AN ABSTRACT OF THE THESIS OF

Alice Gillespie for the degree of Master of Science in Civil Engineering presented on December 17, 2015.

Title: Wave-by-wave Forecasting of Sea Surface Elevation for WEC Applications Utilizing NARX Neural Networks.

Abstract approved:

_____

Merrick C. Haller

Forecasting of ocean waves over a short duration on the order of tens of seconds was approached with the optimization of wave energy conversion in mind. This study outlines the development of an artificial neural network model, specifically the Nonlinear Autoregressive Network with Exogenous Input (NARX), to predict a wave-by-wave surface elevation time series based entirely on previous observations at the site of interest. Such a model would be computationally less intensive than competing deterministic techniques rooted in wave theory. Furthermore, it could potentially fit irregular patterns without some predetermined function as would be necessary for many other stochastic approaches. In principle, neural networks can be trained to learn previous patterns based on a recent wave record and then utilized in a feedback mode to yield multistep predictions for perhaps up to three wave periods (~45 seconds). The challenge of this approach is error accumulation in the intermediary steps that can lead to poor performance for longer prediction horizons. It was hypothesized that filtering the wave record input via wavelet or Fourier transformations could enhance model performance and hence, explorations of these types of input preprocessing were an integral part of the study. The NARX architecture allowed for an exogenous input time series that could conceivably mitigate error accumulation by providing an additional degree of signal correlation. Accordingly, the investigation also included potential exogenous inputs

that could be derived from the original wave record. The work culminated with a band-pass exogenous input delivering a significant forecasting advantage. Yet, providing the zero-phase narrow-band signal posed a challenge when applied in real-time, without the use of future data. A two-prong tactical approach was undertaken to address this challenge but ultimately proved insufficient. Consequently, the success of the NARX wave-by-wave forecasting method under the conditions of a real-world application will depend upon a better solution to the zero-phase filtering challenge.

©Copyright by Alice Gillespie
December 17, 2015
All Rights Reserved

Wave-by-wave Forecasting of Sea Surface Elevation for WEC Applications Utilizing NARX Neural Networks

by
Alice Gillespie

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented December 17, 2015
Commencement June 2016

Master of Science thesis of Alice Gillespie presented on December 17, 2015

APPROVED:

Major Professor, representing Civil Engineering

Head of the School of Civil and Construction Engineering

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Alice Gillespie, Author

ACKNOWLEDGEMENTS

I express sincere appreciation to Dr. Haller for his advice on this project. I also would like to extend a round of thanks to committee members Dr. Istok, Dr. Batten, and Dr. Sharp for generously reviewing this work. It must be said that my time at OSU has been on the whole an enjoyable and productive experience. The faculty has been excellent and I have honestly appreciated each and every class. My fellow students have been especially impressive in their talents and kindness. Thank you in no small measure to my very wonderful husband for supporting my engineering efforts. And my warmest "thank you" is extended to my three daughters, who embraced the presence of their mother at *their* university.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

1 Introduction

At present, there is an interest in forecasting ocean waves in the very near term, as it is a particular requirement of many wave energy systems under development. An efficient conversion of the energy in waves to electrical power is fundamental to the viability of ocean waves as an economical source of renewable energy. A forecast of impending waves would facilitate the tuning of wave energy systems to optimize power conversion.

The aim of this study was to develop a forecasting model utilizing an artificial neural network (ANN) that could reliably predict sea elevation data for wave energy applications. Such a model would be capable of obtaining local forecasts based on a single wave record. In this study, special attention was given to the preprocessing of input. It is hypothesized that forecast accuracy could be improved by modifying the wave elevation data to be more suitable to an ANN multi-step model. Several preprocessing techniques were investigated including filtering in the frequency and time domains, and wavelet transformations.

2 Literature Review

The following review is in support of the development of a wave-by-wave forecasting model utilizing artificial neural networks (ANN) that can be applied in real-time for ocean energy applications. For the purposes of this investigation, the forecast is based solely on a single wave record at the site of interest.

This effort crosses into several different disciplines, some with unique terminology. Consequently this review contains some general discussion with regard to wave energy conversion, stochastic processes, artificial neural networks, filtering, and wavelet analysis to provide context.

2.1 Wave Energy Converters (WEC)

Commercial-scale wave power operation and the industry as a whole are decades behind the wind and solar industries. Yet, wave energy has many auspicious aspects and is building momentum. Nascent projects are worldwide in areas including Portugal, Australia, Scotland, and in the United States. The multiple efforts have lead to a variety of wave energy converters (WEC) to suit differing circumstances. Some WECs generate electricity locally and transmit via undersea cables. Others harvest the mechanical energy from the waves to be converted to electricity onshore.

There are differing methodologies of wave energy conversion, as well. The oscillating water column (OWC) is a device that operates on the principle of air compression and decompression. The attenuator is usually a long device, oriented perpendicular to the wave front and uses both vertical and horizontal wave motion to drive hydraulic fluid through pumps. The overtopping device diverts water into a basin to create hydraulic head. The point absorber is a relatively small buoy that harnesses the vertical motion of the waves. And another, the submerged pressure differential device, sits on the ocean floor and converts the pressure fluctuations of wave motion into electricity. (Brekken et al., 2009; Drew et al., 2009) Although multiple design types allow for a greater number of potentially suitable wave energy

sites, one disadvantage facing the wave industry is that there is no universally preferred WEC design to focus efforts.

2.1.1 WEC System Controllers

Design-specific system controllers enhance the conversion of wave energy into electrical energy. System controllers are no small part of WEC development, as they are integral to device safety, operational efficiency, and profitability. To illustrate some control strategies, consider the point absorber variety of WEC, for which the system oscillations can be tuned to the incident wave to maximize energy absorption. One way of tuning, linear monochromatic control, adjusts the resonant frequency of the system to match the dominant wave frequency. Assuming operation in one degree of freedom (vertical motion), the maximum power capture is limited to 50% of the incoming wave power. (Brekken et al., 2009) Alternatively, the system oscillations can be stalled at the extremes of movement and released when the waves are in phase to maximize energy extraction. (Drew et al., 2009)

The first is an example of reactive control. The power takeoff system (PTO) is utilizing feedback and reverses power flow during parts of the oscillation cycle to better match the dominant frequency. The second method described (latching) is an example of a resistive control, a suboptimal method to approximate phase alignment without reversing power flow through the PTO machinery. (Falnes & Hals, 2012) A third strategy, model predictive control (MPC), is an example of advanced reactive control. It accounts for constraints on system forces, velocities, and position in the context of desired future velocities and positions. (Brekken et al., 2009; Brekken, 2011)

2.1.2 WEC Forecast Requirements

Generally, optimal and many suboptimal control strategies require some early knowledge of the next impending wave. (Fusco & Ringwood, 2012; Hong et al., 2014; Korde, 1999; Schoen & Hals, 2011; Reikard, 2013) Quantifying the length of forecast necessary is subject to electric control strategy, WEC design, the power

takeoff system, and of the ocean site. Fusco & Ringwood (2012), with a focus on heaving WECs and reactive-based controller systems, identified typical lead times from 30 to 60 seconds (2 - 4 peak periods) depending on the device size. In this study, the focus was on forecasting wave excitation force, which is obtained from wave elevation measurements via a system transfer function set at optimal gain. By lowering the system gain in accordance with the excitation force spectral distribution, the prediction horizon requirement could be reduced to perhaps 1 peak period in many cases, although with a cost to overall performances.

Further, the existence of a critical horizon was discussed as the horizon at which no new wave-by-wave data is helpful to system performance. The critical horizon was deemed dependent on the particular WEC situation and likewise, difficult to quantify in a general sense. Although not explicitly stated within the study, it could be inferred that improvements in power capture asymptotically diminish with given data ~1 peak period beyond an established beneficial forecast. (Fusco & Ringwood, 2012)

In the case of a point absorber type WEC employing bang-bang control (feed-back control that abruptly switches between two states) with both state and input constraints, a two-fold increase in energy output is achieved using sea elevation predictions. (Li et al., 2012) The results were based on simulations, but did utilize real sea wave data gathered off the coast of Cornwall, UK. Surface elevation forecasts were achieved via deterministic sea wave prediction (DSWP). The key finding in this study was that with regard to this control method, no more than 1 second prediction horizon was required to achieve optimal benefit, assuming point measurements could be made locally. However, it was argued that in this case, it would not be possible to acquire local data due to the interference created by the WEC motion. Therefore, point measurements must be made at a distance from the WEC and the corresponding prediction horizons were recommended to be on the order of ten seconds to allow for the propagation delay.

There are attempts to obviate the reliance on forecasting all together. Fusco & Ringwood (2013) propose a simple suboptimal, but robust control that promises to achieve levels of power capture nearly as good as achieved via MPC in most sea

states without the need of predictions. This alternative is specific to bottom-reference oscillating WECs, and relies on a novel method of tuning the magnitudes of velocity and excitation force in real-time.

Clearly, the performance requirement of real-time forecasting is tied to the state of WEC control. Success at optimizing power conversion will depend jointly on improvements in both WEC controller systems and wave-by-wave forecasting. Yet, attempts to sidestep the forecasting requirement signify a frustration with of the lack of a viable forecasting solution, thus far. Of the small number of documented attempts focused on wave-by-wave forecasting, success has remained elusive. The ability to accurately extend the forecast horizons to 3-4 peak wave periods would be valuable, allowing for greater flexibility in controller design. The various methodologies explored in recent literature are discussed below and include both deterministic (theory based) and stochastic (statistics based) approaches.

2.2 Deterministic Forecasting and Remote Sensing

When predicting the near-term wave evolution, a deterministic model is the more conventional choice. It is always desirable to model an environmental system through the governing laws of nature. However in this occasion, a solution based on physics necessitates both multiple sea surface measurements to detect directional variability and mathematical simplifications for complex wave theory. Many solutions have been generally imprecise (phase-averaged) and/or computationally intensive. One approach addresses these constraints by limiting forecasting to quiescent intervals. This has proved useful for maritime tasks such as cargo transfers or recovery/deployment of small watercraft. (Belmont et al., 2014) Other noteworthy deterministic efforts are described in the following paragraphs.

An alternative to phase-averaged spectral models was proposed by Blondel et al. (2008). The authors utilized a High Order Spectral (HOS) method to predict phase-resolved simulations of nonlinear wave dynamics. To begin, the initial wave field was reconstructed from wave elevation measurements via a multi-level iterative process. Analytic calculations for first and second order wave components and HOS simulation for higher order components were utilized in the reconstruction. Once the

initial wave field was obtained, the forecast was made either via an analytic wave solution based on the initial wave components or a HOS simulation using the reconstructed wave field as the initial condition.

In another study, Halliday et al. (2011) applied the fast Fourier transform (FFT) to simulated data to deconstruct the time series into a harmonically related wave vector of real and imaginary components yielding wave amplitudes and phase values. The forecast was realized by using this harmonic representation together with the following dispersion relationship

$$\omega_n^2 = g k_n \tanh(k_n h) \tag{2-1}$$

that relates angular frequency, ω_n, to wave number, k_n. The other parameters are gravitational acceleration, g, and depth, h.

Agreement was poor with regard to forecasting in time. The reconstruction imposed an artificial period onto to the generally non-periodic ocean wave history, based on the record length of the FFT. Better correlation was found by instead forecasting in both space (50 meters ahead) and time (30 seconds into the future). However, the reconstructed wave profile was less accurate at early times since this portion of the forecast included the slower traveling higher frequency content that had not yet in reality reached its destination.

Other notable advances take advantage of remote sensing marine radar technology. Nautical X-band radar of phase-resolved incoming waves may fulfill the spatial input needs of deterministic prediction models. Typically, X-band radar has been used in conjunction with spectral 3D FFT methods to retrieve statistical wave parameters such as mean wave period and wave direction. Lately, there have been strides in relating the X-band radar images to wave elevations. (Gangeskar, 2014 and Wijaya & Van Groesen, submitted) These methods are an improvement over Borge et al. (2004) in that they do not require any reference measurements.

Preliminary work in the reconstruction of sea surface elevation from sequences of synthetic radar observations is detailed in Wijaya, Naaijen, Andonowati, & Groesen, 2015. Furthermore, these synthetic reconstructions were transformed into

the future sea surface elevations with a correlation above 0.9. In this case, Dynamic Averaging-Evolution Scenario (DAES) was employed to reconstruct regions of "shadowed" sea elevation due to imperfect radar imagery. This combined with a linear (dispersive) evolution operator determined the future evolution of the waves. These results are promising, but require further development to realize a potential for real sea applications.

2.3 Stochastic Forecasting

In the case of near-term forecasting, stochastic models are an attractive alternative to deterministic models, since they are the only way of obtaining local forecasts with single point measurements. They are also quick in application and can be relatively straightforward to implement. Speed in application is of particular value to successful and profitable WEC operation, since WEC controllers require forecast updates in real-time. Wave-by-wave forecasting via stochastic models is an on-going effort, as is demonstrated a growing presence in scientific literature that are described in 2.4.1. Prior to discussing such work and the stochastic time series models used for wave-by-wave forecasting, it is useful to start with a brief overview of stochastic analysis.

2.3.1 Stochastic Analysis and Bulk Parameters

As the counterpart to a deterministic system, a stochastic system is characterized by a random pattern that cannot be known precisely. Stochastic ocean-wave forecasting employs probabilistic methods based on wave records. And as a stochastic model predicts a set of possible outcomes weighted by probability, it requires little knowledge of the physical processes that govern the evolution of the ocean surface. Stochastic approaches have been widely employed for decades to analyze the changes of bulk parameters such as significant wave height or peak wave period, or to estimate the likelihood rogue waves. Time lengths typically on the order of seasons, days, or hours are considered because of the imprecise nature of statistics. (Ochi, 2005)

Stochastic analysis of random ocean waves spans the domains of time, frequency, and probability. Statistical analysis from random observation in the time domain has roots in early oceanography. Careful readings of wave records and applying the appropriate probability functions can yield reliable estimates of bulk parameters, if the wave records are long enough. The auto-correlation function in the time domain provides the time-average of wave energy. But more importantly, the autocorrelation function is a mathematical link to the frequency domain; the autocorrelation function and the spectral density function are Fourier transform pairs. Further, the wave spectrum provides a convenient bridge to the probability domain. And it is spectral analysis in the frequency domain combined with the appropriate probability functions that allow for more rigorous probabilistic predictions of bulk properties. (Ochi, 2005)

2.3.2 Time Series Forecasting Models

Stochastic analysis is not limited to bulk parameters and systems need not be purely random to merit stochastic-based forecasting. Forecasting via time series models is generally treated as a stochastic difference problem that involves both deterministic and stochastic processes.

Ocean waves in deep water are considered to be a Gaussian random process in that they are normally distributed with zero-mean and constant variance. All Gaussian time series are considered linear and as such, the best forecast approach is a linear model. Examples of linear time series models are autoregressive (AR), autoregressive with exogenous input (ARX), and autoregressive moving-average (ARMA). These forecasts are linear combinations of past and present values of the time series and are quick and relatively straightforward, but are limited in capability to capture the more irregular aspects inherent in the sea.

In ocean depths less than half a wavelength, waves are often non-Gaussian and may be nonlinear. In this case it may be beneficial to consider a nonlinear forecasting approach. Nonlinear time series models in the form of artificial neural networks (ANN) bring in complexities, but also offer potential for greater accuracy. Nonlinear autoregressive (NAR) and nonlinear autoregressive with exogenous

(external) input (NARX) are two types of ANN specifically suited for nonlinear time series forecasting. The capabilities of NARX to capture irregular patterns in chaotic time series analysis are documented in Kim et al. (1999). The ability of NARX to track longer-term variability was demonstrated by Jain et al. (2011), based on simulated noisy inputs.

2.4 Forecast Error Metrics

Two established values to assess the performance of short-term forecasts are the root relative error, RRE and the correlation coefficient, R. They are defined below and utilized in previous work on a scaled test model of an oscillating water column, for short-term forecasts of airflow and WEC motions. (Sheng & Lewis, 2011)

$$RRE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{2-2}$$

$$R = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{2-3}$$

The predicted and measured values are represented by x and y, respectively.

The RRE value is essentially the root mean square error relative to the root mean square amplitude. It provides a relative measure of the difference in amplitude between the predicted and measured values. Small RRE values are desireable and RRE = 0 indicates perfect agreement. The correlation coefficient R indicates the phase relationship between the predicted and measured values. R values range from 0 to 1, where R = 1 indicates perfect agreement.

The coefficient of efficiency *CE* and the goodness-of-fit index *F(l)* are two metrics also cited in relevant literature. They are the same metric in different form, and for each forecasting horizon *l,* they are related to RRE as follows.

$$CE(l) = 1 - RRE(l)^2 \tag{2-4}$$

$$F(l) = (1 - RRE(l)^2) \cdot 100 \qquad (2\text{-}5)$$

CE may range from -∞ to 1, with values close to 1 indicating better agreement between observed and forecasted values. *F(l)* is *CE* in percentage form, and a value of 100% would indicate perfect agreement.

The combination of the correlation coefficient (*R*) and the coefficient of efficiency (*CE*) gives a rigorous estimate of overall forecast performance. Figure 2.1 provides an illustration of how these values relate to differences in amplitude and phase. The errors were computed using two monochromatic signals of identical frequency. It is clear from the left side of Figure 2.1, that the correlation factor does not measure differences in amplitude. A 100% difference in amplitude is however indicated by a value of 0 with respect to the coefficient of efficiency. From the graph on the right, phase errors are reflected in both error metrics. A phase difference of ~60 degrees corresponds to a coefficient of efficiency of 0 and a correlation factor of 0.5.

