Proceedings of the 12th International Conference on the Stability of Ships and Ocean Vehicles, 14–19 June 2015, Glasgow, UK.

Direct Assessment Will Require Accreditation—What This Means

Arthur M. Reed, *Carderock Division, Naval Surface Warfare Center (NSWCCD)*, arthur.reed@navy.mil

Aurore V. Zuzick, *Carderock Division, Naval Surface Warfare Center (NSWCCD)*, aurore.zuzick@navy.mil

ABSTRACT

With the advent of the second-generation intact stability criteria, IMO has initiated a two-tier performance-based stability assessment process for unconventional hulls. If the design fails the first tier evaluations, it progresses to the second tier, where direct assessment criteria are applied. The design is considered satisfactory if the direct assessment criteria are passed. If these criteria are not passed, operator guidance is needed to provide vessel operators with the information needed to safely operate the vessel in dangerous conditions. Ship motion simulation tools are needed to apply the direct assessment criteria and generate operator guidance, if necessary.

A framework is presented for certification that simulation tools used for direct assessment of stability failures and generation of operator guidance are sufficiently accurate for these purposes. Based on US Navy experience, guidance is provided on the Verification, Validation and Accreditation (VV&A) pro-cess, structure, and participation, and acceptance criteria are given for both quantitative and qualita-tive accreditation approaches. Accreditation acceptance criteria are tailorable to ship-specific VV&A efforts, particularly with regards to definition of critical motions and physical limits.

Keywords: *Verification, Validation and Accreditation; VV&A; Formal VV&A*

1 INTRODUCTION

For commercial vessels, the classical intact stability criteria is based on the work of Rahola (1939) and is incorporated in the International Code on Intact Stability, the 2008 IS Code (MSC 85/26/Add.1[1]). Similar criteria for naval vessels is provide by Sarchin & Goldberg (1962) and codified in the NATO Naval Ship Code (NATO, 2007a,b) and by a US Navy Design Data Sheet (Rosborough, 2007). These criteria are prescriptive—that is they are a set of criteria, defined based on empirical data, which are *assumed* to ensure that a vessel meeting the criteria will have adequate static stability. The history of development and the background of the IMO criteria are described by Kobylinski & Kastner (2003); a summary of the origin of these criteria is also available in chapter 3 of the Explanatory Notes to the International Code on Intact Stability (MSC.1/Circ.1281).

The deficiency of these prescriptive ap-

[1]References to IMO documents such as “MSC 85/26/Add.1” appear in the list of references with an “IMO” prefix, i.e. as: IMO MSC 85/26/Add.1. As there is no ambiguity in the names of the IMO citations, the year will be omitted from the citations.

proaches is that their adequacy is contingent upon vessels and their modes of operation lying within the "design space" of the vessels that define the empirical data used to derive the criteria. However, the design space is not necessarily well defined and modern vessels are more and more tending to lie outside of the traditional design space—the classical intact stability criteria do not apply to these latter vessels.

Beginning in the early 2000's efforts were initiated to develop performance based stability criteria for commercial vessels with the re-establishment of the intact-stability working group by IMO's Subcommittee on Stability and Load Lines and on Fishing Vessels Safety (SLF) (cf. Francescutto, 2004, 2007). Over time, the terminology to describe the new intact stability criteria evolved from "performance based" to "next generation" to "2nd generation," the terminology in use today. This entire evolution is described in the introduction to Peters, *et al.* (2011).

The SLF Working Group decided that the second-generation intact stability criteria should be performance-based and address three modes of stability failure (SLF 48/21, paragraph 4.18):

- *Restoring arm variation* problems, such as parametric roll and pure loss of stability;
- *Stability under dead ship condition*, as defined by SOLAS regulation II-1/3-8; and
- *Maneuvering related problems in waves*, such as surf-riding and broaching-to.

Ultimately, a fourth mode of stability failure was added:

- *Excessive accelerations*.

The deliberations of the Working Group led to the formulation of the framework for the second generation intact stability criteria, which is described in SLF 50/4/4 and was discussed at the 50th session of SLF in May 2007. The key elements of this framework were the distinction between parametric criteria (the 2008 IS Code) and performance-based criteria, and between probabilistic and deterministic criteria. Special attention was paid to probabilistic criteria; the existence of the *problem of rarity* was recognized for the first time and a definition was offered. Also, due to the rarity of stability failures, the evaluation of the probability of failure with numerical tools was recognized as a significant challenge.

"Second-generation intact-stability criteria" are based on a multi-tiered assessment approach: for a given ship design, each stability-failure mode is evaluated using two levels of vulnerability assessment. The two tiers or levels of vulnerability assessment criteria are characterized by different levels of accuracy and computational effort, with the first level being simpler and more conservative than the second.

A ship which fails to comply with the first level is assessed by the second-level criteria. In a case of unacceptable results, the vessel must then be examined by means of a direct assessment procedure based on tools and methodologies corresponding to the best state-of-the-art prediction methods in the field of ship-capsizing prediction. This third-level criteria should be as close to the physics of capsizing as practically possible.

The framework and the concept of vulnerability criteria were first introduced in Belenky, *et al.* (2008a). The state-of-the-art in the assessment of vulnerability is presented in detail in Peters, *et al.* (2011). Criteria for pure loss of stability, parametric roll, and surf riding and broaching were codified in February of this year in SDC 2-WP.4 Annexes 1, 2 and 3, respectively.

Direct assessment procedures for stability failure are intended to employ the most advanced technology available, yet be sufficiently practical so as to be uniformly applied, verified, validated, and approved using currently available infrastructure. Ship motions in waves, used for assessment on stability performance, can be reproduced by means of numerical simulations or model tests (SLF 55/3/11). The process of approval, which we will call accreditation will be the major focus of the remainder of this paper.

The structure of this paper will consist of a definition of Verification, Validation and Accreditation (VV&A), a description of the VV&A process, and accreditation criteria. The VV&A process will be subdivided into the process structure, documentation, specific intended uses, and a description of Verification and Validation (V&V). The acceptance criteria will be split between quantitative and qualitative criteria, where quantitative is the more rigorous and thus more difficult.

2 DEFINITION OF VV&A

If decisions regarding the design and construction of ships, each costing hundreds of millions of dollars, if not a few billion dollars (in the case of naval vessels), are going to be made based on the stability predictions of a simulation tool, there must be a reasonable assurance that the tool provides acceptably accurate results. The process by which a tool may be determined to be sufficiently accurate is known as verification, validation and accreditation.

Quoting from a US Navy VV&A presentation, "Verification, Validation, and Accreditation are three interrelated but distinct processes that gather and evaluate evidence to determine, based on the simulation's intended use, the simulation's capabilities, limitations, and performance relative to the real-world objects it simulates." Beck, *et al.* (1996), AIAA (1998), DoD (1998, 2003, 2007, 2012), McCue, *et al.* (2008), ASME (2009), and Reed (2009) provide different, although consistent, definitions of the three components of VV&A. The U.S. DoD definitions for these three terms are provided below, each followed by a practical commentary relevant to computational tools for predicting dynamic stability.

1. Verification—the process of determining that a model or simulation implementation accurately represents the developer's conceptual description and specification, i.e., does the code accurately implement the theory that is proposed to model the problem at hand?

2. Validation—the process of determining the degree to which a model or simulation is an accurate representation of the real world from the perspective of the intended uses of the model or simulation, i.e., does the theory and the code that implements the theory accurately model the relevant physical problem of interest?

3. Accreditation—the official determination that a model or simulation, . . . is acceptable for use for a specific purpose, i.e., is the theory and the code that implements it adequate for modeling the physics relevant to a specific platform? In other words, are the theory and code relevant to the type of vessel and failure mode for which it is being accredited?

2.1 Verification

Experience with attempting to verify ship-dynamics software has been that the documentation for many hydrodynamic codes, particularly the theoretical basis, is neither complete nor rigorous enough for the verification process to be separated from the validation process. Under these circumstances, when one finds that the computations do not adequately model the physical reality, one is left to ponder whether the code is not accurately modeling the intended physics or whether the intended physics are not adequate for the problem. In this case, the dilemma becomes: should one attempt to debug the code or should one abandon use of the code because its underlying physics model is not adequate? Attempting to resolve this dilemma can be expensive, in terms of both time and money.

Another issue related to verification of software is the actual quality of the code and the documentation of the code itself. Often the coding does not follow any consistent standard and there is often insufficient guidance to link the actual code back to its theoretical basis.

As for the actual verification of the code, this is best done by means of unit tests, where each module and block of modules is exercised against known or expected solutions. When properly constructed, these unit tests will not only test the module against normal execution, but also against unexpected or unanticipated inputs, to determine if the code handles error exceptions correctly via error traps or error returns. Many codes are not designed robustly enough so as to deal with anomalous inputs—they expect that the input will always be correct and that all modules output that is input to other modules provide correct input. Rationally, this is a rather naïve assumption.

2.2 Validation

Validation commences with a series of Elemental Tests (or comparisons to model data), which provide insight into a simulation's ability to capture the overall physics of the ship motions in waves problem. Elemental tests consider such quantities as roll decays, calm water turning circles, calm water zig-zag maneuvers, turning circles in regular waves, and acceleration from rest in calm water. The results of the elemental tests provide evidence that the computational tool is capturing the physics of the problem of a ship maneuvering in waves. They also provide confidence that the quantitative comparison results obtained with available model data may be assumed characteristic of the code and applicable for similar conditions for which model data is not available.

It is reasonable to assume that if a predictive tool is capable of predicting responses in extreme seas, it should be capable of making reasonable predictions of motions in moderate seas. The motions problem in small and moderate seas can be characterized as the seakeeping problem. In the seakeeping problem, the ship's control system should have no difficulty in maintaining the ordered speed and heading—on the average the vessel will maintain a constant heading at a constant speed. These are the standard assumptions of seakeeping theory.

Thus, as a first order validation, the computational tool should be capable of reproducing the single significant amplitude motions that are measured during a model test in moderate seas, where we interpret the term motions in a most liberal way as motions, velocities and/or accelerations—this can also be considered an Elemental Test. This liberal interpretation is necessitated by the fact that, depending on how the experiment is run, it can be very difficult to measure linear (as opposed to rotational) displacements. The major challenge here is that experimental data is required, and the experimental data must be of sufficient duration in irregular seas to have sufficiently small confidence bands for the comparisons to be meaningful (*cf.* ITTC, 2011, Sect. 5; 2014, Sect. 5). The Acceptance Criteria section to follow will discuss some possible statistical means of comparison.

In order to accommodate the validation of simulations for predicting motions in extreme seas and stability failures, situations must be examined that are not easily characterized using techniques that are routinely used for seakeeping validation. Nonlinear dynamics methods appear to show significant promise. There are two aspects of nonlinear dynamics that appear to apply to validation. The first is *nonlinear time-series analysis* and the second is *bifurcation analysis*, these methods are discussed in detail in Reed (2009), summarized here. A third issue is that of *the problem of rarity*, which is also briefly discussed below.

Nonlinear Time-Series Analysis—In nonlinear time-series analysis (*cf.* Kantz & Schreiber, 2004), the same time-series analysis is applied to motions measured on a physical model (or ship) and to simulations of the same vessel, in the same environment, as observed during the measurements. The results of the two sets of analysis are compared to each other, often graphically, to determine whether they have produced similar results.

McCue, *et al.* (2008) provides examples of nonlinear time-series analysis, applied as it might be for validation of simulations. Both qualitative and quantitative metrics that may apply were examined. Some qualitative measures include: reconstructed attractors, correlation integrals, recurrence plots, and Poincaré sampling; possible quantitative measures are: correlation dimension, Lyapunov exponent comparison, system entropy, and approximations to the equations of motion (EoM).

While nonlinear time-series analysis techniques can easily illustrate differences between measurements and predictions, there is still much to be investigated. The range of time-series analysis techniques, which may be applicable to dynamic-stability failure prediction certainly has not been exhausted. However, these comparisons are at best qualitative; quantitative methods, particularly for physical understanding and for comparing experimental and computed results, are needed. Bifurcation analysis techniques may provide this necessary additional insight.

Bifurcation Analysis—There are at least four bifurcations that have been observed in ship dynamics which could be used to analyze whether or not a dynamic-stability code is producing the correct dynamic behavior: Fold bifurcation (Spyrou, 1997; Belenky & Sevastianov, 2007: Sect. 4.5.2 for roll, Sect. 6.5.6 for yaw; Francescutto, *et al.*, 1994), Flip bifurcation (Spyrpou, 1997; Belenky & Sevastianov, 2007: Sect. 4.5.3 for roll, Sect. 6.5.6 for yaw), Hopf bifurcation (Spyrou, 1996; Belenky & Sevastianov, 2007: Sect. 6.5.2; Kan, 1990a,b), and Homoclinic bifurcation (Belenky & Sevastianov, 2007: Sect. 6.3.5). Bifurcation analysis (Spyrou, *et al.*, 2009) would appear to be appropriate for application to the lateral-plane aspects of dynamic stability.

