Example of Validation of Statistical Extrapolation

Timothy Smith [1*]

1. *David Taylor Model Basin (NSWCCD) USA*

Abstract: Statistical extrapolation is a method to predict extreme, rare events from smaller, more common events using relatively short duration data sets. The validation of such methods requires a multi-tier validation approach consistent with the true value data structure. This paper provides a full demonstration of the multi-tier validation approach using roll ship motion with the Generalized Pareto Distribution.

Key words: statistical extrapolation, validation, ship motion

1. Introduction

The validation of numerical simulations is a large field of interest and spans many engineering disciplines. Various professional societies and governmental bodies have outlined verification and validation processes to assist their members [1,2,3]. These processes are often generalized with details left to the engineers actually performing the verification and validation. However, in all cases some comparison is made between the simulation and the "true value," and becomes the basis for a validation decision. The true value comes from scale model testing or higher fidelity simulations.

There are some phenomena, such as large rolling or capsizing, that are at once non-linear and rare. The simulation of these phenomena requires advance, hydrodynamic blended method prediction tools due to the non-linearity involved. The ITTC parametric roll study [4] showed the uncertainty can be quite large due to practical non-ergodicity.

This paper continues Smith and Campbell [5] by providing a worked example to demonstrate the multi-tiered validation approach.

2. **Test Case**

This test case considers ship roll motion for a range of relative wave heading in a high sea state. Extrapolations are made based on a sub-set of time history data and compared to a directly counted true value at a motion level not necessarily seen in the data set.

2.1 Extrapolation Method

The Generalized Pareto Distribution (GPD) can be used to approximate a tail of any distribution that makes use of a scale and shape parameter to fit the data. There are various implementation details in terms of selecting a threshold and determining the scale and shape parameter. This paper uses the GPD as implemented in [6] as the extrapolation method.

The confidence intervals for the extrapolated estimate were calculated using assumption of the normal distribution of the GPD parameters. This follows the confidence interval method from [6], except the logarithm of the scale parameter was used instead of the scale parameters itself. The use of the logarithm of the scale parameter ensures its positive value.

2.2 True Value

The true value was determined by calculating hundreds of thousands of hours of ship motion simulation using a fully coupled, 3 degree of freedom simulation tool based on volume calculation [7]. This model assumes constant radiation and diffraction

---

*** Corresponding author:** Timothy Smith, research field: dynamic stability and seakeeping, E-mail timothy.c.smith1@navy.mil

The work described in this paper has been funded by the Office of Naval Research (ONR) under Dr. Ki-Han Kim.

forces with non-linear hydrostatics on 2D strip hull representation. As such it captures essential hydrostatic non-linearity and maintains very fast computation time.

The peaks were extracted using envelope approach [8]. This method ensures independent data samples as required to apply GPD.

The true value of the exceedance rate is found using a direct counting procedure studied in detail in [8]. It compared favourably with theoretical results available from upcrossing theory [9].

3. Validation Approach

This example uses a multi-tier validation approach consisting of a parameter, condition and set criteria [5, 10]. This reflects typical scale model data structures of individual motion channels, a run condition of speed-heading-seaway, and a test consisting of many conditions. Criteria are set to determine an acceptable parameter comparison, and what constitutes an acceptable condition and set.

A parameter comparison is the elemental comparison between the simulation and true value. It refers to a single motion or response. Typically, a condition is the environmental parameters, speed and heading used to define the simulation and the motion response. So a set of environmental parameters, speed and heading and two motions would be two conditions due to the two motions. Thus, condition can be defined as a deterministic vector:

$$\vec{S} = \left(H_S, T_m, V_S, \beta, i_{dx}\right) \tag{1}$$

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). Repeating multiple parameter comparisons for the same condition makes the second tier.

The third tier, the set comparison, defines how many conditions have to pass for the simulation to pass the validation criteria.

