The shape parameter defines the type of tail: heavy, exponential, or light, as shown in Figure 2[1].

The exponential tail (ξ=0) describes the extreme values of a normal distribution. The heavy tail (ξ>0) is above the exponential tail, while the light tail (ξ<0) is below. As the exponential tail is the smallest infinite tail, the light tail has a limit, which is its right bound. The heavy tail is unbounded.

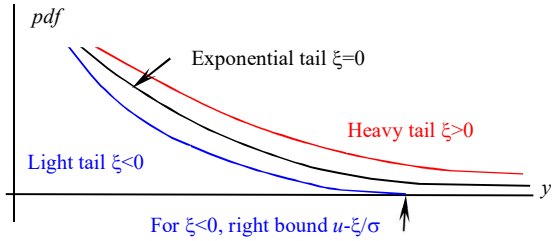

Figure 2: Types of tails (Belenky *et al.* 2018)

3. STRUCTURE OF ROLL TAIL

The roll restoring arm (GZ) curves of most ships have a limited range of stability, leading to the appearance of an unstable equilibria at the angle of vanishing stability, as well a maximum value of GZ. This configuration leads to a heavy tail after the maximum of the GZ curve, which switches to a light tail in the immediate vicinity of the angle of vanishing stability. Figure 3 shows such a distribution, computed for a dynamical system with piecewise linear (PWL) restoring.

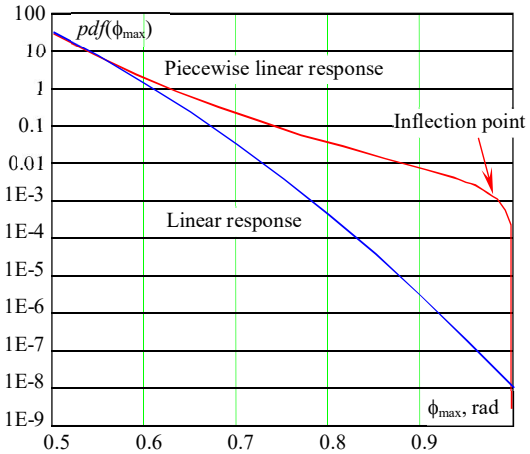

Figure 3: PDFs of peaks of linear response and PWL response (Belenky *et al.*, 2016)

The piecewise linear approximation of the GZ curve allows a closed-form solution for the tail of the distribution of the peaks and instantaneous values of the roll angle (Belenky *et al.* 2016, 2019). Belenky *et al.* (2018) presents an argument for why the piecewise linear result can be generalized for any dynamical system with softening stiffness, including the roll motions of many ships.

The roll angles associated with dynamic stability failures (*e.g.* 50 degrees per paragraph 2.3.1 of the 2008 IS Code, IMO, 2008) are usually located around and beyond the angle of the maximum of the GZ curve. Therefore, to assume a heavy tail appears appropriate for extrapolation problems associated with dynamical stability failures.

4. FITTING HEAVY TAIL FOR ROLL

When the shape parameter ξ>0 (heavy tail) and threshold value $u = \sigma/\xi$, the GPD is equivalent to a Pareto distribution with scale $y_m = \sigma/\xi$ and shape $\alpha = 1/\xi$:

$$\mathrm{pdf}(y) = \frac{\alpha y_m^{\alpha}}{y^{\alpha+1}} \tag{5}$$

The conditional probability of exceedance of a target value y associated with dynamic stability failure is expressed as:

$$\mathrm{P}(Y > y | Y > u) = \left(\frac{u}{y}\right)^{\alpha} = \left(\frac{y}{u}\right)^{-\frac{1}{\xi}} \tag{6}$$

Here, the threshold u does not have to be the same as in the GPD case. A method for finding the threshold and estimating the shape parameter is proposed in Belenky *et al.* (2018), which is based on Beirlant, *et al.* (2004), Dupuis and Victoria-Feser (2006), and Mager (2015).

