Enhancing maritime autonomous surface ship navigation safety: A dynamic risk assessment framework integrating system theoretic process analysis and stochastic Petri nets

Chun Zou [a,b,c], Lijia Chen [a,b,c*], Hanwen Fan [d], Xiaobin Tian [a,b,c], Li Liu [e], Zaili Yang [d*]

[a] School of Navigation, Wuhan University of Technology, Wuhan, China
[b] State Key Laboratory of Maritime Technology and Safety, Wuhan University of Technology, Wuhan, China
[c] Hubei Key Laboratory of Inland Shipping Technology, Wuhan, China
[d] Liverpool Logistics, Offshore and Marine Research Institute, Liverpool John Moores University, Liverpool, UK
[e] Hubei Province Tianmen City Waterway Maintenance Center, Tianmen, China

*Corresponding author.
Corresponding author email addresses:navisky@qq.com(LJ.Chen) and z.yang@ljmu.ac.uk(ZL.Yang)

Abstract: The transition of maritime autonomous surface ships toward complex digital architectures has shifted navigation safety from a component reliability issue to a systemic control problem. Traditional failure-centric models struggle to capture non-linear interactions within complex network-featured control loops. This study proposes a control-centric hybrid framework integrating system-theoretic process analysis and stochastic Petri nets to bridge qualitative hazard identification with quantitative dynamic simulation. Initially, system-theoretic process analysis maps functional dependencies within the control loop, identifying risk-influencing factors and unsafe control actions. Subsequently, three critical propagation paths leading to system-level accidents are mapped onto an isomorphic stochastic Petri nets model to simulate dynamic risk evolution. A case study of the Yangtze River's Taicang section, utilizing automatic identification system records and meteorological data, is conducted to demonstrate the feasibility of the approach. The results reveal significant path dependence and systemic lock-in effects among environmental stressors, perceptual fidelity, and autonomous decision-making. Notably, the intelligent integration platform and autonomous ship control system serve as global risk convergence hubs, accounting for over 70% of steady-state risk in collision scenarios. Sensitivity analysis shows that improving environmental perception can cut overall risk by 55.5%. This framework provides a rigorous quantitative basis for systemic risk characterization and resilience enhancement.

Keywords: Maritime autonomous surface ship; System-Theoretic Process Analysis; Stochastic Petri Nets; Risk assessment; Maritime risk;

1. Introduction

Driven by rapid digital advancements, the development of Maritime Autonomous Surface Ships (MASS) aims to enhance maritime safety and sustainability by reducing reliance on human intervention [1]. In response, the International Maritime Organization (IMO) is prioritizing regulatory frameworks like the MASS Code to establish safety acceptance principles [2]. However, the introduction of autonomous technologies and the subsequent shift in operational models are accompanied by the emergence of novel risks that differ markedly from traditional ship operations [3-5]. While traditional maritime risks are often linked to human cognitive biases [6], hazards in MASS increasingly stem from non-linear interactions between software algorithms, complex sensor fusion, and remote-control feedback loops [7-9].

This transformation in risk typology suggests that navigational safety is no longer merely a matter of preventing component breakdowns but has evolved into a complex control problem. Consequently, a fundamental shift in research perspective is required: moving away from a traditional focus on individual component reliability toward a systematic analysis of functional interdependencies within the perception-decision-control-execution closed-loop. Within this new paradigm, hazards are viewed as emerging from dysfunctional interactions that may remain latent until specific operational conditions are met.

Despite the critical need for a systemic perspective, traditional risk analysis methodologies remain predominantly static and component-based. To shift hazard management from a component-oriented to a system-oriented level for complex autonomous operations, it is essential to capture the non-linear interactions and cascading failures inherent in MASS. These traditional approaches often fail to characterize the dynamically evolving risk profiles in highly concurrent environments [10]. To realise the navigational safety of MASS from a systemic level, this study proposes a "control-centric" hybrid framework integrating System-Theoretic Process Analysis (STPA) and Stochastic Petri Nets (SPN). By operationalizing systemic vulnerabilities into computable network elements, this approach formalizes the time-dependent risk propagation.

The main contributions of this work are summarized as follows:

(1) A closed-loop hybrid analytical framework is developed to bridge the gap between qualitative hazard identification and quantitative stochastic modeling. By directly integrating STPA-based index construction with Markov chain calculations derived through SPN isomorphism, the framework achieves a seamless methodological transition from system-theoretic analysis to formal risk quantification.

(2) A multi-source information fusion method for parameterization is proposed to accurately quantify latent human-system interaction failures. This approach synthesizes objective operational data from Automatic Identification System (AIS) records and environmental variables from meteorological datasets with subjective expert judgment, providing a robust mechanism for estimating transition rates in complex maritime sociotechnical systems.

(3) Empirical experience and risk performance quantification are provided through a comprehensive case study of the Yangtze River’s Taicang section. By applying Monte Carlo simulations to this high-density mixed-traffic environment, the study identifies global hubs of system vulnerability, offering practical insights for safety management and maritime traffic control in restricted waterways.

The remainder of this paper is organized as follows: Section 2 provides a literature review of maritime risk assessment, while Section 3 introduces the proposed STPA-SPN hybrid methodology. Section 4 presents a detailed case study of the Yangtze River's Taicang section to validate the framework, and Section 5 discusses the advantages, limitations, and practical implications of the results. Finally, Section 6 concludes the study by summarizing core findings and future research directions.

2. Literature review

2.1 MASS safety: Evolution from static to dynamic component-based assessment

Collision, grounding, and sinking constitute the core accident typologies defining conventional maritime navigational hazards [11]. These events typically originate from unsafe human actions, hazardous object states, and adverse environmental conditions [12], all of which significantly impact the operational integrity of MASS. Compared to conventional cargo vessels, MASS feature more complex interactions and control processes among system components. As system interactivity and functional coupling intensify, complex behavioral patterns that are difficult to predict in advance may emerge during navigation [13], rendering traditional static assessment methods insufficient. Against this background, researchers are increasingly focusing on how to effectively quantify the uncertainty associated with unexpected events arising from this inherent system complexity [14].

Early risk assessment for MASS primarily utilized component-based methodologies, such as Formal Safety Assessment (FSA) [15], Failure Mode and Effects Analysis (FMEA) [16, 17], and Hazard and Operability Analysis (HAZOP) [18, 19]. While these established a baseline for identifying individual hazards, they struggle to adequately characterize synergistic effects, cascading failures, and the dynamically evolving risk profiles encountered during real-time operations [20]. The underlying nature of dynamic risk is defined by its temporal fluidity and path dependence, where the system's safety state is not a static probability of failure but a time-varying process governed by continuous feedback loops. This mechanism is best illustrated through the propagation of control flaws: a transient perception distortion may appear benign in isolation, yet it can trigger a non-linear escalation of decision-making errors as it propagates through the intelligent integrated platform. In high-density environments, these sequences often result in a systemic lock-in effect, where the risk converges toward a hazardous steady state even in the absence of a primary component failure. To address these temporal limitations, recent advancements have introduced dynamic networks and data-driven models [21-23]. For instance, Luo et al [24, 25] utilized Dynamic Bayesian Networks (DBN) to evaluate multi-phase mission success and temporal correlations during berthing operations, while Nakashima et al. [26] developed a real-time reliability assessment framework integrating machine learning with physical knowledge to predict the health state of marine mechanical equipment.

Recent studies have further expanded these hybrid approaches to better capture the systemic nature of navigation risks through the lens of risk coupling. Fan et al [27] developed an operational risk coupling framework specifically for MASS navigating in restricted waters, while Yang et al. [8] quantified coupling risks during operational mode transitions in autonomous vessels. Furthermore, Yuzui and Kaneko [28] proposed a systematic hybrid review highlighting the critical shift towards proactive systemic analysis to address the inherent limitations of traditional methodologies. However, despite these advancements in dynamic modeling, a critical research gap remains: most existing dynamic models still operate under a "failure-centric" perspective, focusing on the reliability of physical components rather than the dysfunctional control relationships between them. This study argues that even the most advanced DBN or data-driven models often miss the emergent systemic risks arising from the hierarchical control loop, necessitating a transition from analyzing component failures to simulating systemic control flaws.

2.2 From failure-centric to systemic control analysis

Conventional risk assessment frameworks, rooted in the reductionist principle of component reliability, struggle to capture the emergent behaviors inherent in the non-linear architectures of autonomous shipping. As the primary risk driver in MASS shifts from discrete hardware failures to dysfunctional interactions within complex control loops, a methodology is required that can proactively identify latent hazards during the conceptual phase—thereby preempting the need for costly late-stage structural retrofits. Furthermore, such an approach must move beyond "failure-centric" analysis to encapsulate the systemic vulnerabilities arising from the intricate interplay between

software logic, sensor fusion, and human-machine interfaces. To address these imperatives, the introduction of STPA [29] provides an advanced methodology that marks a foundational paradigm shift, redefining safety as a continuous control problem rather than a discrete failure issue. Grounded in the System-Theoretic Accident Model and Processes (STAMP) framework, STPA targets the functional interdependencies within the hierarchical control structure of MASS, specifically the coordination between Ship Autonomous Control Systems (SACS) and Remote Operation Center (ROC), to isolate Unsafe Control Actions (UCAs).

The practical utility of this advanced methodological framework has been validated across diverse maritime scenarios. Ventikos et al. [30] utilized STPA to identify hazards related to increasing autonomy levels, while Yang et al [31] applied it to identify failure scenarios during control mode transitions. Additionally, Cheng et al. [32] employed STPA to analyze the safety of human-system interactions within autonomous systems. Consequently, STPA is recognized as a premier tool for risk analysis in MASS systems [28, 33, 34]. Despite these strengths, the application of STPA still exhibits notable limitations, primarily its predominant reliance on qualitative analysis. The lack of deep integration with quantitative simulation models makes it difficult to quantify the probabilistic characteristics of risk evolution in high-concurrency environments.

2.3 From qualitative logic to stochastic quantitative simulation

To bridge the gap between qualitative logic and quantitative simulation, researchers have attempted to hybridize STPA with various tools. Recent efforts have integrated STPA with Fault Tree Analysis (FTA) [35], Bayesian Networks (BN) [36, 37], Markov Chains [38], Hidden Markov Models (HMM) [39], Genetic Algorithms [40], Success Likelihood Index Method (SLIM) [41] and System Dynamics [42] to quantify navigation risks. However, these approaches often struggle to accurately represent the high degree of concurrency and asynchrony characteristic of MASS operations in restricted waterways. For example, while integration with HMM can describe temporal state transitions, it lacks the graphical formalism to intuitively visualize the network structure of risk propagation.

SPN offer exceptional modeling capabilities for systems exhibiting such stochastic and concurrent behaviors due to their strong representation of asynchronous events [43-47]. It has become a key tool for system modeling and performance evaluation in multiple advanced fields. Specific applications include: Bensaci et al. [48] utilized SPN to model and simulate collision risks in transport tasks within multi-mobile robot systems; Brito et al. [49] applied it to assess the reliability of complex urban systems; and Farias et al. [50] employed SPN for reliability analysis of satellite systems. Furthermore, SPN has demonstrated strong analytical capabilities in performance evaluation for IoT systems [51] and dynamic evolution modeling of urban fires [52]. In the domain of maritime safety, Nyvlt et al. [53] used SPN to model ship collision accidents, systematically simulating the accident evolution process and conducting experimental analysis. Zhou et al. [54] integrated hierarchical timed colored Petri nets with Markov chain theory to develop an evaluation model for ship fire emergency response processes, enhancing the quantitative assessment capability of emergency procedures. Hu et al. [55] proposed an enhanced risk assessment framework for complex maritime traffic systems, introducing SPN to quantify comprehensive risks for ship navigation in Arctic waters. Building on these advantages, this study proposes a hybrid STPA-SPN framework to bridge the disconnection between qualitative hazard identification and quantitative dynamic simulation. By defining SPN places and transitions based on the control loop flaws identified in STPA, the model characterizes the time-dependent risk propagation inherent in MASS operations. This approach enables the formal quantification of how specific control flaws propagate into system-level accidents through critical mechanisms such as path dependence and systemic lock-in.

2.4 Research gaps

Based on the aforementioned review of existing methodologies, several critical research gaps in MASS risk assessment can be synthesized:

(1). While traditional dynamic methodologies, such as DBN, have transitioned from static to temporal analysis, they often operate under a "failure-centric" perspective focused on individual component reliability. Such approaches provide a static snapshot of system safety, failing to characterize the non-linear, emergent systemic risks that arise from dysfunctional interactions within the hierarchical control loop.

(2). In the "perception–decision–control–execution" loop of MASS, the temporal dimension—including control delays and precise sequences of events—is a critical determinant of system integrity. Existing models struggle to quantify how a control action, though logically sound, becomes functionally hazardous when executed with improper timing.

(3). There is a persistent lack of integrated frameworks that bridge qualitative hazard identification with quantitative simulation. Specifically, STPA provides a robust qualitative foundation for identifying structural hazards, yet its predominant reliance on descriptive analysis creates a disconnect between hazard identification and dynamic probabilistic assessment. Conversely, SPN offer exceptional modeling capabilities for asynchronous behaviors through firing delays, but they require a rigorous logical framework to define the state space of control-loop failures accurately.

To address these limitations, our study proposes a "control-centric" hybrid framework integrating STPA and SPN. By operationalizing systemic vulnerabilities into computable network elements, this approach formalizes the time-dependent risk propagation inherent in MASS operations. This allows for the quantification of complex risk behaviors, such as path dependence and systemic lock-in effects, providing a more rigorous basis for navigation safety in high-concurrency environments.

3. MASS navigation dynamic risk assessment: A hybrid framework construction

The proposed "control-centric" hybrid framework is designed to overcome the limitations of traditional "failure-centric" models by bridging the gap between qualitative hazard identification and quantitative dynamic simulation. Grounded in the principle that navigational safety in MASS has evolved into a complex control problem, this approach focuses on the functional interdependencies within the system's closed-loop architecture. By operationalizing systemic vulnerabilities into computable network elements, the framework formalizes the time-dependent risk propagation inherent in autonomous operations.

As illustrated in Fig. 1, the framework proposed in this study consists of three key phases integrated into a purpose-built methodological pipeline. Phase 1 defines the hierarchical control structure and utilizes STPA to outline the structural indices and causal logic constraints in detail. Phase 2 extends these structural outputs by identifying the involved physical and logical states, leading to the construction of a hybrid STPA-SPN model to capture the functional interdependencies within the "perception–decision–control–execution" closed-loop. Phase 3 applies isomorphic Markov chain calculations to compute the risk performance, which represents the dynamic safety state quantified from a systemic control perspective. The phases are described in detail from Section 3.1 to Section 3.3.

