Error Estimation of Buoy, Satellite, and Model Wave Height Data

PETER A. E. M. JANSSEN, SALEH ABDALLA, HANS HERSBACH, AND JEAN-RAYMOND BIDLOT

ECMWF, Shinfield Park, Reading, United Kingdom

(Manuscript received 23 June 2006, in final form 13 November 2006)

ABSTRACT

Triple collocation is a powerful method to estimate the rms error in each of three collocated datasets, provided the errors are not correlated. Wave height analyses from the operational European Centre for Medium-Range Weather Forecasts (ECMWF) wave forecasting system over a 4-yr period are compared with independent buoy data and dependent *European Remote Sensing Satellite-2* (*ERS-2*) altimeter wave height data, which have been used in the wave analysis. To apply the triple-collocation method, a fourth, independent dataset is obtained from a wave model hindcast without assimilation of altimeter wave observations. The seasonal dependence of the respective errors is discussed and, while in agreement with the properties of the analysis scheme, the wave height analysis is found to have the smallest error.

In this comparison the altimeter wave height data have been obtained from an average over *N* individual observations. By comparing model wave height with the altimeter superobservations for different values of *N*, alternative estimates of altimeter and model error are obtained. There is only agreement with the estimates from the triple collocation when the correlation between individual altimeter observations is taken into account.

The collocation method is also applied to estimate the error in *Environmental Satellite* (*ENVISAT*), *ERS-2* altimeter, buoy, model first-guess, and analyzed wave heights. It is shown that there is a high correlation between *ENVISAT* and *ERS-2* wave height error, while the quality of *ENVISAT* altimeter wave height is high.

1. Introduction

Satellite observations have resulted in considerable improvements to weather and wave forecasting. Because these observations have such a large potential value, it is important to validate them. A common procedure to do this is as follows. As soon as a satellite is launched and the instruments on board are performing in a stable manner, the observed products are compared with analyzed fields to check on gross errors and, if needed, to retune geophysical algorithms. The advantages of a comparison against analyses are that the quality of an analysis is fairly well known and in a relatively short period many collocations between observed quantities and the analyzed counterparts are available. Thus, a rapid assessment of the quality of satellite observations may be given. Also, the collocation between analysis and observations is of vital importance to develop geophysical algorithms such as the C-band Geophysical Model function version 4 (CMOD4; Stoffelen and Anderson 1997) and the National Aeronautics and Space Administration (NASA) Scatterometer (NSCAT; Wentz and Smith 1999) for C- and Ku-band scatterometers. Nevertheless, a check against in situ observations is an important addition to the quality assurance of satellite products, although the number of collocations is lower by typically two orders of magnitude.

However, when comparing several types of data it is desirable to have an idea about the size of the errors. Note that these errors consist of several components. The instrumental measurement error usually only gives a small contribution to the total error. More significant are representativeness errors and errors caused by the finite distance and the time between two observations.

For example, when calibrating one instrument against another it is important to know their error because the calibration constants depend on them. The example of linear regression is discussed by Marsden (1999) and Tolman (1998).

Furthermore, data assimilation requires knowledge of the weights given to the data and the first-guess (FG) field. These weights depend on the ratio between the

Corresponding author address: Peter Janssen, ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.
E-mail: p.janssen@ecmwf.int

DOI: 10.1175/JTECH2069.1

first-guess and observation errors. In wave forecasting these errors are usually not known, and one assumes, as in the optimum interpolation (OI) scheme of the European Centre for Medium-Range Weather Forecasts (ECMWF) wave forecasting system, that the errors are equal. Hence, first guesses and observations get equal weight during the analysis.

The need for estimates of errors of different data sources was realized by Stoffelen (1998). He proposed the use of a triple-collocation method to calibrate observations of winds from a scatterometer using winds from buoys, model analysis, and the *European Remote Sensing Satellite* (*ERS*)*-1* scatterometer. In his approach it was assumed that error and truth were not correlated. In a similar vein, Caires and Sterl (2003) applied a triple-collocation method to estimate and calibrate analyzed winds and wave heights from the 40-yr ECMWF Re-Analysis (ERA-40) analysis effort. Quilfen et al. (2001) followed a different approach, proposed by Freilich and Vanhoff (1999), to estimate and calibrate ERS scatterometer wind measurements over the period from 1992 to 1998. However, in this methodology the true wind speed was assumed to be Weibull distributed and the datasets were not independent because, through the data assimilation, the analyzed wind depends on both buoy and scatterometer winds. In a somewhat different context, Tokmakian and Challenor (1999) estimated errors in model and *ERS-2* and Ocean Topography Experiment (TOPEX)/Poseidon satellite mean sea level anomalies using a method that only assumes that there is no correlation between the respective errors. However, a calibration is then not possible.

It is straightforward to show that with three datasets that have uncorrelated errors, the error of each data type can be estimated from the variances and covariances of the datasets. However, unless additional assumptions are being made, it is not possible to perform a calibration among the datasets, simply because there are not enough equations. A possible way out of this dilemma is to use a minimization procedure. We therefore propose the following method to estimate errors and to calibrate datasets. We assume that the errors are not correlated and that the errors of the three datasets are estimated using the triple-collocation method. Given these estimated errors, calibration is then performed using the neutral regression approach of Deming (Mandel 1964; Marsden 1999), which is based on the minimization of the error in both variates. Using the calibrated datasets a new estimate of the errors may be obtained, resulting in new calibration constants. Thus, an iteration procedure is started and continued until convergence of the results is obtained. This is discussed in some detail in section 2.

In section 3 we apply this approach to the estimation of the wave height error of the ECMWF wave analysis. Operationally, we have available joint estimates of the true state from buoys,[1] the *ERS-2* altimeter, and the wave analysis. However, these data sources are not independent because the wave analysis uses altimeter data. One would expect, and this is common practice in meteorological data assimilation, that the first-guess field and altimeter observations may be regarded as independent. But, we argue and show that this is not an appropriate assumption in the case of *ERS-2* altimeter data. For this reason we generated a fourth independent dataset by rerunning the wave forecasting system with ECMWF-analyzed winds, but without the assimilation of *ERS-2* wave height data (this is called a hindcast). With four datasets, in which there is one independent triplet, all relevant variances and error covariances may be obtained. Monthly root-mean-square (rms) errors are obtained over a 4-yr period, and we discuss seasonal variations in the errors and calibration constants for buoys, *ERS-2* satellite, and wave analysis. The buoy errors are found to be the largest, followed by the altimeter error, while the wave analysis has the smallest error. This last finding follows from the properties of the OI scheme, which results in analysis errors being the smaller of the first-guess and observation errors. That buoy errors are the largest may perhaps come as a surprise. It is believed that a considerable part of the buoy error is related to the representativeness error, but there are also issues with quality control. By inspecting buoy time series and comparing them with wave heights from the wave analysis, a number of buoys of questionable quality could be identified. Using the collocation data with stricter quality control the resulting buoy error was reduced by 10%, while altimeter and analysis error were hardly affected. Irrespective of the quality control procedure, which is based on the consistency and visual inspection of the time series, we suspect that buoy data of questionable quality might have infiltrated our dataset. Nevertheless, it was decided that the collocation dataset, obtained by applying the stricter quality control, should be used because it gave the best results for the buoy errors.