To extrapolate with equation (6), the threshold is found from applicability considerations so only one parameter needs to be fitted. Decreasing the number of parameters from two (in case if the GPD is used) to one decreases the statistical uncertainity. This is how the physical understanding of the process propagates into a statistical model.

The input data for fitting is designated ϕ_e and consists of N independent peaks extracted from the envelope of the roll time histories (Figure 1). The method is applied to a sample sorted in descending order – a.k.a. order statistics:

$$\mathrm{Y} = \mathrm{sort}_{\mathrm{desc}}(\phi_e) \tag{7}$$

[1] No universally accepted definition of heavy and light tail exists. Other souces may use heavy/light tail in a different context.

The Hill estimator provides the shape parameter ξ:

$$\hat{\xi}_k = \frac{1}{k}\sum_{i=1}^{k} \log\left(\frac{Y_i}{Y_k}\right) \quad (8)$$

where the index k refers to the number of upper order statistics used in the estimation. Mager (2015) suggests the first index $k = \min(40, 0.02N)$, while the last (largest) value for the index taken as $0.2N$.

The threshold u is found by an index that corresponds to a minimum of the mean squares prediction error function (Figure 4):

$$\hat{\Gamma}(k) = \frac{1}{\hat{\xi}_k^{\,2} k^1}\sum_{i=1}^{k} \frac{\left(\log\left(\frac{Y_{i-1}}{Y_{k-1}}\right) + \hat{\xi}_k \log\left(\frac{i}{k+1}\right)\right)^2}{\left(\sum_{j=i}^{k} j^{-2}\right)} + \frac{2}{k}\sum_{i=1}^{k} \frac{\left(\log\left(\frac{i}{k+1}\right)\right)^2}{\left(\sum_{j=i}^{k} j^{-2}\right)} - 1 \quad (9)$$

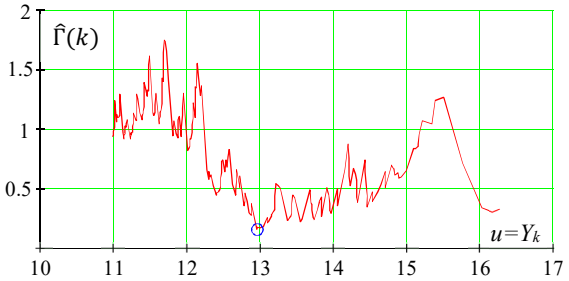

Figure 4: Mean squares prediction error function

Once the index k corresponding to a minimum of $\hat{\Gamma}$, is found, the threshold is set as:

$$u = Y_k \quad (10)$$

The confidence interval for the extrapolated value is computed assuming a normal distribution for the estimate of the shape parameter $\hat{\xi}$. Its variance estimate is expressed as:

$$\widehat{Var}(\hat{\xi}) = \frac{\hat{\xi}^2}{k} \quad (11)$$

The boundries of the confidence interval of the estimate are:

$$\hat{\xi}_{up,low} = \hat{\xi} \pm K_\beta \sqrt{\widehat{Var}(\hat{\xi})} \quad (12)$$

where K_β is a half of a non-dimensional confidence interval computed as a normal quantile of $0.5(1+\beta)$, where β is the confidence probability. For β=0.95, K_β=1.96.

The extrapolated estimate of the exceedance rate of target value c can be computed as:

$$\hat{\lambda}(c) = \hat{\lambda}(u)\left(\frac{c}{u}\right)^{-1/\hat{\xi}} \quad (13)$$

where $\hat{\lambda}(u)$ is the rate of upcrossing of threshold u, estimated through the direct counting procedure with its confidence interval, as described in paragraph 5.4 of Annex 1 of SDC 6 /WP.6 (IMO, 2019).

