Figure 1:

1. Introduction

The forecasting of wave heights is important for a range of maritime and coastal applications, from port operations through to beach management and surfing. One of the challenges is the ability to downscale predictions from oceanographic wave models to the coast. This is required as additional physical processes occur as the waves approach local features, such as depth induced wave breaking, refraction and defraction. Typically this involves the development of computationally expensive physics based models, which also require good quality input data, including bathymetry to perform well. Data-driven approaches have increasingly been applied in order to help overcome the computational cost and allow the rapid downscaling of wave forecasts and hindcasts. Most have utilised wave parameters ««REFS HERE»» , but this can have it's challenges as wave parameters are summaries of the wave spectrum (which is a more full description of wave conditions in the form of energy spread over a range of frequencies)«<REFS»>. Machine learning techniques are being increasingly applied in this space

2. Method

2.1. Datasets

Training of the machine learning models can require a substantial amount of data, two key datasets were used for training. Offshore wave spectrum from a wave watch 3 model (in this case the CAWCR wave hindcast) and wave observations at the point of interest. The CAWCR wave hindcast spectrum consists of a 2d dimensional wave spectrum with 30 directional bins and X frequency bins at y spacing.

Wave observations sourced from the Queensland Government Wave Monitoring program for sites shown in figure 4. The length of the wave monitoring data for each site was varied and training was therefore undertaken on the available data.

2.1.1. Data preparation

Extracting features from training data is key element in training machine learning models. As well as extracting features each dataset needed to be cleaned to ensure consistency, remove errors and fill gaps where possible.

To simplify the 2 dimensional wave spectrum for each of the offshore points two processing steps were undertaken, first wave parameters were extracted, second the spectrum was integrated into a 1 dimensional spectrum (integrating directions together). Wave parameters calculated including, Tp, pDir, Hs, Tz and zDir. Directional values were split into x and y values.

The wave monitoring data also required a range of pre-processing techniques, as the data is based on observational equipment there were several issues present in the dataset. These included error values, small data gaps, large data gaps and spurious values. Error values were removed, small gaps (< 3 hours) were linearly interpolated and spurious data was removed using a range filter. Directional values were converted into x and y values.

The datasets were then merged to create a continuous dataset, and for feature extraction. Features were then extracted, particularly focussed on identifying seasonality in the time series data, months, days, year and hours were all added to the training dataset.

2.2. Training, Testing, Validation

The dataset was split into validation (1 year of data) with testing and training each being 80% and 20% of the remainder. Scaling was undertaken on the entire dataset to ensure that scaled data included all the data within the minimum and maximum of the dataset. Time series cross validation was configured for model training, with a growing training window of 7 days.

3. Machine Learning Model

An Long-Term Short-Term memory model was selected....

4. Discussion

5. Conclusion

6.

Leo is a PhD candidate at the Griffith Univeristy School of Engineering and Built Environment, and the Marine and Coastal Research Centre