Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria
 Published by Research Publishing, Singapore.
doi: 10.3850/ESREL2026061419_esrel26-p29145-cd

Modelling Human–Machine Interaction Risks in Remote Maritime Autonomous Surface Ship Operations Using System Theoretic Process Analysis and Bayesian Networks

M Irsyad Hasbullah

Naval Architecture, Ocean & Marine Engineering, University of Strathclyde, Glasgow, United Kingdom. E-mail: irsyad.hasbullah@strath.ac.uk

G Theotokatos

Naval Architecture, Ocean & Marine Engineering, University of Strathclyde, Glasgow, United Kingdom. E-mail: gerasimos.theotokatos@strath.ac.uk

The safe operation of next-generation autonomous ships relies on effective Human–Machine Interactions (HMIs), particularly in remote operations where operators supervise from Remote Operation Centres (ROCs). The increasing interaction complexity among operators, automation, and stakeholders within Maritime Autonomous Surface Ships (MASS) ecosystem introduces emerging risks that remain challenging to capture using conventional risk assessment methods. The aim of this paper is to demonstrate a framework for analysing and quantifying HMI-related risks in remote ship operations for MASS degree of autonomy (DoA) level 3 by combining System-Theoretic Process Analysis (STPA) with Bayesian Network (BN) modelling. A concept of operations (ConOps) based case study of a remotely controlled ship was developed to capture critical interaction scenarios between ROC operators, autonomous ship systems, and operational stakeholders within the MASS ecosystem such as vessel traffic services (VTS), tugboat operators and port authorities. STPA was applied to systematically identify hazards, unsafe control actions, and interaction-related causal pathways in the remote operation. The causal factors identified by STPA were then organised into a set of Human-Machine Interaction Risk Influence Factors (HMI-RIFs), which formed the basis for constructing the Bayesian Networks. Model outputs are evaluated using a Normalized Risk Index (NRI) which converts probabilistic outcomes into a scalar metric for comparing operational scenarios. The results demonstrate that the proposed approach provides a structured and transparent representation of HMI-related risk. Sensitivity analysis indicates that Information Quality and Operator Workload are the dominant contributors to HMI-related risk. The novelty of this work lies in demonstrating a STPA–BN framework for analysing human–machine interaction risks in remote MASS operations, enabling a transparent transition from qualitative hazard analysis to quantitative probabilistic modelling and scenario-based risk comparison.

Keywords: Maritime Autonomous Surface Ships, Human-Machine Interactions, System Theoretic Process Analysis, Bayesian Networks, Risk Modelling

1. Introduction

The maritime industry is undergoing a gradual transition toward autonomous and remotely operated vessels, driven by technological advances and ongoing regulatory developments for Maritime Autonomous Surface Ships (MASS) (Aylward et al., 2020; Hogenboom et al., 2020). While fully autonomous operations remain a long-term goal, current industrial implementation focuses primarily on Degree of Autonomy (DoA) Level 3, where vessels are remotely controlled while human operators continue to supervise and intervene when necessary (Wróbel & Montewka, 2020).

In this operational context, human operators remain central to safe system performance, shifting from onboard execution to remote supervision through digital interfaces and communication networks (Veitch et al., 2024). Consequently, safety increasingly depends on the quality of human–machine interactions (HMIs), where reduced situational awareness, limited sensory feedback, and communication delays can introduce new operational challenges (J. Li & Wang, 2023; Z. Li et al., 2023; Porathe, 2022).

1.1. *HMI challenges and limitations of existing risk approaches*

Level 3 operations create distributed control environments in which humans and automation jointly influence system outcomes. The remote operator (RO) is legally responsible for the vessel but is physically detached from the operational environment, mediating strictly through digital interfaces (Porathe, 2020). These sociotechnical challenges such as mode confusion, latency-induced desynchronization, and trust calibration errors are not adequately captured by traditional component-based risk models (Wróbel, 2021). Risks therefore emerge from interaction mechanisms rather than isolated component failures, including issues related to automation coordination, information interpretation, and delayed feedback (Hoem et al., 2022; Le Guillou et al., 2023).

