Non-Gaussian analysis methods for planing craft motion

Abhilash Somayajula[*] and Jeffrey M. Falzarano[a]

Marine Dynamics Laboratory, Texas A&M University, College Station, Texas 77843, USA

Abstract.   Unlike the traditional displacement type vessels, the high speed planing crafts are supported by the lift forces which are highly non-linear. This non-linear phenomenon causes their motions in an irregular seaway to be non-Gaussian. In general, it may not be possible to express the probability distribution of such processes by an analytical formula. Also the process might not be stationary or ergodic in which case the statistical behavior of the motion to be constantly changing with time. Therefore the extreme values of such a process can no longer be calculated using the analytical formulae applicable to Gaussian processes. Since closed form analytical solutions do not exist, recourse is taken to fitting a distribution to the data and estimating the statistical properties of the process from this fitted probability distribution. The peaks over threshold analysis and fitting of the Generalized Pareto Distribution are explored in this paper as an alternative to Weibull, Generalized Gamma and Rayleigh distributions in predicting the short term extreme value of a random process.

Keywords: peaks over threshold; generalized pareto distribution; goodness of fit; extreme value prediction; return level plot

1. Introduction

The design of any structure requires an assessment of the extreme response of the structure during its design life. In the case of boats, ships and offshore structures, the responses are governed by random wave environment. Therefore, it is important to understand the statistics of the response in waves. In the absence of an analytical solution, the statistics of the response are estimated based on previous data sets or experiments or numerical simulations. Prediction of extreme responses based on existing time series data is called extreme value analysis. It is a set of statistical tools developed in the past years which has found a wide application across many disciplines.

The methods of prediction of the response of a structure to random ocean waves can be classified into two types - analytical approaches and fitting distributions. The analytical approaches rely on theoretical formulation of the problem and try to estimate the statistical responses on the basis of this theory. However fitting methods try to fit a particular distribution through the time series obtained through experiments or simulations to estimate the extreme values. While the analytical formulations maybe more accurate, they are difficult to develop and are

---

∗Corresponding author, Ph.D. Student, E-mail: sabhilash@email.tamu.edu
[a] Professor, E-mail: jfalzarano@civil.tamu.edu

subject to change by the problem at hand. The fitting methods are more general but do not have a solid theoretical background supporting the problem.

Some of the analytical approaches for estimating the responses of traditional ships have been explored by Texas A&M University's Marine Dynamics Laboratory. Mulk and Falzarano (1994) studied the coupled heave and pitch motions using a complete six degree of freedom Euler's equations of motion approach to investigate the dynamic bias phenomenon. Jamnongpipatkul (2010) estimated the expected time to capsize for a ship by solving the stochastic differential equation for roll motion using path integral methods. Continuing on this work, Su and Falzarano (2013) compared the mean first passage time as calculated from assuming ship rolling to be a Markov process to that obtained by the application of Melnikov's method. Su and Falzarano (2011) also developed an automatic cumulant neglect tool to close the higher order moment equations to include the non-Gaussian nature of beam sea rolling. Moideen *et al.* (2013) investigated the parametric rolling of ships in regular waves in head seas. Continuing on the previous work, Moideen *et al.* (2014) investigated the stochastic nature of parametric roll of container ships in head seas using a Volterra series expansion method. Somayajula and Falzarano (2014) improved upon the Volterra series approach and used it to investigate the ergodicity of parametric roll of container ships in head seas. Somayajula *et al.* (2014a) also investigated the effects of parametric roll in irregular seas and the mean drift forces on a ship on the EEDI (Energy Efficiency Design Index) performance index of ships.

The fitting approach is more common when the theoretical background of the problem is complex or unknown. Fridsma (1969) performed the first systematic investigation of the planing hull motions, accelerations and forces. He concluded that the peaks of the heave motion followed a "Distorted Rayleigh Distribution" given by Rice (1944) and the pitch acceleration followed an exponential distribution. Ochi (1978) has also suggested that the use of exponential distribution for the prediction of extreme values for a squared Gaussian process. The other works by Ochi include Ochi (1978b), Ochi (1978a) Ochi (1981) and Ochi (2005) describe further details on the prediction of extreme values of both wind waves and responses of structures to wave loading. This paper investigates fitting of the Generalized Pareto distribution of which the exponential distribution is a special case.

The other investigators of similar problems include Lewis (1994) who performed a statistical analysis of the hull girder time response histories and later also investigated the wave impact design pressures (Lewis 2005). A lot of details about the spectral methods of analysis have been given by Priestley (1981) in his books. Priestley (1988) also describes various methods of analysis for nonlinear and non-stationary time series.

Over the years, many approaches to predict the extremes of a response based on fitting distributions to an existing time series have been developed. Some of the most widely used methods may be classified (Vanem *et al.* 2013) into

1. Initial distribution method
2. Block maxima method
3. Peaks over threshold (POT) method
4. Mean Number of Up-crossings (MENU) method

In the initial distribution method the extreme value is estimated as the quantile q_p of the response distribution $F(x)$ with a specified probability p. The block maxima method divides the time series into separate blocks and the extremes from each block are collected together. These

block maxima are assumed to follow one of the three Generalized Extreme Value Distributions (Gumbel, Frechet or Weibull). The peaks over threshold uses the peak values of data exceeding a threshold and models them using a Generalized Pareto Distribution (Wang 2001) (Coles 2001). This method is gaining popularity in the analysis of motions of marine vessels. Guha *et al*. (2013) used the peaks over threshold method combined with Generalized Pareto fitting to analyze the connection forces between modules of an Improved Navy Lighterage System (INLS). More details on this work might be found in Guha *et al*. (2013b). The Mean Number of Up-crossings method estimates the extreme value by imposing a condition that the expected number of up-crossings beyond this extreme value is one in the given time duration.

