# Addressing Challenges in the Validation of Dynamic Stability Simulation Tools

William Belknap, Timothy C. Smith, and Bradley Campbell
*David Taylor Model Basin (NSWC/CD)*

**ABSTRACT**

Significant challenges remain in the validation and formal accreditation of dynamic stability simulation tools, which can have a limiting effect on operations or concept design due to a need for conservatism. These challenges primarily consist of validation metrics and criteria, uncertainty characterization, and defining the scope of conditions (both environmental and operating) that must be examined in order to ensure that the simulation tool is valid for all conditions of interest. In discussing these challenges, this paper proposes solutions to the problem of validation and formal accreditation that can be applied in future efforts.

**KEYWORDS**

Validation; Simulation; Intact Stability

## INTRODUCTION

Given the breadth of possible environmental and operating conditions (wave height, period, and directionality; wind speed and relative direction), numerical simulation tools are an attractive, if not necessary, option for predicting the dynamic stability performance of a ship. However, in order for a simulation tool to be useful, it must be validated and its limitations understood.

The very challenges that lead to the difficulty in modeling the dynamic stability hydrodynamic problem (see Belknap and Reed (2010)) also contribute to the difficulty in validating the simulation tool. Not only are the stability failure events of interest exceedingly rare, but they are governed by a nonlinear dynamical system. Additionally, the conditions of interest are so vast that the validation domain is essentially limitless.

This paper isn't the first to propose an approach to solving the dynamic stability validation problem. In fact, Grochowalski and Jankowski (2009) provide a validation vision that is perhaps the most complete published to date. The present paper addresses three key challenges that are encountered when such a state of the art process is applied. These challenges are:

1. Validation metrics and criteria
2. Uncertainty characterization
3. Validation scope

It is hoped that the following discussion of these three issues will spur debate within the ship stability community that will result in a higher quality validation of dynamic stability simulation tools.

## VALIDATION METRICS AND CRITERIA

While seemingly straight forward, establishing the validation metrics and criteria for what constitutes acceptable agreement is a significant problem. It is generally not sufficient for a group of subject matter experts to simply agree that the simulation results and model data (for example) are "consistently close enough." Rather, given how the numerical tool is going to be used, definitive quantifiable criteria must be established that can be used to defend the conclusions of the decision makers.

### *Specific Intended Uses*

The first step in establishing metrics and criteria is to define the Specific Intended Use (SIU) of the simulation tool. The more precisely stated the SIU is, the more focused the validation activities can be and the more appropriate the validation criteria will be. Ideally, SIUs should establish:

- type of ship (if not specific ship)[2]
- operating environments (waves and wind)
- loading conditions
- speeds and relative wave headings
- motions and stability failure events of interest
- statistical measures of interest
- how the simulation motion data and statistics will be used

SIUs defined in the manner above will serve to affirm the physical phenomena intended to be modeled by the simulation tool and provide the core set of metrics for which quantitative acceptance criteria will be set. However, while it is necessary to have quantitative acceptance criteria, validation activities must also provide for validation tests for which there is no clear accuracy requirement. These validation tests fall under the realm of "qualitative" validation.

### *Qualitative Validation*

As emphasized by Grochowalski and Jankowski (2009), validation of the component physics is necessary because it is not only the large amplitude motion event physics that are important to validate but also the transient behavior that leads to the event in the first place. The component physics can be thought of as both the component force models (presuming the theoretical model employs force superposition) and/or elements of the dynamic stability hydrodynamic problem domain. This domain includes, but is not limited to:

- wind and wave environment modeling
- roll damping
- calm water maneuvering
- seakeeping
  - radiation problem
  - diffraction problem
- nonlinear stiffness
- appendage forces (including propellers and rudders)
- maneuvering in waves
- drift forces (including added resistance)

[2] It should be noted that the problem trying to be predicted is likely to be a *full-scale* ship, which means that the validation activities should address the potential issues associated with using model scale data as the source validation data.

Qualitative validation is then the aggregate of these elemental/component validation tests which provide confidence in the simulation tool.