In a collocation study, the representativeness error is

[1] In situ hourly wave observations are available via the Global Telecommunication System (GTS). They come mostly from moored buoys and platforms. A small number of observations is also available from ships. In this study we have limited our dataset to buoys and platforms, but the bulk of the data comes from buoys. Most of the buoys are in the Northern Hemisphere, and, therefore, the collocation study only concerns properties of the sea state in the Northern Hemisphere.

a serious issue that needs to be addressed. Usually, instruments and the model refer to different scales of the truth, and in order to reduce problems with representativeness, averaging the observations toward the scales as seen by the model is required. Superobservations for buoys are obtained by time averaging 5-hourly observations. The altimeter superobservations are obtained from an average of N individual observations, where for *ERS-2* $N = 8$. It is, of course, of interest to investigate the dependence of the comparison results between the model and altimeter superobservations as a function of N. It turns out that this then gives another estimate for the model and altimeter errors. Agreement with the results from the triple collocation is only obtained when correlation between the individual altimeter measurements is taken into account. For *ERS-2*, we typically find a correlation length scale of 70 km, in agreement with the practice to smooth the tracking results over 10 consecutive observations.

Finally, on 1 March 2002 the *Environmental Satellite* (*ENVISAT*) was launched and maneuvered into almost the same polar orbit as that of *ERS-2*. As a consequence, there are five collocated datasets, namely, from the *ENVISAT* and *ERS-2* altimeters, from buoys, the model's first guess, and analysis. Note that during the initial phase of the *ENVISAT* mission *ENVISAT* data are not assimilated, because first the quality of the new data needs to be monitored. It is shown that there are correlations between *ENVISAT* and *ERS-2* altimeter wave height errors because results from a triple collocation of *ENVISAT*, *ERS-2*, and buoys are not consistent with results from a triple collocation of *ENVISAT*, the model first guess, and buoy data. When correlated errors are taken into account, we find that the relative errors in wave height are respectively 6%, 6.5%, 8%, and 5% for *ENVISAT*, *ERS-2*, buoys, and wave analysis. The errors for the *ERS-2*, buoy, and analysis are in fair agreement with the operational results.

A preliminary account of this work may be found in Janssen (2004).

2. On error estimation

In this section a description is given of the method used to determine the respective errors from three independent estimates of the truth and to calibrate the data by minimization. Note that it is essential that assumptions have to be made regarding the relation between the model and observations on the one hand, and the truth on the other. At the same time this gives an implicit definition of the error. Because of an assumed relation between observation and truth, it follows that in case this relation is incorrect; the error has both a systematic and random component. Therefore, the assumption of uncorrelated errors is by no means evident, and should, if possible, be tested.

Suppose we have three estimates of the truth, denoted by X, Y, and Z, obtained from observations or from simulations of the truth by means of a forecasting system. In the following all these estimates of the truth will be referred to as measurements. Furthermore, it is assumed that the measurements depend on the truth T in a linear fashion,

$$\begin{aligned} X &= \beta_X T + e_X, \\ X &= \beta_Y T + e_Y, \\ Z &= \beta_Z T + e_Z, \end{aligned} \tag{1}$$

where e_X, e_Y, and e_Z denote the residual errors in the measurements X, Y, and Z; while β_X, β_Y, and β_Z are the linear calibration constants. Because we are estimating wave height, which is a quantity that is positive definite, no intercept is included in the model for the measurements. A finite intercept (such as that used by Caires and Sterl 2003) gives rise to negative values of either the mean value of the truth or of the measurement, which is physically impossible because of the definition of significant wave height and the way in which it is measured.

It is emphasized that the linear dependence of the measurement on the truth is an assumption that needs not to be true and, therefore, one cannot assume that the errors are purely random. For example, if there is actually a nonlinear relation between measurement and the truth but the linear calibration model (1) would be taken instead, the error will have a random and a systematic component. Furthermore, if two types of measurements have a similar nonlinear relation with the truth, then in the context of the linear model (1) there is now the possibility of correlated errors. This may be the case when intercomparing two altimeters that share the same measurement principle.

Let us now assume that the linear model (1) is valid and that the measurement results X, Y, and Z have uncorrelated errors,

$$\langle e_X e_Y \rangle = \langle e_X e_Z \rangle = \langle e_Y e_Z \rangle = 0, \tag{2}$$

where the angle brackets denote the average over a sufficiently large sample. To eliminate the calibration constants we introduce the new variables $X' = X/\beta_X$, $e'_{X'} = e_X/\beta_X$, etc., so that

$$\begin{aligned} X' &= T + e_{X'}, \\ Y' &= T + e_{Y'}, \\ Z' &= T + e_{Z'}, \end{aligned} \tag{3}$$

and the primed observations have uncorrelated errors as well. Now we eliminate the truth to obtain

$$\begin{aligned} X' - Y' &= e_{X'} - e_{Y'}, \\ X' - Z' &= e_{X'} - e_{Z'}, \\ Y' - Z' &= e_{Y'} - e_{Z'}. \end{aligned} \tag{4}$$

Then, multiplying the first with the second equation of (4) and utilizing the assumption of independent errors [(2)], one immediately obtains the variance of error in X' in terms of the variance of X' and the covariances of X' and Y', X' and Z', and Y' and Z'. In a similar manner, by multiplying the first with the third equation of (4), one obtains the variance of error in Y', while the variance of error in Z' is obtained by multiplying the second and the third equation. Hence,

$$\begin{aligned} \langle e_{X'}^2 \rangle &= \langle (X' - Y')(X' - Z') \rangle, \\ \langle e_{Y'}^2 \rangle &= \langle (Y' - X')(Y' - Z') \rangle, \\ \langle e_{Z'}^2 \rangle &= \langle (Z' - X')(Z' - Y') \rangle. \end{aligned} \tag{5}$$

Therefore, if the errors are uncorrelated, only three collocated datasets are needed to estimate the variance of the error in each of them.

The next step is to perform a calibration of the measurements. Because the truth is not known, only two of the three calibration constants can be obtained. Therefore, we arbitrarily choose X as the reference; we have the freedom to do this because results on the errors do not depend on this choice. Because the errors in the measurements are now known, the calibration constants for Y and Z may be obtained using neutral regression (Marsden 1999). As a result, the regression constant for Y becomes

$$\beta_Y = (-B + \sqrt{(B^2 - 4AC)})/2A, \tag{6}$$

where $A = \gamma\langle XY \rangle$, $\gamma = \langle e_X^2 \rangle / \langle e_Y^2 \rangle$, $B = \langle X^2 \rangle - \gamma \langle Y^2 \rangle$, and $C = -\langle XY \rangle$. Replacing Y with Z in Eq. (6) then gives the regression constant for Z.