The Problem of Rarity—Another issue for the VV&A of simulations for dynamic stability is the "problem of rarity," where the time between events is long compared to the wave period (Belenky, *et al.*, 2008a,b). Large numbers of realizations may be required to observe dynamic stability failures, either in a simulation or experimentally.

Even if these events are observed, direct comparison between realizations is difficult due to the stochastic nature of the failure event. One method that may help to resolve this problem is the use of deterministic critical-wave groups. This would enable direct comparison of realizations, while also capturing the worst-case conditions of the stochastic environment necessary to assess the ship's stability performance. Themelis & Spyrou (2007, 2008) demonstrated the production of deterministic critical-wave groups using simulation tools, and Clauss (2008) and others have done so experimentally.

2.3 Accreditation

Accreditation is the process by which a computational tool is certified as being sufficiently accurate and thus acceptable for use in a particular case for a particular vessel of class of vessels. In the IMO context, this would be a vessel of a particular size and proportions, which will have a particular mode of operation. In practice this would also be tied to a particular mode of stability failure, and would be defined as a particular Specific Intended Use (SIU).

As much of the rest of this paper will be focused on accreditation, accreditation will not be discussed further here except to state that accreditation can be thought of as validation with acceptance criteria. Depending on the druthers of the Flag Administration, accreditation may require more model data than validation, but this is a detail—albeit a potentially expensive one, that does not affect the process.

3 DESCRIPTION OF THE VV&A PROCESS

The VV&A in the process leading to accreditation by a Flag Administration must be a formal process with structure that is prescribed. The process and structure that will be described is that employed by the US Navy (Navy, 1999, 2002, 2004, 2005). However, some commentary will be provided as to how this process might be modified without compromising the integrity of the process.

3.1 Accreditation Responsibilities and Organizations

This structure includes the identification of an Accreditation Authority (AA) and the establishment of three panels: the Accreditation Review Panel (ARP), the Simulation Control Panel (SCP) and the Modeling and Simulation Proponent (MSP). There are four documents that are produced during this formal process: an Accreditation Plan (AP), a Verification and Validation (V&V) Plan, a V&V Report, and an Accreditation Report, The first three of these are produced by the MSP under the guidance of the

SCP, and the latter is produced by the SCP. All of the VV&A efforts are centered about a statement or set of statements that define what the vessel is that will be assessed, its mode of operation and the stability failures that are considered critical for this type of vessel—these are the Specific Intended Uses. Finally, the process includes verification and validation of the modeling and simulation (M&S) tool.

The AA is the individual representing the Flag Administration who will actually accredit the modeling and simulation tool for use with a particular specific intended use (SIU). The ARP is the panel which recommends to the AA whether or not he should accredit the simulation tool. The group in the middle of this process is the SCP who guide the VV&A process, providing guidance to the MSP review the MSP products and prepare a report based on the resulting simulations for the ARP. The SCP is composed of the individuals who will actually perform most of the work, preparing plans, running the simulations, and preparing the V&V report.

The following material based on "Best Practices Guide for Verification, Validation, and Accreditation of Legacy Modeling and Simulation" (Navy, 2005), describes the role and responsibilities of the AA, ARP, SCP and MSP.

Accreditation Authority—The AA is the senior management level individual directly responsible to approve the use of an M&S capability for a particular application or set of applications. The AA will:

a. Resource the VV&A effort
b. Develop the accreditation process
c. Establish the ARP, approve the chairman and its charter
d. Designate models and/or simulations for VV&A
e. Approve the M&S Accreditation Plan
f. Accredit the models and/or simulations (Approve/Disapprove/Resolve ARP M&S accreditation recommendations and assessment reports)
g. Maintain and disseminate gathered VV&A information

Accreditation Review Panel—The ARP is composed of AA representatives and Subject Matter Experts (SMEs) as needed, and the ARP will include a Flag Administration representative(s). The Flag Administration will reconvene the ARP for each M&S milestone effort and should allow tailoring of approaches and participants to the specific models and simulations under consideration. The AA or his designated representative chairs the ARP. The ARP will:

a. Develop M&S Accreditation Plans with MSP assistance
b. Establish Simulation Control Panels (SCPs) (Report all resource requirements for VV&A activities to the AA prior to execution of tasking)
c. Approve the V&V Assessment Report
d. Review V&V information
e. Prepare the Accreditation Recommendation Letters

The ARP Chair shall:

a. Approve the SCP Charter, establish the SCP, designate the Chair, and approve SCP membership
b. Coordinate development of the Accreditation Plan for the designated M&S
c. Oversee SCP activities
d. Approve the VV&A Assessment Report

Simulation Control Panel—The SCP(s) should consist of technical SMEs from the relevant Flag Administration and supporting organizations. The SCP is not a permanent body. An SCP will be chartered for each model or simulation designated for accreditation. The SCP chairman is designated by and reports directly to the ARP chairman. The SCP will:

a. Provide guidelines for V&V Plan development to the MSP
b. Approve the V&V Plan
c. Guide the gathering of V&V information
d. Provide guidelines for the V&V Report to the MSP
e. Approve the V&V Report
f. Prepare the Accreditation Report and deliver it to the ARP

M&S Proponent—An MSP is a developer, maintainer, modifier, or user of a model or simulation designated for VV&A. The MSP will:

a. Provide a Point of Contact (POC) to the ARP Chairman
b. Assist the ARP in drafting the M&S Accreditation Plan
c. Develop a Configuration Management (CM) Plan for the M&S
d. Develop a V&V Plan and deliver to the SCP
e. Execute the V&V Plan upon approval by the SCP
f. Develop the V&V Report and deliver to the SCP, along with supporting documentation
g. Assist the SCP in determining model capabilities versus requirements
h. Provide VV&A Status to the Flag Administration M&S

With the assistance of the MSP, the SCP will identify model test data that is appropriate for use in the VV&A process and also define the acceptance criteria that the MSP will use in its comparison of computed results to experimental results. There are two substantial challenges related to this, the first and potentially most expensive of these will be identifying sufficient data of acceptable quality for use in the validation effort. As identified in ITTC (2011, Sect. 5; 2014, Sect. 5), this is not something that can be done with a single run of a model in a single sea state. It is conceivable that 10's of runs will be required at each speed and heading in each relevant sea state. If sufficient data is not available, the confidence intervals for the results will be so large as to render the comparisons meaningless.

The second challenge is that of deciding what constitutes an acceptable comparison between experimental results and simulations. This is an area in which there is substantial experience and in which there is significant guidance—see the last section of this paper.

An issue that is often overlooked in the VV&A process is Configuration Management (CM). Because software is seldom static—it tends to change over time. If software changes after it has been accredited, there is no assurance that it is still capable of simulating what it was accredited for correctly. Thus, the necessity of a Configuration Management Plan; the development of a CM Plan is one of the MSP's responsibilities. Although a CM Plan does not contribute directly to the VV&A of a M&S tool, its proper development and implementation assures that the M&S can and will remain accredited over time, quoting from Navy (1999) "A strong CM plan is one of the critical ingredients in ensuring the continued credibility of models and simulations."

The process outlined above has three panels performing the work of the VV&A. This is intended to isolate the panel recommending whether or not the simulation tool should be accredited or not, the ARP, from the individuals performing the computations, the MSP. If it is not felt hat this level of isolation is required, then the process can be simplified by eliminating the SCP. The functions of the SCP would need to be distributed between the ARP and the MSP. As it is unlikely that the AA will have the expertise to make an informed judgment as to the adequacy of an M&S tool, there will need to be an independent panel of subject mater experts between the AA and the MSP, who can advise and make recommendations to the AA. By definition the MSP is not composed of independent individuals, they are experts on the M&S tool being evaluated.

3.2 Formal Accreditation Process

It should be noted that the Flag Administration formal accreditation process for M&S VV&A includes three phases: *designation*, *execution*, and *accreditation*. Preceding these three phases is a *designation process* The *designation process* and *designation phase* are separate activities. The designation process is that process that leads to the selection and formal designation of M&S for accreditation. The designation phase is the initial activity that takes place after the selected model or simulation has been identified for accreditation.

Designation Process

The purpose of M&S VV&A designation is for the user and the owner/developer to agree that the model or simulation selected is capable of satisfying the specified need and that there are sufficient resources to complete accreditation. Each Flag Administration will have specific variations on designating M&S—these guidelines are intended to provide a basic understanding. An external organization, such as a classification society or consulting group, identifies the need to accredit a model or simulation and requests accreditation from the Flag Administration.

The Flag Administration should ensure that an "M&S Accreditation Designation Request Form" be completed and submitted to that Flag Administration. This form will provide the information that is necessary to process the designation request.

Figure 1 provides a process flow diagram for the formal accreditation process, showing the designation, execution, and accreditation phases and their interactions with the Accreditation Authority, Accreditation Review Panel, Simulation Control Panel, and the Modeling and Simulations Proponent. A description of the phases follows.

Designation Phase

During the designation phase, the AA establishes the ARP. The ARP establishes the SCP and documents information from the preceding designation process in an Accreditation Plan. This document will consist of a description of the M&S, an overview of its intended use, M&S requirements and acceptability criteria, the V&V techniques to be used, and the AA's Plan of Action and Milestones (POA&M) for the accreditation effort.

The designation phase is completed when the Accreditation Plan receives AA approval.

Execution Phase

The execution phase of the VV&A process begins with the development of the V&V Plan. The plan should contain the specific qualitative and/or quantitative testing requirements to satisfy the acceptance criteria of the accreditation plan. The SCP provides V&V Plan guidelines to the MSP. These guidelines should consist of an outline, schedule for the execution phase, and clarification of any questions regarding the accreditation plan requirements. V&V Plans may vary greatly based upon previous V&V efforts, the complexity of simulation functionality, length of usage, scope of intended use, and M&S application requirements.

Once the V&V Plan is approved by the SCP, the MSP is tasked with executing that plan. According to the length and complexity of the required V&V, the SCP may have one or more In-Progress Reviews to ensure that the schedule and product development is progressing according to schedule. Prior to completion of V&V testing, the SCP should provide the MSP with guidance for the V&V Report. This guidance should include an outline, inputs on desired formats of information, and distribution formats. When all required V&V efforts and documentation are complete, the MSP provides a final V&V Report to the SCP for evaluation and approval.

The V&V Report should summarize all V&V efforts in accordance with the requirements set forth in the V&V Plan. The SCP can decide to approve the V&V Report with or without modification. As the V&V Report is a critical document in the accreditation process, modification to the report might be necessary to clarify V&V results or to correct deficiencies. Once the V&V Report is approved, the SCP must prepare an Accreditation Report.

