Research Article

Mediterranean Marine Science
Volume 7/1, 2006, 47-62

The Errors-in-Variables approach for the validation of the WAM wave model in the Aegean Sea

T. H. SOUKISSIAN and A. M. PROSPATHOPOULOS

Hellenic Center for Marine Research, Institute of Oceanography
P.O. Box 712, 190 13, Anavissos, Attica, Greece

e-mail: tsouki@ath.hcmr.gr

Abstract

In previous studies, wave fields from the 3rd-generation wave model WAM-Cycle 4 have been validated by using in situ buoy measurements in the Aegean Sea within the framework of the POSEIDON project; however, limitations of the data sets, concerning mainly the short distance of the buoys from the shore and the short length of the data, render those validation studies incomplete. In this work, significant wave height forecasts obtained from WAM-Cycle 4 wave model are validated by means of TOPEX/Poseidon (T/P) data in specific offshore locations in the central part of the North Aegean Sea. The linear structural relationship between the two data sets has been modelled by implementing the Error-In-Variables approach, assuming that both T/P data and WAM results are subjected to errors. The underestimation of significant wave height from WAM, which has been concluded from the comparison with buoys at near-shore points, is also observed from the WAM-T/P comparison at offshore locations, thus being considered of general validity for the Aegean Sea. In addition, a correction relation for the WAM model results, based on the linear structural relationship, is proposed and applied.

Keywords: Structural relation; Regression analysis; WAM model; TOPEX/Poseidon altimeter; Wave climate; Aegean Sea; POSEIDON system.

Introduction

In order to effectively deal with subtle applications such as those related to wave climate assessment and variability or operational wave forecasting, data from various sources are frequently required: visual data, remote-sensing data, buoy data and numerical model data (GULEV *et al.*, 1998; KROGSTAD *et al.*, 1999). Visual observations have been proved very useful in the past, but due to their significant objectiveness and rapid technological developments they have been gradually replaced by data obtained from the other sources. Buoy data are used as reference for reasons of accuracy, but their sparseness intimates the need for denser data, obtained from remote-sensing sources. Remote-sensing data, although of high quality, are often calibrated with respect to buoy measurements (CARTER *et al.*, 1992; COTTON & CARTER, 1994; GOWER, 1996; BARSTOW *et al.*, 1997; YOUNG, 1999; RAY & BECKLEY, 2003; KECHRIS & SOUKISSIAN, 2004). Numerical wave models could cover various geographical areas extending from a small basin to the World Ocean and provide systematic time series of spectral parameters in various spatial and temporal resolutions. However, they suffer from inaccuracies and uncertainties due to errors in model parameterization, initial conditions and forcing terms (i.e., incorrect wind fields), grid discretization, as well as to physical uncertainties of the wave model itself (LIU *et al.*, 2002; BABOVIC *et al.*, 2005; KOBAYASHI & YASUDA, 2004). For this reason, considerable work has been published on the improvement (correction/calibration) of wave model data with respect to buoy measurements and/or satellite data (KOBAYASHI & YASUDA, 2004; SANNASIRAJ *et al.*, 2005; BABOVIC *et al.*, 2005).

As concerns the Hellenic Seas, buoy measurements are limited to a few locations at distances of up to 6 nautical miles from the nearest shore; on top of that, the Aegean Sea is scattered with many islands, rendering the wave propagation patterns very complicated. For this reason, the measured wave data do not always reflect realistic offshore wave conditions. On the other hand, satellite altimetry data cover wide geographical areas, though the corresponding spatial and temporal coverage (strictly connected with the orbit of each satellite) are not always convenient or appropriate for applications in small, closed basins such as the Aegean Sea (KROGSTAD & BARSTOW, 1999). For example, TOPEX/Poseidon (T/P) and its successor Jason-I collect measurements from the same path only every ten days.

An inter-comparison of WAM model results with in-situ buoy wave measurements is presented in SOUKISSIAN & PROSPATHOPOULOS (2003). The comparisons and the obtained conclusions as regards the performance of WAM are valid for near-shore locations where the effects of coastal morphology and bottom topography are of significant importance. Based on that analysis, a major feature of WAM is the systematic underestimation of the high sea-states.

A tool used quite often for revealing the relationship of wave model results to measured wave data is classical regression analysis, which produces a linear relationship between the variables under consideration (SARKAR *et al.*, 1997; MONBALIU *et al.*, 1999; BIDLOT & HOLT 1999; KOBAYASHI & YASUDA, 2004; SANNASIRAJ *et al.*, 2005, BABOVIC *et al.*, 2005). A common misuse of classical regression analysis is the omission of the fact that it requires the independent variable (stochastic or not) to be measured without error, a condition that is very rarely met in practice. For this reason, a more general methodology should be looked for.

In the present work, validation of WAM results is performed in an offshore area of the Aegean Sea. To this end, the WAM results are compared to T/P data along its unique track (No 33) above the North Aegean Sea. Instead of applying classical regression analysis, a more general method, the so-called Errors-in-Variables (EIV) leading to structural relations between variables, is elaborated and implemented for the validation of WAM data. This method permits both variables to be measured with an error, which is actually our case, since the T/P significant wave height is a variable subject to measurement errors (BARSTOW *et al.*, 1997; CAIRES & STERL, 2003; RAY & BECKLEY, 2003). The results of the present work together with results from previous studies concerning WAM-buoy data inter-comparison provide more complete knowledge for the assessment of the wave climate of the North Aegean and the forecast capabilities of the WAM model. Furthermore, it provides evidence for the entire Aegean Sea, given that WAM is validated at a series of other offshore locations.