Having performed the calibration of Y and Z it is clear that the work is not finished yet because this calibration will affect the estimation of the errors in X, Y, and Z, and hence the calibration constants, etc. Therefore, we adopted the following iteration procedure. We start with the initial guess $\beta_Y = 1$, $\beta_Z = 1$; we scale Y and Z with β_Y and β_Z, respectively; and we determine the errors using Eq. (5). A first estimate for the calibration constants follows then from Eq. (6). In the next step we scale Y and Z with the newly found estimates for β_Y and β_Z, determine the errors and the regression constant using (5)–(6), and we continue until convergence is achieved. By comparing results from a different number of iterations it was found that after 10 iterations an accuracy of four significant digits was achieved.

Note that only a relative calibration is possible. Nevertheless, results on errors do not depend on the chosen reference standard. This was checked by choosing Y instead of X as a reference standard; errors were identical up to four significant digits and, as expected, the calibration constant of X was the inverse of the calibration constant of Y when X was chosen as a reference.

In the remainder of this paper we use the triple-collocation scheme [(5)–(6)] as a basic tool to try to understand relations and errors among collocated datasets. In section 3 we apply this approach to the estimation of wave height error in the ECMWF wave analysis, in buoys, and in *ERS-2* wave heights over a 4-yr period; while in section 4 we use wave height results obtained during the *ENVISAT*–*ERS-2* tandem mission and apply the triple-collocation approach to show that there are significant correlations between the altimeters from *ERS-2* and *ENVISAT*.

3. On errors in the operational ECMWF wave analysis

Operationally, joint estimates of the true state from buoys, the *ERS-2* altimeter (as they arrive in near–real time through the GTS), and the wave analysis are available, but these data sources are not independent because the wave analysis uses altimeter data. Therefore, another dataset is needed, which is independent from the observations. A possible choice would be to use the first-guess field as an independent dataset. It is common practice in data assimilation to regard the first-guess field and observations as independent, because these observations have not yet been used in the analysis and forecasting scheme. This is plausible when an instrument has only random errors, but it may be problematic in the case of systematic errors in the observations. For example, it is well known that the so-called fast-delivery significant wave heights from the *ERS-2* altimeter have a systematic error for low wave heights, simply because this altimeter does not produce wave heights lower than about 60 cm. This is illustrated in Fig. 1 where over a 4-yr period a comparison is shown between wave height data from the *ERS-2* altimeter superobservations and buoy data. Note the overestimation of wave height by the altimeter for wave heights below 1.5 m.

Suppose these erroneous low wave height data were assimilated in the model. Because low wave heights usually correspond to swell conditions and swell has a long memory, it is likely that the following first-guess

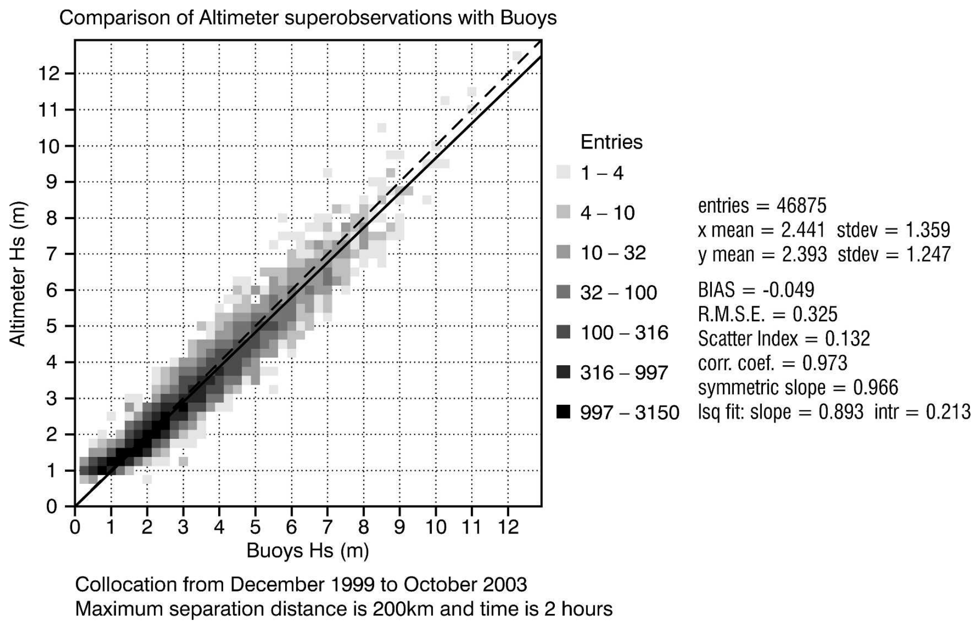

FIG. 1. Comparison of altimeter wave height data with buoy data for the period from 1 Jan 2000 until December 2003.

field is contaminated by the wrong data. Hence, first-guess and altimeter data might have correlated errors.

For this reason, we generated a fourth independent dataset by running the wave forecasting system with 6-hourly ECMWF-analyzed winds over the 4-yr period of January 2000 until December 2003 (after a 2-month warm-up period to eliminate any altimeter impacts). No *ERS-2* altimeter data were assimilated so that the hindcast results are independent of altimeter and buoy wave height data. From the operational results and the hindcast the following collocated dataset was generated: X = hindcast, Y = altimeter, Z = buoy, V = first guess, and W = analysis. We adopt model (1) for these datasets. Noting that the buoy data are not used in the analysis, we allow for error correlations between X and V, X and W, Y and V, Y and W, and finally between V and W. In other words, the covariances

$$\langle e_{X'}e_{V'}\rangle, \langle e_{X'}e_{W'}\rangle, \langle e_{Y'}e_{V'}\rangle, \langle e_{Y'}e_{W'}\rangle, \langle e_{V'}e_{W'}\rangle \qquad (7)$$

are finite. Here, a prime again denotes scaling with the slope β. Together with five error variances this gives 10 unknowns, while by correlating the differences $X' - Y'$, $X' - Z'$, $X' - V'$, $X' - W'$, etc., there are $4 + 3 + 2 + 1 = 10$ equations, so that all the unknowns may be determined. Other choices of finite covariances in (7) do not have any physical support. To test this, we falsely assumed that there is no correlation between altimeter and first-guess errors, while allowing a finite correlation between the hindcast and altimeter errors. This resulted in a negligibly small (and even negative for a few months) correlation between the hindcast and altimeter errors.

This large set of equations may be solved in a straightforward manner by realizing that the first three measurements are independent; hence, the triple-collocation technique may be applied and their errors follow from Eq. (5). The errors for the remaining measurement types, which have correlated errors with each other and the altimeter and hindcast, follow from two steps. For example, the variance of the analysis error follows by correlating $W' - Y'$ and $W' - Z'$. Using (7) we find

$$\langle e^2_{W'}\rangle = \langle e_{Y'}e_{W'}\rangle + \langle (W' - Y')(W' - Z')\rangle, \qquad (8)$$

and the correlation between W and Y follows from correlating $X' - W'$ and $Z' - Y'$. Hence,

$$\langle e_{Y'}e_{W'}\rangle = \langle (X' - W')(Z' - Y')\rangle. \qquad (9)$$

The error and correlation with the altimeter for the first-guess V follows from (8)–(9) by replacing W with V.