Boundaries for the extrapolated value are computed through the lower and upper boundaries of the upcrossing rate estimate $\hat{\lambda}_{low,up}(u)$ and the shape parameter estimate $\hat{\xi}_{low,up}$:

$$\begin{cases} \hat{\lambda}_{low}(c) = \hat{\lambda}_{low}(u)\left(\frac{c}{u}\right)^{-1/\hat{\xi}_{low}} \\ \hat{\lambda}_{up}(c) = \hat{\lambda}_{up}(u)\left(\frac{c}{u}\right)^{-1/\hat{\xi}_{up}} \end{cases} \quad (14)$$

Equations (14) contain a product of the boundaries of two estimates. If the desired confidence probability for the entire extrapolated estimate $\hat{\lambda}(c)$ is to be β = 0.95 as recommended in 5.4 of Annex 1 of SDC 6 /WP.6 (IMO, 2019), then the confidence probabilities for each estimate $\hat{\lambda}(u)$ and $\hat{\xi}$ must be set as:

$$\beta_1 = \sqrt{\beta} = \sqrt{0.95} = 0.975 \quad (15)$$

In order to account for the difference in the confidence probability, K_β is set to 2.236 in equation (12) and the confidence 0.975 is used in the direct counting procedure (paragraph 5.4.4 of Annex 1 of SDC 6 /WP.6, IMO, 2019) instead of 0.95.

5. STATISTICAL VALIDATION

A statistical validation of heavy tails was carried out following the recommendations of section 5.6 of Annex 1 of SDC 6 /WP.6 (IMO, 2019).

Per the recommedation in paragraph 5.6.3, a reduced order mathematical model in the form of volume-based 3-DOF calculations was applied as described in Weems *et al.* (2018). This fast code creates very large samples of data in which large roll angles associated with rare failures are observable. The observations estimates a "true value" from direct counting.

A series of validation data sets was computed for the ONR tumblehome configuration (Bishop *et al.* 2005) with KG = 7.5 m, resulting in GM = 2.2 m. Simulations were performed with independent pseudo-random realizations of a seaway by a Bretschneider spectrum with a significant wave height of 9 m and a modal period of 15 seconds. The ship speed was set to 6 knots. Other simulations parameters, including the total simulation time determined the true value, are presented in Table 1.

Section 5.6 of Annex 1 of SDC 6 /WP.6 (IMO,

2019) does not specifiy the target values, which would be the angle associated with dynamic stability failure. This value may be different for different ships, depending on considerations such as the location of the floodable opening. Thus, a number of target angles will be examined.

Table 1: "True value" calculations

Headings Deg.	Total time, hrs	Number of targets	Largest target	Number of exceedances of largest target
15	570,000	5	20	14
22.5	200,000	7	27.5	16
30	200,000	13	45	9
37.5	200,000	15	60	7
45	690,000	15	70	8
60	600,000	15	70	12
90	690,000	9	37.5	12
135	690,000	3	20	6

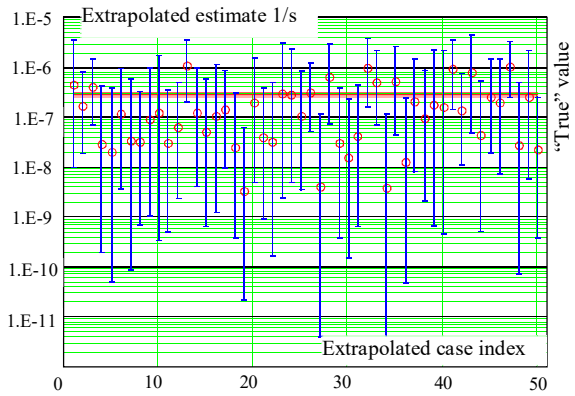

Figure 5: Example of extrapolation validation for a heading of 45 degrees and target value of 45 degrees

The extrapolation procedure was applied to a series of small subsets of this large sample, and the extrapolated estimates were compared with the "true value." Figure 5 shows an example comparison for a 45 degree heading (stern quartering seas) and a target roll value of 45 degrees. Fifty (50) extrapolation estimates occur, each computed from 100 hours of data. The main index of performance is the passing rate, which indicates the percentage of successful extrapolations. An extrapolation is considered successful if the confidence interval of the extrapolated exeedence rate includes the "true value." The example shown in Figure 5 has 45 successful extrapolations, resulting in a passing rate of 90%.

