Journal of Coastal Research	SI 50	pg - pg	ICS2007 (Proceedings)	Australia	ISSN

Application of a POT model to estimate the extreme significant wave height levels around the Balearic Sea (Western Mediterranean)

B. Cañellas†, A. Orfila†, F. J. Méndez‡, M. Menéndez‡, and J. Tintoré†

† IMEDEA (CSIC-. UIB) 070190, Illes Balears, Spain

‡ Ocean and Coastal Research Group. Universidad de Cantabria, 39005, Santander, Spain

ABSTRACT

CAÑELLAS, B.; ORFILA, A.; MÉNDEZ, F.J.; MENÉNDEZ, M. and TINTORÉ, J., 2007. Application of a POT model to estimate the extreme significant wave height levels around the Balearic Sea (Western Mediterranean). Journal of Coastal Research, SI 50 (Proceedings of the 9th International Coastal Symposium), pg – pg. Gold Coast, Australia, ISBN

Extreme value wave climate analysis at a particular site requires predicting long-term wave height levels from short duration records. In the present work we used the Peak Over Threshold (POT) model, assuming the frequency as a Poisson process and the intensity to be Pareto distributed, to characterize the spatial variability of the long-term extreme value wave climate along the Balearic Sea. Wave data used is part of the HIPOCAS database, a 44-years high resolution, spatial and temporal, wave hindcast, covering an area between 38ºN-42ºN and 1ºW-6ºE, of the western Mediterranean Sea. The use of data from a homogeneous grid, instead of a single location wave data record, allows describing the spatial variability of the long-term extreme wave height levels, over the whole Balearic Basin. Results show that extreme values for a 50-year return period level around 11 m are found in the north sector of the Balearic Islands while in the southern part much lower extreme values are found. This is due to the shadow effect of the islands over the severe north-eastern storms.

ADITIONAL INDEX WORDS: *extreme wave climate, HIPOCAS, spatial variability.*

INTRODUCTION

Detailed assessment of wave climate is a previous requirement for all human activities in the coastal zone. Beach nourishment, port design and operability, dispersion and diffusion of pollutants are some examples that require a precise knowledge of the long-term distribution of significant wave height, H_s, and mean period, T_m, as well as the long-term extreme value distribution. On the other hand, wave climate analysis requires a large amount of data to ensure the statistical significance. These data have been collected in the last decades using scalar and directional wave buoys moored at specific locations providing high temporal resolution records. In the last decade, satellites have been used to overcome the spatial lack of data (Krogstad and Barstow, 1999) but the problem of having a large amount of spatial and temporal wave records were still unresolved.

Queffeulou (2005) used altimeter data to perform an analysis of the wave height variability over the Mediterranean Sea but as noted for some authors, altimeter data has as a shortcoming its temporal inhomogeneity and a coarse spatial resolution in areas like the Western Mediterranean, marked by a complex orography.

Statistical analysis of wave climate has been thus, carried out with relatively short data sets, e.g. 10 years of data for the satellite altimeters (TOPEX,ERS-1/2) used in the analysis carried out by Queffeulou (2005).

Alternatively, wave generation models are another option to avoid the usual lack of data in ocean and atmospheric studies. Models are initialized with real conditions and the deviation due to the nonlinearity of the governing equations corrected with the assimilation of data. Numerical models can be now implemented in very fine grids. These hindcast models have become a powerful tool not only for engineering or prediction scales but for climate studies involving large temporal periods. The 44 years of hourly wave data base with 0.125º spatial resolution obtained from HIPOCAS project (Soares et al. 2002) in the Western Mediterranean is used in the present work.

In order to characterize the long-term extreme value distribution of significant wave height in the Balearic Sea, we use the Peak Over Threshold (POT) method, which is widely used in the definition of the extreme behaviour of severe storms. The spatial variability of the extreme wave events is obtained determining the 50-year return period quantile in every node.

The paper is structured as follows. In the first section, we present the wave data as well as the POT model. The next section deals with the estimation of the extreme value return levels. The 50-year return period levels over the Balearic Sea is presented and discussed in the third section. Finally we conclude the work.

