Verification within Wave Resource Assessments. Part 1: Statistical Analysis

Verification within Wave Resource Assessments. Part 1: Statistical Analysis

EC Edwards[1], LC Cradden[1], DM Ingram[1], C Kalogeri[2]

[1]Institute for Energy Systems, University of Edinburgh, Edinburgh EH9 3JL
[2]Atmopsheric Modelling and Forecasting Group, National Kapodistrian University of Athens, Greece

Key Words: wave energy, wave power, resource assessment, model validation, WAM, EMEC test site, numerical model

ABSTRACT

Interest in wave energy as a viable renewable energy has increased greatly in the past couple of decades. To determine the potential that a certain location has to harvest wave energy, a resource assessment must be performed for that location. As wave energy converter technologies get closer to market, it is becoming necessary to undertake more detailed resource assessments in order to determine the optimal location for deployment as well as the design and operating sea states. This study shows the level of sophistication that must be included in the verification process within a wave resource assessment. We describe the methodology in two articles. The first part shows how doing a complete statistical analysis of the fit of the model at the location of interest is essential for determining the reliability of the model data. Part 2 of this study will investigate the systematic trends of the fit of spectral values. In Part 1, it is shown that spatial analysis, the examination of distributions to reveal overall trends, and the careful choice of the appropriate statistical model to describe the fit of the wave model to buoy observations are all critical steps that must be added to verification processes. Part 2 demonstrates that looking closely at the fit of spectral values can reveal potentially vital issues for energy extraction. Better statistical validation gives the predictions of a particular resource assessment greater credibility or reveals areas where model accuracy must be improved.

1. Introduction

Gunn and Stock-Williams [1] estimate a global wave power resource of 2.11 ± 0.05 TW. With the growing interest in this source of energy and the resulting recent increase in the level of detail required for resource assessments comes the need for a more thorough understanding within each assessment of the specific wave model used. The appeal of using wave models to predict resources is very logical; for example, a model can give a prediction over a much longer time period than buoy or satellite measurements. Wave models are extensively validated [e.g. 2, 3, 4], but these validations are often not focused on parameters that are vital for wave energy extraction. For a resource assessment, before the wave model can be deemed reliable, a complete statistical analysis of the fit of the wave model at the location of interest must be completed, and the implications of any errors found must be

thoroughly understood. To verify the fit of the model, output must be compared to *in situ* observations from a buoy or satellite.

Large-scale resource assessments have become more comprehensive in the past couple of years, but overall there is not a large emphasis in these studies on verification of the model used. Gunn and Stock-Williams [1] provide a comparison of previous global and national/regional resource assessment studies. One example, Mørk et al. [5], performs a global resource assessment study, and a global map of correlation coefficients between model and satellite significant wave height is shown by way of verification. The global assessment by Ariange and Cheung [6] includes a slightly more detailed verification process, in which the correlation coefficient, root mean square error (see section 2.2), and 90% confidence intervals are calculated for significant wave height measured against altimetry data and for significant wave height, peak wave period and mean wave period measured against buoy data.

For national or regional studies, the degree of verification increases. For example, in ESBI's Wave Energy Resource Atlas Ireland [7] the verification process involves calculating coefficients of correlation and determination for significant wave height and wave period, and correcting the model data to align with buoy data if the values suggest this is required. In ABPmer's Atlas of UK Marine Renewable Energy Technical Report [8] significant wave height and resultant wave power model values were compared to monthly means and overall distributions of buoy observations. The authors provide hypotheses on why certain locations had better agreement than others and observations on seasonality. EquiMar (Equitable Testing and Evaluation of Marine Energy Extraction Devices) Project Deliverable D2.3 [9] proposes a standard practice for validation of numerical models, following the example of Ris et al. [4], which involves calculating bias, root mean square error, scatter index, model performance index, and operational performance index (see section 2.2).

The main focus of Part 1 of this study is to highlight the benefits of carrying out a comprehensive statistical analysis for *localized* resource assessments. There is limited evidence of this in practice in existing literature. For example, in an assessment of the wave energy potential in El Hierro by Iglesias and Carballo [10], there was no verification and only a model was used. In a study of the U.S. Pacific Northwest by Lenee-Bluhm et al. [11], only buoy measurements were used.

Stopa et al. [12], in an assessment of the wave energy potential in Hawaii, compare model data to both satellite data and buoy data. Statistics such as bias, normalized root mean square error, correlation coefficient, and scatter index (see Section 2.2) are calculated, and time series and quantile-quantile plots are constructed. The fit of the model at different locations are compared, and a structured pattern of agreement in the model associated with buoy locations around the islands is observed. The analysis is, however, limited to significant wave height, and does not include energy period which is also influential in calculating wave energy potential.

Liberti et al. [13] compare the fit of model calculations to satellite and buoy measurements for the Mediterranean. Statistics are calculated for significant wave

height, spectral period (T_{02}), and mean wave direction, including bias, root mean square error, and scatter index. This study does consider periods by examining spectral period, but analyzing energy period would lead to a more targeted validation for wave energy.

In a resource assessment for the Cornish coast (UK), van Nieuwkoop et al. [14] perform the most extensive verification analysis for a local resource assessment at present. The importance of removing outliers is stressed, relevant statistics are calculated for significant wave height, energy period and mean direction, and bias is calculated for different bins of the parameters. It is also openly stated that the verification led to the conclusion that the model data needs to be handled with care. In the discussion, a correction of the parameters is shown, and the impact of this on wave resources and wave extremes is demonstrated. A quadratic correction formula is used for significant wave height, and a linear one is used for energy period. A difference in wave power of 7 kW/m is found, and extreme wave heights are shown to be 10-15% higher than previously predicted.

Inadequate consideration of statistics could cause considerable errors in the resource assessment. For example, it may result in under- or over- prediction of theoretical power potential, the amount of power technically recoverable, the amount of energy coming in as swell vs. wind waves, the extreme waves the devices must withstand, the direction with the most energy, or the extent of directional spread, critical for devices in a fixed orientation. It will be demonstrated in this work that by adding the following steps to the verification process, additional information is obtained that will affect the conclusions of the resource assessment:

- examining the fit of multiple surrounding model grid points for each buoy to determine which grid point should be used to calculate the potential wave resource at a site,
- examining distributions to determine overall trends in the data, and
- choosing the appropriate statistical model to describe the fit of the model to the *in situ* measurements. Specifically, it is essential not to assume normality of error structure and blindly rely on calculations based on linear regressions

The site examined is near the Orkney Islands, Scotland. It is not intended to be a resource assessment of this site, as there already exist such studies [e.g. 15] and the time period examined here is relatively short (2 years). It will be shown that if previous methodologies were used for verification and if the model results were used as they were given, the theoretical average power per crest would be over-estimated by 11.91 kW/m over the two year period of the data (an over-estimate by approximately 40%).

Section 2 will show the methodology and results of the complete statistical analysis. This includes describing what might be concluded if the verification were to stop at certain points. Specifically, 2.1 will introduce the data and show the calculations performed to compare parameters. 2.2 will present the spatial analysis and calculate basic statistics that existing verifications include. 2.3 will show the deeper analysis

and describe what the tests divulge about the fit of the model. Section 3 will then summarize the findings of Part 1 and connect the analysis with Part 2.

2. Methodology and Results

2.1 Overview of the wave model and data used

In this section, an example verification will be performed. Comparison of results from a wave model to *in situ* observations from a buoy will be made. The wave system used for this analysis is the ECMWF Cy331 version [16, 17] of WAve Model (WAM) [18, 19]. WAM is a third generation wave model that has been adopted by many operational and research centres worldwide. The model solves the spectral action balance equation without any presumptions on the shape of the wave spectrum and represents the physics of the evolution of the wave spectrum in accordance with our current knowledge using the full set of degrees of freedom of a two-dimensional wave spectrum. The grid resolution can be arbitrary in space and time. The propagation can be done on a latitudinal – longitudinal or on a Cartesian grid. WAM is able to run in a deep-water or a shallow-water mode and includes the effect of wave refraction caused by changes in depth and by ocean currents.

This particular version, developed at the National Kapodistrian University of Athens (NKUA), incorporates a number of important implementations that increase the potential capabilities of the wave system. Some of the most important improvements include the new advection scheme, which takes corner points into account, and thus provides a more uniform propagation in all directions, using the new Corner Transport Upstream scheme [20], and the new parameterization of the shallow water effects. In particular, in shallow water the four-wave interaction is vanished by the wave induced currents generated by the finite amplitude surface gravity waves. Thus following the work of Janssen and Onorato [21], a parameterization of this shallow water effects is introduced and affects both the evolution of the wave spectra and the determination of the kurtosis of the wave field.

The wave model has been configured to run in computational domain covering the North Atlantic between latitudes 20°N and 75°N and longitudes 50°W and 30°E. The domain extends far beyond the area of interest aiming to capture the important swell propagation. It is discretized in a considerably high resolution of (0.05 x 0.05 degrees) which has been adopted as suitable for capturing the fine-scale features in a credible way. The wave spectrum has been discretized in 25 frequencies (logarithmically spaced from 0.0417 to 0.5476 Hz) and 24 equally spaced directions. A 75-second time-step was selected so as to satisfy the CFL stability criterion. The wave model operated was driven by 3-hourly 10m winds provided by the regional atmospheric modelling system SKIRON [22, 23]. The horizontal resolution of the SKIRON model is also 0.05 x 0.05 degrees extending from surface up to 50hPa in 45 vertical levels using a 15-sec time-step. The wave model provides outputs for wide range of wave parameters and components such as wave height (significant and

swell), directions, energy and peak period. The result is full wave spectrum at preselected grid points.

We extracted information from 12 of the model's grid points, which were located near the European Marine Energy Centre (EMEC) wave test site off the coast of the Island of Mainland in the Orkney Islands, Scotland (see Figure 1). Since the period parameter used in wave resource assessments is energy period (T_e), it is necessary to approximate this using $T_e = 1/\overline{f}$.

