On the Validation of Statistical Extrapolation for Stability Failure Rate

Timothy C. Smith,
David Taylor Model Basin (NSWCCD)

Bradley L. Campbell,
David Taylor Model Basin (NSWCCD)

ABSTRACT

Statistical extrapolation is a method to predict extreme, rare events from smaller, more common events. The validation of such methods requires a true value for comparison. Generation of that true value is quite difficult in itself. Due to the random nature of extreme, rare events, the comparison needs to account for the associated uncertainty. This paper examines the requirements for the true value, a methodology for generating acceptance criteria, and demonstrates that methodology with a numerical example.

KEYWORDS

Statistical Extrapolation; Validation;

INTRODUCTION

Probabilistic theory is a very useful tool in the assessment of the dynamic stability of surface ships, especially for unconventional hull forms or modes of operation when previous experience is only partially applicable. The need for such a probabilistic assessment tool is well understood for both naval and commercial ships (Perrault *et al*, 2010, Alman, 2011, Reed, 2011, Peters, *et al*, 2011).

The development of probabilistic methods is a challenging task; the problem was first stated 50 years ago (Sevastianov, 1963. available in English 1994). Only recently did the practical solution come in view. The challenge comes from the physical nature of the problem; to evaluate a risk of stability failure, one needs to consider a rare event in the response of highly nonlinear dynamical system in a non-stationary random environment during the ship's lifetime.

The consideration of a stability failure as a Poisson event provides an explicit relationship between probability and time. It also provides a means for presenting the life-time probability as a series of short-term problems in a stationary random environment (Sevastianov 1963, 1994). Further development included calculations of capsize probability during a voyage spanning several sea states (Boonstra et al 2004, Themelis & Spyrou, 2007).

The evaluation of the short-term probability of a stability failure inevitably led to the extrapolation problem as both model test and advanced ship motion code cannot be run for sufficient time to observe those rare events in realistic conditions. Three approaches to extrapolation were identified in a state-of-the art review on the subject (Belenky, *et al* 2012).

Prior to practical application of any of those methods, their validation is needed. The validation of general computational hydrodynamic tools is a well-established field;

see e.g. (AIAA, 1998, ASME 2009). However, the validation of numerical tools intended for dynamical stability assessments has a lot of important specific details not seen in general applications (Vassalos, *et al*, 1998, Reed 2008, 2011). One aspect is the different physical mechanisms of stability failure that requires qualitative validation considered by Belknap, *et al* (2011, 2012). Another aspect of dynamic stability validation is statistical as the problem is considered in irregular waves (Smith, 2011, 2012).

Extrapolation tools needed for probabilistic assessment of dynamic stability, in principle, are independent of the simulation tools. That is why Peters, *et al* (2012) considered the validation of extrapolation tools as a separate problem. The general methodology of the validation of an extrapolation tool is the main focus of this paper.

CONCEPT AND TERMINOLOGY

A statistical extrapolation method uses the statistical properties of a data set to predict the probability of events that are too rare to be observed during a model test or reproduced by numerical simulation of reasonable duration. A common approach is fitting a Weibull or Rayleigh distribution to a data set. Both of these are forms of the General Extreme Value Theory (GEVT).

Probability and time are related. Theoretically, if time is long enough (infinite) a rare event surely will occur (probability equals 1). The usual way to relate time and probability is the Poisson assumption, i.e. considering the rare event to follow Poisson flow. Then the only parameter to find is the event rate λ *i.e.* probability of event per unit of time *T*, as the probability of at least one stability failure during time T is expressed as:

$$P(T) = 1 - \exp(-\lambda T) \qquad (1)$$

The extrapolation methods being considered for dynamic stability are based on the Separation Principle (Belenky, *at al* 2012). They use intermediate thresholds that are low enough so their crossings can be observed and sufficient statistical data is available.

A stability failure (partial or total) can usually be associated with the exceedance of a certain level of roll angle. Then the event rate λ in the Equation (1) becomes a crossing rate (or mean crossing rate). The extrapolated mean crossing rate is evaluated at various levels associated with stability failure - levels of interest.

Validation requires a comparison between the extrapolated and true value. However, the true value generally is not available. Thus it has to be substituted by an estimate. This estimate must be evaluated by a commonly accepted method, such as direct counting.