When the response is Gaussian, then the extremes of the process are given by Rayleigh or Rice distribution (Rice 1944) based on the bandwidth parameter $\grave{O}$ and it is possible to arrive at an analytical expression for the extreme value based on the MENU method (Cramér and Leadbetter 1967). Continuing on this approach Tromans *et al*. (1991), Winterstein *et al*. (1998), Torhaug (1996), Taylor *et al*. (1997) and others developed methods to predict the extremes from the responses to shorter time series.

In this paper the motions of a planing craft in random seas are analyzed. We shall first show using multiple methods that unlike the displacement ship responses, the planing craft motions are highly non-Gaussian. Then we shall describe the application of the peaks over threshold approach to the non-Gaussian planing craft motions to estimate the extreme values of the response. A brief description of the data collection system for the planing craft motions is followed by a detailed description and results of the statistical analysis of the motion time series.

2. Planing craft motion data

A planing craft was tested by NSWC Carderock on 27 April 2005 and 05 May 2005. The motion data including the tri-axial accelerations and tri-axial angular rates was collected.

Motion data was collected using IST EDR-6 degree of freedom 50/300 self-contained data recorders at a sampling rate of 512 samples per second for all data channels. Each data recorder contains an internal tri-axial piezo-resistive accelerometer with a nominal full-scale range of 50 g and an internal tri-axial MEMS (Micro Electro-Mechanical Systems) -based angular rate sensor with a nominal full-scale range of 300 deg/sec. The bandwidth of accelerometers and angular rate sensors was 200 Hz and 60 Hz respectively.

Analog signals are converted to discrete time samples by digital sampling. The number of quantization levels available in the data recorders analog-to-digital converter is 1024 levels (10 bits) equally spaced over the full-scale range of the sensor. Consequently, acceleration resolution is 0.1 g (100 g/1024) and angular rate resolution is 0.6deg/s (600deg/s/1024).

The test craft was driven at speeds ranging between approximately 28 and 40 knots (depending on the sea heading) for about 10 minutes at each direction to the sea, starting in a head-sea aspect. The average craft speed was determined using a handheld GPS receiver. The active control system was deactivated during these runs.

The data was collected by the US Navy and was later presented to Marine Dynamics Laboratory at Texas A&M University where further statistical analysis was performed.

2.1 Filtering

The data was collected at intervals of 0.002 seconds and includes a significant amount of noise as seen from Fig. 1. Before performing a statistical analysis on this data, it is necessary to remove the noise form it. The noise was filtered by the use of a 2 Hz low pass Lanczos filter shown in Fig. 2. The filtered time series is shown in Fig. 3. It can be seen that the filtered time series is devoid of all the sudden peaks which are clearly unphysical. The statistical analysis was performed on this filtered data set.
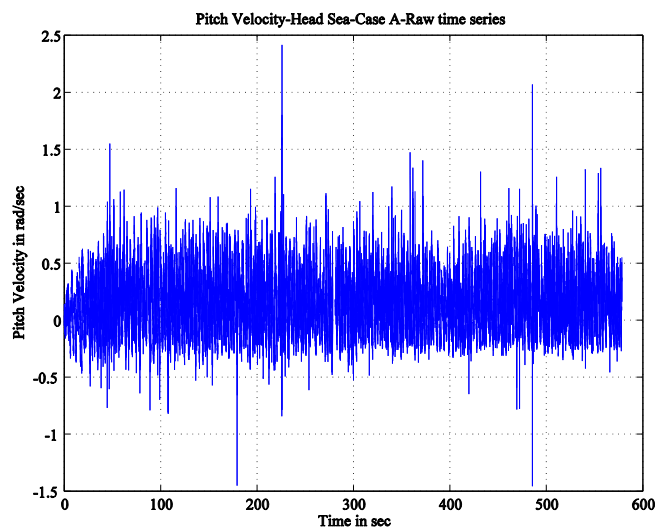

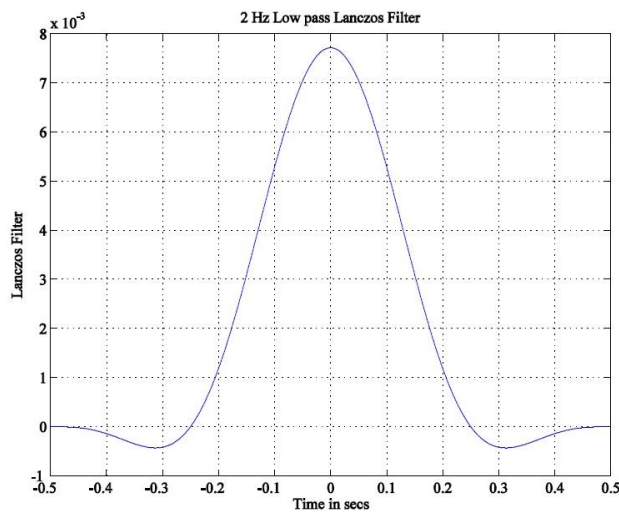Fig. 1 Raw time series of pitch motion in head seas
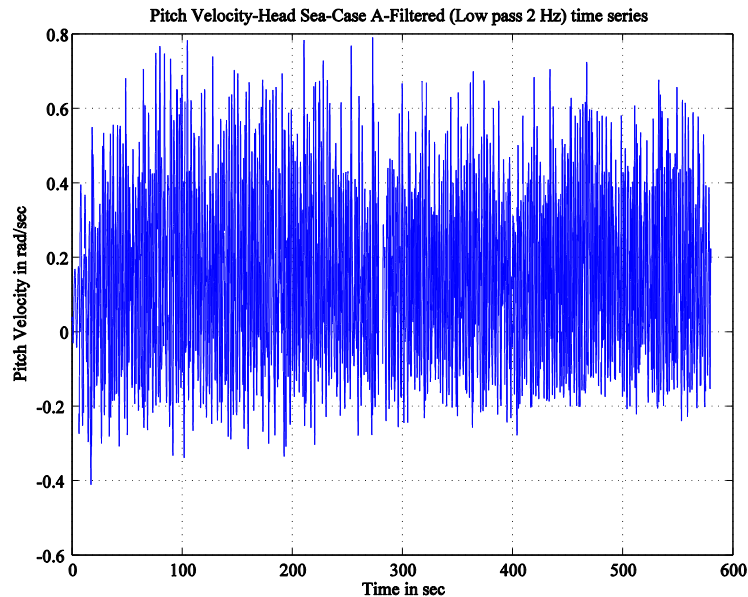

