INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **25**: 963–977 (2005)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/joc.1175

CLIMATOLOGY, VARIABILITY AND EXTREMA OF OCEAN WAVES: THE WEB-BASED KNMI/ERA-40 WAVE ATLAS

ANDREAS STERL[a,*] and SOFIA CAIRES[a,b]
[a] *Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands*
[b] *Meteorological Service of Canada, Climate Research Branch, Downsview, Ontario, Canada*

Received 9 April 2004
Revised 17 August 2004
Accepted 18 August 2004

ABSTRACT

The European Centre for Medium-Range Weather Forecasts (ECMWF) has recently finished ERA-40, a reanalysis covering the period September 1957 to August 2002. One of the products of ERA-40 consists of six-hourly global fields of wave parameters, like significant wave height and wave period. These data have been generated with the centre's WAM wave model. From these results we have derived climatologies of important wave parameters, including significant wave height, mean wave period, and extreme significant wave heights. Particular emphasis is on the variability of these parameters, both in space and time. Besides being important for scientists studying climate change, these results are also important for engineers who have to design maritime constructions. This paper describes the ERA-40 data and gives an overview of the results derived. The results are available on a global $1.5^\circ \times 1.5^\circ$ grid. They are accessible from the Web-based KNMI/ERA-40 wave atlas at http://www.knmi.nl/waveatlas. Copyright © 2005 Royal Meteorological Society.

KEY WORDS: waves; climatology; extremes; variability; reanalysis

1. INTRODUCTION

The European Centre for Medium-Range Weather Forecasts (ECMWF) has recently completed the computations of the ERA-40 dataset, a reanalysis of global meteorological variables, among which are ocean surface wind waves, from September 1957 to August 2002 (45 years). The reanalysis was produced by ECMWF's Integrated Forecasting System (IFS) that uses variational data assimilation. In terms of sea-state data, this reanalysis is the first in which an ocean wind wave model is coupled to the atmosphere. Moreover, its final product consists of the longest and most complete global wave dataset available. It is given on a $1.5^\circ \times 1.5^\circ$ latitude–longitude grid covering the whole globe. The continuous 45 year length of the ERA-40 datasets makes it especially suitable for studying climate variability and for estimating extreme values of certain wave parameters, e.g. the 100 year return wave height.

As part of the ERA-40 project, we have extensively assessed the quality of the wave-related parameters and used them to build the Web-based KNMI/ERA-40 wave atlas describing the global wave climate. This paper gives an overview of the verification work done and highlights the main features of the atlas.

The atlas contains some explanatory text, a basic description of the wind and wave climates in terms of means and variability, and wave statistics that are important in ocean engineering and naval architecture. The objective is twofold. On the one hand, the atlas aims at providing a global description of the ocean climate by means of simple statistical measures. On the other hand, it aims at revealing the existence of decadal variability in the wave climate and showing the extent to which this variability affects the estimates

* Correspondence to: Andreas Sterl, Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands;
e-mail: sterl@knmi.nl

of parameters such as the '100 year return significant wave height'. This is the significant wave height that, on average, is exceeded only once every 100 years. It is used in the design of ships and of coastal and offshore structures. The information on decadal variability is also of great interest for climate (impact) research.

The wave statistics part of the atlas complements, updates and improves the few existing and popular sources of global wave statistics, namely the *Global Wave Statistics* book of Hogben *et al*. (1986) and the *Atlas of the Oceans: Wind and Wave Climate* of Young and Holland (1996). (Other sources are private consulting firms who sell wave information to their customers. Access to these data is limited and expensive.) The information in Hogben *et al*. (1986) was derived from visual observations from voluntary observing ships (VOSs) and, therefore, suffers from the relative unreliability of visual observations and their poor spatial coverage outside the North Atlantic. The statistical information presented in Hogben *et al*. (1986) is widely used in naval engineering. Therefore, an update and extension to global scale is worthwhile. The length and coverage of ERA-40 make such an extension possible.

The atlas of Young and Holland (1996) was created using 3 years of altimeter data. It contains a lot of global wind speed and wave height statistics, such as means and quantiles, but no information on wave period and direction. These quantities cannot be inferred reliably from altimeter observations. The ERA-40 data can provide this information. Furthermore, owing to its much longer period, it provides a more robust estimate of the climatology than Young and Holland (1996). More importantly, owing to the uniform spatial and temporal coverage, extreme value and time series analyses become feasible. The first could not be done with the 3 years of data available from Young and Holland (1996), and the second cannot be done from satellite passes.

Regarding climate variability, the atlas reports and analyses the variability observed during the 45 year period covered by ERA-40, paying special attention to its effects on parameter estimates. For instance, 100-year return wave height estimates based on data from three different decades are significantly different in the North Pacific and the North Atlantic. This part of the atlas is also intended as a complement and addition to current studies of ocean wave variability, such as by the WASA Group (1998) and Wang and Swail (2001), which were confined to the Northern Hemisphere.

The atlas covers significant wave height H_s, 10 m wind speed U_{10} and mean wave period T_m. However, owing to space limitations, this paper focuses on wave height.

2. THE DATA SOURCES

2.1. What is reanalysis?

A weather forecast is essentially an initial value problem. Given the atmospheric state (the 'weather') at one time, the state at a later time can be calculated. The problem with this simple view, however, is that the atmospheric state is never known exactly. For large parts of the atmosphere observations are not available (remote areas, upper air), and available measurements necessarily contain errors. To overcome this problem in operational weather forecasting, the initial state for a forecast is obtained by a combination of the latest forecast and all new observations. The latest forecast has usually been initialized 6 h earlier and gives a good *first guess* for the initialization of a new forecast. Most important, it provides a complete description of the atmosphere, as by definition it has values of all relevant quantities at all grid points. The first guess is then combined with the newly available observations in a way not violating physical laws. The observations 'push' the first guess towards 'reality'. This step, by no means trivial, is called *analysis*. At ECMWF, the analysis costs about half of the total CPU-time needed to make a 10 day forecast, the other half being used for the time integration.

As a consequence, operational forecast centres naturally produce a complete description of the atmosphere's state, usually four times a day. In principle, these data could be a valuable source of information for all kinds of investigation into the long-term variability of the atmosphere. However, weather forecast models (including the analysis procedure) are continually improved. Therefore, variability in the analyses is dominated by model changes rather than by natural variability, making them unsuitable for variability studies (e.g. see Siefridt *et al*. (1999)). The aim of *reanalysis* is to overcome this problem of inhomogeneity. A state-of-the-art analysis system is used to repeat the analysis procedure for the past. One can go back in time as far as the available

data coverage allows. As a result, one obtains a complete description of the atmosphere over a long period of time that is free of inhomogeneities due to model changes. Unfortunately, inhomogeneities due to changes in data coverage remain (e.g. Sturaro, 2003; Sterl, 2004).

2.2. ERA-40

For the production of ERA-40, a version of the IFS has been used that was operational in June 2001. To make a 45 year integration possible, the horizontal resolution of the model has been decreased to T_L159 ($\sim$125 km) instead of the T_L511 ($\sim$40 km) that is currently used in operations, and the 4DVAR data assimilation procedure has been replaced by the cheaper 3DVAR. A complete description of the IFS can be found at http://www.ecmwf.int/research/ifsdocs/index.html.

A distinguishing feature of ECMWF's model is its coupling to a wave model. The coupling is needed because, over sea, the roughness length depends on the sea state (Janssen, 1989, 1991). Specifically, the Charnock parameter (Charnock, 1955) is not taken as constant, but is a function of the whole wave spectrum. Thus, wave information is a natural product of ERA-40. Starting in 1991, wave height data obtained from the altimeters on board of ERS-1 and ERS-2 are assimilated. The impact of the assimilation will be discussed in Section 3.1.

The wave model used in IFS is the well-known WAM (Komen *et al*., 1994). It is a so-called third-generation model, in which the wave spectrum is computed by integration of the energy balance equation without any prior restriction of the spectral shape. From the full spectrum, integral quantities like significant wave height or mean wave frequency are calculated. The model resolution is $1.5^\circ \times 1.5^\circ$, and the time step is 15 min. At each fourth time step the actual 10 m wind from the atmosphere model, which has a time step of 20 min, is passed to WAM. The new roughness length is then computed and passed back to the atmosphere model, where it is used to calculate the air–sea fluxes of momentum, heat and moisture.

