SWRL Net: A Spectral, Residual Deep Learning Model for Improving Short-Term Wave Forecasts

JONNY MOONEYHAM

Computer Science, Western Washington University, Bellingham, Washington

SEAN C. CROSBY

Geology, Western Washington University, Bellingham, Washington

NIRNIMESH KUMAR

Civil and Environmental Engineering, University of Washington, Seattle, Washington

BRIAN HUTCHINSON

Computer Science, Western Washington University, Bellingham, and Computing and Analytics Division, Pacific Northwest National Laboratory, Richland, Washington

(Manuscript received 18 January 2020, in final form 8 July 2020)

ABSTRACT

Skillful nearshore wave forecasts are critical for providing timely alerts of hazardous wave events that impact navigation or recreational beach use. While typical forecasts provide bulk wave parameters (wave height and period), spectral details are needed to correctly predict wave and associated circulation dynamics in the nearshore region. Currently, global wave models, such as WAVEWATCH III (WW3), make spectral predictions, but do not assimilate regional buoy observations. Here, Spectral Wave Residual Learning Network (SWRL Net), a fully convolutional neural network, is trained to take recent WW3 forecasts and buoy observations, and produce corrections to frequency-directional WW3 spectra, transformed into directional buoy moments, for up to 24 h in the future. SWRL Net is trained with 10 years of collocated NOAA's WW3 CFSR reanalysis predictions and buoy observations at three locations offshore of the U.S. western coast. At buoy locations SWRL Net residual corrections result in wave height root-mean-square error (RMSE) reductions of 23%–50% in the first 6 h and 10%–20% thereafter. Sea frequencies (5–10 s) show the most improvement compared to swell (12–20 s). SWRL Net reduces mean direction RMSE by 28%–54% and mean period RMSE by 20%–56% over 24 forecast hours. While each model is trained and tested at independent locations, SWRL Net exhibits generalization when introduced to data from other locations, suggesting future development may be composed of training sets from multiple locations.

1. Introduction

Surface gravity waves in the coastal and deeper ocean impact commercial and recreational boating, and are source for marine renewable energy. In the nearshore region, extending from the coastline to few kilometers offshore, surface gravity waves shoal and subsequently

break, driving sediment transport (Komar and Inman 1970) and coastal erosion. Circulation driven by surface wave breaking has important consequences for offshore transport of pollutants and contaminants, maintenance of nearshore water quality, and ecological balance for epifaunal organisms (Halpern et al. 2012; Boehm et al. 2017). Extreme wave events interacting with local bathymetry may lead to strong offshore directed rip currents and/or sneaker waves (García-Medina et al. 2018), which are some of the most common causes of lifeguard rescues. Furthermore, design and maintenance of coastal structures depend on magnitude and forecasts of extreme wave events. Rapid and accurate wave predictions could

Denotes content that is immediately available upon publication as open access.

Corresponding author: Jonny Mooneyham, jonathan.mooneyham@gmail.com

DOI: 10.1175/WAF-D-19-0254.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](https://www.ametsoc.org/PUBSReuseLicenses).

improve public and boating safety and allow for timely alerts of high wave events and corresponding rip and alongshore currents. Additionally, accurate wave forecasts could drive local tracer models (larvae, sediments, pollutants), providing critical forecasts to coastal managers.

Operational global and regional wave model skill continues to improve with increasingly accurate wind data from satellites and improving parameterization of wind-wave generation and dissipation mechanisms in state-of-the-art models (e.g., [Cavaleri et al. 2007](#); [Ardhuin et al. 2010](#)). Prior efforts to assimilate wave observations have demonstrated some success with synthetic aperture radar satellite observation ([Walker 2006](#)) and buoy observations ([O'Reilly and Guza 1998](#); [Veeramony et al. 2010](#); [Orzech et al. 2013](#); [Crosby et al. 2017, 2016](#)). Recently a simple wave assimilation framework has been developed through a network of free-drifting satellite-connected surface weather buoys ([Smit et al. 2020](#)). Yet, assimilation remains challenging and operational products such as the National Oceanic and Atmospheric Administration's WAVEWATCH III (WW3; [Tolman 2009](#)) do not currently assimilate buoy observations.

Multiple characteristics associated with surface wave dynamics make numerical wave modeling and wave data assimilation challenging. Wave generation, dissipation, and some aspects of propagation like quadruplet and triad wave interaction are nonlinear and vary across time scales. Errors in surface wind forcing input on regional and global scales ([Stopa 2018](#)), inaccuracies in boundary wave forcing ([Crosby et al. 2017](#); [Kumar et al. 2017](#)), inadequate spatial resolution (e.g., [Crosby et al. 2019](#)), complications associated with model tuning and calibrations ([Allahdadi et al. 2019a](#)), sensitivity associated with physics parameterizations for processes like whitecapping ([Allahdadi et al. 2019b](#)), and exclusion of wave-current interaction processes in global and regional model applications limit accurate wave predictions. Given the existing limitations with data-assimilative wave modeling, other techniques are needed for correct forecasts.

From historical observation and prediction records, simple bias adjustments can be developed and applied to improve future predictions. Methods vary, and in some cases may allow corrections to wave height or energy to be a function of incoming wave direction ([Camus et al. 2013](#)). Recently more sophisticated techniques have been developed to correct forecasts. Application of a Kalman filter to wave height forecasts and collocated buoy observations suggests errors may be reduced up to 10%–30% for a 12-h forecast ([Emmanouil et al. 2010](#)). Combination of a similar Kalman filtering approach and optimal interpolation with both buoy and satellite

altimeter observations shows further forecast improvement up to 24 h ([Emmanouil et al. 2012](#)). Such filtering techniques have been previously applied successfully across a wide array of fields. However, if wave prediction errors are highly nonlinear, the aforementioned methods are expected to be of limited use, and machine learning techniques via a neural network (NN) framework may further improve the forecasts.

Before model forecasts were widely available (e.g., WW3), earlier NN implementations showed skillful predictions for up to 6 h based solely on previously observed wave conditions ([Londhe and Panchang 2006](#)). NNs have also been developed to replace computationally costly phase-averaged numerical models with training on wind and wave observations ([Deo et al. 2001](#)). More recent implementations have trained NNs with many iterations of dynamic model runs alongside measured winds and predicted currents ([James et al. 2018](#)). When wave forecasts are available, prior studies have shown a shallow NN (consisting of one hidden layer) can make corrections to wave heights and period, though improvement was marginal ([Makarynsky 2004](#)). A more recent application of a wavelet NN reduces wave height and peak period prediction errors substantially (up to 50%) over 3–24 forecast hours ([Deshmukh et al. 2016](#)). Additionally, bagged regression tree approaches ([Ellenson et al. 2020](#)) applied to buoy locations in California–Oregon border have been used to determine deviations in the output of significant wave height from WW3, compared to buoy measurements.

