Quantifying Good Seamanship For Autonomous Surface Vessel Performance Evaluation

Paul Stankiewicz[1,2], Michael Heistand[1], and Marin Kobilarov[2]

Abstract— **The current state-of-the-art for testing and evaluation of autonomous surface vehicle (ASV) decision-making is currently limited to one-versus-one vessel interactions by determining compliance with the International Regulations for Prevention of Collisions at Sea, referred to as COLREGS. Strict measurement of COLREGS compliance, however, loses value in multi-vessel encounters, as there can be conflicting rules which make determining compliance extremely subjective. This work proposes several performance metrics to evaluate ASV decision-making based on the concept of "good seamanship," a practice which generalizes to multi-vessel encounters. Methodology for quantifying good seamanship is presented based on the criteria of reducing the overall collision risk of the situation and taking early, appropriate actions. Case study simulation results are presented to showcase the seamanship performance criteria against different ASV planning strategies.**

I. Introduction

As autonomous vehicle software matures, these systems will be expected to safely navigate in increasingly complex scenarios. However, as the complexity of the missions increases, the possibility of emergent behavior and unexpected performance increases as well. Therefore, robust testing and evaluation (T&E) methods are needed to ensure safety and build trust in the underlying decision-making engine.

With regards to autonomous surface vessel (ASV) technology, robust performance includes a combination of many competing objectives including mission completion, collision avoidance, and predictable maneuvering in a manner that an experienced ship captain would exhibit. For this last expectation, the International Regulations for Prevention of Collisions at Sea [1], referred to as COLREGS, contain a set of maneuvering protocols which all vessels must follow to reduce confusion in a potential collision encounter. This paper assumes that ASVs must comply with these COLREGS maneuvering rules in a manner identical to human ship drivers. The difficulty for ASVs, however, is that while COLREGS provide guidance on expected behavior, the rules were intentionally left vague so as to let human intuition and common sense ultimately drive the decision-making to avoid a collision.

This paper does not approach ASV decision-making from a planning perspective, but rather from a performance evaluation perspective. Evaluating whether fielded ASV software is compliant with COLREGS maneuvering requirements is a challenging problem, where it becomes necessary to perform an *objective* evaluation based on *subjective* COLREGS protocols. Some recent attempts at quantifying COLREGS compliance are provided in [2], [3]. This problem becomes even more challenging when performing evaluation for multi-vessel encounters. In these instances, it is understood that COLREGS are interpreted more as guidelines to good seamanship as opposed to a set of rules which must be strictly followed. To accommodate performance evaluation on multi-vessel scenarios, this work aims to generalize the notion of quantifying COLREGS compliance to the notion of quantifying good seamanship. Practicing good seamanship is based on how much the ASV reduces the overall collision risk and whether the system takes early, appropriate action.

To the best of the authors' knowledge, this paper offers some of the first research into quantifying proper seamanship behavior for the purposes of ASV performance evaluation in complicated multi-vessel scenarios. Additionally, some novel future collision risk indices are provided based on maximum mutual ship domain violation. We envision the performance metrics proposed in this paper can be used not only in a standalone fashion, but also as a key component within a larger evaluation suite for ASV maneuvering performance, perhaps supplementing some of the rule-based evaluation provided in [2], [3]. The remainder of this paper is organized as follows: Section II provides related work, Section III describes the approach to quantifying good seamanship for ASV evaluation, and Section IV provides a case study analysis in evaluating different ASV planning strategies in a multi-vessel scenario.

II. Related Work

A. Navigation Safety Assessment

There have been many studies over the past two decades that assess the collision risk between two oncoming vessels, with general surveys of maritime risk assessment provided in [4], [5]. Traditional methods for assessing collision risk consist of algorithms derived from the closest point of approach (CPA), defined as the location where two moving objects with fixed velocity vectors reach their minimum separation distance. For example, the works of Bukhari et al. [6] and Perera et al. [7] both assess collision risk based on fuzzy inference of CPA-based indices. Meanwhile, the work in [8] applies Dempster-Shafer theory to combine multiple CPA-based indices into an overall collision risk assessment.

A parallel line of research has focused on performing maritime safety assessment using the concept of ship domain.

[1]P. Stankiewicz and M. Heistand are with the Johns Hopkins Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel MD, 20723, USA. paul.stankiewicz@jhuapl.edu, michael.heistand@jhuapl.edu

[2]P. Stankiewicz and M. Kobilarov are with the Department of Mechanical Engineering, Johns Hopkins University, 3400 N Charles Street, Baltimore, MD 21218, USA. pstanki2@jhu.edu, marin@jhu.edu

The traditional definition of ship domain as defined by Goodwin [9] is “the surrounding effective waters which the navigator of a ship wants to keep clear of other ships or fixed objects.” With regards to safety assessment, various authors have applied ship domain in a binary fashion (i.e. any object within the vessel’s domain is considered a threat to navigational safety) [10], [11], while others have adopted a more continuous risk assessment, known as a fuzzy ship domain, based on the object’s degree of penetration into the vessel’s domain [12], [13], [14]. He et al. [15] use multiple ship domains that depend on the COLREGS situation to determine collision avoidance maneuvers. An excellent review of different ship domain models and their applications is presented in [16].