Output of results takes place at the common synoptic hours 00, 06, 12, and 18 UTC. A large subset of the complete ERA-40 data set, including H_s, mean wave period and mean wave direction, can be downloaded free of charge from http://data.ecmwf.int/data/d/era40_daily/. This subset is available on a $2.5^\circ \times 2.5^\circ$ grid. The complete dataset at the full model resolution ($1.5^\circ \times 1.5^\circ$) is available through ECMWF's Data Service. This service is not free.

2.3. Validation data

2.3.1. Buoy measurements. Buoy observations are the most reliable wave observations, but they are limited in space and time. Most buoys are located along the coast in the Northern Hemisphere, and are available only after 1978. We use buoy observations from the American National Data Buoy Center (NDBC-NOAA), which are freely available from http://www.nodc.noaa.gov/BUOY/buoy.html. The buoys are situated along the coasts of North America.

From the available NDBC-NOAA buoy locations, 20 have been selected for the validations. Selection criteria were the 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 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 surface wind measurements are available hourly from 20 min and 10 min length records respectively. Although these measurements have gone through some quality control, they are processed further using a procedure similar to the one used at ECMWF (Bidlot *et al*., 2002) and described in Caires and Sterl (2003a). Wind speeds are adjusted to 10 m height using a logarithmic profile under neutral stability (e.g. Bidlot *et al*., 2002). In order to compare the ERA-40 results with the observations, time and space scales must be made compatible. The reanalysis results are available at synoptic times (every 6 h) and each value is an estimate of the average condition in a grid cell, whereas the buoy measurements are local. Therefore, the reanalysis data are compared with 3 h averages of buoy observations, 3 h being the approximate time a long wave would take to cross the diagonal of a $1.5^\circ \times 1.5^\circ$ grid cell at mid latitudes. To get ERA-40 data at the buoy location, the reanalysis data at the appropriate synoptic time are interpolated bilinearly to the buoy location.

2.3.2. Altimeter measurements. Whereas buoys provide high-quality continuous point measurements, satellite-borne altimeters provide near-global coverage, but every point is sampled only once in several (typically 10) days. We use along-track quality-checked deep-water altimeter measurements of H_s and the normalized radar cross-section σ_0 from GEOSAT, TOPEX, ERS-1, and ERS-2. The data are obtained from the Southampton Oceanography Centre (SOC) GAPS interface (http://www.soc.soton.ac.uk/ALTIMETER; Snaith, 2000). The drift observed in TOPEX wave heights during 1997 to 1999 (cycles 170 to 235) is corrected according to Challenor and Cotton (1999), and the relation $H_s^{\text{buoy}} = 1.05 H_s^{\text{topex}} - 0.07$ (Caires and Sterl, 2003a) is used to make the TOPEX observations compatible with the buoy observations. The GEOSAT altimeter wave height data are increased by a factor of 1.065 according to Cotton and Carter (1996). No corrections are applied to the data from ERS-1 and ERS-2. The surface wind speed up to 20 m/s is obtained from σ_0 using the algorithm of Gourrion *et al*. (2002), whereas for wind speeds above 20 m/s the relation of Young (1993) is used. More details can be found in Caires and Sterl (2003a).

The satellite measurements are performed about every second with a spacing of about 5 to 7 km. 'Super observations' are formed by first grouping together consecutive measurements crossing a $1.5^\circ \times 1.5^\circ$ region. The satellite observation is then taken as the mean of these grouped data points. A quality control similar to the one applied to the buoy data is done. The reanalysis data are linearly interpolated in space and time to the mean location and the mean time of the altimeter observation.

3. ASSESSMENT OF THE ERA-40 WAVE PRODUCT

3.1. Validation

Before using the ERA-40 wave data to produce derived quantities they have been extensively validated against buoy and altimeter data. Figure 1 shows the time series of H_s as measured at buoy 46 001 in the Gulf of Alaska (148.3 °W, 56.3 °N) during 1988, together with the corresponding ERA-40 data. Three properties of the ERA-40 data can easily be recognized: (a) the two curves are nearly perfectly in phase; (b) low wave heights tend to be overestimated by ERA-40; (c) high waves tend to be substantially underestimated.

These three features are not a peculiarity of the special location, but a general property of the ERA-40 wave data. Among the reasons for these deficiencies are resolution (P. Janssen, personal communication) and a slight underestimation of high wind speeds (Caires and Sterl, 2003a). These reasons are, however, not sufficient to explain fully the severe underestimation of the wave heights in ERA-40.

Owing to changes in the data assimilated, the characteristics of the data are not homogeneous in time. Four different periods have to be distinguished.

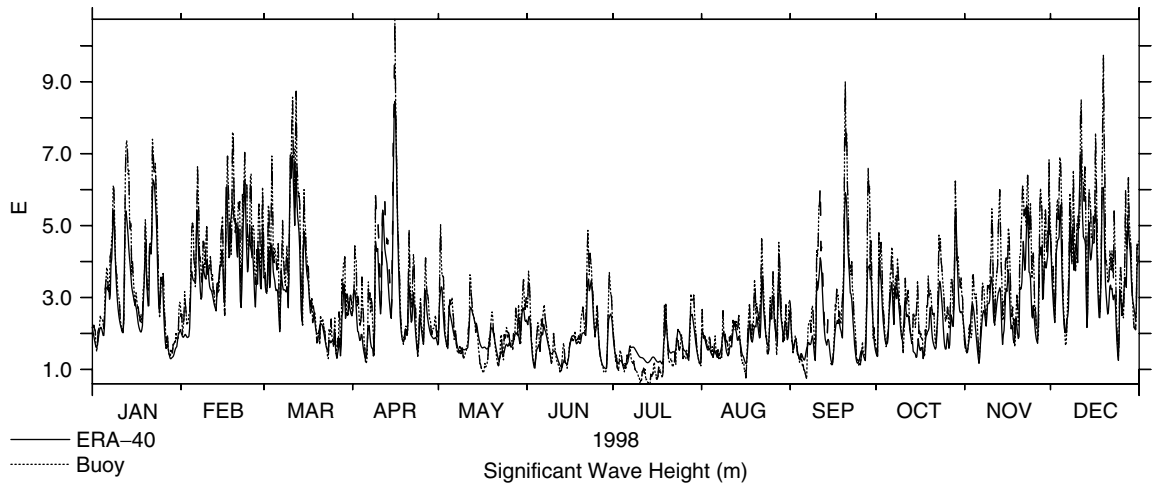

Figure 1. Measured (dotted) and modelled (solid) H_s at buoy 46 001 (148.3 °W, 56.3 °N) in 1988

P1. September 1957 to November 1991 (P1a) and June 1993 to December 1993 (P1b): no assimilation of altimeter wave height data.
P2. December 1991 to May 1993: assimilation of faulty ERS-1 fast delivery product (FDP) wave height data.
P3. January 1994 to May 1996: assimilation of good but uncalibrated ERS-1 FDP wave height data.
P4. June 1996 onwards: assimilation of ERS-2 FDP wave height data.

Figure 2 shows the time series of the globally averaged monthly mean H_s from ERA-40. There is no physical interpretation of this global average, nor are there climatological values indicating the range of values such an average should take. It is presented here to give a synthesized picture of the data. The four periods identified above are clearly visible. This is especially true for period P2. The faulty data that were assimilated have a density function with two peaks. One of them is sharp and located at around 2 m (Bauer and Staabs, 1998) and corresponds to a systematic overestimation of wave heights around that value. Period P4 can also be easily identified: it starts with a positive trend and then levels off.

Caires and Sterl (in press) have analysed the error characteristics of the four periods in more detail. The following is a short summary of their findings illustrated with quantile–quantile (Q–Q) plots of buoy wave heights against ERA-40 wave heights (Figure 3). Corresponding plots using altimeter wave heights instead of the buoy measurements give essentially the same picture.