For this example, the parameter comparison is the comparison of a statistical extrapolation to the true value at a specified critical motion level. The parameter comparison passes the test if the extrapolation confidence interval captures the true value. Multiple extrapolations are made from different data sets all representing the same condition, that is speed-heading-seaway-motion combination. That condition passes if the true value is captured by the confidence interval at a percentage roughly equal to the confidence level. This is repeated for a number of different conditions. The extrapolation method is considered valid if a high percentage of conditions pass.

Other parameter comparisons besides confidence interval capture to the true value may be used depending on what is important to the application. For instance, the amount of conservatism or absolute distance may be used as a metric. A change of the parameter comparison could change the condition criteria. The multi-tier validation definition used in this study provides a check on both the extrapolation and the confidence interval formulation as both are included in the parameter comparison.

It is also possible to consider the response independently of the environmental conditions in Tier II, the condition level. Then the number of passing responses becomes a criterion to condition passing. This is essentially a bookkeeping issue, but may be conceptually more appealing to some in formulating acceptance criteria.

4. Results

One hundred extrapolations based on different 100 hours of simulation were made using GPD. They were compared to the true value at an evaluation level corresponding to a high motion level in the true data set. The comparison was based on overlap of the 95% confidence interval with the true value.

The evaluation level was selected as the highest level in the true data set that had more than 30 data samples. Thirty samples are enough to have meaningful uncertainty. With less than 30 samples, the uncertainty becomes very large and the true value has not stabilized.

Fig. 1 shows an example of the parameter comparisons for near following seas, 15 deg heading. In this figure, the true value is represented by a solid line ($1.47\text{x}10^{-8}$). Each extrapolation confidence interval is represented by a vertical line. The extrapolation captures the true value if these lines cross. The mean crossing rate is denoted by a circle and the most probable crossing rate is denoted by a cross. The confidence intervals are asymmetric relative to the mean or most probable crossing rate. This is a property of GPD and different than the symmetric confidence intervals more commonly seen with the Normal distribution.

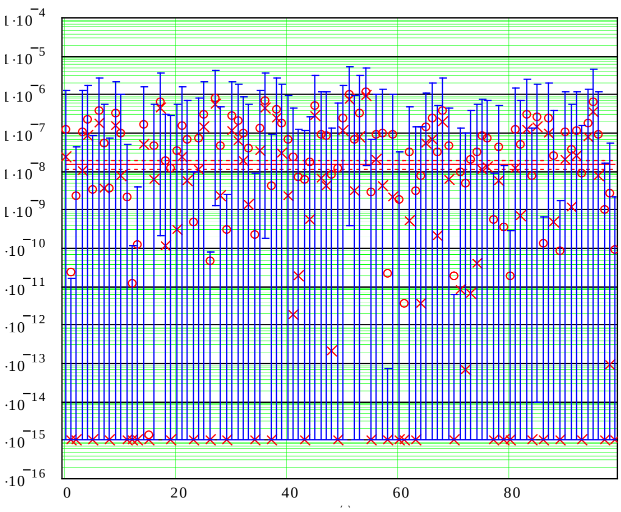

Fig. 1 – Extrapolation-true value comparison plot for roll, 15 deg heading using log normal confidence interval method.

The results for roll are presented in Table 1. The passing rate is the percentage of data sets that pass the parameter comparison, that is, capturing the true value. This table shows an average passing rate, 93.35, that is near the expected confidence level of 95%. Each condition, except for near following seas, 15 deg heading, was also acceptably close to the expected confidence interval. Given only 100 samples, it is unreasonable to expect exactly 95% due to the statistical uncertainty [5]. The 15 degree heading had the lowest roll angles, and the low passing rate could indicate a change in non-linearity at the evaluation level that is not represented at the GPD threshold level.

Most of conditions (6/7) have acceptable passing rates indicating a successful validation of both the mean value extrapolation and the confidence interval. The large confidence intervals certainly contributed to capturing the true value. The fact that the passing rate is not universally 100% indicates the confidence interval is not too large. A complete validation of the extrapolation method would consider many more conditions and motions representing the expected operational scenarios.

