How to Train Your Ship Traffic Model

Lessons from developing data-driven microscopic maritime traffic simulation models as a design tool for the Houston Ship Channel Gate Complex

Jasper van den Broek

Delft University of Technology

TU Delft

MARIN

How to Train Your Ship Traffic Model

Lessons from developing data-driven microscopic maritime traffic simulation models as a design tool for the Houston Ship Channel Gate Complex

by

Jasper van den Broek

to obtain the degree of Master of Science in Transport, Infrastructure, and Logistics
at Delft University of Technology,
Faculty of Civil Engineering and Geosciences.

to be defended publicly on Thursday, March 26th, 2026 at 12:45.

Student number:	5262887		
Course code:	TIL5060		
Project duration:	September 3rd, 2025 – March 26th, 2026		
Thesis committee:	Prof.dr.ir. M. (Mark) van Koningsveld	TU Delft CEG	Chair
	Dr.ir. W.M. (Winnie) Daamen	TU Delft CEG	Supervisor
	Dr. F. (Frederik) Schulte	TU Delft ME	Supervisor
	Y. (Yvonne) Koldenhof	MARIN	External Supervisor
	Dr. N. (Nadia) Pourmohammadzia	TU Delft CEG	Additional Examiner

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Document Version:	Final
Document Date:	March 19th, 2026
Photography:	Jasper van den Broek
Report Style:	TU Delft Report Style by K.P. Hart

Preface

Texas has long held a special place in my heart and my upbringing. As my time as a student comes to a close, and I am about to step into the “adult” world, I am lucky to have had the opportunity to write a thesis that relates to this place that holds personal significance to me. Therefore, after over six months of hard work, I proudly present this thesis.

This project would not have been possible without the many people that supported me throughout. Therefore I would like to take this opportunity to acknowledge and thank those who have helped me.

As is customary, I acknowledge my supervisors from Delft University of Technology: Mark, Winnie, and Frederik. I appreciate the many substantive discussions I had with Winnie and the substantive feedback she provided on my work. I also want to note Frederik’s encouragement throughout the project, and Mark’s role in serving as the chair of the thesis committee. I also want to acknowledge Nadia for stepping in as additional examiner.

I further want to thank Yvonne for her role as supervisor on behalf of MARIN. I very much appreciate the many discussions we had around the whiteboard throughout the project, and the time you spent on providing feedback on my work. Yvonne helped a lot in placing this research into the much needed practical perspective, and by generally showing her appreciation for the work I was doing.

While this document showcases my thesis project from an academic perspective, I am sad to note that very little of what actually motivated me to do this project can be reflected in an academic document like this. I was much more motivated by the experience of working at MARIN, building my professional networks, and experiencing the site I studied first-hand, rather than doing (and writing) academic research. I want future students to know that I believe it is okay to not be hyper-focused on academia during a project like this, and that it is very valuable to see your research in the world beyond the university.

For example, I was especially lucky to actually be able to do part of this project on location in Texas, and believe that this was a once-in-a-lifetime experience that made the problems I studied tangible, and provided much of the practical perspective in this project too. I want to thank Professor Jonkman for making this possible. I also want to thank Arjan Voogt (MARIN USA) for helping me with the interviews while I was in Houston, and of course for inviting me to his family’s Thanksgiving. Finally, I also want to thank the experts who took the time to talk to me about their experiences in navigating Bolivar Roads, and about their perspectives on the gate complex.

I also want to take this opportunity to thank those who I hold dear, to thank those who have kept me sane throughout the process: thank you to my friends I made in Delft and my friends back home for being there for me throughout my studies. Whether we have only recently met or have known each other since were little kids, I want you all to know that I appreciate the time we spend together. I especially want to express gratitude to my friends in the thesis room, who helped me maintain discipline throughout the project and gave me a reason to show up every day. I hope I did the same for you.

I further want to specifically thank my friends in Scouting. Thank you for providing a home away from home, and for giving me something that I can look forward to every Saturday, to get some release from the daily grind.

But most of all, I want to thank my girlfriend Madeline. Thank you for your love, your kindness, and putting up to me when I got frustrated throughout the process. Thank you for the many delicious snacks, lunches, and dinners you made for me, thank you for picking me up from Schiphol, and thank you for being there for me throughout the process. Your love, patience, and care made all the difference too.

And last but not least, I want to thank my parents. Thank you for supporting me throughout my studies, and thank you for always being there for me and supporting me in my dreams and ambitions. Thank you for always being proud of me and for appreciating the things I do.

