100-Year Return Value Estimates for Ocean Wind Speed and Significant Wave Height from the ERA-40 Data

S. CAIRES AND A. STERL

Royal Netherlands Meteorological Institute, De Bilt, Netherlands

(Manuscript received 2 March 2004, in final form 3 September 2004)

ABSTRACT

In this article global estimates of 100-yr return values of wind speed and significant wave height are presented. These estimates are based on the ECMWF 40-yr Re-Analysis (ERA-40) data and are linearly corrected using estimates based on buoy data. This correction is supported by global Topographic Ocean Experiment (TOPEX) altimeter data estimates. The calculation of return values is based on the peaks-over-threshold method. The large amount of data used in this study provides evidence that the distributions of significant wave height and wind speed data belong to the domain of attraction of the exponential. Further, the effect of the space and time variability of significant wave height and wind speed on the prediction of their extreme values is assessed. This is done by performing detailed global extreme value analyses using different decadal subperiods of the 45-yr-long ERA-40 dataset.

1. Introduction

The design of ship, offshore, and coastal structures requires a good knowledge of the most severe wind and wave conditions that they need to withstand during their lifetime. This knowledge is often difficult to infer because the amount of data available is usually small, and often there is no data at all on which to base inferences.

The European Centre for Medium-Range Weather Forecasts (ECMWF) has recently completed ERA-40, a global reanalysis of meteorological variables, among which are ocean winds and waves from 1957 to 2002. The data consist of 6-hourly fields on a 1.5° × 1.5° latitude/longitude grid covering the whole globe. The time and space coverage of this dataset makes it ideal for the study of extreme wind and wave phenomena over the whole globe. Initial validations of the data reveal a generally good description of variability and trends (Caires et al. 2004) but some underestimation of high wave heights and wind speeds (Caires and Sterl 2003b). The objective of this article is to use the ERA-40 data to compute global estimates of return values of significant wave height (H_s) and near-surface wind speed (U_{10}). Particular attention will be paid to the effect the ERA-40 underestimation has in the estimation of return values by comparing parameter estimates obtained from the ERA-40 dataset with those obtained from in situ buoy and global altimeter measurements.

Corresponding author address: Sofia Caires, KNMI, P.O. Box 201, NL-3730 AE De Bilt, Netherlands.
E-mail: caires@knmi.nl

Following Ferreira and Guedes Soares (1998, 2000) and Anderson et al. (2001), we compute the return value estimates using the peaks-over-threshold (POT) method rather than the widespread approach of fitting a distribution (such as the lognormal, Weibull, beta, etc.) to the whole dataset and extrapolating from it. Although these authors only looked at H_s data, their objections apply equally to U_{10} data. Among the arguments given by these authors against the latter method we mention the following:

- Due to dependence and nonstationarity, H_s and U_{10} series violate the assumptions of independence and identity in distribution, which invalidates the application of the common statistical methods used (confidence intervals and tests) as well as the definition of return value.
- There is no scientific justification for using one particular distribution to fit to H_s or U_{10} data, and the usual goodness-of-fit diagnostics are not able (on the basis of realistic sample sizes and given the length of the required "prediction horizon") to distinguish data with type I (exponential) tail, say, from data with type II (heavier than exponential) tail. In contrast, if for example one concentrates on averages, maximum values, or excesses over a high threshold of very general variables, then statistical theory provides a scientific basis for the use of, respectively, the normal, generalized extreme value and generalized Pareto distributions.

Another statistically sound approach to obtain return values from the data would be to use the annual maxima method (see, e.g., Coles 2001), in which the

Generalized Extreme Value Distribution is fitted to the sample of annual maxima. This approach is not considered here because the size of the samples obtained from 45 or 10 years of data is too small for reasonably accurate inferences.

The POT approach (e.g., Coles 2001) consists of fitting the generalized Pareto distribution to the peaks of clustered excesses over a threshold, the excesses being the observations in a cluster minus the threshold, and calculating return values by taking into account the rate of occurrence of clusters. Under very general conditions this procedure ensures that the data can have only three possible, albeit approximate, distributions (the three forms of the generalized Pareto distribution) and, moreover, that observations belonging to different peak clusters are (approximately) independent.

Ferreira and Guedes Soares (1998) described an application of the POT method to predict return values of H_s. The method was applied to 10 years of H_s data from one location off the Portuguese coast, and it was concluded that the exponential distribution (the generalized Pareto distribution with shape parameter $\kappa = 0$; see below) fitted the data well. Moreover, the closely exponential character of the data was explained theoretically: the H_s estimate of each sea state is a realization of a random variable whose approximate distribution is a mixture of Rayleigh distributions with an unspecified number of parameters, which has a type I or exponential tail. There is also a lot of literature hypothesizing the exponential character of U_{10} data; see Simiu et al. (2001) and references therein. The exponential character of both ERA-40 and buoy H_s and U_{10} observations will be extensively assessed in this article using data from several locations. As we shall see, the establishment of the exponential distribution as an appropriate model for the peak excesses not only points to a general pattern but also simplifies the problem of predicting extreme values since one parameter less needs to be estimated.

Many past studies of extreme values of ocean winds and waves were based on a rather limited number of years of data, usually no more than a decade. The climate, however, changes, and it is interesting to investigate how well estimates obtained from certain data periods compare with estimates from other periods. With this in mind, we will not only compute estimates using the whole 45-yr dataset but also using three decadal subsets (1958–67, 1972–81, and 1986–95) and analyze the time and space variability of these estimates. However, we will not look at the effect that within-year variability and within-decade trends may have on the return value estimates; for a study of these effects the reader is referred to Anderson et al. (2001).

2. Data description

a. ERA-40 data

ERA-40 is the name of ECMWF's most recent reanalysis of global meteorological quantities, including ocean winds and waves, from 1957 to 2002. The reanalysis used ECMWF's Integrated Forecasting System (IFS), a coupled atmosphere–wave model with variational data assimilation (see Simmons 2001). This is a state-of-the-art model very similar to the one used operationally for weather forecasts, though with lower resolution. The aim of the reanalysis was to produce a dataset with no inhomogeneities, as far as the technique of analysis is concerned, by reconstructing the 45 years of data using the same numerical model throughout. This is the fourth reanalysis performed, but the first in which a wave model is coupled to the atmosphere model (see Janssen et al. 2002). In previous reanalyses, wave data had to be generated offline by forcing a wave model by the reanalyzed winds; an overview of these efforts can be found in Caires et al. (2004). In terms of the ocean wave data, the present reanalysis has the largest time and space coverage. A large subset of the ERA-40 fields (among which H_s and U_{10}) can, for research purposes, be freely downloaded from ECMWF's Web page online at http://data.ecmwf.int/data/d/era40. The H_s and U_{10} data consist of 6-hourly fields on a 1.5° $\times$ 1.5° latitude/longitude grid covering the whole globe.

b. Buoy measurements

So far, buoy observations are considered the most reliable wave observations, but they are limited to some locations along the coast, mainly in the Northern Hemisphere, and are available only at a small number of locations before 1978. From 1978 onward buoy observations from the National Oceanic and Atmospheric Administration (NOAA)/National Data Buoy Center (NDBC) off the coast of North America are available online at http://www.nodc.noaa.gov/BUOY/buoy.html, and can be retrieved for free. Due to their high quality they will be used here to assess the return value estimates obtained from the ERA-40 data.

From all the NDBC data buoy locations available, we have selected a total of 20 locations for the validations: 1 off the coast of Peru, 4 around the Hawaiian Islands, 3 in the Gulf of Mexico, 5 in the Northwest Atlantic, 3 off the coast of Alaska, 3 in the Northeast Pacific, and 1 off the coast of California; see Fig. 1. 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 since the buoys should not be too close to the coast in order for the corresponding grid points to be located at sea. The buoy H_s and U_{10} measurements are available hourly from 20- and 10-min-long records, respectively. 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) and described in Caires and Sterl (2003b, 43.2–3). When the anemometers of the buoys are not at a height of 10 m, the wind speed measurements are adjusted to that height using a logarithmic

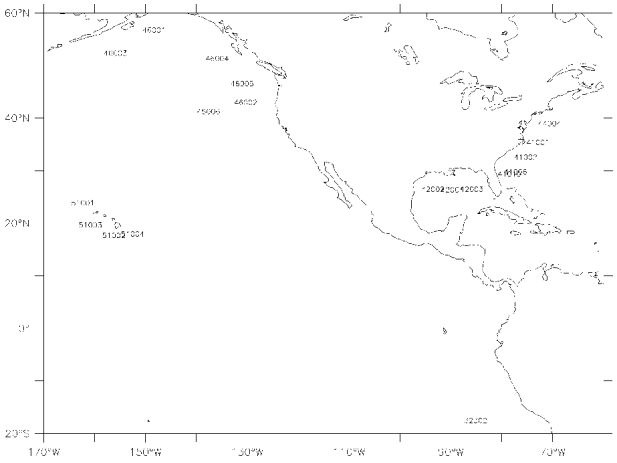

FIG. 1. Buoy codes and locations.

profile under neutral stability (see, e.g., Bidlot et al. 2002, p. 291). To compare the ERA-40 results with the observations, time and space scales must be brought as close to each other as possible. The reanalysis results are available at synoptic times (every 6 h, at 0000, 0600, 1200, and 1800 UTC) and each value is an estimate of the average condition in a grid cell; on the other hand, the buoy measurements are local. To make the time and space scales of the data compatible, the reanalysis data are compared with 3-h averages of buoy observations centered around the synoptic times, 3 h being the approximate time a long wave would take to cross the diagonal of a 1.5° × 1.5° grid cell at midlatitudes. To get ERA-40 data at the buoy location, the reanalysis data at the appropriate synoptic time are interpolated bilinearly to the buoy location.

c. Altimeter measurements

The Topographic Ocean Experiment (TOPEX) along-track quality checked deep water altimeter measurements of H_s and the normalized radar cross section (σ_0) were obtained from the Southampton Oceanography Center (SOC) Global Altimetry Processing Scheme (GAPS) interface online at http://www.soc.soton.ac.uk/ALTIMETER (Snaith 2000). There are several corrections available to bring the altimeter measurements closer to the buoy's. The TOPEX wave height observations for 1997–99 (cycles 170–235) have drifted; the drift is corrected according to Challenor and Cotton (1999). Caires and Sterl (2003b), using a functional linear relationship model, found that TOPEX data relate to the buoy data according to $H_s^{\text{buoy}} = 1.05H_s^{\text{topex}} - 0.07$. We have made the TOPEX observations used here compatible with the buoy observations by applying this relationship.

Although altimeters do not measure U_{10} directly, the altimeter backscatter depends on (i.e., correlates highly with) the sea surface wind speed. There are several empirical algorithms available to compute the wind speeds up to 20 m s^{-1} from σ_0. The most recent algorithm is due to Gourrion et al. (2002) and is used here. Caires and Sterl (2003b) have compared the data produced using this algorithm and those due to Witter and Chelton (1991), which are used operationally for the TOPEX/Poseidon satellite altimeters; the results were inconclusive as to which of the algorithms should be preferred. For wind speed above 20 m s^{-1} the relation of Young (1993) is used.

The satellite measurements are performed about every second with a spacing of about 5.8 km. From these we form altimeter "observations" by grouping together the consecutive observations crossing a 1.5° × 1.5° latitude–longitude region (observations at most 30 s or $1.5\sqrt{2}$° apart). The altimeter observation is taken as the mean of these grouped data points after a quality control similar to the one applied to the buoy data. The reanalysis data at the synoptic times before and after the time of the altimeter observation are interpolated bilinearly to the mean observation location, and these two data points are then linearly interpolated in time to the mean time of the observation.

