Near-shore swell estimation from a global wind-wave model: spectral process, linear, and artificial neural network models

Matthew Browne [*,1] Bruno Castelle [1] Darrell Strauss [1]

Rodger Tomlinson [1] Michael Blumenstein [2] Chris Lane [3]

Abstract

Estimation of swell conditions in coastal regions is important for a variety of public, government, and research applications. Driving a model of the near-shore wave transformation, from an offshore global swell model such as NOAA WaveWatch3, is an economical means to arrive at swell size estimates at particular locations of interest. Recently, some work (e.g. Browne et al. (2006)) has examined an artificial neural network (ANN) based, empirical approach to wave estimation. Here, we provide a comprehensive evaluation of two data driven approaches to estimating waves nearshore (linear and ANN), and also contrast these with a more traditional spectral wave simulation model (SWAN). Performance was assessed on data gathered from a total of 17 near-shore locations, with heterogenous geography and bathymetry, around the continent of Australia over a 7 month period. It was found that the ANNs out-performed SWAN and the non-linear architecture consistently out-performed the linear method. Variability in performance and differential performance with regard to geographical location could largely be explained in terms of the underlying complexity of the local wave transformation.

Preprint submitted to Elsevier Science *13 October 2006*

Key words: artificial neural networks, near-shore wave transformation, wave modeling, wave estimation

1 Introduction

Knowledge of swell conditions at specific nearshore locations is important for coastal research, marine engineering, and policy development. Although global swell models are an effective approximation of open swell conditions, they become less accurate in the nearshore zone. Many forms of remote sensing suffer from similar issues; satellite observations, for example, due to the manner in which they are sensed and averaged, normally pertain only to deeper ocean locations, typically 30km away from the the shoreline (Kalra et al., 2005). Variations in nearshore bathymetry, local wind-generated seas and the effects of artificial structures transform deep water swell due to reflection, shoaling, refraction, diffraction and breaking (Londhe and Deo, 2004). At a particular location, local topography may lead to attenuation or accentuation of long or short period swells, either directly, or by the contribution of local wind conditions. This makes the estimation of onshore wave heights, even given reliable offshore swell measurements, a non-trivial exercise.

The ability to estimate and predict onshore wave heights at the shore is of significant public interest in countries such as Australia, due to the high level of public activity in or near the break zone. There is a clear need for this

* corresponding author: m.browne@griffith.edu.au
[1] Griffith Centre for Coastal Management, Gold Coast campus, Griffith University PMB 50 Gold Coast Mail Centre QLD 9726
[2] School of Information and Communications Technology, Griffith University
[3] CoastalWatch Australia, Suite 3, 66 Appel Street, Surfers Paradise QLD 4217

data both as a general advisory to the public concerned with recreational use of the near-shore zone, and for life-guards concerned with providing advice in order to ensure public safety. Much of the Australian coastline is characterised by both high density structures, and highly dynamic sediment evolution, and monitoring of onshore wave heights is important for monitoring and managing coastal development. Finally, scientific and engineering evaluations of particular locations and structures in this zone depend on reliable data on wave action.

Currently, estimation of onshore wave heights from visual inspection is carried out on a regular basis, and this information is broadcast through private and commercial channels. The conventional approach is for this estimation to be performed and recorded manually by human observers, using a combination of heuristics, local knowledge, and offshore wind and wave models published by government agencies. Numerical models based on wave propagation theory have been neglected as a practical onshore wave height estimation tool due to a combination of factors: complexity of implementation, high amounts of processor time required, the need for accurate local bathymetric surveys, and general inaccuracy, even when the previous conditions are met. However, there is a clear need to substitute an objective and automated approach for human observations of onshore wave activity, which are expensive, time consuming, and prone to the usual forms of human failure and error.

In academic research, the propagation of swell in nearshore areas is conventionally studied by running either an actual, or a virtual simulated physical model (Londhe and Deo, 2004). Physical scale models require a significant investment of resources for their construction and simulation. For this reason, physical modeling using numerical computer simulation incorporating the lo-

cal physical environment and local swell conditions is often used. The Simulating Waves NearShore (SWAN) numerical model is an example of a popular approach for modeling wave propagation in the near-shore zone. However, numerical models themselves require care and expertise in their implementation. For example, a wave prediction system based on a numerical model must often contend with physical processes on a wide range of scales. These techniques are also sensitive to accuracy of the bathymetric data for the study area, and the quality of driving data at the model boundaries. Typically, hours of processing time is required for simulation of a region with adequate temporal and spatial resolution. Due to technical issues that may arise, results are sometimes unsatisfactory, especially in the break-zone.

Empirical wave recordings and observations are used to calibrate and validate theoretical (Booij et al., 1999a) and empirical (Komar and Gaughan, 1972; Caldwell, 2004) approaches to modelling the wave transformations that occur as they progress from deep to shallow water. The method of Caldwell (2004), for instance, was based on comparisons between buoy-measured H_s in deep water in close proximity to locations where visual observations were taken of trough to crest vertical wave height at the break point. There has been interest in developing empirical, data-driven models of nearshore wave characteristics for many years (Booij et al., 1999a). For example, an empirical method of obtaining surf height at the breakpoint from offshore wave data was derived from observations by Komar and Gaughan (1972). Following this approach, an alternative measure of surf height, H_{surf}, was developed for coastal zones with narrow shelves, steep bottom slopes and high refraction,

$$H_{\text{surf}} = H_b K_r(H_b), \tag{1}$$

where K_r is the empirical estimation of the refraction coefficient as a function of shoaling and buoy-estimated breaker height, H_b:

$$K_r = -0.0013{H_b}^2 + 0.1262H_b + 0.3025. \quad (2)$$

The method is based on comparisons between buoy-measured H_s in deep water in close proximity to locations where visual observations were taken of trough to crest vertical wave height at the break point. We note the particular operational definitions of wave height in the break zone (e.g. tough-crest height, assessed visually): it should be recognised that it is usually impossible to measure breaking wave height using a fixed wave gauge, since the spatial point of breaking varies with incident wave conditions. Estimating near-shore wave height through modelling the shallow-water wave transformation via simplified equations that are optimised using real-world data may be considered a 'semi-empirical' approach.

In their application to analysis in the physical sciences, artificial neural networks (ANNs) can be regarded as strongly empirical, as opposed to model-based approaches to estimate and predict wave behaviour (Deo and Jagdale, 2003). ANNs are a flexible learning architecture which rely on the presentation of input and target data, rather than a theoretical model, for the estimation of an underlying physical relationship. As general purpose function approximators, ANNs purposely impose no constraints on the final model generated, although the size of the neural network necessarily limits the overall complexity of the modelled function. In their role of associating temporally, spatially, or modally distinct measurements of wind-wave activity, they must approximate the physical propagation of wave energy. The effect of local geography or bathymetry is inferred from the co-variation of input-target pairs, rather

than explicitly determined. Thus, a representative corpus of training data is essential for the function approximation potential of an ANN to be realised.

The use of ANNs has been reported for numerous applications in the geological and marine sciences, and in particular have been used for forecasting wave climate time series (Deo et al., 2001; Agrawal and Deo, 2002). Tsai and Lee (1999) utilised neural networks for forecasting tidal variability and Tsai et al. (1999) used neural networks for forecasting wave heights at near-shore locations, using measurements from other locations as input, finding that using multiple sites as input increased the accuracy of predictions. Some work has tested ANNs for specific oceanic structural and engineering tasks: (Mase and Kitano, 1999) used feed-forward networks to estimate wave force impact on a marine structure, and (Mase et al., 1995) found that a similar ANN architecture accurately predicted damage levels on a breakwater resulting from wave action. ANNs have recently shown to produce superior estimates of wave spectra from wave parameters than those provided by theoretical or statistical predictions (Naithani and Deo, in print, 2005). Non-linear empirical models have been shown to approximate well the underlying wave physics (Tolman et al., 2005).

ANNs have been applied to estimating missing wave buoy data (Balas et al., 2004), and recently Kalra et al. (2005) have detailed an ANN-based effort to map offshore wave data to coastal locations, reporting superior performance of ANNs compared to a linear statistical approach. The authors estimated wave activity in the near-shore region from satellite monitoring at offshore locations. Scotto and Soares (2000) concluded that in estimating significant wave height of sea-states using non-linear autoregressive (AR) models, linear models satisfactorily modelled the lower statistical moments, but non-linear

models better approximated the higher moments such as skewness and kurtosis. Also using an AR model, Ho and Yim (2005), recently demonstrated the feasibility of interpolating missing buoy data between two wave measuring stations.

Neural networks have been demonstrated to be superior to conventional approaches for forecasting significant wave height in open water, in a variety of situations (Deo et al., 2002, 2001; Deo and Kumar, 2000; Balas et al., 2004). Rao and Mandal (2005) focused on ANNs as an alternative to numerical modelling for estimating wave-fields generated by cyclone events. Browne et al. (2006) emphasised the use of ANNS for bridging of modes of observation (i.e. global model output / logged buoy data / visual observations) in the context of bringing offshore estimates to estimate activity in the near-shore zone. It has been shown that non-linear ANN approaches outperformed linear statistical and numerical modelling, in the estimation of both human observations at the beach, and wave-rider buoy data.

Apart from the welcome comparison of ANN performance with that of linear regression by Kalra et al. (2005), there have been insufficient comparisons of ANNs with other forms of swell estimation, such as linear predictors, and numerical modelling. Work reported by Browne et al. (2006) represented a first step towards a systematic comparison between ANNs, statistical, and numerical modelling approaches. However, this work needs to be extended to include a greater variety of geographical areas, and longer study time periods.

As neural networks are unconstrained general-purpose function approximators, with potentially thousands of degrees of freedom, care must be taken in their application in order to ensure that the underlying function is in fact well

approximated, allowing in turn for good generalisation to new data. *Over-fitting*, which results from a high ratio of model degrees-of-freedom to presented data, is a particularly common pitfall for engineering and scientific applications of ANNs. These issues have led to concerns being raised regarding the validity of previous work in applying ANNs to wave estimation. For example, a method of wave forecasting using neural networks was recently reported by Makarynskyy (2004), and subsequently challenged by Medina (2005) due to issues related to over-fitting, lack of baseline performance comparisons, and an insufficient degree of cross-validation. Inappropriate application of ANNs often leads to spuriously high performance estimates. It must be also noted that as a strongly empirical approach, ANNs do not provide the insight into wave propagation processes that is provided by full-scale numerical modelling. However, the advantages include computational efficiency and potentially greater predictive power, without the need for detailed geographic information, or the laborious testing of a range of physical model parameters.

Recent research is increasingly focusing on the use of neural networks in the role of an unconstrained empirical function approximation tool for estimating the relationship between geographically or temporally displaced observations of wave height. However, to-date little attention has been paid to applying them to approximate the deep-to-shallow water wave transformation. The current paper is focused on mapping offshore global wind-wave observations to activity observed at particular onshore locations. This requires the model to incorporate the effects of a number of physical processes such as bottom friction, diffraction and refraction, generated by an interaction of local geography and bathymetry with the offshore wave field. Recent studies by Kalra et al. (2005) and Browne et al. (2006), have reported superior performance

by ANNs. However, each study considered only a single geographical location for a relatively short time, which prevented strong conclusions regarding the relative efficacy of ANNs to be drawn.