The Accreditation Report summarizes the overall V&V execution, provides an assessment of the demonstrated functionality's support of the specific intended use, and makes a recommendation to the ARP for action on the results. This recommendation could be any one of the following:

a. he model or simulation can be used as is for the specific intended use
b. The model or simulation can be used for the specific intended use with recommended modifications
c. The model or simulation requires additional V&V to be considered suitable for accreditation
d. The model or simulation should not be used

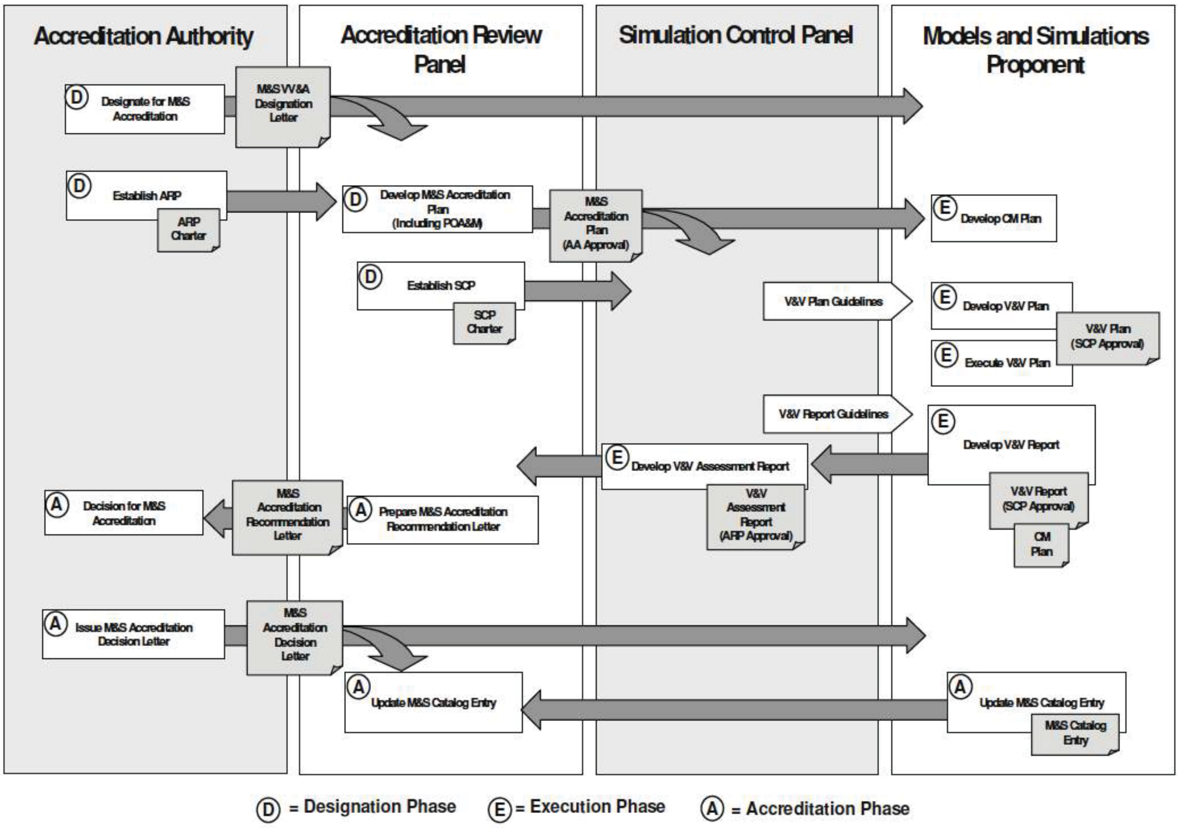

Fig. 1 M&S VV&A Process (Navy, 2005)

for the specific intended use

A major challenge of the VV&A process for a dynamic stability code is that of determining acceptable V&V techniques. The DoD VV&A Recommended Practices Guide provides information and guidance on many V&V techniques and statistical methods. However, they do not seem to be tailored to dealing with the predictions from stochastic processes. Thus the section on Acceptance Criteria that follows.

Accreditation Phase

Upon completion of the Accreditation Report, the ARP evaluates the report for consistency, correctness, and completeness. Once the ARP is satisfied that the V&V information provided meets the stated accreditation requirements, the ARP prepares an M&S Accreditation Recommendation Letter.

This recommendation provides all the M&S information required to support accreditation, such as version and intended use. The AA can approve the recommendation, deny the recommendation, or request additional information. Upon approval by the AA, an M&S Accreditation Decision Letter is sent to the MSP and the ARP. The SCP is dissolved at this time. If the recommendation is denied or if additional information is required, the AA should provide written notification to the ARP and MSP. The SCP may be retained if the ARP decides that further V&V is required for accreditation.

The accreditation remains in effect as long as the intended use or limitations/assumptions of the model or simulation do not change, or until revoked by the AA. If the functionality or the intended use of the model or simulation defined in the M&S Accreditation Decision Letter change, the AA must submit the model or simulation for re-accreditation.

Governing Principles of Accreditation

One governing principal of the accreditation process is to leverage from other VV&A

effort of the Flag Administration (and other Flag Administrations) to the greatest degree possible. Therefore the group seeking accreditation should strive to capture and use other VV&A efforts performed by the Flag Administration. The group seeking accreditation at a minimum should request information about existing VV&A from the applicable Flag Administration(s) and should invite representatives from the Flag Administration to participate in the ARP and/or SCP of the new accreditation effort.

Another governing principle of this process is to place authority in M&S matters consistent with the accountability for the proper use of M&S. M&S is accredited for a specific purpose or a specific use. This specific use or specific purpose drive M&S requirements, which have to be demonstrated by proper V&V techniques before the M&S can be accredited. M&S requirements should be levied on the MSP by the Accreditation Authority. M&S requirements should be imposed on the Flag Administration by IMO.

3.3 Documentation

There are four core documents that are produced during the VV&A process. They are the Accreditation Plan, the V&V Plan, the V&V Report and the Accreditation Report. These documents are produced over time, used at different times by different groups. Thus they must all be complete and independent. As much of the information included in each document is common, it should be shared for consistency and efficiency.

The following material describes the four core reports, it is based on information extracted from: Department of Defense Standard Practice: Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations (DoD, 2012).

The *Accreditation Plan* focuses on: defining the criteria to be used during the accreditation assessment; defining the methodology to conduct the accreditation assessment; defining the resources needed to perform the accreditation assessment; and identifying issues associated with performing the accreditation assessment.

The *V&V Plan* focuses on defining the methodology for scoping the V&V effort to the application and the acceptability criteria; defining the V&V tasks that will produce information to support the accreditation assessment; defining the resources needed to perform the V&V; and identifying issues associated with performing the V&V.

The *V&V Report* focuses on documenting the results of the V&V tasks; documenting M&S assumptions, capabilities, limitations, risks, and impacts; identifying unresolved issues associated with V&V implementation; and documenting lessons learned during V&V.

The *Accreditation Report* focuses on documenting the results of the accreditation assessment; documenting the recommendations in support of the accreditation decision; and documenting lessons learned during accreditation.

Table 1, from DoD (2012), shows the outlines of the four core VV&A documents. The appendices of DoD (2012) provide detailed templates for these four documents.

3.4 Specific Intended Uses

SIUs are the statements that define the scope of the problem or simulation that is to be modeled, and for which the M&S will be accredited. In the context of direct assessment under second-generation intact stability, this will need to include a definition of the vessel for which the M&S tool is to be accredited—accreditation for small fishing vessels may well not apply to a RO/PAX vessel; as well as the mode of stability failure that is anticipated to be an issue. There can, and in fact would likely be multiple SIUs for the same VV&A activity.

The SUIs are used to determine what needs to be characterized and analyzed from the perspective of the V&V process. This is accomplished by the development of a *Requirements Flow-Down Table*. In the Requirements Flow-Down Table, each SIU is decomposed in to several high level requirements (HLRs),

Table 1 Outlines of four core VV&A documents, report sections in *italic* text are common and shared across all four documents. (DoD, 2012)

Accreditation Plan	V&V Plan	V&V Report	Accreditation Report
Executive Summary	Executive Summary	Executive Summary	Executive Summary
1 *Problem Statement*	1 *Problem Statement*	1 *Problem Statement*	1 *Problem Statement*
2 *M&S Requirements and Acceptability Criteria*	2 *M&S Requirements and Acceptability Criteria*	2 *M&S Requirements and Acceptability Criteria*	2 *M&S Requirements and Acceptability Criteria*
3 *M&S Assumptions, Capabilities, Limitations & Risks/Impacts*	3 *M&S Assumptions, Capabilities, Limitations & Risks/Impacts*	3 *M&S Assumptions, Capabilities, Limitations & Risks/Impacts*	3 *M&S Assumptions, Capabilities, Limitations & Risks/Impacts*
4 Accreditation Methodology	4 V&V Methodology	4 V&V Task Analysis	4 Accreditation Assessment
5 Accreditation Issues	5 V&V Issues	5 V&V Recommendations	5 Accreditation Recommendations
6 *Key Participants*	6 *Key Participants*	6 *Key Participants*	6 *Key Participants*
7 Planned Accreditation Resources	7 Planned V&V Resources	7 Actual V&V Resources Expended	7 Actual Accreditation Resources Expended
		8 V&V Lessons Learned	8 Accreditation Lessons Learned
Suggested Appendices A *M&S Description* B *M&S Requirements Traceability Matrix* C *Basis of Comparison* D *References* E *Acronyms* F *Glossary* G Accreditation Programmatics H Distribution List	Suggested Appendices A *M&S Description* B *M&S Requirements Traceability Matrix* C *Basis of Comparison* D *References* E *Acronyms* F *Glossary* G V&V Programmatics H Distribution List I Accreditation Plan	Suggested Appendices A *M&S Description* B *M&S Requirements Traceability Matrix* C *Basis of Comparison* D *References* E *Acronyms* F *Glossary* G V&V Programmatics H Distribution List I V&V Plan J Test Information	Suggested Appendices A *M&S Description* B *M&S Requirements Traceability Matrix* C *Basis of Comparison* D *References* E *Acronyms* F *Glossary* G Accreditation Programmatics H Distribution List I Accreditation Plan J V&V Report

which characterize important aspects of the SIU. The HLRs are each further mapped into several detailed-functional requirements (DFRs). A comparison metric and acceptance criteria are then identified for each DFR.

The SUIs are used to determine what needs to be characterized and analyzed from the perspective of the V&V process. This is accomplished by the development of a Requirements Flow-Down Table. In the Requirements Flow-Down Table, each SIU is decomposed in to several high level requirements (HLRs), which characterize important aspects of the SIU. The HLRs are each further mapped into several detailed-functional requirements (DFRs). A comparison metric and an acceptance criterion are identified for each DFR. Additional clarification is provided by the definition of the comparison metrics and their associated acceptance criteria. High-level requirements reflect the technical specifications provided by SME-opinion. Detailed-functional requirements provide additional specifications as necessary to more fully-describe each HLR. Requirements Flow-Down Tables are useful tools in high-level assessment of the appropriateness of the proposed accreditation criteria as well as required components of the Accreditation Plan (DoD, 2012).

To clarify this, an example of an SIU and its accompanying Requirements Flow-Down Table,

Table 2, are provided. The prototype SIU is:

"The XYZ simulation tool will be used to generate operator guidance polar plots for all applicable speeds and headings against pure loss of stability for RO/PAX vessels in the 11,000–13,000 t displacement range, lengths of 130–150 m, and with beam-to-draft ratios of 4.5 to 5.5. These polar plots will enable the vessel operators to avoid situations where pure loss of stability could be an intact stability issue. The information used to generate the operator guidance polar plots will be developed using numerical data generated by the XYZ simulation tool."

4 VALIDATION APPROACH AND ACCEPTANCE CRITERIA

Following are proposed validation acceptance criteria, which could be applied when seeking accreditation for a numerical simulation tool to be used for direct assessment of stability failure. Two types of accreditation are examined: *Quantitative Accreditation* and *Qualitative Accreditation*. *Quantitative Accreditation* is achieved only if the simulation tool successfully passes all elemental tests and quantitative validation criteria. *Qualitative Accreditation* results from quantified measures of simulation tool accuracy being assessed as "good enough" and is only achieved if the tool passes all elemental tests. For the purpose of this discussion, we treat each type of accreditation as a separate SIU.

The code accreditation is based on comparison to non-rare and rare model-scale data representative of the conditions the vessel would be expected to operate in. It is generally considered that model-scale data captures the relevant physics and scale effects can be accounted for through accepted scaling laws. Utilizing data from multiple scales of models will help to demonstrate the validity of this assumption. Correlation with full-scale trials data will occur prior to certification of the Quantitative Accreditation. Model-scale motion data are collected for a set speed, relative wave heading, and seaway using a model that matches the geometry and anticipated mass properties of the full-scale ship.

Validation is accomplished by comparing statistical properties calculated from model test and simulation data sets for a given speed-heading-seaway combination; these properties are known as condition statistics. Methods for calculating a desired condition statistic from the available data vary depending on the lengths of the motion time histories.

In the case of scale-model test data, run lengths are limited by the size of experimental facilities and statistical properties are calculated from a series of repeated shorter runs. Multiple runs are collected for each speed-heading-seaway combination to form an ensemble of data. The ensemble of data provides enough exposure time (data samples) to accurately represent the statistics of the ship motion at the given speed-heading-seaway combination. Multiple simulation realizations are made at the model-scale test conditions to generate an ensemble of simulation data with the same number of runs and exposure time as the model test.

Non-rare motions will be compared using the motion standard deviation and its uncertainty interval. Rare motions will be compared using the 90th percentile of peak amplitudes and its uncertainty interval. Rather than compare statistically-extrapolated motions for rare motion comparison, the proposed acceptance criteria utilize the most rare motion characteristics of the available model test data which are considered repeatable and not subject to significant variation due to sampling.

Table 2 Example Requirements Flow-Down Table.

