVCs whose median-filtered wind vector is less than 3 m/s, or are non-data WVCs. Note that the outside cross track row is assigned state 1 because retrieval is not performed there. State 2 is assigned to all WVCs that are flagged as "inconsistent" (see section 3.2.3). These are the edges of the regions of ambiguity selection error. Because states 1 and 2 are defined by the characteristics of the selected wind flow and not by the states of the surrounding cells, they never change during the dilation and erosion steps.

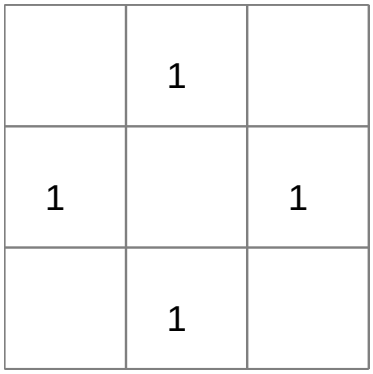


Figure F.1: The "nearest neighbor" structuring element. A value of "1" indicates a nearest neighbor. The center pixel is the origin.

Cells not assigned state 1 or 2 are initially assigned state 0. Through dilation and erosion, state 3 is assigned to all isolated regions of ambiguity selection error. We redefine dilation as the changing from state 0 to state 3, and erosion as the changing from state 3 to state 0. Our neighborhood is all "nearest neighbors."

During the dilation step, the rules for states change are as follows: State 1 and 2 do not change. State 0 changes to state 3 when the neighborhood contains at least one state 2 cell, or contains a state 3 cell accompanied by at least one other state 1 or 3 cell. After iterating, these rules allow the inconsistent edges to dilate until they come in contact with WVCs of state 1, 2 or 3, filling isolated regions. The dilation step is iterated 20 times.

During the erosion step, the rules on state changes are modified. State 3 changes to state 0 when the neighborhood contains one state 0 cell and no state 2 cells, or contains at least two state 0 cells. Again, cells of state 1 or 2 never change. Thus, the non-isolated cells erode away, leaving only the isolated regions. The erosion step is iterated 40 times. All WVCs with non-zero state are flagged as isolated regions of ambiguity selection error. We demonstrate the dilation and erosion steps in Fig. F.2.

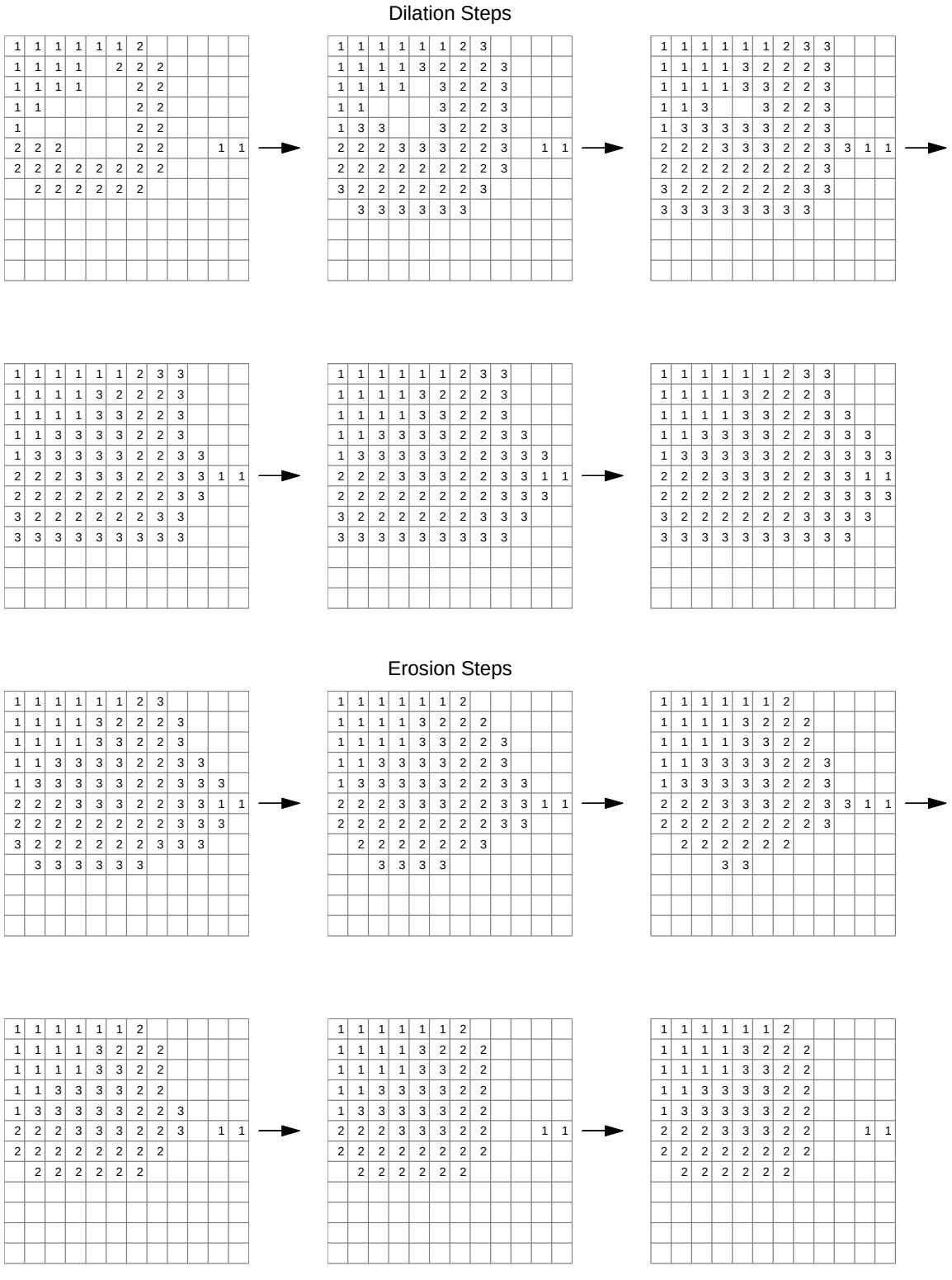


Figure F.2: (a) Dilation and (b) erosion steps, demonstrating how an isolated region is "filled in." The initial state matrix is derived from the example in Fig. 3.5. To enhance readability, state 0 cells are left blank.

