Validation of ocean wind and wave data using triple collocation

S. Caires and A. Sterl

Royal Netherlands Meteorological Institute, De Bilt, Netherlands

Received 24 May 2002; revised 13 December 2002; accepted 16 December 2002; published 27 March 2003.

[1] Significant wave height and wind speed fields from ERA-40 are validated against buoy, ERS-1, and Topex altimeter measurements. To do so, we propose and apply a triple collocation statistical model. The model takes into account the random errors in observations and model results and allows the estimation of the variances of the errors. We first examine the case where the random errors of the different systems are independent, but situations where independence is not strictly observed are also considered. The results show that the ERA-40 predictions underestimate high values of significant wave height and, contrary to what would be obtained by less sophisticated statistical methods, wind speed, that the variance of the errors associated with the ERA-40 system is much higher than that of the errors of the measurements, and that the former shows a dependence on the value of the observations not present in the latter. The altimeter measurements of significant wave height are very precise, in contrast to the large uncertainty associated with the altimeter retrieved wind speeds. *INDEX TERMS:* 3299 Mathematical Geophysics: General or miscellaneous; 3339 Meteorology and Atmospheric Dynamics: Ocean/atmosphere interactions (0312, 4504); 4504 Oceanography: Physical: Air/sea interactions (0312); 4560 Oceanography: Physical: Surface waves and tides (1255); 4594 Oceanography: Physical: Instruments and techniques; *KEYWORDS:* significant wave height, wind speed, errors in variables, ERA-40

Citation: Caires, S., and A. Sterl, Validation of ocean wind and wave data using triple collocation, *J. Geophys. Res.*, *108*(C3), 3098, doi:10.1029/2002JC001491, 2003.

1. Introduction

[2] Currently, the European Centre for Medium-Range Weather Forecasts (ECMWF) is conducting ERA-40, a reanalysis of global meteorological wind, temperature and humidity fields, stratospheric ozone, deep water sea states and soil conditions from 1957 to 2002. The reanalysis uses ECMWF's Integrated Forecasting System (IFS), a coupled atmosphere-wave model with variational data assimilation. This is a state-of-the-art model very similar to the one used operationally, although with lower resolution. The aim of this reanalysis is to produce a data set with no inhomogeneities as far as the technique of analysis is concerned, by using the same numerical model throughout. However, since the availability of the observations to be assimilated varies in time and space, a certain amount of inhomogeneity will result. The assimilation of data also reduces the number of independent data sets available to validate the results. On the other hand, these two shortcomings are countervailed by the increase in the quality of the data provided by the assimilation; most of the reliable data sets of observations will be used in the assimilation, leading to the best reanalysis possible.

[3] This is the fourth reanalysis project performed, but the first in which a wave model is coupled to the atmosphere model. For previous reanalysis, wave data had to be generated off-line by forcing a wave model by the reanalyzed winds. An overview of these efforts is given by *Caires et al.* [2002]. In terms of the ocean-wave data, the present reanalysis will be the most complete and consistent reanalysis data set available. It will be the most complete because it will cover more than 40 years of data on a global 1.5° by 1.5° latitude/longitude grid, and the most consistent since the atmosphere model is coupled with the wave model, allowing the atmosphere to react to the waves. If successfully validated, this data set could be used to study the climatology of ocean waves and to compute extreme wave statistics over the whole globe. This would be a valuable outcome of the project, since studies of this kind are usually confined to the Northern Hemisphere [e.g., *WASA Group*, 1998; *Wang and Swail*, 2001].

[4] Validation is possible only if good independent (i.e., not used in the assimilation) wave and wind measurements are available. For a short period of time, from June to December 1993, the period considered here, the ERA-40 significant wave height data can be assessed against independent altimeter data from two distinct satellites, the Topex/Poseidon and the ERS-1, as well as against independent buoy measurements. The wind speed ERA-40 data for this period can also be assessed, but only against altimeter-derived data, because some of the buoy wind speeds are used in the assimilation. After this period, the ERS-1 altimeter wave heights are no longer independent because they are assimilated from January 1994 until May 1996. Topex and ERS-1 measurements are available since the beginning of 1992, but erroneous Fast Delivery Product (FDP) ERS-1 significant wave height measurements were assimilated into ERA-40 from January 1992 until May

0148-0227/03/2002JC001491

1993. This period will be rerun with no data assimilation, but the data was not yet available for this study.

[5] Since all the observing systems (numerical models included) describe the reality not only with some offsets but also random errors, comparison between systems is possible only using errors-in-variables models. In this paper we propose an errors-in-variables model to make triple data set comparisons. The model assumes that each data set corresponds to observations of a linear systematic deviation from the underlying reality plus a random error and allows one to estimate the variances of the errors and the coefficients of the linear functional relationships relating the three data sets. Initially, we assume that the errors of the data sets are independent. But we also consider cases where the errors of two of the systems are dependent, in which case the problem is no longer closed, and the covariance of the two systems' errors needs to be known.

[6] This paper is divided into five sections. Section 2 describes the data sets and their collocation. There are several algorithms available to obtain wind speed measurements from the radar backscatter; three of these will be considered in this paper, and their description is also given in section 2. The errors-in-variables model used in the comparisons is presented in section 3. Section 4 presents the significant wave height and wind speed comparisons. We finish in section 5 with some conclusions.

2. Data Description and Preparation

[7] The buoy, satellite and ERA-40 data represent different time and space scales. The ERA-40 reanalysis data comes on a 1.5° by 1.5° grid at synoptic times. The data is representative of the average condition in the area occupied by a grid box. Buoy measurements are available hourly and come from the processing of single location 20-min records. Altimeter measurements are available every second and at distances of about 7 km apart. In the following we will describe the data sets used in this paper and explain how the data was processed in order to make the time and space scales of the different systems compatible and how they are collocated.

2.1. ERA-40

[8] After the success of ERA-15 [*Gibson et al.*, 1997; *Sterl et al.*, 1998], ECMWF is now performing their second reanalysis, ERA-40. It will span the time from 1957 until now. The computations are divided into three partially overlapping streams (1957–1972, 1972–1987, 1987–now) that are run in parallel. Only for the last stream do enough independent measurements of wave height and surface wind exist to make a thorough validation possible. At the time of preparation of this paper, data from 1987 to 1996 were available.

[9] The ERA-40 output comprises the full directional wave spectrum. However, as the spectrum is rarely measured we will concentrate on significant wave height (H_s) and 10-m wind (U_{10}). (It should be noted that there is more than one 10-m wind speed parameter available from the ERA-40 reanalysis, namely the "10-m atmospheric wind speed" and the "10-m wave model wind speed," the one used in this study. The differences between these two U_{10} products have to do with way the coupling of the wave model with the atmosphere is done and with the 3D-var assimilation scheme used in ERA-40. Roughly speaking, the wave model is forced by hourly winds from the latest 6-hour forecast instead of by the analyzed winds [see *Janssen et al.*, 2002].) Most observations of these quantities are used in the assimilation process of ERA-40 and are therefore not suited for a validation of the results.

[10] In situ wind measurements from ships are used throughout the whole ERA-40 period. In July 1987 the use of SSM/I 1D-Var winds starts. From June 1990 onwards, buoy winds are used, and scatterometer winds over the ocean are used from April 1992 onwards. For wind, the only large data set remaining for the first half of the 1990s are winds derived from altimeter measurements.

[11] Wave height measurements are available from a few buoy locations from 1977 onward, and from the altimeters flown onboard GEOSAT (1985–1989), ERS-1/2 (December 1991 onward), and Topex (October 1992 onward). Buoy and GEOSAT waves are not assimilated in ERA-40 and can be used to validate the results. However, for the GEOSAT period no independent wind measurements exist. Therefore, as we are interested in both winds and waves, the GEOSAT measurements will not be considered here. Wave height measurements from ERS-1 were assimilated in ERA-40 starting in December 1991. However, Fast Delivery Product (FDP) data were used. These data are faulty due to a processing error [*Bauer and Staabs*, 1998]. When this error was recognized, assimilation of the FDP data was halted in May 1993 and the rest of 1993 was run without wave height assimilation. Assimilation was resumed in January 1994 with good data, and the period December 1991 to May 1993 will be rerun later. Thus there was a 7-month period (June 1993 through December 1993) in which no wave height data were assimilated, but altimeter-derived wave heights are available. Note that only the FDP data contain an error, while the original altimeter measurements do not. Correctly reprocessed wave height data for this period are available and are used in this study. Finally, the Topex altimeter measurements are not used in ERA-40 and are also used in this study.

2.2. Buoy

[12] The buoy data to be used in this study come from the NOAA database (National Data Buoy Center, http://seaboard.ndbc.noaa.gov/). This data set was selected because its quality is quite high and it is freely available on the internet. From all the NOAA data buoy locations available during this period, we have selected a total of 17 for these comparisons. The selection of the locations took into account their distance from the coast and the water depth. Only deep water locations can be taken into account since no shallow water effects are accounted for in the wave model, and the buoy should not be too close to the coast in order for the corresponding grid point to be located at sea. The buoy significant wave height measurements are available hourly from 20-min-long records. These measurements have gone through some quality control; we do, however, still process the time series further using a procedure similar to the one used at ECMWF [*Bidlot et al.*, 2002]. All the observations for which $H_s > 25$ m or $H_s < 0.15$ m are discarded. Observations that deviate more than 6 times the standard deviation of the monthly data from its mean, or more than 2 times the standard deviation of the monthly

data from the previous observation, are identified as outliers and removed from the data. This procedure is executed three times. Sometimes buoys report every 2 or 3 hours rather than hourly. When such gaps occur they are filled in by linear interpolation. The hourly time series resulting from the application of the three above procedures are used to produce a new time series at synoptic times by averaging the data over 3 hours around synoptic times. This averaging over time is expected to bring the temporal and spatial scales closer to each other: The model output at a given synoptic time is an estimate of the average condition at a grid cell of 1.5° by 1.5° (the wave model resolution), which long waves take an average 3 hours to cross. The synoptic time series still goes through another quality control: the removal of measurements corresponding to the 24 hours immediately preceding a gap of 18 hours or more. Experience has shown that before these gaps occur there is usually a sudden and unrealistic increase in wave height.

2.3. Satellites

[13] The Topex and ERS-1 along track quality checked deep water altimeter measurements of significant wave height (H_s) and the normalized radar cross section (σ_0) were obtained from Southampton Oceanography Centre (SOC) (GAPS interface, http://www.soc.soton.ac.uk/ALTIMETER/ [*Snaith*, 2000]). Although altimeters do not measure wind speeds directly, the altimeter backscatter depends on and correlates highly with the sea surface wind speed. There are several empirical algorithms available to compute the wind speed from σ_0. Three of them will be considered here and are presented next.

[14] The most widely used algorithm to compute the wind speed from σ_0 is the *Witter and Chelton* [1991] tabular algorithm, which is the operational altimeter wind speed algorithm for the Topex/Poseidon satellite altimeters. This algorithm was devised by modifying the *Chelton and Wentz* [1986] Seasat altimeter wind speed model. *AVISO* [1996] obtained a least squares fit of a degree 4 polynomial to the *Witter and Chelton* [1991] wind speed tabular model, given by

$$U_{10} = c_0 + c_1\sigma_0' + c_2\sigma_0'^2 + c_3\sigma_0'^3 + c_4\sigma_0'^4, \quad (1)$$

where the polynomial coefficients are given in Table 1, and σ_0' is equal to σ_0 for the ERS-1 data and to $\sigma_0 - 0.63$ for the Topex data [*Cotton et al.*, 1997].