The direct counting is a procedure of statistical estimation of the mean crossing rate. It was studied in details in (Belenky & Campbell, 2011) and compared favourably with theoretical results available from upcrossing theory (Kramer & Leadbetter, 1967). A brief description of this procedure is available later in the paper.

The exposure time, or duration, is the length of the data set time history. Indeed with a very long exposure times, 100,000 or more hours, it is possible that the crossing of the level of interest may actually occur. When the exposure time is millions of hours, the direct counting answer approaches the actual true value and is used as a surrogate for “true value” for comparison. The actual true value is the direct counting answer for infinite time.

A condition is defined as the combination of independent parameters that result in a unique set. Typically, a condition is the environmental parameters, speed and heading used to make the simulation of a particular motion or acceleration, and the particular motion or acceleration. So a set of environmental parameters, speed and heading and three motions (or combination of motions and accelerations) would be three conditions due to the three motions. Thus, condition can be defined as a deterministic vector:

$$\vec{S} = \left(H_S, T_m, V_S, \beta, i_{dx}\right) \qquad (2)$$

where H_S is a significant wave height, T_m modal frequency, V_S, forward speed, β-heading, i_{dx} –motion index (say, i_{dx}=4 corresponds to roll).

A validation comparison is made for each condition. First the "true" value is to be estimated from a large data set (we discuss its generation later in the paper). The true-value validation set is defined as

$$\begin{aligned} &W(\vec{S}) := \{N_C \geq N_{ST}\} \\ &\downarrow \mathit{Estimation} \\ &\hat{\lambda}^T(\vec{S}, P_\beta) := \left[\lambda^T_{low}; \lambda^T_{up}\right] \end{aligned} \qquad (3)$$

Where N_C is the number of stability failures associated with crossing of the level of interest; N_{ST} is the number of events considered to be statistically significant; and $\hat{\lambda}^T$ is the "true" value substituted by its estimate with confidence interval $\left[\lambda^T_{low}; \lambda^T_{up}\right]$ with the confidence probability P_β. Then a subset is used as the data set for extrapolation and the result of extrapolation is compared with the true value.

$$\begin{aligned} &U(\vec{S}) := \{N_C < N_{ST}\} \subseteq W(\vec{S}) \\ &\downarrow \mathit{Extrapolation} \\ &\hat{\lambda}^E(\vec{S}, P_\beta) = \left[\lambda^E_{low}; \lambda^E_{up}\right] \end{aligned} \qquad (4)$$

where superscript E stands for the extrapolated estimate. As the simulations produce a random process, multiple extrapolations N_{EX} should be made for each condition to account for that randomness.

$$U_i(\vec{S}) \quad : \quad i = 1, \ldots N_{EX} \qquad (5)$$

VALIDATION TECHNOLOGY

As it was stated above the extrapolation method is considered valid if, a probability (rate of events) of large excursion, for example exceedances of 50 degrees of roll, can be predicted from the time series that does not contain such events.

As such, the validation data set representing the true value for comparison simply needs to have stability failures in statistically representative quantity. The actual duration of the validation data set is immaterial; the validation data set merely provides a mean crossing rate for comparison. It is possible to generate the validation data set as a long duration data set with an expected large event rarity using typical conditions or as a shorter duration data set with more frequent large events in an artificially severe condition.

For a true-value validation data set to represent realistic rarity of a stability failure in a storm a long duration is on the order of millions of hours and short duration is on the order of 100,000 - 500,000 hours full scale. In this case, long and short duration are relative considering most simulations are on the order of a half hour to three hours full scale. Experimental data for such long durations is impractical. High fidelity simulations are similarly impractical. There is a need for a simulation tool that is computationally fast and contains enough of the physics to be representative of the higher fidelity simulations.

For statistically extrapolating the ship motions used in this study, a fully coupled 3 degree of freedom simulation tool, based on volume calculation was used (Weems & Wundrow, 2013). It is a simple time domain, strip theory simulation using body exact hydrostatics and Froude-Krylov forces. The radiation and diffraction coefficients are considered constant. The exponential pressure decay of the Froude-Krylov forces is ignored. The ship is free to heave, roll, and pitch. The ship is considered on constant speed and heading with respect to the waves. The motions tend to be larger than higher fidelity simulation tools because there is no decay of wave energy and the reduced degrees of freedom include energy that would ordinarily be shed to the restrained degrees of freedom. For instance, roll tends to be higher with 3DOF models than 6DOF models as wave energy that would appear as sway and yaw motion can only be accounted for as more roll. However, all these simplifications lead to calculation efficiency sufficient to reproduce true-value validation data set of long duration.