Appendix G

Alternative Method of Determining The Low Wind Speed Threshold

In this appendix, we describe an alternative method to estimate the low wind speed threshold using the relationship of the wind to σ^o (the GMF). To first order, the GMF has a power law dependence on wind speed,

$$\sigma^o = A(f, \theta, \chi, p) U^{\gamma(f,\theta,\chi,p)} \tag{G.1}$$

where f is the frequency, θ is the incidence angle, χ is the relative incidence angle, and p is the polarization [39]. Here, γ is referred to as the wind exponent which gives a measure of how sensitive σ^o is to the wind.

The power law dependence of the GMF generally holds for moderate wind speeds but falls off at low winds speeds and may saturate at high wind speeds. In order to characterize the roll-off of the wind at low wind speeds, we take the following simple model for the GMF:

$$\sigma^o = \begin{cases} A_1(f, \theta, \chi, p) U^{\beta(f,\theta,\chi,p)} & U < \Psi \\ A_2(f, \theta, \chi, p) U^{\gamma(f,\theta,\chi,p)} & U \geq \Psi \end{cases} \tag{G.2}$$

or

$$\sigma^o_{dB} = \begin{cases} A_1' + \beta 10 \log_{10}(U) & U < \Psi \\ A_2' + \gamma 10 \log_{10}(U) & U \geq \Psi \end{cases} \tag{G.3}$$

constrained by

$$\beta \geq \gamma \tag{G.4}$$

and

$$A_1 \Psi^\beta = A_2 \Psi^\gamma. \tag{G.5}$$

This model assumes one power law dependence at wind speeds below the low wind speed threshold (Ψ), and another for wind speeds above Ψ. The two constraints insure that the slope of σ^o below Ψ is greater than the slope of σ^o above Ψ and that the model is continuous across the threshold.

The approach taken to find the optimum Ψ for each Bragg wavelength is to determine a least-squares fit of the model to a range of possible threshold wind speeds and then find which threshold generates the lowest RMS error between the model and the data.

The model is applied to the YSCAT data through a constrained least squares optimization where Lagrangian multipliers [26] are used to force the constraints. To develop the least squares solution, we introduce the matrices, B_1 and B_2 which are defined as:

$$B_1 = \begin{bmatrix} 1 & 10\log_{10}(U_1) \\ 1 & 10\log_{10}(U_2) \\ \vdots & \vdots \\ 1 & 10\log_{10}(U_M) \end{bmatrix} = [\mathbf{1} \;\; \mathbf{u}_{1\ldots M}] \tag{G.6}$$

and

$$B_2 = \begin{bmatrix} 1 & 10\log_{10}(U_{M+1}) \\ 1 & 10\log_{10}(U_{M+2}) \\ \vdots & \vdots \\ 1 & 10\log_{10}(U_{M+N}) \end{bmatrix} = [\mathbf{1} \;\; \mathbf{u}_{M+1\ldots M+N}] \tag{G.7}$$

where $U_1 \ldots U_M$ are the speeds of the measurements below Ψ and $U_{M+1} \ldots U_{M+N}$ are the speeds of the measurements above Ψ. In addition, we introduce the vectors:

$$\sigma_1 = \begin{bmatrix} \sigma_1^o \\ \sigma_2^o \\ \vdots \\ \sigma_M^o \end{bmatrix}, \tag{G.8}$$

$$\sigma_2 = \begin{bmatrix} \sigma^o_{M+1} \\ \sigma^o_{M+2} \\ \vdots \\ \sigma^o_{M+N} \end{bmatrix} \tag{G.9}$$

and

$$\mathbf{x} = \begin{bmatrix} A_1 \\ \beta \\ A_2 \\ \gamma \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{1,2} \\ \mathbf{x}_{3,4} \end{bmatrix}. \tag{G.10}$$

where $\sigma^o_1 \ldots \sigma^o_M$ and $\sigma^o_{M+1} \ldots \sigma^o_{M+N}$ are the measured σ^o values below and above Ψ respectively. Now, we find the least squares solution to the equations

$$B_1\mathbf{x}_{1,2} = \sigma_1 \tag{G.11}$$

$$B_2\mathbf{x}_{3,4} = \sigma_2. \tag{G.12}$$

with respect to the constraints given in equations (G.4) and (G.5). Thus, we minimize

$$\|B_1\mathbf{x}_{1,2} - \sigma_1\|_2^2 + \|B_2\mathbf{x}_{3,4} - \sigma_2\|_2^2 \tag{G.13}$$

$$= \sum_{m=1}^{M}(x_1 + u_m x_2 - \sigma^o_m)^2 + \sum_{n=M+1}^{M+N}(x_3 + u_n x_4 - \sigma^o_n)^2 \tag{G.14}$$

with respect to

$$x_1 + \Psi_{dB}x_2 = x_3 + \Psi_{dB}x_4 \quad \text{or} \quad x_1 - x_3 + \Psi_{dB}(x_2 - x_4) = 0 \tag{G.15}$$

and

$$x_2 \geq x_4 \ \text{ or } \ x_4 - x_2 \leq 0 \tag{G.16}$$

These equations are formed into the Lagrangian,

$$\begin{gathered} L(\mathbf{x}, \lambda, \mu) = \textstyle\sum_m (x_1 + u_m x_2 - \sigma^o_m)^2 + \sum_n (x_3 + u_n x_4 - \sigma^o_n)^2 + \\ \lambda[x_1 - x_3 + \Psi_{dB}(x_2 - x_4)] + \mu(x_4 - x_2) \end{gathered} \tag{G.17}$$

where λ is the Lagrangian multiplier of the equality constraint and μ is the Lagrangian multiplier of in the inequality constraint. The gradient of the Lagrangian with respect

to $\mathbf{x}$, λ and μ is:

$$\nabla L(\mathbf{x},\lambda,\mu) = \begin{bmatrix} \sum_m 2(x_1 + u_m x_2 - \sigma_m^o) + \lambda \\ \sum_m 2(x_1 + u_m x_2 - \sigma_m^o)u_m + \Psi_{dB}\lambda \\ \sum_n 2(x_3 + u_n x_4 - \sigma_n^o) - \lambda \\ \sum_n 2(x_3 + u_n x_4 - \sigma_n^o)u_n - \Psi_{dB}\lambda \\ x_1 - x_3 + \Psi_{dB}x_2 - \Psi_{dB}x_4 \\ x_4 - x_2 \end{bmatrix}. \tag{G.18}$$