Materials and Methods

Sources of wave data

a. Satellite data

The satellite data set, recorded during the time period from 1/1993 to 9/2002, was derived from TOPEX/Poseidon. The information about the mission of T/P is gathered from the official JPL website (www.jpl.nasa.gov) and the official eoPortal website (www.eoportal.org). From the collected T/P data – obtained from the Merged Geophysical Data Records (GDR-M), processed and provided by AVISO (AVISO/ALTIMETRY, 1996) – wave measurements with a recording interval of 10 days along the T/P track no. 33 were extracted. Three points were selected from a segment of the track, lying in the central part of the North Aegean Sea, as being the more representative of the prevailing wave conditions; (Fig. 1). The distance between points A and B is about 18 km and between points B and C about 24 km. In order to collocate the altimetry data with the results obtained from the WAM model, the Ku band data derived from the GDR-M products were examined and records were filtered out using the following criteria:

C1. Values of $H_{m_0,T} > 11$ or $H_{m_0,T} < 0$ (suggested in AVISO/ALTIMETRY (1996));

C2. Values of $H_{m_0,T}$ that differ significantly from neighbouring points along the track,

where $H_{m_0,T}$ denotes the significant wave height, as obtained from T/P. The data obtained from T/P will be called hereafter '**T data**'.

b. WAM model data

The model data set of this work consisted of forecasts from WAM-Cycle 4 during the period 5/2000–12/2002. The 3rd-generation wave model WAM has operated in the Hellenic Centre for Marine Research (HCMR) since 1999 and provides daily wave forecasts

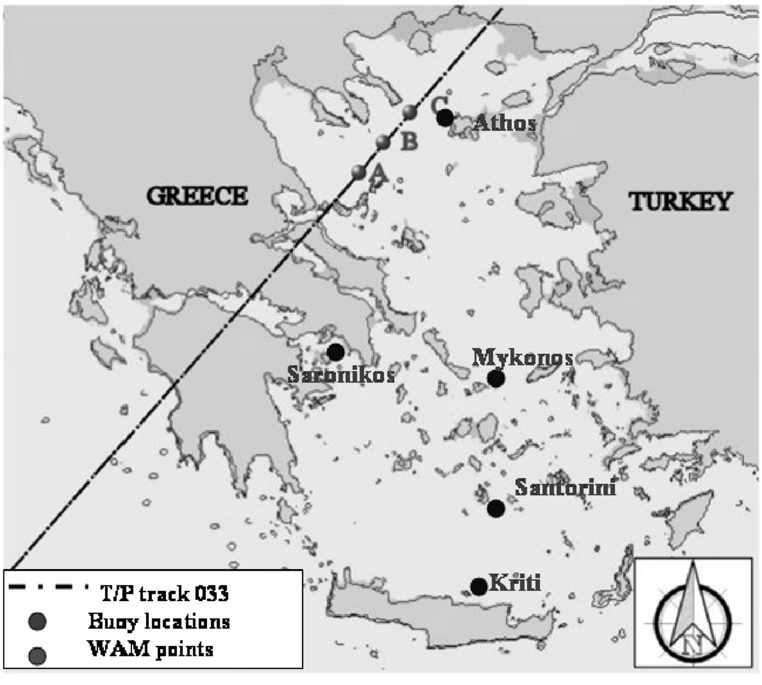

Fig. 1: The T/P track no. 33, the points selected for comparison of altimetry data with WAM results and the buoy locations in the Aegean Sea.

for the Aegean and East Ionian Sea. The complete theory, on which WAM is based, is described in detail in WAMDI GROUP (1988), while the physics and the numerical schemes used by WAM-cycle 4 can be found in KOMEN *et al*. (1994). In the implementation presented herein, WAM is fed by the wind fields from a weather forecasting model, based on the SKIRON system, developed at the University of Athens (KALLOS *et al*., 1997). The meteorological model is executed in two cycles per day: i) the coarse one (resolution 0.24^0 or about 23km) with area of application 24.2^0W–51.8^0E and 12.9^0N–53.4^0N, and NCEP initial and boundary conditions; ii) the results of this version provide initial conditions for the finer-grid cycle (resolution 0.10^0 or about 10km) with area of application 2.6^0E–38.4^0E and 27.4^0N–49.5^0N and boundary conditions updated every one hour. WAM is also executed in two steps: i) the whole Mediterranean (5.75^0W–36.25^0E, 30.25^0N–46^0N) with resolution 0.25^0, and ii) the Aegean (20^0E–29^0E, 34^0N–41^0N) with resolution 0.05^0, using the high-resolution wind field (with spatial interpolation from 0.10^o to 0.05^o) and as boundary conditions the results of the Mediterranean version, updated every three hours. The propagation time step is 720 sec for the Mediterranean version and 180 sec for the Aegean. For both cases 24 discrete directions are considered, and the estimation of spectral density is performed within the range 0.05054– 0.66264 Hz, which is discretized logarithmically into 28 frequencies. More details concerning how WAM is implemented and relative results can be found in SOUKISSIAN & PROSPATHOPOULOS (2003) and SOUKISSIAN *et al.* (2001). From a 72-hour daily forecasting period obtained regularly from the WAM model, only the first 24-hour forecasts (per 3-hour intervals) were taken into account for the T/P-WAM comparison. The wave data obtained from the WAM model will be called hereafter **'M data'** and the corresponding significant wave height will be denoted as $H_{m_0,M}$.

c. Buoy measured data

Within the context of the POSEIDON project the Hellenic Center for Marine Research (HCMR) operates a real time monitoring and forecasting system for the Hellenic marine environment (POSEIDON system); see SOUKISSIAN *et al.* (1999). The POSEIDON monitoring network consists of oceanographic buoys that measure meteorological, environmental and oceanographic parameters. For the validation of the obtained correction relation for WAM model forecasts, measured in-situ wave data from a buoy located near the Athos Peninsula (39.96^0 N - 24.72^0 E, water depth 220 m, see Fig. 1) will be also used. The recording interval of the measurements is 3 h and the sampling period of the free surface elevation is 1024 sec. These wave data will be called hereafter **'B data'** and the corresponding significant wave height will be denoted as $H_{m_0,B}$.

Some elements from structural relation theory

Linear regression (LR) analysis consists of a family of statistical techniques, used for making predictions and determining the corresponding bands of error, through the modelling and assessment of the linear relationship between a dependent variable (also called response) and a single variable or multiple independent variables (also called regressors or predictors). The type of regression most frequently used in practice is the so-called Simple Linear Regression (SLR) or classical regression, where the predictor is a mathematical variable and an additional error term, totally attributed to the response and assumed to be normally distributed, is involved in the statistical model. Classical regression is a special case of the more general class of linear functional/structural relationships between mathematical or random variables. In this class – known in the literature as *Errors-in-Variables* (EIV) – both regressors and response are measured with error and, in general, techniques other than ordinary least-squares should be used for estimation of the regression parameters and establishing the possibly linear relationship between the variables. EIV reduces to SLR only under specific assumptions, a fact which is widely ignored in practical problems (WEBSTER, 1997). In this paper, the more general EIV method will be described in some detail, and implemented for $H_{m_0,T}$ and $H_{m_0,M}$ in the next section.