Jasper van den Broek
Delft, March 19th, 2026

Abstract

The proposed Houston Ship Channel Gate Complex (HSCGC) near Galveston, Texas, is a central element of the Texas coastal protection plan and is intended to reduce societal and economic risks from hurricanes and sea-level rise for the Greater Houston Port System. At the same time, the gate introduces a major new intervention in a heavily used waterway, with previous studies indicating that the proposed layout may create navigational hazards and act as a chokepoint that affects traffic well beyond the immediate gate location. This thesis situates the HSCGC as an example of a broader trend: increasing implementation of flood-protection and other fixed structures in navigable waterways that must continue to accommodate growing maritime transportation demand, raising the need for tools that can evaluate traffic impacts early in design.

The objective of this research is to contribute to the development of data-driven microscopic maritime traffic simulation models as a design tool for new maritime infrastructure, using the HSCGC as the case context. The main research question asks: "*What requirements and characteristics must such a data-driven maritime traffic simulation model have to assess the impact of prospective maritime infrastructures on maritime traffic patterns in the context of designing the HSCGC?*"

To answer this question, a mixed methodology is applied. A literature review establishes the state of the art in microscopic ship-traffic modelling and motivates the selection of AIS-based learning approaches, because they can reproduce complex manoeuvring behaviour without fully prescribing rules or equations. However, because purely data-driven models generalize poorly to unseen infrastructure, the thesis justifies the exploration of Safe Reinforcement Learning extensions, specifically safety filtering layers that can enforce collision and obstacle-avoidance constraints while deviating minimally from learned behaviour. Empirically, AIS data from 2024 is processed into trajectories to derive baseline traffic structure, kinematics, and interaction hotspots, while semi-structured interviews with expert navigators complement AIS by identifying operational constraints and anticipating behavioural changes under an HSCGC scenario. Finally, the selected simulator (ShipNaviSim) and extensions [1] are evaluated on historical realism and situational adaptability using trajectory- and behaviour-focused performance indicators.

Results show that maritime traffic in the study area is highly structured yet interaction-rich: dominant channel-aligned flows coexist with frequent crossings (notably the Galveston-Point Bolivar ferry corridor), producing localized encounter hotspots and heterogeneous manoeuvring demand. The evaluated data-driven simulator reproduces goal-seeking motion and qualitatively plausible transit classes, but does not consistently match observed kinematic distributions (speed, drift, curvature, and acceleration), limiting quantitative realism. Among tested extensions, intermediate goals improve channel-following substantially, while an MPC-based safety layer reduces obstacle entry violations and supports scenario execution under modified geometries, though robustness remains challenging in head-on and high-density encounters.

The thesis concludes that a design-capable data-driven maritime traffic model must be a validated microscopic multi-agent AIS-driven simulator that reproduces site-specific route structure, interaction dynamics, and vessel heterogeneity, while explicitly accepting scenario inputs for new obstacles and changing demand patterns. Critically, it must incorporate a robust safety mechanism, such as safety filtering within a safe reinforcement learning framework, to enable credible and safe behaviour in previously unseen infrastructure configurations.

Keywords: Safe Reinforcement Learning, Maritime Traffic Model, Microsimulation, AIS Data, Design Tool, Model Development

[1] Source code is available via https://github.com/TUDelft-CITG/traffic-behaviour-cloning

Contents

List of Figures

List of Tables

Nomenclature

Abbreviation	Definition
AIS	Automatic Identification System
ATB	Articulated Tug and Barge
BC	Behaviour Cloning
COG	Course Over Ground
COLREGs	International Regulations for Preventing Collisions at Sea 1972
DOF	Degree(s) of Freedom
GC-ADE	Goal-Conditioned Average Displacement Error
GIWW	Gulf Intracoastal Waterway
HDG	Heading
HSC	Houston Ship Channel
HSCGC	Houston Ship Channel Gate Complex
IMO	International Maritime Organization
KPI	Key Performance Indicator
MMSI	Maritime Mobile Service Identity
MPC	Model Predictive Control
NOAA	National Oceanic and Atmospheric Administration
OD	Origin-Destination
ODD	Overview, Design, and Detail
RL	Reinforcement Learning
ROT	Rate of Turn
SOG	Speed Over Ground
US	United States
USA	United States of America
USACE	United States Army Corps of Civil Engineers
UTC	Universal Coordinated Time