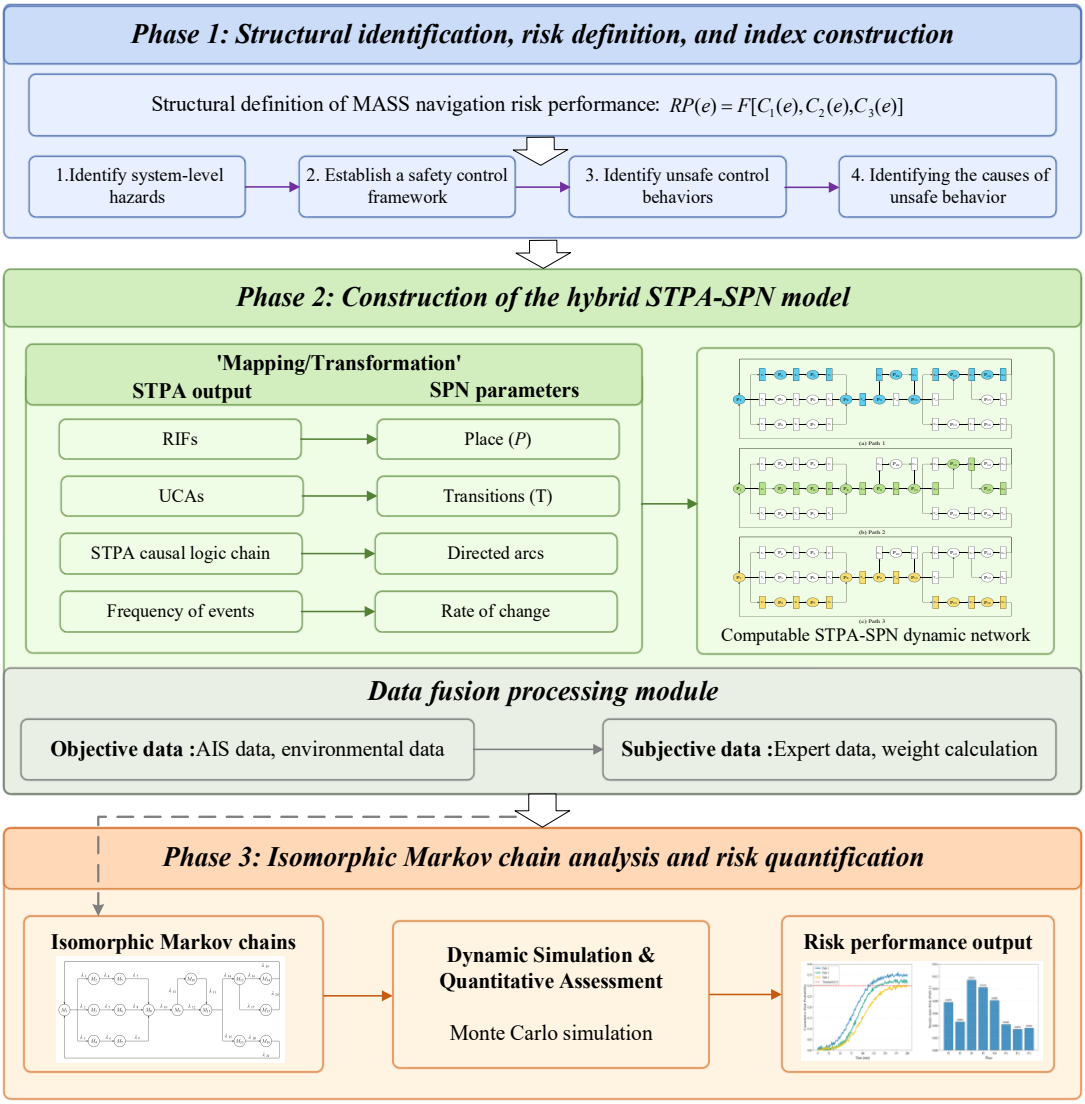


Fig. 1. Methodological framework.

3.1 Phase 1: Structural identification, risk definition, and index construction

The logic of this phase is primarily rooted in the STAMP, which re-evaluates maritime safety from a cybernetic perspective. This approach fundamentally treats navigation safety as a complex control problem [56] rather than a simple matter of component reliability, positing that accidents arise from unsafe interactions among system components rather than from independent component failures.

Within this STPA-based control perspective, the meaning of risk itself must be reformulated. Traditional Probabilistic Risk Assessment (PRA) reports risk as a static scalar likelihood on a 0 to 1 scale [57, 58], detached from the causal chain that drives the system toward an accident, and therefore cannot capture how risk accumulates along the perception-decision-control-execution loop in highly dynamic operational environments [28, 59]. To address this limitation, this study redefines risk as risk performance , the steady-state probabilistic behavior of the system along a specific risk-propagation path. By the isomorphism between the SPN and its associated continuous-time Markov chain (CTMC) $X(t)$, risk performance is equivalently the steady-state probability mass that the chain accumulates on the markings induced by that path. Let $R(M_1)$ be the reachable marking set from the initial marking M_1 and $\pi(M) = \lim_{t\to\infty} P(X(t) = M)$ the stationary distribution. For a risk-propagation path e identified by STPA, its induced reachable subset is denoted $Re \subseteq R(M_1)$, and risk performance is expressed in Eq. (1):

$$RP(e) = \sum_{M \in Re} \pi(M) \tag{1}$$

where $RP(e)$ denotes the risk performance of risk-propagation path e identified by STPA; M is a reachable marking of the SPN; $R(M_1)$ is the set of markings reachable from the initial marking M_1; $R_e \subseteq R(M_1)$ is the subset of markings induced by path e, namely those whose occupied places lie on the perception-decision-control segment of e and that terminate at the accident marking with $M(P_1)=1$; and $\pi(M) = \lim_{t\to\infty} P(X(t) = M)$ is the steady-state probability of marking M under the stationary distribution of the isomorphic CTMC $X(t)$.

$RP(e)$ thus admits a single, unambiguous interpretation: the stationary probability that the system resides on path e under the long-run behavior of the CTMC, fully determined by the SPN structure, its firing rates λ, and the initial marking M_1. Any node-level contribution ranking or place-level occupancy profile derived from $RP(e)$ is not a separate definitional component of risk performance, but an auxiliary diagnostic that decomposes $RP(e)$ to localize systemic vulnerabilities along path e.

To localize risk accumulation within a path, this study defines the node contribution of place P_i along path e as the proportion of $RP(e)$ that co-resides at P_i in its risk state:

$$C_i = \frac{\sum_{M \in R_e} \pi(M) \cdot 1\{M(P_i) = 1\}}{RP(e)} \tag{2}$$

where $1\{M(P_i) = 1\}$ is the indicator that place P_i carries a token (risk-active state) under marking M, and all other symbols follow Eq. (1). By construction, $C_i(e) \in [0, 1]$ and $\Sigma_i\, C_i(e) = 1$. Unlike the AHP-derived importance weights, which are a priori expert elicitations used to calibrate firing rates λ_j, $C_i(e)$ is an a posteriori diagnostic computed from the converged stationary distribution; it measures the fraction of realized steady-state risk that accumulates at node P_i, and is therefore independent of subjective weighting.

To operationalize this dynamic definition, the framework utilizes STPA to define system boundaries, encompassing the ROC and the SACS within a mixed-traffic environment. A hierarchical safety control structure is constructed to visualize the functional dependencies within the perception-decision-control-execution closed-loop by mapping critical feedback paths between controllers, actuators, and sensors. Building upon this structural topology, the core task of index construction is performed by systematically identifying UCAs. These inappropriate actions are categorized into four types[60].

(1) A required control action is not provided or not properly executed;

(2) An incorrect or unsafe control action is provided;

(3) A correct action is applied at the wrong time (too early or too late);

(4) A correct action is stopped too soon or applied for too long.

By investigating the root causes of these UCAs, the analysis maps abstract control flaws to specific physical and logical components, culminating in the synthesis of a systematic set of key RIFs across three dimensions: component failures, unsafe system interactions, and external environmental disturbances. Ultimately, Phase 1 transforms descriptive causal scenarios into a rigorous index of structural logic and causal constraints, providing the essential parameter inputs for the subsequent quantitative stochastic simulation.

3.1.1. Development of the safety control structure

System-level hazards for MASS during navigation primarily include ship collisions, groundings, and loss of control. Within the ship-shore integrated operational architecture, navigation control is collaboratively implemented by the ROC and the SACS, specifically encompassing the following three typical scenarios.

(1) The autonomous navigation mode dominated by the SACS;

(2) The remote control mode led by ROC operators;

(3) The navigation mode involving collaborative decision-making between the ROC and the SACS.

Based on these hazard types and operational scenarios, system-level hazards can be further clarified, and corresponding safety constraints can be defined. Subsequent analysis, following the STPA methodology, will involve the sequential development of a hierarchical control structure, analysis of UCAs, and identification of causal scenarios to enable a systematic and in-depth investigation into the navigation safety of MASS.

In the STPA framework, the safety control structure primarily comprises four core components: the controller, the actuator, the controlled process, and sensors. Within this structure, the controller issues control commands to the actuator based on mission objectives, and the actuator acts upon the controlled object. Sensors are responsible for the real-time monitoring of the operational state of the controlled process and transmit feedback information to the controller. Through this closed-loop feedback mechanism, the control system can dynamically adjust its control strategies, thereby achieving precise control and continuous monitoring of the controlled object.

Based on the operational characteristics of MASS navigation, a corresponding safety control structure was developed, as illustrated in Fig. 2. This model systematically represents the control and feedback relationships among various subsystems or functional modules within the ship navigation system.

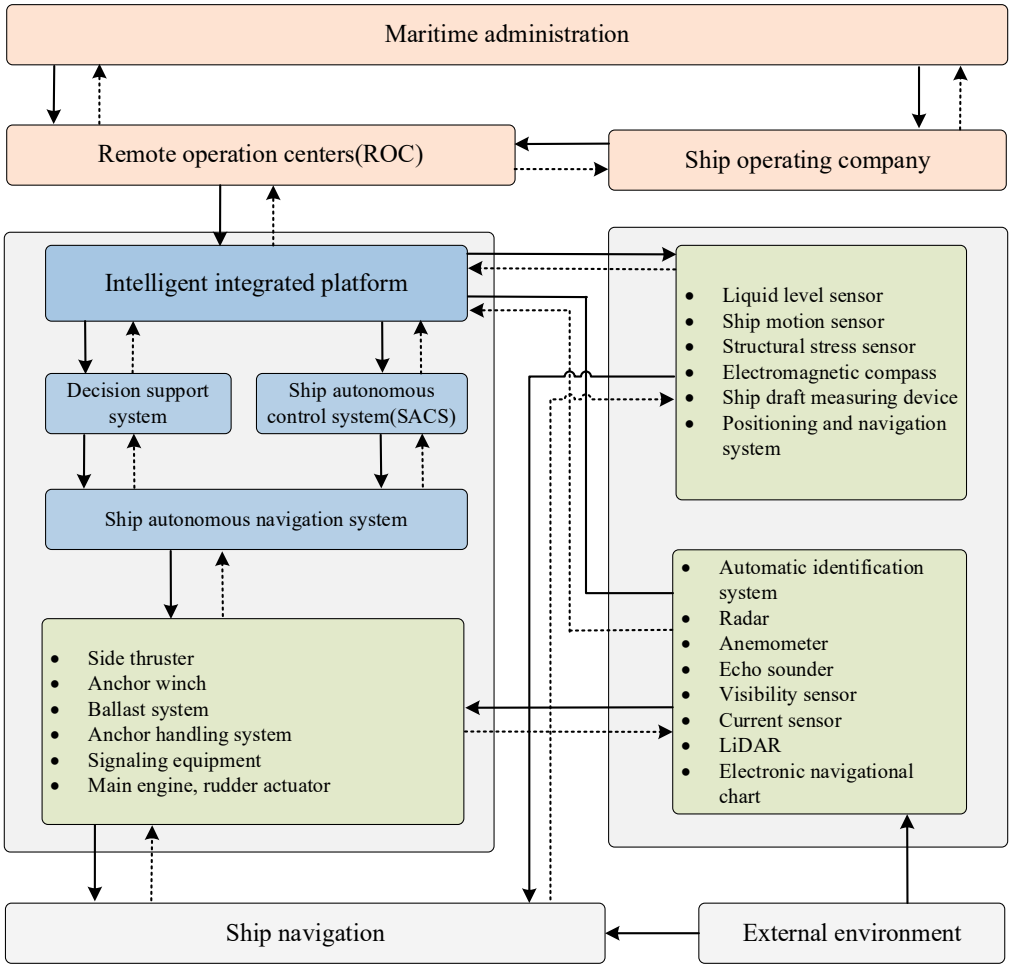


Fig. 2. System safety control feedback structure for MASS navigation.

3.1.2. Identification of unsafe control actions and causes

Based on the safety control structure, the study systematically identifies UCAs across key behaviors such as collision avoidance, route planning, and propulsion control. Following the STPA process, typical UCAs and their causal scenarios were identified and summarized in Table 1. These UCAs represent critical control failures where actions are not provided when needed, provided incorrectly, or executed with improper timing or duration.

The final step of this phase involves mapping these abstract control flaws onto specific physical and logical components to identify root causes. To ensure the objectivity and reproducibility of this process, a structured mapping

protocol was adopted, wherein the synthesis of RIFs was iteratively validated through cross-verification by a panel of maritime safety experts. This approach ensures that the identified factors are not a product of individual analyzer bias but represent a consensus-based systemic perspective. By refining these constraint relationships and integrating insights from historical accident reports and existing literature [61-66], a comprehensive index of key RIFs was synthesized (for details, see Table 2, places P_2-P_{16}). The reliability of the constructed index is further verified through the sensitivity analysis in Section 4.4, confirming its effectiveness in capturing critical risk drivers. This structured index construction provides a validated and robust foundation for the subsequent quantitative dynamic simulation.

Table 1. Analysis of UCAs and causal scenarios in MASS navigation

Causal scenario	Control behavior	Type of unsafe control action
SACS	Collision avoidance decision-making	Not provided: A collision risk is detected, but no evasive action is taken.
		Provided incorrectly: An incorrect collision avoidance command is issued.
		Provided too early/too late: The avoidance maneuver is executed too early (disrupting traffic flow) or too late.
	Route planning	Not provided: System failure results in no execution of route tracking.
		Provided incorrectly: The planned or executed route deviates from the safe channel, heading towards shallow waters or a restricted zone.
ROC	Propulsion and maneuvering control	Not provided: The system hangs, failing to send control commands to the actuators.
		Provided incorrectly: Conflicting commands are issued.
	Remote control command	Not provided: The ROC fails to issue a control command when intervention is required.
		Provided incorrectly: Operator error leads to an incorrect heading or speed command being sent.
		Provided too early/too late: The command is delayed, mismatching the current vessel state and environment.
	Monitoring and alarm handling navigation	Not provided: The operator fails to notice or ignores a critical alarm issued by the system.
		Not provided: The ROC does not initiate timely takeover when the SACS exhibits abnormal behavior.
	Mode switching	Provided incorrectly/at wrong time: Control mode is forcibly switched during a complex or critical navigation phase, causing system confusion.
SACS / ROC (Collaboration)	Ship-shore communication coordination	Stopped providing: The control command or feedback information is completely lost due to communication link interruption.
		Provided incorrectly: Communication is interfered with, transmitting corrupted or erroneous data.

3.2 Phase 2: Hybrid model and mapping mechanism

Building upon the qualitative structural foundation, this phase constructs a hybrid model by applying a rigorous mapping mechanism that transforms the structural outputs of STPA into a computable SPN. SPN, as a probabilistic extension of classical Petri Nets, is uniquely suited for this study due to its exceptional capacity to represent systems

characterized by concurrency, asynchrony, and stochastic behavior. In SPN, transitions are associated with randomly distributed firing delays, enabling the precise characterization of dynamic system behavior and state transition probabilities. This provides an effective mechanism for describing state evolution and event scheduling in discrete-event stochastic systems. As illustrated in Fig. 3, SPN possesses a natural graphical representation, which facilitates the modeling of such complex systems.