The buoy is a Datawell Waverider Buoy situated at the EMEC wave test site (58.98˚N, 3.39˚W). Hourly buoy data was available from 1/1/2006 to 26/12/2007, with 674 records missing (3.8% of total). The missing records were fairly evenly spread over the time period, and both buoy and model values at these records were registered as missing so as not to affect the statistical computations. The buoy measures a directional spectrum 8 times every 30 minutes, and then averages the values to output one spectrum. Only the first record every hour, representing the first 30 minutes, is used in this study for comparison to the model. For each time record, the spectral file includes the maximum power spectral density (PSD) and 64 frequencies (ranging from 0.0225 Hz to 0.585 Hz), along with normalized PSD, mean direction, and directional spread for each frequency. Parameters calculated from the spectra are also provided, including significant wave height (H_{m0}), spectral period (Tdw2), and mean period (T_1). Energy period (T_e) is defined as

$$T_e = \frac{m_{-1}}{m_0} \tag{1}$$

Where m_n is the n^{th} moment of the spectrum. Since T_e is not a direct output of the buoy measurements, it is calculated by combining the equations for spectral period (2) and mean period (3).

$$Tdw2 = \sqrt{\frac{m_{-1}}{m_1}} \tag{2}$$

$$T_{01} = \frac{m_0}{m_1} \tag{3}$$

Inserting equations (2) and (3) into equation (1), T_e can be defined as

$$T_e = (Tdw2)^2 \cdot \frac{1}{T_{01}} \tag{4}$$

[24, 25]

Mackay et al. [26] states that spatial and temporal scales must be compared when evaluating different sources of data. Buoy and model measurements and the derived

quantities correspond to specific points in space. They are not averaged over an area, indicating that the spatial scales here are compatible. In contrast, the buoy and model differ temporally. As described above, buoy observations are averages of the first 30 minutes of the hour, whereas the model calculations are essentially samples from “on the hour” (± 75 seconds). It is also important in this study to compare spectral scales, for both directions and frequencies, which is done in Part 2.

2.2 Basic statistics and spatial analysis

The locations of the buoy and the twelve grid points at which the model is run are represented in Figure 1.

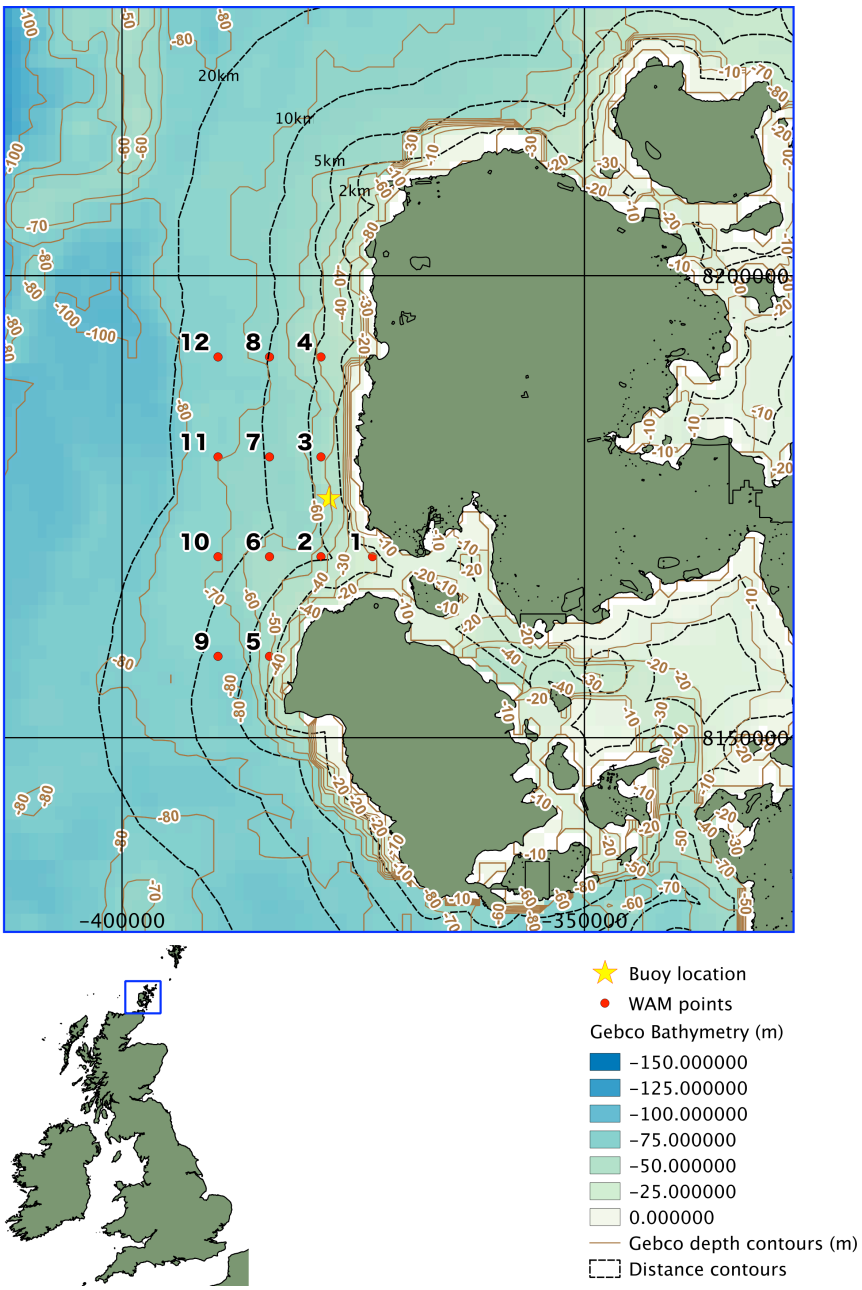

Figure 1: Map of the location of the Waverider Buoy (star) and the 12 grid points (red dots) at which the wave model was run off the coast of the Orkney Islands, Scotland

To determine the model grid point to which the buoy data will be compared, the most common procedure is to consider the closest grid points and then select the most appropriate based on physical restrictions. In this case, we must select between points 1,2,3,6, and 7. Point 1 should be excluded because it is in a sheltered area near the boundaries. Points 2 and 3 should also be excluded because they are boundary points and consequently miss all of the wave action from the eastern sector.

To confirm these physical notions, very basic statistics are calculated by comparing the buoy dataset against each model dataset. The fit of the model can be visually represented by Taylor diagrams (Figures 2 and 3). In a Taylor diagram [28], each grid point's correlation coefficient (*r*), root mean square error (RMSE), and the ratio of the standard deviations are combined into one point. Appendix A describes how the statistical parameters are calculated. Appendix B describes how to produce a Taylor diagram. These calculations are performed at each grid point for both significant wave height (H_{m0}) and energy period (T_e).

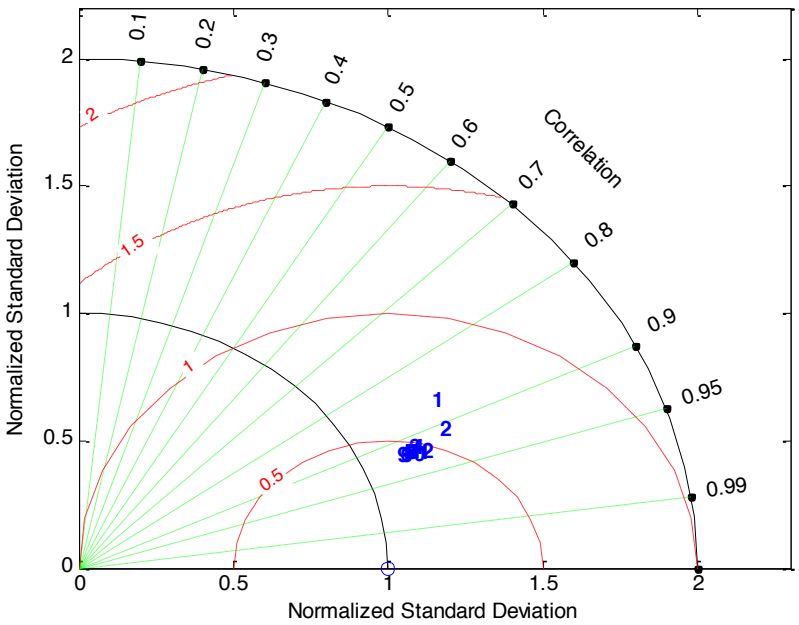

Figure 2: Taylor Diagram for H_{m0}, showing grid points 1 and 2 are least correlated with the buoy data, and all other points are very close together

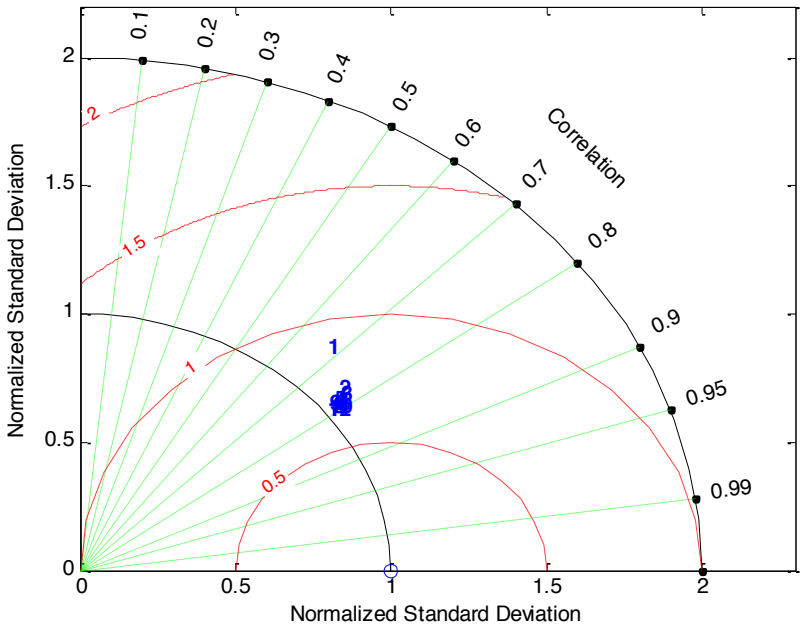

Figure 3: Taylor Diagram for T_e, showing grid point 1 is least correlated with buoy data, and all other points are very close together

Figures 2 and 3 reinforce what was found by considering the physical constraints of the location and emphasize the importance of spatial analysis. The closest grid points to the buoy are 1,2, and 3, but it is shown that points 1 and 2 are much less similar to the model than points 3-12. Because of what is seen in the Taylor diagrams, as well as what was seen by looking at physical restrictions of the location, point 6 will be the focus of the rest of the investigation.