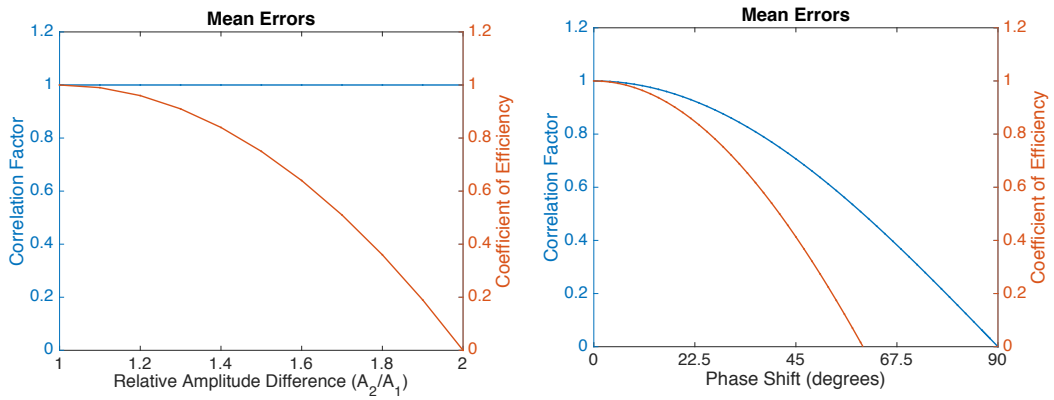


Figure 2.1. Monochromatic example of the effect of amplitude difference (left) and phase shift (right) on the correlation factor and coefficient of efficiency.

2.4.1 Stochastic Time Series Model Comparisons

The following review of previous work regarding wave-by-wave forecasting underscores the small number of documented efforts. First, the work by Fusco et al. (2010) is discussed, which offers a good overview of the potential performance of

many stochastic contenders. Other studies addressed thereafter include Boren et al. (2014), Shoori, et al. (2014), Paparella et al. (2015), and Schoen et al. (2011).

The principal stochastic approaches were surveyed by Fusco et al., (2010) using measured data from two locations. One set was from the West Coast of Ireland and generally represented a narrow-banded sea state (sampling frequency of 2.56 Hz and a depth of 20m). The other was from the Azores Archipelago, and generally represented a broader-banded sea state (sampling frequency of 1.28 Hz and a depth of 40m). As an additional note, input data was low-pass filtered prior to forecasting, which would present a challenge in real-time implementation. See 2.6 for further discussion on input preprocessing.

The major approaches considered are listed below. The prediction accuracy was measured with the goodness-of-fit index *F(l)* for each forecasting horizon *l*, which is defined the preceding section. Note that a 100% value for *F(l)* is a perfect *l*-step ahead prediction.

- Cyclical models are noted as the most straightforward. Wave elevation is expressed as a superposition of linear harmonic components so that the forecast is effectively achieved from a multivariate linear regression onto these fixed harmonic frequencies. The choice of frequencies is crucial and problematic.

Broad-banded: $F(l = 10s) \cong 30\%$

Narrow-banded: $F(l = 12s) \cong 5 - 29\%$

- Sinusoidal extrapolation with the extended Kalman filter proved more challenging, yet offered little actual improvement over the cylical models. The wave elevation was treated as a single cyclical component with a time varying frequency.

Broad-banded: $F(l = 10s) \cong 5\%$

Narrow-banded: $F(l = 12s) \cong 24\%$

- The AR model provided forecasts of 1 to 1.5 peak wave periods and maintained accuracy for more than 2 hours.

Broad-banded: $F(l = 10s) \cong 95\%$

Narrow-banded: $F(l = 12s) \cong 94\%$

- The extended Kalman filter allowed the nonlinear prediction equations to be approximated by a linearized form. Extending the model to multiple cyclical components of variable frequency is under study.

Broad-banded: $F(l = 5s) \cong 5\%$

Narrow-banded: $F(l = 12s) \cong 24\%$

- ANNs were unfavored in part since the models are completely enclosed in a black box and do not allow inspection of inherent characteristics to assess process behaviors. Preliminary attempts found that ANNs did not generate the superior results necessary to justify the added model complexities.

Broad-banded: $F(l = 10s) \cong 68 - 84\%$

Narrow-banded: $F(l = 12s) \cong 80 - 93\%$

Of all of the approaches considered by the authors, AR was deemed the most capable. However, it was cautioned that AR performance was still not adequate to meet most current WEC controller needs.

Beyond the above survey, there are a few other investigations of stochastic wave-by-wave forecasting. Boren et al. (2014) conducted an informal feasibility study using an ANN for wave-by-wave forecasting. Results were optimistic enough to lead to a second, more detailed effort. In this follow-up study, Shoori et al. (2014) presented two methods of training a neural network to forecast wave elevations. The data was from the Western Coast of Ireland at the Belmullet Wave Energy Test Site (sample rate of 1.28 HZ). The two models shared similar FFNN architecture, but were trained on different datasets. The first dataset contained the measured wave elevation. The second contained wave elevations reconstructed from the spectral analysis of the first dataset. The primary aim was to assess the scalability and

generality of forecasting, and the secondary aim was to assess the capability of training from spectral data, which are widely available. Only single-step forecast horizons (0.78 s) were evaluated.

Both methods of training yielded predictive capability with correlation coefficients (R, as defined in 2.4) of 0.761 for the reconstructed data and 0.845 for the measured model, averaged over all considered sea states. However, the model trained on reconstructed data did not scale nearly as well to a significantly different sea state as the one trained on measured data.

In a study specific to an oscillating water column (OWC), Paparella et al. (2015) used WEC up-wave input to aid in the prediction of wave profiles within the OSC chamber. Four forecasting methods were investigated. An AR method was used first and served as the baseline approach. The second method was a finite impulse response model (FIR) that assumed the chamber wave elevation is a linear combination of past up-wave measurements alone. The third and fourth methods were ARX and NARX, where the primary input was the past values of chamber elevations and the up-wave input served as the exogenous component.

Performance was gauged using the goodness-of-fit (as defined in 2.4) relative to the AR accuracy. The FIR model demonstrated far inferior performance relative to AR for all prediction horizons. The ARX model demonstrated similar accuracy for shorter horizons, but reduced accuracy at horizons longer than 15 seconds. NARX offered no benefit over ARX, as the sea state was dominated by swell conditions and the surface within the chamber was generally linear. In conclusion, the use of up-wave exogenous input on the whole was deemed not useful.

Some other stochastic examples of wave-by-wave forecasting can be found as integral parts of various controller strategies. For example, Schoen et al. (2011) focused primarily on WEC controller development, but does address the forecasting within. The authors proposed a hybrid Kautz/AR model, but also considered Kautz alone, AR, and ARMA to address forecasting needs of the fuzzy logic/robust controller that required only very short forecast horizons (2 s).

2.4.2 Related Stochastic Forecasting Efforts

There are a number of other forecasting studies for different but related applications. Sheng & Lewis (2011) used an ANN to predict airflow and motions of an oscillating water column WEC using a scaled test model in a wave tank. Correlation R and root relative error RRE were approximately 0.7 and 0.8, respectively for 2 second forecasts of column water height. These values corresponded to a sea state with a peak wave period of 10 s and a wave height of 5m. The model performance was deemed promising but still inadequate for practical device control.

Ling (2015) proposed the use of the WEC itself as both a measurement and prediction device via software and motion sensors. Specifically, the WEC motions were used to estimate excitation force by way of the Kalman filter. Predictions of future excitation forces were then obtained with an AR model. Heave, surge and pitch for horizons of up to 20 seconds were predicted, but good performance was generally limited to 5-6 seconds, beyond which correlation coefficients were well below 0.5 for wave excitation forces. Prediction accuracy was closely correlated with estimation accuracy and was better for high-energy swell dominated sea states as opposed to low-energy wind dominated sea states.

Notable success was found regarding multi-step forecasting of wave power. (Hatalis et al., 2014) It should be stated that forecasting power output is fundamentally different from forecating wave elevation in that the power time series are non-sinusoidal with hourly time scales. However, the analysis had some parallels, considering the multi-step NAR method. This study utilized data collected from Nov. 2011 to Dec. 2011 at Belmullet Berth B, a high wave potential region off the coast of Ireland. Multi-step forecasts were conducted using an ANN nonlinear autogregressive (NAR) model and yielded correlations of $R = 0.98$ at 24 hours into the future. In this case, success was only achievable once the input data was smoothed via an exponential moving average, highlighting the importance of the preprocessing of input. Further discussion of input preprocessing is given in Section 2.6.

There are two other nonlinear stochastic studies mentioned for completeness. The first employed ANN, genetic programming (GP) and model tree (MT) for hourly forecasting of significant wave height based on wind time histories. (Jain et al., 2011) All three models demonstrated similar forecast performances. The second compared several stochastic models against the physics-based wave prediction model WAM for the short-term forecasting of hourly wave energy flux (Ibarra-Berastegi et al., 2015). The random forests (RF) method was found to outperform both the empirical orthogonal functions (extended EOFs) and the analogues stochastic methods, and in certain conditions RF also out-performed WAM.

2.4.3 Stochastic versus Deterministic Forecasting

ANN and AR models were assessed relative to Wavewatch III, a physics-based prediction model, for short term forecasting of simulated power output for five types of WECs. (Reikard, 2013) Again note that a wave power time series differs in shape and time scale from a wave elevation time series. The accuracy associated with the stochastic models was highly dependent on horizon size while the forecast accuracy of the physics-based model was less so. Relatively speaking, the stochastic models showed better performance for shorter forecast horizons (less than 8 hours), the physics-based model faired better on longer forecast horizons (more than 11 hours), and both types demonstrated similar performance in the mid horizon range (8-11 hours).

2.5 Artificial Neural Networks (ANN) as Time Series Models

As noted earlier, ANNs (Artificial Neural Networks) are a viable contender in wave-by-wave forecasting as they have the potential to “learn” model complexities of an irregular time series and return results with speed. They can shortcut the design effort in that they do not depend on a physical understanding of the process at hand. Yet, ANN modeling comes with its own set of challenges. The process of optimizing the learning algorithm requires familiarity with ANN techniques. This is usually done by hand by testing out different algorithm settings and assessing the resulting

performance. A poorly designed ANN is unstable and inaccurate and can lead to the premature conclusion that any ANN approach is unsuitable. (Deo, 2010) Fortunately, some recent inroads toward automating this time consuming process for ANN models bode well for the future. Flores et al. (2012) demonstrate an optimization process, Evolutionary Design of Forecasting Models (EDFM), that successfully identifies both the specific architecture and the weights of the ANN with no human intervention.

2.5.1 ANN Architectures

A common approach for time series analysis is a feedforward neural network (FFNN) trained with backpropagation. These networks are relatively easy to use and converge quickly. One drawback of FFNNs is the lack of short-term memory in that they repond similarly to a given input, regardless of former inputs. Time delay neural networks (TDNN) are FFNNs with delayed input and can be viewed as autoregressive models. Recurrent neural networks (RNN) are used in similar applications as TDNNs, but have the advantage that the neurons send feedback signals to each other giving it a short term memory. Nonlinear autoregressive (NAR) and nonlinear autoregressive with exogenous output (NARX) are examples of RNNs with time delayed input. NARX models are typically more robust than TDNNs, but require more processing time. (Hulthén, 2004; Dorffner, 1996; Kondratenko & Kuperin, 2003)

2.5.2 Training Algorithms

ANN training algorithms fall into two basic categories according to the type of optimization operator: Deterministic and Stochastic. (Masters, 1995; Prudêncio, 2003) Deterministic operators are fast, but are sensitive to fall into local minima. The most common training algorithms use determinstic search operators and include back propagating, scaled conjugate gradient method and other gradient-based algorithms. Stochastic algoritms (or evolutionary algorithms) perform global searches according to probabilistic criteria. Stochastic algorithms are less vulnerable

to local minima, but can be slow to converge. Genetic Algorithms (GA) and Simulated Annealing (SA) fall in this category.

There has been some investigations into combining algorithm types to achieve a desireable mix of qualities. Prudencio (2003) utilized hybrid learning by combining GA and the gradient-based Levenberg-Marquardt (LM) to yield lower prediction errors in shorter executions in preditions of monthy river flow using a time delay neural network (TDNN). Another example of hybrid learning is given by Salami et al. (2013) for NARX using differential evolution and GA for the purpose of identification (as opposed to time series forecast).

An incomplete list of optimal state estimators is given below. (Simon, 2006; Wilamowski, 2003)

Linear Systems

- Least Squares Estimation: The estimation of a constant vector on the basis of linear but noisy measurements. Least squares serves as the basis of many optimal state estimators.
- Kalman filters (KF): the optimal state estimator for unconstrained, linear systems subject to Gaussian noise.

Nonlinear Systems

- Extended Kalman filters (EKF): Nonlinear state estimator that relies on linearization to propagate the mean and covariance state. The unscented Kalman filter (UKF) is a modified version of EKF with reduced linearization errors.
- Moving-horizon Estimation (MHE): A constrained, nonlinear optimization that reduces to the Kalman filter for unconstrained, linear systems. Moving-horizon estimation proved better capable than EKE to handle multiple optima and is posited as a feasible alternative to the EKE in the context of chemical reaction networks. (Haseltine & Rawlings, 2005)
- Particle filter: A completely nonlinear state estimator at the price of increased computational effort.

- Levenberg-Marquardt optimization: Second-order gradient based technique and good choice for many time series problems. (A stock Matlab function for nonlinear regression)
- Bayesian Regulation: a form of Levenberg-Marquardt that minimizes a linear combination of squared errors and weights. Preferred for noisy data, but requires longer processing. (A stock Matlab function for nonlinear regression)
- Scaled conjugate gradient method: Utilizes gradient calculations that are more memory efficient than the Jacobian calculations used in Levenberg-Marquardt. (A stock Matlab function for nonlinear regression)
- Resilient back propagation: Eliminates the effects of harmful derivative functions. Relies only on the sign of the derivative (not the magnitude) to determine weight updates. It is generally faster than the standard steepest descent algorithm, in exchange for a modest increase in memory requirements. (A stock Matlab function for nonlinear regression)

2.5.3 ANN Forecasting in Related Ocean Applications

Most previous investigations of ANNs with respect to ocean waves have targeted the evolution of bulk parameters such as significant wave height (H_s) and peak wave period (T_p) on an hourly time scale. (Balas et al., 2004; Deo et al., 2001; Deo & Naidu, 1998; Makarynskyy et al., 2005; Mandal & Prabaharan, 2006)

Despite the large time scales and non-sinusoidal patterns, recommendations based on these efforts can be related to the current application. In particular, refining forecasts by using ANNs in series is suggested in Deo, 2010. The ability of NARX to retain information 2-3 times as long, with respect to conventional recurrent networks, is concluded in Lin et al., 1998. Khalil, 2011 provides a good reference for the inner-workings of the NARX architecture. The impact of exogenous input on the overall performance of NARX (and ARX) for the application of a reactor-exchanger is investigated in Chetouani, (2008). Further issues particular to many ANN applications such as over-fitting, architecture bottlenecks, and qualifying input are documented in similar studies. (Huang et al., 2003; Kim et al., 1999; Londhe & Deo, 2004; Mandal & Prabaharan, 2010)

2.6 Input and Preprocessing

Successful implementation of autoregressive methods such as AR, ARX, and NARX, typically requires the preprocessing of input. The removal of the time series mean, and the removal of linear or nonlinear trends often accomplished by differencing, are important prior to forecast. These processes can be reversed post forecast. Smoothing of input is also recommended, which can be done via filtering in the time or frequency domain. (Dorffner, 1996) Time series applications that demonstrate the effectiveness of input preprocessing include the use of a moving average for forecasting foreign exchange rates (Kondratenko & Kuperin, 2003), an exponential moving average for ocean power production forecasts (Hatalis et al., 2014), and wavelet transforms for ocean wave height forecasting (Dixit et al., 2015). While these time series differ from wave elevation histories in signal profile and time step resolution, the commonalities within the prediction model architectures suggest the preprocessing techniques offer potential benefits for wave-by-wave forecasting.

2.6.1 Signal Filtering

Stochastic forecasting methods exploit signal correlations, and are thus limited by the intrinsic correlation of the signal. And since the intrinsic correlation is in general proportional to the signal bandwidth, the ability to forecast waves is highly related to sea state. (Fusco & Ringwood, 2013) Narrow-banded signals are more highly correlated than broad-banded signals. On this basis, there may be potential to improve forecast performance through filtering.

Filtering can be done in the time domain by averaging over adjacent data points or in the frequency domain by limiting high frequency components. Moving average and exponential moving average are two examples of time domain filtering. In the case of a moving average, data is averaged evenly over a window of a specified length. The smoothed signal is obtained by sliding the window forward step by step. This type of moving average would preserve the phase of the time series and thus, is a zero-phase filter. Filtering in the frequency domain is another example of a zero-phase filter. The there will not be a practical method of zero-phase filtering in real-

time without further research. (Fusco & Ringwood, 2010) The specifics of real-time filtering limitations of zero-phase filters will be addressed in 3.2.1.