Mathematical Notation

Symbol	Definition	Unit
A_Ω	Area of zone Ω	m^2
$\mathcal{A}_{safe}$	Set of safe actions.	
$\mathbf{a}_t^k$	Action of ship k at timestep. t	
$\mathbf{a}_t^{k\prime}$	Corrected action of ship k at timestep. t	
$dist$	Euclidean distance.	m
dh	Change in heading h.	°
dt	Timestep size.	s
dx	Change in x (easting, longitude).	m
dy	Change in y (northing, latitude).	m
$f(\mathbf{s}_t^k, \mathbf{a}_t^k)$	Kinematic mapping function.	
G_k	Ordered set of goal vectors of ship k.	
$\mathbf{g}_t^k$	Goal vector of ship k at timestep t.	
h_t^k	Heading of ship k at timestep t.	°
I	Identity matrix	
K	Set of active ships in the simulation. $k \in K$	
L	Gate line for traffic metrics.	
M	Set of rays cast from a ship. $m \in M$	
$N_{\Omega t}$	Number of ships present in zone Ω at timestep t	
n	Goal setting interval.	steps
$\mathbf{o}_t^k(S_t)$	Observation of environment state S at timestep t from the perspective of of ship k.	
P	Set of obstacle polygons. $p \in P$	
q_L	Flow of traffic across gate L	ships/h
q_Ω	Flow of traffic in zone Ω	ships/m^2/h
$q_{\Omega,\max}$	Capacity of zone Ω	ships/m^2/h
R	Weighting matrix.	
$\hat{\mathbf{r}}_m$	Unit vector along ray m.	
S_t	Environment state at timestep t.	
$\mathcal{S}_{safe}$	Set of safe states.	
$\mathbf{s}_t^k$	State of ship k at timestep t.	
$u_{\Omega,agg}$	Time-aggregated space-mean speed in zone Ω at timestep t	m/s
$u_{\Omega,t}$	Instantaneous space-mean speed in zone Ω at timestep t	m/s
t	Timestep.	
T	Number of steps in a trajectory.	
v_t^k	Speed of ship k at timestep t.	m/s
x	Position (easting, longitude) relative to reference coordinates.	m
x_t^k	Position (easting, longitude) of ship k at timestep t relative to reference coordinates.	m
y	Position (northing, latitude) relative to reference coordinates.	m
y_t^k	Position (northing, latitude) of ship k at timestep t relative to reference coordinates.	m
Z	Objective value.	
ζ_{mt}	Slack penalty along ray m at timestep t	m
θ	Weights and biases of the policy π_θ.	
λ_{mt}	Weight corresponding to ζ_{mt}.	
π	Behaviour Cloning Policy.	
φ	Angle relative to ship heading.	°

Symbol	Definition	Unit
$\rho_{\Omega t}$	Density of traffic in zone Ω at timestep t	ships/m^2
τ	Timestep at which a the linearization of Constraint 6.17 is performed.	
Ω	Zone for the purposes of traffic metrics.	


IT'S NOT A RACE
LEAVE SOME
SPACE
EAST BEACH & STEWART BEACH

1 Introduction

As part of the Texas coastal protection plan, the construction of a storm surge barrier in the Bolivar Roads strait near Galveston, Texas has been proposed by the United States Army Corps of Civil Engineers (USACE) (Greater Houston Port Bureau, 2023b). This barrier, known as the Houston Ship Channel Gate Complex (HSCGC) should protect the city of Houston, the most populous city in Texas and fifth most populous in the United States, and its port, a major petroleum hub and second-busiest port in the United States, as well as surrounding ports in Galveston and Texas City, from the economic and societal risks caused by hurricanes and rising sea levels along the Texas coast. The HSCGC is one of the most important and costly elements in the proposed coastal protection.

A recent study commissioned by the Houston and Galveston pilots showed that the current proposed design, as shown in Figure 1.1, is perceived as undesirable from a marine navigation point of view (Burkley et al., 2022). Specifically, both the siting and the layout of the current proposed design are considered a hazard that complicate safe passage of Bolivar Roads. The human-in-the-loop simulations performed in the study by Burkley et al. showed that the proposed design requires ships to perform turns at severe angles, in an environment with strong (2.5 knot) currents that sees heavy ship traffic. In addition to that, the gate complex introduces a “chokepoint for ship traffic in the area” which impacts the “spacing, timing, and navigation of ships for miles on either side of the HSCGC” (Burkley et al., 2022, p. 19). Burkley et al.’s report also shows that with the possibility of collisions with other ships, or allisions with the gate complex structure(s), it is evident that the introduction of the HSCGC would significantly alter the maritime navigational conditions in the Bolivar Roads strait.

