SNAME Maritime Convention 2024
14-16 October, Norfolk, VA

www.sname.org

Validation Methodology for Time-Domain Simulation of Maneuvering and Seakeeping

F.H.H.A. Quadvlieg[1] (FL), **P. Bandyk**[1] (M), **N. Carette**[1] (V), **P. de Jong** (V)[1]
1. MARIN Maritime Research Institute Netherlands

A time-domain ship simulation program is intended to model nonlinear motions, including combined maneuvering and seakeeping. The program is a reduced-order model capable of carrying out thousands of simulations so that it can be used to evaluate the probability of extreme events. The results of the thousands of calculations are used to quantify the capsize risk, for example. The challenge to validate the capsize risk is that an occurrence of capsizing is the result of a chain of (hydrodynamic) events, each of which needs validation. To validate prediction of rare events, we claim validation on many relevant subevents up to a satisfactory level.
The program FREDYN is used in this paper, though the approach is valid for similar tools. Initially, FREDYN was meant to be used for naval monohulls: frigates and destroyers. Recently there has been a desire to extend the tool for other ship types as well. Extending the applicability of the tool also extends the more formal validation process, in addition to the verification. The distinction is that verification results in an objective comparison where pass/fail is possible. Validation evaluates the quality of the prediction for a given set of reference cases, usually model tests or full-scale measurements. The validation of the code therefore requires some subjective assessment of the results, which can obfuscate the process. The objective of this paper is to discuss some of these details. After various iterations in this validation methodology, this process is now automated. The validation framework can be repeated efficiently, allowing frequent use when developing or updating the underlying code or adding new validation data, such that the overall performance can be quantified.

KEY WORDS: Capsize; maneuvering; seakeeping; stability.

INTRODUCTION

A time-domain ship simulation program is intended to model nonlinear motions. The program is a reduced-order model capable of carrying out thousands of simulations so that it can be used to evaluate the probability of extreme events such as capsize risk.

Such a program needs to be fast and robust, with sufficiently high accuracy to model the important physics. It needs to be validated for a range of ships, not just for one ship. The code consists of an eclectic mix of semi-empirical relations to quantify forces and moments. A balance is to be found between accuracy in predicting extreme events and the accuracy of prediction of frequent events. One such tool commonly used within MARIN and various navies is FREDYN. Initially, the program was meant to be used for naval monohulls: frigates and destroyers. Recently there has been a desire to extend the tool for other ship types as well. Extending the applicability of the tool also extends the more formal validation process, in addition to the verification. The distinction is that verification results in an objective comparison where pass/fail is possible. Validation evaluates the quality of the prediction for a given set of reference cases, usually model tests or full-scale measurements.

The validation should be performed for several ships using reliable and well documented reference data. But, when a (reduced-order) simulation program is based on empiricism, a perfect match cannot be expected in all circumstances and for all test cases. The validation should however reveal whether "sufficient accuracy" can be achieved over the range of possible applications. For example, modification of the empirical formulae to improve accuracy in predicted maneuvers in calm water could mean decreased accuracy for course keeping in waves. To find a balance in predictive quality, a system of aggregated quality numbers is developed.

A validation framework has been developed, broken into several key subtopics that play a part in the overall simulation of the behavior in waves, including extreme waves:

- Roll decay at zero and forward speed;
- Maneuvers in calm water;
- Motions in regular waves while course keeping;
- Motions in irregular waves while course keeping;
- Course keeping and motions in extreme waves.

We want to draw conclusions on the validity of the code for each of these subtopics.

Each simulation results in a set of representative time traces

for ship motions. We derive a quality number (which can be seen as the prediction error) to compare the simulation with validation data. This quality number not only depends on "how close the simulation is to the experiment" but takes into account 1) which data is obtained from a time trace in order to be representative for the behavior of the simulation, 2) the quality of the experimental data, and 3) what is "good enough".

After various iterations in this validation methodology, this process is now automated at MARIN for use with FREDYN. The validation framework can be repeated efficiently, allowing frequent use when developing or updating the underlying code (for example a new roll damping module) or adding new validation data, such that the overall performance can be quantified and compared.

SIMULATION TOOL

Global Description

FREDYN is a time-domain seakeeping and maneuvering code developed by the CRNavies. FREDYN has been under constant development since the late 1980s (see Hooft and Pieffers, 1988) and is optimized for frigate-like ships, though the code can handle other types of vessels, and even progressive flooding. The code has been used to study the probability of capsize of intact ships (see de Kat et al, 1994, De Kat et al., 1998 and McTaggart and de Kat, 2000). Also progressive flooding is implemented which allows to study damaged frigates, as shown by de Kat and Peters (2002).

The time-domain code is based on the instantaneous calculation of a range of forces: maneuvering reaction forces, rudder forces, propeller forces, wave excitation forces, damping forces (also called retardation forces in combination with added mass), diffracted wave forces. Instantaneous submergence allows to calculate the so-called non-linear Froude Krilov force and therefore the non-linear hydrostatics. In the terms of ITTC, this is hence a non-linear seakeeping code.

Maneuvering Reaction Forces

The so-called maneuvering hull forces are the sum of three components: the lift forces (the forces and moments due to an angle of attack on the hull), the lift damping forces (the lift that is caused by the rate of turn) and the cross-flow drag, which is occurring due to a transverse velocity on every section of the ship's hull.

The lift generated on the hull of a ship is traditionally expressed in linear maneuvering coefficients, in particular the terms $Y'_{v'}$, $N'_{v'}$, $Y'_{r'}$ and $N'_{r'}$, which are the linear maneuvering coefficients in side force and yawing moment due to sideways velocity v and yaw rate r. For FREDYN, these terms are not calculated directly. When a ship is sailing in waves, the relative wave elevation along the hull is constantly varying and consequently the lift and damping terms are varying also. Since FREDYN is developed for fine ship forms, formulae that predict linear maneuvering coefficients have been derived for frigate-type hull forms. They are derived for the ship at a range of draughts and pitch angles. This allows prediction of important nonlinear effects, such as decrease of directional stability as the ship's bow plunges into a wave, while maintaining relatively simple input including the time-varying wetted hull geometry.

The linear maneuvering force and moment (which are in fact non-linear) for sway Y, roll K, and yaw N are calculated every time step:

$$\begin{aligned} Y_{lin} &= Y_{uv}\left(T_f, T_a, L_{PP}, B, C_B, Fn\right) \cdot \frac{1}{2}\rho u v L_{PP} T \\ &+ Y_{ur}\left(T_f, T_a, L_{PP}, B, C_B, Fn\right) \cdot \tfrac{1}{2}\rho u r {L_{PP}}^2 T \\ K_{lin} &= z_Y \cdot Y_{uv}\left(T_f, T_a, L_{PP}, B, C_B\right) \cdot \tfrac{1}{2}\rho u v L_{PP} T^2 \\ N_{lin} &= N_{uv}\left(T_f, T_a, L_{PP}, B, C_B, Fn\right) \cdot \frac{1}{2}\rho |u| v {L_{PP}}^2 T \\ &+ N_{ur}\left(T_f, T_a, L_{PP}, B, C_B, Fn\right) \cdot \frac{1}{2}\rho u r {L_{PP}}^3 T \end{aligned} \quad (1)$$

These forces and moments are based on the instantaneous velocity of the ship (u, v, and r) and the instantaneous local draught. The wave orbital velocities are not taken into account in the calculation of these velocities.

