



# CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

## Rogue ocean waves and the St. Petersburg paradox

Paul Johns, Jake Fontana, and Peter Palffy-Muhoray

Phys. Rev. E **105**, 025103 — Published 10 February 2022

DOI: [10.1103/PhysRevE.105.025103](https://doi.org/10.1103/PhysRevE.105.025103)

# Rogue Ocean Waves and the St. Petersburg Paradox

Paul Johns and Jake Fontana\*  
*U.S. Naval Research Laboratory  
4555 Overlook Ave. Southwest,  
Washington, DC 20375, USA*

Peter Palfy-Muhoray†  
*Advanced Materials and Liquid Crystal Institute  
Kent State University, Kent, OH 44242, USA  
(Dated: January 3, 2022)*

Powerful rogue ocean waves have been objects of fascination for centuries. Elusive and awe-inspiring, with the potential to inflict catastrophic damage, rogue waves remain unpredictable and imperfectly understood. To gain further insight into their behavior, we analyzed 3.4 billion ocean surface waves in order to determine the statistical height distribution of the largest waves. We found that the distribution of rare events which resolves the St. Petersburg paradox also describes the relative height distribution of the largest waves. This result is expected to contribute to the modeling of ocean surface dynamics and improve the accuracy of marine weather forecasts.

Rogue waves are exceptionally large naturally occurring waves, ranging from ocean surface waves [1, 2] to waves in optics [3] Bose-Einstein condensates [4], plasmas [5] and even finance [6]. Although their emergence is infrequent, encounters with them are often catastrophic, underpinning the need to anticipate them.

Considerable progress has been made over the last 70 years in understanding and forecasting rogue waves. Yet, ambiguities still exist regarding how well the established models fit observations, particularly when the number of waves becomes large [3, 7–11].

To predict rogue waves, an effective description is needed to relate the heights of waves to the number of waves  $N$  in a set. It is standard practice to monitor the maximum normalized wave height,  $H_{max}/H_s$ , where  $H_{max}$  is the maximum wave height and  $H_s$  is the significant wave height. The significant wave height  $H_s$  is the average height of the highest  $1/3^{rd}$  of the waves in the set. Rogue waves are those whose height  $H$  is at least twice the significant wave height, that is, where  $H/H_s > 2$ . Typically, rogue waves appear in sets where  $N$  is large and increase in height with increasing  $N$ .

While many models have been developed to predict  $H_{max}/H_s$  for ocean surface waves, at least two models have emerged over time which agree well with observations. The Rayleigh model [12], developed by Longuet-Higgins, is based on the postulate that rogue waves arise as the sum of contributions from a large number of smaller waves with random phases. Making use of the Rayleigh distribution, it proposes a square root and logarithm dependence to predict the most probable value,  $H_{max}/H_s = \alpha\sqrt{\ln(N)} + \beta$  where  $\alpha$  and  $\beta$  are constants. The Weibull model [7], introduced and empirically justified for ocean waves by Forristall, predicts a generalized

combination of a power law and logarithmic dependence,  $H_{max}/H_s = \alpha(\ln(N))^\beta$ .

Rather than pursuing a detailed model here, we follow the approach of Weibull [13] and Forristall [7] and provide here a simple empirical form which gives excellent fit to the data. Based on our analysis of the largest dataset reported to date, we demonstrate that the relative heights of ocean surface waves appear to accumulate similarly to profits in the St. Petersburg game [14]. These results may also have applications in other scientific and engineering fields where rogue waves are prevalent.

In the St. Petersburg game [15] a true coin is flipped until it lands heads. If this happens on the  $n$ th flip, the player receives  $2^n$  dollars. The paradox is that although the expectation value of the payout is infinite, the typical payout is small. A lucid resolution of the paradox, provided by Feller [16], is that the length of longest run of tails, and hence the expected payout per game, depends linearly on the logarithm of the number of flips.