Table 1: Basic statistics calculated to this point for grid point 6

Statistic	H_{m0} (m)	T_e (s)
r	0.92	0.79
Bias	0.23	1.19
RMSE	0.61	1.71
SI	0.29	0.20
s(model)	1.48	1.91
s(buoy)	1.31	1.83

The summary statistics for the grid point of interest (Point 6) are shown in Table 1. Appendix A describes how these statistics are calculated. Some validation studies stop at this point. If this study were to finish here, it would be concluded that there is strong correlation between model and buoy H_{m0} values and relatively strong correlation between model and buoy T_e values. Overall, the bias for H_{m0} is insignificant, although there is a slight tendency to over-estimate the value of H_{m0}. Since RMSE is higher than bias, some of the differences between model and buoy values cancel each other, and there are both over-estimates as well as under-estimates present. The bias for T_e is large, so there is an overall tendency of the model to over-estimate the buoy measurement for T_e. The RMSE is also higher than the bias for T_e, again suggesting that there are both under-estimates as well as over-

estimates in the model dataset. The SI values suggest a narrow scatter for both parameters, and the standard deviations are relatively similar.

Whilst this summary suggests good agreement, it does not show the whole picture. All of the statistics are stationary and do not express changes of fit in time. Consequently, many conclusions are masked.

2.3 Further Analysis of wave parameters

In this section, significant wave height and energy period predictors will be compared to buoy observations in more detail to assess the reliability and suitability of the model. The analysis includes time series, error scatter plots, kernel density distributions, tests for similarity of distributions, cross- and auto-correlation, and linear and generalized linear models.

Time series are often examined to look for trends, serial dependence, and stationarity [29]. Qualitatively assessing model fit over a long period of time in this manner is difficult. Here, results for December 2006 are presented as the representation of the most volatile month in the entire time history.

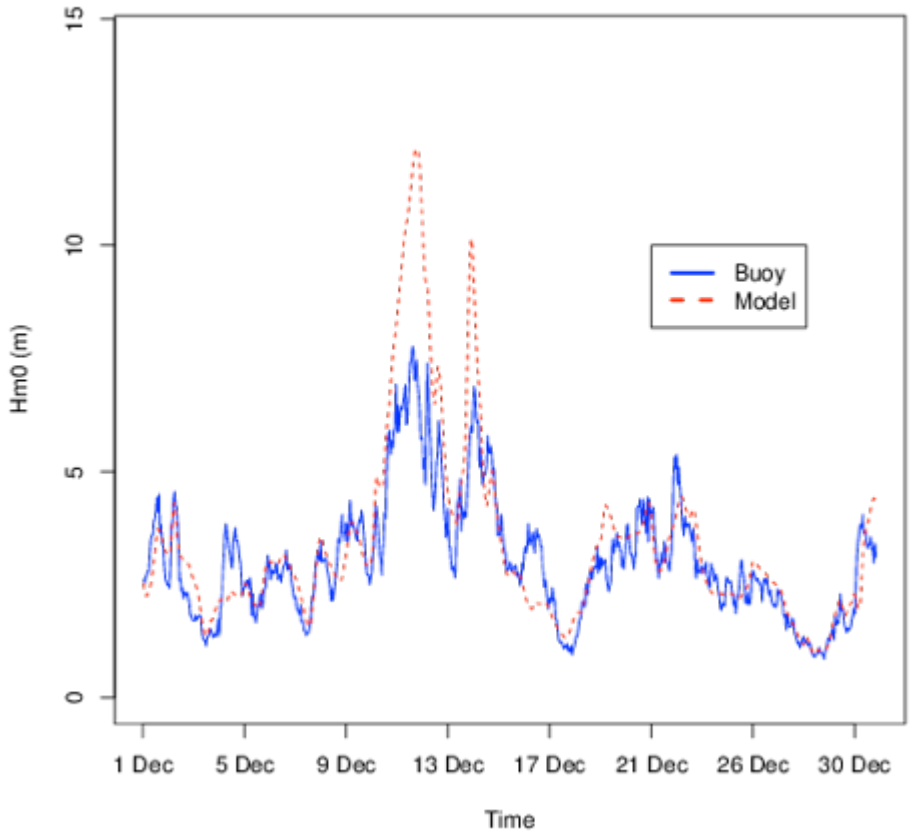

Figure 4: Time series of hourly H_{m0} values from the wave model and the buoy from 1 December 2006 to 31 December 2006

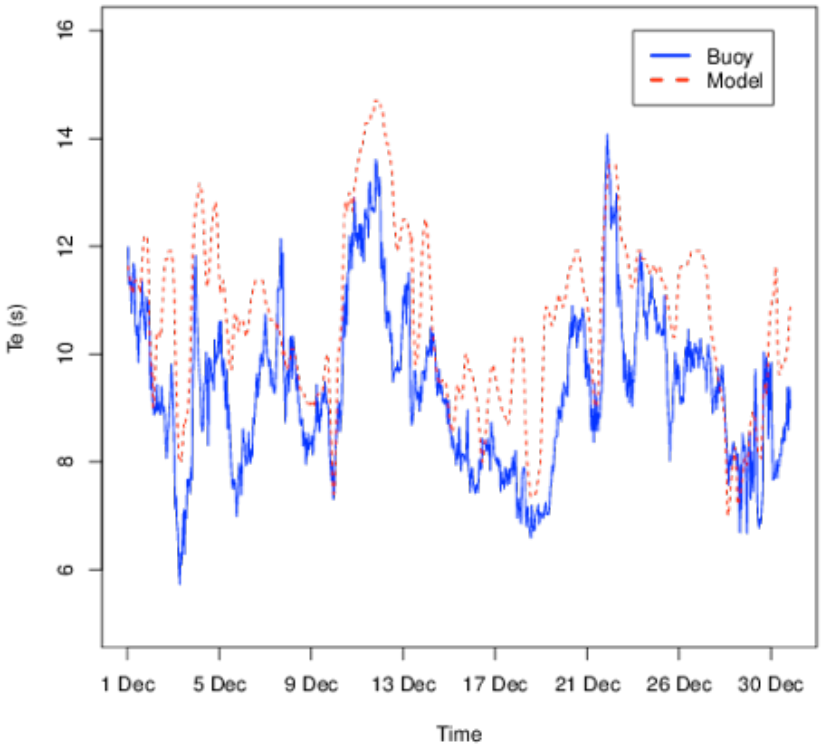

Figure 5: Time series of hourly T_e values from the wave model and the buoy from 1 December 2006 to 31 December 2006

Figure 4 shows that the model H_{m0} values follow the general patterns of the corresponding buoy values well, but during the two periods of time where the buoy values are at their highest, the model over-estimates the H_{m0} by a significant amount. Figure 5 shows that the model over-estimates the value of T_e frequently and consistently. There is clearly insufficient information to make sensible conclusions from these plots.

Rather than look at scatter plots of the model parameters vs. buoy parameters [as seen in 14, 30, 12, 6, 13, 31], here the scatter plot of error (buoy – model values) against buoy measurements is used. This enables a better understanding of the fit of the model. Specifically, it reveals problem areas based on parameter value. Figure 6 shows an error scatter plot for H_{m0}, and Figure 7 shows a similar plot for T_e.

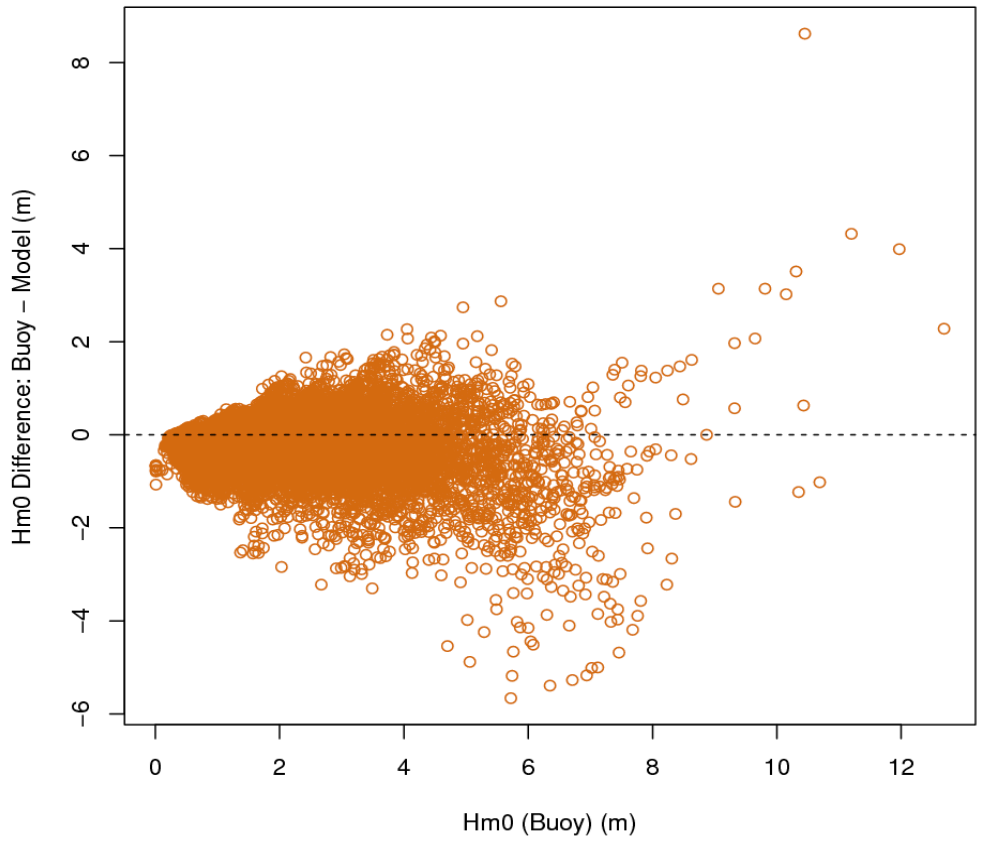

Figure 6: Error scatter plot of H_{m0}, defined as Buoy – Model against Buoy value