P1. In this period, the monthly mean wave fields compare well with observations; but, as shown in Figure 1, ERA-40 underestimates high wave heights and overestimates the low ones. In particular, the underestimation is clearly visible in the upper left panel of Figure 3.
P2. In this period, the H_s values below 3 m are overestimated and those above are underestimated. The quality of the waves with heights above 3 m is similar to that in period P1. The Q–Q plot of ERA-40 data versus buoy data (upper right panel of Figure 3) clearly shows the overestimation for values between 1 and 3 m that is due to the peak in the distribution function of the ERS-1 FDP mentioned above.
P3. In this period, the known calibration correction to the ERS-1 FDP data was not applied because, although it would have improved the H_s data analysed, it would have given poorer, too high, mean wave periods. The quality of the wave height data is, therefore, similar to that of the data in period P1, though it has a lower scatter index (root-mean-square error normalized by the mean; not shown).
P4. The assimilation of the ERS-2 FDP measurements of wave height during P4 has improved the H_s analysed, especially in the tropics. The underestimation of high wave heights and the slight overestimation of low wave heights by the ERA-40 dataset, however, continues in this period, as is clearly seen in the lower right panel of Figure 3.

3.2. Estimation of extreme significant wave heights

For safety considerations it is important to know extreme wave heights, i.e. wave heights that are, on average, exceeded only once per 20, 50, or 100 years. Despite the ERA-40 wave heights' inability to capture high waves, the data set proved an invaluable basis to obtain global estimates of these extremes. Note that only

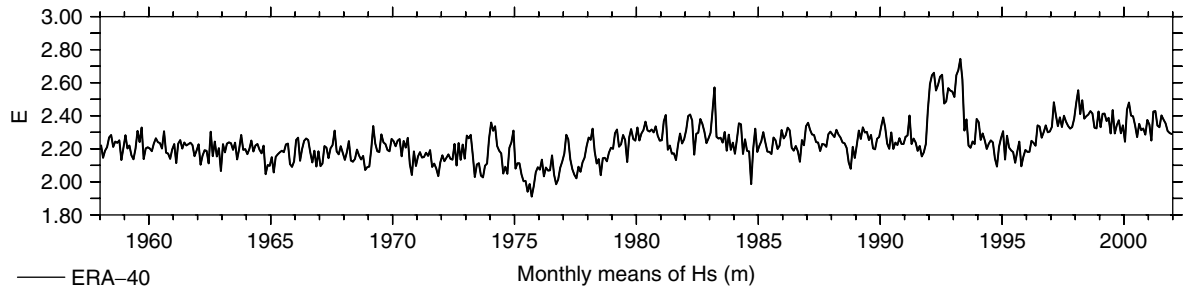

Figure 2. Time series of the monthly mean, globally averaged H_s from ERA-40. Monthly means are computed from the 6 h fields between 81 °S and 81°N, and a latitudinal correction has been applied

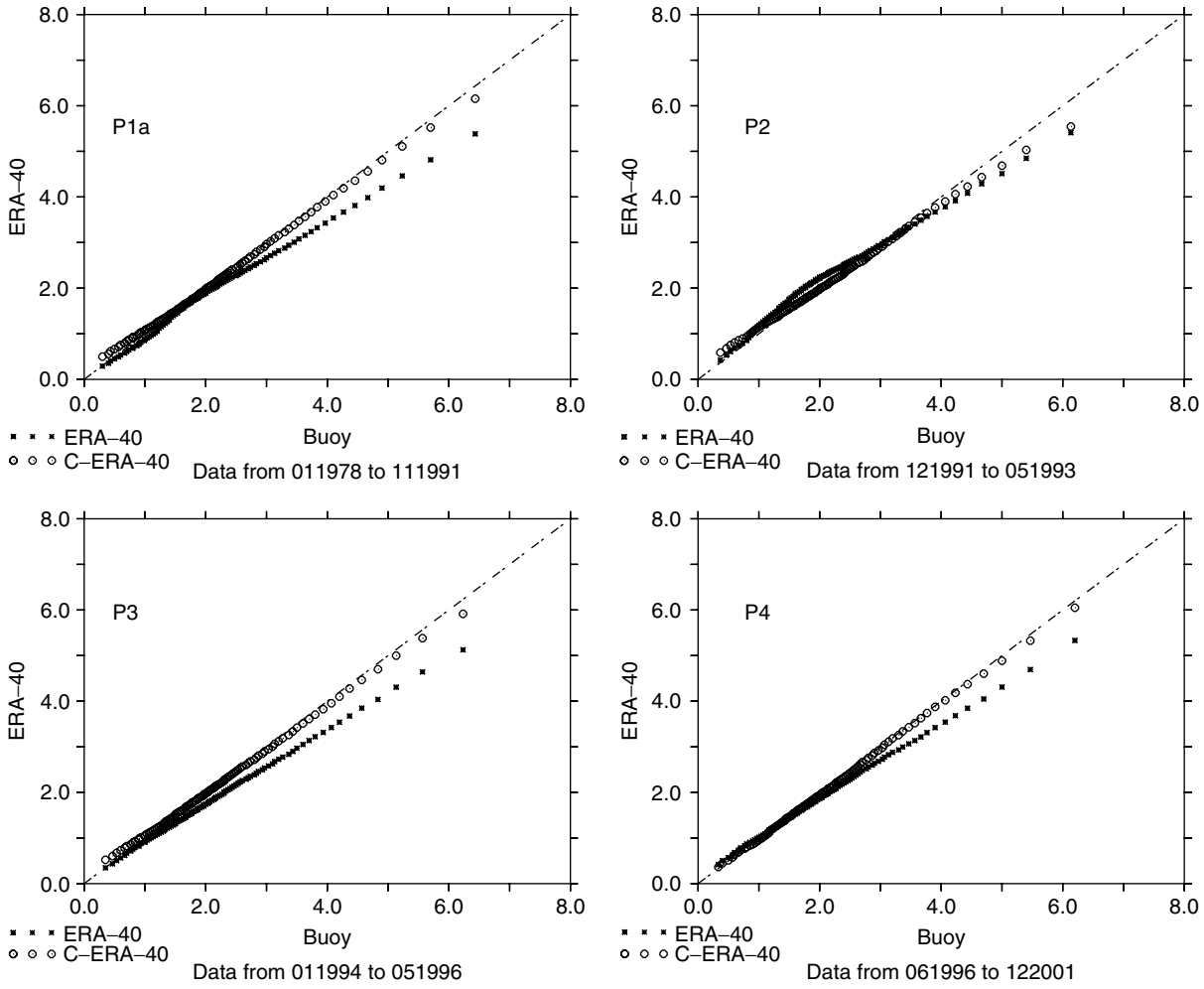

Figure 3. Quantile–quantile plots of H_s from the 20 selected buoys (see Section 2.3.1) against collocated ERA-40 values for the different periods indicated. Asterisks: raw ERA-40 data; circles: corrected data (to be discussed in Section 3.3)

extremes of significant wave height, rather than those of individual waves, can be obtained from the ERA-40 data. Estimates of the 100 year return significant wave height X_{100} obtained from buoy measurements and from the ERA-40 data appear to be linearly related by

$$X_{100}^{\text{buoy}} = 0.52 + 1.30 X_{100}^{\text{ERA}-40} \tag{1}$$

This relation is illustrated in Figure 4.

To estimate return wave heights we used the peak-over-threshold (POT) method (e.g. Coles, 2001), rather than the usual fit of a distribution function. There are no good theoretical arguments as to what distribution that fit should be. In the POT method, a threshold is chosen. Whenever wave height exceeds that threshold for a period of time, the highest exceedence within that period is recorded. On theoretical grounds, the exceedences must fit the two-parameter generalized Pareto distribution (GPD). As a special case, the GDP contains the exponential distribution, which has only one parameter. Statistical tests show that the observed distribution of the exceedences cannot be distinguished from an exponential one. Therefore, an exponential distribution was fitted to the exceedences to obtain the return values. After some trial and error, the 93% quantile was chosen as the threshold at each grid point. Doing so both for buoy measurements and for the ERA-40 data resulted in Equation (1). More details can be found in Caires and Sterl (2003b, 2005).