Before we present results on error estimation, it is important to discuss the collocation method. First, each estimate represents a different aspect of the truth. The global ECMWF wave prediction system has a spatial resolution of 55 km, and is forced by atmospheric winds that before 21 November 2000 had a spatial resolution of 65 km, but after that date increased to 40 km. However, horizontal diffusion in the atmospheric model reduces activity at the short scales considerably, and also the first-order advection scheme in the wave model may give rise to smoother wave fields. Hence, in practice the model wave height fields only properly represent spatial scales larger than about 100 km. The *ERS-2* altimeter measures significant wave height every 7 km, while buoys typically produce hourly measurements that are 20-min averages. Clearly, instruments and the model refer to different scales of the truth, and in order to reduce problems with representativeness, averaging of the observations toward the scales as seen by the model is required. Superobservations for buoys are obtained by time averaging 5-hourly observations following the quality control procedure of Bidlot et al. (2002), except that the time window is centered on the altimeter superobservations time. The altimeter superobservations are obtained from an average of N individual observations, where operationally $N = 8$. In addition, an along-track quality control is also used to remove all spurious altimeter observations.

Second, each estimate refers to a slightly different location within 200 km and time within 2 h. To alleviate this problem, the wave model field is linearly interpolated in space and time toward the altimeter and buoy observations. Hence, we deal with two model counterparts—one referring to the altimeter observation, denoted by $X_{\rm alt}$, etc., and one referring to the buoy observation, denoted by $X_{\rm buoy}$, etc. In this collocation study the model value is taken as the mean of $X_{\rm alt}$ and $X_{\rm buoy}$. Nevertheless, a collocation error between the altimeter and buoy remains. Using the difference $X_{\rm alt} - X_{\rm buoy}$, the collocation error can be estimated, however. To ensure that the error estimation is not affected by the collocation error, only collocations satisfying a relative difference of at most 5% are considered. This corresponds to a relative collocation error, defined as the rms difference normalized with the average model wave height of 1%–2%. Because of this restriction, the number of collocations reduces by 50% from about 30 000 to 16 000.

a. Results

Results over the period from January 2000 until November 2003 of monthly relative error (obtained from a 3-monthly running average) for first-guess, analyzed, *ERS-2* altimeter, and buoy wave height are shown in Fig. 2. Here, the relative error is defined as the ratio of rms error to the mean wave height. This ratio is usually called the scatter index (SI). It is striking that these errors are relatively small, with the buoy errors being the largest while the analysis errors are the smallest. The high-quality analysis is a consequence of the properties of the OI scheme that is used to produce the wave analysis.

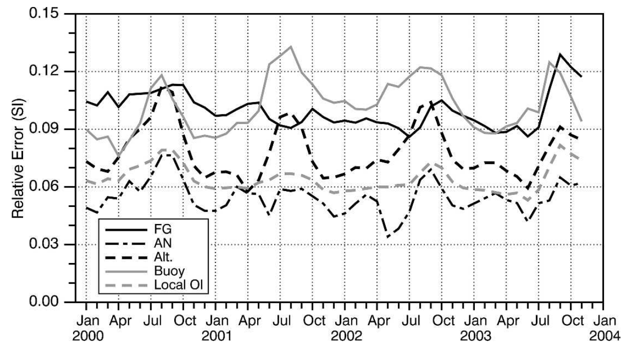

FIG. 2. Monthly relative error of first-guess (FG), analyzed (AN), *ERS-2* altimeter (Alt), and buoy wave height. Maximum relative collocation difference is 5%. For comparison the analysis error according to a local OI scheme is shown as well.

To see this we discuss the well-known case of the assimilation of a single measurement. Results can be generalized straightforwardly to the case of the assimilation of many observations. Let us denote the analysis by A, the observation by O, the first guess by F, and their corresponding errors by e_A, e_O, and e_F. According to OI, the analysis is given by a linear combination of model first guess and observation,

$$A = wO + (1 - w)F, \tag{10}$$

and the weight w is chosen in such a way that the analysis error is minimal. By subtracting the truth and assuming no systematic errors, hence the error model (1) with $\beta = 1$ is adopted, the analysis error becomes

$$e_A = we_O + (1 - w)e_F, \tag{11}$$

and assuming that there is no correlation between the first-guess and observation errors, $\langle e_O e_F \rangle = 0$, the mean-square analysis error becomes

$$\langle e_A^2 \rangle = w^2 \langle e_O^2 \rangle + (1 - w)^2 \langle e_F^2 \rangle. \tag{12}$$

The analysis scheme is optimal when the mean-square analysis error is minimal. By differentiating the mean-square error with respect to the weight, one finds for w at the minimum

$$w = \frac{\langle e_F^2 \rangle}{(\langle e_O^2 \rangle + \langle e_F^2 \rangle)}, \tag{13}$$

while the analysis error in terms of observation and first-guess error becomes

$$\frac{1}{\langle e_A^2 \rangle} = \frac{1}{\langle e_O^2 \rangle} + \frac{1}{\langle e_F^2 \rangle}. \tag{14}$$

Therefore, the analysis error is the smaller of the observation and first-guess error.

Note that the result in (14) is based on the assumption that there is no bias and no correlation between first-guess and observation errors, while it is also assumed that observation error and first-guess errors are known. We will see in a moment that the bias is relatively small (at least with respect to the buoys). Also, the correlation between first-guess and observation error is at most 20%. However, the reason for the present study is to obtain information on the first-guess and observation errors, and therefore in the ECMWF wave analysis scheme these errors are not known. Instead, for wave heights larger than 1.5 m, observations and the first guess are given equal weight, therefore $w = 0.5$, while for wave heights below 1.5 m the weight w is reduced because of the known problems with the altimeter (cf. Fig. 1). In other words, the analysis error is given by Eq. (12), and the results of this expression are plotted in Fig. 2. The errors from the local OI scheme are slightly larger than the analysis error as obtained through the triple-collocation method. This is most likely caused by the fact that the actual analysis is not local, that is, the quality of the analyzed wave height benefits from remote observations as well.

By comparing Eq. (12) with given weight, say $w = 0.5$, and Eq. (14), the importance of knowing $\langle e_F^2 \rangle$ and $\langle e_O^2 \rangle$ becomes apparent. This is most easily seen by taking the extreme case of large observation error. In the OI scheme with weights given by Eq. (13) the observations get zero weight, and hence the analysis error is given by the first-guess error. In contrast, in the usual application of OI with given weights the analysis error is dominated by the observation error. Evidently, (14) with the choice of weight (13) is optimal, while (12) with constant weight is not.