DATA AND METHODOLOGY

The HIPOCAS data

Wave data is part of the HIPOCAS Project (Hindcast of Dynamic Processes of the Ocean and Coastal Areas of Europe).This database consists on a high resolution, spatial and temporal, long-term hindcasted data set (Soares et al. 2002). This reanalysis, covers in an hourly basis a period ranging from 1958 to 2001 providing a 44-years of wave data over an homogeneous grid. This dataset was produced by means of dynamical downscaling from the NCEP/NCAR global reanalysis using the regional atmospheric model REMO. Hourly wind fields from the REMO (U_{10}) were used as forcing for the third generation wave model WAM. As a result, wave data used are the output of the WAM model implemented in a 1/8º resolution mesh over the western Mediterranean Sea. In this work we cover 1387 nodes in an area between 38ºN-42ºN and 1ºW-6ºE (Figure 1). This dataset provides the opportunity to perform a significant analysis of return period levels at deep waters and their spatial distribution, that can help to understand the severity of the storms in this particular area.

The Generalized Extreme Value and Generalized Pareto Distribution

The classical approach to perform an extreme value analysis is to fit the annual maxima values with the Generalized Extreme Value (GEV) cumulative distribution function,

$$G(x) = \exp\left\{ -\left(1+\xi\frac{x-\mu}{\psi}\right)_{+}^{-1/\xi} \right\} \quad (1)$$

where μ is the location parameter, $\psi > 0$ is the so-called scale parameter and ξ is a shape parameter which determines the tail of the distribution. When $\xi \approx 0$ the GEV distribution corresponds to the Gumbel family, conversely for $\xi > 0$ the Fréchet form is adopted and for $\xi < 0$ the Weibull form is adopted. The annual maxima method developed by GUMBEL (1960) considers only the largest value for each year. There is some criticism in the use of this approach, because using only the maximum value per year leads to the loss of information contained in other large–sample values for a given period (CASTILLO, 1997).

To solve the problem of working only with a data per year the Generalized Pareto Distribution (GPD) was introduced (PICKANDS, 1975). The GPD method models all values larger than a given threshold u. The differences between these values and the threshold u are called exceedances over the threshold and it is assumed to follow a GPD(σ, ξ) distribution whose Cumulative Distribution Function is defined by,

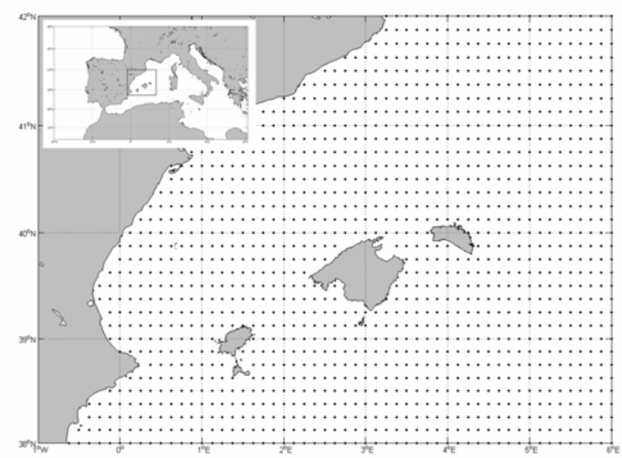

Figure 1. Geographic location of the study area and the 1/8º resolution HIPOCAS grid over the Balearic Sea.

$$G(y;\sigma,\xi) = \begin{cases} 1-(1+\xi y/\sigma)^{-1/\xi}; \xi \neq 0, \sigma > 0 \\ 1-\exp(-y/\sigma); \xi = 0, \sigma > 0 \end{cases} \quad (2)$$

where σ>0 is the location parameter, $-\infty<\xi<\infty$ is the shape parameter and y are the exceedances over the threshold u ($y = x-u>0$).

The Poisson-GPD model

A modification to the model defined in Eq.(2) is the Poisson-GPD model for exceedances, originally developed by hydrologists, which is closely related to the Peaks Over Threshold (POT) method. This model is a joint distribution, the GPD, for the exceedances values y, and a Poisson distribution for the number of exceedances over a level u in any given year. With this model, one can estimate not only the intensity of the exceedances but also the frequency of these events.

Therefore we assume that the number, N, of exceedances of the level u in any one year has a Poisson distribution with mean λ, and the exceedances $\{y_i\}_{i=1}^{N}$ are independent and identically distributed from the GPD.

Under these hypothesis, the probability that the annual maximum of the GPD-P process is lower than a value x, with $x>u$, is given by,

$$F(x) = \exp\left\{ -\lambda\left(1+\xi\frac{x-u}{\sigma}\right)_{+}^{-1/\xi} \right\} \quad (3)$$

where the Poisson parameter λ, the scale parameter σ and the shape parameter, ξ are to be determined.

