Estimating extreme wave height probabilities from observations and the ERA-40 reanalysis

Andreas Sterl and Sofia Caires

Introduction

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, however, with this simple view is that the atmospheric state is never known completely. For large parts of the atmosphere observations are not available (remote areas, upper air), and available measurements inevitably contain errors. To overcome this problem, the initial state for a new weather forecast is obtained by a combination of the latest forecast and all new observations. The latest forecast has usually been initialised six hours earlier and gives a good first guess for the initialisation of the new forecast. Most importantly, 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 respecting 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. However, weather forecast models and analysis procedures are continually improved. Variability in the operational analyses is dominated by these changes rather than by natural variability, making them unsuitable to study long-term changes. The aim of re-analysis is to overcome this problem of inhomogeneity. A state-of-the-art analysis system is used to repeat the analysis procedure for the past. As a result one obtains a complete description of the atmosphere over a long period of time, which is free of inhomogeneities due to model changes. Unfortunately, inhomogeneities due to changes in data coverage remain [1].

ERA-40

The first global reanalyses were produced in the first half of the 1990s [2,3]. They are widely used in climate and meteorological research. Progress in modelling and data assimilation as well as the availability of new data sets led ECMWF to conduct a new reanalysis, ERA-40, covering the 45 years from September 1957 to August 2002. It uses a version of ECMWF's Integrated Forecasting System (IFS) that was operational in June 2001, albeit on a coarser grid ($\approx$ 125 km instead of $\approx$ 40 km). A large subset of the complete ERA-40 data set is available at *http://data.ecmwf.int/data/d/era40_daily* on a 2.5° × 2.5° grid.

A distinguishing feature of the IFS is that over the oceans the surface roughness depends on the sea state[4,5], and the sea state is obtained from the WAM wave model[6]. Thus wave information is a natural product of ERA-40. One of the most important wave parameters is the significant wave height (H_S), a measure of the severity of the sea state. (To be precise it is the 20 minute-average of the upper third of a wave height record).

The length of the ERA-40 data set makes it especially suitable to study variability and extremes of weather-related quantities. Information about decadal variability of climate quantities and their extremes is of great interest for climate (impact) research. An example of an extreme parameter is the 100-year return wave height (H_{100}). 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 maritime structures.

Validation and correction of the ERA-40 wave product

We have thoroughly validated the raw ERA-40 wave data against buoy and altimeter data. Buoys provide high-quality continuous point measurements at a very limited number of sites. Satellite-born altimeters provide near-global coverage, but every point is sampled only once in several (typically 10) days.

Figure 1 shows the time series of H_S as derived from measurements at buoy 46001 in the Gulf of Alaska (148.3°W, 56.3 °N) during 1988 (blue), together with the corresponding ERA-40 data (red). 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 slightly overestimated by ERA-40, and (c) high waves tend to be substantially underestimated. These three features are not a peculiarity of the special location, but a general

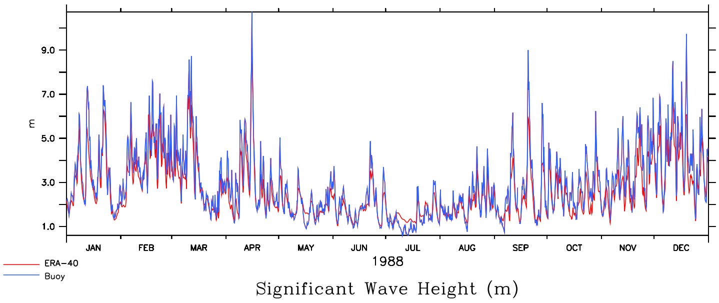

Figure 1. Measured (blue) and modelled (red) H_S at buoy 46001 (148.3° W, 56.3 ° N) in 1988.

The length of the ERA-40 data set makes it especially suitable to study variability and extremes of weather-related quantities

property of the ERA-40 wave data. Among the reasons for these deficiencies are spatial resolution (P. Janssen, pers. communication) and a slight underestimation of high wind speeds 7).

Apart from the underestimation of large wave heights ERA-40 waves also suffer from inhomogeneities due to changes in the data that were assimilated. As a synthesized picture of the data Figure 2 shows the time series of the globally averaged monthly mean H_S from ERA-40 (blue). From 1991 onwards wave height data from altimeters flown onboard satellites became available and were assimilated. The impact of these data is clearly seen, especially for the period from December 1991 to May 1993, when erroneous data were used. Another inhomogeneity is visible in 1996, when the altimeter data changed from ERS-1 to ERS-2.

Fortunately, it was possible to devise a non-parametric correction method for the ERA-40 data. A corrected dataset was created 8) which 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 (Figure 2, red). Furthermore, reliable estimates of the 100-year return wave height could be obtained from the raw ERA-40 data by a calibration against buoy measurements. These and other results from the ERA-40 wave data have been incorporated into the web-based KNMI/ERA-40 Wave Atlas (*http://www.knmi.nl/waveatlas*).