Let us now study some further results of our statistical analysis. In Fig. 3 monthly time series of the calibration constant β for first-guess, analyzed, and altimeter wave heights are shown. As a reference we have chosen the in situ buoy wave height data. The largest correction is required for the first-guess wave height, but over this 4-yr period improvements are clearly visible. This change is most likely to be caused by the introduction of the T_l 511 atmospheric model, together with a doubling of the angular resolution of the wave spectrum on 21 November 2000. In particular, the increased spatial resolution has resulted in stronger surface winds, giving higher wave heights. The smallest correction, of the order of 4%, is needed for the *ERS-2* altimeter wave height data. In an operational wave forecasting system, such as the one at ECMWF, normally fast-delivery products from the European Space

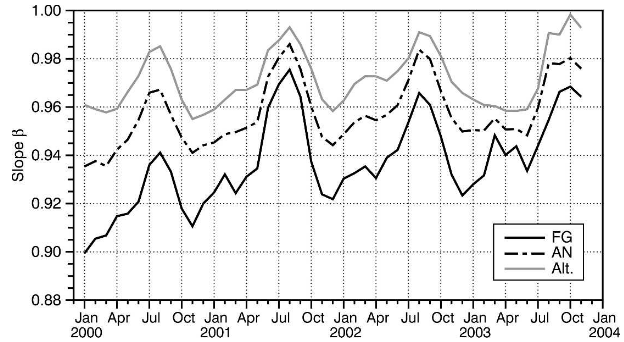

FIG. 3. Monthly calibration constants β for altimeter, first-guess, and analyzed wave height over the period from January 2000 until November 2003. Reference is the buoy wave height.

Agency (ESA) are assimilated. The fast-delivery wave heights are usually about 7%–8% lower than the buoy wave heights (Janssen 2000). However, at ECMWF, the fast-delivery wave height data are corrected for the non-Gaussian nature of the ocean surface, resulting in increased wave height by 3%–4% (Janssen 2000).

As noted before, in analysis schemes it is usually assumed that first guesses and observations have uncorrelated errors. With the present collocation study it is possible to estimate the correlations [cf. Eq. (9)]. In Fig. 4 monthly time series of the correlation between first-guess and altimeter errors is shown, and as a reference the correlation between first-guess and hindcast errors is also shown. Note that, in particular, during the Northern Hemisphere (NH) summer months when low wave heights and therefore swells prevail, there is a considerable correlation between the first guess and altimeter.

Last, it is noted that, with the exception of the first-guess error, there is a clear seasonal cycle in the statistical results. One needs to keep in mind that, because of the location of the buoy measurements, the present collocation study is restricted to the NH only. Typically, the relative error or scatter index is the largest during

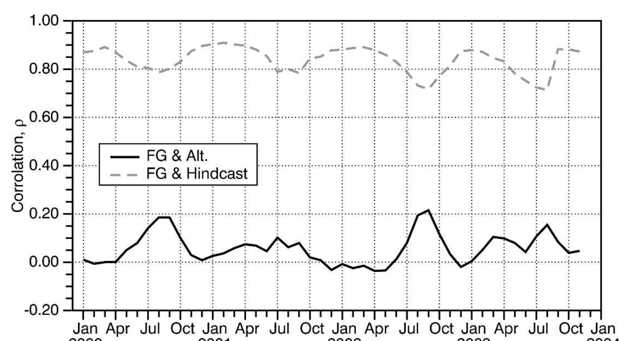

FIG. 4. Monthly correlation between first-guess and altimeter wave heights errors as compared to the correlation between first-guess and hindcast errors.

the NH summertime when, on average, wave height is low. This can be understood for the altimeter data (and, as a consequence, for the analysis) because of the already mentioned problems at low wave height. However, it is not clear why buoy data have larger relative errors during the summertime compared to wintertime. Nevertheless, one would expect the large buoys to have problems with estimating small wave heights.

b. Correlation between individual altimeter observations

An important issue to address is how the statistical results given in the previous section depend on the procedure to obtain, for example, the altimeter superobservations. Recall that individual altimeter wave height observations are obtained every 7 km and that operationally a superobservation involves an average over $N = 8$ observations. Hence, formally, the altimeter superobservations refer to a smaller spatial scale (56 km) than the model analysis or first guess (~100 km). Therefore, the statistical results could depend on the number N of observations used to make the superobservation, but it turns out that the dependence of results on N is weaker than expected.

Assume for the moment that the individual altimeter observations are independent from each other; that is, there is no correlation between the errors. Also, assume that first-guess and altimeter wave height errors are not correlated (this can be achieved by choosing a winter month, cf. Fig. 4). Let σ be the rms error resulting from the comparison of first-guess and altimeter wave height, then

$$\sigma^2 = \langle e_F^2 \rangle + \frac{1}{N}\sigma_a^2, \tag{15}$$

where $\langle e_F^2 \rangle$ is the variance of the first-guess error, while σ_a is the error of an individual altimeter wave height measurement. Note that $\langle e_A^2 \rangle = \sigma_a^2/N$ in case the individual altimeter observations errors are uncorrelated.

Therefore, Eq. (15) suggests a sensitive dependence of the rms error σ on the number of observations used in the averaging. At the same time, it was realized that Eq. (15) suggests another method to estimate first-guess error and the altimeter error: do the comparison between first-guess and altimeter data for different N and plot the results for σ^2 as function of $1/N$. The intercept at $1/N = 0$ then gives the first-guess error while the slope gives the error of the individual altimeter measurement. We therefore redid the intercomparison between the first-guess and altimeter wave height data for the month of December 2001 for different values of

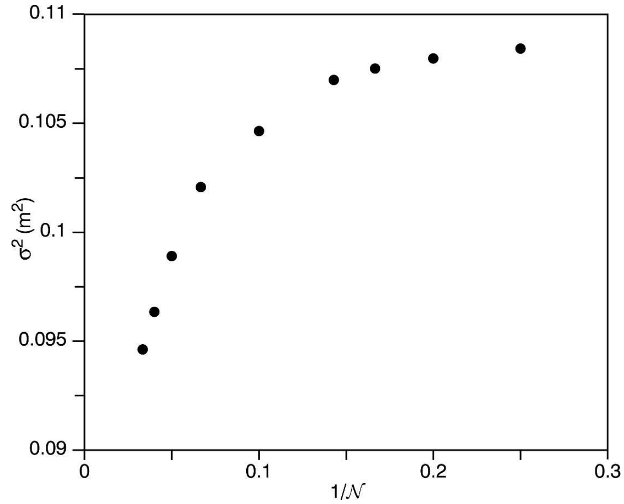

FIG. 5. Comparison of *ERS-2* altimeter superobservations with the first-guess wave heights for December 2001 for the global area. Dependence of the square of the rms error σ^2 on the inverse of the number N of observations used in the altimeter superobservations. If individual altimeter observations are independent, a linear dependence is expected.

N. For the global area the rms error as function of $1/N$ was determined. The results of this exercise are shown in Fig. 5. In contrast to expectation, the error variance σ^2 is not increasing linearly with $1/N$, and there are clear signs of saturation when only a few observations are used in the averaging. This is an indication that the error in the individual altimeter observations is spatially correlated.