High Level Requirements	Detailed Functional Requirement	Comparison Metric	Acceptance Criteria
HLR 1.a Simulation must demonstrate good correlation to model data for ship responses to elemental tests to suggest that underlying physics are sound.	DFR 1.a.1 Simulation must demonstrate the ability to successfully predict critical motion values in a large number of Quantitative Accreditation conditions for which model test data is available for comparison.	CM 1.a.1 Check-list of quantifiable metrics defining "reasonable" correlation for elemental tests used to inform SME opinion	AC 1.a ARP will vote using SME opinion informed by elemental test comparisons whether to assess subsequent acceptance criteria.
	DFR 1.a.2 Collective SME judgment shall ultimately decide whether or not this requirement is met (regardless of the code's ability to meet the suggested quantifiable metrics).	CM 1.a.2 SME opinion/judgment	
HLR 1.b The simulation and model-scale data must show consistently good correlation ranging from the more simple conditions to the more complex conditions. Good correlation must be demonstrated for the range of operational, environmental, and loading conditions defined in the Quantitative Accreditation scope for which comparison model data are available.	DFR 1.b.1 Parameters which characterize the ship's operating condition relative to the seaway, and identify the corresponding critical motion, must be assessed.	CM 1.b.1 Mean values, μ, of achieved speed and heading	AC 1.b.1 Differences between mean achieved speed and mean achieved heading for each validation condition must be less than specified amounts.
	DFR 1.b.2 All comparisons must take into account all known sources of uncertainty (sampling, instrument, condition, etc.).	CM 1.b.2 90% uncertainty intervals on the each parameter (model and simulation)	AC 1.b.2 The 90% confidence intervals on each parameter value (and A90%) for a given motion and condition must overlap in order to suggest that the underlying populations (model and simulation) may be the same.
	DFR 1.b.3 Parameters that are used to define Quantitative Accreditation polar plots risk values and lifetime risk calculation must be assessed. If direct validation of these quantities is not achievable, a sufficient substitute quantity shall instead be assessed. (rare motion metrics)	CM 1.b.3 The 90th percentile of peak amplitudes, A90%, of motions (in lieu of exceedance rates of physical limit thresholds which are not expected to be available for validation)	
	DFR 1.b.4 Parameters that are used to evaluate the Quantitative Accreditation system health must be assessed. (non-rare motion metrics)	CM 1.b.4 Mean standard deviation, σ, of motions	

Table 2 (Cont'd) Example Requirements Flow-Down Table.

High Level Requirements	Detailed Functional Requirement	Comparison Metric	Acceptance Criteria
HLR 1.c Necessary accuracy of the simulation shall be influence by an appropriate balance between technical excellence and judiciousness	DFR 1.c Thoughtful engineering judgment shall be applied in the determination of permissible differences between simulation and model test results.	CM 1.c Margin applied to observed sample parameter values (defined in CM 1.b.2 and CM 1.b.3)	AC 1.c The observed values of compared sampled parameters may be deemed acceptable if the difference between the values is less than a specified amount. (margin)
HLR 1.d The safety of the ship and sailor must be prioritized and reflected in the criteria established for validation.	DFR 1.d.1 Reasonable conservatism on the part of the simulation solution should be endorsed to promote the overall safety of the sailor.	CM 1.d.1 Margin applied to observed sample parameter values (defined in CM 1.b.2 and CM 1.b.3)	AC 1.d.1 The margin allowed by AC 1.c shall be increased by 50% in the case of over-prediction on the part of the simulation to allow for additional conservatism on the part of the simulation. (additional conservative margin)
	DFR 1.d.2 Determination of simulation tool success must only be reached using reasonably high-fidelity validation data sets.	CM 1.d.2 Combined uncertainty in the comparison, calculated as a function of the 90% uncertainty intervals (CM 1.b.2) on both data sets, model and simulation	AC 1.d.2 Successful validation comparisons for both rare and non-rare motions (and A90%) may only be accepted if the combined uncertainty in both data sets is sufficiently small.
HLR 1.e Simulation must be deemed usable for conditions within the current scope of the Quantitative Accreditation for which comparison model test data is not available.	DFR 1.e.1 Simulation must demonstrate the ability to successfully produce critical motion values in a large number of Quantitative Accreditation conditions for which model test data is available for comparison.	CM 1.e.1 Number of conditions which successfully pass the following criteria: AC.1.b.1 through AC 1.d.	AC 1.e 70% of Quantitative Accreditation conditions for which model data are available for comparison must pass criteria (AC 1.a through AC 1.d) for 100% of critical motion parameter values. (rare and non-rare motion assessments calculated independently)

4.1 Elemental Tests

Elemental tests (or comparisons to model data) provide insight into the code's ability to capture the overall physics of the ship motion problem. They also provide confidence that the quantitative comparison results obtained with available model data may be assumed characteristic of the code and applicable for similar conditions for which model data is not available. The results of the elemental tests provide evidence to the ARP to inform their final decision making. Subject matter experts on the SCP will provide the ARP with general guidelines about the comparisons; this guidance will include both qualitative and quantitative characteristics of good correlation.

The code will simulate the following elemental tests in support of validation:

- Roll decays
- Zig-zag maneuvers
- Calm water turning circles
- Turning circles in regular waves
- Acceleration from rest tests
- Generation of response amplitude operators (RAO) for comparison with model data (if available)
- Integrity values

Standard maneuvering and seakeeping analyses of the time histories will be performed on the code and model data time histories in order to provide comparison quantities for SCP guidance. Integrity values will be plotted on polar and surface plots to investigate the code's ability to capture the ship's capsize boundary. An integrity value is a ratio between the number of runs which did not include a dynamic stability event divided by the total number of runs examined. This metric allows for comparisons between model test and simulation in which the ship response is highly sensitive to initial conditions. Since the initial conditions under which each model test was performed cannot be known precisely, a range of simulations is performed in an attempt to cover the range of possibilities.

This elemental test is included on the list above to specifically address the known dynamic stability concerns associated with a ship operating in stern quartering seas. Characterization

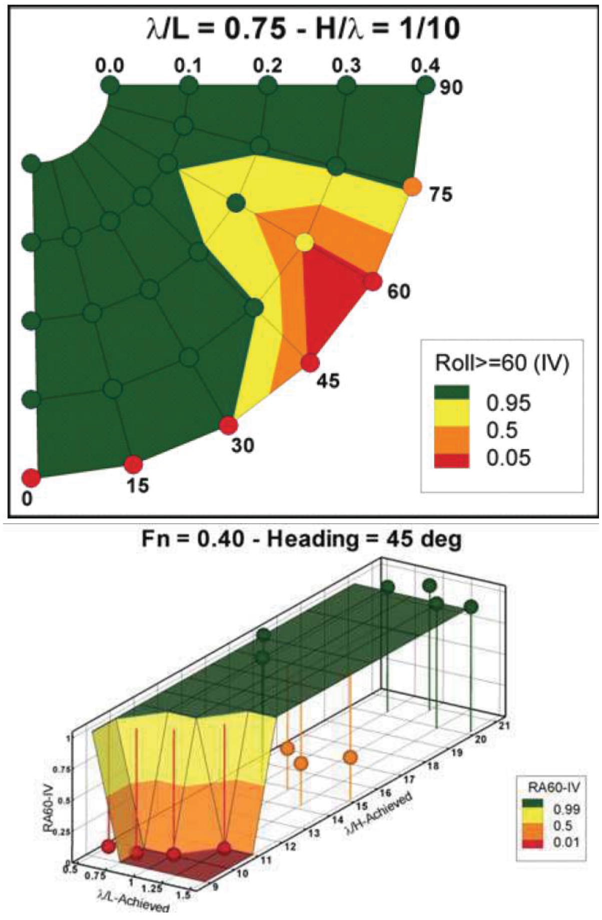

Fig. 2 Notional Integrity Value Polar Plot (top) and Surface Plot (bottom)

of the ship's response in these conditions from irregular seas model data is challenging, so integrity value plots (using regular waves model test results) provide the necessary additional insight into the code's ability to capture this aspect of the physics. Figure 2 shows an example of integrity value surface and polar plots.

4.2 Quantitative Validation

Beyond successful demonstration that the general ship motion physics are captured by the code, it will be assessed for its suitability for each of the specific intended uses. These assessments are quantitative in nature, although ARP opinion will ultimately be included in all final accreditation recommendations. Following are recommended quantitative acceptance criteria for Quantitative Accreditation and Qualitative Accreditation.

Definitions

The acceptance criteria described in this section for Quantitative Accreditation utilize statistical quantities and their uncertainty intervals calculated for a single motion and condition (speed, heading, wave height, wave period); these quantities are referred to as condition statistics.

Scale-model tests are characterized by multiple repeated runs of short run lengths. For each comparison to model data, an equivalent number of runs and run durations will be performed by the code. The condition statistics will be calculated from the model data time histories and the code time histories in the same manner. The condition statistic varies by SIU and rare or non-rare motion. The statistical quantities examined are: condition standard deviation (non-rare motion), condition 90th percentile amplitude (rare motions), and condition mean.

Mean values of speed and heading are used to compare the results of achieved speed and heading in a seaway. Standard deviation values are used to compare non-rare motion responses. 90th percentile of peaks values are used to compare rare motion ship responses. Direct assessment of very rare ship motions is typically prohibited by the limitations of available model test data, and this condition statistic was selected as the peak amplitude threshold for comparison because analysis has suggested that it is the highest motion magnitude (most rare quantity) that is statistically stable for typical model data sets. Higher percentiles of the peaks showed great variation in repeated simulations, suggesting that statistical sampling combined with the non-linear system led to instability in the values above the 90th percentile provides the analysis used to determine this threshold. Figure 3 illustrates relationship between peak distributions and percentiles of peaks for two data sets.

Uncertainty associated with the value of the condition statistic (mean, standard deviation, or percentile) is captured by intervals applied about the condition statistic. The size of these intervals is influenced by sampling statistics, instrumentation uncertainties, and variations in the conditions under which the model was tested.

Uncertainty due to statistical sampling is captured by a confidence interval. The confidence interval is a conventional mathematical quantity which NIST (2014) defines as a range of values which is likely to contain the population parameter of interest. Its purpose is to account for the possible difference between a discreet value derived from limited population samples from the underlying population value. The level of confidence associated with the interval defines its length and corresponds to the probability that the sampled value and intervals encompass the true population value. When defined relative to a mean value and assuming a large sample size, the confidence interval is defined as

$$CI_{\mu} = z_{1-\alpha/2}\frac{\sigma}{\sqrt{N}}$$

where σ is the sample standard deviation, N is the number of samples, α is the desired significance level (corresponds to confidence level), and z is the two-tailed Gaussian distribution factor with significance level, α. The upper and lower bounds of the confidence intervals applied to the sample mean are defined as

$$\mu_{\text{sample}} \pm CI_{\mu}$$

where μ_{sample} is the sample mean. Belenky, *et al.* (2013) provides an extension of this theory to calculate the confidence interval on the ensemble mean standard deviation value from a set of time histories of ship motions for one parameter and one condition. Calculation of the confidence interval for a quantile or percentile is a standard statistical process, which utilizes the binominal distribution.

It should be noted that the terms “confidence” and “uncertainty” are often used interchangeably. This document uses the term uncertainty to include all sources of uncertainty. The confidence level is 90-percent for comparisons involving confidence intervals. Figure 4 shows the relationship between the condition statistic value, intervals and uncertainty limits used in motion comparisons.