In the ***Errors-in-Variable case*** (FULLER, 1987) it is considered that both variables X and Y are measured with error, i.e.

$$X_i=\xi_i+\delta_i \tag{1a}$$

and

$$Y_i=\eta_i+\varepsilon_i, \tag{1b}$$

where δ_i and ε_i are the measurement errors for X_i and Y_i, i=1, 2,..., n respectively. Assuming that a straight-line relationship holds for the unobserved values ξ_i and η_i, after some algebra, we have the following linear relationship between X and Y:

$$\begin{aligned} Y_i &= \beta_0+\beta_1\left(X_i-\delta_i\right)+\varepsilon_i \Leftrightarrow \\ Y_i &= \beta_0+\beta_1 X_i+\underbrace{\left(\varepsilon_i-\beta_1\delta_i\right)}_{\text{Error term}}. \end{aligned} \tag{2}$$

The above relation resembles the classical regression equation $Y=\beta_0+\beta_1 x+\varepsilon$ with the exception that the error term is slightly, but essentially, different: in the regression equation X is independent of the error term, but in Eq. 2 is not. Now if:

i) the errors δ and ε are uncorrelated amongst themselves and with each other and normally distributed with a constant zero mean value and constant variance,
ii) $Cov(\xi,\delta)=0$, i.e. the covariance of the variables ξ and δ is zero (which is typically the case),

then the ordinary least squares method for estimating β_1 results in a biased estimator of β_1; KENDALL & STUART (1961), DRAPER & SMITH (1998), p. 90. Therefore, we have to turn our attention to estimate β_0 and β_1 by the maximum likelihood method. In this case we face an identifiability problem due to additional information needed for the estimation. As suggested by BARNETT (1967) and WONG (1989), this additional information is the parameter λ, defined as the ratio of the variances of the measurement errors:

$$\lambda=\frac{\sigma_\varepsilon^2}{\sigma_\delta^2}. \tag{3}$$

In the case that λ is known, the maximum likelihood method results in the estimates

$$b_1=\frac{S_{YY}-\lambda S_{XX}+\sqrt{\left(S_{YY}-\lambda S_{XX}\right)^2+4\lambda S_{XY}^2}}{2S_{XY}} \tag{4a}$$

and

$$b_0=\overline{y}-b_1\overline{x}, \tag{4b}$$

where

$$S_{XX}=\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2, \tag{5a}$$

$$S_{XY}=S_{YX}=\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right), \tag{5b}$$

$$S_{YY}=\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2, \tag{5c}$$

and $\overline{x}$ and $\overline{y}$ denote the (sample) mean values of variables X and Y respectively.

Most usually in practical applications σ_ε^2 and σ_δ^2 are unknown. Then, some meaningful a-priori choices for λ can lead to tractable results. For λ=1 the above described procedure is quite often met in the relevant literature as *orthogonal distance regression* and the so-obtained estimators as *orthogonal estimators* (KENDALL & STUART, 1961). Let us note that ordinary linear regression aims at the minimization of the sum of the squared vertical distances between the values of the dependent variable and the corresponding values on the fitted regression line, while orthogonal regression aims at the minimization of the orthogonal distances from the observations to the fitted regression line. In this case the estimates of β_0 and β_1 are given by Eq. 4a and Eq. 4b for $\lambda = 1$. For $\lambda = S_{YY}/S_{XX}$, Eq. 4a is called the *geometric mean functional relationship* (DRAPER & SMITH, 1998, p. 92), and b_1 is simplified to

$$b_1 = \sqrt{S_{YY}/S_{XX}} \ . \tag{6}$$

An interesting property of the second case is that the estimator b_1 is the geometric mean of the slopes of the regression equations of Y on X and of X on Y; so the geometric mean is a sort of the average value between the two slopes. In both cases described above, b_0 and b_1 are biased but consistent estimators of β_0 and β_1 respectively.

Simple linear or ***classical regression*** is a special case of EIV case corresponding to $\lambda \to \infty$, which implies that $\sigma_\delta^2 \to 0$, i.e that the independent variable is measured without error. In this case the (unbiased) estimators of β_1 and β_0 are calculated by ordinary least-squares and given by the following relationships:

$$b_1 = \frac{S_{XY}}{S_{XX}}, \tag{7a}$$

$$b_0 = \bar{y} - b_1\bar{x} \tag{7b}$$

Thus, the obtained statistical model is

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \tag{8a}$$

which after estimation of the regression parameters takes the form

$$\hat{y} = b_0 + b_1 x \ . \tag{8b}$$

More details on classical regression can be found in RAWLINGS *et al.* (1998), SNEDECOR & COCHRAN (1989) and DRAPER & SMITH (1998).

Results and Discussion

Statistical analysis of wave data

In this section the basic statistics of the two data sets, the T and M data will be derived. The main statistical parameters of the significant wave height data are the (sample) mean value m_H, the standard deviation s_H, the skewness g_H, the kurtosis k_H, the median M_H and the maximum value $\max_H$. In addition, the Pearson correlation coefficient r for the collocated data and the coefficient of variation CV will be also calculated. The latter parameter is the ratio of the standard

Table 1
Statistics of collocated data from WAM model (M) and TOPEX/Poseidon (T) significant wave height in the North Aegean Sea.

Parameter	M	T
N	264	
m_H	0.62954	0.94572
M_H	0.50000	0.70000
s_H	0.59876	0.78527
$\max_H$	4.80000	5.20000
g_H	4.43189	2.69357
k_H	24.40258	9.69169
CV	0.95110	0.83030
r	0.81910	

deviation to the corresponding mean value

$$CV = \frac{s}{m}, \tag{9}$$

and can be considered as a measure for comparing the relative variation between two or more samples.