Let us explore the consequences of spatially correlated errors. Consider a superobservation A_{sup} defined as the spatial average of a number of N individual observations a_i taken at location i, or,

$$A_{\mathrm{sup}} = \frac{1}{N}\sum_i a_i. \tag{16}$$

Suppose that each individual observation obeys the error model (1) with $\beta = 1$. Then the variance of the error in the superobservation becomes, in general,

$$\langle e_A^2 \rangle = \frac{1}{N^2}\sum_{i,j} \langle e_{a_i} e_{a_j} \rangle. \tag{17}$$

In the special case for which the individual observations have no spatial correlation, the correlation matrix $\langle e_{a_i} e_{a_j} \rangle$ has only diagonal elements, which all equal to σ_a^2, where σ_a is the error of the individual observation. In that event the sum in Eq. (17) equals N and the variance of error in the superobservation becomes $\langle e_A^2 \rangle = \sigma_a^2/N$, in agreement with Eq. (15). Assuming a positive definite correlation function, spatially correlated errors will give rise to a sum that is larger than N and therefore, as expected, the effective number of degrees of freedom will be less than N.

To make progress, a simple correlation model will be adopted. Initially, we tried a Markovian model, which corresponds to a correlation function that decays exponentially with distance. This model worked fairly well but was deficient in cases of small and large numbers of observations involved in the superobservation. Better results were obtained with a correlation model that resembles a Gaussian

$$\langle e_{a_i} e_{a_j} \rangle = \sigma_a^2 c^{(i-j)^2}, \tag{18}$$

where c is the constant correlation coefficient between two neighboring observations. The corresponding correlation length scale l_c, defined as the distance over which the correlation drops to e^{-1}, is then given by $l_c^2 = -1/\log(c)$. Substitution of (18) in (17) gives

$$\langle e_A^2 \rangle = \frac{\sigma_a^2}{N^2}[N + 2(N-1)c + 2(N-2)c^4 + \cdots]. \tag{19}$$

After some rearrangement the final result becomes

$$\langle e_A^2 \rangle = \frac{\sigma_a^2}{\mathcal{N}}, \tag{20}$$

where $\mathcal{N}$ is the effective number of degrees of freedom,

$$\frac{1}{\mathcal{N}} = \frac{1}{N}\left[1 + 2\sum_{l=1}^{N}\left(1 - \frac{l}{N}\right)c^{l^2}\right], \tag{21}$$

which in case of positive correlation is less than the number of observations N used in the averaging. Hence, in case of correlated errors Eq. (15) is replaced by

$$\sigma^2 = \langle e_F^2 \rangle + \frac{1}{\mathcal{N}}\sigma_a^2. \tag{22}$$

One still needs to determine the correlation coefficient c. This was done by trial and error, insisting that in agreement with (22) a linear relation exists between σ^2 and the inverse of the effective number of degrees of freedom $1/\mathcal{N}$. Best results are obtained with a correlation coefficient c of 0.99; in other words, the spatial correlation scale is about 70 km ($l_c = 10$ observations). Results for this choice of correlation coefficient are shown in Fig. 6. Fitting the results with a linear function it is then found that with a mean wave height of 2.5 m the first-guess SI is 0.11, which is in fair agreement with the result from the quintuple-collocation study of the previous section that gave a first-guess SI of about 0.094 for December 2001. The SI for the individual *ERS-2* altimeter measurements is found to be 0.067. In the quintuple-collocation study the altimeter superobservation consisted of an average of $N = 8$ individual observations. With a correlation of 0.99 the effective number of degrees of freedom $\mathcal{N}$ becomes about 1.10, giving an SI of the altimeter superobservation of 0.064, which compares favorably with the result from the quintuple study where an SI of 0.065 was found. Note that results for the SI of the altimeter superobservation are fairly insensitive to the averaging number N. In view of matching spatial scales with the model it would have been more appropriate to use $N = 15$ in the averaging of the altimeter data. However, this only decreases the SI of the altimeter superobservation from 0.064 to 0.058. There is no need to emphasize that this weak dependence on N is caused by the significant spatial correlation between individual altimeter observations.

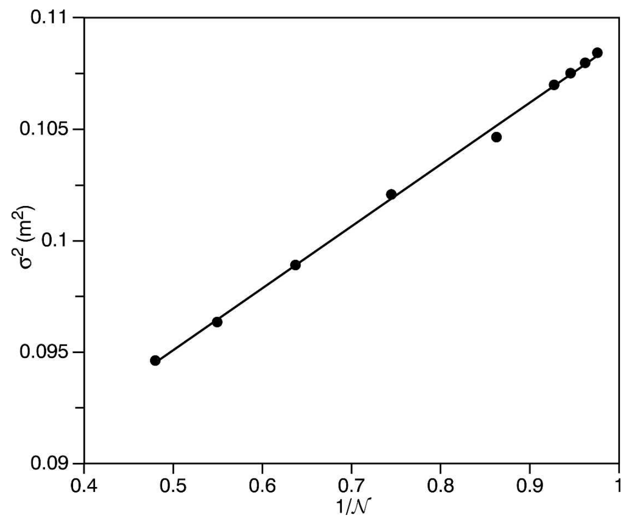

FIG. 6. The same as in Fig. 5, but as a function of the effective number of degrees of freedom $\mathcal{N}$, where a correlation of 0.99 is taken between neighboring altimeter observations. A best fit gives a scatter index of 11% for first guess and 6.7% for the individual altimeter observations.

It is important to try to understand why there is such a large spatial correlation between errors in the individual *ERS-2* altimeter observations. To be able to make a wave height observation an altimeter needs to track the ocean surface. In the case of the *ERS-2* altimeter, the tracking results are smoothed over 10 individual observations. This results in spatially correlated errors in significant wave height and makes it plausible as to why we find a correlation scale of about 70 km. It would be desirable to test whether indeed the smoothing of the tracker results causes spatial error correlation by varying the length of the smoothing filter. Now, the altimeter on board Jason does not average the tracker results but averages wave height results over five indi-

TABLE 1. Scatter index results of several triple collocations obtained during the *ENVISAT* commissioning phase for wave height larger than 1 m. Here, *N* denotes the number of collocations.

	N	*ENVISAT*	*ERS-2*	First guess	Analysis	Buoy
ENVISAT–*ERS-2*–buoy	6062	2.5%	3.2%	—	—	9.9%
ENVISAT–first guess–buoy	6062	6.1%	—	9.7%	—	8.2%
ENVISAT–analysis–buoy	6062	5.4%	—	—	4.0%	8.6%

vidual observations. Jean-Michel Lefèvre (2002, personal communication) studied results from the collocation between Jason altimeter wave height data and ECMWF wave height analysis, and he plotted the dependence of the variance of the total error σ^2 as function of the effective number of degrees of freedom $\mathcal{N}$. In the case of Jason, he found a spatial correlation scale of about 30 km (five observations).