[15] *Gourrion et al.* [2002] have devised a new algorithm for the computation of wind speed altimeter measurements. It is the operational algorithm for the recently launched JASON-1 satellite. The method was devised by training an artificial neural network with collocated Topex/Poseidon altimeter observations and wind estimates of NASA scatterometer (NSCAT) data from September 1996 to June 1997. The devised algorithm not only takes into account the altimeter normalized radar cross section measurements, but also the altimeter significant wave height measurements. The reason for accounting also for significant wave height is that swell, which is not closely coupled the local wind field, also influences the backscatter measurements. The *Gourrion et al.* [2002] algorithm is given by

$$U_{10} = \frac{U - 0.1}{0.0284394}, \quad (2)$$

Table 1. Polynomial Coefficients to be Used in Equation (1)

c_0	c_1	c_2	c_3	c_4	σ_0' Limits, dB
51.04531	10.98280	1.89571	−0.17483	0.00544	$\sigma_0' < 10.8$
317.47430	−73.50790	6.41120	−0.24867	0.00361	$10.8 \le \sigma_0' \le 19.6$
0	0	0	0	0	$\sigma_0' > 19.6$

where

$$U = (1 + \exp(-W_2 * X + 2.2838729))^{-1},$$

$$W_2 = [0.541201 \quad 10.4048140],$$

$$X = \begin{bmatrix} (1 + \exp(-W_{11} * P - 18.0637810))^{-1} \\ (1 + \exp(-W_{12} * P + 0.3722814))^{-1} \end{bmatrix},$$

$$W_{11} = [-33.9506170 \quad -11.0339400],$$
$$W_{12} = [-3.9342847 \quad -0.058344],$$

$$P = \begin{bmatrix} 0.0690915 * \sigma_0 - 0.3433598 \\ 0.0637450 * H_s + 0.0872510 \end{bmatrix}.$$

This algorithm was devised using the Topex data, and a correction of +0.63 dB should be applied to the ERS-1 σ_0 data when applying this method. Both methods presented above were devised to estimate wind speeds ranging for 0–20 m/s. For higher wind speeds *Young* [1993] provides the following equation,

$$U_{10} = -6.4\sigma_0 + 72. \quad (3)$$

This equation was derived using Geosat data from passes that transverse the eyes of mature, stable tropical cyclones and a "ground truth" being inferred from model cyclone winds.

[16] Since equation (3) was devised using Geosat data, in principle a −0.63 dB correction should be applied to the Topex σ_0 measurements. However, *Gourrion et al.* [2002] have shown that this equation fits the data quite well for wind speeds above 18 m/s with no σ_0 correction. This equation will therefore be here applied to ERS-1 and Topex data for wind speeds above 18 m/s without any σ_0 correction. Since both *Witter and Chelton* [1991] and *Gourrion et al.* [2002] are used operationally, they will be also used in the comparisons.

[17] Figure 1 shows for 1 month of Topex data the wind speeds obtained by equation (1) (thick black line), the corresponding wind speeds obtained by applying equation (2), where the shading scale corresponds to H_s, and equation (3) (thin black line). The wind speeds obtained by equation (1) are close to the ones obtained using equation (2) for low values of H_s, while for higher H_s values equation (2) provides lower U_{10} estimates. Note that, although this is not clear from the plot, occurrences of high H_S in conjunction with low U_{10} are not very common.

2.4. Data Collocation

[18] The satellite measurements are performed about every second with a spacing of about 5.8 km for Topex and 6.7 km for ERS-1. From these we form satellite "super observations" by grouping together the consecutive observations crossing a 1.5° by 1.5° latitude-longitude region (observations at most 25 s for ERS-1 and 30 s for Topex or $1.5\sqrt{2}°$ apart).The satellite super observation is the mean of

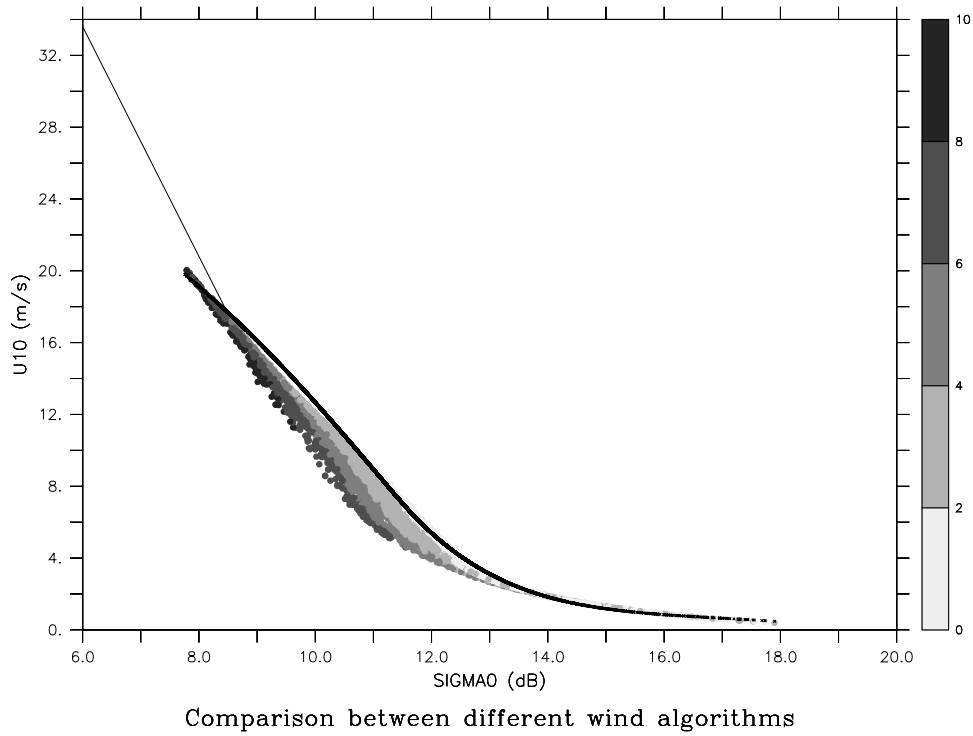

Figure 1. Comparison between different wind speed altimeter algorithms: *Young* [1993] (thin black line), *Witter and Chelton* [1991] (thick black line), and *Gourrion et al.* [2002] (H_s values in m given by shading).

these grouped data points after a quality control similar to the one applied to the buoy data is run through the data.

[19] When collocating buoy, satellite and ERA-40 data, we chose the satellite super observations created from altimeter observations within a 1.5° by 1.5° latitude-longitude region centered at the buoy location. The ERA-40 data at the synoptic times before and after the time of the satellite super observation is interpolated bilinearly to the buoy location and these two data points are then linearly interpolated in time to the time of the super observation. The buoy synoptic data is also linearly interpolated to the time of the super observation. The top panel of Figure 2 shows, for the period from June to December 1993, the locations and the number of observations per location of the triple collocated ERA-40, buoy and Topex data set; the bottom panel gives the same information for the ERA-40, buoy and ERS-1 data set.

[20] For the Topex, ERS-1 and ERA-40 collocation, satellite super observations are collocated if they are less than 0.75° and 1 hour apart. The ERA-40 data is interpolated to the mean space and time location of the two satellite super observations. The locations at which these triple collocations were found, for the period from June to December 1993, is presented in Figure 3. The number of observations per location is five at most, but in most of the locations only one observation is found.

[21] Having explained how the buoy and satellite "super observations" are obtained, in the rest of this article we will refer to them merely as observations.

3. Triple Collocation FR Model

[22] There are several published studies comparing model results with measurements. In such works, the measurements are usually considered to represent "physical truth"; that is, all discrepancies between the measurements and the model data are assumed to be attributable to the model, and standard statistical methods and concepts, such as linear regression analysis (LR), scatter-indices, mean errors, correlation coefficients, etc., are used to study the relationship between the two [e.g., *Cox and Swail*, 2001]. There are also a few papers comparing different measuring instruments, and in these the choice of the "physical truth" depends on which instrument is thought to be more reliable, usually buoys.

[23] The limitations of using standard statistical methods for comparing data with inherent random errors has been acknowledged in several works [e.g., *Bauer and Staabs*, 1998; *Sterl et al.*, 1998] where authors have resorted to variations of the classical linear regression model, estimating "symmetric regression slopes" or principal component analysis (PCA) slopes. Although these methods take into account the fact that the variables have inherent random errors, they do not account for the fact that they may have different error magnitudes.

[24] Analyses of this type are adequate for calibrating one measuring/hindcasting system relative to the other, but inappropriate for comparing them irrespective of which (if any) is the "physical truth" and of their noise levels, which are in principle different. The inadequacy of the above mentioned approaches is due to the fact that both measurements and numerical model predictions are subject to errors, and possibly offsets, and none of them can be considered "truth." More formally, each of the observations/predictions corresponds to the measurement of an unobserved or underlying quantity, the "reality" of the measurements and the "reality" of the model. Equivalently, each of the observations/predictions is subject to an unobserved error.

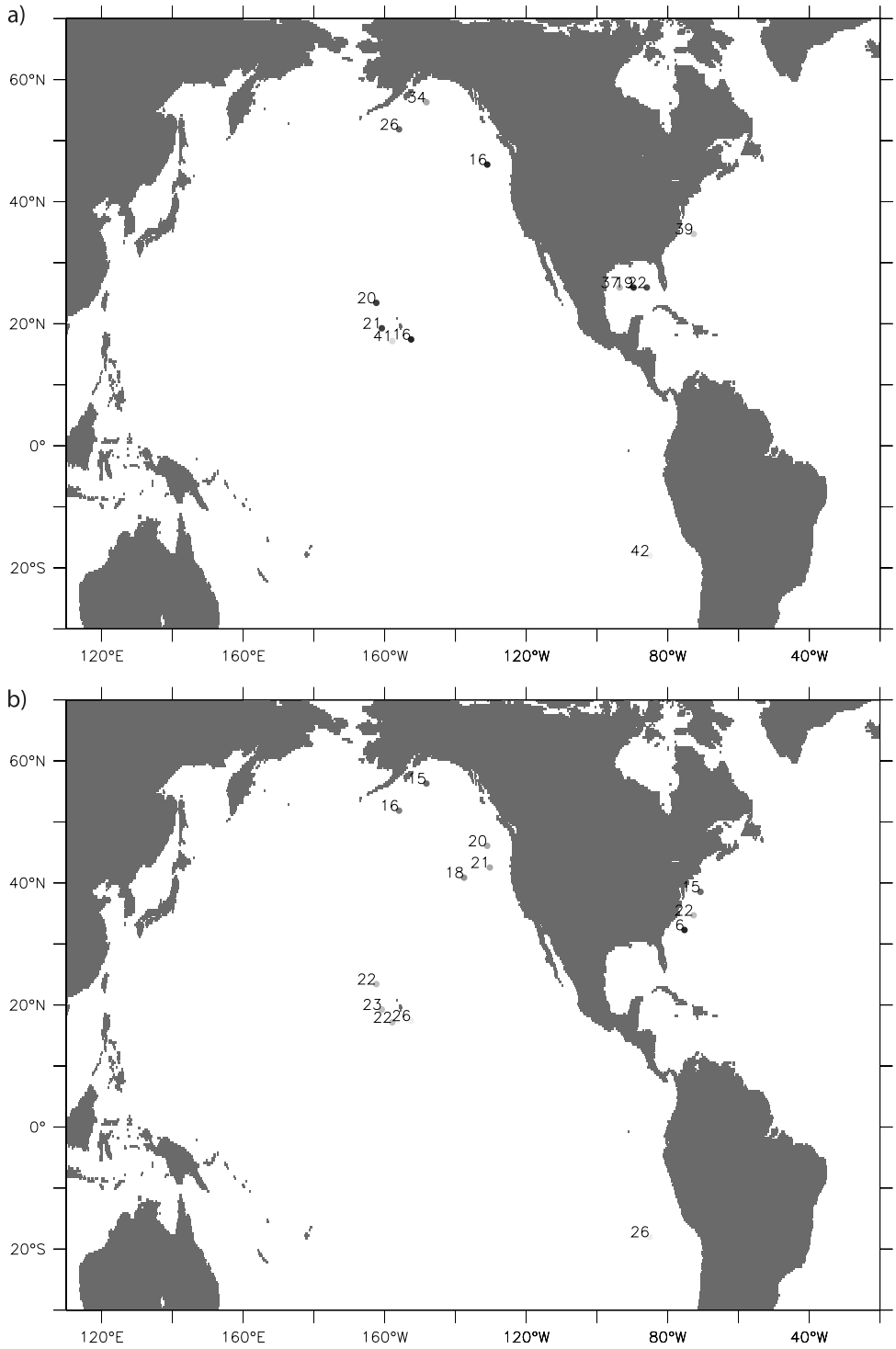

Figure 2. Locations and number of observations per location for triple altimeter, buoy and ERA-40 collocation. (top) ERA-40, buoy and Topex. (bottom) ERA-40, buoy and ERS-1.

