



HAL
open science

Deep Learning for Predicting Significant Wave Height From Synthetic Aperture Radar

Brandon Quach, Yannik Glaser, Justin Stopa, Alexis Mouche, Peter Sadowski

► **To cite this version:**

Brandon Quach, Yannik Glaser, Justin Stopa, Alexis Mouche, Peter Sadowski. Deep Learning for Predicting Significant Wave Height From Synthetic Aperture Radar. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59 (3), pp.1859-1867. <10.1109/TGRS.2020.3003839>. <hal-04203389>

HAL Id: hal-04203389

<https://hal.science/hal-04203389v1>

Submitted on 26 Sep 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Deep Learning for Predicting Significant Wave Height From Synthetic Aperture Radar

Brandon Quach¹, Yannik Glaser², Justin Edward Stopa³, Alexis Aurélien Mouche⁴, and Peter Sadowski

Abstract—The Sentinel-1 satellites equipped with synthetic aperture radars (SARs) provide near-global coverage of the world’s oceans every six days. We curate a data set of collocations between SAR and altimeter satellites and investigate the use of deep learning to predict significant wave height from SAR. While previous models for predicting geophysical quantities from SAR rely heavily on feature-engineering, our approach learns directly from low-level image cross-spectra. Training on collocations from 2015 to 2017, we demonstrate on test data from 2018 that deep learning reduces the state-of-the-art root mean squared error by 50%, from 0.6 to 0.3 m when compared to altimeter data. Furthermore, we isolate the contributions of different features to the model performance.

Index Terms—CWAVE, deep learning, machine learning, neural networks, Sentinel-1, significant wave height, synthetic aperture radar (SAR).

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) enables us to measure submesoscale phenomena with unprecedented coverage, resolution, and frequency. By measuring the backscatter from the ocean surface, SAR captures information about ocean swells and sea surface roughness at high spatial resolutions (<10 m) [1], from which many oceanic, atmospheric, and biologic phenomena can be identified [2]. The two Sentinel-1 satellites of the European Space Agency (ESA) take regular SAR measurements of the ocean surface, together covering the entire globe every six days [3], and have already accumulated more than 600 TB of level-1 (L1) wave mode data. However, in order to take full advantage of this technology and the torrent of data being produced, new methods are needed to extract useful information from the high-dimensional measurements.

Sea state information extracted from SAR has been instrumental in understanding swell decay [1], [4], [5], improving swell propagation in numerical models [6], and predicting swell amplitudes and arrivals times by assimilation into numerical models [7]. SAR can also be used to estimate extreme

sea states in extra-tropical and tropical cyclones [8]–[10]. A geophysical quantity of particular interest is the *significant wave height*, H_s , defined as the mean of the top third of a wave height distribution, and estimating H_s from SAR has immediate practical uses in alerting ships to dangerously large waves. Traditional “inverse” algorithms for inferring H_s from SAR are slow and perform poorly in windy conditions typical of most storms [11], [12] because of the complex nonlinear mechanism involved in the image synthesis when observing moving scenes. As a result, several recent studies have focused on data-driven statistical models [8]–[10], [13].

Previous data-driven approaches for predicting H_s from SAR used small data sets of buoy observations as targets for training (<5000 examples) [14]–[16], or numerical models of global wave generation such as WAVEWATCH3 [8], [10], [13], [17]. The current state-of-the-art method uses a neural network trained on the latter, and predicts H_s with 0.6-m root mean squared error (RMSE) [10]. However, the WAVEWATCH3 targets are only an *estimate* of H_s and are known to be unreliable in high sea states [18]–[20].

Furthermore, the neural network in [10] relies on a reduced representation of the modulation cross-spectra: a set of 22 engineered features known as CWAVE [13]. Such dimensionality-reduction methods can be very useful, but often come at the cost of discarding relevant information. We hypothesize that the SAR image modulation spectra contains *additional* information about H_s that is lost by the CWAVE dimensionality-reduction step. We propose to *learn* the relevant intermediate data representations using deep learning with artificial neural networks, similar to what has been done in other fields from computer vision [21] to high-energy physics [22]–[24].

In this work, we address both limitations of current data-driven H_s prediction models. First, we curate a data set containing *direct observations* of ocean wave heights by identifying 750,000 collocations of SAR and altimeter satellites. Second, we train a statistical model to extract information directly from low-level SAR image spectra using deep learning. Finally, we analyze the importance of the different inputs to this model, and its performance in different settings.

II. DATA AND METHODS

A. Sensors, Collocations and Preprocessing

Our first contribution is a data set of historical measurements from two types of polar-orbiting satellites: Sentinel-1 SAR satellites and altimeter satellites. Because the satellites are in different orbits, their paths intersect, providing

Manuscript received February 14, 2020; revised May 22, 2020; accepted June 8, 2020. Date of publication July 17, 2020; date of current version February 25, 2021. (Corresponding author: Justin Edward Stopa.)

Brandon Quach is with the Computing and Mathematical Sciences Department, California Institute of Technology, Pasadena, CA 91125-0002 USA, and also with the Information and Computer Sciences Department, University of Hawai’i at Mānoa, Honolulu, HI 96822 USA.

Yannik Glaser and Peter Sadowski are with Information and Computer Sciences Department, University of Hawai’i at Mānoa, Honolulu, HI 96822 USA.

Justin Edward Stopa is with Ocean Engineering Department, University of Hawai’i at Mānoa, Honolulu, HI 96822 USA (e-mail: stopa@hawaii.edu).

Alexis Aurélien Mouche is with the Univ. Brest, CNRS, IRD, IFREMER, Laboratoire d’Océanographie Physique et Spatiale (LOPS), IUEM, 29280 Brest, France.

Color versions of one or more of the figures in this article are available online at <https://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2020.3003839