It is concluded that the present method and the triple-collocation technique give consistent results for first-guess and altimeter superobservations, provided a significant spatial correlation between the errors of individual altimeter observations is taken into account. The reason for the correlation is most likely the smoothing of the results, because the correlation scales for *ERS-2* and Jason are found to correspond with the length of the smoothing filter.

4. Validation of *ENVISAT* altimeter wave height data

On 1 March 2002 the ESA launched *ENVISAT*, which carries on board nine instruments, two of which are relevant for understanding ocean waves, namely, the RA-2 altimeter and the advanced synthetic aperture radar (ASAR). We discuss herein the quality of the altimeter wave height results. The RA-2 altimeter is a dual-frequency altimeter, but only Ku-band results will be studied.

ENVISAT was maneuvered in such a manner that it has an almost identical orbit as the *ERS-2* satellite. The time difference between observations from the two satellites is, therefore, only 20 min. Hence, this provides a unique opportunity for validation, because we have five collocated datasets available, namely, from the *ENVISAT* and *ERS-2* altimeter, from buoys, from model first guess and analysis (quintuple collocation). The period is from 18 July until 20 October 2003. During this period *ERS-2* data were assimilated into the ECMWF wave forecasting system, but *ENVISAT* data were not because their quality was being monitored and assessed. In June 2003 the tape drives on *ERS-2* packed in and observations could not be stored on board any more. Only when *ERS-2* is in view of a ground station it is possible to transmit observed data to earth. From then onward, *ERS-2* has only provided observations for part of the North Atlantic area and for the west coast of North America. Fortunately, because many buoys are located in these areas the number of collocations during the summer period in 2003 did not reduce dramatically (except for a period of 3 weeks after the incident), so that it was still possible to do a valid calibration study. However, *ERS-2* lost global coverage, which prompted ECMWF to commence with the assimilation of *ENVISAT* data in October 2003. Clearly, after this date the calibration method described below cannot be applied because of correlations between the *ENVISAT* data and the first-guess wave height data.

Let us first discuss how it was realized that errors in *ENVISAT* and *ERS-2* altimeter are correlated. Because there are five collocated datasets available there are several opportunities to apply the triple-collocation method. For example, one might assume that the errors in *ENVISAT* and *ERS-2* altimeter wave height are not correlated and that they are not correlated with the buoy errors. Other possibilities are the triplet *ENVISAT*, buoy, and first guess, and the triplet *ENVISAT*, buoy, and analysis. All of these triplets have *ENVISAT* and buoy wave height data in common. Results of the three triple-collocation exercises are given in Table 1. It is evident from Table 1 that, in particular, the results for the relative error in the *ENVISAT* altimeter wave height data are not consistent. Also, the relative error in the *ERS-2* wave height data is much smaller than found during the 4-yr period analyzed in the previous section.

The inconsistency in Table 1 is plausible when it is realized that our assumption of uncorrelated *ENVISAT* and *ERS-2* errors might be incorrect. Clearly, additional information is needed to determine correlations between the errors of different observations. With five collocated datasets and assumptions on which observation type is correlated it is just possible to obtain all relevant variances and covariances. To this end it is assumed that buoy errors are not correlated with errors of any other observation type. We have seen in the previous section that first-guess and *ERS-2* errors are correlated because of the systematic prob-

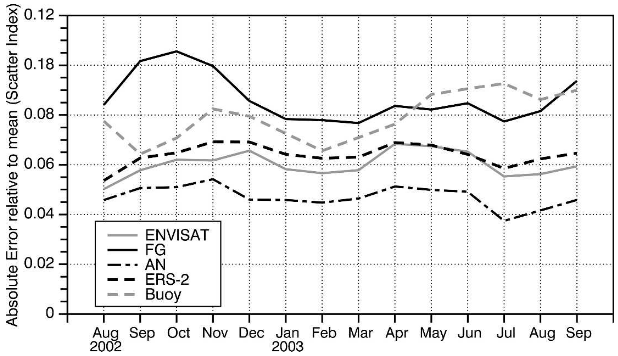

FIG. 7. Monthly relative error of *ENVISAT*, first-guess, analyzed, *ERS-2* altimeter, and buoy wave height over the period from August 2002 until September 2003. The maximum relative collocation difference is 5%.

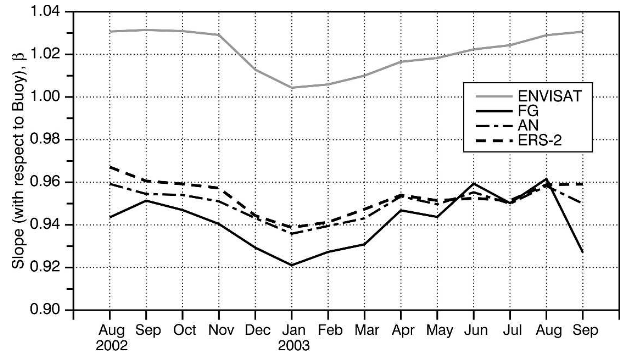

FIG. 8. Monthly calibration constants β for *ENVISAT* and *ERS-2* altimeter, first-guess, and analyzed wave height over the period from August 2002 until September 2003. Reference is the buoy wave height.

lems in the *ERS-2* altimeter at low wave height. But, it is unlikely that the first-guess error is correlated with the *ENVISAT* error because the *ENVISAT* altimeter does not suffer from the *ERS-2* problems at low wave height. We therefore assume that the triplet *ENVISAT*, buoy, and model first guess is independent, and therefore their errors can be determined by means of the triple-collocation method.

Furthermore, correlation between *ENVISAT* and *ERS-2* errors is allowed. This implies that *ENVISAT* and analysis errors are correlated as well, because due to the assimilation, analysis and *ERS-2* errors are correlated. Last, first-guess and analysis error are evidently correlated. As a result we have 10 unknowns, namely, 5 variances and 5 covariances, and just as in the previous section there are 10 equations. Hence, the problem is solvable.

The collocation dataset is obtained with the same procedure as the operational dataset from the previous section, except that the *ENVISAT* superobservations are averaged over 11 individual observations. Only results for wave heights larger than 1 m are presented so that *ERS-2* altimeter wave height data will not be penalized because of the low wave height problems. When correlated errors are taken into account we find that for the whole period the relative errors in wave height are 6.1%, 6.4%, 8.2%, and 4.9% for, respectively, *ENVISAT*, *ERS-2*, buoys, and wave analysis. The results for the correlations are, in descending order, *ENVISAT* and *ERS-2* = 79%, first guess and analysis = 73%, analysis and *ERS-2* = 39%, analysis and *ENVISAT* = 26%, and first guess and *ERS-2* = 3%. Note the high correlation between *ENVISAT* and *ERS-2* altimeter wave height errors. A likely reason for the high correlation is that both instruments share the same measurement principle, although the *ENVISAT* altimeter has an improved treatment of the wave form. However, a thorough study is required to understand better the systematic part of the altimeter wave height error.