B. *COLREGS Maneuvering Compliance Evaluation*

COLREGS maneuvering compliance can be thought of as a subcategory to navigational safety assessment that evaluates not the collision risk of an encounter, but whether the actions of each vessel were appropriate based on the COLREGS protocols of [1]. While there have been a significant number of studies into ASV path planning approaches that consider COLREGS [17], [18], [19], [20], there is much less literature on techniques for evaluating the COLREGS compliance of different planning algorithms. A more recent and complete analysis of COLREGS evaluation was performed by Woerner et al. [3], [2]. In this work, algorithms are defined that evaluate the actions taken by a vessel for each COLREGS rule specified in [1]. This procedure provides a more principled analysis for comparing observed behavior against the expected behavior in different COLREGS scenarios. This analysis framework was used by Minne [21] to compare multiple ASV navigation strategies and by Stankiewicz et al. [22] to determine failure modes and performance boundaries in ASV decision-making. While these works provide a path forward for one-versus-one COLREGS encounters, the rule-based nature of the scoring criteria does not generalize well to multi-vessel scenarios.

III. Approach

The approach presented here aims to evaluate ASV decision-making by determining whether its actions constitute good seamanship. An appropriate quantification of good seamanship should capture the essence of COLREGS maneuvering compliance without subjecting the analysis to rule-based heuristics that may not be appropriate in multi-vessel encounters, e.g. when there are conflicting give-way and stand-on expectations. As such, it should be emphasized that this methodology is targeted towards multi-vessel encounters as a supplement to existing COLREGS evaluation [2], [3], where specific COLREGS rule classification loses value due to potentially conflicting rules for each target ship.

Seamanship evaluation is quantified here based on the combination of two categories: (1) how the ASV reduces the overall collision risk to all vessels involved in an encounter and (2) whether the ASV takes early, appropriate action. The first category is rooted in the collision avoidance requirements of all mariners while the second category captures the essence of responsibility intended by COLREGS.

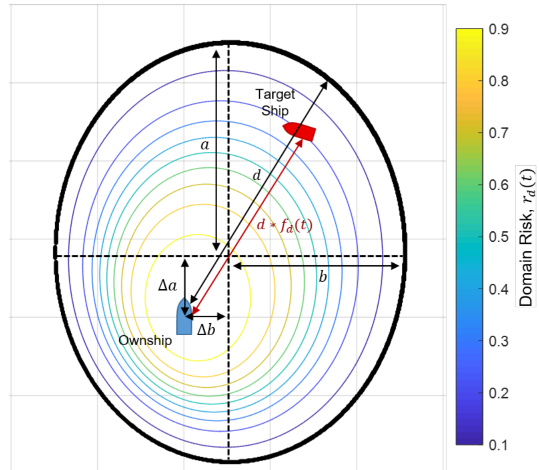

Fig. 1: Decentralized ellipse ship domain, where d is the distance to the edge of ownship’s domain in the direction of the target ship. The fuzzy ship domain which translates domain violation $f_d(t)$ into a risk value is shown by the inner contour plots.

A. *Collision Risk Quantification*

1) Ship Domain: This paper makes use of the concept of ship domain to quantify the overall collision risk between multiple vessels. As briefly discussed in Section II, ship domain describes an area around each vessel which should be kept free of other vessels and has been widely used in the maritime research community for ship safety assessment. An appropriate definition of the ship domain geometry can overcome several of the weaknesses associated with CPA-based risk assessment. This is because ship domain geometry can be described by any polygon, typically one which emphasizes keeping the fore and starboard sectors of the vessel clear, as illustrated in Fig. 1.

Existing ship domains proposed in the literature generally define the geometry based on three different methods: analytical [13], [23], empirical from ship movement data [11], and those based on artificial intelligence from expert knowledge [14], [24]. Because ship domain has been used in various applications, different authors adopt slightly varying definitions of ship domain and there is no universally accepted geometry. This work utilizes ship domain with the following assumptions:

- The ship domain is designed for an open water setting, although it could be modified for use in traffic separation schemes or waters with constricted maneuvering.
- Following [16], the domains of all vessels in an encounter should be kept clear, i.e. ownship should not violate the domain of any target ship and likewise each target ship should not violate ownship’s domain.