Fig. 2 2 Hz low pass Lanczos Filter

Fig. 3 Filtered time series of pitch motion in head seas

3. Non-Gaussian nature of the data

The motions of the planing craft are not governed by the standard linear equations of motion as the lifting forces on the hull make the problem non-linear. Due to this non-linearity, the responses of the craft would not follow a Gaussian distribution. Thus the standard techniques to estimate the extreme values are not applicable to this problem and a different approach is needed to calculate the extreme values for these responses. However, for only small non-linearities, it is still possible to use a Gaussian assumption and get approximate results for the extreme values. Therefore, there is a need to check for the non-Gaussian nature of motion time series. Two methods are described below to check for the non-Gaussian nature of the planing craft motions.

3.1 Quantile-Quantile plot

The quantile-quantile plot (QQ plot) is a plot of the quantiles of the data plotted against the same quantiles from a specified distribution. If the data follows the specified distribution, then the quantile-quantile plot will be a straight line passing through the origin. Therefore, generating a quantile-quantile plot for normal distribution will clearly indicate if the data is Gaussian or not. Figs. 4(a) and 4(b) show the quantile-quantile plot for the pitch motion time series shown in Fig. 1 and Fig. 3 respectively.

It can be clearly seen that the outliers in the original QQ plot have been eliminated in the filtered QQ plot. However, the non-Gaussian nature of both the time series is evident from the divergence from the straight line.
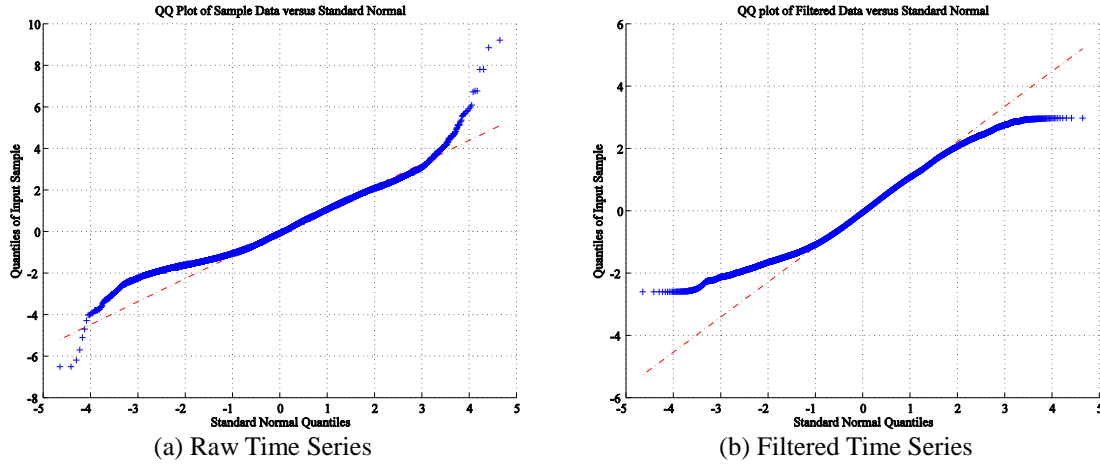
(a) Raw Time Series   (b) Filtered Time Series

Fig. 4 QQ plot of data quantiles versus normal quantiles

3.2 Hypothesis testing

Statistical hypothesis testing is a statistical inference tool which uses the data from a scientific study to ascertain if an assumption about the data (hypothesis) can be accepted or rejected with a statistical significance. This is the alternative approach to quantile-quantile plots to check for the Gaussianity of the data set. There are two different hypothesis tests - Kolmogorov-Smirnov Test and Chi-Squared Test - which can be used to determine if the data set is Gaussian or not.

3.2.1 Kolmogorov-Smirnov test

The null hypothesis in this test is chosen that the data follows normal distribution. The test statistic D_n for the Kolmogorov-Smirnov test is given by Eq. (1).

$$D_n = \sup_x | F_n(x) - F(x) | \tag{1}$$

where $F_n(x) = \dfrac{1}{n} \sum_{i=1}^{n} 1_{X_i \leq x}$ is the empirical distribution for n independent identically distributed

(iid) observations $X_1, X_2, ..., X_n$ and $F(x)$ is the standard normal distribution.

The null hypothesis is rejected at a significance level α if $\sqrt{n}D_n > K_\alpha$, where K_α is calculated from the cumulative distribution function of the Kolmogorov distribution as shown in Eq. (2).

$$P(K \leq K_\alpha) = \frac{\sqrt{2\pi}}{K_\alpha} \sum_{i=1}^{\infty} e^{-\frac{(2i-1)^2 \pi^2}{8K_\alpha^2}} = 1 - \alpha \tag{2}$$

The application of the Kolmogorov Smirnov test to the time series shown in Fig. 3 results in the rejection of the null hypothesis for a chosen $\alpha = 0.05$ (corresponding to a 95% statistical

significance). Thus the test confirms that the pitch time series is non-Gaussian.

3.2.2 Chi-squared test

The null hypothesis in this test is chosen that the data follows normal distribution. The first step is to construct an n-bin histogram of the data where n is a large number. The test statistic χ_n^2 for the Chi-squared test is given by Eq. (3).

$$\chi_n^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}$$

(3)

where o_i is the observed frequency for the i-th bin of the n-bin histogram of the data such that $\sum_{i=1}^{n} o_i = N$ where N is the total number of data points and e_i is the expected frequency for the i-th bin based on the null hypothesis (Gaussian distribution in this case).