The difference between condition statistics

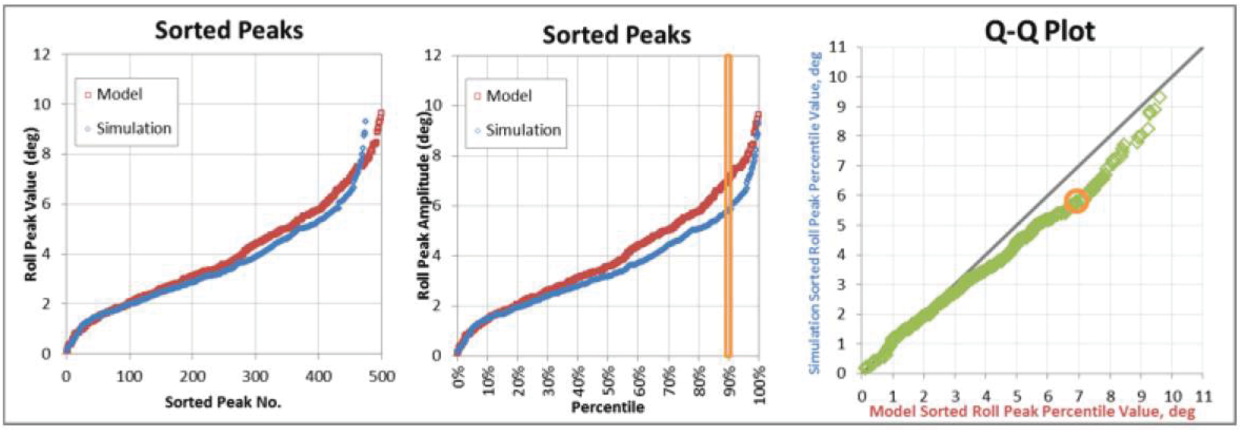

Fig. 3 Sorted Peak Amplitudes for Two Data Sets [by number (left), by percentile (center), percentiles plotted against one another (right)]

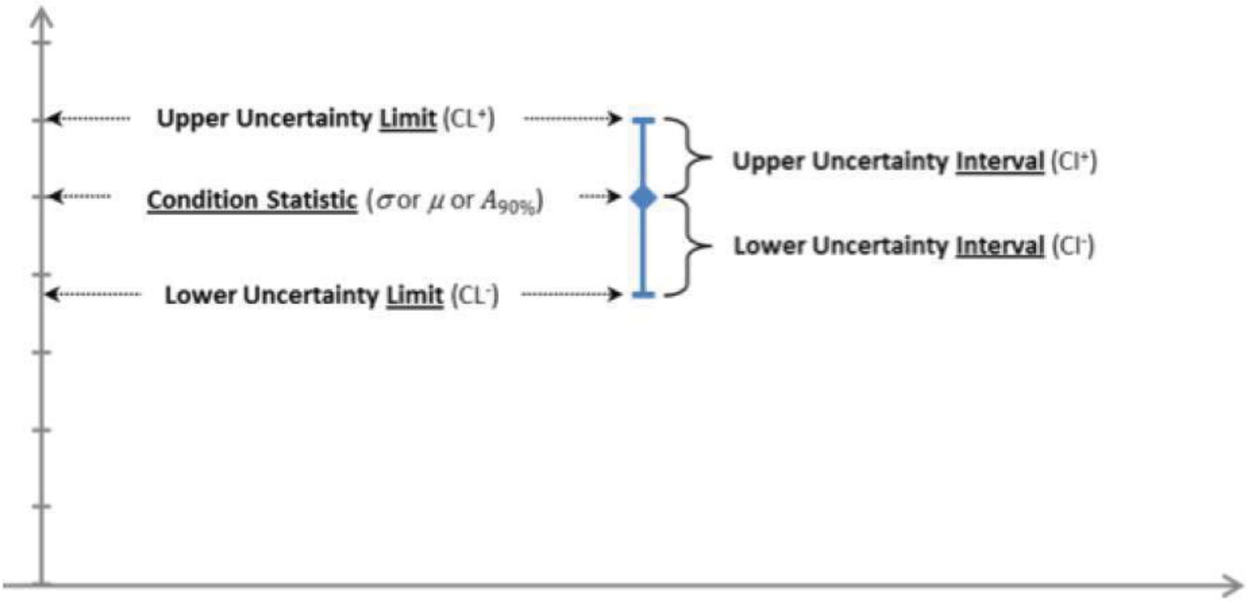

Fig. 4 Metric Nomenclature (condition statistic, interval, and limit)

is the primary metric used for quantitative validation and is defined as the model test value subtracted from the simulation value. A positive value is associated with simulation overprediction, and a negative value denotes simulation under-prediction. This concept is certainly not new to the field of validation, but its use is often associated with largely deterministic processes. The use of the difference between data sets as a foundation for validation acceptance criteria is consistent with industry practice. (*cf.* Oberkampf & Barone, 2006; AIAA, 1998; ASME, 2009; Eça & Hoekstra, 2012). Both Oberkampf & Barone (2006) and ASME (2009) refer to this quantity as the error between model and experimental results, noting that the experimental results are only an estimated measure of the "true" parameter value.

The confidence interval on the difference between condition statistic values of a model and simulation result can be formulated as a function of the confidence intervals on each set. The confidence interval on the difference between mean values is defined as

$$CI_{\Delta\mu} = z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{N_1}+\frac{\sigma_2^2}{N_2}} \tag{1}$$

where the subscripts 1 and 2 distinguish between data sets.

Additional sources of uncertainty may be applicable to the sample value, including uncertainty due to instrumentation limitations and uncertainty due to variability of the conditions under which the data was generated. Combined uncertainty intervals constructed from multiple sources of uncertainty are typically the root sum of squared intervals calculated separately for each source. While confidence intervals (based only on sampling characteristics) are symmetric, combined uncertainty intervals may be asymmetric.

To compare two data sets with equal number of samples (i.e. $N_1 = N_2$) and symmetric confidence intervals, (1) can be rearranged and described in terms of the confidence intervals associated with each data set value as

$$CI_{\Delta\mu} = z_{1-\alpha/2}\sqrt{\left(\frac{CI_{\mu\_1}}{z_{1-\alpha*/2}}\right)^2 + \left(\frac{CI_{\mu\_2}}{z_{1-\alpha*/2}}\right)^2} \quad (2)$$

where $\alpha*$ refers to the level of significance associated with the sample intervals and α refers to the level of significance associated with the uncertainty in the difference.

Equation (2) lends itself to a definition of the combined uncertainty (e.g. statistical, instrument, etc.) in the difference between samples which is agnostic to the methods used to define the combined uncertainty intervals associated with each data set, assuming the uncertainties of each set are Gaussian distributed. Further, (2) can be adapted to account for asymmetric intervals by distinguishing between the upper and lower intervals associated with each set.

For validation purposes, consider the definition of the difference (simulation minus benchmark) to compare two ensemble mean standard deviation quantities. Given combined uncertainty intervals associated with each data set of significance level $\alpha*$, the upper and lower combined uncertainty intervals on the difference

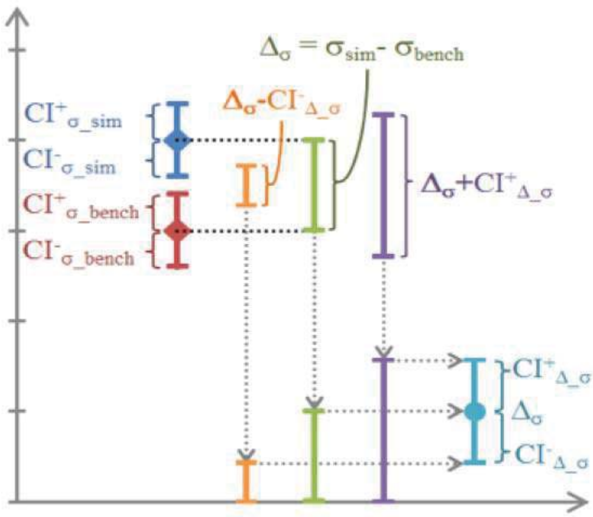

Fig. 5 Uncertainty Intervals On Two Data Sets and On the Difference Between Data Sets

can be calculated as

$$CI_{\Delta} = z_{1-\alpha/2}\sqrt{\left(\frac{CI^-_{\text{bench}}}{z_{1-\alpha*/2}}\right)^2 + \left(\frac{CI_{\text{sim}^+}}{z_{1-\alpha*/2}}\right)^2}$$

and

$$CI_{\Delta} = z_{1-\alpha/2}\sqrt{\left(\frac{CI^+_{\text{bench}}}{z_{1-\alpha*/2}}\right)^2 + \left(\frac{CI_{\text{sim}^-}}{z_{1-\alpha*/2}}\right)^2}$$

where the subscripts "bench" and "sim" refer to the benchmark (or model test) and simulation data sets, respectively. Figure 5 illustrates the relationships between the uncertainty intervals on both data sets and the uncertainty interval on the difference. The formulation of the confidence interval on the difference based on the confidence intervals on both samples is applicable to comparisons of mean, standard deviation, and amplitude percentile quantities.

The combined uncertainty intervals surrounding a difference between simulation and benchmark statistics enclose the region within which the "true" difference between populations is found. The level of confidence associated with interval calculations corresponds to the probability that the true difference is within the interval limit. For a 90-percent level of confidence, there is a 90-percent probability that the difference between the simulation and benchmark results is between the lower and the upper interval extents.

Positive values denote a simulation value which is greater than the benchmark (over-prediction) while negative values denote under-prediction. A zero-crossing of an interval denotes the possibility that there is no difference between the underlying. It should be noted, however, that the confidence level associated with the interval does not equal the probability that the difference is zero. In fact, there is equal likelihood that the true difference falls anywhere else within the interval extents.

As noted above, when the uncertainty interval on the difference crosses zero, there may be no difference between the two populations. A zero-crossing of difference intervals is most analogous to an overlap of uncertainty intervals associated with two data sets. Note, however, that zero-crossing is a more "strict" measure of similitude than interval overlap. For the same level of significance, it is mathematically possible for the intervals to slightly overlap without the corresponding interval on the difference crossing through zero.

A particularly useful attribute of the difference between statistics is its ability to convey information about the simulation's accuracy for a given parameter across a range of conditions. This utility forms the foundation of acceptance criteria for quantitative validation.

4.3 Quantitative Accreditation (Acceptance Criteria)

The Quantitative Accreditation acceptance criteria are a tiered series of *channel*, *condition*, and *code* criteria. An evaluation of each critical motion is made to assess a speed-heading-seaway condition. The *channel criteria* are applied to the statistical properties calculated from model test and simulation time histories. The *condition criteria* are applied to the results of the *channel criteria* for each unique environmental and operational condition combination within the validation data domain space. Finally, the *code criteria* are applied to the results from the *condition criteria* to determine the final accreditation outcomes. The code acceptance is based on passing over 70-percent of the conditions.

Figures 6 and 7 provide an overview of acceptance criteria for Quantitative Accreditation of non-rare motions and rare motions, respectively.

Channel Criteria

Condition statistics (standard deviation and 90th percentile values) calculated from model and simulation time histories are used (with their associated uncertainty intervals) to assess the code's ability to provide the required non-rare and rare motion ship response. The motions listed in Figures 6 and 7 are considered "critical channels" for assessment of intact stability-related motions. Channel criteria are defined relative to a physical limit value for each motion. Physical limit definitions may be tailored to address ship-specific hull and machinery requirements. Yaw and yaw rate physical limits are defined relative to the definition of a broach.

Condition statics and uncertainty intervals for both model and simulation data sets are calculated for a single motion and condition from the respective sets of time histories of the motion. The difference between condition statistics (including uncertainty) is then calculating from the results of both data sets

Ordered values of ship speed and heading identify the ship's operational environment for each condition. The average (mean) achieved values of speed and heading resulting from the ordered values and the ship's response to the seaway influence the ship's motions response. Condition mean values are determined from time histories of both simulation and model tests and are represented by the variable, μ.