The term “data reduction factor” (DRF) was introduced in (Belenky & Campbell 2011) to measure efficiency of an extrapolation method. The data reduction factor is simply the inverse of the mean crossing rate of the largest level of interest satisfactorily extrapolated divided by the duration used to make the extrapolation. The key is it depends on the mean crossing rate of the largest acceptable level of interest not the duration of true value validation data set. Also, DRF depends upon the data sets used and acceptance criteria. Thus, DRF cannot provide a universal answer as to how much exposure time is required for validation.

ACCEPTANCE CRITERIA

Acceptance criteria should reflect the Specific Intended Uses (SIU) of the extrapolation. In this case, the SIU could be relatively simple and general – determine the mean crossing rate (the upper boundary of confidence interval is often used in practical cases) at a specified level of interest based on specified exposure time for a given motion or acceleration, speed, heading and environmental conditions. This leads to two approaches: treating each level of interest independently or treating the entire extrapolation over multiple levels of interest as a unit. These two approaches have common elements: acceptable difference between mean crossing rates, percentage of passing extrapolations, and overall acceptance. This paper only addresses the first approach as the most straightforward, treating each level of interest independently.

Acceptance also depends on the data set used to make the extrapolation. As a result, multiple extrapolations from multiple data sets, N_X=100 or more, should be used to evaluate acceptance.

Following the multi-tier approach from Smith (2012) and considering levels of interest independently, a comparison at a single level of interest would be parameter; multiple levels of interest would be conditions; and multiple extrapolations for different data sets would be the overall or set tier. The use of multiple extrapolations is specified in the first tier.

At a single level of interest, the extrapolation would be compared to the “true value” at that level of interest. The comparison could be made based on confidence interval overlap, a more rigorous statistical test (Smith, 2011), or even simply the distance between mean crossing rates. For the confidence interval overlap, one can consider a random variable x:

$$x\left(P_{\beta}, U_i(\vec{S}), \phi_j\right) = \begin{cases} 1 & if \quad \left[\lambda^E_{low};\lambda^E_{up}\right] \cap \\ & \quad \left[\lambda^T_{low};\lambda^T_{up}\right] \neq \varnothing \\ 0 & if \quad \left[\lambda^E_{low};\lambda^E_{up}\right] \cap \\ & \quad \left[\lambda^T_{low};\lambda^T_{up}\right] = \varnothing \end{cases} \qquad (6)$$

where ϕ_j is the j-th level of interest with symbolic logic notation.

Specifying an allowable distance between mean crossing rates requires knowing the effect of the distance on the final use of the extrapolation. A sensitivity study on the final use of the extrapolation can help set these values. Also, the allowable distance should not be less than the “true value” uncertainty (where the “true value” uncertainty is that due to random process error). Using an allowable distance less than this could conceivably not accept another realization of the "true value." This is indicative of overly strict acceptance criteria.

With multiple extrapolations, the percentage of passing extrapolations needs to be specified to pass a level of interest. Even with multiple extrapolations, this is still a random process dependent on the data sets:

$$C1\left(P_\beta, \vec{S}, \phi_j\right) = \frac{1}{N_{EX}} \sum_{i=1}^{N_{EX}} x\left(P_\beta, U_i(\vec{S}), \phi_j\right) \quad (7)$$

Equation (7) represents the tier 1 criterion, as it reflects how good the extrapolation is. If the extrapolation method is ideal and it always recovers the true value, the criterion C1 will equal 1. Real-world extrapolation methods may miss the true value just because they use statistical estimates that are random variables. Thus the acceptance condition is formulated as:

$$C1\left(P_\beta, \vec{S}, \phi_j\right) \geq B1 \quad (9)$$

where B1 is a boundary (or standard) for the acceptance at the tier 1.