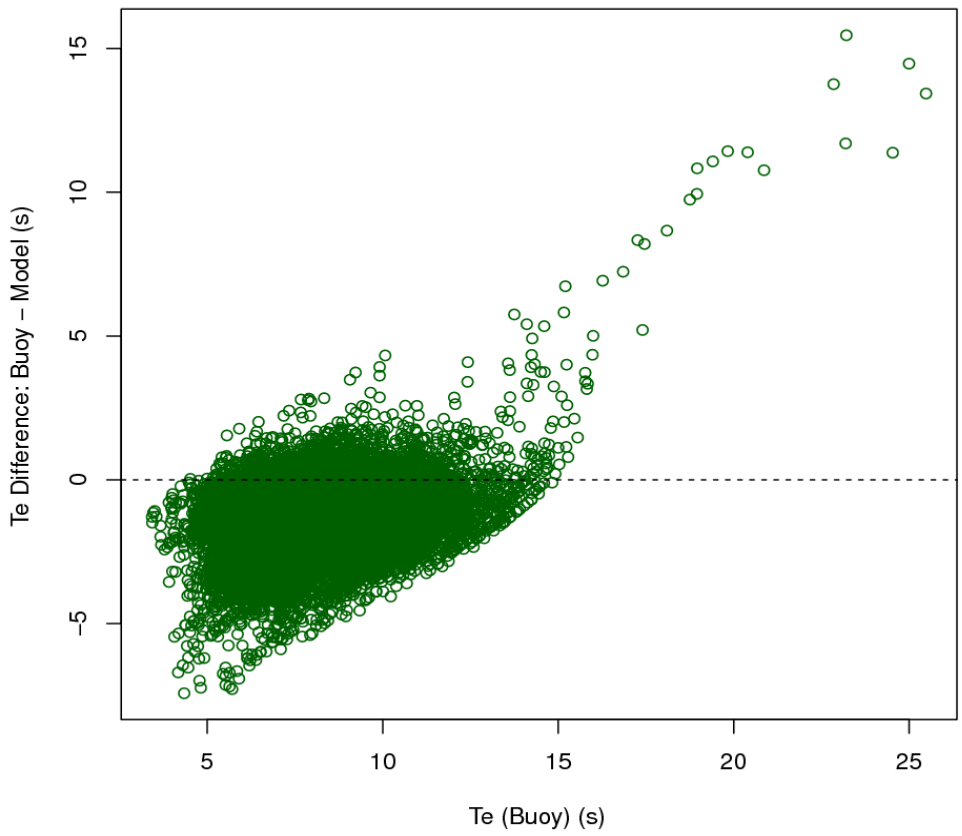

Figure 7: Error scatter plot of T_e, defined as Buoy – Model against Buoy value

If the model is a good fit the errors are expected to be random and normally distributed with a mean of zero. Furthermore, the error scatter plots should demonstrate homoscedasticity. Figure 6 shows a fan-shaped trend—the error becomes larger as the buoy value increases—leading to the conclusions that the errors are not normally distributed and that there is a problem among higher values of H_{m0}. Figure 7 does not show a clear pattern of increasing or decreasing error with buoy value, but it does highlight problems with large energy periods (>15 seconds).

Next, distributions of parameters are examined to inspect overall trends in the datasets. These trends are important because overall bias and differences between buoy and model parameters will cause large errors in the prediction of available wave power. Specifically, kernel methods are used in this study to provide a quantitative analysis of the distributions as a whole. Kernel methods approximate the probability density function of a distribution by

$$f(x) = A\sum_{j=1}^{n} K(x, x_j) \tag{7}$$

where K is the kernel function. A is a constant chosen so that $\int_{-\infty}^{\infty} f(x)dx = 1$. In each kernel method, K depends on the bandwidth, h. The value for h must be such that there is a balance between smoothing unimportant peaks and not smoothing important peaks [32]. The kernel density estimates of buoy and model H_{m0} are shown in Figure 8, and Figure 9 shows the equivalent for T_e.

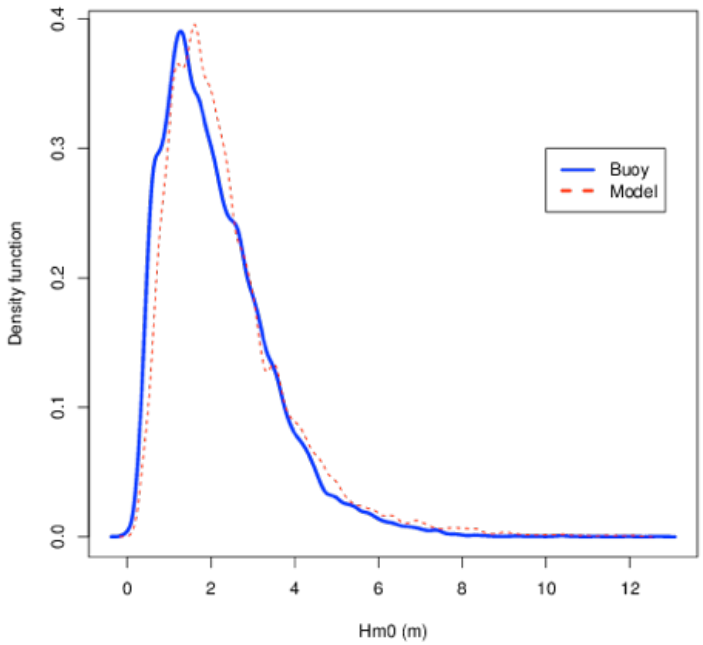

Figure 8: Kernel density estimates of the probability distribution for H_{m0} for both the buoy and model

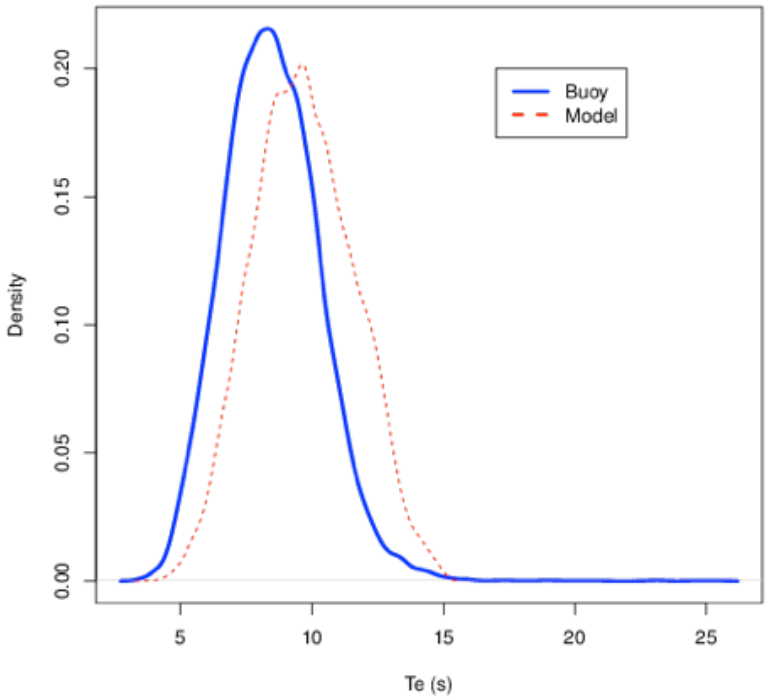

Figure 9: Kernel density estimates of the probability distribution for T_e for both the buoy and model

Figure 8 shows that the model and buoy distributions of H_{m0} are quite similar. There is generally a slight tendency to under-predict values less than about 2 meters and a slight tendency to over-predict values between 2 and 3 meters. Figure 9 suggests that there is a shift in the data for T_e. There is a tendency for the model to under-predict values less than 10 seconds and over-predict values between 10 and 15 seconds.

Next, 95% confidence intervals based on the t-distribution were constructed for the buoy means to see if corresponding model means lie within the intervals. Wang [33] confirms that even when the samples are not normally distributed it is valid to calculate confidence intervals of the mean with a test statistic which requires normality. The results of this test for H_{m0} and T_e are shown in Table 2.

Table 2: One sample t-test confidence intervals for the mean of buoy H_{m0} and T_e kernel density estimator (KDE) values, calculated to see if the respective mean of model KDE values is within this range

	H_{m0}	T_e
95% Confidence Interval	0.0627 – 0.0855	0.0347 – 0.0483
KDE of model mean	0.0783	0.0768
Within 95% CI?	YES	NO

Table 2 shows that the mean of the model H_{m0} kernel density distribution estimates is within the 95% confidence interval of the mean of the buoy H_{m0} kernel density

distribution, whereas the mean of the model T_e kernel density distribution estimates is not within the 95% confidence interval of the mean of the buoy T_e kernel density distribution.