3. Peaks-over-threshold method

a. Basics

In the POT method, the peak excesses over a high threshold u of a time series are assumed to occur according to a Poisson process with rate λ_u and to be independently distributed with a generalized Pareto distribution (GPD), whose distribution function is given by

$$F_u(x) = \begin{cases} 1 - (1 - \kappa x/\alpha)^{1/\kappa} & \text{if } \kappa \neq 0 \\ 1 - \exp(-x/\alpha) & \text{if } \kappa = 0, \end{cases}$$

where the range of x is $(0, \infty)$ if $\kappa \leq 0$ and $(0, \alpha/\kappa)$ if $\kappa > 0$. For $\kappa = 0$ the GPD is the exponential distribution with mean α, for $\kappa < 0$ it is the Pareto distribution, and for $\kappa > 0$ it is a special case of the beta distribution. For $\kappa < 0$ the tail of the GPD; that is, the function $x \to 1 - F_u(x)$ is heavier (i.e., decreases more slowly) than the tail of the exponential distribution, and for $\kappa > 0$ it is lighter (decreases more quickly and actually reaches 0) than that of the exponential. The GPD is said to have a type II tail for $\kappa < 0$ and a type III tail for $\kappa > 0$. The tail of the exponential distribution is called a type I tail.

As stated in the introduction, the excesses over the threshold u of a time series $X_1, X_2, \ldots$ are the observations (called exceedances) $X_i - u$ such that $X_i > u$. A *peak excess* is defined as the largest excess in a cluster of exceedances; hence its definition depends on that of a cluster. In this paper we shall adopt the usual definition of a cluster as a group of consecutive exceedances. The model now outlined can be justified theoretically for a great variety of time series; see, for example, Leadbetter (1991) and the references therein.

One of the main applications of the POT method is the estimation of the m-year return value, $x_m^{(u)}$. If the

number of clusters *per year* is a Poisson random variable with mean λ_u, then the expected number of clusters/peak excesses in m years is $m\lambda_u$ and, if the peak excesses over u are independently distributed with distribution function F_u, then the expected number of observations exceeding x is $m\lambda_u[1 - F_u(x)]$; setting this equal to 1 and solving for x gives

$$x_m^{(u)} = \begin{cases} u + \dfrac{\alpha}{\kappa}[1 - (\lambda_u m)^{-\kappa}] & \text{if } \kappa \neq 0 \\ u + \alpha \log(\lambda_u m) & \text{if } \kappa = 0. \end{cases} \quad (1)$$

b. Estimation and testing

Once a threshold u has been selected and the peak excesses have been extracted from the time series, the scale (α) and shape (κ) parameters of the GPD can be estimated by a variety of methods such as maximum likelihood, the method of moments, etc. The estimates of the GPD presented here were obtained using the Method of Probability-Weighted Moments; see Hosking and Wallis (1987) for this method and some of its advantages over maximum likelihood.

The choice of a threshold should take into account the threshold stability property of the GPD: if the GPD model is valid for peaks over the threshold u_0, it is also valid for peaks over the threshold $u > u_0$ with the same shape parameter κ and an "adjusted" scale parameter. Since higher thresholds will generate fewer peaks with which the GPD parameters can be estimated (see Davison and Smith 1990, p. 395), one should ideally choose the lowest threshold at which the GPD is valid.

The parameter λ_u, the yearly cluster rate, can be estimated by the average number of clusters/peak excesses per year. More generally, for yearly series with different numbers of observations, λ_u can be estimated by[1]

$$\hat{\lambda}_u = k^{-1} \sum_{i=1}^{k} N_i / p_i, \quad (2)$$

where k is the number of years considered, $p_i = n_i/n$, n_i is the number of observations available in the ith year, N_i is the corresponding number of peak excesses, and n is the maximum number of observations in a yearly series. Under the Poisson assumption, $E(\hat{\lambda}_u) = \lambda_u$ and

$$\mathrm{var}(\hat{\lambda}_u) = \lambda_u k^{-2} \sum_{i=1}^{k} p_i^{-1},$$

and, since λ_u is relatively large, we have that $\hat{\lambda}_u$ is approximately normal with mean λ_u and variance $\mathrm{var}(\hat{\lambda}_u)$.

Confidence intervals for the return value estimates can be estimated by using the delta method (see Ferguson 1996); more precisely, the asymptotic variance of the return value estimates can be estimated by

$$\widehat{\mathrm{var}(x_m^{(u)})} = d^{\mathrm{T}} \boldsymbol{\Sigma} d,$$

where d is the vector of derivatives of $x_m^{(u)}$ with respect to the estimated parameters (λ, κ, and α if $\kappa \neq 0$, and λ and α if $\kappa = 0$) and $\boldsymbol{\Sigma}$ the asymptotic covariance matrix of the parameter estimates, both evaluated at the estimates of the parameters.

When the data can be assumed independent, the POT method uses all the observations above a threshold. Since the data we are studying are dependent (especially in the case of H_s), the POT method we shall apply uses only the peak excesses above a threshold. The nature of the data makes it necessary to impose another restriction, namely to take only a single peak excess from two or more neighboring clusters that happen to be too close to each other. It can happen that within a storm there is a somewhat calmer period followed by another rough period, in which case the time series might go below the threshold and then rise again, thus creating two clusters out of the same storm. In order that no more than one observation is taken from the same storm, we shall treat clusters at a distance of less than 48 h apart as a single cluster—as if belonging to the same storm—and hence use only the highest of the cluster excesses.

To test the exponentially of the data we will use the Anderson–Darling statistic (see, e.g., Stephens 1974), which can be used for testing the exponential versus any other distribution. We use a 5% significance level, for which the asymptotic critical value of the Anderson–Darling statistic is 1.341.

c. Application of the POT method to the different datasets

Although the ERA-40 data at each location consist of 6-hourly time series with no missing values (missing values occurring only where the ice coverage changes), the buoy time series, which are also 6-hourly, may have a lot of gaps, and the TOPEX time series at each location are not sampled regularly. The application of the POT method to the measurements must therefore be adapted to this special situation: in order to compare the estimates arising from the measurements with those from the ERA-40 data, the ERA-40 data will be sampled in the same way as the measurements so that the same procedure applies to data from different sources.

1) TIME SERIES WITH GAPS

In section 3b we have outlined the application of the POT method to regularly sampled time series. The possibility of gaps in the time series was taken into account by letting p_i in Eq. (2) vary. When obtaining estimates from the ERA-40 dataset alone, there are no gaps in the time series (so $p_i = 1$) in most of the locations, exceptions being the locations where the ice coverage changes. The time series of buoy measurement, on the

[1] Throughout this article $\hat{x}$ denotes an estimator of x.

contrary, often have gaps, some of which are quite large (of more than a year). The existence of gaps within a cluster creates a problem because it is then impossible to say whether the highest observation in the *censored* cluster is the peak of the cluster or not. There are two ways to deal with this problem: one is to use all clusters and treating the cluster maxima in clusters with gaps as censored observations from the GPD (see Davison and Smith 1990, p. 397); the other is to use only noncensored clusters. For the sake of simplicity and because the number of censored clusters is small, we will use the latter approach. When computing ERA-40 estimates at the buoy locations (in order to compare them with the respective estimates arising from the buoy data), the ERA-40 data will be sampled in the same way as the buoy data, thus having the same gaps. When more than one month is of data missing during a given year, the whole year is excluded.

2) IRREGULARLY SAMPLED DATA

The altimeter data are collected along the satellite trajectory. Therefore, no regularly sampled time series of altimeter observations can be obtained at a given location. The TOPEX trajectory has a cycle of 10 days. Due to the way we have combined the 1-s along-track observations, by averaging the values obtained by the altimeter when crossing a $1.5° \times 1.5°$ grid, some grid locations are crossed more than once per cycle, implying that the number of observations in the same grid point per month can be of up to 12, while other grid locations contain at most 3. Since at these time scales the behavior of H_s and U_{10} can vary considerably, whole clusters can be missed. The problem of censoring also arises here since when identifying a data point above a certain threshold there is no way of saying whether it is the peak over the threshold or not; thus the exceedances are observed at random times in storms/clusters. However, it is known—see the discussion in Anderson et al. (2001, p. 71) and references therein—that the distribution of an observation randomly selected from a cluster and the distribution of cluster peaks is asymptotically the same. This means that by collecting the exceedances over a threshold of the altimeter data we should be able to estimate the parameters of the GPD describing the cluster peaks. This is as far as we can go in terms of direct estimates from TOPEX data. Return values cannot be estimated without further assumptions[2] since the number of exceedances per year, λ, cannot be estimated from such scarcely sampled data. Only the threshold and GPD estimates of the ERA-40 data can be compared with those from altimeter data; no comparison of return values can be made.

[2] Anderson et al. (2001) obtain return value estimates from altimeter H_s data assuming that the typical storm duration is known.

4. Validation

In the following analysis, a "year" is defined as the period from October to September. For Northern Hemisphere data this definition is preferable to that of a calendar year because it avoids breaking the winter period in two.[3]

a. Significant wave height

We started our analysis by trying to find "good" thresholds for the buoy and the ERA-40 time series at different buoy locations. Our approach was to choose the threshold as the smallest value at which the GPD fitted the peak excesses reasonably. The threshold is expected to depend on the location, the period considered, and the dataset. The most important factor is supposed to be the location since locations at high latitudes will be exposed to more severe conditions than those at lower latitudes. From assessments of the ERA-40 data against buoy data (see Caires and Sterl 2003b), the thresholds for the ERA-40 data are expected to be lower than those of the buoy data since ERA-40 underestimates the high values of H_s. We have considered time series of 3 or more years (the longest going from 1978 to 2001, the whole period for which the NOAA buoy data are available) and for different periods. We have tried to fix the threshold by assessing the fit of the GPD—comparing fitted densities with kernel density estimates (the continuous analog of the histogram; Silverman 1986) and examining quantile–quantile plots at different thresholds—and looking at the stability of the estimates in relation to the threshold (see, e.g., Coles 2001; Ferreira and Guedes Soares 1998; Anderson et al. 2001). So far we have not been able to devise an automatic procedure to fix the threshold, but found that in most of the cases a threshold fixed at the 93% quantile of the *whole* data gives good results. There were cases, however, where the threshold had to be set higher, up to the value of the 97% quantile, in order to achieve a good fit.

Fixing the threshold at the 93% quantile of the data, we have applied the POT method to buoy data and to the corresponding ERA-40 data for the periods 1980–89, 1990–99, and 1980–99. The reason for looking at 10-yr periods is that we later want to compare the POT estimates obtained with the ERA-40 data from these three periods with those obtained from other 10-yr periods for which no buoy measurements are available. It is of course impossible to obtain estimates based on a 45-yr period of buoy data, as the buoys have not been deployed for that long. However, in order to check

[3] The study of Hogg and Swail (2002), which was concentrated in the Northern Hemisphere alone, defined a "wave year" from July to June; this definition would have the same disadvantages in the Southern Hemisphere as the calendar year has in the Northern Hemisphere.

TABLE 1. Some results of the application of the POT method to buoy (codes ending with b) and ERA-40 (codes ending with e) H_s data from 1990 to 1999. For an explanation see the text.

