

Wave Height Distribution In Unsaturated Surf Zones

M. Jishad¹, E. M. Yadhunath², and Jaya Kumar Seelam^{3*}

¹Scientist, Oceanic Sciences Division/AOSG/EPISA, Space Applications Centre, Ahmedabad, India;
email: jishad1hm@gmail.com

²Research Scholar, Bharathidasan University, Tiruchirappalli (Presently at Naval Physical and Oceanographic Laboratory, Kochi, India); e-mail: yadhunath90@gmail.com

³Senior Principal Scientist, CSIR - National Institute of Oceanography, Dona Paula, Goa, India;
Professor, School of Oceanography, AcSIR, NIO Dona Paula, Goa, India; email: jay@nio.org
(*Corresponding Author)

Abstract

This paper addresses the performance of different wave height distributions in the surf zone on natural beaches on the central west coast of India. The surfzone water level data for different surf zones were analyzed using Fast Fourier Transformation techniques to obtain the wave characteristics viz., wave height, and wave period. Different probability density functions (pdfs) were fit for the surf zone wave heights to identify the best-fit pdf. The quality of the surf zone wave modeling depends on the skill of the identified best-fit pdf in explaining the measured wave height distribution. This study further reiterates that the Rayleigh and Exponential pdfs are a poor fit for the surf zone wave heights. However, Kernel, GEV, and Logistic pdfs were observed to very well fit the surf zone waves. A constant value of power $p=2$ in the Rayleigh distribution is not suitable for the central west coast of India. The power p varies for different surf zones, and it also varies spatially and temporally. In this study, a simplified approximation for p is estimated using the maximum and the minimum values of significant wave height, root mean square wave height and the total number of waves. This improved Rayleigh pdf was applied for the surf zone waves measured at different locations along the central west coast of India and found to be fit.

Keywords: pdfs, surf zone, wave, FFT, Rayleigh distribution

1.0 Introduction

The effective management of coastal activities highly depends on accurate modeling under changing offshore conditions. The modeling of the depth-limited wave height distribution in the surf zone is one of the challenges in coastal studies. Parametric wave transformation models were the most widely used cross-shore morphological models due to their computational efficiency (Baldock et al., 1998). The wave height distribution function and the fraction of broken waves assumptions must be made in the parametric models to predict random wave behavior (Baldock et al., 1998; Battjes and Janssen, 1978; Ruessink et al., 2003; Thornton and Guza, 1983). Statistical distributions of the sea surface wave heights have drawn considerable attention in the past (Gluhovski, 1968; Goda et al., 1975; Kuo, 1975; Siefert, 1975; Soares et al., 2003; Vanem, 2016). Dattatri, (1973), and Hameed and Baba(1985) have well studied the distributions of wave heights for the Indian west coast. The surf zone is termed as unsaturated when there is an increase in the offshore incident wave height, which in turn results in the increase in wave height at the inner surf zone. Most of the wave transformation models assume that the inner-surf zone waves are always depth limited, and therefore the surf zone is saturated. The surf zones from the central west coast of India are mostly unsaturated (Jishad, 2018). The nearshore slope and offshore wave characteristics play a significant role in the surf zone waves, which influence the transformation and distribution of the waves.

Modeling of shallow water waves using theoretical distribution is challenging due to the non-linear wave-wave interactions, wave transformations, shoaling, and breaking while propagating from offshore to shore. Ochi (1978) had proposed the use of the bivariate-log-normal distribution, an exponential transformation of the bivariate-Normal distribution. The logarithm of the low and moderate significant wave heights (H_s) is mostly normally distributed. However, this may not apply to the large values of H_s . Longuet-Higgins (1975) has discussed the theoretical result of the joint distribution of individual wave heights and periods. Later, Tayfun (1980) explained separate pdfs for wave crests and troughs. Arhan and Plaisted (1981) described the differences between the crests and the troughs for a specific parameter. Recently, researchers tried a combination of two or more pdfs for better representation of the oceanic waves. Haver (1985) used a combination of lognormal distribution (for small H_s), Weibull (for large H_s), and lognormal distributions for zero-crossing wave period (T_z). Mathisen and Bitner-Gregersen (1990) proposed a better joint distribution by using Weibull distribution for H_s and lognormal distributions for T_z . Bitner-Gregersen et al. (1998) used a Gaussian transformation. Ferreira and Guedes Soares (2000; 2002) used the Kernel density function for their studies. Ochi (1998) derived a pdf for waves using the non-linearity concept. Forristall (2000), Wist et al. (2002), and Al-Humoud et al. (2002) studied the importance of the

second-order nonlinearity on the distribution functions. The influence of the higher order nonlinearity was studied by Mori and Yasuda (2002) and they concluded that the third order nonlinearities have a dominant effect on the wave height distribution. The pdf of peaks, troughs, and heights for wind waves has been proposed by Cherneva et al., (2005).

Numerous studies (e.g., Amrutha and Kumar, 2015; Chakrabarti and Snider, 1974; Dattatri et al., 1979; Longuet-Higgins, 1975, etc.) have recommended Rayleigh distribution for wave height distribution studies at outside as well as within the surfzone. But, other studies contradict that the Rayleigh pdf does not fit well with observations for the waves higher than H_s (e.g., Baba, 1983; Dattatri, 1973; Forristall, 1978; Goda, 1974; Haring et al., 1977; Power et al., 2011; Soares, 2003; Thompson, 1974). Rayleigh couldn't describe the non-linear interactions between spectral wave components responsible for the vertical asymmetry in the wave profile (Soares et al., 2004).

This paper addresses the performance of different wave height distributions in the surf zone on natural beaches for the central west coast of India. Most of the wave height pdfs were developed using one-directional wave flume experiments that used two-dimensional (2D) real field measurements. Many pdfs viz., Gamma, Beta, Exponential, Birnbaum-Saunders, Kernel, Inverse Gaussian, Log-logistic, Logistic, Nakagami, Log-normal, Normal, Rician, tLocation Scale, Weibull, and Rayleigh, were proposed in the recent past. It is challenging to develop a single pdf that is capable of simulating all the conditions along the entire coast. Therefore, this is an attempt to identify the most suitable pdf for the surfzone of the central west coast of India.

2.0 Study Region

The surf zone measurements were carried out over six surf zones from the central west coast of India (Figure 1), viz; Miramar beach ($15^{\circ}28'49.26''$ N; $73^{\circ}48'20.08''$ E), Candolim Beach ($15^{\circ}30'31.78''$ N; $73^{\circ}45'53.56''$ E), Keri Beach ($15^{\circ}42'56.44''$ N; $73^{\circ}41'25.09''$ E), Bhatye Beach ($16^{\circ}58'19''$ N; $73^{\circ}17'37''$ E), Ganapatipule Beach ($17^{\circ}08'54''$ N; $73^{\circ}15'50''$ E) and Sagartirth Beach ($15^{\circ}50'54''$ N; $73^{\circ}37'42''$ E). All the beaches experience semidiurnal tides having a tidal range of 2 to 4 m.

Miramar Beach, which lies at the confluence of River Mandovi encompassing the Arabian Sea, is a crescent-shaped beach within a 3km region between the Cabo headland in the south and Aguada headland in the north. The beach towards the south of the river confluence is a low energy beach due to the wide surf zone that waves break at far from shore and dissipate much of their energy before reaching the coast. The waves at the Miramar beach plunges during low tide and spills at high tide (Seelam et al., 2015). Keri beach is 1.5km long beach having River Terekol at the north (Rajasekaran

et al., 2014). Keri beach experienced erosion during the past decade, and hence a protective wall was constructed all along the beach. The waves are of plunging type with occasional spilling breakers (Seelam et al., 2015). Candolim is a long stretch of beach about 7.5km between the headlands Aguada and Baga. The waves at Candolim are of plunging and spilling type (Seelam et al., 2015). The Bhatye beach is a long stretch having a length of about 1.5km and is a flat and straight beach, and the Ganpatipule is one of the beaches on the Konkan coast along the west coast of India and is 20km from Ratnagiri. Sagartirth is located 33km towards the south from the district headquarters Oros. Sagartirth is near Redi headland and Redi port. The surf zone wave characteristics during the rising tide for the above-considered beaches in Goa state were briefly dealt with by Seelam et al. (2015).