The variances of the kernel density distributions are also compared. Since the distributions are not normal, the F-test cannot be used, so instead Levene's test is employed. Levene's test provides "a test for homogeneity of variance that is insensitive to departures from normality" [34]. The test statistic is calculated by

$$W = \frac{n-m}{m-1} \frac{\sum_{j=1}^{m} n_j \left(\bar{z}_j - \bar{z}\right)^2}{\sum_{j=1}^{m} \sum_{k=1}^{n_j} \left(z_{jk} - \bar{z}_j\right)^2} \tag{8}$$

where

$$z_{jk} = \left| y_{jk} - \bar{y}_j \right| ,$$

y_{jk} is the k^{th} observation in the j^{th} of m samples, and $\bar{y}_j$ is the mean of the n_j observations in this sample.

Testing the hypothesis that the variances of model and buoy KDE functions are equal for H_{m0} gives a test statistic of W = 0.0354 and an associated probability of p = 0.8509. Because 0.8509 > 0.05 there is insufficient evidence at the 5% significance level to reject the null hypothesis. We therefore cannot conclude that the variances are different. For T_e, testing a similar hypothesis that the variances are equal gives a test statistic of W = 0.001 and an associated probability of p = 0.975. Again, because 0.975 > 0.05 there is insufficient evidence at the 5% significant level to reject the null hypothesis, so we cannot conclude that the variances are different.

This analysis shows that for both H_{m0} and T_e the distribution variances are similar but that for T_e the mean of the model predictions is significantly higher than the buoy mean. These results give confidence that the model is working well for H_{m0} but is less reliable for T_e.

Considering our findings from above, it is necessary to investigate the possibility that the bias in T_e was due to a time lag using an analysis of cross- and auto-correlation.

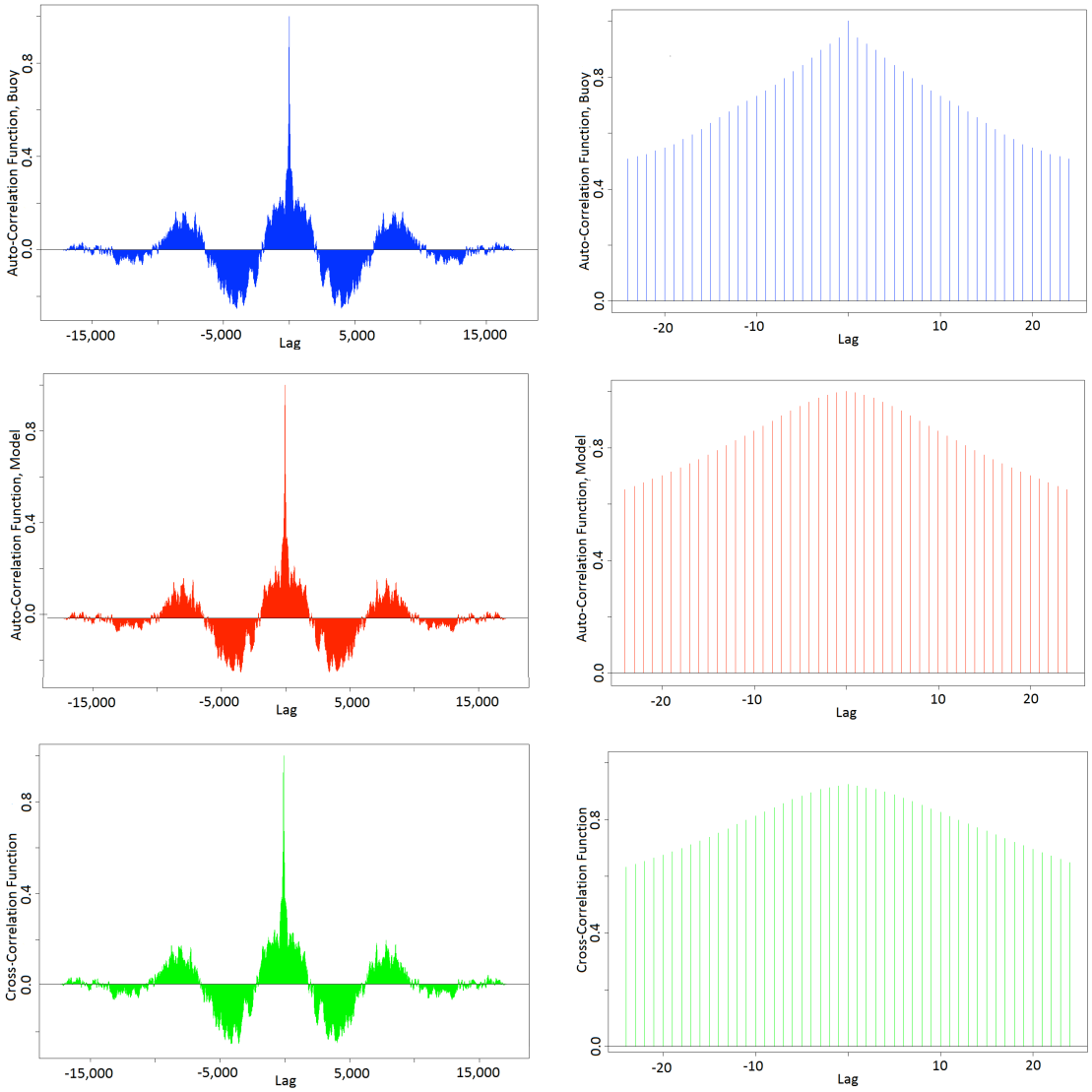

Figure 10: Auto-correlation of H_{m0}: buoy (top), model (middle); Cross-correlation between buoy and model H_{m0} time series (bottom)

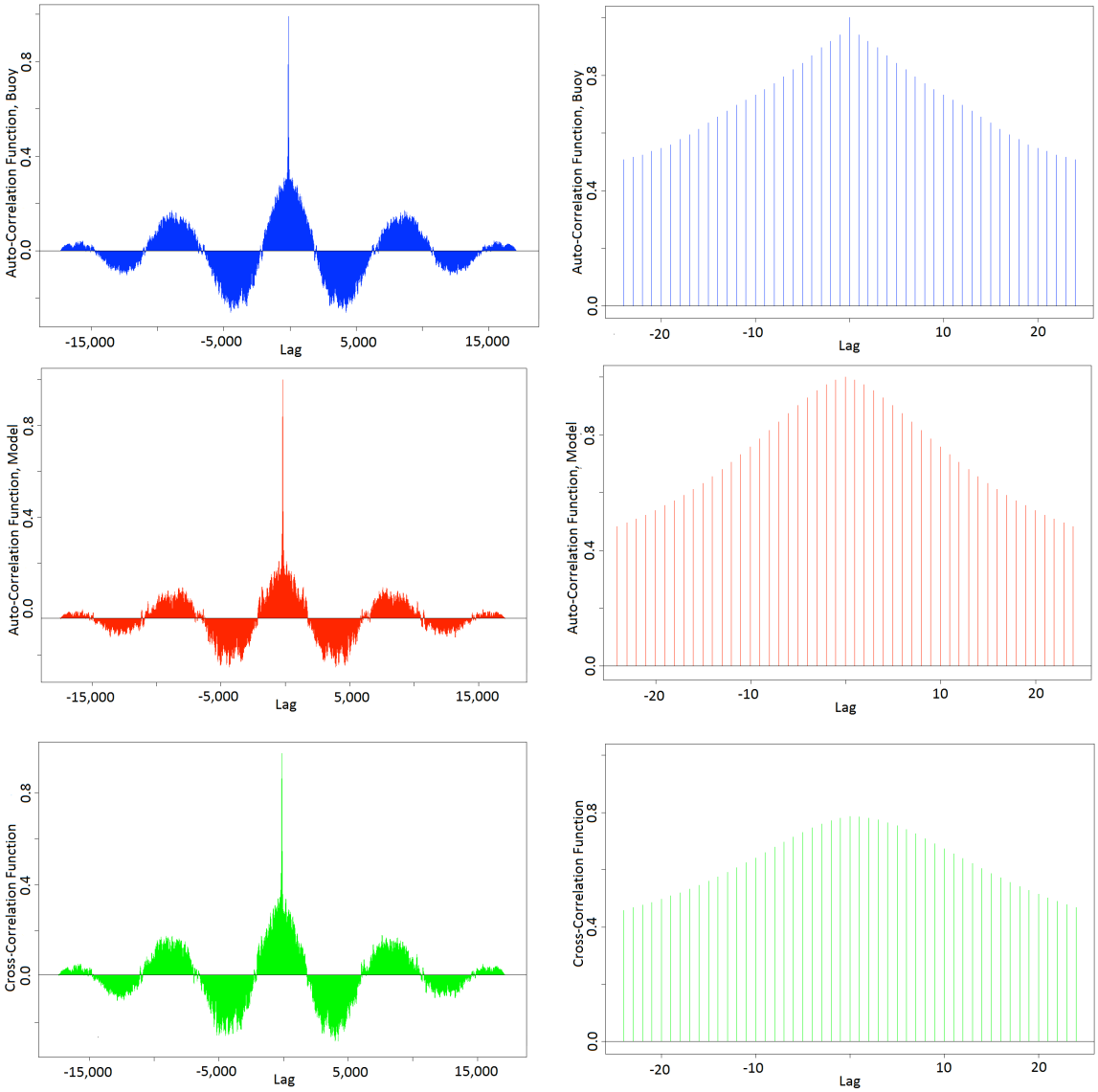

Figure 11: Auto-correlation of T_e: buoy (top), model (middle); Cross-correlation between buoy and model T_e time series (bottom)

Figures 10 and 11 show that there is no time offset in either parameter and that the auto-correlation in the model is sensible. This rules out the possibility that the bias in T_e is due to a time offset.

Finally, statistical models are constructed and analyzed for both H_{m0} and T_e to fully describe the fit of the model to the corresponding buoy values. As mentioned before, the error is not constant for the value of H_{m0}. If a linear model were fitted to this data, the variance would depend on the buoy value, and consequently the errors would not be normally distributed. It is necessary to use a generalized linear model. Error scatter plots must be examined (Figures 6 and 7) before constructing linear models and relying on statistics from these models. If the errors were normally distributed, these plots would be random and consistent, and a linear model would be most appropriate. If the errors are not normally distributed, as seen

in Figure 6 by the fan-shape of data points, a linear model cannot be used, and a generalized linear model must be used.