Buoy	n_t	u (m)	$\hat{\lambda}_u$	AD	$\hat{\alpha}$(m)	$\hat{x}_{100}$ (m)	$\hat{\alpha}$(m)	$\hat{\kappa}$	$\hat{x}_{100}$(m)
32302b	4390	3.20	18.02	*2.16*	0.58	7.58 (6.40, 8.76)	0.88	0.51 (0.10, 0.92)	4.90 (4.24, 5.56)
32302e	4390	3.02	13.99	0.28	0.35	5.58 (4.80, 6.37)	0.37	0.05 (−0.31, 0.40)	5.29 (3.26, 7.32)
51001b	9987	4.00	19.65	1.28	1.15	12.72 (11.18, 14.26)	1.47	0.28 (0.05, 0.51)	8.65 (6.91, 10.39)
51001e	9989	3.49	16.86	0.41	0.55	7.60 (6.84, 8.36)	0.61	0.11 (−0.11, 0.33)	6.63 (5.04, 8.22)
51002b	10 231	3.47	19.94	0.70	0.53	7.49 (6.80, 8.17)	0.51	−0.03 (−0.22, 0.17)	7.81 (5.21, 10.41)
51002e	10 231	2.97	15.70	0.65	0.35	5.54 (5.05, 6.02)	0.37	0.06 (−0.16, 0.27)	5.19 (3.96, 6.42)
51003b	14 492	3.37	19.97	*2.10*	0.67	8.42 (7.70, 9.15)	0.86	0.29 (0.10, 0.47)	6.02 (5.22, 6.82)
51003e	14 493	3.08	15.68	0.63	0.44	6.30 (5.79, 6.81)	0.50	0.13 (−0.06, 0.33)	5.39 (4.44, 6.35)
51004b	11 693	3.40	17.41	0.69	0.62	8.04 (7.26, 8.82)	0.70	0.13 (−0.07, 0.33)	6.76 (5.26, 8.27)
51004e	11 693	3.04	13.87	0.34	0.36	5.62 (5.14, 6.11)	0.38	0.07 (−0.15, 0.29)	5.20 (4.06, 6.34)
42001b	8520	2.23	22.22	0.54	0.81	8.45 (7.35, 9.55)	0.85	0.05 (−0.15, 0.26)	7.61 (4.64, 10.57)
42001e	8522	1.87	20.92	0.98	0.62	6.64 (5.78, 7.50)	0.71	0.14 (−0.07, 0.36)	5.19 (3.60, 6.78)
42002b	13 029	2.40	24.48	*1.77*	0.83	8.85 (7.97, 9.72)	1.03	0.24 (0.07, 0.41)	6.00 (4.87, 7.13)
42002e	13 030	1.92	23.31	0.48	0.58	6.45 (5.83, 7.07)	0.64	0.09 (−0.07, 0.25)	5.48 (4.04, 6.92)
42003b	11 454	2.27	23.66	0.64	0.84	8.81 (7.83, 9.79)	0.80	−0.05 (−0.22, 0.13)	9.75 (5.57, 13.94)
42003e	11 456	1.80	20.67	0.34	0.67	6.91 (6.12, 7.71)	0.68	0.01 (−0.17, 0.19)	6.77 (4.21, 9.33)
41001b	9744	3.90	26.71	1.04	1.19	13.30 (11.84, 14.76)	1.35	0.13 (−0.06, 0.32)	10.53 (7.63, 13.44)
41001e	9747	3.15	24.80	0.44	0.93	10.40 (9.29, 11.51)	1.00	0.08 (−0.10, 0.26)	8.98 (6.26, 11.69)
41002b	8642	3.53	21.55	1.03	1.28	13.39 (11.61, 15.17)	1.42	0.11 (−0.11, 0.32)	10.98 (7.16, 14.81)
41002e	8643	2.94	20.83	1.21	0.92	9.98 (8.72, 11.23)	1.07	0.16 (−0.05, 0.38)	7.64 (5.47, 9.80)
41006b	2805	3.27	18.40	0.86	1.31	13.10 (9.71, 16.48)	1.81	0.38 (−0.10, 0.86)	7.72 (5.01, 10.44)
41006e	2806	2.57	16.85	0.75	0.90	9.23 (6.90, 11.55)	1.16	0.29 (−0.17, 0.75)	6.09 (3.58, 8.59)
41010b	14 125	2.90	20.60	0.48	0.98	10.37 (9.30, 11.45)	0.95	−0.03 (−0.19, 0.14)	10.98 (6.88, 15.09)
41010e	14 128	2.35	18.85	0.68	0.69	7.54 (6.78, 8.30)	0.63	−0.09 (−0.26, 0.08)	9.11 (5.28, 12.94)
44004b	12 661	4.13	27.48	0.42	1.34	14.78 (13.34, 16.21)	1.43	0.07 (−0.09, 0.22)	12.95 (9.21, 16.69)
44004e	12 664	3.35	26.78	0.49	0.92	10.64 (9.70, 11.58)	0.96	0.04 (−0.11, 0.19)	9.82 (7.07, 12.56)
46001b	13 153	5.03	26.01	1.28	1.41	16.09 (14.65, 17.53)	1.67	0.19 (0.03, 0.35)	11.90 (9.62, 14.18)
46001e	13 153	4.37	22.98	1.00	1.00	12.15 (11.08, 13.21)	1.18	0.17 (0.00, 0.34)	9.43 (7.63, 11.22)
46003b	10 119	5.40	26.15	*2.18*	1.45	16.81 (15.11, 18.52)	1.75	0.20 (0.02, 0.39)	12.25 (9.71, 14.79)
46003e	10 120	4.77	24.54	0.59	1.04	12.90 (11.66, 14.13)	1.11	0.06 (−0.11, 0.24)	11.57 (8.36, 14.78)
46002b	10 107	5.03	21.48	0.21	1.36	15.48 (13.79, 17.18)	1.41	0.04 (−0.15, 0.22)	14.50 (9.58, 19.42)
46002e	10 108	4.39	17.12	*1.40*	1.06	12.26 (10.83, 13.69)	1.37	0.30 (0.06, 0.53)	8.51 (7.00, 10.03)
46005b	11 569	5.27	21.42	1.15	1.51	16.83 (15.04, 18.61)	1.84	0.22 (0.03, 0.42)	12.05 (9.59, 14.50)
46005e	11 570	4.68	19.53	*1.63*	1.09	12.92 (11.61, 14.23)	1.37	0.26 (0.05, 0.46)	9.23 (7.65, 10.81)
46006b	7307	5.43	20.29	0.49	1.44	16.39 (14.20, 18.57)	1.64	0.14 (−0.10, 0.38)	13.12 (9.02, 17.22)
46006e	7307	4.78	20.39	0.50	1.00	12.43 (10.93, 13.93)	1.10	0.10 (−0.13, 0.33)	10.66 (7.38, 13.94)
46059b	8645	4.97	23.42	0.53	1.27	14.81 (13.14, 16.48)	1.33	0.05 (−0.15, 0.24)	13.60 (8.94, 18.26)
46059e	8645	4.28	17.66	0.44	0.92	11.19 (9.85, 12.53)	0.98	0.06 (−0.17, 0.28)	10.23 (6.77, 13.70)

whether POT estimates based on larger buoy and ERA-40 datasets reveal a different relationship from that based on 10-yr datasets, we will also be looking at estimates based on 20-yr datasets.

Table 1 presents some results for the period 1990–99. The first column gives the code of the data location (see Fig. 1)—codes ending in "b" refer to buoy data and codes ending in "e" to ERA-40 data, the second column (n_t) gives the number of points in each time series, the third column gives the threshold used, the fourth column the estimates of λ,[4] the fifth column gives the values of the Anderson–Darling statistic, the sixth column gives the estimates of α in the exponential distribution, the seventh column gives the 100-yr return value (x_{100}) estimates of the exponential distribution, the eighth and ninth columns give $\hat{\alpha}$ and $\hat{\kappa}$ in the GPD, and the last column gives $\hat{x}_{100}$ of the GPD. Some of the estimates are given along with 95% confidence intervals. We note that multiplying $\hat{\lambda}$ by $n_t/1460$ (the number of years) yields the approximate number of peak excesses used to fit the distributions.

Looking at the test statistics, we see that the exponentiality of the data is rejected in only four cases for the buoy data and two cases for the ERA-40; the corresponding values of the statistics are italic in Table 1. Since the tests are being done at a 5% significance level, four (two) rejections in 19 tests is above the expected proportion of rejections. We have analyzed the cases of rejection in Table 1 and observed that in all of them the hypothesis of exponentiality was not rejected once the threshold was increased and set at the 97% quantile of the whole dataset.

Figure 2 shows the values of the Anderson–Darling statistic obtained by applying the POT method to the ERA-40 data from the period 1990–99. The regions where exponentiality is rejected at a 5% level are shaded; there are 20% of rejections. The use of different test statistics, such as the exponential-versus-GPD

[4] From now on we drop the u subscript from the parameters and their estimates.

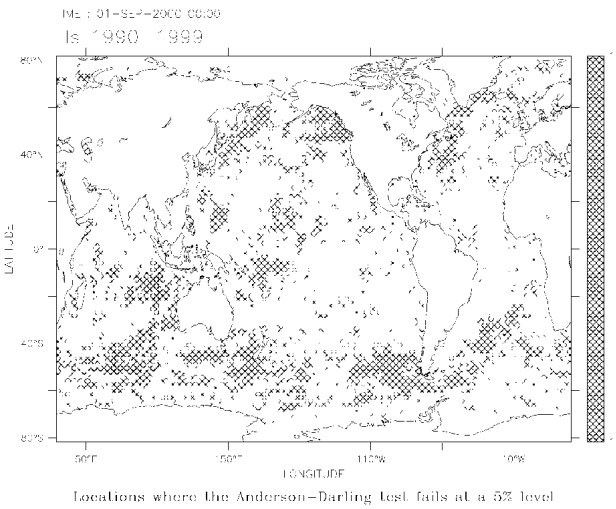

FIG. 2. The shades in the map indicate location where exponentiality is rejected at a 5% level. The rejections are based on the Anderson–Darling statistic results of the application of the POT method to H_s data from 1990 to 1999 using the 93% sample quantile as a threshold.

test of Gomes and van Montfort (1986), yields much the same results. If the threshold is fixed at the 97% quantile of the whole datasets, the percentage of rejections drops to only 10% and the estimates of 100-yr return values based on the exponentiality assumption would not change significantly, by which we mean that the confidence intervals based on the two thresholds always overlap.

Similar rejection percentages are obtained in the three other 10-yr periods considered: 17% for 1958–67; 14% for 1972–81, and 19% for 1986–95. In view of the amount and breadth of the data used, we conclude from these results that the exponential distribution is a rather good model for modeling the peak excesses of both buoy and ERA-40 H_s data and will estimate return values using the fitted exponential distributions.

Clearly, there are certain regions where the exponential assumption does not seem to apply. These occur mainly at high latitudes in the storm track regions, and in most of the cases the GPD estimate of κ is greater than zero, suggesting a type III rather than an exponential tail, and hence, that the return value estimates based on the exponential distribution are actually overestimates; compare, for instance, column 7 with column 10 of buoy location 46003 in Table 1. However, a closer examination of the data reveals that in most of these cases the 100-yr return value estimates based on the GPD are too low, implying that the GPD is really an inappropriate model for the data. For instance, the buoy measurements at location 32302 yield an x_{100} estimate of 4.9 m based on the GDP, but this value is exceeded three times in the period considered (in 1992, 1994, and 1995). A look at the kernel density estimate, which is presented in the left panel of Fig. 3 along with the fitted exponential and GPD densities, indicates that the threshold has not been taken high enough since neither the GPD nor the exponential provide a good fit of the data. That the fitted GPD model is especially unrealistic is clear from the fact that the upper limit of its support (1.73, the estimate of α/κ) is far below the upper range of the data (about 2.75).[5] The right panel of Fig. 3 shows the densities obtained when setting the threshold at the 97% quantile; both GPD and exponential provide reasonable fits to the data, but the "hump" around 1.80 m suggests the presence of two populations of extremes. In spite of the poor fit provided by any of the models when exponentiality of the data is rejected, one can say that the return value estimates based on the exponential distribution are the more realistic and on the conservative side. The exponential x_{100} estimate setting the threshold at the 97% is 6.57 m with confidence interval (5.56, 7.57).

Comparing the buoy estimates with the respective ERA-40 estimates (assuming exponentiality of the data), we see that the ERA-40 threshold and the λ and α estimates are lower than those of the buoy data; consequently, the ERA-40 return value estimates are lower.

