U. S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction
5200 Auth Road Room 207
Camp Springs, MD 20746

Technical Report

Altimeter data for use in wave models at NCEP*.

Hendrik L. Tolman ‡,† Degui Cao† and Vera M. Gerald

Environmental Modeling Center
Marine Modeling and Analysis Branch

October 2006

THIS IS AN UNREVIEWED MANUSCRIPT, PRIMARILY INTENDED FOR INFORMAL EXCHANGE OF INFORMATION AMONG NCEP STAFF MEMBERS

---

* MMAB Contribution No. 252.
‡ e-mail: Hendrik.Tolman@NOAA.gov
† Science Applications International Corporation

This page is intentionally left blank.

Abstract

This report presents an in-house review of wave height and wind speed data from the ERS-2, TOPEX, Jason-1 and ENVISAT altimeters at the Marine Modeling and Analysis Branch (MMAB) of the National Centers for Environmental Prediction (NCEP). These data are intended for use in data assimilation and model validation of the wind wave models at NCEP. The wave height data from the Fast Delivery and Quality controlled sources are shown to be of high quality, and ultimately suitable for assimilation in and validation of wave models. The wind data are of lesser quality. With the exception of the Jason-1 wind data, wind speed retrievals appear to be sensitive to the underlying wave conditions of. Hence, such wind data should not be used in wave model validation or assimilation.

This report is available as a pdf file from

http://polar.ncep.noaa.gov/waves

This page is intentionally left blank.

Contents

This page is intentionally left blank.

1 Introduction

1.1 Background

Altimeter wave data provide the only truly global source of wind wave observations that is operationally available. At NCEP, altimeter data from the ERS-1 and ERS-2 satellites have been used in the past for model validation (Tolman, 1998b; Tolman et al., 2002; Tolman, 2002a,b). More recently, ERS-2 data have been assimilated in the operational global wave model (Chen et al., 2003). Similar approaches are used at many other operational forecast centers.

There are typically two sources of altimeter data, the so-called fast-delivery (FD) data, that are provided through satellite communication in near real time, and the 'science' data, that are provided at a later time and include extensive quality control (QC). Ideally, NCEP uses FD data, because timeliness is essential for use in real time assimilation into operational models as well as for real-time validation purposes. Previous experience with ERS-1 and ERS-2 FD data at NCEP, as well as at many other institutes (e.g., Cotton and Carter, 1994), has shown the need to bias-correct the FD altimeter data. The data corrections can be substantial. For instance, in Tolman (2002b) the corrected altimeter ERS-1 wave height $H_{a,c}$ (in m) was computed from the FD product H_a as

$$H_{a,c} = 0.10 + 1.17 H_a \quad , \tag{1.1}$$

and in Tolman et al. (2002) the corresponding correction for the ERS-2 FD data was found to be

$$H_{a,c} = 0.03 + 1.09 H_a \quad . \tag{1.2}$$

Both corrections included a nonlinear component for low wave heights that is discussed below.

Altimeter winds are not widely used at NCEP, mainly because there appears to be a dependency of the wind speed retrieval on the maturity of the wave fields, which can not be adequately removed (Tolman, 1998b).

NCEP has operationally received FD data from the altimeters on the ERS-1, ERS-2 and TOPEX platforms. Of these only ERS-2 still provides sporadic data, although for all practical purposes, no real-time altimeter data have been available at NCEP since August 2002. Since then several other altimeter data sources have become available, but the majority of these data are not yet available in near real time at NCEP[1]. These data are nevertheless of great interest for model validation and development, and can be used in hindcasts and retrospective studies. Such analyses can fill the global wave model validation gap that has existed since 2002.

[1] With the exception of Jason-1, available at NCEP since January 2006.

Table 1.1: Altimeter data sources used in this study. Fast Delivery (FD) data is obtained through the operational job stream. QC data is obtained from the Naval Research Laboratory (NRL) web sites. Data that are presently available have been used here up to 2005/12.

instrument	months available	type
ERS-2	1997/02 - 2003/06	FD
TOPEX	2002/01 - 2002/08	FD / QC
GFO	2003/01 - present	QC
Jason-1	2003/03 - present	QC
ENVISAT	2003/10 - present	QC

The following sub-sections will describe the various altimeter data sources, sources for observations used to validate and correct altimeter data, and the validations techniques used. In Sections 2 through 6 various altimeter products will be discussed. In Section 7 conclusions are provided.

1.2 Altimeter data

As mentioned above, NCEP needs to receive FD altimeter data, to be used for both assimilation and validation purposes. However, it takes considerable time to include such data in the operational job stream, sometimes several years. NCEP has obtained archived (QC) data from various altimeters that are not operationally available at NCEP from the historical archive at the Naval Research Laboratory (NRL), through their web[2] and ftp[3] servers. The QC data are not available for assimilation, but can still be valuable for validation and model tuning. NCEP has been maintaining wave model archives since Jan. 1997. Any data observed since then can be used for the above purposes. This excludes the above mentioned ERS-1 data, but includes data from ERS-2, TOPEX, GFO, Jason-1 and ENVISAT. Data types and availability are summarized in Table 1.1.

The ERS-2 data window starts with the beginning of the wave model archiving activities. ERS-2 still produces sporadic data, that are used in assimilation. However, because the data became sparse and of lesser quality in early 2002, these data are no longer included in the wave model archive. The TOPEX data became part of the operational job stream at NCEP in January 2002. These data were discontinued in August of the same year. During this period, FD data have been archived and collocated with hind- and forecast results of the wave model. For comparison purposes, the corresponding QC data have also been obtained from the NRL web site. The data from the remaining three instruments are not

[2] http://www7320.nrlssc.navy.mil
[3] ftp://ftp7320.nrlssc.navy.mil

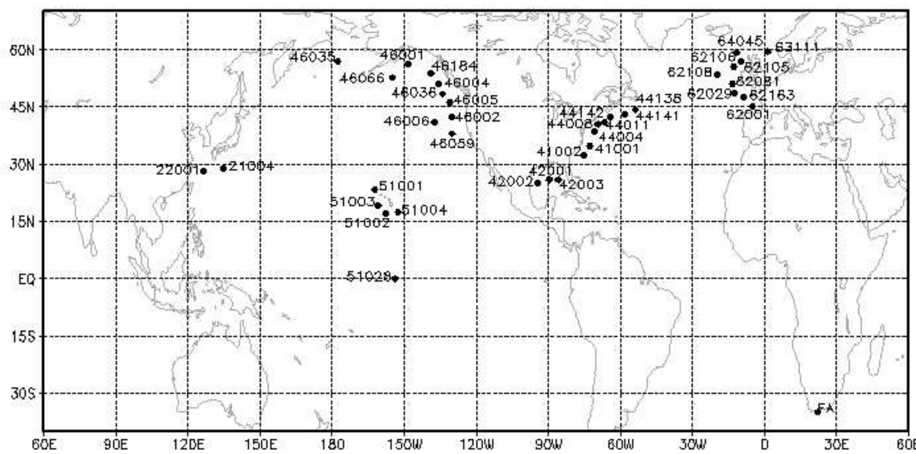

Fig. 1.1 : In-situ data locations used for the validation of the altimeter data.

yet operationally available at NCEP[4]. All available data have been retrieved from the NRL web site up to December 2005.

1.3 Validation data

The altimeter data are validated against deep ocean in-situ data (almost exclusively buoy data). The latter data are obtained in near real time from the operational data stream at NCEP, and are quality controlled using both automated techniques and visual inspection. Coastal buoys have not been used to avoid land contamination in the altimeter data, and also because collocation errors are likely to increase with decreasing distance to the shore. All wind data from the buoys have been converted to 10m height using the known anemometer heights, and assuming neutral stability. The 'FA' data are provided by the South African Weather Service though an exchange program, and are taken from an oil platform with a downward looking altimeter.

[4] Jason-1 FD have become available operationally in 2006/01.

1.4 Validation

Altimeter data are typically available every 7km along the track. With nominal footprints as small as 7km, these data are not (spatially) representative for the deep ocean models at NCEP. Buoy data typically represent a 20min wave data average or a 8min wind speed average. To make the two data sources more compatible, and the altimeter data are averaged along the track in 10s intervals. These averaging intervals are also used to objectively QC the altimeter data, not allowing for records with large standard deviations of data compared to the average data for the interval. The collocated altimeter data are obtained by linear interpolation along the track from two averaged data values, at the track location closest to the buoy.

The present validation is mainly based on linear regression. If the observed parameter is x_o, and the modeled parameter is x_m, the regression describes the relation between measured and observed values as

$$x_m = a + b x_o \quad , \tag{1.3}$$

where a and b are the intercept and the slope, respectively. Traditionally, it is assumed that errors in the observations x_o are much smaller than errors in the model x_m, in which case the slope can be computed as

$$b = \frac{s_{om}}{s_{oo}} \quad , \tag{1.4}$$

where s_{om} represents the conventional covariance of the observations and model data, and where s_{oo} represents the variance of the observations. However, if the observation error is not negligible, systematic errors occur in this slope b (e.g., Draper and Smith, 1981, section 2.14). If the observation error variance s'_{oo} is known, an error corrected slope estimate is given as (Tolman, 1998a)

$$b = \frac{s_{om}}{s_{oo} - s'_{oo}} \quad . \tag{1.5}$$

In the extreme case where the observation error is much larger than the model error, the inverse regression becomes most appropriate, with

$$b = \frac{s_{mm}}{s_{om}} \quad . \tag{1.6}$$

For each estimate of b, the corresponding intercept a can be found from

$$\overline{x}_m = a + b \overline{x}_o \quad , \tag{1.7}$$

where $\overline{x}_m$ and $\overline{x}_o$ are the mean modeled and observed values, respectively.

In the above context, the altimeter corresponds to the 'model' data, and the buoy corresponds to the 'observation' data. The observation error consists of a combination of the direct buoy observation error, and the error incurred by an imperfect collocation. Estimates of such errors for wave heights can be found in (Tolman, 2002b, Appendix A). As is shown in Tolman et al. (2002) and as will be shown in the following sections, the wave observation error in fact is much larger than the remaining error of the averaged altimeter data. Therefore, it is appropriate to use the inverse regression (1.5) for the wave height data. Estimates of observation errors for wind speeds can be found in (Tolman, 1998b, Appendix).