The HSCGC is a flood protection project that introduces a new structural intervention into the Houston Ship Channel. While its primary function is storm surge mitigation, the HSCGC also modifies the navigational environment by changing channel geometry and introducing additional operational constraints. The construction and operation of the HSCGC would affect the traffic capacity and logistics of the Houston Ship Channel (Greater Houston Port Bureau, 2023a). These changes affect vessel manoeuvrability and ship-ship interactions within the the HSCGC immediate surroundings, with implications for traffic flow and capacity. However, the implications of these changes for maritime traffic patterns on the larger scale are not studied by Burkley et al., as their study concerns the local safety in the gate structure. Using their approach to study larger scale maritime traffic patterns by traffic situations through real-time human-in-the-loop simulations would be expensive in terms of time, capital, and human resources. Therefore, it is necessary to assess the impact of the HSCGC on maritime traffic patterns, including how the gate complex influences traffic flow and capacity through adjacent waterways (see Figure 1.2) and the Greater Houston Port System as a whole, through fast-time simulation instead.

The case of the HSCGC is an example of a broader pattern of increasing intervention in navigable waterways. Driven by rising shipping volumes, port development, development of offshore infrastructures, and flood risk management, structures such as storm surge barriers, floodgates, bridges, and tunnels are being introduced into waterways that must continue to support a growing demand for maritime traffic. Such interventions alter the physical and operational characteristics of waterways and therefore influence how vessels navigate and interact with the waterway, as well as with other vessels.

Simultaneously, the assessment of changes in maritime traffic patterns associated with alternative infrastructure designs remains challenging. The evaluation of vessel behaviour and traffic dynamics is often based on observed traffic patterns under existing conditions. Data-driven ship traffic models using Automatic Identification System (AIS) data capture these behavioural patterns effectively, but they are inherently based on historical observations and fixed waterway configurations.

The introduction of the HSCGC creates a navigational configuration that has not previously existed in the Houston Ship Channel, and for which no corresponding traffic data are available. Nevertheless, decisions regarding its design and operation require insight into how vessels and traffic patterns may respond. This situation motivates the need to examine how data-driven maritime traffic simulation models can be applied or extended to assess vessel behaviour and traffic dynamics under modified waterway conditions.

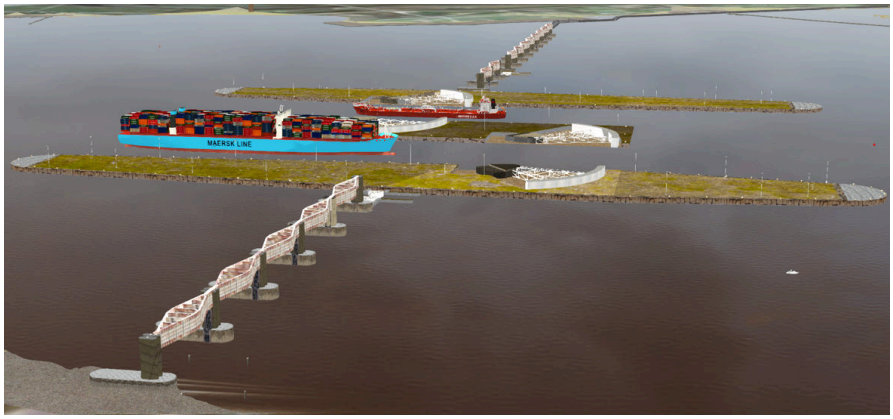


Figure 1.1: Impression of a proposed design of the Houston Ship Channel Gate Complex (Greater Houston Port Bureau, 2023b).