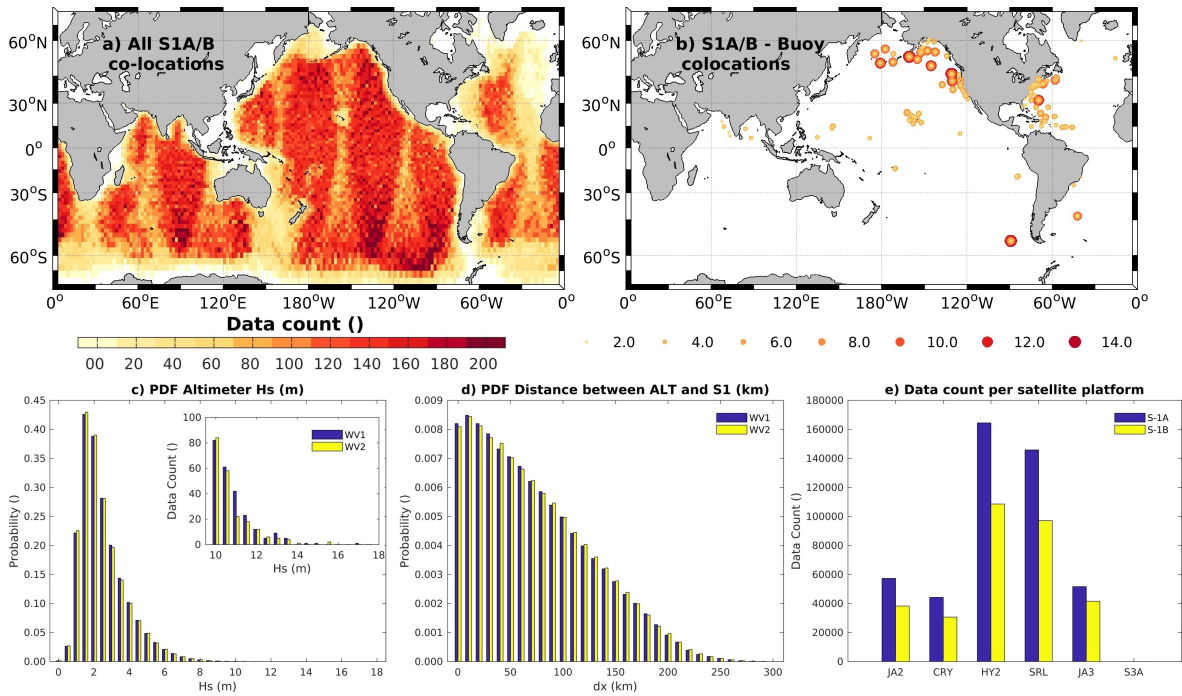


Fig. 1. Collocations between S-1 and satellite altimetry and moored buoys. (a) Total number of S-1/ALT collocations in $2 \times 2^\circ$ bins. (b) Locations of the S-1/buoy collocations, the colors and size of the markers indicating the buoy H_s . (c) Histogram of H_s for S-1/ALT for WV1 and WV2, with the inset showing the tail for extreme sea states. (d) Histogram of distance between the S-1/ALT acquisitions for WV1 and WV2. (e) Bar graph showing number of collocations from each of the altimeter-SAR combinations.

measurements at roughly the same time and location. Specifically, we use pairs of measurements taken within 3 h and 200 km apart; if this condition is satisfied by multiple measurement pairs by the same set of satellites on the same pass, then only the closest pair is used. This process results in 753,777 collocation events from 2015 through 2018 that are well-distributed geographically (Fig. 1). The majority of collocations are within 100 km and 1 h (68% and 76%, respectively). These events have both SAR imaging from Sentinel-1 and H_s from an altimeter, and provide a high-fidelity reference data set.

1) *Sentinel-1*: The S-1 mission is a constellation of two (A&B) polar-orbiting, sun-synchronous satellites equipped with SARs [3]. The two satellites have the same orbit and cross the equator at approximately 0600 or 1800 local time, with a 180° phase difference to provide an effective six-day repeat cycle. The SARs are active sensors that emit and record electromagnetic signals with wavelengths of 5.5 cm (C-band). Although S-1 has four exclusive acquisition modes, in this study, we focus solely on wave mode (WV), the default acquisition mode over open ocean. WV was specifically designed for sensitivity to the modulation of ocean surface gravity waves with footprints of 20×20 km and a 4-m spatial resolution. S-1 WV primarily operates at two incidence angles: 23° (WV1) and 36° (WV2) from nadir. Acquisitions are made every 100 km along the orbit alternating between WV1 and WV2 along its track so that images with the same incidence angle are separated by 200 km. The S-1 WV can image the ocean surface in either VV (default) or HH polarization; in this study, we use VV data only, because this is the default mode

of S-1 and the number of HH acquisitions is insufficient to create a robust model for the global ocean.

We show the spatial coverage of collocations between S-1 WV and altimeters in Fig. 1(a). The spatial coverage is the same as the entire S-1A/B monthly coverage. Every month there are $\sim 32,000$ observations for each incidence angle and S-1A/B, giving $\sim 130,000$ observations per month. There are few WV acquisitions over the Arctic Ocean, Northeast Atlantic, coastal seas, and closed seas (Red, Black, Mediterranean, and Caribbean seas) because S-1 uses other imaging modes in these regions. S-1A and S-1B went into routine data collection after their calibration and validation stages in October 2015 and July 2016, respectively. This work is based on the image modulation spectra, NRCS (or σ_0), and normalized variance (nv) of the image intensity included in the ESA-Copernicus WV level 2 (L2) product.

2) *Altimeter Data*: We leverage a multiplatform altimeter data product which has been calibrated and quality controlled by Young *et al.* [25] and Ribal and Young [26]. The 1-Hz altimeter data set estimates significant wave heights with spatial footprints of 6–10 km and is an updated version of [27] that includes all available altimeter missions. In particular, it contains six overlapping mission from 2014 to 2018: Jason-2 (JA2), CRYOSAT-2 (CRY), Haiyan-2 (HY2), AltiKa SARAL (SRL), Jason-3 (JA3), and Sentinel-3A (S3A). Each mission is quality controlled and calibrated with respect to moored buoys and cross-calibrated between platforms. The cross-calibration is extremely important as it enables the collective merged database to be relatively consistent between platforms and improves consistency over time [25]. An altimeter H_s RMSE

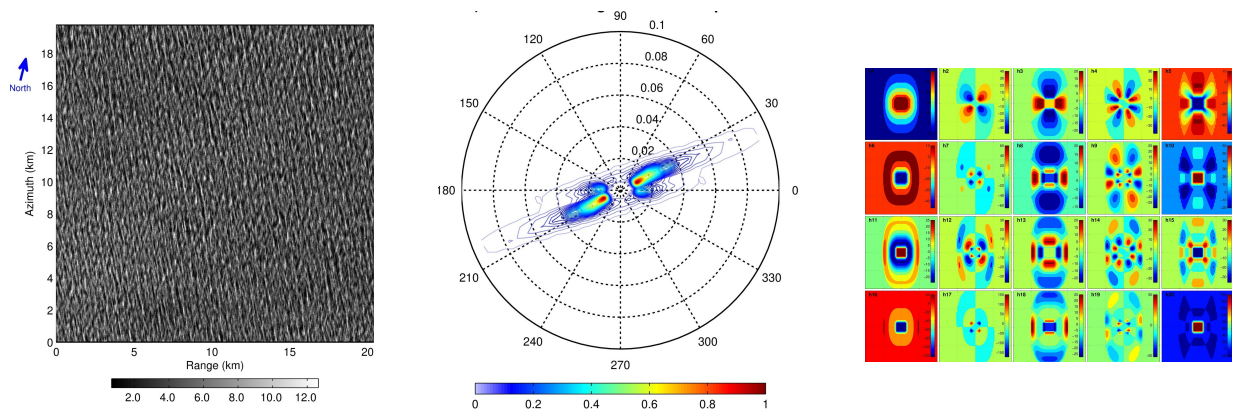


Fig. 2. (Left) Level 1 SAR image covering a square 20×20 km area. (Center) Real component of the image spectra obtained by taking the 2-D Fourier transform. (Right) 20 orthogonal CWAVE basis functions designed to summarize the image spectra. The inputs to the DNN are the real and imaginary components of the image spectra, represented as two 72×60 matrices.

is typically within the range of 10–30 cm [27], [28]. This calibrated multiplatform database is collectively called ALT herein.