To form a correspondence between ocean surface wave heights and the game, one can regard ocean waves as the sum of small wavelets. If the height of waves is taken to be the sum of heights of many small identical wavelets with the same phase, then wave heights accrue similarly to lengths of runs of tails in the St. Petersburg game. The expected number of waves with dimensionless height  $h$  in a set of  $N$  waves is  $(1-p)^2 p^h N$ , where  $p$  is the probability of a wavelet with a given phase. Since there is only one wave with maximum height  $h_{max}$  in the set,  $(1-p)^2 p^{h_{max}} N = 1$ , and the expected maximum wave height  $h_{max} = -\ln[N(1-p)^2]/\ln(p)$ , giving the maximum normalized wave height [17],

$$\frac{H_{max}}{H_s} = \alpha \ln(N) + \beta \quad (1)$$

We call this hypothesis the Petersburg model. To verify the proposed relationship between the maximum normalized wave heights and the number of waves by ob-

---

\* jake.fontana@nrl.navy.mil

† mpalfy@kent.edu

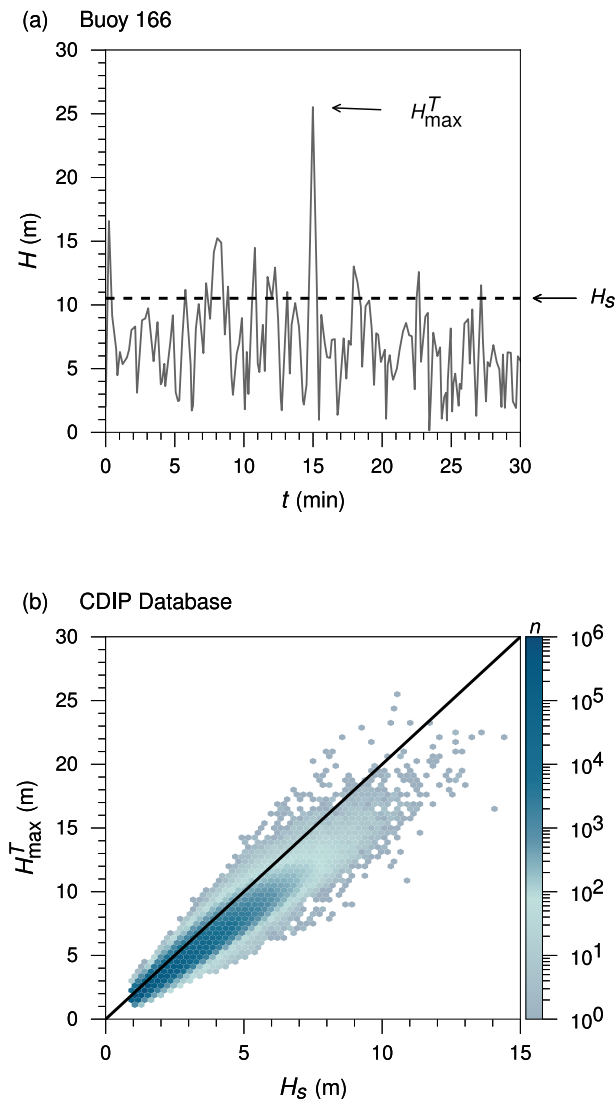


FIG. 1. Maximum and significant wave height per period. (a) Wave height evolution for a 30 minute period from buoy 166 (50.017 N, 145.170 W) on 12/02/2016. (b) Hexbin plot of the maximum wave height per period versus significant wave height for all buoy datasets. The solid line is the rogue wave threshold and  $n$  is the number of waves per hexbin.

servation, data spanning several orders of magnitude is required due to the logarithmic dependence, as demonstrated in other applications of the Petersburg description [17–19].