3.0 Data And Methodology

The measured surfzone waves considered in this study are from different beaches, and at different times (2012 to 2015) on the central west coast of India. Aquistar[®] pressure transducers (PTs) and SeaBird[®] Wave and Tide Gauges (WTGs) were employed to measure the wave data. The PTs and WTGs, have a respective sampling frequency of 8Hz and 4Hz. Due to the intense wave action, tidal level, and the storage memory limit of the instrument, the instrument deployment was for a shorter duration. Hence, we get different lengths of data for each set of observations. Firstly, the wave record is filtered using a Butterworth filter with cut-off frequencies being 0.3 and 2Hz to smoothen the raw data. Later, the tidal fluctuations were removed using the FFT filter. The PT and WTG records were divided into sections, each respectively having 1024 and 512 data points (2.13 minutes). Thus equal wave record lengths were obtained for further analysis. Every data record underwent a thorough check to identify whether the PT is in the surf or swash zone, and data sets with dry values are removed. To get less noisy sea-swell wave information, the raw data were filtered using a low pass filter. The zero-crossing method for every 2.13 minutes of wave data provided the various wave parameters out of which the significant wave height (H_s). The details of the surfzone, number of 2.13min records (n), no of instruments used (nPT), sampling frequency (F_s in Hz), average number of waves analyzed (Nw), beach slope ($\tan\beta$), significant wave height (H_s in m), mean wave period (T_{mean} in s), and root mean square wave height (H_{rms} in m) are presented in Table 1.

There are many pdfs derived and standardized for the ideal fluid condition. The mathematical expressions for each of the distributions were well explained in Forbes et al. (2011). To find a suitable pdf that accurately represents the shallow water waves, Birnbaum-Saunders, Beta, Gamma,

Exponential, Kernel, Inverse Gaussian, Log-logistic, Logistic, Log normal, Rician, Normal, Nakagami, Rayleigh, Weibull and tLocation Scale distribution functions are used. Appendix 1 showcases the mathematical representation and parameters. Each of these sixteen distributions has its advantages and disadvantages.

Out of all the above-considered pdfs, the Rayleigh distribution function is a special case of the Weibull distribution and the earliest and most widely used models (Longuet-Higgins, 1952). It is different from other pdfs because it has only a scale parameter and no shape parameter. The H_{rms} has been calculated with wave data using the assumption that there is a large number of linear waves. However, for wind-driven ocean conditions, the narrow band concept is no longer valid due to nonlinear waves. The Rayleigh distribution performs well for deep water wave heights. However, when a wave propagates from deep water to shallow water, waves tend to steepen on shoaling, which increases the wave height and becomes more nonlinear and eventually breaks.

It is essential to choose a suitable distribution function that explains the distribution of the wave height at a particular location. Often by quantifying the difference between the predicted pdf and pdf of the measured data, the best fit is identified. Five statistical parameters, viz. Correlation(R), Kolmogorov-Smirnov test (KS), Chi-square error (CS), Root mean squared error (RMSE), and R^2 test (R^2). R^2 test is used for hypothesis testing and goodness-of-fit. It is the correlation between the predicted cumulative probabilities and the observed cumulative probabilities of a wave height distribution. The high value of R^2 gives a better result with measurements. CS is used to assess whether the observed pdf is differing from the predicted pdf. RMSE gave a comparison of the deviation between observed probabilities and predicted probabilities. The small value of RMSE, KS, and CS indicates a better distribution function model and best fitting with measurements (Kollu et al., 2012). The best fit distribution is identified by ranking the pdfs for the measured data. The ranking was done in such a way that descending order is for correlation and R^2 and ascending order for KS, CS, and RMSE. For example, a large R^2 value gets the lowest rank (1), and the least R^2 value gets a rank value of 15. Similarly, a smaller RMSE gets a rank of 15, and larger RMSE gets a smaller rank value. The average rank obtained using five tests is estimated, and the highest averaged rank is considered as the best fit with the measurements.

Chi-square error (CS),
$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Kolmogorov-Smirnov (KS) = $max | E_i - O_i |$

$$\text{RMS error (RMSE)} = \sqrt{\frac{\sum_{i=1}^n (E_i - O_i)^2}{n}}$$

$$\text{Correlation (R)} = \frac{\sum_{i=1}^n (E_i - \bar{E})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (E_i - \bar{E})^2 \sum_{i=1}^n (O_i - \bar{O})^2}}$$

4.0 Results And Discussion

The surf zone wave measurements carried out were analyzed to study their characteristics for different tidal levels, different seasons, and different surf zones. The deployment locations of PTs and WTGs in the surf zone over the measured cross-section profile at Miramar, Candolim, Ganapatipule, Bahtye, Sagartirth, and Keri are shown in Figure 2. The zero elevation in Figure 2 represents the mean sea level (MSL).

Even though all the datasets were analyzed (Table 1), the detailed results were shown only for the selected locations for different tide and time. All these locations are situated within a 200km stretch and dominated by the semi-diurnal Meso-tidal regime. This study includes data records from Miramar, Candolim, and Keri, and one time observations from Bhatye, Ganapatipule, and Sagartirth, for different duration of measurements during 2012 -2014. The foreshore slope ($\tan\beta$) is varying with seasons for all the locations. The beach slope drastically increases during monsoon (June-September) due to intense wave action.

The identified pdfs were applied to check for their usefulness in depicting the distribution of the measured waves. The calculation of versatility of each pdf was in terms of R, KS, CS, RMSE, and R^2 . The performance of all the fifteen distribution functions for all the data collected from different surf zones and times were tested (Table 1). Based on the performance of the pdfs for all the six locations, the rank of each pdf was assigned from 1 to 15 (Figure 3).

For Sagartirth, the performance of each pdf is varying with tests. The performances of exponential and Inverse Gaussian were inferior, and Kernal was the best. The results from Keri, Candolim, and Miramar also confirm that the performance of the Exponential and Rayleigh was substandard, and Kernal was the best. Even though the Exponential and Rayleigh could be able to pick the distribution of Hs for Bhatye, the Kernel functions very well fit the surf zone wave-height distribution. Whereas the applicability of Rayleigh and Exponential functions was unfit for the surf zone waves except for Skewed distribution.

For each data record, the observed normalized cumulative wave height distribution was compared with a cumulative Rayleigh distribution. The goodness-of-fit tests show that the Rayleigh

distribution is not useful in the original form and needs to be modified for waves from the central west coast of India. The results showed that the Rayleigh is inadequate for the surf zone, and similar results are reported in Power (2011). The KS, CS, and RMSE were very high for Rayleigh and Exponential than the other thirteen distributions. The R^2 value was less than 1% for Rayleigh and Exponential functions, whereas it is more than 99% for the Kernel distribution function. This significant variation of R^2 can be due to (a) waves generated by non-linear wave interactions, (b) the effect of bottom friction yielding a reduction in the low-frequency components, and (c) the effect of wave breaking which would truncate the distribution and transfer some of the kinetic energy to the high-frequency components (Kuo, 1975).

Power (2011) has modified the Rayleigh distribution by analyzing the field data from 10 different beaches at 27 deployment periods; about 47% of the data was significantly different from Rayleigh distribution. The modified version of the Rayleigh distribution (hereafter modified Rayleigh) was, $P\{H \leq x\}2^2 = 1 - \exp[-(H_n/H_{rms})^p]$. To calculate the value of p , firstly, the wave heights were arranged in the descending order. Then the normalized descending ranked wave height, H_n/H_{rms} , was compared with the probability of each wave height occurring, n/N , where, n is the descending wave height rank, and 'N' is the number of waves in the data record. Finally, p is calculated from the log-log plot of H_n/H_{rms} versus $-\ln(n/N)$. The above procedure is repeated for each data record to obtain the value of p . The value of p is equal to 2 for the classical Rayleigh distribution function. The value of p greater than two indicates that waves are narrowly distributed than that predicted by a Rayleigh distribution. The value of p less than two specifies a broader distribution in comparison with the Rayleigh distribution. The Rayleigh distribution computed using p is referred to as Modified Rayleigh distribution. This process of estimating p is lengthy, and we propose a simplified method of estimating p . The p estimated using the simplified approach is denoted as p' . The p' can be estimated from the relation $((H_{max} - H_{min})/H_{rms})/(\ln(N))$, where H_{max} is the maximum wave height, H_{min} is the minimum wave height, and N is the total number of waves. Hereafter, the Rayleigh distribution using p' is referred to as the Improved Rayleigh distribution. All these three Rayleigh distribution function variants are applied for all the data collected from the study region. Power et al., (2016) reported a mean value of 2.4 for p , indicating that, on average, data are more narrowly distributed than a Rayleigh distribution (exponent of 2). They also observed that the value of p increased in shallower water, with a strong negative correlation between water depth and the value of p .