The coefficients Y_{uv}, Y_{ur} , N_{uv} and N_{ur} are evaluated based on the instantaneous draught fore and aft, T_f and T_a, respectively. The formulae that predict the coefficients are based on a regression of several model scale measurements. The coefficients are speed dependent terms and reflect that the directional stability changes as a function of the forward velocity. Therefore, these coefficients are a function of ship length L_{PP}, waterline breadth B, block coefficient C_B, and Froude number Fn.

Apart from so-called linear maneuvering forces, non-linear maneuvering terms also play a role. These are approximated by the cross-flow drag. Certainly in maneuvers such as a turning circle maneuver, the cross-flow drag is one of the dominant terms. Several papers have been written on cross-flow drag, see Hooft (1994) and Hooft and Quadvlieg (1996). An earlier version of this theory is implemented in FREDYN.

The forces and moments due this cross-flow drag (Y_{CFD}, K_{CFD} and N_{CFD}) are calculated as follows:

$$\begin{aligned} Y_{CFD} &= \frac{1}{2}\rho \int_{-L_{PP}/2}^{L_{PP}/2} v_x \cdot |v_x| \cdot T_x \cdot C_D(x,u,v,r) \cdot dx \\ K_{CFD} &= \tfrac{1}{2}\rho \int_{-L_{PP}/2}^{L_{PP}/2} v_x \cdot |v_x| \cdot T_x \cdot C_D(x,u,v,r) \cdot z_{CFD} \cdot dx \\ N_{CFD} &= \frac{1}{2}\rho \int_{-L_{PP}/2}^{L_{PP}/2} v_x \cdot |v_x| \cdot T_x \cdot C_D(x,u,v,r) \cdot x \cdot dx \end{aligned} \quad (2)$$

In these forces and moments, the local sectional transverse velocity is v_x. This v_x is the local transverse velocity in ship coordinate system, relative to the waves: the wave orbital velocity is taken into account. The formula to calculate v_x is hence:

$$v_x = v + x \cdot r - v_w \quad (3)$$

In which v_w is the wave orbital velocity perpendicular to the ship hull, which is decomposed in the ship-fixed coordinate

system from the wave orbital velocity. Furthermore, T_x is the instantaneous draught of every section, which includes the effect of the instantaneous pitch angle and also the (undisturbed) wave height. z_{CFD} is the supposed height of application of the cross-flow drag force. L_{PP} is the length between perpendiculars of the ship.

The cross-flow drag coefficient is a sectional coefficient, which is a function of the position of every section, the block coefficient and the velocity of the ship. The cross-flow drag coefficient is slightly dependent on the Froude number.

Because the linear maneuvering forces and the cross-flow drag forces are functions of the instantaneous draught, pitch, relative wave height and wave orbital velocities, there are wave excitation forces generated in this part of the code. Due to the nature of these forces and the way they are calculated, they will induce mean wave drift forces. In the terminology of Pinkster (1980), components III and IV can be seen as this contribution.

The calm water resistance at design draught is pre-calculated and given as input to FREDYN. Another longitudinal force is the centrifugal contribution of the added mass, also called the X_{vr} term. The magnitude of this term is based on regression of a number of rotating arm experiments on slender body hull forms. This is a term which is based on the instantaneous draught.

In FREDYN, the direct coupling between the maneuvering and the wave forces is handled by correcting the added mass and damping coefficients to not double count these forces.

Rudder and Propeller Forces

The propeller (or waterjet) are controlled as a function of the RPM. A 4-quadrant curve or a propeller open water curve can be used as input. Based on the instantaneous velocity at the propeller (which includes the wake fraction) and the instantaneous wave orbital velocity, the advance ratio J ($= \frac{V_a}{n \cdot D_P}$) is calculated. In the case a 4-quadrant model is used for the propeller, the advance ratio is in fact the angle of attack $b_P = \mathrm{atan}\left(V_a, 0.7pnD_P\right)$. V_a is the projected speed along the ship's x-axis and is the sum of instantaneous ship velocity, wake fraction, and the longitudinal wave orbital velocity, n is the propeller rate per second and D_P the propeller diameter.

Based on these, the propeller thrust and torque are calculated. In addition, FREDYN calculates whether the propeller is in- or out- of the water. The thrust and torque are decreased based on the amount of submergence of the propeller disc.

The rudders on the ship are modeled as lifting surfaces. Lift and drag are calculated based on the instantaneous velocity over the rudder and the instantaneously submerged rudder area. The velocity over the rudder is calculated taking into account the instantaneous attitude and velocity of the ship, the wave orbital velocity and the thrust of the propeller in front of the rudder.

Additionally, the rudder is steered by an autopilot programmed to keep the ship on track or on heading. The rudder angle, together with the instantaneous longitudinal and lateral velocities determine the angle of attack. Using the locally-resolved velocities and angle of attack, the lift and drag of the rudders can be calculated based on a lift coefficient, which depends on the rudder aspect ratio, sweep back angle, and taper ratio. Lift and drag are decomposed into the ship-fixed frame to obtain forces and moments. Special attention is paid to modelling the post-stall angle behavior.

In addition to the lift, there is a rudder-to-hull interaction. The origin is a carry-over of lift from the rudder towards the hull, which generates a pressure field on the hull which (usually) increases the effective transversal force. This can be expressed as follows:

$$Y_{Rud} = (1 + C_{HRU}) \cdot (Y_{RSB} + Y_{RPS}) \quad (5)$$

Where C_{HRU} is the rudder-to-hull interaction coefficient based on an empirical formula obtained from past model tests. Y_{RSB} and Y_{RPS} are the forces on the starboard and portside rudders and Y_{Rud} is the total force due to the rudder.

The rudder is usually mounted on a headbox, where the lift and drag are calculated in a similar manner.

FREDYN is also able to cope with waterjets. For waterjets, similar formulations are created as propeller open water data. The interesting observation is that the waterjet comes with a steering bucket. Depending on the steering angle of the waterjet, the thrust is influenced, but also the *Y* force, *K* moment, and *N* moment.

Wave Forces

In FREDYN, the first order ship motions are solved by accounting for:

- non-linear hydrostatic and Froude-Krylov forces applied on a 3D mesh, where the underwater part is determined at each time-step based on the current position and undisturbed wave elevation,
- pre-computed wave diffraction forces from a linear frequency-domain 2D strip theory method, using the underwater geometry at rest. The diffraction forces are then converted in a time varying force using direct summation method,
- radiation forces based on retardation functions based on pre-computed added mass and damping, from a linear frequency domain 2D strip theory method, using the underwater geometry at rest.

To account for the maneuvering and seakeeping coupling, the frequency-domain added mass and damping are corrected to avoid double counting of the linear maneuvering reaction forces: the linear maneuvering coefficients correspond to the damping coefficient at zero frequency.

Due to the phasing of the coupled heave-pitch and sway-yaw motions, and due to the way the instantaneous hydrostatic and

Froude-Krylov forces are calculated, mean and slowly varying drift forces are occurring. One of the missing components in FREDYN is that the relative wave elevation around the ship is limited to the incoming wave and the ship motions, meaning that the drift forces are incomplete. The effect of the diffracted and radiated wave on the relative wave elevation is not taken into account in this approach, which has an effect on the mean drifting behavior. Also the steady wave pattern caused by the forward speed of the vessel is not taken into account.