The Coastal Data Information Program (CDIP) at U.C. San Diego [20] has recorded 27 years of ocean surface wave data from buoys located predominantly across the Pacific ocean. We analyzed data from 156 buoys from the database, spanning the period August 27, 1993, 16:32 (UTC) – June 30, 2020, 14:27 (UTC), resulting in 3,441,188,683 waves after filtering, the equivalent of a single buoy collecting data for  $6\frac{1}{2}$  centuries (649 years).

A Python code was used on a desktop computer (128 GB RAM, 6-core 3.4 GHz processors, 32 TB hard drive)

to process all the data presented here [21]. Sensors on the buoys recorded their movement relative to the still-water line (mean value), produced by the waves as a function of time. The code used a zero-upcrossing method to identify and measure the height of each wave. The difference between the maximum and minimum vertical displacements between consecutive zero up-crossings determined the height of each wave. The data collected by each buoy was filtered to remove spurious events (data gaps, electrical noise, anomalous buoy motion, or wave data flagged by CDIP) using standard quality control methods [22]: the wave crest height was restricted to an upper bound of  $< 1.5 \times H_s$  and a lower bound of  $H_s > 1$  m, and the wave kurtosis was restricted to an upper bound of  $< 6$ . Unless otherwise specified, the datasets were analyzed in standard periods of  $T = 30$  minutes, where  $H_s$  was determined every  $T$  [20].

An illustrative example of the maximum and significant wave heights per period,  $H_{max}^T$  and  $H_s$ , is shown in Fig. 1(a), highlighting the largest wave recorded in the CDIP database ( $H = 25.53$  m). The maximum versus significant wave heights per period were determined for all the buoy datasets; these are plotted in Fig. 1(b). The data were organized into hexagonal bins (hexbins), where  $n$  is the number of waves per bin, displayed as a blue color hue. The solid black line corresponds to the rogue wave threshold,  $H_{max}^T/H_s = 2$ , resulting in 615,978 rogue waves.

The maximum normalized wave height,  $H_{max}^N/H_s$ , is the largest  $H(N)/H_s$  encountered in  $N$  waves and is shown for a single buoy (092) in Fig. 2(a) and for all the CDIP buoy datasets in Fig. 2(b). To determine  $H_{max}^N/H_s$ , it is useful to define the normalized wave heights  $\hat{H} = H(N)/H_s$ . These were calculated initially, updating  $H_s$  every period. Normalizing  $H(N)$  by  $H_s$  enables comparison over many time periods, locations, and varying sea states. The normalized waves were then sequentially analyzed per buoy. If  $\hat{H}(N)$  exceeded the heights of all preceding waves, then  $\hat{H}(N)$  was used to replace  $H_{max}^N/H_s$ . These results are plotted in Fig. 2(a) and Fig. 2(b), showing the maximum normalized wave height a stationary observer (buoy) encounters in  $N$  waves.

The Petersburg, Rayleigh, and Weibull models were fit to the data in Fig. 2(a) and 2(b). The adjusted coefficient of determination,  $\bar{R}^2$ , was used as the figure of merit to determine the best fit. As an example for a single buoy (092), with data consisting of over 110 million waves, we find  $\bar{R}^2$  is 0.946, 0.909 and 0.866 for the Petersburg, Rayleigh, and Weibull models, respectively [Fig. 2(a)]. When data including all 3.4 billion waves from the CDIP database were examined [Fig. 2(b)], we found that  $\bar{R}^2$  is 0.892, 0.889, and 0.843 for the three models.