There are two primary objectives in validating at the elemental level:

1. Provide insight into what may be causing the simulation results to differ from the validation data in the SIU-defined conditions → Diagnostic capability
2. Provide confidence that the physics are well-modeled in conditions for which direct validation data are not available

These objectives are satisfied by performing elemental validation tests that isolate the active elements of interest. For example, an elemental validation test to assess roll damping prediction is to run a roll decay comparison, where the qualitative validation metrics are the roll decrement coefficient and roll period as a function of roll amplitude. More examples of elemental validation tests and their qualitative metrics are provided in Table 1. An example series of plots showing the qualitative comparison for the regular wave dynamic stability elemental validation runs is shown in Figure 1. Figure 1a) shows the comparison of roll motion through a phase plot. Figure 1b) shows a joint visualization of the predicted and measured track, speed (scatter symbol color), and roll amplitude (scatter symbol size). Finally, Figure 1c) provides a notional dashboard view of the various time history, track, and phase plots that would facilitate qualitative comparisons of results for key motion channels.

What makes these elemental validation tests part of "qualitative validation" is that there are no obvious requirements for just how close the simulation data must match the validation data to affect the SIUs. If a calm water tactical diameter has 50% error, will that mean the simulation tool is unusable for predicting instances of parametric roll? The answer is likely "no." However, such an error indicates that hull lift and/or rudder and propulsor forces are not well modeled when the ship has a drift angle, which may lead to poor ability to predict broaching.

The elements that must be tested depend on the physical phenomena the SIU call for capturing. Ideally, a systematic procedure is followed that lists the failure mechanisms expected to be captured (e.g. pure loss of stability, parametric roll, surf-ride and broaching, synchronous roll, breaking wave impact) and then validates that component models are appropriate for capturing these. If parametric roll is intended to be captured, the simulation tool must adequately model nonlinear stiffness and roll damping, in addition to regular seakeeping forces.

**Table 1: Example elemental tests and their qualitative metrics**

| Test | Element(s) | Metric(s) |
|---|---|---|
| Roll Decay | Roll damping, nonlinear stiffness | Decrement coefficients and periods |
| Calm water turning circle | "maneuvering" forces, appendage forces | Diameters, steady heel, steady drift |
| Calm water zig-zag | "maneuvering" forces, radiation, appendage forces | Overshoots, max roll |
| Regular wave (non-steep) seakeeping | Radiation, diffraction, roll damping | Motion transfer functions: amp. and phase |
| Regular wave (steep) dyn. stability | (several) | Motion time histories, integrity values, max. value scatter |
| Forced motion | (several, but isolated) | Force time histories: amp. and phase |

If grouped waves are required to initiate the event, does the wave model produce grouped waves? If surf-riding is expected to be captured, does the simulation tool properly model resistance in a zero-encounter frequency steep wave? Is the loss of rudder lift captured in a steep following sea? These are just some examples of the questions that must be asked to ensure that the proper elemental tests are performed.

Finally, it should be noted that the nonlinear time series analysis techniques that have been proposed for use as validation tools (see McCue, et al., 2008)) fall into the realm of qualitative validation. While these techniques can be very illuminating in showing fundamental ways in which the time series are different, there is not a defined value to which these metrics need to agree in order for the simulation tool to be useful.