Casting the gradient of L into matrix/vector form and setting it equal to zero, we obtain:

$$\begin{bmatrix} 2M & 2\sum_m \sigma_m^o & 0 & 0 & 1 \\ 2\sum_m \sigma_m^o & 2\sum_m (\sigma_m^o)^2 & 0 & 0 & \Psi_{dB} \\ 0 & 0 & 2N & 2\sum_n \sigma_n^o & -1 \\ 0 & 0 & 2\sum_n \sigma_n^o & 2\sum_n (\sigma_n^o)^2 & -\Psi_{dB} \\ 1 & \Psi_{dB} & -1 & -\Psi_{dB} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \lambda \end{bmatrix} = \begin{bmatrix} 2\sum_m u_m \\ 2\sum_m u_m \sigma_m^o \\ 2\sum_n u_n \\ 2\sum_n u_n \sigma_n^o \\ 0 \end{bmatrix} \tag{G.19}$$

for μ inactive and

$$\begin{bmatrix} 2M & 2\sum_m \sigma_m^o & 0 & 0 & 1 & 0 \\ 2\sum_m \sigma_m^o & 2\sum_m (\sigma_m^o)^2 & 0 & 0 & \Psi_{dB} & -1 \\ 0 & 0 & 2N & 2\sum_n \sigma_n^o & -1 & 0 \\ 0 & 0 & 2\sum_n \sigma_n^o & 2\sum_n (\sigma_n^o)^2 & -\Psi_{dB} & 1 \\ 1 & \Psi_{dB} & -1 & -\Psi_{dB} & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} 2\sum_m u_m \\ 2\sum_m u_m \sigma_m^o \\ 2\sum_n u_n \\ 2\sum_n u_n \sigma_n^o \\ 0 \\ 0 \end{bmatrix} \tag{G.20}$$

for μ active. Equation (G.19) is first solved. If the inequality condition is violated, equation (G.20) is solved forcing the inequality constraint.

After estimating $\mathbf{x}$ for each Ψ_i examined, the RMS error for that Ψ is determined by

$$E_{\Psi_i} = \sqrt{\frac{\mathbf{e}_1^T\mathbf{e}_1 + \mathbf{e}_2^T\mathbf{e}_2}{M+N}} \tag{G.21}$$

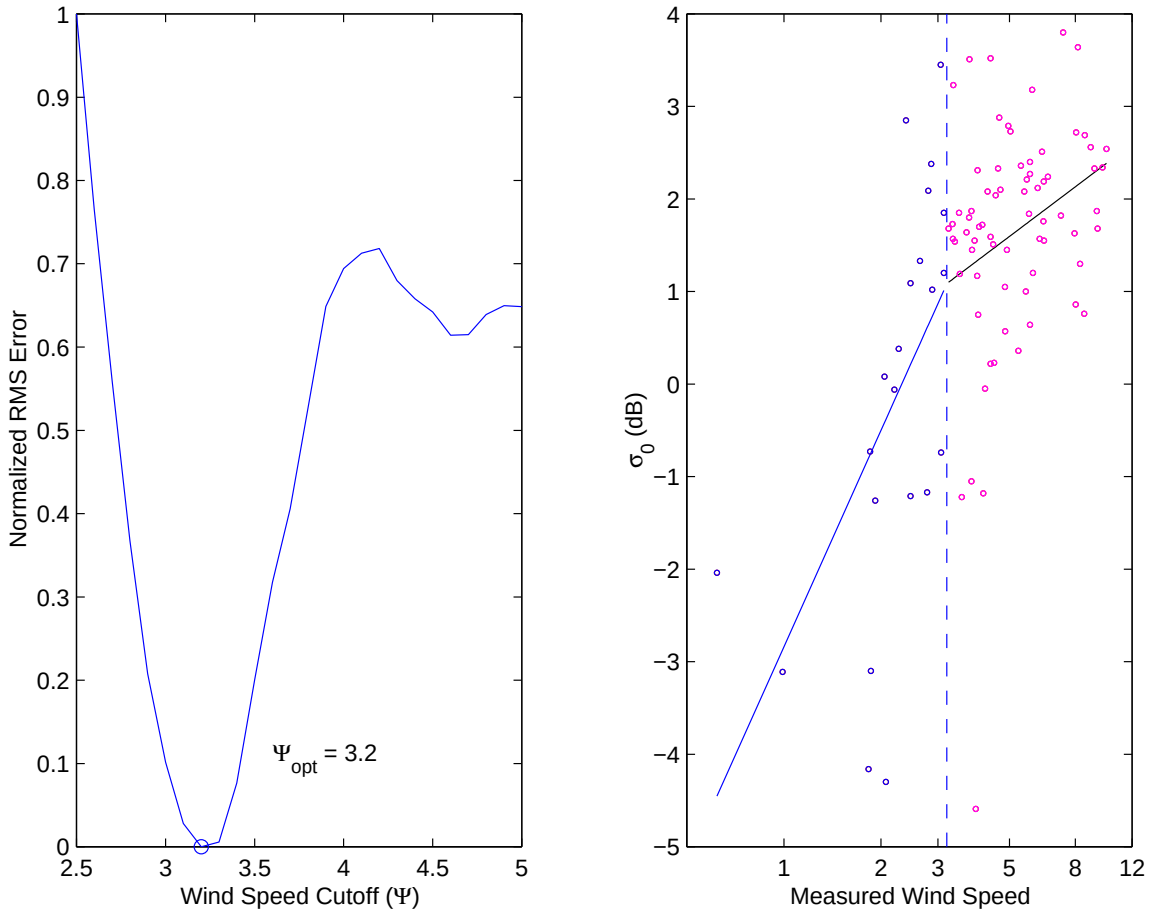


Figure G.1: An threshold wind speed calculation example

where

$$\mathbf{e}_1 = B_1\mathbf{x}_{1,2} - \sigma_1 \tag{G.22}$$

$$\mathbf{e}_2 = B_2\mathbf{x}_{3,4} - \sigma_2. \tag{G.23}$$

The Ψ_i that minimizes the error is chosen as the low wind speed threshold for that Bragg wavelength.