Figure 4 compares 100-yr return value (x_{100}) estimates of the ERA-40 and the buoy data from the 1980–89 and 1990–99 periods. A striking, and for us unexpected, feature of this comparison is that the ERA-40 underestimation of x_{100} can be reliably accounted for using a linear correction. This linear association between the ERA-40 and buoy x_{100} estimates is present for all the periods considered. Moreover, the estimates of its parameters (slope and "constant term") obtained with data from different periods are compatible; that is, they are approximately the same regardless of whether we fit a line to the x_{100} estimates arising from the 1980–89 dataset, from the 1990–99 dataset, from these two datasets pooled together, or from the 1980–89 dataset. To maximize the number of data points used to estimate the linear association, we have put together the x_{100} estimates from the two decades, 1980–89 and 1990–99, which is a total of 38 data points. Since the estimates at different buoy locations are associated with different sample sizes because the availability of data varies and therefore have different confidence intervals (see columns 2 and 7 of Table 1), it would be inappropriate to fit a line by giving equal weight to each estimate. We have therefore opted for fitting a functional linear relationship in which the variance of each estimate is taken into account (Anderson 1984).

[5] Estimates obtained with the maximum likelihood and with the moment methods in this particular case yield lower estimates of the shape parameter ($\hat{\alpha} = 0.72$, $\hat{\kappa} = 0.24$ and $\hat{\alpha} = 0.79$, $\hat{\kappa} = 0.36$, respectively), and any of these yields a fit that is much more plausible from a graphical point of view, but the lack of fit of the GPD model remains. All three estimation methods provide similar estimates when the threshold is increased.

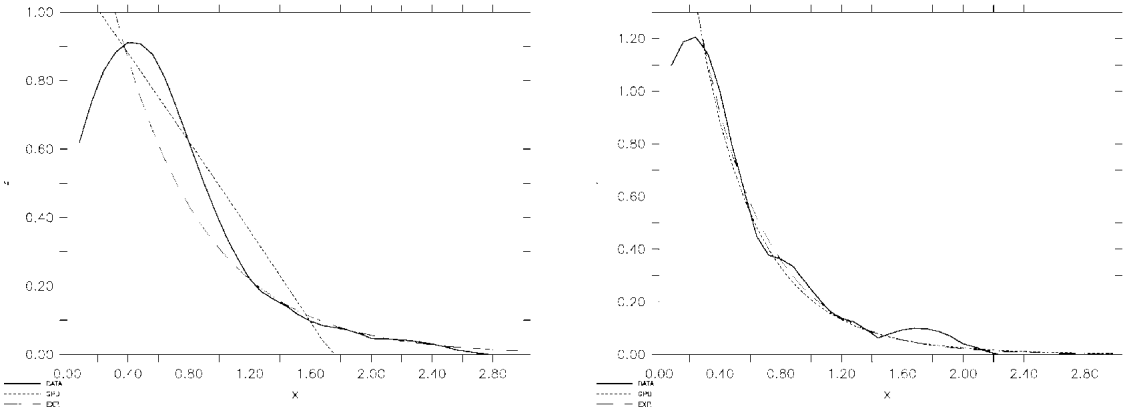

FIG. 3. Kernel density estimates (full line) and fitted exponential (dashed line) and GPD densities (dotted line) from buoy data at location 32302. (left) Threshold fixed at the 93% quantile. (right) Threshold fixed at the 97% quantile.

The following relation between buoy and ERA-40 data 100-yr return values is found:

$$X_{100}^{\text{buoy}} = 0.52 + 1.30X_{100}^{\text{ERA}-40}. \quad (3)$$

Equation (3) is plotted in Fig. 4.

The realization that the ERA-40 x_{100} estimates of H_s can be reliably calibrated is quite a fortunate one. This, however, is based on comparisons with buoy estimates that, though quite reliable, are limited to a restricted number of locations. To consolidate this linear calibration it is desirable to have an idea of how parameter estimates obtained from ERA-40 compare with those obtained from measurements on a global scale. This can only be done by resorting to altimeter data. However, as explained in section 3c(2), these comparisons can only be made in terms of the threshold u and of $\hat{\alpha}$ [two of the parameters used in the estimation of x_{100}; see Eq. (1)]. The u and α estimates were obtained from the TOPEX and collocated ERA-40 data from January 1993 to December 2001. Again, the 93% quantile of the data was used as the threshold, and only data for which the exponentiality of the data was not rejected are considered. The Anderson–Darling statistic gives only 12% of rejections, providing further evidence of the exponential character of H_s data. Figure 5 presents scatterplots of the ERA-40 estimates versus the TOPEX es-

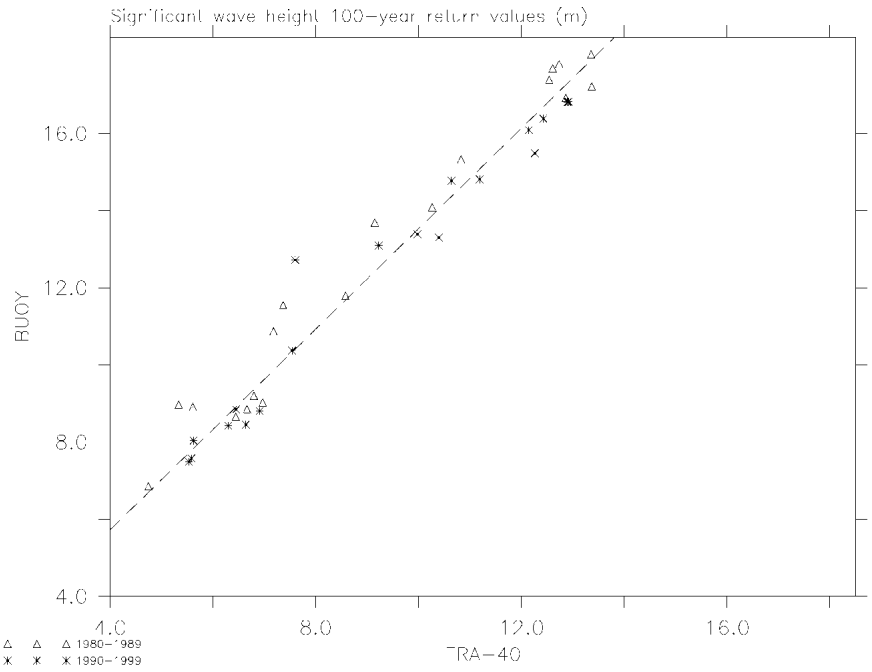

FIG. 4. Illustration of the linear relationship between the H_s 100-yr return values estimated from ERA-40 and buoy data.

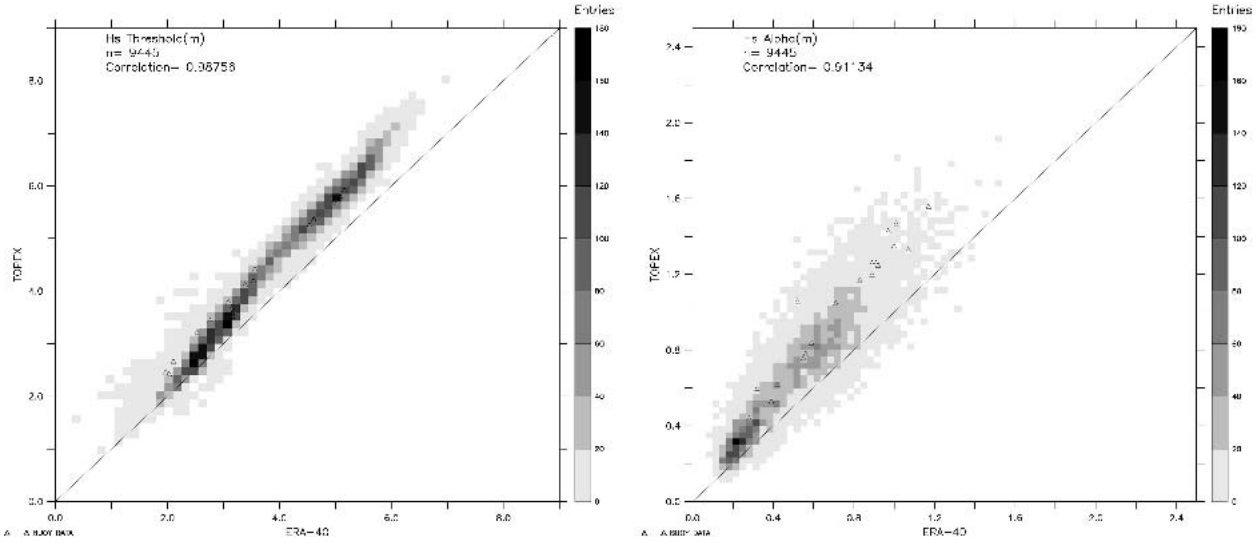

FIG. 5. Scatter diagrams of (left) u and (right) α estimates from TOPEX vs ERA-40 data with the buoy estimates vs those from ERA-40 superimposed. Estimates based on H_s data for Jan 1993–Dec 2001.

timates. To compare the relationships between ERA-40 and TOPEX with those between ERA-40 and buoy, we have computed u and $\hat{\alpha}$ for buoy and ERA-40 data from January 1993 to December 2001; the values of these are superimposed in the figure. Obviously, there is more scatter in the comparisons between ERA-40 and TOPEX estimates than in those between ERA-40 and buoy estimates, but the relationship seems to be the same in both cases. Thus, the results suggest that relation (3) can be applied globally.

It is also interesting to check whether the distribution of an observation at a randomly selected cluster time and the distribution of cluster peaks are approximately equal. This can be done using the ERA-40 data since the complete time series at each location are available. The ERA-40 estimates obtained when sampling the data in the same way as the TOPEX data and those obtained without subsampling the data (i.e., using the complete time series in each location) were compared. Figure 6 shows scatterplots comparing the thresholds and estimates of α of the exponential distribution in the two cases. The majority of the estimates of α arising from the subsampled data are less than or equal to the estimates obtained from the whole time series, the un-

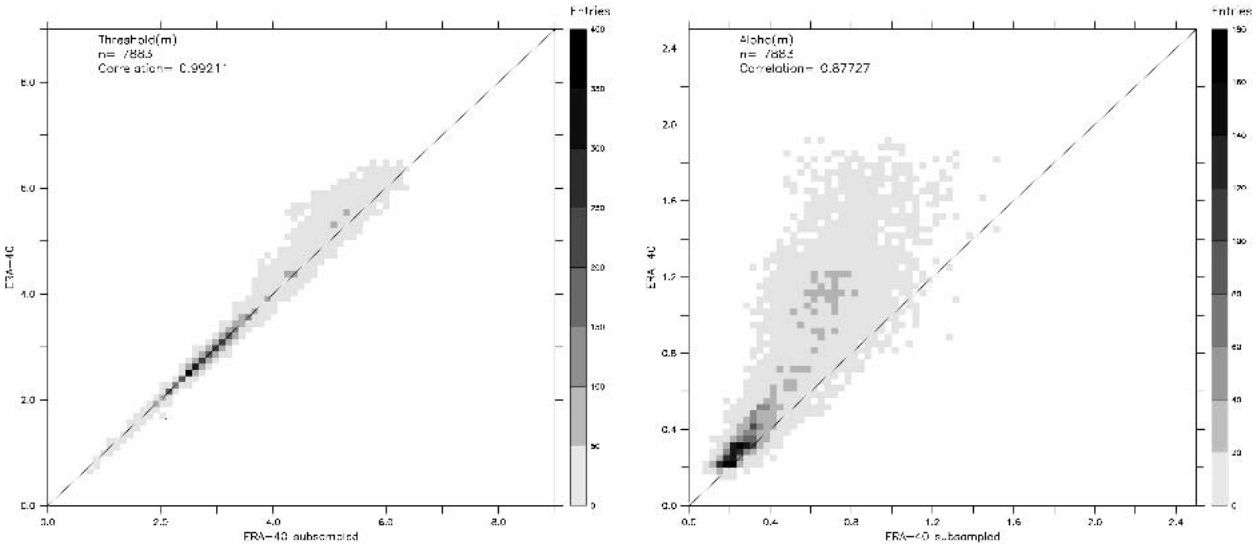

FIG. 6. Scatter diagrams of (left) u and (right) α estimates from ERA-40 vs subsampled ERA-40. Estimates based on H_s data for Jan 1993–Dec 2001.