This work adopts a decentralized ellipse as the ship domain (Fig. 1) similar to [25], albeit with different dimensions. While various authors have proposed complex polygonal domain geometries, many of these can be approximated by a decentralized ellipse. Further, as discussed in [25], the governing equations of a decentralized ellipse can be solved analytically while still creating a domain that emphasizes COLREGS maneuvering compliance, e.g. by favoring port-to-port maneuvers or those which cross astern of the target ship. As shown in Fig. 1, the geometry of this domain is described by the ellipse axis lengths a and b, and displacements from the ellipse center Δa and Δb. Values for these parameters are approximated according to the following:

$$a(t) = \tfrac{1}{2}\big(R_{f,d}(t) + R_{a,d}(t)\big), \tag{1}$$
$$b(t) = \tfrac{1}{2}\big(R_{s,d}(t) + R_{p,d}(t)\big), \tag{2}$$
$$\Delta a(t) = R_{f,d}(t) - a(t), \tag{3}$$
$$\Delta b(t) = R_{s,d}(t) - b(t), \tag{4}$$

where $R_{f,d}(t)$, $R_{a,d}(t)$, $R_{s,d}(t)$, and $R_{p,d}(t)$ are radii for the fore, aft, starboard, and port sectors of the domain, respectively. Values for these radii are configurable and can be affected by factors such as vessel size, vessel speed, environmental conditions, the skill of the mariner, etc. Following [26] and [13] with slight modifications for a more conservative domain size, the values of these radii are chosen based on the ship's evasive maneuvering characteristics including the length of the ship in nautical miles, L, speed in knots, $V(t)$, advance, A_D, and tactical diameter, D_T:

$$R_{f,d}(t) = \left(1 + 1.34\sqrt{k_{A_D}(t)^2 + \big(0.5k_{D_T}(t)\big)^2}\right)2L, \tag{5}$$
$$R_{a,d}(t) = \left(1 + 0.67\sqrt{k_{A_D}(t)^2 + \big(0.5k_{D_T}(t)\big)^2}\right)2L, \tag{6}$$
$$R_{s,d}(t) = \big(0.2 + k_{D_T}(t)\big)2L, \tag{7}$$
$$R_{p,d}(t) = \big(0.2 + 0.75k_{D_T}(t)\big)2L, \tag{8}$$

where

$$k_{A_D}(t) = A_D(t)/L \approx 10^{(0.3591\log_{10}V(t)+0.0952)}, \tag{9}$$
$$k_{D_T}(t) = D_T(t)/L \approx 10^{(0.5441\log_{10}V(t)-0.0795)}. \tag{10}$$

The equations above depend only on $V(t)$ and L which are assumed to be known for ownship and estimated for a target ship through Automatic Identification System (AIS) or other perception systems.

2) Collision Index: Collision risk has traditionally been assessed by utilizing the distance at closest point of approach (DCPA) and the time remaining to reach the closest point of approach (TCPA). While these parameters are intuitive, they do not differentiate between the risk associated with different vessel geometries. For example, consider two CPA configurations with identical DCPA values: one in which ownship is directly off the bow of the target ship and one in which ownship is directly astern of the target ship. Even though these encounters have identical DCPA values, it is obvious that the collision risk is significantly less when ownship passes astern of the target vessel.

To alleviate the problems with CPA-based risk assessment, Szlapczynski introduced a new measure for collision risk derived from the concept of ship domain [12]. At every time instant t, collision risk can be assessed based on a scale factor, $f_d(t)$, of the largest domain-shaped area that is free from other vessels, i.e. after scaling ownship's domain by $f_d(t)$, the target ship will be on the boundary of the scaled domain. Values for $f_d(t) < 1$ can then represent the degree of penetration by a target ship into ownship's domain as shown in Fig. 1. The work in [25] provides analytical formula for calculating $f_d(t)$ for a decentralized ellipse ship domain, however, $f_d(t)$ could also be determined for arbitrary domain geometries as described in [12]. Thus, the reader is free to substitute any ship domain for use with this methodology.

The domain scale factor can then be translated into a domain risk index, $r_d \in [0, 1]$, using a logistic function according to the following:

$$r_d(t) = \frac{1}{1 + e^{k(f_d(t) - f_{50})}}, \tag{11}$$

where k and f_{50} are parameters that define the shape of the logistic curve. Equation 11 was chosen based on [27] which found that the safety perceived by mariners is roughly proportional to the logarithm of the vessel separation distance. This work uses $k = 10$ and $f_{50} = 0.5$ based on tuning experiments and the resulting domain risk is shown by the inner contour plots of Fig. 1.

Several works only consider violations of ownship domain [13], [14], [23] or target ship domain [10], [25] when assessing current and future collision risk. These interpretations, however, are incomplete, as the actual collision risk between two vessels should be identical regardless of each vessel's perspective. Thus, we define the combined mutual domain risk between the ASV and the i-th target ship as such:

$$r_d^{A,i}(t) = r_d^A(t) + r_d^i(t)\left(1 - r_d^A(t)\right), \tag{12}$$

where r_d^A is calculated from the ASV perspective and r_d^i is calculated from the i-th target ship perspective.