The channel criteria are applied to the critical motions as four tests (referred to as Four Box criteria) which result in a "pass," "fail," or "null" conclusion. Figure 8 illustrates the relationship between the Four Box criteria and the determination of the motion comparison for both non-rare and rare channel criteria. Figure 9 shows an example (roll standard deviation) of the relationship between condition statistic difference

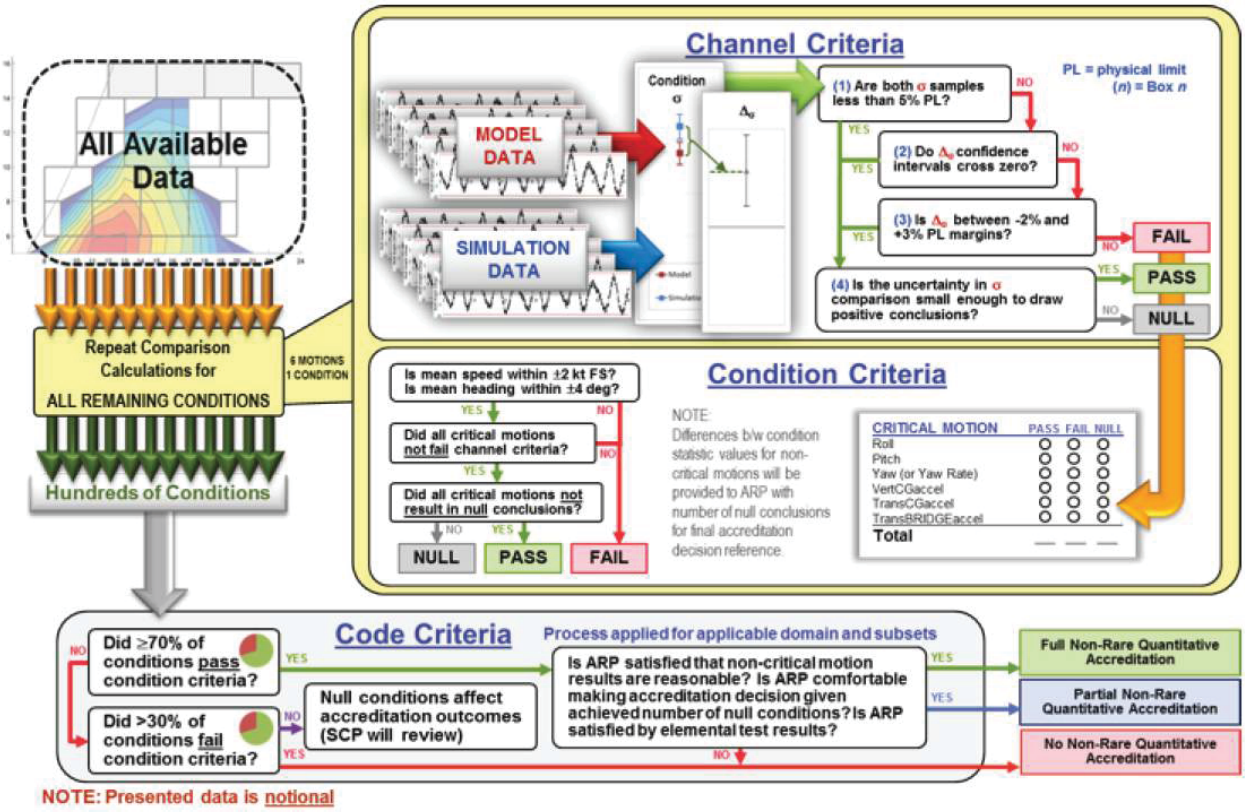

Fig. 6 Acceptance Criteria for Quantitative Accreditation Support (Non-Rare Motions)

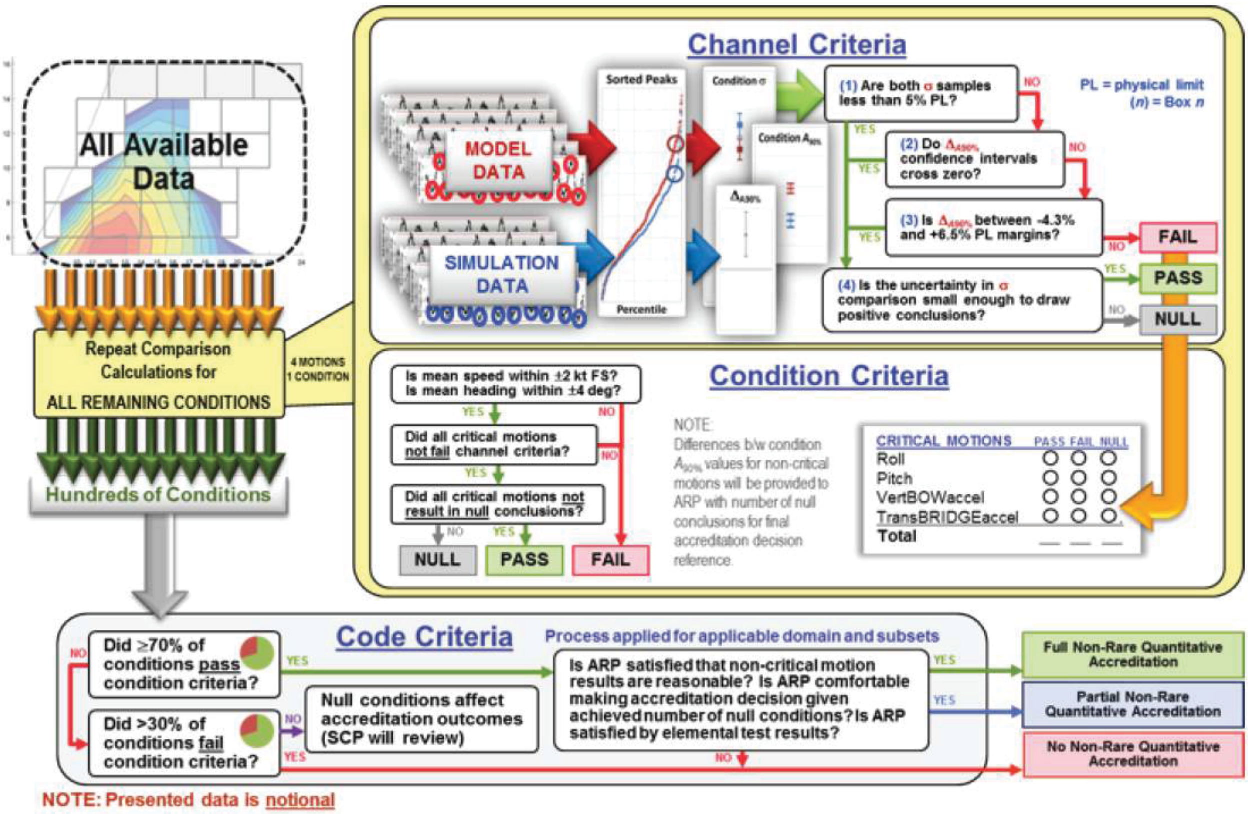

Fig. 7 Acceptance Criteria for Quantitative Accreditation Support (Rare Motions)

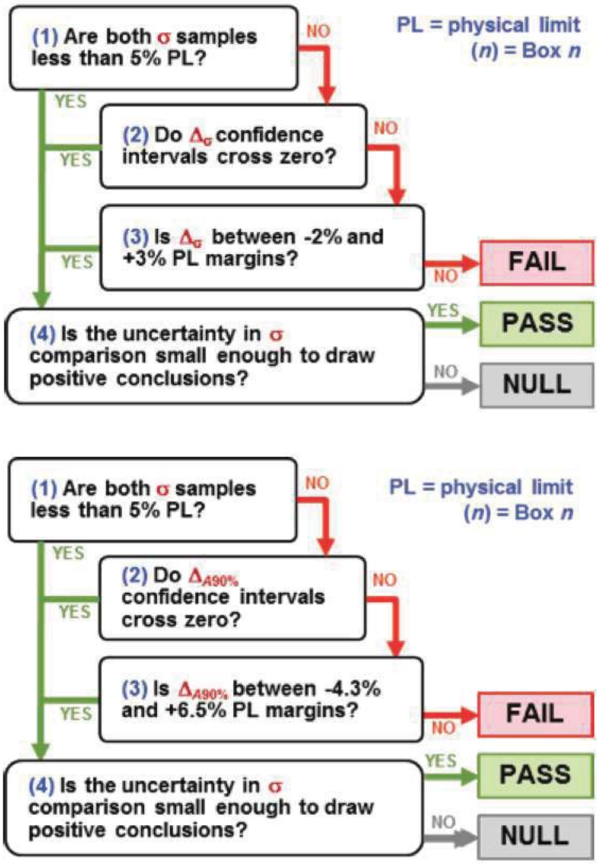

Fig. 8 Overview of Channel Criteria for Quantitative Accreditation for Non-Rare (top) and Rare (bottom) Motions

values and the four-box channel criteria.

Box 1: Very Small Motions

The Box 1 criterion is met if both the model and simulation condition σ-values are less than 5-percent of the physical limit. Passing the Box 1 criterion indicates that the motions are sufficiently small to pose no significant risk to ship operations.

Box 2: Zero Crossing of the Difference Uncertainty Interval

The Box 2 criterion is met if the uncertainty intervals about the difference between condition statistics passes through zero. Demonstration of a zero-crossing indicates a non-negligible statistical probability that the two condition statistics (model and simulation) may come from the same distribution and may be statistically the same (i.e. zero difference).

Box 3: Samples Within Margins

The Box 3 criterion is met if the model and simulation condition statistics differ by a permissibly small amount, or margin. The sample margins are conservatively biased; greater differences are allowed for over-prediction than for under-prediction. The margin values for non-rare motion comparisons are 3-percent of the physical limit for simulation over-prediction and 2-percent of the physical limit for simulation under-prediction. The margin values for rare motion comparisons are 6.5-percent of the physical limit for simulation over-prediction and 4.3-percent of the physical limit for simulation under-prediction. The margin values applied to the condition 90th percentile values are the non-rare motion margins multiplied by 2.15. This factor is based on the relationship between standard deviation and the 90th percentile of peaks for the Rayleigh distribution. Passing the Box 3 criterion addresses cases where the uncertainty intervals are small, but the condition statistic values are sufficiently similar to one another for practical purposes.

Box 4: Limitations on Uncertainty

The Box 4 criterion is met if the overall uncertainty in a comparison is less than a specified amount based on statistical Type II error (accepting what should be rejected due to too much uncertainty). The following equation presents the simplified numerical criterion for this test in terms of the confidence intervals on each data set.

$$\sqrt{(CI_{\sigma_{\text{model}}})^2 + (CI_{\sigma_{\text{code}}})^2} < 5\% \text{ of the physical limit}$$

Note that the characteristic interval length for each data set should be taken as the average of the upper and lower intervals if the intervals are asymmetric.

Failure of the Box 4 criterion does not signify a deficiency on the part of the simulation. Rather, failure of the Box 4 criterion denotes a comparison of poor quality from which no positive conclusions may be drawn.

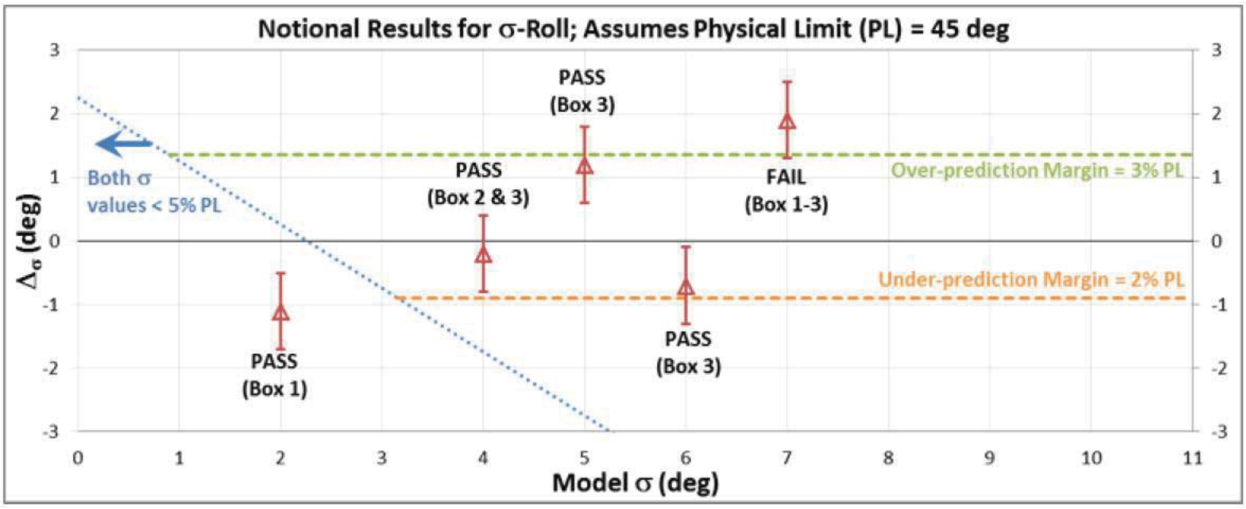

Fig. 9 Illustration of Channel Criteria for Quantitative Accreditation

Condition Criteria

Three outcomes are possible for the condition criteria: “pass,” “fail,” and “null.” The condition criteria are passed if the differences between mean speed and heading are permissibly small and 100-percent of the critical channels pass the channel criteria. The condition criteria are failed if the mean speed or heading differences are excessively large or one or more channels within a condition fail the channel criteria. The condition criteria results in a null conclusion if all of the following criteria are met: 1) mean speed and heading differences are permissibly small, 2) no motions fail the channel criteria, and 3) one or more motions result in a null conclusion of the channel criteria. Figure 10 illustrates the relationship between the condition criteria and the possible outcomes.

The simulation must demonstrate the ability to sufficiently model the mean speed and heading of the condition. The condition mean achieved model and simulation values of speed over ground and heading must fall within 2 knots full-scale and 4 degrees, respectively. Note that these limits should be tailored (based on ship speed and natural pitch and roll periods) to limit permissible deviation from wave encounter frequency.

