ISSN: 2754-4915

Journal of Material Sciences & Manufacturing Research

Research Article

Open Access

Improving Expert Judgment in Scenario Simulation for Container Terminals and Supply Chain

Farshad Shamlu[1,3*], Marco Gotelli[1,3] and Emilio Jimenez[2]

[1]Simulation Team, Via Magliotto 2, 17100, Savona, Italy

[2]Electrical Engineering Department, University of La Rioja, Logrono, 26004 La Rioja, Spain

[3]Department of Mechanical, Energy, Management and Transport Engineering, University of Genoa, Italy

ABSTRACT

This study explores the use of Modeling and Simulation (M&S) and Structured Expert Judgment (SEJ) to handle complex problems in the logistics sector, with an emphasis on container terminals. This paper focuses on expert judgment within the Generator of Logistics Flow (GOLF), a model that uses SEJ to produce synthetic logistics data sets when empirical data are incomplete, uninformative, or unavailable. Combining SEJ with a stochastic generator enables quantification of unknown parameters through uncertainty adjustment and propagating that uncertainty into scenario outputs. Partial validation shows high precision for rail container mode approximately 3%, while larger errors for transshipped containers reveal sensitivity in this sub-model. Furthermore, GOLF highlights the role of strategic engineering, which integrates modeling, simulation, and data analytics in a closed loop to provide decision makers for policy implementation and strategy design. It allows decision makers to perform "What if?" analyzes to obtain a variety of useful data in different scenarios and demonstrate the potential of intelligent systems to address data scarcity, enhance operational efficiency, and foster innovation in supply chain and container terminal management.

***Corresponding author**
Farshad Shamlu, Simulation Team, Via Magliotto 2, 17100, Savona, Italy.

Received: February 16, 2026; **Accepted:** February 20, 2026; **Published:** March 02, 2026

Keywords: Modeling and Simulation, Structured Expert Judgment, Strategic Engineering, Supply Chain Management, Logistics

Introduction

Global trade is heavily dependent on maritime and port logistics. More than 80% of world trade by volume (and approximately 70% by value) is carried by the sea and handled through ports [1]. In 2023, global maritime trade grew by 2.4% to reach 12.3 billion tons, with further growth expected through 2029 [2]. As ports handle most of global trade, their strategic importance in international commerce is undeniable. Ensuring efficient port terminal operations and a deep understanding of logistical flows are essential to minimizing delays and maintain supply chain reliability [3]. The standardization of goods has enabled containers to move seamlessly across rail, road, or sea modes. While this improves connectivity and increases port throughput, it could also lead to congestion, reduce port competitiveness, and negatively affect local economies [4].

Despite living in a data-rich era, the assumption that sufficient data exist to evaluate all future events, risks, and opportunities is a misconception [5]. In practice, data-related challenges persist: shipping companies often hesitate to share container flow data, and the available datasets are frequently incomplete or uninformative. These limitations make the analysis and optimization difficult. Furthermore, existing data and modeling tools often cannot provide decision makers with all the information needed to design and implement effective policies or to make optimal management choices. For this reason, decision makers frequently supplement other sources of information with expert judgment [6]. In such cases, expert judgment is essential for filling the knowledge gaps. Using expert judgment allows decision makers to quantify uncertainty related to unknown or poorly understood parameters. This can involve a wide range of approaches, from asking a single expert for a best estimate, to informally consulting colleagues, or to applying a formal, documented procedure for obtaining and combining probabilistic judgments.

Experts have contributed by guiding model selection, interpreting analysis results, and estimating uncertain parameters when reliable empirical data are lacking. However, relying on expert input raises concerns regarding potential bias. Expert inputs are subjective and could be biased by overconfidence and other cognitive biases [7]. To address this, structured elicitation methods, such as the Investigate, Discuss, Estimate, Aggregate (IDEA) protocol, have been developed. For expert judgments to be recognized as scientific data, they must conform to the principles of the scientific process, namely accountability, neutrality, fairness, and empirical control. Among these principles, empirical control can be ensured through validation. Originating from the Delft University SEJ approach, these methods help to systematically gather and aggregate expert opinions, improving the reliability of inputs. Researchers use SEJ techniques to obtain most expert knowledge. A popular method is the Classical Model, also known as the Delft method [8].

In this case, Strategic Engineering approach is applied to develop a stochastic simulator to generate the container flow called GOLF that leverages M&S, Data Analytics and AI for statistical data analysis and collection regarding container flows [9]. Also, by including expert-derived input, simulations could explore a broader range of scenarios than those suggested by limited historical data. For instance, if no data exists on how a container terminal (CT) might handle an expansion of volume, experts could provide assumptions about potential bottlenecks or resource needs at that scale, and the simulation could then be used to simulate the scenario. This study used SEJ to set up its simulation inputs. It is recognized that sufficient data will never be available to simulate and analyze all future events; therefore, expert judgments are relied upon [8].