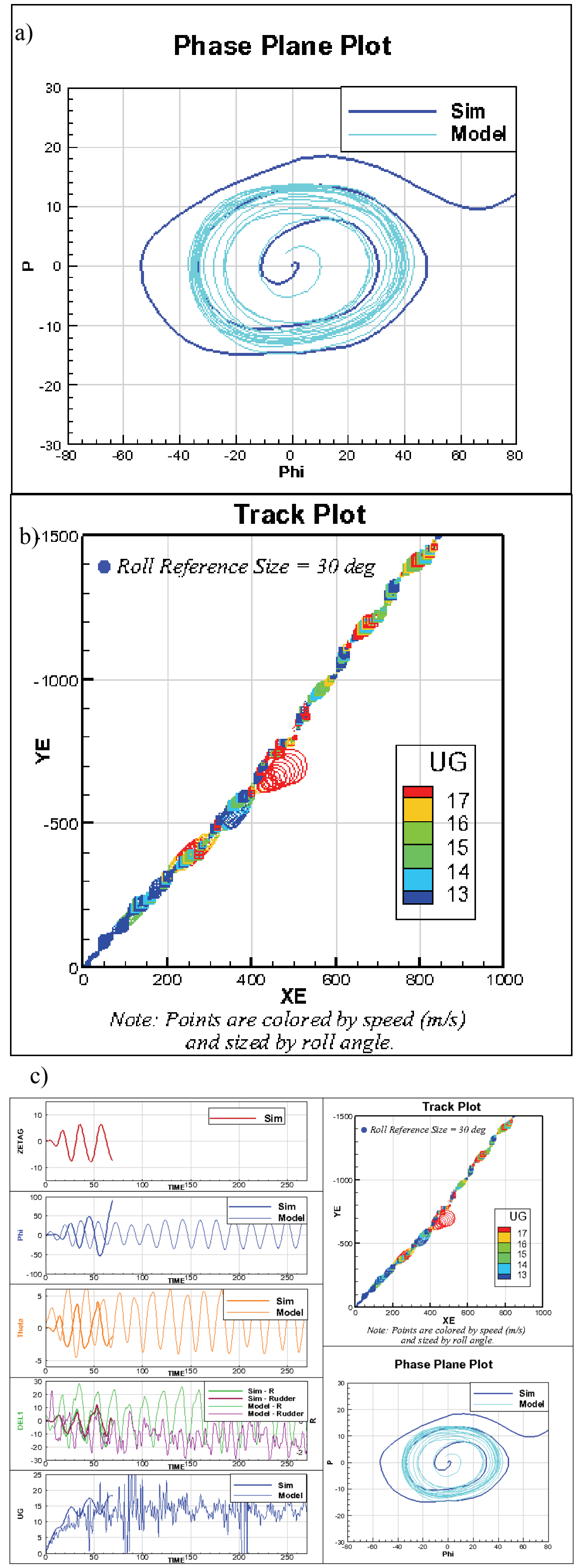


**Fig. 1: Example visualization of regular wave dynamic stability run time history data: (a) roll phase plane plot; (b) track-speed-roll plot; (c) notional "dashboard"**

### *Quantitative Validation*

The key measures that will be used by decision makers to determine whether or not a simulation tool can be accredited (i.e. meets the requirements for a given SIU) are those by which quantitative criteria are established. The only defensible quantitative criteria are those directly tied to the SIU, which means the metrics need to be the very same quantities that the tool is intended to predict. Furthermore, the validation metrics must be obtained for conditions (operational and environmental) as close to those directly defined in the SIU. This requirement is similarly noted by Vassalos, et al. (1998) and Grochowalski and Jankowski (2009).

The problem of establishing acceptability criteria on these metrics is deceptively difficult. In all likelihood, there is an acceptable level of error that can be tolerated and accounted for, but establishing what is acceptable may be an iterative process. The more difficult questions to answer are whether or not all modes of motion must concurrently meet the acceptance criteria, or just certain key channels, and how to define an overall acceptable level of quantitative validation performance where the "passing" trends are inconsistent.

What can be said with certainty is that comparisons must be made to validation data representing the real world conditions of which the tool will be simulating. This ideally means multi-directional seas with wind, but due to difficulty in obtaining and working with that kind of data, long-crested irregular wave (no wind) runs may be all that is available. To the degree that real-world conditions are not directly tested (referred to as "direct validation"), the less-than-ideal validation data set must be used for quantitative validation in conjunction with the elemental validation tests in an approach referred to as "indirect validation."

The last item to be discussed within the context of validation metrics and criteria is the issue of uncertainty characterization. Because this issue represents a significant challenge by itself, it will simply be stated that the validation metrics and acceptance criteria must acknowledge the uncertainty in the system and take steps to address it. The next section discusses this particular challenge.

## UNCERTAINTY CHARACTERIZATION

As stated earlier, the dynamic stability problem represents a stochastic nonlinear process where the events of interest can be extremely rare. Because of this, the quantities of interest are not known exactly, but rather have an associated uncertainty. The validation of dynamic stability simulation tools must therefore properly address this associated uncertainty.

### *Stochastic Process Uncertainty*

The dynamic stability problem represents a particular challenge in defining the variance of mean or variance estimates, due to the fact that ship motion is a continuous random process. Care must be taken to ensure the independence of data sets, which means that the length of runs (both model data and simulation data) and the number of component frequencies used to create the exciting wave fields must be appropriately selected. The approach to addressing the challenges related to stochastic process uncertainty in the dynamic stability problem is discussed in detail by Smith (2011).