The Petersburg model fits the total wave data best [Fig. 2(b)], improving the capability to anticipate the maximum normalized wave height based on the number of waves buoys encounter.  $H_{max}^N/H_s$  is expected to be

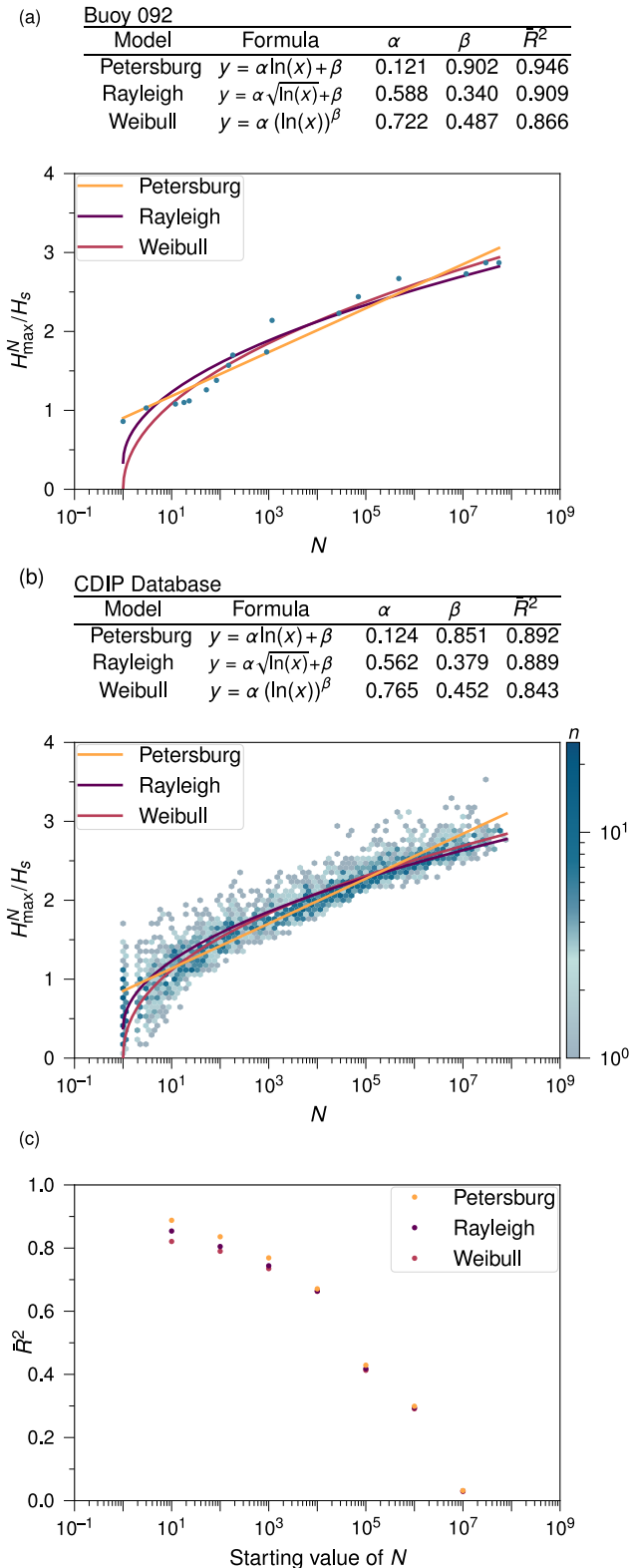


FIG. 2. Maximum normalized wave height versus wave number,  $N$ . (a) Buoy 092 (33.61791 N, -118.31701 E), (b) Hexbin plot for all CDIP buoy datasets, where  $n$  is the number of waves per hexbin and (c) the adjusted coefficient of determination for subranges of  $N$ .

TABLE I. (a) Summary of the individual buoy fits in Fig. 2(b). (b) The effect of varying the period: Petersburg, Rayleigh, and Weibull model fits to and for 60 and 15 minute periods over all of the CDIP database.

