Coastal Engineering 137 (2018) 1–10

Contents lists available at ScienceDirect

Coastal Engineering

journal homepage: www.elsevier.com/locate/coastaleng

A machine learning framework to forecast wave conditions

Scott C. James[a,*], Yushan Zhang[b], Fearghal O'Donncha[c]

[a] Baylor University, Departments of Geosciences and Mechanical Engineering, One Bear Place #97354, Waco, TX 76798-2534, USA
[b] University of Notre Dame, Department of Chemical and Biomolecular Engineering, Notre Dame, IN 46556-5637, USA
[c] IBM Research, Dublin, Ireland

ARTICLE INFO

Keywords:
Machine learning
SWAN wave modeling
Wave-condition forecasting

ABSTRACT

A machine learning framework is developed to estimate ocean-wave conditions. By supervised training of machine learning models on many thousands of iterations of a physics-based wave model, accurate representations of significant wave heights and period can be used to predict ocean conditions. A model of Monterey Bay was used as the example test site; it was forced by measured wave conditions, ocean-current nowcasts, and reported winds. These input data along with model outputs of spatially variable wave heights and characteristic period were aggregated into supervised learning training and test data sets, which were supplied to machine learning models. These machine learning models replicated wave heights from the physics-based model with a root-mean-squared error of 9 cm and correctly identify over 90% of the characteristic periods for the test-data sets. Impressively, transforming model inputs to outputs through matrix operations requires only a fraction ($< 1/1,000^{\text{th}}$) of the computation time compared to forecasting with the physics-based model.

1. Introduction

There are myriad reasons why predicting wave conditions is important to the economy. Surfers aside, there are fundamental reasons why knowledge of wave conditions for the next couple of days is important. For example, shipping routes can be optimized by avoiding rough seas thereby reducing shipping times. Another industry that benefits from knowledge of wave conditions is the $160B (2014) aquaculture industry (FAO, 2016), which could optimize harvesting operations accordingly. Knowledge of littoral conditions is critical to military and amphibious operations by Navy and Marine Corps teams. Also, predicting the energy production from renewable energy sources is critical to maintaining a stable electrical grid because many renewable energy sources (e.g., solar, wind, tidal, wave, etc.) are intermittent. For deeper market penetration of renewable energies, combinations of increased energy storage and improved energy-generation predictions will be required. The US Department of Energy has recently invested in the design, permitting, and construction of an open-water, grid-connected national Wave Energy Test Facility at Oregon State University (US DOE, 2016). Given that America's technically recoverable wave-energy resource is up to 1,230 TW-hr (EPRI, 2011), there is a strong interest in developing this renewable resource (Ocean Energy Systems, 2016). Commercialization and deployment of wave-energy technologies will require not only addressing permitting and regulatory matters, but overcoming technological challenges, one of which is being able to provide an accurate prediction of energy generation. A requirement for any forecast is that an appropriately representative model be developed, calibrated, and validated. Moreover, this model must be able to run extremely fast and to incorporate relevant forecast data into its predictions. A machine learning framework for this capability is developed here.

Because wave models can be computationally expensive, a new approach with machine learning (Goodfellow et al., 2016; LeCun et al., 2015; Schmidhuber, 2015) is developed here. The goal of this approach is to train machine learning models on many realizations of a physics-based wave model forced by historical atmospheric and sea states to accurately represent wave conditions (specifically, significant wave heights and characteristic period). Predicting these wave conditions at locations corresponding to a (potential) wave-energy-converter (WEC) array facilitates accurate power-production forecasts. Given the recent development of a wave-energy-resource classification system (Haas et al., 2017), if the wave conditions at a particular location can be predicted, the power potential for a hypothetical WEC array can be estimated.

Computational expense is often a major limitation of real-time forecasting systems (DeVries et al., 2017; Mallet et al., 2009). Here, we apply machine learning techniques to predict wave conditions with the goal of replacing a computationally intensive physics-based model by

* Corresponding author.
E-mail address: sc_james@baylor.edu (S.C. James).

https://doi.org/10.1016/j.coastaleng.2018.03.004
Received 5 September 2017; Received in revised form 23 February 2018; Accepted 7 March 2018

straightforward multiplication of an input vector by mapping matrices resulting from the trained machine learning models. Because matrix multiplication is an exceedingly rapid operation, the end result is a machine learning technique that can predict wave conditions with comparable accuracy to a physics-based model for a fraction of the computational cost. While machine learning has been used to predict wave conditions (Peres et al., 2015; Makarynskyy, 2004; Etemad-Shahidi and Mahjoobi, 2009; Mahjoobi and Etemad-Shahidi, 2008; Browne et al., 2006, 2007), it has not been used in the context of a surrogate model as defined below.

One of the challenges for machine learning applications is their enormous appetite for data. It is the exception more than the rule that a machine learning approach has what is considered an optimal amount of data. However, when developing a machine learning surrogate for a physics-based model, there is the luxury of being able to run the physics-based model as many times as necessary to develop a sufficiently large data set to train the machine learning model. Here, we define a surrogate model (Razavi et al., 2012) as a data-driven technique to empirically approximate the response surface of a physics-based model. These have alternately been called “metamodels” (Blanning, 1975; Kleijnen, 2009), “model emulators” (O'Hagan, 2006), and “proxy models” (Bieker et al., 2007). To assemble the training dataset required to develop a robust machine learning model, the inputs and outputs of many thousands of wave model runs were accumulated into a suitably large set of input vectors.