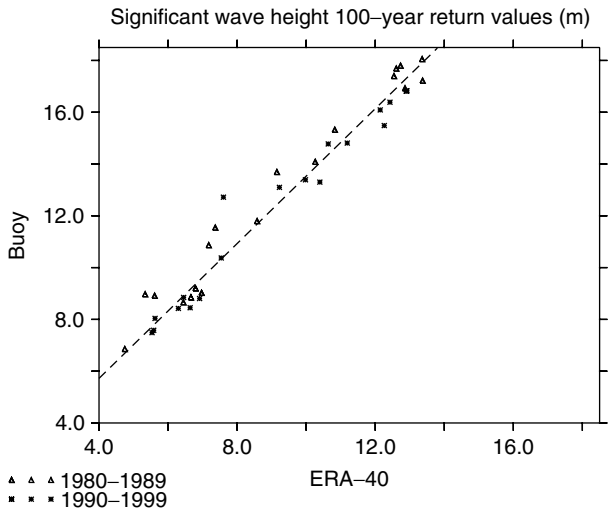

Figure 4. Linear correlation between 100 year return value estimates of H_s from buoy data and from ERA-40. The dashed line is Equation (1)

Buoy locations are very unevenly distributed in space, and the largest value of X_{100} found at the buoy locations is about 17 m (Figure 4). Therefore, it would be preferable to have a relation between X_{100} estimates from ERA-40 and from satellites. However, satellites cross a given point only once in typically 10 days. Together with the relative shortness of the satellite record, this gives too few data for a reliable extreme-value estimate. In particular, the average number of exceedences per year cannot be determined. However, as far as parts of the estimation procedure were possible with satellite data, their results are not incompatible with Equation (1), and we therefore apply Equation (1) globally and for all values of X_{100}. More details can be found in Caires and Sterl (2003b, 2005).

Figure 5 shows the X_{100} values obtained by applying the POT method to the ERA-40 data and correcting the results using Equation (1). It is obvious that the highest values occur in the North Atlantic. One might suspect this to result from observation density being higher in the North Atlantic than in the Southern Ocean. Although this is true for the early years, satellite data density is comparable in both areas towards the end of the ERA-40 period, but still the return values are highest in the North Atlantic (Figure 5, lower right panel). Although mean wave heights are not higher in the North Atlantic than they are in the North Pacific or in the Southern Ocean (see Figure 6 below), the North Atlantic shows the highest variability as measured, e.g. by the inter-monthly standard deviation (not shown). In other words, conditions in the Southern Ocean are always rough, whereas in the North Atlantic you can be lucky and the sea is calm even in winter, or you find yourself between the highest waves possible on Earth.

Some care has to be taken in interpreting the maps in Figure 5. First, the ERA-40 values represent 6 h averages over a $1.5^\circ \times 1.5^\circ$ area. The time and space scales of the buoy data have been made compatible with this model scale as described in Section 2.3.1. Therefore, the extreme values depicted in Figure 5 are for those scales, and much higher waves on smaller spatial and temporal scales must be expected. Second, the version of the WAM model used for ERA-40 does not contain shallow-water effects. Therefore, Figure 5 is not valid along the coasts. Finally, tropical storms are not properly resolved on a T_L 159 grid. In regions of tropical cyclones, extreme waves heights are, therefore, expected to be higher than shown in the figure.

3.3. Correction of ERA-40 data

Two main limitations of the ERA-40 H_s data have been identified. The existence of inhomogeneities in time (see Figure 2) limits the use of the data for studies of climate variability and trends, and the underestimation of high wave heights (Figure 1) discourages the use of the data in design studies, where a good description of the data in all ranges is important.

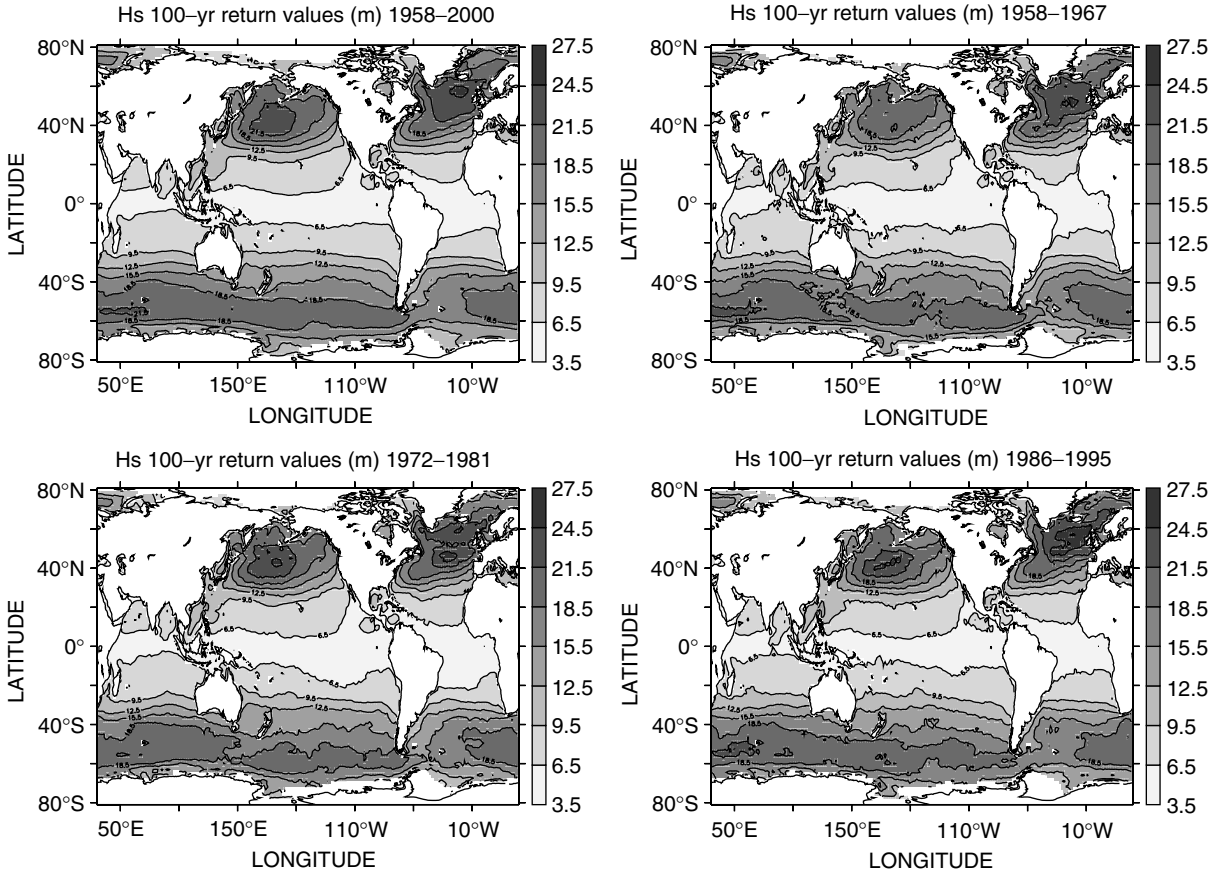

Figure 5. The 100 year return H_s from ERA-40, corrected using the relationship displayed in Figure 4. Note that the results pertain to averages over 1.5° × 1.5°, that shallow-water effects are not included, and that tropical cyclones are not resolved in ERA-40. The upper left panel is for the whole ERA-40 period (1958–2000), and the other panels are derived from three 10 year sub-periods, as indicated. These panels are discussed in Section 4.2

