Journal of Navigation (2025), 1–13
doi:10.1017/S0373463325000219

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

An approach to establishing evaluation criteria for unmanned surface vessel manoeuvres during COLREGs encounters

Michael E. Heistand and Craig M. Payne

Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA
Corresponding author: Michael E. Heistand; Email: michael.heistand@jhuapl.edu

Received: 23 July 2024; **Revised:** 27 November 2024; **Accepted:** 08 April 2025

Keywords: COLREGs; ship behaviour; unmanned surface vessel

Abstract
The question arises when developing and testing Unmanned Surface Vessel (USV) Manoeuvring Autonomy (MA): 'is the performance we are seeing in our current on-water tests better than that of the last autonomy software version we deployed?' An approach to answer this question is inspired by educators' rubrics, in which a teacher grades a student's work to objective criteria and then sums the individual criteria to determine the student's overall grade. Here, individual metrics are used to evaluate a USV manoeuvring within range of another vessel. A weighted average is then applied to determine the overall score. With that objective performance value now obtained, similar manoeuvring tests can be compared between autonomy software versions to determine if the autonomy under development is progressively improving. This paper does not determine the threshold score needed to establish that a USV is safe to operate; thresholding of sufficient performance is recommended for future study.

1. Introduction

When developing Unmanned Surface Vessel (USV) Manoeuvring Autonomy (MA), it is essential to establish a framework for scoring and analysis. As MA is changed and new behaviours are developed, performance measurements demonstrate whether the changes do or do not improve performance, and if changes to one behaviour may have unforeseen effects on other behaviours. A core set of test cases support regression testing following autonomy subsystem updates.

This paper proposes a method for evaluating the performance of MA with the following goals in mind.

- Establish a framework for scoring and analysis.
- Standardise tools, techniques and data collection/format/archival used in assessment of MA.
- Establish a set of metrics that enable performance evaluation and trending of MA versions throughout development and at any stage of maturity.
- Provide objective methods for the comparison of different vessels or different MAs on the same vessel, while adhering to nautical rules of the road.
- Provide guidance on developing tailored test plans/strategies to achieve desired confidence.

This method does not encompass all aspects of MA. Future development of additional metrics and evaluation methods will support new advancements in MA and increases in its sophistication.

2. Scope of evaluation criteria

This paper is an example of the assessment of the performance of MA as it applies to a subset of the International Maritime Organization (IMO) Convention on the International Regulations for Preventing Collisions at Sea (COLREGs) (U.S. Coast Guard, 1999) between two vessels approaching one another on the open ocean.

COLREGs provide specific guidance, 'Rules of the Road', regarding two vessels interacting within sight of one another, and there exists a Risk of Collision, as defined in Rule 7. Rule 8 (Action to Avoid Collison) is a general rule that provides guidance on qualities of good seamanship. Specific rules that cover One-on-One vessel interactions are Rule 13 (Overtaking), Rule 14 (Head-on), Rule 15 (Crossing), Rule 16 (Action by Give-way Vessel), Rule 17 (Action by Stand-on Vessel) and Rule 18 (Responsibility Between Vessels).

Other non-vessel hazards such as non-moving, both charted and uncharted, are not considered. Non-vessels that may be drifting are considered, since rudimentary perception systems classify such objects that have both a speed over ground and a direction as an actual vessel, albeit a relatively slow one in comparison to an underway vessel.[1]

This paper only applies to a subset of manoeuvre operations that ocean-going vessels might be engaged in with emphasis on safety, primarily collision avoidance and hazard avoidance, and efficiency/performance of manoeuvres initiated for safety reasons. The focus is on the manoeuvring rules contained within the IMO COLREGs Rules 7, 8 and 13–18, with certain limitations.

Two aspects of manoeuvring rules, which are intertwined with these COLREGs rules, are not directly scored using the methodology of this paper: Safe Speed and In-Extremis[2] situations. Safe speed is assumed to be continuously evaluated, and the speed of the vessel when the encounter begins is assumed to be considered safe. In Extremis situations, because MA is designed to maintain a minimum Closest Point of Approach (CPA) outside close quarters, In-Extremis manoeuvring is assumed to be handled and evaluated separately.