In this context, classical linear regression and similar methods are not appropriate, and more sophisticated tools are required. The theory of errors-in-variables models provides an appropriate (and, up to now, possibly the only) method.

[25] The application of errors-in-variables models is, however, difficult, when only two sets of data are available. This is because the functional relationship between the variables and the variances of their associated random errors can only be estimated once the ratio of these variances is known, which is often not the case. In some applications it is assumed that this parameter, which some authors (such as *Tan and Iglewicz* [1999]) refer to as the precision ratio, is 1, i.e., that the errors of the two variables have equal variances, but this is often an unreasonable assumption. In other cases numerical methods are used to estimate its value [e.g., *Caires*, 2000]; however, our experience is that such meth-

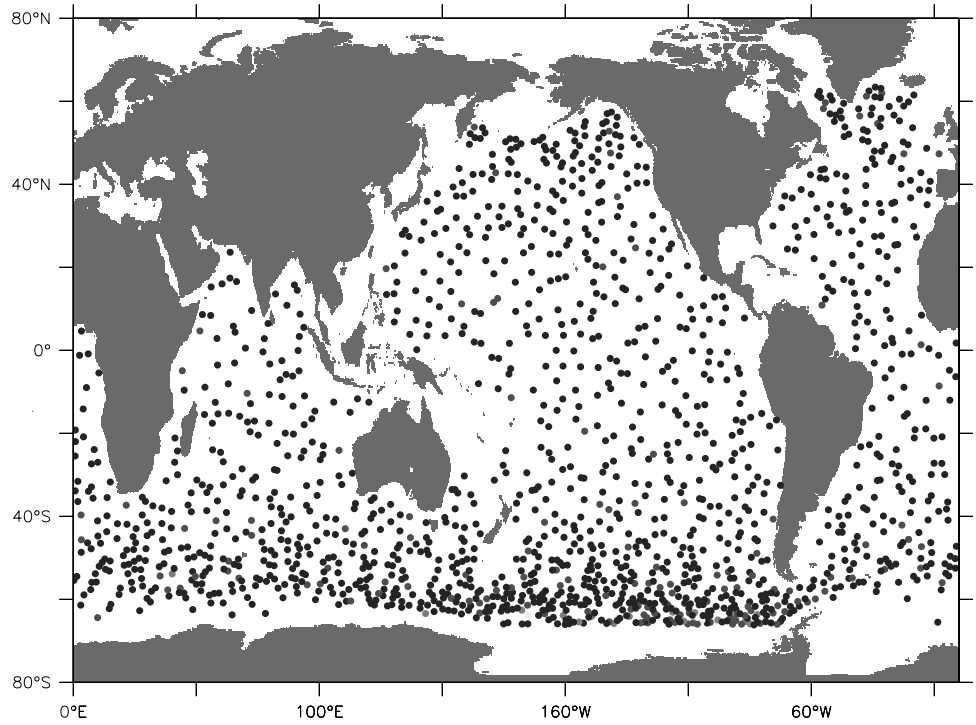

Figure 3. Locations for triple ERA-40, Topex and ERS-1 collocation.

ods are only valid for precision ratios close to one and will do a poor job when the magnitudes of the errors of the two system are very different.

[26] These problems can be solved when triple collocation observations are available, and that is the case we will be considering here.

[27] Assume that we are given three sets of n observations (x_i, y_i, z_i), $i = 1, \ldots, n$, and that these observations correspond to measurements of certain deterministic underlying variables T_i, $i = 1, \ldots, n$, made with certain systematic deviations and subject to zero mean random errors (e_{xi}, e_{yi}, e_{zi}), $i = 1, \ldots, n$. More precisely, we assume that (in what follows we drop the subscripts of the variables, the averages of which are denoted by $\langle x \rangle$, $\langle xy \rangle$, etc.)

$$\begin{aligned} x &= X + e_x \equiv T + e_x \\ y &= Y + e_y \equiv \alpha_1 + \beta_1 T + e_y \\ z &= Z + e_z \equiv \alpha_2 + \beta_2 T + e_z, \end{aligned} \tag{4}$$

and we want to estimate the unknown parameters α_1, α_2, β_1 and β_2, and the variances of the errors. Having computed β_1 and β_2 (see below), we can go back to equation (4) and compute the constant coefficients α_1 and α_2 by taking averages,

$$\alpha_1 = \langle y \rangle - \beta_1 \langle x \rangle \tag{5}$$

$$\alpha_2 = \langle z \rangle - \beta_2 \langle x \rangle. \tag{6}$$

Removing the mean from each of the variables and denoting the result by x^*, y^*, z^*, and T^*, we can simplify our model to

$$\begin{aligned} x^* &= T^* + e_x \\ y^* &= \beta_1 T^* + e_y \\ z^* &= \beta_2 T^* + e_z. \end{aligned}$$

Computing the cross correlations, we have, on noting that $\langle Te_x \rangle = \langle Te_y \rangle = \langle Te_z \rangle = 0$ (the true observation is deterministic), and assuming that the errors are independent, so that $\langle e_x e_y \rangle = \langle e_x e_z \rangle = \langle e_y e_z \rangle = 0$,

$$\beta_1 = \langle y^* z^* \rangle / \langle x^* z^* \rangle, \tag{7}$$

$$\beta_2 = \langle y^* z^* \rangle / \langle x^* y^* \rangle. \tag{8}$$

Using equations (7) and (8), we obtain the variances of the errors,

$$\begin{aligned} \langle e_x^2 \rangle &= \langle x^* 2 \rangle - \langle x^* y^* \rangle \langle x^* z^* \rangle / \langle y^* z^* \rangle \\ \langle e_y^2 \rangle &= \langle y^* 2 \rangle - \langle x^* y^* \rangle \langle y^* z^* \rangle / \langle x^* z^* \rangle \\ \langle e_z^2 \rangle &= \langle z^* 2 \rangle - \langle x^* z^* \rangle \langle y^* z^* \rangle / \langle x^* y^* \rangle. \end{aligned} \tag{9}$$

[28] One feature of this type of errors-in-variables models is their symmetric nature: The result of applying the model to data is independent of which variables are chosen to be x, y or z. This implies, in particular, that one can compute from the above expressions the coefficients in the relationship between Y and Z, $Y = \alpha_3 + \beta_3 Z$, as $\alpha_3 = \alpha_1 - \alpha_2 \beta_1 / \beta_2$ and $\beta_3 = \beta_1 / \beta_2$.

[29] An errors-in-variables model in which the underlying variables T are deterministic (or fixed) is called a functional relationship (FR) model; otherwise the model is referred to as a structural relationship model.

[30] In the above calculations we consider the underlying physical parameters T as deterministic variables, and hence use a FR model. This option demands some justification. Although the underlying physical parameters are random variables, in the sense that they correspond to observations from a hypothetical population of physical scenarios (implying a structural relationship model), we are not interested in

their statistical behavior. Instead, we are interested in studying the relationship between the outcomes, that is, between the x, y and z that happened to occur. In statistical language, we are interested in studying the relationship between what the instruments and the numerical model observe (each with its inherent errors) conditionally on a specific "physical scenario" (the particular occurrence of environmental processes in which measurements took place). A more formal justification can be given on the basis of the conditionality principle [*Cox and Hinkley*, 1974, p. 38], according to which we should condition on the actual observations ("physical scenario"), and thus regard them as fixed (though unknown), and hence use a functional model.

[31] If we had taken T as a random variable, we would have to impose extra conditions, namely that the correlations between T and the errors e_x, e_y and e_z are zero, in order to obtain equations (7) and (8). For the FR model, the disappearance of the correlations is a natural property.

[32] When fitting our model to the collocated data, sets were formed by pooling space and time data together. This is a way of reducing a problem that is originally multivariate (multiple time-space dimensions) to a bivariate problem. An objection that may be raised to the pooling of data from all space-time locations is the possible dependence between measurements at neighboring points and the lack of homogeneity in measurements that are too distant in time or space. However, what our functional relationship model requires is that the errors be independent and homogeneous, not the measurements themselves.

[33] The assumption of independent and homogeneous errors is still strong and will not always be tenable. The violation of the homogeneity assumption (i.e., that the errors e_{xi} have all the same variance, and similarly for e_{yi} and e_{zi}) can be detected by dividing the original data sets into subsets and checking whether the variance estimates obtained with the subsets are compatible with those obtained with the whole data set. We will return to this point when presenting the results.

[34] Another of the assumptions used above is that the errors associated with the different systems are uncorrelated. This is, however, not always the case. The model can be reformulated to account for dependence between the errors of different data sets, but this requires the covariance between the errors of the different data systems to be known. The system is no longer "closed."

[35] If the errors between two of the data sets are correlated, say $\langle e_y e_z \rangle \neq 0$, the value of which is known, we obtain

$$\beta_1 = \big(\langle y^* z^* \rangle - \langle e_y e_z \rangle\big) / \langle x^* z^* \rangle, \tag{10}$$

$$\beta_2 = \big(\langle y^* z^* \rangle - \langle e_y e_z \rangle\big) / \langle x^* y^* \rangle, \tag{11}$$

and then, using equations (10) and (11),

$$\begin{aligned} \langle e_x^2 \rangle &= \langle x^* 2 \rangle - \langle x^* y^* \rangle \langle x^* z^* \rangle / \big(\langle y^* z^* \rangle \langle e_y e_z \rangle\big) \\ \langle e_y^2 \rangle &= \langle y^* 2 \rangle - \big(\langle x^* y^* \rangle \langle y^* z^* \rangle - \langle e_y e_z \rangle \langle x^* y^* \rangle\big) / \langle x^* y^* \rangle \\ \langle e_z^2 \rangle &= \langle z^{*2} \rangle - \big(\langle x^* z^* \rangle \langle y^* z^* \rangle - \langle e_y e_z \rangle \langle x^* z^* \rangle\big) / \langle x^* y^* \rangle. \end{aligned} \tag{12}$$

[36] Besides obtaining point estimates of the parameters, we are interested in determining (estimates of) their standard errors, as these reflect their range of variability. Thus, we require estimates of the variances of our estimators.

[37] Although in some cases the delta method (which is based on Cramér's theorem [see *Ferguson*, 1996, p. 45]) can be used to find explicit expressions for the variances of the estimators, it often appears that such a task is extremely complicated or even impossible to carry out. In situations like this, resampling methods like the bootstrap offer a simple and reliable alternative for estimating standard errors of estimators.

[38] We will now briefly explain how the bootstrap method can be used to estimate the variance of the above estimates. The reader is referred to *Efron and Tibshirani* [1993] for an explanation of how and why the method works.

[39] In many situations, we have a random sample $\mathbf{x} = \{x_i,\ i = 1, \ldots, n\}$ of observations of some random variable or population X, and we wish to estimate a population parameter θ by an estimator $\hat{\theta} \equiv \hat{\theta}(x_1, \ldots, x_n) \equiv \hat{\theta}(\mathbf{x})$ based on $\mathbf{x}$. For instance, θ might be the median of the population (the quantile of probability 0.5), the population mean (the expectation of X), or a functional relationship estimator. If $\hat{\theta}$ has a simple expression and the distribution of X has simple mathematical properties, one can determine the variance of $\hat{\theta}$, or at least an approximation to it. In those cases where this is not possible, one can estimate (rather than determine exactly) Var $\{\hat{\theta}\}$ (the variance of $\hat{\theta}$) by the bootstrap method.