This methodology was tested, and the ratio of H_n/H_{rms} and $-\ln(n/N)$ are plotted for selected data sets, and a linear relationship is observed (Figure 4). The value of p for each location is estimated and found to vary spatially and with time. Similarly, the value of p' was computed for each location and is found to be similar to p . The different probability distributions of H_s viz., (i) Rayleigh distribution function, (ii) Modified Rayleigh distribution function, (iii) Improved Rayleigh distribution function, and (iv) the Kernal distribution functions were shown in Figure 5. The Rayleigh distribution can be used where the distribution of H_s is skewed and can see in the case of Bhatye data (20/02/2014) and Miramar data (05/06/2014). The results show that Improved Rayleigh distribution performs better for all types of surf zones and all periods of measurements.

4.1. Cross-shore and alongshore Measurements

This performance of the pdfs was tested for the measured data from Miramar (for alongshore variation) and Keri (for cross-shore variation). For Miramar, nine sensors were deployed in three cross-sections in 3 depths (0.5m, 1m, and 1.5m) for March, April, and May 2014. The H_s at 1.5m depth varied from 0.54 to 0.19 m, 0.70 to 0.33m, and 0.54 to 0.23m for March, April, and May, respectively. The H_s at 1m depth was 0.53 to 0.21 m, 0.61 to 0.26m, and 0.66 to 0.29m for March, April, and May, respectively. And for 0.5m depth, it was 0.66 to 0.22 m, 0.63 to 0.36m, and 0.78 to 0.34m for March, April, and May, respectively. The mean period T_{mean} was almost uniform and in the range of 5 to 11s for all the depths. The averaged mean wave periods were increasing from 1.5m water depth to 0.5m water depth, and this increase can be attributed to the generation of long waves from wave shoaling and wave breaking.

The normalized ranked wave height (H_n/H_{rms}) was compared with the logarithm of the probability of each wave height occurring ($-\ln(n/N)$) for half tidal cycle at Miramar for 4th March, 2nd April and 16th May 2014. The data from all three locations and three cross-sections are shown in Figure 6. The value of p and p' was obtained and used for modified and improved Rayleigh distribution functions. The results confirm that the modified Rayleigh and improved Rayleigh fit well with the surf zone observations on the central west coast of India (Figure 7). The values of p and p' range from 6.25 to 9.49 and 8.04 to 5.13, respectively.

The instruments deployed at nine cross-shore locations for Keri were named as P1 to P9 from oceanside to beachside. The instruments included five PTs and four WTGs and were deployed between 3m and 0.5m water depth. The wave characteristics from the Keri surf zone obtained using the zero-crossing method showed significant wave height (H_s) in the range of 0.2 to 0.9m and the mean wave period in the range of 6-19s.

The plot of H_n/H_{rms} vs. $-\ln(n/N)$ falls on a straight line for two out of 9 locations (Figure 8), and the remaining 7 locations, there lies a non-linear relationship. The value of estimated p considering a linear best fit line, though it fits only a part of the data, for all the cases showed a variation between 3.36 and 6.15. The value of p' is found to vary between 4.32 and 5.61, which is in the range of p . For the shallowest deployed point (P9), the H_n/H_{rms} is an almost straight line, and the number of outliers decreases from P1 to P9. The outliers correspond to the waves with their height much smaller or larger than H_{rms} . The p values are greater than 2, which indicates that wave heights are narrowly distributed. The probability distributions of H_s using (i) classical Rayleigh distribution function, (ii) Modified Rayleigh distribution function, (iii) Improved Rayleigh distribution function, and (iv) the Kernel distribution functions for locations P1 to P9 (Figure 9) clearly show that classical Rayleigh distribution does not fit well for the surfzone waves. In contrast, the Improved and Modified Rayleigh distributions fit well with the observations and are comparable with the Kernel distribution function. The larger waves break, and small waves dissipate their energy much before the shoreline. Hence, the number of larger waves and small waves, other than locally generated waves due to wave breaking and other interactions, are less in very shallow waters.

The thirteen pdfs considered in this study clearly showed that the majority of wave height distributions in the surf zones on these natural beaches differ from the classical Rayleigh distribution but provides a good relation with Kernel distribution. Observed distributions were mostly narrower, with fewer extreme values than the Rayleigh distribution, as expected with depth controlled wave breaking. There is a statistical difference between most of the observed surf zone wave-height distributions and the Rayleigh distribution, which limits the use of the Rayleigh distribution in the standard parametric models for surf zone regions. Though the observations of Thornton and Guza (1983), show that the Rayleigh distribution was qualitatively acceptable for surf zone wave heights, it is not applicable for the central west coast of India. The Rayleigh distribution performs well for the measured data set if the value of p is calculated using the proposed new method. If the best fit line of the log-log plot of H_n/H_{rms} versus $-\ln(n/N)$ could explain all the points correctly, then the predicted equation for calculating p' can be used and the value of p' will be equal to p . The relation between p and p' for all the study locations shows that for most of the locations considered in this study, p' can be used with a small degree of deviation (Figure 10). The limitation of the proposed Improved Rayleigh distribution is that it matches well with the theoretical equation for the data where the relationship between $((H_{max}-H_{min})/H_{rms})$ and $\ln(N)$ is linear.

5.0 Conclusions

The surf zone wave-characteristic studies were carried out using the time series data collected from the central west coast of India using WTGs and PTs. The time series of wave data are analyzed using FFT techniques, and the spectral wave characteristics viz., wave height, and wave period were estimated. The best-fit distribution function was identified using statistical analysis such as RMSE, KS, R^2 test, CS, and R. These parameters were used as decision criteria to check the suitability of the probability distribution functions. Results show that the Rayleigh and Exponential functions do not fit well for the surfzone waves, whereas the Kernel, GEV, and Logistic distribution functions fit very well for the surfzone wave heights. The goodness-of-fit tests show that there are significant differences between the measured wave height distribution and classical Rayleigh distribution. For each data record, the observed normalized cumulative wave height distribution was compared with a cumulative Rayleigh distribution. The study results suggest that surfzone wave heights can be modeled more accurately using best-fit pdf. The results also showed adopting a Modified Rayleigh method wherein the value of distribution function parameter p is changed would improve the poor fit to a best-fit distribution. Further, an Improved Rayleigh distribution is used wherein the parameter p is estimated using the difference of maximum and minimum wave heights to H_{rms} . This Improved Rayleigh distribution and the Modified Rayleigh distributions were observed to provide a better fit. These results indicate that wave heights can be modeled with a higher degree of accuracy using a modified parameter in Rayleigh distribution. This methodology was tested for the measured data from Miramar (for alongshore variation), and Keri (for cross-shore variation). The results are promising for all the points within the surfzone. Through this study, the feasibility of improved Rayleigh distribution has been validated for different beaches having different characteristics.

6.0 Acknowledgements [provided separately to facilitate anonymous review]

6.0 Acknowledgments

The authors gratefully acknowledge the Director, CSIR-National Institute of Oceanography, Goa. We also place our gratitude to Mr. Ryan Luis, Dr. Prakash Mehra, Project staff, and Dissertation students for their assistance during the data collection program. NIO contribution No. is xxxx.

7.0 References

- Al-Humoud, J., Tayfun, M. and Askar, H., 2002. Distribution of nonlinear wave crests. *Ocean Engineering*, 29(15): 1929-1943.
- Amrutha, M. and Kumar, V.S., 2015. Short-term statistics of waves measured off Ratnagiri, eastern Arabian sea. *Applied Ocean Research*, 53: 218-227.
- Arhan, M. and Plaisted, R., 1981. Non-linear deformation of sea-wave profiles in intermediate and shallow water. *Oceanologica Acta*, 4(2): 107-115.
- Baba, M., 1983. Wave climate at selected locations along the southwest coast of India. Paper presented in the symposium 21A of the 19th General Assembly of IUGG, Hamburg, p.^pp.
- Baldock, T., Holmes, P., Bunker, S. and Van Weert, P., 1998. Cross-shore hydrodynamics within an unsaturated surf zone. *Coastal Engineering*, 34(3): 173-196.
- Battjes, J. and Janssen, J., 1978. Energy loss and set-up due to breaking of random waves. *Coastal Engineering Proceedings*, 1(16).
- Bitner-Gregersen, E., Guedes Soares, C., Machado, U. and Cavaco, P., 1998. Comparison of different approaches to joint environmental modelling. 17th International Conference on Offshore Mechanics and Arctic Engineering, Lisbon, Portugal, Paper No. OMAE98-1495, p.^pp.
- Chakrabarti, S.K. and Snider, R.H., 1974. Wave statistics for march 1968 north Atlantic storm. *Journal of Geophysical Research*, 79(24): 3449-3458.
- Cherneva, Z., Petrova, P., Andreeva, N. and Soares, C.G., 2005. Probability distributions of peaks, troughs and heights of wind waves measured in the black sea coastal zone. *Coastal Engineering*, 52(7): 599-615.
- Dattatri, J., 1973. Waves off Mangalore harbor—west coast of India. *Journal of the Waterways, Harbors and Coastal Engineering Division*, 99(1): 39-58.
- Dattatri, J., Raman, H. and Shankar, N.J., 1979. Height and period distributions for waves off mangalore harbour—west coast. *Journal of Geophysical Research: Oceans*, 84(C7): 3767-3772.
- Ferreira, J. and Soares, C.G., 2000. Modelling distributions of significant wave height. *Coastal Engineering*, 40(4): 361-374.
- Ferreira, J. and Soares, C.G., 2002. Modelling bivariate distributions of significant wave height and mean wave period. *Applied Ocean Research*, 24(1): 31-45.
- Forbes, C., Evans, M., Hastings, N. and Peacock, B., 2011. *Statistical distributions*. John Wiley & Sons.
- Forristall, G.Z., 1978. On the statistical distribution of wave heights in a storm. *Journal of Geophysical Research: Oceans*, 83(C5): 2353-2358.

