WAVE HEIGHT BACKGROUND ERRORS SIMULATION AND FORECASTING VIA STOCHASTIC METHODS IN DEEP AND INTERMEDIATE WATERS

Zoe Theocharis[1], Constantine Memos[1] and Demetris Koutsoyiannis[1]

Wave forecasting is accomplished today via numerical models. In this work we apply stochastic techniques using actual measurements to improve wave height forecast in real time. Application of these techniques in four locations of the Aegean Sea results in significant improvement of the forecast in the time domain retaining the same pattern of modifications, suggesting, thus, this method for operational use in deep and intermediate waters. The improvement is obtained by four regression models, which take into account the variable of the significant wave height as measured and forecasted by the model. Space-wise extension of the method was also investigated and applied to the Aegean Sea and the Indian Ocean, where its performance was remarkable.

INTRODUCTION

The forecast of the significant wave height is valuable in numerous coastal and offshore investigations and activities. This is currently accomplished numerically via the state of the art third generation deterministic wave models that solve the wave energy balance equation. In this work, we are dealing with the improvement of the wave forecast produced by these models. Waves have a deterministic relation to the wind. Nevertheless, the forecasting error can be important. There are two essential aspects regarding this issue:

- The 3rd generation wave model WAM underestimates the wave height forecast in a global level (Janssen, 1997)
- Wave fields predictability can not exceed wind fields predictability (Young, 1999)

In recent years, data assimilation and artificial neural network techniques have been used in a number of wave height forecast improvement efforts.

Data assimilation technique is a widespread method of satellite data operational exploitation. Its aim is the improvement of the forecast using imported observations of the wind and wave field mainly from satellites. Re-run of the model is taking place at the sequel. Usually, the operation wave models produce forecasts for the next 72-hours (1st guess forecast). Meanwhile, the satellites are overpassing the earth surface collecting information that can operationally being used. Henceforth, depending on the data used for the correction, one can obtain wind and wave data assimilation respectively.

In the last decade there is a tendency of predicting wave parameters via Neural Network Techniques. Tsai et al. (2001) for example present a work to

[1] School of Civil Engineering, National Technical University of Athens, Heroon Polytechniou 5, Athens, 15780, Greece

forecast the wave height using a set of previous observations. This is achieved with the application of neural networks which use a background algorithm related to the history of wave realizations. Furthermore, in the framework of this research, the wave height forecast was produced with the use of an artificial neural network, using as input data the wave height observations time series of adjacent buoy stations instead of the wind parameters time series. This is achieved through the application of the ANN's back propagation algorithm learning process, which provides the desirable results. The model computes the weight relations between neighbouring stations, based on short-term data, from which we can produce the forecast wave height time series or the missing data time series of the neighbouring station. The results show that ANN achieves satisfactory records for both aims of application. Deo et al. (2001), also, dealt with the wave height forecast in real time. They applied at the same time series two models, in order to compare the results: an ANN and a 2nd order autoregressive model. The ANN resulted in a coefficient of determination near 0.66 for a one-step forecast (three hours) and in 0.61 for a two steps forecast. The corresponding results for the bivariate autoregressive model were 0.61 and 0.50.

The stochastic applications in the sector of wave forecasting are relatively limited and concern in the majority of cases gap filling of measurements time series or re-analysis of wave hindcasts. Contrary to the stochastic models used for long-term forecasts, which produce probabilities of appearance of extreme sea states, the stochastic models related to short term forecasting or now casting, like the regression models, are using the memory-history to reproduce the following sea states. This fact prompted initially the researchers to use stochastic methods, and specifically the regression models, so that they described sea surface elevation time series as well as its spectrum (Scheffner and Borgman 1992). On the other hand, the stochastic simulation of DelBalzo et al. (2002) uses a twenty years wave height visual observations time series, provided by the COADS database. The simulated time series were compared with the 4-years observations time series of the National Data Buoy Center. The inter-comparison was satisfactory and the deviations occurred are due to the interaction between the buoys and the passing boats.

Finally, Caires and Sterl (2005), proposed a non parametric method to correct ERA-40 re-analysis. This method predicts the error using a background algorithm produced from a learning dataset.