(a)

Model	Number of buoys best fit	Total number of waves	Average number of waves/buoy
Petersburg	74	1,670,245,198	22,570,881
Rayleigh	66	1,516,934,178	22,983,851
Weibull	16	304,220,375	19,013,773

(b)  $T = 60$  min

Model	$\alpha$	$\beta$	$\bar{R}^2$
Petersburg	0.128	0.846	0.884
Rayleigh	0.584	0.351	0.874
Weibull	0.755	0.466	0.837

$T = 15$  min

Model	$\alpha$	$\beta$	$\bar{R}^2$
Petersburg	0.118	0.872	0.888
Rayleigh	0.547	0.394	0.897
Weibull	0.786	0.434	0.850

less than 1 for the first wave, near unity after 10 waves, 2 (rogue) by 10,000 waves and 3 by 30 million waves. Temporally,  $H_{max}^N/H_s$  will be 1, 2, and 3 after approximately a minute, a day (16 hours), and half a decade, respectively, assuming  $3.4 \times 10^9$  waves/649 years.

The Rayleigh and Weibull models fit the full set of wave data nearly as well as the Petersburg model. It is worth noting that even with 3.4 billion measurements there is only a  $\Delta \bar{R}^2$  of 5.5% between the three models, due to the intrinsic scatter of the data. Albeit, one important distinction is as  $N$  becomes large  $> 10^5$  ( $> 1$  week), the Rayleigh and Weibull models diverge from the observed data, capturing only the lower limit of the potentially largest and most destructive rogue waves [Fig. 2(b)].

If the data in Fig. 2(b) is fit for different subranges [Fig. 2(c)], where the starting value of  $N$  varies from  $10^1$  to  $10^7$ , then the Petersburg  $\bar{R}^2$  remains the best fit for each range. The  $\bar{R}^2$  values indicate poorer fit overall for all three models as the subranges decrease in size.

If the buoy datasets are fit individually [Table 1(a)], as opposed to fitting the entire database [Fig. 2(b)], then the Petersburg model also fits better for a larger number of buoys.

The significant wave height serves as a normalization constant, yet  $H_s$  depends on the period  $T$  as does the relationship between  $H_{max}^N/H_s$  and  $N$ . The most commonly used period is 30 minutes [20] and this was used in our analysis above. To probe the dependence of the results on the period  $T$ , the entire CDIP database was analyzed;  $T$  was doubled to 60 minutes and halved to 15 minutes with results in Table 1(b). For the  $T = 60$  minute case, all three models gave a poorer fit overall than in the  $T = 30$  minute case, but the relative  $\bar{R}^2$

rankings did not change. In contrast, when  $T$  was halved to 15 minutes, the Rayleigh model fit the data slightly better ( $\Delta\bar{R}^2 = 0.9\%$ ) than the Petersburg model.

It appears that as the size of the period  $T$  and hence the number of waves per period increases, so does the ratio of adjusted coefficient of determination  $\bar{R}_{Paradox}^2/\bar{R}_{Rayleigh}^2$  from 0.98 at  $T = 15$ , to 1.00 at  $T = 30$  minutes and 1.01 at  $T = 60$  minutes. Distributions which depend on sample size are discussed by Feller [23] and Hayer and Andersen [24].

In addition to providing best fit to the data, the advantage of the Petersburg model is its simplicity. The empirical logarithmic dependence of Eq. (1) is supported by 9 decades (3.4 billion waves) of oceanographic data. This result suggests that the relative height of ocean surface waves may accumulate similarly to profits in the St. Petersburg game. Beyond ocean waves, it is interesting

to ask whether the Petersburg model will have utility in other research fields where rogue waves are encountered. An open question is if the Petersburg model will continue to provide the best description for ocean surface waves as  $N$  continues to increase, assuming other factors remain constant. Regrettably, indications of climate changes make the latter condition unlikely. For now, according to the analysis and statistical indicators used here, the Petersburg model appears to be the best predictor of the relative heights of ocean surface waves.

This work was supported by the Office of Naval Research (N0001420WX00146, N0001421WX00025, and N00014-18-1-2624). We thank the Ocean Engineering Research Group (OERG) and Scripps Institution of Oceanography for the maintenance and access to the CDIP database. We also are grateful to Mark D. Orzech and David W. Wang for useful discussions.