Along with the regression slope, conventional error measures are used such as the bias, the root-mean-square (rms) error, the correlation coefficient and the scatter index (SI). The latter is defined here as the rms error normalized with the mean observation. The definition of all other measures can be found in any textbook, and will not be reproduced here.

In validation studies using collocation techniques, it has been commonplace to use scatter plots of collocated data pairs. Such plots are useful to identify outliers, but can become misleading if large numbers of data are considered. With increasing numbers of data, a scatter plot will suggest a broadening of a data distribution due to optical illusion. For sufficiently large numbers of data, it is therefore more appropriate to consider probability density functions (pdf) of the joint buoy and altimeter data sets. If sufficient data are present to resolve the pdf, additional data will not influence the results. Discretizing the parameter space with increments Δx, the pdf is estimated as

$$\text{pdf} \approx \frac{n_{bin}}{n_{tot} \Delta x_m \Delta x_o} \quad , \tag{1.8}$$

where n_{bin} represents the number of data pairs that fall in the discrete bin considered, and n_{tot} represents the total number of data pairs. A disadvantage of a pdf is its lack of ability to identify outliers. This disadvantage has been mitigated somewhat in the present study by identifying bins without data by gray scaling in all figures. Another disadvantage of pdf plots is that the necessary binning of data may obscure small scale behavior of the collocations. Both pdf plots and scatter diagrams will therefore be used in the present study as appropriate.

1.5 Data correction

Previous literature (e.g., Cotton and Carter, 1994) has identified the need to correct (FD) altimeter wave height data for systematic biases. Typically, a linear correction is used, as in Eqs. (1.1) and (1.2). However, the altimeter measures the wave height as the variance of the surface elevation within the footprint of the altimeter. This variance results in a smearing of the front of the radar pulse. This smoothing is assessed with a discrete signal, and therefore cannot reproduce the full block input signal. Consequently, altimeters generally have a lowest observable wave height, which is significantly larger than 0. Correspondingly, altimeters show a nonlinear bias for low waves. With this in mind, a linear wave height correction is used for raw altimeter wave heights above a critical value H_c. Below this value, a quadratic correction is used

$$\left.\begin{array}{lll} H_{a,c} = a_l + b_l H_a & \text{for} & H_a \geq H_c \\ H_{a,c} = a_q + b_q H_a + c_q H_a^2 & \text{for} & H_a \leq H_c \end{array}\right\} \quad . \tag{1.9}$$

The correction is defined by the linear coefficients a_l and b_l, and the critical wave height H_c. Furthermore, a 'zero wave height' $H_{a,0}$ is defined for which $H_{a,c} = 0$. Finally enforcing continuity of the correction and its first derivative at the connection point, the quadratic correction coefficients a_q, b_q and c_q follow directly from a_l, b_l, H_c and $H_{a,0}$. Note that $H_{a,0}$ can be estimated from the data distribution of H_a as the highest wave height for which no retrievals are found. The choice of H_c is more subjective. Note also that the continuity in the derivative of the correction is important to assure that the Jacobian transformation between the distributions of the corrected and uncorrected wave heights is continuous. If the latter is not the case, discontinuities are introduced in the distribution of the corrected wave heights.

Wind speed errors for altimeter data are generally much larger than for wave height data. For the larger errors, a linear correction is sufficient, and a quadratic correction for low winds represents an unjustifiable level of detail. Hence, the wind correction is simply defined as

$$U_{10,a,c} = a_l + b_l U_{10,a} \quad . \tag{1.10}$$

Negative wind speeds as occasionally occur due to this correction are simply set to zero. In Tolman (1998b) it is shown that the the wind speed retrievals of the ERS-1 altimeter are systematically influenced by maturity of the wave field that coexists with the local wind. This maturity is assessed by the nondimensional wave height $\tilde{H}$, which is defined here as

$$\tilde{H} = 3.33 \frac{gH}{U_{10}^2} \quad , \tag{1.11}$$

which is approximately 1 for mature wind seas, and larger than 1 for swell dominated wave conditions. The effects of the wave maturity on the altimeter wind data is assessed by stratifying collocated wind data with the nondimensional wave height $\tilde{H}$, as computed from the buoy data.

2 ERS-2

The ERS-2 FD data have been used at NCEP for real-time validation for more than seven years (Table 1.1). In this period, more than 10,000 collocated wave height data pairs and nearly 9,000 collocated wind speed data pairs have been obtained, with collocation distances of less than 100km and 30min (see Table A.1 in the Appendices for collocation details). Intercomparison of parts of these data sets indicates that the quality of the data is constant over this time period (figures not presented here). In this section, we consider results for the entire data sets only.

2.1 Wave data

The pdf for the collocated wave heights from ERS-2 FD and buoy data for collocation radii of 100km and 30min are presented in the left panel of Fig. 2.1, together with the corresponding statistical error measures. The pdf shows a narrow linear relationship between the ERS-2 and buoy wave heights. This linear relation extends up to the highest wave heights considered here, as can be inferred for the corresponding 'outliers' distribution as identified by the white background shading. For the lowest wave heights, a somewhat nonlinear relationship appears to exist. This becomes more clear from the corresponding scatter plot for lower wave heights, as presented in the left panel of Fig. 2.2. The latter figure clearly justifies the nonlinear wave height correction (1.9). Note that smaller collocation radii have been used in this figure to highlight the nonlinear behavior in the scatter plot.

Correction parameters according to Eq. (1.9) can be derived iteratively from these collocation data. The resulting correction parameters are gathered in Table 2.1. Note that the correction does not depend on the collocation radii. Furthermore, $H_{a,0}$ is estimated conservatively from non-averaged ERS-2 wave height distributions, which indicated that the ERS-2 altimeter does not produce lower wave heights (figures not presented here). The resulting pdf and scatter plot are presented in the right panels of Figs.2.1 and 2.2. The resulting pdf is linear over the entire data range considered, and the nonlinear bias at low wave height has been eliminated effectively. The remaining error (SI) is 11%, which is close to, or even smaller than the estimated combined buoy observation and collocation error (Tolman, 2002b, Appendix A). This suggests that the error-corrected and averaged ERS-2 wave height data are more accurate than buoy data (with an implicit sampling error of approximately 8%, Donelan and Pierson, 1983), and that the remaining random error of these wave height data is of the order of 5% or perhaps even smaller. This justifies the use of the inverse regression for these data.

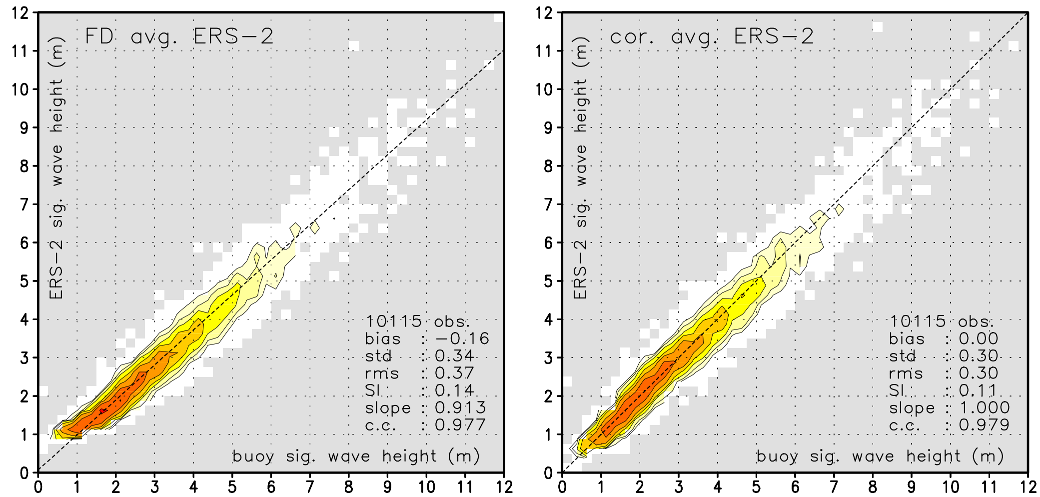

Fig. 2.1 : ERS-2 wave data collocated with buoy observations. Bulk statistics and pdf's computed using wave height increments $\Delta H = 0.25$m. The left panel represent the raw FD data, the right panel represents the data after error correction. The gray background indicates data bins $\Delta H \times \Delta H$ without collocation data. Collocation radii of 100km and 30min.

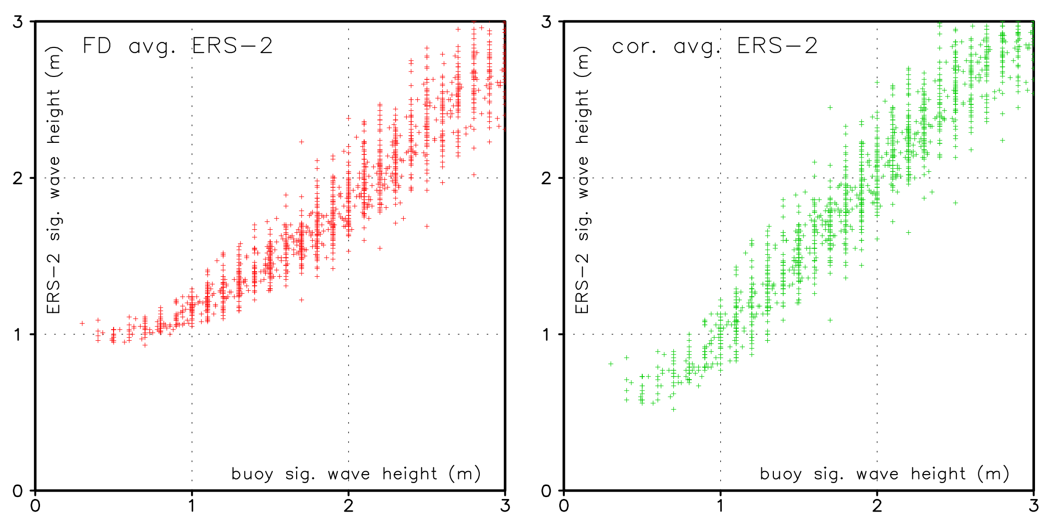

Fig. 2.2 : Scatter plot for lower wave heights corresponding to Fig. 2.1 for collocation radii of 50km and 15min.