The aforementioned studies show promise for applying NNs to make forecast correction, however, NNs have not yet been used to make corrections to wave spectra details. Spectral information is typically available at specific forecast locations and measured by collocated directional wave buoys. Though wave conditions are typically reported in bulk parameters (wave height, period, and direction) directional wave buoys measure energy and four low-order directional moments as a function of frequency ([Longuet-Higgins et al. 1963](#)). These spectral details are needed to accurately estimate infragravity energy, total shoreline run-up ([Guza and Feddersen 2012](#); [Fiedler et al. 2018](#)), Stokes drift ([Kumar et al. 2017](#)), and as model boundary input for nested wave or current models, especially in highly sheltered regions ([Crosby et al. 2017](#)). Additionally, in open coastal settings mixed seas are typical where remotely generated swell is present alongside locally generated seas (e.g., [Adams et al. 2008](#)). Use of spectral observations and predictions reduced ambiguity and may improve assimilation implementations.

Most of the previous studies use shallow NNs (consisting of a single hidden layer), which, given a sufficiently

large number of hidden units, are theoretically as expressive as deep networks. In practice, the inductive bias of deep networks tends to align well with real-world, highly nonlinear input–output modeling tasks, allowing deep networks to learn more efficiently from data and obtain better generalization. Through repeated nonlinear transformations, deep networks have the ability to learn increasingly high-level (abstract) representations at each subsequent layer (Bengio 2009). They are particularly effective when applied to “raw” (i.e., not aggregated or bulk) features and when trained on large amounts of data. Owing to the volume of data and likely nonlinear nature of wave prediction errors, the application of deep NNs is promising. While previous studies have applied deep NNs to forecast significant wave heights (e.g., Cornejo-Bueno et al. 2016; Ali and Prasad 2019), none have considered the spectral approach presented here.

Here, we develop a deep learning model that we call Spectral Wave Residual Learning Network (SWRL Net) to improve numerical model predictions with directional wave buoy observations. Spectral wave predictions at buoy locations are used with collocated directional buoy observations to generate forecast corrections up to 24 h in the future. Frequency-directional spectra are transformed into the observed buoy moments resulting in a large feature set and large number of model parameters. The remainder of this paper is organized as follows. Background information on machine learning is presented in section 2. Methods are defined in section 3, followed by experimental results in section 4. Findings from this study are further discussed in section 5, and a summary is provided in section 6.

2. Machine learning background

Machine learning is the study of algorithms that learn from data. This methodology is playing an increasingly important role in a range of scientific domains, including climate prediction, biology, and chemistry. In supervised machine learning, a training set of input–output pairs is fed to a learning algorithm, which produces a model. The training process typically optimizes a model to yield the best predictions (lowest error on average) on the training set data. Once trained, the resulting model is then capable of producing predictions for novel inputs. Here, for example, the model takes a combination of WW3 forecasts and historical buoy observations as inputs and generates corrected regional forecasts. Deep learning is a subtype of machine learning, involving models in which the input undergoes several nonlinear transformations, before a prediction is made, via the composition of many differentiable functions. One can interpret each successive transformation of the input as

emphasizing the discriminatory information contained in the input, while attenuating the irrelevant information. Neural networks are the dominant model used in deep learning. In a fully connected, feed-forward neural network with K hidden layers, the output $y \in \mathbb{R}^C = h_{(K+1)}$ is a function of the input $x \in \mathbb{R}^D = h_{(0)}$ as follows:

$$h_k = f_{(k)}[\mathbf{W}_{(k)}^T h_{(k-1)} + \mathbf{b}_{(k)}]. \quad (1)$$

Here $\mathbf{W}_{(1)}, \dots, \mathbf{W}_{(K+1)}$ and $\mathbf{b}_{(1)}, \dots, \mathbf{b}_{(K+1)}$ are the model parameters also known as model weights and biases. The terms C and D denote the dimensionality of input and outputs, respectively. The variable h_k denotes the intermediate representation of the input at layer k of the model, with h_{K+1} denoting the model predictions. Functions $f_{(1)}, \dots, f_{(K)}$ are the hidden nonlinear activation functions and $f_{(K+1)}$ is the output activation function; without the hidden activation the overall model would be linear. Training the neural network consists of finding the optimal model parameters. There are many variations of neural networks, adapted for a variety of tasks.

Convolutional neural network (CNNs) are a popular deep learning architecture that are highly suited to grid structured input, such as images. CNNs differ from neural networks by utilizing convolutional and pooling layers in addition to fully connected layers. In our work we use *fully convolutional networks*, which only use convolutional layers. The input to a convolutional layer is a tensor, where one dimension of the tensor corresponds to *channels* (i.e., features) and the remaining dimensions are assumed to be spatial or temporal. For example, image inputs are represented by a third-order tensor: $\mathbf{X} \in \mathbb{R}^{L \times H \times W} = h_{(0)}$. In this case, H is the spatial height of the image, W is the spatial width of the image, and L is the number of channels (e.g., three with red, green, and blue intensities). The output of a convolutional layer is a tensor of the same order as the input (e.g., a convolutional layer with input $h_{(0)} \in \mathbb{R}^{L \times H \times W}$ would have output $z_{(1)} \in \mathbb{R}^{L' \times H' \times W'}$). In the hidden layers, transformations are applied to the input data followed by an activation function, so $h_{(1)} = f_{(1)}(z_{(1)})$. By using convolutional layers instead of fully connected layers, CNNs model have fewer learned parameters than standard deep neural networks, aiding trainability. For further details on CNNs and the convolution operation, we refer the reader to Goodfellow et al. (2016), chapter 9.

3. Methods

a. Buoy observations and processing

A directional wave buoy measuring pitch, roll, and heave, yields estimates of energy e and four directional moments (a_1, b_1, a_2, b_2) across frequency (Longuet-Higgins

TABLE 1. National Oceanic and Atmospheric Administration (NOAA) buoy names, locations, and time frames used in model training, validation, and testing.

NOAA ID	Buoy name	Latitude	Longitude	Training set	Valid set	Test set
46211	Grays Harbor, WA	46.858°N	124.244°W	1 Sep 1993–7 Apr 2003	29 Apr 2003–5 Apr 2005	26 Apr 2005–9 Mar 2007
46214	Point Reyes, CA	37.950°N	123.472°W	6 Dec 1996–30 Oct 2005	17 Nov 2005–28 May 2007	29 Jun 2007–19 Apr 2009
46218	Harvest, CA	34.454°N	120.783°W	1 Dec 1995–8 Jul 2003	8 Jul 2003–22 Feb 2005	22 Feb 2005–14 Dec 2008

et al. 1963). Datawell directional buoys provide accurate observations (O'Reilly et al. 1996), and a large array is maintained by the Coastal Data Information Program (CDIP).¹ Observations at three locations (Table 1) along the U.S. West Coast (Fig. 1) are collected and processed for model training.