Parameter estimation

The GPD-P model reduces to the determination of the three unknown parameters, λ>0, σ>0 and $-\infty<\xi<\infty$. The scale and shape parameters arise from the GPD (PICKANDS, 1975) and λ from the Poisson distribution. These three parameters are estimated using the Maximum Likelihood Method (MLM). The Maximum Likelihood Estimators are the values of the unknown parameters that maximise the log-likelihood function. In practise these are local maxima found by nonlinear optimization.

The log-likelihood function for the GPD-P, if N exceedances are observed over a τ-year period is given by,

$$l(y;\lambda,\sigma,\xi) = N\log\lambda - \lambda\tau + \log\left(\frac{1}{\sigma^N}\right) + \log\prod_{i=1}^{N}\left(1+\xi\frac{y_i}{\sigma}\right)^{\frac{-1-\xi}{\xi}} \quad (4)$$

Maximizing $l(y;\theta_k)$ respect to $\theta = (\lambda, \sigma, \xi)$ in the GPD-P leads to the maximum likelihood estimate $\hat{\theta} = \left(\hat{\lambda}, \hat{\sigma}, \hat{\xi}\right)$.

On the other hand a useful relation between the GEV and GPD-P parameters is found in literature (SMITH, 2003),

$$\sigma = \psi + \xi(u-\mu) \qquad \lambda = \left(1+\xi\frac{u-\mu}{\psi}\right)^{-1/\xi} \quad (5)$$

Then through a simple parameter substitution (eq.5 into eq.4) we could express the model in terms of the GEV parameters (μ, ψ, ξ) and, consequently, fit the GPD-P model.

Threshold selection and time span for the GPD-P

An important issue when modelling threshold excesses with the GPD-P is to choose correctly the threshold value u and the minimum time span Δt between successive extreme events. Then the extreme events are identified by considering all values larger than a given threshold u and with a minimum time span Δt between the storms, to ensure the meteorological independence of the observed excesses. It is not an easy aspect because it requires a balance between bias and variance caused by the selected threshold. If the selected threshold is too low we will violate the asymptotic basis of the model, causing bias. On the other hand a too high threshold will produce few excesses over the selected threshold causing a high variance in the estimated values (MENDEZ et al., 2006).
Some tools are available to choose the "correct" threshold. For example an *aprioristic* test like the mean excess plot (COLES, 2001) can be used (Figure 2). This test leads to a quick estimation of the shape parameter ξ. Assuming that Y follows a GPD distribution, the mean excess over threshold u, is a linear function of u with slope $\xi/(1-\xi)$.
However, sometimes, the mean excess plot can be difficult to interpret, making the decision subjective (COLES, 2001). As an example, in Figure 2, a linear tendency for a threshold above u>5 meters is observed at the HIPOCAS node 1193, located at the north of the Minorca island, but a more precise value is difficult to obtain.
To avoid the subjectivity in the threshold selection, we use an alternative diagnostic method known as the W-statistic plot (SMITH, 2003). The W-statistic is defined as,

$$W_i = \frac{1}{\xi}\log\left\{1 + \frac{\xi Y_i}{\psi + \xi(u-\mu)}\right\} \qquad (6)$$

This method is based on, if all assumptions are correct, including the selected threshold u and the time span Δt, then W_i are also independent and exponentially distributed variables with mean 1. Figure 3 shows the quantile plot (QQ-plot) for the W-statistic at the same HIPOCAS node, for the selected threshold u = 5.2 meters and Δt = 72 hours. As seen in this figure, expected values for W are close to the observed ones with a slope near the unit diagonal, indicating the suitability of the selected parameters u and Δt.

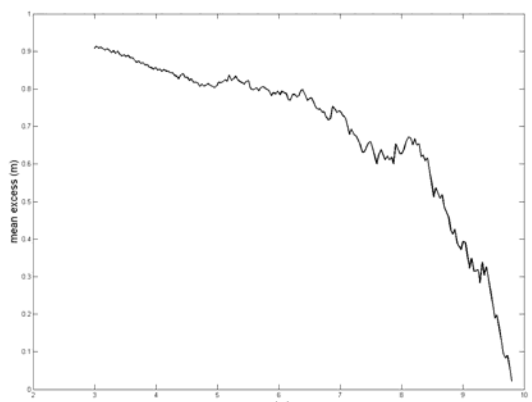

Figure 2. Mean excess plot for different thresholds values.