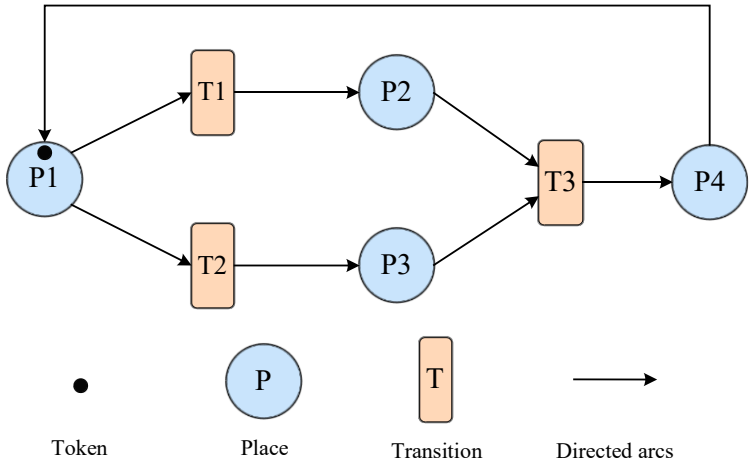


Fig. 3. Structure and transition logic of the SPN module.

The mapping process effectively converts the descriptive causal scenarios and indices identified in Phase 1 into discrete modeling elements to reflect the dynamic evolution of the MASS hierarchical control structure. Specifically, the mapping mechanism is operationalized through the following rules:

Places (*P*): Represent the physical and logical states of system components, such as sensor status, interaction quality, and structural integrity.

Transitions (*T*): Model the occurrence of UCAs and other risk events that cause system state changes.

Directed arcs (*F*): Define the network topology and the direction of system state evolution based on the causal logic chains identified during STPA.

Firing rates (λ): Average rate of successful enactment of each control function; a higher λ means faster processing and lower risk, while the unsafe condition is encoded by a deficient λ together with the RIFs-driven place weights, all calibrated from RIFs frequency and severity.

Building on the element mappings above, the semantics of a token is specified as follows. Consistent with the 1-bounded setting, a token is a binary risk-state indicator rather than a physical entity, information flow, or processing unit, with $M(P_i) \in \{0,1\}$: $M(P_i)=0$ denotes the nominal safe state of a component or factor, and $M(P_i)=1$ denotes a realized control flaw at that node, that is, a hazard-active condition. The system marking $M = (M(P_1), \ldots, M(P_{16}))$ thus encodes the instantaneous distribution of realized control flaws across the perception-decision-control-execution loop. Token presence means risk activation, not active processing: $M(P_8)=1$ indicates that the intelligent integrated platform is operating under a realized perceptual or fusion flaw, not that it is merely handling data, while the initial marking M_1 leaves the hazard place P_1 empty.

Each transition is labelled by the function it governs (for example, T_4, perceive environmental information), and its firing rate λ_j is the competence rate of that function: a higher λ_j denotes faster, more reliable execution that drains probability mass from the hazard markings, whereas a low λ_j encodes the unsafe condition. Adverse RIFs enter at the place level rather than as transitions, through the initial-marking weights and the environmental modifier f_e of Eq. (4) that suppresses competence rates under degraded conditions; calibrated from RIFs frequency and severity, a more severe factor thus lowers the relevant rate and raises the steady-state risk of P_1. For instance, a token reaching P_8 via T_4 and feeding T_{10} instantiates the perception-distortion to decision-error segment of Path 1.

Formally, the STPA-SPN hybrid model is defined as a 6-tuple: $SPN = (P, T, F, W, M_0, \lambda)$, where:

(1) $P=\{P_1,P_2,\cdots,P_m\}$ is a finite set of places, representing all possible states in the system.

(2) $T=\{T_1,T_2,\cdots,T_n\}$ is a finite set of transitions, representing events that cause state changes.

(3) $F\subseteq(P\times T)\cup(T\times P)$ defines the directed arc relation between places and transitions, describing the direction of system state evolution.

(4) $W:F\to N^+$ is the arc weight function, defining the weight on each arc.

(5) M_1 is the initial marking, representing the initial state distribution of the system. The evolution of the system state over time is termed the marking process, denoted as $M=\{M_1,M_2,\cdots,M_j\}$.

(6) $\lambda=\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$ is the set of average firing rates for transitions, where each transition T_i is associated with a stochastic firing rate λ_i.

3.3 Phase 3: Isomorphic Markov chain analysis and risk quantification

3.3.1. Multi-source data fusion and parameterization

In the final phase of the methodological framework, the qualitative structural model is instantiated through quantitative dynamic simulation to evaluate the system's risk performance over time. The rigorous mathematical foundation for this assessment relies on the isomorphism between the SPN and a homogeneous Continuous-Time Markov Chain (CTMC), assuming that transition firing times follow exponential distributions. This isomorphism provides the necessary framework for analyzing steady-state behavior and quantifying long-term risk performance by solving for the state probabilities of the Markov chain [67].

To bridge the gap between qualitative hierarchical control logic and quantitative stochastic simulation, this phase incorporates a multi-source data fusion architecture for model parameterization. As illustrated in Fig. 4, the parameterization process is executed through three standardized steps to ensure the model accurately reflects the non-linear risk evolution in complex maritime environments:

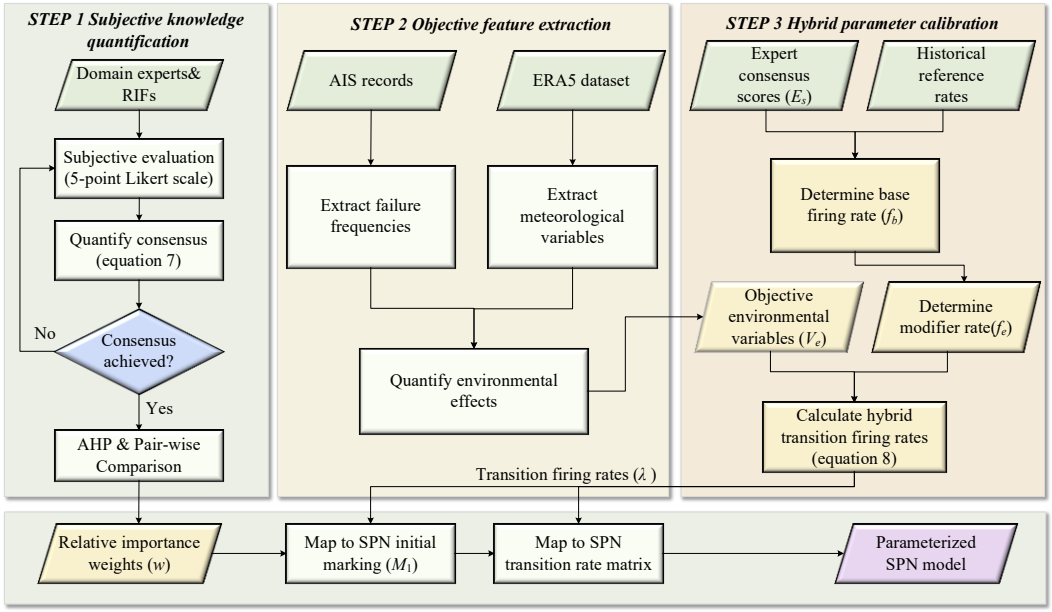


Fig. 4. Framework of multi-source data fusion for model parameterization.

(1) Subjective knowledge quantification

For latent functional interactions within the MASS control loop (e.g., decision-making biases or control inadequacies) that are not directly observable through statistical data, a hybrid Delphi-AHP approach was employed. Initially, domain experts evaluated the relative impact of each RIFs using a five-point Likert scale. To ensure the robustness of these subjective inputs, the consensus degree for each factor *j* was quantified via the standard deviation

(STD_j), see Eq (3) for details:

$$STD_j = \sqrt{\frac{\sum_{i=1}^{n}(S_{ij} - \bar{S}_j)^2}{n}} \tag{3}$$

where S_{ij} denotes the score assigned by expert i to factor j, $\bar{S}_j$ is the mean score, and n is the total number of experts. Upon achieving consensus, the Analytic Hierarchy Process (AHP) was utilized to construct pair-wise comparison matrices, deriving the relative importance weights w for each place. These weights, verified by a Consistency Ratio ($CR < 0.1$), were subsequently used to initialize the state distribution (Initial Marking, M_1) of the SPN model.

(2) Objective feature extraction

To ground the stochastic model in realistic operational contexts, two objective data streams were integrated. First, operational trajectory data, consisting of large-scale high-frequency AIS records, were processed to derive the baseline statistical frequencies of component failures and communication outages, providing a robust probabilistic foundation for transition triggers. Second, environmental reanalysis data—specifically meteorological variables such as wind speed and visibility—were extracted from ERA5 datasets. These data were utilized to quantify the "non-linear modification effects" of external disturbances on the system's perception-decision-execution closed-loop, ensuring the model accurately reflects the dynamic complexities of the maritime environment.

(3) Hybrid parameter calibration

The core of the parameterization lies in the hybrid mapping of transition firing rates λ_j. Rather than employing static values, the model utilizes a dynamic function that couples subjective base rates with objective environmental modifiers:

$$\lambda_j = f_b(\lambda_{jr}, E_s) \bullet f_e(RIFs_e, V_e) \tag{4}$$

where f_b represents the base firing function determined by the expert consensus scores (E_s) and historical reference rates (λ_{jr}). f_e denotes the environmental modifier function based on real-time objective variables (V_e), reflecting the system's vulnerability under extreme environmental coupling.

3.3.2. Computational workflow for isomorphic Markov chains

Once the model is parameterized, the computational workflow is executed through four integrated steps

Step 1: State mapping. Each marking M_i in the SPN is uniquely mapped to a state S_i in the CTMC. The mapping function is defined as Eq (5):

$$S_i = f(M_i), (i = 1, \ldots, N) \tag{5}$$

Step 2: Construction of the CTMC transition rate matrix. If marking M_i in the SPN can transition to M_j via a set of transitions $T(M_i \rightarrow M_j)$, then the transition rate from S_i to S_j in the CTMC equals the sum of the firing rates of all transitions in this set, see Eq (6) for details:

$$q_{ij} = \sum\nolimits_{\tau \in T(M_i \rightarrow M_j)} \lambda_\tau \tag{6}$$

Step 3: Formulation of steady-state balance equations. When the CTMC reaches a steady state, the inflow rate equals the outflow rate for each state, as described by the global balance Eq (7):

$$\begin{cases} \pi_i q_i = \sum\limits_{i \neq j} \pi_j q_{ij} \\ \sum_{i=1}^{N} \pi_i = 1 \end{cases} \tag{7}$$

where q_{ij} denotes the transition rate from state S_i to S_j, and q_i represents the total outflow rate from state S_i.

Step 4: Calculation of steady-state probabilities. The steady-state probability of system risk is derived as the sum of the steady-state probabilities corresponding to all risk-related markings, see Eq (8) for details:

$$\pi P_m = \sum_{M_i \in \Omega} \pi_i \tag{8}$$

where, P_m represents any places, Ω denoting a set of risk identifiers.

Steady-state convergence was evaluated based on the relative variation of risk probabilities. Let $P_i(t)$ denote the probability of risk path (i) at time (t). The system was considered to have reached steady state when:

$$\frac{|P_i(t+\Delta t) - P_i(t)|}{P_i(t)} < \varepsilon \tag{9}$$

where $\varepsilon = 1\%$ [68] and Δt =5 min. The criterion was required to hold for all risk paths over three consecutive observation intervals. The earliest time satisfying this condition was defined as the convergence time of the corresponding risk path.

By executing this dynamic simulation engine, the framework successfully transforms the structural logic of Phase 1 into computable, time-dependent risk performance indicators, identifying steady-state probabilities and critical risk paths and providing an innovative and practical quantitative basis for MASS systemic vulnerability characterization and control optimization in complex maritime environments.

4. Case study

To validate the effectiveness of the proposed framework, the Taicang Port section of the Yangtze River is selected as the study area. While open-sea testing is common for MASS, the ultimate deployment goal involves berth-to-berth autonomy, which necessitates navigating complex inland waterways. This section represents a critical "mixed traffic" environment, characterized by high vessel density (approximately 1,500 vessels per day) and narrow channels. Validating the risk analysis model in such a highly constrained environment ensures its robustness and effectively captures the complexity of future ship-shore coordinated navigation. This scenario is illustrated in Fig. 5.

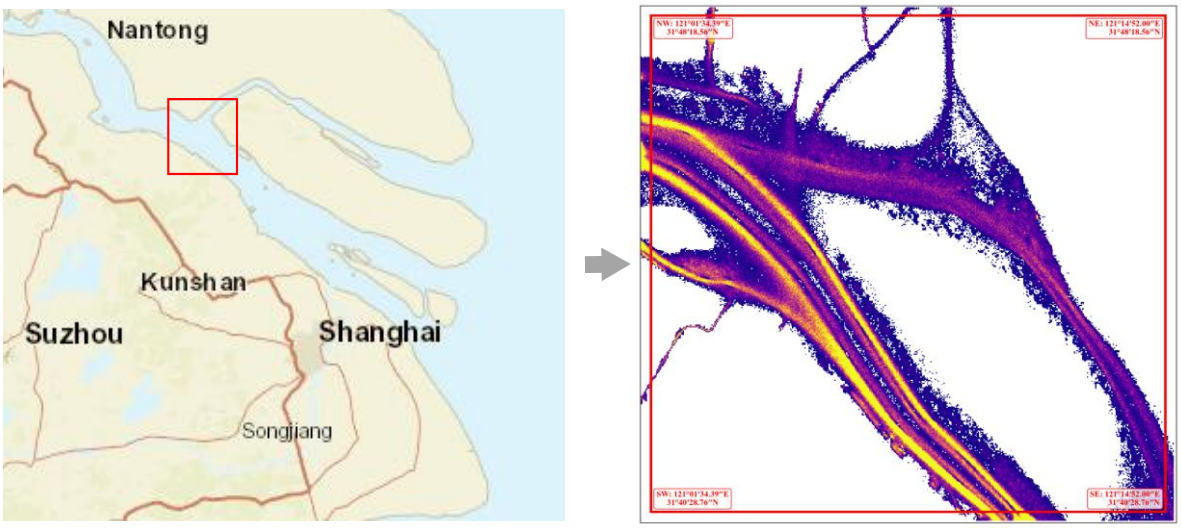


Fig. 5. Taicang port location and ship trajectory distribution, March 2025.

4.1 Data collection

According to the multi-source data fusion framework established in Section 3.3, a comprehensive data collection process was implemented for the case study. The data are categorized into objective and subjective dimensions as follows:

(1) Objective Data: Approximately 12 million high-frequency AIS records were collected from the investigated waters in March 2025, providing detailed dynamic and static information regarding vessel trajectories and behaviors.

To enhance analytical efficiency while preserving accuracy, these data underwent rigorous processing, including outlier removal, trajectory smoothing using a standard Kalman filter, and redundancy compression. Additionally, environmental variables such as wind speed and visibility were extracted from the European Centre for Medium-Range Weather Forecasts (https://cds.climate.copernicus.eu/) ERA5 reanalysis dataset to quantify external disturbances.