Additionally, this methodology does not address COLREGs Rules 19–37, the capabilities and implications of interpreting nautical lights, sound signals, day shapes and Bridge-to-Bridge communications with other vessels. Lastly, the evaluation criteria covered here are not meant to cover the implications of Rule 2, Special Circumstances (Multiple Vessel Encounters), which gives seamen leeway to use good judgment in handling special circumstances rather than forcing strict adherence to the COLREGs rules:

> 'nothing in the Rules shall exonerate any vessel, or the owner, master or crew thereof, from the consequences of any neglect'
>
> 'all dangers of navigation and collision and to any special circumstances, including limitations of the vessels involved, which may make a departure from these Rules necessary to avoid immediate danger'

The target system boundary for assessments is drawn around the software component(s) responsible for hazard and collision avoidance, assuming that is sufficiently separable from components requesting desired manoeuvres without consideration of hazards and contacts. This boundary can be readily established in simulation-based testing. During at-sea testing, it may be difficult to isolate the target system from supporting systems. However, contributions to performance by supporting systems will be consistent and the root cause of poor performance may be determined through analysis. For example, if a perception system has a tracking latency, this is isolated through analysis, and these latency effects can be accounted for when assessing MA performance.

[1] A perception system or human designation is needed to differentiate a drifting hazard from a vessel. Once so designated, MA can perform the appropriate action. If not, MA may treat non-charted, slow-moving objects as vessels.

[2] In-Extremis is defined when a situation exists such that to avoid a collision both vessels shall take action – this is Rule 17.

https://doi.org/10.1017/S0373463325000219 Published online by Cambridge University Press

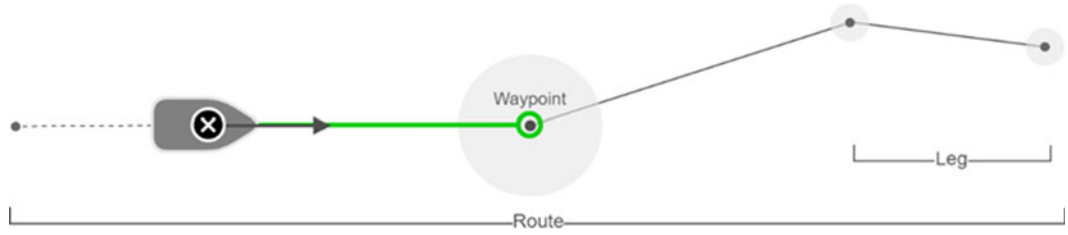


Figure 1. *General route illustration.*

3. Nautical rules covered and explained

3.1 Covering just the manoeuvring rules

Since the general use of terminology tends to be non-standard in this subject area, this section defines the meaning and usage of key terms within this document.

3.1.1. Route or transit behaviour

Route or transit behaviour in MA is used to navigate a vessel from one geographic location to another using waypoints. When navigating for long durations, waypoints are planned in sequence along a route. The vessel will automatically transit between sequential waypoints, achieving each, until reaching the last planned waypoint, as illustrated in Figure 1. Upon achieving the last waypoint, the vessel will automatically transition to the next planned behaviour. In the case when no task is planned following the last waypoint, the vessel would execute a default behaviour.

3.1.2. One-on-one vessel COLREGs interactions

COLGREGs Rules 7, 8 and 13–18 define the 'Rules of the Road' for One-on-One vessel interactions. These rules are then the basis for which vessel MA must be able to follow in practice and spirit for safe operations.

Ranges, relative positions, orientation, and time and distance to the CPA of the contact vessel are extremely important to the mariner for proper manoeuvres to occur in an expected seaman-like manner. As such, these considerations are assumed to be ensconced in the MA design and software code. Relative positions and orientation are specifically accounted for in the COLREGs rules. However, apart from *'in sight of one another'* and '*so close that collision cannot be avoided'*, COLREGs do not generally define ranges in which manoeuvring should occur or ranges for avoidance. Therefore, it is incumbent on the mariner to determine these ranges. Every situation is different and relies on a mariner's experience for the determination of ranges. Ship captains provide standing orders to their manoeuvring officers regarding the ranges expected to be observed, and these ranges can change. Likewise, MA developers are in need of similar range parameters, which are outlined below (see Figure 2). The defined ranges used by the MA are not necessarily the same ranges the other vessel is using; in other words, one mariner's estimation of what is close may not be the same as other mariners.

Reference distances:

- Observation Area: The area within the vessel's sensor Line of Sight (LOS) or Evaluation Range.
- Minimum CPA: Minimum range that should be achieved during an encounter. Vessel will manoeuvre to keep contacts outside of this minimum distance. Minimum CPA is a configurable value set by the operator.