In some studies, focus is shifted to an alternative context to minimize the need for filtering. Quiescent interval prediction of sea elevation is the target of Belmont et al., (2006) for the purposes of using DSWP for marine applications. Non-causal (zero-phase) finite impulse response filter is deemed suitable for the specific application due to the nature of the DSWP process. Some researchers focus their forecasting efforts on the effect of the wave elevation on the device such as the wave excitation force. Depending on the size of the device relative to the higher frequency wavelengths, the wave excitation force is inherently low-pass filtered to some extent, lessening the need for a zero-phase filter. (Brekken, 2011; Ling, 2015)

The exponential moving average weighs the windowed data unevenly, by favoring terms closer to the time of interest. It is not a zero-phase filter, but the phase shift is linear and can be largely corrected post forecast. The implementation of this filter is described in Hatalis et al. (2014), in which a fairly accurate method for short-term forecasting of wave power is developed for the primary goal of operational scheduling. The method utilized NAR and its success was due in large part to the preprocessing of input. The model forecasted exponentially smoothed wave power directly converted from measured SWH from the Irish Marine Institute. The authors improved on other power forecast attempts by using the NAR version of NARX (as opposed to other ANNs), predicting high resolution (ten-minute intervals) exponentially smoothed values of wave power instead of SWH, and forecasting multistep into the future instead of just one step into the future. Refer back to 2.4.2 for performance details. The preprocessing specifics are given below.

Prior to exponential smoothing, linear interpolation was applied to the data set to yield values at every minute for the purpose of correcting missing and irregular data points. From this set they then extracted 10-minute data points. The exponential smoothing formula is

$$s_{i+1} = \alpha y_i + (1-\alpha)s_{i-1} \tag{2-6}$$

The smoothed weighted average s_{i+1} is calculated from the previous observation y_i, and previous smoothed value, s_{i-1}. The smoothing factor α is given by the following expression.

$$\alpha = \frac{2}{(k+1)}, \quad 0 < \alpha < 1 \tag{2-7}$$

The window size is specified by k, which is taken to be 12 hours in this case. The result of the exponential smoothing is shown below. (Hatalis et al., 2014)

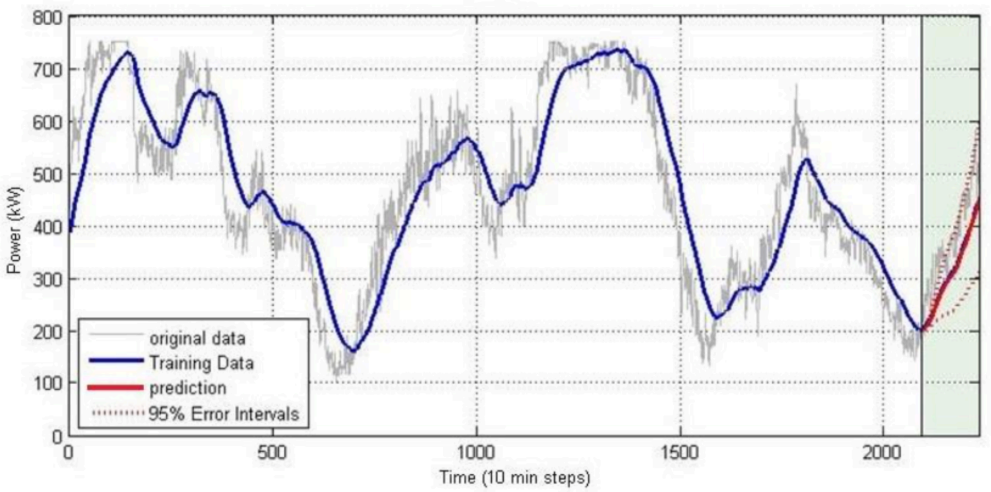


Figure 2.2. Exponential smoothing from (Hatalis et al., 2014).

2.6.2 Wavelet Analysis

Wavelet analysis is an extension of Fourier analysis. In Fourier analysis, a signal is approximated with a Fourier transform, which is a combination of infinite sinusoids of differing frequencies. In wavelet analysis, a signal is approximated with a wavelet transform, which is a combination of finite wavelets of similar shape, but differing in shift and scale. Where Fourier analysis is a windowing technique of the same time and frequency resolution, wavelet analysis is a windowing technique of varying time-frequency resolution. It allows the use of long time intervals for low frequency information and shorter time intervals for high frequency information.

Wavelet analysis is applicable to both stationary and non-stationary signals, as opposed to Fourier analysis, which requires stationarity. There is no artificial periodicity imposed on the wavelet transform since it is a combination of finite waveforms. These finite waveforms are all scaled and shifted versions of one mother wavelet. Time information is not lost with wavelet transforms as it is with Fourier transforms. It is capable of revealing points of discontinuities or certain aspects of trends.

However, the tracking of so much information comes with computational cost. This is addressed with discrete wavelet transforms (DWT). DWT shift and scale at discrete intervals, allow for greater efficiency with respect to continuous wavelet transforms (CWT) that operate at every scale. DWT can provide similar accuracy to CWT, depending on the choice mother wavelet and level of decomposition. There are many dictionaries of mother wavelets, to suit various applications. The mother wavelet is chosen to match the general characteristics of the time series under analysis. (Mallat, 2009; Misiti et al., 1996; Misiti et al., 2015)

There have been several recent accounts of ocean wave height forecasting in which wavelet transforms are applied in a preprocessing step. Three studies are described below. In all three, single-step forecasting models were used (as opposed to multi-step feedback models) based on the forecast horizons desired. In the first study, the wavelet-transformed input is used for a fuzzy logic model. In the latter two studies, the wavelet-transformed input is used for ANN models. Wavelet transforms are discussed in further detail in 3.2.2.

First study: Özger (2010) used a combination of wavelet and fuzzy logic (WFL) to achieve improvements in long-range (>24 h) forecasts of significant wave height and average wave period. Performance was evaluated against three other prediction methodologies: Autoregressive Moving Average (ARMA), Artificial Neural Network (ANN), and Fuzzy Logic (FL). Data, obtained from NOAA's National Data Buoy Center, was from 5 different wave rider buoys off the west coast of the United States.

The evaluations showed that the WFL model outperformed ARMA, ANN, and FL for all forecast horizons (3, 6, 12, 24, 48 h) and especially in higher forecast horizons (48 h). The following table lists CE values as defined in Section 2.4.

Table 2.1. First Study: Coefficient of Efficiency (Özger, 2010)

Horizon (h)	ARMA	ANN	FL	WFL
3	0.944	0.948	0.948	0.970
6	0.873	0.885	0.884	0.933
12	0.721	0.755	0.754	0.877
24	0.433	0.533	0.532	0.880
48	0.188	0.366	0.363	0.718

Second study: Deka & Prahlada (2012) utilized discrete wavelet transforms to decompose the wave height data for use as input for the feed forward neural network (FFNN). The data, obtained off the west coast of India by New Mangalore Port Trust (NMPT), was collected from 2004-2005. The Mallat algorithm with db4 type wavelet was used for the discrete wavelet transform (DWT). The FFNN was trained using Levenberg-Marquardt (LM) method to forecast future wave heights directly, without the need of an inverse transform. The hybrid model (WLNN) was evaluated with respect to the FFNN using raw data. The results confirmed that wavelet preprocessing improved forecasts, especially for longer lead times (forecasting horizons). The root mean squared errors (RMSE) of the two models are compared in the figure below.

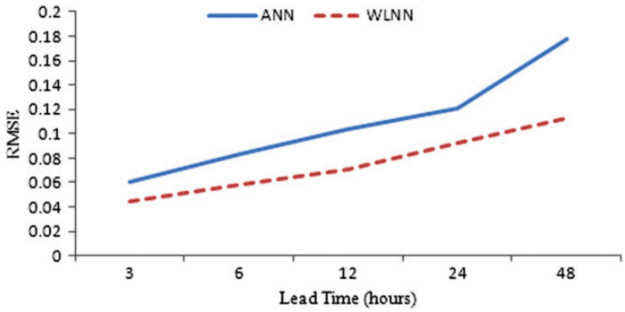


Figure 2.3. Variation of RMSE over lead time (Deka & Prahlada, 2012)

Third study: Removing prediction lags that plague ANN time series forecasting in general, and ANN wave height forecasting in particular, was the focus of the third study by Dixit et al. (2015). In this case, two separate ANNs were trained, one for approximate wavelet coefficients and the other for detail wavelet coefficients. The outputs of the networks were then transformed back to wave heights using an inverse discrete wavelet transform (iDWT).

The ANNs were of FFNN architecture and trained using LM, similar to the previous study. The data sets included seven years of hourly data (2006-2012) from three stations off the coast of Florida, obtained from NOAA. Correlation coefficients (R) as defined in 2.4 are listed in Table 2.2.

Table 2.2. Third Study: Correlation Coefficients of ANN and Hybrid Model MNWT (Dixit et al., 2015)

Horizon (h)	**Station 42040**		**Station 42039**		**Station 41004**	
	ANN	**MNWT**	**ANN**	**MNWT**	**ANN**	**MNWT**
12	0.94	0.98	0.80	0.83	0.73	0.78
24	0.86	0.98	0.60	0.63	0.50	0.51
48	0.78	0.94	0.40	0.42	0.30	0.37

Based on the above analyses, preprocessing the input via wavelet transforms has the potential to extend the prediction horizons of wave-by-wave forecasting. An

investigation into the practicality of implementing the wavelet transforms in a multi-step model is of particular interest.

2.6.3 Exogenous Input

Some forecasting models have the capability for exogenous input (external time series), which can greatly improve forecasting performance. Determining the appropriate exogenous input requires some trial and error. Excess input can compromise the forecast.

Traditionally, a second related time series would be used. For example in the case of a reactor-exchanger, the heat transfer fluid temperature was the exogenous input to aid the NARX forecast of oulet temperature of the reactor-exchanger. (Chetouani, 2008) In an example more relevant to the present goal, up-wave exogenous input was used (with little benefit) for a hybrid Kautz/AR model to forecast the water surface elevation within the oscillating water column. (Paparella et al., 2015) This analysis was previously discussed in greater detail in 2.4.1.

The present investigation is limited to a single time series, but it is conceivable to derive the exogenous input from the primary time series. With regard to foreign exchange rates, daily returns were used to suplement the primary input of a moving average of the daily returns. An exponential moving average was also considered and was found to yield similar results. (Kondratenko & Kuperin, 2003)

3 ANN Model Development

The forecasting method used in this analysis can be separated into three basic components: The preprocessing of input, the assimilation of data with time, and the forecasting of output. The overall success depends on whether the input could be preprocessed in a way that significantly aided forecasting, yet could be employed consistently in real-time. A major concern of integrating the components was that autoregressive techniques of the ANN time series model rely heavily on the last data points measured. Many preprocessing steps that are often beneficial in a hindcast sense potentially compromise the beginning and end of the data record, and thus, complicate performance in the real-time forecasting sense. These components are interdependent, but will be first addressed individually for clarity. The forecasting process was the core of the model and dictated the type of preprocessing required. For that reason, the description of the overall model begins with the NARX design and requirements.

3.1 Forecasting of Output

The forecasting method was based on the nonlinear autoregressive with exogenous inputs (NARX), a type of ANN particularly suited for time series prediction as opposed to other ANNs that may be better suited for classification, clustering, or pattern recognition. The NARX model is stated algebraically in Equation 3-1.

$$y_n = F(y_{n-1}, y_{n-2}, y_{n-3}, \ldots\ u_n, u_{n-1}, u_{n-2}, u_{n-3}, \ldots) + \epsilon_n\,, \qquad (3\text{-}1)$$

where y_n is a value of the time series of interest at time step n. F represents a nonlinear function of the present value of an exogenous time series u, as well as previous values of both y and u. A stochastic error term is given by ϵ.

Neural networks are model-less approximators in that they are capable of modeling without prior knowledge of the nature of the system being modeled. In the

case of the NARX model, input is regressed onto a nonlinear function approximated with a multi-layered perceptron (MLP). A perceptron is a mathematical model of a biological neuron. The MLP serves as the basis of supervised learning where the training input is mapped onto target outputs. NARX is termed a recurrent dynamic network with feedback. The model is recurrent in the sense that global feedback is applied backward and forward within the model. It is dynamic in the sense that the network continues to adapt during operation, as opposed to a static ANN. A static network learns during the training process and once trained, the model is fixed, or static. (Beale et al., 2015)

The NARX model, illustrated in Figure 3.1, is made up of three layers of nodes (or perceptrons): The input layer, the hidden layer, and the output layer. The first layer is the input layer, which acts as an embedded memory in the form of tapped time delays. Each node is connected to a value in the wave record. The number of nodes in the input layer (often termed input order) and delay spacing (tapped delays) can be tuned to suit the character of the input signal. The performance of the network is largely dependent on the choice of the input order and delays. Generally, larger spacing captures lower frequency patterns and tighter spacing captures higher frequency patterns. Insufficient delay spacing leads to redundancy of input, which is undesirable. Too many input nodes will destabilize the network. Typically, a feedback input order of 24 with consecutive delay spacing (15 points per peak period) was used in the Huntington Beach analysis, which was sampled once per second. Exogenous input order was smaller and often spaced by a factor of 2 or 3 (5 - 7 points per peak period), since the signal tended to be more regular. The tapped delay spacing may need to be increased in cases of higher data sample rates.

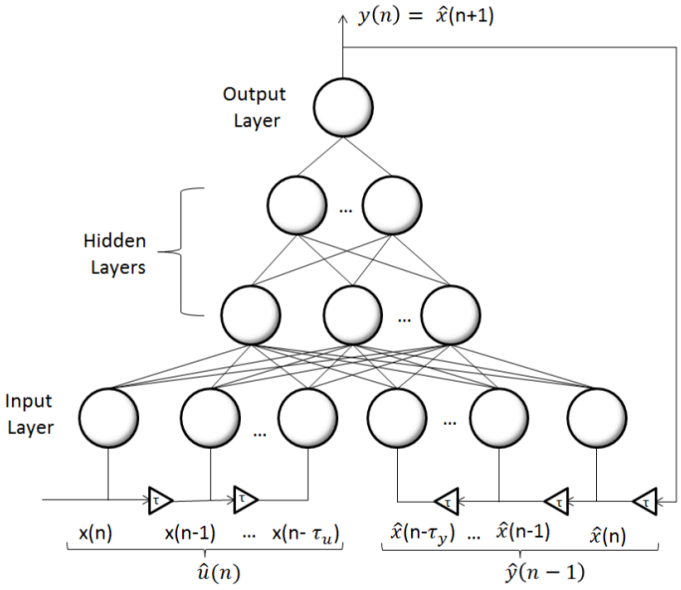


Figure 3.1. NARX architecture (Hatalis, 2014).

The second layer is referred to as the hidden layer. Too many nodes in this layer may lead to over-fitting, yielding a network that is not well-generalized for new input. Too few nodes may lead to under-fitting and reduced accuracy. However, the small variations in size of this layer were not as critical to overall performance as was the sizing of the input layer. The hidden layer contained 3 nodes in all cases of this analysis. Lastly, the output layer comprised of a single node since the only output is the primary time series (forecasted wave record).

3.1.1 Two-Step Forecasting Method

The NARX forecasting model has the capability to manage two input signals in predicting of the future sea surface profile (Eta). A flowchart illustrating the process is given in Figure 3.2. The feedback input is the signal to be forecasted (measured sea-surface elevation). If the measured data were particularly noisy, the feedback input may be smoothed in a preprocessing step. In that case, the network target would actually be the de-noised version of the measured data. Regardless, performance would be gauged against the expected raw measurements.

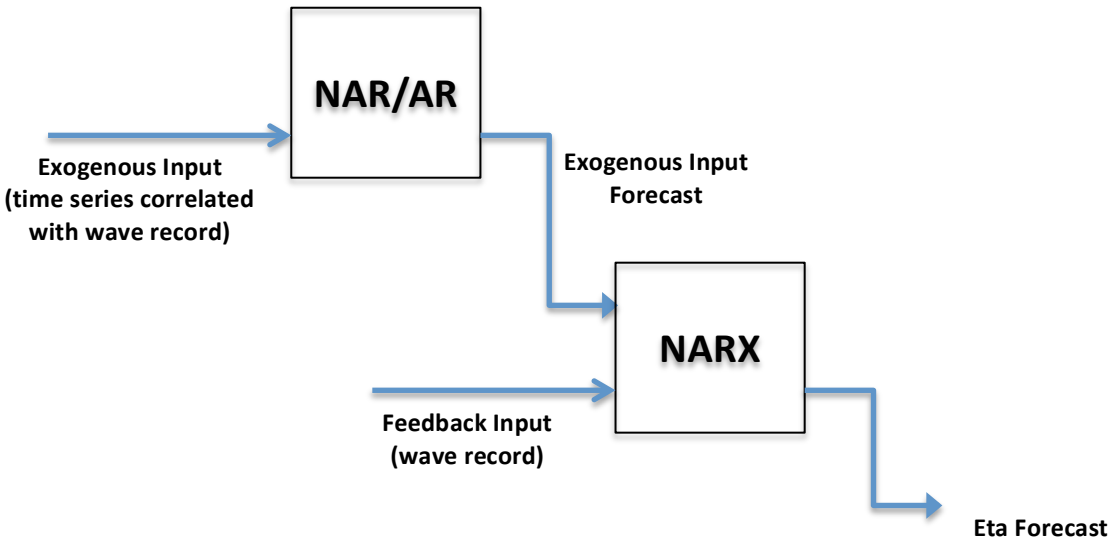


Figure 3.2. Process flow of the 2-step forecasting method.

If an exogenous input were available, Step 1 would be employed in order to forecast future values of this signal. The exogenous time series need not be of the same character as the primary feedback input, but it is important that it be correlated. Additionally, it must be predictable for simple linear AR (Auto-Regressive) or NAR routine. Next, the NARX model would be employed in Step 2 to forecast the wave elevation (Eta) based on both the primary feedback input and the forecasted exogenous input. In the case of no exogenous input, the wave elevation would be forecasted via the Step 2 network configured as NAR.