After the filtering and editing procedures, two collocated – in space and time domains – data sets were produced. The maximum time lag between collocated data is 1.5 hours. The distances between the WAM grid points and the corresponding points of the T/P track are 0.005^0 (about 0.5 km) for point A, 0.008^0 (about 0.8 km) for point B and 0.0015^0 (about 1.5 km) for point C. The statistics of the collocated data, summarized in Table 1, clearly indicate that the description of the wave climate for the North Aegean Sea could exhibit variations with respect to the wave data source. More specifically, the following conclusions can be derived:

i) The collocated statistics of WAM results and T/P measurements are very different regarding all the examined parameters.

ii) The mean bias (0.31618) is high compared to the mean values $m_{HM} = 0.62954$ and $m_{HT} = 0.94572$. On the other hand, although the standard deviation of the T/P measurements is greater than the WAM standard deviation ($s_{HT} > s_{HM}$), the variability according to the coefficient of variation CV is greater for the WAM results.

iii) The correlation coefficient $r_{TM} = 0.8191$ does not indicate a very strong linear relationship between WAM model and T/P measurements, but is clearly a support for proceeding to a structural analysis between $H_{m_0,T}$ and $H_{m_0,M}$.

In Figure 2 the initial histograms of $H_{m_0,T}$ and $H_{m_0,M}$ are presented. The horizontal axis of the histograms is discretized using the relation $k = 1 + 2.2\log(N)$ (LARSON, 1983), where k denotes the number of bins and N the sample size. It is clear that the differences between the $H_{m_0,T}$ - and $H_{m_0,M}$ - populations are important, resulting in different distributions for the significant wave

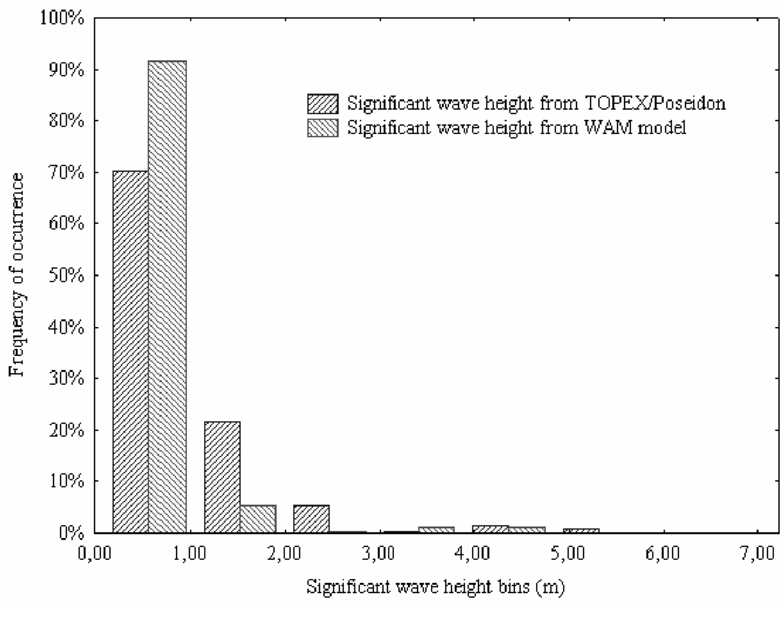

Fig. 2: Histograms of H_{m_0} obtained from TOPEX/Poseidon and WAM model results in the North Aegean Sea.

height. The main deviations are observed for $H_{m_0,T} \in [0.1, 1.0]$, where $H_{m_0,M}$ has a frequency of occurrence of about 93% compared to 70%, which corresponds to $H_{m_0,T}$. In addition, the frequency of occurrence of sea states with $H_{m_0,T} > 1.0\,m$ is underestimated by the WAM model.

In Figure 3, where $H_{m_0,T}$ and $H_{m_0,M}$ are presented in decreasing order of magnitude with respect to $H_{m_0,T}$, more clarifying information is presented: the ordered T/P significant wave height decreases quasi-exponentially from 5.5m; a systematic underestimation of the WAM forecasts (continuous black line) is evident, especially for $H_{m_0,T} > 0.5m$, possibly due to underestimated input wind fields, which affect the computation of the wave conditions (KOMEN *et al.*, 1994); for $0.25 < H_{m_0,T} < 0.5$ the model forecasts are scattered around the dotted line, while for $H_{m_0,T} < 0.25$ the values of the wave model lie almost completely above the dotted line, indicating a tendency towards overestimation in this range of values.

The aforementioned variability of $H_{m_0,M}$ around the low values of $H_{m_0,T}$ is not very important for most of the practical operational applications, in contrast with the fact that for $0.5 < H_{m_0,T} < 1$, the bulk of underestimated cases increases and, for $H_{m_0,T} > 1m$, the forecasts of WAM model are scattered below the line, in some cases presenting large deviations from it.

In SOUKISSIAN *et al.* (2001) and SOUKISSIAN & PROSPATHOPOULOS (2003) it is shown that the above - mentioned behaviour of the WAM model is also found in other locations of the Aegean Sea, where in-situ near-shore measured wave data are available. More precisely, it was found that in general the WAM model follows the trend of the sea state evolution satisfactorily, but it does not describe the sea state intensity adequately enough. Although evaluation of the accuracy of WAM in Aegean Sea implicates a number of parameters (possible calibrations applied, selected grid size, bathymetry, etc.), there is strong evidence that the most influential parameter is the input wind field (PAPADOPOULOS *et al.*, 2000; KOMEN *et*

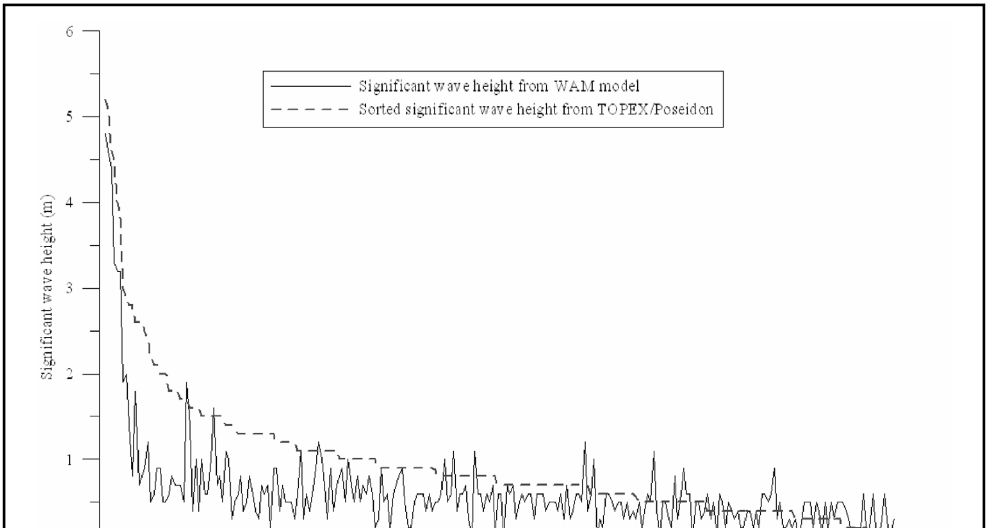

Fig. 3: Diagram of ordered $H_{m_0,T}$ and associated $H_{m_0,M}$.