3) *Buoy Observations*: Several buoy networks have openly accessible data archives, and we provide additional validation of our results using data from the National Data Buoy Center (NDBC), Coastal Data Information Program (CDIP), Marine Environmental Data Service (MEDS), the collaborative project OceanSITES, and the Ocean Observatories Initiative (OOI). Fig. 1(b) shows the locations of the buoys, which are mostly located in the Northern Hemisphere. The buoy data set contains 19,721 collocations with S-1, with 71 having wave heights larger than 8 m (0.36%). Many of the largest wave heights are recorded at the OOI buoy located in the Southern Ocean offshore of Chile, an undersampled region of the ocean.

4) *SAR Image Spectra*: The Sentinel-1 SAR data set consists of the real and imaginary components computed from SAR modulation cross spectra. Each data point within the cross spectra was created by taking the Level 1 SAR image with 5×5 m pixel resolution covering a 20×20 km area and applying a 2-D Fourier transformation to different “looks” within the dwell time [29] to obtain the real and imaginary modulation spectra [30] (see the real part in the middle of Fig. 2). The modulation spectra consists of two matrices (real and imaginary) of shape 72×60 corresponding to wavenumbers and directions. These two matrices were then stacked to form the input tensor with shape $72 \times 60 \times 2$. The SAR image spectra were then preprocessed by centering and scaling the real and imaginary image modulation spectra separately—each pixel was normalized by subtracting the overall mean and dividing by the overall standard deviation of all pixels and all collocations.

5) *High-Level Features*: In addition to the SAR image spectra, our data set includes a number of high-level features that describe the observing conditions, along with the 22 CWAVE features that are derived from the image modulation spectra. The CWAVE features are obtained by mapping the image modulation spectra onto an orthogonal basis set of 20-nondimensional parameters (right of Fig. 2).

TABLE I
HIGH-LEVEL FEATURES

Indices	Description
1-20	CWAVE nondimensional orthogonal basis set.
21	CWAVE normalized radar cross section σ^0 .
22	CWAVE normalized variance of radar cross section.
23	Time of day.
24-25	Latitude (sine and cosine).
26-27	Longitude (sine and cosine).
28	Incidence angle.
29	Incidence angle mode (binary flag, WV1 or WV2).
30	Satellite source (binary flag, S-1A or S-1B).
31	Time difference between ALT and SAR.
32	Spatial difference between ALT and SAR.

CWAVE was originally developed for the second European Remote Sensing Satellite (ERS2), then used by ENVISAT [8] and Sentinel-1A [10] to predict H_s .

The observing conditions are described in terms of time of day, satellite platform, incidence angle, latitude, and longitude. We also include the time and distance between the SAR and altimeter satellite observations for each collocation event. This provides a rough estimate of how much we can trust the altimeter measurement to provide an accurate target since sea states can change faster than our time and space constraints. Obviously, these are only available for the training data, and so they are set to zero at inference time, but in any case we found the impact of this variable to be small in our models (experiments not shown). In total, there are 32 high-level features for each collocation (Table I).

6) *Data Split*: The data are split into training, validation, and test sets based on the year of collection. Collocation events from 2015 to 2016 were used as the training set, events from 2017 was used as a validation set, and events from 2018 was used as held-out test set. The result was 303,574 training examples, 265,052 validation examples, and 185,151 test examples. The validation set was used for learning rate annealing, early stopping, and hyperparameter selection (i.e., optimization parameters such as learning rate), while the test set was only used for the final evaluation of the model.

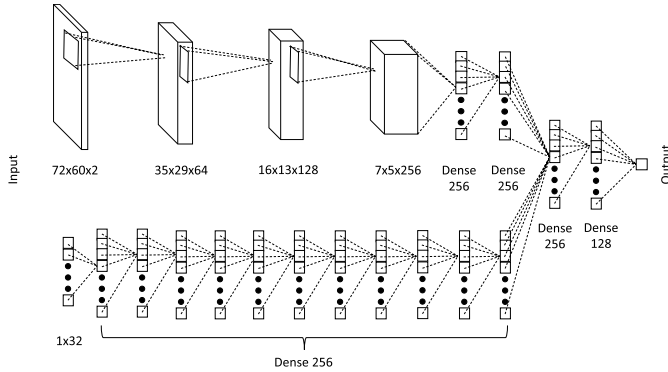


Fig. 3. DNN architecture with two input types. (Top) SAR image spectra comprising one real and one imaginary channel. (Bottom) 32 scalar-valued features. The SAR images are processed by multiple 2-D convolution layers before the two branches of the network are combined by three dense layers at the output. We predict H_s in this work, but we expect that the same model architecture could be used to predict other sea state parameters given an adequate training data set.

B. Deep Learning

1) *Deep Neural Network (DNN) Architecture*: We propose a DNN architecture that predicts H_s from SAR image spectra and high-level features. The model starts as two branches with separate inputs: one which processes the spectral input and another which processes the high-level features (Fig. 3). The spectral input branch takes an input tensor of the shape (72, 60, 2) where the real and imaginary components of the image spectra are stacked along the third axis, analogous to the “colors” of an RGB image. This input tensor is then fed into three sequential convolutional layers containing 64, 128, and 256 filters, respectively. A filter size of 3×3 is maintained at each convolutional layer with a rectified linear unit (ReLU) activation [31]. In addition, each layer is followed by a max-pooling layer with a 2×2 window. The final convolutional layer is fed into a global max-pooling layer which produces a flattened array of size 256. This is then fed into two additional dense layers with 256 hidden units each with ReLU activation.

The high-level features are normalized to have zero mean and unit variance, with the following exceptions. The time-of-day was encoded as a value between -1 and 1 using the function $f(t) = 2 \sin(2\pi t/48) - 1$; this normalization helps stabilize the DNN optimization. The latitude and longitude were encoded by representing each as an angle in the range $[0, 2\pi)$ then taking the sine and cosine. Binary labels were created to specify the SAR satellite (S1-A or S1-B) to account for any small calibration differences.

The 32 high-level features are fed into 11 dense layers with 256 ReLU hidden units each. Both branches yield a flattened array of size 256 which are then concatenated to form a single vector with 512 features. Two hidden dense layers of 256 and 128 hidden ReLU units then integrate the image spectra branch with the second branch. Finally, an output layer with a dropout [32] rate of 0.337 and *softplus* activation [33] makes the final prediction.

This model is trained to minimize the mean squared error (MSE) using the Adam optimizer [34] with a batch size

TABLE II
HYPERPARAMETER SEARCH SPACE FOR DNN

Hyper-parameter	Range
Batch Size	{128, 256, 512, 1024}
Learning Rate	[0.0001, 0.001]
Dropout	[0.2, 0.5]

of 128 and an initial learning rate of 0.0003. The learning rate was decreased by 20% if the validation set MSE did not improve over four epochs (one epoch is defined as a pass through the data set), and training was stopped when the validation set MSE did not improve over ten epochs. The best model was trained for 35 epochs. The dropout rate, initial learning rate, and batch size were optimized using the SHERPA black-box optimization package for machine learning hyper-parameter tuning [35] on a cluster of Nvidia RTX 2080 Ti GPUs. One hundred models were trained using the random search algorithm to optimize over the search space shown in Table II. The code for training the model is available at <https://github.com/hawaii-ai/SAR-Wave-Height>.