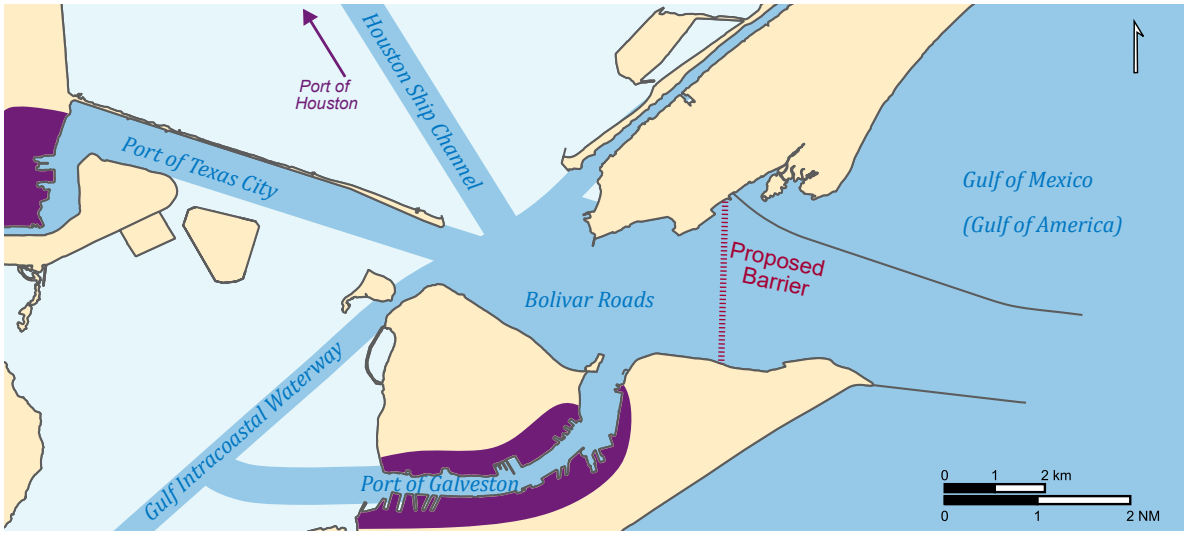


Figure 1.2: Situation map of the waterways around the proposed HSCGC. Major port areas are highlighted in purple. Map data © “OpenStreetMap” (2025).

Current research shows that maritime traffic patterns and ship behaviour can be modelled microscopically, and that such microscopic modelling allows emergent behaviours such as ship-ship and ship-object interaction to be captured (Bellsolà Olba et al., 2019; Li et al., 2013; Mathioudakis et al., 2025; Papandreou et al., 2025; Shu, 2019; Zhou, 2022). Most models rely on either deterministic rule-following logic or mathematical law to determine the motions and interactions of each ship (Desyani, 2019). Recently, scientific attention has shifted to foregoing imposing mathematical law or rules-based logic on ship manoeuvring models, shifting to replicating observed ship manoeuvring and traffic behaviour using machine learning based approaches instead, making use of ship’s self-reported AIS

(Automatic Identification System) transponder data (Basrur et al., 2021; Pan et al., 2025; Pham et al., 2025).

1.1. Problem Statement

Although these machine learning AIS data-driven ship-traffic models provide a useful foundation for analysing vessel behaviour, they lack the ability to account for scenarios in which the waterway infrastructure has changed. The reason for this is that since AIS data reflects only past traffic patterns, no data exists for situations that have not yet occurred. This makes current data-driven models unable to represent traffic responses to modified waterways. In turn, this means that the applicability of data-driven maritime traffic simulations as a tool in the design of maritime infrastructures is limited. Hence, this gap limits the ability to evaluate how infrastructure changes may influence future traffic flows (that is, after realization of an intervention in a waterway) using data-driven maritime traffic simulation models.

Therefore, in order to address this gap, the main objective of this research is to use the case of the design of the HSCGC in order to contribute to the development of data-driven maritime traffic simulation models as a design tool for new maritime infrastructures.

1.2. Research Questions

The objective posed in the previous section leads to the following main research question:

What are the requirements and characteristics of a data-driven maritime traffic simulation model that is useful to assess the impact of prospective maritime infrastructures on maritime traffic patterns in the context of designing the Houston Ship Channel Gate Complex?

In order to systematically investigate this question, this study considers the following sub-questions:

1. *What is the current state of the art in data-driven ship traffic models and their potential adaptations for use in maritime infrastructure design?*
    (a) *What is the state of the art in data-driven maritime traffic simulation models?*
    (b) *What theoretical extensions are identified in literature to adapt these models for infrastructure design?*
    (c) *What theoretical requirements should a data-driven maritime traffic simulation model meet?*
2. *What are the existing characteristic vessel movements, traffic patterns, and behavioural features observed from AIS data at the proposed site of the HSCGC that a maritime traffic model must be able to capture?*
3. *What are the characteristic vessel movements, traffic patterns, and behavioural features as observed under current conditions and predicted under future conditions by expert navigators that a maritime traffic simulation model must be able to capture?*
4. *To what extent does the selected data-driven maritime traffic model and its extensions meet the performance criteria for both historical realism and situational adaptability?*
5. *Based on their performance, which developed model extensions demonstrate the highest efficacy in capturing critical vessel behaviours and should be prioritized for further investigation in order to apply data-driven maritime traffic simulations as design tools for maritime infrastructure design?*

1.3. Definitions and Scoping

In this study, a “manoeuvring model” refers to relatively simple models that describe the movement of ships using three degrees of freedom (2D position and heading). It is important to note that these models generally do not account for hydrodynamic forces that act on ships explicitly, as some other ship manoeuvring models (such as those used to study ship dynamics) do. Consequently, this study does not aim to model individual forces on ship to determine their motions, but rather looks at their motions using trajectories.