A linear model is represented by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{9}$$

where y is the response variable, $x_1, x_2, \ldots, x_k$ are the predictor variables, $\beta_0, \beta_1, \ldots, \beta_k$ are unknown weights, and ε is the random error. For a linear model, ε has a normal distribution with an expected value of 0. However, since the error distributions for H_{m0} and T_e are not normal, it cannot be assumed that the expectation is 0.

A generalized linear model is a model whose errors follow a general distribution of the exponential type. The model is given by

$$g(\mu_i) = g\left[E(y_i)\right] = \underline{x_t'}\,\underline{\beta} \tag{10}$$

Where $\underline{x_t}$ is a vector of predictor variables and $\underline{\beta}$ is a vector of unknown weights. Every generalized linear mode (GLM) has an error structure, a linear predictor and a link function [35]. The error structure is the distribution of errors of the response variable. A linear predictor, η, is "a linear sum of the effects of one or more explanatory variables," [29] and is defined by

$$\eta_i = \sum_{j=1}^{k} x_{ij}\beta_j \tag{11}$$

The link function is a function that "connects the linear predictor to the natural mean of the response variable" [35]. It is defined by

$$\eta_i = g(\mu_i) \tag{12}$$

where μ_i is the i^{th} response variable and g is the link function [29].

In this study, it was observed that the variances increase with the value of the predictor value. Consequently, a gamma distribution of errors and a link function of $g(a)=a$ are most appropriate [35].

To compare the GLM to a linear model, the Akaike's Information Criterion (AIC) is calculated. It is given by

$$AIC = -2 \times \log likelihood + 2(p+1) \tag{13}$$

where p is the number of parameters in the model. Smaller values of AIC imply a better fit [29]. The AIC for a linear model and a generalized linear model were compared for both H_{m0} and T_e, and the results are shown in Table 3. The results

show that a GLM with a gamma distribution is more appropriate for H_{m0}, whereas a regular linear model is more appropriate for T_e.

Table 3: Akaike's Information Criterion (AIC) for a linear model (LM) and generalized linear model with gamma error (GLM) for both H_{m0} and T_e

	H_{m0}	T_e
LM	28462.82	53350.24
GLM	20769.28	53750.58

To check the goodness of fit of a model, deviance is calculated [35]. The deviance for a regular linear model is

$$D(\beta) = \sum_{i=1}^{n} (y - \hat{\mu})^2 \tag{14}$$

where $\hat{\mu}$ is the mean of the response variable. For a GLM with gamma error deviance is

$$D(\beta) = 2\sum_{i=1}^{n} \left[-\ln\left(y/\hat{\mu} \right) + (y - \hat{\mu})/\hat{\mu} \right] \tag{15}$$

Myers et al. [35] states that "a good rule of thumb is that lack of fit may be a problem when deviance/(n-p) exceeds 1.0 by a substantial amount." Deviance, degrees of freedom (df), calculated by (n-p), and deviance/df are shown for the two models of parameters in Table 4. The GLM for H_{m0} is a good fit. While the deviance/df for the LM for T_e is not substantially more than 1, it is not as good a fit. This supports previous tests that suggest that the wave model is much better at predicting H_{m0} values than T_e values. These results also rule out the possibility that errors in T_e are due to changing variance.

Table 4: Calculations for deviance, degrees of freedom (df) and Deviance/df for the GLM for H_{m0} and the LM for T_e

	Deviance	df	Deviance/df
H_{m0}	837.6925	16844	0.0497324
T_e	23368.86	16855	1.386465

An alternative way to assess the fit of the model is to look at residuals. For a GLM with gamma errors the residual is defined as $r_i = \frac{y_i - \mu_i}{\mu_i}$, whereas for linear modes r_i is defined in the usual way.

Residuals are examined to see if there are any systematic trends [29]. In the statistical programming language R, model checking plots are slightly different for linear models and for generalized linear models. The plots made to check the GLM for H_{m0} are shown in Figure 12, and Figure 13 shows the linear model plots for T_e.

The first graph (plot a) shows residuals against predicted values from the model. This shows that the variance decreases slightly as the value of the predicted value increases, but overall the fit works well.

The second graph (plot b) is a Normal Q-Q plot. If the errors are normally distributed then the points will lie on the straight line. For a GLM, it is not expected to stay precisely on the straight line because it is not based on a normal distribution, but the points should stay close to it, like they do in this plot.

The third graph (plot c) is a positive-valued version of the first graph [29]. It confirms that there is no heteroscedasticity, which would show up as a triangular scatter.

The fourth graph (plot d) plots the standardized Pearson residual against the leverage (a measure of a point's influence on the model fit) and also shows Cook's distance. Cook's distance "is an attempt to combine leverage and residuals in a single measure" [29]. The main purpose of plot d is to point out very influential points. In this case, the model checking plots drew attention to point numbers 1284, 6442, and 6468.

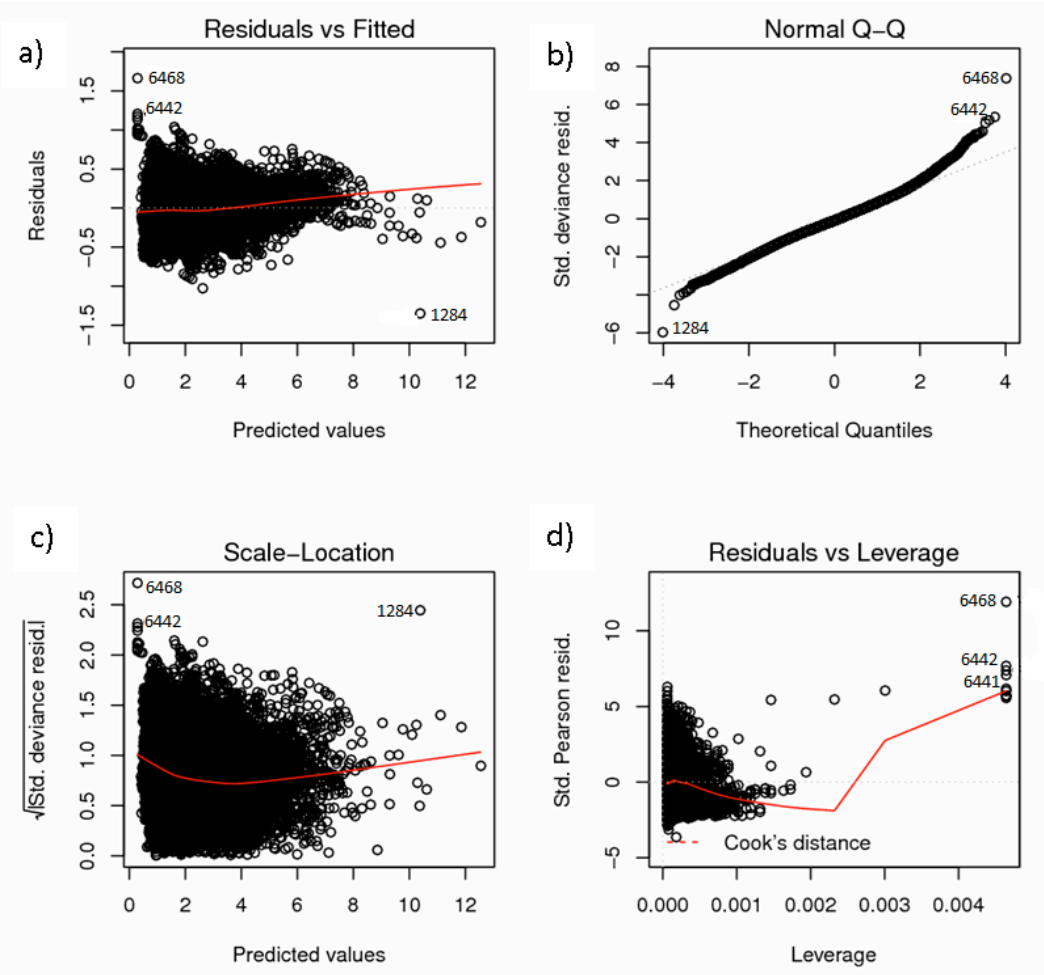

Figure 12: Model checking plots for the GLM for H_{m0}: a) Residuals against predicted values from the model; b) Normal Q-Q plot; c) A positive-values version of plot a; d) Standardized Pearson residual vs. leverage and Cook's distance

The plots to check the linear model for T_e are shown in Figure 13. Plot (a) shows the residuals against the fitted values. Again, this plot checks for unwanted trends or tendencies. It appears that the variance decreases with the fitted values, which is not desirable. This may be because there are extreme outliers so the model tries to match them, consequently causing the low values to be poorly represented. Plot (b) again shows the normal Q-Q plot. This shows large deviations from the straight line at the ends. However, further plots were done using gamma errors and power errors, but these showed even bigger problems and trends. Plot (c) is a standardized, positive version of plot a, and again shows a problem in the extreme outliers Plot (d) shows the standardized residuals against leverage. This model checking draws attention to point numbers 1284, 2295, and 6445. We note that 1284 is also flagged for H_{m0} and is worthy of further analysis. It is necessary to look closely at these flagged points to determine the wave and meteorological conditions corresponding to these points. Part 2 will do an in-depth analysis. It will be shown that at some points, such as 1284, the output from the buoy is unreliable at this time record. Other points reveal systematic problems with the model.