Last, in Fig. 7 we show monthly time series of *ENVISAT*, first-guess, analysis, *ERS-2*, and buoy relative wave height error. It confirms that errors for *ERS-2*, the first guess, and analysis are consistent with the findings of the previous section. It also shows the high quality of the *ENVISAT* altimeter wave height results. In Fig. 8 the monthly time series for the slopes of *ENVISAT*, *ERS-2*, first guess, and analysis are shown. In agreement with the operational results, slopes for *ERS-2*, first guess, and analysis are less than unity when compared with the buoys, and hence, according to this standard, are underestimating wave height. In contrast, the *ENVISAT* altimeter wave heights are higher by about 2% on average.

Note that compared to the study in the previous section there are two important differences. First, in the *ENVISAT* study only wave heights larger than 1 m were considered, and as a consequence the seasonal cycle in the *ERS-2* and buoy wave heights is reduced (cf. Figs. 7 and 2). Second, because in the early period only a limited number of *ENVISAT* wave height observations were received in near–real time, the number of collocations per month is at least smaller by a factor of 2 compared to the study of section 3. This has affected in particular the buoy error.[2] Compared to the altimeter and the model wave height error, the buoy error is sensitive to details such as the number of collocations and geographical location. It should be realized that the buoy network is not homogeneous be-

[2] This was checked by redoing the calibration study of the previous section for those collocations that were present in the *ENVISAT* study only. The buoy error was reduced by about 20% and became similar to the one found in the *ENVISAT* study.

cause there are different type of buoys depending on the location and buoy network operator. Different buoy types have different error characteristics.

To conclude, we have studied the accuracy of *ENVISAT* altimeter wave height. The impression is that the quality of these data is high. In addition, it has been shown that there is a significant correlation between the errors of the *ENVISAT* and *ERS-2* altimeters, presumably because they share the same measurement principle. This requires a study of the systematic error of these altimeters, but it should be emphasized that the error level is small and therefore the systematic error is only a minor problem.

5. Conclusions

We have used a triple-collocation method to estimate the rms error of collocated wave height datasets, and we have combined this with a neutral regression approach to obtain a relative calibration of the datasets. The basic assumption of this method is that the three datasets have uncorrelated errors.

We have applied this approach to the estimation of wave height error in the ECMWF operationally analyzed and first-guess wave fields and in buoy and *ERS-2* altimeter wave height observations, while we also applied this approach to results from the *ENVISAT–ERS-2* tandem mission.

At this time the wave analysis at ECMWF uses *ERS-2* altimeter data and therefore an additional dataset is required, which is independent of the observations. This was provided by a hindcast with the wave model over a 4-yr period, rather than by the first-guess fields, because a correlation with the altimeter data was suspected. Results from the application of the triple-collocation method show that there is indeed a correlation between first-guess error and *ERS-2* altimeter wave height error of at most 20%. Time series of the monthly errors show that the relative error is typically between 5% and 10%, which may be regarded as small. We also studied the dependence of the statistical results on the number N of individual altimeter observations involved in the altimeter superobservation. This resulted in an alternative method to estimate the first-guess error and altimeter error. Results are consistent with the triple-collocation method, provided that one assumes a considerable correlation between the individual altimeter observations. This correlation is plausible when it is realized that *ERS-2* altimeter tracker results are smoothed over 10 consecutive observations. For Jason, wave height results are smoothed over five observations and as a consequence the correlation length scale for Jason altimeter wave height is half of the one from *ERS-2* (J.-M. Lefèvre 2002, personal communication).

Results from the *ENVISAT–ERS-2* tandem mission do suggest that the *ENVISAT* altimeter wave heights are of high quality. It is also found that there is a high correlation between the *ENVISAT* and *ERS-2* altimeter wave height errors.

Finally, although the triple-collocation method is a powerful tool it should be realized that one cannot apply this approach blindly. It is important to point out that this method may only be applied when datasets may be regarded as independent. Otherwise, inconsistencies may result, as is evident when the method was applied to the *ENVISAT–ERS-2* tandem mission. But, the restriction to independent datasets is not a weakness of the triple-collocation method, rather it is a strong point; because of the inconsistencies in results it was realized that there was a correlation between *ENVISAT* and *ERS-2* altimeter wave height error.

Acknowledgments. The authors appreciate the suggestion by Anthony Hollingsworth that individual altimeter observations may be correlated because of the smoothing of the tracker results. Martin Miller is thanked for critically reading the manuscript. Also, discussions with Bertrand Chapron and Peter Challenor are much appreciated. The altimeter data were provided by ESA and the buoy data were provided by the National Data Buoy Center (NDBC), United States, the Marine Environmental Data Services (MEDS), Canada, the Met Office, United Kingdom, Metéo-France, France, Met Eireann, Ireland, DNMI, Norway, and the Icelandic Maritime Administration, Iceland.

REFERENCES

Bidlot, J.-R., D. J. Holmes, P. A. Wittmann, R. Lalbeharry, and H. S. Chen, 2002: Intercomparison of the performance of operational ocean wave forecasting systems with buoy data. *Wea. Forecasting,* **17,** 287–310.

Caires, S., and A. Sterl, 2003: Validation of ocean wind and wave data using triple collocation. *J. Geophys. Res.,* **108,** 3098, doi:10.1029/2002JC001491.

Freilich, M. H., and B. A. Vanhoff, 1999: QuikScat vector wind accuracy: Initial estimates. *Proc. QuikScat Cal/Val Early Science Meeting,* Pasadena, CA, Jet Propulsion Laboratory.

Janssen, P. A. E. M., 2000: ECMWF wave modeling and satellite altimeter wave data. *Satellites, Oceanography and Society,* D. Halpern, Ed., Elsevier Oceanography Series, Vol. 63, Elsevier, 35–56.

——, 2004: *The Interaction of Ocean Waves and Wind*. Cambridge University Press, 300 pp.

Mandel, J., 1964: *The Statistical Analysis of Experimental Data*. Interscience, 410 pp.

Marsden, R. F., 1999: A proposal for a neutral regression. *J. Atmos. Oceanic Technol.,* **16,** 876–883.

Quilfen, Y., B. Chapron, and D. Vandemark, 2001: The ERS scatterometer wind measurement accuracy: Evidence of seasonal and regional biases. *J. Atmos. Oceanic Technol.,* **18,** 1684–1697.

Stoffelen, A., 1998: Towards the true near-surface wind speed: Error modeling and calibration using triple collocation. *J. Geophys. Res.,* **103,** 7755–7766.

——, and D. Anderson, 1997: Scatterometer data interpretation: Estimation and validation of the transfer function CMOD4. *J. Geophys. Res.,* **102,** 5767–5780.

Tokmakian, R., and P. G. Challenor, 1999: On the joint estimation of model and satellite sea surface height anomaly errors. *Ocean Modell.,* **1,** 39–52.

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

Wentz, F. J., and D. K. Smith, 1999: A model function for the ocean-normalized radar cross section at 14 GHz derived from NSCAT observations. *J. Geophys. Res.,* **104,** 11 499–11 514.