Furthermore, this study specifically focuses on maritime traffic simulation models as a design tool. This means that the models studied in this research should be applicable during the design phase of new infrastructures. Whilst other applications of maritime traffic simulation (for example, traffic management or integration with human-in-the-loop simulations) can be imagined, they are left for future research. In addition, this study limits itself to data-driven methods, as these represent a relatively new class of maritime traffic simulation models whose applicability in design contexts has not yet been extensively explored.

The scope of this study is further limited by its reliance on AIS data to numerically describe the motions of vessels. This choice is made because AIS provides a standardized, centrally collected, and publicly accessible source of vessel information, ensuring consistency and comparability across observations. In contrast, incorporating additional data sources, such as video recordings or detailed ship kinematic measurements obtained from on-board sensors, would require substantially greater effort in terms of data acquisition, processing, and harmonization. Given these practical constraints, AIS data offers the most feasible and reliable foundation for the analyses conducted in this study.

1.4. Research Methodology

To answer the research questions, the methodology shown in Figure 1.3 is used. Research Question 1 is answered through a literature review, which is used to determine the state of the art in data-driven ship traffic modelling, and to select the most promising model to test as a design tool. Furthermore, potential approaches to extend the selected model are formulated. The literature review is also used to determine which key performance indicators (KPIs) are suitable to assess the performance of data-driven ship traffic models.

Next, historic AIS data from the proposed site for the HSCGC is analysed in order to determine what characteristic behaviours are observed in the study area, and therefore what characteristic movements and behaviours that data-driven maritime traffic models should be able to capture. This way, Research Question 2 is addressed. The AIS data is also used to formulate test cases which are used to assess the performance of the selected model and the proposed extensions, using the KPIs identified from the literature review.

In addition to that, in order to address Research Question 3, semi-structured interviews with local expert navigators are conducted. The purpose of these interviews is to highlight behaviours that are not (or rarely) observed from AIS data, as well as to gain insights in how experts anticipate their behaviours would change around new infrastructures. Together, these steps provide a list of characteristic manoeuvring behaviours that a data-driven ship traffic model must be able to capture.

The rest of this research focuses on extending and assessing the selected maritime traffic model from the literature review in order to determine their usability as design tools for new maritime infrastructures. Therefore, the ability of these models to replicate manoeuvring behaviour as observed in historical data as well as their ability to interact with unseen obstacles is tested, addressing Research Question 4.

Finally, the results of the model performance assessment are used to provide insights and recommendations in the further development of data-driven maritime traffic simulations as design tools for maritime infrastructure, addressing Research Question 5.

1.5. Document Structure

The rest of this document is structured as follows: Chapter 2 describes the literature review which introduces related work and describes the current state of the art in data-driven ship traffic modelling. Chapter 3 analyses existing behaviours through AIS data and formulates model requirements specific to the case of the HSCGC. Chapter 4 expands these model requirements based on the results of interviews with experienced navigators from the study area. The results from these chapters are then summarized in Chapter 5. After that, Chapter 6 introduces a model structure and extensions to existing models that should theoretically (partially) meet these requirements. The performance of these models is then assessed in Chapter 7. Furthermore, Chapter 8 interprets the results and limitations of this study, and identifies directions for future research. Finally, in Chapter 9 the research questions are answered and concrete recommendations for future research are formulated.