al., 1994, Chapter IV; RESIO & CARDONE, 1999; SIGNELL *et al.*, 2005), among others, highlight the importance of wind field input as a factor that affects the accuracy of a wave prediction system. WAM-Cycle 4 can be reliable in open ocean forecasting (KOMEN *et al.*, 1994), but this is difficult to claim this for enclosed basins, where the effect of the surrounding topography and bathymetry play a considerable role and an adjusted limited area weather model must be used in order to diminish the errors for the wind input fields.

The above analysis implies that appropriate adjustments should be applied to the simulated wave fields from WAM in order for those to be used for quantitative wave climate analysis in enclosed basins. On the other hand, the T/P data could be used for local wave climate analysis (and under specific circumstances for extreme wave analysis) when in-situ measured data are either not available or of limited duration. The analytic validation and adjustment of the WAM results with respect to T/P data by implementing the linear structural theory is made in the following subsection.

Structural analysis of WAM forecasts and T/P measurements in the North Aegean Sea

Classical regression assumes that the regressor $H_{m_0,T}$ is measured without error. As regards the present study, the real fact is that both variables, $H_{m_0,T}$ and $H_{m_0,M}$, are 'measured' with error. More specifically:

1. TOPEX/Poseidon measurements have inherent errors, although the precise behaviour of the measurement error is not known; this fact is restrictive, since we have to adopt assumptions about it. The classical regression analysis seems to be an attractive idea, since on a worldwide basis only satellite data provide a satisfactory spatial coverage and could be accepted as being the universal wave measurement standard. This kind of approach has been followed, for example, by SARKAR *et al.* (1997), MONBALIU *et al.* (1999), BIDLOT & HOLT (1999) and YOUNG (1999) for the validation of WAM model results with significant wave height data obtained from T/P, ERS-1 or buoys. However, this procedure remains, in principle, unjustified.
2. $H_{m_0,M}$ is a variable the realizations of which include errors.

Thus, a structural relation between $H_{m_0,T}$ and $H_{m_0,M}$ should be sought for. According to Eq. (2) a linear relation between these variables, both measured with errors (EIV case), is of the following form:

$$H_{m_0,M} = \beta_0 + \beta_1 H_{m_0,T} + \left(\varepsilon_i - \beta_1 \delta_i\right), \quad (10)$$

where δ_i and ε_i are the corresponding errors of $H_{m_0,T}$ and $H_{m_0,M}$. The additional information, required in this case in order to proceed with the estimation of the parameters β_0 and β_1 using the maximum likelihood method, is the parameter λ defined as

$$\lambda = \frac{\sigma^2_{\varepsilon,M}}{\sigma^2_{\delta,T}}, \quad (11)$$

where $\sigma^2_{\varepsilon,T}$ and $\sigma^2_{\delta,M}$ are the variances of the errors of T/P measurements and WAM results, respectively. After the estimation of the parameters β_0 and β_1, the structural relation between $H_{m_0,M}$ and $H_{m_0,T}$ is

$$\hat{H}_{m_0,M} = b_0 + b_1 H_{m_0,T}\,. \quad (12)$$

Let us now investigate the above equation for the special cases λ=0, λ=1, $\lambda=S_{H_M H_M}/S_{H_T H_T}$ and $\lambda \to \infty$ (see also previous section):

1. The case λ=0 seems to have no physical meaning in the context of the present application, since it implies that $\sigma^2_{\varepsilon,M} = 0$. This case corresponds to the simple functional relationship between the variables $H_{m_0,T}$ and $H_{m_0,M}$.
2. The case λ=1 corresponds to the Orthog-

onal Distance Regression and Eq. 12 becomes

$$H_{m_0,M} = -0.05093 + 0.71953\ H_{m_0,T} \quad (13a)$$

3. The case $\lambda = S_{H_M H_M} / S_{H_T H_T}$ results in the following structural relation, called Geometric Mean Functional Relationship:

$$H_{m_0,M} = -0.09156 + 0.76249\ H_{m_0,T} \quad (13b)$$

4. Finally, the case $\lambda \to \infty$ implies that $\sigma^2_{\delta,T} \to 0$, i.e. that the T/P measurements have no error (or a constant error). This situation corresponds to the classical linear regression relationship:

$$H_{m_0,M} = 0.0389 + 0.62455 H_{m_0,T} \quad (13c)$$

Eqs. 13a, 13b and 13c are plotted in Figure 4. It is evident from this Figure that the scatter is rather large. This is probably due to the fact that there is a complete lack of values of significant wave height in the interval 3 – 4 m for TOPEX/Poseidon (roughly corresponding to 2-3 meters from the WAM model). In addition, the 3 points in the upper right-hand corner of the scatter are influential points, but we can not rationally justify (based on physical evidence) the exclusion of these points from further analysis. Another possible reason for the scatter could be the seasonal bias introduced by the points obtained from the additional 6 months period. However, exclusion of these points did not alter significantly the re sults.

Following the suggestions by BOGGS & ROGERS (1990), DRAPER & SMITH (1998) and JCOMM (2003) the case $\lambda = 1$ will be chosen for the present study. The same choice has been also adopted by RAY & BECKLEY (2003).