2. Wave modeling

2.1. Numerical model

The Simulating WAves Nearshore (SWAN) V 41.62 FORTRAN code is the industry-standard wave-modeling tool developed at the Delft University of Technology that computes wave fields in coastal waters forced by wave conditions on the domain boundaries, ocean currents, and winds (The SWAN Team, 2006). SWAN models the energy contained in waves as they travel over the ocean and disperse at the shore. Specifically, information about the sea surface is contained in the wave-variance spectrum, or energy density $E(\sigma,\theta)$, and this wave energy is distributed over wave frequencies (as observed in an inertial frame of reference moving with the current velocity) with propagation directions normal to wave crests of each spectral component.

Action density is defined as $N = E/\sigma$, which is conserved during propagation along the wave characteristic in the presence of ambient current. Evolution of $N(x,y,t;\sigma,\theta)$ in space, x,y, and time, t, is governed by the action balance equation (Komen et al., 1996; Mei et al., 1989):

$$\frac{\partial N}{\partial t}+\left(\frac{\partial c_x N}{\partial x}+\frac{\partial c_y N}{\partial y}\right)+\left(\frac{\partial c_\sigma N}{\partial \sigma}+\frac{\partial c_\theta N}{\partial \theta}\right)=\frac{S_{\mathrm{tot}}}{\sigma}. \tag{1}$$

The left-hand side represents the kinematic component of the equation. The second term (parenthetical) denotes the propagation of wave energy in a two-dimensional Cartesian space where c is wave celerity. The third term represents the effect of shifting of the radian frequency due to variations in water depth and mean current. The fourth term expresses depth- and current-induced refractions. The quantities c_σ and c_θ are the propagation speeds in spectral space (σ,θ). The right-hand side represents the spatio-temporally variable sources and sinks of all physical processes that generate, dissipate, or redistribute wave energy (i.e., wave growth by wind, nonlinear transfer of energy through three- and four-wave interactions, and wave decay due to white-capping, bottom friction, and depth-induced wave breaking).

Haas et al. (2017) define the wave-energy resource as a function of the significant wave height, H_s and peak wave period, T. This information can be used to compute the wave power density. Hence, estimates of peak period T and, in particular, H_s because J is quadratically related to wave height, are necessary to predict wave energy potential.

2.2. Model verification

The coastal ocean presents a complex modeling challenge, intimately connected as it is to both the deep ocean and the atmosphere (Song and Haidvogel, 1994). Uncertainties in wave forecasting emanate from the mathematical representation of the system, numerical approximations, and uncertain and incomplete data sets. Studies demonstrate that the greatest sources of uncertainty in operational wave forecasting are the model input data. This study simulates wave conditions subject to real forcing conditions at a case-study site, Monterey Bay, California. As summarized in Table 1, the wave model was driven by available NOAA wave-condition data, archived current nowcasts from the Central and Northern California Ocean Observing System (CeNCOOS) (Patterson et al., 2012), and post-processed (i.e., data subject to quality-assurance procedures) wind data from The Weather Company (2017).

Before developing a machine-learning surrogate for the physics-based SWAN model, it is important to demonstrate that SWAN can accurately replicate wave conditions in Monterey Bay so that a training data set can be developed for the machine learning models. The SWAN model validated by Chang et al (2016) was used in this effort because it has a demonstrated track record for accurately simulating wave conditions in Monterey Bay. The bathymetric data shown in Fig. 1 were obtained from the NOAA National Geophysical Data Center. The horizontal resolution in this SWAN model was 0.001°.

Because SWAN discretizes wave frequencies in its calculations, only a user-defined number of discrete T values can be returned by a simulation. Specifically, because this effort is relevant to forecasts for wave-energy converters, we consider the peak wave period (RTP in the SWAN model) because this is used in the calculation for wave power (Cahill and Lewis, 2014). To this end, when building the model the user specifies the minimum, ϕ_1, maximum ϕ_{N}, and number of discrete frequencies, N, which are logarithmically distributed as (The SWAN Team, 2006):

$$\phi_\iota=\phi_{\iota-1}+\phi_{\iota-1}\left(\frac{\phi_{\mathrm{N}}}{\phi_1}-1\right)^{\frac{1}{\mathrm{N}-1}}. \tag{2}$$

Note that the logarithmic distribution yields smaller increments between periods for larger T, which is most relevant for capturing the effects of long-period waves – those most important for energy generation. For these simulations, $\phi_1 = 0.042$ Hz ($T_1 = 23.8$ s) and $\phi_{\mathrm{N}} = 1$ Hz ($T_{\mathrm{N}} = 1$ s), and $\mathrm{N} = 24$ discrete T values were specified (Chang et al., 2016). However, only 11 distinct T values were ever calculated by SWAN

Table 1
Data sources.

Data	Source	URL	Resolution
Wave conditions (H_s, T, D)	NDBC[c,a] Buoy 46042	http://www.ndbc.noaa.gov/station_page.php?station=46042	36°47′29″ N 122°27′6″ W
Wave conditions (H_s, T, D)	WAVE WATCH III ENP NCEP[b]	http://nomads.ncep.noaa.gov:9090/dods/wave/enp	0.25°
Ocean currents	ROMS COPS[d]	http://west.rssoffice.com:8080/thredds/catalog/roms/CA3000m-forecast/catalog.html	3 km
Winds	TWC[d]	https://api.weather.com/	User defined
Bathymetry	NOAA NGDC[e]	https://www.ngdc.noaa.gov/mgg/bathymetry/hydro.html	0.001°

[a] National Data Buoy Center.
[b] Eastern North Pacific National Centers for Environmental Prediction.
[c] Regional Ocean Modeling System Cooperative Ocean Prediction System.
[d] The Weather Company.
[e] National Geophysical Data Center.