The model was designed to refresh multi-step forecasts as it progressed in time, producing a vector of forecasts associated with each newly measured data point. In the case of the Huntington Beach dataset, measured data was provided at every second. A forecast profile was computed for every future second up to a maximum horizon of three peak wave periods (45 seconds).

To understand how the input was incorporated to create each forecast profile, consider one instant in time. Figure 3.3 shows an example of input used with respect to the first and last prediction of the multi-step forecast. The upper graphs represent

Step 1 in which the exogenous input has already been forecasted. The 45-second forecast is marked by the blue circles that are perfectly aligned with the expected values for the purpose of this example. The lower graphs represent Step 2 in which the feedback is currently being forecasted. The red boxes on the left isolate the required input from both time series to make the first 1-second forecast. The red boxes slide forward to accommodate the NARX model as it runs in feedback mode and incorporates more and more previously forecasted values as input. The input required to make the final 45th-second forecast in Step 2 relies heavily on previously forecasted values as shown on the right.

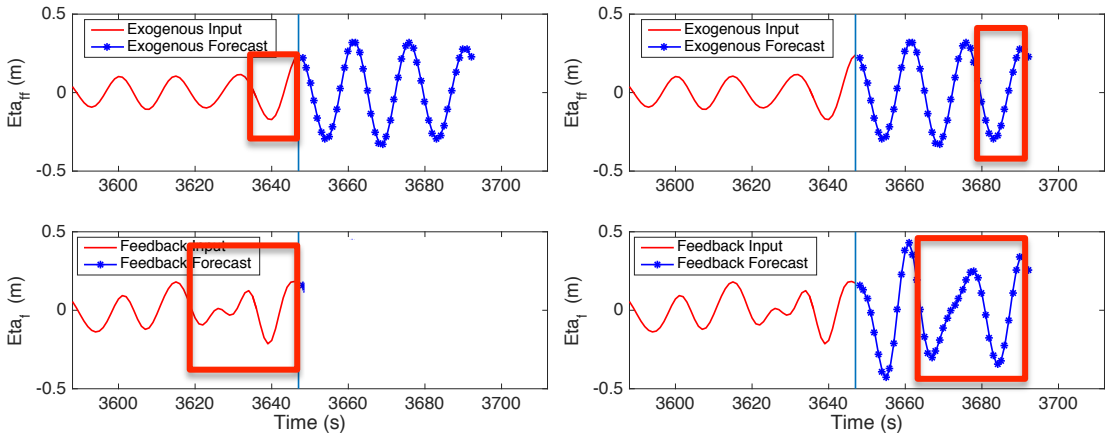


Figure 3.3. Examples of the required Step-2 input indicated by the red boxes for the first (left) and last (right) prediction in the multi-step forecast.

Figure 3.4 depicts the 2-step model as it was implemented in real-time. First, the training phase was conducted on the recent wave record, providing the networks to be used in the Prediction Phase. Next the prediction phase was initiated at the time of interest (the last known data point in the wave record). The input is preprocessed and then submitted to the networks to provide the first forecast profile. At the next time-step, once the newly acquired measurement is incorporated into the wave record, the model marches forward one step to repeat the process. This would continue for perhaps tens of minutes until the networks required retraining.

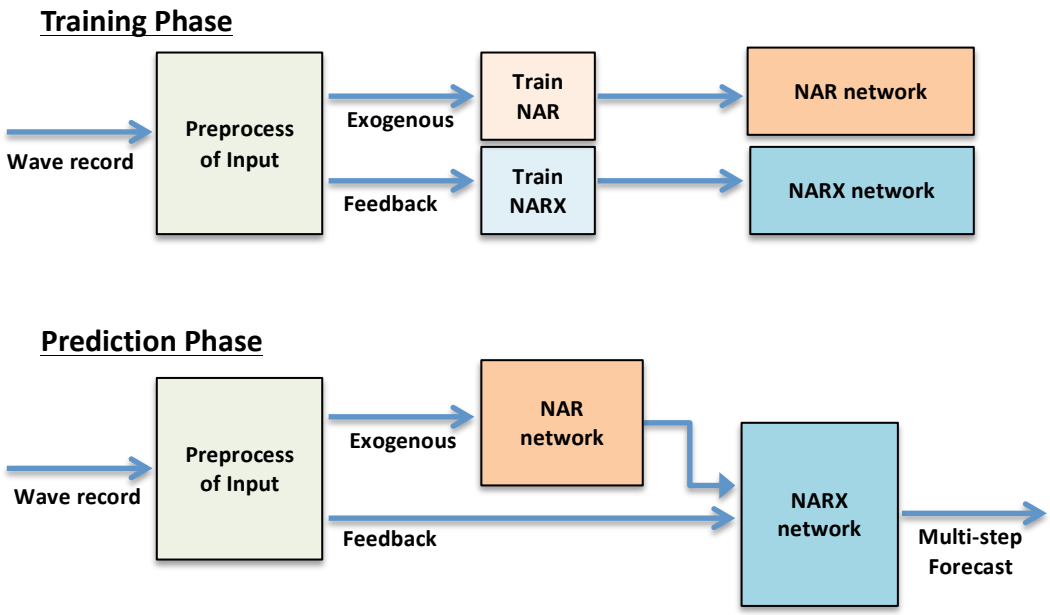


Figure 3.4. Process flow of the training and prediction phases.

3.1.2 Offline Training and Computational Requirements

The NARX model was trained on the recent time history of point measurements using the Levenberg-Marquardt algorithm. Refer back to 2.5.2 for a description of the training options. The network was trained in an open-loop configuration, meaning that actual data was used in leau of feedback. This allowed for a more efficient training with the back-propagation method. Once taining was complete, the network was converted to a closed-loop configuration, allowing for feedback and multi-step forecasts.

The training/validating/testing process was conducted in a 70/15/15 split and typically required 20-30 minutes of data. A sample training record is given in Figure 3.5. The markers indicate training targets (blue), validation targets (green), and testing targets (red). Block division of data, as opposed to the default random division, was utilized to better track longer-term correlations.

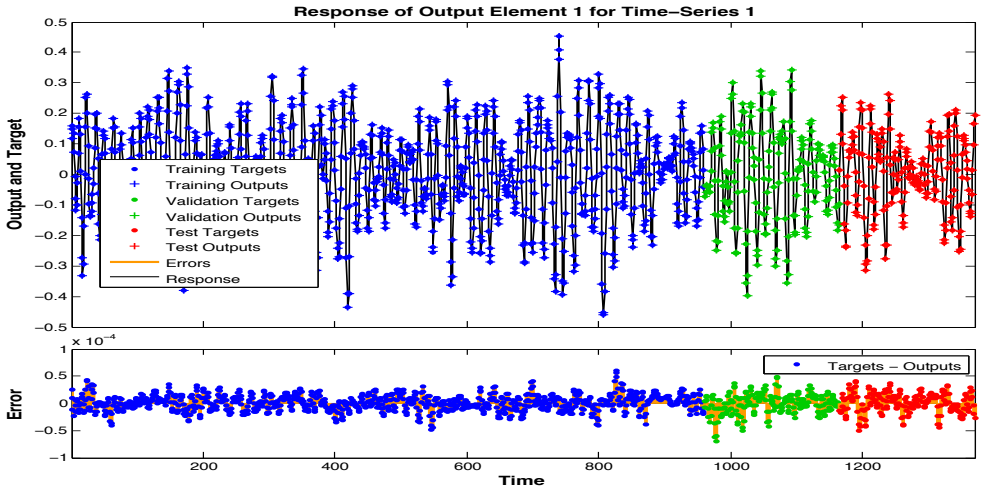


Figure 3.5. A Sample training record.

The training typically required 5-10 seconds per network. Due to the nature of neural networks, each training yields a slightly different output. For all neural networks, occasionally the training goes poorly and the resulting network is unstable or underperforms. Thus the networks were trained multiple times and the network associated with the best open-loop performance determined by minimum mean-squared errors was selected. Since the training process typically required less than 1 minute, and the resulting network could perform multi-step forecasts in less than 0.2 seconds, the method was suitable in this regard for real-time applications.

3.2 Preprocessing of Input

The most important preprocessing step is subtracting the tidal trend and mean elevation from the wave record. This initial processing is common practice in ocean analysis, is easily accomplished in real-time, and was assumed in all datasets. Thus, the preprocessing discussed moving forward, is in addition to de-trending and de-meaning measured data. Data termed raw or unprocessed in the following text refers to the de-trended, zero-mean dataset.

Various methods of preprocessing were explored including filtering in the frequency domain, filtering in the time domain, de-noising using wavelet analysis, and wavelet transformations. The feedback input is the primary input and serves as the target of the forecast. Smoothing of the feedback input requires that the phase and general character of the measured data be conserved. Any steps beyond simple smoothing that significantly alter the input must be reversible post forecast. Exogenous input, on the other hand, serves as an aid to the forecasting network and is not limited in this way. However, the exogenous input must be both correlated to the feedback input and easily forecasted to be of any value to the network.

Preliminary investigations into potential preprocessed signals identified the following approaches worthy of further testing.

Feedback input:

- no preprocessing
- level 2 wavelet filtered input (de-noised)
- low-pass filtering in the frequency domain
- wavelet transformed input

Exogenous input

- band-pass filtering in the frequency domain
- level 3 wavelet input

Other approaches considered but eliminated due to poor predictability were moving average and exponential moving average, two smoothing techniques in the time domain. Also in the time domain is the loess filter, a local regression using weighted linear least squares and a second order polynomial. The loess filter provided similar performance to the low-pass filter in the frequency domain and could be a suitable alternative if needed.

3.2.1 Filtering in the Time and Frequency Domains

The first preprocessing approach considered is the smoothing of feedback input. Smoother input generally facilitates autoregressive forecasting. It is accomplished by the removal of higher frequency variations that contain both measurement noise and shorter, low energy wind waves. Although loss of shorter wind waves may be undesirable, the resulting smoother signal represents the lower frequency waves that carry the greater portion of total wave energy and determine the general shape of the wave profile. Filtering can be achieved in the time domain by averaging over successive data points or the frequency domain by limiting higher frequency components.

In the time domain, symmetrical smoothing conserves signal phase by weighting data on both sides relative to the time of interest. In other words, the windowing of data is centered on the time of interest. Windowed data to the left of center is termed backward-looking data and that to the right of center is forward-looking data. Filtering errors, or edge effects, occur at the beginning of the wave record due to the lack of backward-looking data, and at the end of the record due to the lack of forward-looking data. In the case of forecasting, the end of the wave record is of primary importance. Asymmetrical smoothing in which only backward-looking data is used, eliminates the edge effects at the time series end, but at the price of a phase shift. If the phase shift is linear (does not vary over the time domain), it can be corrected. The exponential filter is one asymmetrical filter that was considered, but eliminated, since it did not produce a signal easier to forecast.

Filtering in the frequency domain by way of discrete Fourier transforms, can produce a very clean signal, but also produces the attendant edge effects. These effects are due to limitations in the Fourier transformation. The discrete Fourier transform is a finite approximation of an infinite mathematical process. The approximation assumes that the finite wave record repeats forward and backward smoothly and infinitely, thus imposing an artificial period of the length of the wave record. The time series end is assumed to continue on from the beginning and so forth. Additionally, any discontinuities within a series produce spurious oscillations

(rippling artifacts due to the Gibbs phenomenon). Band-pass and low-pass filtering create discontinuities that manifest as ripples in the time domain.

Conversely, windowing in the time domain cause ripples upon transformation to the frequency domain. These edge effects are inescapable, but can be mitigated. Window tapering in either domain reduces the edge effects by reducing the sharpness of the discontinuities. End-point matching is a technique in which the wave record end-points are matched by adding a linear trend so as to reduce the discontinuity caused by the imposed Fourier period.

3.2.2 Wavelet Transformations

The present work was conducted using the Matlab R2015b wavelet toolbox and the following descriptions of wavelet transformations are with respect to the Matlab algorithms as explained in Michel Misiti et al. (2015).

For this analysis, the symlet mother wavelet of order 6 (ψ = Sym6*)* was chosen from the dictionaries of wavelet functions based on smoothness and symmetry and some trial and error. The Sym6 wavelet is shown in Figure 3.6. Discrete wavelet transforms achieve greater efficiency by using scales and positions based on powers of two (dyadic scales). Thus, consider a rectangular grid with scale defined on Z and scale-related time defined on N. The mother wavelet is scaled by the factor a, and translated to position b according to Equation 3-2.

$$a = 2^j, \quad \mathrm{b} = \mathrm{k}2^j, \qquad j \in N, k \in Z \tag{3-2}$$

It follows that the discrete wavelet transform yields coefficients in scale-related time according to

$$C(a,b) = C(j,k) = \sum_{n \in Z} s(n) g_{j,k}(n) \quad , \tag{3-3}$$

where the sampled signal is given by s, and the wavelet filter, *g,* plays the role of the mother wavelet, ψ. Reconstruction of the signal s is achieved via the inverse wavelet transform

$$s(t) = \sum_{j \in Z} \sum_{k \in Z} C(j,k) \psi_{j,k}(t) \quad . \tag{3-4}$$

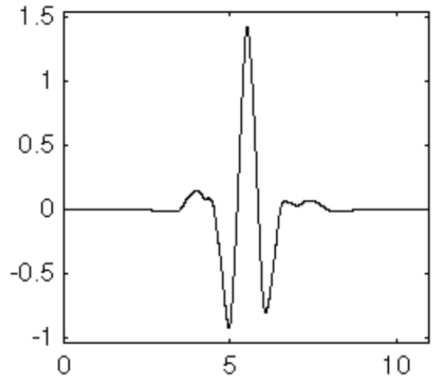


Figure 3.6. Mother wavelet Sym6 (Michel Misiti et al., 2015).

Wavelet details D(t) are defined according to

$$D_j(t) = \sum_{k \in Z} C(j,k)\psi_{j,k}(t) \tag{3-5}$$

and are related to the sampled signal by

$$s = \sum_{j \in Z} D_j \quad . \tag{3-6}$$

Two sorts of details are defined with respect to a reference level J. The approximation of the signal (A) corresponds to the sum of details above J according to Equation 3-7. These are the coarser details responsible for the general character of a signal.

$$A_J = \sum_{j>J} D_j \tag{3-7}$$

Equation 3-6 can be rewritten with respect to the reference level J where the finer variations are within the summation of details below J according to

$$s = A_J + \sum_{j \leq J} D_j \quad . \tag{3-8}$$

Discrete wavelet transformations were utilized in this work in two distinct ways. The first involved de-noising via wavelet approximations. As shown above, a signal can be decomposed into approximate and detailed signals at increasing levels of scale. Generally, D_1 and D_2 contain the noise of the signal and de-noising can be done by omitting or reducing their contributions. Thus, the first method was to use the A_2 approximation as the de-noised input signal for NARX.

The second approach was based on forecasting wavelet coefficients as described in 2.6.2. For the second method, the NAR input was converted to level 2 wavelet coefficients and the forecasted coefficients were reconstructed to the A_2 approximation post forecast.

In both cases the input was decomposed to the second level. Scaling levels 3 and above decompose the signal further to effectively yield progressively more extreme low-pass and high-pass filtered versions of the signal. The general character of level 3 and higher approximations differed significantly from that of the original signal. Consideration for the use of the higher-level decompositions was reserved for the exogenous input. As a final note, the maximum decomposition level is dependent upon mother wavelet type and signal length.

3.3 Assimilation of Data in Time

The assimilation of new measurements into feedback and exogenous inputs as they were made available in time is necessary for real-time application of the model. This assimilation was done as model progressed in time at the rate of data sampling and continued potentially, for as long as the training sustained the network. A new multi-step forecast of wave elevation profiles was incrementally produced corresponding to each step of the progression.

The assimilation of Data in time occurs in the prediction phase as outlined in Figure 3.7. The basic steps include the following.

- Augmenting newly measured data to the wave record

- Implementing input preprocessing on the wave record
- Mitigation of edge effects

The exact nature of these steps is dependent upon the type of preprocessing and the mitigation tactics implemented. A detailed description of the assimilation process is deferred to 5.4.

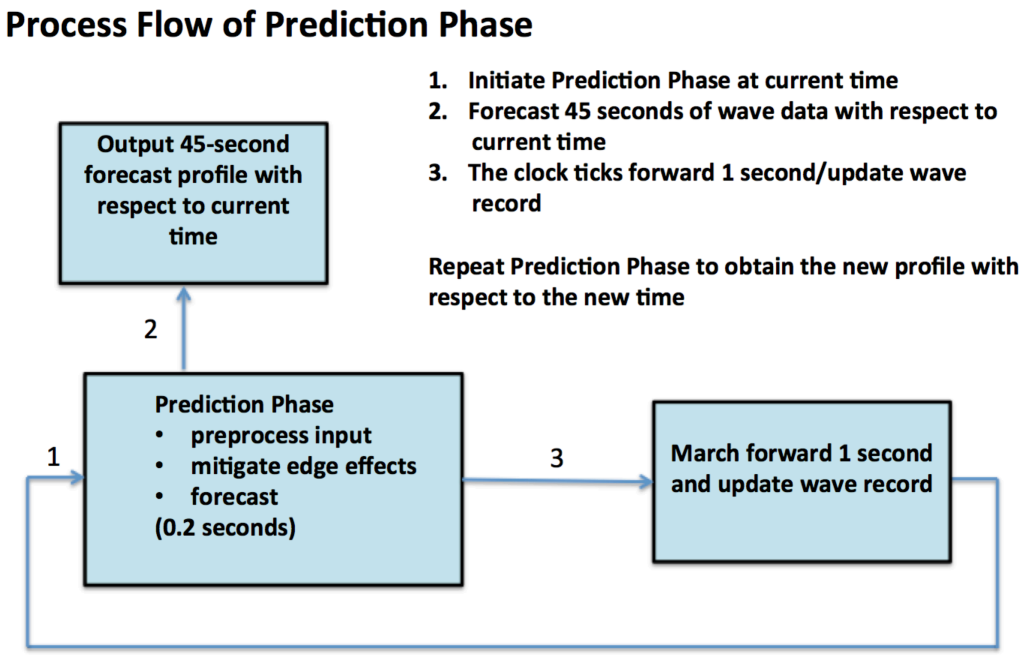


Figure 3.7. Process flow of Prediction Phase

4 Wave Record Datasets

The NARX forecasting model was developed based primarily on one data set that was collected off the coast of Huntington Beach, California. A second data set was considered post development to test the model application in a more general sense.