The present study attempts definitive testing of ANNs for modeling the nearshore wave transformation: bringing global ocean wave model output to nearshore locations, and demonstrating a potentially useful tool for emulating expensive surf reporter observations. A comprehensive evaluation of empirical methods is attempted by considering a total of 17 onshore locations across 5 geographical regions distributed across the continent of Australia, for a period of 8 months. ANNs are compared with baseline, linear and model-based approaches and explanations for the differential performance are provided. Detail is provided on the technical implementation and validation of ANN performance, which as discussed above, is critical for establishing an accurate benchmark of performance. The SWAN numerical model is applied for performance comparisons, and in order to investigate the characteristics of the study areas.

2 Method

2.1 Study Regions

As noted above, in order to achieve a comprehensive evaluation of spectral, linear, and ANN based modeling, the present study considered five regions in Australia, each with distinct properties, in both in terms of bathymetry and wave climate. This is illustrated in the regional maps (Figs 2,3 & 4). The red filled circles denote the onshore locations considered in this study, while the red stars indicate the theoretical location of the NOAA WW3 grid point.

Figure 2 maps the bathymetry of each region, which shows a range of profiles, from open, exposed beaches, to sheltered bays, along with a range of island and headland features. Figure 3 plots the mean wave height, as estimated by the SWAN model from the NOAA data over the time period of interest: this may be considered in conjunction with table 2 to provide a view of regional and intra-regional wave climate variability. Finally, figure 4 shows the degree of linear relationship (correlation r) of the SWAN model output at each point within each region, to that region's NOAA WW3 grid point, over the study period. Lighter values indicate a smaller proportion of the variability being explained by a linear relationship with offshore conditions, and hence indicate a greater proportion of non-linear wave effects in the SWAN model output at that point. We consider this to be a useful guide to the model-estimated complexity in the wave-transformation over each locale with each study region.

2.2 Data

In the following sections we describe the various forms of data used in this study; bathymetric surveys used as input to the numerical physical model, driving variables from NOAA WaveWatch 3 (NWW3), and visual surf reporter estimates of H_s.

2.2.1 NOAA WaveWatch 3

The driving variables for all models tested in this study were drawn from global wave model data gathered from the NWW3 model at 6 grid locations in various regions off the coast of Australia from 12/01/05 to 9/8/05. The NWW3 model provided updates four times daily, at 4am, 12pm (midday), 6pm, and

12am (midnight). The NWW3 variables were; significant wave height, primary swell direction, primary swell period, wind direction, wind speed, secondary swell direction, secondary swell period, wind wave direction, and wind wave period. NWW3 generates swell forecasts at every three hours from +3 to +180 hours ahead of the current time as well as a single analysis at 0 hours, which represents the model state, given current measured data. In the present study, only the 0 hour state information was used. For the purpose of presenting the NWW3 variables to the ANNs for empirical prediction, variables in degree format present an issue because of the discontinuity at the $0^{\circ}/360^{\circ}$ point. Each of the directional variables; primary swell direction, wind direction and wave direction were transformed to Cartesian co-ordinates ('northerliness D^N and 'easterliness' D^E):

$$D^N = cos\Big(\frac{\pi D}{180}\Big), D^E = sin\Big(\frac{\pi D}{180}\Big), \tag{3}$$

The locations of the NWW3 grid points are shown in table 2. Table 2 displays the cross-correlation matrix between H_s at the various grid points. As expected, there are correlations between nearby grid points; i.e. between Queensland (QLD) and North New South Wales (NNSW) regions, and between the South Australia (SA) and Victorian (VIC) grid points. However, 11 of the 15 cross-correlations are below 0.3, indicating that overall, the swell conditions around the continent had a significant degree of statistical independence during the study period. This entails that to a large extent, although the regions were considered over the same time period, the study regions represent independent sources of testing the estimation methods. The univariate statistics of the NWW3 estimated H_s in table 2 also illustrate the heterogeneous nature of the offshore swell conditions around Australia. Southern areas (i.e. VIC,

SA and Western Australia (WA)) had higher and more variable seas, whilst NNSW and QLD were characterised by greater skewness in H_s, primarily due to the large contribution that intermittent tropical storm events make to the overall variability of H_s in these areas.

2.2.2 Bathymetric Data

Bathymetric data is a key parameter to undertake nearshore wave modelling. Bathymetric data were required for 5 areas of the Australian Coast from deep water to shallow water in order to compute the wave propagation from the NWW3 output grid points to the surf reporter locations. The publicly available Geoscience Australia bathymetric database (www.ga.gov.au) was used to generate the whole bathymetry of 4 areas.

The database contains data around the Australian margin since 1963 from a variety of systems and levels of accuracy. A total of 931 surveys are currently registered, whose extent is 34N - 79S, 90 - 180E. The typical point data spacing is 25-200 m. Approximately 20% of these surveys were acquired by Geoscience Australia, the other component being surveys from other institutions, such as oil exploration companies and the National Geophysical Data Centre to which various institutions have contributed. Swath bathymetry in deep water, laser airborne depth sounder (LADS) data, points digitised from Australian Hydrographic Service charts on the shelf and predicted bathymetry from satellite altimetry (Smith and Sandwell, 1997) have been brought together. The database grid was computed at a cell size of 0.01 (36" or 1111 metres) as a compromise between conveying detail and limiting the file size.

For the Gold Coast area, accurate bathymetry data were provided by the Gold

Coast City Council. ETA lines covering the Gold Coast from the Gold Coast Seaway to Tweed Heads and refined bathymetric surveys around Burleigh Heads and Currumbin Creek estuary were combined with the Geoscience Australia database for the offshore information (typically used from 20 to 80 m depth).

2.2.3 Surf Reporter Data

In collaboration with the Australian surf monitoring firm CoastalWatch^{TM}, the study relied on a network of professional surf reporters for performing daily visual estimates of significant wave height at each of the 17 beach locations. Estimates were made using a mobile internet link and time-stamped. Visual inspection was usually performed once per day, between 6am and 8am, but during dynamic periods a second visual estimate was made in the afternoon. Table 1 displays the number of observations and number of days for which observations were recorded for each of the 17 beaches. As the surf reports were the limiting source of data for optimizing the models, this table also summarizes the number of input / target data pairs available for analysis. Table 3 displays the univariate statistics for the surf reporter observations at each beach study site. As expected, both overall significant wave height and variability are significantly less than the open ocean estimates produced by NWW3. With reference to the regional maps and the NWW3 statistics, it may be seen that there is variability between the various study areas corresponding to both offshore swell characteristics between regions, and local geographic features. At certain nearby beaches, where local geography and bathymetry is relatively homogeneous, there is a high degree of similarity between the univariate statistics of the beaches. This is supported by table 4, which displays

the normalised bivariate correlations in H_s between the surf reporter sites. The Sydney, New South Wales (SNSW) beaches in particular, which are grouped in a relatively small region in and around Sydney, and similar exposure to offshore swell, have a high degree of inter-dependence. In general however, there is wide variability in the univariate moments due to the different swell conditions prevailing in the different continental locations, and considerable variety in local geography and bathymetry.

2.3 Analysis and Simulation

2.3.1 ANN Background

ANNs are widely accepted as a valuable tool for modeling, approximation, and classification (Kostanic (2001); Bishop (1995); Ripley (1996)). The common fully inter-connected feed-forward architecture implements a mapping $\mathbf{y} = f(\mathbf{x}) : \Re^m \rightarrow \Re^n$ and is optimized by providing multiple (assumed noisy) paired samples of the input and target output $\{\mathbf{i}_p \in \Re^m, \mathbf{t}_p \in \Re^n\}$. The transfer function g, which generates a unit's output given net activation from connections to units in the previous layer, should generally be smooth and have a well bounded range for any input; e.g. $g : (-\infty, \infty) \rightarrow (-1, 1)$ for the tan-sigmoidal case (logistic and gaussian functions are also common basis functions). Assuming the activation function at each layer $i = \{1, ..., L\}$ is homogenous, an ANN implements the function

$$f(\mathbf{x}) = f(\mathbf{a}_0) = h_L(h_{L-1}(h_{L-...}(h_1(\mathbf{a}_0)))) \tag{4}$$

with the layer transformation h defined by

$$\mathbf{a}_i = h_i(\mathbf{a}_{i-1}) = g_i\Big(\mathbf{W}_i\Big[\mathbf{a}_{i-1}^T\ 1\Big]\Big) \tag{5}$$

where the set of free parameters (termed weights) in the system $\{\mathbf{W}_i\}$ determine the particular non-linear mapping, noting that $\mathbf{a}_0 = \mathbf{x}$ is an m element vector, $\mathbf{a}_L = \mathbf{y}$ an n element vector, and necessarily the dimensions $\{d_i^1, d_i^2\}$ of $\mathbf{W}_i$ have the constraints $d_1^1 = m + 1$, $d_L^2 = n + 1$, and $d_i^1 = d_{i-1}^2$. This notation represents the constant or *bias* term as a unit input appended to the output vector of the previous layer. ANNs are usually conceptualised as a series of neural layers, with forward interconnections between subsequent layers, as shown in figure 1.

A common application of feed-forward networks for estimation involves a fixed architecture or topology, with two or three layers L, and an arbitrary number of neurons (defined by d_i^2) in each layer. Training a neural network usually involves minimising an error function (e.g. for mean-square error (MSE), $\epsilon_n = \sum_p (\mathbf{y}_p - \mathbf{t}_p)^2$), utilising local gradient search algorithms operating on $-\delta\epsilon_n/\delta\{\mathbf{W}_i\}$, the derivative of the error function with respect to the free weight parameters in the network. Sophisticated and efficient search algorithms, such as the Levenburg-Marquardt method or conjugate gradient descent (Marquardt, 1963; Kan and Timmer, 1989), along with modern computational resources, allow for fast optimisation of medium sized networks. Single layer, or linear feed-forward networks converge to a global optimum. The function-approximation power of non-linear ANNs with one or more hidden layers is based on the non-linearity of the basis functions in the hidden layers. However, this property also entails the presence of *local minima* in the error function. ANN implementation requires acknowledgement that optimi-

sation based on local gradients may be expected to yield solutions located in some form of local minima: i.e. good but not optimal solutions. Multiple runs with random initial $\{\mathbf{W}_i\}$ are a straight-forward way to alleiviate this issue. Further, since ANN architectures can involve a large number of free parameters, *over-fitting* of the training data is common: this must be taken into account both in model optimization, and in estimating model effectiveness.

2.3.2 ANN Implementation

In the present study, the basic feed-forward ANN architecture was used to implement three empirical estimates of surf reporter readings of significant wave height, H_s^{sr}. In each case, the 9 NWW3 parameters (given in 2.2.1) were used as inputs to the model, with the four directional angle variables transformed to 2D Cartesian co-ordinates on the unit circle, leading to a total of 13 inputs. Firstly, a simple linear scaling of NWW3 derived H_s, H_s^N was implemented:

$$\widehat{H}_s^{sr} = w_1 H_s^N + w_0 \tag{6}$$

with the model weights w determined empirically. Because of the extreme simplicity of this model, we refer to this as the baseline-scaled estimate. In order to reduce the number of input variables, and hence the number of free weights in the networks, principle components analysis (PCA) was applied as a pre-processing step [4]. Using a form of scree plot, it was decided that retaining the first 6 orthogonal projections of the data retained a reasonable level ($>$ 92%) of total (normalised) variability in the NWW3 data. Given the vector

[4] As PCA is a standard data pre-processing technique for reducing data dimensionality, it will not be discussed. For more details the reader is directed to Joliffe (2002).

of normalised NWW3 driving variables $\mathbf{n}$, and the 13x6 PCA transformation matrix $\mathbf{P}$, the linear empirical model is given by

$$\widehat{H}_s^{sr} = \langle \mathbf{w}^T, \left[\mathbf{n}^T\mathbf{P}\ 1\right]^T \rangle \tag{7}$$

where $\langle\cdot\rangle$ denotes the inner product operation, and the 6-element weight vector $\mathbf{w}$ is optimized with respect to the training data. All feed-forward ANN and non-ANN model optimization was done using the Levenburg-Marqhardt (LM) algorithm implemented in MATLAB. Like other iterative optimization algorithms, the LM method finds parameter values that minimize the sum of squares using local gradient information of the objective function. It is a more robust form of Gauss-Newton algorithm (which utilizes first derivative information in estimate updates).