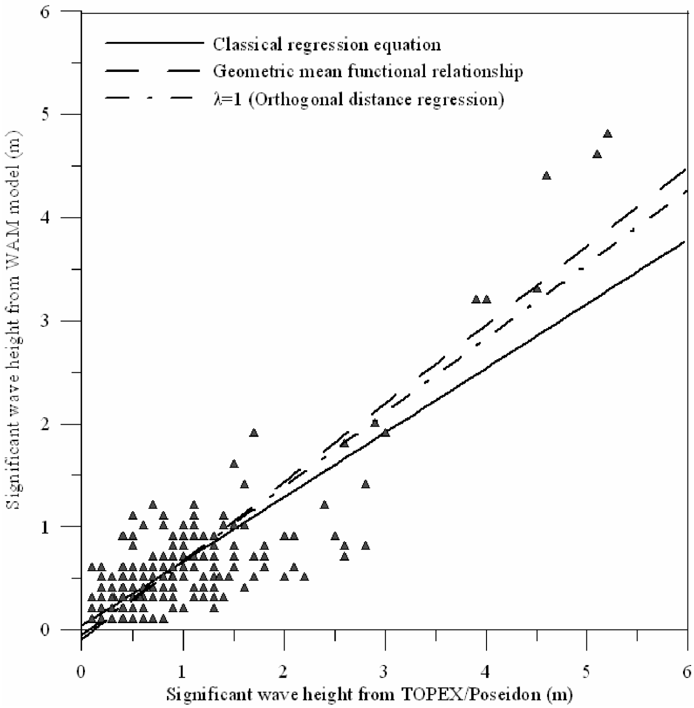

Fig. 4: Plots of classical regression, geometric mean functional and orthogonal distance regression relations between $H_{m_0,M}$ and $H_{m_0,T}$.

Rearranging Eq. 13a we obtain the correction relation[1] for WAM model results

$$\hat{H}_{m_0,M} = -\frac{b_0}{b_1} + \frac{1}{b_1} H_{m_0,M} \Leftrightarrow$$
$$\hat{H}_{m_0,M} = 0.07078 + 1.3898\, H_{m_0,M} \qquad (14)$$

where $\hat{H}_{m_0 M,}$ is the corrected significant wave height and $H_{m_0,M}$ is the significant wave height obtained directly from WAM model. The obtained histogram of $\hat{H}_{m_0 M,}$ along with the histograms of $H_{m_0,T}$ and $H_{m_0,M}$, is depicted in Figure 5.

The differences between the $H_{m_0,T}$ - and $H_{m_0,M}$-populations which were initially significant, have been now reduced: for $H_{m_0,T} \in [0.1, 1.05]$, the initial frequency of occurrence of $H_{m_0,M}$ is reduced from about 93% to 80%, while the corresponding frequency of occurrence of $H_{m_0,T}$ is 70%; for $H_{m_0,T} \in [1.05, 2.00]$, the initial frequency of occurrence of $H_{m_0,M}$ increased from about 5% to 16%, while the corresponding frequency of occurrence of $H_{m_0,T}$ is 22%; the frequency of occurrence of sea states with $H_{m_0,T} > 1.05\,m$, which was underestimated by WAM, has also been increased. In Figure 6, $H_{m_0,T}$ and $\hat{H}_{m_0 M,}$ are presented in $H_{m_0,T}$ -decreasing order of magnitude. It is obvious that the large discrepancies between $H_{m_0,T}$ and $H_{m_0,M}$, observed in Figure 3, have been smoothed out.

The proposed correction relation for WAM results in the North Aegean Sea is Eq. 14, while the corresponding correction relation for WAM results obtained from the clas-

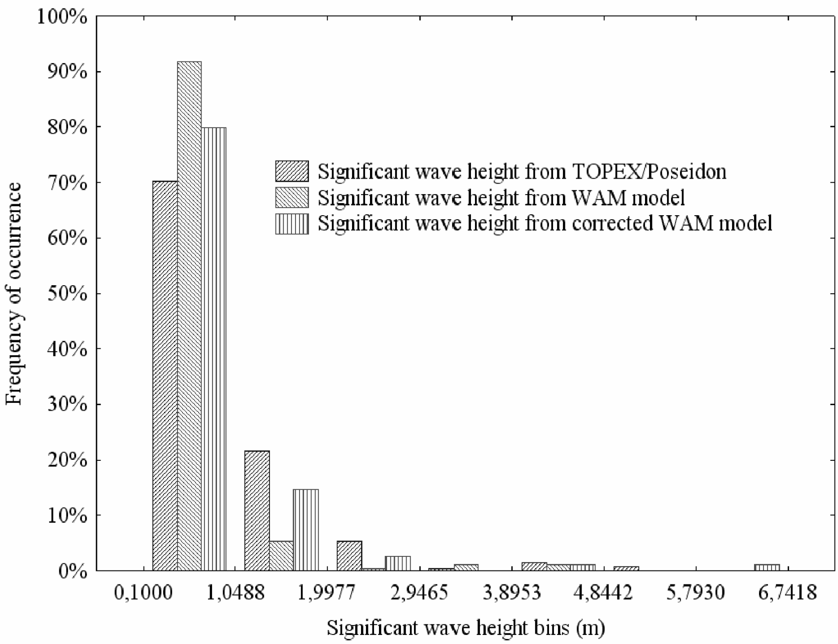

Fig. 5: Histograms of H_{m_0} obtained from TOPEX/Poseidon, WAM model results and corrected WAM model in the North Aegean Sea.

[1] *Notice that, in the case of structural relations between variables, instead of the term 'calibration relation' we use the more appropriate 'correction relation'.*

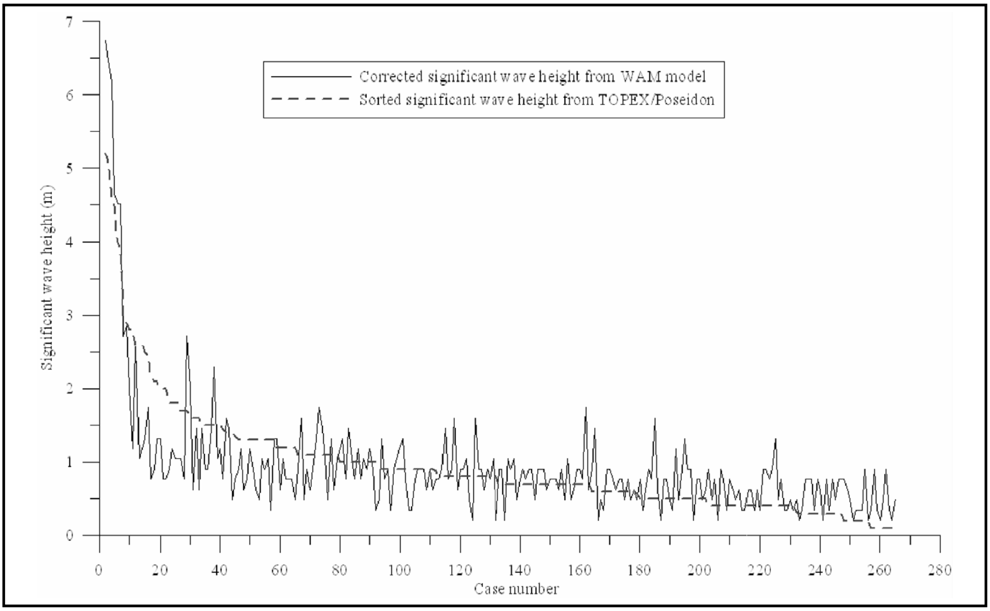

Fig. 6: Diagram of ordered $H_{m_0,T}$ and associated corrected $H_{m_0,M}$ $(\hat{H}_{m_0M})$.