In the three-tiered validation methodology of Smith and Zuzick (2015), the tiers are defined as:

1) all extrapolations for a single target value,

2) extrapolations for all available target values, and

3) extrapolations for all available operational and environmental conditions.

The tier 1 validation is a set of comparisons of extrapolated estimates with the true value. Its example is shown in Figure 5.

The second tier of statistical validation considers all available target angles. The passing rates are shown in Figure 6. An acceptable passing rate for 50 extrapolation data sets is from 0.88 to 1 (Smith, 2019, and paragraph 5.6.7 of Annex 1 of SDC 6 /WP.6, IMO, 2019). This variation of the passing rate can be explained by the natural variability of the statistical estimates. The extrapolations are acceptable for all targets, excluding 50 and 60 degrees, for which the passing rates fell to 0.86. The average passing rate for the 45 degrees heading is 0.90, which is within the acceptable range.

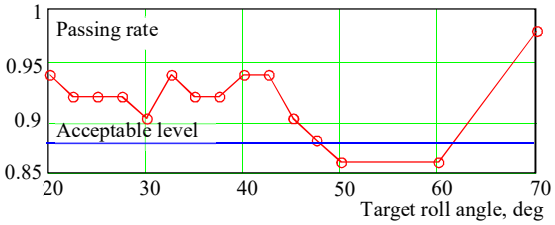

Figure 6: Passing rate for heading of 45 degrees

The third tier of validation assesses the performance over all available conditions. The passing rates are shown in Figure 7. Two lines are shown: one corresponds to an averaged passing rate over all target values, while the other corresponds to the smallest passing rate value encountered among all the target values. For a 45 degree heading, the latter corresponds to a minimum shown in Figure 6. Obviously, the extrapolation did not work for the heading of 135 degrees.

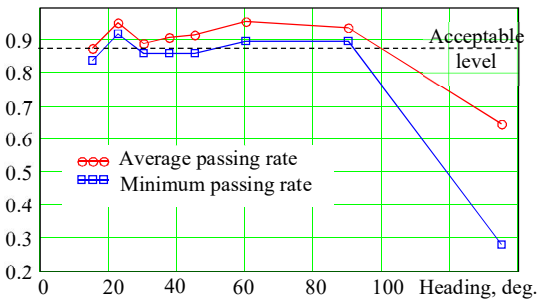

Figure 7: Passing rate for all headings

Belenky *et al.* (2018) looked into the reasons why the validation at the heading of 135 degrees failed. The reason is very likely to be insufficient data in the nonlinear region. That work considers possible indications of sufficient/insufficient data for extrapolation.

Belenky *et al.* (2018) also considered other performance indicators of the EPOT method. The average conservative distance (measure of practical statistical uncertainty) does not exceed an order of magnitude, in terms of the exceedance rate. This performance seems to be sufficient to distinguish between realistic and distant chances of dynamic stability failure.

Overall, the validity of EPOT, except for the 135 degree heading, can be characterized as "almost there." The passing rate falls short of the required 0.88 for a few cases, but not by much. The averaged passing rates stay above 0.88 for all the cases except for the heading of 135 degrees. The decrease of uncertainty in comparison with the two-parameter GPD is substantial and brings EPOT closer to practical application.

6. SUMMARY AND CONCLUSIONS

The paper considers an EPOT extrapolation method for roll motion in the context of direct stability assessment (DSA) and the second-generation IMO intact stability criteria. The main objective was how to make EPOT method a practical tool for DSA.

An envelope of the motion time histories generates a set of independent peak values, and a Generalized Pareto Distribution approximates the tail of the peak distribution above a threshold. The GPD model is completely data-driven, and its statistical uncertainty reflects the volume of the available sample. For rare events and extreme values, the statistical uncertainty of the prediction may be large. To reduce the uncertainty without having to increase the sample volume, physical information can be incorporated into the statistical model.

Based on previous research, the tail of the distribution of roll peaks must be heavy because of the softening nonlinearity of roll stiffness. The application of the tail of the Pareto distribution within the framework of the EPOT method provides an effective physics-informed statistical model.