- 
- [1] V. E. Zakharov, Stability of periodic waves of finite amplitude on the surface of a deep fluid, *Journal of Applied Mechanics and Technical Physics* **9**, 190 (1968).
- [2] G. Dematteis, T. Grafke, M. Onorato, and E. Vandendijnden, Experimental evidence of hydrodynamic instabilities: The universal route to rogue waves, *Phys. Rev. X* **9**, 041057 (2019).
- [3] D. R. Solli, C. Ropers, P. Koonath, and B. Jalali, Optical rogue waves, *Nature (London)* **450**, 1054 (2007).
- [4] Y. V. Bludov, V. V. Konotop, and N. Akhmediev, Matter rogue waves, *Phys. Rev. A* **80**, 033610 (2009).
- [5] H. Bailung, S. K. Sharma, and Y. Nakamura, Observation of peregrine solitons in a multicomponent plasma with negative ions, *Phys. Rev. Lett.* **107**, 255005 (2011).
- [6] Z.-Y. Yan, Financial rogue waves, *Communications in Theoretical Physics* **54**, 947 (2010).
- [7] G. Z. Forristall, On the statistical distribution of wave heights in a storm, *Journal of Geophysical Research: Oceans* **83**, 2353 (1978).
- [8] S. Birkholz, C. Bre, I. Veseli, A. Demircan, and G. Steinmeyer, Ocean rogue waves and their phase space dynamics in the limit of a linear interference model, *Scientific Reports* **6**, 35207 (2016).
- [9] C. Kharif and E. Pelinovsky, Physical mechanisms of the rogue wave phenomenon, *European Journal of Mechanics-B/Fluids* **22**, 603 (2003).
- [10] M. Christou and K. Ewans, Field measurements of rogue water waves, *Journal of Physical Oceanography* **44**, 2317 (2014).
- [11] J. M. Dudley, V. Sarano, and F. Dias, On hokusai's great wave off kanagawa: localization, linearity and a rogue wave in sub-antarctic waters, *Notes and Records of the Royal Society* **67**, 159 (2013).
- [12] M. S. Longuet-Higgins, On the statistical distribution of the heights of sea waves, *Journal of Marine Research* **11**, 245 (1952).
- [13] W. Weibull, A statistical distribution function of wide applicability, *Journal of Applied Mechanics-Transactions of the Asme* **18**, 293 (1951).
- [14] I. Todhunter, *A history of the mathematical theory of probability* (Chelsea, New York, 1949).
- [15] D. Bernoulli, Exposition of a new theory on the measurement of risk, *Econometrica* **22**, 23 (1954).
- [16] W. Feller, Note on the Law of Large Numbers and "Fair" Games, *The Annals of Mathematical Statistics* **16**, 301 (1945).
- [17] J. Fontana and P. Palfy-Muhoray, St. Petersburg paradox and failure probability, *Phys. Rev. Lett.* **124**, 245501 (2020).
- [18] H. Scher, M. F. Shlesinger, and J. T. Bendler, Timescale invariance in transport and relaxation, *Physics Today* **44**, 10.1063/1.881289 (1991).
- [19] P. Palfy-Muhoray, R. Barrie, B. Bergersen, I. Carvalho, and M. Freeman, *J. Stat. Phys.* **35** (1984).
- [20] C. D. I. P. (CDIP), <http://cdip.ucsd.edu> (2020).
- [21] See supplemental material at `**[url will be inserted by publisher]**` for the python code used to process the cdip database.
- [22] M. D. Orzech and D. W. Wang, *Rogue Wave Events in Time Series of Surface Elevation*, Tech. Rep. (U.S. Naval Research Laboratory, Stennis Space Center, MS, 2019).
- [23] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed., Vol. 2 (Wiley, New York, 1957).
- [24] S. Hayer and O. J. Andersen, Freak waves: rare realizations of a typical population or typical realizations of a rare population?, in *The Tenth International Offshore and Polar Engineering Conference* (OnePetro, 2000).