- Forristall, G.Z., 2000. Wave crest distributions: Observations and second-order theory. *Journal of Physical Oceanography*, 30(8): 1931-1943.
- Gluhovski, B., 1968. Distribution characteristics of wave parameters and changes in wave action with depth. *Reports St. Inst. Oceanology*, 93: 98-111.
- Goda, Y., 1974. Estimation of wave statistics from spectral information. *Proc. Ocean Waves Measurement and Analysis Conference*, 1974, p.^pp. 320-337.
- Goda, Y., Kishara, Y. and Kamiyama, Y., 1975. Laboratory investigation on the overtopping rate of seawalls by irregular waves. *Report of the Port and Harbour Research Institute*, vol 14, nr. 4.
- Hameed, T.S. and Baba, M., 1985. Wave height distribution in shallow water. *Ocean Engineering*, 12(4): 309-319.
- Haring, R., Osborne, A. and Spencer, L., 1977. Extreme wave parameters based on continental shelf storm wave records, *Coastal engineering 1976*, pp. 151-170.
- Haver, S., 1985. Wave climate off northern Norway. *Applied Ocean Research*, 7(2): 85-92.
- Kollu, R., Rayapudi, S.R., Narasimham, S. and Pakkurthi, K.M., 2012. Mixture probability distribution functions to model wind speed distributions. *International Journal of Energy and Environmental Engineering*, 3(1): 1-10.
- Kuo, H.-H., 1975. *Gaussian measures in Banach spaces*. Springer, pp. 1-109.
- Longuet-Higgins, M.S., 1952. On the statistical distributions of sea waves. *J. mar. Res.*, 11(3): 245-265.
- Longuet-Higgins, M., 1975. On the joint distribution of the periods and amplitudes of sea waves. *Journal of Geophysical Research*, 80(18): 2688-2694.
- Mathisen, J. and Bitner-Gregersen, E., 1990. Joint distributions for significant wave height and wave zero-up-crossing period. *Applied Ocean Research*, 12(2): 93-103.
- Mori, N. and Yasuda, T., 2002. A weakly non-gaussian model of wave height distribution for random wave train. *Ocean Engineering*, 29(10): 1219-1231.
- Ochi, M.K., 1978. *Wave statistics for the design of ships and ocean structures*.
- Ochi, M.K., 1998. Probability distributions of peaks and troughs of non-gaussian random processes. *Probabilistic Engineering Mechanics*, 13(4): 291-298.
- Power, H., Holman, R. and Baldock, T., 2011. Swash zone boundary conditions derived from optical remote sensing of swash zone flow patterns. *Journal of Geophysical Research: Oceans*, 116(C6).
- Power, H.E., 2011. *Hydrodynamics of surf and swash on natural beaches*, The University of Queensland, 230 pp.

- Power, H.E., Nielsen, P., Hughes, M.G., Aagaard, T. and Baldock, T.E., 2016. Wave height distributions in the surf zone on natural beaches. *Journal of Coastal Research*: 917-921.
- Rajasekaran, C., Jayakumar, S., Gowthaman, R., Jisad, M. and Yadhunath, E.M., 2014. Sand spit and shoreline dynamics near Terekhol river mouth, Goa, India. *Indian Journal of Marine Sciences*, 43(7): 7.
- Ruessink, B., Walstra, D. and Southgate, H., 2003. Calibration and verification of a parametric wave model on barred beaches. *Coastal Engineering*, 48(3): 139-149.
- Seelam, J.K. et al., 2015. Surfzone wave characteristics during flood tide on the central west coast of India. *Coastal Engineering Journal*, 57(04): 1550020.
- Siefert, W., 1975. Analysis of random waves in shallow water and laboratory reproduction.
- Soares, C.G., 2003. Probabilistic models of waves in the coastal zone, Elsevier oceanography series. Elsevier, pp. 159-187.
- Soares, C.G., Cherneva, Z. and Antão, E., 2003. Characteristics of abnormal waves in north sea storm sea states. *Applied Ocean Research*, 25(6): 337-344.
- Soares, C.G., Cherneva, Z. and Antão, E., 2004. Abnormal waves during hurricane Camille. *Journal of Geophysical Research: Oceans*, 109(C8).
- Tayfun, M.A., 1980. Narrow-band nonlinear sea waves. *Journal of Geophysical Research: Oceans*, 85(C3): 1548-1552.
- Thompson, E.F., 1974. Results from the CERC wave measurement program.
- Thornton, E.B. and Guza, R., 1983. Transformation of wave height distribution. *Journal of Geophysical Research: Oceans*, 88(C10): 5925-5938.
- Vanem, E., 2016. Joint statistical models for significant wave height and wave period in a changing climate. *Marine Structures*, 49: 180-205.
- Wist, H.T., Myrhaug, D. and Rue, H., 2002. Joint distributions of successive wave crest heights and successive wave trough depths for second-order nonlinear waves. *Journal of Ship Research*, 46(3): 175-185.

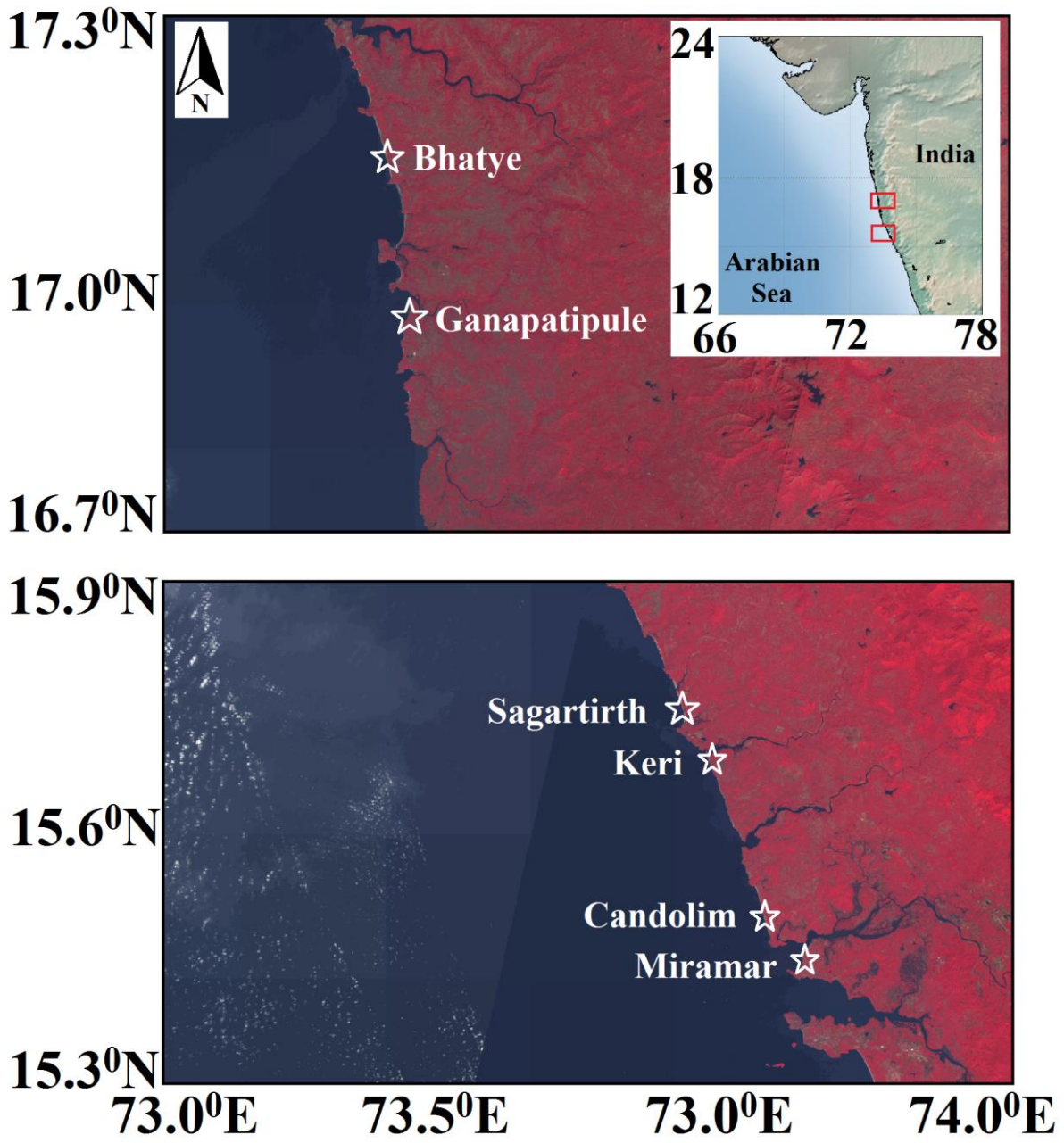


Figure 1. Map of the study region [LISS III image source: <https://mosdac.gov.in/>].

