Deep Sensing of Ocean Wave Heights with Synthetic Aperture Radar

Brandon Quach,[1,2] **Yannik Glaser,**[2] **Justin Stopa,**[3] **Peter Sadowski**[2*]
[1]Computing and Mathematical Sciences, California Institute of Technology
[2]Information and Computer Sciences, University of Hawai'i at Mānoa
[3]Ocean Resources and Engineering, University of Hawai'i at Mānoa

Abstract

The Sentinel-1 satellites equipped with synthetic aperture radars (SAR) provide near global coverage of the world's oceans every six days. We curate a data set of co-locations 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 co-locations from 2015-2017, 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 meters to 0.3 meters.

Introduction

Synthetic aperture radar (SAR) is an important remote sensing technology able to achieve high spatial resolution (< 10 meter). From SAR satellite data, geophysical properties can be predicted using statistical models, enabling researchers to monitor global sea states with unprecedented coverage, precision, frequency, and without the use of complicated SAR modulation transfer functions (Schulz-Stellenfleth, König, and Lehner 2007). Sea state information provides scientific value in understanding the propagation of waves (Collard, Ardhuin, and Chapron 2009; Stopa et al. 2016) and the effects of climate change (Young, Zieger, and Babanin 2011), as well as immediate practical benefits such as alerting ships to dangerously large waves created by storms.

SARs capture sea surface roughness and many other geophysical phenomena (Wang et al. 2019). Therefore, predicting ocean wave signatures from SAR images typically requires *feature engineering* — a dimensionality-reduction step that extracts task-specific information. CWAVE is a common feature set for describing wave properties in SAR as a basis of 20 orthogonal features derived from the SAR modulation spectra. CWAVE has been used to estimate the significant wave height for the SARs aboard: 1) ERS-2 (Schulz-Stellenfleth, König, and Lehner 2007), 2) ENVISAT (Li, Lehner, and Bruns 2011), and Sentinel-1 (Stopa and Mouche 2017; Pleskachevsky et al. 2019) linking SAR imaging to vital sea state information. Such reduced representations of high-dimensional data can be useful when fitting statistical models to relatively small data sets, but they are also limiting; task-relevant information is almost-certainly lost when reducing a high-dimensional SAR image to the low-dimensional CWAVE feature space.

In this work, we attempt to extract *additional* information from SAR images using *deep learning* with artificial neural networks. Deep learning has proven to be an extremely effective approach to representation learning, leading to rapid advances in diverse fields such as computer vision (Krizhevsky, Sutskever, and Hinton 2014), high-energy physics (Baldi, Sadowski, and Whiteson 2014; Sadowski and Baldi 2018), and chemistry (Lusci, Pollastri, and Baldi 2013; Duvenaud et al. 2015). Deep learning has the potential to make similar advances in remote sensing for oceanography by extracting information directly from SAR modulation cross-spectra.

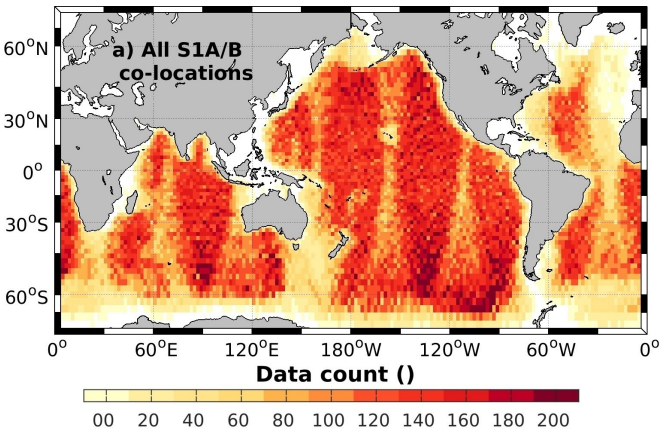

Figure 1: Spatial distribution of co-locations between Sentinel-1 SAR satellites and altimeter satellites, in $2 \times 2°$ bins.

