

# Short-term Wave Forecasting as a Univariate Time Series Problem

Francesco Fusco

December 2009

Research Report: EE/2009/3/JVR  
Department of Electronic Engineering,  
NUI Maynooth  
Maynooth, Co. Kildare, Ireland

# Contents

<b>1</b>	<b>The univariate approach to wave forecasting</b>	<b>2</b>
<b>2</b>	<b>Data analysis</b>	<b>3</b>
2.1	Wave spectra . . . . .	3
2.2	Non-stationarity and the Wavelet transform . . . . .	4
2.3	Linearity analysis . . . . .	7
2.4	Predictability measure . . . . .	10
2.5	Choice of cut-off frequency . . . . .	17
2.6	Choice of the sampling frequency . . . . .	20
<b>3</b>	<b>Methodology</b>	<b>21</b>
3.1	Getting the observations of the signal to predict . . . . .	21
3.2	Forecasting models . . . . .	23
3.2.1	Cyclical models . . . . .	23
3.2.2	Auto Regressive (AR) models . . . . .	25
3.2.3	Sinusoidal extrapolation and the Extended Kalman Filter	28
3.2.4	Neural networks . . . . .	30
3.3	Confidence intervals . . . . .	31
<b>4</b>	<b>Results</b>	<b>32</b>
4.1	Cyclical models . . . . .	32
4.2	Auto-Regressive models . . . . .	39
4.2.1	Choice of model order . . . . .	39
4.2.2	Static AR models . . . . .	39
4.2.3	Adaptive AR models . . . . .	42
4.3	Sinusoidal extrapolation through Extended Kalman Filter (EKF)	47
4.4	Neural networks . . . . .	49
4.5	Effects of sampling frequency on prediction . . . . .	55
<b>5</b>	<b>Further possibilities</b>	<b>57</b>
5.1	Gaussian Processes . . . . .	57
5.2	Auto Regressive Moving Average (ARMA) models . . . . .	59
<b>6</b>	<b>Conclusion</b>	<b>59</b>

# 1 The univariate approach to wave forecasting

This report aims to collect and present the main results and conclusions due to the work done by the author on the problem of forecasting the ocean wave elevation at a specific point of the sea surface based on past observations collected at the same point. The problem is strictly connected to the real time control, on a wave by wave basis, of a Wave Energy Converter (WEC), which can, potentially, significantly increase its efficiency and energy capture ability, but that is well known to be described by non-causal relationships in the time domain [1],[2],[3].

The wave elevation  $\eta(k)$ , observed at a particular point with a certain sampling frequency, is treated as a pure univariate time series, so that the forecasting problem consists of determining the prediction  $\hat{\eta}(k+l/k)$  a number of steps,  $l$ , ahead based on all the information up to the current sampling instant  $k$ . All the solutions proposed so far in the literature deal with the problem by trying to reconstruct the wave field at a certain point of the sea surface based on one or more distant measurements [4],[5],[6],[7], as shown in figure 1(b). This approach, however, requires complex numerical models and a large enough array of measurements in order to properly deal with the non-linearities of wave propagation, including wave refraction and multidirectionality.

Alternatively, the solution discussed here, based only on local measurements of the wave elevation (or of any other related quantity of interest, such as the wave excitation force), as illustrated in figure 1(a), allows for certain significant advantages:

- ↑ multidirectionality and all the associated complications need not to be considered;
- ↑ the wave propagation laws do not need to be modelled and no simplifying assumptions (e.g. linearity, dispersion relationship) are then required about them;
- ↑ if the considered point corresponds to the position of the WEC, the radiated waves do not affect the measurements;
- ↑ all the well established theory about univariate time series forecasting may be exploited;
- ↑ no additional instrumentation around the device is required (cheap solution).

Its validity, however, is limited by the possibility to effectively estimate the wave elevation at the point where the device is located (this problem and some related implications are briefly discussed in section 3.1).

The available data and its detailed analysis through different tools is presented in section 2. Then the actual solution to produce the predictions is proposed in the methodology of section 3 and evaluated on real wave data in the results section 4. Some other possible forecasting models, for which results were not produced, and motivations for their unsuitability are discussed in section 5. Conclusions are finally presented in section 6.

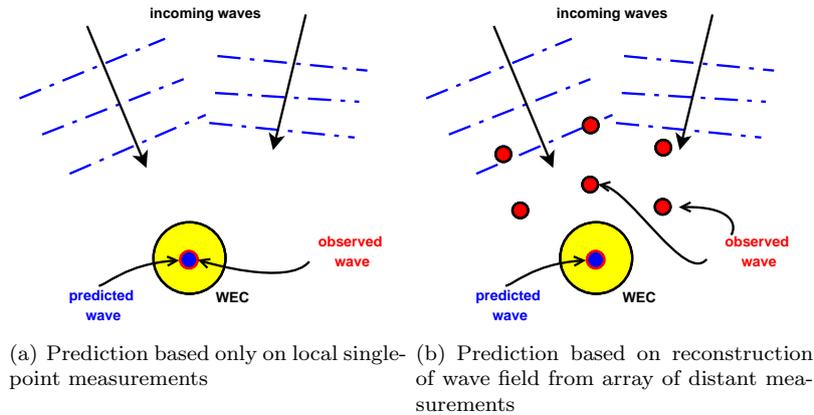


Figure 1: Two main approaches to wave forecasting

## 2 Data analysis

The data available comes from different locations.

The Irish Marine Institute provided real observations from a data buoy located in Galway Bay, on the West Coast of Ireland (at approximately  $53^{\circ}13'N$ ,  $9^{\circ}18'W$ ). These observations consist of 20 minute records sets for each hour, collected at a sampling frequency of  $2.56\text{ Hz}$ , for parts of years 2007 and 2008. The location is sheltered from the Atlantic Ocean so that the wave height magnitude is generally small, which makes it not an ideal site for full size WECs, though a wave energy test site has been established there for 1/4-scale prototypes.

Wave elevation time series are also available from the Atlantic Ocean at the Pico Island, in the Azores archipelago, at approximately  $38^{\circ}33'N$ ,  $28^{\circ}34'W$ . They are collected in the form of two contiguous 30 minute record sets for each hour, with a sampling frequency of  $1.28\text{ Hz}$  (that is 2304 samples for each set).

### 2.1 Wave spectra

The main tool for a first analysis of waves is their spectral distribution, the *wave spectrum*, which shows how the energy is distributed across the different frequency components of the wave, assumed to be completely independent of each other. Although offering limited time-averaged information (a Wavelet transform would offer a more complete information in the time domain, refer to section 2.2) it is still very valuable in order to provide some overall characteristics of the sea conditions in different situations and at different locations.

A first analysis, which is interesting to carry out over the available hourly data sets, concerns the distribution of the significant wave height  $H_s$  and the peak and mean radian frequency of the spectrum, respectively  $\omega_{peak}$  and  $\omega_{mean}$ , and to assess if their behaviors are correlated to each other in some way. The significant wave height is a measure of the mean energy contained in the wave, while  $\omega_{peak}$  and  $\omega_{mean}$  can be a way to represent where the spectrum (and so the energy) of the wave is more concentrated. From figure 2, it is clear how high energy wave systems present a much narrower spread of  $\omega_{peak}$  and  $\omega_{mean}$ , centered at a low frequency (about  $1\text{ rad/s}$  for the Galway Bay data,

even lower for the wave systems from Pico), which means that the energy is more concentrated at the low frequencies and the spectral distribution has a well defined narrow peak (swell). The lower the energy, on the other hand, the more the distance between the peak and the mean frequency, which denotes a much flatter spectrum where the high frequency wind waves have a similar energy content to the low frequency swell. The sample spectra of figure 3 are particularly illustrative in this respect.

## 2.2 Non-stationarity and the Wavelet transform

A better and more complete understanding of the energy distribution in the waves can only be achieved through an analysis in both the frequency domain and the time domain, that is something like a continuous evolution of the wave spectrum over time. An interesting tool that has been successfully applied to this proposal is the Wavelet transform (references [8],[9]), which shows a good resolution both in time and frequency, thus making the application of the short-term Fourier transform, where only a compromise between time and frequency resolutions can be obtained, less attractive for waves analysis.

Wavelets have been successfully implemented in signal and image processing, ordinary and partial differential equation theory, numerical analysis and communication theory [9]. On the other hand, the application of the wavelet transform to ocean engineering and oceanography is not frequent. This is mostly due to the fact that not all such applications provide quantitative results, so that the wavelet transform has been regarded as an interesting tool to produce colorful pictures, yet purely qualitative results [9].

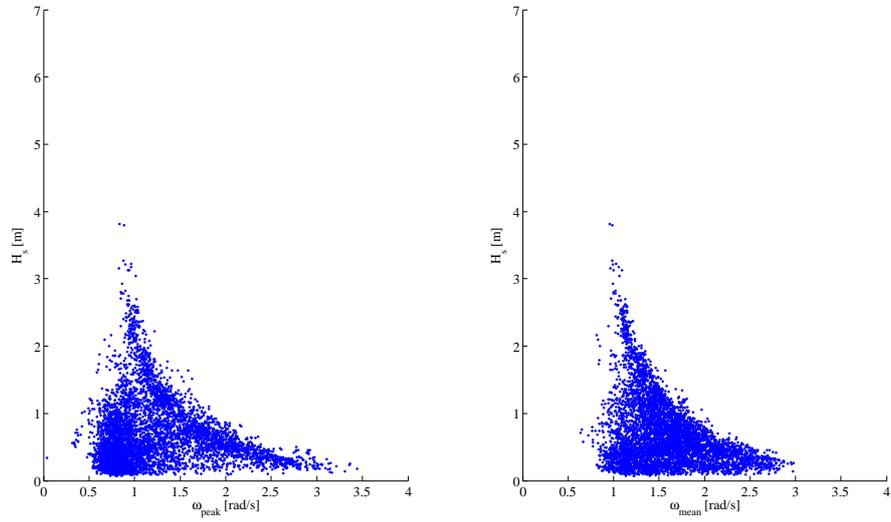
The Wavelet transform of a signal,  $x(t)$ , is defined from the following expression:

$$WT(t, b) = \int_{-\infty}^{+\infty} x(t)g^*(t; \tau, b)d\tau \quad (1)$$

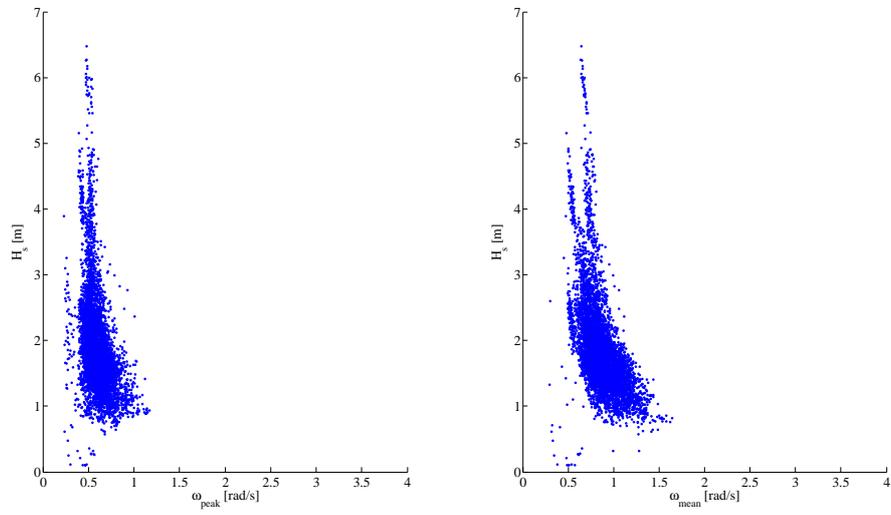
Here,  $g^*(t; \tau, b)$  is the complex conjugate of a continuously translated and dilated mother wavelet function  $g(t)$ :

$$g(t; \tau, b) = \frac{1}{\sqrt{b}}g\left(\frac{t - \tau}{b}\right) \quad (2)$$

where  $t$  is the translation parameter, corresponding to the position of the wavelet as it is shifted through the signal,  $b$  is the scale dilation parameter determining the width of the wavelet. At low frequencies (high value of scale  $b$ ), the frequency resolution is better but the time resolution is poor (more ambiguity regarding the exact time). On the other hand, at higher frequencies (low scale  $b$ ), the frequency resolution is poorer and the time resolution is better. This main characteristic of the Wavelet transform, which is due to the fact that the signal is multiplied with a window whose width is changed as the transform is computed for each spectral component [9], is a significant improvement to the Short Term Fourier Transform (STFT), where the window is constant and an appropriate compromise has to be made between time and frequency resolution. In fact, a finer time resolution at higher frequencies is important because the signal is changing faster, while a poorer time resolution at low frequencies can be acceptable because the signal is changing more slowly.



(a)



(b)

Figure 2: Correspondence between significant wave height  $H_s$ , peak frequency  $\omega_{peak}$  and mean frequency  $\omega_{mean}$  at the locations: (a) Galway; (b) Pico.

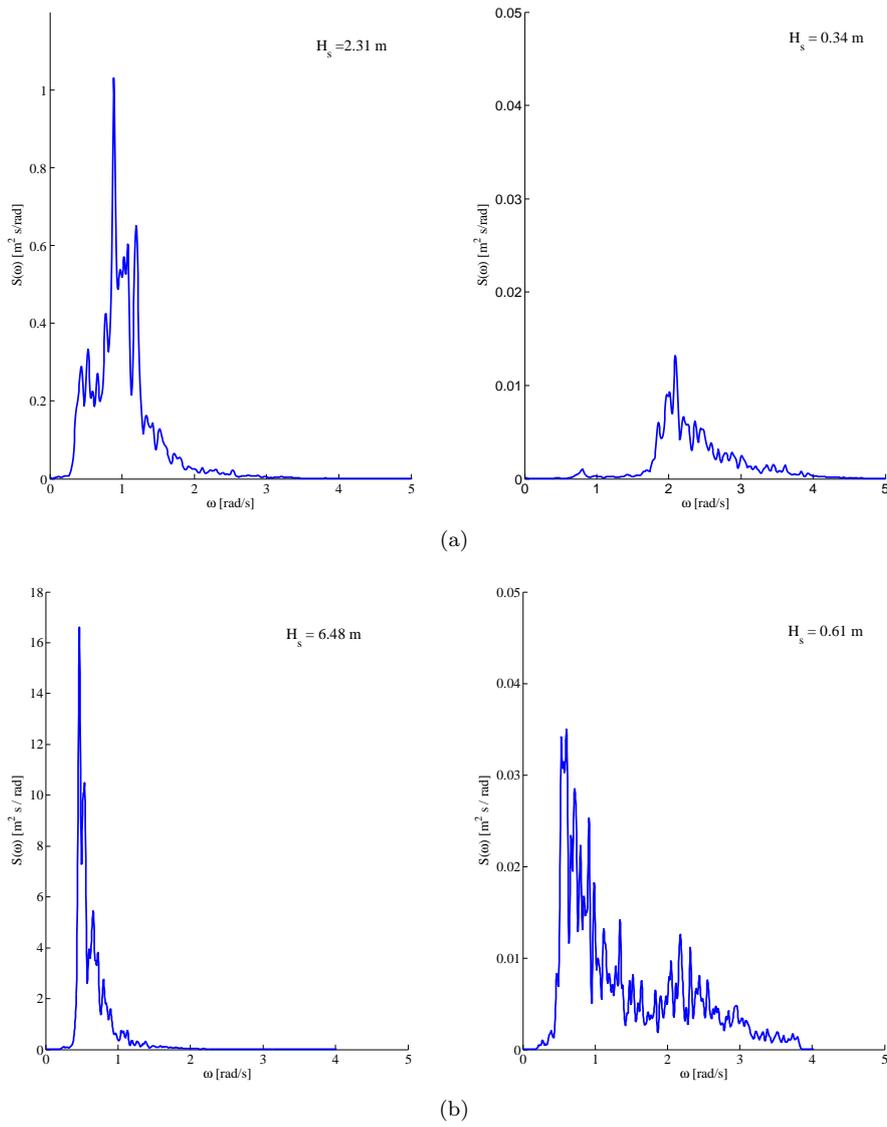


Figure 3: Typical high and low energy spectra at the locations: (a) Galway; (b) Pico.

A correspondence between the scale value  $b$  and the Fourier period  $T$  can be found, and it depends on the specific mother wavelet,  $g(t)$ , chosen. In the case of the very common *Morlet wavelet*, the following expression can be derived [9]:

$$b = \frac{c + \sqrt{c^2 + 2}}{4\pi} T \triangleq \alpha T \quad (3)$$

where  $c$  is a parameter defining the Morlet mother wavelet, also having the nature of a frequency in some sense. Note that the physical dimension of  $b$  is the time (*seconds*). Expression (3) can then be used to have physical meaning (the frequency) of the scale dimension of the transform  $WT(t, b)$ , which can then become a  $WT(t, f)$  or  $WT(t, \omega)$  in a very straightforward way.

Figures 4 and 5 show the Wavelet transform for some wave elevation data sets, respectively from Galway Bay and from Pico. Although, as mentioned, it is not easy to derive any quantitative result from them, it is still possible to get some interesting information out of them. Note, in fact, how the different frequency components of the Fourier spectrum may appear in different moments so that, in the short term, the bandwidth of the wave signal could actually be much narrower than what the Fourier transform suggests.

### 2.3 Linearity analysis

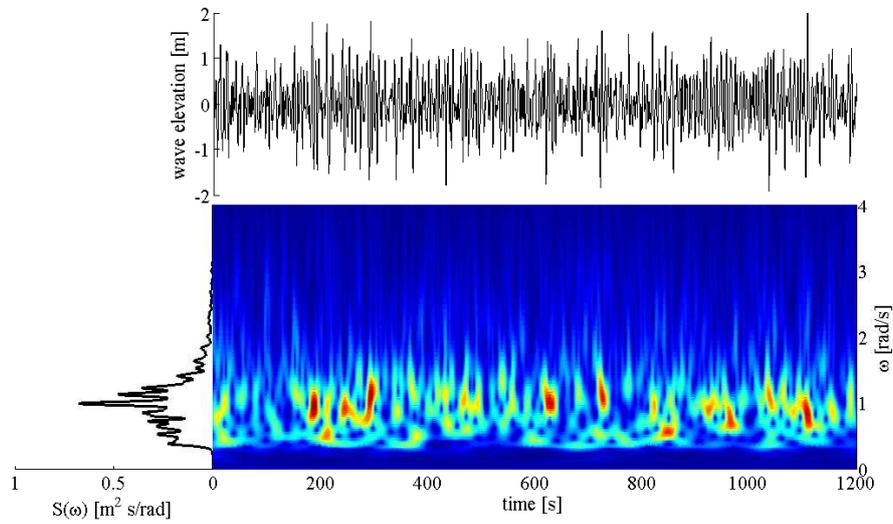
Ocean waves, like most systems in the real world, are not linear, and it would be helpful and valuable to quantify how far from linearity they are so that, in the particular case of wave forecasting, a proper model can be chosen. Linearity in the case of waves means linear superposition of harmonic components (sines and cosines).

The emerging of non-linearities in waves manifests itself, in the first instance, by a non-Gaussian distribution of the wave elevation time series around its mean value (zero, the water surface level), due to the presence of higher and narrower peaks than troughs (modelled by quadratic, cubic, ... and so on, terms of the linear harmonic components, according to Stokes theory [10]). The degree of asymmetry depends on the significance of the water depth with respect to the wavelength (the difference between the greatest elevation and the greatest depression is minimum for  $h \gg \lambda$  and maximum for shallow water). This non-linearity is expected, therefore, to be more consistent at the high energy and low frequency components of the waves when the water depth is not sufficiently large. A higher order statistical analysis, through the indices of kurtosis and skewness, would be useful to assess the extent of this non-linearity. Skewness and kurtosis, in fact, can determine the degree of Gaussianity of the distribution of the wave elevation around the mean water surface level, and statistical analysis of the time history of wave records indicates that the wave profiles are normally distributed apart some very small deviations on rare occasions [11].