VALIDATION METHODOLOGIES

Validation Methodologies for Multiple Codes

Much experience exists for comparing multiple codes against a well-defined validation case. Workshops on validity of predictions help to understand the state of the art for the prediction quality by various prediction methods. The insights from these workshops steers code development and guides the industry in the application of these codes.

These methodologies are developed and used in workshops like the well-known CFD workshops and the SIMMAN workshops (see Stern et al, 2009 and 2011, Quadvlieg et al, 2015 and Quadvlieg et al 2024). The validation exercises have the objective of comparing various codes against the prediction for one ship.

When performing a validation, the results of the prediction will be held against experimentally determined data. It has become common to validate not time traces, but derived quantities. Time traces which are on top of each other are very convincing, especially when they are on top of each other. However, to demonstrates the validity of a code, it is better to quantify the prediction error of a derived quantity. The challenge is two-fold: to properly select the key characteristics and to obtain high-quality data for a ship, preferably from tests at multiple institutes to quantify also the facility bias.

The methodology is to determine the highest possible quality of the experimental data, including uncertainty analysis obtained though repeat experiments, as described by Quadvlieg and Brouwer (2011), Kisjes et al (2019). An example of the quantification of the facility bias is given by Eloot et al (2015).

For every code, the prediction error is determined for a number of key characteristics. This is calculated as S-D, in which D is the experimental value and S the prediction. The average code prediction error is the average of all individual prediction errors. When the average prediction error is smaller than the experimental uncertainty, “validation” is reached, as described by Stern et al (2018).

Validation Methodology for One Code

The validation of a code like FREDYN is different from validations mentioned in the previous section. The objective is that the results of the calculation will be held against multiple tests for multiple ships. For eclectic simulation programs based on mixes of semi-empirical relation and approximative potential flow codes, the following aspects play a role.

- It is necessary to verify the performance of the code for a range of ships. These ships may be on the borders of the validity range of the software, which is interesting to know and observe. The code should also be applied to ships outside the validity range to identify and document the borders of the validity range.
- Although the ultimate goal is the validation of capsize probability, there are not many ships for which capsize tests are available. Moreover, also when tests are reported, they often focus on the occurrence of a capsize in a very specific condition. On the other hand, we are interested in the range of environments of capsizing and the probability of capsizing. Therefore, not every capsize test is valuable for the validation.
- Per ship, different types of tests are available. It may be that for a ship not all scenarios are present: we may have capsize test results, but no zigzag test results. The code should be able to handle multiple speeds, rudder angles, etc., and the validation should indicate where possible improvements could be made to assist code development.
- To aid code development, the validation error of earlier versions of the code should be easily comparable to the validation error of newer versions of the code. Since the codes are empirical, it may be that code validation improves for a certain class of ships while it does not improve or becomes worse for other ships. Clear insight in that is essential. This means that when code modifications are taken place, the validation should be easily redone and there should be a clear report which documents the effectiveness of the modifications over the whole performance range of the software.

To cope with these different types of validation objectives, the following methodology is developed. It is explained using the example of a tactical diameter measured in a turning circle maneuver in calm water.

Database of Experimental Data and Data Conditioning

The idea is to obtain as much trustworthy data for as many ships as possible. From this data, key characteristics are derived. In the case of for example the tactical diameter, we gathered for 11 ships a total of 90 turning circle maneuvers. All 90 cases are ships with different speeds, different rudder angles. On average, there are about 9 turning circle tests per ship.

Determining the Score for a Key Characteristic

This section explains the validation methodology for the tactical diameter (which is a key characteristic for a calm water maneuver).

Compared to other validation methodologies, we are not comparing time traces but derived quantities. Each of the derived quantities has a measurement accuracy (we call this precision). An example is given in for the tactical diameter. The tactical diameter is considered to be a key characteristic of a calm water maneuver with 35° rudder angle. The experimentally determined value is 3.5·LPP. We determined a tolerance of 0.1·LPP. This

would mean that when a prediction is made, and the result of the prediction is between 3.4·LPP and 3.6·LPP, the prediction should be considered "excellent". This tolerance has a close relationship to the U95 as used by other validation studies (see Stern et al 2022), but is not exactly the same.

On the other hand, we will have a limit. When the prediction is outside of the limit, this is considered unacceptable and a score of zero is given. In this example, we consider it unacceptable when the error is beyond 1·LPP.

Fig. 1 explains the way a score is given. A predicted tactical diameter of 4.7 will give a score of zero. A predicted tactical diameter of 3.8 will give a score of 0.8. Any predicted tactical diameter between 3.4 and 3.6 will give a score of 1.

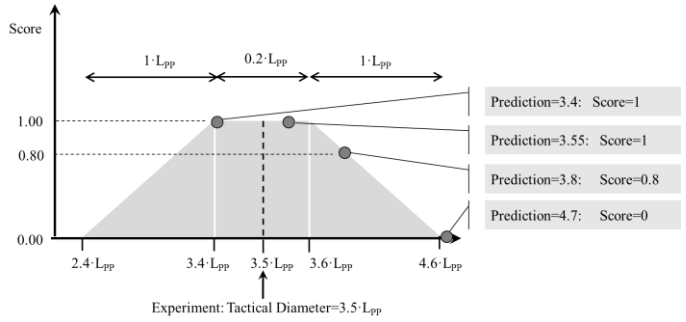

Fig. 1, Methodology for scoring

The Prediction Quality Number for a Ship

For every key characteristic of a ship, we are determining the score. To further condense the error, the errors are averaged in two ways: by averaging the individual scores, the so-called arithmetic mean S_A (see formula (6)) and by using the geometric mean S_{ar} (see formula (7)). A characteristics of the geometric mean is that as soon as one receives a value of zero, this progresses through every score, and that the total score becomes zero. This acts as a warning flag to determine when a code modification results in a code giving unacceptable results.

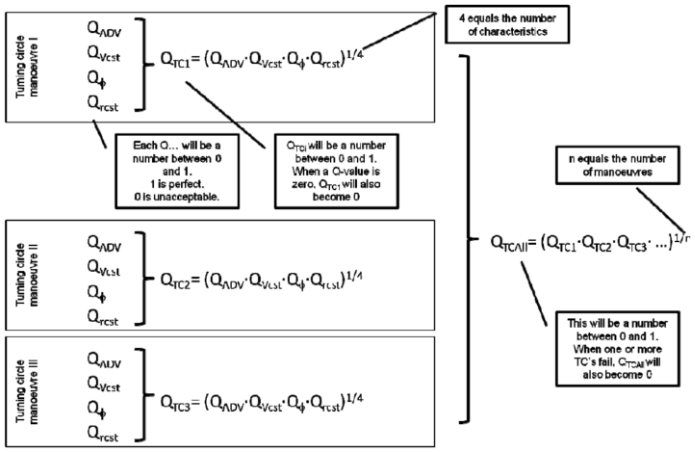

Fig. 2, Summary of prediction score for multiple turning circle test results

$$S_A = \frac{1}{n}\sum_{i=1}^{n} score_i \quad (6)$$

$$S_{ar} = \sqrt[n]{score_1 \cdot score_2 \cdot \ldots \cdot score_n} \quad (7)$$

While this is explained for the tactical diameter, this is accumulated for different key characteristics and for every maneuver (at different speed, rudder angle, etc.). At the end, an average quality can be determined for all turning circle tests. This is visualized in Fig. 2.