Subsequently, the collision index between the ASV and the i-th target ship at time t is defined as the maximum value of the mutual domain risk over a future time horizon T:

$$\Theta_C^i(t) = \max_{\tau\in[t,t+T]} r_d^{A,i}(\tau). \tag{13}$$

This optimization can be easily solved numerically by assuming constant speed and heading for each vessel.

B. Appropriate Action Quantification

1) Ship Arena: Similar to the concept of ship domain is the notion of ship arena, defined as the area around ownship where a mariner should begin maneuvering if a collision risk exists. The ship arena naturally encompasses a larger area than the ship domain, as any evasive action should be planned and executed well before violations of each ship's domain. Additionally, defining a geometry which preferences earlier action to vessels in ownship's fore and starboard sectors, the ship arena can capture the give-way / stand-on expectations governed by COLREGS.

This work again chooses a decentralized ellipse to define the ship arena, albeit with different dimensions of $R_{f,a}$, $R_{a,a}$, $R_{s,a}$, and $R_{p,a}$. It is typical for mariners to internally set predefined distances at which they begin to consider

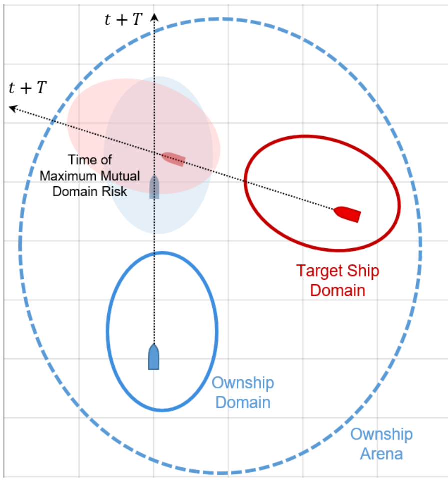

Fig. 2: Illustration of overall risk methodology. Θ_C^i is calculated as the maximum mutual domain risk with the i-th target ship over a future time horizon $t+T$. Θ_A^i is calculated based on the i-th target ship's degree of penetration within ownship's arena.

maneuvering actions. In give-way situations, taking early action is preferred, sometimes right when a target ship is reliably detected on radar or AIS. Conversely, in stand-on situations, mariners are expected to maintain course and speed until the target ship has been deemed noncompliant, i.e. the target ship does not perform its own maneuvering action. Evasive action in these stand-on situations would then only occur at distances much smaller than the detection or give-way range. It should be emphasized that the dimensions of exactly when a ship should begin maneuvering are strictly based on the preferences of the mariner. There is a dearth of literature on more principled methods to make these parameter selections; thus, in the analysis of Section IV, values for $R_{f,a}$, $R_{a,a}$, $R_{s,a}$, and $R_{p,a}$ are set arbitrarily based on the size of the vessels. Future research should certainly guide the community into more accepted standards for defining the ship arena based on the vessel classification (e.g. sailing vessel, etc.) and relative speed, as these values all affect when evasive action should be considered.

2) Action Index: Once the ship arena is defined, the action index with respect to the i-th target ship, $\Theta_A^i(t)$, is calculated in a similar fashion to r_d, i.e.

$$\Theta_A^i(t) = \frac{1}{1+e^{k(f_a^i(t)-f_{50})}}, \tag{14}$$

where now the value $f_a^i(t)$ is the ship arena scale factor such that the i-th target ship lies on the boundary of ownship's scaled ship arena. This representation can be thought of as the degree to which ownship should take action and appropriately penalizes delayed avoidance maneuvers. When compared to the calculation of $\Theta_C^i(t)$, it should be noted that $\Theta_A^i(t)$ is evaluated for each target ship only from the perspective of the ASV.

C. Overall Risk Index

We can now define an overall risk index for the i-th target ship as

$$\Theta_S^i(t) = \Theta_C^i(t)\Theta_A^i(t). \tag{15}$$

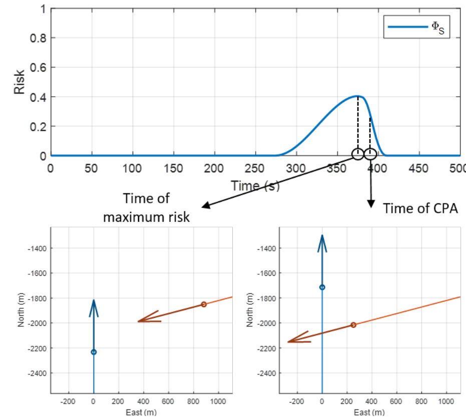

Fig. 3: Overall risk curve for a two vessel crossing scenario. Snapshots of the vessel configuration are given at the time of maximum overall risk and the time of minimum DCPA.