2) *CWAVE Models*: The deep learning approach was compared to simpler models trained on the CWAVE feature set used in [8] and [13], without image spectra. We trained both a linear regression model and a DNN to predict H_s from high-level features alone—either the 22 CWAVE features plus incidence angle and mode, or the extended set of 32 features listed in Table I. The results were nearly the same in each case, suggesting that the additional features (e.g., longitude, latitude, time of day, and so on) contained little additional information.

The CWAVE DNN used for comparison consisted of eleven dense hidden layers of 256 ReLU units, followed by a layer of 64 ReLU units, and a *heteroskedastic Gaussian* output layer [36]. The network outputs two values y_1, y_2 corresponding to the mean and variance of a Gaussian distribution $\mathcal{N}(y_1, y_2)$, where y_2 is restricted to ensure positive variance by defining a custom activation function

$$y_2 = \begin{cases} y_2, & y_2 > 0 \\ \frac{1}{1 - y_2}, & y_2 \leq 0 \end{cases}.$$

Weights were initialized using the scaling suggested by [37], and the conditional log-likelihood of the target values was maximized using that Adam optimizer with minibatches of size 1024. The initial learning rate of 0.003 decayed starting at epoch 300 with decay rate 0.0005 applied at the end of each epoch. A dropout rate of 0.5 was applied to the penultimate layer. Training was stopped when the validation loss did not improve after 15 epochs. The architecture, learning rate, and early stopping were optimized with SHERPA; see the Appendix for details.

III. RESULTS

A. Data Validation

First, we performed experiments testing the validity of combining data from multiple SAR and altimeter satellites into a single data set. We did this through a series of experiments in

TABLE III
PLATFORM COMPARISON

Data	N	RMSE (m)	Standard Deviation (m)
Incidence Angle Sensitivity			
WV1	78,050	0.39	0.98
WV2	75,598	0.42	0.95
SAR Platform Sensitivity			
S1A	63,199	0.40	0.96
S1B	90,449	0.41	0.96
Expecting S1A	90,449	0.41	0.96
Expecting S1B	63,199	0.41	0.97
Altimeter Platform Sensitivity			
JS2	93,638	0.44	0.93
CRY	74,118	0.38	0.97
HY2	268,046	0.43	0.95
SRL	239,167	0.40	0.91
JS3	93,187	0.43	0.92

which a DNN was trained and evaluated on different subsets of the data. For simplicity, we train the 12-layer neural network described above on the 32 high-level, engineered features (no image spectra).

The DNN was trained on the combined collocations from all satellites—six altimeters and two SAR satellites with two viewing modes—and report the performance on the test set. As expected, the model performs slightly better in the WV1 viewing mode compared to WV2, which has a larger incidence angle (Table III) consistent with prior findings [10]. The average model performance is the same on collocations from both platforms—moreover, this remains true *even when we tell the model it is receiving data from the other platform*, which we can do by manipulating the platform flag feature. This confirms that these satellites are well-calibrated for this task.

In the case of the altimeter platforms, five different NN models are trained, excluding one platform at a time. Results on the left-out platform are reported in Table III. While there is a slight increase in RMSE, this is not surprising, and no satellite has a large error increase. Overall, the results indicate that all five platforms tested within the altimetry data set are well-calibrated to each another.

B. Deep Learning From SAR

Next, we tested whether deep learning could extract information from image spectra that is lost by the CWAVE feature representation. We trained each type of model described above on data from 2015 to 2016, tuned hyperparameters on data from 2017, and then measured performance on events from 2018. Table IV compares the RMSE of the different models on the test set. The linear regression model that uses only the 22 CWAVE features plus incidence angle and wave mode achieves an RMSE of 0.64 m, while the NN on the same features achieves 0.41-m RMSE. The latter is already a significant reduction from the 0.6 m of the previous state-of-the-art machine learning models that were trained on WAVEWATCH3 data [10]. However, the DNN model that uses the image spectra achieves another boost in performance,

TABLE IV
RMSE ON TEST SET

Wave Height	CWAVE Linear	CWAVE NN	Deep NN	Percentage of Total Data
All Waves	0.64	0.41	0.33	100%
<1m	0.84	0.52	0.39	1.4%
1m - 3m	0.52	0.33	0.26	66.4%
3m - 8m	0.78	0.52	0.43	31.8%
>8m	3.20	1.34	1.22	0.4%

TABLE V
FEATURE IMPORTANCE STUDY

Wave Height	No CWAVE	No Latitude Longitude	No Satellite Type	No Incidence Angle Or Mode	All Included
All Waves	0.33	0.33	0.33	0.32	0.33
<1m	0.43	0.42	0.41	0.39	0.39
1m - 3m	0.26	0.26	0.25	0.25	0.26
3m - 8m	0.44	0.43	0.43	0.43	0.43
>8m	1.19	1.19	1.20	1.18	1.22

lowering the RMSE to 0.33 m. The performance improvement is consistent across small, medium, and large waves.

A feature importance study was performed to assess the dependence of the model on each set of high-level features. Four additional models were trained with the same architecture and hyperparameters, but with certain features removed: 1) the 20 CWAVE parameters; 2) the latitude and longitude features; 3) satellite type flag; or 4) incidence angle and wave mode. In each experiment, the other high-level features are unchanged. Table V shows no degradation in performance when removing these features, suggesting that they *contain little or no additional information that is relevant to predicting H_s* given the other inputs. When removing the incidence angle and mode features, we actually see a small reduction in error, which is not entirely unexpected—the inclusion of features that convey no useful information contributes to model overfitting.

In Fig. 4(a), we show the well-established directional dependence of the radar cross section, σ_0 , to the wind direction, which is taken from the Climate Forecast System Reanalysis (CFSR) [38]. However, the residuals of the predicted H_s have no clear dependence on the relative wind direction [Fig. 4(b)], or dominant wave direction (not shown).

Finally, we explore the impact of increasing the size of the training set on the discrepancy between including and not including the CWAVE features. In this experiment, we fix the hyperparameters and train DNNs on variable-sized subsets of the 2015–2017 data (568,626 examples total). Testing is performed using 2018 data. These are trained for a fixed 30 epochs where the learning rate is annealed by a factor of 0.4 every ten epochs (no early stopping). For each subset, we train six repetitions of the same model with different random initialization of the weights. Fig. 5 shows the mean performance of the six networks for the different training set sizes.