As $n \to \infty$ the test statistic χ_n^2 follows a χ^2 distribution with $n-1$ degrees of freedom. The null hypothesis is rejected at a significance level α if $p < \alpha$ where p is calculated from χ^2 probability density function as shown in Eq. (4).

$$p = P[\chi^2 \geq \chi_n^2] = \int_{\chi_n^2}^{\infty} p_{\chi^2}(x, n-1)dx = \int_{\chi_n^2}^{\infty} \frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}}$$

(4)

where $\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$ is the Gamma function.

The application of the Chi squared test to the time series shown in Fig. 3 results in the rejection of the null hypothesis for a chosen $\alpha = 0.05$ (corresponding to a 95% statistical significance). Thus the test confirms that the pitch time series is non-Gaussian.

4. Extreme value analysis

When a probability distribution is fit to any data, the regions with high density of data points have a good agreement with the distribution function. However, the fit of the distribution in the tail region with fewer data points is usually not accurate. In many applications like design, the primary interest is to accurately predict the extreme values of the process which requires a good fit in the tail region. The Generalized Pareto Distribution is better suited for such applications than any other distributions. This paper discusses the theory and application of the Generalized Pareto Distribution to the planing craft motion data collected for the US Navy ONR project.

4.1 Peaks over Threshold (POT) - theory

In general, probability distribution models are used to achieve an accurate fit at the high density regions. In extreme value analysis, it is more important that the fitting between the data and model

is more accurate in the tail regions, even if it does not fit accurately in the high density regions. For this reason, Peaks over Threshold (POT) is a popular choice for extreme value analysis. As the name suggests, only the data peaks higher than a particular threshold are chosen while leaving out the smaller peaks. Fitting of the distribution is performed to only these peaks over threshold to get a better fit in the tail region.

Generalized Pareto Distribution (GPD) is based on the peaks over threshold theory and can be used to model the tail regions of a wide variety of distributions (Coles 2001) (Wang and Moan 2004). The GPD models the tail of the data as a conditional distribution given that the peak exceeds the threshold value. Thus only the data points which are beyond the threshold value are of interest for the analysis.

However, for the calculation of the short term extreme value based on a probability of exceedance, the cumulative distribution function of the process is required. This means there is a need to combine the empirical cumulative density function with the smoother tail fitting of Generalized Pareto Distribution. The evaluation of extreme value based on this method is described below.

4.2 Generalized Pareto Distribution – theory

The Generalized Pareto Distribution is a two parameter (k and σ) distribution where, k is the shape parameter and σ is the scale parameter. Fig. 5 shows the probability density of the GPD for positive, zero and negative values of the shape parameter. It is a generalization of both exponential distribution ($k = 0$) and Pareto distribution ($k > 0$). The GPD includes these two distributions in a larger family so that a continuous range of shapes is possible.
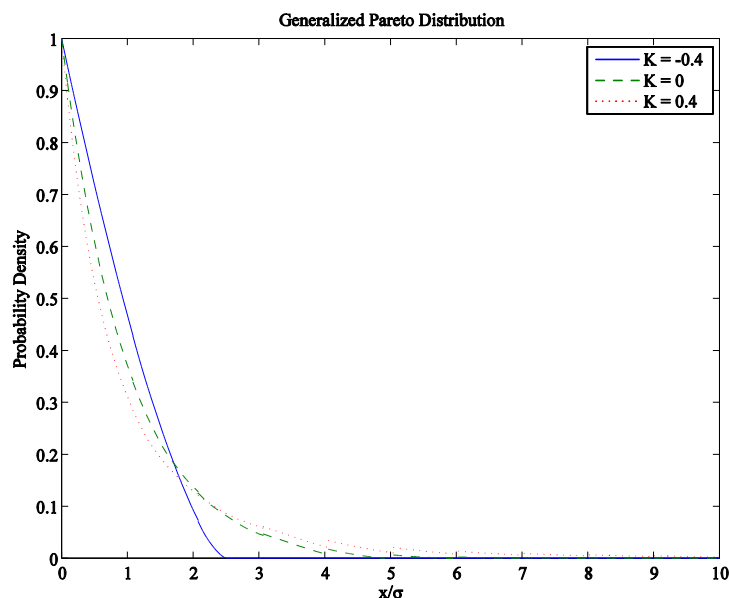

Fig. 5 Variation of GPD density with shape parameter k

When the shape parameter is negative, the distribution has zero probability above an upper limit of $-1/k$. For non-negative values of k, GPD has no upper limit. Higher values of k result in a heavier tailed distributions.

The GPD gives the conditional probability of the exceedances of a process X over a specified threshold value u. The conditional cumulative distribution is given in Eq. (5).

$$Pr(X > x \mid X > u) = \left(1 - k\left(\frac{x-u}{\sigma}\right)\right)^{-\frac{1}{k}}$$ (5)

4.3 Estimation of short term extremes

The analysis of short term extremes of the planing craft motions is vastly different from the analysis of usual offshore structures where the focus is on estimating the n-year return level of response or the survivability in a 1000 year Hurricane. Instead of n-year return value of the response, the primary interest is in the short term extreme experienced during the operation of the planing craft.

Unlike for the long term extreme value which requires simulations for long times (usually 3 hours or more for a hurricane environment), it is not necessary to perform very long simulations to ascertain the short term extreme.

The first step in the estimation of short term extreme is to pick an appropriate threshold and extract all the peaks over threshold. The next step is to fit the GPD through these peaks over threshold. The third step is to find the Cumulative Distribution function in the tail from the conditional GPD and the empirical cumulative distribution of the data. The last step is to find the extreme value based on a given probability of exceedance. Details of the estimation of the short term extreme value are described below.

4.3.1 Threshold selection

Threshold is a very sensitive parameter and must be chosen carefully. It should be appropriately chosen such that it is sufficiently high enough to model the tail data but at the same time should have sufficient number of data points to give a meaningful fit.