Scope of Simulation

To address complexity and build a robust system capable of withstanding political and natural challenges, supply chains should use modeling, simulation, and digital technologies [10]. A new discipline called strategic engineering has been developed to create and implement these solutions within a structured framework. Strategic engineering integrates modeling and simulation with artificial intelligence, data analytics, and intelligent agents in a closed loop to support decision-makers [11]. In this context, there is a clear opportunity to take advantage of M&S as part of the smart CT tool. Simulation provides a safe and cost-effective way to conduct what-if analyses of CT logistics flow without interfering with real operations, and CT managers and policymakers could anticipate the outcomes of changes in infrastructure, policies, and external factors that could cause congestion and impact local and global sustainability.

GOLF is designed to address these challenges [9]. GOLF is a modeling and simulation framework designed to generate synthetic container flow data and simulate logistic processes based on a CT environment. In essence, it creates a stochastic model of CT logistics flow, where uncertain parameters are filled using probability distributions informed by experts and any available data. Notably, the synthetic data generated by GOLF in these scenarios could also be fed into other models and simulations. In model development, the Generator Model calculates the primary and secondary modes of transportation for each container, specifying whether a container is transported by truck, train, oceanic ship, or feeder ship for each scenario. The simulator vectorized the different transportation modes to compute the number of transport modes for each scenario, segmenting the journey into primary and secondary modes of transportation. So far, the container has been transported by the four main ports of Genoa, Savona, La Spezia, and Livorno [9].

In addition, GOLF operates at a strategic level, generating logistics flows through the entire system (origin→container terminal→hinterland/destination). GOLF could evaluate scenarios of trade growth, mode shift, or network changes directly and distinguishes itself by formally integrating expert judgments to set up its input distributions. Table 1 shows GOLF capabilities and limitations to clarify the positioning of the model and to distinguish it from potential misconceptions.

The remainder of this paper is organized as follows. Section 2 reviews related work on logistics simulation, structured expert judgment, and digital-twin concepts. Section 3 describes the GOLF model, SEJ methodology, and the expert elicitation and aggregation procedures used to derive model inputs. Section 4 analyses model validation and scenario results including the train and transshipment flow comparison. Section 5 discusses limitations and recommendations for future work.

Table 1: GOLF Capabilities and Limitations

Aspect	GOLF Is	GOLF Is Not
Definition	A modular stochastic simulation framework for logistics flows in container terminals.	A detailed micro-operation port simulator.
Scope	Strategic-level scenario analysis and decision support under uncertainty.	Real-time operational control or scheduling tool.
Data Handling	Generates synthetic data using SEJ and probability distributions for uncertain parameters.	A system fully replacing empirical data needs.
Integration	Can feed traffic, emissions, and supply chain analyses with synthetic datasets.	A "black box" tool; requires expert configuration.
Alignment	Aligns with the Digital Twin concept for high-level analysis.	Does not simulate individual crane or gate operations.
Validation	Requires calibration and validation with available empirical data.	A plug-and-play simulator without local tuning.

State of the Art

Simulation has long been used to study and optimize port operations. Discrete-Event Simulation (DES) could capture complex interactions and queuing processes in ports such as ship arrivals, berthing, quay cranes, yard storage, and gate processing. Several researchers have developed models and simulations for specific port subsystems. Pereira et al. built a DES model using AnyLogic to predict the behavior of a new port gate facility, allowing planners to foresee bottlenecks and test various gate operation scenarios before implementation [12]. Benghalia et al. used a tool called FlexTerm to model container flows inside the Port of Algiers terminal with high fidelity. The simulation was configured with actual operating parameters and served as a baseline to assess the current operational performance [13]. These micro-level models, once calibrated with empirical data, allow port operators to experiment with changes in a risk-free digital environment and identify potential improvements in throughput and reductions in congestion. A common thread in port simulation literature emphasizes the use of models for scenario analysis and decisions support. In particular, discrete-event simulation is praised for its ability to safely evaluate "what-if" scenarios. Planners could test the impact of infrastructure expansion, new operational policies, or external shocks (such as surges in traffic or labor disruptions) without affecting real-world operations [12]. For example, a simulation study might reveal how adding an extra quay crane could reduce vessel turnaround time, or how implementing truck appointment systems at the gate might mitigate road congestion. The use of the simulation fits well with the digital twin concept. Here, the model is constantly updated and used alongside real operations to improve them. Simulation assists in making investment and planning decisions. This aligns with the digital twin concept, in which a model is constantly updated and used in real operations for better results. Recent studies have described port simulation models as digital twins [14].