VALIDATION PARTS

As explained in previous sections, although the ultimate stability behavior is the main focus, FREDYN as a tool will be used for many studies that do not include ultimate stability. Moreover, when the ultimate stability is predicted less accurately, we need to question the intermediate results: the roll motions, behavior in calm water, and basic behavior in waves. There is a need to quantify that these more common predictions are good. This approach of having many more validation parts has proven to be worthwhile for the development of other predictions programs for rare phenomena, such as the behavior of free fall lifeboats in extreme waves (see Palermo et al, 2015).

We separate the validation into 6 parts, treated in the following sections:

- Roll behavior, captured by roll decay tests;
- Turning behavior, captured by turning circle maneuvers;
- Course keeping and directional stability, and in particular yaw checking behavior, which is captured by zigzag tests;
- The basic motions of the ships, captured by the outcome of tests in regular waves;
- Motions in moderate irregular, unidirectional (long-crested) sea states;
- Motions in extreme waves, in particular stern quartering waves where surf riding, loss of stability, and broaching may occur.

Roll Decay

Roll motions are a vital part of a prediction for software that predicts ultimate stability. The roll damping will have an effect on many other behaviors except the behavior in head waves.

The physical phenomena behind the roll motions, and in particular the roll damping, depend on the ship forward speed. This is illustrated in Fig. 3. The left shows a roll decay experiment and simulation at zero speed, on the right at a speed of 18 knots. The nature of the decay is quite different. At zero speed, the damping is coming from vortex shedding at the bilges and bilge keels, containing cross-flow drag phenomena, while at forward speed, lift generation of the hull, rudder, and other appendages are sources of roll damping. At zero speed, the motions are lightly damped, while the test at higher forward speed show much more damping (although still underdamped). As the speed increases towards speeds of 30 knots, ships tend to be critically damped.

It should be noted that these phenomena are non-linear (meaning that the amount of damping depends on the amplitude of roll angle and roll velocity), and in the case of extreme roll

motions, depend on whether bilge keels are submerged or not, propellers may emerge, etc.

The second aspect of importance is that the motion is a coupled motion. While rolling, a ship will sway, and even heave and yaw. The coupled damping related to these motions will also be present in the time traces of the roll motion. At speed, the roll decay is performed while sailing on a straight line, under autopilot. This means that the rudder reacts on the motions of the ship, and this will have an effect on the roll motions as well.

The above summary illustrates that it is important that all settings of the experiments need to be known to make the proper simulation.

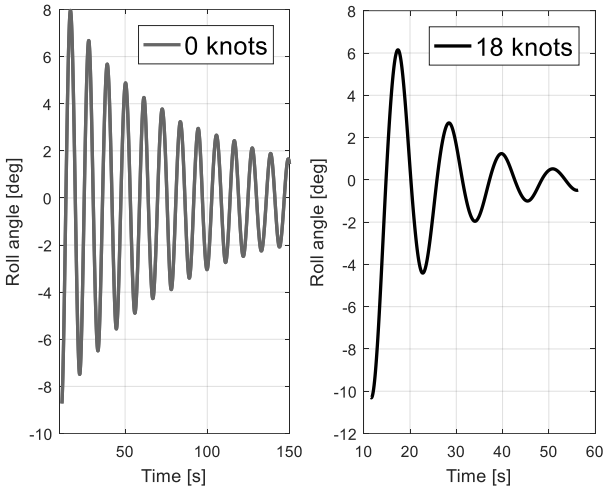

Fig. 3, Time trace of a roll decay test at 0 and at 18 knots

A roll decay motion is often quantified by a PQ analysis in which the proportional and quadratic parts of the damping are derived from a decaying motion. This only works well in the case of a lightly damped motion, which means when the speed is low enough. The motion observed in the left plot of Fig. 3 can be described well by proportional and quadratic parts. For the right plot of Fig. 3, a proportional part can be derived, but a quadratic part would be less meaningful because there are simply not enough oscillations in the signal.

To quantify the accuracy, decay analysis is performed on the experiment and simulation results. This gives three values: the roll period, T_ϕ, and the decay parameters, p and q. Instead of comparing the values of p and q, the value of $p+5°\cdot q$ is compared. This is representative of the damping at a roll angle of 5°. This is stored in an array which also contains values for the initial speed and the initial roll angle, and also notes related to the use of stabilizer fins.

It is concluded that the two key characteristics are the roll damping value $p+5°\cdot q$ and the roll period T_ϕ.

Using the array of results, two types of figures are created. The first one is a QQ plot, showing the experiment versus the simulation (Fig. 4). The second type of plot is where the non-dimensional and objective values are plotted against an important parameter, in this case the ship speed (Fig. 5). The availability of these plots gives important insights, for example the predicted roll damping at speeds above 15 knots is too high, but at speeds below it shows good agreement. These are important leads to future code developments.

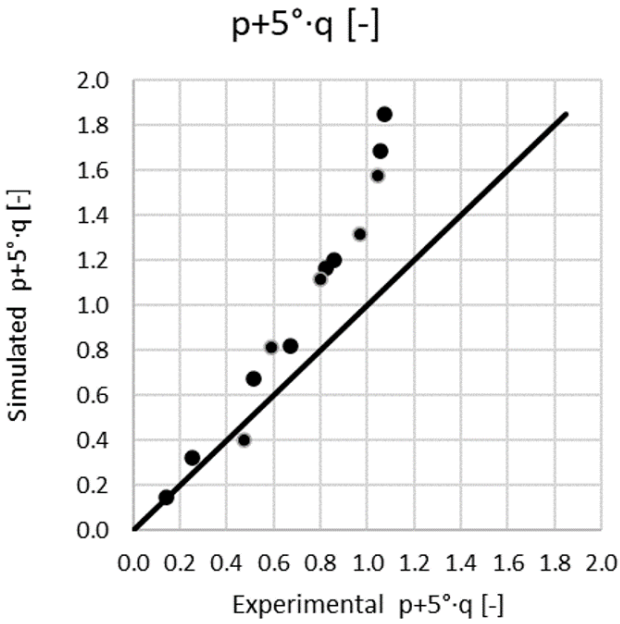

Fig. 4, values of p+5°q: experiment versus simulation

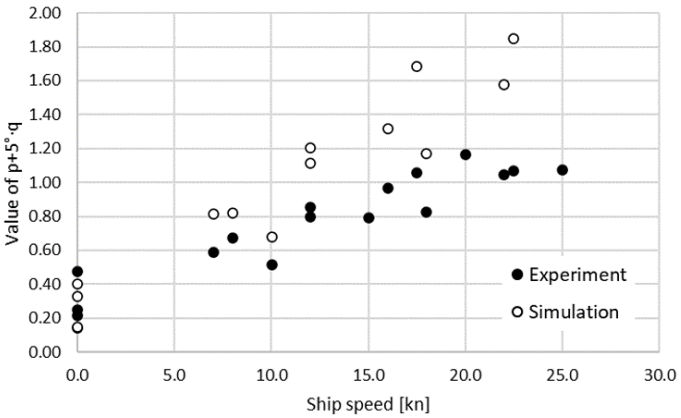

Fig. 5, Example of key characteristic plotted against speed

Based on Fig. 5**Error! Reference source not found.**, it can be observed that the uncertainty of the experimentally determined value of $p+5°\cdot q$ is dependent on the ships speed. Consequently, this is how the values of the limit and tolerance should be set. The tolerance is set to a value of 0.05 (independent of the ship speed). The value of the limit is set to $0.20+V_S\cdot 0.025$ (in which V_S the ship speed in knots).