In this work, we first curate a data set of over 750,000 co-locations of SAR and altimeter satellites, which provides SAR in conjunction with direct measurements of ocean wave heights. The data is used to train deep neural networks

*peter.sadowski@hawaii.edu

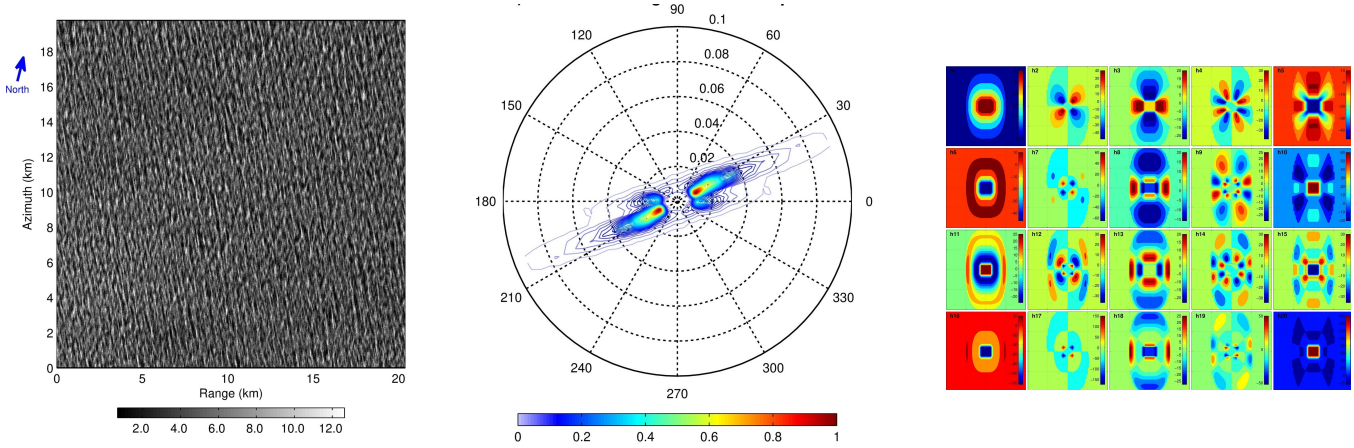

Figure 2: Level 1 SAR image covering a square 20×20 km area (left); real component of the image spectra obtained by taking the 2D Fourier transform (center); and 20 orthogonal CWAVE basis functions designed to summarize the image spectra (right). The inputs to the deep neural network are the real and imaginary components of the image spectra, represented as two 72×60 matrices.

to predict *significant wave height*, H_s, defined as the mean of the top third of a wave height distribution. We compare training on SAR image spectra vs. high-level CWAVE features, and analyze the effect of training data set size.

Methods

Data

We curate a training set of historical data from two types of polar-orbiting satellites: Sentinel-1 SAR satellites and altimeter satellites. Because the satellites are on different trajectories, their paths frequently intersect, providing measurements from roughly the same location at the same time. Specifically, the data set is constructed using measurements that are less than 3 hours apart and with spatial differences less than 300 km, resulting in 753,777 *co-location* events from 2015 through 2018 that are geographically well distributed (Figure 1). These events have both SAR imaging from Sentinel-1 and significant wave height from an altimeter, and provide a high-fidelity reference data set (Ribal and Young 2019).

The data set is split into training, validation, and test sets based on year of data collection. Co-location events from 2015 and 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 hyper-parameter selection, while the test set was only used for the final evaluation of the model.

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 2D Fourier transformation to different "looks" within the dwell time (Engen and Johnsen 1995) to obtain the real and imaginary modulation spectra (Johnsen and Collard 2009) (Figure 2). The modulation spectra consists of two matrices (real and imaginary) of shape 72×60 with one dimension corresponding to wavenumber and the other direction. These two matrices were then stacked to form the input tensor with shape $72 \times 60 \times 2$. The 1-Hz altimeter dataset estimates significant wave heights with spatial footprints of 6 to 10 km. The altimeter dataset consists of data merged from 6 different altimeter missions and has been cross-calibrated between platforms, as in Ribal and Young (Ribal and Young 2019). The SAR image spectra were then pre-processed 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 co-locations.

In addition to the SAR image spectra, we include a number of high-level features in our model. First, we include the time and distance between the satellite co-location measurements, normalized to have zero mean and unit variance — while this information is only available during training. The time (or distance) between satellite observations provides a rough estimate of how much we can trust the altimeter measurement to provide an accurate target because sea states can change faster than our original time and space constraints. These features are simply set to zero at prediction time. Second, the time-of-day was encoded as a value between -1 and 1 using the function $f(t) = 2sin(\frac{2\pi t}{48}) - 1$; this normalization helps stabilize the neural network optimization. Third, latitude and longitude were encoded by representing each as an angle in the range $[0, 2\pi)$ then taking the sine and cosine, resulting in four features total. Fourth, a binary label was created to specify the SAR satellite; S1-A or S1-B are calibrated to produce comparable data, but there could be small differences. Finally, we also include the 20 non-dimensional CWAVE parameters that are derived from the image spectra, each normalized using standard scaling to have zero mean and unit variance over the training examples.