Unlike Caires and Sterl method, the methods used here are parametric and customized to the specific application basin. Thus, in this work we present the application of linear and non-linear stochastic techniques to show that WAM background errors can be reasonably predicted by using a limited number of buoy observations and improve thus its forecasting robustness. Re-run of the wave model is not required. A full description of the third-generation ocean wave model WAM can be found in WAMDIG (1988).

THE METHOD - STOCHASTIC MODELS USED

In this investigation four types of multiregression models were used:

- a bivariate linear model **BLM1** whose explanatory variables are the WAM prediction of the current step and the measured height at the previous step:

$$\hat{\Psi}_{t+\Delta t} = aX_{t+\Delta t} + \beta\Psi_t + \gamma \tag{1}$$

where,

$\hat{\Psi}_{t+\Delta t}$: the estimated value of the significant wave height aty time step t+Δt

$X_{t+\Delta t}$: the value of the WAM forecast for the significant wave height, at time step t+Δt

Ψ_t : the value of the significant wave height measurement at time step t

α : weight parameter corresponding to $X_{t+\Delta t}$

β : weight parameter corresponding to Ψ_t

γ : steady parameter

- a bivariate linear model **BLM2** whose explanatory variables are the WAM prediction of the current step and the measured height at prior to the previous step:

$$\hat{\Psi}_{t+2\Delta t} = aX_{t+2\Delta t} + \beta\Psi_t + \gamma \tag{2}$$

where,

$\hat{\Psi}_{t+2\Delta t}$: the estimated value of the significant wave height at time step t+2Δt

$X_{t+2\Delta t}$: the value of the WAM forecast for the significant wave height, at time step t+2Δt

Ψ_t : the value of the significant wave height measurement at time step t

α : weight parameter corresponding to $X_{t+2\Delta t}$

β : weight parameter corresponding to Ψ_t

γ : steady parameter

- a trivariate linear model **TLM** whose explanatory variables are the WAM prediction of the current step and the measured height of two previous steps:

$$\hat{\Psi}_{t+\Delta t} = aX_{t+\Delta t} + \beta\Psi_t + \gamma\Psi_{t-\Delta t} + \delta \tag{3}$$

where,

$\hat{\Psi}_{t+\Delta t}$: the estimated value of the significant wave height at time step t+Δt

$X_{t+\Delta t}$: the value of the WAM forecast for the significant wave height at time step t+Δt

Ψ_t : the value of the significant wave height measurement at time step t

$\Psi_{t-\Delta t}$: the value of the significant wave height measurement at time step t-Δt

α : weight parameter corresponding to $X_{t+\Delta t}$

β : weight parameter corresponding to Ψ_t

γ : weight parameter corresponding to $\Psi_{t-\Delta t}$

δ : steady parameter

- a bivariate non-linear model **BNLMa** and **BNLMb** with explanatory variables same as in the first bivariate model:

<u>BNLMa</u>:

$$\hat{\Psi}^{\lambda}_{t+\Delta t} = a_1 X^{\lambda}_{t+\Delta t} + \beta_1 \Psi^{\lambda}_t + \gamma_1, \quad \text{if } \Psi_t \le c \tag{4}$$

$$\hat{\Psi}^{\lambda}_{t+\Delta t} = a_1 X^{\lambda}_{t+\Delta t} + \beta_2 \Psi^{\lambda}_t + \gamma_2, \quad \text{if } \Psi_t \ge c \tag{5}$$

<u>BNLMb</u>:

$$\hat{\Psi}^{\lambda}_{t+\Delta t} = a_1 X^{\lambda}_{t+\Delta t} + \beta_1 \Psi^{\lambda}_t + \gamma_1, \quad \text{if } X_{t+\Delta t} \le c \tag{6}$$

$$\hat{\Psi}^{\lambda}_{t+\Delta t} = a_2 X^{\lambda}_{t+\Delta t} + \beta_1 \Psi^{\lambda}_t + \gamma_2, \quad \text{if } X_{t+\Delta t} \ge c \tag{7}$$

where,

$\hat{\Psi}_{t+\Delta t}$: the estimated value of the significant wave height at time step t+Δt

$X_{t+\Delta t}$: the value of the WAM forecast for the significant wave height at time step t+Δt

Ψ_t : the value of the significant wave height measurement at time step t

α : weight parameter corresponding to $X_{t+\Delta t}$

β : weight parameter corresponding to Ψ_t

γ : steady parameter

c : the wave height defining the applicability range of the model equations

λ : exponential index of the non-linear transformation

The parameters listed above are estimated by a weighted least squares procedure and from the continuity equations.