4.1 Huntington Beach, California

The Huntington Beach dataset was collected off the coast of southern California at a depth of about 15 m. The 68-minute wave record was sampled at 1 Hz and had waves on the order of 0.5 meters with a peak wave period of 15 seconds. As can be seen from the spectrum given in Figure 4.1, the data exhibit a bimodal character. The primary peak is likely the result of ocean swell conditions and the broader, secondary peak is more typical of local wind-waves.

4.2 Newport, Oregon

The Newport dataset was collected in the summer of 2013 via acoustic wave and current profiler (AWAC) at the North Energy Test Site. This site was located 2-3 miles off the coast of Oregon and had an average depth of 50 m. The dataset contains 601 time series of heave measurements, each 40 minutes long and sampled at 2 Hz. Wave record 417, selected for this analysis, had waves on the order of 1 meter with peak wave period of 10.5 seconds. The spectrum, given in Figure 4.1 shows a single, relatively broad peak.

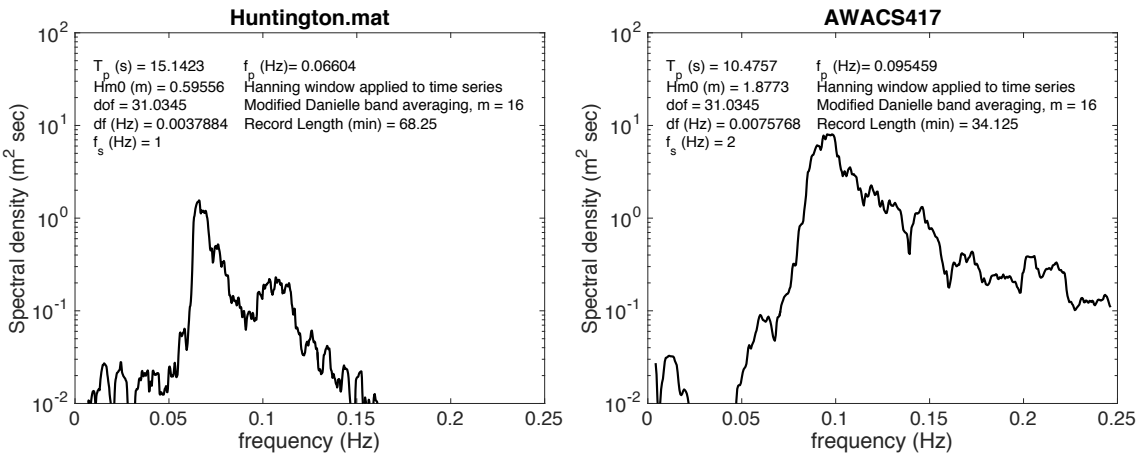


Figure 4.1. Energy spectrum of Huntington Beach dataset (left) and Newport 417 dataset (right).

5 Results and Discussion

Overall performance was evaluated in stages to isolate the potential effectiveness of input preprocessing from the challenges of implementing the preprocessing in real-time. All primary analyses were conducted using the Huntington Beach dataset that was described in 4.1. Some further analyses were conducted using the Newport dataset that was described in 4.2.

The analysis began with an investigation into the value of several feedback-preprocessing methods. The key findings are outlined in 5.1. In the next stage, the benefit of adding an exogenous input was evaluated, which is discussed in 5.2. Upon identifying the most viable set of input conditions, the model testing was expanded to the Newport dataset as described in 5.3. In the final stage, the analysis turned to the implementation of the preprocessing steps in real-time (5.4). Several comparable linear-AR cases were also considered and these results are given in 5.5. The full analysis is represented by a progression through 13 salient cases, with results for each case given in figures of identical format. An explanation of this format is beneficial before evaluating the individual cases.

Consider Figure 5.1 as a reference. On the left side of each figure three forecast profiles are given, which represent random snapshots of performance as the NAR(X) model progresses (or marches) in time. The data are in black, where values shown to the left of the vertical line are known and values to the right are unknown to NAR(X) but are provided as a reference. The input is in red and the multi-step forecast computed at each time step of the march is in blue. In these cases, a forecast maximum is equivalent to 3 peak wave periods (T_p = 15 s) or 45 seconds. Although the profiles shown are separated by 10 seconds, a 45-second forecast was computed with respect to every second of the march.

On the right side of the figures, the corresponding errors are expressed in two forms. The average correlation between the forecasted and expected values is given by a correlation factor shown in blue. As noted in 2.4, this value does not take into account magnitudinal differences, but gauges whether the forecast and reference

signals are in sync. The coefficient of efficiency, shown in red, is an average measure of the relative difference in amplitude at each forecast step. For both metrics, a value of 1 indicates good agreement. The correlation factor and coefficient of efficiency were computed according to Equations 2-3 and 2-4, respectively. Errors corresponding to a forecast length were averaged cumulatively. For example, the errors associated with a 10-second forecast length account for all of the first ten steps of each forecast over the full 5-minute march.

5.1 Determination of Effective Feedback Input

Cases 1 - 4 represent NAR forecasts of the Huntington Beach data set using several different methods of input preprocessing. These forecasts do not utilize exogenous input. The purpose of these forecasts was to assess the potential value of various preprocessing methods with regard to the predictability of the feedback input. Preprocessing was applied to the entire data set prior to use in the forecasting model. Thus at this stage of the analysis, all issues regarding filtering edge effects were put aside. Real-time application would require preprocessing on the wave record as it grows adding potential complications.

Case 1 serves as the base case and is shown in Figure 5.1. Raw data was unprocessed prior to input and no exogenous input was provided. It is clear in a visual inspection of the profiles that the forecast performance fell off after the first steps of the process. This is consistent with the general performance as represented by the error calculations averaged cumulatively with respect to forecast length, over a march of 20 peak wave periods (5 minutes). The error values averaged over 1 peak period were $R_{15} = 0.70$ and $CE_{15} = 0.52$.

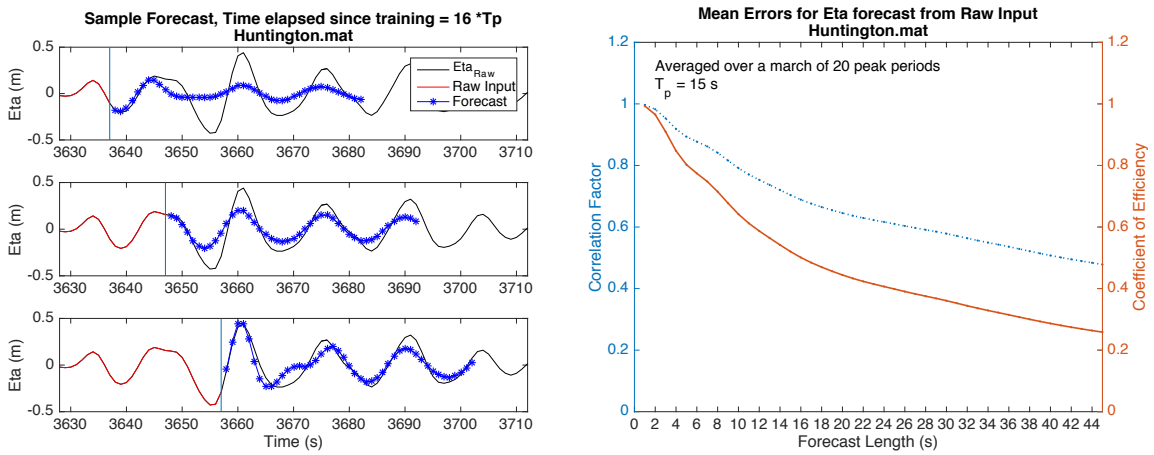


Figure 5.1. Case 1: Sample NAR forecast profiles based on raw input (left) and the corresponding errors (right).

In Case 2, the feedback input was de-noised via a second-level wavelet transform. Note that the input (in red) differed slightly from the reference. Results show no improvement over the base case ($R_{15} = 0.70$ and $CE_{15} = 0.51$), indicating that this preprocessing was unhelpful. In the present analysis, only A_2 approximations were considered for feedback preprocessing. The A_3 signals differed too significantly from the raw data to serve as feedback input. Forecasting A_3 and D_3 in separate networks and then summing the two resulting forecasts is another way to yield an A_2 forecast. However, this approach proved no better than the more direct method.

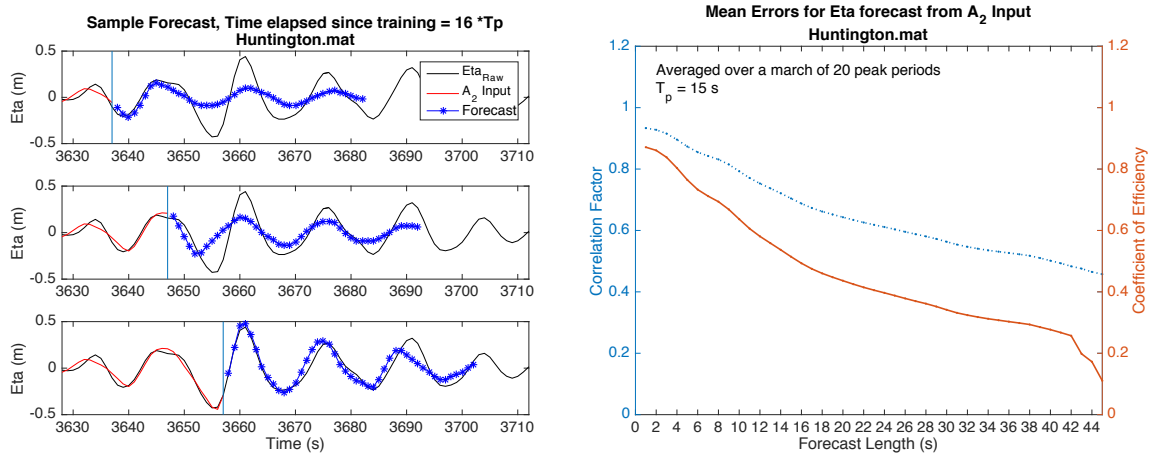


Figure 5.2. Case 2: Sample NAR forecast profiles based on wavelet-filtered input (left) and the corresponding errors (right).

In Case 3 an alternative use of the wavelet transformation was employed. This involved forecasting the wavelet coefficients similar to what was done in previous analyses that are described in 2.6.2. This method involved more intermediated steps. Wavelet coefficients represent the magnitude of the correlation of the scaled mother wavelet with respect to time. Since discrete wavelet transformations employ successive down-sampling at each level of decomposition, missing correlation coefficients corresponding to the original data sampling were approximated using cubic interpolation. The interpolated time series of coefficients was then used as feedback input yielding a forecast in terms of wavelet coefficients. This output had to be down-sampled prior to wavelet reconstruction to ultimately yield the wave elevation forecasts. The NAR performance, shown in Figure 5.3, was fairly level across forecasting length, but generally poor overall. The error values averaged over 1 peak period were $R_{15} = 0.63$ and the $CE_{15} = 0.40$.

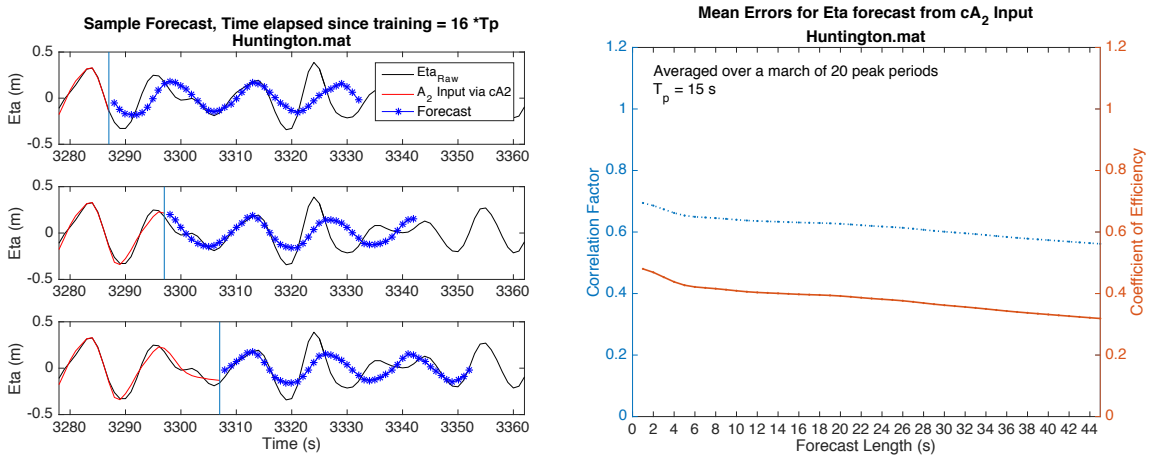


Figure 5.3. Case 3: Sample NAR forecast profiles based on wavelet-transformed feedback input (left) and the corresponding errors (right).

Forecasting improvement was noted in Case 4 (Figure 5.4). Here the input was filtered via a low-pass frequency filter with a cutoff set to 1.9 times the peak frequency. In this case, the forecast was generally good up to 1- 2 peak wave periods (~15-30 seconds). The error values averaged over 1 peak period were $R_{15} = 0.96$ and

CE_{15} = 0.92. For reference, Figure 5.5 provides the Huntington Beach energy spectrum corresponding to both filtered and unfiltered signals.

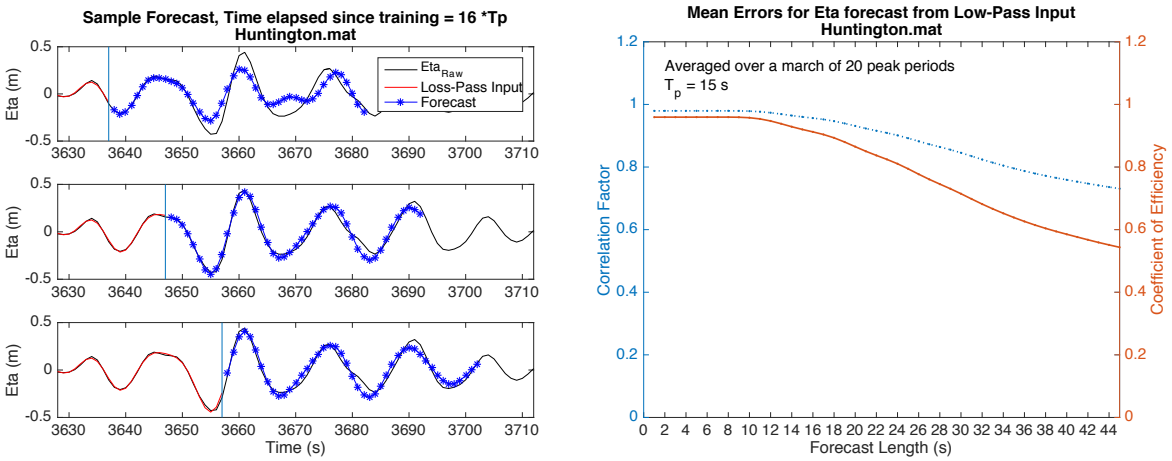


Figure 5.4. Case 4: Sample NAR forecast profiles based on low-pass filtered input (left) and the corresponding errors (right).

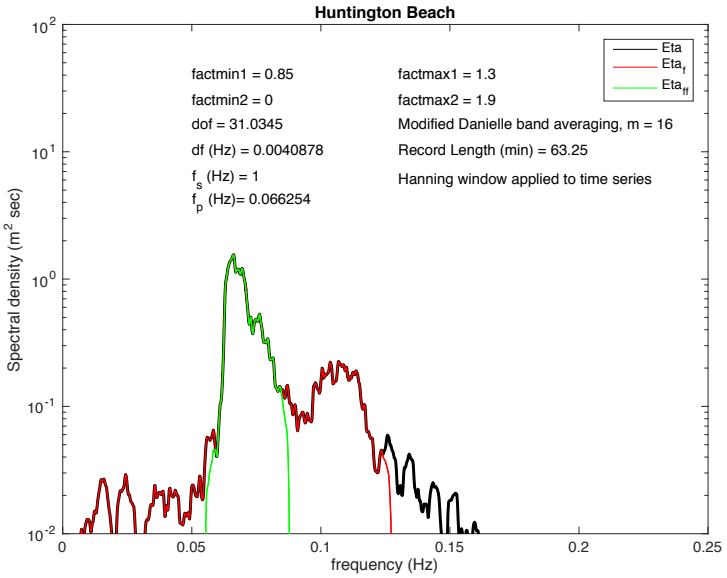


Figure 5.5. Spectrum of Huntington Beach dataset showing low-pass and band-pass filtered signals.

