A Prediction Model for Aggravated, Roguish Sea-States

Dylan [redacted]
AP Statistics
[redacted]

[2270 Words]

***Abstract*—This paper is an exploration into the use of computer modeling of neural networks as a way of predicting 'aggravated roguish sea-states', which are themselves localized sea-states that experience greater than 0.0003 rogue wave probabilities. Networks will attempt to predict this behavior based on a multitude of characteristic descriptors of that localized sea-state, known as characteristic sea-state parameters.**

***Keywords*—rogue waves, neural networks, sea-state, characteristic sea-state parameters**

I. Introduction

Despite an ever growing body of research surrounding the understanding of the behavior of our oceans, a number of phenomena require much more research to be properly understood by science. Specifically, the phenomenon of Rogue Waves is noticeably lacking a fleshed out explanation in our modern understanding of physics. These anomalous waves, known in early research as Freak Waves, are categorized as waves whose height exceed double of that of a sea-state's significant wave height (Hs), which itself is the average of the largest third of wave heights in a sea-state. There exists many explorations into the physics behind these frightful phenomena, with even some delving into the realm of theoretical quantum physics and a multitude of wave pools, conventional theoretical physics examinations, and countless other approaches. Although these have been attempted and have gained moderate insight, they remain incapable of truly describing the criteria responsible for the formation of the waves itself, the means by which they are capable of reaching their height and their force, and any true descriptors of the factors that may maintain ship safety from these Rogues.

That being said, in the past few decades a number of advances in technology have facilitated the ability of both the robust collection of incredibly large databases pertaining to wave behavior and the ability of individual researchers to iterate through such databases. An individual researcher also benefits massively from the incredible depth of existing statistics or learning based libraries. In the case of this paper, were the researcher to attempt to develop neural networks from scratch rather than utilizing existing libraries such as TensorFlow to do the same, the project length would likely more than double and the efficacy of the final model would be massively reduced. With the assistance of these developments, it is feasible for a single researcher with no funds, only access to a semi-competent computer and a moderate background in a programming language to conduct somewhat meaningful research in almost any field.

II. Methods

A. Data Collection and Transformation

The data utilized throughout the course of this project originates from a single database, FOWD. This database itself is a quality controlled and efficiently stored version of CDIP data and includes a number of characteristic sea-state parameters. This data, formatted around single waves, is massively helpful but could be transformed to better accomplish the goal of this project. Since the researcher is after finding and predicting localized sea-states that experience greater than expected rogue wave occurrences, the data should be converted from wave specific to localized sea-state specific. It was decided, in order to maintain a large amount of data but also not have too small a sample size for each sea-state, that each localized sea-state should contain 10,000 waves. The parameters for each of these localized sea-states would then be the average of those parameters for the waves in each sea-state. This process entailed an iteration through each targeted parameter, on each wave, resulting in a total iteration count of 23.46 billion (1.38 billion× 17). This program took several days to fully complete and is available for viewing under the supplemental materials section of this paper preceding references.

TABLE I. Characteristic Parameters and Their Computation

Parameter	***Computation***
Ursell Number	$U = \frac{H}{h}(\frac{\lambda}{h})^2$
Mean Period (Direct)	$\overline{T_{s,0}} = \sqrt{m_0/m_2}$
Mean Period (Spectral)	$\overline{T_{d,0}} = \frac{1}{N}\sum_{i=0}^{N} t_i$
Skewness	$\mu_3 = \frac{\sum_i^N (x_i - \overline{X})^3}{(N-1)\times\sigma^3}$
Kurtosis	$Kurt = \frac{\mu_4}{\sigma^4}$
Steepness	$\epsilon = \sqrt{2m_0 k_p}$
Bandwidth Peakedness	$\sigma_Q = \frac{m_0^2}{2\sqrt{\pi}}[\int_0^\infty fS(f)^2 df]^{-1}$
Bandwidth Narrowness	$\sigma_N = \sqrt{\frac{m_0 m_2}{m_1^2} - 1}$

Parameter	***Computation***
BFI	$BFI = \frac{\epsilon\upsilon}{\sigma}\sqrt{max\{\beta/\alpha, 0\}}$
Crest Trough Correlation	$\lambda = \int_0^\infty S(\varpi) sin(\varpi \frac{\bar{T}}{2}) d\varpi)$

a. All equations are derived from "On the Computation of the Benjamin Feir Index" [2] and FOWD [3].