FIRST ASSESSMENT

Improvement in the time domain

Forecasting of the wave climate in the Aegean Sea is accomplished today via the numerical model WAM, run under the auspices of the National Centre for Marine Research of Greece in the framework of the Poseidon Operational System. A systematic underestimation of the significant wave height is observed for the region of the Aegean Sea, as seen for example in Figure 1, where the observations and WAM prediction for Athos location ($39^0$58' - $24^0$43') in the northern part of this sea are depicted for the year 2001. This work deals with the improvement of the significant wave height forecast in real-time. The results were checked against measurements from four pilot-study monitoring stations of the Aegean Sea. These stations are located in the open sea near the Athos peninsula and offshore the islands of Lesvos, Mykonos and Santorini.

In order to develop an efficient wave forecasting system for the Aegean Sea, taking into account the existing peculiarities such as the complex shore-line and the numerous islets as well as the changeable nature of the wind field that influence WAM's predictive power, a set of stochastic methods was examined: four regression models, three linear and one non-linear that take into account the measured and forecasted significant wave height.

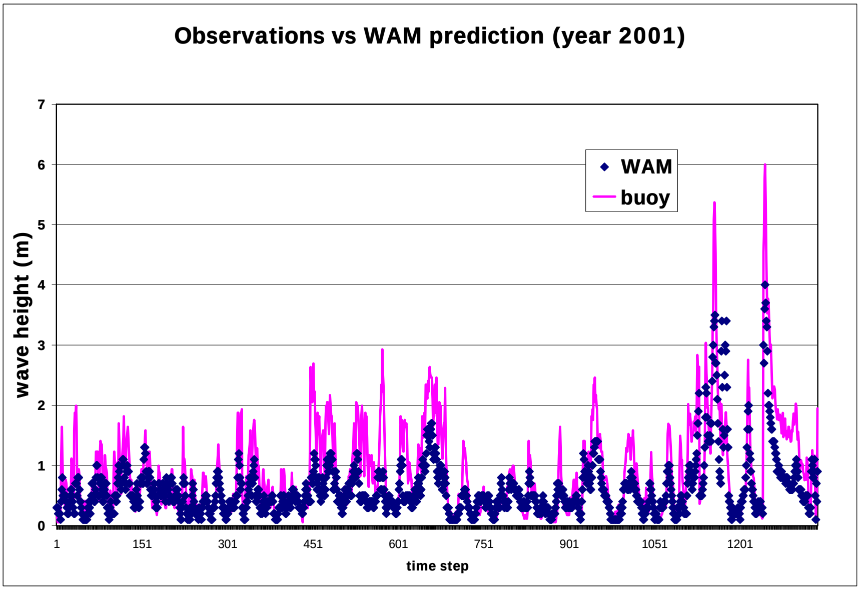

Figure 1: Observed and predicted values of the wave height, Athos station. Severe underestimation observed especially in high sea-states in the year 2001.

The four multiregression models presented in the previous section, resulted in significant improvement of the coefficient of determination, irrespectively of the time period of application. More specifically, the coefficients of determination, that give the proportion of the variance of the predicted variable from the measurements and represent the model adequacy, have increased from approximately 0.7 to over 0.9 as shown in Table 1.

Table 1. Coefficients of determination obtained for each station and each regression model.

	R^2 Athos	R^2 Lesvos	R^2 Mykonos	R^2 Santorini
WAM	0.781	0.713	0.722	0.676
BLM1	0.920	0.892	0.927	0.906
BLM2	0.865	0.813	0.860	0.820
TLM	0.924	0.895	0.931	0.911
BNLMa	0.929	0.897	0.936	0.909
BNLMb	0.929	0.898	0.935	0.908

It is shown that the stochastic models result in a significant forecast improvement, irrespectively of the application time period and of the location of the prediction. Better approximation of the measured wave height as compared to WAM's prediction as well as significant decrease of the standard deviation is achieved in all stations. This can be seen in Figure 2, where Athos station results are presented in descending order of wave heights.