Range definitions:

https://doi.org/10.1017/S0373463325000219 Published online by Cambridge University Press

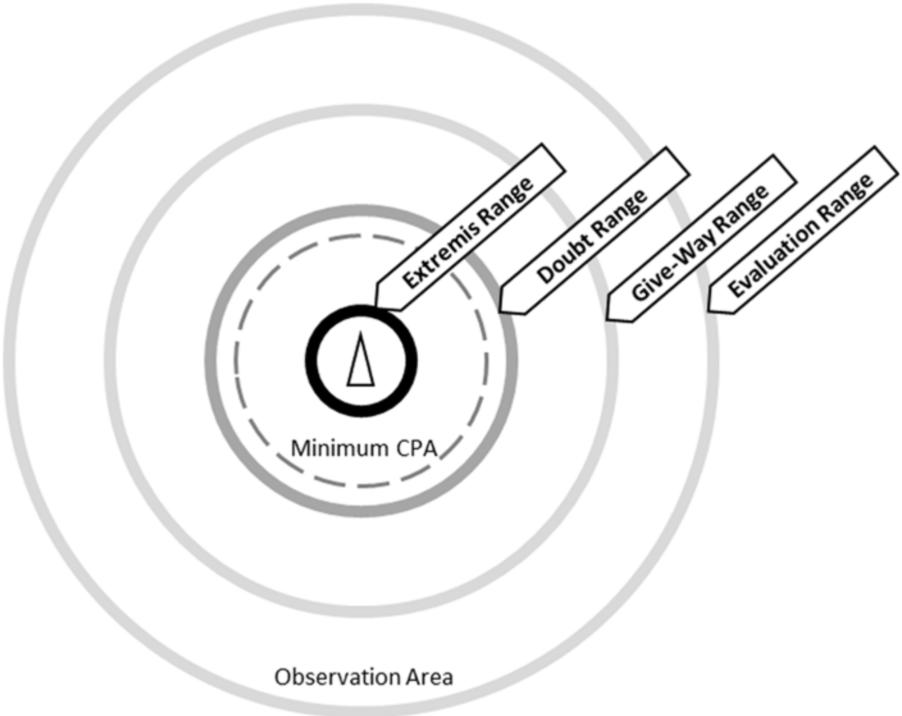


Figure 2. *Range definition.*

- Evaluation Range: Maximum range at which the vessel detects and begins to evaluate contacts. Evaluation Range is variable, based on sensor performance and environmental conditions.
- Give-Way Range: Distance at which a vessel may begin to manoeuvre as the give-way vessel. Give-Way Range is variable, based on the known and/or estimated characteristics of the vessels involved.
- Doubt Range: Distance at which the vessel, despite being the stand-on vessel, will begin to take action to remain outside of Minimum CPA. Doubt Range is variable, based on the known and/or estimated characteristics of the vessels involved.
- Extremis Range: Distance at which the vessel will take action to avoid or reduce the effects of a collision. Extremis Range is a fixed value based on vessel characteristics.

All regions are shown as circular in Figure 2 for the purposes of simple visualisation; however, manoeuvre autonomies are not precluded from establishing complex-curve ranges to account for varying uncertainty in differing meeting situations and relative bearings.

4. Metrics development

4.1 COLREGs manoeuvring rules compliance evaluation

This COLREGs scoring algorithm uses position and vector data from ownship and contact vessels to calculate individual metric scores on a scale of 0 to 1. A final overall run score is then calculated using a weighted average of the metric scores, which results in an overall evaluation between 0 and 1.

https://doi.org/10.1017/S0373463325000219 Published online by Cambridge University Press

Table 1. *COLREGs engagement types*

COLREGs engagements	Description
Overtaking	The vessel is approaching another vessel from behind 'from a direction more than 22.5 degrees abaft her beam'. COLREGs Rule 13 and Multiple Vessel Encounters (U.S. Coast Guard, 1999)
Being overtaken	The vessel is being approached from behind by another vessel. COLREGs Rule 13 (U.S. Coast Guard, 1999)
Head-on	Two vessels are 'meeting on reciprocal or nearly reciprocal course' and 'each shall alter her course to starboard so that each shall pass on the port side of the other'. COLREGs Rule 14 (U.S. Coast Guard, 1999)
Crossing give-way	The vessel is approaching another vessel on the starboard (right) side and is to 'take early and substantial action to keep well clear' of the other vessel. COLREGs Rules 15 and 16 (U.S. Coast Guard, 1999)
Crossing stand-on	The vessel is being approached from the port (left) side and is to 'keep her course and speed'. COLREGs Rules 15 and 17 (U.S. Coast Guard, 1999)
Multiple vessel encounters (not considered in this document)	Encounters with two or more other vessel nearly simultaneously such that normal COLREGs type manoeuvring must not be done sequentially. COLREGs Rule 2 (U.S. Coast Guard, 1999). Multiple Vessel Encounters are discussed by Stankiewicz et al. (2020)

4.1.1. COLREGs manoeuvring metrics

It is important to quantify the performance of the MA beyond strictly subjective subject matter expert (SME) evaluation. Early in the development of a prototype MA, SME evaluation is often used to provide guidance to developers regarding what manoeuvres a mariner would find acceptable or unacceptable. However, as MA matures, it becomes challenging to assess, with SME evaluation alone, whether a newer version of the MA is better than the previous version: a binary, qualitative evaluation may not be adequate. Quantifiable metrics are needed to provide a more granular level of evaluation and a means of comparing identical simulation runs or nearly identical on-water runs between MA versions.