The boundary, $B1$, needs to be set a level the accounts for the uncertainty. For extrapolations carried out with the confidence interval P_β, it is not reasonable to set $B1$ to be more than P_β. Otherwise, the acceptance may not be reached purely because of natural statistical uncertainty of the random variable x, that has nothing to do with validity of the extrapolation method, so

$$B1 < P_\beta \quad (10)$$

However, statistical uncertainty is not the only imperfection of an extrapolation method intended for nonlinear dynamical system under random excitation. Inevitably other assumptions are made. Some of them may be related to use of limit distributions, like Weibull for extreme values or even normal for estimates. Then the result becomes dependent on how quickly the actual distribution converges to its limit, which is not always known. Other assumptions may involve dynamics; such as a response to wave group made out of sinusoidal waves represents the response to a real-world wave group. Thus, it makes sense to set the standard lower than the confidence probability, for example:

$$B1 = 0.9 \quad if \quad P_\beta = 0.95 \quad (11)$$

The averaging in Equation (7) also brings additional statistical uncertainty that can be dealt with by calculation of confidence interval for the estimate of $C1$. This can be done using the binomial distribution for the number of extrapolations, as equation (6) can be considered as a Bernoulli trial:

$$\begin{aligned} C1_{low} &= \frac{Q_B\left(0.5\left(1-P_\beta\right), B1, N_{EX}, p\right)}{N_{EX}} \\ C1_{up} &= \frac{Q_B\left(0.5\left(1+P_\beta\right), B1, N_{EX}, p\right)}{N_{EX}} \end{aligned} \quad (12)$$

Where Q_B is the quantile function or inverse of binomial cumulative distribution function (CDF) for N_{EX} Bernoulli trials with the parameter p:

$$p = C1\left(P_\beta, \vec{S}, \phi_j\right) \quad (13)$$

For example, for 100 data sets with 90% probability and taking the 95% confidence, the lower quantile is 84, or 84%. B1 needs to be adjusted in a similar fashion. Finally the tier 1 acceptance condition is written as:

$$C1_{low}\left(P_\beta, \vec{S}, \phi_j\right) \geq B1 \quad (14)$$

The second tier of the acceptance criteria is the number of levels of interest that need to pass for acceptance. Using a single level of interest is certainly simplest but allows for an extrapolation be unacceptable at all other levels of interest and still be acceptable. However, using all the levels of interest may give to much weight to levels of interest that are unimportant.

With multiple levels of interest, it is also necessary to specify how they are to be consolidated to a single metric and acceptance value. The straightforward approach is specifying an average passing value, minimum passing value, or percentage of passing level of interest in the range of N_{LVL} levels of interest, making the second tier criterion $C2$ and the acceptance condition with the standard B2:

$$C2\left(P_\beta,\vec{S}\right)=\frac{1}{N_{LVL}}\sum_{j=1}^{N_{LVL}}C1\left(P_\beta,\vec{S},\phi_j\right) \tag{15}$$

$$C2\left(P_\beta,\vec{S}\right)\geq B2 \tag{16}$$

For instance, this can be stated for multiple extrapolations and a range of level of interest, the average passing rate from tier 1 needs to be above 90% (*i.e.* $B2$=0.9). This value should not be set so high as to overly constrain the acceptance criteria (Smith, 2012).

The third tier, overall acceptance, deals with how many conditions need to pass for overall acceptance of an extrapolation method:

$$C3\left(P_\beta\right)=\frac{1}{N_{CND}}\sum_{k=1}^{N_{CND}}C2\left(P_\beta,\vec{S}_k\right) \tag{17}$$

To increase confidence in the extrapolation method more than one condition should be examined, so the average passing rate should not be less than B3

$$C3\left(P_\beta\right)\geq B3 \tag{18}$$

Interestingly, the actual environmental conditions used do not matter as many different environmental conditions will produce similar levels of rarity (order of magnitude for the rate of failures). The key is that the extrapolation method is valid for a wide range of rarities or low probability tail behaviours. So the acceptance criteria need to specify simulations that cover the different tail behaviours rather than operational conditions. There may be a correlation between tail behaviours and operational conditions but that is not guaranteed. Thus, all conditions or cases would be required to pass for overall acceptance as the total number of cases can be relatively small.

NUMERICAL EXAMPLE

The following Numerical example shows some elements of validation of extrapolation method using EPOT (Envelope Peak over Threshold, see Belenky & Campbell 2011) as an example. EPOT is based on known Peak over Threshold method (see e.g. Coles, 2001) using a peak envelope to control clustering. EPOT requires relatively large amount of data, but, in principle, can work with any data source, including model test or full-scale trials.