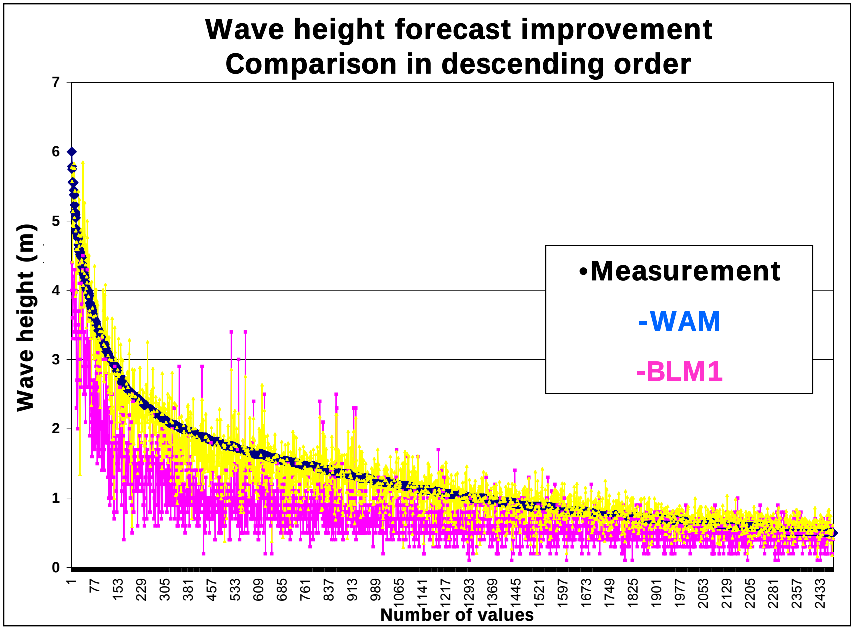

Figure 2: BLM1 results in descending order together with the measured and forecasted values of the wave height in Athos station.

By applying the stochastic model BLM1 for lead times between the three hours time step and the 72 hours step, one can observe that the predictive power of the stochastic model reaches the wave model coefficient of determination in all four stations after about 72 hours following the pattern shown in figure 3, suggesting, thus, that the stochastic tools could be valuable for operational use along with the wave model prediction.

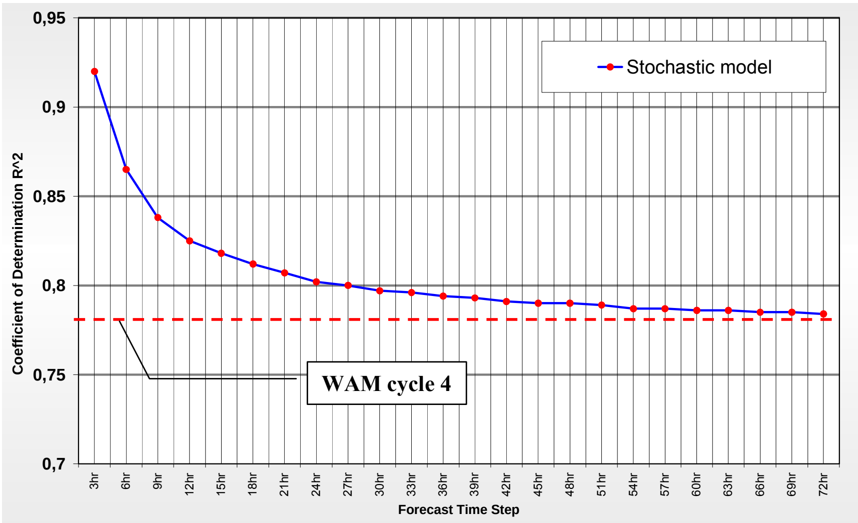

Figure 3: Coefficient of determination decay v.s. lead time

By observing results of the various applications performed, one can see that by applying the same parameters for all seasons and locations, significant improvement can be obtained. The results for all periods of the year showed that there is no need for individual (depending on the season) implementation of the regression models, since the linear and non-linear factors were found to be almost constant regardless of the station or period analyzed. In Table 2 for example one can see a sample of the parameter values produced for different periods of implementation of BLM1 for Athos station. It is remarkable that by using an overall mean value of α,β and γ parameters produced for different periods the same significant improvement can be achieved.

Table 2. BLM1 parameters for different implementation periods. Athos station.