5.2 Determination of Effective Exogenous Input

An exogenous input is a separate time series used to aid the NARX model in forecasting the primary feedback input. Use of an exogenous input would be helpful only if it is both predictable and correlated to the feedback input. Various potential signals were considered, including exponential filtering in the time domain, narrow-band filtering in the frequency domain, and wavelet decomposed approximate and detailed signals.

Two predictable time series were identified as potential exogenous inputs and were implemented in the NARX model. The first signal was a level-3 wavelet approximate signal (A_3). The second signal was a narrow-banded filtered signal isolating frequencies near the peak frequency (f_p) of 0.07 Hz. Refer back to Figure 5.5 for the spectrum of this filtered signal. The narrow-band cutoffs were 0.85 f_p and 1.3 f_p, and thus included frequencies between 0.06 and 0.09 Hz. The low-pass signal included all frequencies below 0.13 Hz. The NAR forecasts for these two time series are given in Figure 5.6.

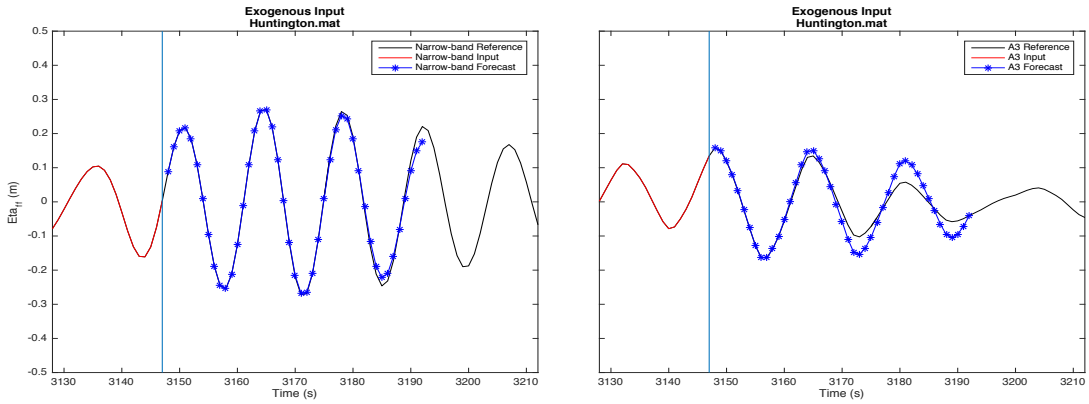


Figure 5.6. Sample Step 1 NAR forecast profiles of narrow-band exogenous input (left) and A3 exogenous input (right).

Implementing Step 1 of the NARX model using the narrow-band exogenous input did, in fact, extend the prediction horizon significantly. Case 5 represents raw

feedback and narrow-band exogenous input and is given in Figure 5.7. Note that the corresponding error values averaged over 3 peak periods (R_{15} = 0.76, CE_{15} = 0.58) exceed the 1 peak period values for Case 1. The results are still better assuming low-pass feedback (Case 6) as is evident in Figure 5.8. The error values averaged over 3 peak periods were R_{45} = 0.88 and the CE_{45} = 0.75.

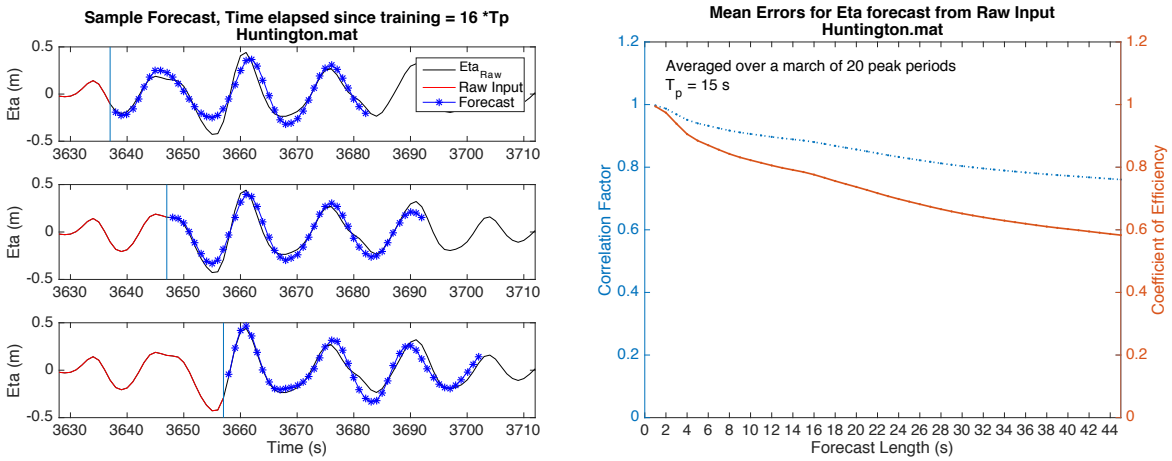


Figure 5.7. Case 5: Sample NARX forecast profiles based on raw feedback input and narrow-band exogenous input (left) and the corresponding errors (right).

The improvements offered by the use of the level-3 wavelet approximation as exogenous input were disappointing. It was anticipated that this type of preprocessing if it were found useful, would have fewer edge-effect concerns under real-time application. In the case of raw feedback input, the error values averaged over 1 peak period were R_{15} = 0.81 and CE_{15} = 0.67, only slightly better than those for Case 1. The 3 peak period values were however, no better (R_{45} = 0.52 and CE_{45} = 0.31). The overall results are not provided in figures. The combination of low-pass feedback and A_3 exogenous input resulted in an unstable forecasts beyond ~1 peak period. This was likely the result of the signals being too similar. Similarity becomes a problem when either of the inputs cannot be well-predicted. It is surmised that since the NARX model is trained on ideal input that are consistently correlated, it is not capable of handling inputs that diverge as errors accumulate.

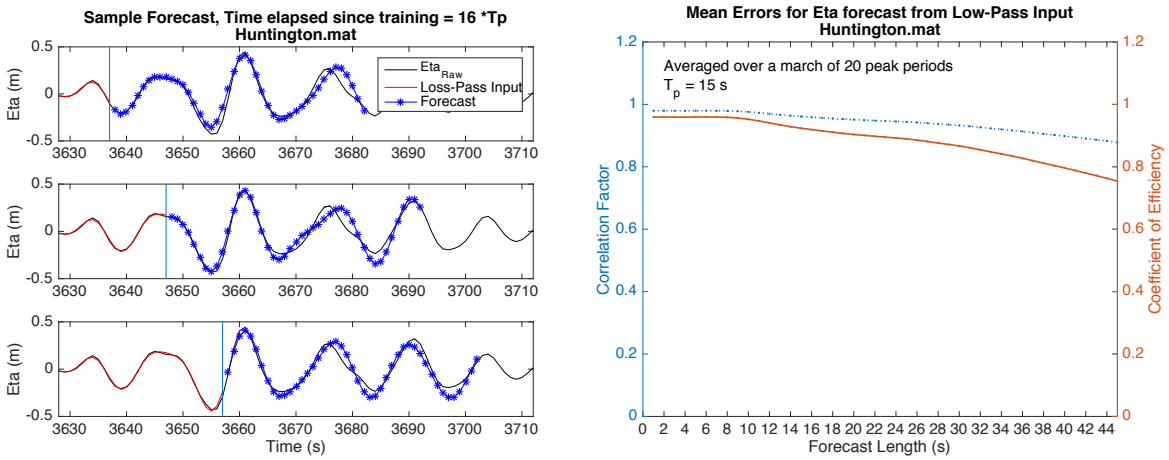


Figure 5.8. Case 6: Sample NARX forecast profiles based on low-pass feedback input and narrow-band exogenous input (left) and the corresponding errors (right).

5.3 Extending Assessment to the Newport Dataset

The next step in the analysis was to assess whether this approach may be extended to the Newport dataset. Refer back to 4.2 and Figure 4.1 for the specifics of this second dataset. The model parameters were unchanged except for a minor tuning of the filtering frequencies with respect to the peak frequency of 0.096 Hz for this wave record. The spectrum of the filtered signals is given in Figure 5.9. The cutoff frequencies are again given as multipliers of the peak frequency, between $0.7f_p$ and $1.1f_p$. Thus, the narrow band includes frequencies between 0.07 and 0.11 Hz. The low-pass input in this case, includes all frequencies below 0.18 Hz.

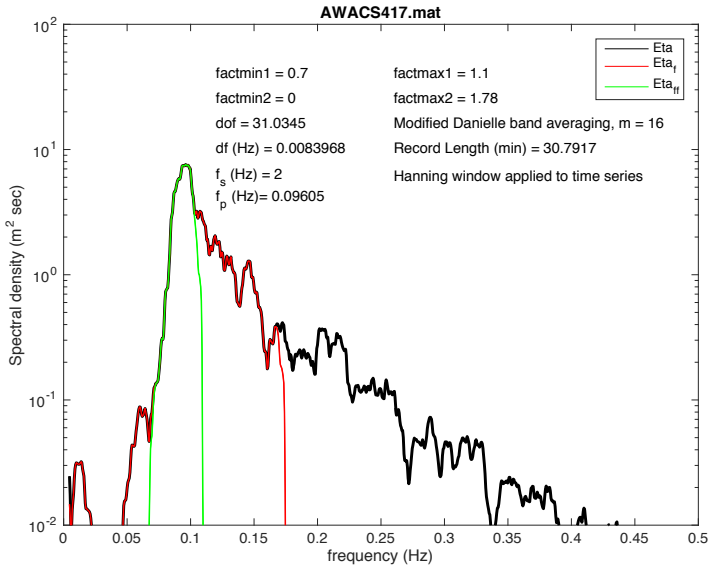


Figure 5.9. Spectrum of Newport 417 dataset showing low-pass and band-pass filtered signals.

Results for Cases 7 and 8 are given in Figures 5.10 and 5.11, respectively. These cases represent the performance corresponding to Cases 5 and 6, using the Newport dataset. It is evident that the exogenous input was beneficial to overall performance as was true for the Huntington Beach dataset. However, the improvement was not quite as significant for the second dataset.

This difference may be attributed to the different sea states. As noted earlier in 2.6.1, stochastic forecasting is limited by the intrinsic correlation of the signal, which is inversely proportional to the bandwidth. Thus, narrow-banded signals are more highly correlated than broad-banded signals. The Huntington Beach dataset is characterized by a bimodal sea state in which there is a clear narrow-banded swell component separate from a local wind-wave component (Figure 5.5). The Newport spectrum (Figure 5.9), characterized by a single, broader peak, has a swell component that was not so easily isolated for the exogenous input. Still, the improvement was substantial enough to suggest general forecasting potential.

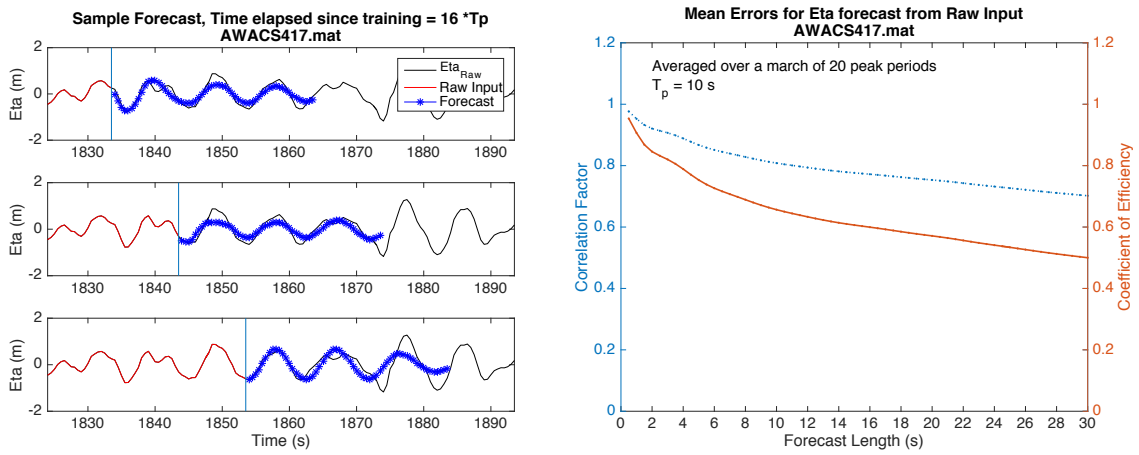


Figure 5.10. Case 7: Sample NARX forecast profiles based on raw feedback input and narrow-band exogenous input (left) and the corresponding errors (right) for Newport 417.

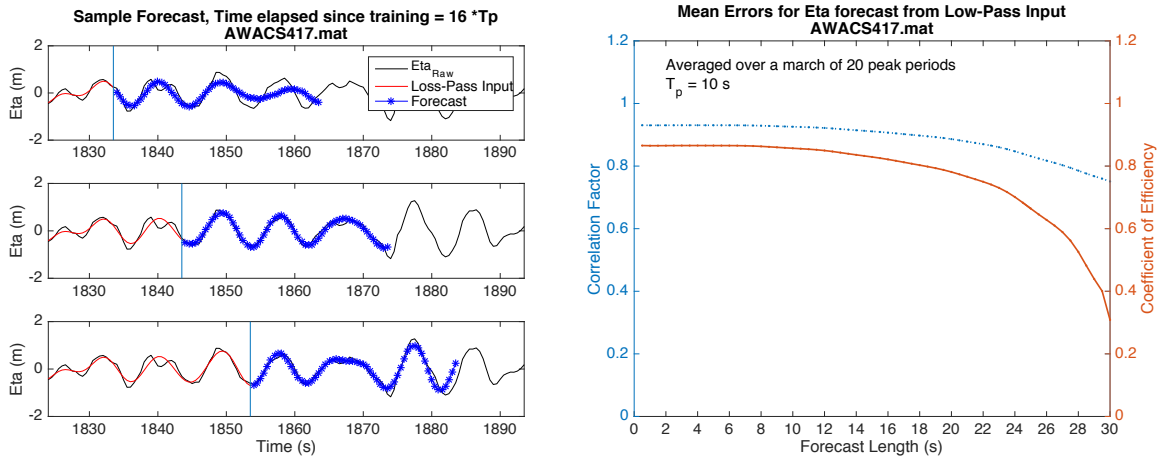


Figure 5.11. Case 8: Sample NARX forecast profiles based on low-pass feedback input and narrow-band exogenous input (left) and the corresponding errors (right) for Newport 417.

5.4 Implementation of Preprocessing Steps in Real-time

Thus far, the model has been developed using a simplification in which the input was preprocessed as a whole prior to use in the forecasting model. Although no forward-looking data was made available to the model in the forecasting stage, it was essentially available in the preprocessing stage due to the symmetrical nature of the filter. This simplification was made in the development stage to separate poor

performance due to useless input from that largely due to edge effects and other real-time filtering challenges that could be addressed later. For this reason, the results cannot be considered realistic at this stage. A realistic application of the forecast model would treat input preprocessing in real-time, using only the data that would be available at the current index.

The particular challenge of the real-time application is the narrow-band exogenous input. This was the exogenous input determined to be the most effective contributor to improved forecasting performance. Unfortunately edge effects are fairly severe in narrow-band filtering and hence, they must be mitigated if there could be any practical use of this forecasting approach. Two mitigation tactics were posited.

The first tactic was to extend the wave record beyond the current index so that the most important data was not at the end of the time series. This was done in real-time by augmenting the wave record with the forecasted wave profile acquired in the previous step. The goal was that the augmented portion of the wave record would serve as a buffer during the filtering process. This could work only if the previous forecast was accurate and long enough that the ripples would largely confined to the buffer. The buffer would then be discarded post filtering to yield an exogenous input of similar quality from middle to end. Further, the Fast Fourier Transform algorithm (FFT) in Matlab operates most efficiently on time series with lengths of a power of 2. After some trial and error, the filtering window was set to 2048 data points to accommodate this constraint. Additionally, endpoint matching was achieved by a linear tapering of the beginning and end of the series to zero.

In theory, the initial forecasting would be launched far enough from the record end to allow a buffer of real data. As the forecasting computations could be conducted quickly, the model would then catch up to the current time index in a matter of seconds.

Beyond forecasting, the NARX network has been commonly used for signal noise reduction. The second tactic would utilize NARX in this capacity in an attempt to repair the end of the exogenous input. By reversing the feedback and exogenous inputs, the NARX model could be trained to target the narrow-band signal using the

measured data as the exogenous input. If the results were good enough, the corrupted end of the filtered signal could be replaced by the narrow-band approximate.

Figure 5.12 shows the Step 1 intermediate forecast of the exogenous input utilizing filtering in real–time. The graph on the left shows the intermediate forecast based on the unmitigated input in green. It is evident that the forecast was inaccurate due to corrupt input. Improvement to this intermediate forecast was achieved by employing both mitigating tactics together as can be seen in the graph to the right. The green signal in the right graph represents improvement using only the first tactic. Here the forecast was based on the red input, which was the combination of both mitigation efforts.