[40] The bootstrap method consists of creating bootstrap samples $\mathbf{x}^*$, each obtained by randomly sampling n times, with replacement, from the original sample $\mathbf{x}$. Given B bootstrap samples, which we denote by $\mathbf{x}_b^*$, $b = 1, \ldots, B$, we can calculate a set of estimates $\hat{\theta}_b^* = \hat{\theta}\big(\mathbf{x}_b^*\big)$, each obtained in the same way $\hat{\theta}$ was obtained, but based on $\mathbf{x}^*$ in place of $\mathbf{x}$. Then the bootstrap estimate of the standard error of the estimate $\hat{\theta}$ is given by

$$\hat{s}_B\big(\hat{\theta}\big) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\big(\hat{\theta}_b^* - \hat{\theta}^*\big)^2}, \quad \text{with} \quad \hat{\theta}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^*.$$

The ideal bootstrap estimate would be $\hat{s}_\infty(\hat{\theta})$, but of course this is not possible to achieve; a limit on B must be stipulated. According to *Efron and Tibshirani* [1993, pp. 50–53], very seldom more than $B = 200$ bootstrap replications are needed for estimating the standard error. This is the value that we have used here.

[41] Estimates of the standard error of the estimate are useful for establishing confidence intervals or regions for the unknown parameters. We will obtain 95% confidence intervals for a parameter θ, by calculating upper and lower limits of the form $\hat{\theta} \pm 1.96 \times \hat{s}_B\big(\hat{\theta}\big)$.

4. Validation

4.1. Wave Data

4.1.1. ERA-40, Buoy and Topex

[42] We start by looking at the comparisons between ERA-40, buoy and Topex significant wave height observations. The scatter diagrams of the comparisons between the

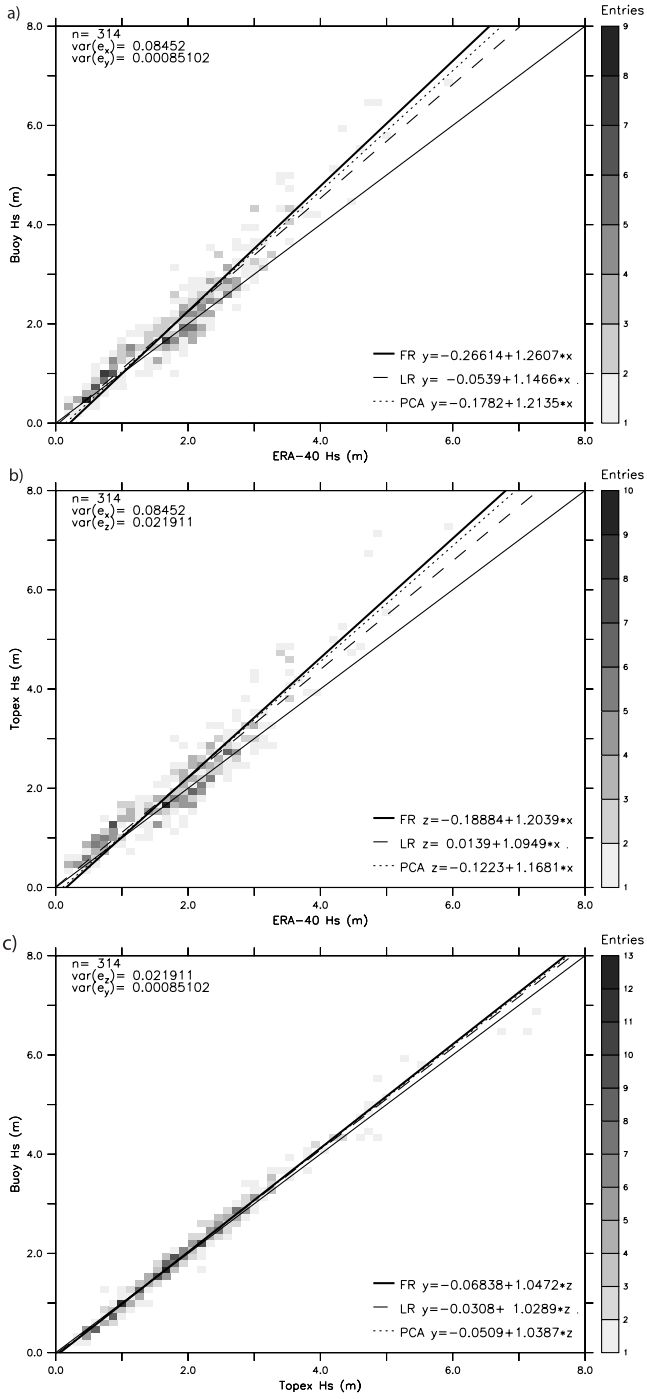

Figure 4. Scatter diagrams with estimated FR, LR and PCA lines for significant wave height triple ERA-40, buoy and Topex collocated data from June to December 1993. (a) ERA-40 versus buoy. (b) ERA-40 versus Topex. (c) Topex versus buoy.

three data sets from June to December 1993 are presented in Figure 4. Superimposed are the functional relationship lines resulting from the application of the method proposed above and assuming no correlations between the errors. The parameters estimated along with their confidence intervals (estimated using the bootstrap method) are given in the first row of Table 2. The results give an underestimation of significant wave height by the wave model when compared with the measurements for high values of H_s and a small overestimation of wave heights below 1 m. The buoy and Topex measurements compare much better ($Y = -0.11 + 1.05 * Z$). Another thing to notice is that the estimate of the variances of the random errors in the model is much higher than those for the buoy and satellite, which can be confirmed by the scatter of the plots. The scatter of the buoy-Topex comparison (Figure 4c) is much smaller than those in the comparisons with the model (Figures 4a and 4b).

[43] To get an idea on how different statistical methods (assumptions) influence the results, we also have performed linear regression and principal component analysis. The results are given in Table 3 and the corresponding lines are drawn in Figure 4. As expected, differences are larger between the triple method and the linear regression and when the random errors are high (comparisons involving the ERA-40 data). The differences between the PCA estimate and the triple method are higher for higher ratios of variances of the errors. This is normal since in the PCA it is assumed that the ratio between the variances of the errors is 1.

[44] It is also interesting to know how dependent these results are on the particular sample we are studying. For this reason we repeated the calculations for 1994 and 1995, during which ERS-1 wave heights were assimilated into ERA-40. The results are also presented in Table 2 along with the results for 1993. The data for 1993 are only from June to December and those for 1994 and 1995 cover the whole year. The variance estimates are mainly the same (comprising confidence intervals), for buoy and Topex, but for the wave model they change from 0.08 to 0.06. This is a consequence of the assimilation of the ERS-1 data in ERA-40, which started in January 1994. Another thing to notice is that the underestimation of high wave heights by the model did not improve, which may be an indication that the data being assimilated are too low.

[45] In order to have a larger sample to obtain more precise estimates, we have pooled the data from June 1993 to December 1995. The results, which are not shown here, are similar to those presented in Table 2, the confidence intervals naturally being narrower; in the case of the altimeter the estimate of the variance of the errors is of 0.01 with confidence interval limits 0.00, 0.01, and in the case of the buoy the estimate of the variance of the errors is of 0.02 with confidence interval limits 0.02, 0.03.

4.1.2. ERA-40, Buoy and ERS-1

[46] We now look at the comparisons between buoy, ERA-40 and ERS-1 data from June to December 1993. The triple collocation estimates are presented in Table 4, and the scatter diagrams are in Figure 5. The results reveal, more pronouncedly than in the previous comparisons, the underestimation of high wave heights by the model when compared with the buoy measurements. Note, however, that the significant wave height average is also higher. The model predictions also underestimate the ERS-1 measurements of high H_s, while the ERS-1 measurements show a negative bias for all values of H_s when compared with the buoy measurements. Since such negative offset is not observed between the buoy and Topex data, and since the offsets we obtained agree with those obtained by *Young* [1999b], we must conclude that the bias is in the ERS-1 data

Table 2. Estimates, for Different Years, of the Functional Relationship Coefficients Between ERA-40(x), Buoy(y), and Topex(z) Significant Wave Height Observations, and of the Variances of the Errors[a]

Year	n	$\langle x \rangle$	α_1	β_1	α_2	β_2	β_3	$\langle e_x^2 \rangle$	$\langle e_y^2 \rangle$	$\langle e_z^2 \rangle$
1993	314	1.83	−0.27	1.26	−0.19	1.20	1.05	0.08	0.00	0.02
			(−0.37, −0.16)	(1.21, 1.31)	(−0.30, −0.07)	(1.14, 1.27)	(1.01, 1.08)	(0.07, 0.10)	(0.00, 0.01)	(0.01, 0.03)
1994	497	1.95	−0.22	1.27	−0.10	1.20	1.06	0.06	0.01	0.02
			(−0.31, −0.13)	(1.22, 1.33)	(−0.20, −0.01)	(1.14, 1.26)	(1.04, 1.09)	(0.04, 0.07)	(0.00, 0.02)	(0.01, 0.03)
1995	442	1.82	−0.11	1.24	0.01	1.16	1.07	0.06	0.01	0.03
			(−0.19, −0.04)	(1.19, 1.28)	(−0.08, 0.09)	(1.10, 1.21)	(1.04, 1.10)	(0.05, 0.07)	(0.00, 0.01)	(0.02, 0.0)

[a]The limits of the confidence intervals are given in parentheses below the estimates. The sample size, n, and the average of the ERA-40 data, $\langle x \rangle$, are also included. Here, 1993 only comprises data from June to December.

and not in the buoy data. This underestimation of the wave height by ERS-1 is the reason why the assimilation of ERS-1 in ERA-40 did not improve the ERA-40 underestimation (see Table 2).

[47] The estimate of the variances of the errors in the buoy measurements remains the same as in the comparisons with Topex. The variance of ERS-1 is found to be around 0.01, slightly lower than the values obtained for Topex. The estimate of the variance of the errors in the ERA-40 data is here higher than the one obtained in the Topex comparisons. This may be an indication that the variance of the errors of the wave model is wave height or location dependent. For this data set the average H_s of the model hindcasts is 2.16 m while in the Topex comparisons it was 1.83 m, and there are more northern buoy locations in these comparisons than there were in the Topex comparisons (see Figure 2). However, the results are not conclusive since the confidence intervals of the estimates obtained here comprise the previous value.

4.1.3. ERA-40, ERS-1 and Topex

[48] The possible sea state or location dependence of the data will be analyzed in more detail in the ERS-1, Topex and ERA-40 global comparisons. Figure 6 shows histograms of the ERA-40, ERS-1 and Topex data from June to December 1993. The ERS-1 and Topex histograms seem to have the same shape, but with the former being shifted to the left relative to the latter. The ERA-40 histogram indicates that the ERA-40 predictions compare well with the Topex measurements for values up to 1.5 m. For higher H_s ERA-40 predictions are mainly concentrated around 2.2 m with less high wave height values than in the Topex observations.

[49] The scatter diagrams of the triple Topex, ERS-1 and ERA-40 comparisons are presented in Figure 7. The triple collocation estimates for the data plotted in the figure are presented in the first line of Table 5.

[50] Qualitatively, the results obtained here are consistent with those obtained at the buoy locations. The model results underestimate the Topex measurements of high significant wave height. This underestimation is not present in the comparisons with ERS-1 data. However, the comparisons between Topex and ERS-1 measurements show negatively biased measurements by ERS-1. Again, the variance of the random errors associated with the ERA-40 significant wave height predictions is much higher that the equivalent satellite variances (about 16 times higher), and this is again clearly visible in the scatter diagrams: the scatter of the ERS-1 versus Topex comparisons is much lower than the scatter of the ERA-40 versus any of the altimeters. However, the estimated variance of the errors of the model prediction (0.16) is higher than the value obtained in the previous comparisons (0.08 and 0.11). This may again be an indication of variance being dependent on the value of H_s. Here the average is 3 m, while in the buoy comparisons it was around 2 m. Indications for an H_s dependence of the random errors in ERA-40 are also present in the scatter diagrams of Figure 7. The scatter diagrams comparing ERA-40 with the altimeter data have a "megaphone shape" with higher scatter for higher values of H_s, while Figure 7c shows no such behavior for the random errors associated with the altimeters measurements.