The final empirical model is a non-linear feed-forward neural network utilising tan-sigmoid activation functions

$$h(x) = \frac{2}{1+e^{-2x}} - 1 \tag{8}$$

in the hidden layer, and linear activation in the output layer. The non-linear ANN model used a single hidden layer of 6 neurons, the equation being:

$$\widehat{H}_s^{sr} = \left\langle \mathbf{w}_2^T, \left[h\Big(\left[\mathbf{n}^T\mathbf{P}\ 1\right]\mathbf{W}_1\Big)\ 1\right] \right\rangle \tag{9}$$

with the 6x7 input-to-hidden connection matrix $\mathbf{W}_1$ and the 1x7 hidden-to-output weight vector $\mathbf{w}_2^T$ being determined empirically.

The multi-layer ANNs used in the study were purposely made as small as possible in order to reduce the potential for over-fitting the data: the non-linear architecture involved 49 free weights. It was assumed that a reasonable

approximation to the shallow-water transformation for a single location would not require an overly complex model. The available data for the surf reporters at each onshore location varied from 800 to 900 data points. The ANNs were trained using a 5-fold combined early-stopping and cross-validation method. That is, 80% of the data was used for optimising the free weights in the model, with 20% used for cross-validation of the data, and early stopping of training. This approach is intended to prevent over-fitting of the training data, and for generating reliable estimates of performance on unseen data. For the non-linear model given in (9), each partition of the data was used to train 10 ANNs with identical topology, but different random weight initialisations, and the best performing neural net used in each case. Finally the *ensemble* ANN output was calculated by averaging the output of the 10 ANNs upon presentation of the entire data set. The ensemble method is known to improve performance and generalisation of ANNs, but the performance in this case is known to be an overestimate, as the performance calculation is necessarily performed on the combined training and test data subsets for all the trained neural networks. Therefore, in tables 5 and 6, the validation performance column may be treated as a conservative estimate of future ensemble ANN performance, whilst the reported ANN ensemble performance is rather more optimistic.

2.3.3 SWAN Background

SWAN (version 40.41) is a spectral wave model based on the action density balance equation (Hasselmann et al., 1973) that describes the evolution of two-dimensional wave energy spectra under specified conditions of winds, currents, and bathymetry (Booij et al., 1999b; Ris et al., 1998).

SWAN can be used on any scale, even if this model is specifically designed for coastal applications. This model requires no restriction on wave approach angle or directional width. SWAN is able to simulate accurately the wave field in coastal areas where reflection and diffraction are not significant. The SWAN Cycle III User Manual by Holthuijsen et al. (2002) provides a detailed account of the theoretical background, program structures, and implementation.

SWAN modeling has been used successfully for storm-induced coastal flooding assessment applications (Cheung et al., 2003), to drive nearshore wave-induced current nearshore models (Castelle et al., 2005), for wind generated waves in lakes (Jin and Z.-G., 2001) and in coastal regions (Ou et al., 2002; Castelle et al., 2006), and to model the evolution of wave spectra in a wave tank (Wood et al., 2001).

2.3.4 SWAN implementation

In the present study, SWAN is used in stationary mode. The model considers a steady state situation that requires the time of propagation of the waves through the domain to be short compared to the variation of water level, currents and changes in offshore wave conditions. Triad interaction is taken into account in the computations. The breaking wave model chosen herein is the bore-based model of Battjes and Janssen (1978), with a constant breaker parameter $\gamma = 0.73$ following Battjes and Stive (1985). The wave forcing provided by the NWW3 nearest output point is to the offshore and lateral boundaries of the model. For Western Australia, South Australia, Victoria and New South Wales areas, the computational grid is concurrent with the Geoscience Australia grid i.e. a regular grid at a cell size of 0.01. For the

Queensland area, a curvilinear grid is used for computations with grid cell size of $O(100$ m). During the simulations, the wave information is requested at each surf reporter location. The outputs are given in 10 m depth in order to avoid an underestimation of wave height due to wave breaking during high energy conditions. The tide level is treated as constant equal to 0 in the Admiralty Height Datum (AHD), i.e. at mid tide. Stationary computations are done every 6 hours, concurrent with the NWW3 output data.

3 Results

Table 5 presents the performance of the various empirical models in terms of mean absolute error (MAE) for the 17 surf reporter locations, organised according to study area. A single NOAA WW3 grid point served as input to models for all the surf reporter locations in each study area. For ease of presentation, the MAEs are provided in cm rather than metres. The first column displays the difference between the raw significant wave heights H^s recorded at the offshore NOAA grid point and the surf reporter observations. This baseline difference ranged from approximately 0.6m at the Sydney, New South Wales beaches, to approximately 2.4m at Trigg Beach, WA and Seaford, SA. The next column displays the MAE after linear scaling of offshore H^s using eq. (6). At all locations, linear scaling resulted in a significant decrease in the MAE. This baseline measure represents the error after linear attenuation of wave energy is accounted for. Because of the insignificant degrees of freedom for this model (i.e. 1), division of the data into training and validation subsets was not performed. For the linear and non-linear ANNs (eq.(7) and (9)), data was split into training and validation and test sets, and the results of both

are shown. The incremental improvement of the linear and non-linear ANNs varied over the test sites. The ensemble ANN performance is also shown. Two sample t-tests (N=5) were performed in order to explicitly test the hypothesis that the non-linear ANN MAE was less than the linear version. Despite relatively small incremental improvements in performance in some cases, the t-statistics were significant at the .01 probability level for all locations except for Warranambool, VIC.

Table 6 shows similar results for the normalised correlation coefficient R between empirical model estimates and surf reporter observations. The R coefficient provides a more sensitive measure of estimation performance. For consistency, raw and scaled correlations are included, although they are identical due to the fact that R is normalised with respect to univariate variance, and linear scaling therefore has no effect of the observed R. Here the improvement is estimation performance over the baseline is more apparent, and so too is the differential improvement over the various locations. Despite the greater degrees of freedom in the non-linear versus linear models, generalisation performance, as measured by the validation columns was consistently better for the non-linear networks. This is confirmed by the t-statistics, which indicate significant improvement in performance in almost every case.

Figures 5-9 compare graphically the output of of the best performing non-linear neural network with the surf reporter observations. A time series over the study time period is shown on the left hand side of each plot. The two error-bar plots on the right hand side summarise the relationship between either H^s offshore baseline (middle plot) or ANN network estimated H^s (right plot). In both cases, the diagonal line represents H^s as reported by the human observer, whilst the mean and standard error at each wave height is summarised by the

error bars.

The full page illustrations provided in figure 2 shows the bathymetric profiles for each of the study regions for which bathymetry was available.

4 Discussion

We first consider the effectiveness of SWAN modeling for nearshore wave estimation. For each area, there is a significant improvement of the wave estimation in terms of the Mean Standard Error (MAE) at the surf-reporter locations in comparison with the NOAA outputs. Thus, the overall degree of wave bottom friction over the continental shelf and wave energy shadowing behind headlands and islands appears to be correctly estimated. However, the results are generally not comparable with either the linear or non-linear empirical models.

The effectiveness of SWAN modeling is, not surprisingly, more effective when accurate bathymetry are available and a more refined computational grid is used. In the Queensland area, for which this was the case, the correlation factor significantly improves in comparison with NOAA output data (from 0.49 to 0.72), as does the Mean Standard Error (0.94 m to 0.32 m). This area is exposed to moderate to high energy southerly to south-easterly swell during the winter period. During these wave conditions, the surf reporter location is partially sheltered by the Coolangatta headlands. This sheltering and resulting wave refraction patterns are mostly responsible for the significant improvement of the correlation factor.

For the other areas, SWAN performance in terms of the correlation statis-

tics is significantly worse in comparison to the ANN estimators, and even the baseline. This can be explained by many factors. The main concern is the bathymetric data, and consequently the coarse computational grid. In the Queensland region, comparison of Geoscience Australia bathymetric data and bathymetric surveys provided by the Gold Coast City Council showed that errors of the order of sometimes a few meters could be measured. The main processes affecting the wave propagation exist in the nearshore zone, usually between 30-40 m depth. For example, in the New South Wales area, there is significant longshore variability of the seabed between 0 and 50 m depth which are poorly revealed by Geoscience Australia database. Mid-scale processes such as refraction are poorly simulated by SWAN in the present study. Furthermore, for areas with a large and shallow continental shelf such as the Western Australia area, calibration of bottom friction is a key parameter for an accurate estimation of wave height to the shore. Finally, one of the main reasons why SWAN results are not in very good agreement with surf reporter wave height estimation is the wave forcing format. Indeed, significant wave height, mean period and mean wave angle are provided by NWW3 outputs in our areas of interest. Directional wave spectra, which were not available in the present study, are necessary for optimal performance of model-based estimation techniques.

The ANN based empirical method used, on the other hand, does not depend on careful adjustment of physical parameters (such as bottom friction) and handles gracefully sub-optimal input data (i.e. swell parameters in lieu of the complete directional spectrum). It was expected that the non-linear ANN architecture would perform better than the linear models, and this was borne out by the results. Tables 4 and 5 show that non-linear networks consistently

out-performed linear networks over the study regions both in terms of standard error, and correlation with the surf reporter targets. This need for a non-linear model was in accordance with predictions, as a non-linear model was expected to be necessary to take into account the interaction of variable offshore forcing conditions and complex local bathymetries. From figures 2-4, it is clear that a significant proportion of the shared variability in offshore and onshore observed swell heights is modulated by non-linear physical processes in the form of refraction, wave interactions, and local wind effects.

The simultaneous prediction of near-shore wave heights over a large number of study regions, along with the use of SWAN modeling of wave propagation in the region (figures 2-4) allows us to consider the reasons for differential prediction performance in each study area. The Queensland region, for instance, is subject to variability in swell direction, biased towards the south-east. For these beaches, the relationship between offshore and onshore swell height is moderated strongly by swell direction, with more southerly swells experiencing a higher degree of sheltering and refraction. In the case of Seaford SA, both linear and non-linear ANNs failed to markedly improve onshore swell estimation. This may be explained by the fact that offshore and onshore activity is highly decoupled: the highly dynamic offshore wave climate has limited propagation to the onshore site due to sheltering. Poor performance at this site also results from the inherent difficulty in monitoring small changes in wave height, and the fact that a higher proportion of this variability is due to unpredictable effects such as localized winds. The utility of non-linear estimation was more apparent at Chiton Rocks SA, which may be explained by its more exposed aspect, yielding a more consistent relationship between offshore and onshore activity, which the empirical models were able to emulate. A non-linear model

was more effective because it was able to take into account differential attenuation due to swell direction. The empirical estimation methods were less useful on Sydney (SNSW) beaches, where the baseline relationship between NOAA and the surf-reporters was very high. The lack of complexity in the offshore-onshore wave transformation meant that there was little further improvement that the linear or non-linear models could make on the baseline prediction. The South Coast NSW study site was a slight exception to this rule, that we explain by its greater distance from the NOAA grid point (it is located further south, off-map). The best performance overall, in terms of ANN improvement over baseline, was observed at Margaret River and Trigg beach, WA. This is consistent with the general relationship observed between local site characteristics and the improvement of estimation performance over baseline. That is, non-linear ANNs are most effective for estimation when there is a systematic yet complex coupling of offshore and onshore wave climates. Best performance is observed when local transformative processes such as shoaling, attenuation and refraction moderate, but do not completely decouple, the onshore from the offshore wave climate. ANNs tended to over-estimate smaller swells. We believe this is due to the fact that the ANN models assume a Gaussian error distribution. However, this is unrealistic since; a) it is impossible to observe negative wave heights, and b) the wave height distribution is positively skewed. This could be addressed in future work by utilising a zero-truncated, or quasi-binomial distribution.