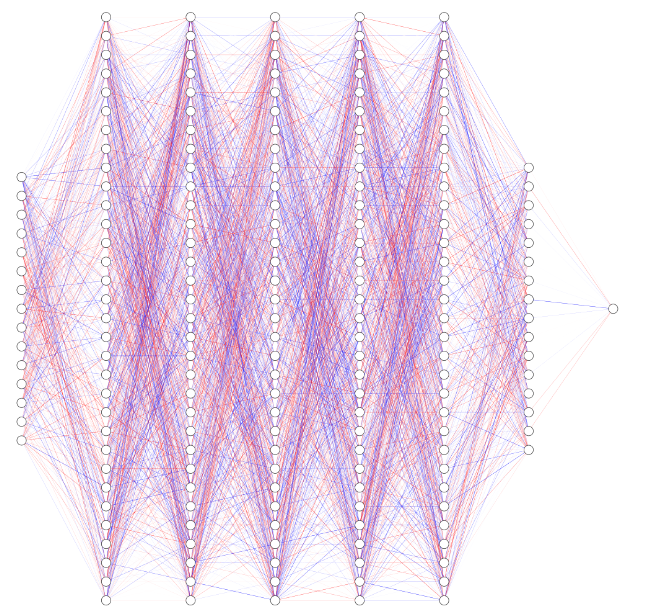
B. Network Development

Now, the data is almost workable for this paper's intended goal. The data originally contained 1.38 billion waves and was then transformed into sea-states of lengths of 10,000 waves, leaving a final dataframe with the shape of (138430, 18). This dataframe was then written to a .csv as a Pandas Dataframe so that it may be readily accessed. This leaves the researcher with an almost comparatively measly 2,214,880 x_training data points and 138,430 y_label data points. From here though, the fun is just getting started.

To properly make use of this data, a number of approaches present themselves. However, for the purposes of this paper, the approach of creating a model based on TensorFlow neural networks is the most appealing. It offers a robust model that may be quickly developed and changed, as well as saved and applied to new data. These deep learning networks however learn deeply and are by proxy difficult to maintain and guide in the right direction. That being said, the easiest of approaches when it comes to a neural network is one that has to guess on the easiest outcome. Since the goal of the model is to identify roguish sea-states, a simple approach entails the idea that a sea-state with a greater than .0003 rogue probability may be converted into a 'roguish sea-state' in the output (represented by 1) and those that are less than would be categorized as non aggravated roguish and represented by a 0.

With this binary representation of roguish states, construction of the architecture for the neural network can begin. The loss algorithm most appropriate for this approach is Keras' Binary_Crossentropy, metric most appropriate being BinaryAccuracy(a measure of if the binary output was correct), and an architecture of 5 layers containing 32 neural network nodes each, followed by a layer of 16 nodes, all utilizing relu activation besides the output layer of a single node activated in sigmoid to normalize the output between 0 and 1. This architecture is visualized below:

FIG. 1. VISUALIZATION OF NEURAL NETWORK



This network yielded a peak epoch binary accuracy of .8615, guessing roguish sea-states at an incredible rate. This sounds incredible until it is realized that roughly 86% of sea-states were non-roguish and it was understood that this deep learning algorithm guessed only zeroes to accomplish this accuracy. Of course, this isn't great and weights were applied to output labels to prevent the algorithm from attempting strategies like this. Finally, the network was able to accomplish a binary accuracy of 0.7625. It is of note that a variety of options may bring up this accuracy and improve the implementation of the algorithm, although most of these options involve testing a change, failing, testing another change, until the optimal selection is made. As a factor of the time available to tweak this network in this way, it is expected that it is capable of much greater than .0.7625 accuracy but it is at this time only capable of that 0.7625 figure.

C. Hypothesis and Validity Test

A 76.25% accuracy on the binary classification of rogue sea-states is hardly a prediction model but could be of use in proving the importance of sea-state parameters in rogue wave occurrence. As such, a hypothesis test to prove this notion is in order. With the hypothesis that this neural network is capable of a greater accuracy than blind guessing the binary output, a one-sample proportion Z test may be conducted:

$$H_o: p = 0.5$$

$$H_a: p > 0.5$$

Conditions:

I) Randomness: The data sampled were randomly selected from the population of interest, which were the localized sea-states that contained CDIP measurement buoys. This is however not a simple random sample, so proceed with caution.

II) 13843 < 10% of all localized sea-states 10,000 wave sea-states over all time in the ocean.

III) np = 10555, > 10

nq = 3288, > 10

Let p represent the probability of correctly guessing the binary output by the model.

Let p_0 represent the probability of correctly guessing the binary output by chance (.5).