Table 2.1: Correction parameters from Eqs. (1.9) for wave heights and (1.10) for wind speeds for the ERS-2 altimeter.

data	collocation radii (km)	(min)	a_l (m) or (ms^{-1})	b_l (-)	H_c (m)	$H_{a,0}$ (m)
waves	100	30	0.14	1.038	2.00	0.70
waves	50	15	0.14	1.038	2.00	0.70
winds	100	30	0.08	1.016	—	—

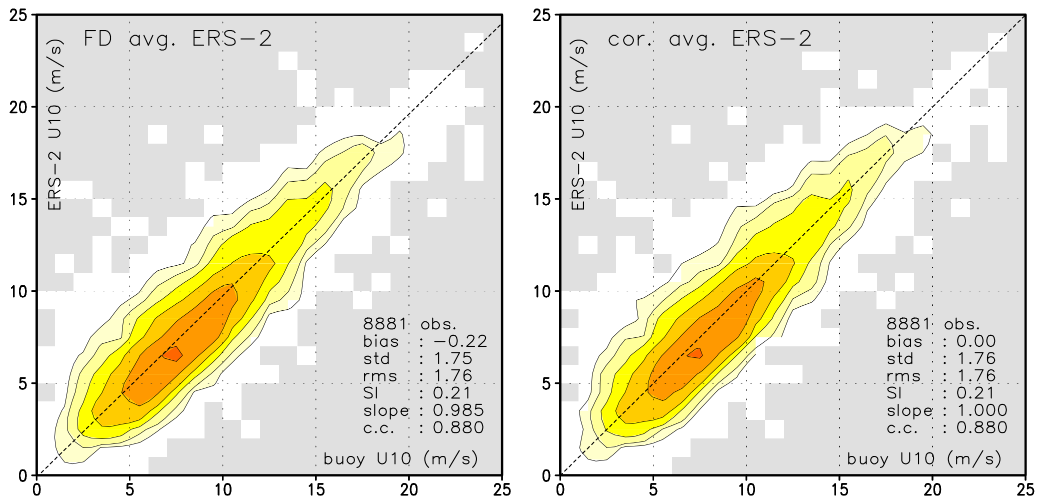

Fig. 2.3 : Like Fig. 2.1 for wind speeds with a bin width of 1ms^{-1}.

2.2 Wind data

The pdf for the collocated 10m wind speeds from ERS-2 FD and buoy data for collocation radii of 100km and 30min are presented in the left panel of Fig. 2.3, together with the corresponding statistical error measures. These data show a small bias and an excellent regression slope, but poorer relative error (SI) and correlation coefficient than the corresponding wave data. These data can be improved upon only mildly, using the error correction of Eq. (1.10) and Table 2.1, as is illustrated in the right panel of Fig. 2.3,

As was mentioned in the previous section, similar data from the ERS-1 satellite were found to be sensitive to the nondimensional wave height $\tilde{H}$ (Tolman, 1998b, Fig. 7). To address if the ERS-2 data have a similar sensitivity, regression lines for collocated data for selected ranges of $\tilde{H}$ are presented for the uncorrected ERS-2 data in Fig. 2.4. The range of $0 < \tilde{H} < 1$ represents (growing)

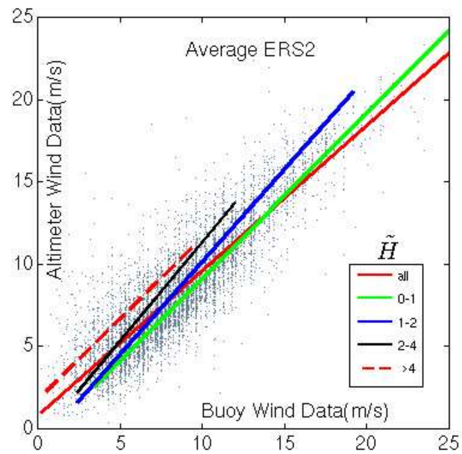

Fig. 2.4 : Regression lines for collocated ERS-2 and buoy wind data for selected ranges of the nondimensional wave height $\tilde{H}$, corresponding to the left panel of Fig. 2.3.

wind seas, $1 < \tilde{H} < 2$ represents overdeveloped wind seas and young swells, and larger values of $\tilde{H}$ represent increasingly mature swells.

Figure 2.4 shows distinctly different regression lines for separate ranges of the the nondimensional wave height $\tilde{H}$. This implies that the validation and error correction of the ERS-2 wind retrieval depends on the wave climate at the validation points. This has two consequences. (i) The validation and error correction cannot be expected to be universally applicable. Particularly, most validation locations are in the northern mid latitudes. The error-corrected winds may be representative for such conditions, but are not likely to be representative for, for instance, the tropics. (ii) Even if the validation data is globally representative, different biases may be expected in different regions with different wave climates. These data are therefore generally not suitable for wind model validation.

3 TOPEX

The TOPEX FD data became operationally available at NCEP only several months before the demise of the satellite. Consequently, only a moderate number of collocations with buoy data could be made (869 for waves and 781 for winds, see Table A.2 in the Appendices for collocation details). Nevertheless, these data are quantitatively sufficient to obtain a reasonable impression of the quality of the TOPEX data.

3.1 Wave data

The pdf for the collocated wave heights from TOPEX FD and buoy data for collocation radii of 100km and 30min are presented in the left panel of Fig. 3.1, together with the corresponding statistical error measures. As with ERS2, the pdf shows a narrow linear relationship between the TOPEX and buoy wave heights, for the entire range of wave heights. For the lowest wave heights, no nonlinear relationship between TOPEX and buoy data appears evident (left panel of Fig. 3.2). This is possibly due to the sparsity of the collocation data.

As with ERS-2, correction parameters for TOPEX according to Eq. (1.9) can be derived iteratively from these collocation data. The resulting correction parameters are gathered in Table 3.1. The resulting pdf and scatter plot are presented in the right panels of Figs.3.1 and 3.2. Conclusions to be drawn from these figures are similar to those obtained for ERS-2. There appears to be no significant impact of the collocation radii on the computed wave height corrections. Narrow linear relations between TOPEX and buoy data are found, suggesting that the averaged TOPEX data are more accurate than the buoy data.

3.2 Wind data

The pdf for the collocated 10m wind speeds from TOPEX FD and buoy data for collocation radii of 100km and 30min are presented in the left panel of Fig. 3.3, together with the corresponding statistical error measures. These data show a significant low bias and a significantly low regression slope. These data can be improved upon significantly by using the linear bias correction of Eq. (1.10) and Table 3.1, as is illustrated in the right panel of Fig. 3.3,

To address if the TOPEX wind data have a sensitivity to the wave conditions, regression lines for collocated data for selected ranges of $\tilde{H}$ are presented for the uncorrected TOPEX data in Fig. 3.4. Figure 3.4 shows distinctly different regression lines for separate ranges of the nondimensional wave height $\tilde{H}$. This implies that the validation and error correction of the TOPEX wind retrieval depends on the wave climate at the validation points, as was also found for ERS-1 and ERS2 wind data in Tolman (1998b), and in the previous section,

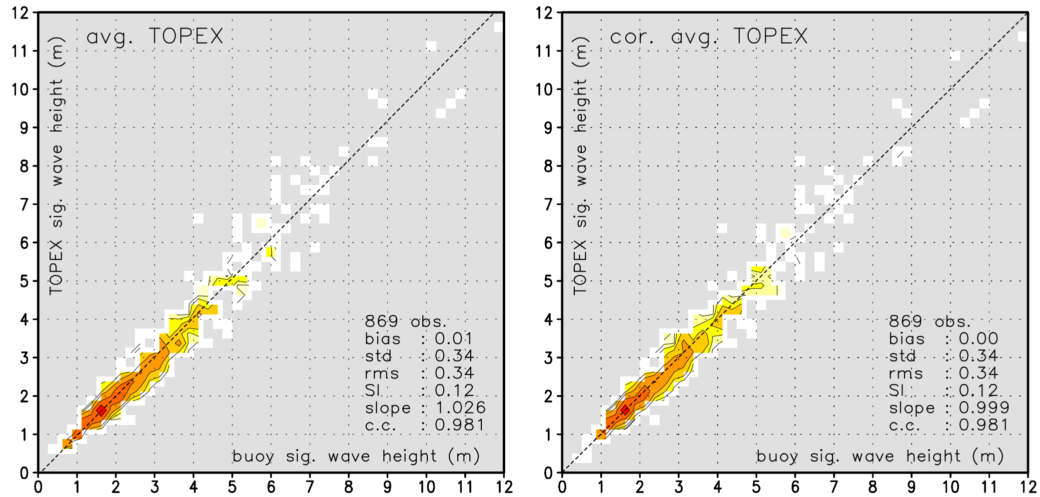

Fig. 3.1 : TOPEX wave data collocated with buoy observations. Bulk statistics and pdf's computed using wave height increments $\Delta H = 0.25$m. The left panel represent the raw FD data, the right panel represents the data after error correction. The gray background indicates data bins $\Delta H \times \Delta H$ without collocation data. Collocation radii of 100km and 30min.

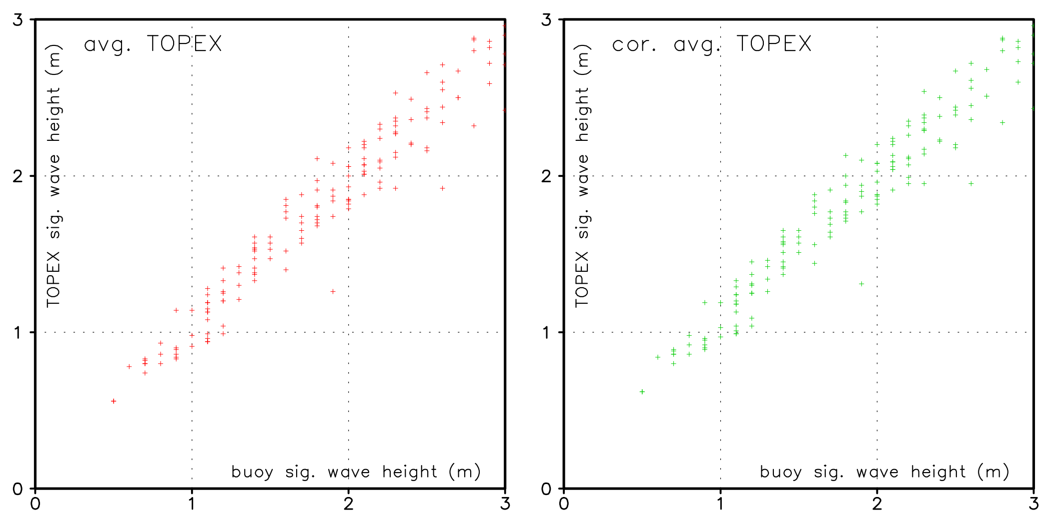

Fig. 3.2 : Scatter plot for lower wave heights corresponding to Fig. 3.1 for collocation radii of 100km and 30min.