Traditional maritime risk assessment methods such as Fault Tree Analysis (FTA) and Event Tree Analysis (ETA) remain widely used, but their reliance on static and binary failure logic limits their ability to represent dynamic and interaction-driven risks common in socio-technical systems (Basnet et al., 2023; Wu et al., 2020). As a result, existing approaches are often insufficient for capturing the complexity of HMI in remote maritime operations.

1.2. *Research gaps*

Systems-Theoretic Process Analysis (STPA) has increasingly been adopted to analyse complex socio-technical systems because of its ability to identify hazards emerging from unsafe interactions rather than component failures (Glomsrud & Xie, 2020; Leveson, 2004). However, STPA remains largely qualitative and does not provide quantitative risk estimates for comparing operational scenarios. Bayesian Networks (BNs), in contrast, enable probabilistic modelling under uncertainty and have demonstrated value in maritime risk applications (Fan et al., 2024; Jiang et al., 2024). Yet, existing STPA-BN integrations often rely on subjective mappings between qualitative findings and probabilistic models, which may reduce structural transparency and traceability.

Furthermore, most quantitative MASS risk studies focus on technical or system-level failures, while human-machine interactions (HMIs) risks such as situation awareness degradation, mode confusion, or trust miscalibration remain insufficiently represented (Johansen & Utne, 2024; Yang et al., 2023). Current models also tend to express outcomes as binary accident probabilities, which do not adequately reflect the gradual and multi-dimensional nature of HMI degradation during remote operations. As a result, there remains a need for a structured framework capable of linking systemic hazard identification with probabilistic modelling while supporting risk comparison across different operational conditions.

1.3. *Aim and objectives*

The aim of this study is to demonstrate a structured framework for analysing HMI risks in MASS Level 3 operations by using STPA-BN based approach. The following high-level objectives support this aim:

(i) To apply STPA to gather HMI-specific losses, hazards, unsafe control actions and causal factors arising from the remote operator automation interaction loop.

(ii) To map STPA-derived HMI factors into a structured Bayesian Network model.

(iii) To demonstrate probabilistic inference through dynamic scenario-based analysis.

2. Methodology

This study applies a methodology combining STPA with a BN to analyse HMI risks in Level 3 MASS operations. The framework consists of three main phases as shown in Figure 1.

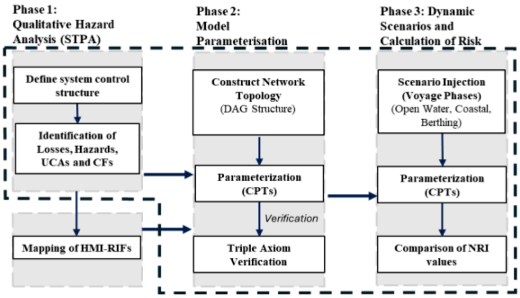

Fig.1.Multi-phase methodology for STPA-based hazard analysis and Bayesian Network risk evaluation.

2.1.*Phase 1: Application of STPA as hazard analysis*

The analysis begins with the definition of the system control structure representing interactions between the remote operator, onboard automation, and the surrounding operational environment. Following the STPA approach, safety is treated as a control problem rather than a consequence of individual component failures.

Based on the defined control structure, losses, hazards, unsafe control actions (UCAs), and associated causal factors (CFs) are systematically identified. This process establishes the qualitative understanding of how unsafe outcomes may emerge from inadequate control actions, delayed feedback, or degraded interaction conditions. The identified causal factors are then organised into Human–Machine Interaction Risk Influence Factors (HMI-RIFs), which provide structured inputs for the subsequent probabilistic modelling stage.

The HMI-RIFs identified describe as conditions in human, interface, and operational contexts that influence operator performance in automated systems. Mapping these factors to causal factors enables the systematic identification of how human–automation interaction may contribute to unsafe control actions and subsequent hazards.