One of the earliest efforts to standardize the use of expert judgment was carried out by the U.S. Nuclear Regulatory Commission (NRC), which formally documented the elicitation process and subjected it to scientific review. This work revealed large differences in expert opinions, raising important questions about how expert judgments should be validated and combined [15]. In response, the nuclear safety community played a leading role in developing expert elicitation methods that address these challenges. Over time, these techniques have been adopted in many other fields, including recent applications to the assessment of greenhouse-gas (GHG) emissions [16,17]. Researchers have applied the classical model to examine the impact of existing invasive species on the economic value of ecosystem services [18]. It has also been used to assess the potential future ecological impact of an Asian Carp invasion in the Great Lakes, as well as to evaluate the effectiveness of various strategies aimed at preventing the establishment of Asian carp in the Great Lakes [19]. One study examined the sustainable supply chain in the Mexican Maquiladora industry. The study proposed ten hypotheses relating the variables that are validated using the partial least squares technique, and information were collected from 187 managers and engineers via a questionnaire working in the Mexican maquiladora industry. The research objective was to determine the relationships among variables to increase performance based on sustainable activities [20].

Expert judgment is not universally applicable to all quantitative questions. It is unnecessary for quantities that are directly observable-such as the speed of light in a vacuum—and unsuitable in domains lacking relevant scientific expertise and empirical measurements, such as inquiries about the behavior of a god [21]. They suggest that expert judgment is most appropriate for targets that are measurable in principle but not feasible to measure directly in practice, like the toxicity of a new substance in humans. Expert judgment is also not needed when adequate historical data are available and there is agreement on methods for converting such data into reliable predictions [22]. Additionally, when outcomes are strongly influenced by human behavior, dependable predictive expertise may be absent [6]. The process of eliciting expert judgments demands significant time and effort from both analysts and experts. Quantifying uncertainty may not be justified if it exerts minimal influence on the final decision or outcome [22]. Consequently, analysts should first determine the principal uncertainties within a problem domain before initiating an expert judgment study.

Table 2 demonstrates a comparison with existing simulation tools, to position GOLF in relation to leading commercial tools, reinforcing the academic and strategic character of the work.

Table 2: Comparison with Existing Simulation Tools

Feature	GOLF	AnyLogic	FlexSim	Simio
Primary Focus	Strategic container terminal and logistics flow simulation using SEJ	Multi-method general-purpose simulation	Discrete-event simulation for logistics and manufacturing	Object-oriented discrete-event simulation
Granularity	Strategic, high-level flows	Any (from high-level to detailed processes)	Detailed operational and process level	Detailed operational and process level
SEJ Integration	Native structured expert judgment integration	Requires manual SEJ integration	Requires manual SEJ integration	Requires manual SEJ integration
Data Generation	Synthetic data for uncertain parameters	Uses real/simulated data inputs	Uses real/simulated data inputs	Uses real/simulated data inputs
Ease of Use	Requires domain expertise and expert elicitation, User-friendly with GUI	User-friendly with GUI	User-friendly with drag-and-drop	User-friendly with process logic
Target Users	Researchers, planners, policy designers	Researchers, industrial engineers, educators	Industrial engineers, planners	Industrial engineers, planners
Integration with Digital Twin	High-level alignment for scenario testing	Full capability depending on model	Full capability depending on model	Full capability depending on model
Typical Application	Port expansions, logistics policy evaluation	Manufacturing, logistics, healthcare, transport	Logistics, manufacturing process improvement	Manufacturing, logistics
Validation Requirement	Essential before use	Required for realistic outputs	Required for realistic outputs	Required for realistic outputs

Methodology

GOLF is a stochastic simulation framework to model container logistics flows through CTs and their hinterland connections in northwest Italy and could be developed to model container terminals worldwide. The GOLF structure is modular, and configuration driven. Each CT or case study is represented by a configuration file or set of input parameters that capture the key characteristics of that context. These include statistics such as the annual container throughput, proportion of containers handled by different transport modes, typical voyage durations, handling rates, and any known operational constraints. Because every CT has unique traits (vessel traffic patterns, rail connectivity, etc.), GOLF's modular architecture allows it to be scaled and adapted by simply changing the configuration inputs for a new site [9].