The quality number can now be determined by calculating the prediction error for every point in Fig. 5: the error is compared to the tolerance and the limit as illustrated in Fig. 6. When this error is smaller than the tolerance (the error is within the green

lines in Fig. 6, and the quality number becomes 1. Between the green and red lines, the value will become between 0 and 1.

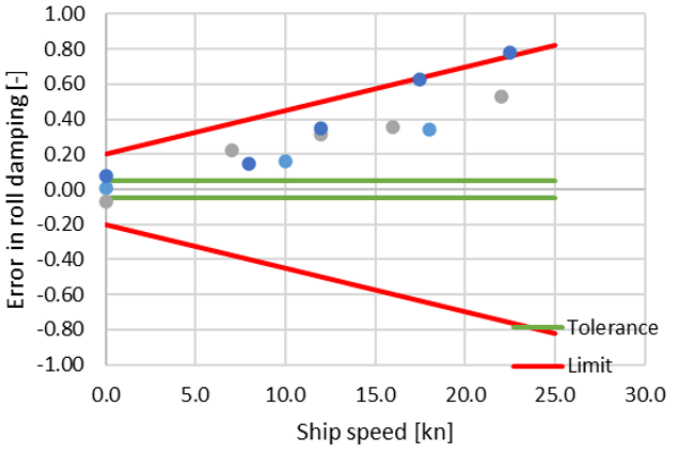

Fig. 6, Prediction error in roll damping compared to tolerance and limit

The average of all the points in Fig. 6 is 0.510, which in this case the quality number in roll damping prediction.

Turning Circle Maneuvers

A turning circle maneuver gives information about the rudder forces and the ability of the ship to turn quickly. Also the response to rudder action is quantified. A large database of turning circle maneuvers for different ships is gathered. These are performed in different institutes, full scale and model scale, and furthermore at different starting speeds and different rudder angles.

For these turning circle maneuvers, the following 10 key characteristics are determined:

- Tactical diameter, expressed in ship lengths; the distance covered by the center of gravity in the direction perpendicular to the original course when the ship has obtained 180 degrees change of heading. The y-position at the first instance in the time trace of the absolute values of the heading which is larger than 180 degrees is defined as the tactical diameter. This characteristic is only calculated for turning circle maneuvers with rudder angles larger than 29°. For a tactical diameter with rudder angles lower than 29º, the experimental uncertainty is higher, and therefore the values of 30º and higher are selected.
- Advance, expressed in ship lengths. The distance covered by the center of gravity (CoG) in the direction of the initial course when the ship has obtained 90 degrees change of heading. Measured as the first point in time where the absolute values of the heading is larger than 90 degrees. This characteristic is only calculated for turning circle maneuvers with rudder angles larger than 29°.
- The maximum inward heel angle ϕ_{in} , expressed in degrees. Immediate after rudder is given, usually a ship heels in the direction of the turn. If the commanded rudder angle is negative (a starboard turn), the maximum of the roll angle between the start of the maneuver and the end of the maneuver. For a positive commanded rudder angle, the minimum is taken. This value is then multiplied with sign of the execute angle. In this way, the values are better comparable in figures.
- The maximum outward heel angle ϕ_{out} , expressed in degrees. After a short while in the turn, the ship will receive an outward heeling angle. Because the speed is still high, this outward heeling angle is usually the largest heeling angle. If the commanded rudder angle is negative (a starboard turn), the minimum of the roll angle between the start of the maneuver and the end of the maneuver. For a positive commanded rudder angle, the maximum is taken. This value is then multiplied with sign of the execute angle. In this way, the values are better comparable in figures.
- The maximum rate of turn r'_{max} (which is made non-dimensional by $r \cdot L_{PP}/V_0$), in which V_0 is the starting speed of the maneuver;
- The non-dimensional rate of turn r'_{stc} (which can also be calculated as $2 \cdot L_{PP}/D_{stc}$, in which D_{stc} is the diameter of the steady turning circle); This value is multiplied with the sign of the execute angle. In this way, the values are better comparable in figures. So, for both a positive (portside turn) or negative rudder angles, a negative rate of turn is obtained.
- The drift angle β_{stc} in the steady part of the turn, expressed in degrees; This value is multiplied with the sign of the execute angle. So, for both a positive (portside turn) or negative rudder angles, a positive drift angle is obtained.
- The pivot point (also calculated as $\sin\beta_{stc} / r'_{stc}$), which is a non-dimensional value;
- The speed loss (calculated as V_{stc}/V_0) , which is a non-dimensional value, in which V_{stc} is the steady speed with the ship in the turn;
- The steady heel angle ϕ_{stc} , expressed in degrees; This value is multiplied with the sign of the execute angle. So, for both a positive (portside turn) or negative rudder angles, a positive drift steady heel angle is obtained.

It should be noted that some of these key characteristics are related and results are not independent. Physically, when V_{stc}/V_0 is low, the values of r'_{stc} will be larger. It is nevertheless kept, as the results will lead to insights which assist in code improvement.

An example of a QQ plot for the tactical diameter is given in Fig. 7. This shows that for one particular ship, the performance is unacceptable, while for the others, the performance is acceptable. The ship with unacceptable performance is outside of the validity range. For the ships inside the validity range, there is still room for improvement.

Similar to roll damping, a tolerance and limit are defined. The tolerance is quite dependent on the quality of the original measurements. In the case of Fig. 7, some measurements are older, some are newer, some come from full scale measurements, and others from model scale measurements. The full scale measurements are often measured in environmental conditions

which were not fully zero. Consequently, the measurement uncertainty is higher than the measurement uncertainty in laboratory circumstances.

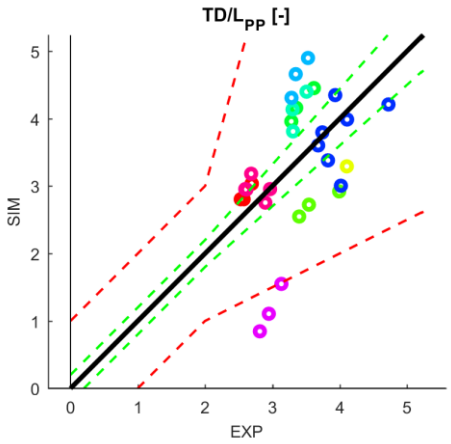

Fig. 7, QQ plot of the tactical diameter obtained for the turning circle maneuver

For the tactical diameter *TD*, an absolute and a relative tolerance is used: the higher the value, the higher the prediction error is allowed to be. The tolerance and limit are set as follows:

If $TD_{EXP} < 2$ then tolerance = 0.2
If $TD_{EXP} > 2$ then tolerance = $0.1 \cdot TD_{EXP}$

For the limit, a similar relationship is defined:

If $TD_{EXP} < 2$ then limit = 1
If $TD_{EXP} > 2$ then limit = $0.5 \cdot TD_{EXP}$

So, when the experimental tactical diameter is 4, the tolerance is 0.4 and the limit is 2. These values are obtained through judgement and experience of the model test and full scale experiments from history.

The quality number can now be calculated. For the calculation at hand, application of the average quality of prediction for 11 ships, 90 turning circle maneuvers is 0.684.

Zigzag Maneuvers

The zigzag maneuver is an important qualification, because it tells us the directional stability and the yaw checking ability. Yaw checking is particularly important for sailing in stern-quartering waves because it gives an impression on how swift a turn in one direction can be stopped by giving a rudder angle order to the other direction.