(2) Subjective Data: To characterize risks arising from human factors and complex system interactions, data were collected from a panel of ten domain experts using the Delphi method (demographic details are provided in Table A2). The panel represents four core professional backgrounds: maritime management, technical R&D, navigation practice, and academic research. Each member possesses over 10 years of professional experience, ensuring the technical depth of the qualitative inputs.

To ensure the scientific rigor of the subjective inputs, the consensus degree for each factor was quantified via standard deviation according to Eq (3), and the assessment consistency was further validated through a correlation heatmap. The expert panel provided critical support in constructing the hierarchical control structure, determining transition firing rates, and formulating state transition matrices. These data constitute the foundation for the quantitative analysis in Section 3.3, providing the basis for parameter calibration in dynamic simulation and subsequent risk performance evaluation.

4.2 Hybrid model mapping and parameterization results

4.2.1. Identification of risk propagation paths

Building upon the identified UCAs and their causal scenarios, this study identifies three typical risk propagation paths leading to distinct system-level accidents. These paths are developed using a dual-driven analytical approach that synergizes qualitative systemic insights with quantitative validation. By first integrating STPA to systematically identify structural hazards and then employing expert validation to calibrate these results, the framework ensures a robust mapping of the logical evolution from component-level failures to system-level accidents. This approach follows the closed-loop logic of perception-decision-control-execution to accurately reflect the functional dependencies of MASS.

The selection of these primary risk propagation paths is governed by three rigorous criteria encompassing severity, representativeness, and expert consensus. First, the severity criterion ensures that priority is given exclusively to paths leading to high-consequence system-level accidents, specifically ship collision, grounding, and loss of control. Second, representativeness requires each path to encompass the complete perception-decision-control-execution functional loop to capture the full scope of potential interaction failures. Finally, expert consensus dictates that the identified paths are validated by a professional panel as the most critical operational edge cases for the Taicang section of the Yangtze River.

This systematic filtering ensures the model captures the most critical causal scenarios while maintaining computational feasibility for dynamic simulation. The structural evolution and causal logic of these scenarios are formalized in Fig. 6, which provides a system-theoretic visualization of the three key risk chains across the perception, decision, and control layers. By mapping abstract control flaws onto specific physical and logical components, this visualization establishes the essential foundation for subsequent quantitative stochastic simulation. The three resulting paths are characterized as follows.

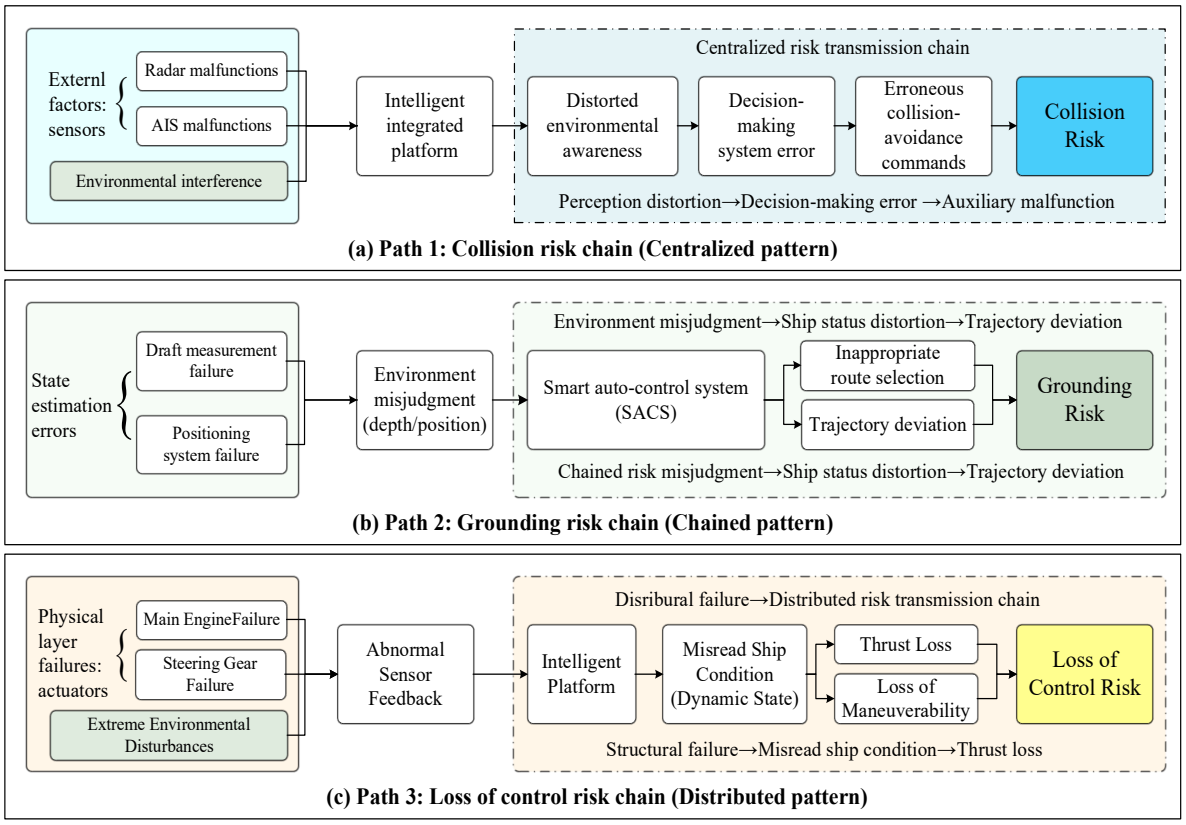


Fig. 6. Systematic evolution and propagation patterns of the three key MASS risk chains.

Path 1 (Collision risk chain): This path primarily originates from failures in the perception phase, such as malfunctions in external sensors like radar or AIS or interference from environmental factors. This leads to distorted environmental awareness within the Intelligent Integrated Platform, which propagates to the decision-making system and triggers erroneous collision-avoidance actions. Simulation results characterize this as a "centralized risk transmission chain" following the pattern of "perception distortion → decision-making error → auxiliary malfunction".

Path 2 (Grounding risk chain): Driven by errors in decision-making and state estimation, this path involves failures in draft measurement or positioning systems that cause the platform to misjudge the ship's position or water depth. Such distortion leads the SACS to select an inappropriate route or fail to maintain a safe trajectory. This path exhibits a "chained risk transmission pattern" described as "environment misjudgment → ship status distortion → trajectory deviation".

Path 3 (Loss of control risk chain): Typically initiating from physical failures in actuators, such as the main engine or steering gear, or extreme environmental disturbances, this path results in abnormal sensor feedback. Consequently, the intelligent platform fails to accurately perceive the vessel's dynamic state, preventing effective control execution and leading to a loss of maneuverability. This reflects a "distributed risk transmission chain" summarized as "structural failure → misread ship condition → thrust loss".

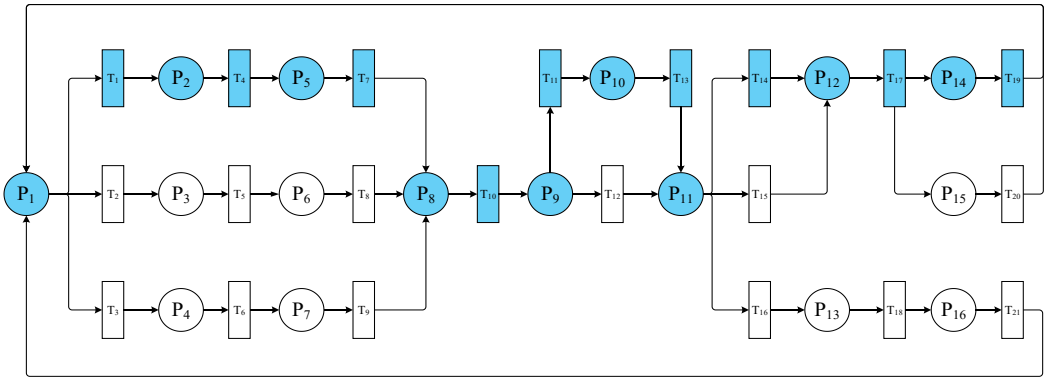
4.2.2. STPA-SPN mapping and network structure

The mapping rules follow those defined in Section 3.2; their application to the MASS control architecture is described below.

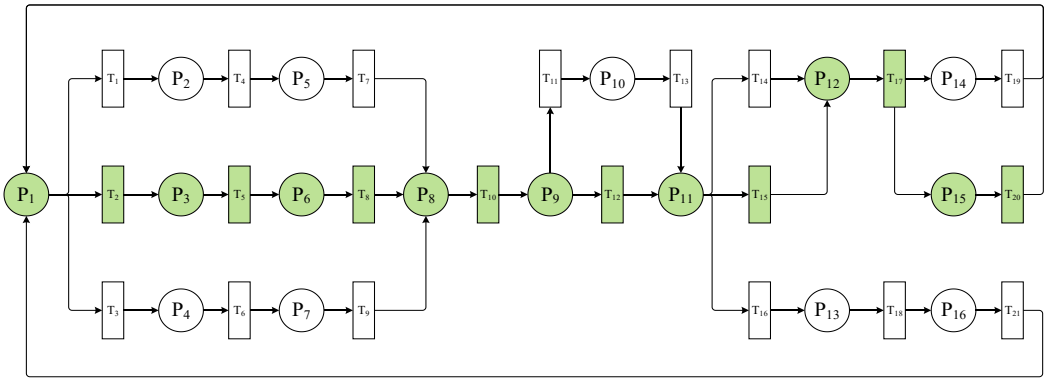
The mapping process converts the descriptive causal scenarios and indices identified in Phase 1 into discrete modeling elements to reflect the dynamic evolution of the MASS control architecture. Specifically, the physical and logical states of components—such as sensor status, interaction quality, and structural integrity—are represented as places. The occurrence of UCAs and other risk events that drive system state changes are modeled as transitions. The network topology and the direction of state evolution are defined by directed arcs based on the causal logic chains

identified during the STPA phase. Finally, the average speed at which these transitions occur is determined by firing rates, which are calibrated based on the frequency and severity of the associated RIFs.

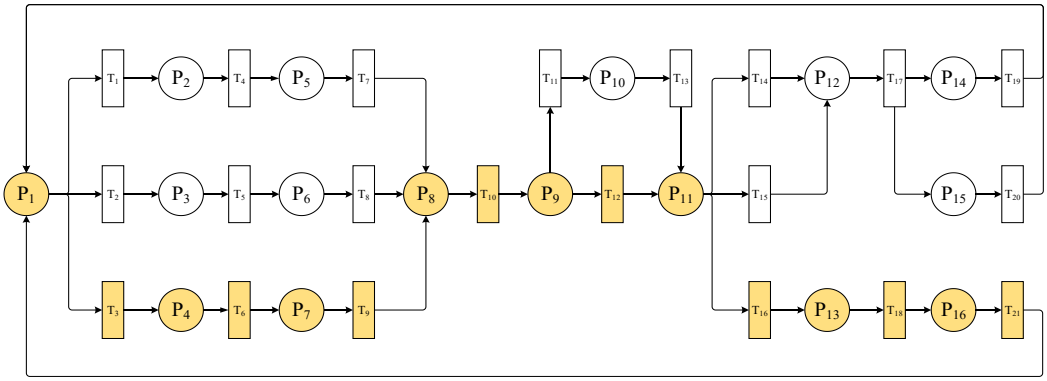
As illustrated in Fig. 7, the final STPA-SPN model for MASS navigation comprises 16 places and 21 transitions. The places represent the binary risk states of the RIFs and the associated system components, whereas the transitions represent the control functions that operate on those factors, with each firing rate interpreted as the competence rate of the corresponding function. Functional descriptions of the 16 places are provided in Table 2, while the 21 transitions are listed in Table 3 together with their correspondence to the STPA unsafe control action types. This correspondence covers all four UCAs categories: a low competence rate captures the UCAs where a required action is not provided, not properly executed, or stopped too soon, since the loop then fails to enforce its safety constraint in time, whereas a flawed upstream marking propagated through an otherwise competent firing captures the UCAs where an incorrect action is provided or provided too early or too late. Unsafe control is thus represented consistently as a deficiency or distortion of competence rather than as a high-frequency adverse event, completing the structural foundation necessary for the subsequent quantitative dynamic simulation.



(a) Path 1



(b) Path 2



(c) Path 3

Fig. 7. STPA-SPN network model for MASS navigation.

Table 2. Place descriptions for the SPN model

Symbol	Indicator description	Symbol	Indicator description
P_1	MASS navigation risk	P_9	SACS
P_2	Natural environment	P_{10}	Decision support system
P_3	Traffic environment	P_{11}	Autonomous ship navigation system
P_4	Ship structure	P_{12}	Steering gear
P_5	Environmental monitoring sensors	P_{13}	Main engine
P_6	Ship status perception sensors	P_{14}	Heading and speed
P_7	Ship condition perception sensors	P_{15}	Navigation track
P_8	Intelligent integrated platform	P_{16}	Lateral thruster

Table 3. Transition descriptions and their correspondence to STPA unsafe control actions

Symbol	Indicator description	Corresponding STPA UCAs type
T_1	Acquire environmental data	Omission / inadequate execution
T_2	Monitor traffic situation	Omission / inadequate execution
T_3	Detect structural status	Omission / inadequate execution
T_4	Perceive environmental information	Omission / inadequate execution
T_5	Read ship status data	Omission / inadequate execution
T_6	Collect ship condition data	Omission / inadequate execution
T_7	Process environmental information	Omission / inadequate execution
T_8	Process ship status information	Omission / inadequate execution
T_9	Process ship condition information	Omission / inadequate execution
T_{10}	Generate control decisions	Incorrect commission
T_{11}	Formulate assistance strategies	Incorrect commission
T_{12}	Plan navigation mission	Incorrect commission
T_{13}	Perform assisted navigation	Wrong timing / duration
T_{14}	Control rudder angle	Wrong timing / duration
T_{15}	Adjust main engine thrust	Wrong timing / duration
T_{16}	Operate lateral thruster	Wrong timing / duration
T_{17}	Adjust heading and speed	Wrong timing / duration
T_{18}	Regulate main engine thrust	Wrong timing / duration
T_{19}	Execute collision avoidance maneuver	Wrong timing / duration
T_{20}	Calculate safe route	Incorrect commission
T_{21}	Trigger emergency response	Wrong timing / duration

4.3 Dynamic simulation and risk performance results

4.3.1. Homogeneous Markov model

The quantitative analysis of MASS navigation risk relies on the mathematical isomorphism between the SPN

and a homogeneous CTMC. The state of the SPN is fundamentally defined by the distribution of tokens within its places. Consequently, an isomorphic Markov chain is constructed by mapping these token distributions to discrete states, with transitions determined by the firing rates of the system. This representation is valid under the assumption that firing times follow an exponential distribution and tokens are countable, allowing each transition firing to trigger a corresponding change in the system state.

To determine the dynamic risk performance indicators, the calculation process follows a rigorous sequence of four integrated steps: Each reachable marking in the SPN is uniquely mapped to a corresponding state in the Markov chain; A transition rate matrix is developed where the rate between states is defined by the sum of firing rates for all transitions that trigger the movement between markings; The system is analyzed at its steady state, predicated on the global balance principle where the total inflow rate for each state equals its total outflow rate; The final steady-state probability of the system risk is calculated by summing the probabilities of all markings that represent specific risk identifiers.