Threshold selection depends a lot on the data and appropriate fit achieved for the exceedances over the threshold. Wang and Moan (2004) suggests that even though there have been some attempts to come up with a specific method to choose the threshold, there is still no uniform accepted method. In the present analysis of Planing Craft Data, the threshold is chosen as the 85% quantile of the peaks of the data. This means that the threshold is chosen such that 85% of the peaks of the data are less than or equal to the threshold.

4.3.2 Generalized Pareto Fitting

Once the peaks over threshold are extracted from the data, the exceedances are obtained by subtracting the threshold value from the peaks over threshold. The generalized Pareto distribution is then fit to the exceedances calculated above.

Fitting of GPD to the data is performed using the inbuilt MATLAB function "gpfit". This function takes as input the exceedances and calculates the best fit GPD parameters (scale parameter k and the shape parameter σ). The fitting is performed by the Maximum Likelihood Method. The Likelihood function is the joint probability density of the given sample for a given

set of parameters. The Maximum Likelihood Method tries to maximize the Likelihood function by iteratively varying the parameters k and σ.

For a sample of size n, x_i and a joint probability density function $f(x_1, x_2, ..., x_n, k, \sigma)$, the likelihood function L is given by Eq. (6).

$$L = f(x_1, x_2, ..., x_n, k, \sigma) \tag{6}$$

If the sample is considered to be statistically independent, then the likelihood function can be expressed as a product of the individual probability densities $f(x_i, k, \sigma)$.

$$L = \prod_{i=1}^{n} f(x_i, k, \sigma) \tag{7}$$

The estimated value for the parameters is obtained by a simultaneous solution of Eq. (8).

$$\frac{\partial \ln L}{\partial k} = 0$$

$$\frac{\partial \ln L}{\partial \sigma} = 0 \tag{8}$$

Estimation of confidence intervals

The estimation of confidence intervals for the shape and scale parameters is also performed internally by the MATLAB function "gpfit". The shape parameter k and the logarithm of the scale parameter $\ln(\sigma)$ are assumed to be normally distributed with their asymptotic variance calculated by the inverse of the Fischer's Information Matrix. For both parameters the mean is assumed to be the best fit value. Based on this information, the confidence intervals are calculated by taking an inverse of the normal cumulative distribution function. The theoretical details of estimating the confidence intervals are quite established and can be found in standard textbooks such as Embrechts *et al*. (1997) and Kotz and Nadarajah (2000). More information may also be found in Bradley (2013).

4.3.3 Goodness of fit
Once the distribution has been fit, it is important to check that the fit is good enough. If the fit is not good, then the extreme values calculated from this fitted distribution will be far off from the actual extremes of the process.

A good way to check if the fit of the distribution to the data is good is to plot data quantiles versus the distribution quantiles. The Quantile-Quantile (QQ) plot comparison of the exceedances versus the GPD would indicate the goodness of fit.

Fig. 3 shows the example of a time series of pitch motion of the planing craft in one of the cases. Similar trends have been obtained for other test cases too. However, only the results of one case are presented here. The results for the other cases may be found in Somayajula *et al*. (2014). A threshold value of 85% quantile (0.44 rad/sec) is used. A GPD fit is performed on the exceedances of the peaks over this threshold. Fig. 6 shows the quantile quantile plot comparison of

the peaks over the threshold versus the fitted distribution. If the data follows the exact distribution, then the quantiles will all be on the same straight line. The deviation from the straight line is a measure of the difference between the data and the fit.
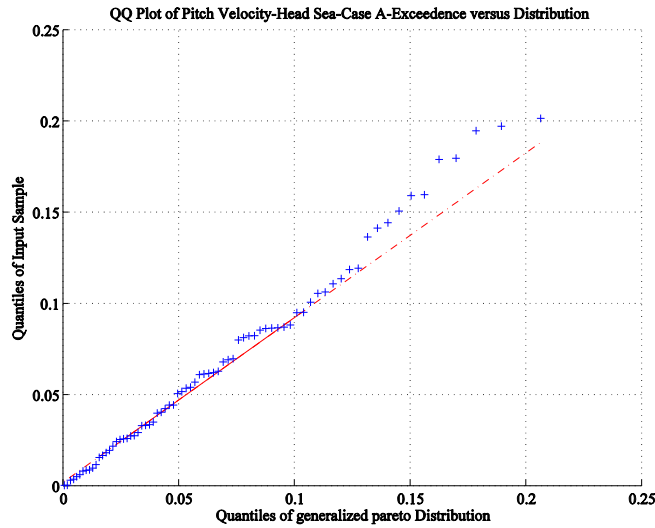