Detailed wave spectra predictions from NOAA WW3 are available at most buoy site locations. Hindcast predictions from NOAA's CFSR phase 2 (Chawla et al. 2013; Tolman 2009) are extracted at Grays Harbor, Pt. Reyes, and Harvest locations. WW3 model predictions are linearly interpolated from 3- to 1-h time sampling to match the 1-h sampling of the buoy observations and converted from log to linear scale in frequency. At each frequency, integrating across direction θ , such that

$$e^{(p)} = \int E(\theta) d\theta \quad (2)$$

and

$$\begin{bmatrix} a_1^{(p)} \\ b_1^{(p)} \\ a_2^{(p)} \\ b_2^{(p)} \end{bmatrix} = \begin{bmatrix} \int E(\theta) \cos\theta d\theta \\ \int E(\theta) \sin\theta d\theta \\ \int E(\theta) \cos 2\theta d\theta \\ \int E(\theta) \sin 2\theta d\theta \end{bmatrix} / e^{(p)} \quad (3)$$

transforms WW3 spectra $E(\theta)$ into total predicted energy $e^{(p)}$ and normalized directional moments $a_1^{(p)}$, $b_1^{(p)}$, $a_2^{(p)}$, and $b_2^{(p)}$, which are in the same physical space as directional buoy observations (i.e., $e^{(o)}$, $a_1^{(o)}$, $b_1^{(o)}$, $a_2^{(o)}$, $b_2^{(o)}$). Time periods with missing data comprise less than 9.2% of overlapping records and are discarded.

b. Data preparation

Neural networks perform best when the input features and output targets have been normalized into a small range near the origin; this avoids the saturation of nonlinearities and improves numerical stability. For SWRL Net, directional moments are normalized

by dividing them by the energy; this bounds them to the range $(-1, 1)$, which is simple to enforce as a model constraint. Energy e is standardized by subtracting its mean and dividing by its standard deviation: $\mathbf{X}'_e = (\mathbf{X}_e - \mu_e)/\sigma_e$. Normalizing all of the inputs in this way also keeps all input features in a similar range, giving them equal footing during training. The normalization process is later reversed for the output layer of SWRL Net model, so that results are reported in the original units.

The frequency range is limited to 0.04–0.25 Hz, typical of surface gravity waves. Each hour in either the buoy observations or WW3 data is represented as a feature map in $\mathbb{R}^{5 \times 28}$, where energy and directional moments

FIG. 1. Domain plot with buoy locations (red circles, see Table 1) used to train models. Gray contours show 500-, 1000-, 2000-, and 4000-m isobaths.

¹ <https://cdip.ucsd.edu/>.

FIG. 2. Model schematic where (a) illustrates input and outputs to the model as third-order tensors and (b) illustrates how model inputs are computed through a series of convolutions with a 3×3 filter and a specific number of channels per layer. Residual outputs of the convolutional layer are added to corresponding WW3 data, the constraints are applied to ensure energy is positive and normalized moments are in the range $[-1, 1]$.

(a_1, b_1, a_2, b_2, e) create five rows, and frequencies create 28 columns.

For each buoy, the data are split into three disjoint datasets: train, valid, and test. The train set is used to train SWRL Net model, the valid set is used to tune the model (i.e., determine which model configurations generalize best to unseen data), and the test set is used to assess the tuned model's skill on unseen data. The data are split along chronological boundaries, with the first 70% in train, the following 15% in valid, and the final 15% in test (Table 1).

Let $\mathcal{X}^{\text{buoy, set}} \in \mathbb{R}^{T \times M \times F}$ be the third-order tensor containing all buoy data in set (out of train, valid or test), where T is the total number of hours in the dataset, M is the number of moments, and F is the number of frequency bins. Then $\mathcal{X}_{\text{tmf}}^{\text{buoy, set}}$ is the m th

moment (in a_1, b_1, a_2, b_2, e) for frequency bin f at time t . Let $\mathcal{X}^{\text{ww3, set}}$ have identical dimensions, but contain the WW3 predictions for the same times and frequencies. Finally, dropping the superscript for notation simplicity, let $\mathcal{X}_{(ij)}^{\text{set}} \in \mathbb{R}^{(j-i+1) \times M \times F}$ denote a slice of $\mathcal{X}$ from times i through j , inclusive, where each hour i yields one data point from set, an input–output pair $(X_{(i)}^{\text{set}}, Y_{(i)}^{\text{set}})$. The input is $X_{(i)}^{\text{set}} = \mathcal{X}_{[(i-\tau^h+1):(i+\tau^p)]}^{\text{ww3, set}} \oplus \mathcal{X}_{[(i-\tau^h+1):i]}^{\text{buoy, set}}$, where $\oplus$ denotes concatenation along the first mode (time) of the tensor, τ^h is the number of hours the prediction is conditioned on, and τ^p is the number of hours to predict into the future. That is, we condition on three distinct types of information: 1) the most recent τ^h hours of WW3 forecasts, 2) the most recent τ^h hours of buoy observations, and 3) the next τ^p hours of WW3 forecasts.

FIG. 3. (a) Significant wave height, (b) mean wave direction, and (c) mean wave period vs time at Grays Harbor buoy location. Observations (black circles) are compared to WW3 predictions (black dashed) and CNN adjusted forecasts (colored lines) that use recent history inputs illustrated by corresponding colored circles.

The corresponding label/output in the data point in set for hour i is $Y_{(i)}^{\text{set}} = \mathcal{X}_{[i:(i+\tau^p)]}^{\text{buoy, set}}$; that is, the true buoy observations for the next τ^p hours.

c. Model architecture

SWRL Net is implemented as a CNN using PyTorch (Paszke et al. 2017). Model attributes are tuned, such as the number of layers and the number of filters per layer. An example architecture is shown in Fig. 2. The output of the ℓ th convolutional layer of the model has dimension $L_\ell \times M \times F$; note that the “spatial” dimensions (M and F) are preserved throughout the model, and L_ℓ is the number of filters in the layer. Each convolutional layer is followed by a leaky rectified linear unit (ReLU; Maas et al. 2013) hidden activation function, except for the final layer. The final output layer generates a tensor with the same dimensions as targets (i.e., $\tau_p \times M \times F$). SWRL Net is trained to output the residual augmenting the WW3 forecast by adding the model output to $\mathcal{X}_{[t:(t+\tau^p)]}^{\text{ww3}}$ before computing loss. In general, modeling residuals has been shown to greatly improve trainability (He et al. 2016).