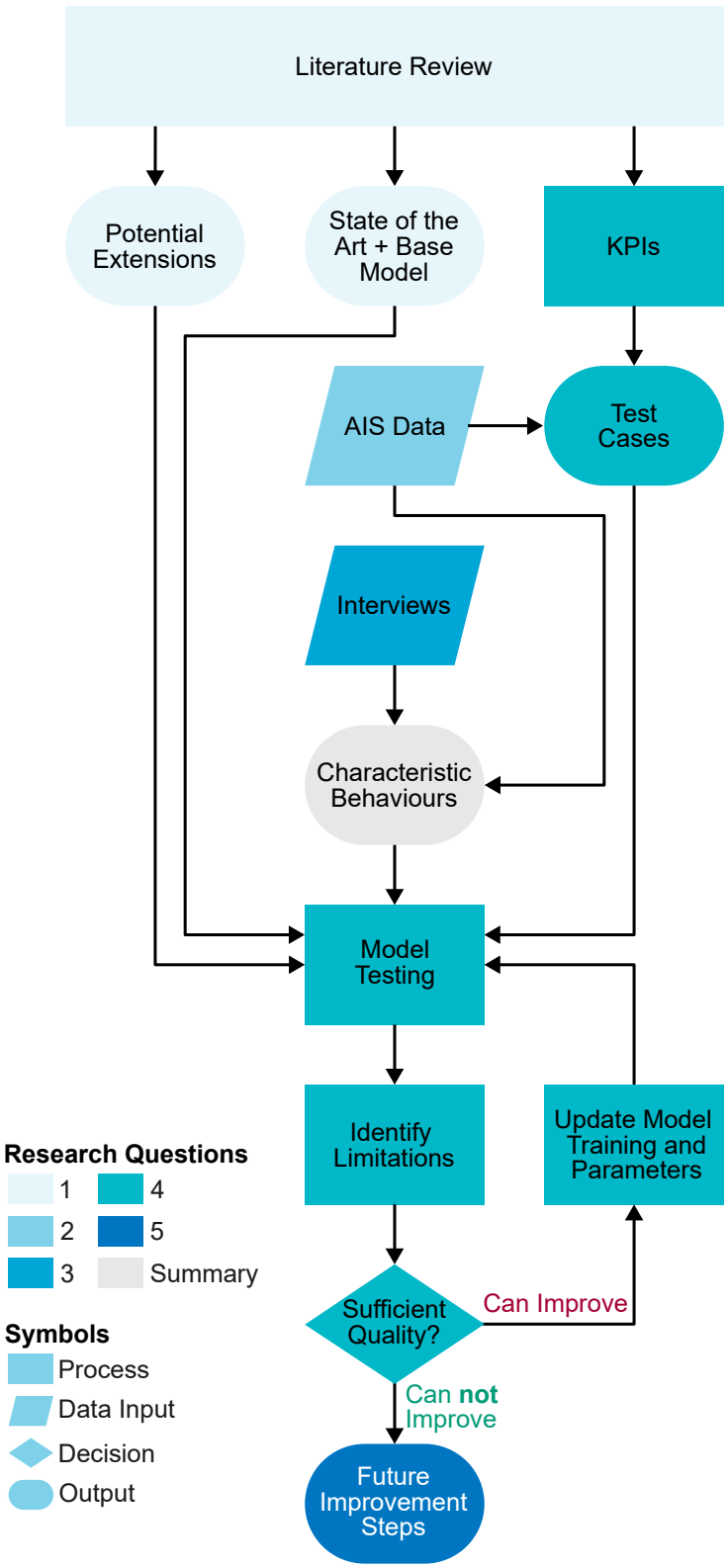


Figure 1.3: Methodological framework used in this study.


GARMIN
APT DL
APT DL

2

Literature Review

As part of this thesis' investigation into how a data-driven ship-traffic model can be extended to account for changes in waterway geometry, and how such an extended model performs in the context of the Houston Ship Channel Gate Complex, this literature review examines the foundations on which such an extension to a data-driven model must build. Although data-driven ship traffic modelling has advanced significantly with growing data availability and computational capabilities, existing approaches are limited to analysing existing situations, instead of new situations for which there is no data available, and as such are not yet suitable as a design tool.

To identify the conditions under which data-driven maritime traffic simulation models are suitable as a design tool, this literature review explores the current state of the art in data-driven ship-traffic modelling. In addition to that, methods for adapting models to altered environmental or infrastructural conditions are also explored. This way, Research Question 1: "*What is the current state of the art in data-driven ship traffic models and their potential adaptations for use in maritime infrastructure design?*" is addressed. This literature review, focuses on the following sub-questions:

- *What is the state of the art in data-driven maritime traffic simulation models?*
- *What theoretical extensions are identified in literature to adapt these models for infrastructure design?*
- *What theoretical requirements should a data-driven maritime traffic simulation model meet?*

This literature review is conducted as an iterative and exploratory process aimed at developing a comprehensive understanding of data-driven modelling approaches in maritime traffic research. The review began with systematic searches in academic databases (most notably, Google Scholar). This search is complemented by the targeted use of generative AI assistants as online search tools to identify relevant publications from keywords and papers.

Keywords such as "ship", "traffic model", "maritime", "data-driven", and "AIS data" are used in various combinations to capture a broad range of studies. These searches yield an initial set of conceptually relevant and frequently cited papers. The review is subsequently expanded using a snowballing approach, examining both the reference lists of key publications and more recent studies citing them. This enables the identification of influential contributions, recurring methodological patterns, and the evolution of research themes within the field.

Rather than applying a strict stopping criterion, the search process continued until thematic saturation was reached, meaning that newly identified publications no longer introduced substantially new concepts, modelling approaches, or data perspectives. At this stage, the literature was considered to provide a sufficiently coherent and representative overview of the field.