COLREGs engagements can be categorised into separate cases in the analysis: Overtaking, Being Overtaken, Crossing Stand-on, Crossing Give-way, Head-on and manoeuvring for Multiple Vessels Encounters, as listed in Table 1. For each, the analysis includes scenarios in which the other vessel is compliant and noncompliant with the rules. Regardless of the type of interaction with another vessel, COLREGs Rule 8 describes *'Action to Avoid Collision'*. These actions represent what would be considered good practice and were used to assess the recommended actions of the MA during all engagements.

To evaluate each situation, the proposed scoring methodology defines 12 metrics across three categories: COLREGs Performance, Mission and Safety. Each metric is scored on a scale of 0 to 1. An overall score is calculated using a weighted average of these scores, as shown in Table 2.

$$CC = min\left(\frac{|Course\ Change|}{30}, 1\right) \tag{1}$$

https://doi.org/10.1017/S0373463325000219 Published online by Cambridge University Press

Table 2. *COLREGs metrics*

Category	Metric name	Definition	Metric motivation
COLREGs performance	Course change (CC) ratio	Magnitude of initial course change as a ratio of 30° [3,4]; See Equation (1)	Metric is used to reflect 'be positive' aspect of Rule 8(a), 'large enough to be readily apparent' aspect of Rule 8(b) and all of Rule 8(c) (U.S. Coast Guard, 1999)
COLREGs performance	Speed change (SC) ratio	Magnitude of first speed change as a ratio of 5 knots[5]; See Equation (2)	Metric is used to reflect 'be positive' aspect of Rule 8(a) and 'large enough to be readily apparent' aspect of Rule 8(b) (U.S. Coast Guard, 1999)
COLREGs performance	Number of course changes (numCC)	Number of course changes during the encounter; cosine function drops in score for 1 to 10 course changes; see Equation (3)	Metric is used to reflect aspect of Rule 8(b) 'a succession of small alterations of course and/or speed should be avoided' (U.S. Coast Guard, 1999)
COLREGs performance	Number of speed changes (numSC)	Number of speed changes during the encounter; cosine function drops in score for 1 to 10 speed changes; see Equation (4)	Metric is used to reflect aspect of Rule 8(b) 'a succession of small alterations of course and/or speed should be avoided' (U.S. Coast Guard, 1999)
COLREGs performance	Course indecision (CI)	Time spent changing between multiple courses as a ratio of duration of encounter; see Equation (5)	Metric is used to reflect aspect of Rule 8(a) 'observance of good seamanship' and aspect of Rule 8(b) 'a succession of small alterations of course and/or speed should be avoided' (U.S. Coast Guard, 1999)
COLREGs performance	Speed indecision (SI)	Time spent changing between multiple speeds as a ratio of duration of encounter; see Equation (6)	Metric is used to reflect aspect of Rule 8(a) 'observance of good seamanship' and aspect of Rule 8(b) 'a succession of small alterations of course and/or speed should be avoided' (U.S. Coast Guard, 1999)
Safety	Minimum range (MinRng) ratio	Shortest range to another vessel as a ratio of risk range; see Equation (7)	Since time, speed and distance are coupled, this metric is used to reflect aspect of Rule 8(a) 'made in ample time', aspect of Rule

(*Continued*)

[3] 'Case law defines apparent course manoeuvres to consist of a minimum of 35 deg turn while common practice often requires no less than 30 deg of heading change' (Woerner, 2016).

[4] A 35-degree minimum turn could easily be substituted here depending upon the type of vessels and the situation.

[5] Speed of 5 knots was chosen here as visual changes to bow wave and wake changes would be very difficult to discern below a change of 5 knots.

https://doi.org/10.1017/S0373463325000219 Published online by Cambridge University Press

Table 2. (*Continued*)