The "True-value" Estimate

The true-value validation set was produced with the fast volume based calculation (Weems & Wundrow, 2013). The sample ship was ONR tumblehome top configuration (Bishop, *et al* 2005). Parameters of the simulations are given in Table 1

Table 1: Condition parameters used to generate "true value" data set.

Parameter	value
Significant wave height, m	9.5
Modal period, s	15
Heading, deg	45
Speed, kn	6
Total duration, hrs	10^6
Duration of one record, min	30

The "true" value was estimated as follows. A data set is represented with N_R records of roll time history, each of which contains N_t data points with time step Δt. The mean crossing rate is estimated as:

$$\hat{\xi}=\frac{N_C}{N_t N_R \Delta t} \tag{19}$$

where N_C is the total number of observed upcrossings of the given level. To apply the assumption of Poisson flow, the upcrossing events must be independent. As a result the number of crossings has binomial distribution with parameter p estimated as:

$$\hat{p} = \frac{N_C}{N_t N_R} \tag{20}$$

The difference between two estimates is just the time step, i.e. a constant coefficient, thus the estimate of mean crossing rate also has a binomial distribution.

In principle, binomial distribution can be used to evaluate the uncertainty of the estimate (19); however, there may be numerical difficulties with calculation of the quantile of the binomial distribution. Thus, a normal distribution can be used instead, with the variance calculated as:

$$V\hat{a}r(\hat{\xi}) = \frac{1}{N_t N_R \Delta t^2} \hat{p}(1-\hat{p}) \tag{21}$$

Figure 1 shows the estimates of mean crossing rate to be used as “true-values.” The mean crossing rate decreases as the level increases, reflecting the rarity of large roll events. The confidence interval also increases as the number of events decreases. This also illustrates that the “true value” is really just an estimate of the actual true value. Interestingly, there is a definite knee in the curve around 30 degrees of roll. This reflects both the non-linear righting arm curve for the ONR tumblehome hull form and a change in response from resonant rolling to loss of stability. The extrapolation should be able to predict motions below the knee from data above the knee to be of practical use.

Validation for One Condition

The validation was carried out by evaluating a single level of interest and a single condition, tiers 1 and 2 respectively, using 10 extrapolations. Tier 3 is redundant to tier 2 as there is only one condition. Each extrapolation was based on 100 hours of simulation.

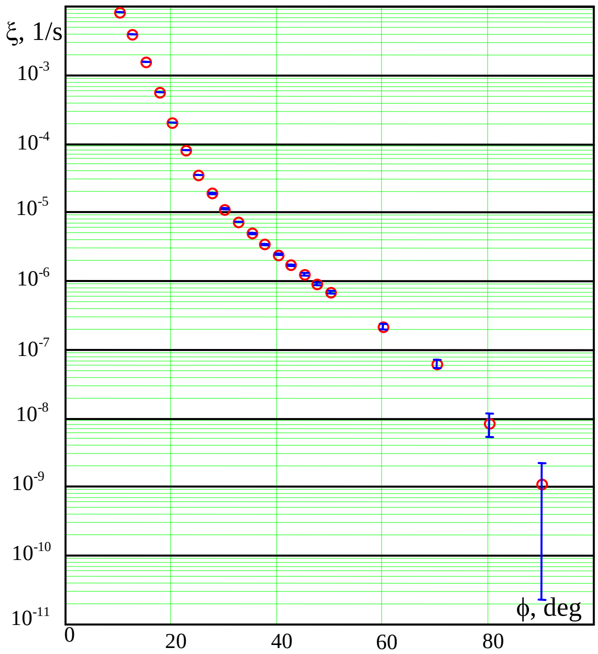

Fig. 1: The “true” values estimates of mean crossing rate (1/sec) by roll angle level (deg)

Tier 1 of the acceptance criteria specifies what is required to accept a single level of interest. In this case, that is the method of comparison and what percentage of the multiple extrapolations needs to be acceptable to consider that level of interest acceptable. The method of comparison is confidence interval overlap. If the confidence intervals of the “true value” and extrapolation overlap, the extrapolation is considered to be the same as the “true value.”