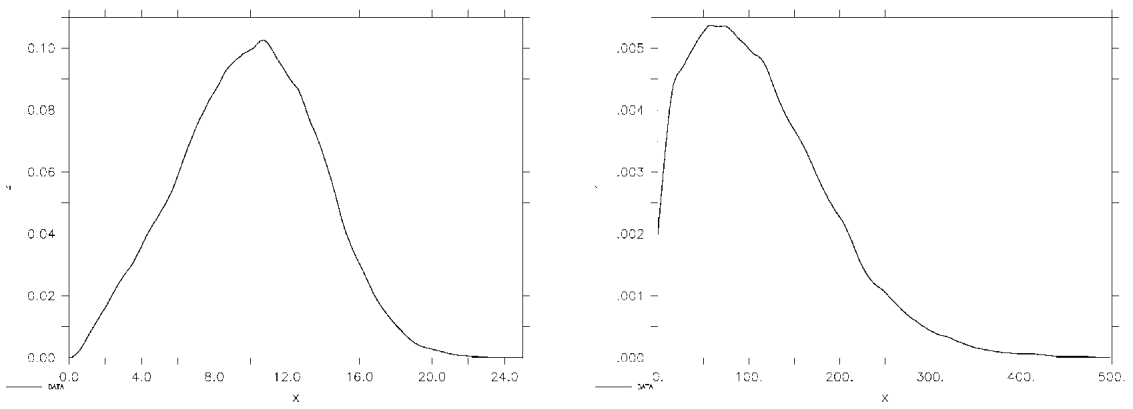

FIG. 7. Kernel density estimates of (left) U_{10} and (right) U_{10}^2 ERA-40 data at a location in the Southern Hemisphere storm track.

derestimation being higher for higher values. This does not invalidate the approximate equality of the distributions, which is an asymptotic property. It shows, however, that the presently available amount of altimeter data is not sufficient for the estimation of extreme values and accentuates the importance of datasets such as ERA-40 in obtaining such estimates. Still, the discrepancy between the estimates could probably be accounted for by estimating upper bounds of the estimates of the undersampled data as described by Robinson and Tawn (2000).

Preliminary H_s 100-yr return value estimates based on the ERA-40 data were presented in Caires and Sterl (2003a). The present estimates, however, are more reliable because in order to obtain Eq. (3) we now use more data and a functional linear relationship instead of simple linear regression. Caires and Sterl (2003a) were also less stringent in their identification of clusters belonging to the same storm, having required clusters to be 18 h apart instead of the 48-h limit used here; while this had no significant effect on the estimates of x_{100}, it did have an effect in the results of the Anderson–Darling tests.

Motivated by deficiencies of the ERA-40 H_s dataset, mainly by some overestimation of low wave heights and underestimation of high wave heights, Caires and Sterl (2005) produced a corrected version of it, the C-ERA-40 dataset, using a nonparametric correction method based on nonparametric regression techniques (e.g., Caires and Ferreira 2005). Although C-ERA-40 represents a considerable improvement of the ERA-40 H_s dataset, it still shows some underestimation of high quantiles, so its return value estimates would also require a linear correction (smaller than the present one, though); for this reason we have chosen to base our analysis on the original ERA-40 H_s data.

b. Wind speed

The same procedures used for analyzing H_s were also used to analyze the U_{10} data. One thing that became immediately clear was that the application of the POT method required a rather high threshold. For example, fixing the threshold at the 97% quantile of the whole data gives good fits in only about 60% of the cases. As in the case of H_s, the most problematic locations are those with higher mean U_{10} climate and the lack of fit is apparently due to the coexistence of two populations of extremes. Although raising the threshold would be an option, the required increase would depend very much on the location, and the corresponding sample sizes would be too small.

The strategy that we adopted was to base our analysis on U_{10}^2 rather than on U_{10}. This can be partly motivated by the fact that the wind velocities subtracted by their means are sometimes assumed normally distributed with mean 0, which implies that the wind speed has a distribution close to the Rayleigh, and hence its square has a distribution close to the exponential. (The assessment of this assumption is difficult and outside the scope of this study.) Our main argument to switch from U_{10} to U_{10}^2, however, rests on the fact Galambos (1987) that the rate of convergence at which the tail of the distribution function of the observations can be approximated by the tail of the GPD in the case of distributions such as the normal, which are rather concentrated around the mean can be very slow, demanding comparatively high thresholds/large sample sizes. Indeed, the histogram of the whole U_{10} dataset reveals a rather concentrated distribution, not dissimilar to a normal one (in contrast with that of H_s data, e.g.), which explains why the application of the POT method to U_{10} will not be very successful; see Fig. 7. The distribution of U_{10}^2, on the other hand, is much more skewed and "nonnormal" than that of U_{10}, and hence more suited to the application of the POT method. Note that this phenomenon matches what happens in theory: although the convergence required by the POT method is slow in the case of normal random variables, it will be rather fast with their squares, which are roughly chi-square variables.

TABLE 2. Some results of the application of the POT method to buoy (codes ending with b) and ERA-40 (codes ending with e) based on U_{10}^2 data from 1990 to 1999. For an explanation see the text.

Buoy	n_t	U (m s^{-1})	$\hat{\lambda}_u$	AD	$\hat{\alpha}$ (m s^{-1})$^{-2}$	$\hat{x}_{100}$ (m s^{-1})	$\hat{\alpha}$ (m s^{-1})$^{-2}$	$\hat{\kappa}$	$\hat{x}_{100}$ (m s^{-1})
32302b	4390	10.53	11.11	*1.44*	18.36	15.48 (14.05, 16.92)	29.36	0.60 (0.04, 1.16)	12.62 (11.82, 13.42)
32302e	4390	10.17	9.99	0.96	18.32	15.16 (13.65, 16.67)	27.15	0.48 (−0.06, 1.02)	12.56 (11.44, 13.68)
51001b	9987	12.04	10.93	0.47	39.40	20.51 (18.94, 22.08)	35.89	−0.09 (−0.36, 0.18)	22.22 (15.80, 28.64)
51001e	9989	11.32	9.80	0.26	31.10	18.50 (17.10, 19.90)	32.96	0.06 (−0.22, 0.34)	17.73 (14.30, 21.16)
51002b	10 231	12.65	10.85	0.59	31.07	19.42 (18.13, 20.72)	26.25	−0.16 (−0.43, 0.11)	22.17 (15.35, 29.00)
51002e	10 231	10.66	10.42	0.79	19.11	15.70 (14.72, 16.68)	16.98	−0.11 (−0.38, 0.15)	17.08 (12.72, 21.44)
51003b	14 492	10.90	12.15	0.93	29.69	18.15 (17.08, 19.23)	28.14	−0.05 (−0.26, 0.16)	18.99 (15.02, 22.96)
51003e	14 493	10.34	10.04	1.21	23.75	16.46 (15.47, 17.45)	29.38	0.24 (−0.01, 0.49)	14.37 (12.94, 15.81)
51004b	11 693	11.86	12.70	*2.29*	24.62	17.79 (16.75, 18.84)	17.56	−0.29 (−0.57, −0.01)	23.55 (13.56, 33.54)
51004e	11 693	10.99	8.49	0.32	21.09	16.22 (15.17, 17.28)	20.20	−0.04 (−0.32, 0.23)	16.69 (13.18, 20.20)
42001b	8520	11.67	16.20	0.24	45.40	21.72 (20.13, 23.30)	44.81	−0.01 (−0.25, 0.22)	21.99 (16.62, 27.36)
42001e	8522	10.76	15.26	0.86	35.63	19.41 (18.00, 20.82)	42.14	0.18 (−0.07, 0.44)	16.91 (14.38, 19.44)
42002b	13 029	12.27	18.01	0.75	45.98	22.25 (21.01, 23.49)	51.46	0.12 (−0.07, 0.31)	20.14 (17.39, 22.89)
42002e	13 030	11.09	16.22	0.88	40.57	20.57 (19.37, 21.76)	46.89	0.16 (−0.04, 0.35)	18.14 (15.79, 20.49)
42003b	11 454	12.30	15.25	0.73	64.54	24.99 (23.18, 26.79)	54.28	−0.16 (−0.38, 0.06)	30.08 (20.00, 40.16)
42003e	11 456	10.67	13.78	0.57	42.74	20.56 (19.13, 21.99)	46.30	0.08 (−0.14, 0.31)	19.11 (15.66, 22.57)
41001b	9744	15.14	18.67	0.60	69.17	27.39 (25.60, 29.18)	76.53	0.11 (−0.12, 0.33)	25.01 (20.85, 29.18)
41001e	9747	13.88	18.86	1.20	59.37	25.31 (23.75, 26.87)	74.48	0.25 (0.03, 0.48)	21.03 (18.74, 23.33)
41002b	8642	13.96	16.67	0.23	67.26	26.34 (24.38, 28.31)	64.00	−0.05 (−0.28, 0.19)	27.68 (20.13, 35.24)
41002e	8643	13.58	17.28	0.61	54.68	24.33 (22.69, 25.97)	59.61	0.09 (−0.14, 0.32)	22.54 (18.57, 26.52)
41006b	2805	14.18	14.48	0.29	63.10	25.07 (22.23, 29.16)	58.96	−0.07 (−0.51, 0.38)	27.41 (13.60, 41.21)
41006e	2806	12.31	13.63	0.63	55.08	23.43 (20.14, 26.72)	70.87	0.29 (−0.22, 0.79)	19.17 (14.80, 23.54)
41010b	14 124	13.08	15.58	0.37	55.29	24.03 (22.63, 25.44)	56.81	0.03 (−0.16, 0.22)	23.44 (19.30, 27.58)
41010e	14 128	11.70	14.76	0.15	46.64	21.84 (20.56, 23.13)	45.22	−0.03 (−0.22, 0.16)	22.50 (17.95, 27.06)
44004b	12 661	15.47	17.75	0.48	73.25	28.06 (26.39, 29.73)	75.35	0.03 (−0.17, 0.23)	27.33 (22.34, 32.31)
44004e	12 664	14.66	19.19	0.45	60.84	25.98 (24.62, 27.33)	66.28	0.09 (−0.09, 0.27)	24.07 (20.73, 27.40)
46001b	13 153	15.69	18.00	*1.48*	69.46	27.69 (26.09, 29.29)	88.01	0.27 (0.05, 0.49)	23.05 (20.80, 25.29)
46001e	13 153	15.08	16.99	1.33	64.62	26.61 (25.16, 28.05)	80.81	0.25 (0.05, 0.45)	22.36 (20.24, 24.48)
46003b	10 119	15.58	18.42	0.63	75.94	28.52 (26.58, 30.46)	89.12	0.17 (−0.06, 0.41)	24.84 (21.19, 28.48)
46003e	10 120	15.68	17.87	0.78	58.68	26.18 (24.69, 27.67)	70.63	0.20 (−0.02, 0.42)	22.74 (20.24, 25.25)
46002b	10 107	14.18	16.34	0.65	73.40	27.28 (25.39, 29.17)	82.99	0.13 (−0.10, 0.36)	24.39 (20.37, 28.40)
46002e	10 108	14.11	16.38	1.10	63.86	25.92 (24.23, 27.61)	67.78	0.06 (−0.15, 0.28)	24.54 (20.07, 29.01)
46005b	11 569	14.77	20.86	0.59	68.52	27.24 (25.57, 28.90)	71.82	0.05 (−0.15, 0.25)	26.03 (21.25, 30.80)
46005e	11 570	14.98	17.16	*1.75*	62.31	26.24 (24.74, 27.74)	80.43	0.29 (0.07, 0.51)	21.66 (19.70, 23.62)
46006b	7307	15.28	15.34	0.63	61.15	26.12 (24.02, 28.21)	69.90	0.14 (−0.15, 0.44)	23.47 (19.27, 27.67)
46006e	7307	14.87	16.79	0.69	55.70	25.20 (23.42, 26.97)	64.65	0.16 (−0.10, 0.42)	22.39 (19.00, 25.79)
46059b	8645	13.89	18.37	0.36	50.74	23.96 (22.37, 25.55)	53.29	0.05 (−0.18, 0.28)	22.94 (18.52, 27.37)
46059e	8645	13.57	15.40	0.28	51.27	23.67 (22.03, 25.32)	55.85	0.09 (−0.15, 0.33)	22.02 (18.08, 25.96)

Cook (1982) also advocates using U_{10}^2 instead of U_{10}, not only because of the higher rate of convergence of the former, but also because in many cases engineers are more interested in dynamic pressure, which is proportional to U_{10}^2.