Table 3.1: Correction parameters from Eqs. (1.9) for wave heights and (1.10) for wind speeds for the TOPEX altimeter.

data	collocation radii		a_l	b_l	H_c	$H_{a,0}$
	(km)	(min)	(m) or (ms^{-1})	(-)	(m)	(m)
waves	100	30	0.07	0.972	1.40	0.20
waves	50	15	0.08	0.973	1.40	0.20
winds	100	30	-1.62	1.319	—	—

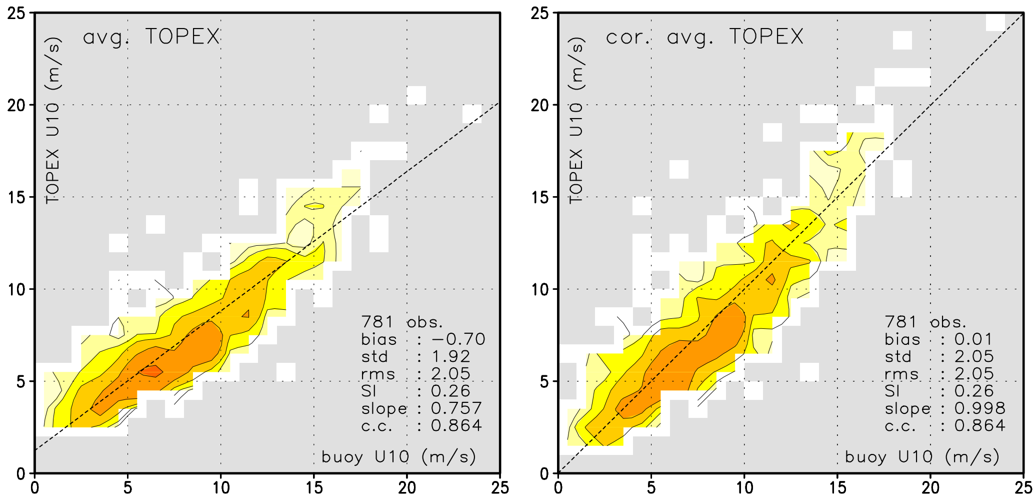

Fig. 3.3 : Like Fig. 3.1 for wind speeds with a bin width of 1ms^{-1}.

respectively.

3.3 QC TOPEX data

TOPEX QC data for the same period for which NCEP obtained FD data have also been obtained from the NRL web site. The QC data result in slightly less collocations than the FD data (775 for waves and 688 for winds, see Table A.3 in the Appendices for collocation details).

Figure 3.5 show the corresponding wave height collocations before and after bias corrections. Collocation results are similar to those of the FD data (Fig. 3.1, however, with some notable differences. In the QC data, there are significantly less collocations for low wave heights. The collocation results for the FD data suggests that the removed low wave height data in the QC data set in fact con-

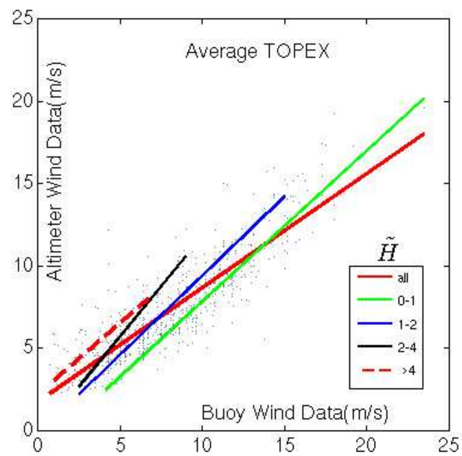

Fig. 3.4 : Regression lines for collocated TOPEX and buoy wind data for selected ranges of the nondimensional wave height $\tilde{H}$, corresponding to the left panel of Fig. 3.3.

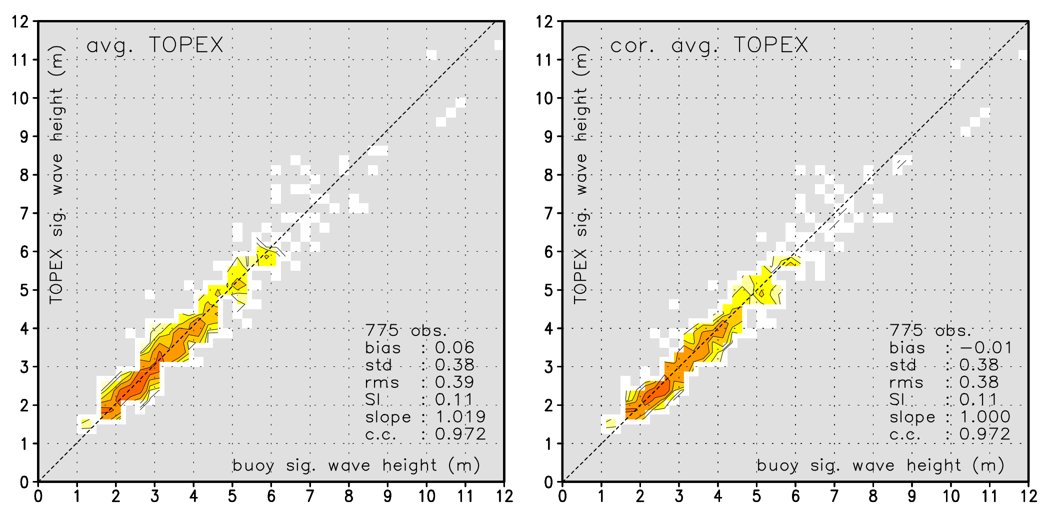

Fig. 3.5 : Like Fig. 3.1 for QC data.

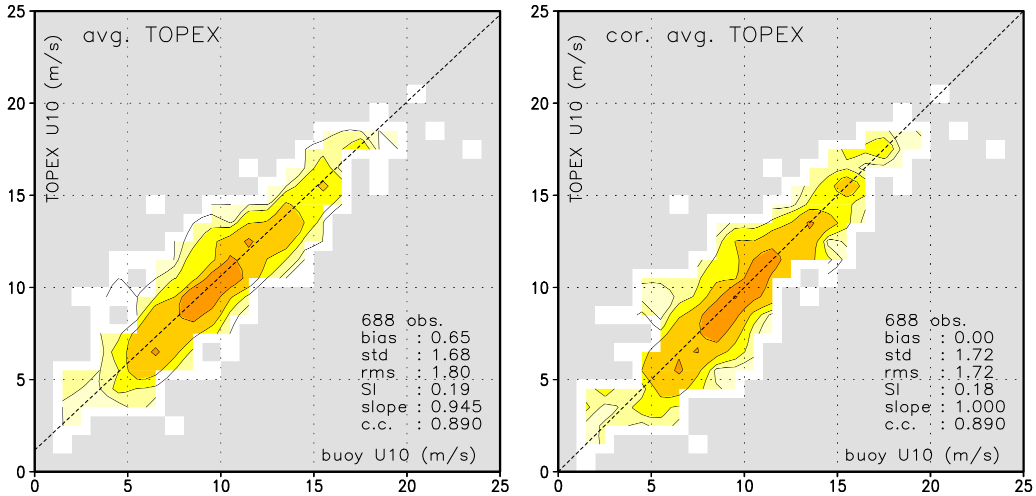

Fig. 3.6 : Like Fig. 3.3 for QC data.

tain useful information, and by and large are accurate. It appears that slightly better bulk statistics for the QC wave height data are attained at the expense of removing useful wave height information for low wave heights.

Figure 3.6 show the TOPEX QC wind speed collocations before and after bias corrections. The QC wind data from TOPEX require much smaller bias corrections, and in general, show much better collocation statistics than the bias-corrected TOPEX FD winds (Compare statistics in Figs. 3.3 and 3.6, respectively). Clearly, the QC process improved dramatically upon the FD wind data for TOPEX.

Finally, Fig. 3.7 presents the collocated TOPEX QC wind data stratified with the nondimensional wave height. The TOPEX QC wind data still show a distinct dependence of the wind retrieval on the wave conditions, which apparently has not been removed in the QC process.

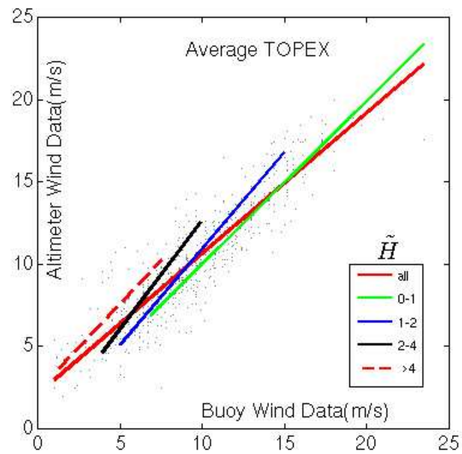

Fig. 3.7 : Like Fig. 3.4 for QC data.

4 GFO

As has been mentioned in the Introduction, GFO data are not yet operationally available at NCEP, and the data have been obtained from the NRL web site for a period of three years (see Table 1.1). This has resulted in 4500 collocation data pairs for wave heights, and 4100 for wind speeds (see Table A.4 in the Appendices for collocation details).

4.1 Wave data

Figure 4.1 presents the pdf's for the uncorrected (left panel) and corrected (right panel) wave height collocations of GFO data with in-situ data. Figure 4.2 shows the corresponding scatter plot for low wave heights. As with the previously discussed altimeter data, a strong correlation and highly linear behavior is found between the collocated altimeter and in-situ data. Like with the TOPEX data, and unlike for the ERS-2 data, there is no clear nonlinear behavior in the collocations for low wave heights. Considering the comparable number of collocations for ERS-2 or GFO, this appears to be due to the actual retrieval algorithm, and cannot be attributed to data sparsity.