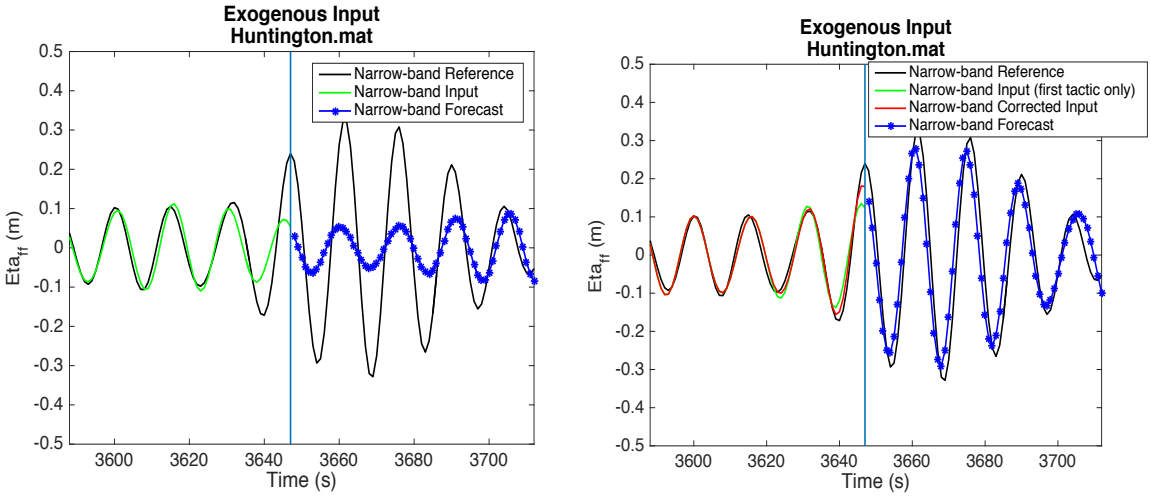


Figure 5.12. Sample exogenous input processed in real-time utilizing uncorrected end values (left) and corrected end values (right).

Unfortunately, the improvement proved too inconsistent to render an ultimately viable forecast. Figure 5.13 shows the resulting forecast for Case 9 in which raw feedback was aided by exogenous input preprocessed with corrections in real-time. Visual inspection of the forecast profiles seems to suggest a performance similar to that of the base Case 1. Yet, the generally poor error values indicate otherwise. This is possibly the result of instabilities at certain steps along the progression.

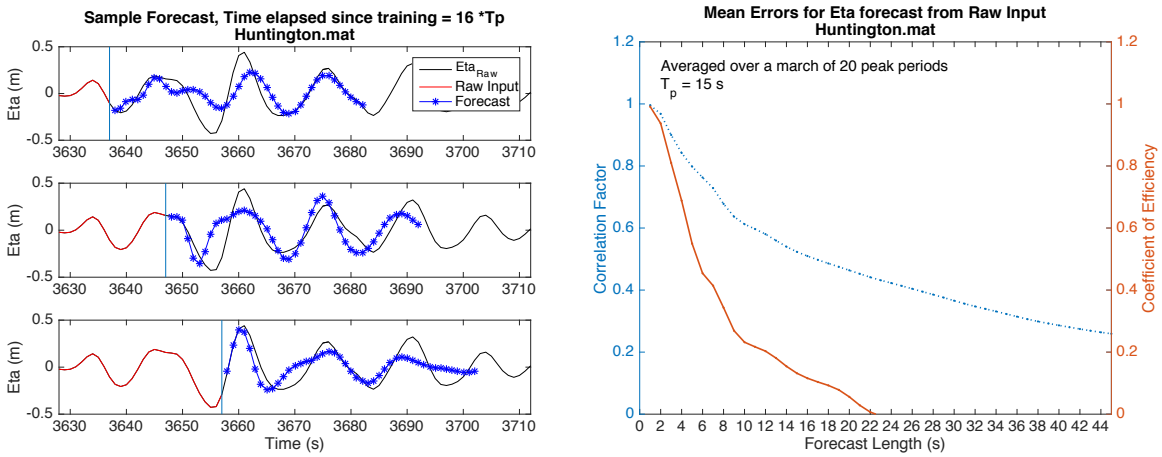


Figure 5.13. Case 9: Sample NARX forecast profiles based on raw feedback input and narrow-band exogenous input (left) and the corresponding errors (right) for Huntington Beach dataset conducted with input processed in real-time.

Case 10 represents the NAR forecast in which the low-pass feedback is processed in real-time without the use of exogenous input and is shown in Figure 5.14. Results represent an improvement over Case 9, but not enough to improve upon the base Case 1. It follows that combining the filtered exogenous input with the filtered feedback input yielded somewhat worse results. These results are not given in a figure.

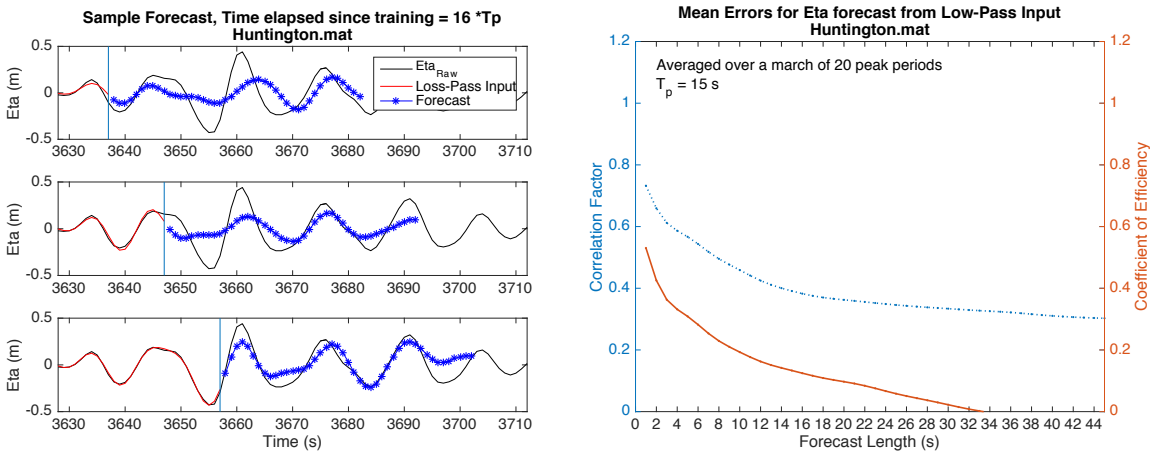


Figure 5.14. Case 10: Sample NAR forecast profiles based on low-pass feedback input (left) and the corresponding errors (right) for Huntington Beach dataset conducted with input processed in real-time.

5.5 NAR(X) Model Comparisons to Linear AR

The final step of this investigation was to compare the NAR(X) performance to that of the linear AR. The linear AR model utilized here was based on the least-squares optimal state estimator. The data was fit over the equivalent of 45 peak wave periods and used an input order of 30 previous steps.

Three AR cases corresponding to NAR Cases 1, 4, and 10 are shown in Figures 5.15 - 5.17, respectively. Figure 5.15 shows the forecast corresponding to Case 1 (Figure 5.1), which was based on raw Huntington Beach input. The results were very similar to that of the NAR model.

In Figure 5.16, the forecast was based on low-pass feedback input processed as a whole. The results yielded good forecasts for the first peak period, consistent to those obtained by Fusco (2010) via AR as summarized in 2.4.1. These results are also similar to the corresponding NAR forecast, unaided by the exogenous input. The NARX model shows superiority only with the use of the narrow-banded exogenous input as seen in Case 8, Figure 5.11. However, it was evident that use of this narrow-banded exogenous input was not reliable in real-time.

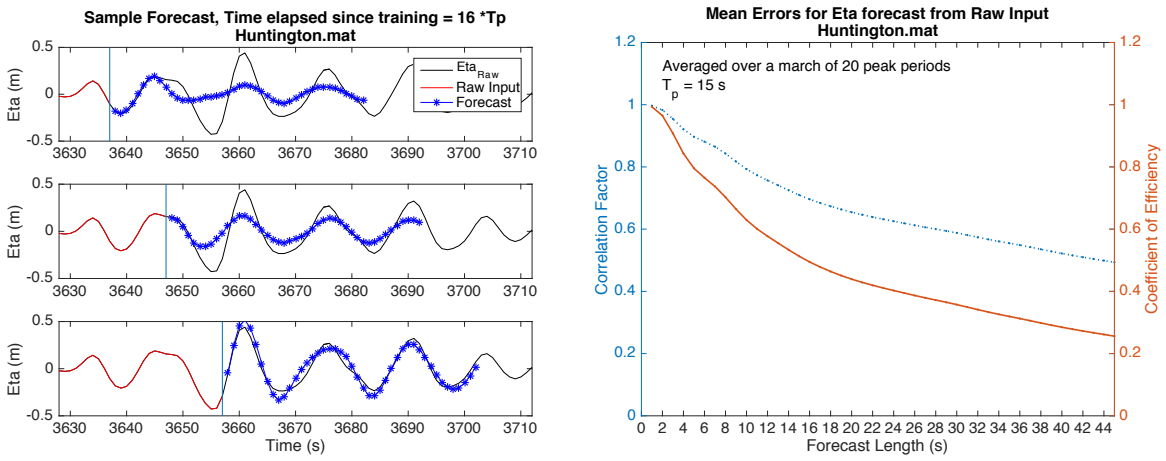


Figure 5.15. Case 11: Sample Linear AR forecast profiles based on raw feedback input and the corresponding errors (right) for Huntington Beach dataset.

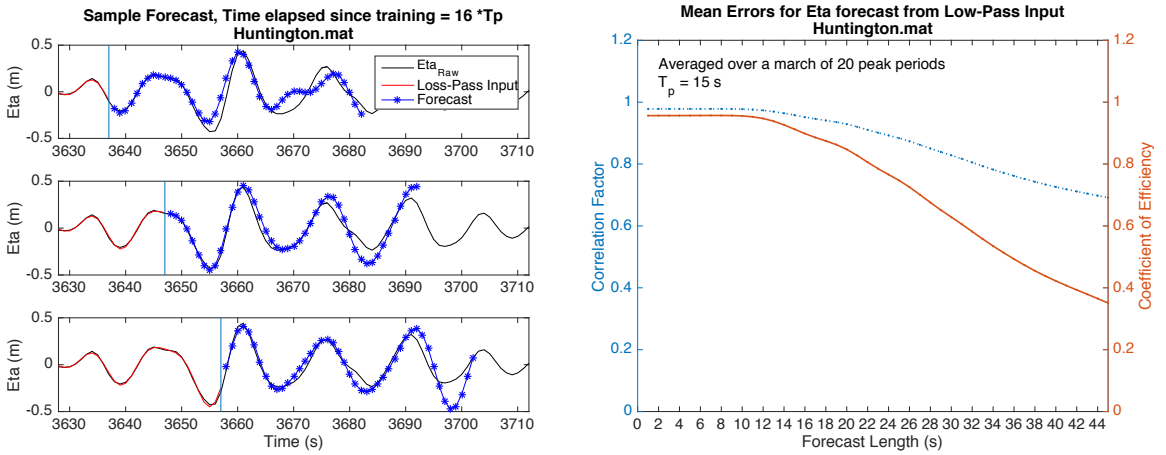


Figure 5.16. Case 12: Sample Linear AR forecast profiles based on low pass feedback input and the corresponding errors (right) for Huntington Beach dataset conducted with input preprocessed as a whole.

Figure 5.17 illustrates that the linear AR method likewise was susceptible to limitations in the real-time application of the low-pass filter, even with the benefit of the mitigation approach utilized for NAR. Results were generally no better than those for the corresponding NAR Case 10 (Figure 5.14).

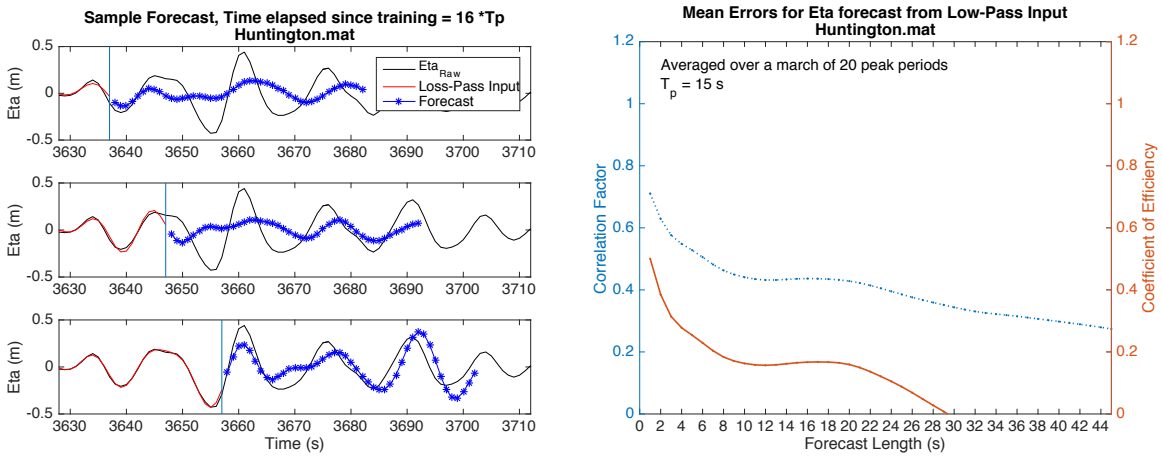
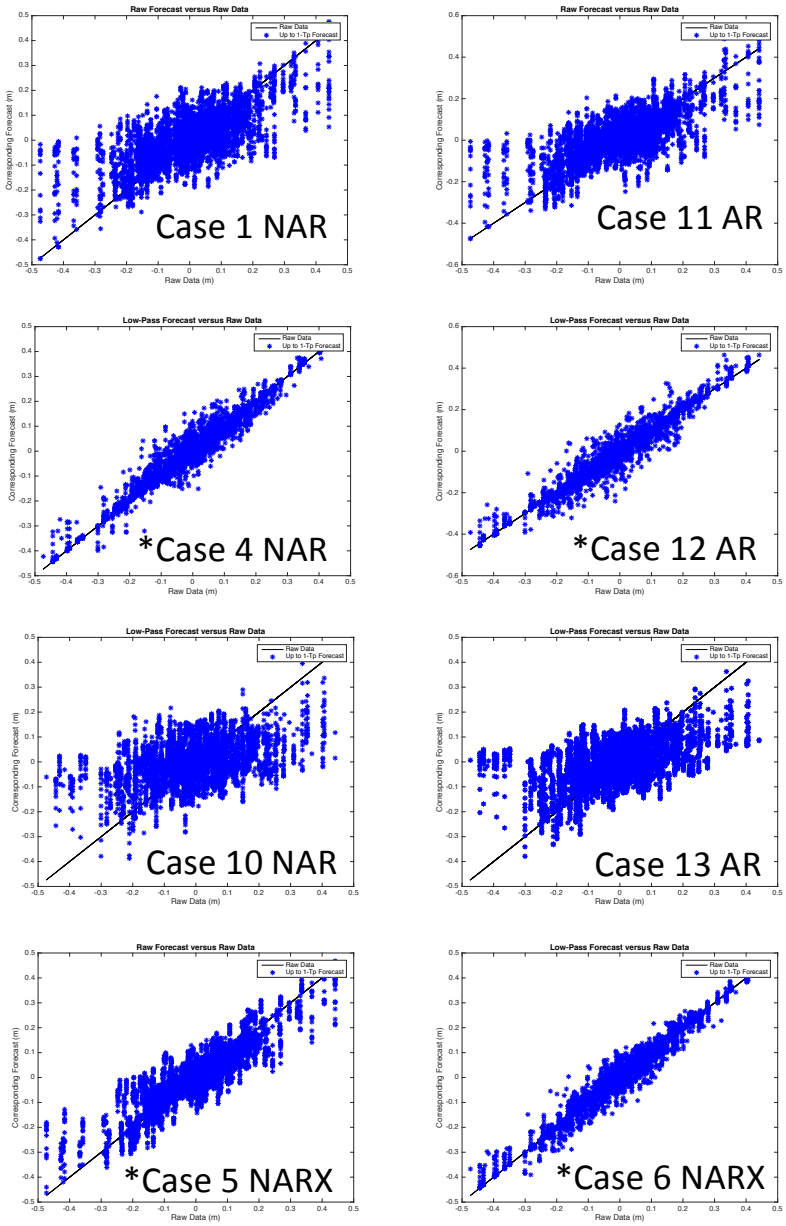


Figure 5.17. Case 13: Sample Linear AR forecast profiles based on low-pass feedback input and the corresponding errors (right) for Huntington Beach dataset conducted with input processed in real-time.

A different comparison of the forecasting results is given in Figure 5.18. Here the one T_p (15 s) forecasts are plotted against measured data for the 5-minute progression for several cases. The predictions, given by the blue symbols, would lie along the diagonal line if accurate. Note that the treatments using forward-looking data for preprocessing input are marked with the asterisk.

The top three graphs on the right represent the AR cases that were just discussed. Case 11, *12, and 13 use raw, low-pass (forward-looking), and low-pass input, respectively. The top three graphs on the left represent the comparable NAR cases. For all three treatments, it is apparent that NAR offers no advantage over AR. However, the bottom two graphs illustrate the advantage of NARX if a narrow-band exogenous input were possible. *Case 5 NARX is an improvement over both Case 1 NAR and Case 11 AR, as is *Case 6 NARX over both *Case 4 NAR and *Case 12 AR, most notably at longer horizons.



*Forward-looking data available during input preprocessing.

Figure 5.18. Forecast vs. Data for predictions up to one peak period (15 sec).

6 Summary and Conclusions

Forecast results of Cases 1–13 are summarized in Table 6.1 below. Although only discussed indirectly in 5.2, two extra input treatments were included in the last entries of the table for completeness.

The base NAR forecast (Case 1) is comparable to the AR forecast (Case 11). Both models are capable of forecasting 1 peak wave period with $R_{15} = 0.7$ and $CE_{15} = 0.5$. In this most basic example, the added complexities of NAR cannot be justified. AR has the advantage over NAR both in speed and simplicity.