As visualized in Fig. 2, the combination of $\Theta_C^i(t)$ and $\Theta_A^i(t)$ is important for evaluating seamanship - even if a target ship is within the ASV ship arena, the overall risk should be low if there is no future collision risk. This overall risk index is able to better capture periods of high risk in the ASV trajectory when compared to CPA-based methods. Specifically, the time and location of maximum collision risk is not necessarily the same as the time and location of CPA. For a vessel crossing the bow of another vessel, it is likely that the point of maximum collision risk occurs well before CPA. Figure 3 illustrates this scenario with two snapshots of vessel geometry: one at the point of maximum risk and another at CPA. It is clear that the overall risk between the two vessels is significantly less in the CPA configuration even though the vessels are closer together.

For a multi-vessel scenario involving $i = 1, \ldots, N$ target ships, the risk associated with each target ship can be combined to capture the risk associated with the overall scenario, $\Phi_S(t)$. As opposed to using the average value of all $\Theta_S^i(t)$, this work proposes the overall scenario risk should be calculated as the union of individual risk indices through the following recursion:

WHILE $i \leq N$

$$\Phi_S(t) = \begin{cases} \Theta_S^i(t) & \text{if } i = 1 \\ \Theta_S^i(t) + \Phi_S(t)\left(1-\Theta_S^i(t)\right) & \text{if } 1 < i \leq N \end{cases} \tag{16}$$

The logic behind this formula is that the overall scenario risk should be *at least* as large as the highest risk from the i-th target ship, with additional risk from other vessels only augmenting the value of $\Phi_S(t)$.

D. Seamanship Evaluation

For the purposes of evaluating ASV decision-making on post-processed trajectories, it is beneficial to define several performance metrics related to measuring the seamanship of the ASV over the scenario. Ideally, for an ASV exhibiting good seamanship, $\Phi_S(t)$ should be kept as close to zero as possible, indicating minimal future collision risk and/or the ASV taking early evasive action. The resulting seamanship performance metric, $\mathcal{S}$, could then potentially be any function of $\Phi_S(t)$ desired by the evaluator, i.e. $\mathcal{S} = g(\Phi_S(t))$.

This section offers one possible procedure for evaluating seamanship performance over the encounter. First, let $\Phi_{S,max}$ be the maximum overall risk of the ASV:

$$\Phi_{S,max} = \max_{t \in [t_s, t_f]} \Phi_S(t), \tag{17}$$

where t_s and t_f are the start and final times of the encounter, respectively. Also, let $\hat{t} = (t - t_s)/(t_f - t_s)$ be the normalized time vector such that $\hat{t} \in [0, 1]$. Two performance metrics can then be defined that measure the maximum risk (Eq. 18) and the cumulative risk (Eq. 19) over the encounter:

$$J_M = 1 - \Phi_{S,max}, \tag{18}$$

$$J_C = 1 - \int_0^1 \frac{\Phi_S(\hat{t})}{\Phi_{S,max}} \, d\hat{t}. \tag{19}$$

The calculation of J_C represents a score for the amount of risk acquired during the encounter, normalized by the maximum risk. Thus, a value of $J_C = 0$ would mean that the ASV maintained the maximum risk for the entirety of the encounter. This metric is able to capture sustained periods of high risk as well as indecision that leads to multiple spikes in the risk.

Finally, the seamanship performance score is calculated as a combination of J_M and J_C:

$$\mathcal{S} = J_M\big(1 + \alpha(2J_C - 1)(1 - J_M)\big). \tag{20}$$

The form of Eq. 20 rewards short periods of low risk and penalizes sustained periods of high risk, where α serves as a tuning parameter that controls the amount of penalty based on the value of J_C. This work uses a value of $\alpha = 0.75$ based on tuning experiments. The seamanship score of Eq. 20 could then be used to supplement existing COLREGS-based ASV evaluation (e.g. [2], [3]) to accommodate general best practices in multi-vessel scenarios.

IV. Case Study Analysis

This section performs a case study to analyze the proposed seamanship evaluation on a canonical scenario for different ASV planning strategies. Because COLREGS protocols become extremely subjective in multi-vessel scenarios, it is difficult to benchmark the evaluation procedure against ground truth performance, as there are typically no "correct" maneuvers that serve as a baseline. Rather, decision-making is more readily evaluated on a relative basis, i.e. comparing one set of actions in a scenario to a different set of actions in the same scenario. For this case study, a three-vessel roundabout scenario was chosen where all vessels are approaching a coincident collision point from 120-deg course offsets at speeds of 20 knots. This scenario was chosen because evaluating COLREGS maneuvering compliance for each vessel individually would be inappropriate due to the fact that, according to COLREGS, each vessel would be expected to give-way to the target ship off its starboard bow, while somehow standing-on to the target ship off its port bow. Additionally, conventional knowledge regarding this scenario does in fact offer an expected resolution, one where each vessel should treat the encounter as a roundabout and proceed in a counterclockwise fashion.