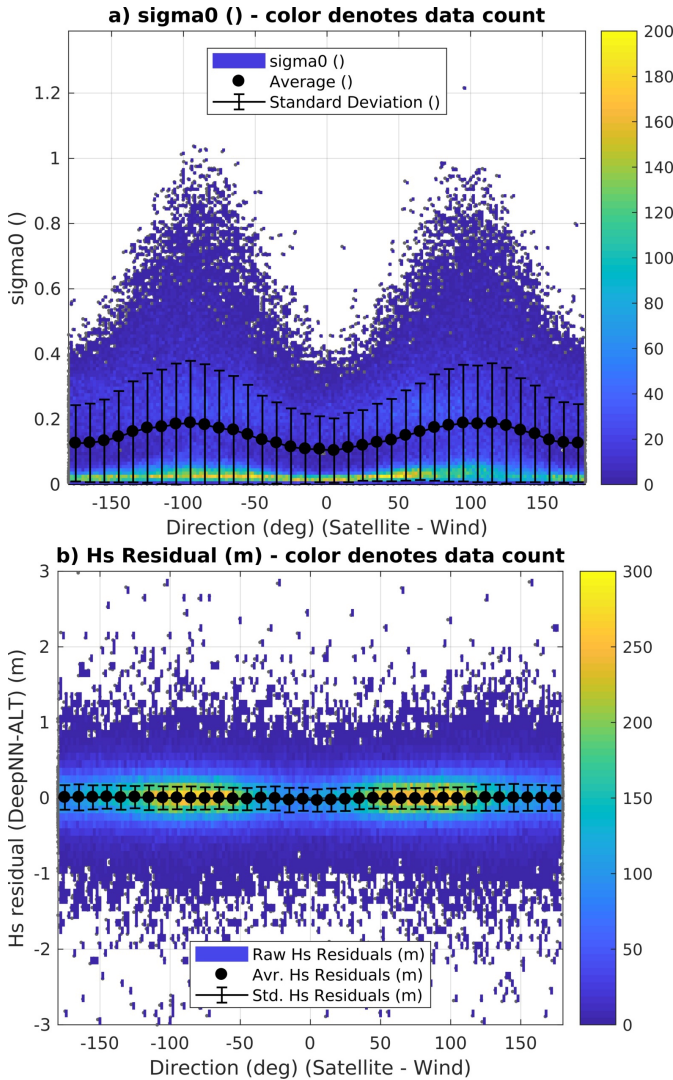


Fig. 4. Plots showing the dependence of (a) radar cross section, σ_0 and (b) H_s residuals (ALT - DNN) with the relative wind direction. The relative direction of the wind (x-axes) is computed by subtracting the wind direction (from CFSR) from the satellite heading.

C. Final Model Performance

We constructed a final ensemble model by averaging the predictions of the six models trained with different random initializations, using all features and the complete training set. This model achieves a test RMSE of 0.307 m compared to altimeter data—a 50% reduction in RMSE from the previous state-of-the-art model of 0.6 m [10]. The H_s RMSEs from our final model are close to the upper limit of the altimeter/buoy errors of 0.30 m.

On the test set of 19,721 buoy collocations, our model achieves 0.53-m RMSE. The error on the buoy data set is larger because the buoys tend to be located at the poles where there are more extreme sea states ($H_s > 8$ m) and larger errors (Fig. 6). Our model has higher error on these extreme sea states because there are fewer of them in the training data set (0.4% of total collocations; Fig. 1).

Fig. 7(a) and (b) shows the relationship between predicted H_s and observations from altimetry and buoys, respectively.

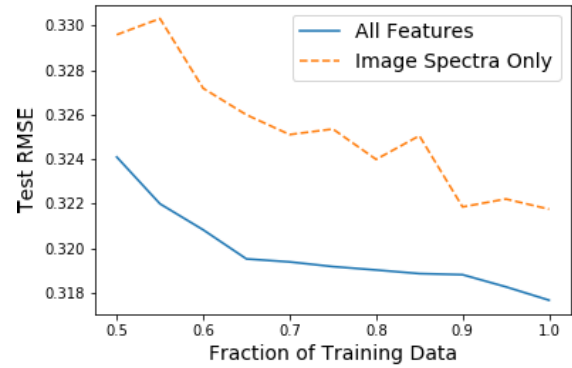


Fig. 5. Performance of DNN with both SAR image spectra and CWAVE parameters (blue line) and only SAR image spectra (orange curve), while varying the training set size. Each point is an average of RMSE objective over six trials with different random weight initializations. We expect the performance gap to decrease and eventually disappear as more training data is acquired.

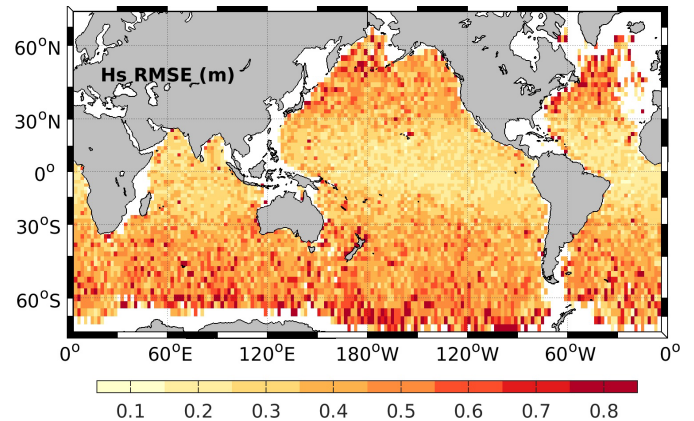


Fig. 6. Map of average test set prediction RMSE (m).

The regression lines have slope less than 1 (0.94 for altimetry), which reflects the tendency of the model to regress toward the mean: overestimating small H_s (< 1 m) and an underestimating large H_s (> 8 m). Fig. 7(c) and (d) shows how the prediction error varies with H_s , with larger errors at the extremes $H_s < 1$ m and $H_s > 8$ m. This is explained by a paucity of training data in this range [see Fig. 7(e) and (f)].

IV. DISCUSSION

The collocation event data set greatly improves the accuracy of data-driven SAR H_s prediction models, and this data set will continue to grow from month to month. We conclude that training on direct observational measurements is more effective than training on predictions from the WAVEWATCH3 numerical model. We have also demonstrated that deep learning can extract useful representations from SAR image spectra that are not captured by the engineered CWAVE features. In a direct comparison between two hyperparameter optimized DNNs, the network with the image spectra information obtained a 20% reduction in RMSE (0.33 versus 0.41 m). This is in keeping with the success of deep learning in other fields, where *learned* features outperform expertly *engineered* features [21], [39]–[42].