5 Conclusions

The methodologies considered here have immediate application for near-shore wave height estimation. This is significant because the vast majority of human activity occurs in the near-shore zone, and swell conditions in this area are therefore of greatest interest to coastal managers and the public. When combined with a near-shore model, existent global wave models such NWW3 can provide a reliable and cost-effective source of offshore data in absence of *in situ* measurements. Compared with linear and spectral modeling, this study has concluded that near-shore conditions can be inferred from WW3 parameters most effectively using ANN-based estimation. As well as explicit and implicit modeling of the near-shore wave transformation, the paper establishes quite strong relationships between a open ocean global wind-wave model and onshore visual estimates of wave height provided by experts. The empirical approach presented here relies on the availability of a corpus of target data for model training: the use of a nation-wide database of expert surf reporter records is a unique characteristic of this study. The practical application rests on the ability to replace the manual ratings, with the automatic estimates generated by the ANN model. The high correlations and relatively small standard-errors obtained by the ANN model on the *validation* data set indicates that 6-12 months of daily observations is sufficient to build a model that generalizes well.

References

Agrawal, J. D., Deo, M., 2002. Online wave prediction. Marine Structures 15, 57–74.

Balas, C. E., Koc, L., Balas, L., Sept/Oct 2004. Predictions of missing wave data by recurrent neuronets. Journal of Waterway, Port, Coastal and Ocean Engineering, 256–265.

Battjes, J., Janssen, J., 1978. Energy loss and set-up due to breaking of random waves. In: Proceeding of the 16^{th} International Conference on Coast. Eng. ASCE, pp. 569–587.

Battjes, J., Stive, M., 1985. Calibration and verification of a dissipation model for random breaking waves. J. Geophys. Res. 90 (C5), 9159–9167.

Bishop, C., 1995. Neural Networks for Pattern Recognition. Clarendon Press.

Booij, N., Ris, R., Holthuijsen, L., 1999a. A third-generation wave model for coastal regions. 1. model description and validation. J. of Geophys. Res. 104 (7), 649–666.

Booij, N., Ris, R., Holthuijsen, L., 1999b. A third-generation wave model for coastal regions, part i: Model description and validation. J. Geophys. Res. 104 (C4), 7649–7666.

Browne, M., Strauss, D., Castelle, B., Lane, C., Blumenstein, M., 2006. Empirical estimation of nearshore waves from a global deep-water wave model. IEEE Geoscience and Remote Sensing Letters, in print, to appear 2006.

Caldwell, P. C., 2004. An empirical method for estimating surf heights from deep water significant wave heights and peak periods in coastal zones with narrow shelves, steep bottom slopes, and high refraction. In: Proceedings of the 8th International Workshop on Wave Hindcasting and Forecasting. North Shore, Oahu, Hawaii. November 14 -19.

Castelle, B., Bonneton, P., SÈnÈchal, N., Dupuis, H., Butel, R., Michel, D., 2005. Dynamics of wave-induced currents over an alongshore non-uniform multiple-barred sandy beach on the aquitanian coast, france. Cont. Shelf Res.In press.

Castelle, B., Cope, M., Abbs, D., Tomlinson, R., 2006. Wave modelling for the gold coast: down-sizing from global to nearshore. In: 13th National Conference: Climate, Water and Sustainability, Newcastle, Feb 6-8. The Australian Meteorological and Oceanographic Society.

Cheung, K., Phadke, A., Wei, Y., Rojas, R. D.-M., Martino, C., Houston, S., Liu, P.-F., Lynett, P., Dodd, N., Liao, S., Nakasaki, E., 2003. Modeling of storm-induced coastal flooding for emergency management. Ocean Eng. 30, 1353–1386.

Deo, M., Godane, D., Kumar, V. S., 2002. Analysis of wave directional spreading using neural networks. Journal of Waterways, Port and Coastal Engineering 128 (1), 30–37.

Deo, M., Kumar, N. K., 2000. Interpolation of wave heights. Ocean Engineering 27, 907–919.

Deo, M. C., Jagdale, S. S., 2003. Prediction of breaking waves with neural networks. Ocean Engineering, 30, 1163–1178.

Deo, M. C., Jha, A., Chaphekar, A. S., Ravikant, K., 2001. Neural networks for wave forecasting. Ocean Engineering 28 (7), 889–898.

Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H., Hasselmann, D., Kruseman, P., Meerburg, A., Miller, P., Olbers, D., Richter, K., Sell, W., Walden, H., 1973. Measurements of wind-wave growth and swell decay during the joint north sea wave project (jonswap). Dtsch. Hydrogr. Z. 12 (A8).

Ho, P., Yim, J. Z., 2005. A study of the data transferability between two wave-measuring stations. Coastal Engineering 52, 313–329.

Holthuijsen, S., Booij, N., Ris, N. Haagsma, I., Kieftenburg, A., Kriezi, E., 2002. Swan cycle iii version 40.11 user manual. Delft University of Technology, The Netherlands.

Jin, K.-R., Z.-G., J., 2001. Calibration and verification of a spectral wind-wave model for lake okeechobee. Ocean Eng. 28, 571–584.

Joliffe, I., 2002. Principal Component Analysis. Springer.

Kalra, R., Deo, M., Kumar, R., Agarwal, V. K., 2005. Artificial neural network to translate offshore satellite wave data to coastal locations. Ocean Engineering, In Press, corrected proof available online 27 June.

Kan, A. R., Timmer, G., 1989. Global Optimization: A Survey, International Series of Numerical Mathematics. Basel: Birkhauser Verlag.

Komar, P. D., Gaughan, M. K., 1972. Airy wave theory and breaker height prediction. In: Proc. of the 13th Conf. on Coastal Eng. Vol. ASCE. pp. 405–418.

Kostanic, F. M. H. . I., 2001. Principles of Neurocomputing for Science and Engineering. McGraw Hill.

Londhe, S. N., Deo, M. C., 2004. Artificial neural networks for wave propagation. Journal of Coastal Research 20 (4), 1061–1070.

Makarynskyy, O., 2004. Improving wave predictions with artificial neural networks. Ocean Engineering 31 (5-6), 907–724.

Marquardt, D., 1963. An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math. 11, 431–441.

Mase, H., Kitano, T., 1999. Prediction model for occurence of impact wave force. Ocean Engineering 26, 949–961.

Mase, H., Sakamoto, M., Sakai, T., 1995. Neural network for stability analysis of rubble mound breakwaters. Journal of Waterway, Coastal and Ocean Engineering 121 (6), 294–299.

Medina, J. R., 2005. Improving wave predictions with artificial neural networks, by o. makarynskyy. Ocean Engineering 32 (1), 101–103.

Naithani, R., Deo, M. C., in print, 2005. Estimation of wave spectral shapes

683 using ann. Advances in Engineering Software.
684 Ou, S.-H., Liau, J.-M., Hsu, T.-W., Tzang, S.-Y., 2002. Simulating typhoon
685 waves by swan wave model in coastal waters of taiwan. Ocean Eng. 29,
686 947–971.
687 Rao, S., Mandal, S., 2005. Hindcasting of storm waves using neural networks.
688 Ocean Engineering 32, 667–684.
689 Ripley, B. D., 1996. Pattern Classification and Neural Networks. Cambridge.
690 Ris, R., Booij, N., Holthuijsen, L., 1998. A third-generation wave model for
691 coastal regions, part ii: verification. J. Geophys. Res. 104 (C4), 7649–7666.
692 Scotto, M., Soares, C. G., 2000. Modelling the long-term time series of sig-
693 nificant wave height with non-linear threshold models. Coastal Engineering
694 313-327.
695 Smith, W., Sandwell, D., 1997. Global seafloor topography from satellite al-
696 timetery and ship depth soundings. Science 277, 1956–1962.
697 Tolman, H. L., Krasnopolsky, V. M., Chalikov, D. V., 2005. Neural network
698 approximations for nonlinear interactions in wind wave spectra: direct map-
699 ping for wind seas in deep water. Ocean Modelling 8, 253–278.
700 Tsai, C., Shen, J. N., Kerh, T., 1999. Wave forecasting using neural network
701 model. In: Kumar, Topping (Eds.), Artificial intelligence applications in civil
702 and structural engineering. Civil-Comp. Press, pp. 123–130.
703 Tsai, C. P., Lee, T. L., 1999. Back-propagation neural network in tidal-level
704 forecasting. Journal of Waterway, Coastal, and Ocean Engineering 125 (4),
705 195–202.
706 Wood, D., Muttray, M., Oumeraci, H., 2001. The swan model used to study
707 wave evolution in a flume. Ocean Eng. 28, 805–823.

Table 1
Counts of surf reporter observations and total number of days monitored from 12/01/05 to 9/8/05. Observations were recorded more than once / day during more dynamic surf conditions.

Region	Beach	Observations	Days
QLD	Surfer's Paradise	231	201
	Sunshine Coast	228	203
WA	Margaret River	229	202
	Perth	223	204
SA	Seaford	222	206
	Chiton Rocks	210	206
NNSW	Mid Coast	361	206
	North Coast	216	206
SNSW	Bondi	374	200
	Cronulla	362	204
	Manly	377	207
	Palm Beach	330	200
	South Coast	236	200
VIC	Woolamai	275	200
	Portsea	268	200
	Torquay	305	206
	Warrombool	243	200
		(N)	(N)

Table 2
Cross-correlation matrix and univariate mean and higher moments of significant wave heights H_s recorded at NOAA WW3 grid points from 12/01/05 to 9/8/05.

	NNSW	QLD	SA	SNSW	VIC	WA
North New South Wales (153.75 -31)	-					
Queensland (153.75, -28)	**.75**	-				
South Australia (136.25,-36)	-.23	-.24	-			
South New South Wales (151.25,-34)	**.80**	**.49**	-.24	-		
Victoria (143.75,-39)	-.16	-.23	**.77**	-.17	-	
Western Australia (115.00,-34)	-.02	.12	.21	-.11	-.02	-
mean $\overline{H}_s$	1.95	1.89	2.91	1.47	2.31	3.06
standard deviation $\sigma^2(H_s)$	0.85	0.65	1.01	0.78	0.84	1.20
skewness $\sigma^3(H_s)$	1.48	0.84	0.65	1.87	0.98	1.06
kurtosis $\sigma^4(H_s)$	6.02	4.26	3.02	8.32	3.88	4.23
(metres)						

Table 3
Mean and higher moments of surf reporter significant wave height H_s observations (in metres) from 12/01/05 to 9/8/05.