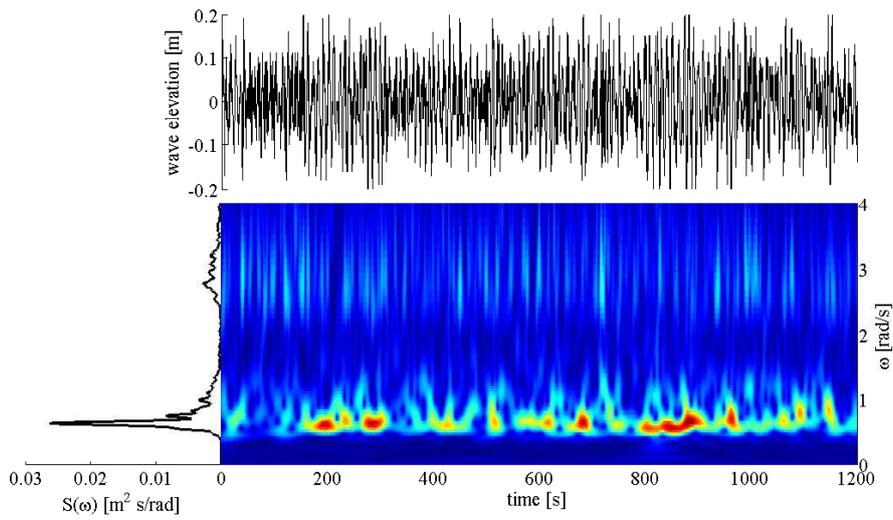
Given the  $n$ th-order central moments,  $\mu_n$ , of a certain random variable  $x$ :

$$\mu_n = E\{(x - \mu)^n\} \quad (4)$$

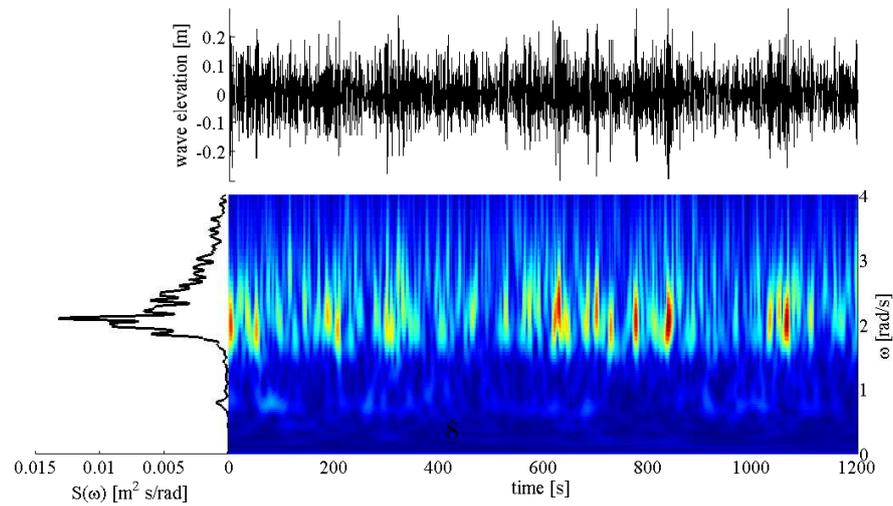
where  $E\{\cdot\}$  is the expectation operator and  $\mu$  is the mean  $E\{x\}$  of the random variable  $x$ , then the two indices of kurtosis,  $\kappa$ , and skewness,  $\gamma$ , are given by



(a)

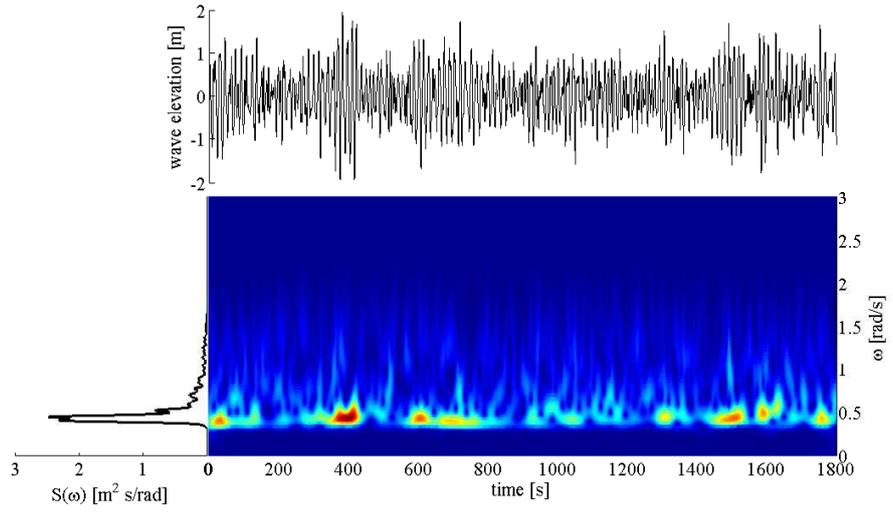


(b)

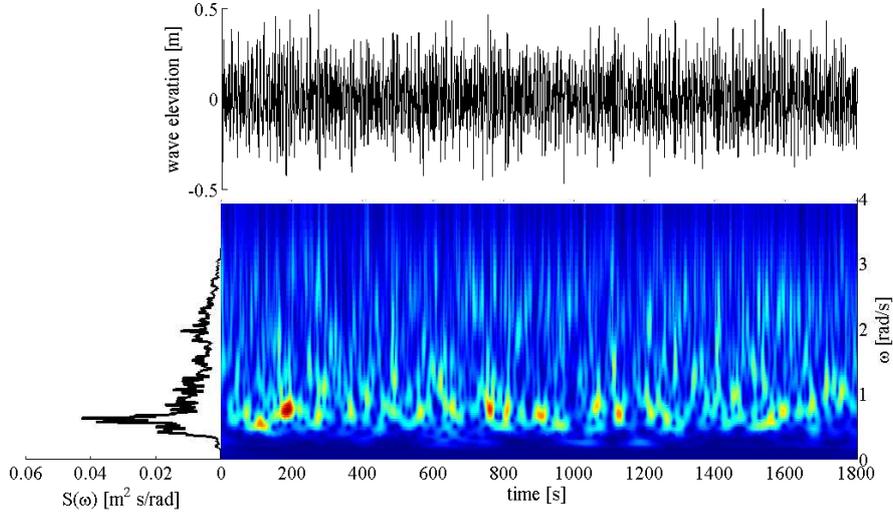


(c)

Figure 4: Wavelet transform for three different data sets from Galway Bay



(a)



(b)

Figure 5: Wavelet transform for three different data sets from the Pico Island.

[12]:

$$\gamma = \frac{\mu_3}{\sigma^3} \quad (5)$$

$$\kappa = \frac{\mu_4}{\sigma^4} \quad (6)$$

where  $\sigma^2 \equiv \mu_2$  is the 2nd-order central moment. The skewness measures the asymmetry of the distribution:  $\gamma = 0$  denotes a symmetric distribution, otherwise, if  $\gamma > 0$ , the distribution is more concentrated around a value greater than the mean and viceversa if  $\gamma < 0$ . The kurtosis represents a degree of peakedness as compared to a Gaussian distribution, in which case  $\kappa = 3$ : if  $\kappa > 3$  the distribution is termed *leptokurtic* (sharp peak), otherwise, if  $\kappa < 3$ , it is termed *platykurtic* (mild peak) [12].

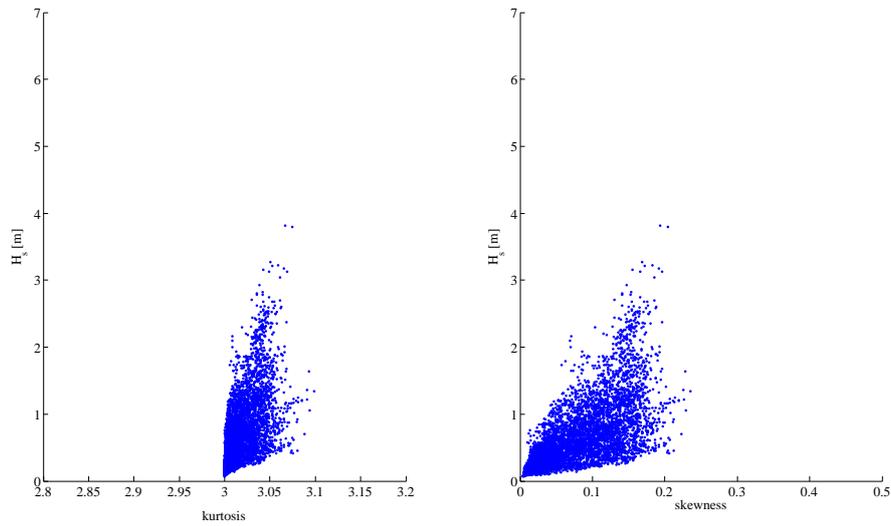
Figures 6 and 7 show this analysis for the two locations of Galway Bay and the Pico Island. As expected, higher energy data sets show a slight deviation of the indices of skewness and kurtosis from the normality condition, particularly in the case of the Galway Bay, where the water depth is smaller (nearly 20 m), while in the Pico Island only very high energy waves move away from normality, as shown in the detailed wave distribution of figure 7(c). From a wave energy point of view, although the interest is obviously focused on high energy waves, the non-symmetry effect may not be an issue if deep water locations are considered.

There is, however, another possible non-linearity, which unfortunately is less quantifiable and can only be analysed through visual inspection. This is due to the interactions occurring between different harmonic components of the wave system, which are neglected in the classical linear wave theory and in the Fourier-Wavelet analysis. A higher order analysis through the Bispectrum (refer to Ochi [11]) revealed to be quite effective in order to detect these interactions; however, as said, a real quantification would be hard to carry out and probably not really significant. This non-linearity is known to be more present in wind waves, which represent high frequency and low energies wave systems and are less interesting from a wave energy point of view. A low-pass filtering of the wave elevation time series, in particular, may help to reduce their effect so that they should not be taken into account by the forecasting model.

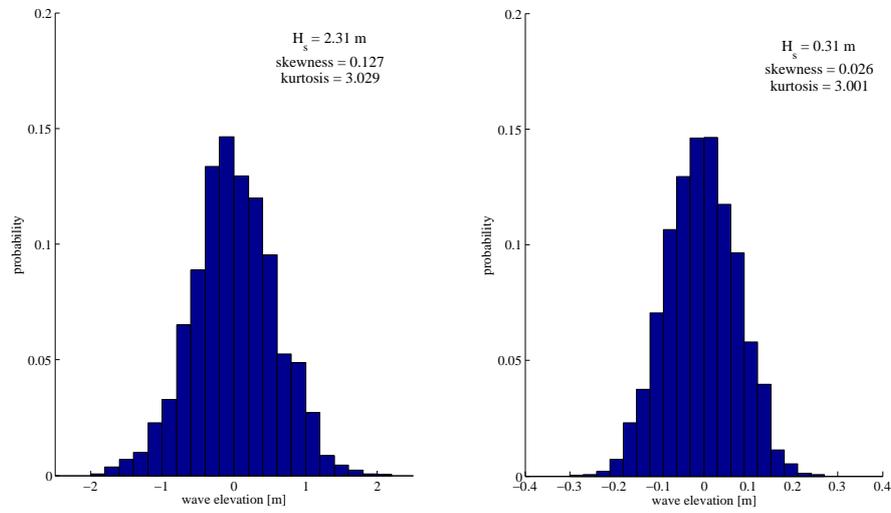
Figures 8 and 9 represent the bispectra calculated for some significant data sets from the Galway Bay and the Pico island, respectively. The off-diagonal components appear if an interaction between the two corresponding frequencies exists, and it is evident how they usually are strong for high frequency wind waves interacting with swell (as in figure 8(c)) or for broad spectra resulting from the superposition of different wave systems (figures 8(a) and 9(b)). The bispectrum is, on the other hand, much more concentrated around the diagonal for narrow banded swell systems, as can be noted from figures 8(b) and 9(a).

## 2.4 Predictability measure

As the focus of this study is on the multi-step-ahead prediction of the wave elevation (or of any connected quantity), one of the striking questions is this: Is there any chance to predict future values of a given signal? Usually, we design a predictor for a special signal or problem and then measure the resulting prediction quality. If there is no *a priori* knowledge on the optimal predictor, the achieved prediction gain will depend strongly on the particular prediction

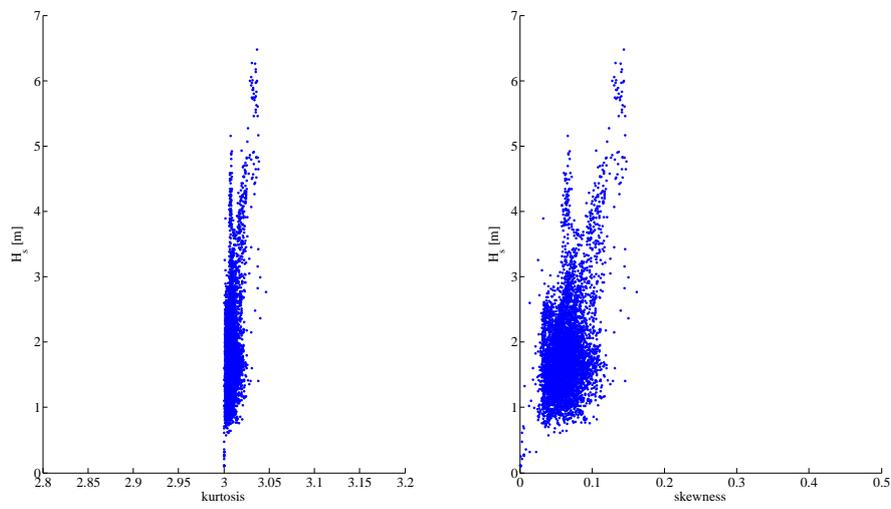


(a) Distribution of skewness and kurtosis

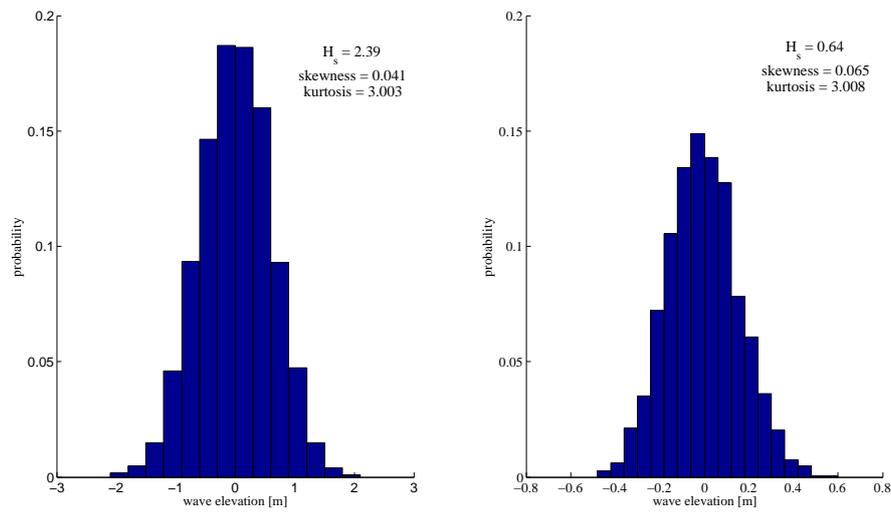


(b) Probability distribution of two sample wave elevation data sets

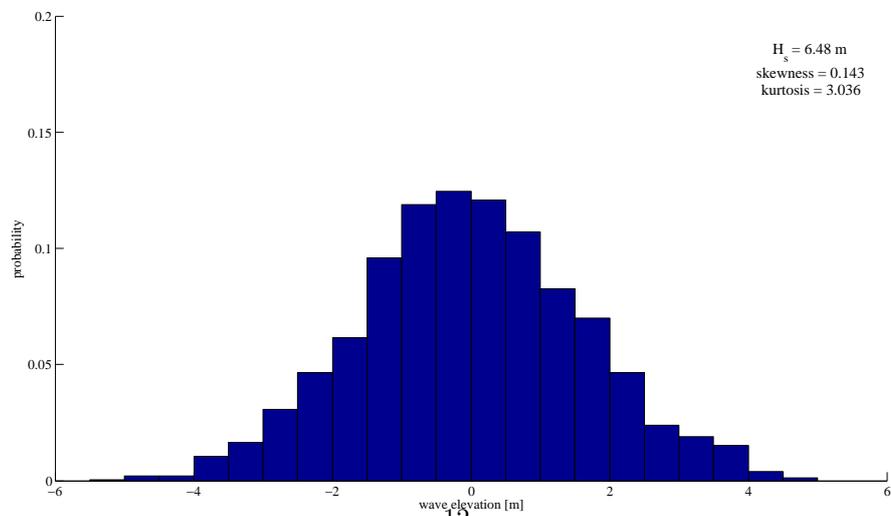
Figure 6: Galway bay Gaussianity analysis



(a) Distribution of skewness and kurtosis

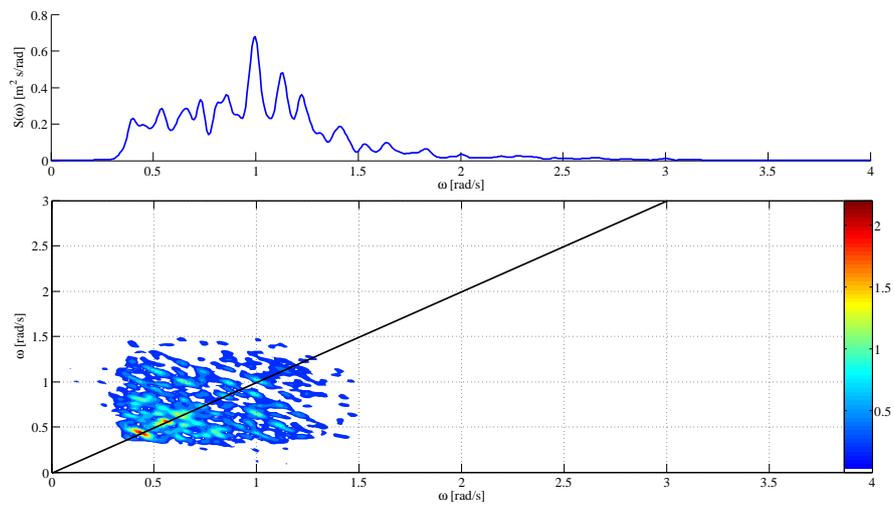


(b) Probability distribution of two sample wave elevation data sets

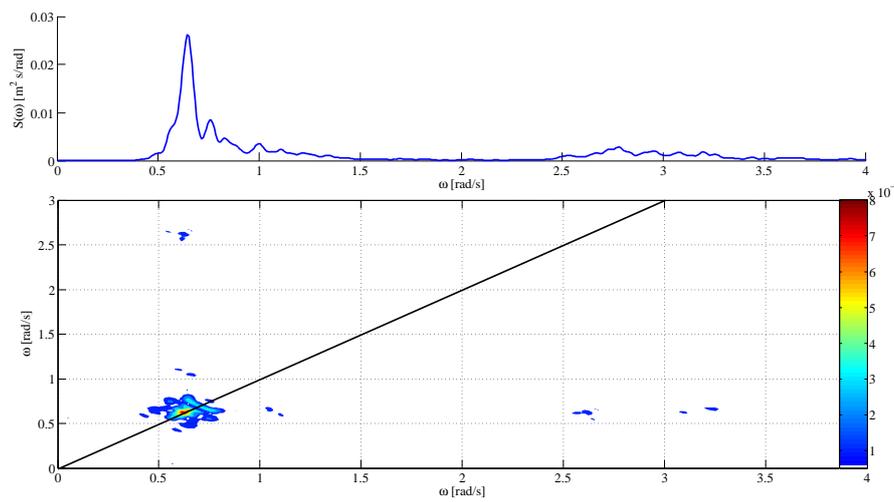


(c) Probability distribution of a very high energy wave elevation data set

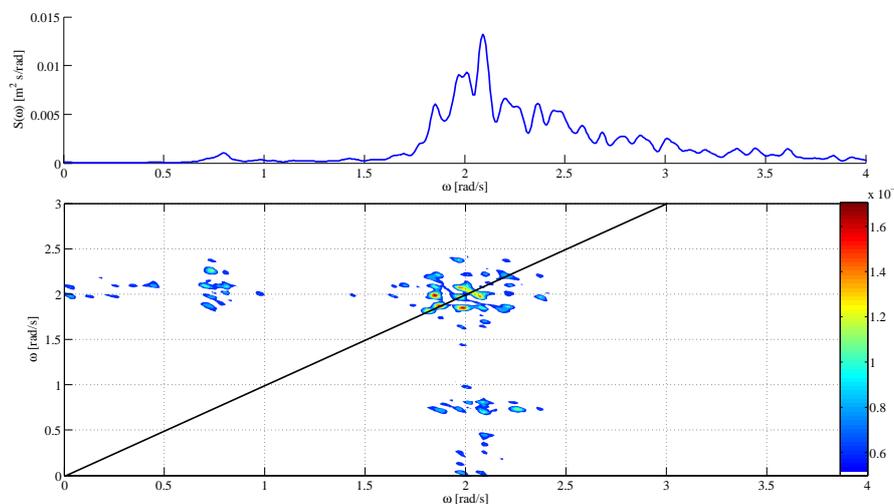
Figure 7: Pico Island Gaussianity analysis



(a)



(b)



(c)

Figure 8: Bispectrum of some data sets in Galway Bay

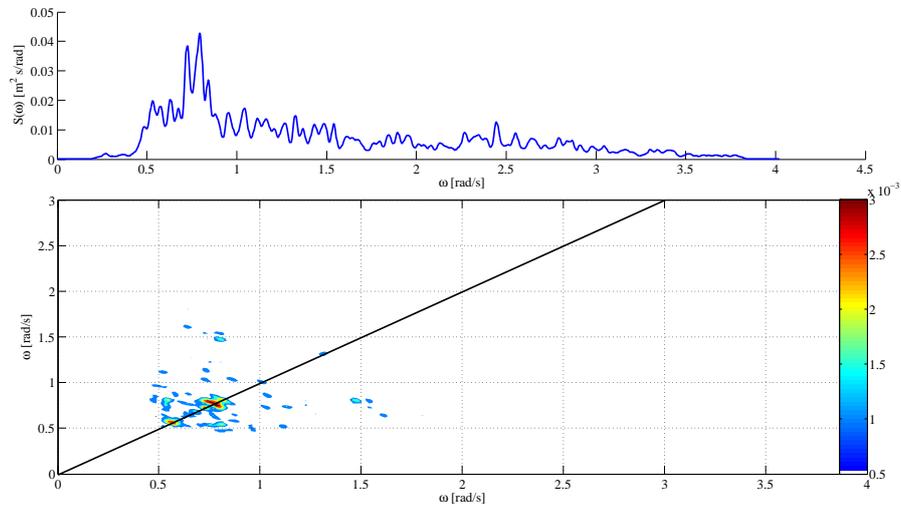
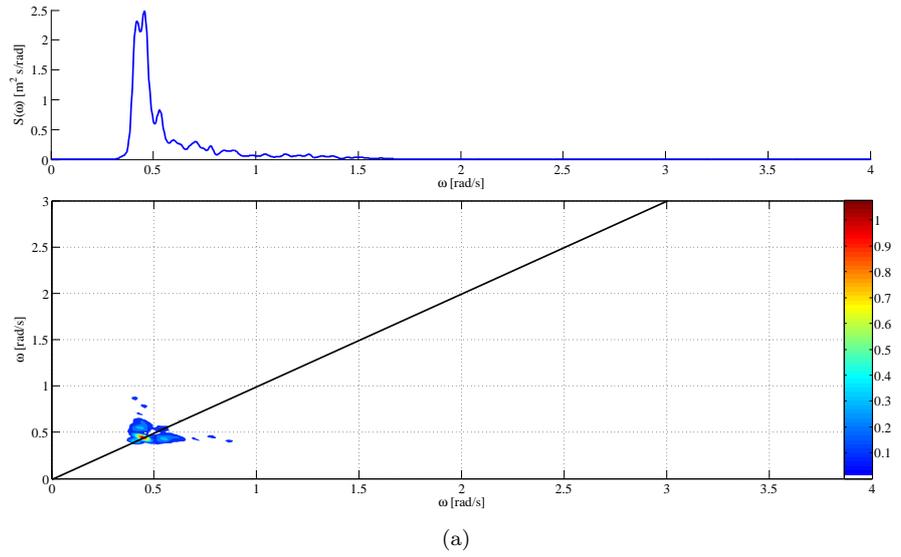


Figure 9: Bispectrum of some data sets in Pico Island

model used. Here, it is argued that, for prediction feasibility analysis, it is not necessary to design any predictors; we just have to know how much information about future signal values can be obtained from the past [13].