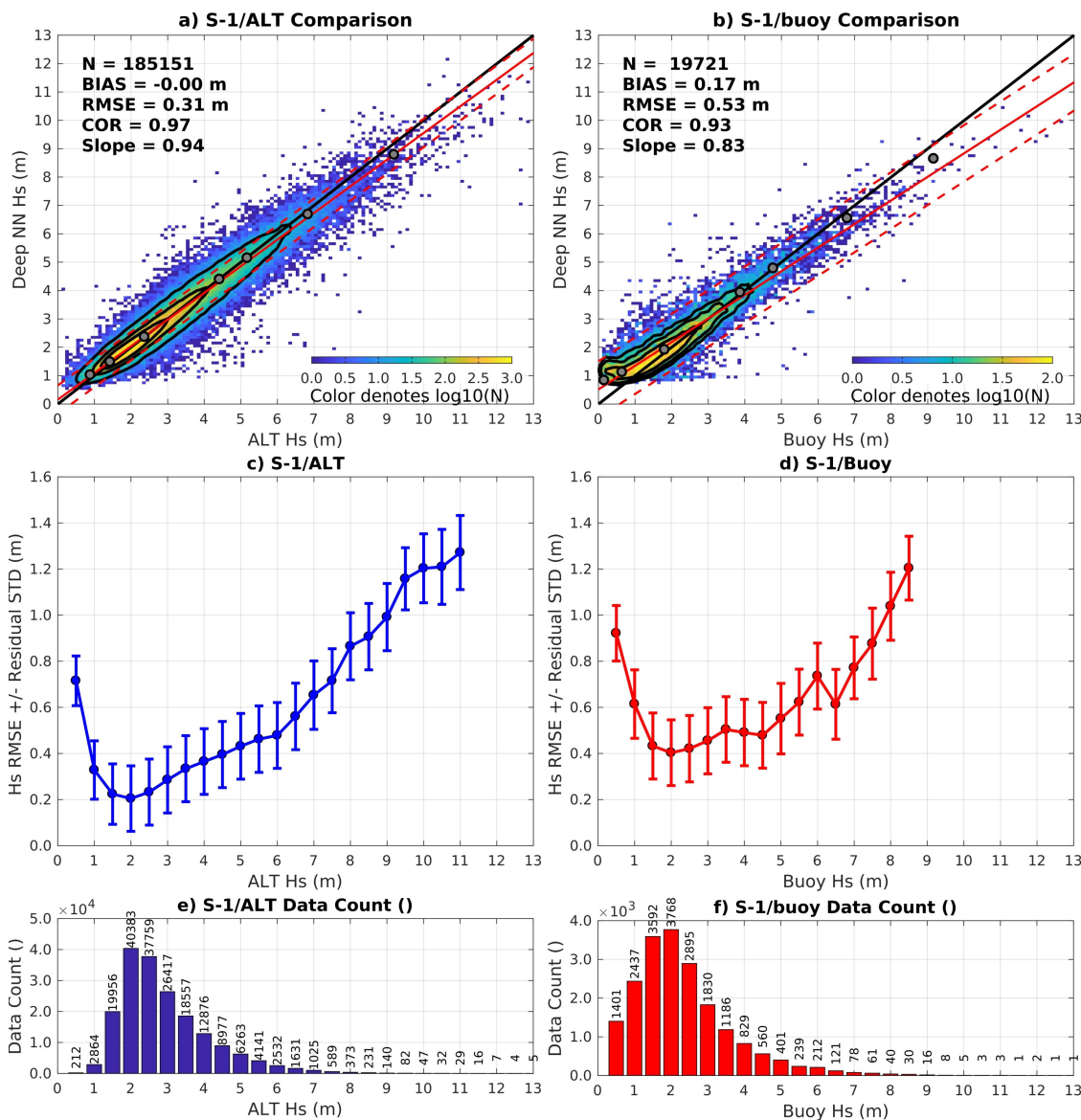


Fig. 7. Comparison of test predictions against (Left) altimeter collocations and (Right) buoy collocations. (a) and (b) Scatter plots of predictions versus measurements. (c) and (d) Plots of prediction RMSE versus measured H_s , with error bars showing the standard deviation. (e) and (f) Histograms of measured H_s where the data count is given in black text. In the top panels, the color denotes data density in 0.1-m bins, solid red lines represent a least square linear regression, and the dashed lines represent 90% of the data. The black contours represent 50%, 75%, and 95% of the data (inner to outer) and the gray dots represent the quantile-quantile points for 1%, 10%, 50%, 90%, 95%, 99%, and 99.9%.

Although our results show that there is still some advantage to including the CWAVE features in the model, this advantage diminishes as the number of training examples increases (Fig. 5). The CWAVE features, being derived from the image data, provide no additional information, and only help bias the model toward a good solution. As the training data set increases, the model is slower to overfit, and the benefit of including the CWAVE features disappears.

Latitude and longitude information can be useful for predicting significant wave heights because there are regional characteristics of the wave climate [43]. However, the feature importance study shows that our model only makes minimal use of this information. This is encouraging, because it implies that the model is relying almost entirely on the direct measurements rather than geographical information. Furthermore, our analysis shows that S-1 satellites are well

calibrated for this product (Table III) meaning our method and model can be used on both platforms to estimate H_s .

The RMSE of our model approaches that between altimeter and buoy measurements. This opens up opportunities to use the SAR data to further study both buoy and altimeter discrepancies through triple collocations (SAR, altimeter, and buoy). Here, we expect that the detailed SAR imagery will provide environmental context such as wave-current interaction or crossing swells which might be difficult for current altimeter algorithms to properly resolve especially in coastal environments.

V. CONCLUSION

We have curated a data set of over 750,000 collocations between radar altimetry and the two Sentinel-1 SAR sensors, and used it to train a DNN regression model that predicts

TABLE VI
HYPERPARAMETER SEARCH SPACE FOR CWAVE DNN

Hyper-parameter	Range
Number of Hidden Layers	{2, 3, 4, 5, 6, 7}
Activation Function	{ReLU, tanh}
Dropout Rate	[0.0, 0.5]
Learning Rate	[0.0005, 0.02]
Batch Size	{128, 256, 512}

H_s from SAR observations with 0.3-m RMSE relative to independent altimeter observations. This reduces the error of the current state-of-the-art approach *by half*. Instead of relying on the set of engineered CWAVE features that capture *most* of the discriminative information, our deep learning approach learns directly from the low-level, high-dimensional image spectra. Furthermore, our results indicate that there is still room for improvement with additional training data, especially in extreme sea states with $H_s > 8$ m. Thus, we expect our model to improve as more collocation events are collected.

APPENDIX CWAVE NEURAL NETWORK TRAINING DETAILS

For a fair comparison with the DNN trained on image modulation spectra, the DNN model trained on only high-level variables underwent its own hyperparameter optimization. The number of hidden layers, the activation function, the dropout rate in the penultimate layer, the learning rate, and the batch size were optimized using the Bayesian optimization algorithm in SHERPA. Overall, 150 models were trained from the hyperparameter space shown in Table VI. The best combination used the maximum number of hidden layers in the specified range (7), so we increased the number of hidden layers in the best model until validation set performance no longer increased. Thus, the final CWAVE DNN used for comparisons in this article has 12 hidden layers.