### *Nonlinear Process Uncertainty*

More widely discussed in dynamic stability research are the issues related to dealing with a nonlinear system. Particularly, it is well acknowledged that initial conditions can have a significant impact on the trajectory of the model (or simulated ship). It is the uncertainty associated with the variation in results that can occur that must be addressed. Free-running model test data that are used for validation purposes more often than not do not provide precise information on the initial conditions relative to the wave field at some arbitrarily stated start time. If the initial conditions are not known precisely for a given model test run (where nonlinear dynamics are believed to be in play), several model test runs at random initial conditions are the minimum requirement for testing initial condition dependency and providing data that is useful for validation.

On the simulation side, the solution is to obtain simulation runs over a wide range of initial conditions to ensure the level of initial condition dependency is determined. It is recommended that, at a minimum, the varied initial conditions include:

- Wave phase
- Roll angle
- Roll rate

An example of this validation process is shown in Figures 2 through 4 for the elemental validation test of steep regular wave dynamic stability runs (free running condition at a constant ordered speed and heading). Figure 2 presents a quarter-polar plot of the validation comparison for a single regular wave condition [wave length ($\lambda$) = 0.75 * ship length (L); wave height (H) = 0.1$\lambda$], where the surface mesh coloring represents the 60-deg roll event integrity value (IV = $N_{runs\_with\_events}$ / $N_{runs}$) for the simulation tool prediction and the scatter data points represent the model data, which are also colored by integrity value. Ideally, the simulation surface mesh will have the same color as the model data scatter points, representing an agreement in integrity values. Where there is a difference, further investigation is warranted. This comes in the form of the data shown in Figures 3 and 4.

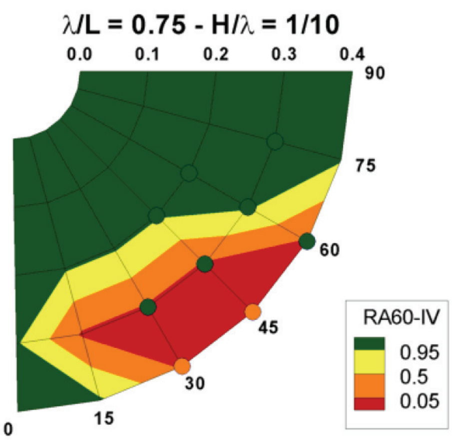


**Fig. 2: Regular wave dynamic stability elemental validation data using contours of integrity values (IV) for 60-degree roll angle exceedance over a range of ship speeds (Fn=0.0 to 0.4) and relative wave headings (□=0 to 90 deg) [sim. – surface; model data – circles)**

Figure 3 shows the actual distribution of maximum roll across relative wave heading for a single speed (Fn=0.4). At the 30-deg relative wave heading, Figure 2 shows that the simulation predicts all runs at this condition to have a roll greater than 60 degrees (as seen by the red surface at that point), whereas the model data point has some cases where roll exceeds 60 degrees and others that do not, depending on initial conditions. Examination of Figure 3 shows that, in fact, all of the simulation runs go well beyond 60 degrees to the point where they all “capsize”, whereas the model data are scattered between 42 degrees and capsize. This would imply that the simulation is not providing a good prediction.

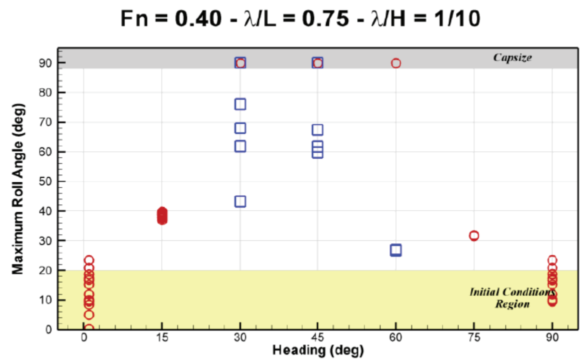


**Fig. 3: Maximum roll values as a function of relative wave heading at a single speed for regular wave dynamic stability case (simulation – red circles; model data – blue squares)**

Clearly, there is a sharp transition across relative wave heading, such that the simulation tool predicts a “green” condition at 15-deg relative wave heading. Therefore, it’s reasonable to wonder about uncertainty in the ordered heading. But what about uncertainty in the actual achieved wave conditions in the model basin? The results plotted in Figures 2 and 3 overlay data such that it is assumed all represent behavior in wave conditions with the ideal wave steepness of 1/10. However, the wavemakers in any model basin are an imperfect system, in that the actual realized wave steepness (and possibly wave length) may be off by some amount. Figure 4 presents this example validation condition (30-deg relative wave heading, Fn=0.4, $\lambda$/L=0.75, H/$\lambda$=1/10) in a different dimension. The simulation data was run at much finer resolution in wave steepness than the model data in order to develop the shown integrity surface. This allows the model data to be plotted at the actual achieved wave conditions. In this example condition, it can be seen that the average achieved wave steepness in the model experiment was actually slightly less than H/$\lambda$=1/10). This places the model result within the steep region of the simulation integrity surface, which would change the validation conclusion to “good agreement.”