Correction parameters according to Eq. (1.9) again have been derived iteratively from these collocation data. The resulting correction parameters are gathered in Table 4.1. The correction again is insensitive to the choice of collocation radii. Note that the correction is dominated by the removal of an absolute bias, and that the correction provides only a minor change in slope of the collocation data. The resulting pdf and scatter plot are presented in the right panels of Figs.4.1 and 4.2. The resulting errors (SI = 11%) are comparable to the errors of the previously discussed altimeter, and are dominated by the errors of the in-situ data and the collocation procedure.

4.2 Wind data

The pdf for the collocated 10m wind speeds from GFO and in-situ data for collocation radii of 100km and 30min are presented in the left panel of Fig. 4.3, together with the corresponding statistical error measures. These data show a much larger error in slope than the ERS-2 altimeter wind data. The slope error is similar to that of the TOPEX FD data. The wind speeds can be improved upon significantly by using the error correction of Eq. (1.10) and Table 4.1, as is illustrated in the right panel of Fig. 4.3. The remaining errors (SI = 25%) are similar to those of the TOPEX data, and slightly larger than those of the ERS-2 data.

Collocation data for uncorrected GFO wind speeds stratified with the nondimensional wave height $\tilde{H}$ are presented in Fig. 4.4. Again, a clear dependence of

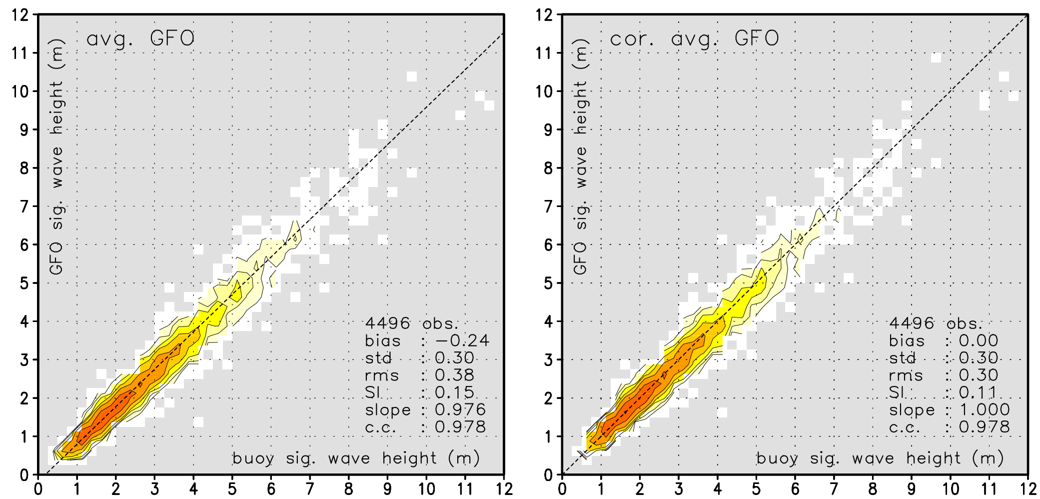

Fig. 4.1 : GFO wave data collocated with buoy observations. Bulk statistics and pdf's computed using wave height increments $\Delta H = 0.25m$. The left panel represent the raw FD data, the right panel represents the data after error correction. The gray background indicates data bins $\Delta H \times \Delta H$ without collocation data. Collocation radii of 100km and 30min.

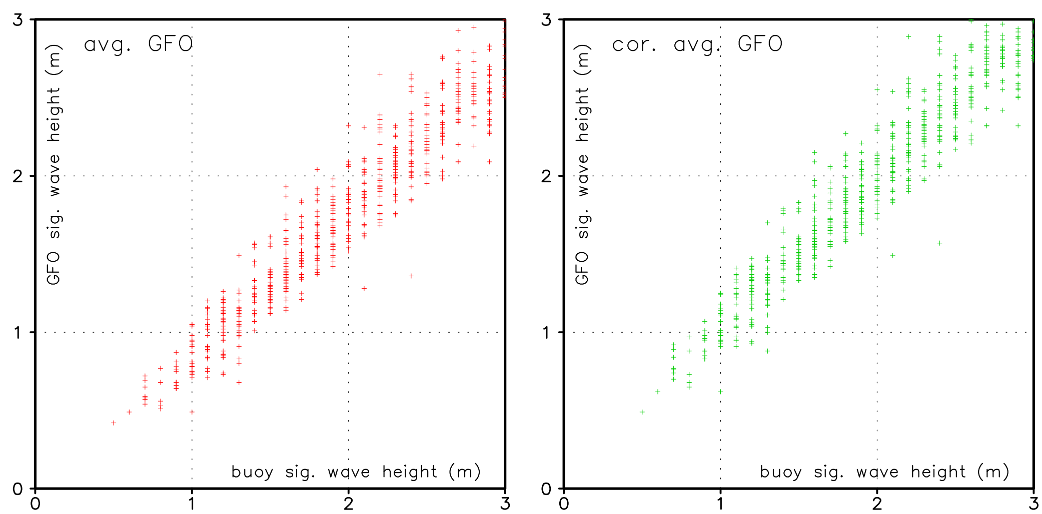

Fig. 4.2 : Scatter plot for lower wave heights corresponding to Fig. 4.1 for collocation radii of 50km and 15min.

Table 4.1: Correction parameters from Eqs. (1.9) for wave heights and (1.10) for wind speeds for the GFO altimeter.

data	collocation radii (km)	(min)	a_l (m) or (ms^{-1})	b_l (-)	H_c (m)	$H_{a,0}$ (m)
waves	100	30	0.18	1.023	1.71	0.20
waves	50	15	0.18	1.028	1.71	0.20
winds	100	30	-2.31	1.437	—	—

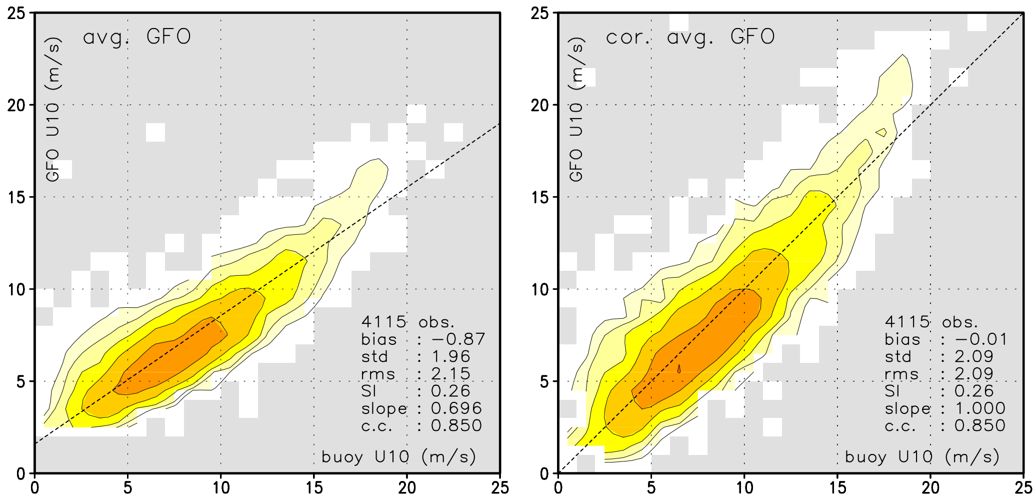

Fig. 4.3 : Like Fig. 4.1 for wind speeds with a bin width of 1ms^{-1}.

the regression behavior on the wave maturity ($\tilde{H}$) is found. This implies similar limitation on the usefulness of the GFO wind data as has been discussed in the context of the ERS-2 data.

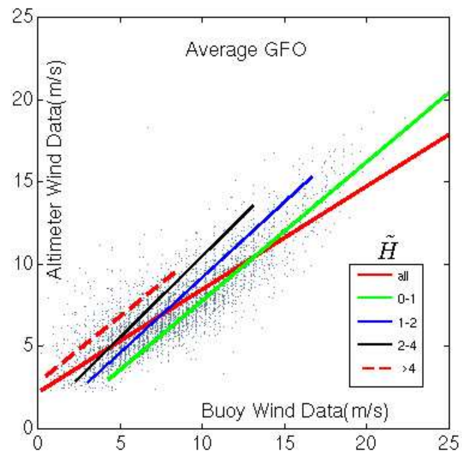

Fig. 4.4 : Regression lines for collocated GFO and buoy wind data for selected ranges of the nondimensional wave height $\tilde{H}$, corresponding to the left panel of Fig. 4.3.

5 Jason-1

Jason-1 FD data have become operationally available at NCEP in January 2006. Previous to Jan. 1, 2006, QC data have been obtained from the NRL web site for a period of nearly three years (see Table 1.1). This has resulted in 4300 collocation data pairs for QC wave heights, and 4000 for QC wind speeds (see Table A.5 in the Appendices for collocation details). This section will concentrate mostly on the QC wave height (Section 5.1) and QC wind speed data (Section 5.2). In Section 5.3, the relation between FD and QC data for the first part of 2006 will be addressed.

5.1 Wave data

Figure 5.1 presents the pdf's for the uncorrected (left panel) and corrected (right panel) wave height collocations of Jason-1 QC data with in-situ data. Figure 5.2 shows the corresponding scatter plot for low wave heights. As with the previously discussed altimeter data, a strong correlation and highly linear behavior is found between the collocated altimeter and in-situ data. Like with the TOPEX and GFO data, and unlike for the ERS-2 data, there is no clear nonlinear behavior in the collocations for low wave heights.

Correction parameters according to Eq. (1.9) again have been derived iteratively from these collocation data. The resulting correction parameters are gathered in Table 5.1. The correction again is insensitive to the choice of collocation radii. Note that the correction is very small. The resulting pdf and scatter plot are presented in the right panels of Figs.5.1 and 5.2. The resulting errors (SI = 11%) are comparable to the errors of the previously discussed altimeter, and are dominated by the errors of the in-situ data and the collocation procedure.