In order to standardize the error metric, an additional "normalized" RMS error is computed defined as

$$\hat{E}_{\Psi_i} = \frac{E_{\Psi_i} - \min\limits_j(E_{\Psi_j})}{\max\limits_j(E_{\Psi_j}) - \min\limits_j(E_{\Psi_j})} \tag{G.24}$$

The normalized RMS error constrains the error to be between 0 and 1. An example of a calculation of the threshold wind speed is shown in Figure G.1. Although this method is not implemented in the analysis of Chapter 4, it is theoretical value, and can be used to form model functions that incorporate the low wind speed thresholds.

Appendix H

Derivation of the Combined Rain/Wind Model Mean and Variance

Assuming that the variability due to atmospheric rain is negligible, the measurement with communication noise, wind model noise, and rain model noise is written as

$$z \cong [\mathcal{M}\alpha_r(1+\eta_1 K_{pm}) + \sigma_e(1+\eta_3 K_{pe})](1+K_{pc}\eta_4). \tag{H.1}$$

The mean of z is taken directly,

$$\begin{aligned} \mathrm{E}\{z\} &\cong \mathrm{E}\{[\mathcal{M}\alpha_r(1+\eta_1 K_{pm}) + \sigma_e(1+\eta_3 K_{pe})](1+K_{pc}\eta_4)\} \\ &\cong \mathrm{E}\{\mathcal{M}\alpha_r + \mathcal{M}\alpha_r\eta_1 K_{pm} + \sigma_e + \sigma_e\eta_3 K_{pe} + K_{pc}\eta_4\mathcal{M}\alpha_r + \\ &\quad K_{pc}\eta_4\mathcal{M}\alpha_r\eta_1 K_{pm} + K_{pc}\eta_4\sigma_e + K_{pc}\eta_4\sigma_e\eta_3 K_{pe}\}. \end{aligned} \tag{H.2}$$

In H.2, all terms containing the independent zero mean, unit variance Gaussian random variables η_1, η_2 and η_4 vanish, leaving

$$\mathrm{E}\{z\} \cong \mathcal{M}\alpha_r + \sigma_e. \tag{H.3}$$

The variance is computed as follows,

$$\begin{aligned} \mathrm{Var}\{z\} &\cong \mathrm{E}\{([\mathcal{M}\alpha_r(1+\eta_1 K_{pm}) + \sigma_e(1+\eta_3 K_{pe})](1+K_{pc}\eta_4) \\ &\quad -\mathcal{M}\alpha_r - \sigma_e)^2\} \\ &\cong \mathrm{E}\{[\mathcal{M}\alpha_r + \mathcal{M}\alpha_r\eta_1 K_{pm} + \sigma_e + \sigma_e\eta_3 K_{pe} + K_{pc}\eta_4\mathcal{M}\alpha_r + \\ &\quad K_{pc}\eta_4\mathcal{M}\alpha_r\eta_1 K_{pm} + K_{pc}\eta_4\sigma_e + K_{pc}\eta_4\sigma_e\eta_3 K_{pe} - \mathcal{M}\alpha_r - \sigma_e]^2\}. \end{aligned} \tag{H.4}$$

Multiplying, we find that all terms with $\eta_i\eta_j$ or $\eta_i\eta_j^2$ go to zero because they are independent, yielding,

$$\begin{aligned}\mathrm{Var}\{z\} &\cong \mathrm{E}\{\mathcal{M}^2\alpha_r^2K_{pm}^2\eta_1^2+\sigma_e^2K_{pe}^2\eta_3^2+\mathcal{M}^2\alpha_r^2K_{pc}^2\eta_4^2+\\ &\quad+\sigma_e^2\eta_4^2K_{pc}^2+\mathcal{M}^2\alpha_r^2K_{pc}^2K_{pm}^2\eta_1^2\eta_4^2\\ &\quad+\sigma_e^2K_{pc}^2K_{pe}^2\eta_3^2\eta_4^2+2\mathcal{M}\alpha_r\sigma_eK_{pc}^2\eta_4^2\}. \qquad \text{(H.5)}\end{aligned}$$

Taking the expectation, with (by definition) $\mathrm{E}\{\eta_i^2\}=1$ and $\mathrm{E}\{\eta_i^2\eta_j^2\}=1$, leaves

$$\begin{aligned}\mathrm{Var}\{z\} &\cong \mathcal{M}^2\alpha_r^2K_{pm}^2+\sigma_e^2K_{pe}^2+\mathcal{M}^2\alpha_r^2K_{pc}^2+\\ &\quad+\sigma_e^2K_{pc}^2+\mathcal{M}^2\alpha_r^2K_{pc}^2K_{pm}^2\\ &\quad+\sigma_e^2K_{pc}^2K_{pe}^2+2\mathcal{M}\alpha_r\sigma_eK_{pc}^2\\ &\cong \mathcal{M}^2\alpha_r^2K_{pm}^2+\sigma_e^2K_{pe}^2+K_{pc}^2(\mathcal{M}\alpha_r+\sigma_e)^2\\ &\quad+\mathcal{M}^2\alpha_r^2K_{pc}^2K_{pm}^2+\sigma_e^2K_{pc}^2K_{pe}^2\\ &\cong (1+K_{pc}^2)(\mathcal{M}^2\alpha_r^2K_{pm}^2+\sigma_e^2K_{pe}^2)\\ &\quad+K_{pc}^2(\mathcal{M}\alpha_r+\sigma_e)^2\\ &\cong (1+K_{pc}^2)(\mathcal{M}^2\alpha_r^2K_{pm}^2+\sigma_e^2K_{pe}^2)+K_{pc}^2\mathcal{M}_r^2. \qquad \text{(H.6)}\end{aligned}$$