The narrative above represents the effect of "input parameter uncertainty," which is the third uncertainty-related challenge to be addressed within the dynamic stability validation problem.

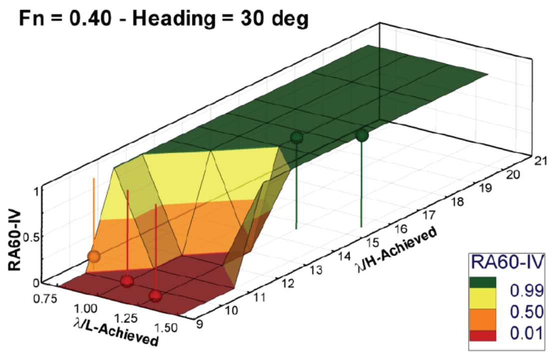


**Fig. 4: Integrity surface for regular wave validation data (integrity values for 60-deg roll) as a function of wave steepness and length (simulation – surface; model data – spheres)**

### *Input Parameter Uncertainty*

In describing the validation uncertainty related to the nonlinear dynamical process of the dynamic stability problem, it was shown in an example case that uncertainty in the knowledge of the input parameters (wave height, in the previous example) can have a dramatic impact on the validation conclusions drawn. Particularly in the quantitative validation stage where direct comparisons are made with real-world data, input uncertainty can be a significant challenge.

There are many parameters that are provided as input to the simulation validation cases that in many cases in the past have been taken as *known* quantities. In reality, these quantities represent best estimates that are based on direct measurement of an unchanging variable (such as mass properties) or statistical analysis of a random process (such as the seaway). As the state of the art in validation of dynamic stability simulation tools is expanded, the issue of input uncertainty must be addressed. In other words, validation conclusions must be based on a comprehension of the propagation of uncertainty due to errors in the assumed input parameters.

The issue of addressing error propagation in the validation of engineering models has been studied by Hills and Trucano (1999). They specifically address the effect of error propagation in statistical validation of nonlinear systems, which means their work is particularly relevant to the validation of dynamic stability simulation tools. One approach offered by Hills and Trucano to assessing the impact of error propagation is to perform Monte Carlo simulations that examine at a system level the impact of perturbations to a matrix of input parameters.

To get a first look at the impact of input parameter uncertainty on the dynamic stability validation problem, a sample Monte Carlo based study was conducted that examined notional input errors in the validation of a stern quartering high sea state condition. The study was not designed to be a comprehensive examination of the impact to be expected across typical validation cases, but rather to provide a snapshot on a single case as an exercise in the process.

The input parameters to be varied within the study were taken as representative sources of error expected in the validation of a real-world scenario. These were taken to be:

- Relative wave heading
- Significant wave height
- Modal period
- GM

The magnitudes of the errors are hypothetical, but are representative of the size of 95% confidence intervals. Using the "best estimate" input values as the basis, the Monte Carlo approach parametrically varied the four sources of error to include simulations at the upper and lower bounds of the input values. These input values are shown in Table 2. Even with this coarse resolution, the resulting matrix of conditions is 3x3x3x3 = 81 discrete simulation conditions to capture the potential error in validating a single validation condition.

**Table 2: Input value uncertainties for Monte Carlo simulation matrix**

| | **Lower Bound** | **Best Estimate** | **Upper Bound** |
|---|---|---|---|
| **Heading** | 35 deg | 45 deg | 55 deg |
| **$H_{1/3}$** | 11.0m | 11.5m | 12.0m |
| **$T_m$** | 13.0s | 14.0s | 15.0s |
| **GM** | $GM_o$ - .05m | $GM_o$ | $GM_o$ + .05m |

The results were processed to examine the variation in the mean up-crossing rate for roll, as this is a typical quantity of interest for dynamic stability analysis. Figure 5 shows all simulation results for the roll mean up-crossing rates as determined from EPOT analysis (see Campbell and Belenky, 2010). The solid blue line represents the simulation results based on the best estimates of the input parameters. The scatter points (at select roll levels of 20, 35, 45, and 60 degrees) show the range of variation in the results that can be expected when the uncertainty in the input values are accounted for.