Typically, these results are obtained in the schematized zigzag tests, of which there are commonly two types: a 10°/10° zigzag test and a 20°/20° zigzag test. Because of the IMO requirements for ship maneuverability, these maneuvers are often carried out, which means there is a good database of test results.

For these zigzag maneuvers, the following key characteristics are determined:

- First overshoot angle, expressed in degrees;
- Overshoot time, made non-dimensional by dividing the time by L_{PP}/V_0;
- Period, made non-dimensional by dividing the time by L_{PP}/V_0;
- Initial turning time (t_{2nd}), made non-dimensional by dividing the time by L_{PP}/V_0;

These values are taken for the 10°/10° and for the 20°/20° overshoot angles. The first overshoot angle in the 10°/10° is typically small for frigate-like naval ships, giving large percentage errors. The values for the 20°/20° are more meaningful for frigate-like ships. However, because FREDYN was to be extended to other ships, the 10°/10° results are kept in the procedure: in the future, ships outside of the validity range shall be included.

Fig. 8 shows the results for the first overshoot angle in the 20°/20° zigzag test. The figure shows the results for the ships inside and outside the validity range.

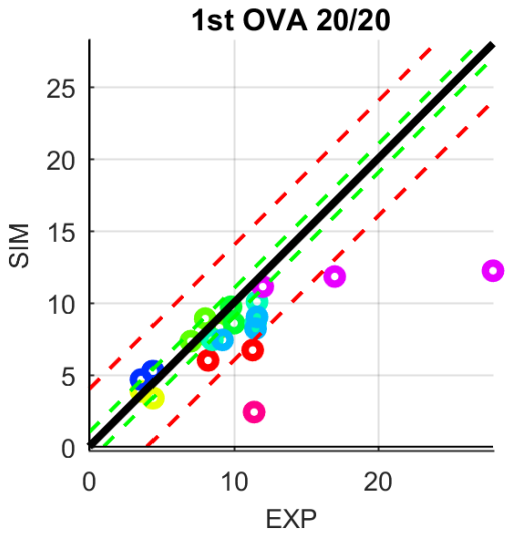

Fig. 8, QQ plot of overshoot angles together with the tolerance and limits.

Tolerance and limits are defined based on similar assumption on the quality of the experimental data. Considering that some of these data were gathered a while ago, the values of tolerance are chosen quite high. The tolerance and limit for the first overshoot angles in the 20°/20° zigzag maneuver *OVA1* are set as follows:

If $OVA1_{EXP} < 10$ then Tolerance = 1°
If $OVA1_{EXP} < 10$ then Tolerance = $0.1 \cdot OVA1_{EXP}$ °

For the limit, a similar relationship is defined:

If $OVA1_{EXP} < 8$ then limit = 4°
If $OVA1_{EXP} < 8$ then limit = $0.5 \cdot OVA1_{EXP}$ °

The ships inside the validity range have an average score of 0.60. The ships outside the validity range have a score of 0.18.

Motions in Regular Waves

When sailing in regular waves, the motions of the ship are usually

expressed in response amplitude operators (RAOs), the ratio between the amplitude of the first harmonic of a motion and the incident wave.

Determining whether a simulation program is able to predict these motions is critical. This should be done at various speeds, various frequencies and various wave directions. To derive the performance in a single score or quality number, there are several challenges.

Fig. 9 shows a tentative example of a RAO of heave, as obtained by linear strip theory (SHIPMO, blue dotted), by model tests (EXPERIMENT, blue line) and FREDYN (green line).

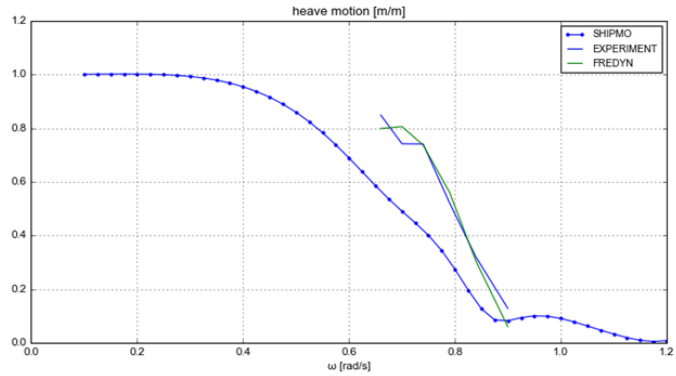

Fig. 9, Example RAO for heave motion

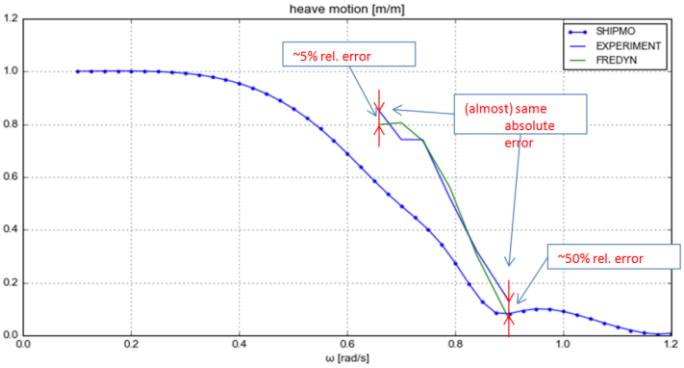

Fig. 10, RAO with indication of prediction error

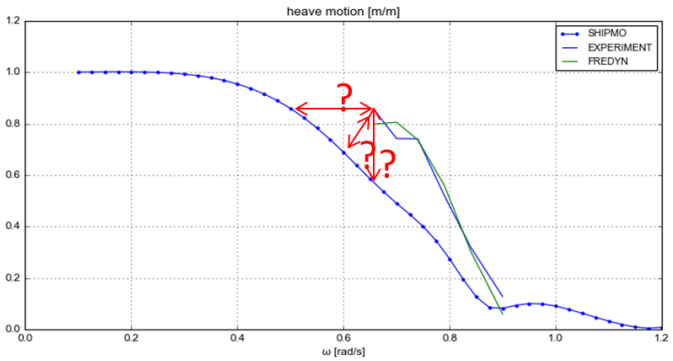

Fig. 11, Alternatives to compare RAO curves

After investigating a number of alternatives, the following procedure was developed:

- The FREDYN simulations were performed at the same wave frequencies as the available experimental data;
- The Mean Absolute Error (MAE) was determined over the entire available frequency range for each motion component and per speed-heading combination:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} = \frac{\sum_{i=1}^{n}|e_i|}{n} \quad (8)$$

Where e_i is the error value of prediction value x_i versus measurement value y_i per frequency component i. The resulting MAE value (RMEA) was divided over the largest value of the maximum of the measured RAO values y and 1.

$$RMAE = \frac{MAE}{max(y,1)} \quad (9)$$

This was done because the rotational motions are expressed as degrees/meter, leading to a quite large range in maximum RAO values from much smaller than 1 for the yaw motions to much larger than one for the roll motions.

In the end, per condition (combination of ship, heading, and speed) and motion mode (surge, sway, heave, roll, pitch, and yaw) one RMAE value is obtained. This value expresses the relative error of the prediction versus the measurement for one RAO curve.