In the output layer, constraints are always applied to suppress forecasts that are out of bounds (Fig. 2). First, forecasts with negative energy in the unnormalized space are set to zero, although negative energy predictions were exceedingly rare. Second, tanh or hardtanh is used to constrain a_1, b_1, a_2, b_2 to a valid range of $(-1, 1)$. The function tanh is the hyperbolic tangent where $\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$, and hardtanh is a simple bounding function where

$$\text{hardtanh}(x) = \begin{cases} -1, & x < -1 \\ x, & -1 \leq x \leq 1 \\ 1, & x > 1 \end{cases} \quad (4)$$

d. Model training

SWRL Net is trained to minimize mean squared error (MSE) between predictions $\hat{Y}$ and the true targets Y :

$$\frac{1}{N_{\text{tr}}} \sum_{i=1}^{N_{\text{tr}}} \sum_{t=1}^{\tau^p} \sum_{m=1}^M \sum_{f=1}^F (\hat{Y}_{(i)\text{tmf}}^{\text{train}} - Y_{(i)\text{tmf}}^{\text{train}})^2. \quad (5)$$

All SWRL Net parameters are jointly optimized using minibatch stochastic gradient descent, where a small set

TABLE 2. WW3 Root-mean-square-errors (RMSE) and SWRL Net percent improvement (positive percentage is an error reduction) on those errors for varying forecast hour window. Errors and improvements are estimated for wave height, mean direction, and mean period at each buoy site for all observations in the test dataset (indicated by mean) as well as for just extreme conditions defined as wave heights above the 90th percentile value.

	Mean					90th percentile wave heights				
	WW3	SWRL Net forecast hours				WW3	SWRL Net forecast hours			
	RMSE	1–3	3–6	6–12	12–24	RMSE	1–3	3–6	6–12	12–24
Grays Harbor										
H_s	0.32	40%	23%	14%	10%	0.65	41%	18%	10%	10%
θ_m (°)	7.68	45%	35%	30%	28%	7.01	41%	30%	28%	27%
T_m	0.78	49%	37%	29%	26%	0.76	50%	31%	22%	19%
Point Reyes										
H_s	0.39	50%	34%	20%	13%	0.67	48%	22%	−1%	−7%
θ_m (°)	12.58	50%	43%	38%	32%	15.31	48%	39%	36%	28%
T_m	0.78	51%	39%	28%	20%	0.63	42%	16%	−3%	−7%
Harvest										
H_s	0.32	52%	36%	24%	18%	0.49	45%	27%	14%	7%
θ_m (°)	10.50	54%	45%	38%	32%	8.18	51%	45%	35%	27%
T_m	0.78	56%	42%	30%	25%	0.73	52%	29%	10%	5%

(minibatch) of input–output pairs $\{(X_{(i)}^{\text{train}}, Y_{(i)}^{\text{train}})\}$ are drawn uniformly at random from the train dataset (without replacement). In this case, each gradient is of a stochastic approximation to Eq. (5), in which the average is taken over minibatch samples, instead of all N_{tr} training points. Training proceeds for many epochs (complete passes over the train set).

Models are trained to produce residuals at each buoy location over a fixed set of model parameter configurations (hyperparameters) for a maximum of 500 epochs. A Bayesian search algorithm (Snoek et al. 2012) interacts with loss results and searches the hyperparameter space to find the ideal configuration to minimize overall loss from experiment-to-experiment. Experiments are broken up by buoy location and history input lengths of 6, 12, 24 and 48 h. The search space also includes minibatch, learning rate, moment constraints (tanh or hardtanh), and model architecture variation (Table 3). For each buoy and history input length, the hyperparameter space search performs 40 experiments. We chose to limit the number of experiments per historical input to 40 as it provided a good empirical trade-off between finding effective hyperparameters and minimizing total computational cost; in our experience, the hyperparameter search algorithm usually converged in 25–35 steps. Sensitivity tests pertaining to these hyperparameters are conducted and discussed in section 4.

4. Results

Hyperparameter search experiments show the best trained model configurations for all buoy locations is achieved with a 3×3 filter and a model with 8 hidden

layers with the following channels per layer: 32, 32, 64, 64, 128, 128, 256, and 256 (see Fig. 2). Each experiment completes training within 2–5 h on a workstation with a consumer-grade Nvidia GPU. (Note: while training a single SWRL Net model takes several hours, once trained, the model generates a single forecast correction in a fraction of a second.) Either optimal moment constraint (tanh and hardtanh) is satisfactory. Here we compare SWRL Net forecasts to observations, and also comment on the persistence of these corrections, and their frequency dependence.

a. SWRL Net comparison to observations

In initial comparisons, SWRL Net–modeled frequency spectra and direction moments are integrated into typical bulk wave parameters, including significant wave height H_s , mean direction θ_m , and mean period T_m . Significant wave height H_s is estimated by the sum of energy, where

$$H_s = 4 \sqrt{\int E(f) df}. \quad (6)$$

Mean wave direction θ_m is estimate from first-order directional moments:

$$\theta_m = \arctan\left(\frac{b_1}{a_1}\right), \quad (7)$$

where a_1 and b_1 are the first moments of the Fourier series (Longuet-Higgins et al. 1963; Kuik et al. 1988). Mean a_1 and b_1 values are integrated across frequency. Mean periods are estimated from the one-dimensional spectra $E(f)$, where

FIG. 4. Autocorrelation ρ of errors in WW3 significant wave height (blue), mean period (red), and mean direction (yellow) at (a) Grays harbor, (b) Pt. Reyes, and (c) Harvest buoy sites. The dashed line indicates the e -folding time scale.

$$T_m = \frac{\int E(f) df}{\int f E(f) df}. \quad (8)$$

Optimized SWRL Net models show improved forecasts during varying conditions throughout the test set. As an illustrative example, Fig. 3 shows 7 days of sequential forecast corrections. With 6 prior hours of observations and 24 forecast hours, corrections show improvement in predicted wave heights, periods and directions in most cases. Model corrections to mean period and mean direction also appear skillful despite any obvious patterns between errors in the recent past and future (Figs. 3b,c). For example, corrections to wave height, direction and period are generally in the correct direction even if history shows a reversed bias as observed for 9 January 2006 (Fig. 3a, dark blue). On average, SWRL Net corrections show substantial reduction in WW3 prediction errors (Table 2). Improvements are highest, 30%–50%, in the first six hours, declining rapidly for wave heights, and persisting longer for mean periods and directions. These findings are discussed further in section 4b. Skill improvements for extreme conditions show similar results to average conditions; however, SWRL Net improvements are lower and persist for fewer forecast hours (Table 2). This is unsurprising because during training SWRL Net was exposed to comparatively fewer extreme events than to average conditions, and as such, is less skillful at correcting them. These results, however, may indicate the need to weight extremes more strongly in future neural network training to improve such corrections.

b. Persistence of SWRL Net corrections

SWRL Net skill in forecasting mean period and direction corrections persists longer than for significant wave height (Table 2). This result is further explored by considering the decorrelation time scale of WW3 prediction errors in significant wave height, mean period and mean direction at the three buoy locations as quantified by the e -folding time scale of lagged correlation.