Let n represent the number of sampled outputs which the model had not seen in training, which were assigned to 10% of all data randomly.

$$\alpha = 0.05$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(q_0)}{n}}}$$

$$= \frac{.7625 - .5}{\sqrt{\frac{.5(.5)}{13843}}}$$

$$= 61.77$$

$p = 0.000$

Given that $p < \alpha$ by a large margin, the null hypothesis may be rejected in favor of the alternate hypothesis.

D. Distribution Analysis

With the efficacy and validity of the model somewhat established, the core part of this project is complete. However, the researcher feels that a number of questions remain unanswered. Among these is a proper understanding of the distribution of rogue wave probabilities amongst the localized sea-states and what this distribution may infer regarding the behavior of rogues. A greater understanding of this may yield a greater understanding of the variability of rogue probabilities, its cause, and improved methods of predicting its occurrence.

Two general approaches were undergone to this end, computerized distribution fitting with the help of the python library distfit, and more conventional graphing of this distribution. Distfit provides the researcher with scores relating to a number of distributions that are difficult to spot with the untrained eye, of which the researcher happens to have two. The full length summary of this library's output is available for download in this paper's supplemental materials.

TABLE II. DISTRIBUTION ANALYSIS

Distr	Distribution Confidence		
	Score	***Loc***	***Scale***
Dweibull	40624171.939259	0.000161	0.000136
Dgamma	41174611.32314	0.000156	0.000063
Foldnorm	45386735.308845	0.0	0.000247
Exponpow	45465962.399524	0.0	0.000351
Halfnorm	45491980.224168	0.0	0.00023

b. These are the top five most likely of the total 80 tested distributions, in order of most to least likely.

c. All distributions are named as in scipy and their respective probability density functions are viewable on scipy's API reference.

FIG. 2. DISTRIBUTION VISUALIZED

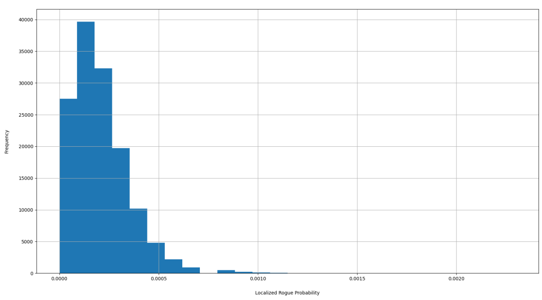


Fig. 2. represents the distribution of localized rogue wave probabilities in the separated 10,000 wave sea-states. It is of note that the *typical* 10,000 wave sea-state will only experience 0-3 rogue waves and as such the rogue wave probability of a given localized sea-state steps very abruptly, making any potential histogram have comparatively large bins. That being said, upon graphing the data, the distribution identified as most likely in TABLE II, a double weighted Weibull distribution, does somewhat resemble what is seen in this figure.

FIG. 3. DOUBLE WEIGHTED WEIBULL VISUALIZED [3]

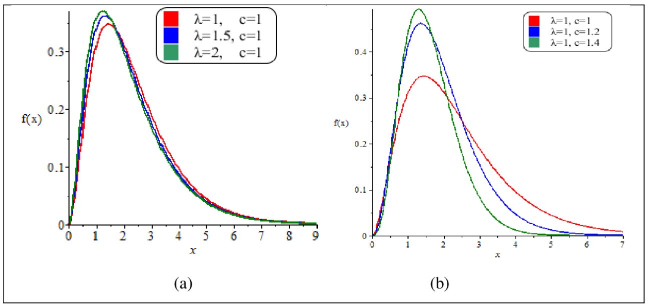
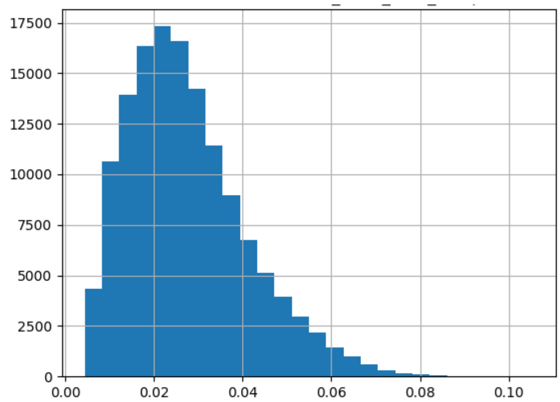


FIG. 4. WAVE STEEPNESS PARAMETER



An interesting offshoot of this development is the striking similarity with which a number of parameter's histograms resemble this extremely specific distribution. In the case of sea_state_30m_steepness, the resemblance to the Double Weighted Weibull distribution is nearly uncanny. It is the opinion of the researcher that continued exploration of the cause of this resemblance, this specific distribution of localized wave steepness, and its application as to the prediction of rogue occurrences is necessary.