5.2 Wind data

The pdf for the collocated 10m wind speeds from Jason-1 and in-situ data for collocation radii of 100km and 30min are presented in the left panel of Fig. 5.3, together with the corresponding statistical error measures. These data show linear behavior between altimeter and buoy data, with a slope close to unity. The error correction of Eq. (1.10) and Table 5.1 has only a minor impact, as is illustrated in the right panel of Fig. 5.3. The rms wind speed errors ($1.55\mathrm{ms}^{-1}$) are slightly smaller than for the previously discussed altimeters. The normalized errors (SI = 18%) are similar smaller than those of previously discussed altimeter.

Collocation data for uncorrected Jason-1 wind speeds stratified with the nondimensional wave height $\tilde{H}$ are presented in Fig. 5.4. Unlike with the other altimeters previously discussed, there appears to be no functional dependency of the regression slope on the nondimensional wave height $\tilde{H}$. This suggests that

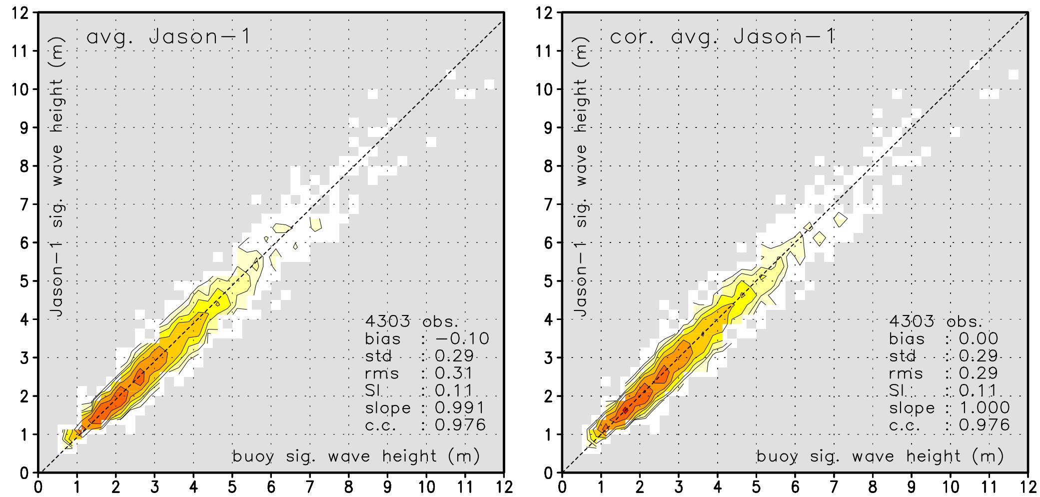

Fig. 5.1 : Jason-1 wave data collocated with buoy observations. Bulk statistics and pdf's computed using wave height increments $\Delta H = 0.25$*m. The left panel represent the raw FD data, the right panel represents the data after error correction. The gray background indicates data bins* $\Delta H \times \Delta H$ *without collocation data. Collocation radii of 100km and 30min.*

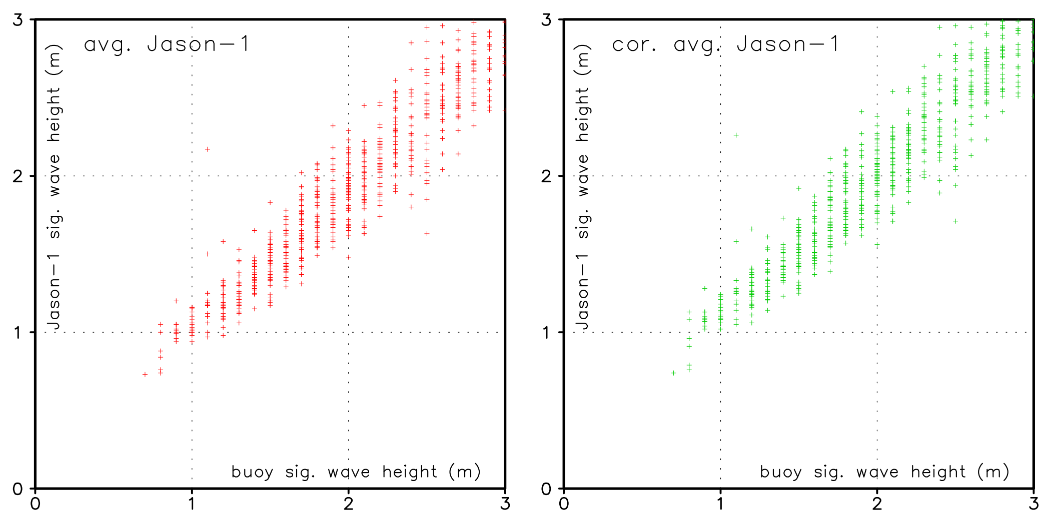

Fig. 5.2 : Scatter plot for lower wave heights corresponding to Fig. 5.1 for collocation radii of 50km and 15min.

Table 5.1: *Correction parameters from Eqs. (1.9) for wave heights and (1.10) for wind speeds for the Jason-1 altimeter.*

data	collocation radii (km)	(min)	a_l (m) or (ms^{-1})	b_l (-)	H_c (m)	$H_{a,0}$ (m)
waves	100	30	0.07	1.009	1.80	0.25
waves	50	15	0.04	1.022	1.80	0.25
winds	100	30	0.36	1.046	—	—

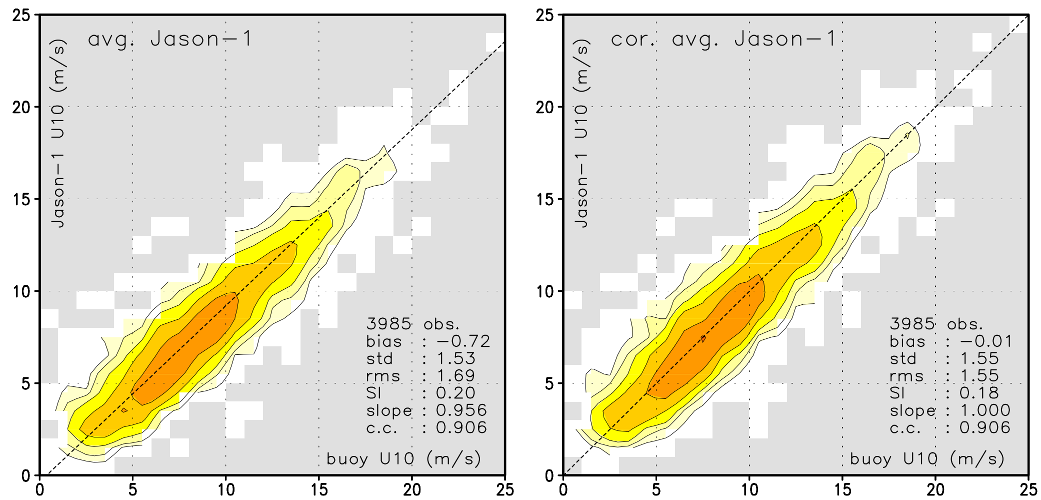

Fig. 5.3 : Like Fig. 5.1 for wind speeds with a bin width of $1ms^{-1}$.

the wind speed retrieval algorithm used for Jason-1 does not display an implicit dependency on the underlying wave field, and is therefore generally or globally applicable, unlike the previously discussed wind speed retrievals from altimeters.

5.3 FD versus QC data

Since January 1, 2006, NCEP has had operational access to both FD and QC Jason-1 data. This period is too short to obtain a sufficient number of wave height and wind speed collocations between Jason-1 and buoy data to assess the quality of the FD data with buoy data only. However, QC and FD data can be compared directly along the satellite ground track, resulting in approximately 100,000 comparisons between FD and QC data per month. From such a comparison, a linear relation between FD and QC data can be derived after collecting

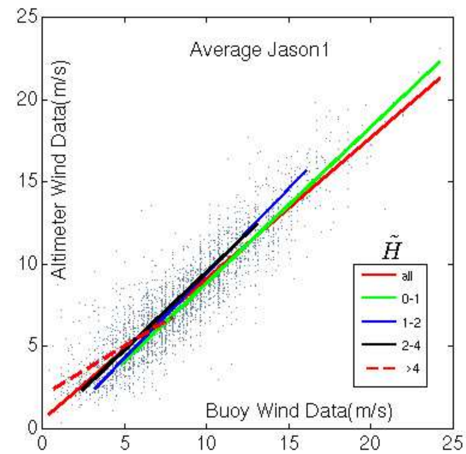

Fig. 5.4 : Regression lines for collocated Jason1 and buoy wind data for selected ranges of the nondimensional wave height $\tilde{H}$, corresponding to the left panel of Fig. 5.3.

only a few days or weeks of data.

Figure 5.5 illustrates the relation between Jason-1 FD and QC data for the first six months of 2006. From these data, the following linear relations between FD and QC data are found.

$$H_{\rm QC} = 0.02 + 1.048 H_{\rm FD} \quad , \tag{5.1}$$

$$U_{10,\rm QC} = 0.51 + 1.004 U_{10,\rm FD} \quad , \tag{5.2}$$

from which the relation between buoy and Jason-1 FD data can be constructed using Eqs. (1.9) and (1.10) and Table 5.1.

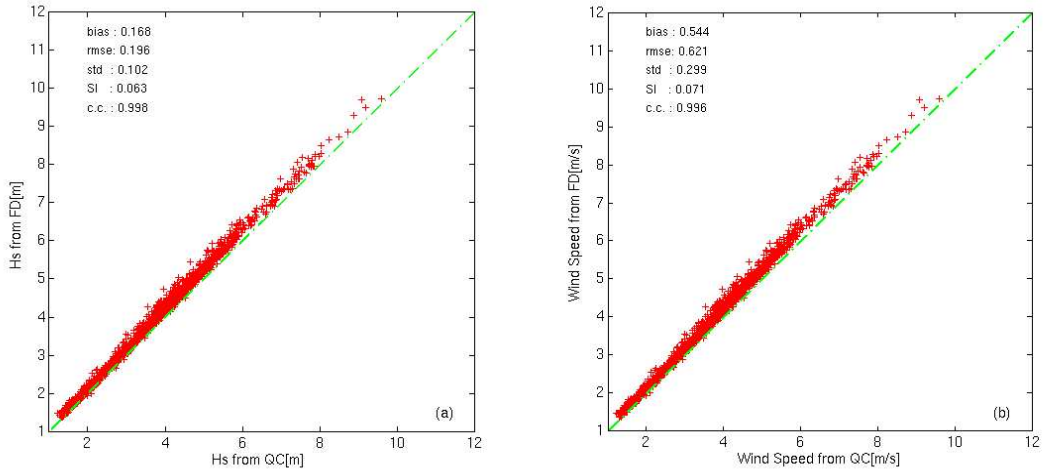

Fig. 5.5 : Jason-1 FD data versus QC data for January 2006. (a) Wave height. (b) Wind speed.