Table 1 – Extrapolation results for roll motion and range of headings.

Wave Heading (deg)	Simulation Time (hours)	GPD Threshold Level (deg)	Evaluation Level (deg)	Average Passing Rate
15	230,000	6.947	15	84
30	100,000	12.877	30	91
45	230,000	17.094	60	94
60	100,000	18.754	50	100
90	230,000	16.055	32.5	99
135	230,000	7.359	17.5	92

4.1 Other Considerations

As a point of further discussion, the evaluation level changed based on what was present in the true value data set. This is appropriate for validation. In practice, the true value is not known, and the evaluation level is set arbitrarily at some critical level.

In order to use a validated extrapolation, a further check is required based on the threshold selected for the GPD fit and the evaluation level or critical level. If this difference is too large, there is an indication that the extrapolation is meaningless and should be ignored.

The determination of "too large" requires some understanding of the physical properties involved. Fortunately, for these conditions the motion levels are low enough to be able to ignore them as "too small to care about." Alternatively, more data may be added to increase the GPD fit threshold in borderline cases. This will also reduce the uncertainty.

For the sake of argument, if the passing rates had not been near the desired 95%, the culprit could be the extrapolation method or the confidence interval formulation. A separate investigation would be required to confirm the confidence interval formulation as the comparison is based on confidence interval. For instance, the confidence interval can be evaluated using synthetic data with a known distribution.

6. Conclusions

This paper demonstrated the applicability of a multi-tier validation approach to a statistical extrapolation method based on the Generalized Pareto Distribution. The first tier, parameter comparison, was made by comparing the 95% confidence interval from a GPD extrapolation to the true value. This was done 100 times to determine a passing rate for the second tier, condition, comparison. Lastly, most of the conditions passed the second tier criterion, which passes the third tier, or set, comparison. The extrapolation method would be considered validated. A rigorous validation effort would specify passing percentages at Tiers I and II.

The use of confidence interval for the Tier I comparison implies a working confidence interval formulation. Other comparison metrics, e.g., mean to true value distance, can be used instead if confidence interval is not useful.

The ratio of the GPD threshold and the evaluation level provides a metric for practical use. The conditions with low motions can either have more data added, in the hope of increasing the GPD threshold level, or ignoring the condition as having negligible motions.

Acknowledgments

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

References

[1] AIAA, Guide for the Verification and Validation of Computational Fluid Dynamics Simulations. American Institute for Aeronautics and Astronautics, Reston, VA (1998).

[2] ASME, Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer, American Society of Mechanical Engineers, New York (2009).

[3] ITTC Loads and Responses Seakeeping, Verification and Validation of Linear and Weakly Nonlinear Seakeeping Computer Code, Procedure 7.5-02-07-02.4 Rev 1 (2011).

[4] The Specialist Committee on Stability in Waves, Final Report and Recommendations to the 26th ITTC, Proc. 26th International Towing Tank Conference, Vol II, Rio de Janeiro, Brazil (2011).

[5] T. Smith and B. Campbell, On the Validation of Statistical Extrapolation for Stability Failure Rate, Proc. 13th Intl. Ship Stability Workshop, Brest, France (2013).

[6] B. Campbell, V. Belenky, and V. Pipiras, On the Application of the Generalized Pareto Distribution for Statistical Extrapolation in the Assessment of Dynamic Stability in Irregular Waves, Proc. 14th Intl. Ship Stability Workshop, Kuala Lumpur, Malaysia (2014).

[7] Weems, K. & D. Wundrow, "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 (2013).

[8] Belenky, V. & B. Campbell,. 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 (2011).

[9] Kramer, H., and Leadbetter, M.R., Stationary and Related Stochastic Processes, John Wiley, New York (1967).

[10] Smith, T. C., "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 (2012).