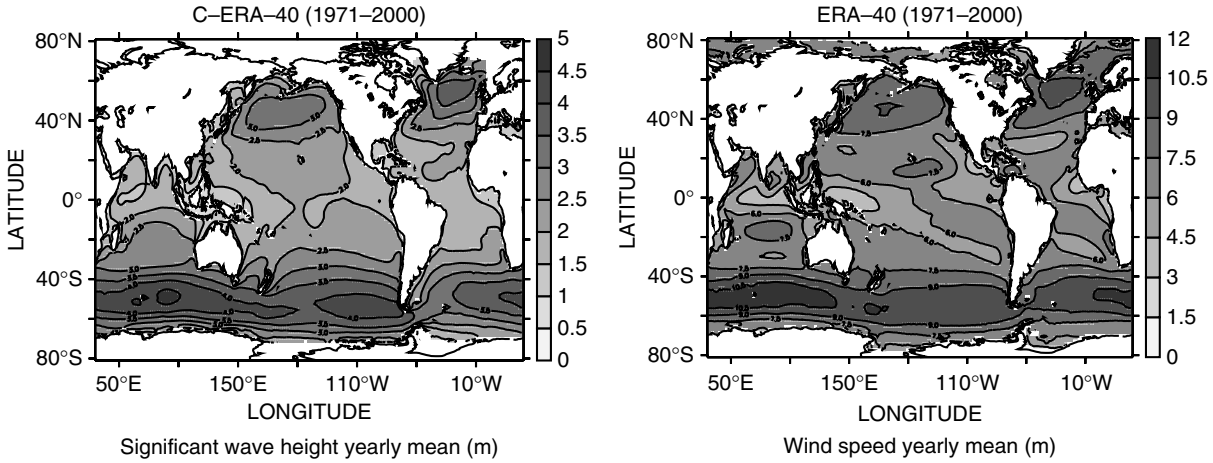

Figure 6. Annual mean climate of H_s (C-ERA-40, left) and U_{10} (right)

Inspection of time series at buoy locations, such as Figure 1, led us to the conclusion that the disagreement between modelled and observed H_s is similar in similar situations. Based on this observation, we proposed a new approach to improve the ERA-40 significant wave height fields. It is based on nonparametric estimation (Caires and Sterl, in press). The idea is to estimate at each time step the error between the ERA-40 H_s value and the 'true' significant wave height value and then correct the data using the estimate. The first step in error estimation is to construct a 'learning dataset' by collecting the discrepancies between model and 'truth', the latter being represented by collocated TOPEX measurements. This is done separately for each of the four periods P1–P4 to account for the inhomogeneities. The second step is to find for each model time step 'similar' situations in the learning dataset and to use the errors to correct the model value. Usually, several similar situations are found, so that a confidence interval around the corrected data can also be given. Trying several possibilities, the most efficient way to define 'similar' was found to require that the last three consecutive values of H_s are close together. See Caires and Sterl (in press) for more details.

Using this method, we created a new 45-year global six-hourly dataset: the C-ERA-40 dataset. Comparisons of the C-ERA-40 data with measurements from *in situ* buoy and global altimeter data show clear improvements in bias, scatter and quantiles in the whole range of values, as well as the removal of the inhomogeneities that are due to changes in altimeter wave height assimilation (see Section 3.1). This can be seen from the Q–Q plots in Figure 3, which contain the results from both the original ERA-40 data and from C-ERA-40, as well as from a comparison of the global-mean H_s from ERA-40 (Figure 2) and from C-ERA-40 (see Figure 8 below). The plots show that the nonparametric corrections work effectively in the whole range of H_s values and for all periods.

4. SOME HIGHLIGHTS FROM THE KNMI/ERA-40 WAVE ATLAS

The atlas is divided into five main parts: introduction and background; description of the data sources; data validation; description of climate; and climate variability. Here, we will describe in some detail how the information on climate and its variability are presented in the atlas and highlight some aspects.

4.1. Climate

Climate is by definition the synthesis of weather conditions in a given area, characterized by long-term statistics (mean values, standard deviations, quantiles, etc.) of the meteorological elements in that area. The World Meteorological Organization (WMO) recommends climate to be based on 30 years of data. The wave climate information provided in the atlas is, therefore, based on the 30 years from 1971 to 2000. It includes monthly and annual means, standard deviations, 90% and 99% quantiles, the annual mean time of exceedence of certain thresholds, namely 3, 6 and 9 m for H_s, and 11, 17 and 24 m/s for U_{10} (chosen in line with the minimum velocities of the WMO 1100 Beaufort scale for strong breeze (Beaufort 6, 10.8 m/s), gale (Beaufort 8, 17.2 m/s) and storm (Beaufort 10, 24.5 m/s)), tabulated frequency histograms of H_s and mean wave period, and estimates of 100-year return values, using the method described in Section 3.2. The return values are based on the whole data set (not only 1971–2000) to increase their accuracy.

Figure 6 shows the annual mean climates of H_s (from the corrected C-ERA-40; see Section 3.3) and U_{10}. They are characterized by high values in the storm track regions of both hemispheres and low values in the tropics. Although the highest means occur in the Southern Hemisphere, the most extreme wave and wind conditions are found in the North Atlantic. Figure 7 shows the annual mean exceedences of the 9 m and 24 m/s thresholds of H_s and U_{10} respectively. Exceedences are highest in the Northern Hemisphere, especially in the North Atlantic. In this region also, the 100-year return value estimates of H_s (Figure 5) and U_{10} (not shown) are highest.

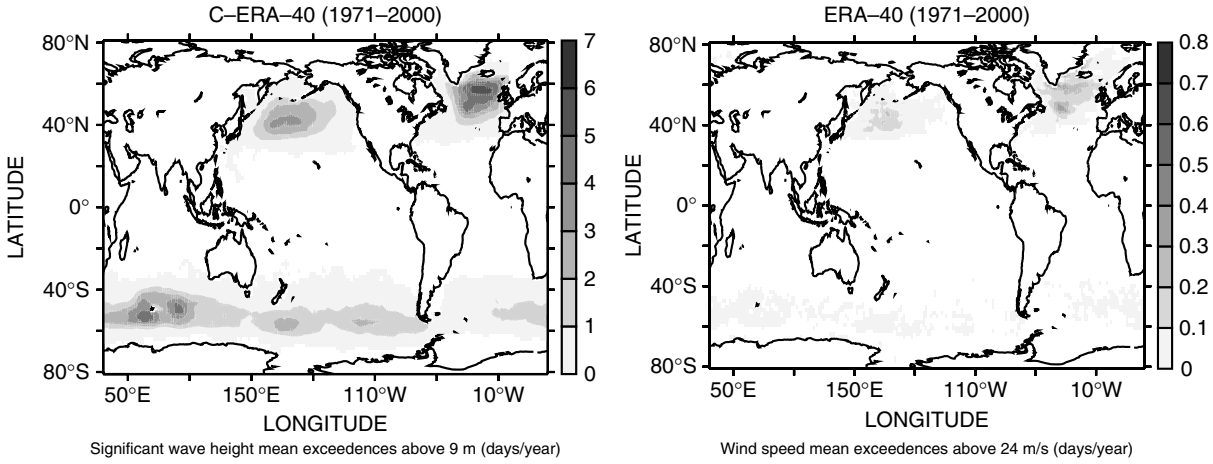

Figure 7. Mean annual exceedences of 9 m of H_s (C-ERA-40, left) and of 24 m/s of U_{10} (right) in days per year. Note that the results pertain to 6 h averages over 1.5° square boxes with no wind gusts included