Imple-mentation period	α	β	γ	R^2 before	R^2 after
All years	0.408	0.879	-0.167	0.781	0.924
1st year	0.394	0.914	-0.192	0.756	0.910
2nd year	0.414	0.869	-0.164	0.801	0.932
5/00 to 9/00	0.402	0.980	-0.254	0.593	0.850
3/01 to 9/01	0.421	0.919	-0.215	0.775	0.914
5/01 to 9/01	0.437	0.915	-0.234	0.603	0.840
10/00 to 2/01	0.412	0.852	-0.152	0.748	0.910
10/00 to 4/01	0.404	0.874	-0.163	0.779	0.922

This fact holds also for all other test stations used, and all other regression models impling that WAM underestimation pattern contains a systematic error source.

Tthe common pattern of the results revealed in all four stations examined, suggested that a space-wise expansion of the present method could be worthwhile.

SPATIAL MODIFICATION

This part of the study consists of a space-wise application including spatial stochastic modeling and wave information transfer aiming at extending the improvement described above in space and especially in coastal regions. To accomplish this, the wind speed and direction effects were included in order:

- to cluster the sea states according to their direction and define a reference station whose observations were used to improve the wave height prediction of the whole basin
- to determine the average time lag required for a sea state to be transferred in different locations of the basin which was used in BLM models

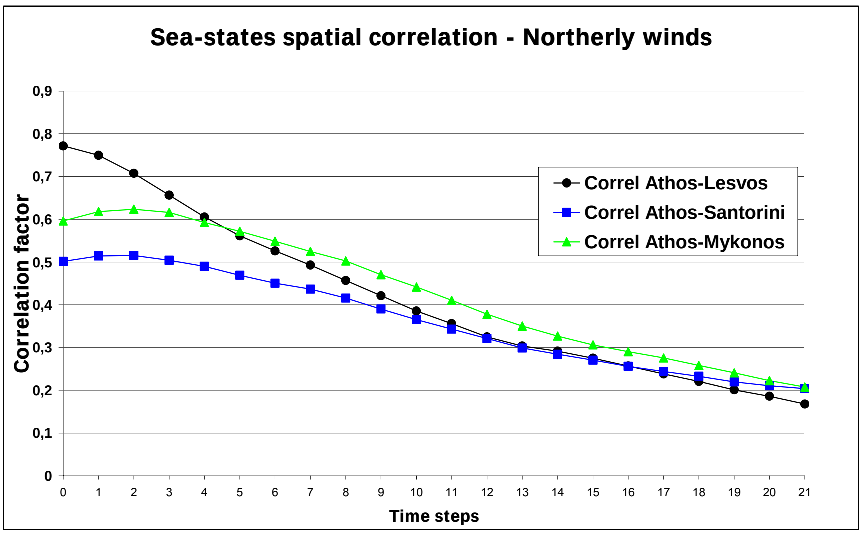

Figure 4: Spatial correlation of sea states for different time lags

For Northern winds for example, Athos station was used as a reference station. Then, by calculating the sea states correlation factor, presented in Figure 4, with respect to the stations to the south of Athos, it was found that depending on the stations distance, the correlation factor values were changing revealing a temporal delay. This temporal delay was found to be equal to the time needed for a sea state to be transferred from one station (Athos) to the other station taking into account the mean wave group velocity. This is the phase lag used to determine the BLM type used in each location.

The results of this space –wise extension of the stochastic methods under consideration, show a steady but somehow limited improvement of WAM's predictability power. This is mainly due to the complexity of the Aegean Sea with its numerous islands and complicated shoreline as well as to the irregular wave filed that is frequent over the Aegean Sea.

SECOND APPLICATION

To avoid the previously mentioned peculiarities of the Aegean Sea, further examination was conducted. Two locations of the Indian Ocean were studied by stochastic techniques. WAM cycle 4 without assimilation schemes along with data from two buoys were used. The scope was to improve the 9 hrs time step wave forecast at the second location which lies in intermediate waters. For this purpose, data of the first location, which lies 900 Km offshore the India peninsula, along with the prediction of the WAM model represent the explanatory variables of the stochastic method.

After examining the correlation factors the reference station was determined and the corresponding time lag was calculated before applying the BNLMa model. It is noted that non linear transformation in the stochastic models is related to the swell content, which can be appreciable in that area. It was found that this technique enhances the improvement of the wave height prediction in intermediate waters by using the offshore measurement. Validation was done by comparing the results with the intermediate waters buoy data. The improvement of the wave height prediction is remarkable as shouwn in Figure 5.