[51] In order to look further into this dependence, we have divided the original global data set of triple collocated ERA-40, ERS-1 and Topex data into subsets according to latitude strips and obtained the functional relationship estimates for each data set. These results are also presented in Table 5. They are indicative of a dependence of the error variance on the average values of H_s, and possibly also on location, as the results obtained for the region between 20°S–20°N seem to have different characteristics than those of the others.

[52] The estimates of the variances of the random errors associated with the altimeters in Table 5 tend to be lower than those obtained from the comparisons at the buoy locations (Tables 2 and 4). This may be a consequence of some covariance between the random errors of the two altimeters that is not being taken into account. The random errors associated with altimeter measurements of sea surface height are assumed to be dependent due to common corrections applied *Tokmakian and Challenor*, 1999]. Since the H_s altimeter values are obtained from slopes of the received signal [see *Young*, 1999a, pp. 247–252], there is no direct dependence on ionospheric or other such correction, that could be common to both altimeters, and thus we think that in the case of significant wave height measurements this is not a source of error dependence. A source of dependence could, however, be an altimeter error dependence on the type of sea being observed, such as pure wind sea, sea or mixed wind sea and swell. So, let us assume that there is some dependence, although we cannot quantify it.

Table 3. Estimates of the Relation Between ERA-40(x), Buoy(y), and Topex(z) Significant Wave Height Data Obtained by the FR, LR and PCA Models[a]

	$y = \alpha + \beta * x$			$z = \alpha + \beta * x$			$y = \alpha + \beta * z$		
	FR	LR	PCA	FR	LR	PCA	FR	LR	PCA
α	−0.27	−0.05	−0.18	−0.19	0.01	−0.12	−0.11	−0.03	−0.05
β	1.26	1.15	1.21	1.20	1.09	1.17	1.05	1.03	1.04

[a]Data from June to December 1993.

Table 4. Estimates of the Functional Relationship Coefficients Between ERA-40(x), Buoy(y), and ERS-1(z) Significant Wave Height Observations, and of the Variances of the Errors[a]

	n	$\langle x \rangle$	α_1	β_1	α_2	β_2	β_3	$\langle e_x^2 \rangle$	$\langle e_y^2 \rangle$	$\langle e_z^2 \rangle$
Estimate	237	2.16	−0.67	1.39	−0.96	1.27	1.09,	0.11	0.01	0.01
Limits			(−0.87, −0.48)	(1.30, 1.49)	(−1.16, −0.77)	(1.18, 1.37)	(1.06, 1.13)	(0.07, 0.15)	(0.00, 0.02)	(0.00, 0.02)

[a]Data from June to December 1993. The limits of the confidence intervals are given in parentheses below the estimates. The sample size, n, and the average of the ERA-40 data, $\langle x \rangle$, are also included.

From the Cauchy-Schwartz inequality ($|\langle xy \rangle|^2 \leq \langle x^2 \rangle \langle y^2 \rangle$) the covariance between the two variables can at most be equal to the square root of the product of the two variances. From the comparisons with buoy data we have that the variances of the errors of Topex and ERS-1 are about 0.02, which means that its covariance can at most be 0.02. Taking into account a covariance of 0.02 between the ERS-1 and the Topex H_s errors ($\langle e_y e_z \rangle = 0.02$) we have obtained new estimates of the FR parameters using equations (5), (6), and (10)–(12). They are given in Table 6.

[53] Assuming that $\langle e_y e_z \rangle = 0.02$ has the effect of increasing the variance of the errors associated with the altimeter measurements, decreasing the variance of the errors associated with ERA-40, although not significantly, and has barely any effect in the estimates of the functional relationship slopes and constants. The increase in the variances of the errors associated with the altimeter measurements is, however, too high. The values now are about 0.03, while 0.02 was the value obtained in the buoy comparisons and 0.01 assuming that there was no correlation between the errors. Thus, while there are indications of covariance between the random errors of the two altimeters, its value is certainly lower than 0.02.

4.2. Wind Data

[54] Having analyzed the ERA-40 significant wave height predictions we will now look at the wind speed values. Unfortunately, the ERA-40 predictions can only be assessed against the altimeter measurements, since the buoy measurements are used in the assimilation. We do not exclude the hypothesis that there is some covariance between the errors associated with the different altimeters. However, we cannot take it into account as we have no idea of how large that covariance can be, and no independent estimates of the variance of the errors associated with each altimeter are available. The reader should be aware that a non-zero covariance between the two would affect results in the same way as it did in the H_s comparisons, i.e., increase of the variance of the altimeter errors and decrease in the variance of the ERA-40 system.

[55] Figure 8 shows the histograms comparing the wind speed of ERA-40 and of the altimeter data using the *Gourrion et al.* [2002] and the *Witter and Chelton* [1991] algorithms as presented in section 2.3. While these histograms are not as smooth as those for H_s, they seem to indicate a better correspondence between the three data sets, especially between the ERA-40 and the Topex data. The ERS-1 histograms have a more pronounced peak than those of ERA-40 and Topex and seem to have a couple of other smaller modes. The wind speeds produced by the *Gourrion et al.* [2002] algorithm are generally lower than those produced by the *Witter and Chelton* [1991] algorithm, their histograms possessing broader peaks. Comparison of the

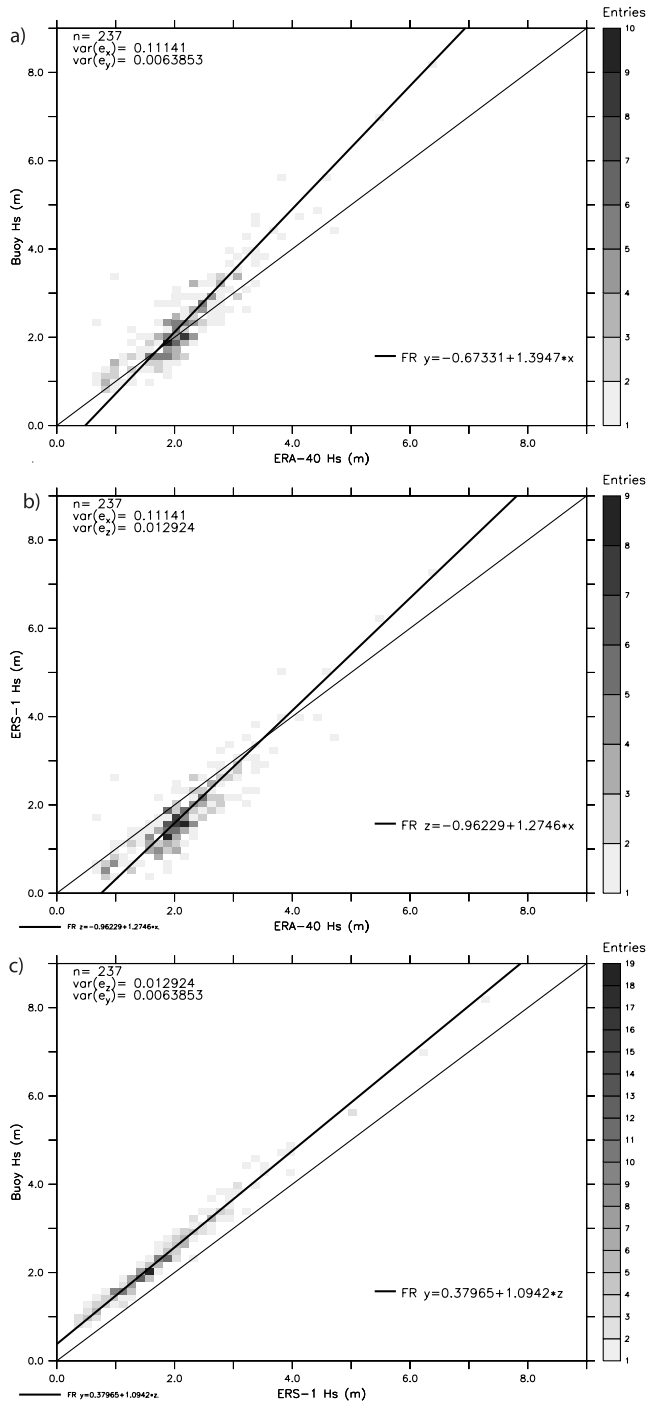

Figure 5. Scatter diagrams with estimated functional relationships for significant wave height triple ERA-40, buoy and ERS-1 collocated data from June to December 1993. (a) ERA-40 versus buoy. (b) ERA-40 versus ERS-1. (c) ERS-1 versus buoy.

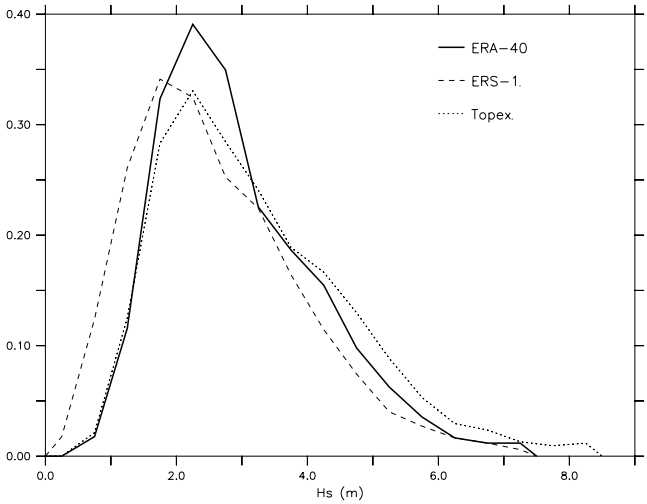

Figure 6. Histograms of ERA-40 (solid line), ERS-1 (dashed line) and Topex (dotted line) significant wave height observations from June to December 1993.

histograms does not indicate a better performance of one relative to the other.

[56] The triple collocation functional relationship comparisons are also not conclusive. Figure 9 shows the comparative scatter diagrams of the ERA-40 wind speeds and Topex and ERS-1 winds speeds obtained using the *Gourrion et al.* [2002] algorithm. Figure 10 shows the corresponding diagrams when the altimeter wind speeds are computed using the *Witter and Chelton* [1991] algorithm. Table 7 shows the corresponding triple collocation functional relationship estimates. The top half of the table gives estimates for data obtained using the *Gourrion et al.* [2002] algorithm, and the bottom half using the *Witter and Chelton* [1991] algorithm. As we did for the H_s comparisons, the original data set was divided into subsets according to latitude strips.

[57] The ERS-1 wind speed obtained with the *Witter and Chelton* [1991] algorithm seems to compare worse with the ERA-40 predictions than those obtained with the *Gourrion et al.* [2002] algorithm (β_1 values in Table 7 closer to 1). However, the ERS-1 and Topex wind speeds obtained with the *Witter and Chelton* [1991] algorithm compare better with each other (less overestimation of Topex) than do those from the *Gourrion et al.* [2002] algorithm. This is because the *Gourrion et al.* [2002] algorithm depends on the altimeter significant wave height, which, as we saw in the comparisons of the H_s ERS-1 measurements, is biased relative to those of Topex.

[58] Independently of the wind speed algorithm used, the FR estimates indicate an underestimation of the high wind speeds by the ERA-40 system. This underestimation would partly explain the underestimation produced by the ERA-40 wave model.