TABLE I: Performance metrics for each ASV planning method

ASV Planning Method	J_M	J_C	$\mathcal{S}$
Ideal Maneuvering (Fig. 4)	0.98	0.99	0.99
VO w/ COLREGS (Fig. 5)	0.62	0.95	0.78
VO w/ COLREGS (Mean of 100 randomized scenarios)	0.52	0.94	0.69

A. Description of ASV Under Test

Because the focus of this study is not a novel ASV planning algorithm, only high-level descriptions of the ASV planning architectures are provided similar to how an evaluator would treat the system as a black box. Two different ASV planning strategies are evaluated for the roundabout scenario: (1) idealized planning where each vessel makes early maneuvers in a counterclockwise fashion and (2) planning which utilizes velocity obstacles (VO) supplemented with COLREGS protocols (similar to the approach in [28]). The COLREGS planning module classifies the COLREGS encounter for each target ship, deems whether ownship should give-way or stand-on, and plans the resulting trajectory based on desired avoidance parameters. For the purposes of this analysis, each vessel in the simulation operates according to the same autonomy logic, albeit with different planning configuration parameters so as to introduce variations in their behaviors. Without loss of generality, the length of each vessel is set to $L = 40$ m and the ship arena radii for the ASV are set to $R_f = R_s = 1$ nmi and $R_a = R_p = 0.5$ nmi, although different values or equations dependent on additional factors could easily be applied instead.

B. Results

The ship trajectories and risk curves for the idealized planning method and the VO with COLREGS planning method are shown in Fig. 4 and Fig. 5, respectively. Additionally, the performance metrics of Section III-D for each ASV planning method are provided in Table I.

For the idealized planning scenario, each vessel makes a 30-deg maneuver to starboard once a target ship is detected at the edge of the ship arena. As shown by the risk plots of Fig 4, this maneuver by the ASV reduces the collision index $\Theta_C^i(t)$ for each target ship and subsequently the overall scenario risk $\Phi_S(t)$ to zero by eliminating future ship domain violations. Thus, as shown at $t = 250$ sec, even though target ship #2 remains within the ASV ship arena, the overall scenario risk is negligible since there is no future domain violation. This desired behavior is reflected by the high seamanship score of $\mathcal{S} = 0.99$ shown in Table I.

Conversely as shown in Fig. 5, the ASV using the VO with COLREGS planner is not able to properly reconcile the conflicting give-way and stand-on objectives from the two target ships. This conflict within the planner causes the ASV to delay its avoidance maneuver, leading to a significant increase in the overall scenario risk as $\Theta_A^i(t)$ increases for

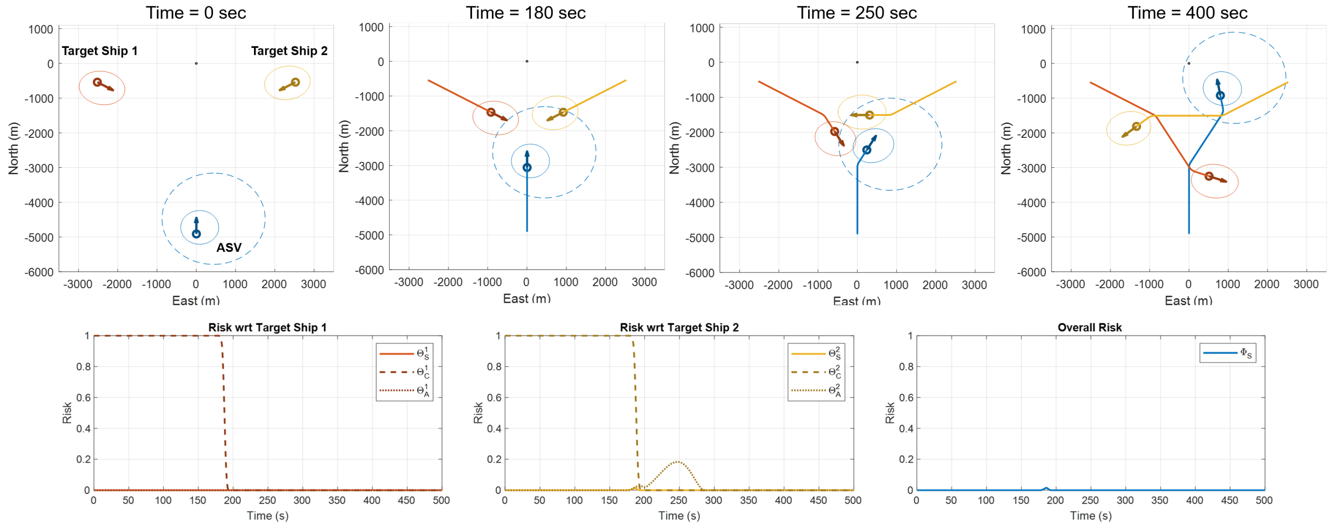

Fig. 4: Three-vessel roundabout scenario with idealized maneuvering. Snapshots of the trajectories are shown at the top. Each ship domain is visualized by the solid ellipse and the ASV ship arena is visualized by the dashed ellipse. The ASV risk curves over time with respect to each target ship are shown at the bottom.