To determine the dynamic risk performance indicators, the calculation process follows a rigorous sequence of four integrated steps. Initially, each reachable marking in the SPN is uniquely mapped to a corresponding state in the Markov chain using the mapping function defined in Eq (5). Subsequently, a transition rate matrix is developed following Eq (6), where the rate between states is defined by the sum of firing rates for all transitions. The system is then analyzed at its steady state, which is predicated on the global balance principle established in Eq (7). Finally, the steady-state probability of the system risk is calculated by summing the probabilities as per Eq (8). Through this methodological pipeline, the framework computes the steady-state probabilities of P_1, quantifying MASS navigation risk performance as defined in Eq (1).

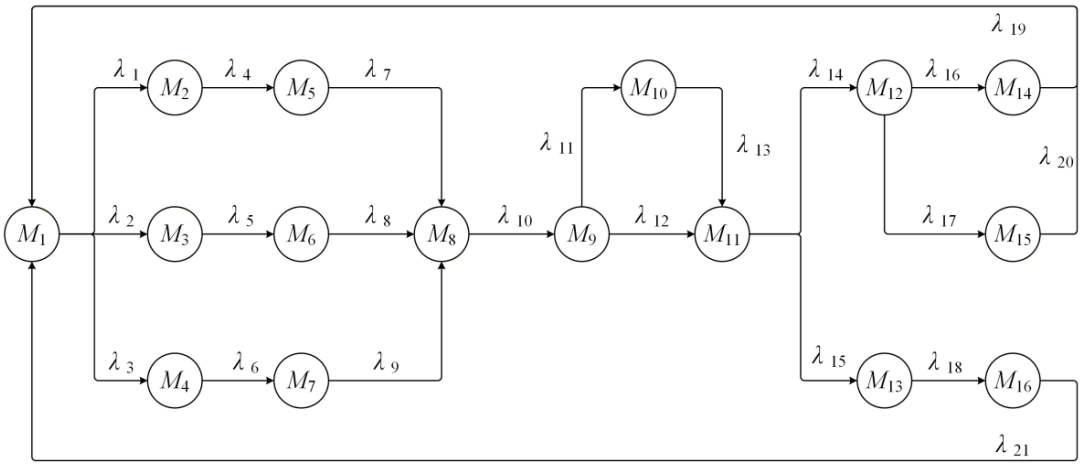


Fig. 8. Homogeneous Markov model.

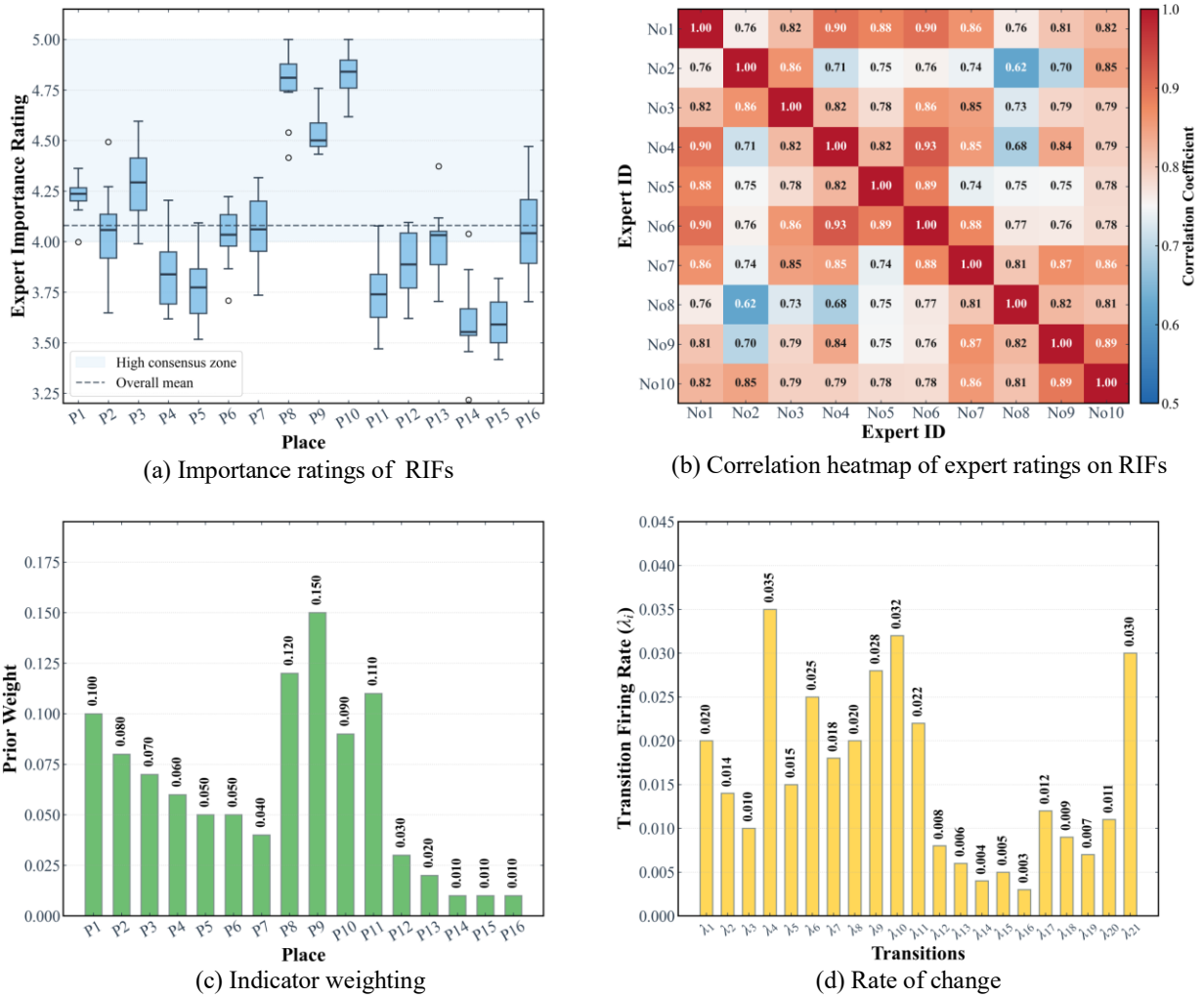
As illustrated in Fig. 8, the resulting homogeneous Markov model starts from the initial state marking M_1 to obtain a complete set of reachable state markings. By defining $P(M_i)$ as the probability of encountering risk node M_i and λ_i as the average firing rate of the i-th state transition, the specific steady-state probability system for the MASS navigation risk model is derived as Eq (10).

$$
\begin{cases}
\lambda_{19}P(M_{14})+\lambda_{20}P(M_{15})+\lambda_{21}P(M_{16})=(\lambda_1+\lambda_2+\lambda_3)P(M_1)\\
(\lambda_1+\lambda_2+\lambda_3)P(M_1)=\lambda_4P(M_2)+\lambda_5P(M_3)+\lambda_6P(M_4)\\
\lambda_4P(M_2)=\lambda_7P(M_5)\\
\lambda_5P(M_3)=\lambda_8P(M_6)\\
\lambda_6P(M_4)=\lambda_9P(M_7)\\
\lambda_7P(M_5)+\lambda_8P(M_6)+\lambda_9P(M_7)=\lambda_{10}P(M_8)\\
\lambda_{10}P(M_8)=\lambda_{11}P(M_9)+\lambda_{12}P(M_9)\\
\lambda_{11}P(M_9)=\lambda_{13}P(M_{10})\\
\lambda_{12}P(M_9)=\lambda_{14}P(M_{11})+\lambda_{15}P(M_{11})\\
\lambda_{14}P(M_{11})+\lambda_{15}P(M_{11})=\lambda_{16}P(M_{12})+\lambda_{17}P(M_{12})+\lambda_{18}P(M_{13})\\
\lambda_{16}P(M_{12})=\lambda_{19}P(M_{14})\\
\lambda_{17}P(M_{12})=\lambda_{20}P(M_{15})\\
\lambda_{18}P(M_{13})=\lambda_{21}P(M_{16})
\end{cases}
\tag{10}
$$

4.3.2. Multi-source parameterization and expert consensus

By integrating specific AIS trajectory records and environmental variables from the Taicang section of the Yangtze River, this study transforms the STPA-SPN framework into a context-aware simulation model tailored for high-density mixed traffic environments. The effective operation of the STPA-SPN model relies heavily on the precise acquisition of experimental parameters (including transition parameters) during the preliminary stage. This process adopts a collaborative method combining expert assessment and simulation validation, aiming to precisely define transition rates, initial markings, place weights, and state transition matrices. These parameters collectively govern the dynamic behavior of the model and the subsequent risk evolution process.

The transition rates (λ), which reflect the frequency of risk events and the rate of system state changes, are derived using the hybrid mapping function defined in Eq (4), which couples subjective base rates with objective environmental modifiers. Navigation data consisting of approximately 12 million high-frequency AIS trajectory records collected in March 2025 provided the basis for deriving sensor failure rates and communication interruption frequencies, establishing a robust probabilistic foundation for transition triggers. Simultaneously, environmental variables such as wind speed and visibility were extracted from the ERA5 reanalysis dataset to quantify the non-linear impact of external disturbances on the system's perception-decision-control-execution closed-loop. Finally, latent behaviors that are difficult to observe directly, such as human operational errors and dysfunctional system interactions, were quantified using the Delphi method with an expert panel of 10 professionals possessing over 10 years of experience across maritime management, technical R&D, navigation practice, and academic research.



(a) Importance ratings of RIFs

(b) Correlation heatmap of expert ratings on RIFs

(c) Indicator weighting

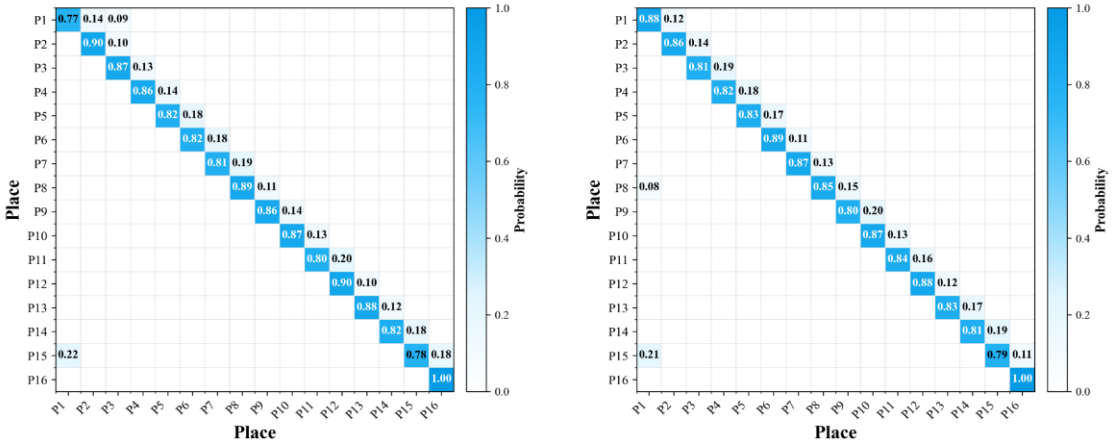
(d) Rate of change

Fig. 9. The assessment results of the experts and the input data of the model.

The importance of the 15 identified RIFs was quantitatively assessed through expert elicitation, as illustrated in Fig. 9(a). The results demonstrated an overall average importance rating of 4.07, with scores significantly concentrated within the high-criticality range of 4.0 to 5.0, underscoring their essential role in the safety assurance of MASS. Assessment consistency was further validated through a correlation heatmap shown in Fig. 9(b), where coefficients between experts ranged from 0.47 to 1.00, confirming a high degree of industry consensus regarding these systemic vulnerabilities.

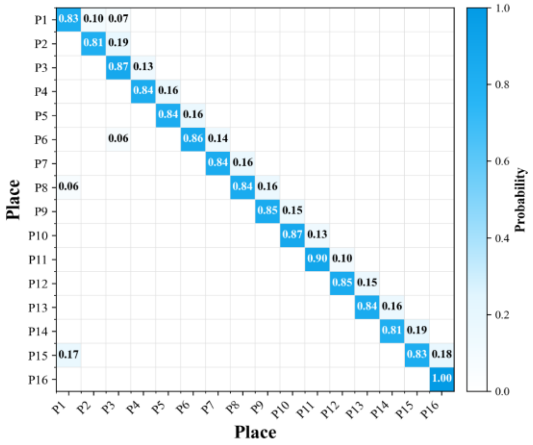
The initial marking (M_1) was established to represent the system's baseline state distribution, where the "MASS navigation risk" place is initialized to a zero-risk state. This study utilized the AHP fused with expert scoring to determine final parameter weights for each place, as summarized in Fig. 9(c). Complementarily, the stochastic firing rates (λ) for the 21 transitions were derived through the multi-source data fusion framework and are presented in Fig. 9(d).

Building upon the UCAs and causal paths identified during the STPA phase, state transition matrices were constructed as shown in Fig. 10(a-c). These matrices specify rigorous logical transitions between places, such as the sequential evolution from perception failure to decision error and eventual control inadequacy, to effectively quantify the dynamic evolution laws of real-world risk coupling.



(a) Path 1 transfer matrix

(b) Path 2 transfer matrix



(c) Path 3 transfer matrix

Fig. 10. The calculation result of the state transition matrix.