The study primarily focuses on integrating SEJ into GOLF. Many of the parameters that drive the GOLF simulation algorithm could be filled in by elicited expert judgment, particularly in situations where historical data are absent. For instance, when data on ship inter-arrival times for a newly launched service are unavailable, experts could estimate the average time between arrivals and the extent to which it varies. Likewise, the share of containers moved by truck or rail could be determined based on current trends and expected changes in transport choices. In practice, the developers followed a process of incorporating expert estimates for uncertain inputs and then refining those inputs as any real data became available.

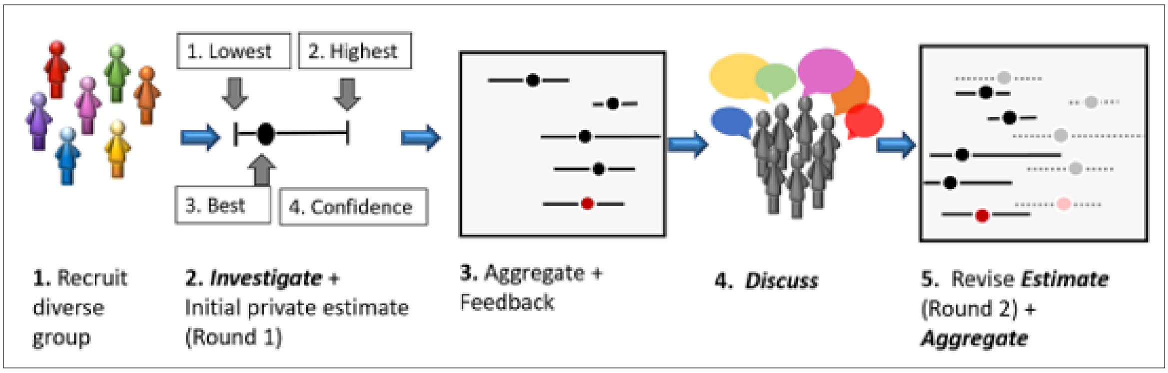

Figure 1: Key Steps of the IDEA Protocol Described in [23].

The classical model provides a structured framework for eliciting and mathematically aggregating expert judgments, with validation embedded as a fundamental component. Within this approach, experts quantify their uncertainty in response to two categories of questions: target questions and calibration questions. The primary focus lies on the target questions, which represent the variables of interest. These are issues that cannot be reliably addressed using alternative methods and therefore require expert judgment [24].

Six experts were asked to help with the modeling process. The elicitation process involved six experts, consistent with common rule-of-thumb guidance suggesting that six participants represent a practical minimum for group of experts. This threshold is often considered sufficient to capture a reasonable diversity of knowledge and perspectives while maintaining feasibility in terms of time and resources. Nevertheless, it is widely recognized that larger group of experts can enhance robustness and consistency by incorporating broader expertise, reducing individual biases, and improving the stability of aggregated results. Before applying their opinions in the simulation, each expert's performance was calculated using two key metrics: the calibration score and the information score. These metrics are commonly used in SEJ methods to quantify the accuracy and informativeness of expert assessments. Also, expert elicitation was conducted before performance scoring to minimize cognitive biases among experts, the IDEA protocol was applied to foster independent reasoning, structured discussion, and iterative refinement of estimates.

Figure. 1 shows the basic steps which could be summarized as follows: A varied group of experts was gathered to respond to the questions with probabilistic or quantitative answers. Initially, these experts were tasked with examining the questions to clarify their meanings and then offering their private, individual best guess point estimates along with corresponding credible intervals.

They then receive feedback on how their estimates compare to those of other experts. With the help of a facilitator, the experts are encouraged to discuss the outcomes, address different interpretations of the questions, cross-examine reasoning and evidence, and subsequently provide a second and final private estimate. Indeed, the goal of discussion in the IDEA protocol is not to achieve consensus but to resolve linguistic ambiguities, foster critical thinking, and share evidence. This approach is supported by evidence indicating that including a single discussion phase in a standard Delphi process enhances response accuracy [10]. Finally, the individual estimates are combined using mathematical aggregation [23].