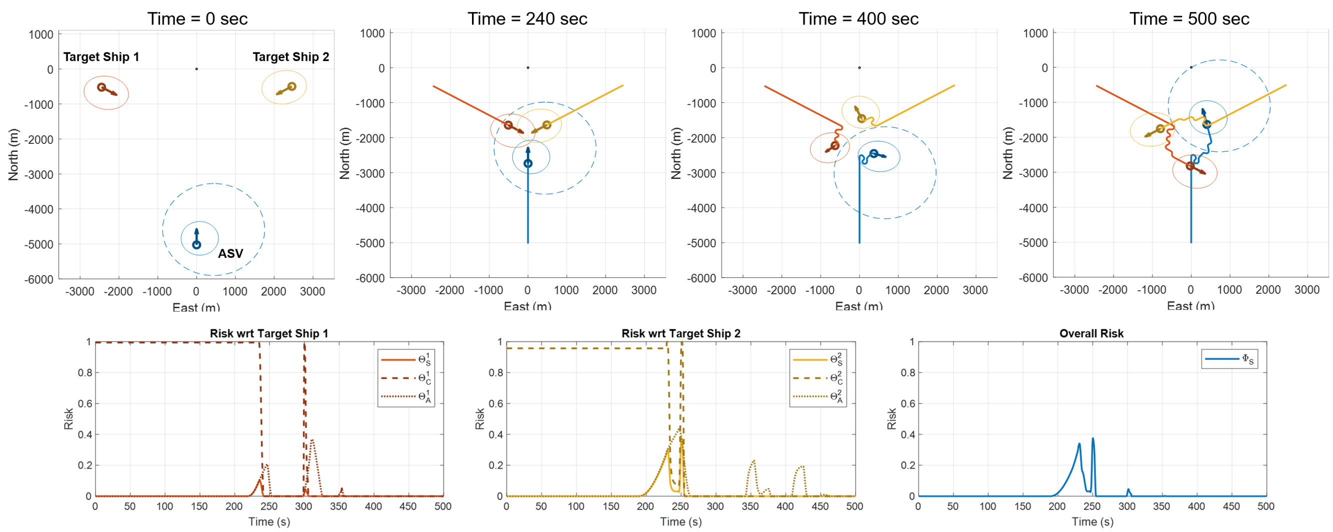

Fig. 5: Three-vessel roundabout scenario with a COLREGS-supplemented velocity obstacle planner. Snapshots of the trajectories are shown at the top. Each ship domain is visualized by the solid ellipse and the ASV ship arena is visualized by the dashed ellipse. The ASV risk curves over time with respect to each target ship are shown at the bottom.

each target ship. It is also evident that there is significantly more indecision in this ASV planner as reflected by the fluctuations in heading and speed of the ASV trajectory. This planning method was further analyzed through 100 scenarios in which the speeds and courses of the vessels were randomly varied within 25% of their nominal values. The mean scores for this Monte Carlo study are shown in the last row of Table I. Overall, the deficiencies of this ASV planner are captured by significantly lower seamanship scores when compared to the desired behavior of Fig 4.

V. Conclusion

This work introduced several performance metrics for evaluating ASV decision-making in multi-vessel scenarios based on the practice of good seamanship. The methodology for quantifying good seamanship is based on the criteria of reducing the overall collision risk of the situation and taking early, appropriate actions with respect to each target ship. Case study simulation results were presented that showcase the seamanship performance criteria against different ASV planning strategies in a three-vessel roundabout scenario. The results indicate that evaluating ASV performance on more general seamanship principles has potential to supplement COLREGS rule-based evaluation by capturing high-level maritime safety criteria.

Future work should include further research into more principled choices for some key evaluation parameters, including the dimensions of the ship arena and the functions used to generate the performance metrics. Additionally, future work could also consider incorporating the performance metrics provided in this paper into a cost function to be used for ASV navigation.

References

[1] U. S. C. Guard, *Navigation Rules*. United States Department of Transportation, 1999.

[2] K. Woerner, "Multi-contact protocol-constrained collision avoidance for autonomous marine vehicles," Ph.D. dissertation, Massachusetts Institute of Technology, 2016.

[3] K. Woerner, M. R. Benjamin, M. Novitzky, and J. J. Leonard, "Quantifying protocol evaluation for autonomous collision avoidance," *Autonomous Robots*, pp. 1–25, 2018.

[4] S. Li, Q. Meng, and X. Qu, "An overview of maritime waterway quantitative risk assessment models," *Risk Analysis: An International Journal*, vol. 32, no. 3, pp. 496–512, 2012.

[5] J. A. Marvel and R. Bostelman, "A cross-domain survey of metrics for modelling and evaluating collisions," *International Journal of Advanced Robotic Systems*, vol. 11, no. 9, p. 142, 2014.