At Grays Harbor, the decorrelation time scale for prediction errors in H_s is 6 h, in comparison to 9 and 14 h for T_m and θ_m , respectively (Fig. 4a). The decorrelation time scale for prediction errors in H_s (θ_m) increases to 18 (25) hours at Pt. Reyes and 20 (28) hours at Harvest, and the mean period T_m decorrelation time scale is 13 h at both these buoy locations (Figs. 4b,c). Overall, compared to total energy, WW3 prediction errors in direction and frequency have higher decorrelation time scales, and therefore SWRL Net is able to predict corrections farther into the future.

c. SWRL Net skill versus frequency

SWRL Net skill also varies across frequency. In Fig. 5, the root mean squared error (RMSE) of energy is compared across frequency for WW3 energy predictions, and the relative improvement by SWRL Net corrections are reported. In this comparison, biases, as a function of frequency, were removed from WW3 RMSE estimates to avoid an inflated effect to SWRL Net corrections owing to a simple learned bias correction. Biases are removed from WW3 RMSE such that

FIG. 5. Percentage change in RMSE with SWRL Net adjustment (see color bar) vs forecast time and frequency at (a) Grays Harbor, (b) Point Reyes, and (c) Harvest buoy locations. The dashed line delineates swell and seas frequencies. (d)–(f) Mean predicted, observed, and RMSE of energy vs frequency at each buoy site.

$RMSE = (1/T)\sqrt{\sum_{i=1}^T (p_i - o_i - b)^2}$, where predictions p and observations o are taken at each time step i . Bias b is defined by, $b = (1/T)\sum_{i=1}^T (p_i - o_i)$. On average, errors are reduced by 30%–50% for locally generated seas frequencies (≥ 0.105 Hz) during the first 6 forecast hours (Figs. 5a–c). Error reductions are smallest for remotely generated swell (0.05–0.105 Hz). Patterns are similar across buoy locations, despite varying wave climatology. An increase in prediction error occurs at just one buoy site (Point Reyes) at 0.05–0.06 Hz after 6 forecast hours (see red in Fig. 4b).

Mean energy and WW3 RMSE indicate where improvements have the most impact on integrated parameters such as H_s and where the most opportunity lies for prediction correction (Figs. 5d–f). Note that WW3 RMSE is sometimes larger than the mean, notably for lower frequencies, indicating poor WW3 skill at these frequencies. However, RMSE is a penalizing metric for intermittent low-frequency energy where small timing

errors in predictions result in large errors in comparison to an overall small mean. Integrated parameters, such as significant wave height tend to smooth out these rapid fluctuations of energy in frequency space and exhibit much lower errors in comparison to mean values. Nonetheless, frequency details in Fig. 5 show where WW3 predictions are poor, where energy is significant, and to what extent SWRL Net corrections improve predictions. Error reductions are greatest for local seas bands and smallest for remotely generated swell (0.05–0.105 Hz), with an exception for very low-frequency, low-energy swell where error reductions are modest. The increase in correction skill at very low frequency is surprising, but may be due to the comparatively low benchmark set by large errors in WW3 predictions (Figs. 5d–f). While corrections to seas frequencies are high, improvement to swell predictions are relatively low at all sites. This may owe to the rapid arrival and decline of swell energy compared to

TABLE 3. Hyperparameter search space used to tune SWRL Net. This space was optimized using a Bayesian hyperparameter search algorithm. (Snoek et al. 2012).

Hyperparameter	Description	Options
Constraint	The constraint applied to moments at the output layer of SWRL Net ($n = 2$)	constraint $\in \{\tanh, \text{hardtanh}\}$
Minibatch size	The number of input–output pairs drawn uniformly at random from the dataset to compute each gradient ($n = 4$)	minibatch size $\in \{16, 32, 48, 64\}$
Learning rate	The step size used during minibatch stochastic gradient descent; selected from a continuous set of values	learning rate $\in [10^{-7}, 10^{-5}]$
Downsampling schematic	Dependent on the history input size ($h \in \{6, 12, 24, 48\}$), this allows the search algorithm to insert hidden layers with the specified, decreasing number of filters prior to the main convolutional network; None skips this phase ($n = 5$ for each history input size)	$h = 6 \in \{\text{None}, [18], [18, 18], [24, 12], [24, 24, 12, 12]\}$ $h = 12 \in \{\text{None}, [24], [24, 24], [32, 16], [32, 32, 16, 16]\}$ $h = 24 \in \{\text{None}, [36], [36, 36], [48, 24], [48, 48, 24, 24]\}$ $h = 48 \in \{\text{None}, [60], [60, 60], [80, 40], [80, 80, 40, 40]\}$
Hidden layer schematic	The number of filters per convolutional hidden layer used in SWRL Net ($n = 9$)	$[8, 8, 16, 16, 32, 32], [8, 8, 16, 16, 32, 32, 64, 64, 128, 128],$ $[8, 8, 16, 16, 32, 32, 64, 64, 128, 128, 256, 256], [16, 16,$ $32, 32, 64, 64], [16, 16, 32, 32, 64, 64, 128, 128], [16, 16,$ $32, 32, 64, 64, 128, 128, 256, 256], [32, 32, 64, 64, 128,$ $128], [32, 32, 64, 64, 128, 128, 256, 256]$

the relatively slow growth and decline of locally generated seas energy on the U.S. West Coast. The autocorrelation of errors in energy at swell frequencies fall off comparatively faster (not shown here) and potentially contribute to SWRL Net's lower skill in swell correction at larger forecast hours.

SWRL Net correction patterns are similar across buoy locations, despite some degree of variability in wave climatology. The wave patterns along the eastern Pacific Ocean are driven by a combination of locally generated waves and offshore swell (Wu et al. 2020; Adams et al. 2008). For a common source of offshore storm, the arrival and decline of swell wave energy is expected to behave similarly across these locations. Subsequently, even though wind forcing varies as a function of latitude, the modeled accuracy is controlled by errors in wind forcing, which is expected to be similar along the eastern Pacific without any particular regional bias.

5. Discussion

In general, SWRL Net training was mostly insensitive to hyperparameter configurations and history length input. Additionally, SWRL Net models were found to generalize well across buoy locations. Details and implications for future application and training are discussed below.

a. SWRL Net sensitivity

SWRL Net was tuned using a guided Bayesian search algorithm to maximize performance on the held out validation set, searching over the space of hyperparameters described in Table 3. In this search, 40

hyperparameter configurations were evaluated for each history length input duration ($h \in \{6, 12, 24, 48\}$). The statistics of these models' performance as a function of hyperparameter are shown for Grays Harbor in Fig. 6. In general, SWRL Net is relatively insensitive to training hyperparameters and history length input (see section 3). Similar patterns are observed across buoy locations and therefore results from just Grays Harbor are present. The hyperparameters, enumerated in Table 3, affect model performance minimally, as shown in sensitivity in overall model skill (MSE in Fig. 6). The constraint hyperparameter option (tanh or hardtanh), minibatch size, and hidden layer schematic (not shown) make little difference on resulting performance. Learning rate, however, does affect performance, with the best results obtained by using larger learning rates. Further increasing the learning rate tended to cause the model to diverge during training.