Forecasting performance based on preprocessed data (filtering via Fourier transformations), and in particular with the use of the exogenous input, suggested great potential for NARX forecasting (Cases 5* and 6*). Credible forecast performance was extended to 3 peak periods using the Huntington Beach dataset, and 2 ½ peak periods with respect to the Newport dataset. If these preprocessing steps could be realistically implemented as new data is being incrementally added, NARX would hold a significant advantage over AR. Unfortunately, real-time implementation of the preprocessing steps proved problematic, as was demonstrated in Cases 9 and 10. Thus, it is apparent that the NARX-based model failed to provide an advantageous solution to the near-term forecasting problem, at least in a practical sense.

All that is left to consider are the reasons for the breakdown. As stated in 2.6.1, stochastic forecasting relies on the intrinsic correlations of the signal, and narrow-banded signals are more highly correlated than broad-banded signals. It followed that forecasting improvement was associated with the inclusion of a narrow-banded exogenous input. However, providing the narrow-banded signal in real-time was susceptible to the limitations of a zero-phase (non-causal) filter, which resulted in imperfect preprocessing.

Table 6.1. Summary of NAR(X) and AR Results. The error metrics R and CE are defined in Equations 2-3 and 2-4, respectively, with subscripts referring to forecast lengths in seconds.

Case (Figure)	Feedback	Exogenous	R_{15}	R_{45}	CE_{15}	CE_{45}
1 (Fig. 5.1)	Raw	N/A	0.70	0.48	0.52	0.26
2* (Fig. 5.2)	A_2	N/A	0.70	0.46	0.51	0.11
3* (Fig. 5.3)	cA_2	N/A	0.63	0.56	0.40	0.32
4* (Fig. 5.4)	Low-pass	N/A	0.96	0.73	0.92	0.54
5* (Fig. 5.7)	Raw	Narrow-band	0.89	0.76	0.78	0.58
6* (Fig. 5.8)	Low-pass	Narrow-band	0.96	0.88	0.92	0.75
9 (Fig. 5.13)	Raw	Narrow-band	0.52	0.26	0.13	-0.22
10 (Fig. 5.14)	Low-pass	N/A	0.39	0.30	0.13	-0.07
11 AR (Fig. 5.15)	Raw	N/A	0.71	0.49	0.51	0.26
12* AR (Fig. 5.16)	Low-pass	N/A	0.96	0.69	0.91	0.35
13 AR (Fig. 5.17)	Low-pass	N/A	0.44	0.27	0.17	-0.20
Extra (no Figures)						
*	Raw	A_3	0.81	0.52	0.67	0.31
	Low-pass	Narrow-band	0.39	0.24	-0.23	-0.34

*Forward-looking data available during input preprocessing.

It was hypothesized in 5.4 that effects of imperfect real-time preprocessing could be mitigated by a combination of filtering adjustments and an input correction algorithm. The edge effects were eased to some degree, and significantly at many points along the time series (Figure 5.12). Unfortunately, at other time steps, the edge effects extended beyond two peak periods and thus, beyond the capability of the correction algorithm to provide improved input. Real-time implementation failure was not limited to the NARX model. Low-pass filtered input that extended the AR forecast to 1.5 or 2 peak wave periods as seen in Case 12* and similarly cited in

Fusco, et al., (2010) was also susceptible to similar failures under real-time implementation (Case 13).

Another consideration concerns the circumstances with regard to training. In the present analysis, the networks were trained using ideal input (no edge effects), and then applied by attempting to approximate the ideal input by mitigating input edge effects. Conversely, if there were a way to train the networks on input exhibiting the same edge effects that they would encounter under practice, the preprocessing dilemma may be sidestepped. Still, it is not clear whether the imperfect input would retain its forecasting value.

While not included in the investigation into real-time implementation, two methods of input preprocessing based on wavelet transformations merit discussion. These methods held promise as alternatives with better potential for real-time implementation due to the finite nature of the mother wavelet. The first method in which wavelet approximations served as low-pass filtered input did not significantly boost forecast potential (Case 2*) when compared to Case 1. The second (Case 3*) relied on forecasting the wave record in the form of wavelet coefficients to be reconstructed post forecast, similar in method to the wave height prediction studies (2.6.2). In Case 3*, forecasts of wave elevation did demonstrate a performance more consistent with respect to forecast length. However, the forecast generally suffered in the early steps relative to corresponding forecasts using raw input (Case 1). Perhaps the forecasts were somewhat compromised in the down-sampling/up-sampling conversions necessary in dealing with DWT formats. Yet, it was also observed that the coefficient time series were not noticeably more predictable than the wave record itself.

The failure of wavelet transformations in boosting NARX performance was especially disappointing considering that the transforms showed potential for improvement in the single-step ANN forecasting of wave heights as described in 2.6.2. It was anticipated that this potential would translate to the multi-step ANN model that would provide a forecast profile appropriate for this application. The discrepancy may be attributed in part to the fact the models are inherently different. The single-step ANN models did not use feedback, allowing for input to be entirely

different from output. The previous wave height studies differed from the present study in other aspects as well. For instance, the time series were generally decomposed at higher levels (~6), and predictions for both approximations and details were used in the final reconstructions. It is also notable that a wave height and wave elevation time series are dissimilar in shape and time-scale.

Bibliography

Balas, C. E., Koç, L., & Balas, L. (2004). Predictions of Missing Wave Data by Recurrent Neuronets. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *130*(October), 256–265. doi:10.1061/(ASCE)0733-950X(2004)130:5(256)

Beale, M. H., Hagan, M. T., & Demuth, H. B. (2015). Neural Network Toolbox™ User's Guide.

Belmont, M. R., Christmas, J., Dannenberg, J., Hilmer, T., Duncan, J., Duncan, J. M., & Ferrier, B. (2014). An examination of the feasibility of linear deterministic sea wave prediction in multidirectional seas using wave profiling radar: Theory, simulation, and sea trials. *Journal of Atmospheric and Oceanic Technology*, *31*, 1601–1614. doi:10.1175/JTECH-D-13-00170.1

Belmont, M. R., Horwood, J. M. K., Thurley, R. W. F., & Baker, J. (2006). Filters for linear sea-wave prediction. *Ocean Engineering*, *33*, 2332–2351. doi:10.1016/j.oceaneng.2005.11.011

Blondel, E., Ducrozet, G., Bonnefoy, F., & Ferrant, P. (2008). Deterministic reconstruction and prediction of non-linear wave systems. *23rd Int. Workshop on Water Waves and Floating Bodies*.

Boren, B., Ling, B., & Batten, B. (2014). Using artificial neural networks for prompt and accurate wave prediciton at Northwest National Marine Renewable Energy center's north energy test site: A feasibility analysis. In *Grand Renewable Energy 2014 Proceedings*.

Brekken, T. K. A. (2011). On Model Predictive Control for a point absorber Wave Energy Converter. *2011 IEEE Trondheim PowerTech*, 1–8. doi:10.1109/PTC.2011.6019367

Brekken, T. K. A., Jouanne, a. Von, & Han, H. Y. H. H. Y. (2009). Ocean wave energy overview and research at Oregon State University. *2009 IEEE Power Electronics and Machines in Wind Applications*. doi:10.1109/PEMWA.2009.5208333

Chetouani, Y. (2008). Using ARX and NARX approaches for modeling and prediction of the process behavior: application to a reactor-exchanger. *Asia-Pacific Journal of Chemical Engineering*, *3*(6), 597–605. doi:10.1002/apj.118

Deka, P. C., & Prahlada, R. (2012). Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time. *Ocean Engineering*, *43*, 32–42. doi:10.1016/j.oceaneng.2012.01.017

Deo, M. C. (2010). Artificial neural networks in coastal and ocean engineering. *Indian Journal of Geo-Marine Science*, *39*(December), 589–596. Retrieved from http://nopr.niscair.res.in/handle/123456789/10807

Deo, M. C., Jha, A., Chaphekar, A. S., & Ravikant, K. (2001). Neural networks for wave forecasting, *28*(August 1999), 889–898.

Deo, M. C., & Sridhar Naidu, C. (1998). Real time wave forecasting using neural networks. *Ocean Engineering*, *26*(3), 191–203. doi:10.1016/S0029-8018(97)10025-7

Dixit, P., Londhe, S., & Dandawate, Y. (2015). Removing prediction lag in wave height forecasting using Neuro - Wavelet modeling technique. *Ocean Engineering*, *93*, 74–83. doi:10.1016/j.oceaneng.2014.10.009

Drew, B., Plummer, a. ., & Sahinkaya, M. . (2009). A review of wave energy converter technology, *223*, 887–902. doi:10.1243/09576509JPE782

Falnes, J., & Hals, J. (2012). Heaving buoys, point absorbers and arrays. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *370*(1959), 246–277. doi:10.1098/rsta.2011.0249

Flores, J. J., Graff, M., & Rodriguez, H. (2012). Evolutive design of ARMA and ANN models for time series forecasting. *Renewable Energy*, *44*, 225–230. doi:10.1016/j.renene.2012.01.084

Fusco, F., Gilloteaux, J. C., & Ringwood, J. (2010). A study on prediction requirements in time-domain control of wave energy converters? *IFAC Proceedings Volumes (IFAC-PapersOnline)*, *3*(1), 372–377. doi:10.3182/20100915-3-DE-3008.00075

Fusco, F., & Ringwood, J. (2010). A study on Short-Term Wave Forecasting for time-domain Control of Wave Energy Converters. *IEEE Transactions on Sustainable Energy*, *1*(2), 99–106.

Fusco, F., & Ringwood, J. V. (2012). A study of the prediction requirements in real-time control of wave energy converters. *Sustainable Energy, IEEE Transactions on*, *3*(1), 176–184. doi:10.1109/TSTE.2011.2170226

Fusco, F., & Ringwood, J. V. (2013). A Simple and Effective Real-Time Controller for Wave Energy Converters. *Sustainable Energy, IEEE Transactions on*. doi:10.1109/TSTE.2012.2196717

Fusco, F., & Ringwood, J. V. (2010). Short-term wave forecasting for real-time control of wave energy converters. *IEEE Transactions on Sustainable Energy*, *1*(2), 99–106. doi:10.1109/TSTE.2010.2047414

Haseltine, E. L., & Rawlings, J. B. (2005). Critical Evaluation of Extended Kalman Filtering and Moving-Horizon Estimation. *Industrial & Engineering Chemistry Research*, *44*(8), 2451–2460. doi:10.1021/ie034308l

Hatalis, K., Pradhan, P., Kishore, S., Blum, R. S., & Lamadrid, A. J. (2014). Multi-step forecasting of wave power using a nonlinear recurrent neural network. *PES General Meeting | Conference & Exposition, 2014 IEEE*. doi:10.1109/PESGM.2014.6939370

Hong, Y., Waters, R., Boström, C., Eriksson, M., Engström, J., & Leijon, M. (2014). Review on electrical control strategies for wave energy converting systems. *Renewable and Sustainable Energy Reviews*, *31*, 329–342. doi:10.1016/j.rser.2013.11.053

Huang, W., Murray, C., Kraus, N., & Rosati, J. (2003). Development of a regional neural network for coastal water level predictions. *Ocean Engineering*, *30*, 2275–2295. doi:10.1016/S0029-8018(03)00083-0

Hulthén, E. (2004). Improving Time Series Prediction using Recurrent Neural Networks and Evolutionary Algorithms.

Ibarra-Berastegi, G., Saénz, J., Esnaola, G., Ezcurra, A., & Ulazia, A. (2015). Short-term forecasting of the wave energy flux: Analogues, random forests, and physics-based models. *Ocean Engineering*, *104*, 530–539. doi:10.1016/j.oceaneng.2015.05.038

Khalil, R. A. (2011). Comparison of Four Neural Network Learning Methods Based on Genetic Algorithm for Non-linear Dynamic Systems Identification ةيبطخ لاا صخلم, 8–12.

Kim, H. S., Eykholt, R., & Salas, J. D. (1999). Nonlinear dynamics, delay times, and embedding windows. *Physica D: Nonlinear Phenomena*, *127*, 48–60. doi:10.1016/S0167-2789(98)00240-1

Kondratenko, V. V, & Kuperin, Y. a. (2003). Using Recurrent Neural Networks To Forecasting of Forex. *Time*, 23. Retrieved from http://arxiv.org/abs/cond-mat/0304469

Korde, U. a. (1999). Efficient primary energy conversion in irregular waves. *Ocean Engineering*, *26*, 625–651. doi:10.1016/S0029-8018(98)00017-1

Li, G., Weiss, G., Mueller, M., Townley, S., & Belmont, M. R. (2012). Wave energy converter control by wave prediction and dynamic programming. *Renewable Energy*, *48*, 392–403. doi:10.1016/j.renene.2012.05.003

Lin, T., Horne, B. G., & Giles, C. L. (1998). How embedded memory in recurrent neural network architectures helps learning long-term temporal dependencies. *Neural Networks*, *11*(5), 861–868. doi:10.1016/S0893-6080(98)00018-5

Ling, B. (2015). *Real-time Estimation and Prediction of Wave Excitation Forces for Wave Energy Control Applications*. Oregon State University.

Londhe, S. N., & Deo, M. C. (2004). Artificial Neural Networks for Wave Propagation. *Journal of Coastal Research*, *20*(4), 1061–1069. Retrieved from http://proxy.library.oregonstate.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=15223088&site=ehost-live

Makarynskyy, O., Pires-Silva, a. a., Makarynska, D., & Ventura-Soares, C. (2005). Artificial neural networks in wave predictions at the west coast of Portugal. *Computers and Geosciences*, *31*, 415–424. doi:10.1016/j.cageo.2004.10.005

Mallat, S. (2009). *A Wavelet Tour of Signal Processing: The Sparse Way* (3rd ed.). Burlington, MA: Academic Press. doi:2009

Mandal, S., & Prabaharan, N. (2006). Ocean wave forecasting using recurrent neural networks. *Ocean Engineering*, *33*(10), 1401–1410. doi:http://dx.doi.org/10.1016/j.oceaneng.2005.08.007

Mandal, S., & Prabaharan, N. (2010). Ocean Wave Prediction Using Numerical and Neural Network Models. *The Open Ocean Engineering Journal*, *3*(1), 12–17. doi:10.2174/1874835X01003010012

Marco P. Schoen, Jørgen Hals, and T. M. (2011). Wave Prediction and Robust Control. *IEEE TRANSACTIONS ON ENERGY CONVERSION*, *26*(2), 627–638. doi:JUNE 2011

Masters, T. (1995). *Advanced algorithms for neural networks : a C++ sourcebook*. New York: John Wiley & Sons, Ltd.

Misiti, M., & Misiti, Y. (1996). Wavelet toolbox. *The MathWorks Inc., ….* Retrieved from http://feihu.eng.ua.edu/NSF_TUES/w7_1a.pdf

Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J.-M. (2015). Wavelet Toolbox ™ User ’ s Guide. Natick, MA: The MathWorks, Inc. Retrieved from www.mathworks.com

Ochi, M. K. (2005). *Ocean Waves: The Stochastic Approach* (Cambridge .). Cambridge University Press. doi:2005

Özger, M. (2010). Significant wave height forecasting using wavelet fuzzy logic approach. *Ocean Engineering*, *37*(16), 1443–1451. doi:10.1016/j.oceaneng.2010.07.009

Paparella, F., Monk, K., Winands, V., Lopes, M. F. P., Conley, D., & Ringwood, J. V. (2015). Up-Wave and Autoregressive Methods for Short-Term Wave Forecasting for an Oscillating Water Column. *Sustainable Energy, IEEE Transactions on*. doi:10.1109/TSTE.2014.2360751

Prudêncio, R. (2003). Neural Network Hybrid Learning: Genetic Algorithms & Levenberg-Marquardt. *Of the 26th Annual Conference of*, (2). Retrieved from http://books.google.com/books?hl=en&lr=&id=Zm5MZ9w8VUEC&oi=fnd&pg=PA464&dq=Neural+Network+Hybrid+Learning+:+Genetic+Algorithms+&+Levenberg-Marquardt&ots=PA0EP-vmP3&sig=2NGwKFgGZ2i9kvZjKnOUNb_9_5s

Reikard, G. (2013). Integrating wave energy into the power grid: Simulation and forecasting. *Ocean Engineering*, *73*, 168–178. doi:10.1016/j.oceaneng.2013.08.005

Salami, M. J. E., Tijani, I. B., Abdullateef, a I., & Aibinu, M. a. (2013). Hybrid algorithm for NARX network parameters' determination using differential evolution and genetic algorithm. *IOP Conference Series: Materials Science and Engineering*, *53*, 012062. doi:10.1088/1757-899X/53/1/012062

Schoen, M. P., Hals, J., & Moan, T. (2011). Wave Prediction and Robust Control of Heaving Wave Energy Devices for Irregular Waves. *Energy Conversion, IEEE Transactions on*. doi:10.1109/TEC.2010.2101075

Sheng, W., & Lewis, A. (2001). Short-Term Prediction of an Artificial Neural Network In an Oscillating Water Column. *International Journal of Offshore and Polar Engineering*, *21*(4), 248–255. doi:December 2011

Shoori, J. H., Ling, B., & Batten, B. (2014). Use of Artificial Neural Networks for Real-Time Prediction of Heave Displacement in Ocean Buoys.

Simon, D. (2006). *Optimal State Estimation*. *Library* (Vol. 2). doi:10.1088/1741-2560/2/3/S07

Wijaya, a. P., Naaijen, P., Andonowati, & Groesen, E. Van. (2015). Reconstruction and future prediction of the sea surface from radar observations. *Ocean Engineering*, *106*, 261–270. doi:10.1016/j.oceaneng.2015.07.009

Wilamowski, B. M. (2003). Neural Network Architectures and Learning. *International Conference on Industrial Technology*, 1–12.