Region	Beach	$\overline{H}_s$	$\sigma^2(H_s)$	$\sigma^3(H_s)$	$\sigma^4(H_s)$
QLD	Surfer's Paradise	0.96	0.35	0.0076	0.0475
	Sunshine Coast	0.91	0.33	0.0086	0.0540
WA	Margaret River	2.02	0.66	0.0009	0.0177
	Perth	0.60	0.28	0.0092	0.0354
SA	Seaford	0.45	0.22	0.0068	0.0324
	Chiton Rocks	1.37	0.31	0.0003	0.0231
NNSW	Mid Coast	0.66	0.43	0.0128	0.0419
	North Coast	1.03	0.33	0.0116	0.0627
SNSW	Bondi	0.88	0.38	0.0132	0.0592
	Cronulla	0.88	0.39	0.0139	0.0608
	Manly	0.88	0.38	0.0126	0.0577
	Palm Beach	0.89	0.38	0.0128	0.0570
	South Coast	0.87	0.45	0.0081	0.0349
VIC	Woolamai	1.40	0.44	0.0002	0.0259
	Portsea	1.46	0.46	0.0034	0.0344
	Torquay	0.85	0.39	0.0085	0.0366
	Warranambool	0.89	0.41	0.0066	0.0360

Table 4
Cross-correlation matrix r of significant wave heights H_s at NOAA WW3 grid points from 12/01/05 to 9/8/05.

		Queensland		Western Aus.		South Aus.		N^{th} NSW		S^{th} NSW					Victoria			
		SP	SC	MR	PE	SF	CR	MC	NC	BI	CR	MY	PB	SC	WO	PO	TO	WL
QLD	Surfer's Paradise SP	-																
	Sunshine Coast SC	**.96**	-															
WA	Margaret River MR	.13	.08	-														
	Perth PE	.12	.09	**.89**	-													
SA	Seaford SF	.15	.15	.12	.14	-												
	Chiton Rocks CR	.12	.12	-.10	-.09	**.45**	-											
NNSW	Mid Coast MC	**.41**	**.37**	-.01	-.03	-.11	-.05	-										
	North Coast NC	**.77**	**.71**	.27	.24	.05	.03	**.39**	-									
SNSW	Bondi BI	.20	.11	.04	.01	-.31	-.23	**.65**	**.36**	-								
	Cronulla CR	.16	.07	.01	-.01	-.32	-.23	**.64**	**.33**	**.98**	-							
	Manly MY	.21	.11	.03	.01	-.31	-.23	**.65**	**.37**	**.99**	**.97**	-						
	Palm Beach PB	.21	.12	.04	.02	-.30	-.22	**.66**	**.36**	**.99**	**.97**	**.99**	-					
	South Coast SC	.12	.05	.06	.07	-.20	-.23	**.59**	**.37**	**.74**	**.76**	**.74**	**.73**	-				
VIC	Woolamai WO	-.01	.01	-.20	-.22	.30	**.62**	-.10	-.07	-.12	-.12	-.12	-.12	-.18	-			
	Portsea PO	-.01	-.01	-.13	-.19	.27	**.59**	-.13	-.06	-.10	-.11	-.11	-.10	-.17	**.91**	-		
	Torquay TO	-.00	.00	-.21	-.25	.15	**.67**	-.04	-.02	-.06	-.06	-.08	-.07	-.10	**.83**	**.83**	-	
	Warranambool WL	-.08	-.08	-.24	-.30	.10	**.54**	-.01	-.07	-.02	-.01	-.04	-.03	-.07	**.71**	**.69**	**.79**	-

boldface indicates correlation $r > .30$

Table 5
Mean absolute error (MAE) in cm of NOAA WW3 derived significant height estimates with respect to surf reporter observations

Region	Beach	Baseline		Linear ANN				Non-linear ANN				Ensemble		SWAN
				Training		Validation		Training		Validation				
		Raw	Scaled											
		MAE	MAE	$\overline{MAE}$	$\sigma_{(MAE)}$	$\overline{MAE}$	$\sigma_{(MAE)}$	$\overline{MAE}$	$\sigma_{(MAE)}$	$\overline{MAE}$	$\sigma_{(MAE)}$	MAE	t^{**}	
QLD	Surfer's Paradise	91.77	23.30	17.70	*0.12*	17.84	*1.32*	14.93	*0.96*	16.12	*1.05*	13.60	3.22*	32.31
	Sunshine Coast	96.33	22.78	16.75	*0.07*	16.77	*0.73*	14.17	*0.93*	15.51	*0.96*	12.40	3.32*	
WA	Margaret River	102.89	36.64	34.20	*0.17*	34.74	*1.30*	23.75	*1.46*	26.69	*1.64*	21.34	12.17*	65.51
	Trigg Beach	244.11	13.32	13.46	*0.13*	13.80	*0.55*	11.15	*0.67*	12.05	*0.89*	10.16	5.25*	62.77
SA	Seaford	245.64	13.82	14.28	*0.06*	14.49	*0.98*	11.92	*0.48*	12.83	*0.96*	11.12	3.83*	15.20
	Chiton Rocks	152.84	21.15	20.56	*0.04*	20.64	*0.52*	16.17	*0.76*	17.42	*0.76*	15.12	11.02*	40.93
NNSW	Mid Coast	127.92	22.91	23.42	*0.06*	22.97	*1.19*	18.65	*1.03*	19.62	*1.13*	17.19	6.46*	
	North Coast	90.93	21.41	20.28	*0.22*	19.98	*0.67*	17.13	*1.35*	18.15	*1.17*	15.08	4.30*	
SNSW	Bondi	59.54	19.47	21.64	*0.13*	21.73	*1.31*	18.12	*0.52*	19.41	*1.04*	16.96	4.38*	
	Cronulla	60.26	19.94	22.22	*0.15*	21.85	*1.09*	18.76	*0.75*	20.25	*0.93*	17.26	3.53*	
	Manly	60.18	19.86	21.71	*0.09*	21.40	*1.50*	18.41	*0.65*	19.49	*1.19*	17.33	3.15*	50.06
	Palm Beach	58.97	19.72	21.96	*0.19*	22.26	*0.88*	17.69	*0.65*	19.75	*1.34*	16.42	4.95*	
	South Coast	64.73	28.06	26.97	*0.11*	27.45	*1.33*	22.14	*1.70*	24.20	*1.53*	20.56	5.06*	
VIC	Woolamai	89.94	29.62	27.88	*0.09*	28.30	*1.05*	23.93	*0.86*	25.64	*1.47*	22.70	4.67*	38.72
	Portsea	83.22	28.81	28.61	*0.14*	28.72	*1.44*	24.49	*0.89*	26.61	*1.78*	22.62	2.92*	
	Torquay	143.57	24.18	23.73	*0.13*	24.34	*1.15*	19.16	*1.00*	21.03	*1.45*	17.83	5.66*	34.98
	Warranambool	141.49	27.93	26.00	*0.11*	25.96	*1.07*	23.90	*0.72*	25.29	*1.23*	22.78	1.30	

$*$ indicates 1-tailed t-statistic significant (N=5, $p < 0.01$) $**$ t-statistic: two sample, one-tailed t-test testing $\mathbf{H}(\overline{MAE}_{nl}^{val} < \overline{MAE}_{lin}^{val})$

Table 6
Correlation R of NOAA WW3 derived surf height estimates with surf reporter observations

Region	Beach	Baseline		Linear ANN				Non-linear ANN				Ensemble		SWAN
		Raw	Scaled	Training		Validation		Training		Validation			$\mathbf{H}(\overline{r}_{nl}^{val} > \overline{r}_{lin}^{val})$	
		$\overline{r}$	$\overline{r}$	$\overline{r}$	$\sigma(r)$	$\overline{r}$	$\sigma(r)$	$\overline{r}$	$\sigma(r)$	$\overline{r}$	$\sigma(r)$	$\overline{r}$	t-statistic	
QLD	Surfer's Paradise	.49	.49	.73	.008	.71	.036	.83	.027	.79	.037	.86	4.84*	.72
	Sunshine Coast	.46	.46	.72	.008	.71	.026	.83	.025	.80	.029	.87	7.42*	
WA	Margaret River	.77	.77	.79	.005	.78	.018	.88	.011	.86	.022	.90	8.23*	.73
	Trigg Beach	.79	.79	.80	.015	.80	.028	.87	.018	.85	.023	.89	4.82*	.72
SA	Seaford	.67	.67	.63	.030	.61	.045	.75	.020	.70	.027	.78	5.81*	.63
	Chiton Rocks	.58	.58	.62	.008	.64	.029	.77	.021	.74	.028	.81	8.63*	.40
NNSW	Mid Coast	.72	.72	.74	.004	.73	.040	.83	.021	.80	.039	.86	3.90*	
	North Coast	.34	.34	.52	.014	.52	.042	.74	.058	.68	.070	.81	6.28*	
SNSW	Bondi	.81	.81	.75	.005	.76	.019	.84	.0093	.82	.021	.86	6.86*	
	Cronulla	.80	.80	.75	.005	.75	.026	.84	.011	.79	.016	.87	4.63*	
	Manly	.80	.80	.75	.004	.75	.036	.84	.0093	.81	.025	.86	4.55*	.76
	Palm Beach	.81	.81	.75	.006	.75	.035	.85	.011	.81	.036	.87	4.15*	
	South Coast	.69	.69	.72	.006	.70	.047	.80	.029	.76	.032	.84	3.16*	
VIC	Woolamai	.67	.67	.69	.003	.68	.033	.78	.018	.74	.034	.80	3.97*	.62
	Portsea	.70	.70	.69	.007	.69	.029	.78	.015	.74	.030	.81	3.53*	
	Torquay	.68	.68	.70	.007	.71	.016	.81	.019	.78	.030	.84	7.35*	.70
	Warranambool	.59	.59	.63	.004	.64	.040	.71	.022	.69	.031	.75	2.91*	

$*$ indicates 1-tailed t-statistic significant (N=5, $p < .01$)

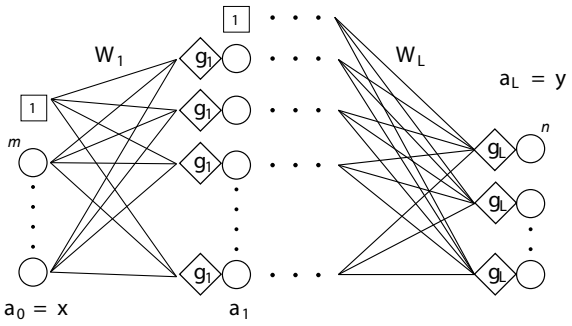


Fig. 1.