Category	Metric name	Definition	Metric motivation
			8(c) 'made in good time', aspects of Rule 8(d) 'passing at a safe distance', 'past and clear', and aspect of Rule 8(f)(i) 'allow sufficient sea room' (U.S. Coast Guard, 1999)
Safety	Crossing the bow (CB) ratio	Range of a bow crossing as a ratio of the distance the contact would travel at current speed in 6 minutes (0.1 × contact speed in knots)[6]; see Equation (8)	Metric is used to account for aspects of Rule 8(f) 'not to impede the passage or safe passage' and 'sufficient sea room' (U.S. Coast Guard, 1999)
Mission	Distance travelled (DIS) ratio	Actual distance travelled as a ratio of the optimal distance to travel; see Equation (9)	Metric is used to reflect the distance efficiency of the chosen solution
Mission	Time late (T) ratio	Actual time to get to next waypoint as a ratio of expected travel time; see Equation (10)	Metric is used to reflect the time efficiency of the chosen solution

$$SC = min\left(\frac{|Speed\ Change\ (kts)|}{5}, 1\right) \tag{2}$$

$$\#CC = \max\left(\cos\left(\frac{Num\ Course\ Changes}{6}\right), 0\right) \tag{3}$$

$$\#SC = \max\left(\cos\left(\frac{Num\ Speed\ Changes}{6}\right), 0\right) \tag{4}$$

$$CI = \cos\left(\frac{2 * \pi * Indecisive\ Time}{Min\ Range\ Time}\right) \tag{5}$$

$$SI = \cos\left(\frac{2 * \pi * Indecisive\ Time}{Min\ Range\ Time}\right) \tag{6}$$

$$DIS = \frac{Straight\ Line\ Distance}{Actual\ Distance\ Traveled} \tag{7}$$

[6] Six minutes was used in this example as it is a standard nautical 'Rule of Thumb'; there is no codified time to pass another ship's bow, except as provided in Rule 8(f). Other times can certainly be applied depending upon the types of vessels and the situation.

https://doi.org/10.1017/S0373463325000219 Published online by Cambridge University Press

If goal waypoint not achieved, $DIS = 0$

$$T = \frac{\textit{Straight Line}\ \text{Travel Trime}}{\textit{Actual Travel Time}} \tag{8}$$

$$MinRng = \frac{Range}{Range\ Threshold} \tag{9}$$

$$CB = \min\left(\frac{Range\ (nm)}{Contact\ Speed\ (kts) * 0.1}, 1\right) \tag{10}$$

4.1.2. COLREGs manoeuvring measures of performance

Safety is emphasised by increasing the weight of the minimum range metric and crossing bow metric by a factor of 3 and 2, respectively. All other metrics are weighted equally. The overall score is described in Equation (11). Additionally, the score will be automatically failed (0) if the T, MinRng or CB score is 0.

$$Score_{1v1} = \frac{(\mathrm{CC} + \mathrm{SC} + \mathrm{numCC} + \mathrm{numSC} + \mathrm{CI} + \mathrm{SI} + \mathrm{DIS} + \mathrm{T} + 3 * \mathrm{MinRng} + 2 * \mathrm{CB})}{13} \tag{11}$$

5. Design of at-sea testing

When designing at-sea tests, every effort was made to create repeatable controlled situations that would enable direct comparisons of the scores for successive MA versions. The vehicle's world model, or its view of everything around it, was free of other contacts and hazards except for those controlled and included in the test. Test days were chosen based on forecasted environmental conditions that would have minimal impact on vessel actions, i.e. low sea state and light winds.

Test scenarios included head-on, crossing and overtaking engagements of Ownship (OS)[7] in both give-way and stand-on positions as well as scenarios in which the Contact Vessel (CV) is compliant and non-compliant. Additional scenarios included the edge cases[8] in which these engagement types converge (i.e. where crossing becomes head-on). Care was taken in the test setup as small variations in parameters can have dramatic effects on the outcome of the test.

The initial relative bearings, headings and speeds of OS and CV were chosen to result in a CPA that violates the OS minimum CPA threshold. To isolate MA from perception, the initial range from OS to CV were distant enough that the SUT had a fully populated world model and had a stable track on CV before it enters the COLREGs range.

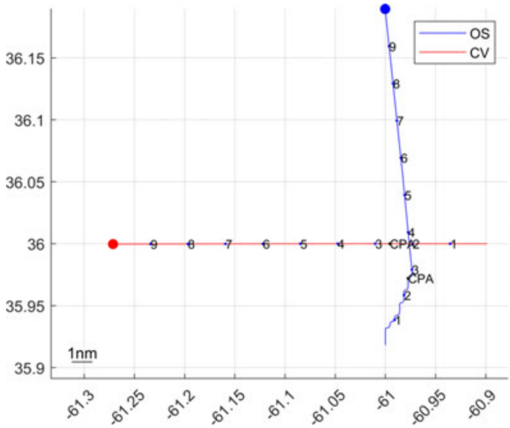
6. Data collection

Guidance for data collection during on-water testing is necessary to ensure accurate scoring can be performed. Omitted or missing data for contact vessels that may have impacted MA decision making will lead to inaccurate scoring results. Therefore, the following data were collected for all vessels within detection range of the OS or within the full operational area, whichever is larger.