Appendix I

Gradient of the Rain/Wind Model

I.1 Gradient of the MLE mean

Simplifying notation, we let $w_1 = u$ and $w_2 = d$ represent the speed and directional components of the true wind. The gradient of $\mathcal{M}_r$ is obtained directly via differentiation, yielding

$$\frac{\partial \mathcal{M}_{rk}}{\partial w_i} = \frac{\partial \mathcal{M}_k}{\partial w_i}\alpha_{rk} \tag{I.1}$$

$$\frac{\partial \mathcal{M}_{rk}}{\partial R} = \mathcal{M}_k \frac{\partial \alpha_{rk}}{\partial R} + \frac{\partial \sigma_{ek}}{\partial R} \tag{I.2}$$

Since the GMF is tabular, $\frac{\partial \mathcal{M}_k}{\partial w_i}$ is obtained numerically.

The derivatives of α_{rk} and σ_{ek} can be written explicitly. The derivative of the attenuation term α_{rk}, given in Eq. (7.2), is obtained via successive applications of the chain rule, yielding

$$\begin{aligned}\frac{\partial \alpha_{rk}}{\partial R} &= \ln(10)10^{(-10^{(f_a(R_{\mathrm{dB}})/10)}/10)} \\ &\quad \times \left(-\frac{\ln(10)}{10}\right) 10^{(f_a(R_{\mathrm{dB}})/10)} \\ &\quad \times \frac{1}{10}\frac{\partial f_a(R_{\mathrm{dB}})}{\partial R_{\mathrm{dB}}} \times \frac{10}{\ln(10)R} \\ &= \frac{\alpha_{rk}\ln\alpha_{rk}}{R}\frac{\partial f_a(R_{\mathrm{dB}})}{\partial R_{\mathrm{dB}}}\end{aligned} \tag{I.3}$$

where

$$\frac{\partial f_a(R_{\mathrm{dB}})}{\partial R_{\mathrm{dB}}} = \sum_{n=1}^{2} n x_{ak}(n) R_{\mathrm{dB}}^{(n-1)}. \tag{I.4}$$

The derivative the effective rain backscatter given in Eq. (7.3) is

$$\begin{aligned}\frac{\partial \sigma_{ek}}{\partial R} &= \ln(10)10^{(f_e(R_{\mathrm{dB}})/10)} \\ &\quad \times \frac{1}{10}\frac{\partial f_e(R_{\mathrm{dB}})}{\partial R_{\mathrm{dB}}} \times \frac{10}{\ln(10)R} \\ &= \frac{\sigma_{ek}}{R}\frac{\partial f_e(R_{\mathrm{dB}})}{\partial R_{\mathrm{dB}}} \end{aligned} \tag{I.5}$$

where

$$\frac{\partial f_e(R_{\mathrm{dB}})}{\partial R_{\mathrm{dB}}} = \sum_{n=1}^{2} n x_{ek}(n) R_{\mathrm{dB}}^{(n-1)}. \tag{I.6}$$

I.2 Gradient of the MLE variance

The gradient of the combined wind/rain variance ς_{rk} is calculated directly using the previously defined derivatives,

$$\frac{\partial \varsigma_{rk}^2}{\partial p_i} = (1+\alpha_k)\frac{\partial \epsilon_k}{\partial p_i} + (2\alpha_k \mathcal{M}_{rk} + \beta_k)\frac{\partial \mathcal{M}_{rk}}{\partial p_i}. \tag{I.7}$$

The derivatives of ϵ_k are calculated as

$$\begin{aligned}\frac{\partial \epsilon_k}{\partial w_i} &= 2(K_{pm}\alpha_{rk}\mathcal{M}_{ak} + K_{pe}\sigma_{ek}) \\ &\quad \times (K_{pm}\alpha_{rk}\frac{\partial \mathcal{M}_k}{\partial w_i}) \end{aligned} \tag{I.8}$$

$$\begin{aligned}\frac{\partial \epsilon_k}{\partial R} &= 2(K_{pm}\alpha_{rk}\mathcal{M}_{ak} + K_{pe}\sigma_{ek}) \\ &\quad \times (K_{pm}\mathcal{M}_{ak}\frac{\partial \alpha_{rk}}{\partial R} + K_{pe}\frac{\partial \sigma_{ek}}{\partial R}). \end{aligned} \tag{I.9}$$

Appendix J

Rain Model Bias Correction

The rain/wind identifiability problem (most evident with cross-swath winds) causes a mean bias in the rain rate estimate [56]. The mean bias can be corrected by adjusting the rain backscatter model. We let R_q represent the QuikSCAT-retrieved rain rate (in decibels). The forward effective rain backscatter model

$$f_e(R_q) = a + bR_q + cR_q^2 \tag{J.1}$$

maps the QuikSCAT retrieved rain rate to a backscatter value. However, this rain rate is biased. We parameterize the bias as a quadratic function of rain rate h that maps TRMM PR rain rates R_t to QuikSCAT rain rates (in a least-squares sense),

$$R_q \approx h(R_t) = d + eR_t + fR_t^2. \tag{J.2}$$

Table J.1: Corrected model parameters for the quadratic log-log model of the effective rain backscatter σ_e.