Fig. 6 Quantile-Quantile plot of peaks over the threshold

(a) Rayleigh Distribution
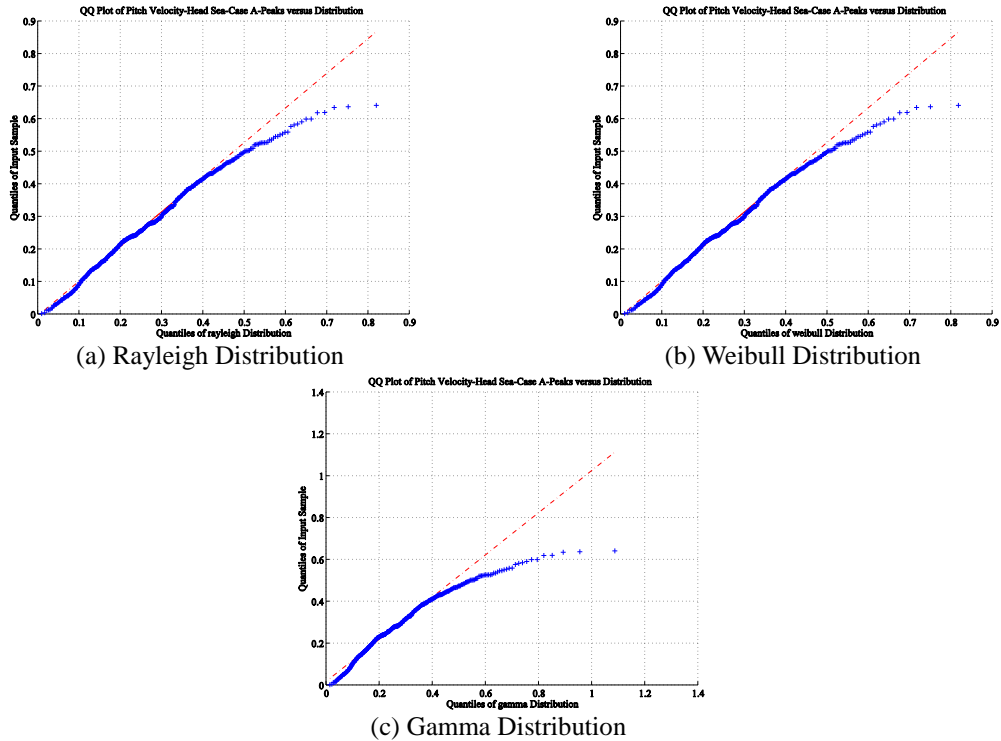
(b) Weibull Distribution

(c) Gamma Distribution

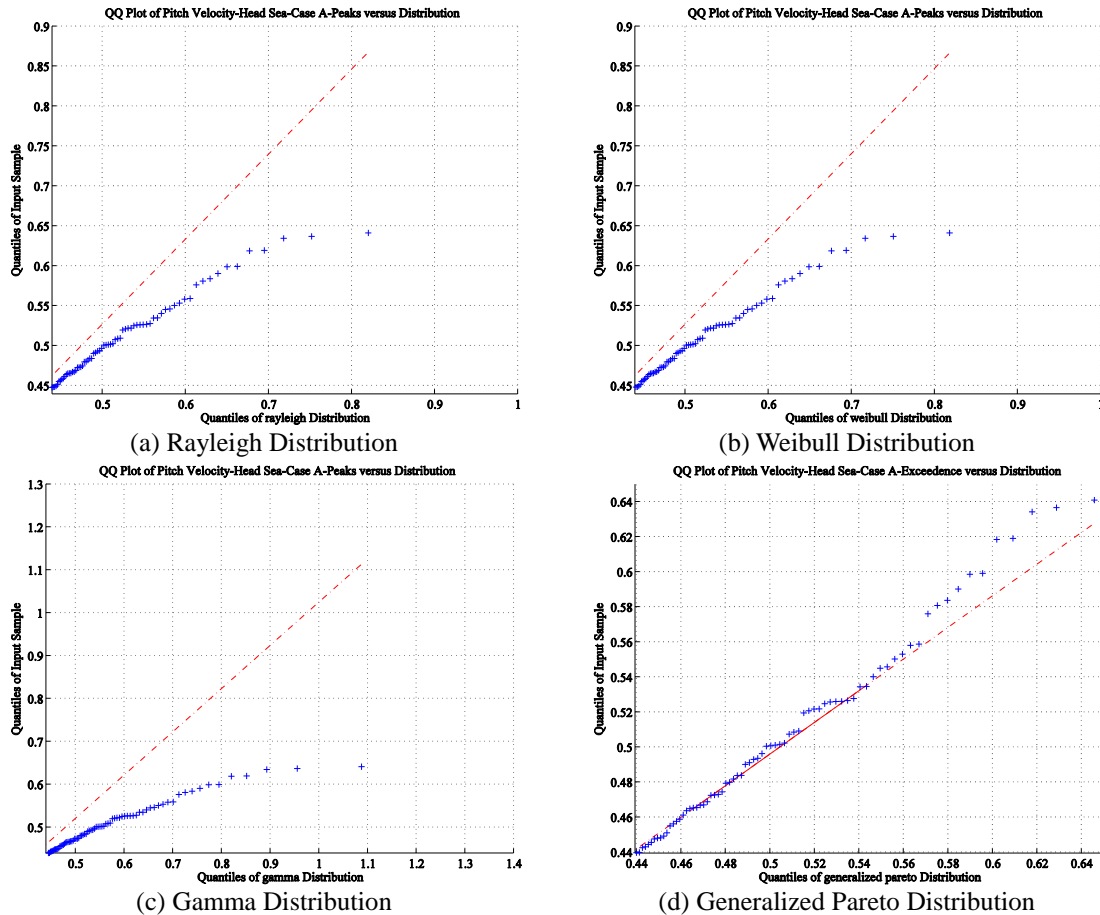Fig. 7 Comparison of peaks of data with other distributions

Fig. 8 Comparison of peaks of data with other distributions

Fig. 7 shows the corresponding quantile-quantile plots of the data peaks compared with these different distributions i.e., Rayleigh, Generalized Gamma and Weibull distributions. It can be seen that the fitting in the tail region for all the three distributions is poor. Note that while Fig. 6 considers the exceedances beyond the threshold of 0.44 radians/sec, Fig. 7 shows the fitting of all the peaks of the data to Rayleigh, Generalized Gamma and Weibull distribution. In order to draw a comparison with the Generalized Pareto fit, only the tail region beyond the threshold of the quantile-quantile plots of all the distributions are shown in Fig. 8. Fig. 8 shows how the Generalized Pareto Distribution gives a better fit than other distributions in the tail regions. Clearly, the extreme value estimated by the other distributions would not be correct.