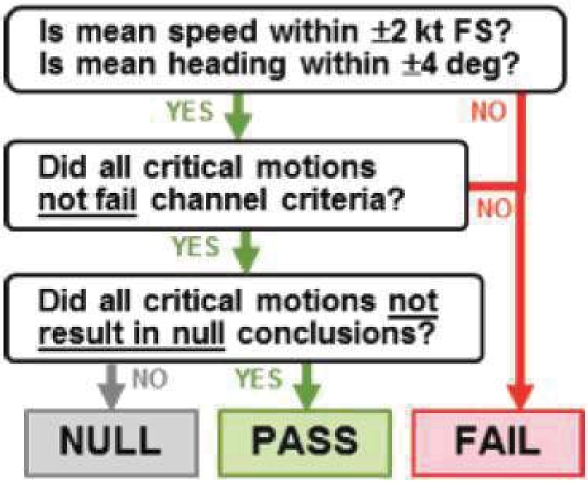

Fig. 10 Condition Criteria for Quantitative Accreditation (Rare and Non-Rare Motions)

Code Criteria

The code will pass the quantitative criteria for either rare or non-rare motions if at least 70-percent of conditions pass the respective quantitative condition criteria. The code will fail the code criteria for either rare or non-rare motions if more than 30-percent of the conditions fail the respective quantitative condition criteria. Otherwise, further review by the SCP is required due to the influence of null conditions on pass/fail outcomes. Further, the ARP must be satisfied with the percentage and locations within the domain space of non-null conditions ultimately available for the code comparison. Table 3 summarizes the quantitative code criteria, which are applied separately for rare and non-rare results.

The ARP must also be satisfied by the accuracy reports for the non-critical rare and non-rare motions (not included in the channel comparisons). A description of the accuracy reports calculated for these channels is given below in the Qualitative Accreditation section.

The 70-percent criterion will be applied, and accreditation recommendations determined by the ARP, for non-rare motions across the following domain spaces:

- Across domain space
- Across defined operational conditions (speed and heading combinations)
- Across defined environmental conditions (wave height and period combinations)

4.4 Qualitative Accreditation

Qualitative Accreditation recommendations for the code's ability to simulate non-rare and rare motions is accomplished by generating accuracy reports (indicating differences between simulation and model results) for each channel across the relevant domain spaces, following the methodology presented in Zuzick, *et al.* (2014). Figures 11 and 12 provide an overview of the non-rare motion and rare motion, respectively, Quantitative Accreditation validation process. Statistical properties and the differences between these values are calculated from model test and simulation time histories. These values are calculated for each motion and unique environmental and operational condition combination within the validation data domain space. Finally, measures of overall accuracy are calculated from the observed difference values and provided to the ARP in the accuracy report.

The main difference between Qualitative and Quantitative Accreditation is the result of the effort. While Quantitative Accreditation provides "pass", "fail", or "null" outcomes to comparisons, Quantitative Accreditation provides statements about the simulation tool's accuracy (e.g. "The simulation over-predicts roll by 1.5 degrees across the validation domain."). These quantified measures of accuracy are contained in accuracy reports and can be used to establish margins on simulation results for ship-specific operator guidance generation.

Accuracy Reports

Qualitative Accreditation results in quantified measures of accuracy of critical and non-critical rare and non-rare motions results produced by the simulation tool across the domain and for subsets of the domain. For Qualitative Accreditation accuracy reporting, the 90-, 95- and 99-percent confidence intervals will each be calculated on the difference. The condition statistics examined through accuracy reports are standard deviation (for non-rare motions), 90th-percentile of amplitude peak (for rare motions), and mean values (for achieved speed and heading).

In addition to calculating the difference between condition statistics, the percent difference between values (difference divided by the model data condition statistic) will be calculated for each motion and condition. Within the maneuvering and seakeeping simulation community, a 20-percent difference (or smaller) is a generally-accepted measure of good correlation of standard deviation. The ARP will be provided with the percentage of channels compared whose percent difference was less than or equal to 20-percent as an additional measure of the code's overall accuracy.

To quantify the code's overall ability to capture rare and non-rare motions, generalized accuracy reports will be generated for each motion using the differences (and associated uncertainties) between the code and model test condition statistics over a range of conditions.

Figure 13 provides a notional representation of a non-rare and rare motion accuracy report for one mode of motion. Each accuracy report will contain the following quantities:

- Arithmetic mean of the difference (including arithmetic means of upper and lower uncertainty limits)
- Weighted mean of the difference (including weighted means of upper and lower uncertainty limits)

Table 3 Quantitative Code Criteria

PASS	FAIL	NULL	Comment
70%	N/A	N/A	Code Passes
N/A	> 30%	N/A	Code Fails
< 70%	< 30%	> 0%	Further examination of null conditions

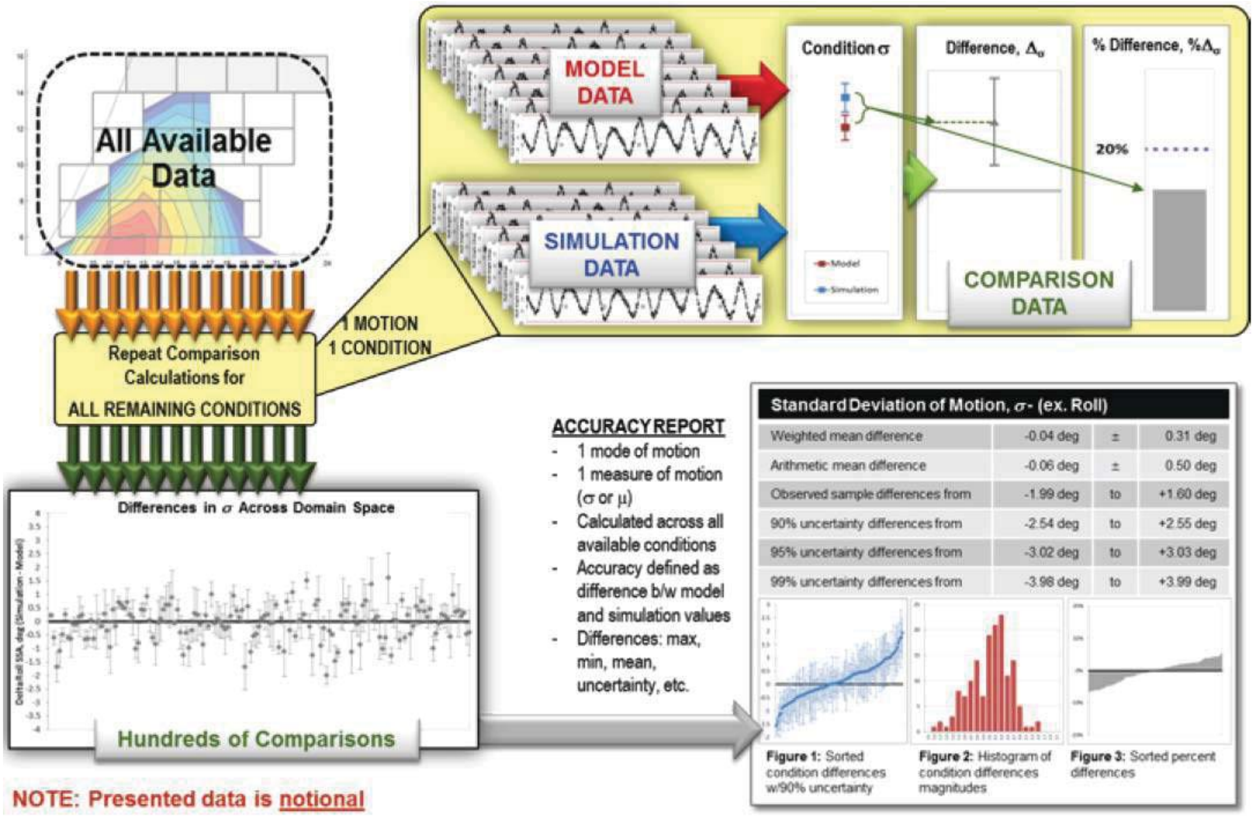

Fig. 11 Acceptance Criteria (Accuracy Reports) for Qualitative Accreditation Support (Non-Rare Motions)

  – Weighting of each comparison condition is determined by the inverse of the combined length of the uncertainty intervals

- Range of observed sample differences
- Range of observed upper and lower uncertainty limits for 90%, 95%, and 99% confidence intervals
- Plot of sample differences (including 90-percent uncertainty limits) sorted from smallest to largest sample differences
- Histogram of sample difference magnitudes
- Quantile-quantile plot of motion peak amplitudes showing all conditions in the domain

A non-rare and rare motion accuracy report will be generated for each motion using individual comparison results from conditions categorized by several domain spaces. Quantified measures of accuracy will be calculated for each motion for the following domains:

- Across domain space
- Across defined operational conditions (speed and heading combinations)
- Across defined environmental conditions (wave height and period combinations)

5 CONCLUSIONS

With the advent of the second-generation intact stability criteria, IMO has initiated a two-tier

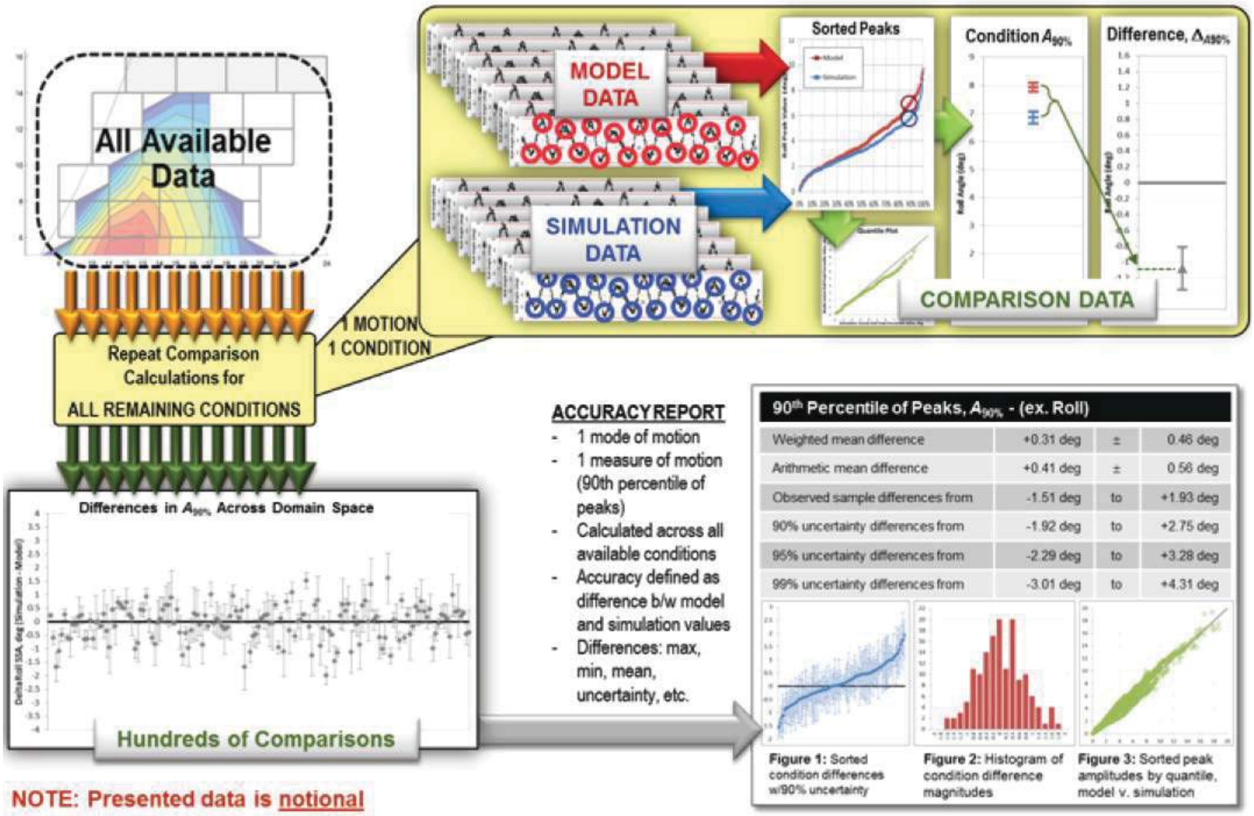

Fig. 12 Acceptance Criteria (Accuracy Reports) for Qualitative Accreditation Support (Rare Motions)

performance-based stability assessment process for unconventional hulls with a risk of intact stability failure. The first tier has two levels where simplified physics-based algorithms are used to assess a design. If the design fails the first level test, which is very simple but quite conservative, the design is then assessed using the second level criteria. The second level test is also simple, but it is more involved and less conservative than the first level method. If the design fails these first tier evaluations, it then progresses to the second tier, where direct assessment criteria are applied.