2.2.*Phase 2: BN parameterisation*

The STPA-derived HMI-RIFs and causal pathways are translated into a Bayesian Network structure represented as a Directed Acyclic Graph (DAG) $U = \{X_1, \ldots, X_n\}$, (see Eq. (1)). Within this structure, HMI-RIFs act as causal parent nodes, unsafe control actions are represented as intermediate nodes, and the terminal node captures overall operational risk.

Bayesian Network forms a Directed Acyclic Graph where each node represents an STPA-derived state. The joint probability distribution follows standard Bayesian factorisation:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\big(X_i \big| Pa(X_i)\big) \qquad (1)$$

where $Pa(X_i)$ represents the parent nodes of X_i.

To ensure structural consistency, verification is conducted using triple axiom verification to confirm causal direction, dependency consistency, and probability coherence across the network topology.

2.3.*Phase 3: Scenarios Injection and Risk Comparison*

The final phase introduces operational scenarios into the BN through evidence injection representing different voyage phases, including open-water transit, coastal operations, and berthing. Scenario-specific conditions are applied through the parameterised Conditional Probability Tables (CPTs) to simulate variations in human-machine interaction performance.

Model outputs are evaluated using the Normalized Risk Index (NRI) (see Eq.(2)) which converts multi-state probabilistic results into a scalar metric suitable for comparison across scenarios. The resulting NRI values allow systematic assessment of how different operational conditions and interaction factors influence overall risk levels.

$$NRI = \sum_{j=1}^{m} P\big(S_j \big| E\big) \cdot w_j \qquad (2)$$

where S_j represents risk states, w_j the associated severity weights, and E the operational evidence introduced into the model.

3. Results

3.1.*STPA Outputs and HMI-RIF Development*

The STPA analysis began with defining the control structure for MASS Level 3 remote operations. Figure 2 presents a condensed representation of the control structure, illustrating the key controllers and feedback loops identified during the analysis.

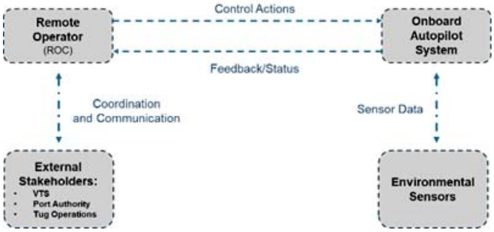

Fig. 2.Condensed control structure

The analysis produced a set of 27 UCAs and associated causal factors related to communication, interface conditions, and operational context, which form the qualitative basis for subsequent modelling.

UCA-1: Operator does not execute collision avoidance due to latency.

UCA-6: Operator provides hazardous waypoints due to mode confusion.

Rather than treating these as isolated events, the causal factors were grouped into structured Human–Machine Interaction Risk Influence Factors (HMI-RIFs) to support systematic probabilistic modelling. Table 1 presents the mapping between STPA-derived HMI-RIFs and their corresponding BN nodes, demonstrating the traceable linkage between qualitative hazard analysis and quantitative probabilistic modelling.

Table 1. Illustrative example of STPA-derived HMI-RIF mapping to Bayesian Network structure

STPA Source Category	Derived HMI-RIF	BN Layer Role
Communication-related CFs	Latency	Causal parent node
Interface-related CFs	Interface Quality	Causal parent node
Cognitive overload CFs	Workload	Causal parent node
Procedural deviation CFs	Procedures	Causal parent node

3.2.*Bayesian Network structure*

The HMI-RIFs formed the foundational layer of the Bayesian Network, with causal dependencies derived from STPA analysis pathways. Figure 3 illustrates the complete network structure, showing how HMI-RIFs propagate through intermediate nodes to influence the overall risk state (see Figure 3).

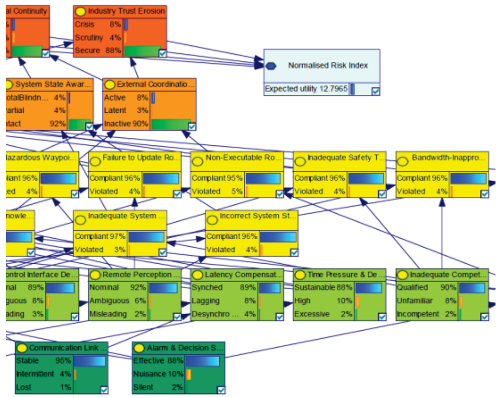

Fig. 3.Bayesian Network structure linking HMI-RIFs to the terminal Normalized Risk Index (NRI) with baseline value 12.80.