	h-pol	v-pol
$x_e(1)$	-26.02	-28.01
$x_e(2)$	0.82	0.86
$x_e(3)$	-0.0012	-0.0039

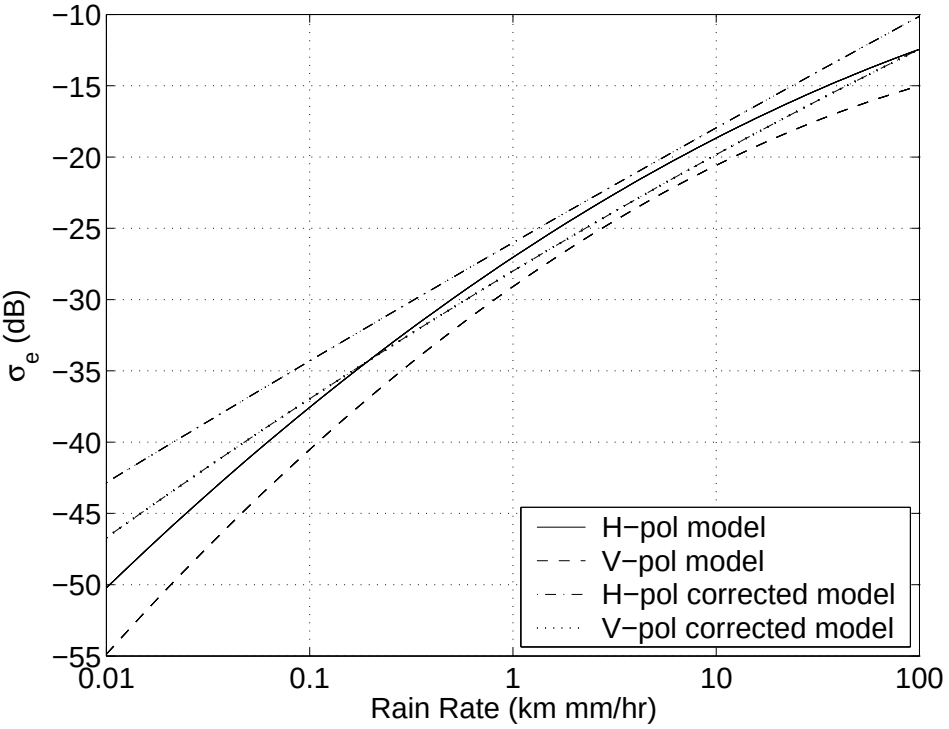


Figure J.1: Validation of QuikSCAT rain rates with TRMM PR antenna average rain rates using the corrected rain model.

The bias-corrected model is calculated directly by inserting the bias correction of Eq (J.2) into the model of Eq. (J.1)

$$\begin{aligned}
\sigma_e &= f_e(h(R_t)) \\
&= a + b(d + eR_t + fR_t^2) + c(d + eR_t + fR_t^2)^2 \\
&\approx (a + bd + cd^2) + (be + 2cde)R_t + \\
&\quad (bf + 2cdf + ce^2)R_t^2. \qquad \text{(J.3)}
\end{aligned}$$

Examining the corrected model in Figure J.1, the corrected model is only slightly higher for moderate rain rates. However, at high and low rain rates, the corrected model is significantly higher. In fact, the corrected model has an almost zero second-order term (see Table J.1). Further, comparison with the TRMM PR is given in Figure J.2. The bias is nearly eliminated.

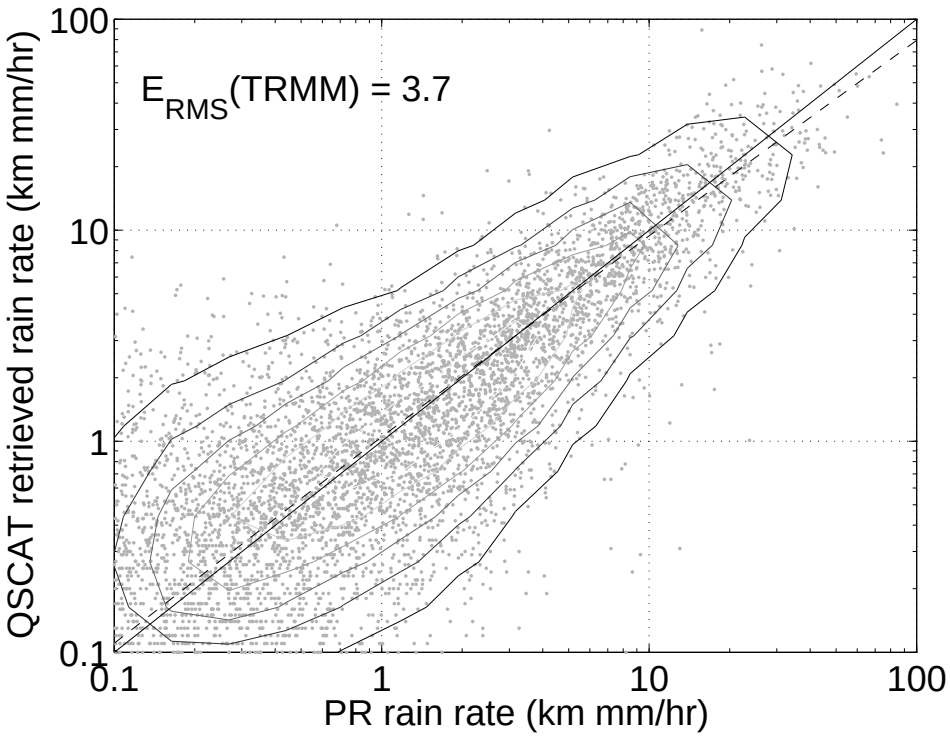


Figure J.2: Validation of QuikSCAT rain rates with TRMM PR antenna average rain rates using the corrected rain model.

Bibliography

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

[2] "A history of scatterometry," http://winds.jpl.nasa.gov/scatterometry/history.html, December 2001.

[3] "Earth observation earthnet online," http://earth.esa.int/, May 2003.

[4] I. S. F. Jones and Y. Toba, *Wind Stress over the Ocean*, Cambridge University Press, 2001.

[5] F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing Active and Passive*, Artech House, 1981.

[6] M. H. Freilich and R. S. Dunbar, "Derivation of satellite wind model functions using operational surface wind analyses - an altimeter example," *J. Geophys. Res.*, vol. 98, no. C8, pp. 14,633–14,649, Aug. 15 1993.