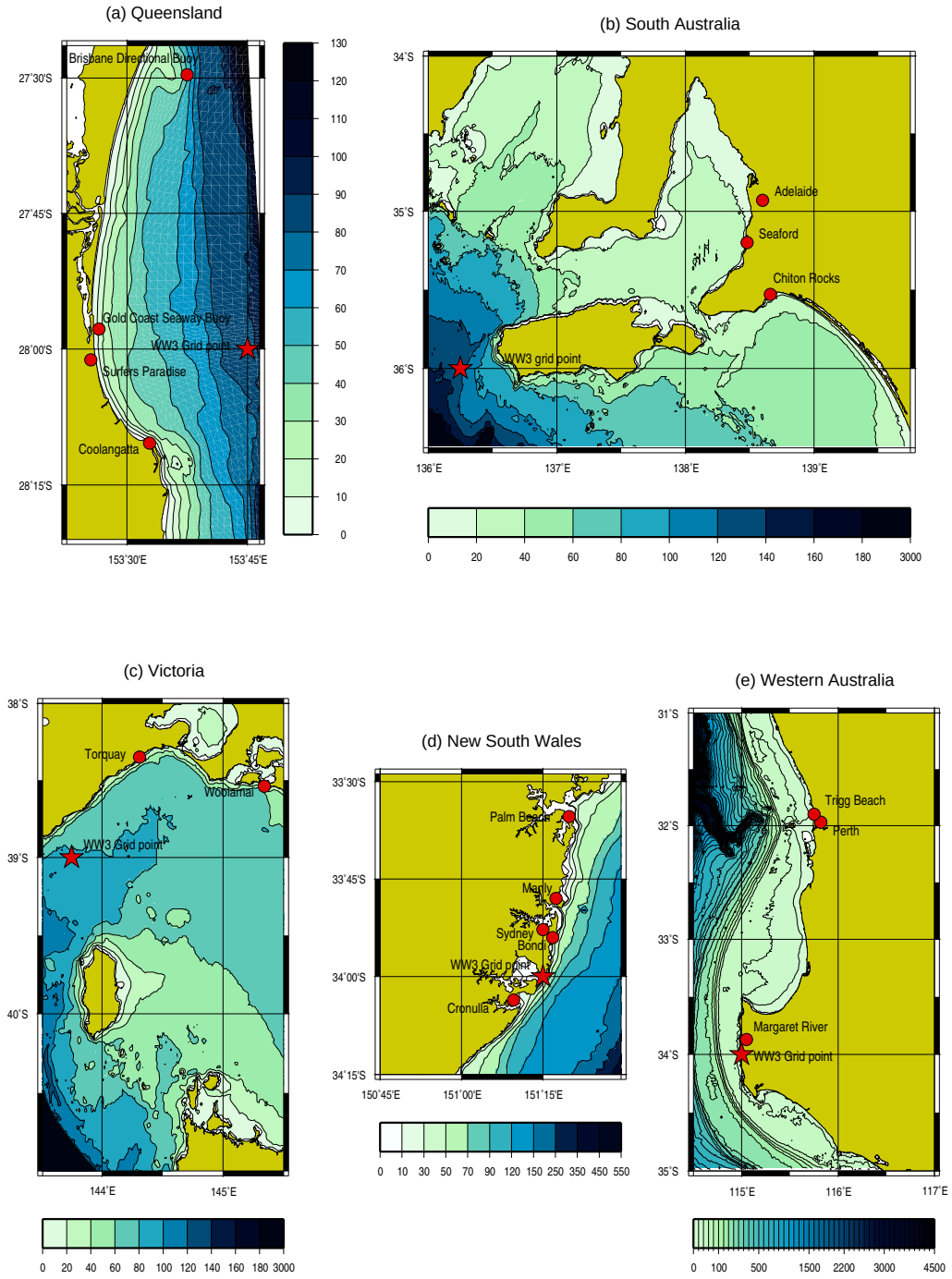

(a) Queensland
Brisbane Directional Buoy
Gold Coast Seaway Buoy
WW3 Grid point
Surfers Paradise
Coolangatta
27°30'S
27°45'S
28°00'S
28°15'S
153°30'E
153°45'E
130
120
110
100
90
80
70
60
50
40
30
20
10
0
(b) South Australia
Adelaide
Seaford
Chiton Rocks
WW3 grid point
34°S
35°S
36°S
136°E
137°E
138°E
139°E
0 20 40 60 80 100 120 140 160 180 3000
(c) Victoria
Torquay
Wobiarnal
WW3 Grid point
38°S
39°S
40°S
144°E
145°E
0 20 40 60 80 100 120 140 160 180 3000
(d) New South Wales
Palm Beach
Manly
Sydney
Bondi
WW3 Grid point
Cronulla
33°30'S
33°45'S
34°00'S
34°15'S
150°45'E
151°00'E
151°15'E
0 10 30 50 70 90 120 150 250 350 450 550
(e) Western Australia
Trigg Beach
Perth
Margaret River
WW3 Grid point
31°S
32°S
33°S
34°S
35°S
115°E
116°E
117°E
0 100 500 1500 2200 3000 4500

Fig. 2.

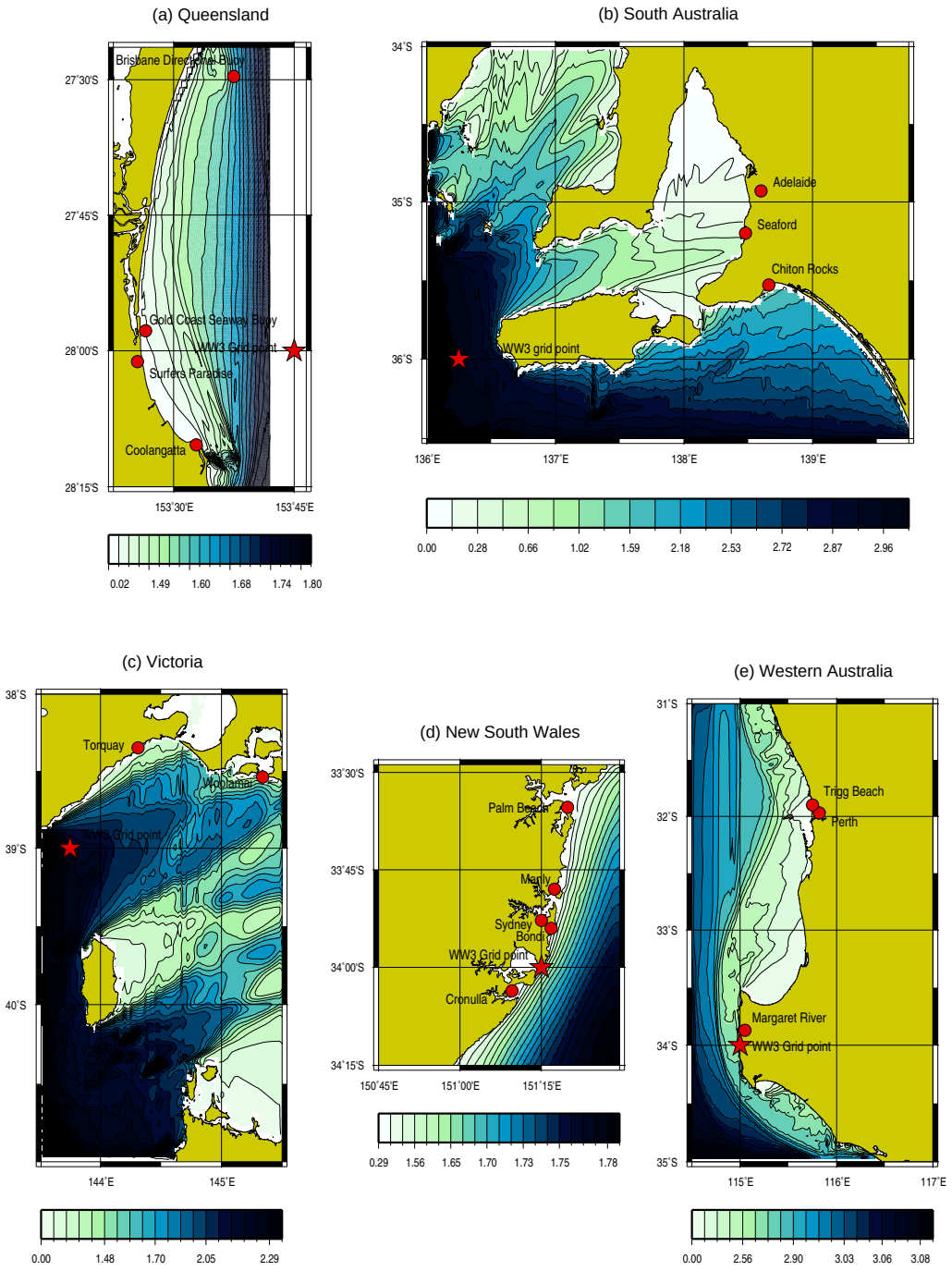

(a) Queensland
Brisbane Directional Buoy
Gold Coast Seaway Buoy
WW3 Grid point
Surfers Paradise
Coolangatta
27°30'S
27°45'S
28°00'S
28°15'S
153°30'E
153°45'E
0.02 1.49 1.60 1.68 1.74 1.80
(b) South Australia
Adelaide
Seaford
Chiton Rocks
WW3 grid point
34°S
35°S
36°S
136°E
137°E
138°E
139°E
0.00 0.28 0.66 1.02 1.59 2.18 2.53 2.72 2.87 2.96
(c) Victoria
Torquay
Wollamai
WW3 Grid point
38°S
39°S
40°S
144°E
145°E
0.00 1.48 1.70 2.05 2.29
(d) New South Wales
Palm Beach
Manly
Sydney
Bondi
WW3 Grid point
Cronulla
33°30'S
33°45'S
34°00'S
34°15'S
150°45'E
151°00'E
151°15'E
0.29 1.56 1.65 1.70 1.73 1.75 1.78
(e) Western Australia
Trigg Beach
Perth
Margaret River
WW3 Grid point
31°S
32°S
33°S
34°S
35°S
115°E
116°E
117°E
0.00 2.56 2.90 3.03 3.06 3.08

Fig. 3.