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. The ERA-40 data set has proved an invaluable basis to derive global estimates of these extremes. Note that extremes of significant wave height rather than those of individual waves are obtained. Estimates of the 100-year return significant wave height H_{100} are obtained using the Peak-Over-Threshold (POT) method 9), in which the tail of the wave height distribution is fitted to the Generalized Pareto Distribution (GPD), the limit distribution for extremes.

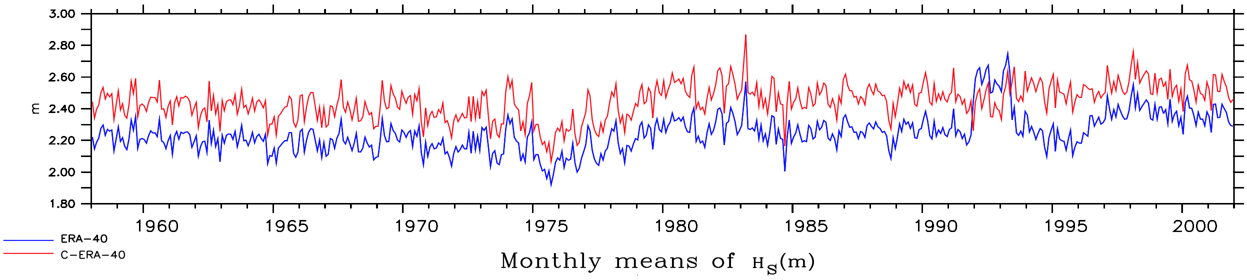

Figure 2. Time series of the monthly mean, globally averaged H_S from the raw (blue) and the corrected (red) ERA-40 data. Monthly means are computed from the 6 hourly fields.

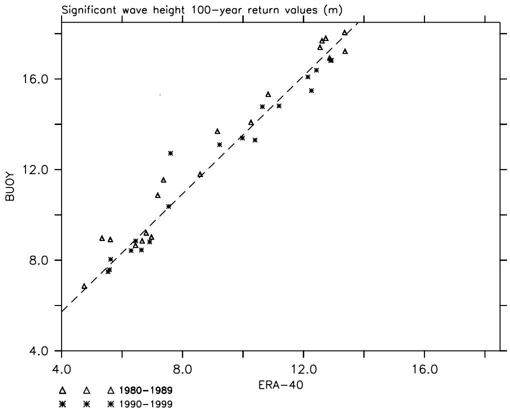

Figure 3. Linear correlation between 100 year return value estimates of H_S from buoy data and from ERA-40. The dashed line is eq. (1).

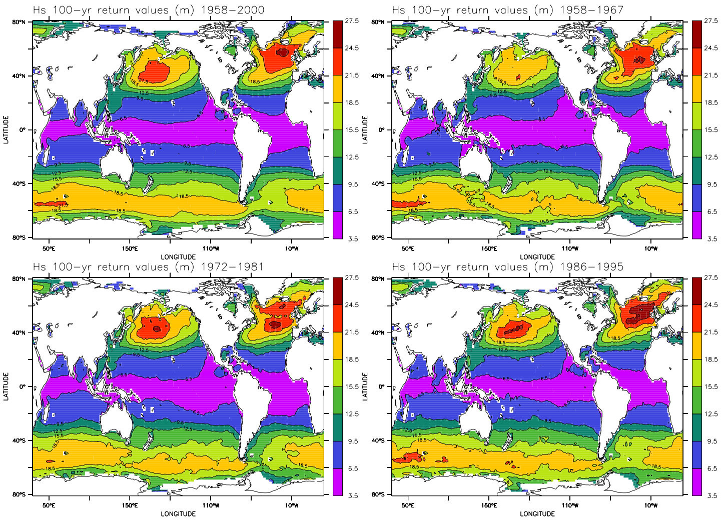

Figure 4. 100 year return H_S from ERA-40, corrected using the relationship displayed in Figure 3. Note that the results pertain to averages over 1.5° × 1.5° and 6 hours, that shallow water effects are not included, and that tropical cyclones are not resolved in ERA-40. The upper left panel is for the whole ERA-40 period (1958-2000), while the other panels are derived from three 10 year sub-periods as

Estimating H_{100} both from buoy measurements and from the raw ERA-40 data yields a linear relationship between the two [10]:

$$H_{100}(\text{buoy}) = 0{,}52 + 1{,}30\ H_{100}(\text{ERA-40}) \quad (1)$$

This relation is illustrated in Figure 3.

Buoy locations are very sparse and unevenly distributed in space, and the largest value of H_{100} found at the buoy locations is about 17 m (Figure 3). Therefore it would be preferable to have a relation between H_{100} estimates from ERA-40 and from satellites, respectively. 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 at a given location. However, as far as parts of the estimation procedure were possible with satellite data their results are not incompatible with (1). We therefore apply this equation globally and for all values of H_{100}.