Calibration Score

The calibration score, also known as statistical accuracy, measures how well the experts' uncertainty assessments correspond to actual realizations. It is calculated using a statistical test that quantifies the extent to which experts' assessments capture the true values at the expected frequency. The score is the p-value of this test, a standard output in classical statistical settings, and ranges between zero and one. For each quantity, each expert partitioned the uncertainty range into four inter-quantile intervals, for which they provided corresponding probability estimates, namely p1 = 0.05, less than or equal to the 5% value; p2 = 0.45, greater than the 5% value and less than or equal to the 50% value; p3 = 0.45, greater than the 50% value and less than or equal to the 95% value; and p4 = 0.05, greater than or equal to the 95% value. The classical model defines the statistical accuracy or calibration scores as follows:

$$\mathrm{Cal}(e)=1-F(2mI(s(e),p)) \qquad (1)$$

Here, F is asymptotically distributed as a chi-square variable with 3 degrees of freedom. This is the likelihood ratio statistic, and I(s(e), p) is the relative information of distribution s with respect to p also known as the Kullback-Leibler divergence, which measures the difference between the two distributions and is defined as:

$$\mathrm{I}(s_{(e)}, p) = \sum_{i=1}^{N} s_i \ln\left(\frac{s_i}{p_i}\right) \qquad (2)$$

where p is the theoretical probability vector that gives the expected proportion of realizations in each interval and s is the vector of observed proportions [21]. A higher calibration score indicates greater statistical accuracy in expert judgments.

Information Score

Intuitively, the information score measures the degree to which the distribution is concentrated. The information is then measured with respect to a background measure, which is typically chosen to be a uniform or log-uniform distribution. The information score could take, in principle, any positive value, but is typically lower than 3. The classical model implements the so-called k overshoot rule. For each item, the smallest interval I = [L, U] that contains all assessed quantiles from all experts, as well as the realization if known, is considered. This interval is extended to:

$$[L^*,U^*]=[L-k(U-L),U+k(U-L)] \qquad (3)$$

The value of k is typically chosen by the analyst as 0.1. A large value of k tends to make all experts look informative and tends to suppress the relative differences in information scores. The information score is defined as:

$$I_j(e) = 0.05\ln\frac{0.05}{q_5 - L} + 0.45\ln\frac{0.45}{q_{50} - q_5} + 0.45\ln\frac{0.45}{q_{95} - q_{50}} + 0.05\ln\frac{0.05}{U - q_{95}} + \ln(U^* - L^*) \quad (4)$$

The information score for a given question j is computed using the following formula, which depends on the three quantiles as well as on L* and U*. For all calibration questions, the information score is defined as:

$$\mathrm{Inf}_{(e)} = \frac{\sum_{j=1}^{M} I_j(e)}{M} \quad (5)$$

where, M denotes the number of questions [21].

Decision Maker

A combination of expert assessments is referred to as a Decision Maker (DM). Mathematical techniques for aggregating expert judgments include axiomatic and Bayesian approaches. Axiomatic methods apply predefined combination rules to generate a single probability distribution, such as the linear and logarithmic opinion pools. In contrast, Bayesian methods rely on likelihood functions to update and combine beliefs. Behavioral methods, however, may fail to adequately address problematic group dynamics [8]. Although Bayesian approaches are theoretically rigorous, they are often difficult to implement in practice. Axiomatic approaches, by comparison, are generally more transparent and easier to apply [25]. Therefore, this study focuses on axiomatic combination methods. The DM discussed herein exemplifies linear pooling. The classical model fundamentally serves as a method for determining weights in a linear pool. "Good expertise" is characterized by effective calibration (high statistical likelihood, high p-value) and high information. We want weights that reward good expertise and pass these virtues to the DM. The objective is to identify a set of weights that maximize the product of the calibration and the information scores. So, the following equation could be defined:

$$W_\alpha(e_i) = \frac{cal(e_i) * Inf(e_i) * 1\{cal(e_i) \geq \alpha\}}{\sum_{j=1}^{N} cal(e_j) * Inf(e_j) * 1\{cal(e_j) \geq \alpha\}} \quad (6)$$

The performance-based weight of expert i is given by Wα, which is the ratio of the combined score of experts i to the sum of the combined scores of all experts. Note that the weights are normalized; that is, the sum of the weights of all experts is one.

The term α is the significance level, which must be positive and smaller than the largest calibration score. An indicator function takes the value of 1 if the statistical accuracy is greater than the significance level α and 0 otherwise. This threshold α allows us to select the experts used in our linear pooling. It is possible to compute the calibration and information score of any proposed DM. The DM expected to perform better than the result of simple averaging, called the equal weight DM, and that the proposed DM is not worse than the best expert in the panel [8]. Finally, the newly obtained weights, which depend on α, could be used to aggregate the experts' distributions.

In practical applications, the classical model validates experts' assessments and evaluates expert performance using calibration questions, also referred to as seed questions. Within this framework, calibration questions fulfill three main functions: they assess expert performance, support performance-based weighting of experts, and provide a basis for comparing and evaluating different combinations of experts' assessments (16). Although experts are not expected to know the exact true values of the calibration questions, they are required to provide well-reasoned quantifications of their uncertainty. Also, identifying suitable calibration questions is a challenging task that requires a thorough understanding of the subject matter of the elicitation. Calibration questions should be closely aligned with the variables of interest and appropriately reflect the domain expertise of the participating experts.