The paper describes statistical validation carried out following the requirements of the draft interim IMO guidlines on the specification of direct stability assessment procedures.

ACKNOWLEDGEMENT

The work described in this paper has been funded by the Office of Naval Research (ONR) under Dr. Woei-Min Lin and by the NSWCCD Independent Applied Research (IAR) program under Dr. Jack Price. The participation of Prof. Sapsis was facilitated by the NSWCCD Summer Faculty Program, while the participation of Prof. Pipiras was facilitated by NSWCCD Summer Faculty and Sabbatical Programs, both of which are also managed by Dr. Jack Price.

The authors are very grateful for the support that made this work possible.

REFERENCES

Beck, R. F., and A. M. Reed (2001). "Modern Computational Methods for Ships in a Seaway". *SNAME Trans.* Vol. 109, pp. 1–51.

Beirlant, J., Y. Goegebeur, J. Teugels, and J. Segers (2004). *Statistics of Extremes*, Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

Belenky, V., D. Glotzer, V. Pipiras, and T. Sapsis (2016). "On the Tail of Nonlinear Roll Motions". *Proc.* 15th Intl. Ship Stability Workshop, Stockholm (ISSW2016), Sweden, pp. 109-114.

Belenky, V., K. Weems, V. Pipiras, D. Glotzer, and T. Sapsis (2018). "Tail Structure of Roll and Metric of Capsizing in Irregular Waves". *Proc.* 32nd Symp. Naval Hydrodynamics, Hamburg, Germany.

Belenky, V., D. Glotzer, V. Pipiras, and T. Sapsis (2019). "Distribution tail structure and extreme value analysis of constrained piecewise linear oscillators". *Probabilistic Engineering Mechanics* (to appear).

Bishop, R. C., W. Belknap, C. Turner, B. Simon, and J. H. Kim (2005). "Parametric Investigation on the Influence of GM, Roll damping, and Above-Water Form on the Roll Response of Model 5613", Report NSWCCD-50-TR-2005/027.

Campbell, B., V. Belenky, and V. Pipiras (2016). "Application of the Envelope Peaks over Threshold (EPOT) Method for Probabilistic Assessment of Dynamic Stability", *Ocean Engineering,* Vol. 120, pp. 298-304.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

Dupuis, D. J., and M.-P. Victoria-Feser (2006). "A robust prediction error criterion for Pareto modelling of upper tails," *Canadian Journal of Statistics*, Vol. 34, 2006, pp. 639–658.

International Maritime Organization (2008). *International Code on Intact Stability, 2008* ("2008 IS Code").

International Maritime Organization (2019). IMO SDC 6 /WP.6 Finalization of Second Generation Intact Stability Criteria. Report of the Expert's Group on Intact Stability, 7 February 2019.

Mager, J. (2015). "Automatic threshold selection of the peaks over threshold method", Master's Thesis, Technische Universitat Munchen.

Pickands, J. (1975). "Statistical Inference Using Extreme Order Statistics," *The Annals of Statistics*, Vol. 3, No. 1, pp. 119-131.

Reed, A. M., and R. F. Beck (2016). "Advances in the Predictive Capability for Ship Dynamics in Extreme Waves", *SNAME Trans.,* Vol. 124.

Smith, T. C., and A. Zuzick (2015). "Validation of Statistical Extrapolation Methods for Large Motion Prediction," *Proc.* 12th Intl. Conf. on Stability of Ships and Ocean Vehicles (STAB2015), Glasgow, UK, pp. 1127-1140.

Smith, T.C. (2019). "Validation Approach for Statistical Extrapolation", Chapter 34 of *Contemporary Ideas on Ship Stability*, Belenky, V., Neves, M., Spyrou, K., Umeda, N., van Walree, F., eds., Springer, ISBN 978-3-030-00514-6, pp. 573-589.

Weems, K., V. Belenky, and K. Spyrou (2018). "Numerical Simulations for Validating Models of Extreme Ship Motions in Irregular Waves" *Proc.* 32nd Symp. Naval Hydrodynamics, Hamburg, Germany.