Fixing the threshold at the 97% quantile of the data we have applied the POT method to buoy and ERA-40 U_{10} data from the periods 1980–89, 1990–99, and 1980–99. Table 2 presents some results for the period 1990–99. The information in the table is organized in the same way as in Table 1. Although the estimates are based on U_{10}^2, the values of u, the estimates of x_{100}, and the confidence intervals associated with the latter are, for U_{10}, given in meters per second. It is straightforward to convert quantities pertaining to U_{10}^2 into quantities pertaining to U_{10} (simply take the square root). To convert the variance of an estimate based on U_{10}^2 into the variance of an estimate related to U_{10}, we compute the approximate variance of $\hat{x}_{100}$ using the relation $\mathrm{var}(\hat{x}_{100}) \approx \mathrm{var}(\hat{x}_{100}^2)/(4x_{100}^2)$. No conversion was applied to the estimates of α since they have no interpretation in terms of U_{10}.

The exponentiality of the data is rejected in three cases for the buoy data and in one case for the ERA-40 data; the corresponding values of the Anderson–Darling statistic are italic in Table 2. As in the case of H_s data the rejection rate (at least in the case of buoy observations) is above the expected proportion of rejections. In the application of the POT method to the ERA-40 data from the four 10-yr periods of 1958–67, 1972–81, 1986–95, and 1990–99, the Anderson–Darling statistic gives about 20% of rejections in each of the periods. Just as in the case of H_s data, the exponentiality of U_{10}^2 data is rejected in the regions with higher U_{10} (Southern and Northern Hemisphere storm tracks). Moreover, in most of the cases where exponentiality is rejected the estimates of κ in the GPD are above zero (with confidence intervals in most cases not including zero, as in the cases in Table 2). To conclude, the amount of rejections is small and seems to occur due to a wrong choice of threshold combined with the possible coexistence of two populations of extremes in those locations. In the estimates presented from now on we will therefore assume exponentiality of the U_{10}^2 data,

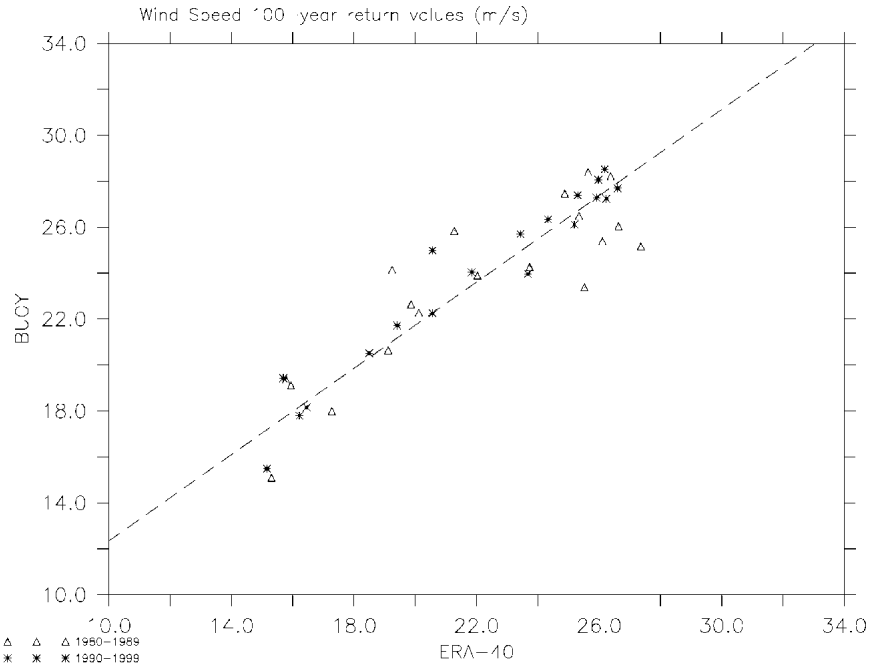

FIG. 8. Illustration of the linear relationship between the U_{10} 100-yr return values estimated from ERA-40 and buoy U_{10} data.

since this assumption seems to apply to the majority of the data, with the caveat that our estimates at locations of high U_{10} may be conservative.

The return value estimates computed from the buoy are in most cases higher than those computed from ERA-40, which is consistent with the underestimation of high values of U_{10} by ERA-40 reported by Caires and Sterl (2003b). Figure 8 compares the two types of estimates obtained with the data from the 1980–89 and 1990–99 periods. Fitting a functional linear relationship to these estimates yields

$$x_{100}^{\text{buoy}} = 2.94 + 0.94 x_{100}^{\text{ERA}-40}. \quad (4)$$

This straight line is represented in Fig. 8 together with the pairs of estimated return values. As with H_s data, it is clear that the return values computed from ERA-40 data can be reliably calibrated using Eq. (4). The main difference relative to the straight line for H_s is that the scatter around Eq. (4) is somewhat greater, and not all ERA-40 estimates of x_{100} for U_{10} are underestimates—in some locations, the correspondence between the ERA-40 and buoy estimates is quite good.

This last observation suggests that the quality of the ERA-40 estimates of U_{10} x_{100} depends on the buoy location. An explanation for this dependence may lie in the data assimilated into ERA-40. ERA-40 benefited from the assimilation of some NOAA/NDBC buoy wind speed measurements present in the Comprehensive Ocean–Atmosphere Data Set (COADS; Woodruff et al. 1998). Although it is difficult to pinpoint the precise buoy measurements that were, if at all, used in the creation of the ERA-40 data, it is likely that the locations at which the ERA-40 estimates compare very well with the buoy estimates are those at which the original buoy data were used in the creation of ERA-40.[6]

To investigate whether the linear correction given by Eq. (4) can be reliably applied to *global* estimates based on the ERA-40 data, we resort to comparing the POT estimates obtained with ERA-40 with those obtained with altimeter data. The POT method was applied to TOPEX and collocated ERA-40 data from January 1993 to December 2001, using again the 97% quantile of the data as the threshold, and values of u and estimates of α were obtained. Figure 9 presents scatterplots of the ERA-40 estimates versus the TOPEX estimates, with the corresponding buoy and ERA-40 pairs of estimates superimposed. The plots only present data for which exponentiality was not rejected for both the altimeter and ERA-40 data, which make up 87% of the data. It is clear from the plots that the pairs of buoy and ERA-40 estimates lay within the scatter of the pairs of TOPEX and ERA-40 estimates, so (4) does apply globally. This assessment justifies the application of (4) to obtain the global ERA-40 U_{10} 100-yr return value estimates used in the sequel.

[6] It should be noted that no H_s buoy measurements were used in the production of the ERA-40 data and therefore this problem did not arise in the comparison of ERA-40 and buoy H_s data.

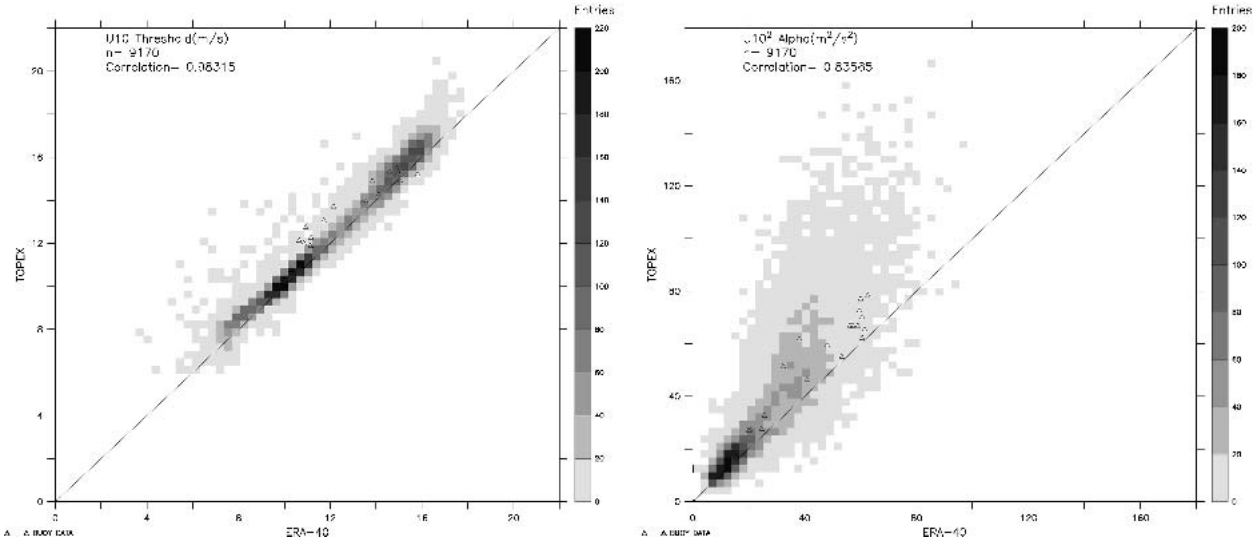

FIG. 9. Scatter diagrams of (left) u and (right) α estimates from TOPEX vs ERA-40 data, with the buoy estimates vs those from ERA-40 superimposed. Estimates based on U_{10} data for Jan 1993–Dec 2001.

5. ERA-40 estimates

a. Significant wave height

Figure 10 presents global maps of the 100-yr return values of H_s computed using different 10-yr periods as well as the whole dataset, assuming the exponentiality of the peak exceedences and corrected by Eq. (3). The storm tracks of the Southern and Northern Hemispheres can be easily identified; the highest return value estimates from all the decadal datasets occur in those regions. The width of the 95% confidence intervals of the estimates is about 20% of the estimate in those coming from 10-yr periods and about 10% when considering the whole dataset.

Statistically significant differences between the return values estimated from the three different decades occur only in a small number of regions: in the North Atlantic (an increase in the region around 51°–56°N, 20°W and a decrease in the region around 42°N, 28°W when comparing the estimates with data from 1986 to 1995 relative to those obtained with data from 1972 to 1981) and North Pacific (an increase in the region around 40°N, 150°E–180° when comparing the estimates with data from 1972 to 1981 with those obtained with data from 1958 to 1967) storm tracks and the western tropical Pacific. The changes in the storm tracks mirror the decadal variability in the Northern Hemisphere.