Results

Table 1 presents the scores and performance measures of each expert. Among them, Expert 2 achieved the highest combined score, which is the product of the calibration and information scores, approximately 0.636, indicating superior overall performance in terms of both calibration and information scores. These combined scores were then transformed into performance-based weights, which incorporates these combined scores to quantify the relative influence of each expert on the aggregated DM. Based on this weighting scheme, Experts 1 and 2 received the highest weights, reflecting their comparatively superior performance. The final decision maker's weights, determined at a significance level (DMSL) of α = 5% via Equ.6, are also reported in Table.1.

Table 3: Expert Scores and Performance Measures

Expert	Calibration Score	Information Score	Combined Score	Performance Based Weights	DMSL
Expert 1	0.1917	1.765	0.385	0.341	0.348
Expert 2	0.5504	1.155	0.636	0.641	0.652
Expert 3	5.4e-8	2.248	1.2e-7	1.2e-7	0.0
Expert 4	0.014	1.147	0.016	0.016	0.0
Expert 5	0.001	1.073	0.001	0.001	0.0
Expert 6	9.9e-5	0.97	9.6e-5	9.6e-5	0.0

Using DM significance level weights and three quantiles corresponding to the 5%, 50%, and 95% with 90% confidence level for each question of interest (target question), the expert assessments were combined into a single probabilistic representation. This approach could obtain both the probability distribution function (PDF) and cumulative distribution function (CDF) for the DM.

Fig. 3 shows the CDF for the three questions of interest. Expert 1 CDF is shown in blue, Expert 2 CDF is depicted in green, and the CDF for the performance-based decision maker is illustrated in red. It is important to note that the performance-based weight of Expert2 pulls the decision-maker's CDF closer to that of Expert 2. This combined CDF offers not only a central estimate but also a full representation of uncertainty. Using the entire available distribution, the best estimate, which is the median was determined. Accordingly, the median values for the three questions of interest are 38.8%, 14.1%, and 34.5%, respectively, representing the most likely outcomes supported by expert judgment and quantified uncertainty.

The validation of the GOLF proceeded in phases. In the initial development phase, the face validity of the model was assessed by subject-matter experts. Port operation experts reviewed the model logic to ensure that they were reasonable and reflected reality. This qualitative validation leveraged expert opinions to identify any structural errors and fine-tune the parameters. Once the GOLF prototype was operational, the validation phase involved comparing its outputs to empirical data that could be obtained for the container terminals [11]. The authors noted that the official statistical data were insufficient for complete validation. Key indicators such as total annual container volumes were known, but more granular data, such as the number of containers left by train or the number of transshipment containers, were not readily published. To overcome this problem, the validation team is going to do data mining using multiple sources.

GOLF was executed for the corresponding historical period, and its outputs were adjusted to ensure comparability with observed data. The results showed encouraging agreement for several performance metrics. In particular, the number of trains, as predicted by GOLF, differed by approximately 3% from the actual recorded number of trains during the same period. This close alignment shows that the rail flow data generated by GOLF were highly consistent with observed operations, thereby validating this component of the model. However, the model's forecast for the number of transshipped containers exhibited a larger error margin. This discrepancy led to further examination of the transshipment logic implemented in the model. It should be noted that, in the case studied, transshipment represented only a small proportion of the total containers. Consequently, although the percentage error for transshipment was relatively high, its overall effect on the container terminal (CT) performance metrics remained limited.