The results show that, for at least this hypothetical case, the variation in the mean up-crossing rates can vary on a scale considerably greater than the statistical process uncertainty (shown by the "best estimate" 95% confidence intervals).

Based on this single example result, it is reasonable to consider input value uncertainty when interpreting validation results. However, it must be emphasized that the present study is a single hypothetical example and the size of the effect of input error is likely to be a function of the physics. It is anticipated that certain parameters will have more of an effect than others depending on the dynamics of the problem (e.g., resonant conditions vs. pure excitation, etc.), which are condition dependent.

More work must be undertaken to relate the joint probabilities of the various input errors to the final total confidence interval size. The example calculations presented in this paper represent the first attempt by the authors to investigate other degrees of uncertainty in the validation problem.

## VALIDATION SCOPE

The final validation challenge addressed in the present paper involves answering a question that must always be made in a validation task: When is the validation complete?

The only defensible answer to the above question is that there is no good answer. Furthermore, the answer should not be "when the available validation data are exhausted."

It has been argued that validation tasking should include qualitative validation and quantitative validation. The scope of the qualitative validation is more easily defined in that it should include the full range of ship speeds and address all components of interest (as defined by the SIU). If the qualitative validation is incomplete, the ability to perform "indirect validation" will be limited or considered higher risk. The scope for quantitative validation is considerably more challenging to define. This is generally because the number of environmental conditions a ship will see is infinite. Not only are there combinations of significant wave height and modal period (as well as spectral shape) to examine, but if bi-directional seas and relative wind direction are part of the environment defined in the SIU statement, the matrix of conditions quickly grows in size.

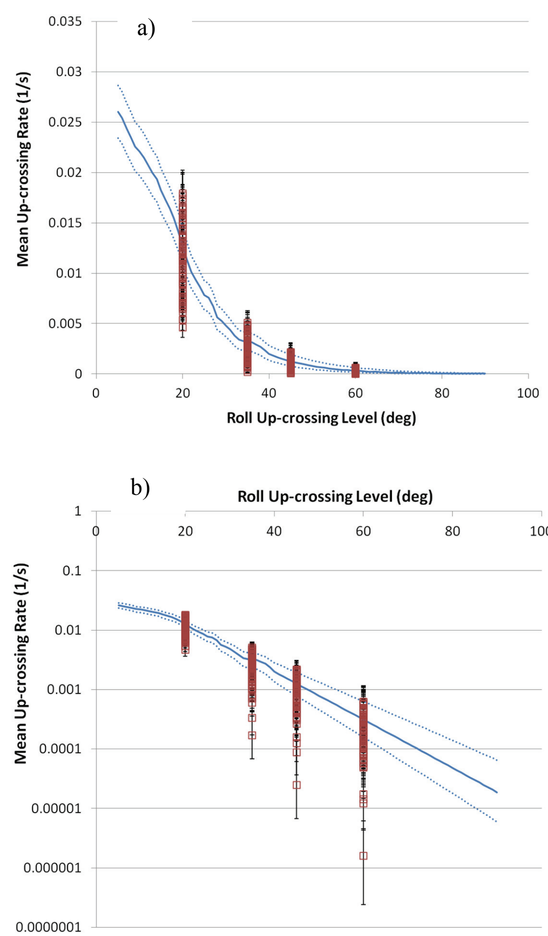


**Fig. 5: Simulation results for the mean up-crossing rates for roll using the best estimate of input parameters (solid blue line with dashed blue lines for 95% confidence intervals) and the parametrically varied input parameters (red scatter points with 95% confidence intervals) – presented on: a) straight scale and b) log scale**

It is not reasonable to expect that direct validation (with quantitative metrics) will be performed for all conditions for which the simulation tool will be accredited. First and foremost, a simulation tool would not be required if trusted data existed for every condition of interest. Also, the variability of the physics between directly validated conditions may not be significant. Even so, a sufficient range of directly validated cases must be obtained. For those conditions that are not directly validated, an argument must be made that the physics in play for the non-tested conditions are not unique. The elemental validation cases and the neighboring quantitative validation cases will be relied upon to provide confidence that the simulation tool is valid at these “interior” conditions.