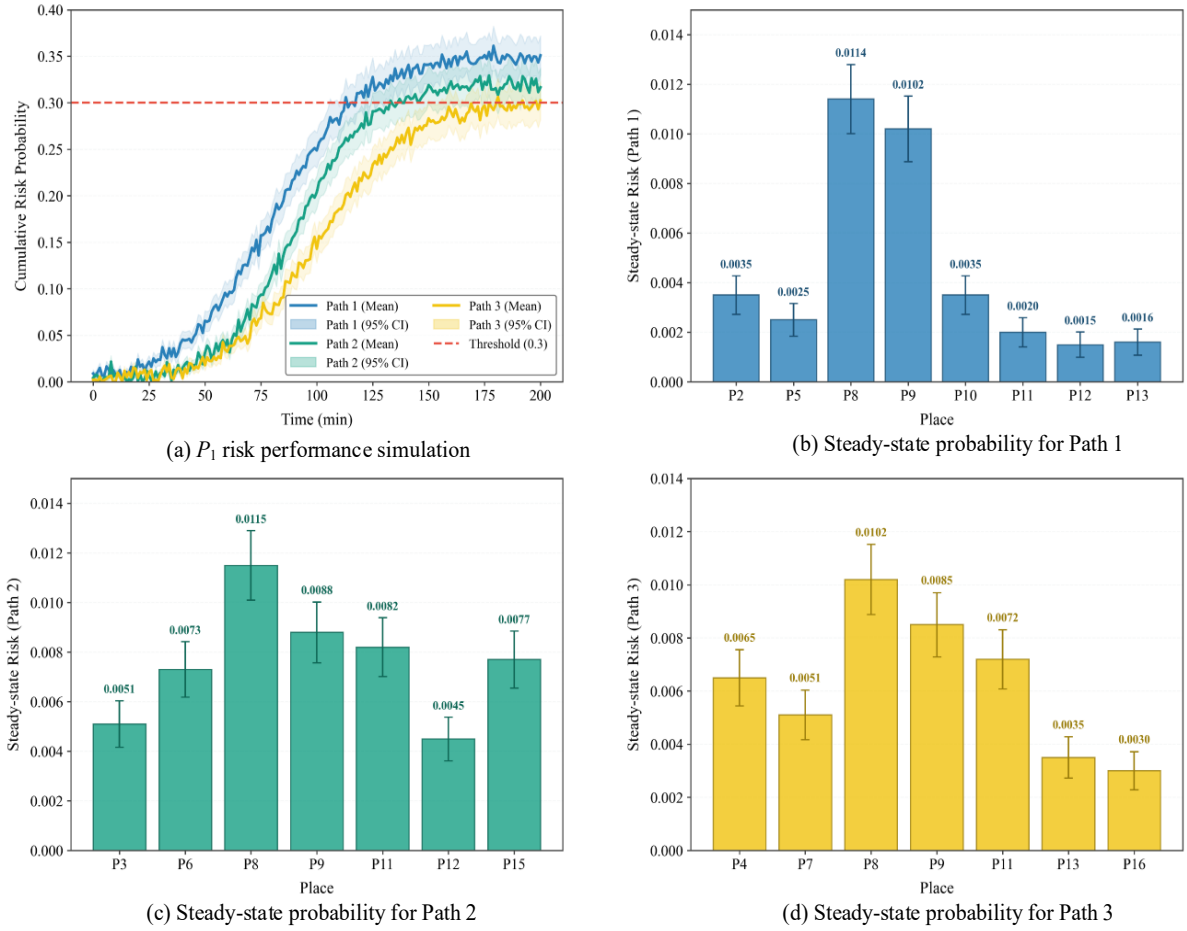
The synthesis of these multi-source parameters, encompassing the 21 calibrated transition rates (λ) and 16 weighted places (P), effectively establishes a digital footprint of the Taicang waterway's operational environment. By mapping qualitative causal scenarios onto these quantitative matrices, the model accurately captures the interaction patterns between system components under actual maritime stressors. Consequently, this parameterization process constitutes the core input for subsequent dynamic simulations, from which specific risk performance results and steady-state probabilities for the Taicang waterway case study are derived.

4.3.3. Navigation risk performance of the MASS

Building on the STPA-SPN model, this study characterized the dynamic evolution of MASS navigation risk through 2000 Monte Carlo replications over a 200 min simulation horizon. The selection of this horizon is justified on both operational and mathematical grounds. Operationally, it is intended to capture the complete evolution of the port-entry and maneuvering scenario considered in this study, covering the progression from hazard initiation to system stabilization. Mathematically, the slowest risk propagation path (Path 3) reaches its statistical steady state at approximately 160 min, and the selected horizon therefore provides an additional margin of about 25% to ensure that transient effects have fully dissipated. Consequently, the reported steady-state probabilities represent stabilized system behavior rather than artifacts of an insufficient simulation duration. Solving the steady-state system defined in Eq. (8) subsequently yields the steady-state risk probability of P1, as illustrated in Fig. 11.

The simulation results clearly reveal that the evolution of navigation risk for the MASS is not random but exhibits significant path dependence and system lock-in characteristics. Different initial risk triggers, decision-

making logic, and environmental interaction patterns collectively shape distinctly different risk evolution trajectories and final steady-state levels.



(a) P_1 risk performance simulation

(b) Steady-state probability for Path 1

(c) Steady-state probability for Path 2

(d) Steady-state probability for Path 3

Fig. 11. Dynamic risk evolution for MASS from SPN simulation results.

Specifically, the three risk paths exhibit distinct evolutionary patterns:

Path 1, which leads ultimately to collision risk, demonstrates the fastest risk convergence rate, the highest risk level, and the most significant fluctuations. It reaches a steady state approximately 120 minutes into the simulation, with a final steady-state risk probability stabilizing at 0.35. As shown in the steady-state probability distribution in Fig. 10(b), this path reveals a notable risk aggregation effect. As quantified by Eq. (2), the node contributions for Path 1 are C_8 = 37.75% and C_9 = 33.77%, yielding a combined contribution of $C_8 + C_9$ = 71.52% > 70%. The intelligent integration platform (P8) and SACS (P9) thus collectively account for over 70% of the steady-state risk probability, forming a centralized risk transmission chain characterized by "perception distortion → decision-making error → auxiliary malfunction." System vulnerability is primarily concentrated in the core control modules.

Path 2, ultimately leading to grounding risk, shows a risk convergence rate and risk level intermediate between those of Path 1 and Path 3. It stabilizes after about 130 minutes, with a steady-state risk probability of 0.32. Fig. 10(c) indicates that key nodes such as intelligent integration platform (P_8), SACS (P_9), and navigation trajectory (P_{15}) together contribute over 65% of the steady-state probability, forming a chained risk transmission pattern described as "environment misjudgment → ship status distortion → trajectory deviation." System vulnerabilities are distributed across environmental perception and trajectory control modules, with risks mainly concentrated in the interactive link between "perception and trajectory control."

Path 3, oriented toward loss-of-control risk, exhibits a slower convergence rate and a lower risk level compared to Path 1 and Path 2. It enters a steady state after approximately 160 minutes, with the final steady-state risk

probability stabilizing at 0.30. According to the steady-state probability distribution in Fig. 10(d), key nodes including intelligent integration platform (P_8), SACS (P_9), and thruster device (P_{16}) collectively account for more than 60% of the steady-state probability, forming a distributed risk transmission chain summarized as "structural failure → misread ship condition → thrust loss." System vulnerabilities are scattered across structural perception and power control modules, with risks exhibiting a decentralized aggregation pattern throughout the "perception–decision–execution" loop.

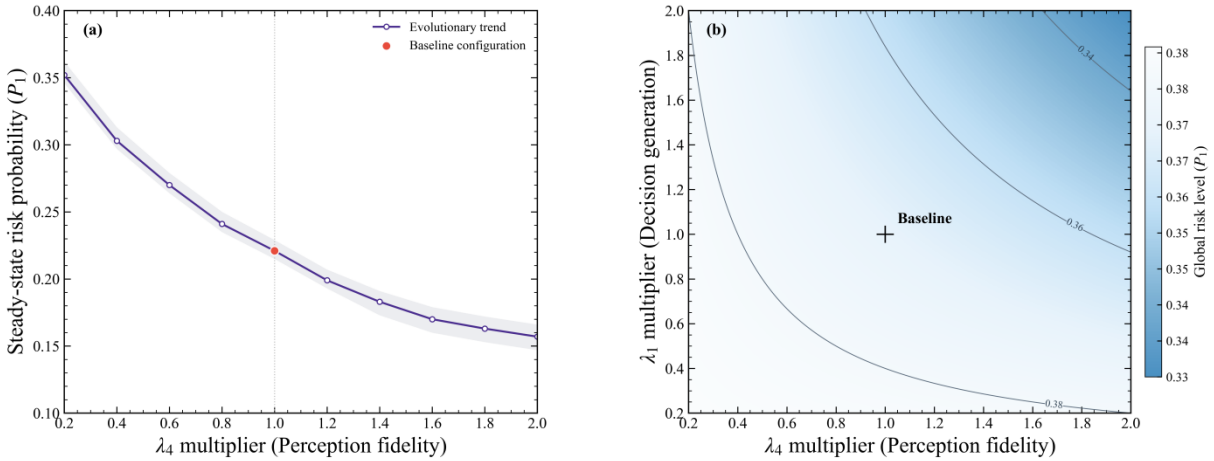
According to the convergence criterion defined in Section 3.3.2, Path 1, Path 2 and Path 3 reached steady state at approximately 120, 130 and 160 min, respectively, with subsequent probability variations of 0.57%, 0.94% and 0.67% over the remaining horizon, all below the predefined 1% threshold. This confirms that the reported steady states are intrinsic to the CTMC rather than artifacts of the chosen time window. To quantify statistical uncertainty, an analysis was conducted over the 2000 replications. Because the model outputs are steady-state probabilities, the 95% confidence interval of each metric was obtained from the central limit theorem using the standard error of a proportion. For the three path-level steady-state risk probabilities this yields 0.35 (95% CI [0.329, 0.371]), 0.32 (95% CI [0.300, 0.340]), and 0.30 (95% CI [0.280, 0.320]), corresponding to relative standard errors of 3.0%, 3.3%, and 3.4%, all below the 5% threshold commonly adopted for Monte Carlo reliability [69]. These narrow intervals confirm that 2000 replications balance computational efficiency against statistical reliability.

In summary, the simulation experiments confirm that the evolution of navigation risks in MASS is jointly determined by initial RIFs and system interaction logic. The steady-state levels, convergence rates, and nodal distribution characteristics of different risk paths provide a quantitative basis for precisely identifying risk control priorities: Path 1, with the most rapid evolution, requires interruption through enhanced system redundancy and decision error-correction mechanisms. Path 2 calls for improved coordination robustness between perception and control loops. Path 3 necessitates a resilient defense system covering monitoring, maintenance, and functional backup. Together, the three paths reveal that the intelligent integration platform and the SACS system constitute the central hubs of global system vulnerability. Future efforts should shift toward a "resilience-adaptation" design philosophy, enabling dynamic risk perception and intelligent fault-tolerant control at the system level, thereby enhancing overall navigation safety.

4.4 Sensitivity and robustness verification

4.4.1. Sensitivity analysis

To validate the sensitivity of the constructed SPN risk assessment model to its key parameters, a combined local and global sensitivity analysis was conducted. This approach examines the influence of both individual parameters and multi-parameter interactions on the system's steady-state risk probability. Two critical parameters, perceive environmental information (λ_4) and generate control decisions (λ_{10}), were selected to represent core performance indicators of the perception and decision-making modules, respectively. As competence transitions, λ_4 and λ_{10} exert a monotonically risk-reducing effect: a higher firing rate reflects more efficient and reliable functional processing, thereby attenuating rather than amplifying the steady-state risk probability of P_1. These parameters were prioritized because they represent the primary "bottleneck" transitions in the ship-shore control loop and exhibited high variance in expert ratings, suggesting they are the main drivers of global uncertainty.



(a) Local sensitivity analysis (b) Global sensitivity analysis

Fig. 12. Model sensitivity verification results.

As illustrated in Fig. 12(a), the local sensitivity analysis independently adjusted these firing rates, which are parameterized through the functional coupling in Eq (4), across a range from 0.2x to 2.0x of the baseline to record changes in the steady-state probability of the system risk (P_1). The results indicate that as λ_4 was incrementally increased, the steady-state probability of P_1 demonstrated a non-linear decreasing trend, declining from a maximum of 0.3512 to 0.1563, which represents a reduction of 55.5%. Notably, the marginal benefit of increasing λ_4 for risk reduction was more pronounced when the value was below the baseline; once λ_4 exceeded the baseline, the risk mitigation effect gradually approached saturation. In comparison, individually increasing λ_{10} also reduced system risk, but the magnitude of its influence was smaller than that of λ_4, with maximum single-parameter reductions being 38.0% for λ_4 and 27.4% for λ_{10}. This identifies that improving environmental perception and data transmission capability holds a higher priority for overall risk control.

The global sensitivity analysis results, presented in Fig. 12(b), reveal a significant synergistic effect on system risk when λ_4 and λ_{10} were adjusted simultaneously. When both parameters were in the low-performance range, the steady-state probability of P_1 was as high as 0.3876, indicating a high-risk system state. In contrast, when both parameters were elevated to the high-performance range, P_1 could be reduced to 0.1232, demonstrating a substantial risk control effect. Further analysis showed that the synergistic optimization of λ_4 and λ_{10} could maximize risk control effectiveness. Specifically, within the interval where λ_4 was between 0.9 and 1.1 times the baseline and λ_{10} was between 1.0 and 1.2 times the baseline, the system risk P_1 could be effectively controlled within the range of 0.18 to 0.22, providing a clear target range for parameter optimization in engineering practice.

4.4.2. Robustness verification

To further validate the robustness of the weight derivation process, a stochastic sensitivity analysis was conducted by introducing $\pm 5\%$ and $\pm 10\%$ fluctuations to the initial input scores of the 15 RIFs (P_2-P_{16}). As illustrated in the Fig. 13, despite the introduced noise, the resulting normalized weights remain highly stable with minimal variance (standard deviation remains below 0.005). This confirms that the hybrid Delphi-AHP and multi-source data fusion architecture effectively filters out local data inconsistencies. The structural stability of the weights ensures that the identified 'risk convergence hubs' are not artifacts of subjective bias but are consistent systemic vulnerabilities, thereby strengthening the reliability of the overall risk assessment framework.

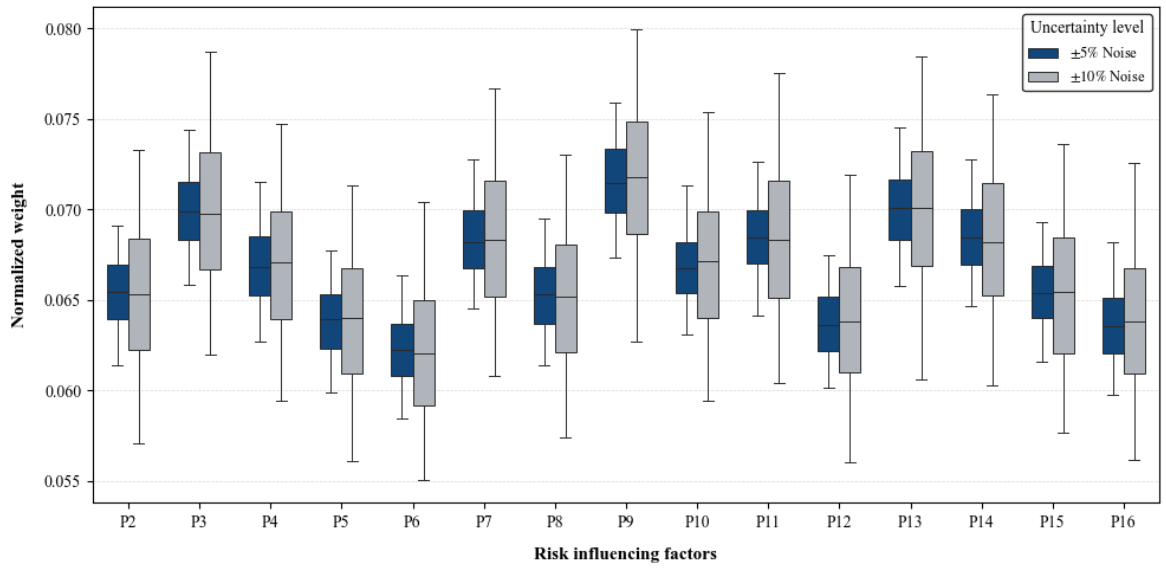


Fig. 13. Robustness verification of weighting derivation under input noise

In summary, the comprehensive evaluation of sensitivity and robustness confirms the technical validity and operational reliability of the proposed STPA-SPN framework. The local and global sensitivity analyses successfully identified environmental perception (λ_4) as the primary "leverage point" for risk mitigation, revealing that its optimization yields a more significant reduction in steady-state risk probability compared to decision-making modules. Furthermore, the robustness verification demonstrated that the model's weight distribution remains remarkably stable even when subjected to $\pm 10\%$ stochastic noise in input data. This dual validation ensures that the framework is not only highly responsive to critical safety parameters but also resilient against the inherent uncertainties of maritime big data and expert elicitation. Consequently, the model provides a high-fidelity decision-support tool capable of pinpointing systemic vulnerabilities while maintaining consistent performance in complex, dynamic navigation environments.