[7] F. J. Wentz and D. K. Smith, "A model function for the ocean-normalized radar cross section at 14 GHz derived from NSCAT observations," *J. Geophys. Res.*, vol. 104, no. C5, pp. 11,499–11,514, May 1999.

[8] F. J. Wentz, M. H. Freilich, and D. K. Smith, "NSCAT-2 geophysical model function," in *1998 Fall AGU Meeting*, 1998.

[9] L. C. Schroeder, D. H. Boggs, G. Dome, I. M. Halberstam, W. L. Jones, W. J. Pierson, and F. J. Wentz, "The relationship between wind vector and normalized radar cross section used to derive SEASAT-A satellite scatterometer winds," *J. Geophys. Res.*, vol. 87, no. C5, pp. 3318–3336, April 1982.

[10] F. M. Naderi, M. H. Freilich, and D. G. Long, "Spaceborne radar measurement of wind velocity over the ocean–an overview of the NSCAT scatterometer system," *Proc. IEEE*, vol. 79, pp. 850–866, 1991.

[11] D. G. Long, "Wind field model-based estimation of Seasat scatterometer winds," *J. Geophys. Res.*, vol. 98, no. C8, pp. 14,651–14,668, August 15 1993.

[12] C. Chi and F. K. Li, "Comparative study of several wind estimation algorithms for spaceborne scatterometers," *IEEE Trans. Geosci. Rem. Sens.*, vol. 26, pp. 115–121, 1988.

[13] M. G. Wurtele, P. M. Woiceshyn, S. Peteherych, M. Borowski, and W. S. Appleby, "Wind direction alias removal studies of SEASAT scatterometer-derived wind fields," *J. Geophys. Res.*, vol. 87, no. C5, pp. 3365–3354, April 1982.

[14] B. W. Stiles, 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. Rem. Sens.*, vol. 40, pp. 79–89, 2002.

[15] T. E. Oliphant and D. G. Long, "Accuracy of scatterometer derived winds using the Cramér-Rao bound," *IEEE Trans. Geosci. Rem. Sens.*, vol. 37, pp. 2642–2652, 1999.

[16] D. G. Long and M. W. Spencer, "Radar backscatter measurement accuracy for a spaceborne pencil-beam wind scatterometer with transmit modulation," *IEEE Trans. Geosci. Rem. Sens.*, vol. 35, no. 1, pp. 102–114, Jan. 1997.

[17] S. J. Shaffer, R. S. Dunbar, S. V. Hsiao, and D. G. Long, "A median-filter-based ambiguity removal algorithm for NSCAT," *IEEE Trans. Geosci. Rem. Sens.*, vol. 29, pp. 167–174, 1991.

[18] M. A. Donelan and W. J. Pierson, "Radar scattering and equilibrium ranges in wind generated waves with application to scatterometry," *J. Geophys. Res.*, vol. 92, no. C5, pp. 4971–5029, May 1987.

[19] W. J. Donnelly, 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.*, vol. 104, pp. 11,485–11,497, 1999.

[20] B. W. Stiles and S. Yueh, "Impact of rain on spaceborne Ku-band wind scatterometer data," *IEEE Trans. Geosci. Rem. Sens.*, vol. 40, no. 9, pp. 1973–1983, Sep. 2002.

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

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

[23] M. A. Vershell, M. A. Bourassa, D. E. Weissman, and J. J. Obrien, "Ocean model validation of the NASA scatterometer," *J. Geophys. Res.*, vol. 104, no. C5, pp. 11,359–11,373, May 1999.

[24] Anil K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.

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

[26] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice Hall, 2000.

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

[28] J. N. Huddleston and B. W. Stiles, "A multi-dimensional histogram rain flagging technique for SeaWinds on QuikSCAT," in *Proc. IEEE Int. Geoscience and Remote Sensing Sympos.*, Honolulu, 2000, vol. 3, pp. 1232–1234.

[29] L. F. Bliven and J. P. Giovanangeli, "Experimental study of microwave scattering from rain- and wind-roughened seas," *Int. J. Rem. Sens.*, vol. 14, pp. 855–869, 1993.

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

[31] D. W. Draper and D. G. Long, "An advanced point-wise ambiguity selection algorithm: Application to SeaWinds," in *Proc. IEEE Int. Geoscience and Remote Sensing Sympos.*, Sydney, AU, July 9-13 2001, vol. 5, pp. 2190–2192.

[32] D. G. Long and J. M. Mendel, "Model-based estimation of wind fields over the ocean from wind scatterometer measurements, -II: Model parameter estimation," *IEEE Trans. Geosci. Rem. Sens.*, vol. 28, pp. 361–373, 1990.

[33] K. Shankaranarayanan and M. A. Donelan, "A probabilistic approach to scatterometer model function verification," *J. Geophys. Res.*, vol. 106, no. C9, pp. 19,969–19,990, 2001.

[34] S. H. Yueh, 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. Rem. Sens.*, vol. 39, pp. 2601–2611, 2001.

[35] D. W. Draper and D. G. Long, "An advanced ambiguity selection algorithm for SeaWinds," *IEEE Trans. Geosci. Rem. Sens.*, vol. 41, no. 3, pp. 538–547, March 2003.

[36] D. W. Draper and D. G. Long, "An assessment of SeaWinds on QuikSCAT wind retrieval," *J. Geophys. Res.*, vol. 107, no. C12, pp. (5)1–(5)14, Dec. 7 2002.

[37] K. K. Kahma and M. A. Donelan, "A laboratory study of the minimum wind speed for wind wave generation," *J. Fluid Mech.*, vol. 192, pp. 339–364, 1988.

[38] W. J. Plant, W. C. Keller, V. Hesany, K. Hayes, K. W. Hoppel, and T. V. Blanc, "Measurements of the marine boundary layer from an airship," *J. Atmos. Ocean. Tech.*, vol. 15, no. 6, pp. 1433–1458, Dec. 1998.