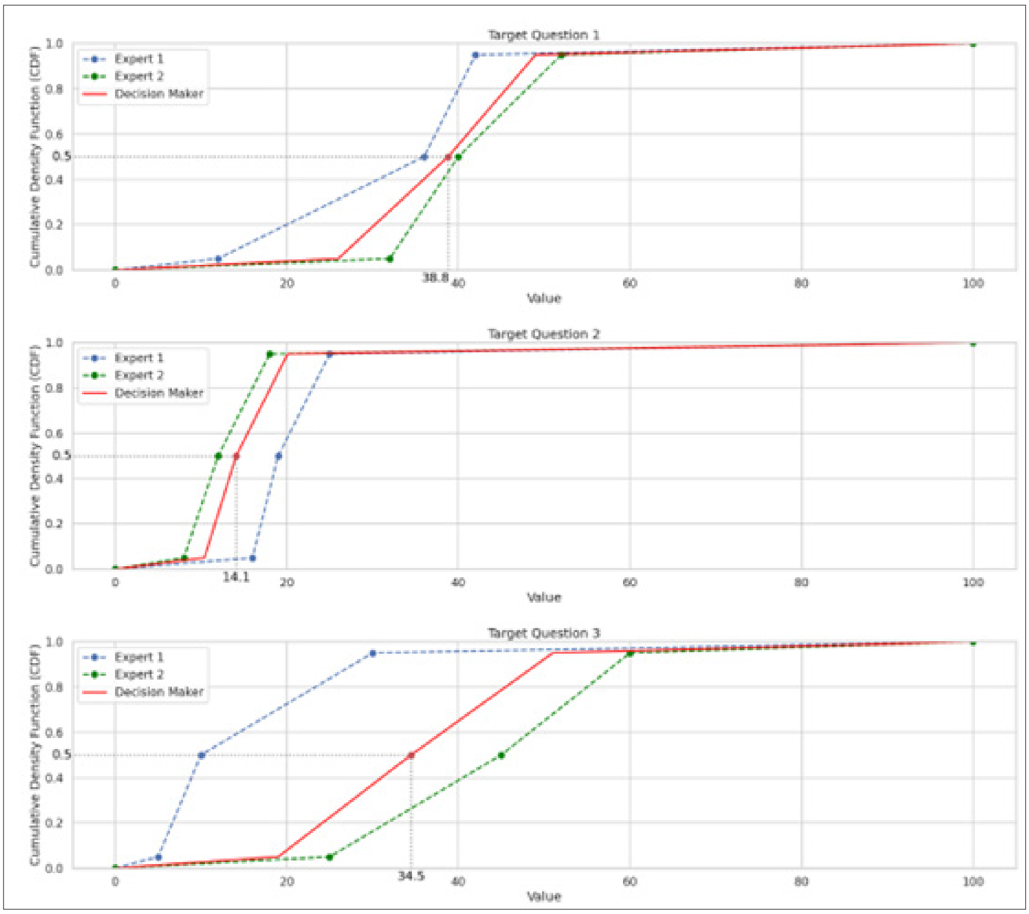

Figure 2: CDF of Decision-Maker for the Three Questions of Interest

Discussion and Future Works

In closing, GOLF represents an advancement in the modeling and simulation of complex logistic systems under uncertainty. It demonstrates that even when faced with unavailable, incomplete, and uninformative data – a common situation in the real world – robust quantitative assessments can still be provided to decision-makers by integrating expert knowledge with advanced simulation techniques. As ports continue to evolve into smart, digitalized hubs of global trade, tools like GOLF will play a crucial role in guiding their growth, ensuring efficiency, and bolstering robustness and sustainability. It requires a calibration-validation loop for each deployment, aligning with how most simulation models are used in practice. The model is as good as the information fed into it.

As a future extension, GOLF's synthetic logistics flows generator could be assessed with Life Cycle Assessment (LCA) models to evaluate the environmental impacts of different port and supply chain strategies under uncertainty. This would allow the tool to not only support operational and policy decision-making but also provide a sustainability perspective aligned with the goals of Industry 4.0 and smart logistics. While GOLF demonstrates significant potential for supporting strategic decision-making under uncertainty in port logistics, its effectiveness depends on the quality of expert input and the availability of validation data. Future improvements could focus on expanding GOLF's application to a wider variety of port configurations, refining scenario sensitivity analyses, and exploring integration with LCA models to assess environmental impacts. These steps will strengthen GOLF's role as a comprehensive decision-support tool for sustainable and efficient port and supply chain management.

Acknowledgments
On 12 September 2025, my dear supervisor, Professor Agostino Bruzzone, passed away unexpectedly. He not only shared his knowledge but also taught us how to think critically and approach complex problems with rigor. His expertise and insightful guidance made a significant contribution to this research, particularly in model development, problem analysis, and experimentation. We remember with gratitude his generosity, dedication, and tireless work.