- Position: Latitude and Longitude [World Geodetic System 1984 (WGS84)] in positive and negative decimal degrees to at least six decimal places.
- Velocity: Speed over ground in meters per second and course over ground in clockwise positive degrees relative to True North.

[7] Ownship here is the system under test.
[8] The boundaries of edge cases are best explored in simulation in which the variables can be controlled.

https://doi.org/10.1017/S0373463325000219 Published online by Cambridge University Press



CC	1.00
SC	1.00
numCC	0.24
numSC	0.67
CI	0.00
SI	0.00
Dis	0.95
T	0.89
MinRng	0.95
CB	1.00
Score	0.74

Figure 3. *Crossing give-way Scenario 1.*

In the absence of position/vector data, Automatic Identification System (AIS) data were used. Additionally, the locations of buoys and other static hazards were recorded. However, in certain scenarios, AIS position and reporting intervals may not be precise enough or frequent enough to be relied upon to generate an accurate score.

7. Example results

Example results are shared below to demonstrate the application of the scoring algorithm. These are simulated results using a basic ‘if-then’ decision tree for autonomy and ‘cookie cutter’ sensors.

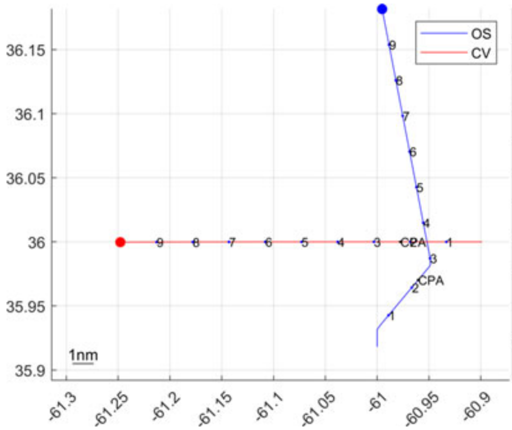
Figure 3 shows a crossing scenario example in which OS is the give-way vessel and CV is the stand-on vessel. The small numbers on the track lines indicate time steps and the CPA is noted. The individual metric scores and overall score are shown to the right of plot.

In this example, a collision is avoided and the OS correctly gives way, but the behaviour is not ideal. OS initiates several manoeuvres back and forth, and several speed changes between the start and time point 3 before passing behind the contact and coming back on course. The exhibited behaviour could cause confusion for the other vessel, injecting more risk into the situation, which shows unfavourably in the COLREGs score. The large number of course and speed changes in a short time affects the ‘number of course changes’ (numCC), ‘course indecision’ (CI), ‘number of speed changes’ (SC) and ‘speed indecision’ (SI) metrics. Additionally, OS violates the 2-nm standoff range and comes within 1.89 nm of CV, resulting in a lower MinRng metric score. The overall score of 0.74 suggests there is room for improvement in the handling of the scenario.

Figure 4 shows a similar crossing scenario. In this example, OS makes a single, obvious course change, maintains speed, passes behind the contact and then comes back on course. OS does violate the 2-nm standoff range by coming within 1.97 nm of CV, but the overall score of 0.99 reflects a more favourable behaviour performance.

The metrics and overall score calculations were developed for the purpose of evaluating results relative to each other. For example, assume scenario 1 above is version 1.0 of an MA. The overall score reveals a problem in the MA, but the individual metrics reveal the deficiencies in the course and speed change logic of the MA. Changes are made to the MA and OS is tested again in scenario 2. The overall score shows that the issues have been addressed and the individual metric scores confirm that the changes did not affect the other metrics.

https://doi.org/10.1017/S0373463325000219 Published online by Cambridge University Press



CC	1.00
SC	1.00
numCC	1.00
numSC	1.00
CI	1.00
SI	1.00
Dis	0.93
T	0.93
MinRng	0.99
CB	1.00
Score	0.99

Figure 4. *Crossing give-way scenario 2.*

8. Next steps

8.1. Establishing 'passing' threshold

Defining thresholds for acceptance is a necessary next step for evaluation of MA. This method of evaluation was intended to provide relative scoring between runs or different MAs. Further research is needed to define a 'passing' threshold.

One option for defining a passing threshold is to determine if the MA is 'as good as a human'. The first step is to align the existing COLREGs evaluation methods with the scoring criteria used by Navy instructors at the Surface Warfare School Command (SWSC). Once the scores from the evaluation method correlate with the grades from the instructors, the resulting evaluation methodology will then be applied to data from SWSC training simulators to score the navigation proficiency of a large distribution of certified human ship handlers. An MA can be tested in the same scenarios from the SWSC simulation. Using the newly aligned COLREGs evaluation methods, the MA can then be quantitatively compared with this distribution of human performance, ultimately resulting in data-driven justifications for whether USV MA is 'as good as a competent mariner'.