A simpler measure of predictability than the very general approach proposed in the literature (based on the mutual information notion [13],[14],[15],[16]) will be adopted here, which supposes that a linear relationship exists that relates the future values of the wave elevation to the past. This is, of course, a limiting assumption but it is still very valuable to provide at least a qualitative study over the predictability of the wave elevation.

In particular, a predictability index  $R^2(l)$  is estimated, defined as the ratio of the variance of the optimal  $l$ -step-ahead prediction,  $\hat{\eta}(k+l/k)$ , to the variance to the real wave elevation,  $\eta(k)$ :

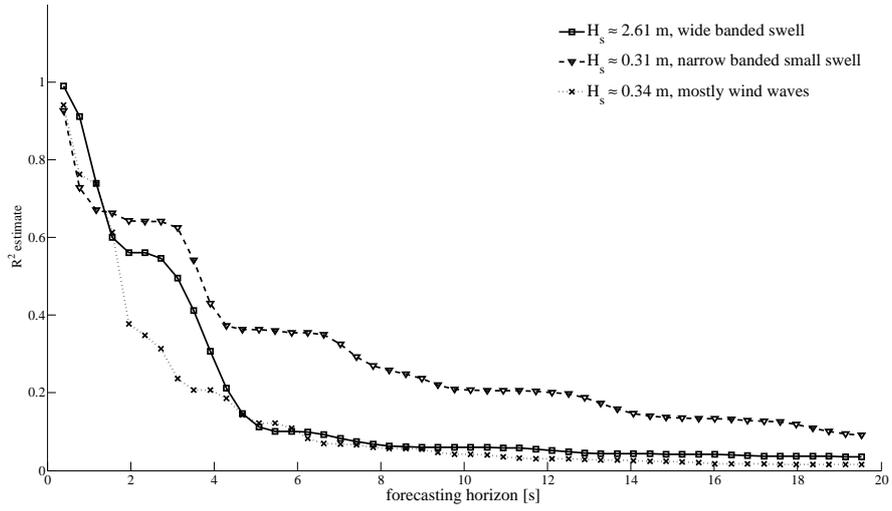
$$R^2(l) \triangleq \frac{E\{\hat{\eta}(k+l/k)^2\}}{E\{\eta(k)^2\}} = 1 - \frac{\hat{\sigma}_l^2}{E\{\eta(k)^2\}} \quad (7)$$

where it is supposed that the wave elevation  $\eta(k)$  has a zero mean and the optimal  $l$ -step-ahead prediction error variance is defined as  $\hat{\sigma}_l^2 \triangleq E\{\hat{\epsilon}(k+l/k)^2\}$ .

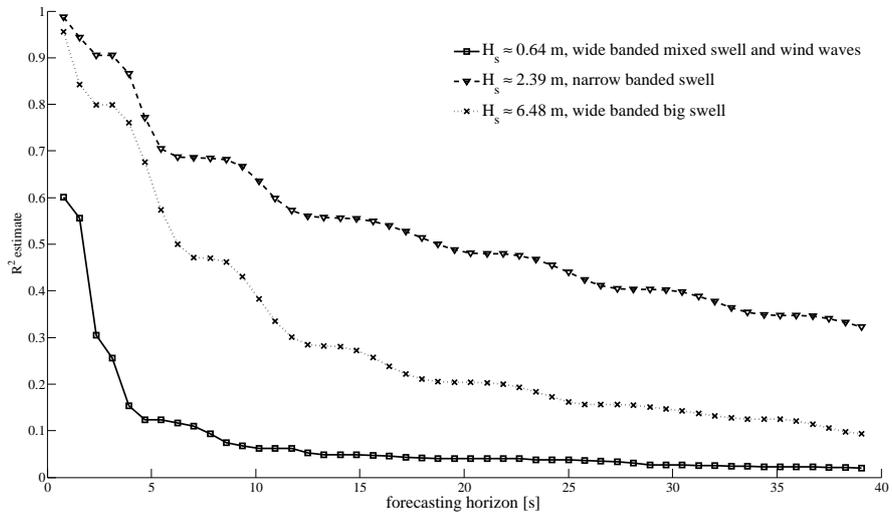
A very efficient algorithm for the estimation of  $R(l)^2$ , under the assumption of a linear univariate time series, was proposed in [17] and it is adopted here for the analysis of the available wave data. Figure 10 shows the estimated predictability index  $R^2(l)$ , for a forecasting horizon of 50 samples of different wave systems at the two locations of Galway Bay and Pico Island. As expected from any real-world time series, it is a non-increasing function of the prediction horizon. All the wave systems considered for the Galway Bay location, figure 10(a), show a relatively poor predictability, which dies out very quickly after 2 – 4 seconds (5 – 10 samples), with a slightly better behavior of the narrow banded (although low energy) wave system and of the high energy one. A much better predictability (index  $R^2$  is relatively high for more than 5 seconds) results for high energy and narrow banded wave systems at Pico Island, figure 10(b), mostly due to the smaller influence of the non-linearities analysed in section 2.3, consisting of either asymmetry in the wave distribution or non-linear interactions between different frequency components.

In a wave energy context, however, one might be interested in forecasting only the high energy components, so that a low-pass filter can be applied to the time series and a focus would be put exclusively on the low frequency components.

In figure 11, the estimated predictability index  $R(l)^2$  is shown for the pre-filtered wave systems at Galway Bay and when different cut-off frequencies,  $\omega_c$ , are applied. It is clear, by comparison with figure 10, how the overall predictability significantly improves with respect to the non-filtered waves. Moreover, the smaller the cut-off frequency, i.e. the lower the frequencies we limit the analysis to, the better the predictability of the time series, when a swell at the low frequencies is present, and more accurate predictions, further in the future, can be expected. Figure 11(c), in fact, referring to the wind waves system of figure 3(a), shows only a significant improvement for the highest cutoff frequency considered,  $\omega_c = 2 \text{ rad/s}$ . This shall, however, not be a big issue in a wave energy context, when these wave systems are not of interest for their very low energy content and their high frequency content (as compared to lower frequency dynamics of wave energy converters).



(a) Galway Bay data



(b) Pico Island data

Figure 10: Predictability indices  $R(l)^2$  estimated for datasets with different energy

Figure 12 depicts the situation for the Pico Island wave elevation data sets. In this case, the narrow banded swell and the low energy mixed wave systems of figures 12(a) and 12(b) derive the best improvement from considering only the low frequency swell. The same improvement is not achieved for the wide banded swell system of figure 12(c), whose spectrum can be seen in figure 3(b), and is also more strongly affected by the non-linearities analysed in section 2.3, in particular the non-Gaussianity.

Note that this analysis does not depend on any actual forecasting technique that might be implemented, so that the prediction accuracy will also depend on the chosen method, but some sort of upper bound for the attainable accuracy, irrespective of the utilised algorithm, is set here (although limited to the range of the possible linear forecasting models).

## 2.5 Choice of cut-off frequency

From the analysis carried out in the previous sections, particularly in 2.1 and 2.4, it emerged how low frequency components are the most interesting from a wave energy point of view and, at the same time, with respect to high frequency wave components, have a more regular behavior so that they are predictable more accurately and further into the future. It was stated, therefore, that one might focus the forecasting algorithm exclusively on the low frequency components, completely neglecting the rest of the signal.

If it is considered, however, that the prediction shall be utilised by a controller in order to improve the WEC ability to extract energy from the waves, we may expect that the energy contained in the frequency components neglected by the forecasting procedure represents a loss of extracted energy. That is, the controller is not able to improve the system response to those frequency component not considered in the prediction algorithm. Note that very high frequencies may be lost anyway due to the lowpass filtering dynamics of the WEC device itself but, in general, this depends on its operating principle and its design parameters.

The choice of the cut-off frequency of the prediction system can therefore be seen as a compromise between the accuracy improvement in the forecasts (which should improve the energy extraction of the WEC) and the loss of the energy carried by higher frequency components of the incident wave.

If the exact relationship between extracted energy and prediction accuracy was known, then a cost functional quantifying the compromise may be calculated and an optimal cut-off frequency may be found. Of course, such a function depends on so many variables (kind of device, control architecture, sea state, etc...) that it would be very hard and, at the same time, not really worth getting a real and complete model of it.

A rough quantification of this cost functional may, however, be carried out at this stage by making no assumptions neither on the forecasting algorithm nor on the device. It can be still a very valuable approach, in the author's opinion, as it would not require any significant effort or accurate knowledge of the problem. It can, moreover, be easily extended and become more and more accurate as new pieces of information are known about the overall problem and are then included in the cost functional. For the moment, it will be supposed that the energy extracted by a general non-specified device equals the energy contained in the forecasted wave weighted by the accuracy of the prediction. The accuracy

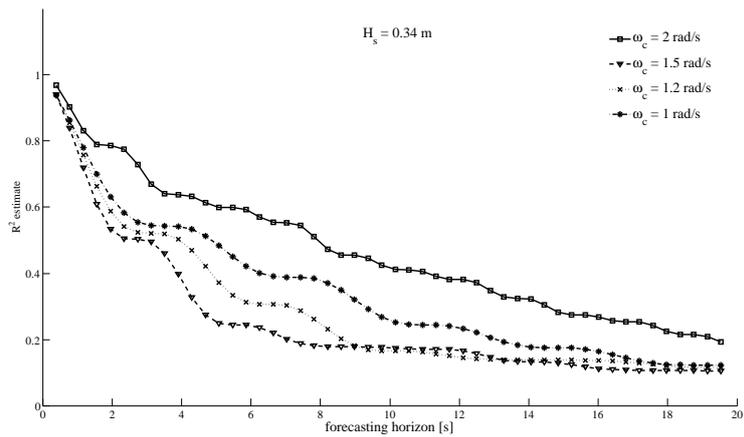
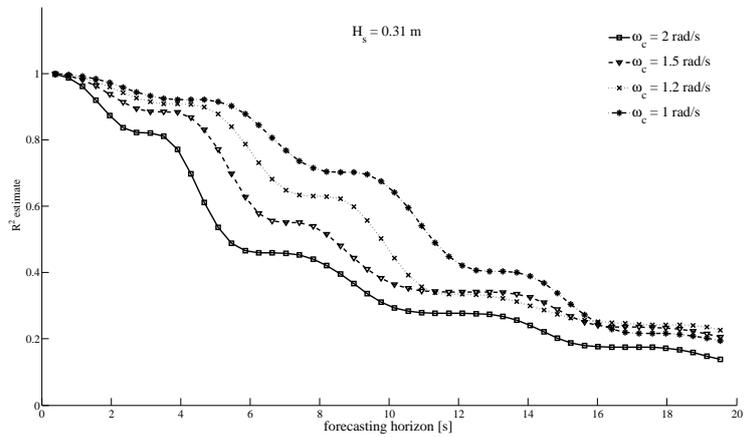
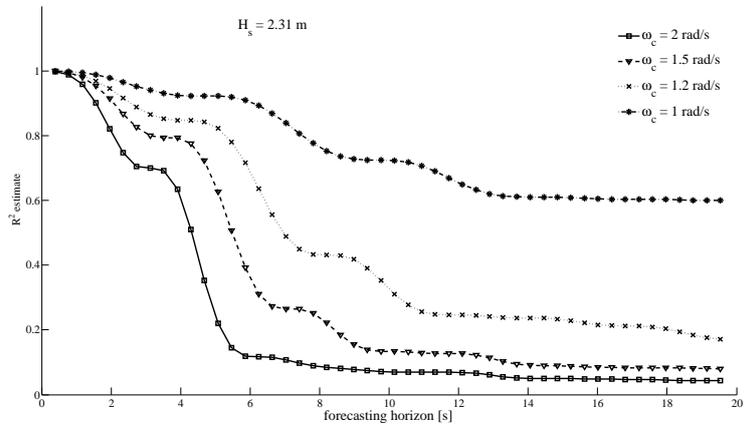
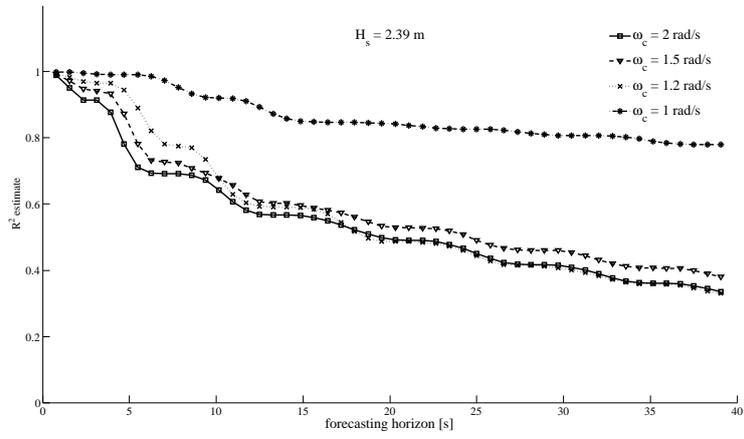
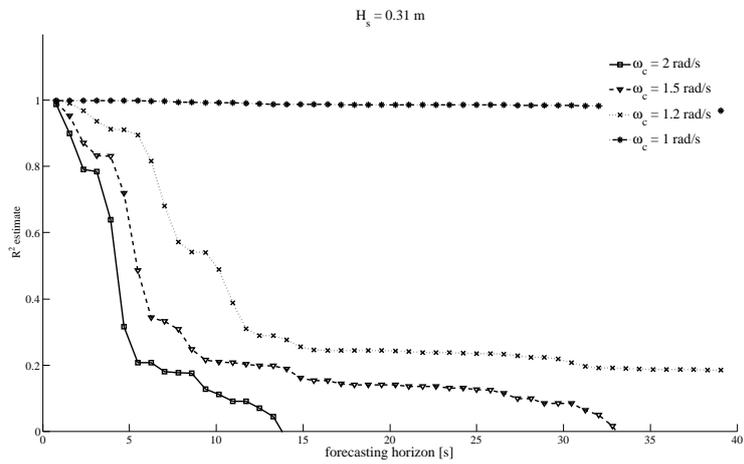


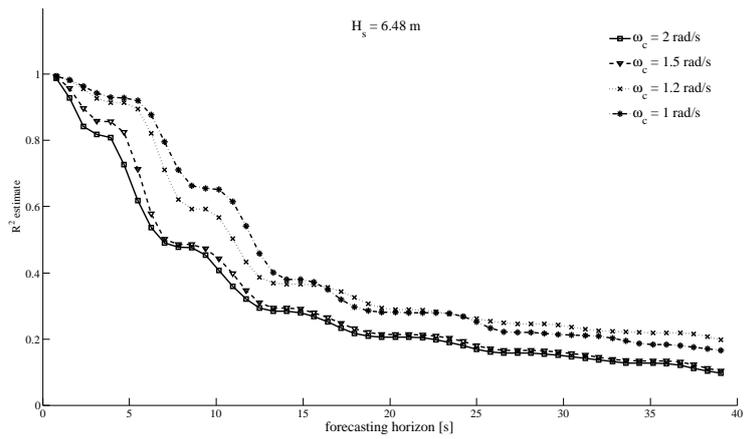
Figure 11: Predictability of some wave elevation data sets from Galway Bay when low-pass filtering with different cut-off frequencies  $\omega_c$  is applied.



(a)



(b)



(c)

Figure 12: Predictability of some wave elevation data sets from Pico Island when low-pass filtering with different cut-off frequencies  $\omega_c$  is applied.

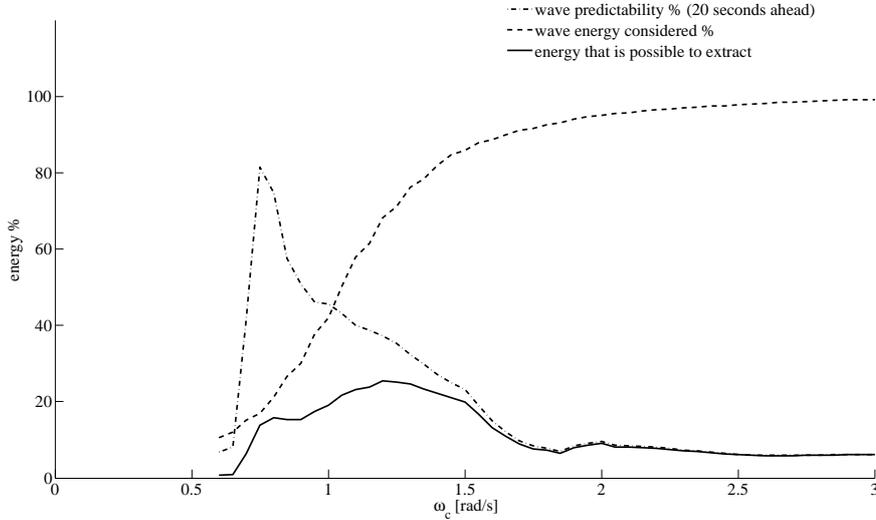


Figure 13: Energy extracted by the Wave energy converters thanks to the wave prediction and dependant on the cutoff frequency

of the prediction, here, is substituted with the achievable accuracy as given by a predictability index such as the one of equation (7).

The theoretical absorbed energy,  $E_a$ , is then calculated as:

$$E_a = P_r \times (E_{tot} - E_{neg}) \quad (8)$$

where  $E_{neg}$  is the energy neglected by the prediction algorithm,  $E_{tot}$  is total energy in the waves and  $0 \leq P_r \leq 1$  is a measure of the predictability. In figure 13 it is clear how the cut-off frequency is a compromise between prediction accuracy and energy cutoff, and if the functional (8) would be exact or reasonably accurate, it can be chosen at the maximum of the  $E_a$  curve.

The quantitative results here presented are not really significant because of the many simplifying (and also unrealistic) assumptions. It is, however, an interesting approach to the choice of the appropriate cut-off frequency when other parts of the problem will be better understood (utility of the prediction for the WEC performance, prediction error influence on the control, etc...).

## 2.6 Choice of the sampling frequency

In general, if the spectrum of a signal has a limited support  $[0 \ \omega_m]$ , then all the information is maintained if the signal is discretised with any sampling radian frequency  $\omega_s \geq \omega_m/\pi$ . Lower sampling frequencies give rise to the *aliasing* phenomenon, thus causing the sampled time series not to be uniquely representative of the original signal.

If a wave elevation time series is low-pass filtered before the prediction, this means that it can be sampled without any loss of information with  $\omega_s \geq \omega_c/2$ , where  $\omega_c$  is the cut-off frequency of the filter (assuming an ideal filter with instantaneous transition). Intuitively, a certain time span of the wave elevation signal is represented by fewer samples if the sampling frequency is lower. It

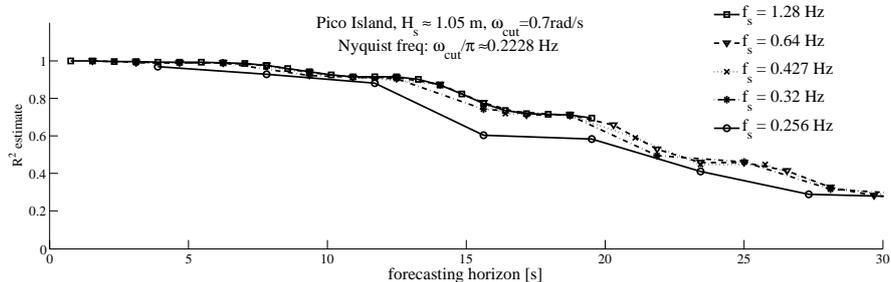


Figure 14: Estimated predictability for a wave elevation time series from Pico Island, when different sampling frequencies are adopted.

may be, therefore, that the choice of the sampling frequency could affect the performance of a prediction algorithm, in particular its forecasting horizon in terms of seconds.

In theory, the information that the past of a signal has about its future, will not be affected by the sampling frequency, if aliasing is avoided. Therefore, a proper forecasting model that manages to extract all the information to produce the prediction, should not perform differently by changing the sampling frequency. Figure 14 shows, in fact, how the predictability of a certain wave elevation data set is not affected by a change in the sampling frequency, when this is greater than the Nyquist frequency.

In practice, however, when performing the prediction, some differences might arise, so it is interesting to assess the effect of the sampling frequency on the prediction accuracy of the forecasting models that will be presented in section 3. This analysis will be carried out in section 4.5.

### 3 Methodology

Here a range of possible forecasting models is presented. Firstly, in section 3.1, some considerations are given about how the wave elevation can be actually measured at the same point of the sea where a device is located. Then, the models are presented in section 3.2, and a methodology to derive proper confidence intervals along with the predictions is proposed in section 3.3.

#### 3.1 Getting the observations of the signal to predict

The prediction of the incident wave (or any related physical quantity) on a WEC, based only on its past history, presents the main issue of getting the actual measurement of the signal being predicted. In the case of the incident wave elevation, there is no direct access to it at the actual point where the device is located (the device being an oscillating body or an oscillating water column).

There is, however, a concrete possibility of estimating the variable of interest from the measurements of other more accessible variables. The accuracy of this estimation will, of course, depend on the accuracy of the mathematical/numerical model relating the measures to the signal to be deduced. Some

considerations relevant to this problem will be provided here for a general oscillating body in single mode of motion.

A basic approximation of the equations of motion of a WEC is given by the following expression:

$$[m + m(\infty)] \ddot{x} + \int_0^t k(t - \tau) \dot{x}(\tau) d\tau + Sx = f_e(t) + f_{ext}(x, \dot{x}, t) \quad (9)$$

where  $x(t)$  is the displacement of the body along the considered degree of freedom,  $m$  and  $m(\infty)$  are the mass and added mass of the body,  $k(t)$  is the impulse response function relating the oscillation velocity of the body to its radiation force,  $S$  is the buoyancy coefficient,  $f_e(t)$  is the wave excitation force and  $f_{ext}(x, \dot{x}, t)$  represents any external, for example provided by the power take-off system.