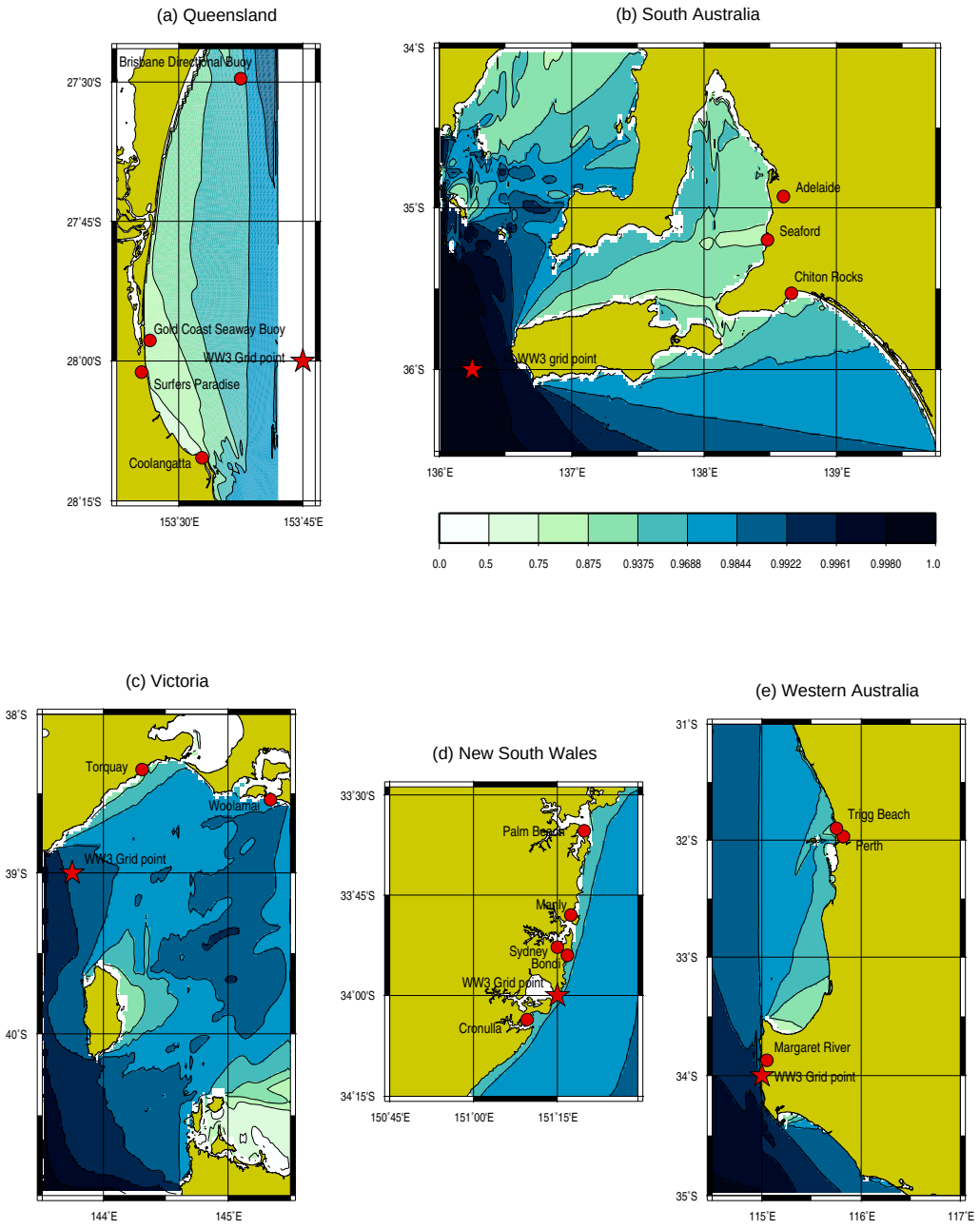

(a) Queensland
Brisbane Directional Buoy
Gold Coast Seaway Buoy
WW3 Grid point
Surfers Paradise
Coolangatta
27°30'S
27°45'S
28°00'S
28°15'S
153°30'E
153°45'E
(b) South Australia
Adelaide
Seaford
Chiton Rocks
WW3 grid point
34°S
35°S
36°S
136°E
137°E
138°E
139°E
0.0
0.5
0.75
0.875
0.9375
0.9688
0.9844
0.9922
0.9961
0.9980
1.0
(c) Victoria
Torquay
Wonthaggi
WW3 Grid point
38°S
39°S
40°S
144°E
145°E
(d) New South Wales
Palm Beach
Manly
Sydney
Bondi
WW3 Grid point
Cronulla
33°30'S
33°45'S
34°00'S
34°15'S
150°45'E
151°00'E
151°15'E
(e) Western Australia
Trigg Beach
Perth
Margaret River
WW3 Grid point
31°S
32°S
33°S
34°S
35°S
115°E
116°E
117°E

Fig. 4.

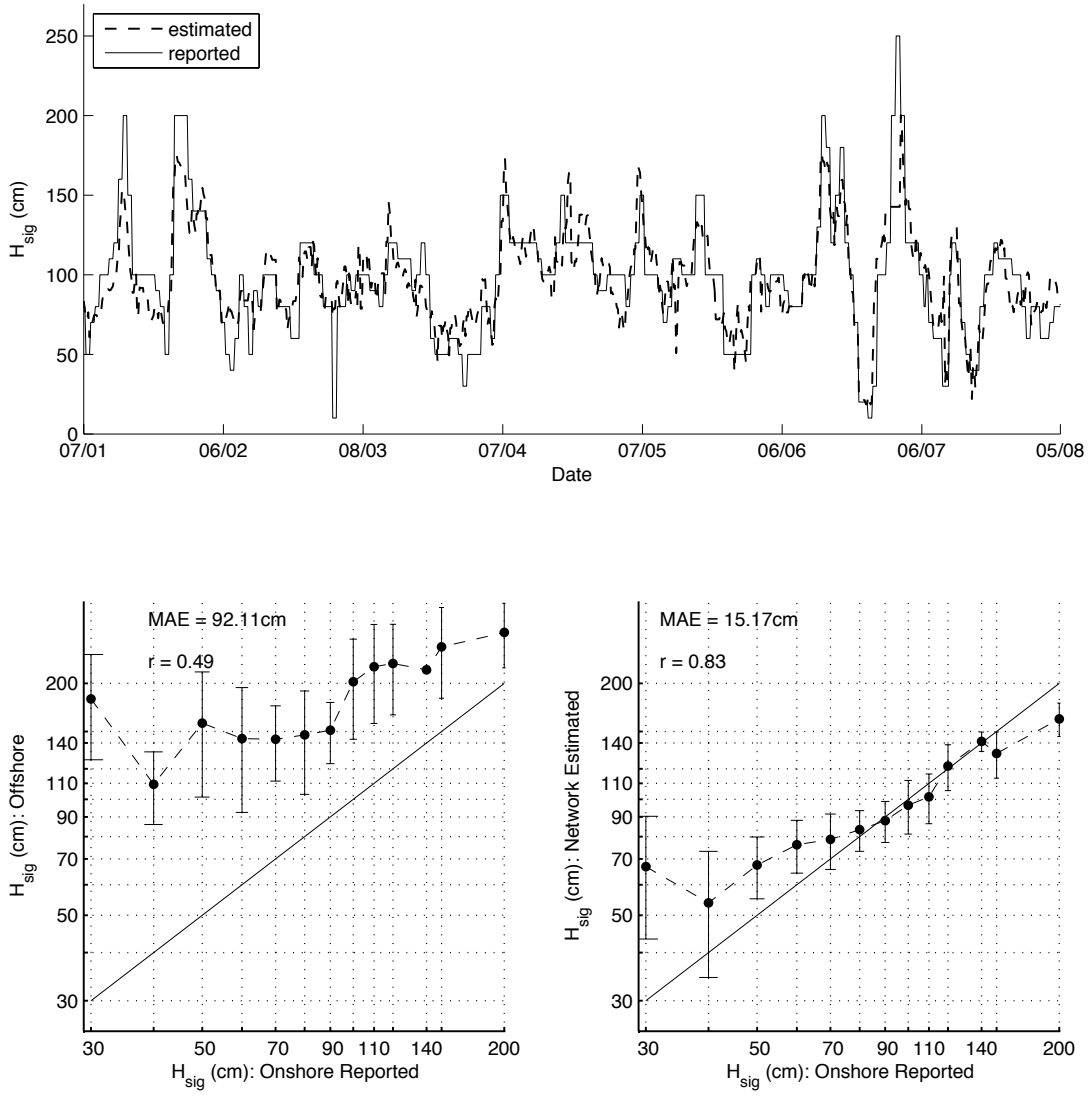
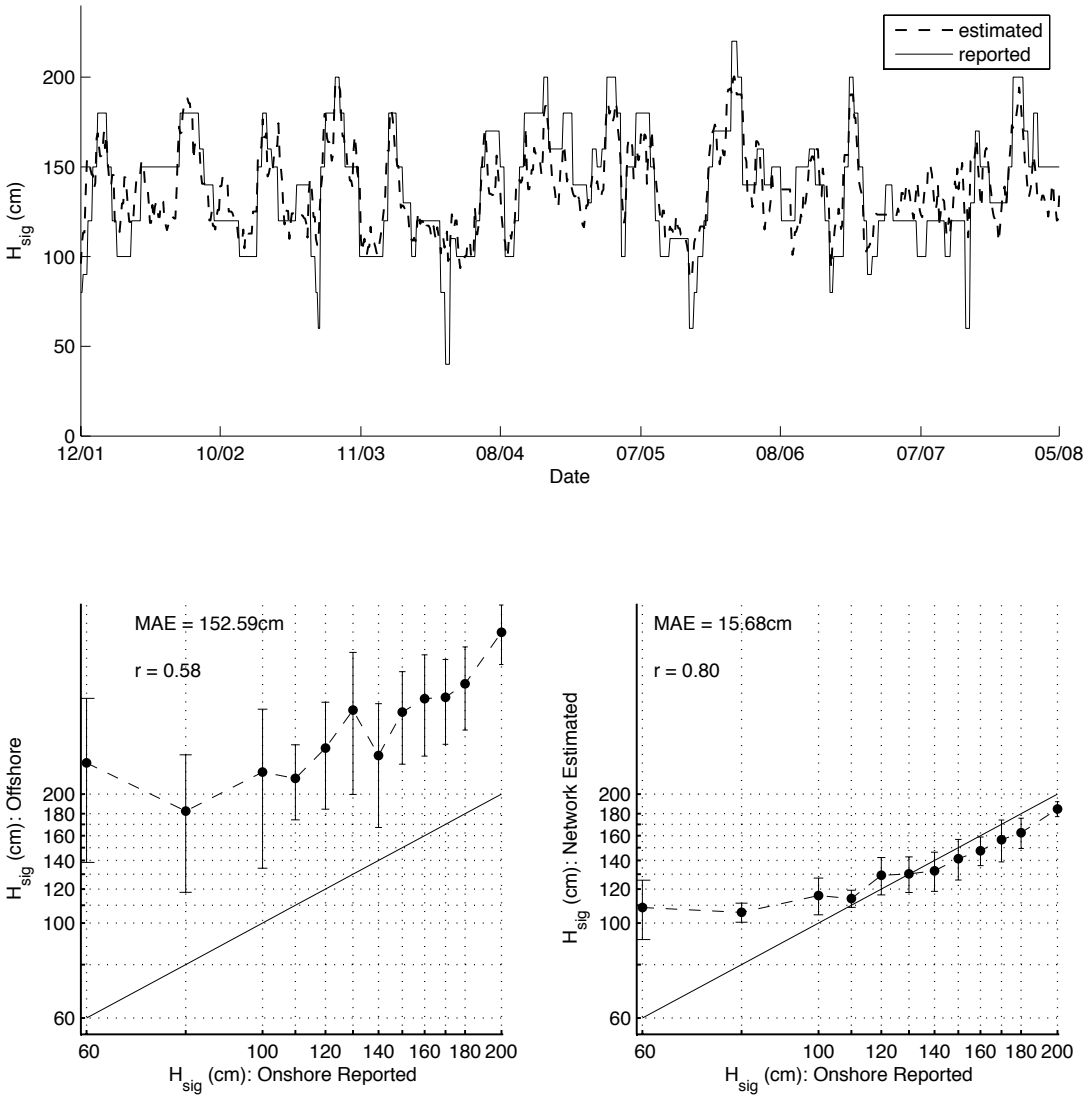
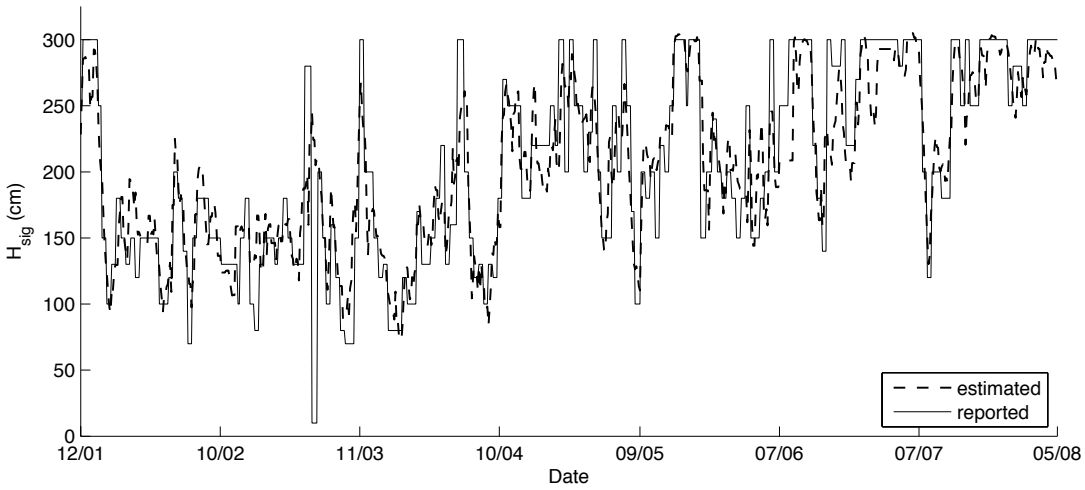

estimated
reported
H_sig (cm)
Date
07/01
06/02
08/03
07/04
07/05
06/06
06/07
05/08
MAE = 92.11cm
r = 0.49
H_sig (cm): Offshore
H_sig (cm): Onshore Reported
MAE = 15.17cm
r = 0.83
H_sig (cm): Network Estimated
H_sig (cm): Onshore Reported