Model verification and validation were conducted to ensure logical consistency and behavioural realism of the BN. Verification focused on confirming structural correctness, including causal direction consistency and probability normalisation:

(i) Axiom 1 (Causal Direction): All parent-child relationships followed logically from STPA causal pathways, with no reverse dependencies.

(ii) Axiom 2 (Dependency Consistency): Conditional independence relationships were consistent with the control structure hierarchy.

(iii) Axiom 3 (Probability Coherence): No contradictory probability assignments were identified during sensitivity testing.

3.3.*Inference results based on scenarios injection*

Following verification, scenario-based simulations were performed to evaluate how changes in HMI conditions influence the terminal risk index. Three operational scenarios were developed representing distinct voyage phases and failure modes, as summarised in Table 2.

Table 2. Comparison of NRI value with different scenarios

Scenario	Voyage Phase	Key Condition	NRI Result
Scenario 1	Open Water Transit	Sensor degradation (zero visibility due to fog)	29.67
Scenario 2	Coastal Operation	Communication latency (>500 ms)	32.78
Scenario 3	Berthing	Increased manual intervention (reduced trust)	23.57

Each scenario was implemented by setting evidence on relevant parent nodes (e.g., sensor state, communication quality, operator intervention mode) and propagating this evidence through the network using the GeNIe inference engine. The resulting posterior probabilities at the terminal risk node were then transformed into NRI values using Equation (2)

A static sensitivity analysis was conducted using the GeNIe Academic 4.0 inference engine to identify which variables exert the greatest influence on the terminal Loss of Control risk node (see Figure 4).

Fig. 4.Sensitivity Analysis of the Normalized Risk Index

Result reveals that information quality and operator workload emerge as the dominant contributors to risk variation, together accounting for 45% of total influence. This finding underscores the critical importance of data provision and operator cognitive load in remote MASS operations.

4. Discussion

4.1.*Interpretation of HMI Influence*

The results demonstrate that the proposed STPA-BN framework successfully captures the multi-factorial nature of HMI risk in remote MASS operations. The increasing NRI values across operational scenarios reflect the escalating interaction demands placed on remote operators as voyage phases become more complex. Notably, the baseline NRI of 12.80 suggests that even under nominal conditions, a measurable probability of degraded HMI performance exists, reinforcing the need for proactive risk management rather than reactive incident response.

The dominance of Information Quality (24%) and Operator Workload (21%) as risk-influencing factors carries important implications for system design. Information quality encompasses not only sensor accuracy but also the timeliness, relevance, and presentation of data through remote operator interfaces. This finding supports the argument that HMI design in MASS should prioritise cognitive support over raw data provision (Veitch et al., 2024). Similarly, the significant influence of workload highlights the vulnerability of remote operations to task saturation, particularly during high-demand phases such as berthing where multiple vessels and stakeholders require simultaneous attention.

The relatively lower influence of Trust Calibration (3%) may appear counterintuitive given the emphasis on trust in human-automation interaction literature. However, this result likely reflects the operational context of Level 3 MASS, where operators retain active supervisory control rather than passive monitoring. Under active control conditions, trust manifests through intervention decisions rather than automation reliance, potentially reducing its direct influence on unsafe outcomes compared to information and workload factors.

4.2. *Methodological Contribution*

Beyond the specific HMI insights discussed above, the proposed STPA-BN framework offers several methodological advances for risk analysis in remote MASS operations. The primary contribution lies in the structured and traceable translation of qualitative hazard analysis into a probabilistic model through the introduction of HMI-RIFs as an intermediate layer.