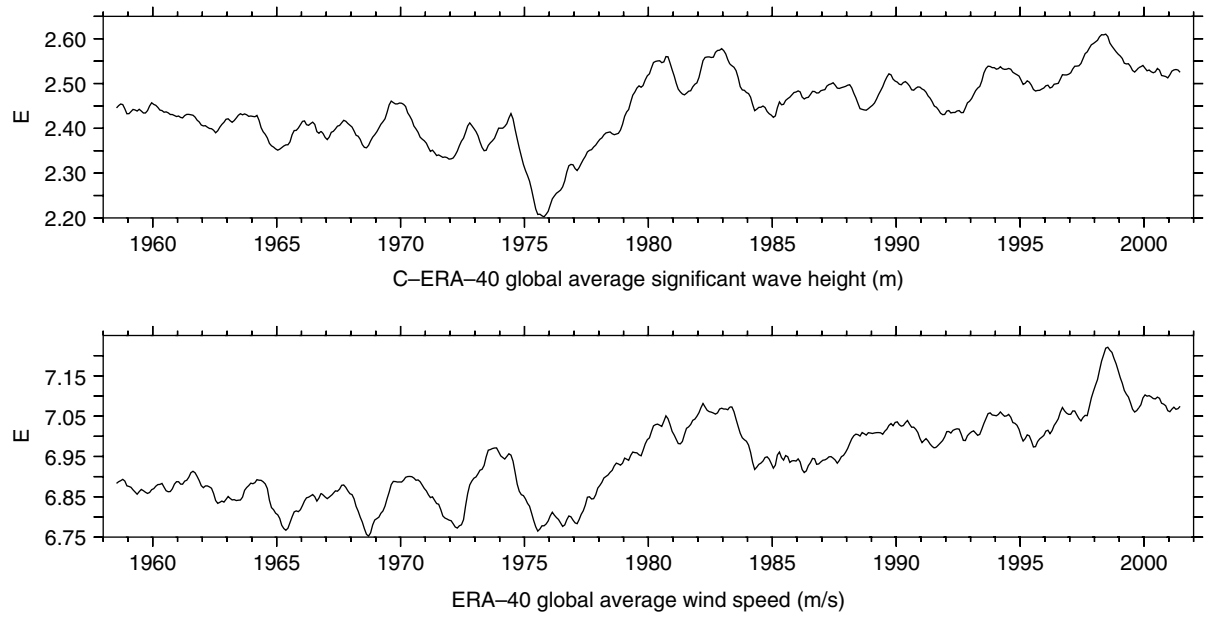

Figure 8. Time series of globally averaged H_s (C-ERA-40, top) and U_{10} (bottom) using latitude correction and a smoothing of 12 months to remove the annual cycle

4.2. Climate variability

The atlas describes the wind and wave climate variability in several ways. A short summary is provided by basin-averaged monthly mean time series for seven ocean basins (global, Antarctic Ocean, Indian Ocean, South and North Pacific, South and North Atlantic). Figure 8 shows the C-ERA-40 H_s and U_{10} average of monthly means over the globe using latitude correction and a smoothing of 12 months to remove the annual cycle. The most prominent feature of the H_s time series is a dip in September 1975, which also seems to signal a change in regime, since the level of the time series after the dip is higher than that before. This feature is also present in the U_{10} time series and can be traced to the Pacific sector of the Antarctic Ocean (between 120 °E, 65 °S and 120 °W, 25 °S), where U_{10} obtains its minimum. Owing to swell propagation, it

affects the average of H_s over all basins with the exception of the North Atlantic (see Figure 10). We cannot trace the 1975 minimum in the time series to changes in the observation system of ERA-40; therefore, it is possible that it is real feature of the climate system. However, the change in the level of the time series before and after the minimum is most likely due to the assimilation of satellite data from 1979 onwards.

The variability is described in more detail by maps of monthly and annual anomalies of the mean and the 90% and 99% quantiles. Anomalies are calculated with respect to the period 1971 to 2000. One of the ways in which variability can be revealed is through the detection of trends. Therefore, the atlas contains maps of trends of the monthly means and of the 90% and 99% quantiles. The trends vary per calendar month and from location to location, with some regions characterized by negative trends and others by positive trends. The trends in the 90% and 99% quantiles show the same spatial patterns as those in the mean, but they have higher slopes. Maximum trends in the mean H_s are of about 4 cm/year and in the 99% quantiles of about 7 cm/year. For wind speed the upper limits are about 6 (cm/s)/year for the mean and 12 (cm/s)/year for the 99% quantiles. As an example, Figure 9 shows the trends in the February monthly means and 99% quantiles from the C-ERA-40 H_s data. Note that, except for the North Atlantic, the trends are dominated by the increase between the 1970s and the 1980s discussed above (Figure 8). The trend found in the North Atlantic and its spatial pattern are in line with the results of Günther *et al*. (1998).

We have used empirical orthogonal function (EOF) analysis to obtain the main patterns of variability, since these patterns may be linked to possible dynamic mechanisms. The atlas presents, for each ocean basin considered, the two most important EOF patterns and their coefficient time series. Some interesting observations arise from the EOF analysis:

- Figure 10 shows the pattern of the first global EOF of C-ERA-40 significant wave height. This EOF explains 15% of the global variability and clearly represents swell propagating from the Southern Hemisphere storm track region into the Indian and Pacific Oceans. Its coefficient has a correlation of about 0.8 with the global mean of C-ERA-40 significant wave height (Figure 8). In particular, it has the same dip around September 1975 as has the global curve. It illustrates the importance of the Southern Hemisphere in governing the variability of the global mean H_s.
- The coefficient time series of the first North Pacific EOF of H_s has a correlation of about -0.76 with the Pacific–North American (PNA) index (Wallace and Gutzler 1981). It explains 31% of the variability in that basin.

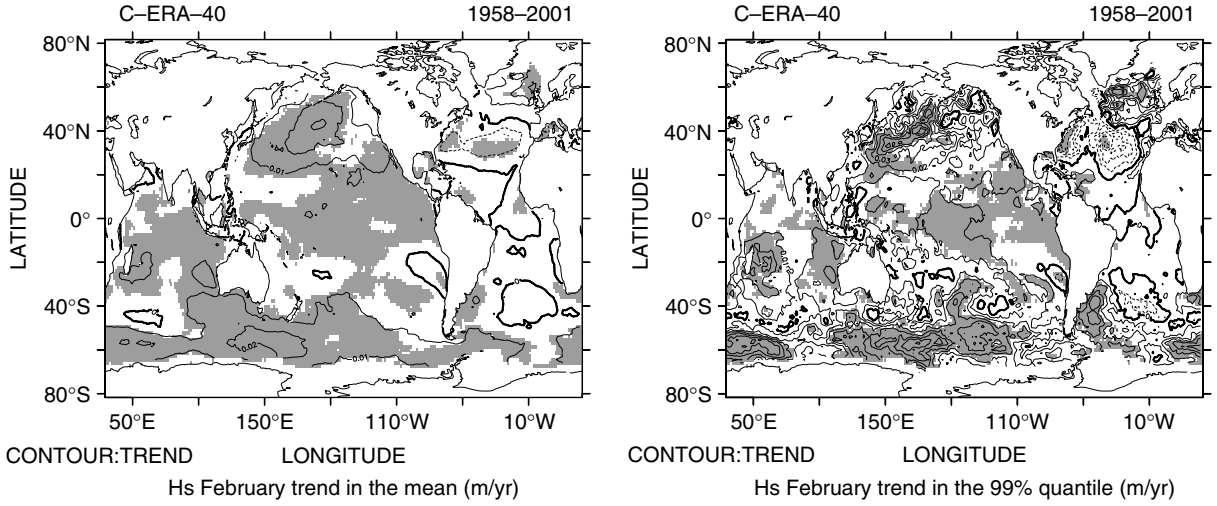

Figure 9. Trends in the February monthly mean H_s (C-ERA-40, left) and the 99% quantiles (right). Areas where the trend is significant at the 5% level are shaded. The significance is estimated using the nonparametric Mann–Kendall test, as described in Wang and Swail (2001)

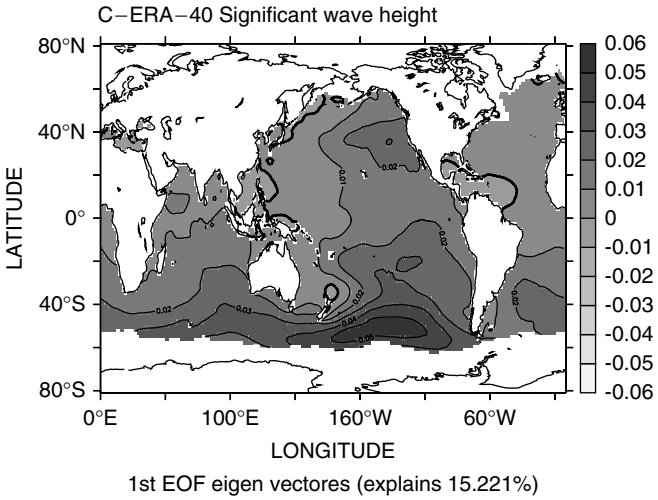

Figure 10. Pattern of the first global EOF of H_s (C-ERA-40)