Figure 4 shows the H_{100} values obtained by applying the POT method to the ERA-40 data and correcting the results using (1). It is obvious that the highest values occur in the North Atlantic. While mean wave heights are not higher in the North Atlantic than they are in the North Pacific or in the Southern Ocean (Figure 5), the North Atlantic shows the highest variability (not shown). In other words, conditions in the Southern Ocean are always rough, while 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.

Besides an estimate based on the whole ERA-40 period, Figure 4 also contains estimates of H_{100} for three different 10-year periods. The estimates obtained from these periods differ in the Northern Hemisphere storm tracks. Specifically, the estimates in the North Pacific storm track region have increased, and in the North Atlantic the pattern has changed. These differences can be attributed to decadal variability in the Northern Hemisphere, especially to changes in the phase of the NAO[8]. This example shows that it is important to take account of climate changes when designing maritime structures. A more detailed investigation reveals that in the North Atlantic changes in estimates of H_{100} are due to changes in the intensity of storms, while in the Southern Hemisphere they are mainly due to changes in the number of storms. In the North Pacific both factors contribute.

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 and the ocean surface four times a day. A subset of the raw

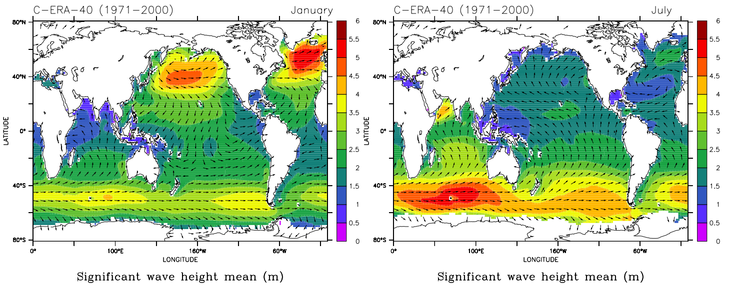

Figure 5. January (left panel) and July (right panel) averages of HS (colours) and propagation direction of waves (arrows) for the whole ERA-40 period.

ERA-40 data can be obtained 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. A non-parametric correction method was devised that eliminates most of these problems.

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 reveal a linear relationship that can be exploited to obtain global reliable return value estimates based on the ERA-40 data. Maps of the ERA-40 data and derived quantities can be found in the web-based KNMI/ERA-40 Wave Atlas (*http://www.knmi.nl/waveatlas*).

Acknowledgments
We are indebted to many persons 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 where always open to our comments and provided valuable help in dealing with the technical aspects of the ERA-40 system. Helen Snaith helped with the altimeter data. The buoy data were obtained from NDBC-NOAA (http://www.nodc.noaa.gov/BUOY/buoy.html). This work was funded by EU as part of the ERA-40 project (no. EVK2-CT-1999-00027).

---

1) Sterl, A., 2004. *On the (in-)homogeneity of reanalysis products* . J. Climate, **17**, 3866-3837.
2) Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, M.I. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, A. Leetmaa, R. Reynolds, R. Jenne and D. Joseph, 1996. *The NCEP/NCAR 40-year reanalysis project* . Bull. Amer. Meteor. Soc., **77**, 437-471.
3) Gibson, R., P. Kållberg, S. Uppala, A. Nomura, A. Hernandez and E. Serrano, 1997. *ERA description*. ECMWF Re-Analysis Project Report Series, **1**, ECMWF, Reading, UK, 73 pp.
4) Janssen, P.A.E.M., 1989. *Wave-induced stress and the drag of air flow over sea waves* . J. Phys. Oceanogr., **19**, 745-754.
5) Janssen, P.A.E.M., 1991. *Quasi-linear theory of wind wave generation applied to wave forecasting* . J. Phys. Oceanogr., **21**, 1631-1642.
6) Komen, G.J., L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann and P.A.E.M. Janssen, 1994. *Dynamics and Modelling of Ocean Waves* . Cambridge University Press, Cambridge, UK, 532 pp.
7) Caires, S. and A. Sterl, 2003. *Validation of ocean wind and wave data using triple collocation* . J. Geophys. Res., **108**(C3), 3098, doi:10.1029/2002JC001491.
8) Caires, S. and A. Sterl, 2005. *Validation and non-parametric correction to significant wave height data from the ERA-40 reanalysis*. J. Atmos. Oceanic. Technol., **22,** 443-459.
9) Coles, S., 2001. *An introduction to statistical modeling of extreme values* . Springer Texts in Statistics, Springer-Verlag, UK, 228 pp.
10) Caires, S. and A. Sterl, 2005. *100-year return value estimates for wind speed and significant wave height from the ERA-40 data* . J. Clim., **18,** 1032-1048.