1. If the measurements of the body motion are available, together with an accurate model of the system, the excitation force  $f_e(t)$  can directly be estimated from equation (9). The records of this estimate,  $\hat{f}_e(t)$ , can then be utilised to provide predictions of its future behavior. Alternatively, the excitation force can be used to derive the actual incident wave elevation and the forecasting problem could be focused on the latter. Note, however, that the excitation force and the wave elevation,  $\eta(t)$ , are related by a non-causal relationship:

$$f_e(t) = \int_{-\infty}^{+\infty} h(t) \eta(t - \tau) d\tau \quad (10)$$

with  $h(t) \neq 0$  for  $t < 0$ , so predictions of the wave elevation are also required to estimate the current excitation force  $\hat{f}_e(t)$ . Consider also that it is the wave excitation force that is actually required in order to compute an optimal reference for the optimal control of the system.

2. If the total wave force  $f_w(t)$  is measured instead, by means for example of pressure transducers on the body surface, then the excitation force could be determined through the following expression:

$$f_e(t) = f_w(t) + \int_0^t k(t - \tau) \dot{x}(\tau) d\tau + Sx - f_{ext}(x, \dot{x}, t) \quad (11)$$

which still needs the motion measurements in order to calculate the radiation and the buoyancy forces. Note that equation (11) directly derives from (9) if we consider that the total force  $f_w(t) \equiv [m + m(\infty)] \ddot{x}$ .

For the oscillating bodies, therefore, it seems reasonable to focus on the measurement and prediction of the wave excitation force, rather than the incident wave elevation. In this view, the approach followed in this report to forecast the wave elevation is, however, still valuable, particularly in view of the lowpass filtering applied to the signal prior to the prediction. The wave excitation force, in fact, is nothing else than the wave elevation filtered by the dynamics of the body (which obviously has a lowpass characteristics).

## 3.2 Forecasting models

### 3.2.1 Cyclical models

From linear wave theory [2], a real ocean sea state may be modelled as a linear superposition of waves with different frequencies and propagating in different directions:

$$\eta(x, y, t) = \int_0^{+\infty} d\omega \int_{-\pi}^{+\pi} A(\omega, \beta) \cos(\omega t - kx \cos \beta - ky \sin \beta + \varphi_i(\omega)) d\beta \quad (12)$$

where  $k$  is the wave number and  $\beta$  represents the direction of propagation in the  $x$ - $y$  plane. If a specific location  $(x_0, y_0)$  is considered, the following simplified expression can then be obtained:

$$\eta(x_0, y_0, t) = \int_0^{+\infty} d\omega \int_{-\pi}^{+\pi} A(\omega, \beta) \cos(\omega t + \phi(\omega, \beta)) \quad (13)$$

where the directionality information is obviously lost and the constant terms  $kx_0 \cos \beta$  and  $ky_0 \sin \beta$  are included in the phase  $\phi(\omega, \beta)$ .

From this knowledge about the real process it is quite straightforward to choose, as a forecasting model for the wave elevation, a simple cyclical model, where the frequency domain is of course discretised [18],[19]:

$$\eta(t) = \sum_{i=1}^m a_i \cos(\omega_i t) + b_i \sin(\omega_i t) + \zeta(t) \quad (14)$$

An error  $\zeta(t)$  has been introduced and the phase and amplitude information for each harmonic component is now contained in the parameters  $a_i$  and  $b_i$ .

The model (14) is completely characterised by the parameters  $a_i, b_i$  and by the frequencies  $\omega_i$ . It could then be fitted to the data through some non-linear estimation procedure (the model is non-linear in the frequencies) and utilised to predict the future behavior of the wave elevation time series. It needs, however, to be adapted to the time variations of the wave spectrum (amplitudes and phases of the frequency components are non-constant), so that a first approach [19] has been considered, where the frequencies are chosen in the model design phase and then kept constant during its utilisation and estimation. In this way the model becomes perfectly linear in the parameters  $a_i, b_i$  and can be easily estimated and on-line adapted to the spectral variations of the sea.

The problem of choosing the frequencies can be divided in two sub-problems:

1. **Choice of the range:** This is a quite easy matter, as statistical information about the location can be utilised to properly define an upper and lower bound for the range. At this point, one may decide to include the range of higher frequencies where the low energy wind waves are, or to simply consider a narrower range including only the swell.
2. **Distribution of the frequencies in the range:** A robust choice would be a constant spacing between the frequencies over all the range, but a more

efficient non-homogeneous distribution was also proposed in [19]. The latter however suffers from the problem of specificity, so that if the wave spectrum changes the frequencies might not be appropriate any more. If the frequencies are kept constant, then it would not be a proper choice.

Once the frequencies are determined, a model for the amplitudes has to be chosen. In [18], [19] it was pointed out how they have to be adaptive to the wave, as constant amplitudes gave very poor results. Two adaptive models are proposed here, in particular:

**Structural model:** based on Harvey's structural model [20], the model (14) is expressed in the following discrete time form:

$$\eta(k) = \sum_{i=1}^m \psi_i(k) + \zeta(k) \quad (15)$$

$$\begin{bmatrix} \psi_i(k+1) \\ \psi_i^*(k+1) \end{bmatrix} = \begin{bmatrix} \cos(\omega_i T_s) & \sin(\omega_i T_s) \\ -\sin(\omega_i T_s) & \cos(\omega_i T_s) \end{bmatrix} \begin{bmatrix} \psi_i(k) \\ \psi_i^*(k) \end{bmatrix} + \begin{bmatrix} w_i(k) \\ w_i^*(k) \end{bmatrix}, \quad i = 1, \dots, m \quad (16)$$

where it can be verified that  $\psi_i(0) = a_i$  and  $\psi_i^*(0) = b_i$ . From equation (16), then, the following state space form, which is more familiar to work with, is easily derived:

$$\begin{aligned} x(k+1) &= Ax(k) + w(k) \\ \eta(k) &= Cx(k) + \zeta(k) \end{aligned} \quad (17)$$

where

$$x(k) \triangleq [\psi_1(k) \quad \psi_1^*(k) \quad \dots \quad \psi_m(k) \quad \psi_m^*(k)]^T \in \mathbb{R}^{2m \times 1} \quad (18)$$

$$w(k) \triangleq [w_1(k) \quad w_1^*(k) \quad \dots \quad w_m(k) \quad w_m^*(k)]^T \in \mathbb{R}^{2m \times 1} \quad (19)$$

$$A \triangleq \text{diag} \left\{ \begin{bmatrix} \cos(\omega_i T_s) & \sin(\omega_i T_s) \\ -\sin(\omega_i T_s) & \cos(\omega_i T_s) \end{bmatrix} \right\} \in \mathbb{R}^{2m \times 2m} \quad (20)$$

$$C \triangleq [1 \quad 0 \quad 1 \quad 0 \quad \dots \quad 1 \quad 0] \in \mathbb{R}^{1 \times 2m} \quad (21)$$

**Dynamic Harmonic Regression (DHR):** Introduced by Young [21], it expresses a cyclical model of the type of eq. (14), where the  $a_i$  and  $b_i$  parameters evolve according to a Generalised Random Walk:

$$\begin{aligned} \begin{bmatrix} x_i(k+1) \\ x_i^*(k+1) \end{bmatrix} &= \begin{bmatrix} \alpha & \beta \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} x_i(k) \\ x_i^*(k) \end{bmatrix} + \begin{bmatrix} \delta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_i(k) \\ \epsilon_i^*(k) \end{bmatrix} \\ x_i &= a_i \quad \text{for } i = 1, \dots, m \\ x_{i-m} &= b_i \quad \text{for } i = m+1, \dots, 2m \end{aligned} \quad (22)$$

where  $x_i^*$  models a slope for the evolution of each parameter  $x_i$ . The disturbance terms  $\epsilon_i$  and  $\epsilon_i^*$  are still assumed to be Gaussian noises and introduce the variability in the model. A particular form of (22) was implemented in this study where the dynamic matrices are chosen in order to represent Harvey's local linear trend [20]:

$$\begin{bmatrix} x_i(k+1) \\ x_i^*(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_i(k) \\ x_i^*(k) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_i(k) \\ \epsilon_i^*(k) \end{bmatrix} \quad (23)$$

for  $i = 1, 2, \dots, 2m$ . A state space form, then, can easily be derived, resulting in the following model:

$$\begin{aligned} x(k+1) &= Ax(k) + \epsilon(k) \\ \eta(k) &= C(k)x(k) + \zeta(k) \end{aligned} \quad (24)$$

where

$$x(k) \triangleq [x_1(k) \quad x_1^*(k) \quad \dots \quad x_{2m}(k) \quad x_{2m}^*(k)]^T \in \mathbb{R}^{4m \times 1} \quad (25)$$

$$\epsilon(k) \triangleq [\epsilon_1(k) \quad \epsilon_1^*(k) \quad \dots \quad \epsilon_{2m}(k) \quad \epsilon_{2m}^*(k)]^T \in \mathbb{R}^{4m \times 1} \quad (26)$$

$$A \triangleq \text{diag} \left\{ \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right\} \in \mathbb{R}^{4m \times 4m} \quad (27)$$

$$\begin{aligned} C(k) &\triangleq [\cos(\omega_1 T_s) \quad 0 \quad \dots \quad \cos(\omega_m T_s) \quad 0 \\ \sin(\omega_1 T_s) \quad 0 \quad \dots \quad \sin(\omega_m T_s) \quad 0] \in \mathbb{R}^{1 \times 4m} \end{aligned} \quad (28)$$

Both the models have the advantage of a state space representation, which is particularly suited to the application of the Kalman filter for a recursive on-line adaption. The initialisation is provided through means of regular least squares on a number of past observations and then the Kalman filter is applied on-line, once a proper covariance matrix for the state and output disturbances is provided. When the estimate of the model's parameters,  $\hat{x}(k/k)$ , is available at any instant  $k$ , the  $l$ -steps-ahead prediction  $\hat{\eta}(k+l/k)$ , based on the information up to  $k$ , is obtained through the free evolution of the model:

$$\hat{\eta}(k+l/k) = C(k+l)A^l \hat{x}(k/k) \quad (29)$$

There are, however, some strong limitations to this approach with cyclical models[18],[19], that will be highlighted also in the results, section 4:

- ↑ The use of constant frequencies requires, for the sake of robustness, a dense and complete set, which adds considerable complexity to the model, and
- ↑ It is not clear how to choose the covariance matrices for the Kalman filter implementation

In the next section 3.2.2, it will be shown how AR models implicitly overcome these difficulties in a very effective, and simple, way.

### 3.2.2 Auto Regressive (AR) models

As a pure time series problem is under study, there is the advantage of the existence of a well established theory, from the time series field, which it is possible to utilise. In a comparison with the cyclical models, where the *a priori* knowledge that we have about the real system is explicitly taken into account, it is particularly interesting to analyse the properties of classical AR models.

The wave elevation  $\eta(k)$  is supposed to be linearly dependent on a number  $n$  of its past values:

$$\eta(k) = \sum_{i=1}^n a_i \eta(k-i) + \zeta(k) \quad (30)$$

where a disturbance term  $\zeta(k)$  has been included. If the parameters  $a_i$  are estimated and the noise is supposed to be Gaussian and white, the best prediction of the future wave elevation  $\hat{\eta}(k+l/k)$  at instant  $k$  is then given by:

$$\hat{\eta}(k+l/k) = \sum_{i=1}^n \hat{a}_i(k) \hat{\eta}(k+l-i/k) \quad (31)$$

where, obviously,  $\hat{\eta}(k+l-i/k) \equiv \eta(k)$  if  $k+l-i \leq k$  (i.e. the information is already acquired and there is no need for prediction).

The properties of such a very simple forecasting model become clearer if an explicit solution of the difference equation (31) is provided [22]:

$$\hat{\eta}(k+l/k) = \sum_{i=1}^n b_i(k) f_i(l) \quad (32)$$

Here, the coefficients  $b_i(k)$  depend only on the forecasting origin (so they stay constant at each instant for the complete prediction time horizon) and are a function of the initial conditions (the past  $n$  observations), whereas  $f_i(l)$  are functions of the lead time  $l$  and, in general, they include damped exponential and damped sinusoidal terms completely determined by the roots  $p_i$  of the transfer function  $\varphi(z)$  describing equation (30) in the  $Z$ -domain:

$$\eta(z) = \frac{\zeta(z)}{\varphi(z)} \triangleq \frac{\zeta(z)}{\prod_{i=1}^n (z - p_i)} \quad (33)$$

The general shape of the prediction function is therefore completely determined by the poles,  $p_i$ , while the particular realisation of this general structure is determined, at each sampling instant, by the past values of the time series. It is particularly interesting to analyse the shape of the forecasting function (32) in the case of  $m/2$  (when  $m$  is even) couples of complex-conjugate poles,  $p_i$  and  $p_i^*$ :

$$\hat{\eta}(k+l/k) = \sum_{i=1}^{m/2} c_i(k) |p_i|^l \sin(\angle p_i k + \varphi_i(k)) \quad (34)$$

Thus, an AR model with only complex-conjugate poles is *implicitly* a cyclical model, where the frequencies are related to the phase,  $\angle p_i$ , of each pole and the amplitudes and phases of the harmonic components are related to the last  $n$  observations of each time instant  $k$ , so that they adapt to the observations.

Note, then, that an adaptivity mechanism is already present even if the AR model parameters are only estimated once on a batch data set. A classical estimation approach is to minimise the squared errors sum, which has a linear solution provided by regular least squares. Only the frequencies, in this case, are fixed, while amplitudes and phases are automatically updated on the basis of the recent past information. A further degree of adaptivity can be introduced with an on-line estimation of the AR model parameters,  $a_i$ , which would introduce an on-line adaptivity of the frequencies as well. If we express equation (30) in vectorial form:

$$\eta(k) = \psi(k)^T \vartheta(k) + \zeta(k) \quad , \quad (35)$$

where

$$\psi(k) \triangleq [\eta(k-1) \ \eta(k-2) \ \dots \ \eta(k-n)]^T \in \mathbb{R}^{n \times 1} \quad (36)$$

$$\vartheta(k) \triangleq [a_1(k) \ a_2(k) \ \dots \ a_n(k)]^T \in \mathbb{R}^{n \times 1} \quad , \quad (37)$$

then a general recursive estimation of the time-varying parameters vector,  $\vartheta(k)$ , is given as:

$$\hat{\vartheta}(k+1) = \hat{\vartheta}(k) + K(k) \left[ \eta(k+1) - \psi(k)^T \hat{\vartheta}(k) \right] \quad (38)$$

as a combination of free evolution and innovation. How the gain vector  $K(k) \in \mathbb{R}^{n \times 1}$  is chosen depends on which specific estimation algorithm is utilised. The most common approaches are the *recursive least squares with forgetting factor* and the *Kalman Filter*, which are outlined in the following subsections:

**Recursive Least Squares (RLS):** With this approach the following functional is minimised:

$$J(\vartheta(k)) = \sum_{j=1}^k \lambda^{k-j} [\eta(j) - \psi(j)^T \vartheta(k)]^2, \quad (39)$$

where more weight, via the forgetting factor  $\lambda < 1$ , is given to recent observations according to an exponential law. The recursive algorithm for the optimal value of  $\vartheta(k)$  that minimises the functional  $J(\vartheta(k))$  respects the general form of equation (38), with the gain  $K(k)$  given as:

$$K(k) = P(k)\psi(k) \quad (40)$$

$$P(k) = \frac{P(k-1)}{\lambda + \psi^T(k)P(k-1)\psi(k)} \quad (41)$$

The forgetting factor  $\lambda$  is typically chosen in the range [0.97, 0.995].

The matrix  $P(k) \in \mathbb{R}^{n \times n}$  represents the covariance matrix of the estimate  $\hat{\vartheta}(k)$ :

$$P(k) \equiv E\{\hat{\vartheta}(k)\hat{\vartheta}^T(k)\} \quad (42)$$

It is interesting to introduce also the *information matrix*  $R(k) \in \mathbb{R}^{n \times n}$ , defined as:

$$\begin{aligned} R(k) \triangleq \psi(k)\psi(k)^T + \lambda\psi(k-1)\psi(k-1)^T + \dots + \lambda^{k-1}\psi(1)\psi(1)^T + R(0) = \dots \\ \dots = \lambda R(k-1) + \psi(k)\psi(k)^T \end{aligned} \quad (43)$$

, which can also be shown to correspond to [23]:

$$R(k) \equiv P(k)^{-1} \quad (44)$$

One main problem of recursive least squares with a forgetting factor is that, if the measurements do not add new information to the system, that is  $\psi(k)$  is approximately zero for a certain time. Therefore, the information matrix decreases until it can get close to the null matrix (or only some of its eigenvalues tend to zero). At the same time, some of the elements of the corresponding gain  $K(k)$  may significantly increase. When  $\psi(k)$ , then, increases in magnitude, the estimate  $\hat{\vartheta}(k)$  can experience a very large growth, known as the phenomenon of *blow-up*.

Some regularisation solutions have been proposed to cope with this problem [24]. One possibility is to monitor the matrix  $P(k)$  ( $R(k)$ ), and reset its values to *acceptable* ones when its eigenvalues assume too large (small) values [23]. Another approach may be a variable forgetting factor, based on the state of the process (steady or transient) [23], or regularisation of the information matrix  $R(k)$  [25].

**Kalman Filter:** If the evolution of the state vector  $\vartheta(k)$  is assumed to be a random walk:

$$\vartheta(k) = \vartheta(k-1) + \varepsilon(k) \quad (45)$$

where  $\varepsilon(k)$  is a Gaussian random process, then the Kalman Filter may be applied, resulting in the recursive form of (38), where the gain  $K(k)$  is now given by [26]:

$$K(k) = Q(k)\psi(k) \quad (46)$$

with

$$Q(k) = \frac{P(k-1)}{R_2 + \psi^T(k)P(k-1)\psi(k)} \quad (47)$$

$$P(k) = P(k-1) + R_1 - \frac{P(k-1)\psi(k)\psi(k)^T P(k-1)}{R_2 + \psi^T(k)P(k-1)\psi(k)} \quad (48)$$

where  $R_1(k) = E\{\varepsilon(k)\varepsilon(k)^T\}$  and  $R_2 = E\{\zeta(k)^2\}$ . Here,  $P(k)$  still represents the covariance matrix of the estimate  $\hat{\vartheta}(k)$ .

### 3.2.3 Sinusoidal extrapolation and the Extended Kalman Filter

The main problem with the cyclical model is the choice of the frequencies, that must be kept constants in order for the model to be linear in the parameters. Efficient linear algorithms for the recursive estimation of the cycles amplitudes (and phases) can therefore be exploited. This means that the accuracy of the model is strictly connected to its capacity to cover as much as possible of the typical range of frequencies where the wave systems at the considered location are mostly concentrated. This approach lacks of efficiency and requires a much more complex method than what would actually be required. In reality few regular waves in the typical range will be active at each time instant.

A more intelligent solution, then, would be to consider a few (or even one) cyclical component with an *adaptive frequency* which is updated on-line with the real observations.

We propose, therefore, to model the wave elevation as a single cyclical component as in the Harvey structural model, equation (16), but with a time-varying frequency  $\omega(k)$ :

$$\begin{cases} \begin{bmatrix} \psi(k+1) \\ \psi^*(k+1) \end{bmatrix} = \begin{bmatrix} \cos(\omega(k)T_s) & \sin(\omega(k)T_s) \\ -\sin(\omega(k)T_s) & \cos(\omega(k)T_s) \end{bmatrix} \begin{bmatrix} \psi(k) \\ \psi^*(k) \end{bmatrix} + \begin{bmatrix} \varepsilon(k) \\ \varepsilon^*(k) \end{bmatrix} \\ \eta(k) = \psi(k) + \zeta(k) \end{cases} \quad (49)$$

where  $\varepsilon(k)$ ,  $\varepsilon^*(k)$  and  $\zeta(k)$  are random disturbances and  $\eta(k)$  is the wave elevation.

Along with the components  $\psi(k)$  and  $\psi^*(k)$ , the frequency  $\omega(k)$  also needs to be estimated. A state vector  $x(k)$ , composed of the quantities that need to be estimated, is then defined as:

$$x(k) \triangleq [\psi(k) \quad \psi^*(k) \quad \omega(k)]^T \in \mathbb{R}^{3 \times 1} \quad (50)$$

The system (49) can then be redefined in terms of  $x(k)$  as:

$$\begin{cases} \begin{bmatrix} \psi(k+1) \\ \psi^*(k+1) \\ \omega(k+1) \end{bmatrix} = \begin{bmatrix} \cos(\omega(k)T_s) & \sin(\omega(k)T_s) & 0 \\ -\sin(\omega(k)T_s) & \cos(\omega(k)T_s) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \psi(k) \\ \psi^*(k) \\ \omega(k) \end{bmatrix} + \begin{bmatrix} \varepsilon(k) \\ \varepsilon^*(k) \\ \kappa(k) \end{bmatrix} \\ \eta(k) = \psi(k) + \zeta(k) \end{cases} \quad (51)$$

In (51) a model for the adaptivity of the frequency  $\omega(k)$  has been introduced, where a simple random walk is proposed, driven by the additional white noise  $\kappa(k)$ . The model, of course, is non-linear in the frequency and an explicit state space structure cannot be formulated, so the following form is adopted:

$$\begin{cases} x(k+1) = f(x(k), w(k)) \\ \eta(k) = [1 \ 0 \ 0]x(k) + \zeta(k) \end{cases} \quad (52)$$

where  $f(x(\cdot), w(\cdot)) \in \mathbb{R}^{3 \times 1}$  is a vectorial non linear function and  $w(k) \triangleq [\varepsilon(k) \ \varepsilon^*(k) \ \kappa(k)]^T \in \mathbb{R}^{3 \times 1}$  is the vectorial form of the state disturbance.

The optimal estimate (in the sense of minimising the variance) for  $x(k)$ , based on the observations  $Z_{k-1} = \{\eta(0), \dots, \eta(k-1)\}$ , is known to be given by the conditional mean value:

$$\hat{x}(k/k-1) = E\{x(k)/Z_{k-1}\} \quad (53)$$

An explicit expression which is usable in practise is difficult to obtain, though, because of the non-linearity of the system. It is well known, however, that a very efficient algorithm for linear (and Gaussian) systems can be adopted, which is the Kalman filter. However, an application of the latter to non-linear problems can also be found in literature, known as the Extended Kalman Filter (EKF).