[59] The point estimates of the variances of the ERA-40 wind speed random errors are not exactly the same when comparing them with the altimeter wind speed from the different algorithms, but the differences are not statistically significant. These values testify to the robustness of the model and give an idea of how sensitive the results are to the choice of the comparing sets. The estimates of the variances of the ERA-40 wind speed random errors are different for the different latitude regions considered. They suggest a wind speed dependence as observed also for the wave height data. This is expected since the noise inherent to the wave model predictions can be regarded as a result of the integration of the input geophysical quantities (in this

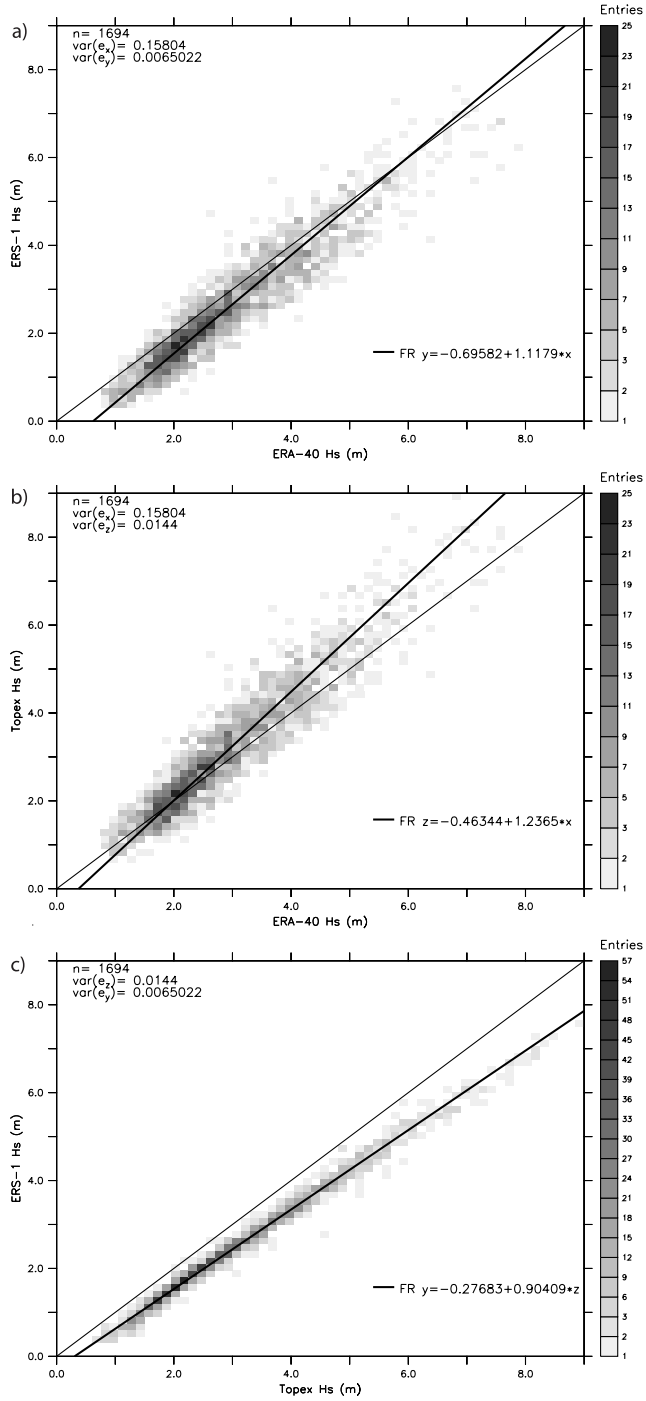

Figure 7. Scatter diagrams with estimated functional relationships for significant wave height triple ERA-40, ERS-1 and Topex collocated data from June to December 1993. (a) ERA-40 versus ERS-1. (b) ERA-40 versus Topex. (c) ERS-1 versus Topex.

Table 5. Estimates, for Different Latitude Strips, of the Functional Relationship Coefficients Between ERA-40(x), ERS-1(y), and Topex(z) Significant Wave Height Observations, and of the Variances of the Errors[a]

	n	$\langle x \rangle$	α_1	β_1	α_2	β_2	β_3	$\langle e_x^2 \rangle$	$\langle e_y^2 \rangle$	$\langle e_z^2 \rangle$
80°S–80°N	1694	2.95	−0.70	1.12	−0.46	1.24	0.90	0.16	0.01	0.01
			(−0.76, −0.63)	(1.09, 1.14)	(−0.54, −0.39)	(1.21, 1.26)	(0.90, 0.91)	(0.14, 0.17)	(0.00, 0.01)	(0.01, 0.02)
80°S–40°S	893	3.60	−0.90	1.15	−0.81	1.30	0.88	0.19	0.01	0.01
			(−1.01, −0.78)	(1.11, 1.18)	(−0.94, −0.68)	(1.25, 1.34)	(0.87, 0.90)	(0.17, 0.22)	(0.00, 0.02)	(0.00, 0.02)
40°S–20°S	226	2.60	−0.82	1.24	−0.65	1.39	0.89	0.09	0.01	0.01
			(−1.00, −0.65)	(1.17, 1.32)	(−0.84, −0.47)	(1.32, 1.47)	(0.87, 0.91)	(0.07, 0.12)	(0.00, 0.02)	(0.00, 0.02)
20°S–20°N	279	2.01	−1.10	1.36	−0.76	1.43	0.95	0.06	0.00	0.01
			(−1.26, −0.95)	(1.28, 1.44)	(−0.93, −0.58)	(1.34, 1.52)	(0.92, 0.98)	(0.04, 0.07)	(0.00, 0.01)	(0.00, 0.02)
20°N–40°N	138	1.91	−0.91	1.27	−0.48	1.31	0.97	0.10	0.00	0.01
			(−1.31, −0.51)	(1.06, 1.49)	(−0.88, −0.08)	(1.10, 1.53)	(0.93, 1.01)	(0.07, 0.13)	(0.00, 0.01)	(0.00, 0.03)
40°N–80°N	158	2.41	−0.84	1.21	−0.61	1.35	0.90	0.13	0.01	0.01
			(−1.05, −0.63)	(1.10, 1.32)	(−0.84, −0.38)	(1.23, 1.47)	(0.87, 0.93)	(0.10, 0.17)	(0.00, 0.02)	(0.00, 0.03)

[a]Data from June to December 1993. The limits of the confidence intervals are given in parentheses below the estimates. The sample size, n, and the average of the ERA-40 data, $\langle x \rangle$, are also included.

case the wind speed) and of the choice of certain empirical coefficients. Thus, if there is such an error structure in the wind speed, it is to be expected that it will also be present in the wave model results.

[60] The variance estimates of wind speed altimeter measurements for the different wind speed algorithms are statistically the same, indicating that each algorithm has the same associated random uncertainty. The error variance estimates, however, do not exclude a dependence of the random error on the geographical location. For the ERS-1 measurements the values obtained in the 40°S–20°S region are statistically different from the estimates in other regions and the same is true for the Topex measurements in the 20°S–20°N region. There is, however, no evidence of a dependence on the values of U_{10}.

[61] It is interesting to see what would result from these wind speed comparisons if the variances of the random errors were not taken into account. In Table 8 we compare the slopes obtained in the comparisons between the ERA-40 wind speeds and the altimeter wind speeds resulting from the application of the *Gourrion et al.* [2002] algorithm using the FR, LR and PCA models. In the scatter diagrams comparing the ERA-40 predictions with the altimeters measurements (Figure 9), we have superimposed the estimated FR, LR and PCA lines. As expected, the larger discrepancies are between the estimates of the linear relationship model and those of the functional relationship model. Both the LR and PCA estimates, contrary to the FR estimates, indicate an overestimation of the high wind speeds by the ERA-40 system when compared with ERS-1. For the ERA-40/Topex comparisons the FR estimate gives an underestimation of about 17% by ERA-40, while according to the PCA estimate it is only 8%, and according to the LR estimate ERA-40 overestimates the measured wind speed by 4%. No differences between the statistical properties occur in the altimeter comparisons. This is due to the fact that the variances of the errors of the altimeter data sets are small and the ratio between the variances is close to 1.

5. Conclusions

[62] We have proposed a functional relationship model for comparing three data sets which allows the estimation of the parameters of the functional relationship and of the variances of the random errors of the three systems. Two situations were considered: a situation where the random errors of the three systems can all be assumed independent, and a situation where random errors of two of the three systems are dependent with known covariance.

[63] The model was used to compare H_s data from ERA-40, buoy, ERS-1 and Topex, covering the period from June to December 1993. For the comparisons of triple collocation ERA-40, buoy and Topex data additional results for 1994 and 1995 were also presented. Comparisons were also made between U_{10} data from ERA-40, ERS-1 and Topex of the 7-month period in 1993.

[64] There are several algorithms available to compute wind speeds from altimeter σ_0 measurements for low to moderate wind speeds. Two of those used operationally, of *Witter and Chelton* [1991] and *Gourrion et al.* [2002], have

Table 6. As in Table 5, but Assuming a Covariance Between the Errors of ERS-1 and Topex of $\langle e_y e_z \rangle = 0.02$

	n	$\langle x \rangle$	α_1	β_1	α_2	β_2	β_3	$\langle e_x^2 \rangle$	$\langle e_y^2 \rangle$	$\langle e_z^2 \rangle$
80°S–80°N	1694	2.95	−0.66	1.11	−0.42	1.22	0.90	0.14	0.02	0.04
			(−0.72, −0.59)	(1.08, 1.13)	(−0.50, −0.35)	(1.20, 1.25)	(0.90, 0.91)	(0.13, 0.16)	(0.02, 0.03)	(0.03, 0.04)
80°S–40°S	893	3.60	−0.85	1.13	−0.75	1.28	0.88	0.18	0.03	0.04
			(−0.96, −0.74)	(1.10, 1.17)	(−0.88, −0.62)	(1.24, 1.32)	(0.87, 0.90)	(0.15, 0.20)	(0.02, 0.04)	(0.03, 0.05)
40°S–20°S	226	2.60	−0.76	1.21	−0.58	1.36	0.89	0.08	0.02	0.03
			(−0.93, −0.58)	(1.14, 1.29)	(−0.76, −0.40)	(1.29, 1.44)	(0.87, 0.91)	(0.05, 0.11)	(0.01, 0.03)	(0.02, 0.04)
20°S–20°N	279	2.01	−0.97	1.29	−0.62	1.36	0.95	0.04	0.02	0.03
			(−1.13, −0.82)	(1.22, 1.37)	(−0.80, −0.43)	(1.27, 1.46)	(0.92, 0.98)	(0.02, 0.06)	(0.01, 0.03)	(0.02, 0.04)
20°N–40°N	138	1.91	−0.82	1.23	−0.38	1.26	0.97	0.09	0.02	0.03
			(−1.19, −0.45)	(1.02, 1.43)	(−0.76, −0.01)	(1.06, 1.47)	(0.93, 1.01)	(0.06, 0.11)	(0.01, 0.03)	(0.02, 0.05)
40°N–80°N	158	2.41	−0.81	1.20	−0.58	1.33	0.90	0.12	0.03	0.04
			(−1.02, −0.60)	(1.09, 1.31)	(−0.80, −0.35)	(−0.80, −0.35)	(0.87, 0.93)	(0.08, 0.16)	(0.01, 0.04)	(0.02, 0.05)

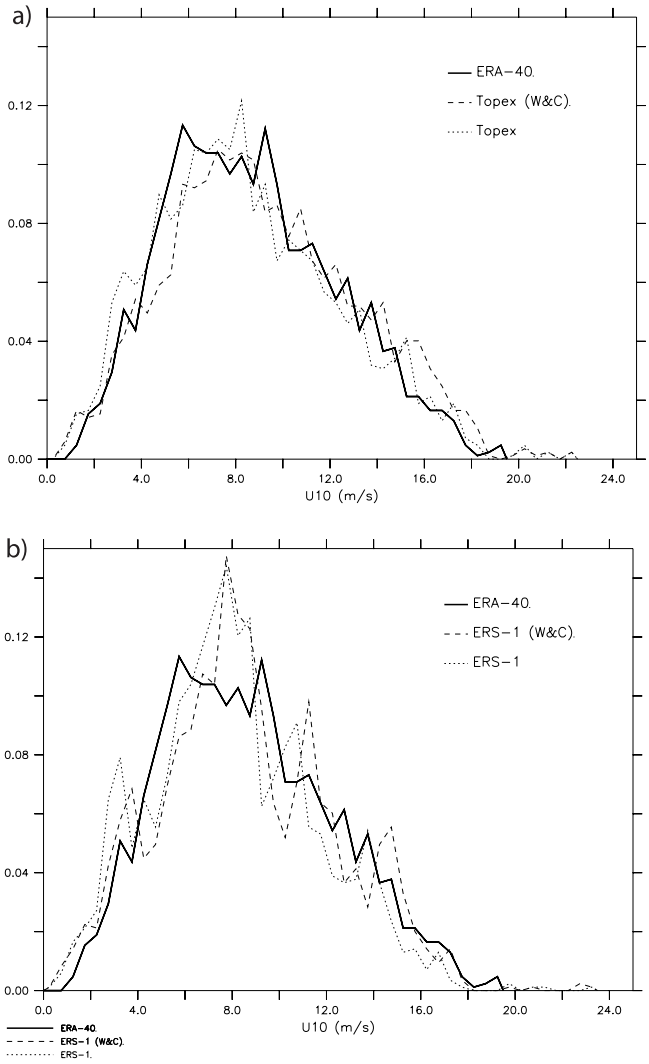

Figure 8. (top) Histograms of ERA-40 wind speed estimates (solid line), Topex altimeter wind speed estimates using the *Witter and Chelton* [1991] algorithm (dashed line), and using the *Gourrion et al.* [2002] algorithm (dotted line). (bottom) Histograms of ERA-40 wind speed estimates (solid line), ERS-1 altimeter wind speed estimates using the *Witter and Chelton* [1991] algorithm (dashed line), and using the *Gourrion et al.* [2002] algorithm (dotted line). Data are from June to December 1993.

been considered here. For high wind speeds ($U_{10} > 18$ m/s) we have considered the algorithm of *Young* [1993].