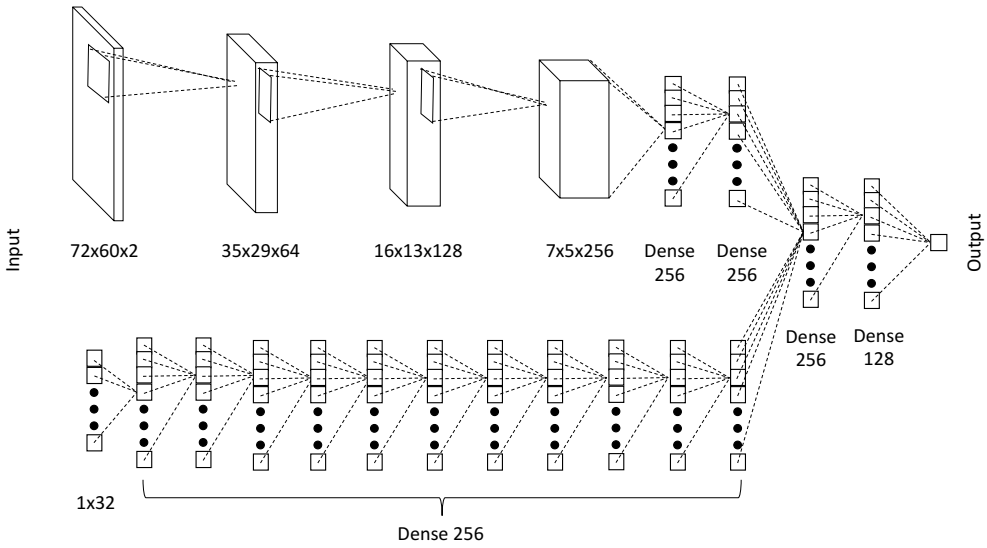

Figure 3: Deep neural network architecture with two input types: SAR image spectra comprising one real and one imaginary channel (top), and 32 scalar-valued features (bottom). The SAR images are processed by multiple 2D convolution layers before the two branches of the network are combined by three dense layers at the output.

Deep Learning

Deep Neural Network Architecture We propose a deep neural network architecture that predicts significant wave height by using the input data from SAR image spectra. The model starts as two branches with separate inputs: one which processes the spectral input and another which extracts information from the high-level features (Figure 3). The spectral input branch takes an input tensor of the shape (72, 60, 2) where the real and imaginary values of the Fourier transform are stacked along the third axis, analogous to the 'colors' of an RGB image. This input tensor is then fed sequentially into three 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. 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 non-spectral data branch consists an input layer of the following 32 features:

- 20 CWAVE features
- 1 time of day feature
- 2 latitude features (sine and cosine)
- 2 longitude features (sine and cosine)
- 1 incidence angle feature
- 1 incidence angle mode feature (binary flag representing WV1 or WV2)
- 1 satellite source feature (binary flag representing Sentinel-1A and Sentinel-1B)
- 1 time difference between altimeter and sentinel measurements feature
- 1 spatial difference between altimeter and sentinel measurements feature
- 1 normalized radar cross section σ^0 feature
- 1 normalized variance of radar cross section feature.

These 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 of 0.337 and softplus activation makes the final prediction.

This model is trained to minimize the mean squared error (MSE) using the Adam optimizer (Kingma and Ba 2014) with a batch size 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 4 epochs, and training was stopped when the validation set MSE did not improve over 10 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 (Hertel et al. 2018) on a cluster Nvidia RTX 2080 Ti GPUs. One hundred models were trained using the random search algorithm to optimize over the search space shown in Table 1.

CWAVE Models To measure the advantage of the deep learning approach over simpler models, we also trained two models that predict the significant wave height from the 32

Table 1: Hyper-parameter Search Space

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

high-level features alone: a simple linear regression model and a deep neural network. The neural network consisted of eleven dense hidden layers of 256 ReLU units, followed by a layer of 64 ReLU units, and two outputs. The two outputs correspond to a heteroskedastic Gaussian distribution $\mathcal{N}(y_1, y_2)$, where y_2 is restricted to ensure positive variance by defining a custom activation function:

$$y_2 = \left\{ \begin{array}{ll} y_2 & y_2 > 0, \\ \frac{1}{1-y_2} & y_2 \leq 0 \end{array} \right\}$$

Weights are initialized using the scaling suggested by (He et al. 2015), and the conditional negative log-likelihood of the target values is minimized using the Adam optimizer (Kingma and Ba 2014) with mini-batches of size 1024. The initial learning rate of 0.003 decays starting at epoch 300 at a decay rate of 0.0005 applied at the end of each subsequent epoch. A dropout rate of 0.5 is applied to the penultimate layer. Training is stopped when the validation loss doesn't improve after 15 epochs. The architecture, learning rate, and early-stopping were optimized with SHERPA.