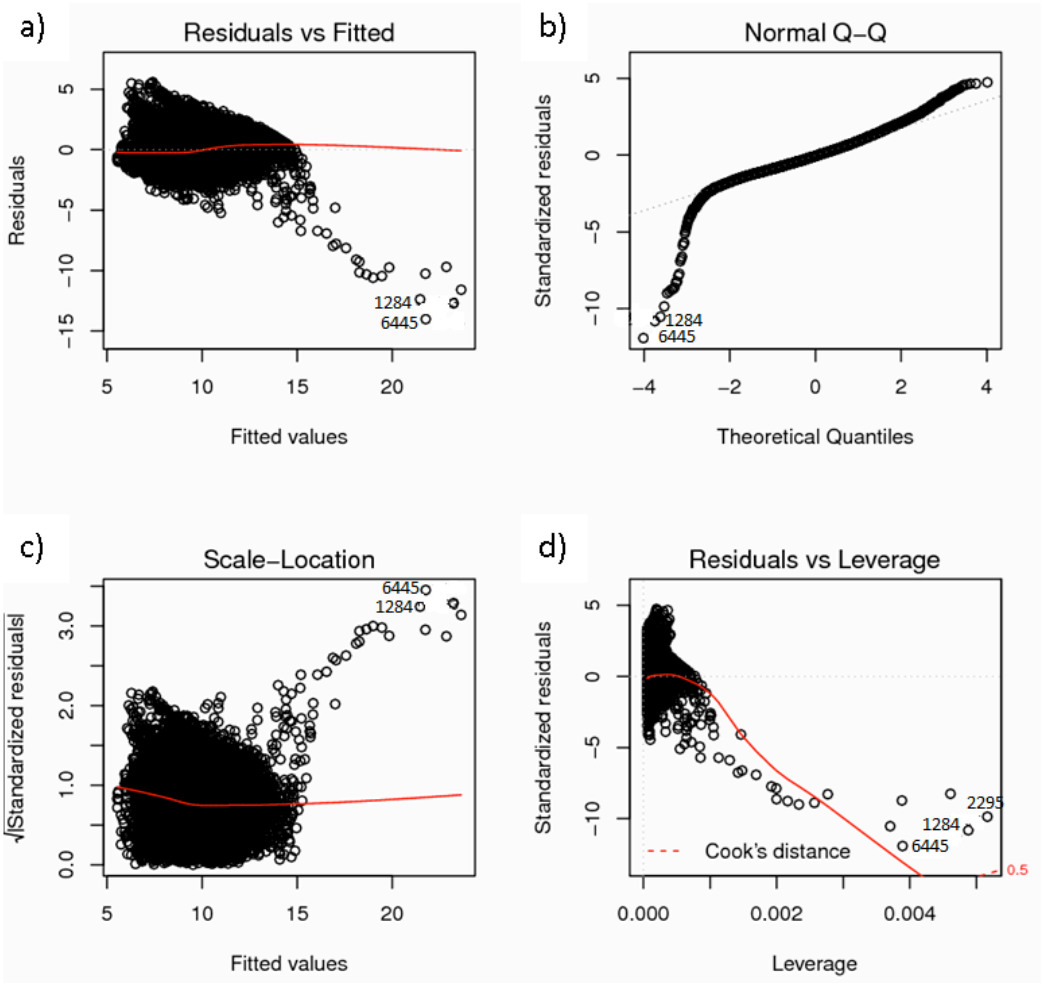

Figure 13: Model checking plots for the LM for T_e: a) Residuals against fitted values from the model; b) Normal Q-Q plot; c) A positive-values version of plot a; d) Standardized Pearson residual vs. leverage and Cook's distance

3. Discussion and Conclusion

If we were to stop the analysis after the basic statistics were calculated in Section 2.2, the conclusion would have been that overall the fit of the model was adequate. However, the results from our further analysis showed that

- the H_{m0} variance is not constant,
- the entire distribution of the model T_e is shifted from the corresponding distribution of the buoy,
- none of the observed differences are due to a timing offset,
- the model does not fit the largest values of H_{m0} or the largest values of T_e well,
- a GLM with gamma error structure is required to fit the model predictions of H_{m0} to the buoy data, and calculations based on linear regression will not be valid, and
- there are problems in the fit of T_e that cannot be explained by any sort of GLM, but the linear model provides a reasonable fit.

Table 5 shows a summary of the methodology used in this study for statistical analysis of wave parameters. The purpose of Part 1 is to show the importance of doing a full statistical analysis of the fit of the wave energy model to the buoy observations.

Table 5: Methodology for the statistical analysis of the fit of our model (WAM3 CY331) to Waverider buoy measurements

- Basic Statistics
  - Correlation coefficient, r
  - Bias
  - Root mean square error, RMSE
  - Scatter Index, SI
  - Standard deviation, s
- Spatial Analysis
  - Taylor diagram
- Further Statistical Analysis
  - Time series
  - Error scatter plots
  - Kernel density distributions
  - Tests to compare distributions
  - Cross-correlation & auto-correlation
- Statistical Modelling
  - Choosing appropriate model
  - Analyze goodness-of-fit and implications
    - Akaike's Information Criterion, AIC
    - Deviance
    - Model checking
      - Analysis of residuals vs. fitted
      - Quantile-Quantile, QQ, plots
      - Positive-valued version of residuals
      - Finding influential points through leverage and Cook's distance

In the statistical analyses done in Section 2.2, it was shown that it is not always best to use the model grid point that is closest to the buoy to calculate the energy potential at the site. From the data used, the closest model grid points to the buoy were the grid points that did not perform as well, probably due to physical restrictions of the location. None of the localized studies mentioned in the introduction (Stopa et al., Liberti et al., and van Nieuwkoop et al.) performed any sort of spatial analysis. Some studies look at model performance locations, such as Stopa et al., who compare different locations around the Hawaiian Islands to different buoys. However, it is important for *each* buoy to look at model results from several surrounding grid points and identify the best fit.

The overall distributions of the parameters and tests of comparisons for these distributions were presented in Section 2.3. The H_{m0} distributions are quite similar, but the T_e model distribution seems to be a shifted version of the buoy T_e distribution. None of the *localized* studies mentioned in the introduction compared distributions of H_{m0} or any period. Liberti et al. [13] showed a frequency distribution for direction but not for the other parameters. A constant over-estimation of the

energy period was observed by examining the distributions. There are multiple reasons why this might be the case. For example, the wave dissipation issues especially in nearshore coastal areas have been discussed in [39]. Since energy period is an important component when assessing the theoretical and technically recoverable available wave energy at a location, it is important to address this problem.

It is important to emphasize how influential the error scatter plots were to this methodology. If regular "model vs. buoy" scatter plots had been employed instead of these error scatter plots, it may have been concluded that most of the errors hover around the identity line and that the model performs reasonably. However, the error plots show heteroscedasticity in the fit of the model H_{m0} to buoy H_{m0} and that the variance was not constant. It was determined that a generalized linear model with gamma error distribution fits better than a linear model, and because of this, calculations based on linear regressions, such as the slope of a linear regression line, would be statistically invalid. For T_e, the model checking plots further confirmed that the predictions of T_e are unreliable. This will again be discussed further in Part 2. All three of the localized studies mentioned in the introduction do at least one scatter plot of model parameters against collocated buoy values. Further, some of these studies find the slope of the linear regression line on these scatter plots and make conclusions from that. This is statistically invalid where the error distribution has not been established to be normal. It is also important to identify the influential points of a statistical model, such as point number 1284, and to perform a detailed investigation. Part 2 of this paper discusses the process of such an investigation.

In performing a resource assessment, the most important outcome is the prediction of power potential for the location. Considering the depth at the buoy, it may be most appropriate to use an intermediate assumption. However, the standard procedure within a wave resource assessment is to use a deep water assumption, and the net wave power per meter crest length (kW/m) can be estimated as

$$P = 0.49 {H_{m0}}^2 T_e \tag{16}$$

For this study, over the two years examined the average power based on buoy measurements was 29.68 kW/m, whereas the average power based on model calculations was 41.59 kW/m. A less complete investigation would have concluded that the model proved a good fit despite this difference in the overall power prediction of 11.91 kW/m. Over-estimating the power is obviously detrimental in resource assessments because it over-estimates the value of the location, leading to a very optimistic estimation of levelised cost of energy. In contrast an under-prediction of wave power could compromise the integrity of the energy converter. It is necessary to accurately predict high-energy sea states to understand the extreme cases the wave energy converter must withstand.

Part 2 of this study closely examines the areas of the model which this investigation revealed to be problematic. It considers the error for different parts of the wave

spectra in order to determine exactly where the wave model is a poor fit to the location data.

Acknowledgements
This work was supported by the EU FP7 project, "MARINA Platform" (Grant agreement number 241402). The authors would like to thank in particular the partners at the Atmospheric Modeling & Weather Forecasting Group at NKUA, for provision of data and advice on analysis. The European Marine Energy Centre (EMEC) are gratefully acknowledged for provision of the Waverider buoy data.

References

[1] Gunn K, Stock-Williams C. Quantifying the global wave power resource. Renewable Energy 2012;44:296-304.

[2] The WAMDI Group. The WAM model—a third generation wave prediction model. Journal of Physical Oceanography 1988;18:1775-1810.

[3] Cavaleri L, Bertotti L, Lionello P. Shallow water application of the third-generation WAM wave model. Journal of Geophysical Research 1989;94:8111-8124.

[4] Ris RC, Holthuijsen LH, Booij N. A third-generation wave model for coastal regions: 2. verification. Journal of Geophysical Research 1999;104:7667-7681.

[5] Mørk G, Barstow S, Kabuth A, Pontes MT. Assessing the global wave energy potential. Proceedings of the 29th International Conference on Ocean, Offshore Mechanics and Arctic Engineering 2010.

[6] Arinaga RA, Cheung KF. Atlas of global wave energy from 10 years of reanalysis and hindcast data. Renewable Energy 2012;39:49-64.

[7] ESBI Engineering & Facility Management Ltd. Accessible Wave Energy Resource Atlas: Ireland: 2005. Available at: http://www.marine.ie/NR/rdonlyres/90ECB08B-A746-4247-A277-7F9231BF2ED2/0/waveatlas.pdf; 2005 [accessed 19.8.2013].

[8] ABP Marine Environmental Research Ltd. Atlas of UK Marine Renewable Energy Resources: Technical Report. Available at: http://www.renewables-atlas.info/downloads/documents/R1432_Final_15May08.pdf; 2008 [accessed 19.8.2013].