The EKF assumes that the discrete time steps ( $T_s$  in our case) are sufficiently small to permit the prediction equations to be approximated by a linearised form, based on the truncation of the Taylor expansion of the model (52) at the first order [27]:

$$x(k+1) \approx f(\bar{x}(k), w) + \left[ \frac{d}{dx} f(x(k), w(k)) \right]_{x(k)=\bar{x}(k)} (x(k) - \bar{x}(k)) \quad (54)$$

where  $\bar{x}(k)$  represents an opportune working point for the state of the system, whose variation is supposed to be small within the time step  $T_s$  considered. Based on this approximation, the following state space form can be derived for the model (52):

$$\begin{cases} x(k+1) \approx A(k)x(k) + F(k) + w(k) \\ \eta(k) = Cx(k) + \zeta(k) \end{cases} \quad (55)$$

where

$$A(k) \triangleq \left[ \frac{d}{dx} f(x(k), w(k)) \right]_{x(k)=\bar{x}(k)} \in \mathbb{R}^{3 \times 3} \quad (56)$$

$$F(k) \triangleq f(\bar{x}(k), w) - A(k)\bar{x}(k) \in \mathbb{R}^{3 \times 1} \quad (57)$$

$$C \triangleq [1 \ 0 \ 0] \in \mathbb{R}^{1 \times 3} \quad (58)$$

The iterative equations for computing the gain  $K(k)$  of the Kalman filter [26] can directly be derived from this linear state space form (55), where the operating point at each step is set as the optimal one step-ahead estimate of the state,  $\bar{x}(k) = \hat{x}(k + 1/k)$ , which is updated based on the *non linear* equation:

$$\hat{x}(k + 1/k) = f(\hat{x}(k/k), [0 \ 0 \ 0]^T) \quad (59)$$

$$\hat{x}(k + 1/k + 1) = \hat{x}(k + 1/k) + K(k) [\eta(k + 1) - C\hat{x}(k + 1/k)] \quad (60)$$

The direct extension to an  $n$ -frequencies model may be given by the following:

$$\mathbf{x}(k) = \begin{bmatrix} \mathbf{x}_1(k + 1) \\ \dots \\ \mathbf{x}_n(k + 1) \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1(\mathbf{x}_1(k), \mathbf{w}_1(k)) \\ \dots \\ \mathbf{f}_n(\mathbf{x}_n(k), \mathbf{w}_n(k)) \end{bmatrix} \quad (61)$$

$$\eta(k) = [1 \ 0 \ 0 \ \dots \ 1 \ 0 \ 0] \mathbf{x}(k) + \zeta(k)$$

As will be shown in the results of section 4.3, this model has the problem that the Kalman filter would update all the frequencies in the same way, which makes it not really effective.

### 3.2.4 Neural networks

It was shown in section 2.3 how the non-linearities appearing in the big low frequency waves, due to the relatively small water depth, are not really relevant. It is however interesting, in the authors opinion, looking at a comparison of the other models with a most widespread tool for time series modelling and forecasting such as neural networks.

In spite of either the great modelling capability and the easiness of building up a suitable structure, neural networks have the great disadvantage of offering a model completely enclosed in a black box where any analysis and properties evaluation is prevented. So, while in the cyclical and AR models an analysis of the estimated parameters and frequencies and their variations in an adaptive structure can provide indications about the real process behavior and its main characteristics, this would not be possible with neural networks.

For the problem under study, a non-linear relationship of the following type is created through a multilayer perceptron [28]:

$$\eta(k) = \mathbf{NN}(\eta(k - 1), \eta(k - 2), \dots, \eta(k - n)) \quad (62)$$

so that the dependance between the current wave elevation and  $n$  past values is realised. The model is then trained through the back propagation algorithm on a set of batch data and utilised for multi-step-ahead prediction.

This is, of course, not the only possibility and many others could be considered. For example, a priori knowledge about the process (which would always be a more appropriate approach) may be included and a non-linear relationship of the following type may be considered instead:

$$\eta(k) = \mathbf{NN}(\cos(\omega_1 T_s k + \varphi_1), \dots, \cos(\omega_n T_s k + \varphi_n)) \quad (63)$$

but some of the limitations outlined in section 3.2.1, when cyclical models were considered, due to an appropriate choice of the frequencies, are still present.

In section 4.4 results will be shown and compared with the cyclical and AR models, for different neural network topologies, with two hidden layers and different numbers of inputs (regression order  $n$ ).

### 3.3 Confidence intervals

The predictions alone, as computed by any of the models presented through section 3.2.1 to 3.2.4, do not give a complete enough information about the future of the signal, because they are inevitably affected by an estimation error, so that it would be fundamental to have an indication about the entity of this error and about the confidence that we can put in the forecasts computed by the prediction algorithm. If the  $l$ -steps ahead prediction error is Gaussian:

$$\hat{e}(k+l/k) = \eta(k+l) - \hat{\eta}(k+l/k) \sim \mathcal{N}(0, \sigma_l^2) \quad (64)$$

then the variance is all we need in order to define its probability distribution and we can assume a certain *confidence interval* where the error is contained with a probability  $\alpha$ , as follows:

$$-z_{\frac{\alpha}{2}} \leq \hat{e}(k+l/k) \leq +z_{\frac{\alpha}{2}} \quad (65)$$

Here  $z_{\frac{\alpha}{2}}$  is the value of the probability distribution such that:

$$P\{\hat{e}(k+l/k) \geq +z_{\frac{\alpha}{2}}\} = P\{\hat{e}(k+l/k) > -z_{\frac{\alpha}{2}}\} = \frac{1-\alpha}{2} \quad (66)$$

where  $P\{\cdot\}$  is the probability function:

$$P\{\hat{e}(k+l/k) \leq z\} \triangleq \int_{-\infty}^z p(y)dy \quad (67)$$

As the distribution of the forecasting error is considered to be zero-mean Gaussian, the probability density function  $p(\cdot)$  assumes the following structure:

$$p(\hat{e}(k+l/k)) = \frac{1}{\sqrt{2\pi}\sigma_l} e^{-\frac{\hat{e}(k+l/k)^2}{2\sigma_l^2}} \quad (68)$$

The estimate of the variance  $\sigma_l^2$  could be calculated from the specific model parameters and its estimation algorithm, which however is not really easy and also could be misleading if the model is not accurate enough. A more straightforward alternative, however, is adopted at this stage, where the estimate of the variance of the forecasting error is based on the past history of the predictions:

$$\hat{\sigma}_l^2 = \frac{1}{N-1} \sum_{k=1}^N [\hat{e}(k+l/k)^2] \quad (69)$$

where  $N$  is the number of past observations available.

The estimate can also be recursively updated as soon as new observations become available [29]:

$$\hat{\sigma}_l^2(k+l) = \frac{k+l-2}{k+l-1} \hat{\sigma}_l^2(k+l-1) + \frac{1}{k} [\eta(k+l) - \hat{\eta}(k+l/k)]^2 \quad (k+l \geq 2) \quad (70)$$

An iterative estimation where more weight is given to the recent past can be obtained through an *exponential forgetting*, represented by a forgetting factor  $\lambda < 1$  [29]:

$$\hat{\sigma}_l^2(k+l) = \frac{2\lambda-1}{\lambda} \hat{\sigma}_l^2(k+l-1) + (1-\lambda) [\eta(k+l) - \hat{\eta}(k+l/k)]^2 \quad (71)$$

## 4 Results

The possible forecasting models proposed in section 3 were tested on some significant sample data sets, appropriately chosen among all those available (refer to section 2) as representative of different sea conditions. In particular, refer to figure 15, one wide banded and one narrow banded sea state from the two sites (Galway Bay and Pico) are considered. Then, a situation where wind waves predominate is picked up from the Galway Bay data and a very high energy wave system, where the sea bottom slightly affects the wave symmetry (this was analysed through higher order spectral analysis and skewness and kurtosis indices, section 2), is chosen from the Pico Island data. For each of the data sets two different cutoff frequencies are applied in the lowpass filtering preprocessing of the data, based on their spectral shape. Each of the data sets is split up into a training and validation set. For the Galway bay data, training and validation consists of two consecutive data sets of 3072 samples each (20 minutes at a sampling frequency of  $2.56 Hz$ ). In the case of the Pico Island data, because the consecutive data sets are actually contiguous in time, training and validation data consist of 4 consecutive sets each (9216 samples, equivalent 2 hours at a sampling frequency of  $1.28 Hz$ ), and the validation data follows the training data *continuously* in time.

The prediction accuracy is measured with the following goodness-of-fit index, which depends on the forecasting horizon  $l$ :

$$\mathcal{F}(l) = \left( 1 - \frac{\|\eta(k+l) - \hat{\eta}(k+l/k)\|_2}{\|\eta(k)\|_2} \right) \cdot 100 \quad (72)$$

Here  $\|\cdot\|_2$  is the Euclidean norm operator (root sum squared) over all the sampling instants  $k$  of the simulation,  $\eta(k+l)$  is the wave elevation and  $\hat{\eta}(k+l/k)$  is its prediction based on the information up to instant  $k$ . A 100% value for  $fit(l)$  means that the wave elevation time series is perfectly predicted  $l$  steps into the future.

Note that the quantity  $\mathcal{F}(l)$  has a direct correspondence with the variance of the prediction error, that, as discussed in section 3.3, is utilised to characterise confidence intervals of the forecasts. In particular, if  $\hat{\sigma}_l^2$  is an estimate of the  $l$ -step ahead prediction error, then:

$$\hat{\sigma}_l^2 = \left( 1 - \frac{fit(l)}{100} \right) \|\eta(k)\|_2 \quad (73)$$

### 4.1 Cyclical models

The two structures for the cyclical models outlined in section 3.2.1, Harvey's structural model and the Dynamic Harmonic Regression (DHR), were tested against the data sets of figure 15. The frequencies of the models were chosen as constantly spaced in a range between  $0.3 rad/s$  (practically no waves appear below this frequency at the considered locations) and the applied cutoff frequency  $\omega_c$ , and different spacings  $d\omega$  are compared. Another critical choice for the models are the initial mean value and covariance matrix of the state for the Kalman Filter algorithm, and the variance of the output equation disturbance, specifically  $\zeta(k)$  in equations (16) and (24). Initial expected value and covariance matrix of the state vector are determined from regular least squares on

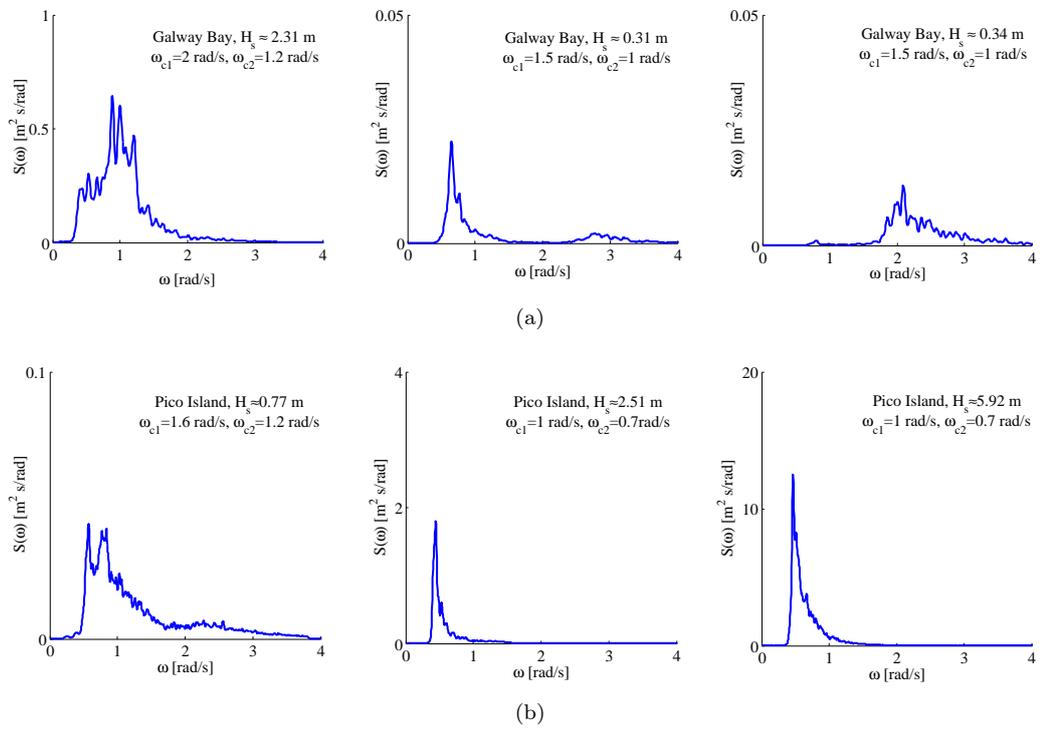
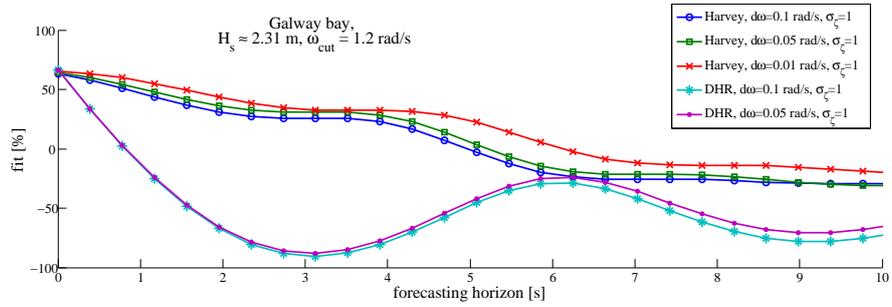


Figure 15: Data sets utilised to test the models of section 3.2: (a) Galway; (b) Pico.

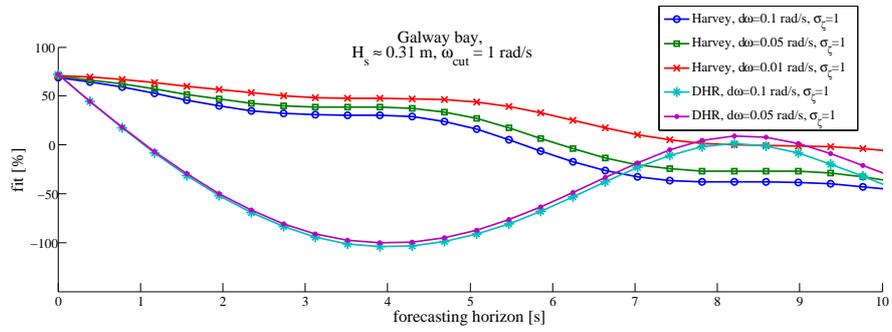
the training data. The variance of the output disturbance  $\zeta(k)$ , namely  $\sigma_\zeta^2$ , on the other hand, is determined by trial and error, a reasonable value found to be  $\sigma_\zeta^2 = 1$ .

Figure 16 shows the performance of different cyclical models on the Galway Bay data sets, when the lowest cutoff frequency is considered. The accuracy is above 50% only for predictions more than 2 – 3 seconds into the future (around 5 – 8 samples) and drops down particularly quickly for the sea state constituted by mostly wind waves, figure 16(c). Slightly better results are obtained for the Pico Island data, figure 17, when the  $\mathcal{F}(l)$  is above 50% until even 5 – 6 seconds ahead (around 6 – 8 samples), although this is not the case for the wide-banded sea state in figure 17(a). In all the cases, the Harvey models give better results, which improve by decreasing the frequency spacing, that is by considering a model with a more dense range of frequencies. The main problem of these models, however, lies in their complexity: a spacing of  $d\omega = 0.01$  with a range  $[0.3, 1.2]$  rad/s generates a state space model of order 182 for Harvey’s cyclical model and order 364 for the Dynamic Harmonic Regression (DHR) model. This sets a strict limit, in practical applications, on how small the spacing can be.

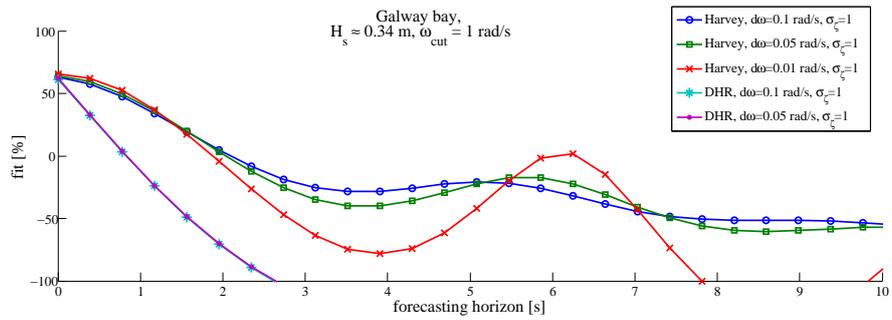
Some sample detailed time series plots are shown in figure 18, representing the predicted wave with a certain fixed lead time against the real filtered wave. Note that a 90% confidence interval is also shown, and it is calculated under the assumptions that the forecasting error is Gaussian. Its variance is estimated in real time according to equation (70). Figure 19 shows that, effectively, in nearly 70% of the cases, the error distribution has a *kurtosis*  $< 3.01$  and a *skewness*  $< 0.05$ , which means it is very close to Gaussian.



(a)

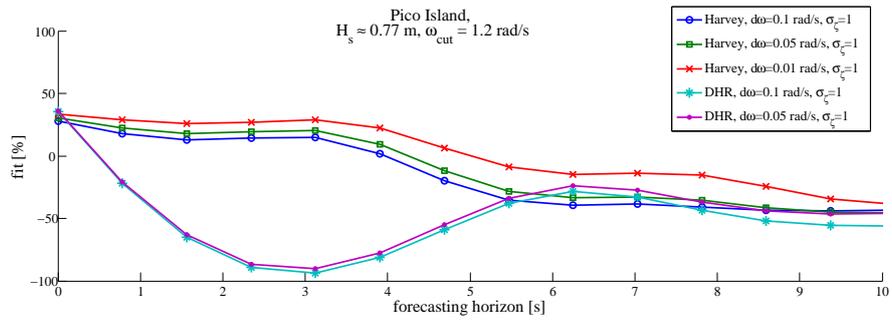


(b)

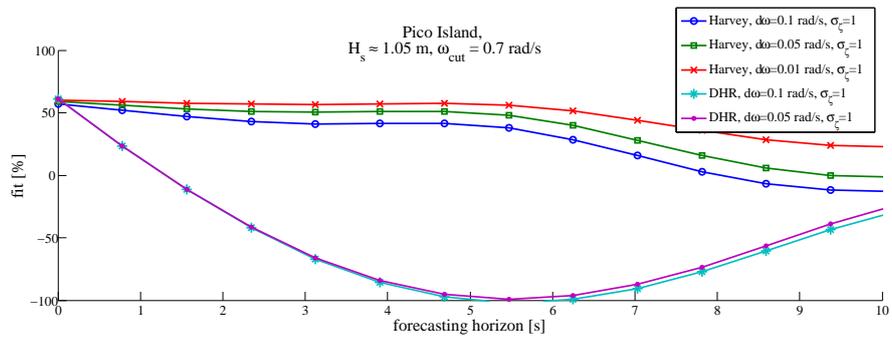


(c)

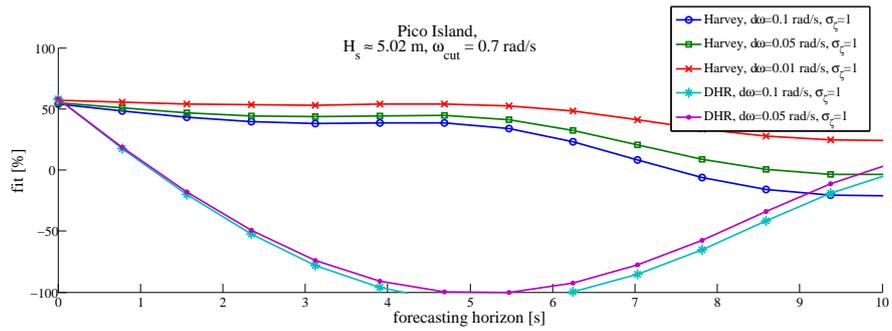
Figure 16: Prediction accuracy of cyclical models on wave data from Galway Bay



(a)

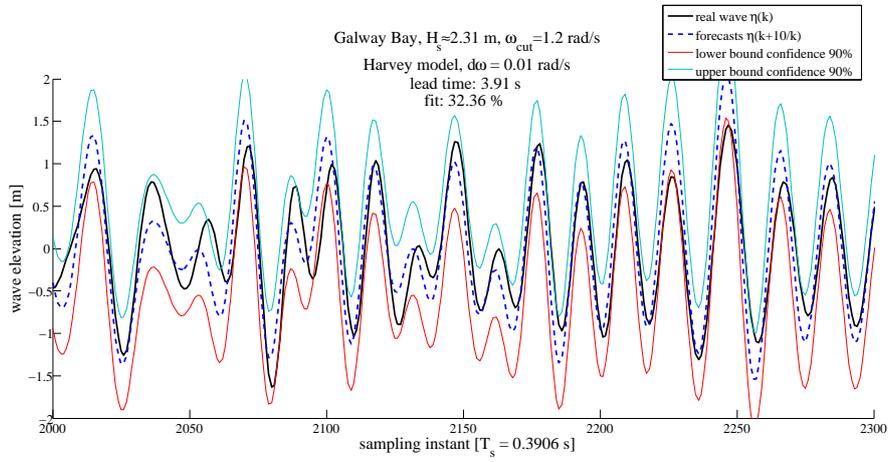


(b)

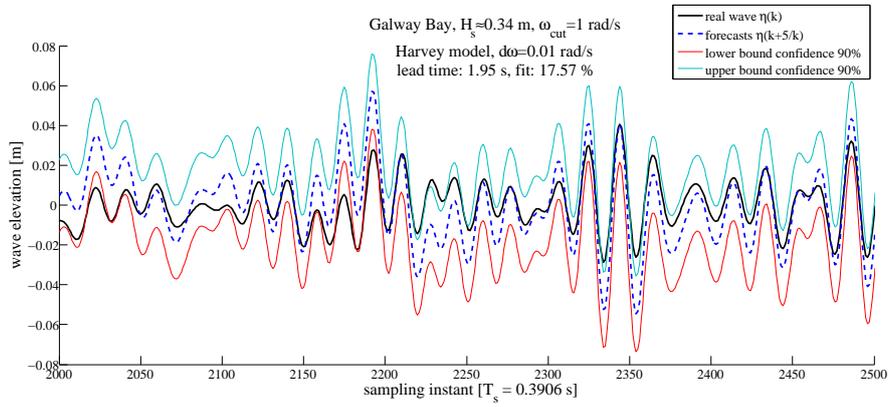


(c)