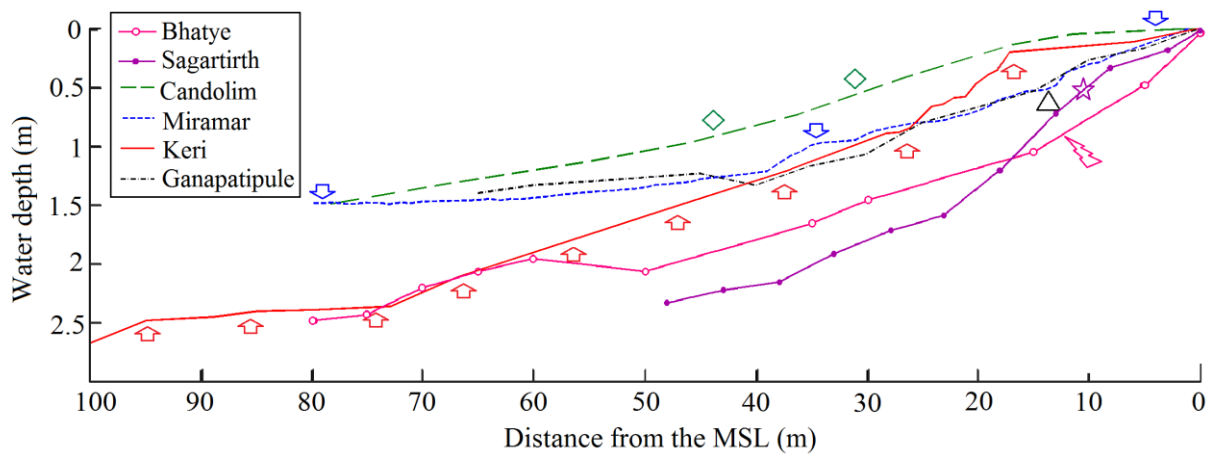


Figure 2. Plot showing beach and nearshore profiles and instrument deployment locations at different locations (Candolim (diamond shape), Keri (Up arrow), Miramar (Down arrow), Ganapatipule (Triangle), Bhatye (flash), and Sagartirth (star)).

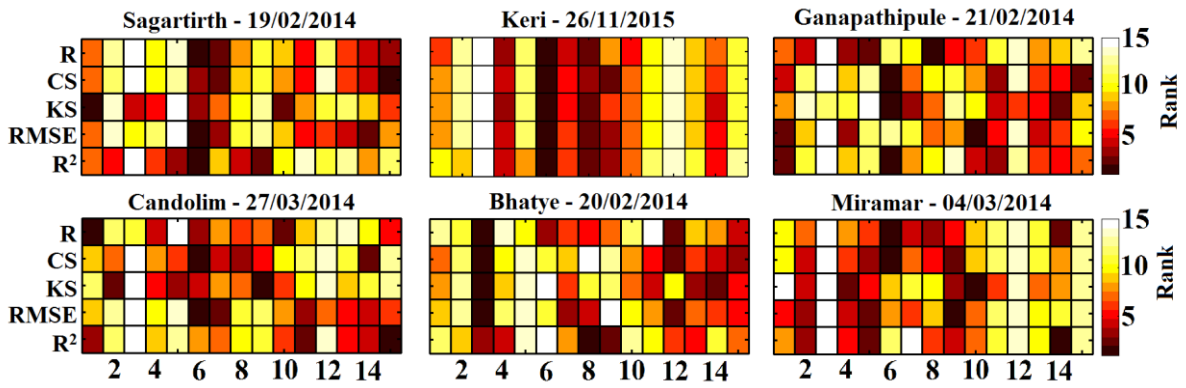


Figure 3. Plot showing the rank obtained from best fit statistical analysis of pdfs with surfzone data for different locations. The numbers along the x-axis indicate the serial number corresponding to different pdfs: 1. Beta, 2. Birnbaum-Saunders, 3. Exponential, 4. Gamma, 5. Inverse Gaussian, 6. Kernel, 7. Logistic, 8. Log-logistic, 9. Log-normal, 10. Nakagami, 11. Normal, 12. Rayleigh, 13. Rician, 14. tLocation Scale, and 15. Weibull, respectively.

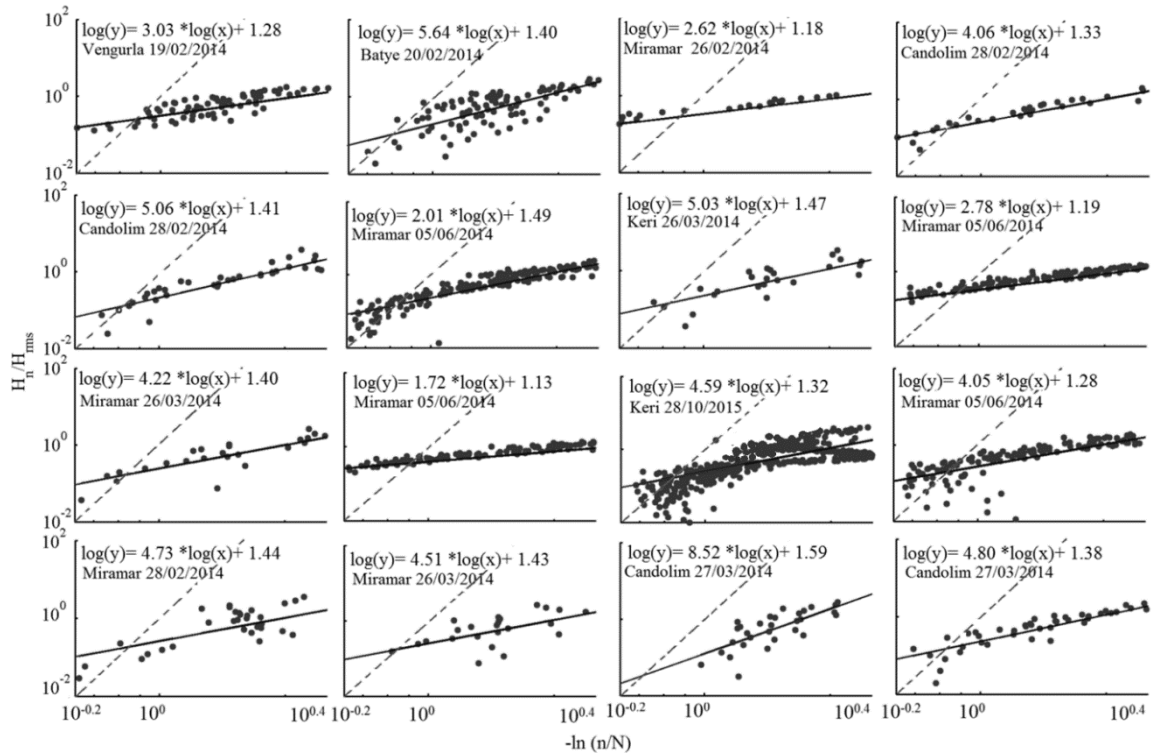


Figure 4. Normalized ranked wave height (H_n/H_{rms}) compared with the probability of each wave height occurring (n/N , \bullet) for different surf zones collected from a different time and the average best fit line (smooth line). The dates on each subplot are showing the deployment date and equation of the best fit line.