This page is intentionally left blank.

6 ENVISAT

ENVISAT data are not yet operationally available at NCEP, and the data have been obtained from the NRL web site for a period of just over two years (see Table 1.1). This has resulted in 2000 collocation data pairs for wave heights, and 1800 for wind speeds (see Table A.6 in the Appendices for collocation details).

6.1 Wave data

Figure 6.1 presents the pdf's for the uncorrected (left panel) and corrected (right panel) wave height collocations of ENVISAT data with in-situ data. Figure 6.2 shows the corresponding scatter plot for low wave heights. As with the previously discussed altimeter data, a strong correlation and highly linear behavior is found between the collocated altimeter and in-situ data. Only minor nonlinear behavior is found for the lowest wave heights.

Correction parameters according to Eq. (1.9) again have been derived iteratively from these collocation data. The resulting correction parameters are gathered in Table 6.1. The correction again is insensitive to the choice of collocation radii, and the resulting errors (SI = 11%) are comparable to the errors of the previously discussed altimeters, and are dominated by the errors of the in-situ data and the collocation procedure.

6.2 Wind data

The pdf for the collocated 10m wind speeds from ENVISAT and in-situ data for collocation radii of 100km and 30min are presented in the left panel of Fig. 6.3, together with the corresponding statistical error measures. Only minor statistical corrections are needed, and the remaining wind speed errors are similar to those found for other altimeters.

Collocation data for uncorrected ENVISAT wind speeds stratified with the nondimensional wave height $\tilde{H}$ are presented in Fig. 6.4. Again, a clear dependence of the regression behavior on the wave maturity ($\tilde{H}$) is found. This implies similar limitation on the usefulness of the GFO wind data as has been discussed in the context of the ERS-2 data.

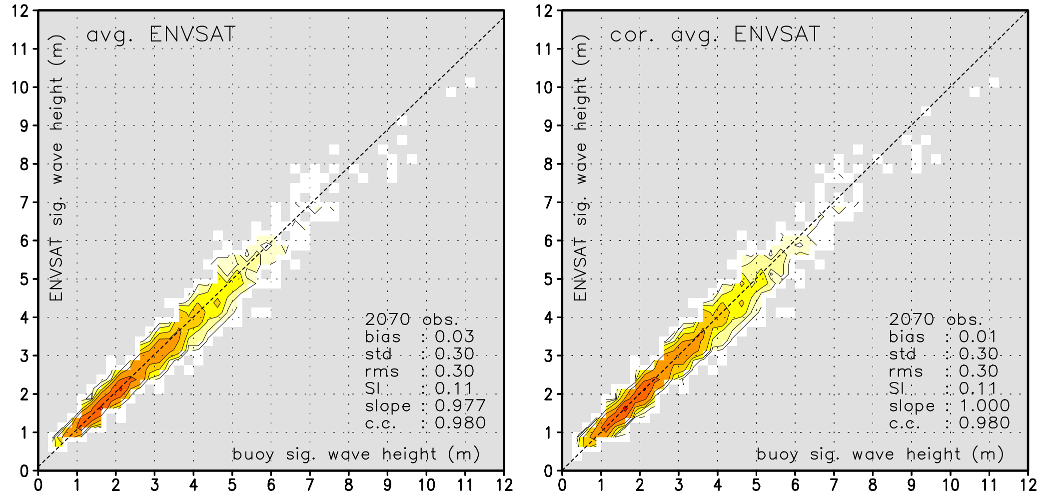

Fig. 6.1 : ENVISAT wave data collocated with buoy observations. Bulk statistics and pdf's computed using wave height increments $\Delta H = 0.25m$. The left panel represent the raw FD data, the right panel represents the data after error correction. The gray background indicates data bins $\Delta H \times \Delta H$ without collocation data. Collocation radii of 100km and 30min.

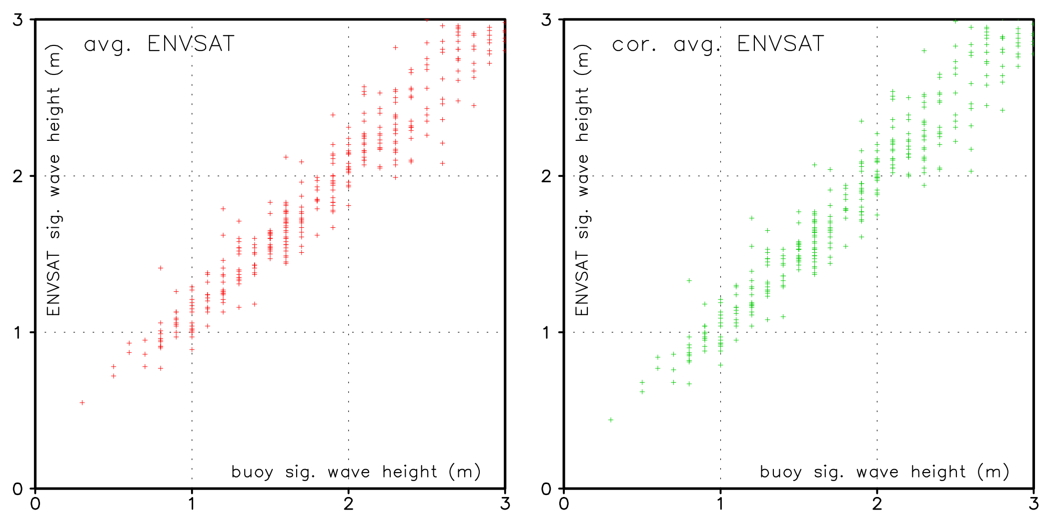

Fig. 6.2 : Scatter plot for lower wave heights corresponding to Fig. 6.1 for collocation radii of 50km and 15min.

Table 6.1: Correction parameters from Eqs. (1.9) for wave heights and (1.10) for wind speeds for the ENVISAT altimeter.

data	collocation radii (km)	(min)	a_l (m) or (ms^{-1})	b_l (-)	H_c (m)	$H_{a,0}$ (m)
waves	100	30	-0.09	1.023	1.40	0.35
waves	50	15	-0.13	1.040	1.40	0.35
winds	100	30	0.24	1.013	—	—

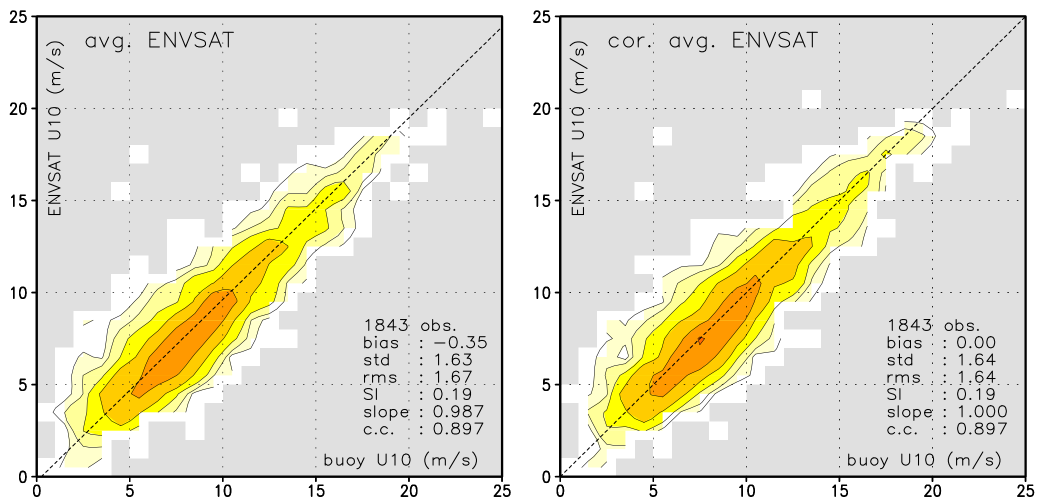

Fig. 6.3 : Like Fig. 6.1 for wind speeds with a bin width of 1ms^{-1}.

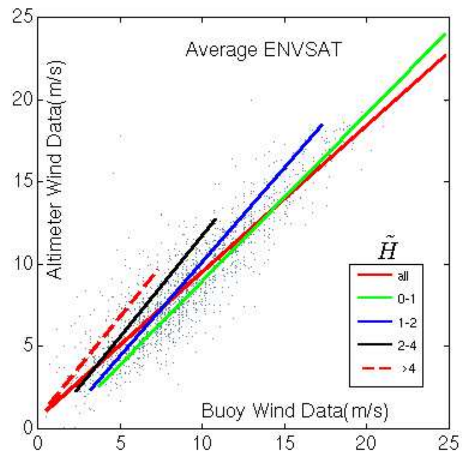

Fig. 6.4 : Regression lines for collocated ENVISAT and buoy wind data for selected ranges of the nondimensional wave height $\tilde{H}$, corresponding to the left panel of Fig. 6.3.

7 Conclusions

The present report presents a study into the quality of altimeter wave and wind data from various sources. These data are considered or used for validation of wind wave models or assimilation into wind wave models at the National Centers for Environmental Prediction (NCEP). The present study considers data from the ERS-2, TOPEX, GFO, Jason-1 and ENVISAT instruments. Some of these data are received at NCEP in near real time (so-called fast delivery or FD data), historical data have been retrieved from the NRL web sites as outlined in Table 1.1.

Altimeters directly measure the variance of the sea surface. Hence, wave height observations from altimeters can be considered as direct measurements. For all platforms and data sources (FD or QC) considered here, wave heights from altimeters were found to be highly accurate when compared to buoy data. Bias corrections as established in the present study are moderate (less than 5%). Altimeter wave height data have a possible nonlinear bias for low wave heights, related time resolution limitations of the corresponding observations. In the present study, a nonlinear bias correction for low wave heights was found to be necessary for the ERS-2 data. For all other platforms, however, such a nonlinear correction appears not to be essential.