Similarly, there was interest in not only knowing how well the RAOs were predicted but also whether there was an over- or underestimation of the RAO. Therefore, not only the absolute error was of interest, but also the bias in the prediction. This was done by calculating the mean error (*ME*): the value of the prediction error with sign (without the absolute value):

$$ME = \frac{\sum_{i=1}^{n} y_i - x_i}{n} = \frac{\sum_{i=1}^{n} e_i}{n} \quad (10)$$

And then the relative mean error (*RME*):

$$RME = \frac{ME}{max(y,1)} \quad (11)$$

This resulted in a 'signed' error. On its own this error is not that interesting, as a combination of over- and under-prediction could still lead to a (close to) zero RME value. But combined with the RMAE it does reveal if there is a bias in the predicted RAO curve.

For 411 RAO points (5 ships inside the validity range, and in total for 9 different wave directions), the performance is quantified in this way. An illustration is given for one ship (the well-known 5415M) in Table 1. This shows that the average quality number is 0.79. It also shows that roll is the least well predicted (with an average score of 0.49), while heave, pitch, and surge are predicted well.

Table 1. Validation results for one ship (5415M)

Motion response		RMAE	Quality Number	Bias
Surge	x_a/ζ_a	0.08	0.87	-0.03
Sway	y_a/ζ_a	0.14	0.72	-0.13
Heave	z_a/ζ_a	0.07	0.96	-0.03
Roll	ϕ_a/ζ_a	0.21	0.49	-0.09
Pitch	θ_a/ζ_a	0.10	0.87	0.06
Yaw	ψ_a/ζ_a	0.11	0.82	-0.05
Average			0.79	

A QQ plot of this is presented in Fig. 12, showing in this case only heave (the best score) and roll (the worst score). Apart from the numbers in Table 1, the plots in Fig. 12 show how close the simulations and experiments are to the bisector line, but also which ships are responsible for a worse performance. In the case of the roll RAO, this will be the yellow ship and the dark blue ship, while other ships are well-predicted.

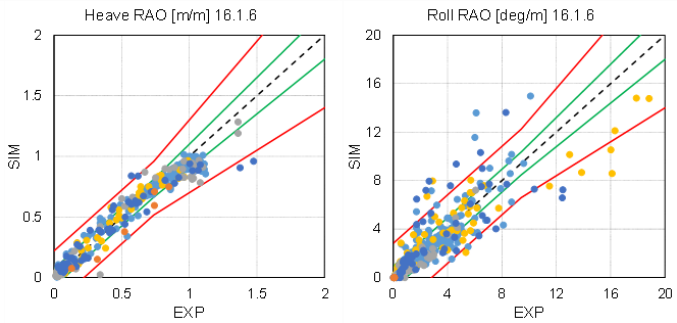

Fig. 12, QQ plot for the heave (left) and roll (right) RAOs

Motions in Irregular Waves

Simulation of sailing in irregular seas comes close to the real operation of ships. This is the end-to-end validation on the motions in a seaway.

A wealth of data was gathered from the results of experiments in irregular seas. Some of the considerations to select key characteristics are given below:

- Not all data is robust enough for comparison with simulations. An example is the surge motion versus the surge velocity and the mean velocity. When the standard deviation of surge velocity is taken, this neglects the mean (2nd-order) shift. For a code like FREDYN, aiming at mainly 1st-order wave frequency behavior, this is desired.
- The wave height is dominant in the response of a value. To make the results less subjective, the values should be made non-dimensional with the wave.
- For motions in stern-quartering waves, we are especially interested in motions that lead to loss of stability, broaching, and course keeping ability (or lack thereof). Therefore, motions related to the heading variations and the related rudder angles are of importance. This means it is also important that the simulations are performed using exactly the same autopilot as used in the experiments. This requires proper documentation, which is not always available.
- The results are restricted to unidirectional waves and, also to conditions where the exact wave height is known. This requires that the tests are performed in a suitable seakeeping basin. Full scale measurements are usually not sufficient to register all these phenomena with a low measurement uncertainty.

Points of attention are the following: the waves in the basin have a target wave height *Hs* and wave period *Tp* (for example *Hs* = 4 m, *Tp* = 9.7 s) and an achieved/measured wave (for example *Hs* = 3.9 m, *Tp* = 9.4 s). All this is within 5%. However, when performing simulations, the actually measured wave conditions should be used, because the results are sensitive to this when performing a validation study.

Based on this consideration and after several iterations, the following list of 11 key characteristics are selected:

- Standard deviation of the surge velocity [m/s].
- Standard deviation of the sway velocity [m/s].
- Standard deviation of the heave motion [m].
- Standard deviation of the roll motion [deg].
- Standard deviation of the pitch motion [deg].
- Standard deviation of the yaw motion [deg].
- Standard deviation of the rudder angle [deg].
- Average rudder angle [deg].
- Average heading angle [deg].
- Mean speed [knots]
- Standard deviation of the speed [knots]

Noted is that these are all dimensional quantities. This is done on purpose.

QQ plots for each of the key characteristics are produced. Fig. 13 shows this for 2 of the 11 key characteristics. As in the previous section, the values of Fig. 14 are for 5415M, but the figures in Fig. 13 are for all ships and all wave directions.

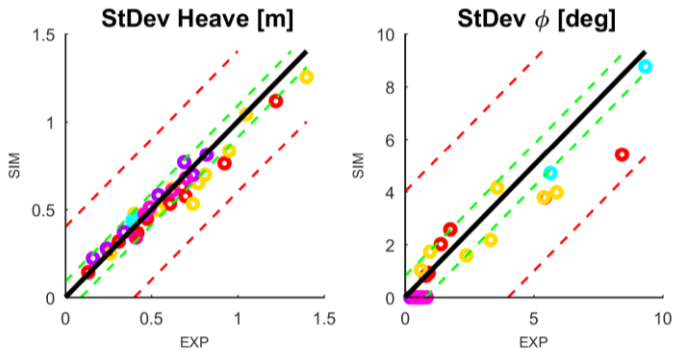

Fig. 13, Standard deviation of heave motion and standard deviation of roll angle, all directions, all speeds

Data like Fig. 13 are insightful as they tell us for which ships the predictions are satisfactory. It helps in defining the validity

range of the simulations because it shows for which ships the predictions are good.

However, to gain insight in the physics to aid code development and code improvement, other plots are needed which show the prediction of key characteristics against meaningful parameters. In the case of motions in irregular waves, these are the wave directions. Such a plot is shown in Fig. 14.

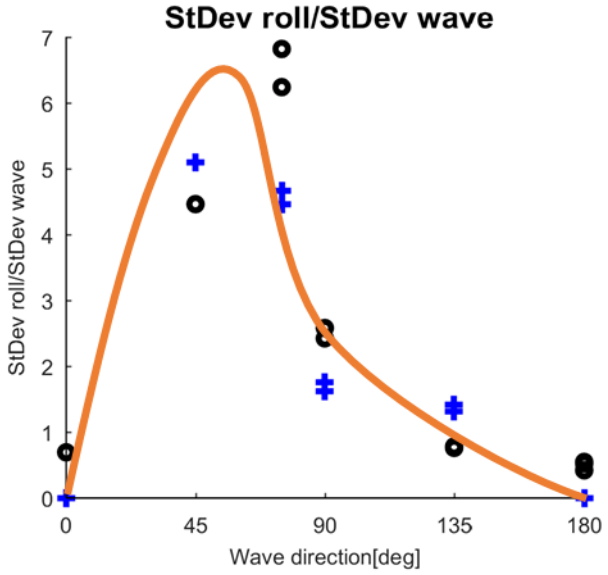

Fig. 14, Experimental and simulated results of the standard deviation of roll plotted against wave direction

The tolerance and limit are now determined in a slightly different way. For the motions heave and roll (the cases of Fig. 13) the values are not a function of the experiments. The following values are used for 4 of the 11 key characteristics:

Table 2. Used values for tolerance and limit for motions in irregular waves

	Tolerance	Limit
Heave	0.09	0.40
Roll	0.8	4.0
Pitch	0.15	0.80
Yaw	0.40	1.50

Table 3. Validation results for motions in irregular waves for one ship (5415M)

Motion response	Quality Number
Head seas	0.887
Bow quartering	0.861
Beam	0.738
Stern quartering	0.927
Stern	1.000
Average	0.883

Using these tolerances and limits, the quality number is calculated. To obtain better insight in the performance in various wave directions, the summation is not only carried out globally, but also per wave direction. For the 5415M, this quality number is shown in Table 3.