The design is considered satisfactory if the direct assessment criteria are passed. If these criteria are not passed, operator guidance is needed to provide vessel operators with the information needed to safely operate the vessel in dangerous conditions. Ship motion simulation tools are needed to apply the direct assessment criteria and generate operator guidance, if necessary.

A framework is presented for certification that simulation tools used for direct assessment of stability failures and generation of operator guidance are sufficiently accurate for these purposes. Based on US Navy experience, guidance is provided on the VV&A process, structure, and participation, and acceptance criteria are given for both quantitative and qualitative accreditation approaches. Accreditation acceptance criteria are tailorable to ship-specific VV&A efforts, particularly with regards to definition of critical motions and physical limits.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Philip Alman of NAVSEA, and David Drazen, Christopher Kent, Joel Park and Timothy Smith of DTMB for their many hours of discussion and contributions relating to acceptance criteria. We would also like to thank Vadim Belenky, William Belknap and Brad Campbell of DTMB for their discussions on the VV&A process—from both a Navy and IMO perspective, during

Standard Deviation of Motion, σ - (ex. Roll)			
Weighted mean difference	-0.04 deg	±	0.31 deg
Arithmetic mean difference	-0.06 deg	±	0.50 deg
Observed sample differences from	-1.99 deg	to	+1.60 deg
90% uncertainty differences from	-2.54 deg	to	+2.55 deg
95% uncertainty differences from	-3.02 deg	to	+3.03 deg
99% uncertainty differences from	-3.98 deg	to	+3.99 deg

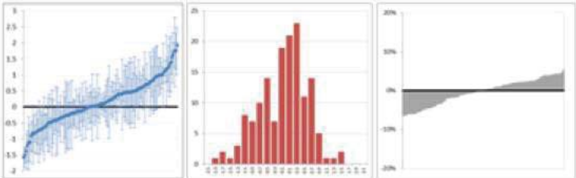

Figure 1: Sorted condition differences w/90% uncertainty **Figure 2:** Histogram of condition differences magnitudes **Figure 3:** Sorted percent differences

90th Percentile of Peaks, $A_{90\%}$ - (ex. Roll)			
Weighted mean difference	+0.31 deg	±	0.46 deg
Arithmetic mean difference	+0.41 deg	±	0.56 deg
Observed sample differences from	-1.51 deg	to	+1.93 deg
90% uncertainty differences from	-1.92 deg	to	+2.75 deg
95% uncertainty differences from	-2.29 deg	to	+3.28 deg
99% uncertainty differences from	-3.01 deg	to	+4.31 deg

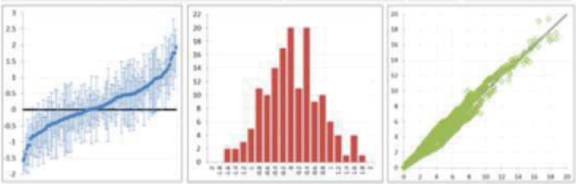

Figure 1: Sorted condition differences w/90% uncertainty **Figure 2:** Histogram of condition difference magnitudes **Figure 3:** Sorted peak amplitudes by quantile, model v. simulation

Fig. 13 Notional Non-Rare (top) and Rare (bottom) Motion Accuracy Report for Roll (NOTE: Data presented is notional)

the preparation of this paper. Finally, we would like to thank Joe Gorski of DTMB for his thorough review of the paper.

REFERENCES

AIAA (1998) Guide for the Verification and Validation of Computational Fluid Dynamics Simulations. AIAA G-077-1998 Guide, Reston, VA: American Institute of Aeronautics and Astronautics, viii+19 p.

ASME (2009) V&V 20-2009, Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer. New York, NY: American Society of Mechanical Engineers, x+87 p.

Beck, R. F., A. M. Reed & E. P. Rood (1996) Application of modern numerical methods in marine hydrodynamics. *Trans. SNAME*, 104:519–37, Jersey City, NJ.

Belenky, V., J. O. de Kat & N. Umeda (2008a) Towards Performance-Based Criteria for Intact Stability. *Marine Tech.*, 45(2):101–123.

Belenky, V., V. Pipiras, C. Kent, M. Hughes, B. Campbell, T. Smith (2013) On the Statistical Uncertainties of Time-domain-based Assessment of Stability Failures: Confidence Interval for the Mean and Variance of a Time Series. *Proc. 13th Int'l Ship Stability Workshop*, Brest, France, 8 p.

Belenky, V. L. & N. B. Sevastianov (2007) *Stability and Safety of Ships: Risk of Capsizing* (2nd ed.). SNAME, Jersey City, NJ, xx+435 p.

Belenky, V., K. M. Weems, W-M. Lin (2008b) Numerical Procedure for Evaluation of Capsizing Probability with Split Time Method, *Proc. 27th Symp. Naval Hydro.*, Seoul, Korea, 25 p.

Clauss, G. (2008) The Taming of the Shrew: Tailoring Freak Wave Sequences for Seakeeping Tests. *J. Ship Res.*, 52(3):194–226.

DoD (1998) DoD Modeling and Simulation (M&S) Glossary. DoD 5000.59-M, U. S. Department of Defense, 175 p.

DoD (2003) DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A). DoD Instruction 5000.61, U. S. Department of Defense, 10 p.

DoD (2007) DoD Modeling and Simulation (M&S) Management. DoD Directive 5000.59, U. S. Department of Defense, 7 p.

DoD (2012) Department of Defense, USD (AT&L) MIL-STD-3022 w/Change 1: Department of Defense Standard Practice: Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations, 55 p.

Eça, L. & M. Hoekstra (2012) Verification and Validation for Marine Applications of CFD, Gothenburg, Sweden: *Proc. 29th Symp Naval Hydro.*

Francescutto, A. (2004) Intact Ship Stability—The Way Ahead. *Marine Tech.*, 41:31–37.

Francescutto, A. (2007) Intact Stability of Ships-Recent Developments and Trends. *Proc. 10th Int'l Symp. Practical Design of Ships and Other Floating Struct* (PRADS '07). Houston, TX, Vol. 1, pp. 487–496.

Francescutto, A., G. Contento & R. Penna (1994) Experimental Evidence of Strong Nonlinear Effects in the Rolling Motion of a Destroyer in Beam Seas. *Proc. 5th Int'l Conference of Stability of Ships & Ocean Vehicles* (STAB 94), Florida Institute of Technology, Melbourne, FL, Vol. 1, 13 p.

IMO MSC 85/26/Add.1 (2008) International Code on Intact Stability, 2008 (2008 IS Code). London, 96 p

IMO MSC.1/Circ.1281 (2008) Explanatory Notes to the International Code on Intact Stability, 2008, London, 30 p.

IMO SDC 2-WP.4 (2015) Development of Second Generation Intact Stability Criteria, Development of Amendments to Part B of the 2008 IS Code on Towing, Lifting and Anchor Handling Operations, Report of the working group (Part 1), London, UK, 48 p.

IMO SLF 48/21 (2005) Report to Maritime Safety Committee, London, UK, 65 p.

IMO SLF 50/4/4 (2007) Framework for the Development of New Generation Criteria for Intact Stability, submitted by Japan, the Netherlands and the United States, London, UK, 6 p.

IMO SLF 55/3/11 (2013) Development of second generation intact stability criteria, comparison study of draft level 2 vulnerability criteria for stability under dead ship condition. Submitted by Italy and Japan. London, UK, 4 p.

ITTC (2011) The Specialist Committee on Stability in Waves: Final Report and Recommendations to the 26th ITTC. *Proc. 26th ITTC*, Rio de Janeiro, Brazil, 36 p.

ITTC (2014) The Committee on Stability in Waves: Final Report and Recommendations to the 27th ITTC. *Proc. 27th ITTC*, Copenhagen, Denmark, 67 p.

Kan, M. (1990a) Surging of large-amplitude and surf-riding of ships in following seas. *Naval Arch. & Ocean Engin.*, 28:49–62.

Kan, M. (1990b) A Guideline to Avoid the Dangerous Surf-riding. *Proc. 4th Int'l Conf. Stability of Ships & Ocean Vehicles*, University Federico II of Naples, Naples, Italy, pp. 90–97.

Kantz, H. & T. Schreiber (2004) *Nonlinear time series analysis*. Cambridge University Press, Cambridge, UK, xvi+369 p.

Kobylinski, L. K. & S. Kastner (2003) *Stability and Safety of Ships: Regulation and Operation*. Elsevier, Amsterdam, 454 p.

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,* Seoul, South Korea., 10 p.

NATO (2007a) Buoyancy, Stability and Controllability. Chapter III of Naval Ship Code, NATO Naval Armaments Group, Maritime Capability Group 6, Specialist Team on Naval Ship Safety and Classification, Allied Naval Engineering Publication ANEP-77, vii+121 p.

NATO (2007b) Guidance on NSC Chapter III Buoyancy and Stability, Part B: Application. Chapter 3, Guide to the Naval Ship Code, NATO Naval Armaments Group, Maritime Capability Group 6, Specialist Team on Naval Ship Safety and Classification, 91 p.

Navy (1999) SECNAV Instruction 5200.40: Verification, Validation, and Accreditation (VV&A) of Models and Simulations, 21 p.

Navy (2002) SECNAV Instruction 5200.38A: Department of the Navy Modeling and Simulation Management, 8 p.

Navy (2004) Department of the Navy, Modeling and Simulation Verification, Validation, and Accreditation Implementation Handbook, Volume I, VV&A Framework, 39 p.

Navy (2005) Approved Navy Modeling and Simulation Standard: Best Practices Guide for Verification, Validation, and Accreditation of Legacy Modeling and Simulation. Navy Modeling and Simulation Standards Project, 48 p.

NIST (2014) National Institute of Standards and Technology, "NIST/SEMATECH e-Handbook of Statistical Methods," [Online]. Available: `http://www.itl.nist.gov/div898/handbook/`. [Accessed 1/2/2014]

Oberkampf, W. L., & M. F. Barone (2006) Measures of agreement between computation and experiment: Validation metrics. *J. Comp. Physics*, 217, 5–36.

Peters, W., V. Belenky, C. Bassler, K. Spyrou, N. Umeda, G. Bulian, B. Altmayer (2011) The Second Generation Intact Stability Criteria: An Overview of Development. *Trans. SNAME*, 119:225–264.

Rahola, J. (1939) The judging of the stability of ships and the determination of the minimum amount of stability especially considering the vessel navigating Finnish waters. PhD Thesis, Technical University of Finland, Helsinki, viii+232 p.

Reed, A. M. (2009) A Naval Perspective on Ship Stability. *Proc. 10th Int'l Conf. Stability of Ships & Ocean Vehicles* (STAB '09), St. Petersburg, Russia, pp. 21–44.

Rosborough, J. M. (2007) Stability and buoyancy of U.S. Naval surface ships. Design Data Sheet DDS 079-1, Version: 2.01, 30 Jan 2008, Carderock Division, Naval Surface Warfare Center Report NSWCCD-20-TR–2007/05, 107 p.

Sarchin, T. H. & L. L. Goldberg (1962) Stability and buoyancy criteria for U. S. Naval surface ships. *Trans. SNAME*, 72:418–58.

Spyrou, K. J. (1996) Dynamic Instability in Quartering Seas: The Behavior of a Ship During Broaching. *J. Ship Res.*, 40(1):46–59.

Spyrou, K. J. (1997) Dynamic Instability in Quartering Seas—Part III: Nonlinear Effectson Periodic Motions. *J. Ship Res.*, 41(3):210–223.

Spyrou, K. J., K. M. Weems & V. Belenky (2009) Patterns of Surf-Riding and Broaching-to Captured by Advanced Hydrodynamic Modelling. *Proc. 10th Int'l Conf. Stability of Ships and Ocean Vehicles* (STAB 09), St. Petersburg, Russia, 15 p.

Themelis, N. & K. J. Spyrou (2007) Probabilistic Assessment of Ship Stability. *Trans. SNAME*, 117:181–206.

Themelis, N. & K. J. Spyrou (2008) Probabilistic Assessment of Ship Stability Based on the Concept of Critical Wave Groups. *Proc. 10th Int'l Ship Stability Workshop*, Daejeon, S. Korea, 11 p.

Zuzick, A. V., A. M. Reed, W. F. Belknap & B. L. Campbell (2014) Applicability of the Difference Between Population Statistics as an Acceptance Criteria Metric for Seakeeping Validation. *Proc. 14th Int'l Ship Stability Workshop*, Kuala Lumpur, Indonesia, 9 p.