[65] Care was taken so that the collocated data had compatible time and space scales by averaging the buoy measurements in time and the altimeter measurements in time and space. The results show that in terms of significant wave height the Topex and buoy observations compare well with each other. The ERS-1 measurements are lower than the buoy and Topex observations. This is in line with linear corrections obtained by calibrating ERS-1 data proposed in the literature and that were not applied here. The ERA-40 values of H_s slightly overestimate low (<1.5 m) values of Topex and buoy, and underestimate high values by more than 20%. The estimated variances of the random errors associated with the ERA-40 H_s observations are much higher than those of the measuring instruments. The standard deviation of the ERA-40 H_s random errors vary, depending on H_s, from about 30 cm for $H_s \approx$ 2m to 45 cm for $H_s \approx$ 3.5 m. The standard deviations of the errors of the measuring instruments do not seem to depend on the value of H_s

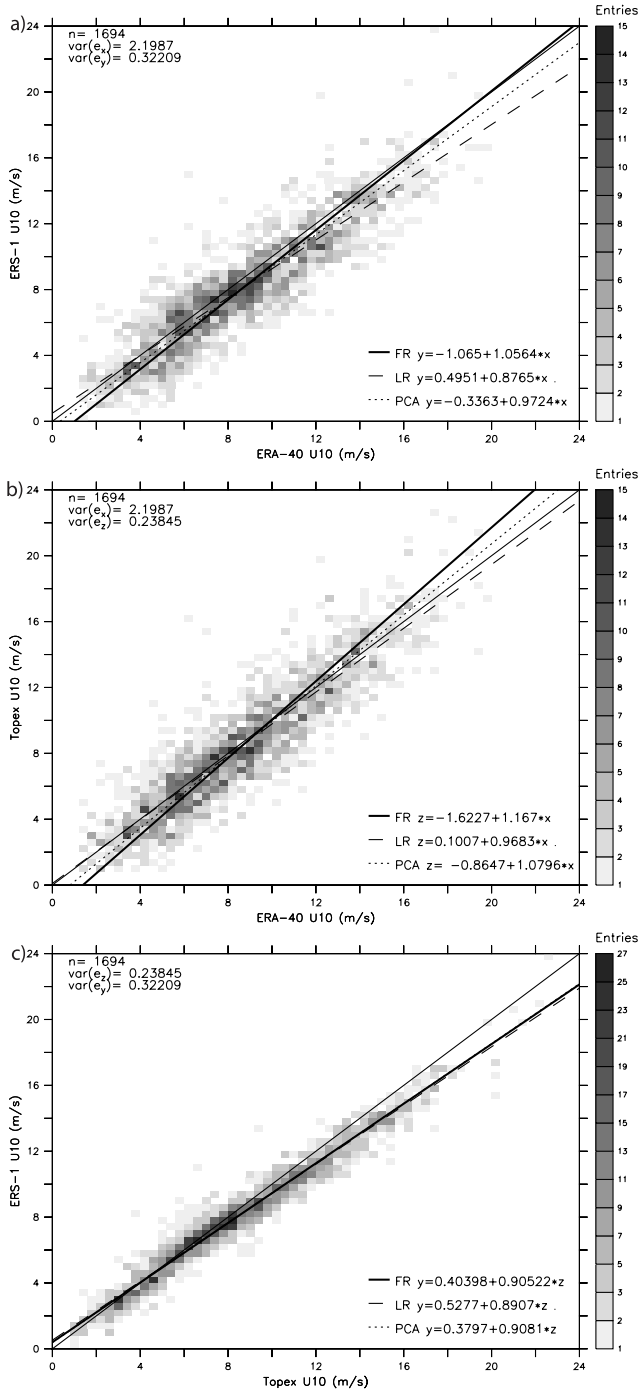

Figure 9. Scatter diagrams with estimated FR, LR and PCA lines for wind speed triple ERA-40, ERS-1 and Topex collocated data from June to December 1993. (a) ERA-40 versus ERS-1. (b) ERA-40 versus Topex. (c) ERS-1 versus Topex. The altimeter wind speed estimates were obtained using the *Gourrion et al.* [2002] algorithm.

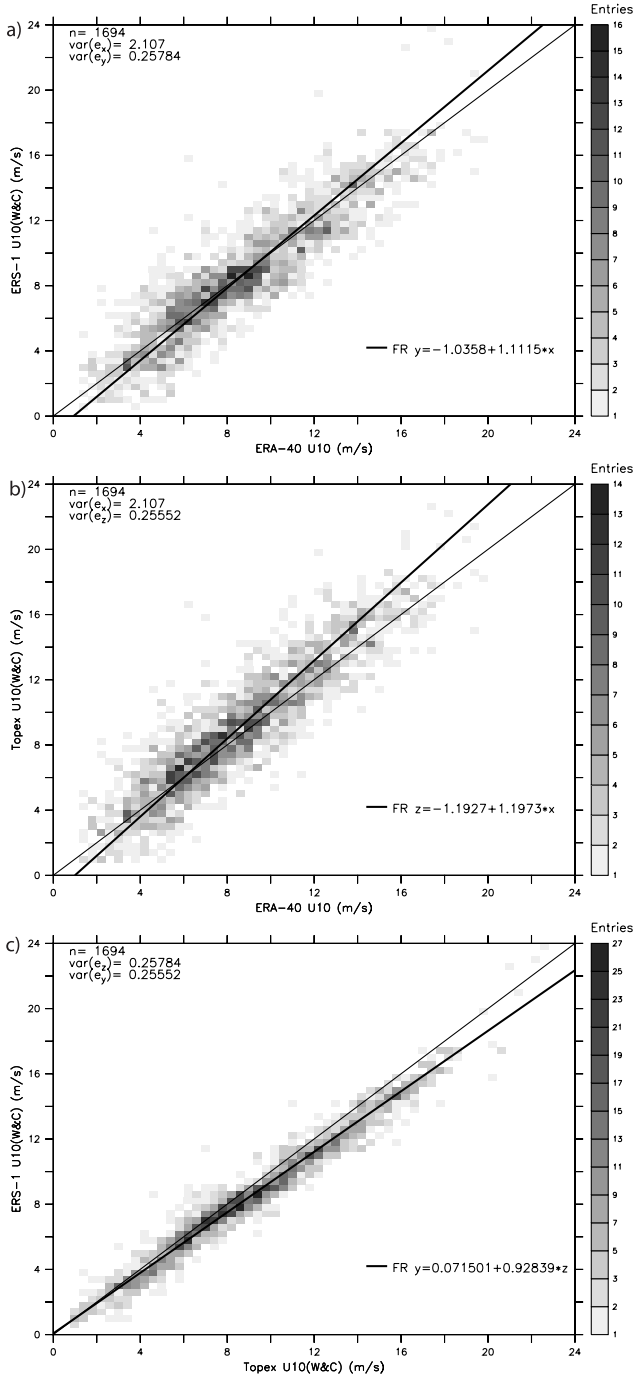

Figure 10. The same as Figure 9, but for winds obtained using the *Witter and Chelton* [1991] algorithm.

and are lower than 15 cm. Our results indicate a positive correlation between the random errors associated with the two altimeter measurements.

[66] The comparisons of wind speed are inconclusive about which of the two altimeter algorithms for low to moderate wind speeds behaves better, since they do not alter the error structure or calibration of the data. Compared with the ERS-1 data, the Topex wind speed observations show a positive bias, increasing with U_{10} and of about 2 m/s for $U_{10} \approx 22$ m/s. The ERA-40 data show a positive bias when compared with ERS-1 data below about 8 m/s and a negative one above it. Again, in these comparisons the random errors associated with the ERA-40 observations are higher than those of the measuring instruments, and show a dependence on the value of U_{10}. For ERA-40 the standard deviations of the errors vary from 1.1 to 1.6 m/s. For the altimeters the value is of about 0.5 m/s.

[67] The ERA-40 overestimation (underestimation) of low (high) wave heights can be partially explained by the corresponding overestimation (underestimation) of low (high) wind speeds. There are, however, other factors, such as resolution and model deficiencies, contributing to these mismatches. In this respect our assessment of the ERA-40 data is in line with those of *Sterl et al.* [1998] and *Bauer and Staabs* [1998], where WAM predictions forced with ECMWF winds were thoroughly assessed; of course no such detailed study was intended nor possible here. Other wave validation studies (such as those of *Bauer and Heimbach* [1999] and *Heimbach et al.* [1998]) manage to divide the samples into pure wind sea and swell cases; the sample sizes used here do not allow such a distinction. However, in order to get an idea of how the quality of the ERA-40 H_s data depends on the mean wave period we have computed the bias between the ERA-40 and the Topex H_s data as a function of the ERA-40 mean wave period. The plot (not shown here) has a small trough of -0.25 m for a mean wave period of about 4.5 s, and the bias decreases uniformly from -0.12 m for 5.3 s to about -0.48 m for a mean wave period of 12 s. This allows no deductions in terms of wind sea and swell, since most of the sea states considered here are mixed, but allows us to say that the underestimation of the ERA-40 wave height tends to increase with the mean wave period.

[68] The estimates of the triple collocation model were, in some cases, compared with the corresponding PCA and LR estimates. As expected, the larger differences were found between the FR and LR estimates. In the case of wind speed data a comparative study based on LR estimates would lead to the conclusion that there is no underestimation of the high wind speeds by ERA-40, in contrast with our conclusions.

[69] Let us note that our estimates of the variances of the errors in the measurements naturally depend on our way of collocating the data. If instead of considering a 1.5° by 1.5° square region centered at the buoy location we use, for example, a circular region of 200 km radius centered at the buoy location, the estimates for the variance of the altimeter errors are about 3 times as large as those shown here, while those of the variances of the buoy and model errors would remain the same. This is understandable because an increase in the collocation area represents an increase in the sampling variability of the altimeter data. Also, if instead of computing the buoy super observations by averaging 3-hourly observations, we average over 1 (5) hours, the estimates of the buoy errors would be closer to zero (respectively closer to 0.01). Finally, if instead of interpolating the ERA-40 data to the buoy location we interpolate it to the midpoint between the buoy and the altimeter observation, we slightly increase (decrease) the variances of the errors of the buoy (respectively altimeter) measurements.