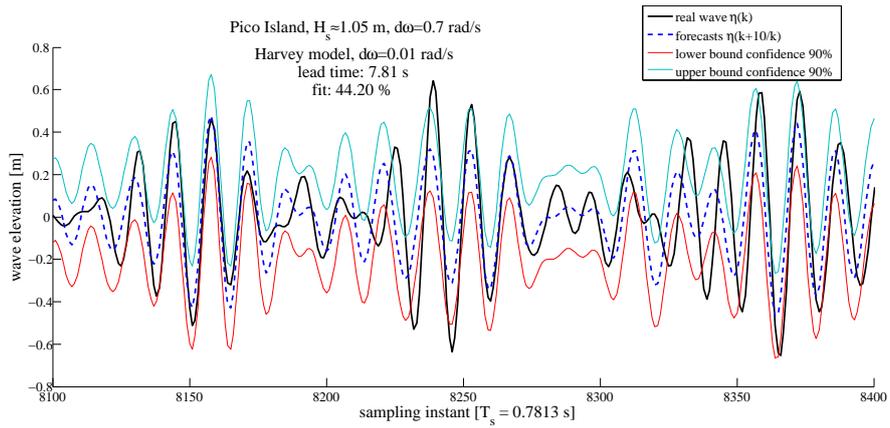
Figure 17: Prediction accuracy of cyclical models on wave data from Pico Island



(a)



(b)



(c)

Figure 18: Some time series details of the multi-step ahead wave prediction, and its 90% confidence, with cyclical models

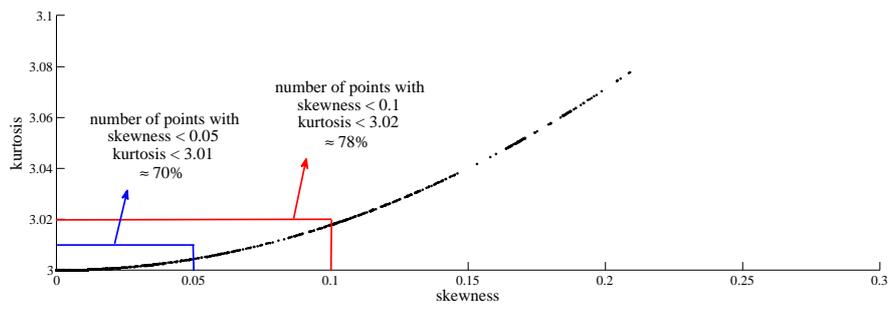


Figure 19: Distribution of skewness and kurtosis of the  $l$ -step ahead prediction error, over different  $l$  and for different cyclical models over all the considered data sets from both Pico and Galway

## 4.2 Auto-Regressive models

Before presenting the actual results achieved with static and adaptive AR models for the problem of wave forecasting, given in sections 4.2.2 and 4.2.3, the estimation procedure is described in detail, because the nature of the problem makes it different from the usual applications. The choice of an appropriate model order is, in fact, performed through the application of slightly modified standard criteria on the basis of a multi-step ahead prediction error, instead of the normally adopted one-step ahead prediction error, section 4.2.1.

### 4.2.1 Choice of model order

An indication about the appropriate order  $n$  of an AR model to predict the wave elevation, can be obtained through minimisation of the classical criteria Akaike Information Criterion (AIC), proposed by Akaike [30], and Bayesian Information Criterion (BIC), proposed by Schwarz [31]:

$$AIC = \log(\hat{\sigma}_\zeta^2) + r \frac{2}{N} \quad (74)$$

$$BIC = \log(\hat{\sigma}_\zeta^2) + r \frac{\log(N)}{N} \quad (75)$$

where  $\hat{\sigma}_\zeta^2$  is the estimate of the variance of the disturbance  $\zeta(k)$ , as in equation (30),  $r = n + 1$  is the number of parameters of the model and  $N$  is the number of observations utilised in the estimation procedure.

Because, however, we are interested in multi-step ahead predictions, the two criteria are evaluated also with respect to the variance of the  $l$ -step ahead prediction error,  $\hat{\sigma}_l^2$ :

$$AIC_l = \log(\hat{\sigma}_l^2) + r \frac{2}{N} \quad (76)$$

$$BIC_l = \log(\hat{\sigma}_l^2) + r \frac{\log(N)}{N} \quad (77)$$

Figures 20 and 21 show the resulting values of the two indices for a range of AR models (from  $n = 1$  to  $n = 50$ ) applied to the different data sets, respectively of Galway Bay and of the Pico Island. As expected the answers are different depending on the lead time (i.e. forecasting horizon) considered. So, accurate predictions very far in the future require relatively high order models (more than order 30), while good predictions for mid-range lead times, between 5 and 20 samples, can be obtained with reduced order AR models (order between 12 and 20). The model order depends also on the spectral range considered and on the specific sea state or location but, in general, an order slightly higher than 30 would be a good choice for any situation. In the following, therefore, a number of AR models are chosen based on these results, with the orders  $n = 12, 16, 24, 32$ .

### 4.2.2 Static AR models

For a proper comparison with the cyclical models results discussed in section 4.1, the performance, in terms of  $\mathcal{F}(l)$ , of different order AR models on the same sea states is shown in figure 22, for the Galway Bay data, and in figure

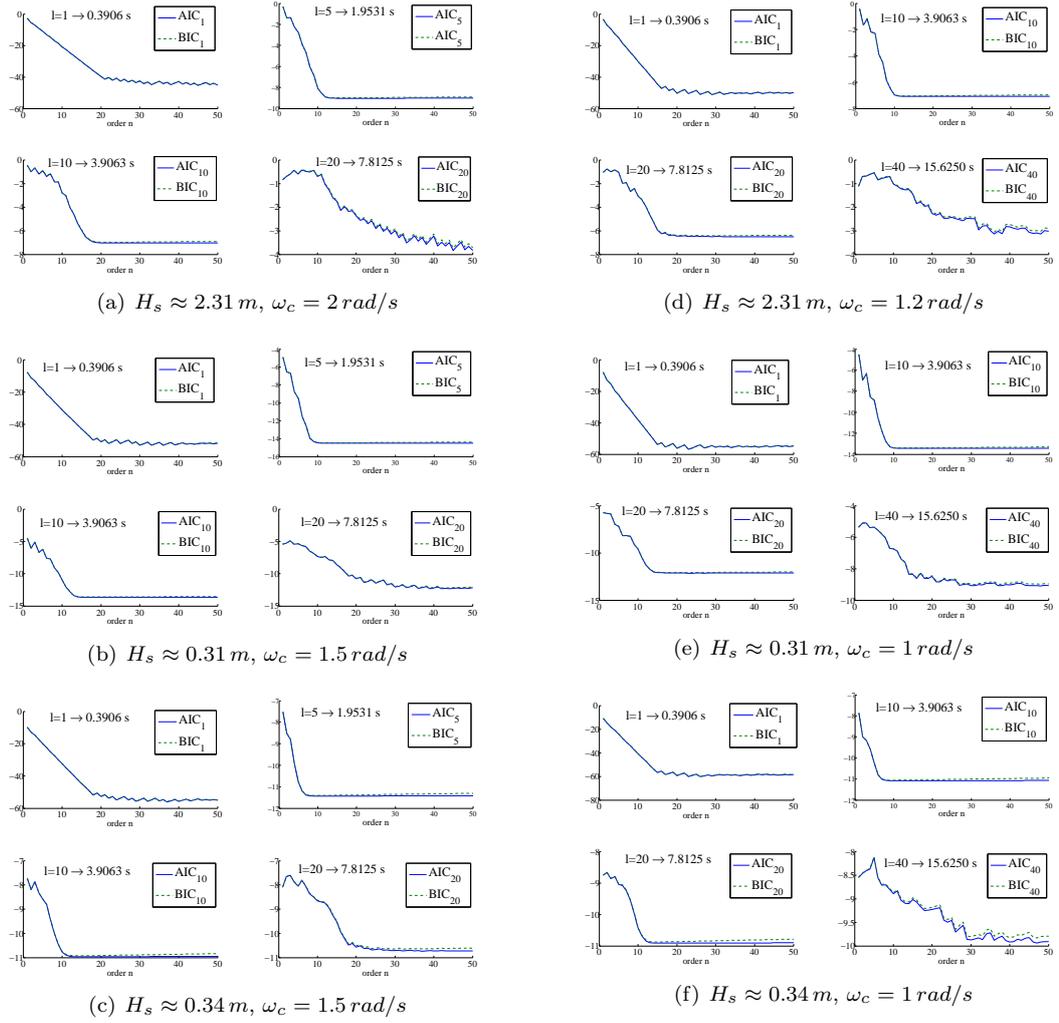


Figure 20:  $AIC_l$  and  $BIC_l$ , for some  $l$ , evaluated for the Galway bay data sets when different cutoff frequencies  $\omega_c$  are considered.

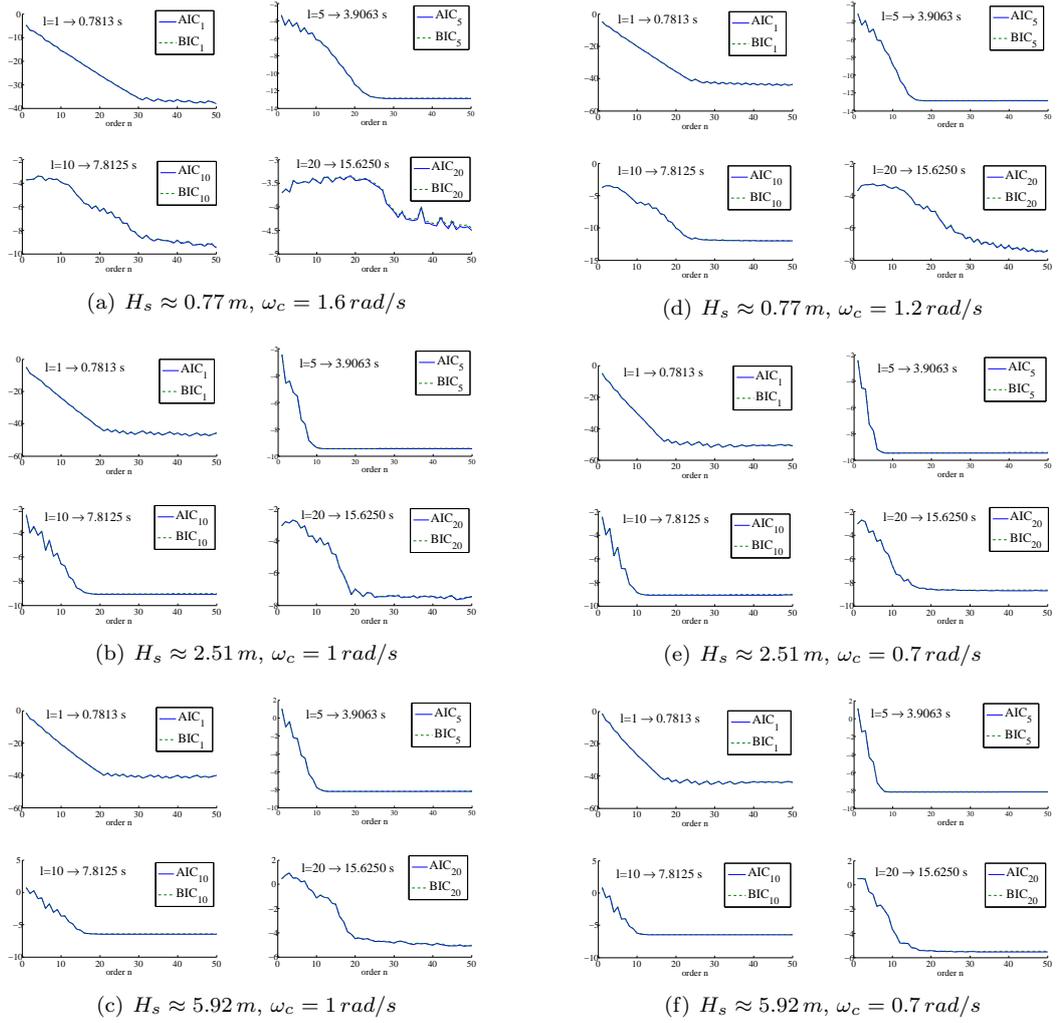


Figure 21:  $AIC_l$  and  $BIC_l$ , for some  $l$ , evaluated for the Pico island data sets when different cutoff frequencies  $\omega_c$  are considered.

23, for the Pico Island data, using constant parameter AR models. In any case, the AR models significantly outperform any of the considered cyclical models. PredictionS with an accuracy of more than 70% are obtained, in the case of Galway Bay, for a forecasting horizon of up to 15 seconds, and for almost 20 seconds in the case of the narrow-banded sea state, as seen in figure 22(b). The sea state composed mostly of wind waves still presents some difficulties with respect to the others, as seen in figure 22(c), but its wave energy significance is quite poor, and the prediction accuracy is still around 70% for almost 10 seconds into the future.

Even better results are obtained for the narrow-banded high energy sea states of the Pico Island (see figures 23(b) and 23(c)), where an accuracy of more than 90% is maintained for predictions up to 20 seconds in the future, even with lower order models with  $n = 26$  and  $n = 24$ , as it was expected from the  $AIC_i$  and  $BIC_i$  criteria for the Pico data, plotted in figure 21(e).

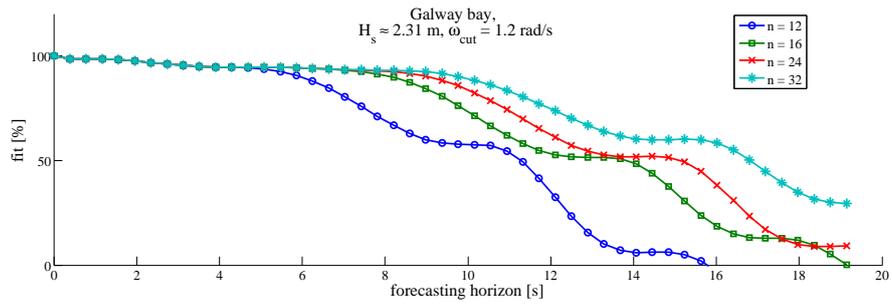
The same time series plots as for the cyclical models are shown in figure 24 for comparison, along with the confidence of the predictions. Note that the assumption of Gaussianity for the prediction error is even stronger in this case, as it can be seen from figure 25.

### 4.2.3 Adaptive AR models

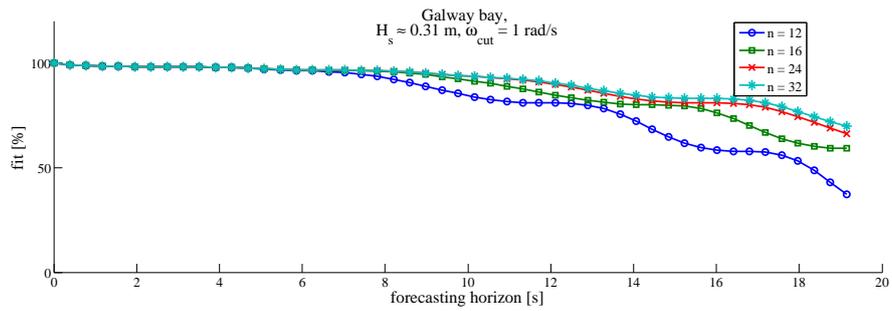
The results shown so far considered only *static* AR models estimated through a one-off batch least squares parameters estimation on a fixed training data set. From the discussion about AR models in section 3.2.2, we know that they can be seen as cyclical models, where the frequencies are strictly dependant on the parameters, so that they might not be suitable to model all the different sea states that may occur at a certain location. For example, consider figure 26, where the estimated AR model spectrum is shown against the training wave data spectrum, 26(a), and against the spectrum of the initial waves, 26(b), and of the final part of the wave data, 26(c), of the validation set. It can be seen how the AR model spectrum can become unrepresentative of the actual current wave climate, and so it might need to be able to track its variations. As regards the actual prediction results, the accuracy of the model seems to be quite robust to lots of changes in the spectral shape, but the models have not been tested long enough to give a definitive answer.

One solution to this problem might be to simply implement a recursive estimation algorithm to adapt the AR model to each new observation, through recursive least squares or through the Kalman filter, as explained in section 3.2.2. Results obtained with this approach, however, are not as good as if the parameters are kept constant after a batch estimate.

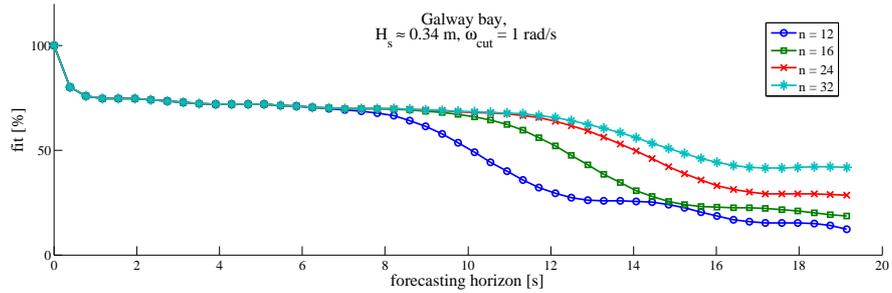
We have shown that static AR models keep a great accuracy for the full 2 hours of simulation, once they are trained on the basis of the previous 2 hours data, even if the sea state undergoes some major changes. Therefore, one practical possibility might be to run a batch estimation of the AR model, through least squares, every certain amount of time (e.g. every 2 hours) so to be sure that the model is always valid for the current sea state. The real-time adaptivity of the AR model does not seem, therefore, to be a crucial problem to solve at the moment and is not further investigated here.



(a)

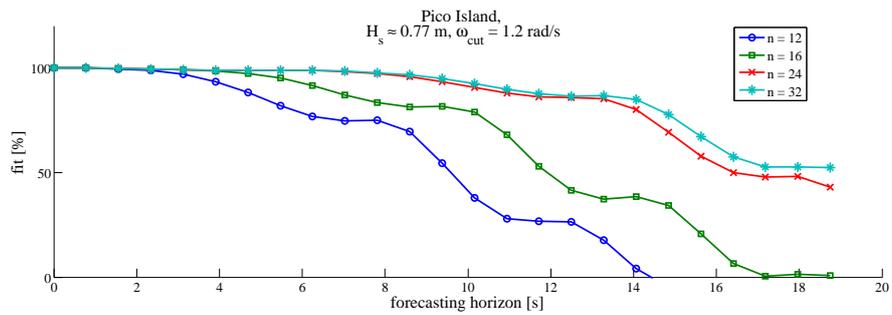


(b)

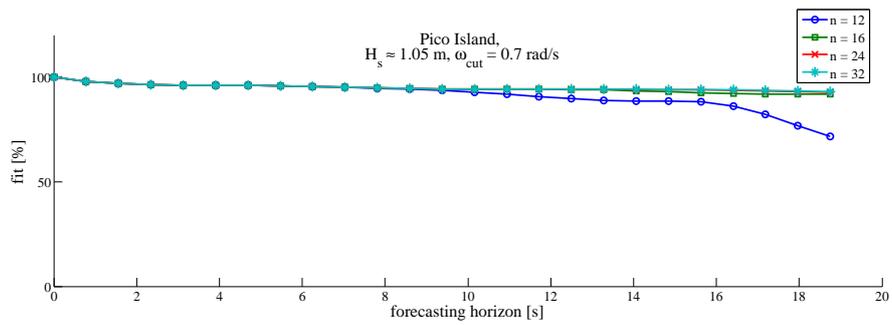


(c)

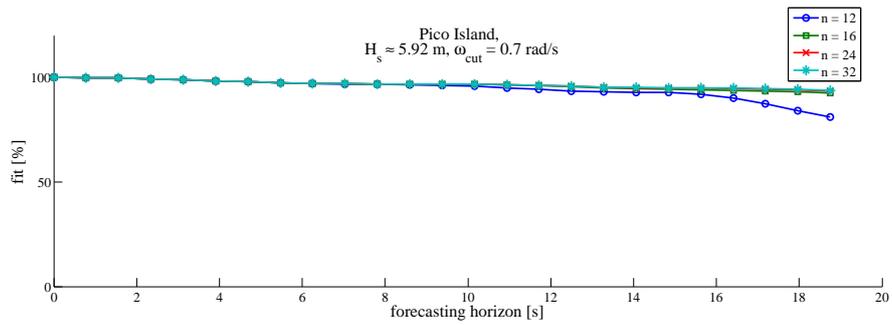
Figure 22: Prediction accuracy of AR models on wave data from Galway Bay



(a)

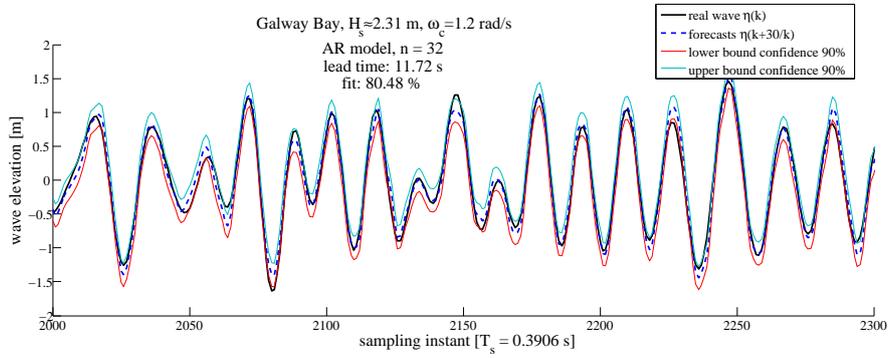


(b)