For all platforms considered here, 10s averaged bias corrected altimeter wave heights show a rms error of 0.30m, a scatter index of 11% and a correlation coefficient of 0.98 when compared to buoy data. These results appear independent of the data source (FD or QC). Considering the sampling error in the buoys and a collocation error due to a mismatch of buoy and altimeter data in space and time, the 11% SI is close to the minimum attainable total collocation error. Subsequently, the resulting random error of the averaged and bias corrected altimeter wave data is significantly smaller. It can only be estimated roughly as less than 5%. Consequently, collocated buoy data are not accurate enough to distinguish between the quality of wave heights from different altimeters, and all these altimeter data may be considered as more accurate than buoy data, which have an inherent sampling error of approximately 8% (Donelan and Pierson, 1983). Simmilar conclusions were reached in, for instance Tolman (2002b) and in Tolman et al. (2002).

For the TOPEX altimeter, both FD and QC data have been assessed. It appears that the QC process for the TOPEX data removes a significant part of lower wave heights. Although this appears to improve error statistics, it also appears to eliminate valuable data in low wave conditions. For this reason, bias corrected FD data appears more valuable that the QC data for wave model validation and assimilation.

Altimeter winds are in principle valuable for validation of wave models, because they are collocated with wave data. Hence, they can tentatively be used to estimate the impact of local wind errors on wave model errors. However, altimeter wind data are not widely used at NCEP, mainly because there appears to be a dependency of the wind speed retrieval on the maturity of the wave fields, which can not be adequately removed (Tolman, 1998b). This implies that the quality of altimeter wind data depends on the local wave conditions both directly, and in relation to the wave climate to which the algorithm is tuned. This makes the data less reliable in general, and for application in or comparison with wave models in general. In the present study altimeter wind data from ERS-2, TOPEX, GFO and ENVISAT are all shown to share the dependency of wind retrieval on local wave conditions. Jason-1 wind data, however, does not appear to show such a dependency. Hence, Jason-1 altimeter wind data appears to be the only wind data that can be used reliably in assimilation, or on connection with wave model data assimilation or wave model validation. The different behavior of the wind speed retrievals of Jason-1 are most likely due to differences in retrieval algorithms. However, further investigation into this behavior is not considered to be a subject of the present study.

References

Chen, H. S., L. D. Burroughs and H. L. Tolman, 2003: Ocean surface waves. Technical Procedures Bulletin 494, NOAA/NWS, online[1].

Cotton, P. D. and D. J. T. Carter, 1994: Cross calibration of TOPEX, ERS-1 and Geosat wave heights. *J. Geophys. Res.*, **99**, 25,025–25,033.

Donelan, M. and W. J. Pierson, 1983: The sampling variability of estimates of spectra of wind-generated gravity waves. *J. Geophys. Res.*, **88**, 4381–4392.

Draper, N. R. and H. Smith, 1981: *Applied regression analysis*. Wiley.

Tolman, H. L., 1998a: Effects of observation errors in linear regression and bin-average analyses. *Quart. J. Roy. Meteor. Soc.*, **124**, 897–917.

Tolman, H. L., 1998b: Validation of NCEP's ocean winds for the use in wind wave models. *The Global Atmosphere and Ocean System*, **6**, 243–268.

Tolman, H. L., 2002a: Testing of WAVEWATCH III version 2.22 in NCEP's NWW3 ocean wave model suite. Tech. Note 214, NOAA/NWS/NCEP/OMB, 99 pp.

Tolman, H. L., 2002b: Validation of WAVEWATCH III version 1.15 for a global domain. Tech. Note 213, NOAA/NWS/NCEP/OMB, 33 pp.

Tolman, H. L., B. Balasubramaniyan, L. D. Burroughs, D. V. Chalikov, Y. Y. Chao, H. S. Chen and V. M. Gerald, 2002: Development and implementation of wind generated ocean surface wave models at NCEP. *Wea. Forecasting*, **17**, 311–333.

[1] http://polar.ncep.noaa.gov/mmab/tpbs/tpb494/tpb494.htm

This page is intentionally left blank.

APPENDICES

This page is intentionally left blank.

Table A.1: Collocations per buoy for ERS-2 data.

Buoy	wave heights 100km 30min	wave heights 50km 15min	wind speeds 100km 30min
21004	0	0	0
22001	1	1	0
51001	188	82	176
51002	269	161	266
51003	273	45	265
51004	271	42	264
51028	130	18	122
42001	262	112	258
42002	219	44	220
42003	205	64	216
46001	502	90	396
46002	291	0	308
46004	247	91	268
46005	382	101	311
46006	224	110	191
46035	349	44	342
46036	370	167	351
46059	299	0	284
46066	93	28	80
46184	362	75	409
41001	236	73	199
41002	282	135	234
44004	209	31	205
44008	266	0	206
44011	335	0	269
44138	92	0	69
44141	276	69	190
44142	301	41	234
62001	28	6	23
62029	325	85	296
62081	311	76	325
62105	434	94	369
62106	450	172	350
62108	397	76	263
62163	347	0	313
64045	188	0	101
62163	468	46	328
FA	127	0	104

Table A.2: Collocations per buoy for TOPEX FD data.

Buoy	wave heights 100km 30min	wave heights 50km 15min	wind speeds 100km 30min
21004	0	0	0
22001	0	0	0
51001	32	10	32
51002	33	20	32
51003	16	10	16
51004	31	0	31
51028	9	5	9
42001	18	0	18
42002	15	6	16
42003	13	9	13
46001	67	10	67
46002	29	0	29
46004	28	7	29
46005	26	9	27
46006	30	0	19
46035	22	8	21
46036	21	4	21
46059	22	10	22
46066	36	6	13
46184	29	4	50
41001	32	19	31
41002	0	0	0
44004	29	0	28
44008	27	0	26
44011	12	0	12
44138	14	2	14
44141	33	14	15
44142	31	8	31
62001	0	0	0
62029	31	16	27
62081	3	1	28
62105	20	3	19
62106	29	15	17
62108	58	0	29
62163	31	14	28
64111	0	0	0
62045	62	10	36
FA	27	7	25

Table A.3: Collocations per buoy for TOPEX QC data.

Buoy	wave heights 100km 30min	wave heights 50km 15min	wind speeds 100km 30min
21004	0	0	0
22001	0	0	0
51001	31	10	30
51002	27	17	27
51003	17	12	17
51004	33	0	33
51028	4	3	4
42001	2	0	2
42002	6	1	6
42003	4	1	4
46001	52	12	50
46002	27	0	27
46004	15	3	25
46005	26	6	26
46006	25	0	8
46035	22	6	21
46036	24	2	24
46059	15	6	15
46066	22	0	1
46184	45	7	43
41001	27	18	26
41002	0	0	0
44004	15	0	14
44008	15	0	14
44011	4	0	5
44138	4	0	4
44141	29	0	4
44142	19	6	19
62001	0	0	0
62029	32	9	28
62081	0	9	38
62105	16	6	14
62106	30	10	18
62108	65	0	39
62163	29	15	28
64111	0	0	0
62045	70	3	51
FA	23	4	23

Table A.4: Collocations per buoy for GFO data.

Buoy	wave heights 100km 30min	wave heights 50km 15min	wind speeds 100km 30min
21004	0	0	0
22001	0	0	0
51001	61	30	61
51002	83	38	83
51003	131	0	132
51004	116	20	122
51028	111	18	64
42001	96	11	107
42002	139	8	144
42003	84	8	83
46001	184	55	163
46002	94	27	119
46004	113	27	109
46005	135	31	134
46006	147	15	146
46035	243	43	233
46036	161	61	139
46059	79	36	78
46066	216	41	213
46184	150	22	228
41001	118	17	117
41002	97	29	99
44004	109	34	105
44008	129	37	110
44011	111	14	83
44138	79	20	32
44141	55	17	56
44142	114	25	68
62001	159	41	128
62029	198	39	178
62081	158	38	136
62105	155	51	140
62106	34	5	25
62108	167	59	140
62163	197	33	190
64111	0	0	0
62045	152	46	37
FA	121	14	113

Table A.5: Collocations per buoy for Jason-1 data.

Buoy	wave heights 100km 30min	wave heights 50km 15min	wind speeds 100km 30min
21004	0	0	0
22001	0	0	0
51001	121	10	120
51002	153	20	149
51003	67	10	64
51004	139	0	145
51028	73	5	38
42001	44	0	58
42002	49	6	63
42003	41	9	53
46001	253	10	219
46002	79	0	101
46004	165	7	164
46005	124	9	121
46006	127	0	124
46035	214	8	209
46036	213	4	177
46059	140	10	143
46066	195	6	198
46184	145	54	198
41001	141	19	147
41002	0	0	0
44004	119	0	120
44008	98	0	95
44011	45	0	34
44138	68	2	22
44141	50	14	47
44142	89	8	74
62001	136	0	113
62029	148	16	137
62081	214	1	189
62105	128	3	112
62106	9	15	8
62108	301	0	261
62163	138	14	136
64111	0	0	0
62045	159	10	42
FA	118	0	104

Table A.6: Collocations per buoy for ENVISAT data.

Buoy	wave heights 100km 30min	wave heights 50km 15min	wind speeds 100km 30min
21004	0	0	0
22001	0	0	0
51001	39	19	38
51002	48	0	46
51003	60	10	58
51004	57	8	52
51028	42	20	29
42001	52	6	56
42002	51	19	49
42003	50	11	50
46001	75	32	73
46002	47	21	45
46004	85	18	80
46005	68	18	63
46006	61	0	57
46035	102	29	97
46036	77	0	61
46059	63	24	62
46066	81	16	76
46184	71	14	65
41001	58	10	54
41002	49	0	55
44004	47	18	44
44008	54	18	49
44011	58	28	51
44138	35	22	5
44141	34	0	30
44142	44	18	35
62001	73	29	50
62029	65	17	51
62081	68	16	60
62105	84	23	81
62106	0	0	0
62108	98	23	87
62163	70	33	62
64111	0	0	0
62163	52	13	30
FA	52	18	42