Previous STPA-BN integrations have often relied on direct or subjective mappings between STPA outputs and network structures, which can obscure the reasoning behind model architecture and limit transparency. By contrast, this study systematically derived HMI-RIFs from STPA causal pathways before constructing the Bayesian Network. This approach ensures that each node in the probabilistic model corresponds to a specific factor identified during hazard analysis, preserving the causal logic established in the qualitative phase. The resulting model provides what Leveson (2004) describes as a "controlled process" view of safety, now rendered in quantitative form suitable for operational decision-making.

The triple axiom verification applied during model construction further strengthens methodological rigour. By confirming causal direction, dependency consistency, and probability coherence, the framework addresses a common criticism of Bayesian applications in safety science, that probabilistic dependencies may be specified without adequate theoretical or empirical foundation. The verification process documented in Section 3.2 demonstrates that the model's structure remains faithful to the underlying control-theoretic principles of STPA.

A further methodological innovation is the use of the NRI as a scalar metric for scenario comparison. Traditional risk metrics in maritime applications often express outcomes as binary accident probabilities, which are ill-suited to capturing the graduated nature of HMI degradation. The NRI, by weighting probabilistic outcomes by severity, enables more nuanced comparisons across operational conditions while maintaining interpretability for practitioners. This aligns with recent calls in the literature for risk metrics that better reflect the multi-dimensional nature of socio-technical system performance (Johansen & Utne, 2024).

However, several limitations should be acknowledged. First, the current model parameters are based on expert elicitation and STPA-derived causal structures rather than empirical operational data. While this is purposely for a demonstration, future validation against incident reports, simulator studies, or in-service data would strengthen confidence in the quantitative outputs. Second, the HMI-RIFs developed here, though systematically derived, represent a particular framing of interaction risks; alternative groupings or additional factors may emerge as operational experience with MASS accumulates. Third, the model currently assumes static dependencies between factors, whereas real-world HMI risks may involve adaptive behaviours and learning effects that unfold over time.

Despite these limitations, the framework offers practical value for system designers and safety managers. By making explicit the causal pathways through which HMI factors contribute to operational risk, the model supports targeted interventions such as in interface design, procedure development, or operator training. The scenario injection capability demonstrated in Section 3.3 illustrates how the framework can be used to compare risk profiles across voyage phases or system configurations, informing resource allocation and risk mitigation strategies.

In summary, the methodological contributions of this work are:

(i) The introduction of HMI Risk Influencing Factors (HMI-RIFs) to enable traceable mapping from STPA results to the Bayesian Network while preserving the underlying causal logic

(ii) The development of a comparative metric, NRI designed to capture the varying severity levels of HMI-related risk.

Together, these elements address limitations identified in prior STPA-BN integrations and provide a foundation for more systematic analysis of human-machine interaction risks as MASS technologies continue to evolve.

5. Conclusion

This study developed and demonstrated a structured framework for analysing HMIs in MASS Level 3 remote operations by combining System-Theoretic Process Analysis with Bayesian Network modelling. The three-phase approach STPA hazard analysis and HMI-RIFs mapping, BN construction, and scenario-based inference provides a systematic pathway from qualitative hazard identification to quantitative risk representation.

The findings indicate that HMI risks in remote MASS operations arise from interacting factors, with information quality and operator workload emerging as influential contributors within the analysed scenarios. The structured mapping between STPA causal pathways and BN structure enhances model transparency and addresses limitations identified in prior STPA–BN based approach. The Normalized Risk Index (NRI) enables comparative evaluation across operational scenarios, supporting informed consideration of interface design, procedural development, and operator workload management.

As maritime autonomous systems evolve, structured approaches for analysing HMIs emerging risks will remain essential. The proposed framework provides a transparent and adaptable modelling structure that can be refined and extended as empirical data from remote operations become available. Future work should focus on empirical validation and application to broader operational contexts as MASS deployment progresses.

Acknowledgment

This work was supported by Innovate UK under the Clean Maritime Demonstration Competition Round 6 (CMDC6) project Virtual Reality-based Immersive Education for Next-Generation Zero-Emission Vessels (VRIVEN).