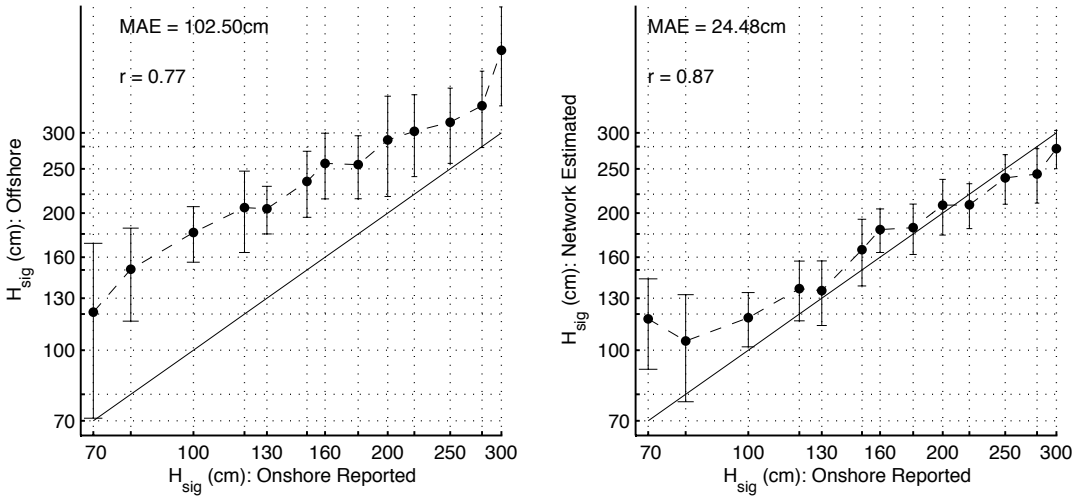
Fig. 5.



(a) Chiton Rocks

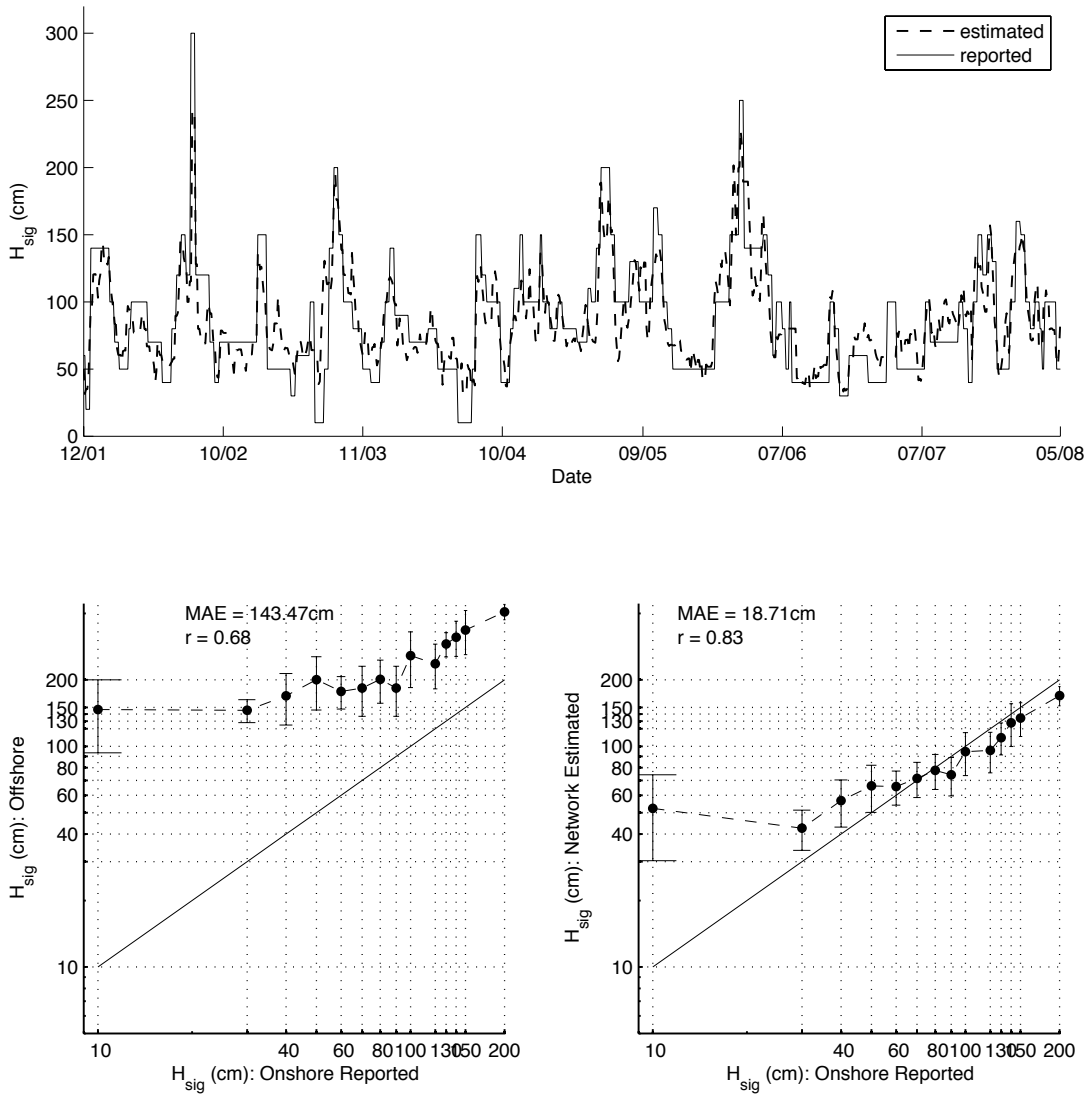
Fig. 6.





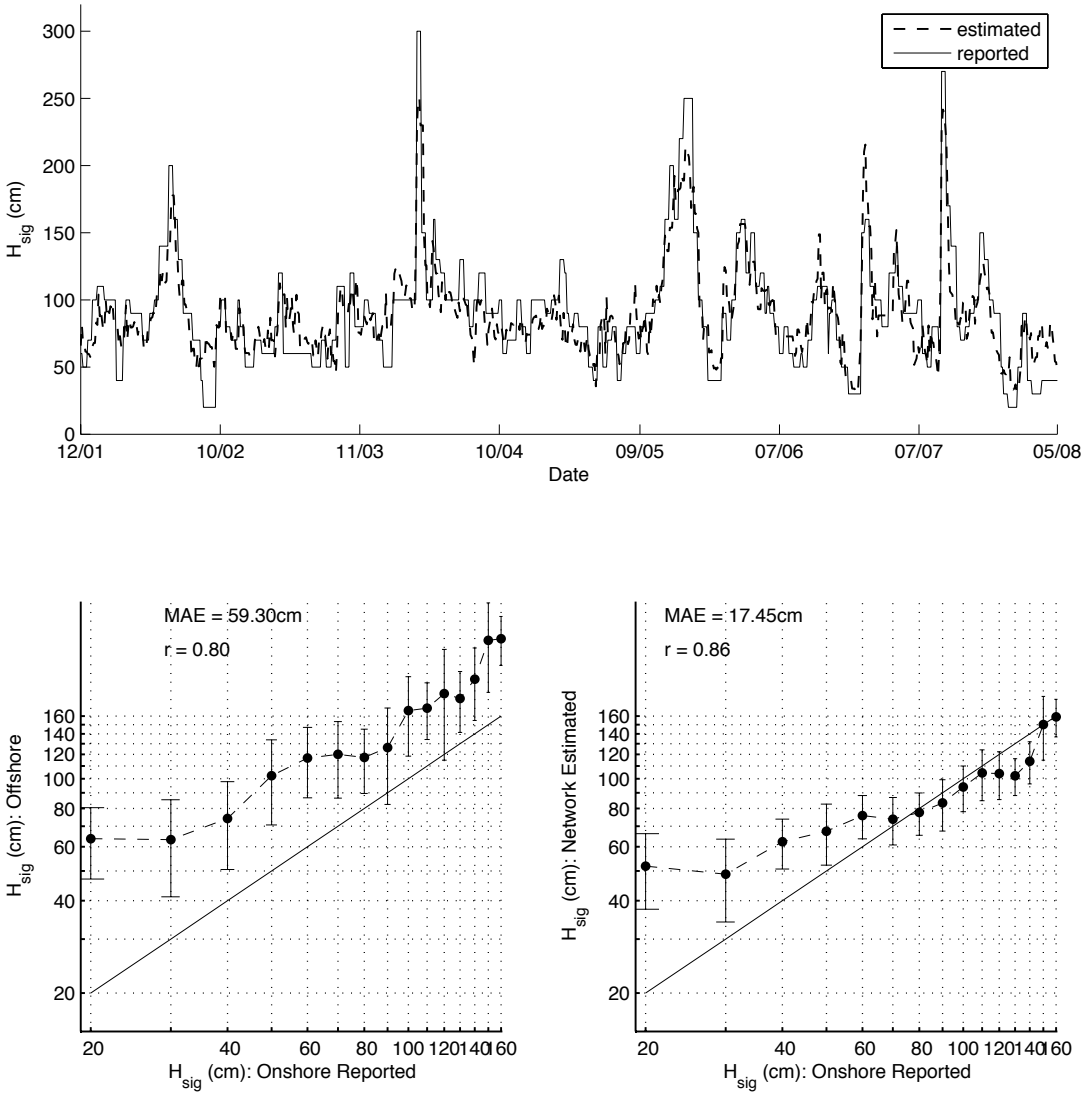
(a) Margaret River

Fig. 7.



(a) Torquay

Fig. 8.



(a) Palm Beach, Sydney

Fig. 9.

Fig. 1 Standard feed-forward ANN architecture

Fig. 2 Bathymetry for five of the study regions

Fig. 3 Mean significant wave height as estimated by the SWAN model when driven by NWW3 over the study period

Fig. 4 Normalised correlation coefficient r between the NOAA driving input and SWAN output significant wave height over the study period. Lighter shades indicate local regions affected by more complex wave transformations: i.e. a greater proportion of observed wave height variability is not linearly related to the wave height at the model boundaries

Fig. 5 ANN estimation performance at Surfer's Paradise, Queensland

Fig. 6 ANN estimation performance on South Australian beaches

Fig. 7 ANN estimation performance on Western Australian beaches

Fig. 8 ANN estimation performance on Victorian beaches

Fig. 9 ANN estimation performance on New South Wales beaches