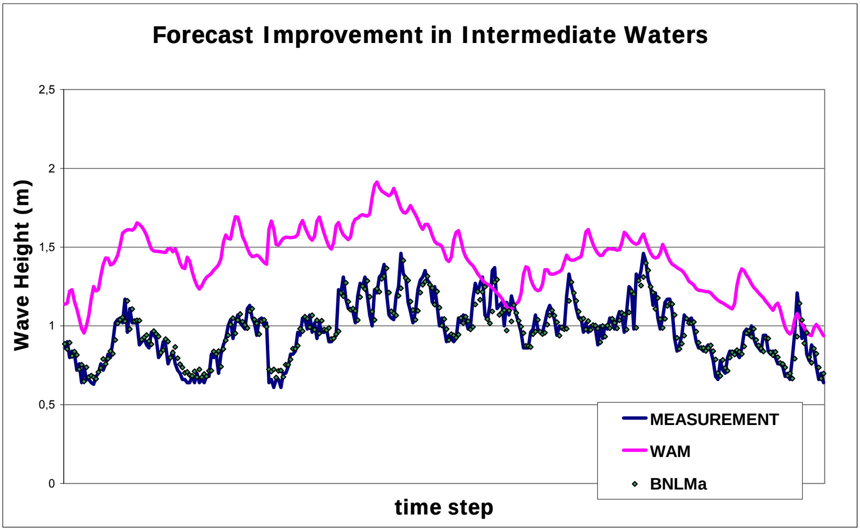

Figure 5: Wave height time series in Indian Ocean intermediate waters. WAM versus buoy and stochastic model

CONCLUSIONS

The results show that:

- A systematic error source exists in the wave models prediction judging by the stochastic parameters behavior irrespectively of the wave height value and season of application
- Measurements utilization in real time in combination with the use of stochastic methods, can significantly improve the wave model forecast in time and space
- Intermediate and shallow waters wave forecast can be produced given that an offshore station exists

Due to the results that seem rather encouraging, the authors have already implemented the method in other regions and tried specifically to focus on the possibility of using the stochastic tools for direct exploitation of satellite data in order to improve the wave forecast. This study has already been applied in a 20^0 by 20^0 degrees box in the Southern Indian Ocean, the so called "wave energy storehouse". This case study will be published in the near future.

ACKNOWLEDGEMENTS
The authors would like to thank Professor Kallos and the National Center for Marine Research for providing the data used in this study.

REFERENCES

Caires, S. and A. Sterl. (2005). “A new non-parametric method to correct model data: Application to significant wave height from the ERA-40 reanalysis”, *J. Atmospheric and Oceanic Tech.,* 22(4), pp.443-459.

DelBalzo, D., R., Schultz, J., R., Earle, M., D., 2002, Stochastic time-series simulation of wave parameters using ship observations. Ocean Engineering 30, pp. 1417-1432.

Deo, M.C., Jha, A., Chaphekar, A.S., Ravikant, K., 2001, Neural networks for wave forecasting. Ocean Engineering, 26, pp. 191-203.

Janssen, P.A.E.M., 1991, Quasi-linear theory of wind-wave generation applied to wave forecasting. J. Phys. Oceanogr., 21, pp. 1631-1642.

Scheffner, N., M., Borgman, L., E., 1992, Stochastic time series representation of wave data. Journal of Waterways Port Coastal Ocean Engineering, ASCE, 118(4), pp. 1955-2012.

Theocharis, Z. Memos C. and Koutsoyiannis D., 2006. Wave forecasting improvement in deep and intermediate waters, WISE 2006, Venice.

Theocharis, Z., Koutsoyiannis D., Memos C., and Soukissian T., 2004, Improvement of the wave height real-time forecast in the Aegean Sea using stochastic methods, 1st General Assembly of the European Geosciences Union, Geophysical Research Abstracts, Vol. 6.

Tsai, C., P., Lin, C., Shen, J., N., 2001, Neural network for wave forecasting among multi-stations. J. Ocean Engineering, 29, pp. 1683-1695.

WAMDIG 1988: The WAM model - A third generation ocean wave prediction model. Journal of Physical Oceanography, 18, pp. 1775-1810.

Young, I.R. 1999. *Wind Generated Ocean Waves*. Elsevier Science, Oxford, U.K. 288pp