[39] D. G. Long, R. S. Collyer, R. Reed, and D. V. Arnold, "Dependence of the normalized radar cross section of water waves on Bragg wavelength-wind speed sensitivity," *IEEE Trans. Geosci. Rem. Sens.*, vol. 34, no. 3, pp. 656–666, 1996.

[40] A. Bentamy, E. Autret, P. Queffeulou, and Y. Quilfen, "Intercomparison of ERS-2 and QuikSCAT winds," in *Proc. IEEE Int. Geoscience and Remote Sensing Sympos.*, Honolulu, HI, 2000, vol. 1, pp. 234–236.

[41] Z. Jelenak, L. N. Connor, and P. S. Chang, "The accuracy of high resolution winds from QuikSCAT," in *Proc. IEEE Int. Geoscience and Remote Sensing Sympos.*, Toronto, CA, 2002, vol. 2, pp. 732–734.

[42] M. Portabella and A. Stoffelen, "Rain detection and quality control of SeaWinds," *J. Atmos. Ocean. Tech.*, vol. 18, no. 7, pp. 1171–1183, 2001.

[43] C. A. Mears, D. Smith, and F. J. Wentz, "Detecting rain with QuikSCAT," in *Proc. IEEE Int. Geoscience and Remote Sensing Sympos.*, Honolulu, HI, 2000, pp. 1235–1237.

[44] D. E. Weissman, M. A. Bourassa, and J. Tongue, "Effects of rain rate and wind magnitude on SeaWinds scatterometer wind speed errors," *J. Atmos. Ocean. Tech.*, vol. 19, no. 5, pp. 738–746, May 2002.

[45] L. F. Bliven, P. W. Sobieski, and C. Craeye, "Rain generated ring-waves: measurements and modelling for remote sensing," *Int. J. Rem. Sens.*, vol. 18, no. 1, pp. 221–228, 1997.

[46] L. F. Bliven, J. P. Giovanangeli, and G. Norcross, "Scatterometer directional response during rain," in *Proc. IEEE Int. Geoscience and Remote Sensing Sympos.*, Vancouver, CA, 1989, vol. 3, pp. 1887–1890.

[47] R. F. Contreras, W. J. Pland, W. C. Keller, K. Hayes, and J. Nystuen, "Effects of rain on Ku-band backscatter from the ocean," *J. Geophys. Res.*, vol. 108, no. C5, pp. (34)1–(34)15, 2003.

[48] "QuikSCAT operational standard data products release; data release addendum; data quality assessment overview," http://podaac.jpl.nasa.gov/quikscat/qs-rlsadd.doc, May 2000.

[49] T. Kozu, T. Kawanishi, H. Kuroiwa, M. Kojima, K. Oikawa, H. Kumagai, K. Okamoto, M. Okumura, H. Nakatsuka, and K. Nishikawa, "Development of precipitation radar onboard the Tropical Rainfall Measuring Mission (TRMM) satellite," *IEEE Trans. Geosci. Rem. Sens.*, vol. 39, no. 1, pp. 102–116, Jan. 2001.

[50] T. Iguchi, T. Kozu, R. Meneghini, J. Awaka, and K. Okamoto, "Rain profiling algorithm for the TRMM precipitation radar," *J. Appl. Meteor.*, vol. 39, no. 12, pp. 2038–2052, 2000.

[51] M.P. Wand and M.C. Jones, *Kernel Smoothing*, Chapman & Hall, Inc., 1995.

[52] C. Craeye, P. W. Sobieski, and L. F. Bliven, "Scattering by artificial wind and rain roughened water surfaces at oblique incidences," *Int. J. Rem. Sens.*, vol. 35, no. 3, pp. 532–539, 1997.

[53] D. W. Draper and D. G. Long, "Evaluating the effect of rain on SeaWinds scatterometer measurements," *J. Geophys. Res.*, to appear, 2003.

[54] M. W. Spencer and M. Shimada, "Effect of rain on Ku-band scatterometer wind measurements," in *Proc. IEEE Int. Geoscience and Remote Sensing Sympos.*, 1991, vol. 3, pp. 1285–1288.

[55] L. Zeng and R. A. Brown, "Scatterometer observations at high wind speeds," *J. Appl. Meteor.*, vol. 37, pp. 1412–1420, 1998.

[56] D. W. Draper and D. G. Long, "Simultaneous wind and rain retrieval using SeaWinds Data," *J. Geophys. Res.*, in review, 2003.

[57] C. Kummerow, J. Simpson, O. Thiele, W. Barnes, A. T. C. Chang, E. Stocker, R. F. Adler, A. Hou, R. Kakar, F. Wentz, P. Ashcroft, T. Kozu, Y. Hong, K. Okamoto, T. Iguchi, H. Kuroiwa, E. Im, Z. Haddad, G. Huffman, B. Ferrier, W. W. Olson, E. Zipser, E. A. Smith, T. T. Wilheit, G. North, T. Krishnamurti, and K Nakamura, "The status of the Tropical Rainfall Measuring Mission (TRMM) after two years in orbit," *J. Appl. Meteor.*, vol. 39, pp. 1965–1982, Dec. 2000.

[58] L. L. Scharf, *Detection, Estimation and Time Series Analysis*, Addison-Wesley, 1999.

[59] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Image Processing*, vol. 5, no. 3, pp. 493–506, March 1996.

[60] A. T. C. Chang, L. S. Chiu, J. Meng, and TR. Wilheit, "First results of the TRMM Microwave Imager (TMI) monthly oceanic rain rate: Comparison with SSM/I," *Geophys. Res. Lett.*, vol. 26, pp. 2379–2382, 1999.

[61] D. W. Draper and D. G. Long, "Evidence of a threshold wind speed in tower-mounted scatterometer data," in *Proc. IEEE Int. Geoscience and Remote Sensing Sympos.*, Toronto, CA, 2002, vol. 4, pp. 1977–1979.

[62] D. G. Long, "High resolution wind retrieval from SeaWinds," in *Proc. IEEE Int. Geoscience and Remote Sensing Sympos.*, Toronto, CA, 2002, vol. 2, pp. 751–753.