III. CONCLUSIONS

A. Aggravated Sea-States

It is perhaps a shortcoming of this paper in the lack of acknowledgement of the frightening distribution of localized rogue probabilities. In sea-states of 10,000 waves, nearly a third of sea-states experience rogue probabilities more than double than that of opposing sea-states. For whatever reason, rogue occurrences can sharply increase in specific localized environments. The intensified occurrence of waves whose height exceeds that of double of the mean upper third of wave heights is a terrifying prospect. In the interest of the safety of the traversal of our oceans, the assurance of proper strength in ship construction and better understanding the never ending physics questions of our oceans, this intensification should be explored.

B. *Future Prediction Models*

Despite the brevity of this project and its explorations, the inexperience with which the researcher attempted to optimize and fit neural networks to only marginally correlated data, the lack of heavy computing power and funding, a prediction model's efficacy over random guessing has been statistically verified. There remains trends in characteristic sea-state parameters that are of use in predicting the aforementioned localized aggravated sea-states. The prediction of these sea-states represents another necessary layer of safety in the prediction of weather events on our oceans and protecting the vessels that traverse them. Future models could utilize this technique of the separation of databases into localized sea-states in order to verify the findings of this paper, better optimize their neural networks, create more effective network architectures using more optimal activations and optimizations and otherwise improve this work to establish a better model.

The essential goal of this project was to verify the validity of the researcher's past findings and apply these findings to the development of rudimentary models capable of the prediction of these intensified sea-states with which the researcher was familiar. This end has been reached. It is the hope of the researcher that those reading this paper may take this information and develop their own presumptions regarding the behavior of this phenomenon to physics and test them. The only way this problem may be conquered is with continual and labored examination.

Acknowledgments

The data for this project was supplied originally by the Coastal Data Information Project of the University of California San Diego, at which point it was converted in a paper by Dr. Dion Hafner [3] into a database that the researcher was then able to utilize to write this paper. Without the work of those researchers responsible for this data's generation explorations like this would be impossible.

Supplemental Materials

All python scripts utilized in the final production of this paper are available for download at the linked GitHub repository. All applicable output.txt files, created .csv data files, original full scale graphs and a number of graphs not included on the final paper are also available for download in that location.

References

[1] Saghir, A., & Saleem, M. (2016). Double Weighted Weibull Distribution Properties and Application. Mathematical theory and modeling, 6, 28-46.

[2] Serio, Marina & Onorato, Miguel & Osborne, A & Janssen, Peter. (2005). On the computation of the Benjamin-Feir Index. Nuovo Cimento della Società Italiana di Fisica C. 28. 893-903. 10.1393/ncc/i2005-10134-1.

[3] Häfner, D., Gemmrich, J., & Jochum, M., 2021: FOWD: A Free Ocean Wave Dataset for Data Mining and Machine Learning. Journal Of Atmospheric And Oceanic Technology, 38, 1305-1322. doi: 10.1175/JTECH-D-20-0185.1

[4] TensorFlow Developers. (2022). TensorFlow (v2.8.2). Zenodo. https://doi.org/10.5281/zenodo.6574269

[5] Casas-Prat, M., Holthuijsen, L., 2010: Short-term statistics of waves observed in deep water. Journal Of Geophysical Research, 115(C9). doi: 10.1029/2009jc005742

[6] Cattrell, A., Srokosz, M., Moat, B., & Marsh, R., 2019: Seasonal intensification and trends of rogue wave events on the US western seaboard. Scientific Reports, 9, 4461. doi: 10.1038/s41598-019-41099-z

[7] Virtanen, P.,Gommers, R., Oliphant, T., 2020: SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature, 17, 261-272. doi: 10.1038/s41592-019-0686-2

[8] Mori, N., Janssen, P., 2006: On Kurtosis and Occurrence Probability of Freak Waves. Journal Of Physical Oceanography, 36(7), 1471-1483. doi: 10.1175/jpo2922.1

[9] Häfner, D., Gemmrich, J. & Jochum, M., 2021: Real-world rogue wave probabilities. Sci Rep 11, 10084. doi: 10.1038/s41598-021-89359-1

[10] Behrens, J., Thomas, J., Terrill, E., Jensen, R., 2019: CDIP: Maintaining a robust and reliable ocean observing buoy network. IEEE, 19260974. doi: 10.1109/CWTM43797.2019.8955166