The percentage needed to be acceptable is taken as the lower binomial 95% quantile for 10 trials of overlapping a 95% confidence interval. This value is 8 of 10 or 80%.

Tier 2 defines the number of levels of interest to consider how they are consolidated, and an acceptance metric. In this case, only a single level of interest, 50 degrees of roll, was considered. The levels of interest will be consolidated using the minimum passing percentage set the same as the tier 1 passing percentage, 80%. The use of a single level of interest makes for a simple case, though the application to multiple levels of interest is straightforward.

Figure 2 shows the 10 extrapolations with confidence intervals compared to the “true value” evaluated at 50 degrees of roll. All 10 extrapolations overlap the “true value,” denoted by the horizontal line. As 10 is greater than or equal 8 (the passing value) this level of interest is considered acceptable at tier 1.

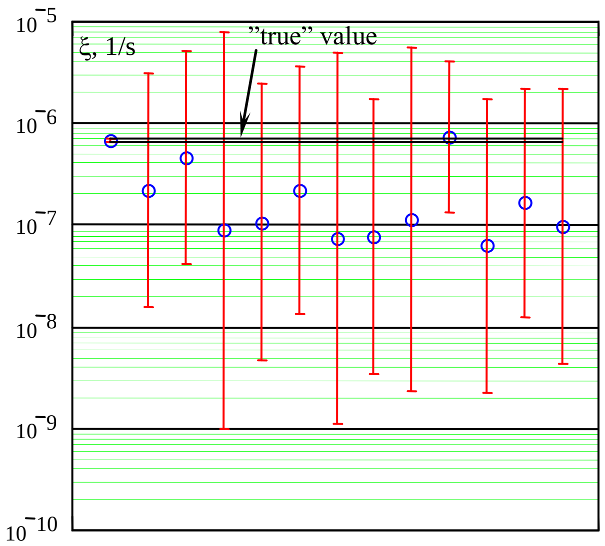

Fig. 2: Results of validation using 10 extrapolations. The level of interest 50 degrees

Proceeding to tier 2, there is a single level of interest with a 100% pass rate. The pass rate, 100%, is greater or equal to the minimum value, 80%. This passes tier 2; this condition has an acceptable extrapolation.

With only one condition, which passed, tier 3 is passed automatically. With more conditions, there would be a passing condition percentage to exceed.

FUTURE WORK

This paper only covered one acceptance criteria approach in a limited fashion. To further understand the acceptance criteria, more extrapolations, more levels of interest, and more conditions need to be considered. Other approaches to formulating the acceptance criteria tiers or making comparisons should be examined to assess the acceptance criteria’s applicability for general use.

Student’s t- Test

The Student's t-test could be used to compare mean crossing rate values, rather than confidence interval overlap. The Student's t-test requires a mean, a variance, and the number of independent data points for the extrapolation and the direct counting answer. These values for the direct counting answer are easily obtained assuming a binomial distribution and using the number of peaks at the level of interest as the number of independent degrees of freedom. For the extrapolation, the mean is the mean crossing rate, the variance can be determined from the confidence interval. The number of independent degrees of freedom is the mean crossing rate multiplied by the duration used for the “true value” approximation.

CONCLUSIONS

This paper discusses validation of statistical extrapolation tools by providing a numerical example using the EPOT extrapolation method. This example reiterates a three tier approach to formulating acceptance criteria and specifies the required elements to have a complete acceptance criteria definition. The numerical example also included the method to determine a “true value” estimate with confidence interval. This indicated the need for computationally fast, reduced order simulation to generate the needed data set. The numerical example showed EPOT was able to predict 50 degrees of roll using 100 hour extrapolations.

ACKNOWLEDGEMENTS

The work described in this paper has been funded by the Office of Naval Research under

Dr. Patrick Purtell and Dr. Ki-Han Kim. This support is gratefully acknowledged by the authors.

The authors are grateful to Dr. V. Belenky for providing the data and analysis tools for the numerical example and thorough editing of the manuscript.

REFERENCES

American Institute for Aeronautics and Astronautics. 1998. *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*. Reston, Va.: American Institute for Aeronautics and Astronautics.

Alman, P.R. (2011) "Thoughts on Integrating Stability into Risk Based Methods for Naval Ship Design", *Proc. of 12th Intl. Ship Stability Workshop*, Washington DC, USA, pp 359-369.