[70] Although fulfilling the objectives of our study, this assessment of ERA-40 data provides no guidelines on the

Table 7. Estimates, for Different Latitude Strips, of the Functional Relationship Coefficients Between ERA-40(x), ERS-1(y), and Topex(z) Wind Speed Observations, and of the Variances of the Errors[a]

	n	$\langle x \rangle$	α_1	β_1	α_2	β_2	β_3	$\langle e_x^2 \rangle$	$\langle e_y^2 \rangle$	$\langle e_z^2 \rangle$
					U_{10}					
80°S–80°N	1694	8.67	−1.06	1.06	−1.62	1.17	0.91	2.20	0.32	0.24
			(−1.31, −0.82)	(1.01, 1.10)	(−1.86, −1.39)	(1.12, 1.21)	(0.89, 0.92)	(2.01, 2.39)	(0.24, 0.40)	(0.15, 0.33)
80°S–40°S	893	10.21	−1.40	1.06	−2.00	1.18	0.90	2.47	0.25	0.36
			(−1.80, −1.00)	(1.02, 1.11)	(−2.43, −1.57)	(1.14, 1.23)	(0.88, 0.92)	(2.14, 2.79)	(0.13, 0.37)	(0.24, 0.48)
40°S–20°S	226	7.25	−1.67	1.14	−1.87	1.22	0.94	2.22	0.18	0.23
			(−2.33, −1.00)	(1.05, 1.24)	(−2.59, −1.15)	(1.11, 1.33)	(0.90, 0.97)	(1.65, 2.80)	(0.03, 0.34)	(0.08, 0.38)
20°S–20°N	279	6.54	−1.87	1.26	−2.05	1.29	0.97	1.40	0.32	0.11
			(−2.64, −1.09)	(1.14, 1.38)	(−2.78, −1.32)	(1.18, 1.40)	(0.93, 1.02)	(1.09, 1.70)	(0.10, 0.54)	(0.00, 0.24)
20°N–40°N	138	6.20	−1.55	1.18	−1.53	1.20	0.98	1.40	0.31	0.14
			(−2.37, −0.72)	(1.05, 1.32)	(−2.29, −0.76)	(1.07, 1.34)	(0.92, 1.05)	(1.01, 1.79)	(0.05, 0.57)	(0.00, 0.48)
40°N–80°N	158	7.94	−1.76	1.18	−2.45	1.28	0.92	1.80	0.45	0.23
			(−2.80, −0.72)	(1.03, 1.33)	(−3.43, −1.47)	(1.14, 1.42)	(0.87, 0.97)	(1.23, 2.36)	(0.20, 0.70)	(0.00, 0.49)
					U_{10}(*W & C*)					
80°S–80°N	1694	8.67	−1.04	1.11	−1.19	1.20	0.93	2.11	0.26	0.26
			(−1.30, −0.78)	(1.06, 1.16)	(−1.44, −0.94)	(1.15, 1.24)	(0.90, 0.95)	(1.91, 2.30)	(0.16, 0.36)	(0.16, 0.35)
80°S–40°S	893	10.21	−0.99	1.09	−1.04	1.17	0.93	2.43	0.22	0.33
			(−1.43, −0.55)	(1.04, 1.14)	(−1.48, −0.59)	(1.12, 1.22)	(0.91, 0.95)	(2.12, 2.74)	(0.08, 0.36)	(0.19, 0.48)
40°S–20°S	226	7.25	−1.51	1.19	−1.44	1.26	0.95	2.17	0.18	0.22
			(−2.27, −0.76)	(1.08, 1.30)	(−2.23, −0.65)	(1.14, 1.38)	(0.91, 0.98)	(1.70, 2.65)	(0.02, 0.33)	(0.05, 0.38)
20°S–20°N	279	6.54	−1.95	1.30	−2.01	1.35	0.96	1.38	0.28	0.10
			(−2.72, −1.17)	(1.18, 1.41)	(−2.86, −1.17)	(1.22, 1.47)	(0.92, 1.00)	(1.07, 1.69)	(0.08, 0.49)	(0.00, 0.24)
20°N–40°N	138	6.20	−1.51	1.21	−1.39	1.24	0.97	1.33	0.26	0.16
			(−2.21, −0.81)	(1.08, 1.33)	(−2.12, −0.66)	(1.12, 1.37)	(0.91, 1.02)	(0.91, 1.76)	(0.02, 0.50)	(0.00, 0.52)
40°N–80°N	158	7.94	−1.91	1.23	−2.27	1.32	0.93	1.68	0.30	0.34
			(−2.90, −0.93)	(1.10, 1.37)	(−3.22, −1.32)	(1.19, 1.45)	(0.89, 0.98)	(1.22, 2.15)	(0.09, 0.50)	(0.10, 0.58)

[a]Data from June to December 1993. The limits of the confidence intervals are given in parenthesis below the estimates. The sample size, n, and the average of the ERA-40 data, $\langle x \rangle$, are also included. The upper half corresponds to data obtained using the *Gourrion et al.* [2002] algorithm, and the lower half corresponds to data obtained using the *Witter and Chelton* [1991] algorithm.

use of climatology statistics, such as return values, that may be computed from the ERA-40 data. To obtain these, a more extensive assessment of the data for different periods, not restricted to locations with at least two measurements from different instruments available, is needed, together with an in-depth extreme values analysis. The results of this study do, however, give a good measure of the characteristics of the ERA-40 data for the period without altimeter assimilation. In principle, the fact that the statistical method we propose is able to account for dependence between the random errors of different data sets suggests that triple collocation of ERS, Topex and ERA-40 data could also be used to assess the ERA-40 data in the periods when ERS data is assimilated into ERA-40. Such a study, however, requires a reliable estimate of the correlation between the ERA-40 data and the ERS data, which we are not able to provide.

[71] The statistical method developed and applied here helps give insight into the way random errors in the data affect comparative linear functions. The fact that estimates of the variances of the random errors can also be estimated suggests other areas of applicability, as, for instance, data assimilation. Most of the assimilation techniques require estimates of the variance of the random errors for both model data and measurements, and so far there is no method to provide such estimates, most of data assimilation applications using ad hoc estimates of variances, especially for the model data.

Table 8. Estimates of the Slope Between ERA-40(x), ERS-1(y), and Topex(z) Wind Speed Data Obtained by the FR, LR and PCA Models[a]

	$y = \alpha + \beta * x$			$z = \alpha + \beta * x$			$y = \alpha + \beta * z$		
	FR	LR	PCA	FR	LR	PCA	FR	LR	PCA
α	−1.07	0.50	−0.34	−1.62	0.10	−0.87	0.40	0.53	0.38
β	1.06	0.88	0.97	1.17	0.97	1.08	0.91	0.89	0.91

[a]Data from June to December 1993.

[72] **Acknowledgments.** We would like to thank Helen Snaith of SOC for the altimeter data, the American National Oceanographic Data Center for the available buoy data, and the ECMWF ERA-40 team for the ERA-40 data and their prompt reaction to all our queries. Special thanks to Peter Janssen and Jean Bidlot at ECMWF, and to our colleagues Gerrit Burgers and Gerbrand Komen, for fruitful discussions and comments, and to Camiel Severijns for the Field library and technical support. This work was funded by the EU-funded ERA-40 Project (EVK2-CT-1999-00027).

References

AVISO, AVISO CD ROM User Manual: Merged TOPEX/POSEIDON Products, *Avi-nt-02-100-cn*, Cent. Natl. d'Etudes Spatiales, Toulouse, France, 1996.

Bauer, E., and P. Heimbach, Annual validation of significant wave heights of ERS-1 synthetic aperture wave mode spectra using TOPEX/Poseidon and ERS-1 altimeter data, *J. Geophys. Res.*, *104*(C6), 13,345–13,357, 1999.

Bauer, E., and C. Staabs, Statistical properties of global significant wave heights and their use for validation, *J. Geophys. Res.*, *103*(C1), 1153–1166, 1998.

Bidlot, J.-R., D. J. Holmes, P. A. Wittmann, R. Lalbeharry, and H. S. Chen, Intercomparison of the performance of operational wave forecasting systems with buoy data, *Weather Forecasting*, *17*(2), 287–310, 2002.

Caires, S., Comparative study of HF radar measurements and wave model hindcasts of waves in shallow waters, Ph.D. thesis, Univ. of Sheffield, Sheffield, England, 2000.

Caires, S., A. Sterl, J.-R. Bidlot, N. Graham, and V. Swail, Climatological assessment of reanalysis ocean data, paper presented at 7th International Workshop on Wave Hindcasting and Forecasting, U.S. Army Eng. Res. and Dev. Cent., Banff, Alberta, Canada, 21–26 October 2002.

Chelton, D. B., and F. J. Wentz, Further development of an improved altimeter wind speed algorithm, *J. Geophys. Res.*, *91*, 14,250–14,260, 1986.

Cotton, P. D., P. G. Challenor, and D. J. T. Carter, An assessment of the accuracy and reliability of GEOSAT, ERS-1, ERS-2 and TOPEX altimeter measurements of significant wave height and wind speed, *ESA WPP-147*, Eur. Space Agency, Paris, France, 1997.

Cox, A. T., and V. R. Swail, A global wave hindcast over the period 1958–1997: Validation and climate assessment, *J. Geophys. Res.*, *106*(C2), 2313–2329, 2001.

Cox, D. R., and D. V. Hinkley, *Theoretical Statistics*, Chapman and Hall, New York, 1974.

Efron, B., and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, New York, 1993.

Ferguson, T. S., *A Course in Large Sample Theory*, Chapman and Hall, New York, 1996.

Gibson, J. K., P. Kållberg, S. Uppala, A. Nomura, A. Hernandez, and E. Serrano, ERA description, *Re-Anal. Final Rep. Ser. 1*, 71 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, England, 1997.

Gourrion, J., D. Vandemark, S. Bailey, B. Chapron, C. P. Gommenginger, P. G. Challenor, and M. A. Srokosz, A two parameter wind speed algorithm for Ku-band altimeters, *J. Atmos. Oceanic Technol.*, *19*(12), 2030–2048, 2002.

Heimbach, P., K. Hasselmann, and S. Hasselmann, Statistical analysis and intercomparison of WAM model data with global ERS-1 SAR wave mode spectral retrievals over 3 years, *J. Geophys. Res.*, *103*(C4), 7931–7977, 1998.

Janssen, P. A. E. M., J. D. Doyle, J. Bidlot, B. Hansen, L. Isaksen, and P. Viterbo, Impact and feedback of ocean waves on the atmosphere, in *Advances in Fluid Mechanics*, vol. 1, edited by W. Perrie, pp. 155–197, WIT Press, Southampton, England, 2002.

Snaith, H. M., Global altimeter processing scheme user manual, vol. 1, technical report, 44 pp., Southampton Oceanogr. Cent., Southampton, England, 2000.

Sterl, A., G. J. Komen, and P. D. Cotton, Fifteen years of global wave hindcasts using winds from the European Centre for Medium-Range Weather Forecasts reanalysis: Validating the reanalyzed winds and assessing the wave climate, *J. Geophys. Res.*, *103*(C3), 5477–5494, 1998.

Tan, C. Y., and B. Iglewicz, Measurement-methods comparisons and linear statistical relationship, *Technometrics*, *41*(3), 192–201, 1999.

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

Wang, X. L., and V. R. Swail, Changes of extreme wave heights in Northern Hemisphere oceans and related atmospheric circulation regimes, *J. Clim.*, *14*, 2204–2221, 2001.

WASA Group, Changing waves and storms in the Northeast Atlantic?, *Bull. Am. Meteorol. Soc.*, *79*, 741–760, 1998.

Witter, D. L., and D. B. Chelton, A Geosat altimeter wind speed algorithm and a method for wind speed algorithm development, *J. Geophys. Res.*, *96*(C5), 18,853–18,860, 1991.

Young, I. R., An estimate of the Geosat altimeter wind speed algorithm at high wind speeds, *J. Geophys. Res.*, *98*(C11), 20,275–20,285, 1993.

Young, I. R., *Wind Generated Ocean Waves, Ocean Eng. Book Ser.*, vol. 2, Elsevier, New York, 1999a.

Young, I. R., An intercomparison of GEOSAT, TOPEX and ERS-1 measurements of wind speed and wave height, *Ocean Eng.*, *26*, 67–81, 1999b.

S. Caires and A. Sterl, Royal Netherlands Meteorological Institute, P.O. Box 201, NL-3730 AE De Bilt, Netherlands. (caires@knmi.nl)