Several earlier studies have found high correlations between the North Atlantic wave pattern and the North Atlantic Oscillation (NAO) index[7]; see, for example, Lozano and Swail (2002). More precisely, the North Atlantic storm track varies according to the NAO index. During periods when the NAO index is positive the storms tend to move from North America in the direction of the Norwegian Sea. On the other hand, when the NAO index is negative the storms move in the direction of the Mediterranean Sea (see Rogers 1997; Fig. 3), and the wave conditions are milder (see Wang and Swail 2001). From the beginning of the 1940s to the beginning of the 1970s the NAO index exhibited a downward trend, the index being negative from 1958 to 1967 (see, e.g., Lozano and Swail 2002; Fig. 2). From the beginning of the 1970s the trend was positive, the period between 1972 and 1981 being characterized by both positive and negative NAO index years. From 1986 to 1995 the index was always positive. The change in the pattern and intensity of the 100-yr return values in the North Atlantic basin is completely in line with these decadal changes of the NAO index. The higher estimates from the period 1958–67 are lower and located to the south of those of the later periods. The pattern of the estimates for the period 1972–82 is characterized by two high lobes, one due to the positive NAO index years and another due to the negative NAO index years. The highest estimates, which are also those with the most northerly peak, are from the period 1986–95, the period during which the NAO index was at its highest.

The plots in Fig. 10 show a clear and strong increasing trend in the estimates of the 100-yr return values from the three decadal periods in the North Pacific storm track, especially from the first to the second period. This is in line with the results of Graham and Diaz (2001). There is some discussion about the reasons for this increase. Graham and Diaz suggest the increasing

[7] A measure of the difference of the sea surface pressure between Reykjavik (Iceland) and Ponta Delgada (Portugal).

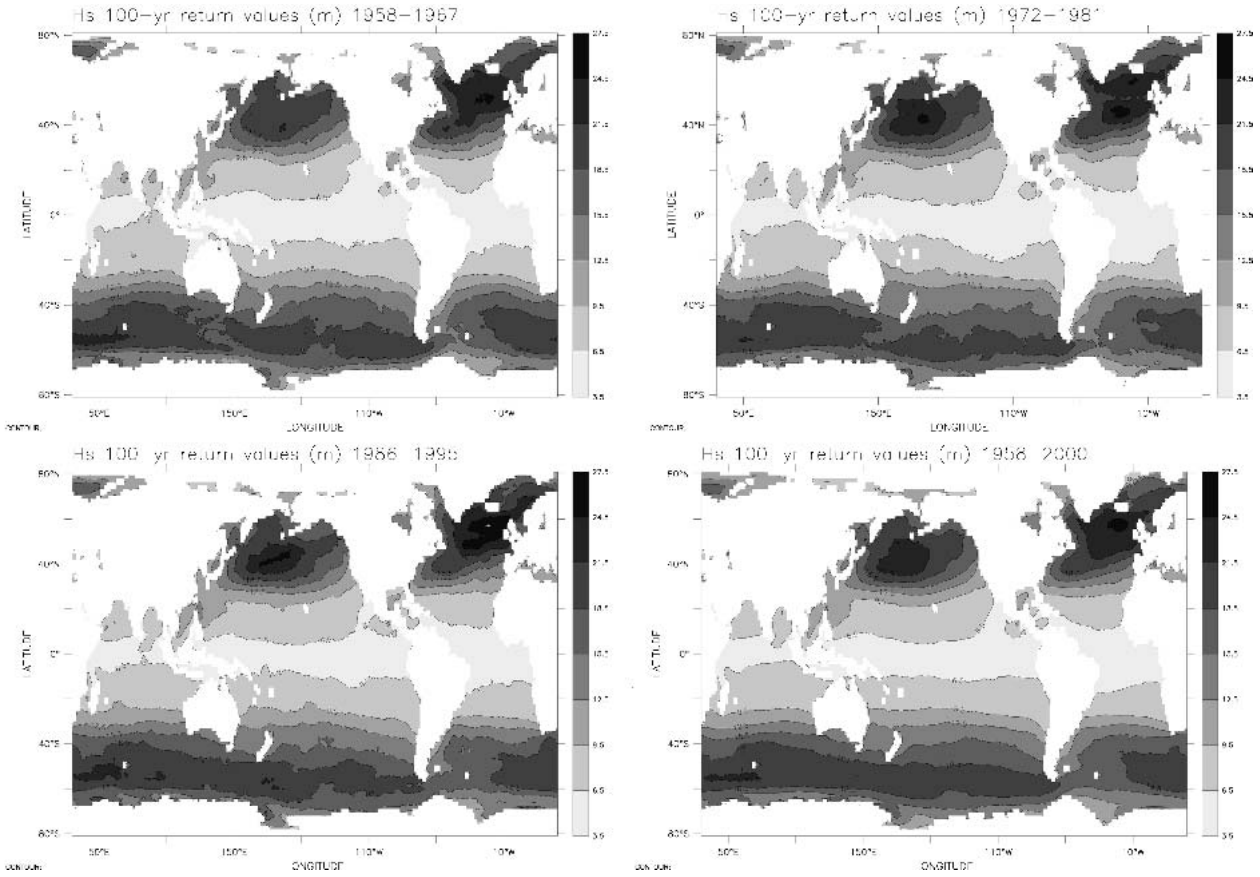

FIG. 10. Corrected 100-yr return value estimates of H_s based on ERA-40 data from three different 10-yr periods and the whole ERA-40 period as indicated.

sea surface temperatures in the western tropical Pacific (region between 20°S and 5°N, 155°E and 180°) as a plausible cause. However, further research is still needed to determine the exact causes.

In the region between 20°S and 5°N, 155°E and 180° there is a statistically significant increase in the return value estimates obtained with data from the period 1972–81 relative to those obtained with data from 1958–67. This is the same region where Graham and Diaz (2001) report an increase in the sea surface temperatures and, as we will see in the next section, it is due to an increase in wind speed from one period to the next.

Figure 10 also presents H_s x_{100} estimates based on the whole ERA-40 data from 1958 to 2001. These estimates are also not compatible, in the sense that the corresponding confidence intervals do not intersect, with those from the different decadal periods in the regions mentioned above, especially when compared with the estimates from the first period. In accordance with the estimates obtained from the different periods, this plot shows that the most extreme wave conditions are clearly in the storm track regions and that the highest return values occur in the North Atlantic. The latter fact may be surprising since some readers might expect the highest return value estimates to be in the Southern Hemisphere storm track region, where average conditions are higher. The explanation for this apparent contradiction is that the variance of the data determines to a certain extent the character of extremes and, even though waves in the Southern Hemisphere storm track region are usually higher, they have smaller standard deviations than those in the North Atlantic storm track.

b. Wind speed

Figure 11 presents global maps of 100-yr return value estimates of U_{10} computed using data from the periods 1957–67, 1972–81, 1986–95, and 1958–2001, assuming an exponential tail for U_{10}^2 and calibrating the estimates using Eq. (4). The width of the 95% confidence intervals of the estimates is about 20% of the estimate in those coming from 10-yr periods and about 10% when considering the whole dataset. The spatial pattern of the x_{100} estimates is quite similar to that of the H_s. The highest values are found in the storm tracks and the lowest in the Tropics. Again, the highest values are in the North Atlantic. There are some high x_{100} estimates in the regions close the ice boundaries; these are probably spurious results due to the variability of the ice coverage in those regions. The highest U_{10} x_{100} esti-

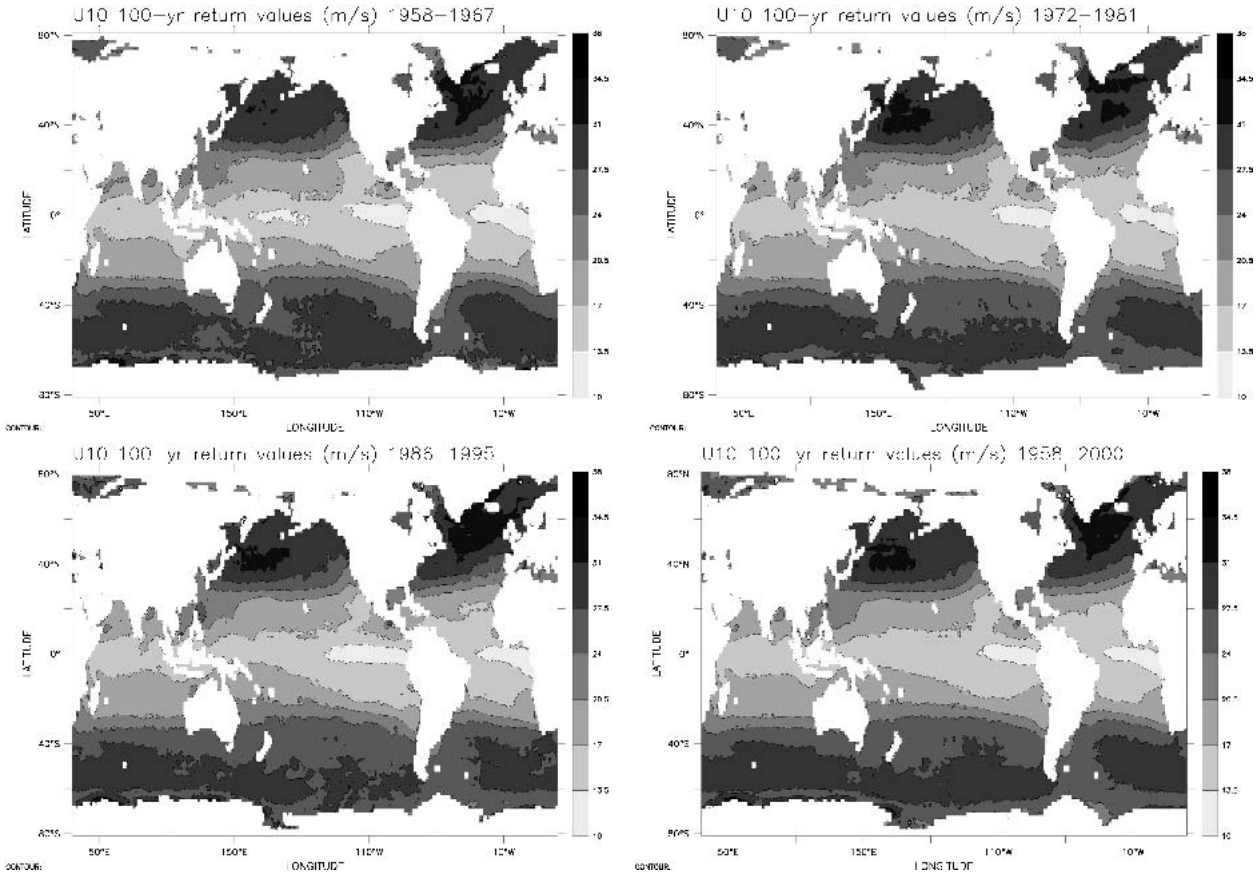

FIG. 11. Corrected 100-yr return value estimates of U_{10} based on ERA-40 data from three different 10-yr periods and the whole ERA-40 period as indicated.

mates are upwind of the corresponding H_s x_{100} estimates. The x_{100} estimates from the various 10-yr periods are significantly different in the same regions where significant differences were obtained for the return value estimates of H_s. The variation of x_{100} estimates of U_{10} is consistent with the explanations we gave for the variation of x_{100} estimates of H_s. In the region between 20°S and 5°N, 155°E and 180° there is a significant increase in the U_{10} x_{100} estimates from data from the period 1972–81, relative to those from data from 1958–67. This can be clearly seen comparing the top-left panel with the top-right panel of Fig. 11. The tongue of low winds in the western tropical Pacific has clearly shrunk. We have looked at the ERA-40 data in this area and there is a clear increase in the mean U_{10} over this area from 1958 to 1972 and we have found similar increases in other datasets. The reason for this increase is, however, unknown to us.

6. Discussion and conclusions

We have presented global 100-yr return value estimates of H_s and U_{10} based on the ERA-40 dataset. These estimates have been directly assessed against estimates from buoy measurements and indirectly against estimates from TOPEX altimeter measurements. We have also looked at the effects of climate variability on the estimates.