References
1. Mengqiao Xu, Pan Q, Muscoloni A, Haoxiang Xia, Cannistraci CV (2020) Modular gateway ness connectivity and structural core organization in maritime network science. arXiv: Physics and Society 11: 1-15.
2. Review of Maritime Transport 2024. United Nations Conference on Trade and Development https://unctad.org/system/files/official-document/rmt2024_en.pdf.
3. Daniel M Bernhofen, Bernhofen DM, Zouheir El-Sahli, El-Sahli Z, Richard Kneller, et al. (2016) Estimating the effects of the container revolution on world trade. Journal of International Economics 98: 36-50.
4. Francesco Parola, Parola F, Anna Sciomachen, Sciomachen A (2005) Intermodal container flows in a port system network: Analysis of possible growths via simulation models. International Journal of Production Economics 97: 75-88.
5. Hanea AM, Nane GF, Bedford T, French S (2021) Expert Judgement in Risk and Decision Analysis. Springer International Publishing 293 https://link.springer.com/10.1007/978-3-030-46474-5.
6. Morgan MG (2014) Use (and abuse) of expert elicitation in support of decision making for public policy. Proc Natl Acad Sci USA 111: 7176-7184.
7. Daniel Kahneman, Kahneman D (2011) Thinking, Fast and Slow. https://mlsu.ac.in/econtents/2950_Daniel%20Kahneman%20-%20Thinking,%20Fast%20and%20Slow%20(2013).pdf.
8. Roger M Cooke (1991) Experts in Uncertainty: Opinion and Subjective Probability in Science. New York : Oxford University Press https://archive.org/details/expertsinuncerta0000cook/page/n5/mode/2up.
9. Agostino G Bruzzone, Anna Sciomachen, Marco Gotelli, Antonio Giovannetti, Alessia Giulianetti, et al. (2024)Scenario Simulation to Investigate new Solutions for Port Terminals and Plants based on Intelligent Data Fusion and Supply Chain Modeling. Proceedings of the 26th International Conference on Harbor, Maritime and Multimodal Logistic Modeling & Simulation 2-5.
10. Agostino G Bruzzone, Marco Gotelli, Antonio Giovannetti, Alberto De Paoli, Roberto Ferrari, et al. (2023) Strategic Engineering for Decision Making during Urban Crises. Proceedings of The 13th International Defence and Homeland Security Simulation Workshop 2-5.
11. Bruzzone AG, Massei M, Frosolini M (2022) Redesign of Supply Chain in Fashion Industry based on Strategic Engineering. Procedia Computer Science 200: 1913-1918.
12. Felipe Gana, Yu (Alan) Zhang, Juliana Cordeiro, George Saad (2025) Enhancing Port Logistics through Discrete Event Simulation: A Focus on Rail Operations. Ports https://ascelibrary.org/doi/10.1061/9780784486153.019.
13. Abderaouf Benghalia, Amani Ferdjallah, Mustapha Oudani, J. Boukachour (2025) Machine Learning and Simulation for Efficiency and Sustainability in Container Terminals. Sustainability 17: 1-22.
14. Kan Wang, Hang Xu, Hao Wang, Rui Qiu, Qianqian Hu, et al. (2024) Digital twin-driven safety management and decision support approach for port operations and logistics. Frontiers in Marine Science 11: 1-23.
15. Oppenheimer M, Little CM, Cooke RM (2016) Expert judgement and uncertainty quantification for climate change. Nature Clim Change 6: 445-451.
16. Cooke RM (2013) Uncertainty analysis comes to integrated assessment models for climate change…and conversely. Climatic Change 117: 467-479.
17. Puig D, Morales-Nápoles O, Bakhtiari F, Landa G (2018) The accountability imperative for quantifying the uncertainty of emission forecasts: evidence from Mexico. Climate Policy 18: 742-751.
18. Rothlisberger JD, Finnoff DC, Cooke RM, Lodge DM (2012) Ship-borne Nonindigenous Species Diminish Great Lakes Ecosystem Services. Ecosystems 15: 1-15.
19. Wittmann ME, Cooke RM, Rothlisberger JD, Lodge DM (2014) Using Structured Expert Judgment to Assess Invasive Species Prevention: Asian Carp and the Mississippi-Great Lakes Hydrologic Connection. Environ Sci Technol 48: 2150-2156.
20. Garcia Alcaraz JL, Reza RD, Macias EJ, Vidal RPI, Montalvo FJF, et al. (2022) Effect of the Sustainable Supply Chain on Business Performance- The Maquiladora Experience. IEEE Access 10: 40829-40842.
21. Cooke RM, Goossens LLHJ (2008) TU Delft expert judgment data base. Reliability Engineering & System Safety 93: 657-674.
22. Hora SC, Fransen BR, Hawkins N, Susel I (2013) Median Aggregation of Distribution Functions. Decision Analysis 10: 279-291.
23. Hemming V, Burgman MA, Hanea AM, McBride MF, Wintle BC (2018) A practical guide to structured expert elicitation using the IDEA protocol. Methods Ecol Evol 9: 169-180.
24. Colson AR, Cooke RM (2018) Expert Elicitation: Using the Classical Model to Validate Experts' Judgments. Review of Environmental Economics and Policy 12: 113-132.
25. Clemen RT, Winkler RL (1999) Combining Probability Distributions From Experts in Risk Analysis. Risk Analysis 19: 187-203.