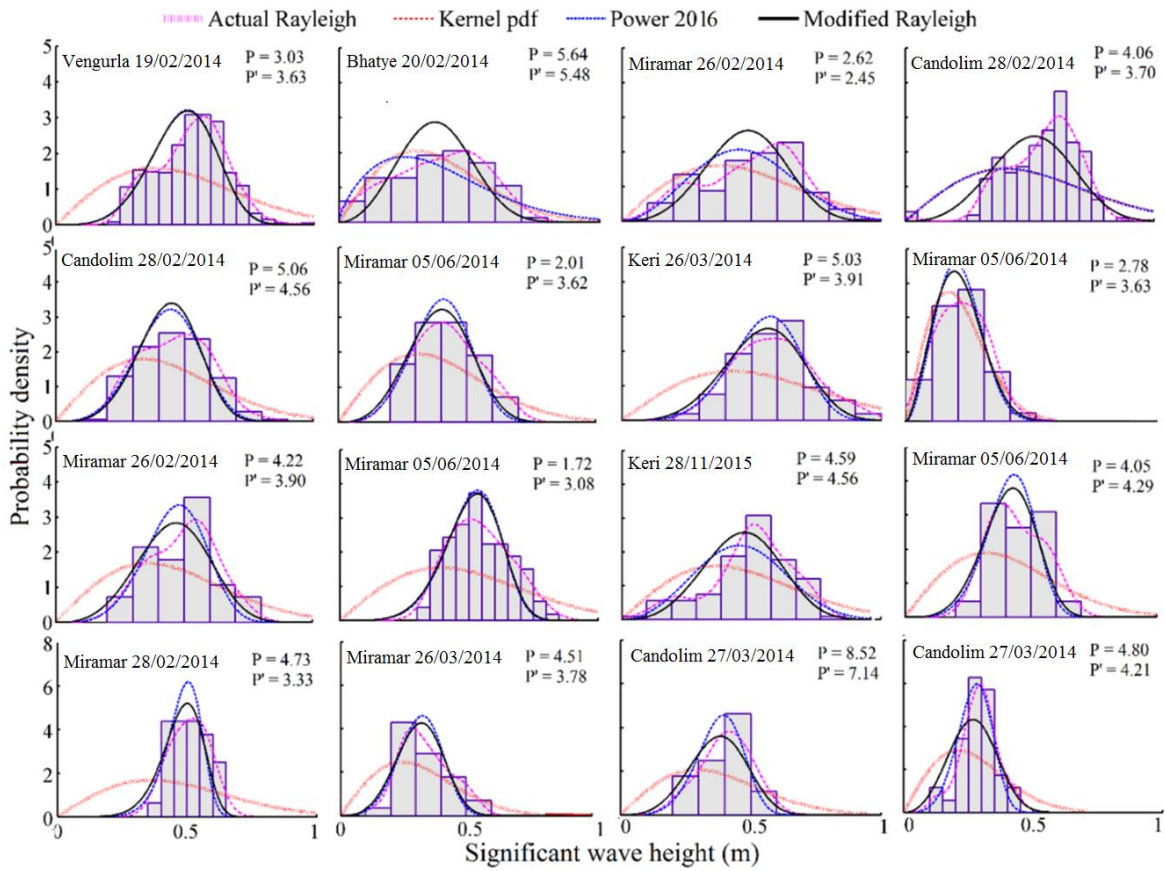


Figure 5. Comparison of the probability distribution of H_s (classical Rayleigh distribution function (magenta dotted line), modified Rayleigh distribution function (blue dot-dashed line) with an optimized value of p , improved Rayleigh distribution function p' (thick black line), best-fit Kernel pdf (red dashed line)). Each subplot shows the deployment date and corresponding values of p and p' .

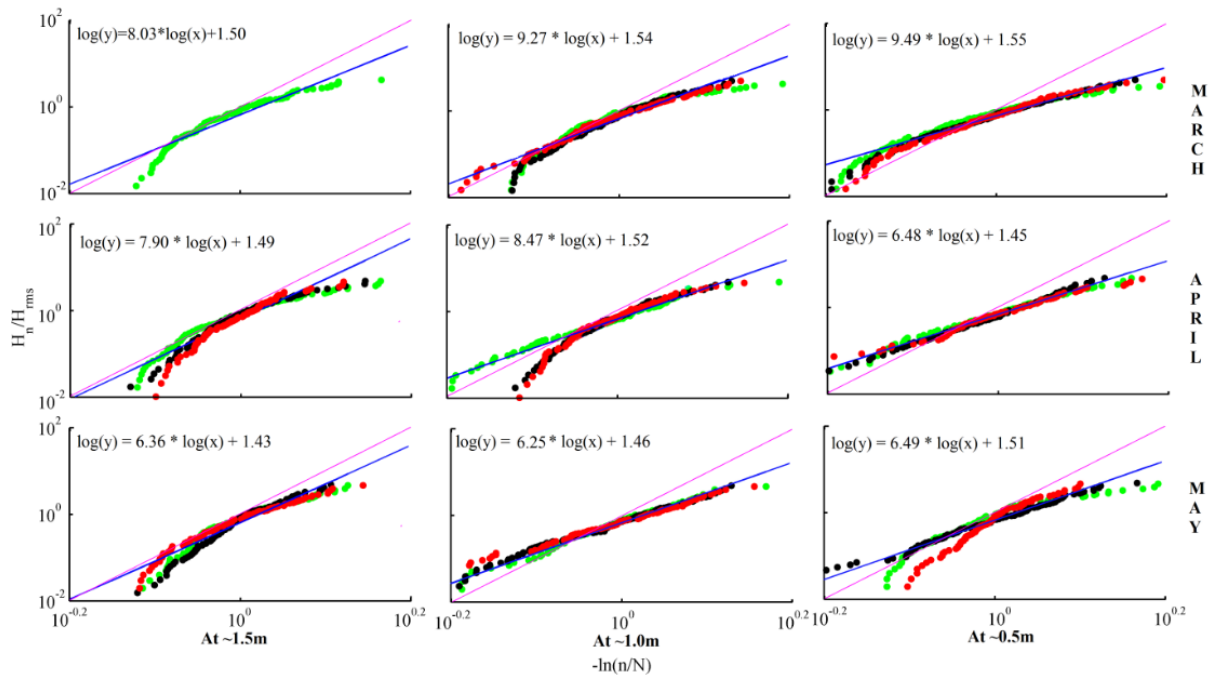


Figure 6. Normalised ranked wave height (H_n/H_{rms}) compared with the probability of each wave height occurring (n/N , \bullet), for a 4-hour data record from Miramar for 4th March, 2nd April and 16th May 2014 (Green indicates from the north side, black for middle and red for south) and the average best fit line (blue line) and 1:1 line (magenta line)

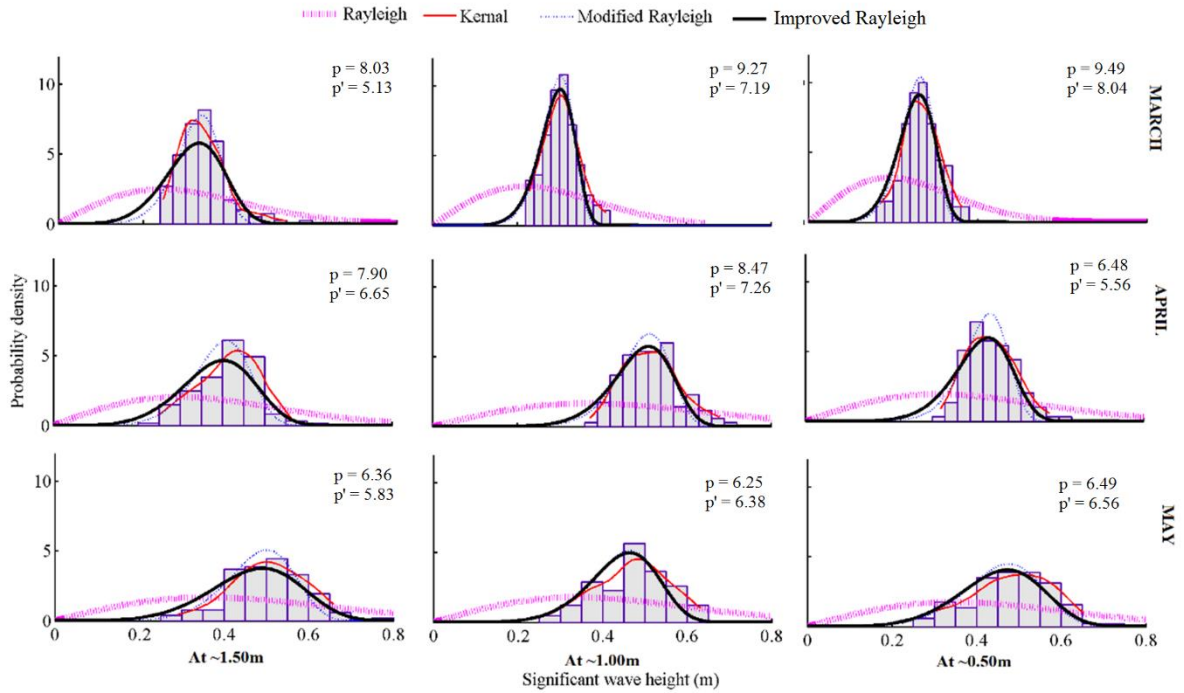


Figure 7. The probability distribution of H_s compared to the Rayleigh distribution function (magenta dotted line), the modified Rayleigh distribution function (blue dot-dashed line) with an optimized value of p and p' is explained in terms of equation $p' = (H_n - H_1)/H_{rms} / (\log(n_1/n_n))$ is plotted on the thick black line, and the best-fitted Kernel pdf is in red dashed line for Miramar