8.2. Application to simulation

Another next step is to apply the MA scoring to simulation. A significant benefit to this scoring approach is that the method applies to simulation-based testing in addition to real-world testing. Simulations of MA, regardless of being a real-time or non-real-time simulation, have the potential to generate large amounts of runs in comparatively less time than on-water testing. Rapid evaluation using simulation supports continuous development and improvement of a COLREGs capability, particularly if simulated tests and metric scoring can be automated to score performance directly upon updating the software. Individual runs of concern may be flagged for review, evaluation, improvement and retesting, calling on qualified mariner review when needed. The scoring approach is well-suited to adaptive sampling methods, which may lead to finding performance limitations or areas for improvement. Improvement in the COLREGs score over time provides an indicator of when the hazard avoidance software may be ready for installation and on-water testing.

8.3. Evaluation of safe speed and in-extremis

Lastly, scoring will need to be developed to apply to other manoeuvring situations such as Safe Speed and In-Extremis. In general, if two vessels come close to each other at sea, adherence by both to the

https://doi.org/10.1017/S0373463325000219 Published online by Cambridge University Press

COLREGs manoeuvring rules should prevent a collision. There are certainly exceptional circumstances, such as equipment casualties, reduced visibility, confined waters or combinations of such factors. When only one of the vessels is following the rules, the probability of collision is much higher. The COLREGs accounts for this possibility; in particular, Rule 17(a)(ii) authorises a stand-on vessel to take action as soon as it becomes apparent that the give-way vessel is not taking appropriate action. Moreover, Rule 17(c) states *'A power-driven vessel which takes action in a crossing situation in accordance with 17(a)(ii) of this Rule to avoid collision with another power-driven vessel shall if the circumstances of the case admit, not alter course to port of a vessel on her own port side'*. At a later point in the developing encounter, collision can only be avoided by simultaneous action from both vessels. At that point, the vessels are said to be 'In-Extremis', and both vessels have the responsibility to do whatever it takes to avoid the collision. This In-Extremis situation is covered by Rule 17(b): *'When, from any cause the vessel required to keep her course and speed finds herself so close that collision cannot be avoided by the action of the give-way vessel alone, she shall take such action as well best aid to avoided collision'*. This essentially means that both vessels must manoeuvre to avoid collision, and presumably there is a coordinated pairing of best courses and speeds by the two vessels to do so.

The range at which the vessels are considered to be In-Extremis depends on the *'circumstances of the case'* with factors such as manoeuvrability, size and relative speed of the vessels being critical. Since we are considering the problem from the perspective of the vessel with MA, we assume that the vessel with MA is the stand-on vessel and the other vessel is the give-way vessel. While the other vessel's manoeuvring characteristics are unknown, the vessel with MA can quickly compute the outcomes of multiple course and speed combinations using the assumption that the other vessel does not manoeuvre. By definition, if there is a course and speed combination for the vessel with MA that avoids collision when the other vessel maintains course and speed, the vessels are not yet In-Extremis. Once no such course and speed combination exists, the vessels are In-Extremis.

An alternate method is to assume conservatively that the other vessel's tactical diameter is five[9] times that vessel's length and that the turning characteristics of both vessels are symmetrical. Additionally, avoiding collision requires that no part of either vessel will touch during the manoeuvre. Given these assumptions, Extremis range can be computed by Equation (12). Figure 5 illustrates the outer boundary of Extremis range.

$$\text{Extremis Range} = 5 \times \frac{\textit{Own ship length}}{2} + \textit{Own ship length} + \textit{Contact length} \tag{12}$$

Guidance on how an In-Extremis manoeuvre is conducted is not entirely clear. The practice of not turning to port for a vessel on the port side has in common practice and training been applied to the In-Extremis situation as described in Naval Ship handling[10] (Crenshaw, 1984). Court rulings have been mixed with a majority of the rulings that have not found at fault the previous stand-on vessel, which is now In-Extremis, for initiating a port turn to a vessel on her port side. However, in some cases, a turn to port was deemed a contributing cause. See Farwell's Rules of the Nautical Road, pages 101–103 and 223–225 (Allen, 2020).

Rule 6 of COLREGs is 'Safe Speed', and states essentially a safe speed is necessary to reduce the risk of collision. This should be evaluated prior to any manoeuvre to determine if the entering speed is acceptable. If the speed is too high or 'unsafe', any action after that is already insufficient to avoid collision. Rule 6 states *'Every vessel shall at all times proceed at a safe speed so that she can take*

[9] Naval Architecture 'rule of thumb' and IMO guideline.