[9] Vengatesan V, Davey T, Girard F, Smith H, Smith G, Cavaleri L, Bertotti L, Lawrence J. Equitable Testing and Evaluation of Marine Energy Extraction Devices in terms of Performance, Cost and Environmental Impact (EquiMar): Deliverable D2.3 Application of Numerical Models. Available at: https://www.wiki.ed.ac.uk/display/EquiMarwiki/EquiMar;jsessionid=E74D286B9C35837A252BD33B4836D848; 2010 [accessed 19.8.2013].

[10] Iglesias G, Carballo R. Wave resource in El Hierro—an island towards energy self-sufficiency. Renewable Energy 2011;36:689-698.

[11] Lenee-Bluhm P, Paasch R, Özkan-Haller HT. Characterizing the wave energy resource of the US Pacific Northwest. Renewable Energy 2011;36:2106-2119.

[12] Stopa JE, Filipot JF, Li N, Cheung KF, Chen YL, Vega L. Wave energy resources along the Hawaiian Island chain. Renewable Energy 2013;55:305-321.

[13] Liberti L, Carillo A, Sannino G. Wave energy resource assessment in the Mediterranean, the Italian perspective. Renewable Energy 2013;50:938-949.

[14] van Nieuwkoop JCC, Smith HCM, Smith GH, Johanning L. Wave resource assessment along the Cornish coast (UK) from a 23-year hindcast dataset validated against buoy measurements. Renewable Energy 2013;58:1-14.

[15] Bertotti L, Cavaleri L. Modelling wave at Orkney coastal locations. Journal of Marine Systems 2012;96-97:116-121.

[16] P. Janssen, “Chapter 3 ECMWF wave modeling and satellite altimeter wave data,”in in *Satellites, oceanograghy and society*, vol. Volume 63, D. H. B. T.-E. O. Series, Ed. Elsevier, 2000, pp. 35–56.

[17] Janssen PAEM. The interaction of ocean waves and wind. Cambridge University Press; 2004.

[18] WAMDIG, The WAM-Development and Implementation Group: Hasselmann S, Hasselmann K, Bauer E, Bertotti L, Cardone CV, Ewing JA, Greenwood JA, Guillaume A, Janssen PAEM, Komen GJ, Lionello P, Reistad M, Zambresky L. The WAM model - a third generation ocean wave prediction model. Journal of Physical Oceanography 1988;18:1775-1810.

[19] Komen G, Cavaleri L, Donelan M, Hasselmann K, Hasselmann S, Janssen PAEM. Dynamics and modelling of ocean waves. Cambridge University Press; 1994.

[20] Bidlot J, Janssen P, Abdalla S, Hersbah H. A revised formulation of the ocean wave dissipation and its model impact. ECMWF Tech Memo. Reading, United Kingdom; 2007; 509.

[21] Janssen PAEM, Onorato M. The Intermediate Water Depth Limit of the Zakharov Equation and Consequences for Wave Prediction. J.Phys.Oceanogr. 2007;37:2389-2400.

[22] Kallos G, 1997: The Regional weather forecasting system SKIRON. Proceedings of the Symposium on Regional Weather Prediction on Parallel Computer Environments, 15-17 October 1997, Athens, Greece.

[23] Spyrou C, Mitsakou C, Kallos G, Louka V, Vlastou G. An improved limited area model for describing the dust cycle in the atmosphere, Journal of geophysical research 2010, 115, D17211, doi: 10.1029/2009JD013682.

[24] SWAN user manual. Delft University of Technology; 1993-2010.

[25] Datawell. Datawell File Formats. Available at: http://www.datawell.nl/Portals/0/Documents/TechnicalNotes/datawell_technicalnote_fileformats_2013-03-14.pdf; 2013 [accessed 19.8.2013].

[26] Mackay EBL, Bahaj AS, Challenor PG. Uncertainty in wave energy resource assessment. Part 1: Historic data. Renewable Energy 2010;35:1792-1808.

[27] U.S. Army Engineer Waterways Experiment Station, Coastal Engineering Research Center. Coastal engineering technical note: Directional wave spectra using normal spreading function. Available at: http://chl.erdc.usace.army.mil/library/publications/chetn/pdf/cetn-i-6.pdf; 1985 [accessed 19.8.2013].

[28] Taylor KE. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research 2011;106:7183-7192.
[29] Crawley MJ. The R book. Chichester, West Sussex, UK: John Wiley & Sons Ltd; 2007.

[30] Portilla J, Sosa J, Cavaleri L. Wave energy resources: Wave climate and exploitation. Renewable Energy 2013;57:594-605.

[31] Janssen PAEM, Abdalla S, Hersbach H, Bidlot JR. Error estimation of buoy, satellite, and model wave height data. Journal of Atmospheric and Oceanic Technology 2007;24:1665-1676.

[32] Duong T. Package 'ks' Available at: http://cran.r-project.org/web/packages/ks/ks.pdf; 2013 [accessed 19.8.2013].

[33] Wang FK. Confidence interval for the mean of non-normal data. Quality and Reliability Engineering International 2001;17:257-267.

[34] Upton G, Cook I. Dictionary of Statistics. Oxford: Oxford University Press; 2002.

[35] Myers RH, Montgomery DC, Vining GC. Generalized Linear Models. New York: John Wiley & Sons, Inc.; 2002.

[36] Devore J, Farnum N. Applied Statistics for Engineers and Scientists. Pacific Grove, CA, USA: Brooks/Cole Publishing Company; 1999.

[37] Hayter AJ. Probability and Statistics for Engineers and Scientists. Boston, USA: PWS Publishing Company; 1996.

[38] Pontes MT, Aguiar R, Pires HO. A nearshore wave energy atlas for Portugal. Proceedings of the 22nd International Conference on Offshore Mechanics and Arctic Engineering 2003.

[39] The WISE Group. Wave modelling – the state of the art. Progress in Oceanography 2007;75:603-674.

Appendix A: Statistical Parameters

The first statistic calculated is the correlation coefficient, ***r***. r^2, the coefficient of determination, is not used because normality of errors and linearity cannot be assumed [27, 34].When considering the model as an estimator of the buoy, it is helpful to determine if it is a biased estimator. The average overall difference, or the ***bias***, is defined by

$$Bias = \frac{1}{n}\sum_{j=1}^{n}(y_j - x_j) \tag{A.1}$$

where x_j is the j^{th} buoy value, y_j is the j^{th} model value, and n is the number of observations [34]. A high bias indicates that there is a tendency in the predicted data to either over- or under-estimate the observed data. Random errors would cancel out if they were present, but an overall *tendency* for an error is represented by the bias.

Root mean square error, ***RMSE***, is widely used in verification sections [35]. In contrast to bias, RMSE just represents the actual distances between observed and predicted values, rather than both distance and direction (positive or negative) of the distance. A low value would indicate that, in general, the collocated observed and predicted values are close to one another. If distances between corresponding values were generally large but equally large on either side of observed values, the bias would be low but RMSE would still be high.

Ris et al. [4] use the scatter index, ***SI***, to "quantify the performance of ocean wave models" by normalizing the RMSE with respect to the mean of the buoy observations ($\bar{x}$). Because it is also used in multiple other validation studies [e.g. 14, 36, 12, 13], it is included here, even though Ris et al. [4] mentions that it appears "to understate the skill of the model."

Finally the standard deviations, ***s***, of each dataset are compared [34].

Appendix B: Taylor Diagrams

For the RMSE, it is necessary to separate "the differences in the patterns from the differences in the means of the two fields" [22] by separating it into two components:

$$E^2 = \overline{E}^2 + E'^2 \tag{B.1}$$

Where

$$\overline{E} = \overline{y} - \overline{x} \tag{B.2}$$

Is the difference between the mean of the model dataset and the mean of the buoy dataset, and

$$E' = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left[\left(yj - \overline{y}\right) - \left(x_j - \overline{x}\right)\right]^2} \tag{B.3}$$

Is the "centred pattern RMS difference" [22]. *E'* approaches zero as the model and buoy become more alike.

A ratio of variances, or sample standard deviations in this case, is also included in the Taylor diagram. These are represented by s_y and s_x for the model and buoy, respectively. The main concept that enables a Taylor diagram to work is the relationship between these standard deviations, the pattern RMSE, and the correlation coefficient:

$$E'^2 = {s_y}^2 + {s_x}^2 - 2 s_y s_x r \tag{B.4}$$

And the law of cosines. These are combined to show the geometric relationship in Figure B.1.

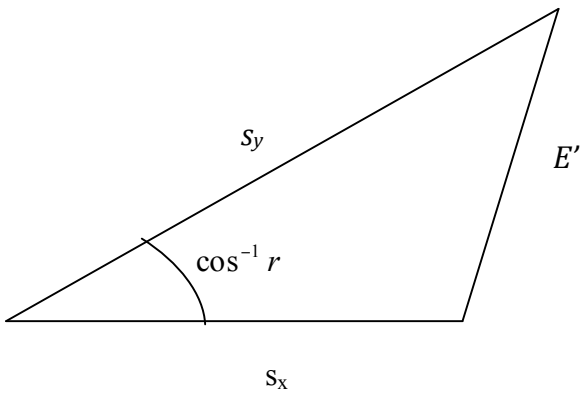

Figure B.1: Geometric relationship that enables a Taylor diagram to work

From this geometric relationship, the Taylor diagram is made by comparing two fields. One is the reference field, representing the state of the buoy, and the other is the test field, representing the state of the model. In the Taylor diagram, “the radial distances from the origin to the points are proportional to the pattern standard deviations, and the azimuthal positions give the correlation coefficient between the two fields. The radial lines are labelled by the cosine of the angle made with abscissa, consistent with Figure B.1. The dashed lines measure the distance from the reference point and, as a consequence of the relationship shown in Figure B.1, indicate the [RMSE] (once any overall bias has been removed)” [22].