[6] A. C. Bukhari, I. Tusseyeva, Y.-G. Kim *et al.*, "An intelligent real-time multi-vessel collision risk assessment system from vts view point based on fuzzy inference system," *Expert systems with applications*, vol. 40, no. 4, pp. 1220–1230, 2013.

[7] L. Perera, J. Carvalho, and C. G. Soares, "Fuzzy logic based decision making system for collision avoidance of ocean navigation under critical collision conditions," *Journal of marine science and technology*, vol. 16, no. 1, pp. 84–99, 2011.

[8] B. Li and F.-W. Pang, "An approach of vessel collision risk assessment based on the d–s evidence theory," *Ocean Engineering*, vol. 74, pp. 16–21, 2013.

[9] E. M. Goodwin, "A statistical study of ship domains," *The Journal of navigation*, vol. 28, no. 3, pp. 328–344, 1975.

[10] Y. Fuji and K. Tanaka, "Traffic capacity," *The Journal of Navigation*, vol. 24, pp. 543–552, 1971.

[11] T. Coldwell, "Marine traffic behaviour in restricted waters," *The Journal of Navigation*, vol. 36, no. 3, pp. 430–444, 1983.

[12] R. Szlapczynski, "A unified measure of collision risk derived from the concept of a ship domain," *The Journal of navigation*, vol. 59, no. 3, pp. 477–490, 2006.

[13] N. Wang, "An intelligent spatial collision risk based on the quaternion ship domain," *The Journal of Navigation*, vol. 63, no. 4, pp. 733–749, 2010.

[14] Z. Pietrzykowski and J. Uriasz, "The ship domain–a criterion of navigational safety assessment in an open sea area," *The Journal of Navigation*, vol. 62, no. 1, pp. 93–108, 2009.

[15] Y. He, Y. Jin, L. Huang, Y. Xiong, P. Chen, and J. Mou, "Quantitative analysis of colreg rules and seamanship for autonomous collision avoidance at open sea," *Ocean Engineering*, vol. 140, pp. 281–291, 2017.

[16] R. Szlapczynski and J. Szlapczynska, "Review of ship safety domains: Models and applications," *Ocean Engineering*, vol. 145, pp. 277–289, 2017.

[17] J. Zhang, X. Yan, X. Chen, L. Sang, and D. Zhang, "A novel approach for assistance with anti-collision decision making based on the international regulations for preventing collisions at sea," *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, vol. 226, no. 3, pp. 250–259, 2012.

[18] S. Campbell, W. Naeem, and G. W. Irwin, "A review on improving the autonomy of unmanned surface vehicles through intelligent collision avoidance manoeuvres," *Annual Reviews in Control*, vol. 36, no. 2, pp. 267–283, 2012.

[19] W. Naeem, G. W. Irwin, and A. Yang, "Colregs-based collision avoidance strategies for unmanned surface vehicles," *Mechatronics*, vol. 22, no. 6, pp. 669–678, 2012.

[20] M. R. Benjamin, J. J. Leonard, J. A. Curcio, and P. M. Newman, "A method for protocol-based collision avoidance between autonomous marine surface craft," *Journal of Field Robotics*, vol. 23, no. 5, pp. 333–346, 2006.

[21] P. K. E. Minne, "Automatic testing of maritime collision avoidance algorithms," Master's thesis, NTNU, 2017.

[22] P. Stankiewicz and G. Mullins, "Improving evaluation methodology for autonomous surface vessel colregs compliance," in *MTS/IEEE Oceans Conference*, 2019.

[23] J. Liu, F. Zhou, Z. Li, M. Wang, and R. W. Liu, "Dynamic ship domain models for capacity analysis of restricted water channels," *The Journal of Navigation*, vol. 69, no. 3, pp. 481–503, 2016.

[24] X. Zhu, H. Xu, and J. Lin, "Domain and its model based on neural networks," *The Journal of Navigation*, vol. 54, no. 1, pp. 97–103, 2001.

[25] R. Szlapczynski and J. Szlapczynska, "An analysis of domain-based ship collision risk parameters," *Ocean Engineering*, vol. 126, pp. 47–56, 2016.

[26] K. Kijima and Y. Furukawa, "Automatic collision avoidance system using the concept of blocking area," *IFAC Proceedings Volumes*, vol. 36, no. 21, pp. 223–228, 2003.

[27] Z. Pietrzykowski, "Ship's fuzzy domain–a criterion for navigational safety in narrow fairways," *The Journal of Navigation*, vol. 61, no. 3, pp. 499–514, 2008.

[28] Y. Kuwata, M. T. Wolf, D. Zarzhitsky, and T. L. Huntsberger, "Safe maritime autonomous navigation with colregs, using velocity obstacles," *IEEE Journal of Oceanic Engineering*, vol. 39, no. 1, pp. 110–119, 2013.