ACKNOWLEDGMENTS

This work was made possible, thanks to SAR data access granted by ESA projects: Sentinel-1 A Mission Performance Center (4000107360/12/I-LG) and Sentinel-1 Ocean Study (S1-4SCI-16-0002). All Sentinel-1 L2 data used in this study can be obtained from the Copernicus Data Hub (cophub.copernicus.eu). The buoy data can be obtained from the respective centers: NDBC (nodc.noaa.gov/BUOY/), MEDS (meds-sdmm.dfo-mpo.gc.ca), and OceanSITES (www.oceansites.org/). NSF Ocean Observatories Initiative Data Portal, <http://ooinet.oceanobservatories.org>, Surface Wave Spectra (CE02SHSM, CE04OSSM, CE07SHSM, CE09OSSM, GA01SUMO, GI01SUMO, GS01SUMO, CP01CNMS: -SBD1205-WAVSSA000) data from September 10, 2014 to July 31, 2018. Downloaded on July 14, 2018. The altimetry data was sourced from the Integrated Marine Observing System (IMOS) - IMOS is a national collaborative research infrastructure, supported by the Australian Government. IMOS 2014–2018, IMOS - SRS Surface Waves Sub-Facility - altimeter wave/wind, <https://portal.aodn.org.au>, accessed

January 24, 2018. The authors would like to thank NVIDIA for a hardware grant to PS. The technical support and advanced computing resources from the University of Hawai'i Information Technology Services Cyberinfrastructure are gratefully acknowledged.

REFERENCES

- [1] F. Collard, F. Ardhuin, and B. Chapron, "Monitoring and analysis of ocean swell fields from space: New methods for routine observations," *J. Geophys. Res.*, vol. 114, 2009, Art. no. C07023, doi: [10.1029/2008JC005215](https://doi.org/10.1029/2008JC005215).
- [2] C. Wang *et al.*, "A labelled ocean SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode," *Geosci. Data J.*, vol. 6, no. 2, pp. 105–115, Nov. 2019.
- [3] R. Torres *et al.*, "GMES Sentinel-1 mission," *Remote Sens. Environ.*, vol. 120, pp. 9–24, May 2012.
- [4] F. Ardhuin, B. Chapron, and F. Collard, "Observation of swell dissipation across oceans," *Geophys. Res. Lett.*, vol. 36, no. 6, Mar. 2009, doi: [10.1029/2008GL037030](https://doi.org/10.1029/2008GL037030).
- [5] J. E. Stopa, F. Ardhuin, R. Husson, H. Jiang, B. Chapron, and F. Collard, "Swell dissipation from 10 years of Envisat advanced synthetic aperture radar in wave mode," *Geophys. Res. Lett.*, vol. 43, no. 7, pp. 3423–3430, Apr. 2016, doi: [10.1002/2015GL067566](https://doi.org/10.1002/2015GL067566).
- [6] F. Ardhuin *et al.*, "Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation," *J. Phys. Oceanogr.*, vol. 40, no. 9, pp. 1917–1941, Sep. 2010, doi: [10.1175/2010JPO4324.1](https://doi.org/10.1175/2010JPO4324.1).
- [7] L. Aouf and J.-M. Lefevre, "On the impact of the assimilation of saral/altika wave data in the operational wave model mfwam," *Mar. Geodesy*, vol. 38, no. 1, pp. 381–395, Feb. 2015, doi: [10.1080/01490419.2014.1001050](https://doi.org/10.1080/01490419.2014.1001050).
- [8] X.-M. Li, S. Lehner, and T. Bruns, "Ocean wave integral parameter measurements using Envisat ASAR wave mode data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 155–174, Jan. 2011, doi: [10.1109/TGRS.2010.2052364](https://doi.org/10.1109/TGRS.2010.2052364).
- [9] R. Romeiser, H. C. Graber, M. J. Caruso, R. E. Jensen, D. T. Walker, and A. T. Cox, "A new approach to ocean wave parameter estimates from C-band scansar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1320–1345, Mar. 2015, doi: [10.1109/TGRS.2014.2337663](https://doi.org/10.1109/TGRS.2014.2337663).
- [10] J. E. Stopa and A. Mouche, "Significant wave heights from Sentinel-1 SAR: Validation and applications," *J. Geophys. Res., Oceans*, vol. 122, no. 3, pp. 1827–1848, Mar. 2017.
- [11] V. Kerbaol, B. Chapron, and P. W. Vachon, "Analysis of ERS-1/2 synthetic aperture radar wave mode images," *J. Geophys. Res., Oceans*, vol. 103, no. C4, pp. 7833–7846, Apr. 1998, doi: [10.1029/97JC01579](https://doi.org/10.1029/97JC01579).
- [12] J. E. Stopa, F. Ardhuin, B. Chapron, and F. Collard, "Estimating wave orbital velocity through the azimuth cutoff from space-borne satellites," *J. Geophys. Res., Oceans*, vol. 120, no. 11, pp. 7616–7634, Nov. 2015, doi: [10.1002/2015JC011275](https://doi.org/10.1002/2015JC011275).
- [13] J. Schulz-Stellenfleth, T. König, and S. Lehner, "An empirical approach for the retrieval of integral ocean wave parameters from synthetic aperture radar data," *J. Geophys. Res.*, vol. 112, Mar. 2007, Art. no. C03019, doi: [10.1029/2006JC003970](https://doi.org/10.1029/2006JC003970).
- [14] M. Bruck and S. Lehner, "TerraSAR-x/TanDEM-X sea state measurements using the XWAVE algorithm," *Int. J. Remote Sens.*, vol. 36, no. 15, pp. 3890–3912, Jul. 2015, doi: [10.1080/01431161.2015.1051630](https://doi.org/10.1080/01431161.2015.1051630).
- [15] M. J. Collins, M. Ma, and M. Dabbar, "On the effect of polarization and incidence angle on the estimation of significant wave height from SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4529–4543, Jul. 2019.
- [16] A. Pleskachevsky, S. Jacobsen, B. Tings, and E. Schwarz, "Estimation of sea state from Sentinel-1 synthetic aperture radar imagery for maritime situation awareness," *Int. J. Remote Sens.*, vol. 40, no. 11, pp. 4104–4142, Jan. 2019.
- [17] The WAVEWATCH, "User manual and system documentation of WAVE WATCH III version 5.16," NOAA/NWS/NCEP/MMAB, College Park, MD, USA, Tech. Rep. 329, 2016.
- [18] J. A. Hanafin *et al.*, "Phenomenal sea states and swell from a North Atlantic storm in February 2011: A comprehensive analysis," *Bull. Amer. Meteorol. Soc.*, vol. 93, no. 12, pp. 1825–1832, Dec. 2012.
- [19] J. E. Stopa, F. Ardhuin, A. Babanin, and S. Zieger, "Comparison and validation of physical wave parameterizations in spectral wave models," *Ocean Model.*, vol. 103, pp. 2–17, Jul. 2016.