ASME, 2009, "Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer," American Society of Mechanical Engineers, New York.

Belenky, V. & B. Campbell (2011). Evaluation of the Exceedance Rate of a Stationary Stochastic Process by Statistical Extrapolation Using the Envelope Peaks over Threshold (EPOT) Method. Naval Surface Warfare Center Carderock Division Report NSWCCD-50-TR-2011/032, 216 p.

Belenky, V., Weems, K.M., Bassler, C.C., Dipper, M.J., Campbell, B. & K. Spyrou (2012) "Approaches to Rare Events in Stochastic Dynamics of Ships", *Probabilistic Engineering Mechanics* 28 30–38.

Belknap, W.F., Smith, T.C., & Campbell, B.L. (2011), "Addressing Challenges in the Validation of Dynamic Stability Simulation Tools," *Proc. of the 12th Intl. Ship Stability Workshop*, Washington, DC. pp 81-90.

Belknap, W, Belenky, V., Campbell, B. and T. Smith (2012) "A Guide for Validation of Simulation-Based Approach for Predicting Ship Stability Failure Probabilities" *Proc. 29th Symp. Naval Hydrodynamics*, Gothenburg, Sweden

Boonstra, H., de Jongh, M.P., & L. Palazzi (2004) "Safety Assessment of Small Container Feeders", *Proc. of PRADS 2004*, Lübeck-Travemünde, Germany

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

Kramer, H., and Leadbetter, M.R., *Stationary and Related Stochastic Processes*, John Wiley, New York, 1967.

Perrault, D., Hughes, T. & S. Marshall (2010) "Developing a Shared Vision for Naval Stability Assessment", *Proc. of 11th Intl. Ship Stability Workshop*, MARIN, Wageningen, the Netherlands, pp. 115-120.

Peters, W., Belenky, V., Bassler C., Spyrou, K., Umeda, N., Bulian, G. & B. Altmayer (2011) "The Second Generation of Intact Stability Criteria an Overview of Development", *SNAME Trans.* Vol. 119**.**

Peters, W., Belenky, V. Reed, A. (2012) "On Regulatory Framework of Direct Stability Assessment", *Proc. of the 11th Int. Conference on Stability of Ships and Ocean Vehicles*, Athens, Greece.

Reed, A. M. (2008) Discussion of: Belenky, V., J. O. de Kat & N. Umeda (2008) "Towards Performance-Based Criteria for Intact Stability." Marine Tech., 45(2):122–123.

Reed, A. M. (2011) "A Naval Perspective on Ship Stability" in *Contemporary Ideas on Ship Stability and Capsizing in Waves* , Neves, et al, editors, Springer, pp.793-826

Sevastianov, N. B. (1963) "On probabilistic approach to stability standards" *Transactions of Kaliningrad Institute of Technology*, Vol. 18, Kaliningrad, pp. 3-12 (in Russian).

Sevastianov, N. B. (1994) "An algorithm of probabilistic stability assessment and standards" *Proc. of STAB'94: 5th International Conference on Stability of Ships and Ocean Vehicles*, Vol. 5, Melbourne, Florida.

Smith, T.C. (2011) "Statistical Data Set Comparison for Continuous, Dependent Data" *Proc. of 12th Intl. Ship Stability Workshop*, Washington DC, USA, pp 75-80.

Smith, T. C. (2012) "Approaches to Ship Motion Simulation Acceptance Criteria" *Proc. of 11th Intl Conference of the Stability of Ships and Ocean Vehicles*, Athens, Greece pp 101-114.

Themelis, N., & K.J. Spyrou (2007) "Probabilistic assessment of ship stability," *Trans*," SNAME, Vol. 115, pp. 181-206.

Vassalos, D., M. Hamamoto, J. O. de Kat, D. Molyneux & A. Papanikolaou (1998), "The State of the Art in Modeling Ship Stability in Waves," *Proc. 25th ATTC*, , Iowa City, Iowa. 8 p.

Weems, K. & D. Wundrow (2013) "Hybrid Models for Fast Time-Domain Simulation of Stability Failures in Irregular Waves with Volume-Based Calculations for Froude-Krylov and Hydrostatic Force", *Proc. 13th Intl. Ship Stability Workshop*, Brest, France