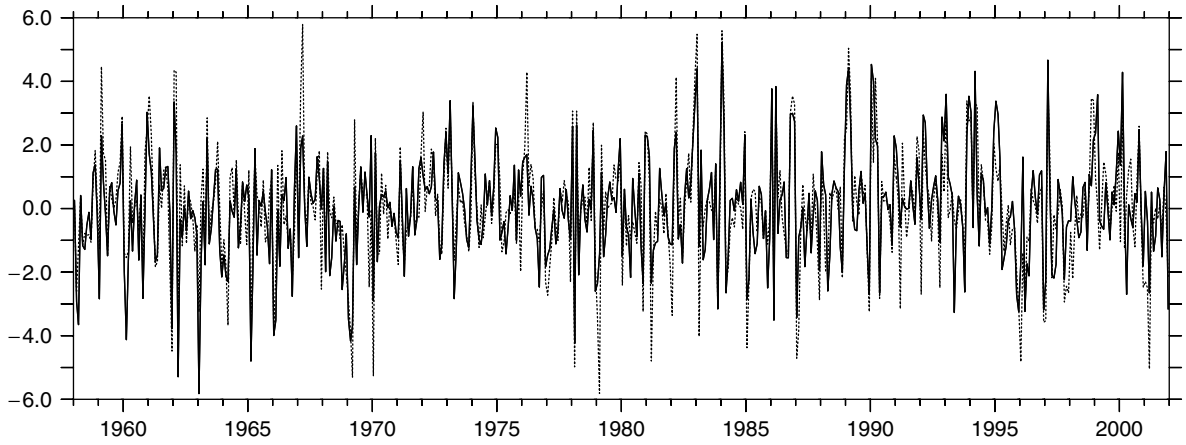

Figure 11. Time series of the ERA-40 NAO index (solid) and the coefficients of the second North Atlantic EOF of H_s (dashed)

- The coefficients of the second North Atlantic EOF, which explains 24% of the variability in that basin, has a correlation of about 0.8 with the North Atlantic oscillation (NAO) index (e.g. see Rogers (1984)). Figure 11 shows the NAO index computed from the ERA-40 mean sea-level pressures in the Azores (26 °W, 38 °N) and Iceland (18 °W, 66 °N) and the coefficients of the eigen vectors of the second EOF for the North Atlantic. The first EOF is dominated by swell and, therefore, not related to the NAO.

Finally, the effect of decadal climate variability on the extreme statistics, namely on the annual mean time of exceedence of certain thresholds and on the 100-year return values, is described. An example is given in Figure 5 for the corrected H_s 100-year return-value estimates. Besides an estimate based on the whole ERA-40 period it also contains estimates based on three different 10 year periods. The estimates obtained from these periods differ in the Northern Hemisphere storm tracks. Specifically, the estimates in the roughest part of the North Pacific storm track region have increased, and in the North Atlantic the pattern has changed. These differences can be attributed to the decadal variability in the Northern Hemisphere, especially to changes in the phase of the NAO (Caires and Sterl, 2005). This example shows that it is important to take account of climate changes when designing maritime structures.

As we have seen, the H_s data reveal trends in both the monthly means and, even more pronounced, the high quantiles. There are also differences in the return values estimated with data from different decades. It is interesting to investigate whether changes in the monthly means and the return values arise from *more* or from *more-intense* storms. Using the POT method we calculated the average number of clusters above a threshold per year (λ, number of storms) and the average of the peak excesses (α, intensity of storms) from the data (see Caires and Sterl (2005) for details). Using the 90% quantile of all data at a given location from 1971 to 2000 as the threshold, we computed estimates of α and λ for the three decades considered above. The results are shown in Figure 12. From this figure we conclude that:

- There are more storms in the storm track region than in the tropics. As we have defined storms as the roughest 10% of the time at a given place, this can be interpreted as storms in the tropics lasting longer. Note, however, that the 90% quantile is much lower in the tropics, so that a typical storm there is much less severe (remember that tropical cyclones are not resolved!) than at higher latitudes. Therefore, it is more appropriate to conclude that the tropics are less variable than the higher latitudes.
- Significant changes in the number of storms λ occur over much larger areas than do changes in intensity α.
- In the North Atlantic, only the intensity of storms changes significantly, and it does so only in small regions.
- In the North Pacific, both the number of storms and their intensity changes.
- Changes in the Southern Hemisphere are mainly due to changes in the number of storms. This change, however, may be an artifact of satellite observations becoming available in 1979 (see also discussion of Figure 8).

5. SUMMARY AND CONCLUSIONS

The ERA-40 reanalysis carried out at ECMWF produced 45 years (September 1957–August 2002) of data describing the state of the atmosphere four times a day. ECMWF's operational model has been used to carry out the reanalysis. In this model, the exchange coefficients for momentum and turbulent energy are dependent on the sea state. To achieve this, the atmosphere model is coupled to the WAM wave model. Therefore, the ERA-40 data also contain information about waves. A subset of the raw ERA-40 data can be downloaded freely from ECMWF's Website at http://www.ecmwf.int/data/d/era40_daily/.

A thorough assessment of the ERA-40 wave height data revealed that they (a) capture very well the variability of the true wave heights on all time scales, (b) slightly overestimate low wave heights, and (c) severely underestimate high wave heights. Furthermore, inhomogeneities due to the assimilation of different data sources are clearly present.

Despite the underestimation of high wave heights, it is possible to give reliable estimates of extreme significant wave heights ('100-year-return values'). Estimates based on the raw ERA-40 wave data and those from buoy measurements revealed a linear relationship that could be exploited to obtain global reliable return-value estimates based on the ERA-40 data. Furthermore, it was possible to devise a nonparametric correction method to the ERA-40 data, resulting in a corrected dataset that has no bias with respect to altimeter-based wave height retrievals and which is free of obvious inhomogeneities resulting from differences in wave-height data that were assimilated.

The ERA-40 wave data have been used to create the Web-based KNMI/ERA-40 wave atlas (http://www.knmi.nl/waveatlas). This atlas contains the comparisons between the ERA-40 data and observations, both from buoys and from satellite altimeters, the climatology of waves as deduced from both the raw and the corrected ERA-40 data, maps of exceedences, as well as of return values, and an assessment of the variability of the wave climate. The latter is especially important for the derivation of the extreme statistics, as the outcome of an extreme-value analysis can depend very much on the period used, with corresponding consequences for decisions based on this analysis.