In the end, the validation scope is likely to be a judgment call by the subject matter experts. The resolution of the validation domain may need to be adjusted as the validation process proceeds.

## CONCLUSIONS

This paper has addressed the key challenges in the validation process for dynamic stability simulation codes. These challenges are:

- Developing the validation metrics and criteria
- Characterizing and incorporating uncertainty
- Defining the validation scope

The key point that is made relative to successful establishment of validation metrics and criteria is that the Specific Intended Use (SIU) of the simulation tool must be completely and precisely defined. The metrics and the acceptance criteria should flow directly from the SIU. The metrics and criteria can be divided between qualitative and quantitative categories. While the quantitative metrics are the most valuable for the decision makers in determining the usefulness of the simulation tool, the qualitative validation tests are essential to use for diagnostics and in support of “indirect” validation. And while the quantitative metrics are the most useful, they must be well thought out and, if possible, incorporate an allowable error that is tied to the SIU.

The challenges associated with uncertainty characterization involve stochastic process uncertainty, nonlinear dynamics uncertainty, and input parameter uncertainty. All three can factor heavily into the validation process and interpretation of results. At the moment, the influence of input parameter uncertainty remains an area of research, though an example study using hypothetical errors in the input parameters has demonstrated that the magnitude of the scatter due to input error can plausibly be larger than stochastic process uncertainty alone.

Finally, difficulty arises in determining when the body of validation work is sufficiently complete. Particularly when the vast matrix of possible environmental conditions is considered, it becomes obvious that a direct validation cannot be performed for all operational (speed, heading, load condition) and environmental conditions. It is concluded that indirect validation will fill the void, but that direct validation must be done for conditions where the physics are sufficiently unique.

### *Future Challenges*

This paper does not represent an exhaustive list of challenges remaining in the validation of dynamic stability simulation tools. Many difficulties will need to be addressed in future efforts, particularly as the “realism” of the validation data increases. For example, the non-stationarity of real-world environmental conditions will require special treatment when these data are used for validation. Another challenge to be addressed is validation for full-scale ships, when the vast majority of available validation data comes from model experiments. Also, there is the very real scenario of dealing with validation results where 100% of validation conditions do not pass the criteria, and there is not a clear trend as to where the simulation is not properly modeling the physics, is the simulation invalidated? Is there an acceptable passing rate across all conditions and what should it be?

These and other future challenges will be discussed in the coming years, but at present, the validation of dynamic stability simulation tools stands to benefit from an improved understanding of metrics and criteria, uncertainty characterization, and proper establishment of validation scope.

## ACKNOWLEDGMENTS

The authors would like to thank Dr. Vadim Belenky, Dr. Art Reed, and Mr. Jim Webster for their thoughtful insights into the challenges discussed within this paper, as well as countless others who have participated over the past years in

lively discussions surrounding validation of dynamic stability simulation tools.

## REFERENCES

Belknap, W. F. & A. M. Reed (2010), "TEMPEST — A New Computationally Efficient Dynamic Stability Prediction Tool," *Proceedings of the 11th International Ship Stability Workshop*, pp. 185-197, Wageningen, The Netherlands.

Campbell, B. L. & V. Belenky (2010), "Assessment of Short-Term Risk with Monte-Carlo Method," *Proceedings of the 11th International Ship Stability Workshop*, pp. 85-92, Wageningen, The Netherlands.

Grochowalski , S. & J. Jankowski (2009), "Validation Methodology for Simulation Software of Ship Behaviour in Extreme Seas," *Proceedings of the 10th International Conference on Stability of Ships and Ocean* Vehicles, pp. 409-420, St. Petersburg, Russia.

Hills, R. G. & T. G. Trucano (1999), "Statistical Validation of Engineering and Scientific Models: Background," SAND99-1256, Sandia National Laboratories, Albuquerque, NM.

McCue, L. S., W. R. Story & A. M. Reed (2008) "Nonlinear Dynamics Applied to the Validation of Computational Methods." *Proc. 27th Symp. on Naval Hydro*., 10 p, Seoul, South Korea.

Smith, T. C. (2011), "Statistical Data Set Comparison for Continuous, Dependent Data," *Proc. of the 12th International Ship Stability Workshop*, 7 p, Washington, DC.

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

## DISCLAIMER

The opinions expressed within this paper are strictly those of the authors and do not represent an official position of the U.S. Navy.