sical regression analysis is

$$\hat{H}_{m_0,M} = -0.06228{+}1.60115\, H_{m_0,M}\ . \quad (15)$$

Denoting as x_0 the obtained value of significant wave height from WAM model, the deviation of the corresponding corrected values of the two correction relations, Eq. 14 and Eq. 15, is $\delta\,\hat{x}_0 = 0.13306{+}0.21135\; x_0$. It is evident that this difference is not negligible, suggesting that the adoption of the classical linear regression, at least for the present application, could lead to correction (calibration) errors.

The final assessment of the validity of the proposed correction relation is performed taking into account measured wave data from the POSEIDON buoy located near the Athos Peninsula, (Fig. 1). Measured $H_{m_0,B}$ time series from this location extending into the period 6/2000 – 12/2002 were utilized for the construction of the appropriate collocated (in time and space) datasets for the buoy measurements and the WAM model forecasts. The statistics of the collocated data are summarized in Table 2. The mean absolute bias between the WAM model data and the buoy data was found initially 0.34265. After applying the proposed correction Eq. (14) to the WAM model data, the mean value and standard deviation of the corrected results became much closer to the corresponding ones of $H_{m_0,B}$, while the skewness, kurtosis and correlation coefficient remain unchanged, as was expected. In addition, the absolute mean bias has been reduced to 0.2912 i.e., a relative improvement of the order of 15%. On the contrary, applying the correction relation (15) obtained from classical regression analysis, resulted in a mean absolute bias between the WAM model data and the buoy data equal to 0.8370. It is clear that the proposed correction relation (14) is much more effective than the Eq. (15) obtained from classical regression analysis. Thus the proposed correction relation was found to improve the WAM wave model results in the area of the North Aegean Sea, even for data not used in the analysis presented above.

Table 2
Statistics of collocated data from WAM model (M), in-situ measured (B) and corrected WAM model (M_{COR}) significant wave height at the Athos location.

Parameter	M	B	M_{COR}
N		5685	
m_H	0.5617	0.8225	0.8514
M_H	0.4000	0.5770	0.6267
s_H	0.4927	0.7815	0.6847
$\max_H$	4.4000	5.9981	6.1859
g_H	3.0997	2.2605	3.0997
k_H	14.0066	7.0765	14.0066
CV	0.8771	0.9501	0.8042
r		0.8636	

Conclusions

In this work, validation of results from the 3rd-generation wave model WAM is performed based on TOPEX/Poseidon data in three offshore locations of the North Aegean Sea. The statistical inter-comparison of the collocated data samples of significant wave heights, collected during a 31-month period (January 2000 – July 2002), revealed that the corresponding populations present significant differences, at least as concerns the main statistical parameters. The mean bias was found to be high compared to the corresponding mean values of the two samples and the variability of the WAM results greater than the corresponding T/P data. However, the correlation coefficient of the two data sets supported the establishment of a linear structural relationship between the two variables. Given the realistic assumption that both T/P measurements and WAM results are subject to errors, the Errors-in-Variables (EIV) regression approach was implemented, the linear relationship between the two variables was produced and a correction relation of WAM results for the North Aegean Sea was proposed. The analysis resulted in a non-negligible difference with classical linear regression, the implementation of which could lead to correction (calibration) errors. On the other hand, the use of the EIV approach leads to improvements of the WAM model forecasts in the North Aegean Sea, a fact that is confirmed even if in-situ buoy measurements are used.

Finally, the underestimation trend of the WAM model, which has been pointed out by previous studies for relatively nearshore locations, was once again verified for the examined offshore locations, and could be considered as a result with general validity for closed basins with the particularities of the Aegean Sea.

Acknowledgements

1. This work has been partially supported by the Hellenic General Secretariat of Research and Technology within the context of a project entitled 'An enhanced operational system for wave monitoring and prediction with applications in Hellenic navigation'.
2. The authors are grateful to the two reviewers for their valuable comments and suggestions.

References

AVISO/ALTIMETRY, 1996. AVISO User Handbook for Merged TOPEX/POSEIDON products. AVI-NT-02-101, Edition 3.0.

BABOVIC, V., SANNASIRAJ, S.A. & CHAN, E.S., 2005. Error correction of a predictive ocean wave model using local model approximation. *Journal of Marine Systems*, 53: 1-17.

BARNETT, V.D. 1967. A note on linear structural relationships when both residual variances are known. *Biometrica*, 54: 670-672.

BARSTOW, S., HAUG, O. & KROGSTAD, H., 1997. Satellite Altimetry data in wave energy studies. *Proceedings of the Third International symposium on Ocean Wave Measurement and Analysis – WAVES' 97*, ASCE, Virginia, 339-354.

BIDLOT, J.-R. & HOLT, M. W., 1999. Numerical wave modelling at operational weather centres. Coastal Engineering, 37(3-4): 409-429.

BOGGS, P.T. & ROGERS, J.E., 1990. Orthogonal distance regression. *Contemporary Mathematics*, 112: 183-194.

CAIRES, S. & STERL, A., 2003. Validation of ocean wind and wave data using triple collocation. *Journal of Geophysical Research*, C, 108 (3): 3098-3113.

CARTER, D.J.T., CHALLENOR, P.G. & SPOKOSZ, M.A., 1992. An Assessment of Geosat wave height and wind speed measurements. *Journal of Geophysical Research*, 97 (C7): 11383-11392.

COTTON, P.D. & CARTER, D.J.T., 1994. Cross calibration of TOPEX, ERS-1 and Geosat wave heights. *Journal of Geophysical Research*, 99: 25025–25033.