TAKEAWAY

The methodology described in this paper has been applied to the FREDYN code developed within CRNav, though it is applicable to any similar time-domain ship motion prediction tool.

The evaluation of the quality score can be implemented in an automated way such that:

- A baseline can be determined for a given set of validation data, key characteristics, (subjective) tolerance levels, etc.
- The baseline can be used to track progress in terms of overall accuracy, for example when new numerical methods are developed or input data are changed.
- The baseline can be used in regression testing, when code changes are made but the implementation is intended to be the same, for example during a refactor.
- A new baseline can be determined when validation data is added (or removed), or the key characteristics/tolerance levels are changed.

These aspects are useful when managing a large, complex code base with many persons contributing. It also gives some confidence to the many organizations depending on reliable and stable code, or indications of improvement when changes are implemented.

CONCLUSIONS

Performing a validation of a prediction code for multiple ships is quite different than a validation for one ship, or even multiple codes and one ship. To be able to do this, a methodology is defined and described in this paper. This methodology is applied to FREDYN, and the authors are quite pleased with this methodology. It is designed such that it can be applied on other prediction codes and can be extended to include other and more ships.

The methodology consists of:

1. Selection of appropriate maneuvers;
2. Selection of key characteristics;
3. Creation of QQ plots;
4. Quantifying tolerance and limits for each key characteristic and prediction error;
5. Definition of the quality number.

This methodology is instrumental to:

- Define the validity range of a prediction code;
- Define objectively of the improvement of the code with increasing version numbers (regression testing).

The validation methodology offers a comprehensive, consistent, and traceable view of the accuracy of the simulation tool, offering significant value for developers, users, and other stakeholders.

ACKNOWLEDGEMENTS

The authors acknowledge the permission of CRNav community to share the insights of the validation of FREDYN with the international community.

The results presented in this paper are of a slightly older version of FREDYN. The values of tolerances and limits are in use at this moment, but may change in the future. For the demonstration of the methodology however, this is not of importance.

REFERENCES

K. Eloot, G. Delefortrie, M. Vantorre and F. Quadvlieg, *Validation of Ship Manoeuvring in Shallow Water Through Free-running Tests.* OMAE ASME 34th International Conference on Ocean, Offshore and Arctic Engineering, St. John's, Newfoundland, Canada, 2015

Hooft, J. P., and Pieffers, J. B. M., "*Maneuverability of frigates in waves*". Marine Technology, Vol. 25, No. 4. 1988

Hooft, J.P. "The cross flow drag on a manoeuvring ship" Ocean Engineering, Vol. 21, pp 329-342, 1994

Hooft, J.P. and Quadvlieg, F.H.H.A.. Non-linear hydrodynamic hull forces derived from segmented model tests. International Conference on Marine Simulation and Ship Manoeuvrability (MARSIM'96), Copenhagen, Denmark, 1996

De Kat, J.O., Brouwer, R., McTaggart, K. and Thomas, W.L., *Intact Snip Survivability in Extreme Waves: New Criteria from a Research and Navy Perspective*, Proc. International Conference on Stability of Ships and Ocean vehicles STAB "94, Melbourne, FL, Nov. 1994

De Kat, J.O. and Thomas, W.L., *Model Tests for Validation of Numerical Capsize Predictions*, Proc. 4th Stability Workshop, St. John's, Newfoundland, Oct. 1998

De Kat, J.O. and Peters, A.; *Model experiments and simulations of a damaged frigate*. 10th Congress of International Maritime Association of the Mediterranean (IMAM 2002), Crete, Greece, 2002

Kisjes, A., Quadvlieg, F., Ferrari, V.; *Uncertainty analysis of free running manoeuvring model test on a modern ferry, with emphasis on heel angles*; International Conference on Ocean, Offshore and Artic Engineering (OMAE2019), Glasgow, Scotland, UK, 9 June, 2019

McTaggart, K. and de Kat, J.O.; *Capsize Risk of Intact Frigates in Irregular Seas*. SNAME Annual Meeting, Vancouver, B.C., Canada, 2000

Palermo, E., Tonelli, R., Quadvlieg, F.H.H.A., Scharnke, J., Drummen, I. and Teigen, P.; *Validation of a software tool (DROPSIM) to predict the drop and sailaway behavior of a lifeboat*. OMAE2015-42144. Proceedings of OMAE 2015 the ASME 34th International Conference on Ocean, Offshore and Arctic Engineering May 31 – June 5, 2015, St. John's, NL, Canada

Quadvlieg, F.H.H.A. and Brouwer, J., *KVLCC2 Benchmark Data Including Uncertainty Analysis To Support Manoeuvring Predictions*. IV International Conference on Computational Methods in Marine Engineering (ECCOMAS MARINE), Lisbon, Portugal, 2011

Quadvlieg, F., Simonsen, C., Otzen, J., Stern, F.; *Review of the SIMMAN2014 Workshop on the state of the art of prediction techniques for ship maneuverability*. international Conference on Ship Maneuverability and Maritime Simulation (MARSIM2015), Newcastle, UK, September 2015

Quadvlieg, F.H.H.A. and Rapuc, S.; *A pragmatic method to simulate manoeuvring in waves*. SNAME SMC, November 2019, Tacoma, WA, USA

Quadvlieg, F.H.H.A., Stern, F., Kim, Y-G, Flensborg-Otzen, J., Yasukawa, H., Wang, W. and Han, Y. *Proceedings of SIMMAN 2020 workshop on validation on manoeuvring prediction methods*, Korea, May 2024.

Stern, F., Agdrup, K., Kim, S.Y., Cura Hochbaum, A., Rhee, K.P., Quadvlieg, F.H.H.A., Perdon, P., Hino, T., Broglia, R., Gorski, J.; Lessons learnt from the workshop on verification and validation of ship manoeuvring simulation methods - SIMMAN2008. International Conference on Marine Simulation and Ship Maneuverability (MARSIM 2009), Panama City, Panama, 17 August 2009

Stern, F., Agdrup, K., Kim, S.Y., Cura Hochbaum, A., Rhee, K.P., Quadvlieg, F.H.H.A., Perdon, P., Hino, T., Broglia, R. and Gorski, J., Experience from SIMMAN 2008-The First Workshop on Verification and Validation of Ship Maneuvering Simulation Methods Journal of Ship Research, 2011

Stern, F., Diez, M., Sadat-Hosseini, H., Yoon, H., and Quadvlieg, F., 2018, "Statistical Approach for Computational Fluid Dynamics State-of-the-Art Assessment: N-Version Verification and Validation," ASME J. Verif. Validation Uncertainty Quantif., Sep 2017, 2(3) p. 031004. (https://doi.org/10.1115/1.4038255)