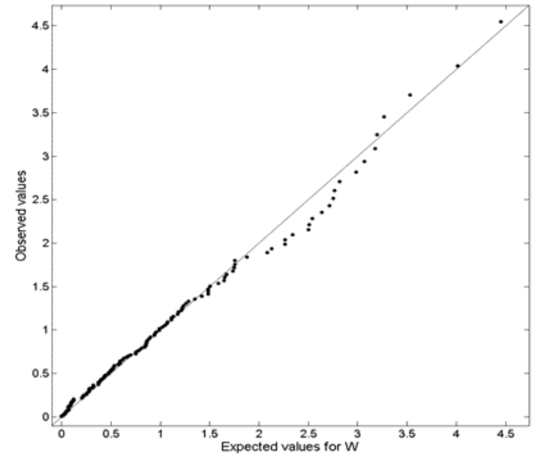

Figure 3. W statistic quantile plot for u=5.2 meters and a 72 hours time span at the HIPOCAS node 1193.

The W-statistic was applied at each grid point for different time spans, between 12 and 144 hours. Finally, a time span Δt = 72 hours was selected for the whole area and a value for the threshold u corresponding to the 99.5% percentile of the empirical distribution was chosen at each grid point.

Model selection

The selection of the simplest possible model that fits the data sufficiently well is important. Therefore, we check for every point if the contribution of the shape parameter ξ is statistically significant. This is performed using the likelihood ratio test (COLES, 2001). With nested models $M_A \subset M_B$ (M_B including the shape parameter and M_A with the shape parameter $\xi = 0$), we can assure that model M_B explains substantially (at the α -level of significance) more variability in the data than M_A if $2[l_B(M_B) - l_A(M_A)] > \chi^2_{k,1-\alpha}$, where $l_B(M_B)$ and $l_A(M_A)$ are the maximized log-likelihood functions under models M_B and M_A, respectively, and $\chi^2_{k,1-\alpha}$ is the 1-α quantile of the χ^2 distribution with k degrees of freedom.

Figure 4(a) shows the estimated shape parameter for the whole area and 4(b) shows the statistical significance of the inclusion of the shape parameter. As see, we can distinguish three different areas, depending on the value of the shape parameter ξ. The dark color in Figure 4(b) reveals that for this area, it is not significant the inclusion of the shape parameter (which corresponds to the Gumbel family for the GEV distribution or the exponential for the GPD distribution). The western area corresponds with a Weibull tail ($\xi < 0$) and the extreme waves in the area around Spain has a Frechet tail ($\xi > 0$). Therefore, for the area where the inclusion of the shape parameter ξ is not significant we fix the value ($\xi = 0$).

Return levels for Hs_{50}

The *N*-year return level is the average time interval in years between successive events of an extreme significant wave height being equalled or exceeded. So, the probability that H_s will be exceeded in any given year is,

$$F(H_s \geq x) = 1 - \frac{1}{N} \tag{7}$$

being *F(x)* the values of a cumulative probability distribution function and *N* the return period. Therefore a fifty-year return period is equivalent to $F(H_s \geq x) \approx 0.98$.
For the 1387 HIPOCAS grid points the Hs_{50} is calculated for the GPD-P as,

$$\hat{Hs}_N = u - \frac{\left(\hat{\sigma}\hat{\lambda}\right)^{\hat{\xi}}}{\hat{\xi}\ln\left(1-1/N\right)^{\hat{\xi}}} - \frac{\hat{\sigma}}{\hat{\xi}} \tag{8}$$

Results for the 50-year return period significant wave height are shown in Figure (5). These values are around 11 meters in the northern quadrant of the islands while in the southern part are less than 8 meters. This is the result of the shadow effect of the islands over the intense north fetch produced by the storms. Extreme wave heights over the Catalan coast are significantly lower than those obtained in the north of the Islands due to the angular spreading of the more severe storms.