Results

To compare the three types of models after hyper-parameter tuning, we trained each on data from 2015-2016, and tested on events from 2018, enabling us to explore the relative benefits of deep learning and the use of image spectra features. Table 2 shows that the deep neural network trained on image spectra achieves a significantly lower root mean squared error (RMSE) of 0.33 meters compared to the other methods that rely only on the high-level features: 0.64 m for the linear model and 0.43 m for the deep neural network trained on CWAVE alone. Furthermore, this performance improvement is consistent across small, medium, and large waves.

Table 2: Root Mean Squared Error on Test Set

Wave Height	CWAVE Linear	CWAVE NN	Deep NN	Percentage of Total Data
All Waves	0.642	0.433	**0.327**	100%
<1m	0.827	0.443	**0.392**	1.4%
1m - 3m	0.515	0.377	**0.255**	66.4%
3m - 8m	0.781	0.514	**0.426**	31.8%
>8m	3.226	1.512	**1.216**	0.4%

A feature importance study (Table 2) shows the dependence on each set of features. Two additional models were trained with an identical DNN architecture and hyper-parameters, but with specific high-level features removed: the 20 CWAVE parameters removed or the latitude and longitude features. In both models, time of day, incidence angle, incidence angle mode, satellite type, time and distance difference between altimeter and sentinel data, normalized radar cross section σ^0 and normalized variance of radar cross section are still included. The results show that the high-level CWAVE features are still used by the model, but only slightly — despite containing no additional information, these features add implicit bias to the model. The location features *do* contain additional information — they essentially allow the model to learn a prior over the wave heights at different locations — but these too only have a small effect on performance.

Table 3: Feature Importance Study

Wave Height	No CWAVE	No Latitude Longitude	All Included	Percentage of Total Data
All Waves	0.334	0.329	0.327	100%
<1m	0.439	0.421	0.392	1.4%
1m - 3m	0.263	0.255	0.255	66.4%
3m - 8m	0.432	0.429	0.426	31.8%
>8m	1.145	1.187	1.216	0.4%

Finally, we explore the impact of increasing the size of the training set on the discrepancy between including and not including CWAVE parameters in our final model. In this experiment, we fix the hyper-parameters, train on data from 2015-2017 (568,626 examples), then test on 2018. The models are trained for a fixed 30 epochs where the learning rate is annealed by a factor of 0.4 every 10 epochs. Initial learning rate and dropout are identical to that of our optimal deep neural network architecture. Figure 4 shows the mean performance of six randomly-initialized networks trained with different fractions of the data set. An ensemble (arithmetic mean) of the 6 models using all features and the complete training set gives a test RMSE of 0.307 — a 50% reduction in RMSE from the previous state-of-the-art of 0.6 m (Stopa and Mouche 2017). This also approaches the RMSE of satellite altimetry compared to buoy observations (Ribal and Young 2019).

Discussion

Our results demonstrate that a deep convolutional neural network can extract useful representations from SAR image spectra that is not captured by engineered CWAVE features. In a direct comparison between two hyper-parameter optimized deep neural networks, the network with the image spectra information obtained a 29% reduction in RMSE (0.33 m vs. 0.43 m). This is in keeping with the success of deep learning in other fields, where the expertly *engineered* features are discarded in favor of *learned* features.

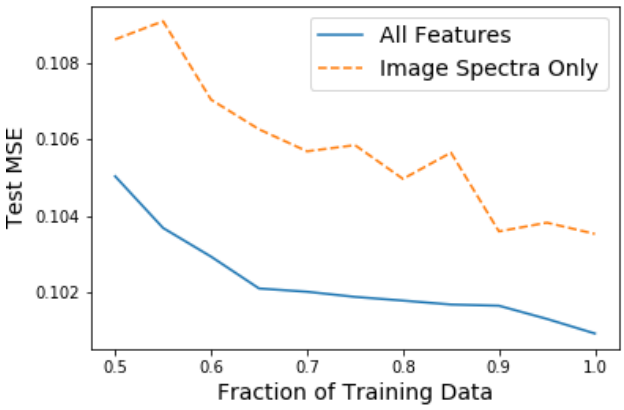

Figure 4: Performance of with both SAR image spectra and CWAVE parameters (blue line) and only SAR image spectra (orange curve), while varying the training set size. Each point is an average of MSE objective over 6 trials with different random weight initializations.