[10] Page 444. The stand-on vessel shall 'keep her course and speed', but not into the jaws of catastrophe. She may manoeuvre when it appears that the other vessel is not taking appropriate action, but she should not turn to port if the other vessel is to port.

https://doi.org/10.1017/S0373463325000219 Published online by Cambridge University Press

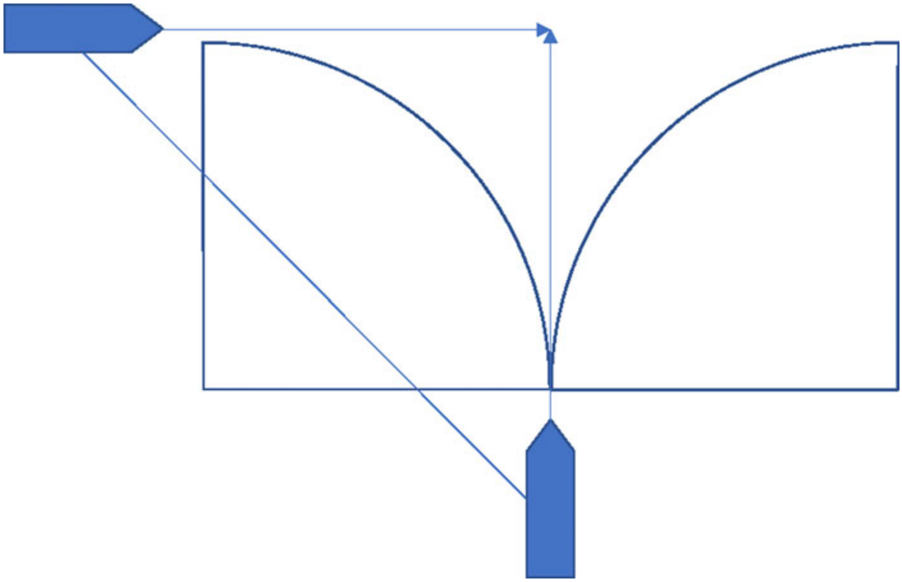


Figure 5. *Scaled illustration of extremis range.*

proper and effective action to avoid collision and be stopped within a distance appropriate to the prevailing circumstances and conditions'. Also, Rule 19 (b) is '*Every vessel shall proceed at a safe speed adapted to the prevailing circumstances and conditions of restricted visibility*'. A rule of thumb is that safe speed is 'fast enough to manoeuvre but slow enough to stop in time'.

Adjusting to a safe and reasonable speed should be initiated prior to enacting any manoeuvring; therefore, it is recommended that COLREGs Rule 6 'Safe Speed' be assessed continually by MA.

9. Conclusion

This paper shows an approach to evaluation of USV manoeuvring based on objective criteria that provides repeatable results such that successive versions of software can be assessed. Additionally, this methodology can be applied to large-scale simulations of autonomy to assess potential problem areas quickly. Because of the transparency of this approach to evaluation, the underlying functions that score the individual metrics can be changed, as can the overall weighted scoring scheme, and additional metrics can be applied and new scores calculated with the same underlining test data.

A cautionary caveat: these metrics are meant to evaluate the shortfalls or failures of USVs to follow basic manoeuvring rules contained in COLREGs. They are not intended to be an all-encompassing set of metrics to evaluate COLREGs compliance. Further development is needed to address the full set COLREGs rules.

Funding statement. The authors would like to acknowledge the support and funding for this work from the United States Navy Office of Naval Research, Advanced Naval Platforms Division, Program Officer Dr. Robert Brizzolara. Under NAVSEA contract N00024-22-D-6404, TA SSQAH and NAVSEA Contract N00024-13-D-6400, Task Order 0771.

Competing interests. The authors declare none.

https://doi.org/10.1017/S0373463325000219 Published online by Cambridge University Press

References

Allen, C. H. (2020). *Farwell's Rules of the Nautical Road*, *Ninth Edition*. Naval Institute Press.

Crenshaw, R. S. (1974). *Naval Shiphandling*, *Fourth Edition*. Naval Institue Press.

Stankiewicz, P. G., Heistand, M. and Kobilarov, M. (2020). Quantifying Good Seamanship for Autonomous Surface Vessel Performance Evaluation. *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 8309–8315.

U.S. Coast Guard. (1999). *Navigation Rules*. United States Department of Transportation.

Woerner, K. (2016). Multi-contact protocol-constrained collision avoidance for autonomous marine vehicles. *Doctoral dissertation,* Massachusetts Institute of Technology.

Cite this article: Heistand, M.E. and Payne, C.M. (2025). An approach to establishing evaluation criteria for unmanned surface vessel manoeuvres during COLREGs encounters. *Journal of Navigation* 1–13. https://doi.org/10.1017/S0373463325000219