DISCUSSION AND CONCLUSIONS

Deep water wave climate over the Balearic Sea has in general a complex pattern as a result of the complex orography of the surrounding area. The Mediterranean Sea is well known to be one of the most active cyclogenesis areas in the world where the climate is mainly conditioned by severe atmospheric forcing during winters. The mountains range in the vicinity is a key factor controlling the storm track. The role of the Pyrenees in the west part and the Alps in the northeast area are decisive boundaries for the wind and pressure distribution over the whole basin. The north western part and central part of the Balearic Sea are forced by northerly winds (mistral) during the main part of the year, while the eastern part is generally modulated by a seasonal variability.

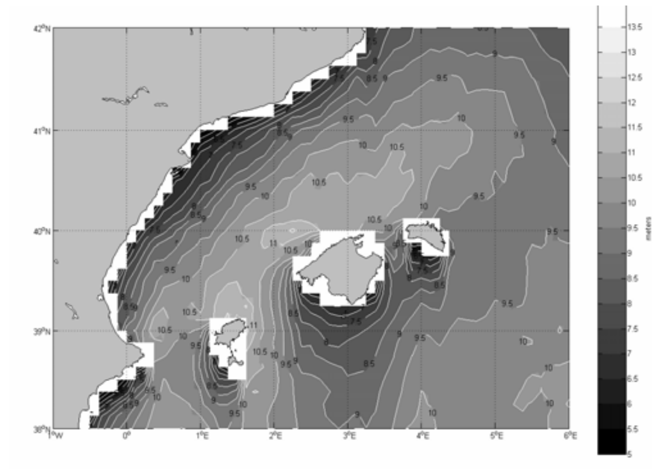

Figure 5. Spatial distribution of Hs50

Gale forced mistrals often develop over the Gulf of Genova when the passage of the 500 mb thought cross the south eastern part of France extending the effects over the whole basin.
In order to have a rough idea of the behaviour of the storms, the intensity and direction of the maximum significant wave height for the 44 years data are shown in Figure 6. As seen, prevailing directions are from the northeast along the Balearic Sea. This result was previously observed by Sotillo et al. (2006) where a high wind area was identified along the Western Mediterranean from the Gulf of Lions to Northern Algeria and Tunisia. Wind speeds for a 100 years return period shows a maximum located in the Gulf of Lions with levels of winds up to 30m/s. In the eastern part, differences in wave directions are obtained as a result of the different storm track pattern over this area. As seen, maximum significant wave heights are reached in the Balearic Channel as was obtained from the GPD-P distribution.

From the analysis we can conclude that the GPD-P distribution provides a good estimation for wave climate analysis. The analysis performed provided the spatial variability of the 50-year return period significant wave height. The shadow effect of the islands and the angular spreading of the storms produce a reduction in the magnitude of the higher return levels.
Results also show a spatial variability of the tail of the extreme

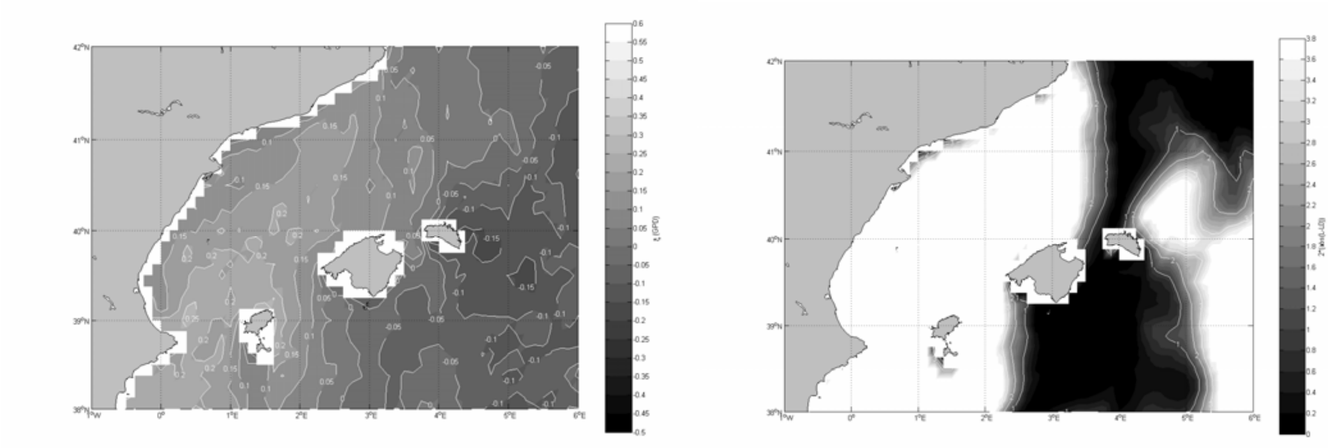

Figure 4. Spatial distribution of (a) shape parameter ξ and (b) statistical significance of the shape parameter ξ