Our results show that there is still some advantage to including the CWAVE features in the model, even with a training set of over 500,000 examples. However, we also show that this advantage diminishes as the number of training examples increases (Figure 4). The CWAVE features, being derived from the image data, provide no additional information, but they help bias the model towards 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 is useful for predicting significant wave heights because there are regional characteristics of the wave climate (Stopa et al. 2013). 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.

Conclusion

Our results demonstrate that deep learning provides a 50% decrease in RMSE compared to the previous state-of-the-art in predicting significant wave height from SAR. 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. Thus, we should we should expect the performance of our model to increase as more co-location events are collected over the next couple years.

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). The altimetry data was sourced from the Integrated Marine Observing System (IMOS). All Sentinel-1 L2 data used in this study can be obtained from the Copernicus Data Hub (cophub.copernicus.eu). 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

Baldi, P.; Sadowski, P.; and Whiteson, D. 2014. Searching for exotic particles in high-energy physics with deep learning. *Nature Communications* 5.

Collard, F.; Ardhuin, F.; and Chapron, B. 2009. Monitoring and analysis of ocean swell fields from space: New methods for routine observations. *Journal of Geophysical Research* 114(C7).

Duvenaud, D. K.; Maclaurin, D.; Iparraguirre, J.; Bombarell, R.; Hirzel, T.; Aspuru-Guzik, A.; and Adams, R. P. 2015. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems*, 2215–2223.

Engen, G., and Johnsen, H. 1995. SAR-ocean wave inversion using image cross spectra. *IEEE Transactions on Geoscience and Remote Sensing* 33(4):1047–1056.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, 1026–1034.

Hertel, L.; Collado, J.; Sadowski, P.; and Baldi, P. 2018. Sherpa: Hyperparameter optimization for machine learning models.

Johnsen, H., and Collard, F. 2009. Sentinel-1 ocean swell wave spectra (osw) algorithm definition. Technical report, Northern Research Institute.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. 2014. Imagenet classification with deep convolutional neural. In *Neural Information Processing Systems*, 1–9.

Li, X.-M.; Lehner, S.; and Bruns, T. 2011. Ocean wave integral parameter measurements using envisat asar wave mode data. *IEEE Transactions on Geoscience and Remote Sensing* 49(1):155–174.

Lusci, A.; Pollastri, G.; and Baldi, P. 2013. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. *Journal of Chemical Information and Modeling* 53(7):1563–1575.

Pleskachevsky, A.; Jacobsen, S.; Tings, B.; and Schwarz, E. 2019. Estimation of sea state from sentinel-1 synthetic aperture radar imagery for maritime situation awareness. *International Journal of Remote Sensing* 40(11):4104–4142.

Ribal, A., and Young, I. R. 2019. 33 years of globally calibrated wave height and wind speed data based on altimeter observations. *Scientific Data* 6(1).

Sadowski, P., and Baldi, P. 2018. Deep learning in the natural sciences: applications to physics. In *Braverman Readings in Machine Learning. Key Ideas from Inception to Current State*. Springer. 269–297.

Schulz-Stellenfleth, J.; König, T.; and Lehner, S. 2007. An empirical approach for the retrieval of integral ocean wave parameters from synthetic aperture radar data. *Journal of Geophysical Research* 112(C3).

Stopa, J. E., and Mouche, A. 2017. Significant wave heights from sentinel-1 sar: Validation and applications. *Journal of Geophysical Research: Oceans* 122(3):1827–1848.

Stopa, J. E.; Cheung, K. F.; Tolman, H. L.; and Chawla, A. 2013. Patterns and cycles in the climate forecast system reanalysis wind and wave data. *Ocean Modelling* 70:207–220.

Stopa, J. E.; Ardhuin, F.; Husson, R.; Jiang, H.; Chapron, B.; and Collard, F. 2016. Swell dissipation from 10 years of envisat advanced synthetic aperture radar in wave mode. *Geophysical Research Letters* 43(7):3423–3430.

Wang, C.; Mouche, A.; Tandeo, P.; Stopa, E. J.; Longepe, N.; Erhard, G.; Foster, R.; Vandemark, D.; and Chapron, B. 2019. A labeled Ocean SAR Imagery Dataset of Ten Geophysical Phenomena from Sentinel-1 Wave Mode. *Geoscience Data Journal*.

Young, I. R.; Zieger, S.; and Babanin, A. V. 2011. Global trends in wind speed and wave height. *Science* 332(6028):451–455.