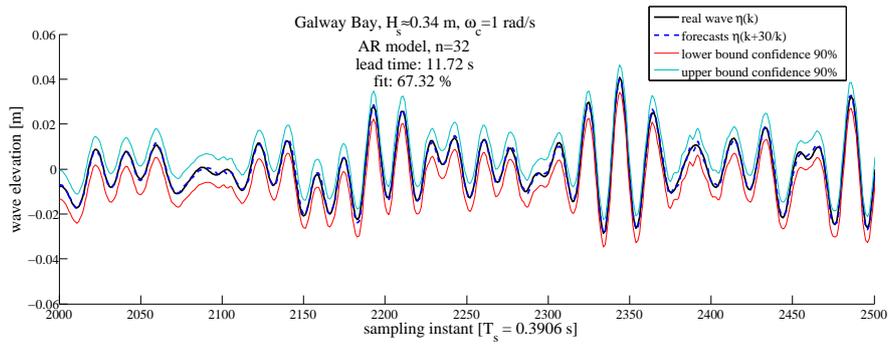


(c)

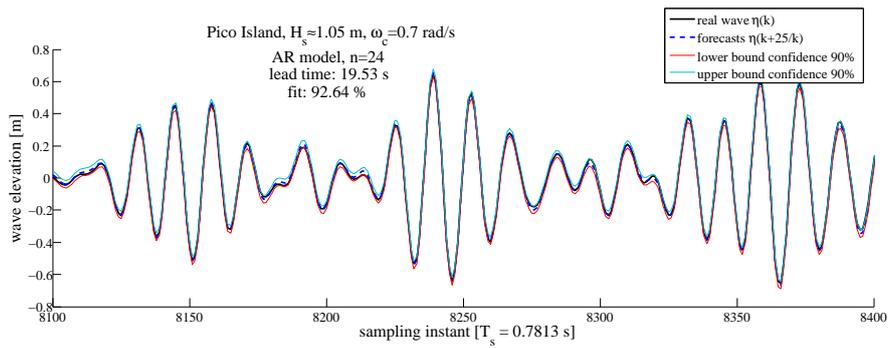
Figure 23: Prediction accuracy of AR models on wave data from Galway Bay



(a)



(b)



(c)

Figure 24: Some time series details of the multi-step ahead wave prediction, and its 90% confidence, with AR models

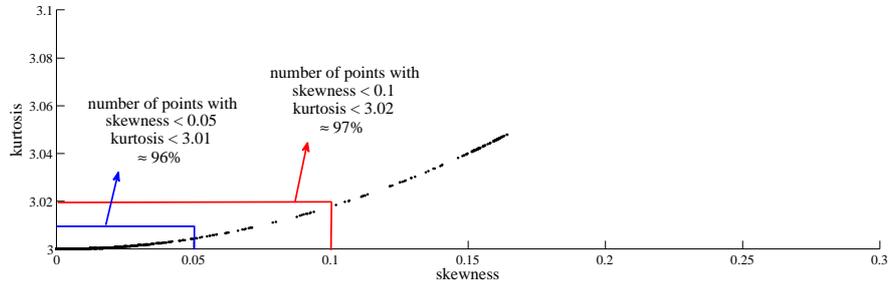


Figure 25: Distribution of skewness and kurtosis of the  $l$ -step ahead prediction error, over several  $l$  and for different fixed AR models over all the considered data sets from both Pico and Galway

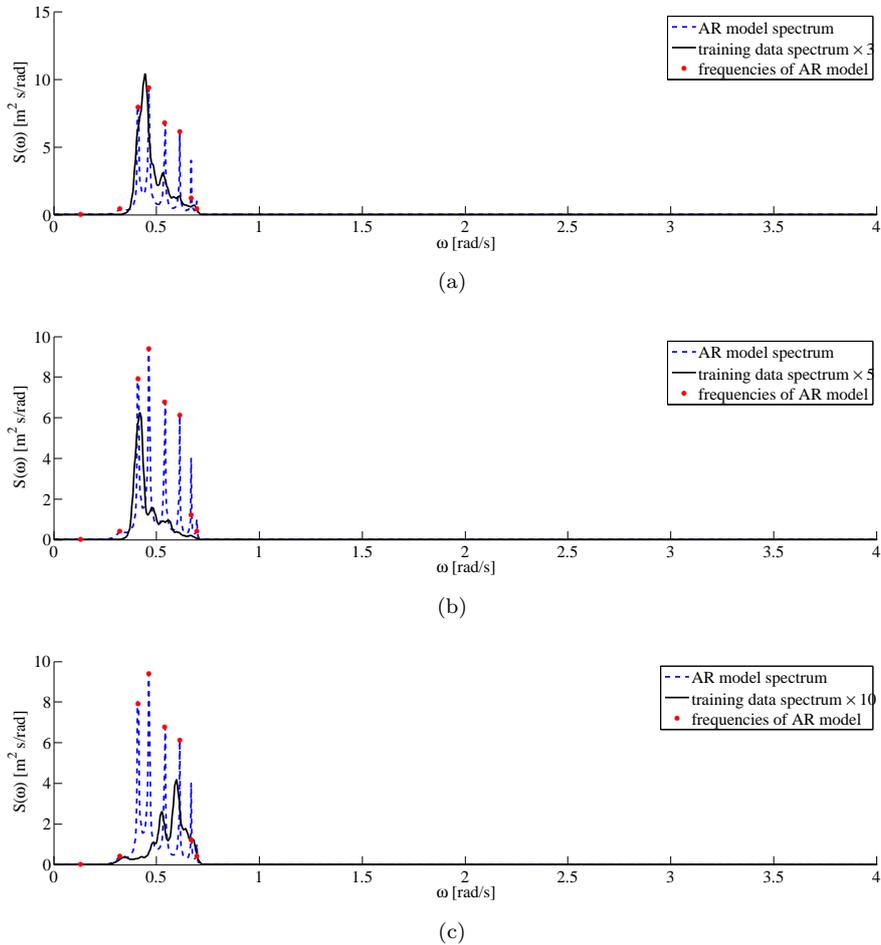
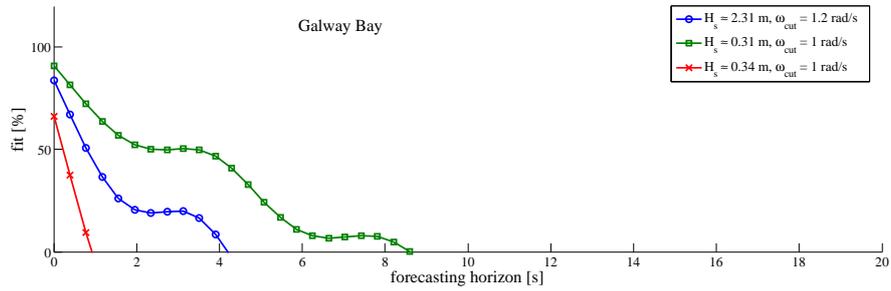
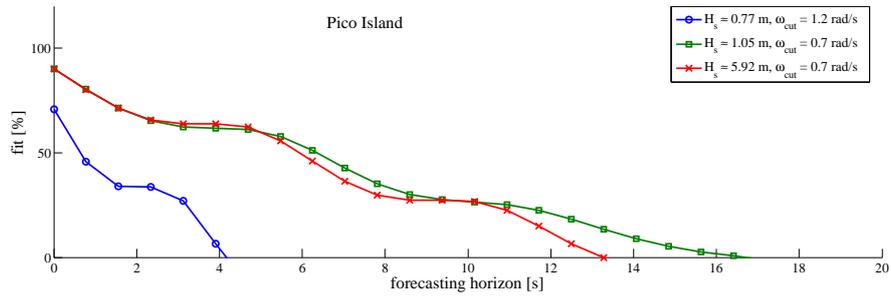


Figure 26: AR spectrum compared to wave spectra, which are amplified accordingly for clarity of exposition.



(a)



(b)

Figure 27: Prediction accuracy of single component cyclical model with variable frequency estimated through the Extended Kalman Filter

### 4.3 Sinusoidal extrapolation through EKF

In section 3.2.3, the possibility to deploy a harmonic model with a single variable frequency adapted on-line with the observations by means of the Extended Kalman Filter, was described. The advantages of this approach are its simplicity (state variable only of dimension 3: frequency, amplitude and phase) and the straightforward physical meaning of the model components. However, as is shown in the results of figure 27, it offers acceptable predictions only in the case of narrow-banded sea states, while it is completely inaccurate in other situations. This is of course due to the fact that the model is capable of tracking only a single dominant frequency in the waves. The approach is, though, a very interesting one and it would become an important solution if a proper way to integrate multiple frequencies is found. In section 3.2.3, in fact, it was stated that a pure superposition of single frequency sub-models of this kind is not an attractive solution, because the Kalman Filter weights the innovation of all the single frequencies with the same Kalman gain, as it can be seen in figure 28, so that the prediction effectiveness of the model drops significantly. A different modelling of the multiple frequency components is required so that the Kalman Filter is able to estimate them according to different dynamics.

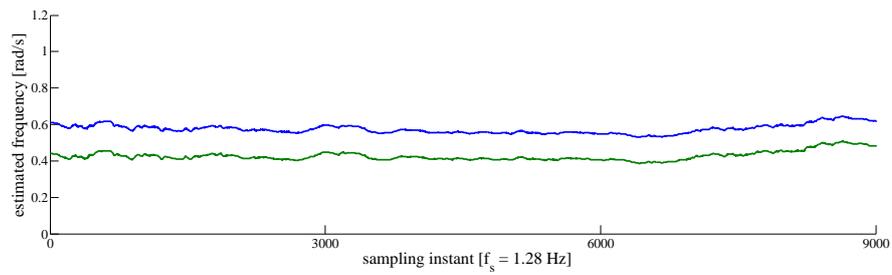


Figure 28: Frequencies of a two components cyclical model, estimated through the Extended Kalman Filter. The data set utilised is from Pico Island,  $H_s = 1.05 m$ , from the spectrum represented in figure 28.

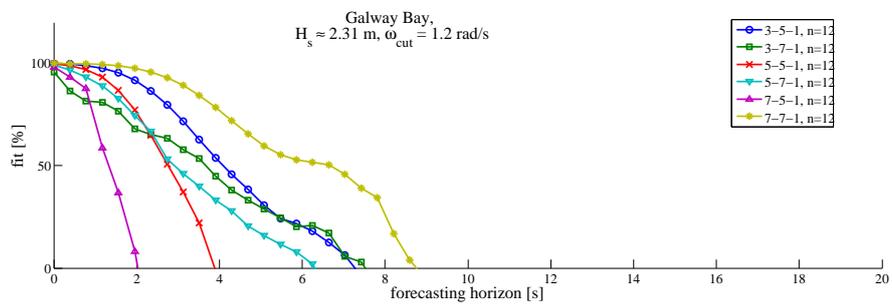
#### 4.4 Neural networks

Several architectures for a feedforward neural network were compared using the wave data of figure 15. Only structures with 1 linear output neuron and 2 hidden layers, made of a number of non-linear neurons (hyperbolic tangent sigmoid transfer function is applied) varying between 3 and 7 each, were considered. Several orders of regression  $n$  (i.e. the number of inputs of the network) were also considered, ranging from 12 to 32, based on the results relative to the linear AR models.

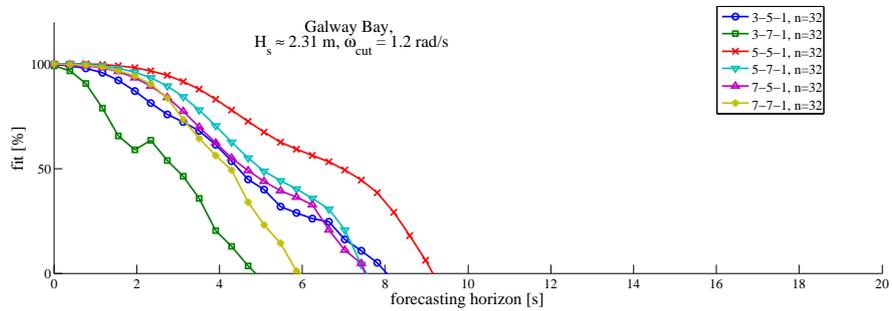
Figures 29 to 32 show  $\mathcal{F}(l)$  obtained using some of the structures. The first thing to notice is that, like all the other forecasting models considered, the neural networks perform much better for narrow-banded sea states (figures 30 and 32). None of the considered structures, though, achieves the same level of accuracy as the AR models, for relatively long forecasting horizons (more than 6 – 7 seconds). In the case, however, of the highly non-linear sea state of Galway Bay,  $H_s \approx 0.34 m$ , mostly consisting of wind waves, figure 31 shows how neural networks give a much better accuracy,  $\mathcal{F}(l) \approx 100\%$  than the AR models,  $70\% < \mathcal{F}(l) < 80\%$ , as from figure 22(c)) for up to 6 second ahead predictions. This kind of sea state, however, has a very low importance from a wave energy perspective, so that a significantly better behavior in such situations can not be really considered as a major decision variable in favor of neural networks.

In figure 33, some detailed prediction on two data sets are represented along with the confidence interval, determined under the assumption of Gaussian prediction error, according to the methodology outlined in section 3.3. Note, however, that these confidence intervals are not really accurate as most of the times the true signal lies outside. This is due to the fact that the prediction error, in this case, unlike for AR models, is not really close to having a Gaussian distribution, as it can be seen from the distribution of skewness and kurtosis of the prediction error, shown in figure 34.

The conclusion is that neural networks, with respect to AR models, do not seem to be capable to offer an improvement in wave forecasting, sufficient to justify their adoption. It has to be considered, in fact, that neural networks also introduce a much higher computational burden and represent a completely black-box approach with no physical meaning of its components.

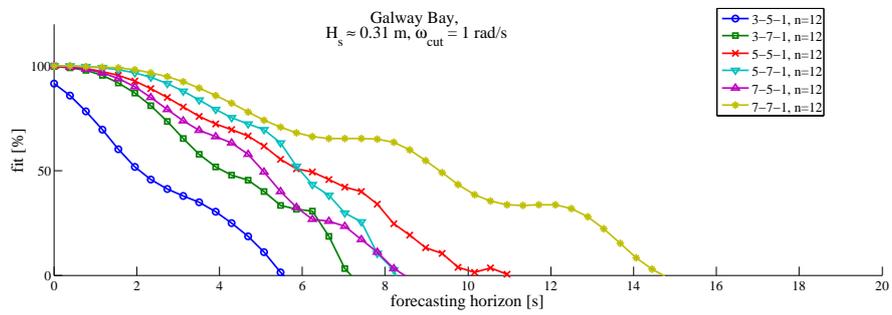


(a)

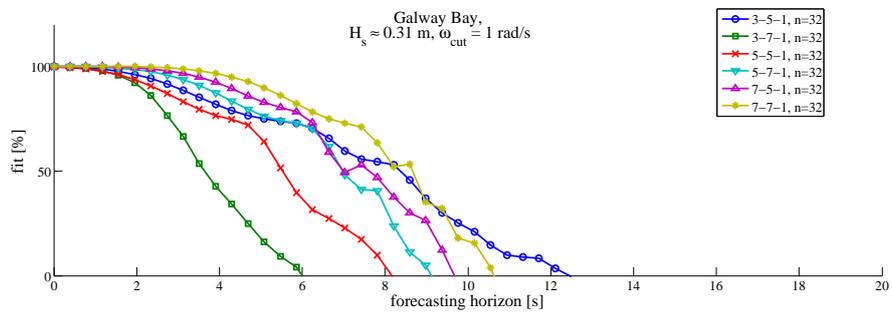


(b)

Figure 29: Prediction accuracy of neural networks

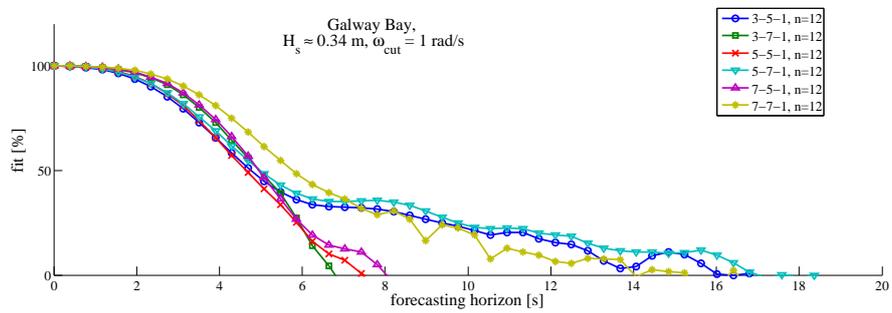


(a)

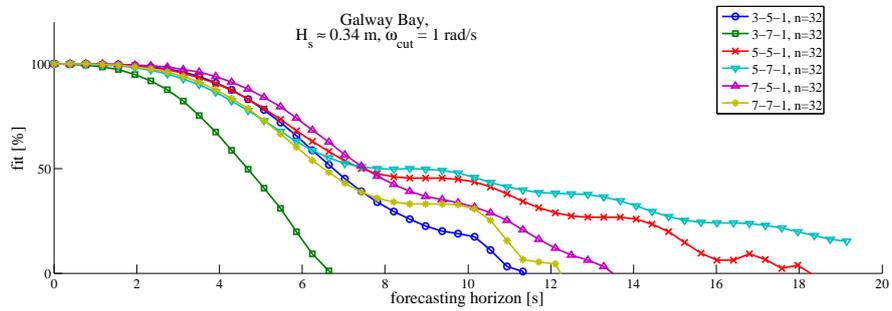


(b)

Figure 30: Prediction accuracy of neural networks

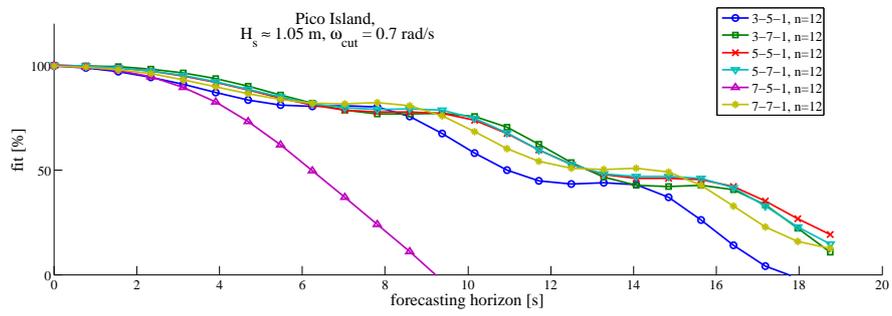


(a)

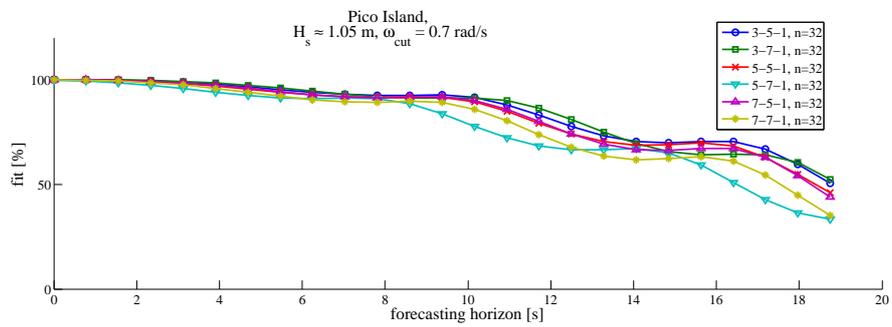


(b)

Figure 31: Prediction accuracy of neural networks

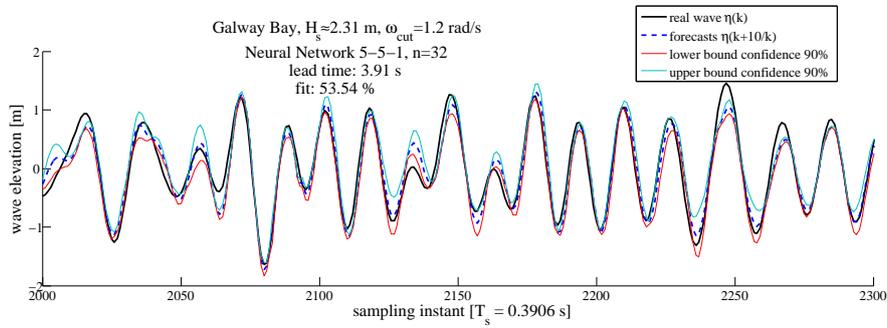


(a)

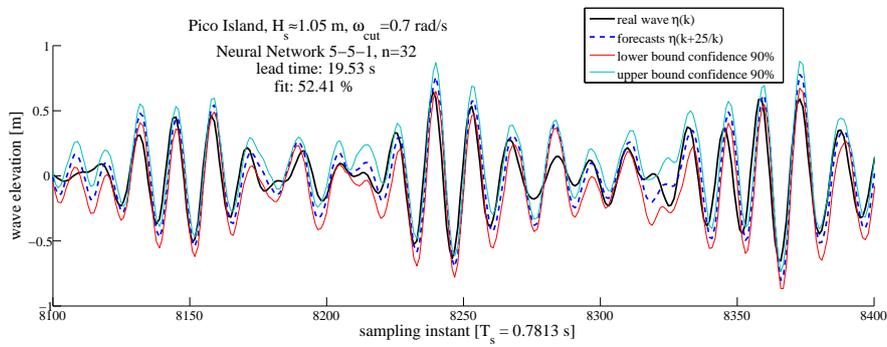


(b)

Figure 32: Prediction accuracy of neural networks



(a)



(b)

Figure 33: Some screenshots of the multi-steps ahead wave prediction, and its 90% confidence, with neural networks

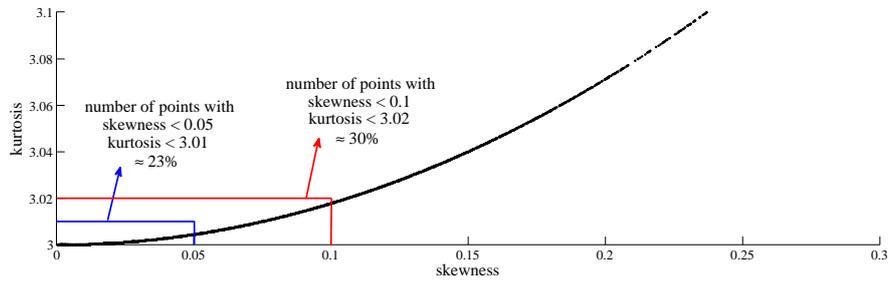
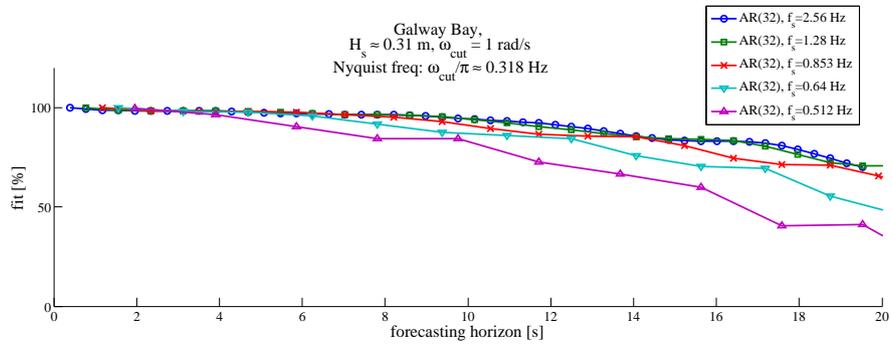


Figure 34: Distribution of skewness and kurtosis of the  $l$ -step ahead prediction error, over several  $l$  and for different neural networks over all the considered data sets from both Pico and Galway

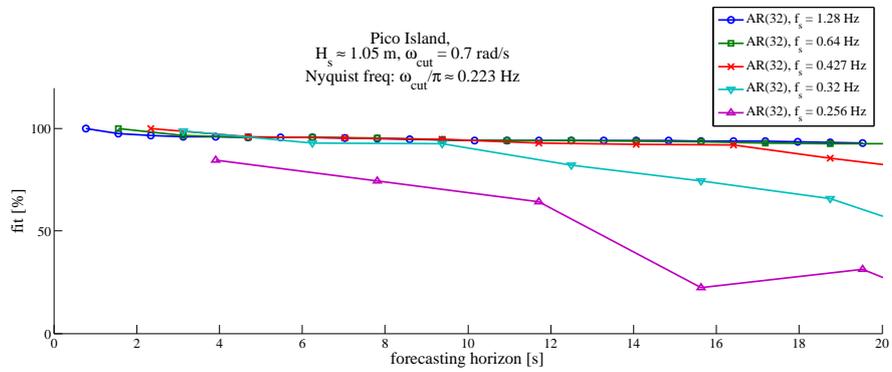
## 4.5 Effects of sampling frequency on prediction

Following the discussion in section 2.6, it is interesting to analyse how the sampling frequency of the wave elevation time series can affect the results achieved with some of the models tested throughout this section. In 2.6 it was pointed out how a model capable of extracting all the information about the future evolution from past observations, shall not be affected by changing the sampling frequency, because, within the limits allowed by the Nyquist theorem, the amount of information in the signal does not change.