Based on the application of the POT method to buoy, TOPEX, and ERA-40 H_s data we have concluded that in the majority of cases the tail of H_s data is exponential. Using this as a model, we have then obtained global x_{100} estimates from the ERA-40 data. Since ERA-40 data underestimates the high values of H_s, it is necessary to apply a correction based on a linear relationship to the return value estimates. Although the determination of this correction was based only on buoy x_{100} estimates, the estimates and thresholds obtained from the POT analysis of TOPEX altimeter data support the validity of the proposed relationship. Corrected global estimates based on three different 10-yr periods of ERA-40 data and on the whole dataset were given, and it was shown that estimates obtained from the different periods differ in the Northern Hemisphere storm tracks and in the western tropical Pacific. The differences in the storm tracks can be attributed to the decadal variability in the Northern Hemisphere, and we have linked them to changes in the global circulation patterns, most

notably to the NAO index. As expected, the x_{100} estimates of H_s are higher in the storm tracks and lower in the Tropics. The most extreme wave conditions are expected to occur in the North Atlantic.

We have tried to analyze the ERA-40 U_{10} data using the same procedures applied to the H_s data. However, we concluded that, due to the distributional characteristics of U_{10} data, return value estimates should be based on U_{10}^2 rather than on U_{10} data. In most cases, the exceedances of U_{10}^2 data were found to be approximately exponentially distributed. Systematic differences were found between the x_{100} estimates of U_{10} from buoy and ERA-40 data, and the latter were also corrected through a linear relationship. Estimates based on TOPEX data indicate [from what was said in the paragraph following Eq. (3)] that this linear correction is valid globally. There were significant differences between the U_{10} return value estimates arising from the three 10-yr periods of data, and these differences are consistent with those detected in the case of H_s.

About 80% of data comply with the exponentiality assumption. The 20% of the cases in which exponentiality does not apply occur mainly in locations of moderate to high wind and wave conditions. In those locations the POT method yields tails that are lighter than exponential, but these are a consequence of poor fits, and not a sign that a light-tailed GPD is an appropriate model for the peak excesses. Our analysis indicates that the required threshold for a valid POT analysis in these regions is too high for the amount of data available and that there may actually be two coexisting extreme populations.

The establishment of the exponential distribution as the approximate distribution of the peak excesses of H_s and U_{10} data may be useful in a number of problems. For instance, it could be used to assess the capacity of numerical models to reproduce extreme winds and waves. The reader, however, should be aware that there are people in the extremes community that (in certain situations) advocate fitting the GPD rather than the exponential even when the exponential seems a plausible model (see Coles and Pericchi 2003).

There are some caveats about the return value estimates we present here:

- The assumption of exponentiality of the data does not apply in some locations. In such cases the estimates tend to be conservative.
- Choosing the threshold at a fixed percentile of the data is a somewhat crude method. A case-by-case analysis would improve the fits somewhat, but with this amount of data this is an expedient method that still gives rather good results.
- Due to resolution, tropical cyclones are not resolved by the ERA-40 system. Therefore, the estimated return values in the regions of tropical storms may be underestimated.
- The ERA-40 model does not account for shallow water effects, and therefore the estimates may not be valid in coastal regions.
- The estimates given here are based on data averaged on a $1.5° \times 1.5°$ region or equivalently a period of about 3 h; these values can be exceeded at short time/space scales.

This research shows once more that the estimation of extremes is not a straightforward business and that more work on general as well as more specific techniques needs to be done:

- The choice of the appropriate threshold is a well-known problem that needs to be addressed from both mathematical and modeling points of view.
- Although a 45-yr global dataset is definitely a considerable amount of data, still more data are needed to answer questions about the possible existence of different extreme populations in regions with high mean conditions and to be able to model climate variability.
- Reliable ways to obtain estimates from altimeter data need to be devised—still in the framework of the POT method since the amount of data will continue to be too small for the annual maxima method. The estimates obtained here from the TOPEX data suggest that the amount of data is not enough for asymptotic assumptions such as the equality of distributions of randomly sampled excesses and peak excesses to hold in a satisfactory way: But how much more data are needed? Would it be enough to pool altimeter measurements from different satellite missions? There would still be the problem of how to go from the GPD parameter estimates to the 100-yr return value estimates. One way would be to obtain estimates of λ from the reanalysis, or other data, and combine these with the altimeter GPD parameter estimates to obtain x_{100} estimates. Another would be to define a point process, as done by Anderson et al. (2001), and obtain estimates of mean storm duration by modeling the variability of H_s in each region using the method suggested by Baxevani et al. (2004).

Acknowledgments. We thank our data sources, namely the ERA-40 team for the ERA-40 data, Helen Snaith of SOC for the altimeter data, and NOAA/NDBC for the buoy data. We are indebted to José Ferreira, Gerrit Burgers, Adri Buishand, and Val Swail for fruitful discussions and comments, and to Camiel Severijns for the Field library and technical support. We also thank INTAS (Grant 96-2089) for facilitating discussion with Sergey Gulev, David Woolf, and Roman Bortkovsky. This work was sponsored by the EU-funded ERA-40 Project (EVK2-CT-1999-00027).

REFERENCES

Anderson, C. W., D. J. T. Carter, and P. D. Cotton, 2001: Wave climate variability and impact on offshore design extremes.

Shell International and the Organization of Oil and Gas Producers Rep., 99 pp.
Anderson, T. W., 1984: Estimating linear statistical relationships. *Ann. Stat.,* **12,** 1–45.
Baxevani, A., I. Rychlik, and R. J. Wilson, 2004: Modelling space variability of H_s in the North Atlantic. *Extremes*, in press.
Bidlot, J.-R., D. J. Holmes, P. A. Wittmann, R. Lalbeharry, and H. S. Chen, 2002: Intercomparison of the performance of operational wave forecasting systems with buoy data. *Wea. Forecasting,* **17,** 287–310.
Caires, S., and A. Sterl, 2003a: On the estimation of return values of significant wave height data from the reanalysis of the European Centre for Medium-Range Weather Forecasts. *Safety and Reliability: Proceedings of the European Safety and Reliability Conference,* T. Bedford and P. van Gelder, Eds., Lisse, Swets & Zeitlinger, 353–361.
——, and ——, 2003b: Validation of ocean wind and wave data using triple collocation. *J. Geophys. Res.,* **108,** 3098, doi:10.1029/2002JC001491.
——, and J. A. Ferreira, 2005: On the nonparametric prediction of conditionally stationary sequences. *Stat. Inf. Stochastic Processes,* **8,** 151–184.
——, and A. Sterl, 2005: A new nonparametric method to correct model data: Application to significant wave height from the ERA-40 reanalysis. *J. Atmos. Oceanic Technol.,* in press.
——, A. Sterl, J.-R. Bidlot, N. Graham, and V. Swail, 2004: Intercomparison of different wind wave reanalyses. *J. Climate,* **17,** 1893–1913.
Challenor, P., and D. Cotton, cited 1999: Trends in TOPEX significant wave height measurement. [Available online at http://www.soc.soton.ac.uk/JRD/SAT/TOPtren/TOPtren.pdf.]
Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values.* Springer-Verlag, 208 pp.
——, and L. Pericchi, 2003: Anticipating catastrophes through extreme value modelling. *Appl. Stat.,* **53,** 405–416.
Cook, N. J., 1982: Towards better estimates of extreme winds. *J. Wind Eng. Ind. Aerodyn.,* **9,** 295–323.
Davison, A. C., and R. L. Smith, 1990: Models for exceedances over high thresholds (with discussion). *J. Roy. Stat. Soc.,* **52B,** 393–442.
Ferguson, T. S., 1996: *A Course in Large Sample Theory.* Chapman and Hall, 245 pp.
Ferreira, J. A., and C. Guedes Soares, 1998: An application of the peaks over threshold method to predict extremes of significant wave height. *J. Offshore Mech. Arct. Eng.,* **120,** 165–176.
——, and ——, 2000: Modelling distributions of significant wave height. *Coast. Eng.,* **40,** 361–374.
Galambos, J., 1987: *The Asymptotic Theory of Extreme Order Statistics.* 2d ed. Krieger, 414 pp.
Gomes, M. Y., and M. A. J. van Montfort, 1986: Exponentiality versus generalized pareto, quick tests. *Third Int. Conf. on Statistical Climatology,* Vienna, Austria.
Gourrion, J., D. Vandemark, S. Bailey, B. Chapron, C. P. Gommenginger, P. G. Challenor, and M. A. Srokosz, 2002: A two parameter wind speed algorithm for Ku-band altimeters. *J. Atmos. Oceanic Technol.,* **19,** 2030–2048.
Graham, N. E., and H. F. Diaz, 2001: Evidence of intensification of North Pacific winter cyclones since 1948. *Bull. Amer. Meteor. Soc.,* **82,** 1869–1893.
Hogg, W. D., and V. R. Swail, 2002: Effects of distributions and fitting techniques on extreme value analysis of modeled wave heights. *Proc. Seventh Int. Workshop on Wave Hindcasting and Forecasting,* Banff, AB, Canada, U.S. Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory, the Fleet Numerical Meteorology and Oceanography Center, and the Meteorological Service of Canada, 140–150.
Hosking, J. R. M., and J. R. Wallis, 1987: Parameter and quantile estimation for the Generalized Pareto Distribution. *Technometrics,* **29,** 339–349.
Janssen, P. A. E. M., J. D. Doyle, J. Bidlot, B. Hansen, L. Isaksen, and P. Viterbo, 2002: Impact and feedback of ocean waves on the atmosphere. *Advances in Fluid Mechanics*, W. A. Perrie, Ed., Vol. 1, Kluwer, 155–197.
Leadbetter, M. R., 1991: On a basis for "peaks over threshold" modeling. *Stat. Prob. Lett.,* **12,** 357–362.
Lozano, I., and V. Swail, 2002: The link between wave height variability in the North Atlantic and the storm track activity in the last four decades. *Atmos.–Ocean,* **40,** 377–388.
Robinson, M. E., and J. A. Tawn, 2000: Extremal analysis of processes sampled at different frequencies. *J. Roy. Stat. Soc.,* **62B,** 117–135.
Rogers, J. C., 1997: North Atlantic storm track variability and its association to the North Atlantic oscillation and climate variability of northern Europe. *J. Climate,* **10,** 1635–1647.
Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 175 pp.
Simiu, E., N. A. Heckert, J. J. Filliben, and S. K. Johnson, 2001: Extreme wind load estimates based on Gumbel distribution of dynamic pressures and assessment. *Struct. Saf.,* **23,** 221–229.
Simmons, A. J., 2001: Development of the ERA-40 data assimilation system. *Proc. ECMWF Workshop on Re-analysis,* Reading, United Kingdom, ECMWF, ERA-40 Project Report Series 3, 11–30.
Snaith, H. M., 2000: Global Altimeter Processing Scheme user manual: v1. Southampton Oceanography Centre Tech. Rep. 53, 44 pp.
Stephens, M. A., 1974: EDF statistics for goodness of fit and some comparisons. *J. Amer. Stat. Assoc.,* **69,** 730–737.
Wang, X. L., and V. R. Swail, 2001: Changes of extreme wave heights in Northern Hemisphere oceans and related atmospheric circulation regimes. *J. Climate,* **14,** 2204–2221.
Witter, D. L., and D. B. Chelton, 1991: A Geosat altimeter wind speed algorithm and a method for wind speed algorithm development. *J. Geophys. Res.,* **96,** 18 853–18 860.
Woodruff, S. D., H. F. Diaz, J. D. Elms, and S. J. Worley, 1998: COADS release 2 data and metadata enhancements for improvements of marine surface flux fields. *Phys. Chem. Earth,* **23,** 517–527.
Young, I. R., 1993: An estimate of the Geosat altimeter wind speed algorithm at high wind speeds. *J. Geophys. Res.,* **98,** 20 275–20 285.