5. Discussion and Implications

5.1 The dynamic risk evolution and path dependence of MASS

The risk performance during MASS navigation constitutes a complex system behavior that evolves dynamically based on mission phases, environmental stressors, and the internal coupling of control states. Simulation analysis utilizing the proposed STPA-SPN hybrid model reveals that risk performance exhibits significant time-varying characteristics and profound path dependency across both temporal and functional dimensions. This finding aligns with the systems-theoretic perspective established by [56], which posits that accidents in highly automated sociotechnical systems are emergent properties arising from the inadequate enforcement of safety constraints rather than isolated component malfunctions. While traditional maritime safety assessments often treat risk as a linear accumulation of failure probabilities [70], the framework captures the non-linear interactions within the hierarchical control loop. By modeling the transition between discrete mission states as a stochastic process, we effectively characterize how risk propagates from the initial environmental sensing layer to the high-level management layers.

A core methodological contribution that enhances the credibility of these findings is the mitigation of epistemic uncertainty through what we define as the empirical anchoring effect. A persistent challenge in autonomous ship reliability engineering is the heavy reliance on subjective expert elicitation, which can introduce cognitive bias into risk parameters. This study addresses addresses this limitation by integrating 12 million raw AIS trajectory records

and high-resolution ERA5 meteorological variables to calibrate the SPN transition firing rates (λ). This data-driven anchoring ensures that the model parameters are grounded in historical operational evidence and actual physical environmental constraints. By explicitly distinguishing between aleatory uncertainty, representing the inherent randomness in maritime environments, and epistemic uncertainty, representing knowledge gaps in autonomous control logic, the model provides a high-fidelity representation of risk evolution during mode transitions. Such calibration is essential for accurately identifying precursors to grounding or collision accidents in complex waterways.

Crucially, the findings refine the understanding of risk drivers during the critical task execution phase, offering a quantitative challenge to the prevailing decision-centric paradigm in contemporary MASS research. While much of the current literature prioritizes the optimization of autonomous decision-making algorithms as the ultimate safety solution, our sensitivity analysis identifies environmental perception and data transmission (λ_4) as the true systemic bottleneck. The simulation results indicate that enhancing the reliability and fidelity of the perception module yielded a 55.5% reduction in the steady-state risk probability, which significantly outweighs the 27.4% reduction achieved through the refinement of decision logic alone. This suggests that in the current stage of MASS technological development, information deficit characterized by sensor fusion errors and communication latency poses a more critical threat to system integrity than algorithmic processing errors [71]. This finding highlights a strategic priority for maritime safety management, suggesting that for a ship-shore control loop to remain resilient, the quality of the feedback perception is more vital than the sophistication of the autonomous command.

Furthermore, the simulation clarifies that risk performance is governed by a systemic lock-in effect, emphasizing a profound path dependency within the navigation mission. During the initialization phase, risk levels are primarily sensitive to the baseline configuration and environmental uncertainty. However, as the mission progresses into the execution phases, the safety state becomes intrinsically tied to the cumulative erosion of safety margins and preceding control configurations. This path dependency implies that the system possesses a memory of its safety state. Once the control loop enters a high-risk trajectory, such as a scenario where environmental stressors overwhelm perception limits, the probability of returning to a stable state decreases exponentially. This technical insight necessitates a paradigm shift from reactive hazard mitigation to proactive, resilience-based control interventions. It underscores the importance of maintaining high perception fidelity throughout the mission duration, as any early-stage degradation in sensing accuracy can lock the system into a failure path that autonomous algorithms may eventually be unable to reverse.

5.2 Advantages of the STPA-SPN hybrid framework

The proposed STPA-SPN framework offers a set of methodological capabilities that are not individually available in existing hybrid approaches and that collectively address the core analytical challenges of MASS navigation risk assessment. To substantiate this positioning, the following discussion provides a structured comparison against three representative methods evaluated across six analytical dimensions.

The most fundamental advantage lies in the structural grounding of the risk model. By commencing from the hierarchical control structure of the MASS perception-decision-control-execution loop, STPA ensures that every modeled state transition corresponds to a causally identified control flaw, specifically a UCAs or RIFs, rather than a statistically inferred correlation. This causal anchoring is absent from Dynamic Bayesian Network approaches [24, 25], which, despite their strong temporal modeling capability, derive risk dependencies from data co-occurrence and therefore cannot distinguish whether a statistical association reflects a genuine control-loop mechanism or a confounding operational pattern. Consequently, DBN-based models are inherently limited in their capacity to identify the structural origin of systemic vulnerabilities, a limitation that becomes particularly consequential in the novel and data-sparse operational context of MASS.

A second distinguishing capability is the native representation of concurrency and asynchrony. The SPN formalism models multiple RIFs as simultaneously active tokens whose interaction is governed by competing

transition firing rates, capturing the parallel triggering dynamics that characterize complex autonomous ship operations. Neither STPA-BN [36] nor STPA-HMM [39] provides this capability: the former maps STPA outputs onto a static directed acyclic graph that precludes concurrent state evolution, while the latter encodes temporal transitions through hidden state sequences that cannot represent the network topology of multi-path risk propagation. The absence of concurrency modeling in these approaches means they cannot reproduce the systemic lock-in effect identified in Section 4.3.3, wherein simultaneous activation of perceptual and decisional flaws drives the system toward a hazardous steady state that persists independently of any single initiating event.

A third advantage is the empirical anchoring of model parameters through multi-source data fusion. By calibrating SPN transition firing rates against 12 million AIS trajectory records and high-resolution ERA5 meteorological variables, the framework grounds its steady-state outputs in historical operational evidence rather than relying solely on subjective expert elicitation. This architecture explicitly distinguishes aleatory uncertainty, representing the inherent stochasticity of the maritime environment, from epistemic uncertainty arising from knowledge gaps in autonomous control logic, thereby producing risk probability estimates that are both physically interpretable and empirically credible. The robustness verification reported in Section 4.4.2, which demonstrates weight stability under $\pm 10\%$ and $\pm 20\%$ input perturbations, further confirms that the identified convergence hubs reflect genuine systemic vulnerabilities rather than artifacts of parameterization bias.

TTaken together, these three capabilities, namely causal structural grounding, native concurrency representation, and empirically anchored parameterization, endow the proposed STPA-SPN framework with the broadest coverage across the six analytical dimensions. To provide cross-study context, the steady-state risk probabilities obtained in this study (Path 1: risk performance = 0.35; Path 2: risk performance = 0.32; Path 3: risk performance = 0.30) are discussed in relation to the SPN-based results of Hu et al. [55], who report values in the range of 0.15 to 0.19 for Arctic Northeast Passage navigation. The higher values obtained here reflect genuine differences in operational hazard profiles: the Taicang section involves high-density mixed traffic under constrained navigational conditions, whereas the Arctic context is characterized by low traffic density with ice concentration as the dominant risk driver. Both studies confirm that SPN-based steady-state analysis produces stable, interpretable risk estimates under the CTMC isomorphism. While the present comparison is necessarily contextual rather than numerical, the maturation of MASS operational databases is anticipated to enable full quantitative benchmarking in future work, representing a natural extension of the validation agenda established here.

5.3 Practical implications

Building on the systemic discussions and the quantitative findings in Section 4, this study suggests several practical insights for the safety management of MASS operations. These findings provide a multi-dimensional support system from which various maritime stakeholders can derive significant safety and operational benefits.

(1) Operational monitoring and technical prioritization

For ROC and fleet managers, it is essential to establish real-time, hub-centric monitoring mechanisms because the intelligent integrated platform (P_8) and autonomous ship control system (P_9) serve as global hubs where over 70% of steady-state risk converges. By implementing dynamic state-monitoring at these specific convergence points, operators can achieve early detection of systemic risks before they escalate through the perception-decision-control-execution loop. Simultaneously, for ship designers and technical R&D teams, the sensitivity analysis reveals that optimizing environmental perception and data transmission capability (λ_4) offers a more significant leveraging effect on risk reduction of up to 55.5% compared to the 27.4% reduction achieved through the refinement of decision logic alone. This identifies a clear technical priority where engineering efforts should focus on the robustness of sensor fusion algorithms and communication stability to mitigate the non-linear risk escalation triggered by environmental stressors.

(2) Architectural resilience and regulatory governance

System architects should prioritize the implementation of functional resilience redundancy to dismantle the systemic lock-in effect identified in Section 4.3.3. By deploying proactive fail-safe architectures, including the multi-source perception cross-validation suggested by the risk performance defined in Eq (1), stakeholders can effectively disrupt the path-dependent propagation of control flaws. Simultaneously, maritime authorities and regulatory bodies benefit from the rigorous quantitative basis provided by this framework to transition toward a "resilience-adaptation" design philosophy. The identification of specific parameter ranges in Section 4.4.1 offers a scientific foundation for evidence-based policy making and the establishment of safety acceptance principles in high-density waterways like the Taicang section.

5.4 Limitations

While the proposed STPA-SPN framework demonstrates promising capabilities for dynamic risk assessment in MASS navigation, several limitations warrant acknowledgement to support a balanced interpretation of the present findings.

First, the framework relies on the isomorphism between the SPN and a homogeneous CTMC, which assumes that transition firing times follow exponential distributions. While this assumption guarantees the mathematical tractability of steady-state analysis, it may not fully capture failure processes characterized by heavy-tailed or non-exponential distributions. Future work may consider the adoption of non-Markovian SPN formulations to relax this constraint and better represent rare but consequential failure events.

Second, despite the multi-source data fusion architecture employed to reduce epistemic uncertainty, a subset of latent functional interaction parameters, particularly those governing decision-making biases and human-system dysfunctions within the MASS control loop, remains partially dependent on expert elicitation for calibration. This reflects a broader challenge in the field: at the current stage of MASS development, fully objective quantification of human- and system-related risk factors remains difficult to achieve. As MASS operational databases continue to mature, transitioning toward objective weight generation from historical near-miss records would represent a valuable direction for enhancing parameterization objectivity.

Third, the present validation of the framework remains primarily internal, relying on convergence analysis of the steady-state distributions and sensitivity checks of the AHP-derived weights, rather than external validation against observed traffic safety outcomes. A rigorous external validation would require comparing the steady-state risk rankings against long-run statistics of near-miss or conflict events derived from AIS data over an extended observation period. Such a longitudinal dataset is not yet available for the Taicang Port case study, and constructing one represents an important direction for future work.

Finally, the methodological design of the proposed framework reflects the current stage of MASS technological development, integrating the analytical strengths and operational realities available at this juncture. As autonomous shipping technology matures and new operational data accumulate, progressive refinement of the model structure, data sources, and assessment methods is encouraged to maintain the framework's relevance and analytical fidelity across diverse waterway environments.

6. Conclusions

This study proposes a "control-centric" hybrid framework that bridges qualitative hazard identification with quantitative dynamic simulation for MASS navigation. By mapping systemic control flaws—identified via STPA-onto an isomorphic SPN, the model effectively captures the functional interdependencies and dynamic evolution within the perception-decision-control-execution closed-loop.

The empirical application to the Taicang section of the Yangtze River provides several critical quantitative findings and practical insights:

(1) Simulation results reveal that MASS navigation risks exhibit significant path dependency and systemic lock-in effects rather than random fluctuations. The collision risk path demonstrates the most rapid escalation, reaching a steady-state risk probability of 0.35 within 120 minutes, confirming that risk evolution is governed by the cumulative reinforcement of control-loop flaws rather than isolated failure events.

(2) Across all three risk paths, the Intelligent Integration Platform and the Autonomous Ship Control System consistently exhibit the largest combined nodal contribution, collectively accounting for over 70% of steady-state risk in the collision-risk path and remaining the most influential risk accumulation nodes across the remaining paths, identifying these two components as the primary convergence hubs of systemic vulnerability.

(3) Sensitivity analysis indicates that optimizing the environmental perception rate can reduce the global steady-state risk probability by up to 55.5%, substantially exceeding the 27.4% reduction achieved through refinement of the decision-making module alone. This result underscores the strategic importance of functional resilience redundancy, including multi-source perception cross-validation, as a means of disrupting the risk aggregation chain at convergence nodes and preventing systemic lock-in under degraded perceptual conditions.

Still, several aspects of the proposed STPA-SPN framework could be improved and further studied. First, regarding the data requirements, the methodology relies heavily on comprehensive, high-fidelity inputs to ensure analytical accuracy. Although this study integrates multi-source information, characterizing latent human cognitive biases and extreme unforeseen events remains challenging, potentially leading to identification blind spots in complex sociotechnical systems. Incorporating real-time data fusion mechanisms and deep learning architectures would be beneficial for enhancing model responsiveness and improving the understanding of risk evolution mechanisms in highly dynamic navigation environments. Second, while the model currently utilizes periodically updated AIS and meteorological datasets, it is advisable to expand the scope of high-frequency sensor integration to mitigate temporal lags in responding to emergent hazards. Finally, although expert elicitation was cross-validated through consistency ratios, the subjective nature of the weighting process remains a source of epistemic uncertainty. Transitioning toward objective weight generation from historical near-miss data and the application of fuzzy-AHP would be a valuable direction for future development to further enhance the objectivity and robustness of the analytical results.

Appendix

Table A1. Summary of common abbreviations and their full names.

MASS	Maritime Autonomous Surface Ships
SPN	Stochastic Petri Nets
STPA	System-Theoretic Process Analysis
STAMP	Systems Theoretic Accident Model and Processes
RIFs	Risk-Influencing Factors
UCAs	Unsafe Control Actions
DBN	Dynamic Bayesian Networks
CTMC	Continuous-Time Markov Chain
SACS	Ship Autonomous Control System
ROC	Remote Operation Center
HMM	Hidden Markov Models
FSA	Formal Safety Assessment
PRA	Probabilistic Risk Assessment
HAZOP	Hazard and Operability Analysis
IMO	International Maritime Organization
FMEA	Failure Mode and Effects Analysis

Table A2 Demographic details of the expert panel

Expert ID	Core Background	Current Title	Experience (Years)	Specific Relevance to Study
Exp. 1	Maritime management	Senior VTS supervisor	22	VTS management in Yangtze River
Exp. 2	Maritime management	MSA Safety official	18	Maritime traffic safety regulation
Exp. 3	Technical R&D	Senior MASS engineer	12	Intelligent navigation system design
Exp. 4	Technical R&D	System Architect	15	Autonomous perception algorithms
Exp. 5	Technical R&D	Technical director	14	Smart ship communication systems
Exp. 6	Navigation practice	Senior captain (unlimited)	25	Traditional & Mixed traffic navigation
Exp. 7	Navigation practice	Senior pilot	20	Pilotage in Taicang/Nantong waterway
Exp. 8	Navigation practice	Safety superintendent	16	Fleet operational risk management
Exp. 9	Academic research	Professor	21	Maritime traffic engineering & Risk
Exp. 10	Academic research	Assoc. Professor	13	Intelligent ship control models

Table A3. Symbols and definitions used in the proposed methodology.