In theory, the model history length balances two competing interests: 1) providing the most information to SWRL Net while training, and 2) avoiding overfitting to specific historic conditions. Here, despite long training sets (~ 10 years), shorter history lengths yield slightly lower error during training, and consequently train faster due to reduced computational requirements. Given these findings, a 6-h history length was used for the experiments in this paper. Bulk parameters, integrated across swell frequencies, show the lowest overall RMSE for SWRL Net swell predictions (Fig. 7) and similarly for seas frequencies (not shown).

b. Model generalization

Long-term (~ 10 years) wave observations and predictions were used in SWRL Net model training,

FIG. 6. Box-and-whisker plots showing the distribution of model mean squared error (MSE) as a function of (a) history length, (b) directional moment constraint, (c) learning rate, and (d) minibatch size. Statistics are computed over the 160 models trained on the Grays Harbor buoy, resulting from the four directed hyperparameter searches. The red bar indicates mean, the blue box spans the 25th–75th percentiles, the black lines span the minimum and maximum values, and the red crosses denote outliers.

but in practice concurrent observations and historical spectra predictions may be sparse. It is therefore relevant to determine if models trained on data in one location are skillful with data from another. If the models are generalizable to some extent, training sets could in practice be constructed from multiple locations and used throughout a region. Additionally, for buoy locations with minimal data, a model can be pretrained with regional data composed of nearby datasets to adapt initial weights before model training.

Across the three locations examined here, spanning the continental U.S. West Coast, SWRL Net generalizes surprisingly well. The wave climate at each site varies substantially, though in general each site is exposed to both long period swell from North and South Pacific storms, and locally generated wind seas. Model corrections to swell appear least generalizable, with predictions at Grays Harbor using Pt. Reyes- or Harvest-trained models exhibiting poor skill (Figs. 8a,d,g). However, Harvest and Pt. Reyes models appear nearly interchangeable, with skill only slightly better for models

trained and implemented at the same location (Fig. 8 center and right panels). Such results may owe to their relatively closer proximity (~ 450 km) compared to Grays Harbor (1000+ km). With the exception of the first 1–3 h, the model is not well generalized for swell energy between California (Pt. Reyes, Harvest) and Washington (Grays Harbor); however, the similarities in Pt. Reyes and Harvest models show promise for regional generalization.

Model generalization for seas predictions shows similar results, but with more potential. Though Pt. Reyes- and Harvest-trained models perform much worse at Gray's Harbor, errors do not exceed uncorrected WW3 predictions, and additionally, improvement holds through the first six hours (Figs. 9a,d,g). This is similarly true for Grays Harbor-trained models implemented on Pt. Reyes and Harvest datasets. Error results suggest improvement at all prediction hours for wave height and direction (Figs. 9b,c,e,f). Though local winds likely vary across buoy locations, the SWRL Net models appear more generalizable across these frequencies suggesting that

FIG. 7. RMSE improvement for Swell frequency bulk parameters: (a)–(c) significant wave height H_s , (d)–(f) mean direction θ_m , and (g)–(i) mean period T_m . SWRL Net predictions (colored lines) are compared to WW3 predictions (black) and WW3 bias removed predictions (black dashed).

model errors may be simpler, and less nuanced compared to swell frequencies.

Overall, initial results suggest that trained SWRL Net models are not site specific and one model trained at one buoy location may be applied in the local region. Additionally, where long training sets are unavailable, multiple locations could be combined to train a local model. Here, analysis was restricted to consistent buoy types and products (CDIP, Datawell buoys) and observations are from relatively deep water (≥ 40 m). Caution, however, may be needed if training with varying buoy types (some contain biases,

O'Reilly et al. 1996), strongly varying depths, or partially sheltered observations.

c. Future application

While SWRL Net represents an important contribution to the wave modeling community, further study is warranted. Owing to a lack of archived spectral WW3 predictions, hindcast predictions were used as a proxy. WW3 hindcasts, driven by reanalysis meteorological hindcasts likely have higher skill than real-time forecasts predicted without ground-truth observations. This higher skilled proxy for forecasts

FIG. 8. Comparison of swell bulk parameters for SWRL Net trained at each location and introduced to data from all other locations: (a)–(c) significant wave height H_s RMSE, (d)–(f) mean direction θ_m RMSE, and (g)–(i) mean period T_m RMSE. SWRL Net predictions (colored lines) are compared to WW3 predictions (black) and WW3 bias removed predictions (black dashed).

leaves less room for improvement, and likely serves as a comparatively more difficult benchmark for SWRL Net. Future application to a growing archive of spectral forecasts may result in improved corrections extending farther in time. Additionally, given observed model generalizability, models pretrained on hindcast may be skillful despite short forecast training sets.

Here, SWRL Net is trained on directional moments, measured by directional wave buoys and integrated from predicted directional WW3 spectra. In turn, the model predicts corrections to these parameters. The ideal network would train with observed frequency-directional

spectra, offering detailed spectra corrections. However, high-resolution directional spectra observations are rare while buoy observations commonplace. Future model development may incorporate detailed spectra corrections from training with observed buoy moments but will require an additional subjective constraint on spectral smoothness (or other properties). Various constraints have been developed to estimate directional spectra from buoy moments (e.g., [Lygre and Krogstad 1986](#)) though each fit the same data ([Ochoa and Delgado-González 1990](#)) and are not necessarily optimal ([Crosby et al. 2016](#)). Nonetheless, when 2D frequency-directional spectra are needed to force local

FIG. 9. As in Fig. 8, but for sea frequencies.

models, corrected SWRL Net forecasts may either be transformed under a selected constraint, or corrected directional information discarded and uncorrected direction predictions may be combined with corrected energy. Depending on the complexity of the directional wave field, uncorrected WW3 directional predictions may or may not be more accurate than directional predictions estimated from buoy directional moments.