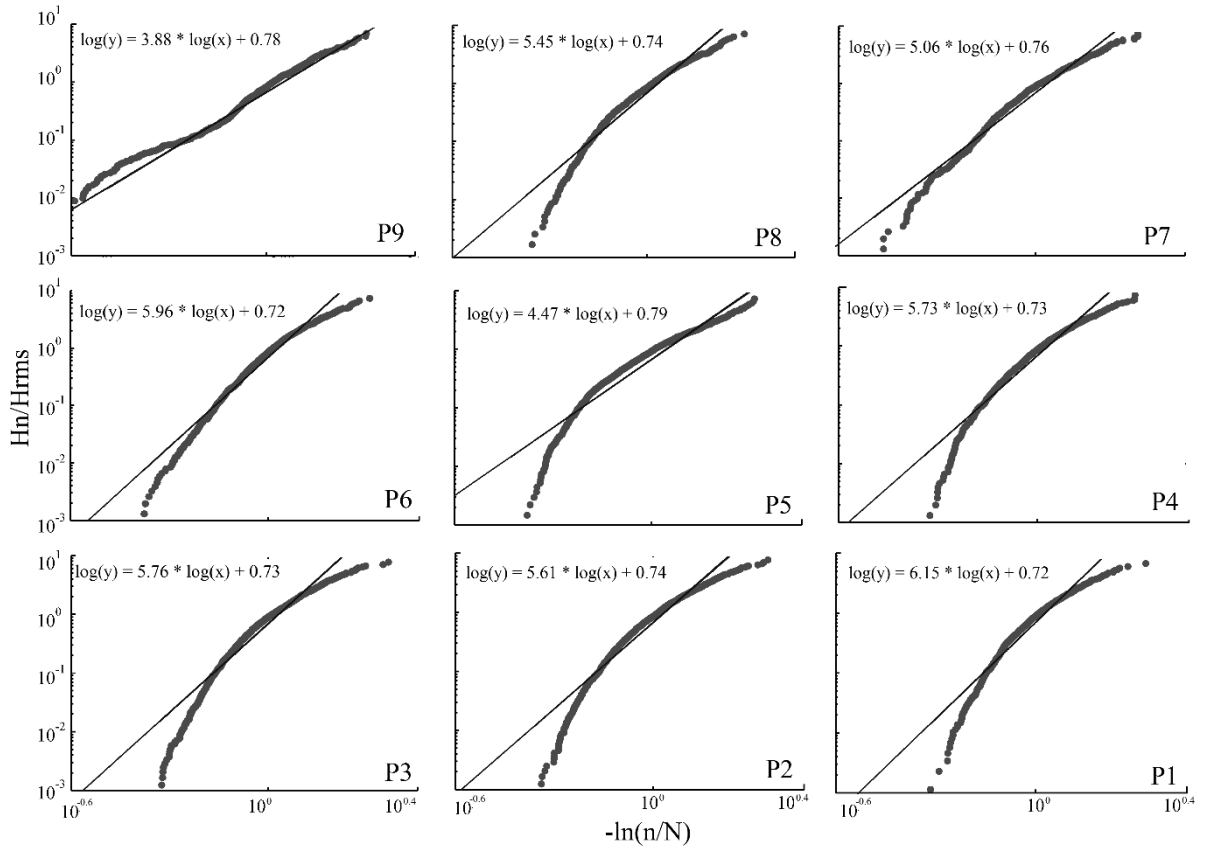


Figure 8. Normalized ranked wave height (H_n/H_{rms}) compared with the probability of each wave height occurring (n/N , \bullet), for Keri offshore (P1) to shallow (P9) and the average best fit line (smooth line).

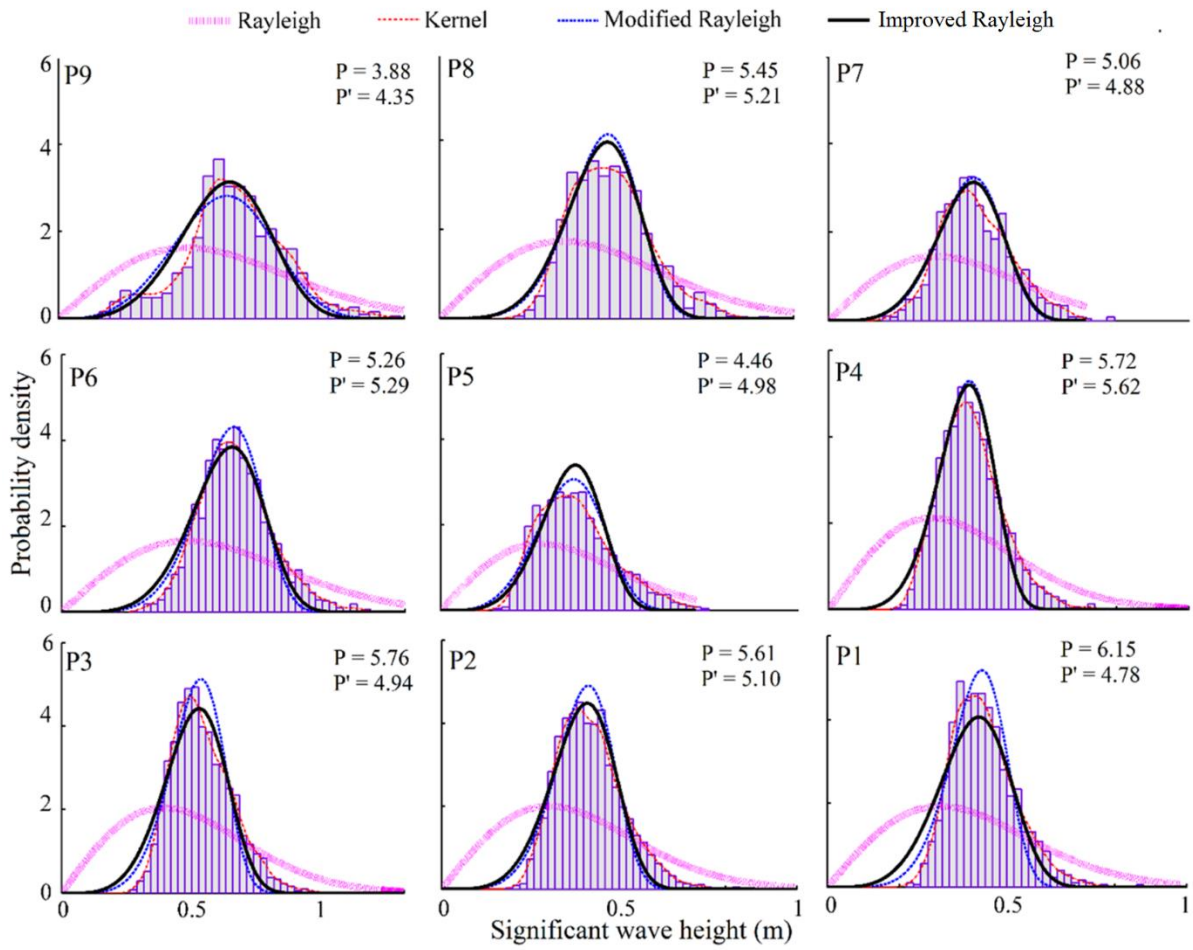


Figure 9. The probability distribution function for H_s for Keri location (classical Rayleigh distribution function (magenta dotted line), modified Rayleigh distribution function with an optimized value of p (blue dot-dashed line), and improved Rayleigh distribution function with a value of p' (thick black line), best-fit Kernel pdf (Red dashed line).