Symbol	Description
$RP(e)$	Dynamic risk performance of event e
$C_1(e)$	Steady-state probability of accident markings
$C_2(e)$	Contribution of primary RIFs based on firing rates
$C_3(e)$	Functional impact based on token residence time
P	Finite set of places representing system states
T	Finite set of transitions representing UCAs or risk event
F	Set of directed arcs defining network topology
W	Arc weight function
M_1	Initial marking (initial state distribution)
λ_i	Average firing rate of transition T_i
STD_j	Standard deviation of expert scores for factor j
S_{ij}	Individual score assigned by expert i to factor j
E_s	Expert consensus scores for parameterization
f_b	Base firing function
f_e	Environmental modifier function
V_e	Objective environmental variables
M_i	Marking in SPN uniquely mapped to a Markov state
S_i	Corresponding state in the CTMC
q_{ij}	Transition rate from state S_i to S_j
q_i	Total outflow rate from state S_i
P_1	Specific place representing navigation risk

Acknowledgements

This work is supported by the European Research Council project (TRUST CoG 2019 864724).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Chun Zou: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lijia Chen**: Writing – review & editing, Validation, Supervision, Investigation. **Hanwen Fan**: Writing – review & editing, Methodology, Validation, Supervision, Investigation. **Xiaobin Tian**: Formal analysis, Validation. **Li Liu**: Data curation, Supervision. **Zaili Yang**: Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

References

[1] Zou YQ, Xiao GN, Li QJ, Biancardo SA. Intelligent Maritime Shipping: A Bibliometric Analysis of Internet Technologies and Automated Port Infrastructure Applications. Journal of Marine Science and Engineering. 2025;13.

[2] Bolbot V, Bergström M, Rahikainen M, Banda OAV. Investigation into safety acceptance principles for autonomous ships. Reliability Engineering & System Safety. 2025;257.

[3] IMO. IMO Takes First Steps to Address Autonomous Ships. Available online. 2018.

[4] IMO. IMO and Autonomous Shipping. Available online. 2024.

[5] IMO. Maritime Safety Committee—109th Session (MSC 109). Available online. 2024.

[6] Fan SQ, Yang ZL. Towards objective human performance measurement for maritime safety: A new psychophysiological data-driven machine learning method. Reliability Engineering & System Safety. 2023;233.

[7] Fan SQ, Shi K, Weng JX, Yang ZL. Letting losses be lessons: Human-machine cooperation in maritime transport. Reliability Engineering & System Safety. 2025;253.

[8] Yang X, Zhou T, Utne IB, Zhang WJ. A framework for quantification of coupling risk in the transition between operational modes of MASS. Reliability Engineering & System Safety. 2025;264.

[9] Qiao RT, Fan SQ. Human-machine interaction for maritime transport using a data-driven approach. Reliability Engineering & System Safety. 2026;267.

[10] Yang KS, Yang D, Lu YX. Enhancing risk perception by integrating ship interactions in multi-ship encounters: A Graph-based Learning method. Reliability Engineering & System Safety. 2025;261.

[11] Fan CL, Montewka J, Zhang D. A risk comparison framework for autonomous ships navigation. Reliability Engineering & System Safety. 2022;226.

[12] Fan HW, Jia HY, Wu ZM. Natural disaster risks in seaports for global petroleum trades. Transportation Research Part E-Logistics and Transportation Review. 2026;210.

[13] Bolbot V, Theotokatos G, Bujorianu LM, Boulougouris E, Vassalos D. Vulnerabilities and safety assurance methods in Cyber-Physical Systems: A comprehensive review. Reliability Engineering & System Safety. 2019;182:179-93.

[14] Wang K, Xiong L, Xue RD. Real-time data stream learning for emergency decision-making under uncertainty. Physica a-Statistical Mechanics and Its Applications. 2024;633.

[15] Sun ML, Zheng ZY, Gang LH. Uncertainty Analysis of the Estimated Risk in Formal Safety Assessment. Sustainability. 2018;10.

[16] Chang CH, Kontovas C, Yu Q, Yang ZL. Risk assessment of the operations of maritime autonomous surface ships. Reliability Engineering & System Safety. 2021;207.

[17] Fan CL, Montewka J, Zhang D. Towards a Framework of Operational-Risk Assessment for a Maritime Autonomous Surface Ship. Energies. 2021;14.

[18] Chou CC, Wang CN, Hsu HP. A novel quantitative and qualitative model for forecasting the navigational risks of Maritime Autonomous Surface Ships. Ocean Engineering. 2022;248.

[19] Kurt YB, Akyuz E, Arslan O. A Quantitative HAZOP Risk Analysis Under Extended CREAM Approach for Maritime Autonomous Surface Ship (MASS) Operation. Marine Technology Society Journal. 2022;56:59-73.

[20] Hu SP, Zou C, Wu JJ, Zhang WF, Wang ZC, Fei JG. Dynamic resilience modeling of maritime traffic systems: A hybrid HMM-DBN approach for analyzing LNG-Fueled vessel interactions. Ocean Engineering. 2025;330.

[21] Fan HW, Wang JX, Chang Z, Lyu J, Jia HY. Embracing imperfect data: A novel data-driven Bayesian network framework for maritime accidents severity risk assessment. Ocean Engineering. 2025;329.

[22] Wang JX, Fan HW, Chang Z, Lyu J. Unleashing data power: Driving maritime risk analysis with Bayesian networks. Reliability Engineering & System Safety. 2025;264.

[23] Wang JX, Fan HW, Wu ZM, Wang NX, Chang Z, Lyu J. Navigating straits/canals safety: a data-driven Copula Bayesian network risk assessment framework. Reliability Engineering & System Safety. 2026;270.
[24] Luo XF, Ling H, Xing MX, Bai X. A dynamic-static combination risk analysis framework for berthing/ unberthing operations of maritime autonomous surface ships considering temporal correlation. Reliability Engineering & System Safety. 2024;245.
[25] Luo XF, Guo LH, Bai X, Li YS, Zan YF, Luo JX. A multi-phase mission success evaluation approach for maritime autonomous surface ships considering equipment performance degradation and system composition changes. Reliability Engineering & System Safety. 2025;254.
[26] Nakashima T, Kureta R, Khastgir S. Addressing systemic risks in autonomous maritime navigation: A structured STPA and ODD-based methodology. Reliability Engineering & System Safety. 2025;261.
[27] Fan CL, Montewka J, Bolbot V, Zhang Y, Qiu YH, Hu SP. Towards an analysis framework for operational risk coupling mode: A case from MASS navigating in restricted waters. Reliability Engineering & System Safety. 2024;248.
[28] Yuzui T, Kaneko F. Toward a hybrid approach for the risk analysis of maritime autonomous surface ships: a systematic review. Journal of Marine Science and Technology. 2025;30:153-76.
[29] Leveson N. A new accident model for engineering safer systems. Safety Science. 2004;42:237-70.
[30] Ventikos NP, Chmurski A, Louzis K. A systems-based application for autonomous vessels safety: Hazard identification as a function of increasing autonomy levels. Safety Science. 2020;131.
[31] Yang X, Zhou T, Zhou XY, Zhang WJ, Mu CR, Xu S. A framework to identify failure scenarios in the control mode transition process for autonomous ships with dynamic autonomy. Ocean & Coastal Management. 2024;249.
[32] Cheng TT, Utne IB, Wu B, Wu Q. A novel system-theoretic approach for human-system collaboration safety: Case studies on two degrees of autonomy for autonomous ships. Reliability Engineering & System Safety. 2023;237.
[33] Thieme CA, Utne IB, Haugen S. Assessing ship risk model applicability to Marine Autonomous Surface Ships. Ocean Engineering. 2018;165:140-54.
[34] Zhou XY, Liu ZJ, Wang FW, Wu ZL, Cui RD. Towards applicability evaluation of hazard analysis methods for autonomous ships. Ocean Engineering. 2020;214.
[35] Correa-Jullian C, Ramos M, Mosleh A, Ma JQ. Operational safety hazard identification methodology for automated driving systems fleets. Proceedings of the Institution of Mechanical Engineers Part O-Journal of Risk and Reliability. 2025;239:310-43.
[36] Qiao WL, Huang EZ, Guo HTY, Lian CP, Chen HQ, Ma XX. On the causation analysis for hazards involved in the engine room fire-fighting system by integrating STPA and BN. Ocean Engineering. 2023;288.
[37] Zhu XM, Hu SP, Li Z, Wu JJ, Yang X, Fu SS, et al. Risk performance analysis approach for convoy operations via a hybrid model of STPA and DBN: A case from ice-covered waters. Ocean Engineering. 2024;302.
[38] Li W, Chen WJ, Hu SP, Xi YT, Guo YL. Risk evolution model of marine traffic via STPA method and MC simulation: A case of MASS along coastal setting. Ocean Engineering. 2023;281.
[39] Li W, Chen WJ, Guo YL, Hu SP, Xi YT, Wu JJ. Risk Performance Analysis on Navigation of MASS via a Hybrid Framework of STPA and HMM: Evidence from the Human-Machine Co-Driving Mode. Journal of Marine Science and Engineering. 2024;12.
[40] Hu SP, Li WJ, Xi YT, Li W, Hou ZQ, Wu JJ, et al. Evolution pathway of process risk of marine traffic with the STAMP model and a genetic algorithm: A simulation of LNG-fueled vessel in-and-out harbor. Ocean Engineering. 2022;253.
[41] Ahn SI, Kurt RE, Turan O. The hybrid method combined STPA and SLIM to assess the reliability of the human interaction system to the emergency shutdown system of LNG ship-to-ship bunkering. Ocean Engineering. 2022;265.
[42] Jiao J, Jing YF, Pang SJ. An Integrated Quantitative Safety Assessment Framework Based on the STPA and System Dynamics. Systems. 2022;10.

[43] Ghazel M. Using Stochastic Petri Nets for Level-Crossing Collision Risk Assessment. Ieee Transactions on Intelligent Transportation Systems. 2009;10:668-77.
[44] Kamil MZ, Taleb-Berrouane M, Khan F, Ahmed S. Dynamic domino effect risk assessment using Petri-nets. Process Safety and Environmental Protection. 2019;124:308-16.
[45] Liu ZY, Jia LC, Dong SH. Refined Oil Loading and Unloading Process Risk Assessment using Stochastic Colored Petri Nets Integrated with Risk Factors. Tehnicki Vjesnik-Technical Gazette. 2024;31:70-8.
[46] Kabir S, Papadopoulos Y. Applications of Bayesian networks and Petri nets in safety, reliability, and risk assessments: A review. Safety Science. 2019;115:154-75.
[47] Taleb-Berrouane M, Khan F, Amyotte P. Bayesian Stochastic Petri Nets (BSPN) - A new modelling tool for dynamic safety and reliability analysis. Reliability Engineering & System Safety. 2020;193.
[48] Bensaci C, Zennir Y, Pomorski D, Innal F, Lundteigen MA. Collision hazard modeling and analysis in a multi-mobile robots system transportation task with STPA and SPN. Reliability Engineering & System Safety. 2023;234.
[49] Brito C, Araújo G, Fé I, Rego P, Callou G, Borges I, et al. Sensitivity-centered preventive maintenance strategies for enhanced urban surveillance: a stochastic petri net analysis. Computing. 2025;107.
[50] Farias D, Nogueira B, Farias IF, Jr., Andrade E. A modeling-based approach for dependability analysis of a constellation of satellites. Software and Systems Modeling. 2025;24:209-24.
[51] Bezerra TV, Callou G, Airton F, Tavares E. A Modeling-Based Approach for Performance and Availability Assessment of IoMT Systems. Electronics. 2025;14.
[52] Hui MH, Ni F, Liu WC, Liu J, Chen NN, Zhou XJ. SPN-Based Dynamic Risk Modeling of Fire Incidents in a Smart City. Applied Sciences-Basel. 2025;15.
[53] Nyvlt O, Haugen S, Ferkl L. Complex accident scenarios modelled and analysed by Stochastic Petri Nets. Reliability Engineering & System Safety. 2015;142:539-55.
[54] Zhou CH, Qin SX, Zhong JH, Du L, Zhang F. Modeling and analysis of external emergency response to ship fire using HTCPN and Markov chain. Ocean Engineering. 2024;297.
[55] Hu SP, Fang CW, Wu JJ, Fan CL, Zhang XX, Yang X, et al. Enhanced risk assessment framework for complex maritime traffic systems via data driven: A case study of ship navigation in Arctic. Reliability Engineering & System Safety. 2025;260.
[56] Leveson NG. Applying systems thinking to analyze and learn from events. Safety Science. 2011;49:55-64.
[57] Fu SS, Zhang Y, Zhang MY, Han B, Wu ZD. An object-oriented Bayesian network model for the quantitative risk assessment of navigational accidents in ice-covered Arctic waters. Reliability Engineering & System Safety. 2023;238.
[58] Guo YL, Hu SP, Jin YX, Xi YT, Li W. A Hybrid Probabilistic Risk Analytical Approach to Ship Pilotage Risk Resonance with FRAM. Journal of Marine Science and Engineering. 2023;11.
[59] Zou C, Hu SP, Chen LJ. Environmental data driven dynamic Bayesian network: Spatiotemporal evolution of coastal shipping risk performance in China. Reliability Engineering & System Safety. 2026;271.
[60] Leveson N. A systems approach to risk management through leading safety indicators. Reliability Engineering & System Safety. 2015;136:17-34.
[61] Islam R, Guo YT, Adumene S, Abdussamie N. Reliability Assessment of Marine Propulsion Systems for MASS: A Bibliometric Analysis and Literature Review. Journal of Marine Science and Engineering. 2025;13.
[62] Jugovic A, Sirotic M, Oblak R, Schiozzi D. Integration of Maritime Autonomous Surface Ships into Coastal Waters Supply Chains: A Systematic Literature Review of Safety and Autonomy Challenges. Journal of Marine Science and Engineering. 2025;13.
[63] Na S, Lee D, Baek J, Kim S, Choung C. Qualitative Risk Assessment Methodology for Maritime Autonomous Surface Ships: Cognitive Model-Based Functional Analysis and Hazard Identification. Journal of Marine Science and Engineering. 2025;13.

[64] Park H, Kim J. STPA analysis for safe operation of maritime autonomous surface ship under degradation state. Frontiers in Marine Science. 2025;12.
[65] Fan CL, Qiu YH, Luo LH, Wang XM, Wang B, Yang Z, et al. Navigational risk criteria design for MASS operations using survey data in China. Ocean Engineering. 2026;347.
[66] Zhang ZW, Wang XJ, Feng YW, Xin XR, Liu ZJ, Yang ZL. Safety analysis of human-machine interaction of autonomous ships using system theory and complex networks. Ocean Engineering. 2026;351.
[67] Liu ST, Sun QY. Modeling and performance analysis of emergency intelligence service process based on stochastic petri nets: A case study of the "7.20" zhengzhou rainstorm in China. Reliability Engineering & System Safety. 2026;269.
[68] Chen JH, Chen ZC, Xu Y, Li H. Efficient reliability analysis combining kriging and subset simulation with two-stage convergence criterion. Reliability Engineering & System Safety. 2021;214.
[69] Agrell C, Dahl KR. Sequential Bayesian optimal experimental design for structural reliability analysis. Statistics and Computing. 2021;31.
[70] Wróbel K, Montewka J, Kujala P. Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. Reliability Engineering & System Safety. 2017;165:155-69.
[71] Fan CL, Wróbel K, Montewka J, Gil M, Wan CP, Zhang D. A framework to identify factors influencing navigational risk for Maritime Autonomous Surface Ships. Ocean Engineering. 2020;202.