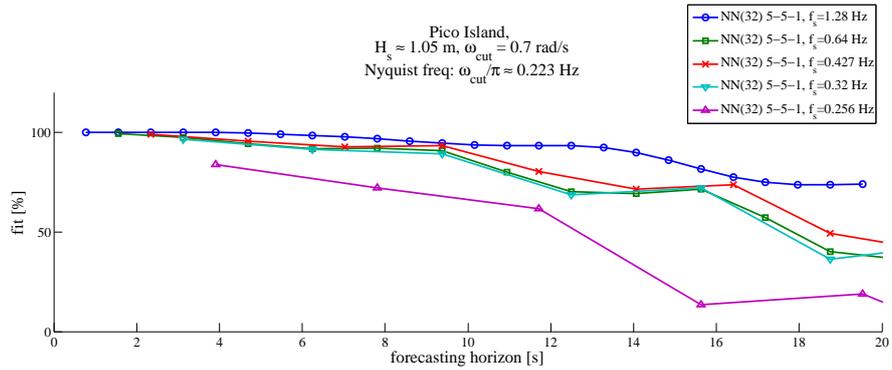
Figure 35 confirms this view, as regards AR models and neural networks, with their ability to predict not affected by a change in the sampling frequency. It can be concluded that reducing the sampling frequency does not increase the forecasting horizon achievable with these prediction algorithms.



(a)



(b)



(c)

Figure 35: Prediction accuracy of AR models and neural networks when the sampling frequency of the data is decreased

## 5 Further possibilities

Some other possible wave forecasting models were considered along with the ones presented in section 3. Gaussian Processes are discussed in section 5.1, and it will be pointed out how they share some of the properties of cyclical models with fixed frequencies, so that they are not seen as a viable solution. Then, the properties of ARMA models are analysed in section 5.2 and the reasons why they were not considered, as compared to AR models, are discussed.

### 5.1 Gaussian Processes

Assuming that a certain function underlies the observed data sets, in our case a function relating the past and future values of the wave elevation time series, the idea behind a Gaussian Process (GP) model is to place a prior directly on the space of functions [32], without assuming any particular function parametrisation. The only hypothesis assumed is that this prior is considered to be Gaussian and is completely specified by its mean and covariance function.

In particular, the prior distribution,  $P(y)$ , of the function  $y$  performing the input-output mapping is given as a zero-mean (a mean value different from zero would not affect any of the following considerations) Gaussian process with covariance matrix  $Q$ , denoted as  $\aleph(0, Q)$ :

$$P(y) \sim \aleph(0, Q) \quad (78)$$

Denoting the input by  $\{u_n\}$ ,  $n = 1, \dots, N$  and the corresponding target values by  $\{t_n\}$ ,  $n = 1, \dots, N$ , and assuming that each target value  $t_n$  differs by additive Gaussian white noise of variance  $\sigma_\epsilon^2$  from the corresponding function value  $y(x_n)$ , then the target values follow a Gaussian prior distribution as well:

$$P(t) \sim \aleph(0, C) \quad (79)$$

$$C \hat{=} Q + \sigma_\epsilon^2 I \quad (80)$$

where  $I$  is an identity matrix of the same dimension as  $Q$ .

By means of the covariance matrix  $C$ , it is then possible to infer a target  $y(x_{N+1})$ , given all the observed targets  $T_N = [t_1, \dots, t_N]^T$ , according to the relationship:

$$P(t_{N+1}/T_N) = \frac{P(t_{N+1}, T_N)}{P(T_N)} \quad (81)$$

where  $P(t_{N+1}/T_N)$  is the probability of  $t_{N+1}$  conditioned on all the observations  $T_N$  and  $P(t_{N+1}, T_N)$  is the joint probability density.

It can be shown [32] that:

$$P(t_{N+1}/\mathbf{t}_N) = \frac{1}{Z} e^{\left[ -\frac{(t_{N+1} - \hat{t}_{N+1})^2}{2\sigma_{\hat{t}_{N+1}}^2} \right]} \quad (82)$$

$$\hat{t}_{N+1} = \mathbf{k}^T C_N^{-1} \mathbf{t}_N \quad (83)$$

$$\sigma_{\hat{t}_{N+1}}^2 = \kappa - \mathbf{k}^T C_N^{-1} \mathbf{k} \quad (84)$$

where  $\hat{t}_{N+1}$  is the mean value of the prediction at the new point  $x_{N+1}$  and  $\sigma_{\hat{t}_{N+1}}^2$  gives the error bars on this prediction.  $Z$  is a scalar coefficient, while

the matrix  $C_N \in \mathbb{R}^{N \times N}$ , the vector  $k \in \mathbb{R}^{N \times 1}$  and the scalar  $\kappa$  come from the following form of the covariance matrix  $C_{N+1} \in \mathbb{R}((N+1) \times (N+1))$  of the vector  $T_{N+1} = [t_1, \dots, t_{N+1}]^T$ :

$$\begin{bmatrix} C_N & k \\ k^T & \kappa \end{bmatrix} \quad (85)$$

The crucial aspect in building a GP model is therefore the choice of the covariance function  $C[y(x_i), y(x_j)]$  from which the covariance matrix  $C = \{C_{ij}\}$  can be calculated, according to the equations:

$$C_{ij} = C[y(x_i), y(x_j)] + \delta_{ij}N(x_i), \quad i, j = 1, \dots, N \quad (86)$$

where  $N(x_i)$  is a noise model, with  $\delta_{ij} = 1$  for  $j = i$ , otherwise  $\delta_{ij} = 0$ .

The covariance function usually has a particular parametric structure [33], and can result from the combination of elemental covariance functions expressing certain peculiar characteristics, such as a trend component, a cyclical component, etc... (a wide overview is given in [32]). The parameters in the covariance function (termed *hyperparameters*), once its form has been defined, are optimised with respect to the available observations through a maximum likelihood procedure [32]. The covariance function can then be utilised to calculate the covariance matrix  $C_{N+1}$  according to relation (86), by which the mean and variance of the estimate can be found from equations (83) and (84).

The attractiveness of GPs lies in the fact that models can be determined using a relatively small number,  $N$ , of observations and the covariance functions can be easily synthesised from standard components representing particular features in the data. GPs perform, moreover, in a safe manner when extrapolating outside the training data, by giving confidence intervals (on the basis of the variance of the estimate) which help to indicate where the model is unreliable (e.g. for forecasts too far ahead in time). One more positive feature is that on-line data addition, that is the introduction of new observations into the model, can be performed in a really straightforward manner (see [32] for details), assuming that it does not imply a change in the hyperparameters, thus permitting the model to adapt to the real process evolution.

The implementation of a GP model, however, presents some numerical problems in at least two ways. First of all, the optimisation of the hyperparameters is not a convex problem, and second (and more critical), the on-line estimation requires an  $N \times N$  matrix inversion, which is an ill-conditioned problem. This latter problem puts, of course, an upper limit to the quantity  $N$  of training data which can be included in the model.

The major drawback, in our specific case of wave forecasting, is that, in order to model the cyclical characteristics of the sea, the harmonic components of the covariance function, and in particular their frequency, have to be permanently assigned in the initial hyperparameter estimation procedure, which is the same problem of cyclical models with fixed frequencies, as articulated in section 3.2.1. A covariance function without explicit cyclical components could still be able to represent the harmonic behavior of the wave elevation, just like a simple AR model without explicit seasonal components, but then the extraction of the physical characteristics of the real process from the estimated model becomes barely possible. This particular reason led the author not to consider Gaussian Processes among the possible candidates for a proper model for wave forecasting.

## 5.2 ARMA models

An ARMA model is the equivalent of an AR model, where the noise is not assumed to be sequentially uncorrelated (i.e. white noise) but it is modelled by a regression:

$$\eta(k) = \sum_{i=1}^{n_a} a_i \eta(k-i) + \sum_{i=1}^{n_c} c_i \zeta(k-i) + \zeta(k) \quad (87)$$

and, in the  $Z$ -domain:

$$\eta(z) = \frac{\vartheta(z)}{\varphi(z)} \zeta(z) \quad (88)$$

The shape of the forecasting function  $\hat{\eta}(k+l/k)$  is the same as for an AR model, equation (32), and it is determined only by the autoregressive part, i.e. the poles of the corresponding transfer function. The moving average terms change the way in which the forecast function is fitted to the past observations and the forecasts. For a purely autoregressive model, the forecast function is that unique curve, of the form required by  $\varphi(z)$ , which passes through the  $n_a$  *pivotal* points  $\eta(k), \eta(k-1), \dots, \eta(k-n_a+1)$  [22]. In the presence of some moving average terms,  $n_c > 0$ , the forecast function is determined by the points  $\hat{\eta}(k+n_c/k), \dots, \hat{\eta}(k+1/k), \eta(k), \dots, \eta(k+n_c-n_a+1)$ , so that the way in which it is fitted to the observations is modified by the moving average coefficients through the first  $n_c$  predictions.

An ARMA model would not, therefore, present significantly different modelling capabilities than pure AR models. In practice, however, to obtain a parsimonious parametrisation, it is sometimes necessary to include both autoregressive and moving average terms in the model [22]. In fact, a finite moving average process can be written as an infinite autoregressive process, and vice-versa, so that if a process were really a Moving Average (MA) of order 1, an autoregressive model would necessarily be a non parsimonious representation of it, and conversely for a AR of order 1 process.

As a conclusion, an ARMA model would not allow us to obtain better results than simple AR models, but the introduction of some moving average terms may be considered at a later stage in order to reduce the complexity of the forecasting model.

## 6 Conclusion

This study was focused on the problem of short-term wave prediction, which is a central topic in the wave energy field, in order to allow a better effectiveness and economic viability of any WEC. A pure univariate time series forecasting approach was followed and several possible solutions were proposed. Real data from Galway bay and Pico island were available for testing the proposed solutions, and some interesting analysis was provided in section 2. In particular, a valuable tool for the predictability analysis, independent of any particular solution, was proposed in section 2.4. The quantification of this predictability showed how lower frequency waves are easier to predict and, from a wave energy point of view, high frequency components, which carry lower energy (as revealed by the Fourier analysis provided in section 2.1), may be filtered out to improve the prediction.

The most straightforward models outlined were harmonic models where the wave elevation is explicitly represented as a sum of sines and cosines, on the basis of linear wave theory. It was underlined how many issues (particularly the high complexity of the resultant models) arise due to the problem of the choice of frequencies when they are kept constant, so that reasonable predictions are only achieved for a maximum of 5-6 seconds in the future (even less, only 2-3 seconds, for the Galway bay data), if only low frequencies are predicted. Cyclical models with adaptive frequencies could have been considered, but then they become non-linear and the complexity will be even higher, so that other solutions should be explored first.

An analysis of AR models, in section 3.2.2, highlighted how they implicitly represent cyclical models, where the frequencies are easily estimated with linear least squares (as they are related to the regression coefficients). The amplitudes and phases of each harmonic component are, moreover, implicitly adaptive to the recent observations due to the regression terms of the model, so that only a batch estimate of the model offered very good accuracy up to 15 seconds (in some cases even 20 seconds) predictions for the low frequency components of the waves. It was also shown how the frequencies are automatically estimated in the significant range of the sample spectrum of the training data set. The possibility to adapt the model in order to track variations of the wave spectrum through variable frequencies in the model was also analysed, and it was concluded that no real benefit is obtained through real-time adaptive AR modelling. Because, however, static AR models were shown to maintain their prediction ability for long times (no performance decrease for 2 hours simulations), their adaptivity is not seen as a main issue at the moment, and also a simple periodic batch estimate may be a feasible solution, or the use of a set of AR models, estimated from different sea conditions, and a switching logic deciding which one is the more appropriate in real time.

A cyclical model with a single variable frequency was also presented in section 3.2.3, where the real time frequency estimation is realised through the EKF. The methodology revealed such a model to be reasonably effective for narrow-banded sea states, with good predictions for 5-10 seconds, but it is completely ineffective for wider-banded wave systems. It is, however, a very light and computationally simple solution (only 3 states), so that if a proper way to integrate multiple frequencies is found, some improvement in performance may be obtained. More work should be done in this direction, as simple aggregation of the single frequency models resulted in some problems and poor results.

Finally, a comparison with neural networks, section 4.4, showed how, although they offer an accuracy comparable with the *AR* models, it would not be very appealing to further undertake this more computationally expensive direction, as no significant improvement is expected. Moreover, neural networks they do not offer any possibility of analysis and extraction of the characteristics of the real process from the model, which would instead be very straightforward with AR models.

As a conclusion, Auto Regressive (AR) models (or more parsimonious ARMA models, from the discussion in section 5.2) seem to be well suited to compute wave predictions for more than one wave period into the future, when a focus is put on lower frequency waves. One issue, however, which was not considered at this stage, includes the lowpass filtering preprocessing procedure applied to the wave elevation time series. In the current work this filtering was performed of-

fine, so that ideal zero-phase filters could be implemented. In real applications, a certain phase distortion and a certain transition band should be considered. This problem is not trivial to treat and is therefore left for future studies.

It is fundamental, moreover, that further work is carried out in order to provide some indications and constraints about the required accuracy of the forecasts (and required prediction horizon), so that the capability of the proposed models can be properly judged. Such work will involve a study of the interconnections between wave absorbers, wave excitation and the control architecture, and will be fundamental before any further attempt to improve the results of this work is eventually undertaken.

## Acronyms

- AIC** Akaike Information Criterion  
**AR** Auto Regressive  
**ARMA** Auto Regressive Moving Average  
**BIC** Bayesian Information Criterion  
**DHR** Dynamic Harmonic Regression  
**EKF** Extended Kalman Filter  
**GP** Gaussian Process  
**MA** Moving Average  
**STFT** Short Term Fourier Transform  
**WEC** Wave Energy Converter

## Nomenclature

- $\mathcal{N}(\mu, \sigma^2)$  Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$   
 $|\cdot|$  module  
 $\angle$  phase  
 $\|\cdot\|_2$  Euclidean norm (root sum squared)  
 $\hat{\sigma}_l^2$  variance of  $l$ -step-ahead prediction error  
 $\hat{x}(k+l/k)$  optimal  $l$ -step-ahead prediction of  $x(k)$  based on information until instant  $k$   
 $\mathcal{F}(l)$   $l$ -step-ahead prediction goodness-of-fit  
 $\omega_c$  cut-off frequency in  $rad/s$   
 $\omega_s$  sampling radian frequency in  $rad/s$   
 $a^T$  vector  $a$  transposed  
 $AIC_l$  AIC based on  $l$ -step-ahead prediction error  
 $BIC_l$  BIC based on  $l$ -step-ahead prediction error  
 $E\{\cdot\}$  expected value  
 $H_s$  significant wave height  
 $R^2(l)$   $l$ -step-ahead predictability

## References

- [1] U. A. Korde, "Control system applications in wave energy conversion," *Proceedings of the OCEANS 2000 MTS/IEEE Conference and Exhibition*, vol. 3, pp. 1817–24, 2000.
- [2] J. Falnes, *Ocean Waves and Oscillating Systems*. Cambridge University Press, 2002.
- [3] J. Falnes, "A review of wave-energy extraction," *Marine Structures*, vol. 20, pp. 285–201, 2007.
- [4] K. Budal, J. Falnes, *et al.*, "The norwegian wave-power buoy project," *The Second International Symposium on Wave Energy Utilization*, Trondheim, June 1982.
- [5] M. R. Belmont *et al.*, "Filters for linear sea-wave prediction," *Ocean Engineering*, vol. 33, pp. 2332–2351, 2006.
- [6] J. Tedd and P. Frigaard, "Short term wave forecasting, using digital filters, for improved control of wave energy converters," *Proceedings of International Offshore and Polar Engineering (ISOPE)*, pp. 388–394, 2007.
- [7] V. Voronovich, B. Holmes, and G. Thomas, "A preliminary numerical and experimental study of wave prediction," *Proc. 6th European Wave and Tidal Energy Conference*, pp. 535–542, 2005.
- [8] G. Nolan, J. Ringwood, and B. Holmes, "Short term wave energy variability off the west coast of Ireland," *Proceedings of the 7th European Wave and Tidal Energy Conference, Porto, Portugal (EWTEC)*, 2007.
- [9] S. R. Massel, "Wavelet analysis for processing of ocean surface wave records," *Ocean Engineering*, vol. 28, pp. 957–987, 2001.
- [10] G. Stokes, "On the theory of oscillatory waves," *Transactions of the Cambridge Philosophical Society*, vol. 8, pp. 441–455, 1847.
- [11] M. K. Ochi, *Ocean waves: The stochastic approach*. Cambridge University Press, 1998.
- [12] M. K. Ochi, *Applied probability and stochastic processes*. Wiley Interscience, 1990.
- [13] H. P. Bernhard, "A tight upper bound on the gain of linear and nonlinear predictors for stationary stochastic processes," *IEEE Transactions on Signal Processing*, vol. 46, pp. 2909–2917, November 1998.
- [14] C. Shannon, "The mathematical theory of communication," *Bell System Technical Journal*, no. 27, pp. 379–423, 1948.
- [15] D. S. Jones, *Elementary Information Theory*. Oxford Applied Mathematics and Computing Science Series, Oxford: Clarendon Press, 1979.
- [16] A. Dionisio, R. Mendezes, and D. A. Mendes, "Mutual information: A measure of dependency for nonlinear time series," *Physica*, no. 344, pp. 326–329, 2004.

- [17] X. Hong and S. A. Billings, “Time series multistep-ahead predictability estimation and ranking,” *Journal of Forecasting*, vol. 18, pp. 139–149, 1999.
- [18] F. Fusco and J. Ringwood, “Linear models for short term wave forecasting,” *Proc. World Renewable Energy Congress (WRECX)*, pp. 1030–1035, 2008.
- [19] F. Fusco and J. Ringwood, “A study on short-term sea profile prediction for wave energy applications,” *Proceedings of the 8th European Wave and Tidal Energy Conference (EWTEC)*, pp. 756–765, Uppsala, Sweden, 2009.
- [20] A. C. Harvey, *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1989.
- [21] P. C. Young, D. J. Pedregal, and W. Tych, “Dynamic harmonic regression,” *Journal of Forecasting*, vol. 18, pp. 369–394, 1999.
- [22] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*. Prentice-Hall, 1994.
- [23] S. Bittanti, *Identificazione dei modelli e controllo adattativo*. Pitagora Editrice, Bologna, 2000.
- [24] G. Kreisselmeier, “Stabilized least-squares type adaptive identifiers,” *IEEE Transactions on Automatic Control*, vol. 35, no. 3, pp. 306–310, 1990.
- [25] S. Gunnarsson, “Combining tracking and regularization in recursive least squares identification,” *Proceedings of the 35th Conference on Decision and Control, Kobe, Japan*, pp. 2551–2552, 1996.
- [26] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME, Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.
- [27] B. Quine, J. Uhlmann, and H. Durrant-Whyte, “Implicit jacobians for linearised state estimation in nonlinear systems,” *Proceedings of the American Control Conference*, pp. 1645–1646, June 1995.
- [28] M. Norgaard, O. Ravn, N. Poulsen, and L. Hansen, *Neural Networks for Modelling and Control of Dynamic Systems*. Springer, 2000.
- [29] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer, 2006.
- [30] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, 1974.
- [31] G. Schwarz, “Estimating the dimension of a model,” *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [32] D. J. Mackay, “Introduction to Gaussian processes,” *Proceedings of Neural Networks and Machine Learning, Cambridge, UK*, pp. 133–165, 1997.
- [33] D. J. Leith, M. Heidl, and J. Ringwood, “Gaussian process prior models for electrical load forecasting,” *Proceedings of the 8th International Conference on Probabilistic Methods Applied to Power Systems, Ames, Iowa*, 2004.