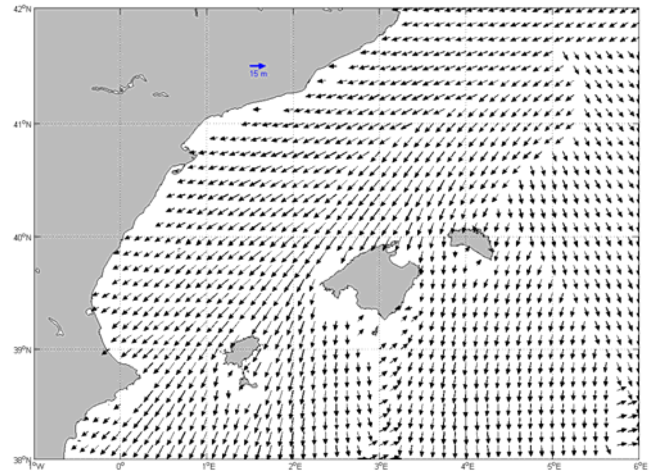

Figure 6. Intensity and direction for the highest value at each point during the 44 years data.

value distribution: in the western area a bounded tail (Weibull) is detected. This can be associated to a homogeneous distribution in the intensity of the extreme events. On the other hand, along the Spanish littoral, the extreme waves tend to be heavy tail distributed (Frechet). This aspect can be related on the two main storms that affected this area in September 1983 and November 2001.

The use of the data from the HIPOCAS reanalysis provides a powerful tool for the estimation of the extreme events for a risk analysis in the western Mediterranean Sea.

LITERATURE CITED

- CASTILLO, E. and HADI, A.S., 1997. Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association*, 92, 1609-1620.
- COLES, S., 2001. *An introduction to statistical Modeling of Extreme Values*. London: Springer-Verlag, 208 p.
- DAVISON, A.C. and SMITH, R.L., 1990. Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society*, B 52, 393-442.
- DUAN, Q.; SOROOSHIAN, S., and GUPTA, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28, 1015-1031.
- GUMBEL, E.J., 1958. *Statistics of Extremes*. New York: Columbia University Press, 375 p.
- KROGSTAD, H.E.; BARSTOW, S., 1999. Satellite wave measurements for coastal applications. *Coastal Engineering*, 37, 283-307.
- MÉNDEZ, F. J.; MENÉNDEZ, M; LUCEÑO, A., and LOSADA, I. J., 2006. Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model. *Journal of Geophysical Research, Volume 111, Issue C7*.
- PICKANDS, J., 1975. Statistical inference using extreme order statistics. *Annals of Statistics*, 3,119-131.
- QUEFFEULOU, P., 2005. Wave height variability over the Mediterranean Sea, using altimeter data. *Proceedings of the 5th International Symposium on Ocean Wave Measurements and Analysis* (Madrid, Spain).
- SMITH, R.L., 2003. Statistics of extremes, with applications in environment, insurance and finance. Chappel Hill, North Carolina: University of North Carolina, Department of Statistics.
- SOARES, C.G.; CARRETERO, J.C.; WEISSE, R., and ÁLAVAREZ, E., 2002. A 40 years hindcast of wind, sea-level and waves in European waters. *Proceedings of OMAE 2002: 21st international conference on offshore mechanics and arctic engineering* (Oslo, Norway).
- SOTILLO, M.G.; AZNAR, R., and VALERO, F., 2006. Mediterranean offshore extreme wind analysis from the 44-year HIPOCAS database: different approaches towards the estimation of return periods and levels of extreme values. *Advances in Geosciences*, 7, 275-278.

ACKNOWLEDGEMENTS

The authors thank to Puertos del Estado (Ministerio de Fomento) for the use of HIPOCAS data base. The authors would also like to thank financial support from Govern Balear Project (UGIZC). AO would like to thank financial aid from CSIC (2006-3-01-068). FJM and MM are indebted to the Spanish Ministry of Education and Science for the funding provided in the “Ramon y Cajal” Program and FPI program, respectively. FJM and MM acknowledge the support of the Spanish Dirección General de Investigación under grants ENE2004-08172 and CGL2005-05365/CLI.