- [20] J. E. Stopa, "Wind forcing calibration and wave hindcast comparison using multiple reanalysis and merged satellite wind datasets," *Ocean Model.*, vol. 127, pp. 55–69, Jul. 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1463500318301458>
- [21] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural," in *Proc. Neural Inf. Process. Syst.*, 2014, pp. 1–9.
- [22] P. Baldi, P. Sadowski, and D. Whiteson, "Searching for exotic particles in high-energy physics with deep learning," *Nature Commun.*, vol. 5, no. 1, Sep. 2014, Art. no. 4308.
- [23] P. Baldi, P. Sadowski, and D. Whiteson, "Enhanced higgs boson to $\tau^+ \tau^-$ search with deep learning," *Phys. Rev. Lett.*, vol. 114, no. 11, Mar. 2015, Art. no. 111801.
- [24] P. Sadowski, J. Collado, D. Whiteson, and P. Baldi, "Deep learning, dark knowledge, and dark matter," in *Proc. NIPS*, 2015, pp. 81–87.
- [25] I. R. Young, E. Sanina, and A. V. Babanin, "Calibration and cross validation of a global wind and wave database of altimeter, radiometer, and scatterometer measurements," *J. Atmos. Ocean. Technol.*, vol. 34, no. 6, pp. 1285–1306, Jun. 2017.
- [26] A. Ribal and I. R. Young, "33 years of globally calibrated wave height and wind speed data based on altimeter observations," *Scientific Data*, vol. 6, no. 1, pp. 1–5, May 2019.
- [27] S. Zieger, J. Vinoth, and I. R. Young, "Joint calibration of multiplatform altimeter measurements of wind speed and wave height over the past 20 years," *J. Atmos. Ocean. Technol.*, vol. 26, no. 12, pp. 2549–2564, Dec. 2009, doi: [10.1175/2009JTECHA1303.1](https://doi.org/10.1175/2009JTECHA1303.1).
- [28] P. Queffelec, "Long-term validation of wave height measurements from altimeters," *Mar. Geodesy*, vol. 27, nos. 3–4, pp. 495–510, Jul. 2004.
- [29] G. Engen and H. Johnsen, "SAR-ocean wave inversion using image cross spectra," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 4, pp. 1047–1056, Jul. 1995, doi: [10.1109/36.406690](https://doi.org/10.1109/36.406690).
- [30] H. Johnsen and F. Collard, "Sentinel-1 ocean swell wave spectra (OSW) algorithm definition," Northern Research Institute, Tromsø, Norway, Tech. Rep. S1-TN-NRT-52-7450, MPC-0469, Feb. 2020, no. 1.3.
- [31] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
- [32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [33] C. Dugas, Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia, "Incorporating second-order functional knowledge for better option pricing," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 472–478.
- [34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [35] L. Hertel, J. Collado, P. Sadowski, J. Ott, and P. Baldi, "Sherpa: Robust hyperparameter optimization for machine learning," Cornell Univ., Ithaca, NY, USA, Tech. Rep., 2020. [Online]. Available: <https://arxiv.org/abs/2005.04048>
- [36] C. M. Bishop, "Mixture density networks," Tech. Rep., 1994.
- [37] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
- [38] S. Saha *et al.*, "The NCEP climate forecast system version 2," *J. Climate*, vol. 27, no. 6, pp. 2185–2208, Mar. 2014, doi: [10.1175/JCLI-D-12-00823.1](https://doi.org/10.1175/JCLI-D-12-00823.1).
- [39] P. Sadowski, B. Radics, Ananya, Y. Yamazaki, and P. Baldi, "Efficient antihydrogen detection in antimatter physics by deep learning," *J. Phys. Commun.*, vol. 1, no. 2, Sep. 2017, Art. no. 025001. [Online]. Available: <http://stacks.iop.org/2399-6528/1/i=2/a=025001>
- [40] A. Lusci, G. Pollastri, and P. Baldi, "Deep architectures and deep learning in chemoinformatics: The prediction of aqueous solubility for drug-like molecules," *J. Chem. Inf. Model.*, vol. 53, no. 7, pp. 1563–1575, Jul. 2013.
- [41] D. K. Duvenaud *et al.*, "Convolutional networks on graphs for learning molecular fingerprints," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2215–2223.
- [42] R. Beck, Y. Glaser, P. Sadowski, and I. Szapudi. (2019). *Refined Redshift Regression in Cosmology With Graph Convolution Networks*. [Online]. Available: https://ml4physicalsciences.github.io/files/NeurIPS_ML4PS_2019_80.pdf
- [43] J. E. Stopa, K. F. Cheung, H. L. Tolman, and A. Chawla, "Patterns and cycles in the climate forecast system reanalysis wind and wave data," *Ocean Model.*, vol. 70, pp. 207–220, Oct. 2013, doi: [10.1016/j.ocemod.2012.10.005](https://doi.org/10.1016/j.ocemod.2012.10.005).



Brandon Quach was born in Honolulu, HI, USA. He is pursuing the bachelor's degree with Computing and Mathematical Sciences Department, Caltech, Pasadena, CA, USA.

He completed this project under Caltech's Summer Undergraduate Research Fellowship Program as a Visiting Researcher at the University of Hawai'i at Mānoa, Honolulu. His interests lie in machine learning applications, computer vision, and probabilistic models.



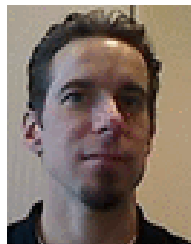
Yannik Glaser received the bachelor's degree in computer science from the University of North Georgia, Dahlonega, GA, USA, in 2018. He is pursuing the master's degree with the University of Hawai'i at Mānoa, Honolulu, HI, USA, in 2020, where he will then continue his Ph.D. degree in computer science.

Since beginning his studies at the University of Hawai'i at Mānoa. He has begun working as part of the UH Data Science Fellows on applied machine learning in the physical sciences and medical imaging.



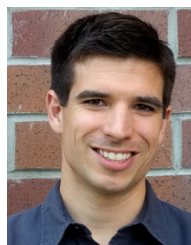
Justin Edward Stopa received the M.S. and Ph.D. degrees from the University of Hawai'i at Mānoa, Honolulu, HI, USA, in 2007 and 2013, respectively. His dissertation focused on the wind and wave climate.

His post-doctoral research was performed at IFREMER and focused on ocean applications with SAR. He is an Assistant Professor of ocean and resources engineering with the University of Hawai'i at Mānoa. His research interests and expertise in physical oceanography, wind and wave climate, spectral wave modeling, oceanic remote sensing, and applied statistics.



Alexis Aurélien Mouche received the master's degree in physics for remote sensing from the University of Pierre et Marie Curie, Paris, France, in 2002, and the Ph.D. degree in physics with a focus on ocean remote sensing from CETP/IPSL/CNRS (National Research Center), Paris, in 2005.

In 2006, he joined the Laboratoire d'Océanographie Spatiale, IFREMER, Brest, France, for two years with a post-doctoral position granted by the French spatial agency (CNES) to work on the ocean dynamics as observed at high resolution. From 2008 to 2013, he was a Research Engineer with BOOST-Technologies and CLS (CNES and IFREMER subsidiary) to develop algorithms for wind, waves, and current measurements with synthetic aperture radar from space. Since January 2014, he has been with the Laboratoire d'Océanographie Physique and Spatiale, IFREMER, where he is a Research Scientist. His principal research interest includes the interactions of electromagnetic and oceanic waves for ocean remote sensing applications.



Peter Sadowski received the bachelor's degree from the California Institute of Technology, Pasadena, CA, USA, in 2009, and the Ph.D. degree from the University of California Irvine, Irvine, CA, USA, in 2016.

He is currently an Assistant Professor of computer science at the University of Hawai'i at Mānoa, Honolulu, HI, USA. His research focuses on machine learning, artificial intelligence, and data science.