6. Summary

Assimilation techniques for wave observations, notably directional buoy observations, do not yet exist in

current state of the art models. Machine learning, therefore, provides an alternative where short-term forecast corrections may be learned from historical observations and hindcasts. Spectral Wave Residual Learning Network (SWRL Net) is a fully convolutional deep learning network developed to improve numerical wave forecasts with directional wave buoy observations. SWRL Net is trained with directional buoy observations (energy and directional moments as a function of frequency) and collocated frequency-directional spectral wave predictions transformed into buoy moments. The network generates forecast corrections up to 24 h in the future. SWRL Net is tested at three locations

along the U.S. West Coast where long-term buoy records are available. Lacking long-term archives of spectral forecast, hindcast predictions from NOAA's CFSR phase 2 are used as a forecast proxy. Wave frequencies are limited to 0.04–0.25 Hz, typical of surface gravity waves. SWRL Net is trained to output the residuals augmenting the WW3 forecasts, conditioned on recent WW3 forecasts and buoy observations. Modeled corrections have skill across all frequencies, though corrections to wind waves (>0.1 Hz) are skillful farthest into the future. Forecast errors are reduced up to 50% in the first three forecast hours, and 10%–40% thereafter. Improvements to direction and wave period estimates remain longest, with reductions of 30%. Model training exhibited little sensitivity to the length of history input, slightly favoring shorter histories (i.e., 6 h of prior wave predictions and observations). Trained SWRL Net models were tested interchangeably at the three locations along the U.S. West Coast and showed skill even at locations for which the model had never seen. The observed generalizability suggests that future training sets could be composed of data from multiple locations, critical for locations with short observational records.

REFERENCES

- Adams, P. N., D. L. Inman, and N. E. Graham, 2008: Southern California deep-water wave climate: Characterization and application to coastal processes. *J. Coastal Res.*, **244**, 1022–1035, <https://doi.org/10.2112/07-0831.1>.
- Ali, M., and R. Prasad, 2019: Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. *Renewable Sustainable Energy Rev.*, **104**, 281–295, <https://doi.org/10.1016/j.rser.2019.01.014>.
- Allahdadi, M. N., B. Gunawan, J. Lai, R. He, and V. S. Neary, 2019a: Development and validation of a regional-scale high-resolution unstructured model for wave energy resource characterization along the U.S. East Coast. *Renewable Energy*, **136**, 500–511, <https://doi.org/10.1016/j.renene.2019.01.020>.
- , R. He, and V. S. Neary, 2019b: Predicting ocean waves along the U.S. East coast during energetic winter storms: Sensitivity to whitecapping parameterizations. *Ocean Sci.*, **15**, 691–715, <https://doi.org/10.5194/os-15-691-2019>.
- Ardhuin, F., and Coauthors, 2010: Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation. *J. Phys. Oceanogr.*, **40**, 1917–1941, <https://doi.org/10.1175/2010JPO4324.1>.
- Bengio, Y., 2009: Learning deep architectures for AI. *Found. Trends Mach. Learn.*, **2**, 1–127, <http://doi.org/10.1561/22000000006>.
- Boehm, A. B., N. S. Ismail, L. M. Sassoubre, and E. A. Andruszkiewicz, 2017: Oceans in peril: Grand challenges in applied water quality research for the 21st century. *Environ. Eng. Sci.*, **34**, 3–15, <https://doi.org/10.1089/ees.2015.0252>.
- Camus, P., F. J. Mendez, R. Medina, A. Tomas, and C. Izaguirre, 2013: High resolution downscaled ocean waves (DOW) reanalysis in coastal areas. *Coastal Eng.*, **72**, 56–68, <https://doi.org/10.1016/j.coastaleng.2012.09.002>.
- Cavaleri, L., J. Alves, F. Ardhuin, and A. Babanin, 2007: Wave modelling—The state of the art. *Prog. Oceanogr.*, **75**, 603–674, <https://doi.org/10.1016/j.pocean.2007.05.005>.
- Chawla, A., D. M. Spindler, and H. L. Tolman, 2013: Validation of a thirty year wave hindcast using the climate forecast system reanalysis winds. *Ocean Modell.*, **70**, 189–206, <https://doi.org/10.1016/j.ocemod.2012.07.005>.
- Cornejo-Bueno, L., J. Nieto-Borge, P. García-Díaz, G. Rodríguez, and S. Salcedo-Sanz, 2016: Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm—extreme learning machine approach. *Renewable Energy*, **97**, 380–389, <https://doi.org/10.1016/j.renene.2016.05.094>.
- Crosby, S. C., W. C. O'Reilly, and R. T. Guza, 2016: Modeling long-period swell in Southern California: Practical boundary conditions from buoy observations and global wave model predictions. *J. Atmos. Oceanic Technol.*, **33**, 1673–1690, <https://doi.org/10.1175/JTECH-D-16-0038.1>.
- , B. D. Cornuelle, W. C. O'Reilly, and R. T. Guza, 2017: Assimilating global wave model predictions and deep-water wave observations in nearshore swell predictions. *J. Atmos. Oceanic Technol.*, **34**, 1823–1836, <https://doi.org/10.1175/JTECH-D-17-0003.1>.
- , N. Kumar, W. C. O'Reilly, and R. T. Guza, 2019: Regional swell transformation by backward ray tracing and SWAN. *J. Atmos. Oceanic Technol.*, **36**, 217–229, <https://doi.org/10.1175/JTECH-D-18-0123.1>.
- Deo, M., A. Jha, A. Chaphekar, and K. Ravikant, 2001: Neural networks for wave forecasting. *Ocean Eng.*, **28**, 889–898, [https://doi.org/10.1016/S0029-8018\(00\)00027-5](https://doi.org/10.1016/S0029-8018(00)00027-5).
- Deshmukh, A. N., M. C. Deo, P. K. Bhaskaran, T. M. Balakrishnan Nair, and K. G. Sandhya, 2016: Neural-network-based data assimilation to improve numerical ocean wave forecast. *IEEE J. Oceanic Eng.*, **41**, 944–953, <https://doi.org/10.1109/JOE.2016.2521222>.
- Ellenson, A., Y. Pei, G. Wilson, H. T. Özkan-Haller, and X. Fern, 2020: An application of a machine learning algorithm to determine and describe error patterns within wave model output. *Coastal Eng.*, **157**, 103595, <https://doi.org/10.1016/j.coastaleng.2019.103595>.
- Emmanouil, G., G. Galanis, and G. Kallos, 2010: A new methodology for using buoy measurements in sea wave data assimilation. *Ocean Dyn.*, **60**, 1205–1218, <https://doi.org/10.1007/s10236-010-0328-9>.
- , —, and —, 2012: Combination of statistical Kalman filters and data assimilation for improving ocean waves analysis and forecasting. *Ocean Modell.*, **59–60**, 11–23, <https://doi.org/10.1016/j.ocemod.2012.09.004>.
- Fiedler, J. W., P. B. Smit, K. L. Brodie, J. McNinch, and R. Guza, 2018: Numerical modeling of wave runup on steep and mildly sloping natural beaches. *Coastal Eng.*, **131**, 106–113, <https://doi.org/10.1016/j.coastaleng.2017.09.004>.
- García-Medina, G., H. T. Özkan-Haller, P. Ruggiero, R. A. Holman, and T. Nicolini, 2018: Analysis and catalogue of sneaker waves in the U.S. Pacific Northwest between 2005 and 2017. *Nat. Hazards*, **94**, 583–603, <https://doi.org/10.1007/s11069-018-3403-z>.
- Goodfellow, I., Y. Bengio, and A. Courville, 2016: *Deep Learning*. MIT Press, 800 pp., <http://www.deeplearningbook.org>.
- Guza, R., and F. Feddersen, 2012: Effect of wave frequency and directional spread on shoreline runup. *Geophys. Res. Lett.*, **39**, L11607, <https://doi.org/10.1029/2012GL051959>.
- Halpern, B. S., and Coauthors, 2012: An index to assess the health and benefits of the global ocean. *Nature*, **488**, 615–620, <https://doi.org/10.1038/nature11397>.