4.3.4 Cumulative distribution function
The cumulative distribution function of the peaks can be calculated from the conditional generalized pareto distribution. Let X be the random variable denoting the peaks of the time

series shown in Fig. 3, let $F(x)$ be its cumulative distribution function and $G(y)$ be the conditional Generalized Pareto distribution. Then for a threshold value of u the relation between the two distributions would be given by Eq. (9).

$$F(x) = G(x-u)(1 - F(u)) + F(u) \tag{9}$$

Eq. (9) is valid for all $x > u$. $F(u)$ is estimated empirically from the data. For n sample data points the empirical cumulative density function is given by Eq. (10).

$$F(u) = P(X \leq u) = \frac{1}{n}\sum_{i=1}^{n} 1_{X \leq u} \tag{10}$$

4.3.5 Short term extreme value

The short term extreme value x_p is usually calculated for a specified probability of exceedance p. Eq. (11) gives the extreme value x_p as a function of p.

$$1 - F_e(x_p) = p$$
$$F_e(x_p) = F^n(x_p) = 1 - p$$
$$F(x_p) = (1-p)^{\frac{1}{n}} \tag{11}$$
$$x_p = F^{-1}\left((1-p)^{\frac{1}{n}}\right)$$

The calculated extreme pitch angular rate for the time series shown in Fig. 3 for a probability of exceedance of 0.05 was found to be 0.66391 radians/sec. This value is in agreement with the return level plot described in the next section.

4.3.5 Return level plots

In most cases, the fitting of a distribution is represented by its cumulative distribution function. But sometimes, it is more useful to plot the distribution in terms of quantiles or return time between occurrences. Here we are also interested in the return times of a particular short term extreme response. The return time of a particular tension value is defined as the expected time between occurrences of response time series exceeding this particular value.

The unconditional cumulative distribution of the peaks of the data is given by Eq. (12).

$$Pr(X > x) = Pr(X > x \mid X > u).Pr(X > u) = \zeta_u\left(1 - k\left(\frac{x-u}{\sigma}\right)\right)^{-\frac{1}{k}} \tag{12}$$

In Eq. (12), $\zeta_u = Pr(X > u)$ is calculated empirically from the data. Thus a value that is exceeded on an average once every m observations would be given by Eq. (14).

$$Pr(X > x_m) = \frac{1}{m} = \zeta_u \left(1 + k \left(\frac{x - u}{\sigma} \right) \right)^{-\frac{1}{k}} \tag{13}$$

$$x_m = u + \frac{\sigma}{k} \left[(m\zeta_u)^k - 1 \right] \text{for } k > 0$$

$$x_m = u + \sigma \ln(m\zeta_u) \text{ f o } k = 0 \tag{14}$$

As in this case the interest is on the short term extreme value, it is convenient to express the return plots in terms of return times in seconds instead hours. If there are n_y observations in a sec, then N-sec return period corresponds to $m = N \times n_y$ observation return. Thus the N-sec return level is defined by Eq. (15).

$$x_N = u + \sigma \ln(N n_y \zeta_u) \tag{15}$$

Fig. 9 shows the plot of return period versus the return values of the response. The time is plotted in the logarithmic scale. The green line represents the best fit model of GPD for the given data. The red and cyan lines provide the 95% confidence interval over the parameters of the distribution. The blue crosses represent the data points beyond the threshold in the given time series data. The return periods of the data are calculated from the empirical distribution function obtained from the data. It can be seen from Fig. 9 that the best fit model is quite good in agreement with the extremes of the data. Also the extreme value of 0.6639 radians/sec predicted in the previous section is in agreement with the asymptotic value of the best fit return plot.
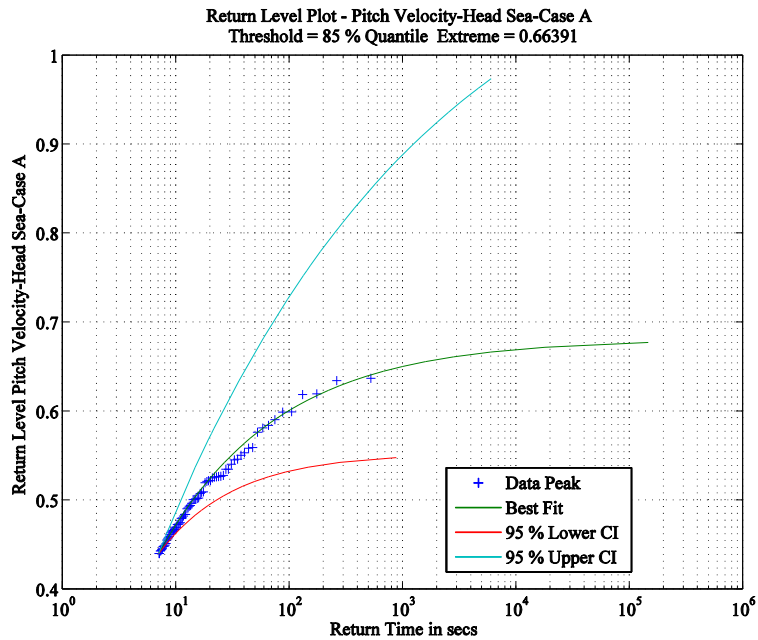

Fig. 9 Return level plot for the pitch time series (using GPD)

5. Conclusions

The analysis of the motion time series of a planing craft was performed. The raw time series was filtered using a 2 Hz low pass Lanczos filter to remove the high frequency noise. Three different tests - QQ plot test, Kolmogorov-Smirnov test and Chi squared test - were performed to test for the Gaussian nature of the time series. All three tests confirmed that the time series was non-Gaussian.

The extreme value analysis of the data was performed using the peaks over threshold approach and fitting a Generalized Pareto Distribution. Goodness of fit of the tail of the data was compared for Generalized Pareto, Rayleigh, Gamma and Wiebull distributions. It was found that the Generalized Pareto distribution performed the best fit to the tail of the data. The Generalized Pareto distribution fit was then used to predict the extreme value based on the probability of exceedance of 0.05. A return level plot was also generated which agreed well with the extreme value predicted. The comparison of the extreme value predicted and the asymptotic value obtained from the return plot are in agreement.

The results of the statistical analysis show that the Generalized Pareto distribution is a good alternative to analyze planing craft motion. The results are also in general agreement with the observations by Fridsma (1969), that the pitch accelerations follow an exponential distribution as the Generalized Pareto distribution includes the exponential distribution as a special case.