DRAPER, N.R. & SMITH, H., 1998. Applied Regression Analysis, Wiley Series in Probability and Statistics (3^{rd} Edition), 736p.

FULLER, W.A., 1987. *Measurement error models*, Wiley, New York.

GOWER, J.F.R., 1996. Intercalibration of wave and wind data from TOPEX/POSEIDON and moored buoys off the west coast of Canada. *Journal of Geophysical Research*, 101 (C2): 3817-3829.

GULEV, S.K., COTTON, D. & STERL, A., 1998. Inter-comparison of the North Atlantic Wave Climatology from Voluntary Observing Ships, Satellite Data and Modelling, *Physics and Chemistry of the Earth*, 23 (5-6): 587-592.

JCOMM, 2003. Advances in the applications of marine climatology. The Dynamic Part of the WMO Guide to the Applications of Marine Meteorology, WMO/TD-No. 1081, JCOMM Technical Report No. 13.

KALLOS, G., NICKOVIC, S., PAPADOPOULOS, A., JOVIC, D., KAKALIAGOU, O., MISIRLIS, N., BOUKAS, L., MIMIKOU, N., SAKELLARIDIS, G., PAPAGEORGIOU, J., ANADRANISTAKIS, E. & MANOUSAKIS, M., 1997. The regional weather forecasting system Skiron: An overview. *Proceedings of the Symposium on Regional Weather Prediction on Parallel Computer Environments*: 109-122.

KECHRIS, C. & SOUKISSIAN, T., 2004. Inter-comparison of Topex/Poseidon, wave buoy and WAM model results in the North Aegean Sea. *Proceedings of the 14^{th} International Offshore and Polar Engineering Conference,* 3: 163-168. Toulon, France, 23-28 May 2004.

KENDALL, M.G. & STUART, A., 1961. The Advanced Theory of Statistics, Vol. 2, Inference and Relationship. Hafner Publishing Company, 3rd Edition, p. 676.

KOBAYASHI, T. & YASUDA, T., 2004. Nearshore wave prediction by coupling a wave model and statistical methods. *Coastal Engineering*, 51: 297-308.

KOMEN G., CAVALERI L., DONELAN M., HASSELMANN K., HASSELMANN S. & JANSEN, P., 1994. Dynamics and Modelling of Ocean Waves. Cambridge, Cambridge University Press, 532p.

KROGSTAD, H. & BARSTOW, S., 1999. Satellite wave measurements for coastal engineering applications. *Coastal Engineering*, 37: 283-307.

KROGSTAD, H., WOLF, J., THOMPSON, S. & WYATT, L., 1999. Methods for inter-comparison of wave measurements. *Coastal Engineering*, 37 (3-4): 235-257.

LARSON, H.J., 1983. Statistics: An introduction. Krieger Pub. Co., 418 p.

LIU, P., SCHWAB, D. & JENSEN, R., 2002. Has wind–wave modelling reached its limit? Ocean Engineering, 29 (1): 81-98.

MONBALIU, J., HARGREAVES, J.C., CARRETERO, J.-C., GERRITSEN, H. & FLATHER, R., 1999. Wave modelling in the PROMISE project. Coastal Engineering, 37 (3-4): 379-407.

PAPADOPOULOS, A., NITTIS, K., KATSAFADOS, P. & THE POSEIDON PROJECT TEAM, 2000. Evaluation of the meteorological forecasting of the POSEIDON system. *Proceedings of the Sixth Hellenic Symposium, Oceanography and Fisheries*, 392-396. Chios, May 23-26, 2000.

RAWLINGS, J.O., PANTULA, S.G. & DICKEY, D.D., 1998. Applied Regression Analysis, A Research Tool. Springer (2nd Ed.), 657p.

RAY, R. D. & BECKLEY, B. D., 2003. Simultaneous ocean wave measurements by the Jason and Topex satellites, with buoy and model comparisons. *Marine Geodesy,* 26: 367-382.

RESIO, D. & CARDONE, V., 1999. Wave Climate Modelling. *Shore & Beach*, 67 (1): 61-65.

SANNASIRAJ, S.A., BABOVIC, V. & CHAN, E.S., 2005. Local model approximation in the real time wave forecasting. Coastal Engineering, 52 (3): 221-236.

SARKAR, A., MOHAN, M. & KUMAR, R., 1997. Inter-comparison of model-predicted wave heights with satellite altimeter measurements in the north Indian Ocean, *Ocean Engineering*, 24 (9): 879-885.

SIGNELL, R. P., CARNIEL, S., CAVALERI, L., CHIGGIATO, J., DOYLE, J. D., PULLEN, J. & SCLAVO, M., 2005. Assessment of wind quality for oceanographic modelling in semi-enclosed basins. Journal of Marine Systems, 53 (1-4): 217-233.

SNEDECOR, G. & COCHRAN, G., 1989. Statistical Methods. Ames, IA, Iowa State University Press (8th ed.), 503p.

SOUKISSIAN, T., CHRONIS G. & NITTIS, K., 1999. POSEIDON: Operational Marine Monitoring System for Greek Seas, *Sea Technology*, 40(7): 31-37.

SOUKISSIAN, T., PERIVOLIOTIS, L., PROSPATHOPOULOS, A. & PAPADOPOULOS, A., 2001. Performance of three numerical wave models on the Aegean Sea. First results. *Proceedings of the 11th International Offshore and Polar Engineering Conference*, III: 40-45. Stavanger, Norway, 17-22 June 2001.

SOUKISSIAN, T. & PROSPATHOPOULOS, A., 2003. Implementation of the 3rd generation Wave Model WAM - cycle 4 in the Aegean Sea. *Technika Chronika*, IV, 23 (1-2): 7-19. Extended summary in English.

WAMDI GROUP, 1988. The WAM model - a third generation ocean wave prediction model. *Journal of Physical Oceanography,* 18: 1776-1810.

WEBSTER, R., 1997. Regression and functional relations. *European Journal of Soil Science*, 48: 557-566.

WONG, M.Y., 1989. Likelihood estimation of a simple linear regression model when both variables have error. *Biometrika*, 76: 141-148.

YOUNG, I.R., 1999. An inter-comparison of GEOSAT, TOPEX and ERS1 measurements of wind speed and wave height, *Ocean Engineering*, 26: 67-81.

Accepted in March 2007