- He, K., X. Zhang, S. Ren, and J. Sun, 2016: Deep residual learning for image recognition. 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, IEEE, 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
- James, S. C., Y. Zhang, and F. O'Donncha, 2018: A machine learning framework to forecast wave conditions. *Coastal Eng.*, **137**, 1–10, <https://doi.org/10.1016/j.coastaleng.2018.03.004>.
- Komar, P., and D. Inman, 1970: Longshore sand transport on beaches. *J. Geophys. Res.*, **75**, 5914–5927, <https://doi.org/10.1029/JC075i030p05914>.
- Kuik, A. J., G. P. van Vledder, L. H. Holthuijsen, A. J. Kuik, G. P. van Vledder, and L. H. Holthuijsen, 1988: A method for the routine analysis of pitch-and-roll buoy wave data. *J. Phys. Oceanogr.*, **18**, 1020–1034, [https://doi.org/10.1175/1520-0485\(1988\)018<1020:amftra>2.0.co;2](https://doi.org/10.1175/1520-0485(1988)018<1020:amftra>2.0.co;2).
- Kumar, N., D. L. Cahill, S. C. Crosby, and G. Voulgaris, 2017: Bulk versus spectral wave parameters: Implications on Stokes drift estimates, regional wave modeling, and HF radars applications. *J. Phys. Oceanogr.*, **47**, 1413–1431, <https://doi.org/10.1175/JPO-D-16-0203.1>.
- Londhe, S. N., and V. Panchang, 2006: One-day wave forecasts based on artificial neural networks. *J. Atmos. Oceanic Technol.*, **23**, 1593–1603, <https://doi.org/10.1175/JTECH1932.1>.
- Longuet-Higgins, M., D. Cartwright, and N. Smith, 1963: Observations of the directional spectrum of sea waves using the motions of a floating buoy. *Proc. Conf. on Ocean Wave Spectra*, Easton, MD, Prentice Hall, 111–136.
- Lygre, A., and H. E. Krogstad, 1986: Maximum entropy estimation of the directional distribution in ocean wave spectra. *J. Phys. Oceanogr.*, **16**, 2052–2060, [https://doi.org/10.1175/1520-0485\(1986\)016<2052:meeotd>2.0.co;2](https://doi.org/10.1175/1520-0485(1986)016<2052:meeotd>2.0.co;2).
- Maas, A. L., A. Y. Hannun, and A. Y. Ng, 2013: Rectifier nonlinearities improve neural network acoustic models. *Proc. 30th Int. Conf. on Machine Learning*, Atlanta, GA, JMLR, Vol. 28, 6 pp., http://robotics.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf.
- Makarynskyy, O., 2004: Improving wave predictions with artificial neural networks. *Ocean Eng.*, **31**, 709–724, <https://doi.org/10.1016/j.oceaneng.2003.05.003>.
- Ochoa, J., and O. E. Delgado-González, 1990: Pitfalls in the estimation of wind wave directional spectra by variational principles. *Appl. Ocean Res.*, **12**, 180–187, [https://doi.org/10.1016/S0141-1187\(05\)80025-2](https://doi.org/10.1016/S0141-1187(05)80025-2).
- O'Reilly, W. C., and R. T. Guza, 1998: Assimilating coastal wave observations in regional swell predictions. Part I: Inverse methods. *J. Phys. Oceanogr.*, **28**, 679–691, [https://doi.org/10.1175/1520-0485\(1998\)028<0679:acwoir>2.0.co;2](https://doi.org/10.1175/1520-0485(1998)028<0679:acwoir>2.0.co;2).
- , T. H. C. Herbers, R. J. Seymour, and R. T. Guza, 1996: A comparison of directional buoy and fixed platform measurements of Pacific swell. *J. Atmos. Oceanic Technol.*, **13**, 231–238, [https://doi.org/10.1175/1520-0426\(1996\)013<0231:acodba>2.0.co;2](https://doi.org/10.1175/1520-0426(1996)013<0231:acodba>2.0.co;2).
- Orzech, M. D., J. Veeramony, and H. Ngodock, 2013: A variational assimilation system for nearshore wave modeling. *J. Atmos. Oceanic Technol.*, **30**, 953–970, <https://doi.org/10.1175/JTECH-D-12-00097.1>.
- Paszke, A., and Coauthors, 2017: Automatic differentiation in PyTorch. *31st Conf. on Advances in Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, 4 pp.
- Smit, P. B., I. A. Houghton, K. Jordanova, T. Portwood, E. Shapiro, D. Clark, M. Sosa, and T. T. Janssen, 2020: Assimilation of distributed ocean wave sensors. arXiv: 2003.04435, <https://arxiv.org/abs/2003.04435>.
- Snoek, J., H. Larochelle, and R. P. Adams, 2012: Practical Bayesian optimization of machine learning algorithms. *25th Conf. on Advances in Neural Information Processing Systems (NIPS 2012)*, Lake Tahoe, NV, 2951–2959, <https://papers.nips.cc/paper/4522-practical-bayesian-optimization-of-machine-learning-algorithms.pdf>.
- Stopa, J. E., 2018: Wind forcing calibration and wave hindcast comparison using multiple reanalysis and merged satellite wind datasets. *Ocean Modell.*, **127**, 55–69, <https://doi.org/10.1016/j.ocemod.2018.04.008>.
- Tolman, H. L., 2009: User manual and system documentation of WAVEWATCH-III, version 3.14. National Oceanic and Atmospheric Administration Tech. Rep., 220 pp.
- Veeramony, J., D. Walker, and L. Hsu, 2010: A variational data assimilation system for nearshore applications of SWAN. *Ocean Modell.*, **35**, 206–214, <https://doi.org/10.1016/j.ocemod.2010.07.008>.
- Walker, D. T., 2006: Assimilation of SAR imagery in a nearshore spectral wave model. Office of Naval Research Tech. Rep., 39 pp., <https://apps.dtic.mil/dtic/tr/fulltext/u2/a445814.pdf>.
- Wu, W.-C., T. Wang, Z. Yang, and G. García-Medina, 2020: Development and validation of a high-resolution regional wave hindcast model for U.S. west coast wave resource characterization. *Renewable Energy*, **152**, 736–753, <https://doi.org/10.1016/j.renene.2020.01.077>.