Acknowledgments

References

Bradley, J. (2013), *MATLAB Statistical Toolbox User's Guide*, MathWorks.
Coles, S. (2001), *An introduction to statistical modeling of extreme values*,Springer-Verlag, London, UK.
Cramér, H. and Leadbetter, M.R. (1967), *Stationary And Related Stochastic Processes: Sample Function Properties And Their Applications*, Courier Dover Publication, Mineola, New York, USA.
Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Statistical methods for extremal events*, Modelling Extremal Events: For Insurance and Finance. Springer-Verlag, Berlin, 283-370.
Fridsma, G. (1969), *A systematic study of the rough-water performance of planing boats (Irregular waves - part II)*, (No. 1275), Davidson Laboratory, Stevens Institute of Technology, New Jersey.
Guha, A., Somayajula, A. and Falzarano, J.M. (2013a), "Analysis of causeway ferry dynamics for safe operation of improved navy lighterage system", *Proceedings of the 13th International Ship Stability Workshop*, Brest, September.
Guha, A., Somayajula, A. and Falzarano, J.M. (2013b), *Analysis Framework for Improved Navy Lighterage System (INLS) Dynamics*, Texas A&M University, College Station, Texas, USA.
Jamnongpipatkul, A. (2010), *Ship rolling motion subjected to colored noise excitation*, MS Thesis, Texas A&M University, College Station.
Kotz, S. and Nadarajah, S. (2000). *Extreme value distributions: Theory and Applications*, Imperial College Press, London, UK.

Lewis, R.R. (1994), *Statistical analysis of ship hull girder response time histories*, MS Thesis, George Washington University, Washington D.C.

Lewis, R.R. (2005), *A systems-level approach for estimating wave impact design pressures*, PhD. Dissertation, George Washington University, , Washington D.C.

Moideen, H., Somayajula, A. and Falzarano, J.M. (2013), "Parametric roll of high speed ships in regular waves", *Proceedings of ASME 2013 32nd International Conferences on Ocean, Offshore and Arctic Engineering*, ASME.

Moideen, H., Somayajula, A. and Falzarano, J.M. (2014), "Application of volterra series analysis for parametric rolling in irregular seas", *J. Ship Res.*, **58**(2), 1-9.

Mulk, M. and Falzarano, J. (1994), "Complete six-degrees-of-freedom nonlinear ship rolling motion", *J. Offshore Mech. Arct.*, **116**(4), 191-201.

Ochi, M. (1978a), "Generalization of Rayleigh probability distribution and its application", *J. Ship Res.*, **22**(4), 259-265.

Ochi, M. (1978b), *Wave statistics for the design of ships and ocean structures*, *SNAME Annual Meeting*, New York.

Ochi, M. (1981), "Principles of extreme value statistics and their application", *Extreme Loads Response Symposium*, SNAME, Arlington, VA.

Ochi, M. (2005), *Ocean waves: The stochastic approach*, Cambridge university press, Cambridge, UK.

Priestley, M. (1981), *Spectral analysis and time series*, Academic Press, London, UK.

Priestley, M.B. (1988), *Non-linear and non-stationary time series analysis*, Academic Press, San Diego, California, USA.

Rice, S. (1944) "Mathematical analysis of random noise", *Bell Syst.Tech. J.*, **23**(3), 282-332.

Somayajula, A. and Falzarano, J.M. (2014), "Non-linear dynamics of parametric roll of container ship in irregular seas", *Proceedings of the ASME 2014 33rd International Conferences on Ocean, Offshore and Arctic Engineering,* San Francisco, USA, June.

Somayajula, A., Guha, A., Falzarano, J., Chun, H.H. and Jung, K.H. (2014a), "Added resistance and parametric roll prediction as a design criteria for energy efficient ships", *Ocean Syst. Eng.*, **4**(2), 117-136.

Somayajula, A., Guha, A. and Falzarano, J.M. (2014), *Analysis of the planing craft motions data*, Texas A&M University, College Station, Texas, USA.

Su, Z. and Falzarano, J.M. (2011), "Gaussian and non-Gaussian cumulant neglect application to large amplitude rolling in random waves", *Int. Shipbuild. Prog.*, **58**, 97-113.

Su, Z. and Falzarano, J.M. (2013), "Markov and Melnikov based methods for vessel capsizing criteria", *Ocean Eng.*, **64**, 146-152.

Taylor, P., Jonathan, P. and Harland, L. (1997), "Time domain simulation of jack-up dynamics with the extremes of a Gaussian process", *J. Vib. Acoust.*, **119**(4), 624-628.

Torhaug, R. (1996), E*xtreme response of nonlinear ocean structures: Identification of minimal stochastic wave input for time-domain simulation*, Ph.D. Dissertation, Stanford University, Palo Alto.

Tromans, P., Anaturk, A. and Hagemeijer, P. (1991), "A new model for the kinematics of large ocean waves-application as a design wave", *Proceedings of the 1st International Offshore and Polar Engineering Conference*, Edinburgh, August.

Vanem, E., Bitner-Gregersen, E. and Wikle, C. (2013), *Bayesian Hierarchical Space-Time Models with Application to Significant Wave Height, Ocean Engineering & Oceanography*, Springer, Berlin Heidelberg, Germany.

Wang, L. (2001), *Probabilistic analysis of nonlinear wave-induced loads on ships*, Ph.D. Dissertation, Norwegian University of Science and Technology, Trondheim.

Wang, L. and Moan, T. (2004), "Probabilistic analysis of nonlinear wave loads on ships using weibull, generalized gamma, and pareto distributions", *J. Ship Res.*, **48**(3), 202-217.

Winterstein, S., Torhaug, R. and Kumar, S. (1998), "Design seastates for extreme response of jackup structures", *J. Offshore Mech. Arct.*, **120**(2), 103-108.