References

Aylward, K., Johannesson, A., Weber, R., MacKinnon, S. N., & Lundh, M. (2020). An evaluation of low-level automation navigation functions upon vessel traffic services work practices. *WMU Journal of Maritime Affairs*, *19*(3), 313–335.

Basnet, S., BahooToroody, A., Chaal, M., Lahtinen, J., Bolbot, V., & Valdez Banda, O. A. (2023). Risk analysis methodology using STPA-based Bayesian network- applied to remote pilotage operation. *Ocean Engineering*, *270*.

Fan, C., Bolbot, V., Montewka, J., & Zhang, D. (2024). Advanced Bayesian study on inland navigational risk of remotely controlled autonomous ship. *Accident Analysis and Prevention*, *203*.

Glomsrud, J. A., & Xie, J. (2020). A structured STPA safety and security co-analysis Framework for autonomous ships. *Proceedings of the 29th European Safety and Reliability Conference, ESREL 2019*, 38–45.

Hoem, Å. S., Veitch, E., & Vasstein, K. (2022). Human-centred risk assessment for a land-based control interface for an autonomous vessel. *WMU Journal of Maritime Affairs*, *21*(2), 179–211.

Hogenboom, S., Rokseth, B., Vinnem, J. E., & Utne, I. B. (2020). Human reliability and the impact of control function allocation in the design of dynamic positioning systems. *Reliability Engineering and System Safety*, *194*.

Jiang, H., Zhang, J., Wan, C., Zhang, M., & Soares, C. G. (2024). A data-driven bayesian network model for risk influencing factors quantification

based on global maritime accident database. *Ocean and Coastal Management*, *259*.
Johansen, T., & Utne, I. B. (2024). Human-autonomy collaboration in supervisory risk control of autonomous ships. *Journal of Marine Engineering and Technology*, *23*(2), 135–153.
Le Guillou, M., Prévot, L., & Berberian, B. (2023). *Trusting Artificial Agents: Communication Trumps Performance*.
Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, *42*(4), 237–270.
Li, J., & Wang, H. (2023). Modeling and analyzing multiteam coordination task safety risks in socio-technical systems based on FRAM and multiplex network: Application in the construction industry. *Reliability Engineering and System Safety*, *229*.
Li, Z., Zhang, D., Han, B., & Wan, C. (2023). Risk and reliability analysis for maritime autonomous surface ship: A bibliometric review of literature from 2015 to 2022. *Accident Analysis and Prevention*, *187*.
Porathe, T. (2020). Deconflicting maritime autonomous surface ship traffic using moving havens. *30th European Safety and Reliability Conference, ESREL 2020 and 15th Probabilistic Safety Assessment and Management Conference, PSAM 2020*, 1043–1049.
Porathe, T. (2022). Remote Monitoring of Autonomous Ships: Quickly Getting into the Loop Display (QGILD). *Human Factors in Transportation*, *60*.
Veitch, E., Dybvik, H., Steinert, M., & Alsos, O. A. (2024). Collaborative Work with Highly Automated Marine Navigation Systems. *Computer Supported Cooperative Work: CSCW: An International Journal*, *33*(1), 7–38.
Wróbel, K. (2021). Searching for the origins of the myth: 80% human error impact on maritime safety. In *Reliability Engineering and System Safety* (Vol. 216). Elsevier Ltd.
Wróbel, K., & Montewka, J. (2020). Comments to the article by Ramos et al. 'Collision avoidance on maritime autonomous surface ships: Operators' tasks and human failure events' (Safety Science Vol. 116, July 2019, pp. 33–44). In *Safety Science* (Vol. 121, pp. 603–605). Elsevier B.V.
Wu, Q., Wang, T., Diaconeasa, M. A., Mosleh, A., & Wang, Y. (2020). A comparative assessment of collision risk of manned and unmanned vessels. *Journal of Marine Science and Engineering*, *8*(11), 1–24.
Yang, R., Bremnes, J. E., & Utne, I. B. (2023). Online risk modeling of autonomous marine systems: Case study of autonomous operations under sea ice. *Ocean Engineering*, *281*.