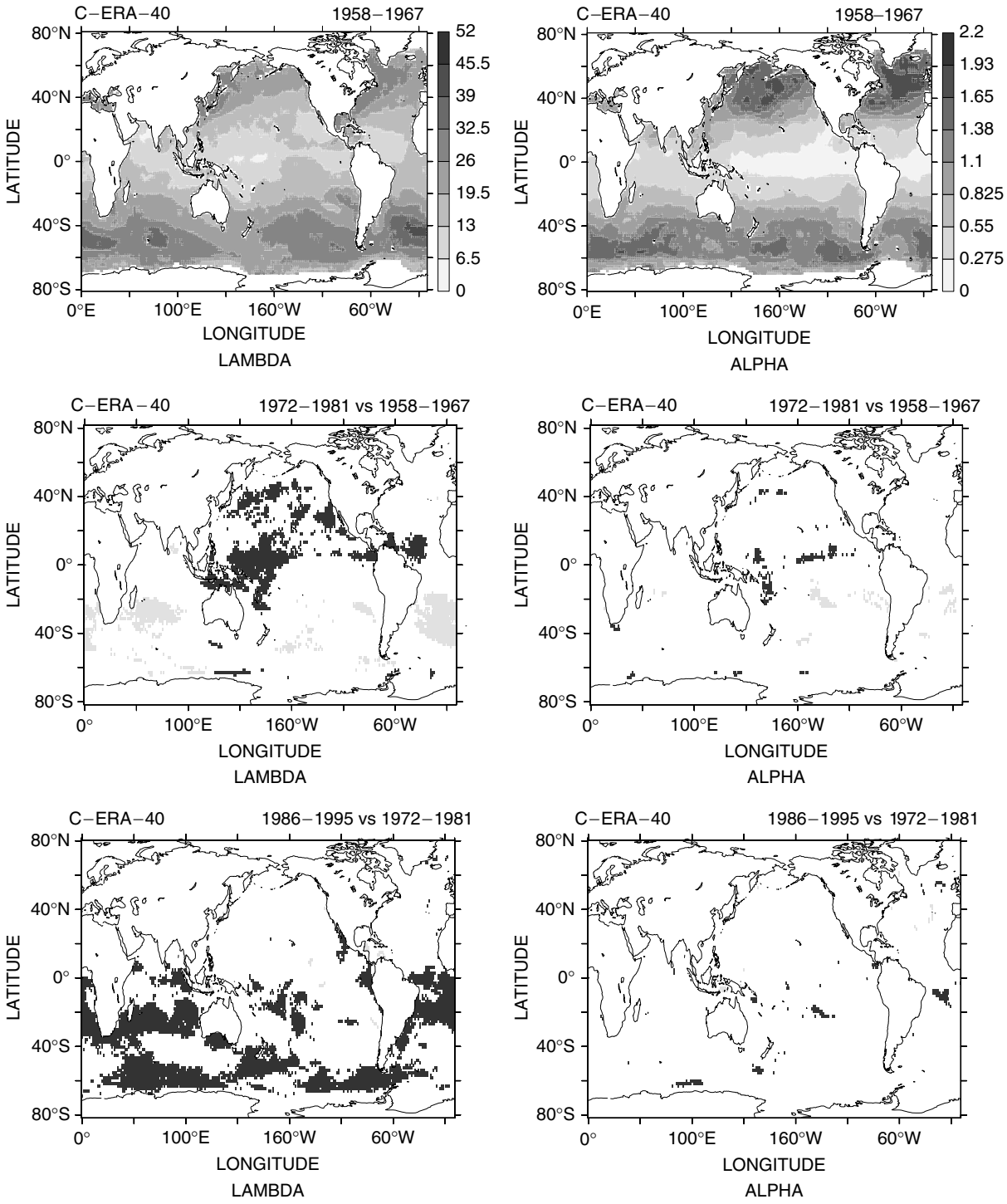

Figure 12. Top: number of storms (left) and average of the peak excesses (right) above the climate 90% quantile (severeness of storms) using data from 1958–67. Middle: changes in both quantities during 1972–81 with respect to 1958–67. Bottom: changes in 1986–95 with respect to 1972–81. Only significant changes are indicated. Dark (light) shading indicates increase (decrease)

ACKNOWLEDGEMENTS

We are indebted to many people for their help and pleasant collaborations. Jean-Raymond Bidlot and Peter Janssen provided valuable suggestions and comments and helped with advice. Sakari Uppala and Per Kållberg, as leaders of the ERA-40 production team, were always open to our comments and provided valuable help in dealing with the technical aspects of the ERA-40 system. Val Swail suggested to us the production of the KNMI/ERA-40 wave atlas and together with Gerbrand Komen advised on its content. Helen Snaith helped with the altimeter data. The buoy data were obtained from NDBC–NOAA (http://www.nodc.noaa.gov/BUOY/buoy.html). The plotting was done with the free Ferret software developed by NOAA/PMEL/TMAP. Camiel Severijns provided software support. An INTAS grant (01-2206) facilitated discussions with Sergey Gulev, David Woolf and Roman Bortkovskii. This work was funded by the EU as part of the ERA-40 project (no. EVK2-CT-1999-00027).

REFERENCES

Bauer E, Staabs C. 1998. Statistical properties of global significant wave heights and their use for validation. *Journal of Geophysical Research* **103**(C1): 1153–1166.

Bidlot J-R, Holmes DJ, Wittmann PA, Lalbeharry R, Chen HS. 2002. Intercomparison of the performance of operational wave forecasting systems with buoy data. *Weather and Forecasting* **17**: 287–310.

Caires S, Sterl A. 2003a. Validation of ocean wind and wave data using triple collocation. *Journal of Geophysical Research* **108**(C3): 3098. DOI: 10.1029/2002JC001491.

Caires S, Sterl A. 2003b. On the estimation of return values of significant wave height data from the reanalysis of the European Centre for Medium-Range Weather Forecasts. In *Safety and Reliability*, Bedford T, van Gelder PHAJM (eds). Swets & Zeitlinger: Lisse; 353–361.

Caires S, Sterl A. 2005. 100-year return value estimates for wind speed and significant wave height from the ERA-40 data. *Journal of Climate* **18**: 1032–1048.

Caires S, Sterl A. In press. Validation and non-parametric correction to significant wave height data from the ERA-40 reanalysis. *Journal of Atmospheric and Oceanic Technology*.

Challenor P, Cotton PD. 1999. Trends in TOPEX significant wave height measurement. http://www.soc.soton.ac.uk/JRD/SAT/TOPtren/TOPtren.pdf [accessed 22 March 2005].

Charnock H. 1955. Wind stress on a water surface. *Quarterly Journal of the Royal Meteorological Society* **81**: 639–640.

Coles S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. *Springer Texts in Statistics*. Springer-Verlag: UK.

Cotton PD, Carter DJT. 1996. Calibration and validation of ERS-2 altimeter wind/wave measurements. Southampton Oceanography Centre, Internal Document 12, unpublished manuscript (D.R.A. I.T.T. CSM/078).

Gourrion J, Vandemark D, Bailey S, Chapron B, Gommenginger CP, Challenor PG, Srokosz MA. 2002. A two parameter wind speed algorithm for Ku-band altimeters. *Journal of Atmosphere and Oceanic Technology* **19**(12): 2030–2048.

Günther H, Rosenthal W, Stawarz M, Carretero JC, Gomez M, Lozano I, Serano O, Reistad M. 1998. The wave climate of the northeast Atlantic over the period 1955–94: the WASA wave hindcast. *The Global Atmosphere and Ocean System* **6**: 121–163.

Hogben N, Da Cunha NMC, Oliver GF. 1986. *Global Wave Statistics*. Unwin Brothers: London.

Janssen PAEM. 1989. Wave-induced stress and the drag of air flow over sea waves. *Journal of Physical Oceanography* **19**: 745–754.

Janssen PAEM. 1991. Quasi-linear theory of wind wave generation applied to wave forecasting. *Journal of Physical Oceanography* **21**: 1631–1642.

Komen GJ, Cavaleri L, Donelan M, Hasselmann K, Hasselmann S, Janssen PAEM. 1994. *Dynamics and Modelling of Ocean Waves*. Cambridge University Press: Cambridge.

Rogers JC. 1984. The association between the North Atlantic oscillation and the southern oscillation in the Northern Hemisphere. *Monthly Weather Review* **112**: 1999–2015.

Siefridt L, Barnier B, Béranger K, Roquet H. 1999. Evaluation of operational ECMWF surface heat fluxes: impact of parameterization changes during 1986–1995. *Journal of Marine Systems* **19**: 113–135.

Snaith HM. 2000. *Global Altimeter Processing Scheme User Manual*. Southampton Oceanography Centre.

Sterl A. 2004. On the (in-)homogeneity of reanalysis products. *Journal of Climate* **17**: 3866–3873.

Sturaro G. 2003. A closer look at the climatological discontinuities present in the NCEP/NCAR reanalysis temperature due to the introduction of satellite data. *Climate Dynamics* **21**: 309–316. DOI: 10.1007/s00382-003-0334-4.

Wallace JM, Gutzler DS. 1981. Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Review* **109**: 784–812.

Wang XL, Swail VR. 2001. Changes of extreme wave heights in Northern Hemisphere oceans and related atmospheric circulation regimes. *Journal of Climate* **14**: 2204–2221.

WASA Group. 1998. Changing waves and storms in the northeast Atlantic? *Bulletin of the American Meteorological Society* **79**: 741–760.

Young IR. 1993. An estimate of the Geosat altimeter wind speed algorithm at high wind speeds. *Journal of Geophysical Research* **98**(C1): 20 275–20 285.

Young IR, Holland GJ. 1996. *Atlas of the Oceans: Wind and Wave Climate*. Pergamon.