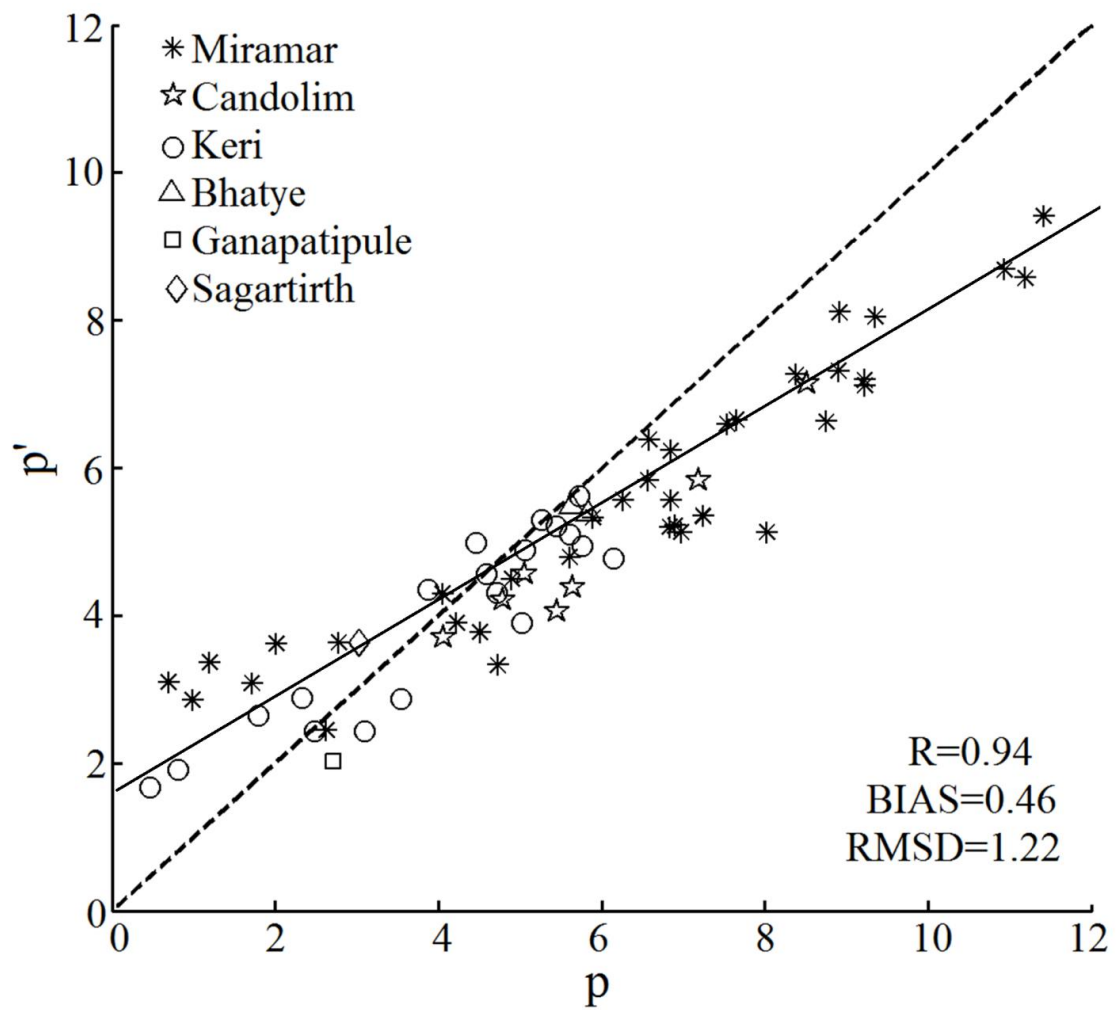


Figure 10. The relation between $p=(H_w/H_{rms})/(\ln(n/N))$ and $p'=((H_{max}-H_{min})/H_{rms})/(\ln(N))$ for different beaches (see text inside for description of the symbols). The best data fit line is plotted (solid line) having linear equation ($p'=0.6p+1.8$)

Table 1. Data records for each deployment: Location, date, number of 2.13 min data records (n), number of PTs in the surf zone (nPT), sampling frequency (Fs), Average number of waves analysed (Nw), beach slope in the inner surf zone ($\tan\beta$), Significant wave height (H_s), mean wave period (T_{mean}) and root mean square wave height in the inner surf zone (H_{rms}) for each deployment.

Location	Date	n	nPT	Fs(Hz)	Nw	Tan β	H_s (m)	T_{mean} (s)	H_{rms} (m)
Miramar	18/01/2012	85	1	4	664	0.036	0.34	17.50	0.40
Miramar	26/02/2014	170	2	8	1424	0.275	0.42	4.72	0.43
Miramar	28/02/2014	170	3	8	810	--	0.48	7.45	0.49
Miramar	04/03/2014	280	7	8	4390	--	0.53	10.35	0.53
Miramar	02/04/2014	280	9	8	4877	0.042	0.63	7.99	0.63
Miramar	16/05/2014	280	9	8	4189	0.233	0.65	8.81	0.65
Miramar	05/06/2014	280	4	8	3783	0.597	0.85	6.15	0.88
Candolim	19/01/2012	85	1	4	829	0.052	0.42	14.50	0.43
Candolim	28/02/2014	26	2	8	771	0.834	0.55	15.38	0.55
Candolim	27/03/2014	91	4	8	919	0.805	0.86	11.66	0.87
Keri	17/01/2012	85	1	4	969	0.656	0.30	19.50	0.31
Keri	27/02/2014	28	2	8	820	0.391	0.44	21.28	0.44
Keri	26/03/2014	28	5	8	332	0.395	0.49	19.02	0.50
Keri	06/06/2014	166	2	8	2190	0.141	1.11	19.52	1.12
Keri	20/08/2014	156	2	8	2495	0.089	0.73	12.80	0.74
Keri	27/10/2015	495	3	8	7489	0.155	0.98	11.40	0.98
Keri	26/11/2015	6809	9	8	20253	0.090	0.73	17.80	0.74
Bhatye	20/02/2014	91	4	8	2387	0.0218	0.86	11.66	0.87
Ganapatipule	21/02/2014	14	2	8	183	0.0125	0.37	16.45	0.38
Sagartirth	19-02-2014	94	4	8	1770	--	0.81	13.75	0.88

Appendix: 1

Mathematical representations of the Probability Distribution Functions

Sl. No.	Probability distribution functions	Equations	Parameters
1	Beta	$y = f(x a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} I_{[0,1]}(x); 0 < x < 1$	a and b
2	BimbaumSaunders	$y = f(x) = \frac{1}{2\sqrt{2\pi\alpha\beta}} \left[\left(\frac{\beta}{x}\right)^2 + \left(\frac{\beta}{x}\right)^{3/2} \right] \exp \left[-\frac{1}{2\alpha^2} \left(\frac{x}{\beta} + \frac{\beta}{x} - 2 \right) \right]$	α and β
3	Gamma	$y = f(x a, b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}}$	a and b
4	Exponential	$y = f(x \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$	μ
5	Inverse Gaussian	$y = \sqrt{\frac{\lambda}{2\pi x} \exp \left\{ \frac{-\lambda}{2\mu^2 x} (x - \mu)^2 \right\}}$	λ and μ
6	Kernel	$y = \frac{1}{nh} \sum_{i=1}^n K \left(\frac{x-x_i}{h} \right) ; -\infty < x < \infty$	Zh
7	Logistic	$f(x \mu, \sigma) = \frac{\exp \left\{ \frac{x-\mu}{\sigma} \right\}}{\sigma \left(1 + \exp \left\{ \frac{x-\mu}{\sigma} \right\} \right)^2}; -\infty < x < \infty$	μ and σ

8	Log-logistic	$f(x \mu, \sigma) = \frac{1}{\sigma x (1+e^z)^2} ; x \geq 0 ; z = \frac{\log(x) - \mu}{\sigma}$	μ and σ
9	Log-normal	$f(x \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}; ; x > 0$	μ and σ
10	Nakagami	$y = 2\left(\frac{\mu}{\omega}\right)^\mu \frac{1}{\Gamma(\mu)} x^{(2\mu-1)} e^{-\frac{\mu}{\omega}x^2}$	μ and ω
11	Normal	$y = f(x \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ and σ
12	Rayleigh	$y = f(x b) = \frac{x}{b^2} e^{-\frac{x^2}{2b^2}}$	Ab
13	t-LocationScale	$y = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left[\frac{\nu + \left(\frac{x-\nu}{\sigma}\right)^2}{\nu}\right]^{-\left(\frac{\nu+1}{2}\right)}$	ν and σ
14	Weibull	$f(x a, b) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} e^{-(x/a)^b}$	a and b
15	Rician	$y = \left(\frac{x}{\sigma^2}\right) \exp\left(-\frac{(x^2 - \nu^2)}{2\sigma^2}\right) I_0\left(\frac{x\nu}{\sigma^2}\right)$	σ and ν