The reviewed literature provides both conceptual and methodological grounding for the present work. It clarifies how existing modelling strategies are positioned within the broader academic landscape and informs the selection of approaches and assumptions adopted in this study. In addition, this literature review highlights limitations and open challenges in current research, thereby identifying areas where the present work seeks to contribute.

The following sections present the results of this literature review, focusing first on the state-of-the-art in maritime traffic models, followed by a more in-depth look at data-driven models. Next, safe reinforcement learning is explored a a method to adapt data-driven models to be suitable for new situations. Finally, the findings from the literature review are used to synthesize a list of theoretical requirements that data-driven maritime traffic simulation models must meet in order to be useful to assess the impact of prospective maritime infrastructures.

2.1. Features of Maritime Traffic Models

When it comes to the state of the art in data-driven maritime traffic simulation models, Shu (2019) found that in constrained waterways, vessel speed and manoeuvring behaviour is influenced by the following factors: size and type of the vessel, the geometry of the waterway, and the direction of navigation. Furthermore, Shu also found that the main external influences on vessel manoeuvring behaviour are wind, visibility, and current, as well as head-on and overtaking encounters with other traffic. Zhou et al. (2019) affirm that these factors are non-negligible.

In addition to considering the movement of individual ships, the manoeuvring behaviour of ships is also determined by the environment in which the movement takes place, and other traffic within this environment.

The density of maritime traffic can be measured by taking the average density of ships across a water region (Wang et al., 2025). Furthermore, Wang et al. stipulate that the flow of traffic can be quantified as the number of ship crossings across a predefined line (a “gate”) within a certain time window. Finally, traffic volume is commonly calculated as the product of density and speed. From this, the capacity of a channel, which is the “maximum number of ships that can negotiate a channel in a given unit of time” (Goodwin, 1975) can be determined.

Since Yip’s work shows that macroscopic traffic flow and capacity emerge from individual ship movements, simulating traffic at the microscopic scale enables conclusions to be drawn about patterns at the network level. This principle underlies the framework proposed by Bellsolà Olba et al. (2017, 2019), where network indicators such as traffic flow and waterway capacity are constructed from microscopic variables including ship velocity, position, and headway.

Within this framework, infrastructure characteristics, navigator behaviour, external conditions, and transport demand interact at the level of individual vessels. Microscopic simulation models explicitly represent these local interactions, making them particularly suitable for analysing the effects of infrastructure changes in specific network sections, such as intersections or channels (Daguano et al., 2023). By capturing how vessels manoeuvre and interact around a new or modified infrastructure element, these models make it possible to assess its local operational impact and translate these effects into network-level performance indicators. This allows the model to be used to quantify how infrastructural changes at a specific location affect traffic flow and capacity, which in turn can be used to assess the changes in traffic characteristics on a network level. Therefore, the microscopic simulation paradigm is appropriate to study the effects of local infrastructure interventions, which can be used to inform the implications of the change in traffic conditions for the broader maritime network.

2.2. Data-Driven Maritime Traffic Models

Data-driven maritime traffic models are models where the behaviour of ships in the traffic model is inferred from real-world observations, instead of imposed on the model through rules-based logic or mathematical law. Such ship control agents are often the product of reinforcement learning processes, and therefore often only have their model structure imposed on them, leaving it to the model training process to determine the policy that transforms an input to an output.

Using reinforcement learning to construct such data-driven models has two distinct advantages. Firstly, reinforcement learning (RL) explicitly models sequential decision-making under dynamic and interactive conditions, which aligns closely with the objective of simulating vessel behaviour in response to infrastructure configurations and surrounding traffic, since “the agent perceives the state of the environment and takes an action” (Buşoniu et al., 2010), after which the environment transits to the new state. Unlike purely predictive models that estimate future vessel states based on historical data, reinforcement learning frameworks aim to learn behavioural policies that adapt to changing environmental conditions. Second, reinforcement learning provides a natural interface with microscopic traffic simulation, as it allows vessel-level agents to make navigational decisions within a simulated environment. Given that this work research seeks to analyse how local infrastructure interventions in-

fluence vessel interactions and network performance, reinforcement learning-based approaches offer the most conceptually and methodologically relevant foundation. Therefore, this section concentrates on this specific class of models.

In imitation learning (IL) models, the agents controlling the ships in the simulation can be trained to imitate the actions of actual humans by being "shown" the actions taken by human decision-makers, and tasked to imitate their actions through a reward function. Furthermore, the reward function can be used to account for other environmental metrics as well. In this case, it is possible to use observations from AIS data to train a model to imitate the behaviour of human-controlled ships. Using this data provides a commonality with other models that also use historical AIS data (Desyani, 2019), while also accounting for complex behaviours that are not properly captured by behavioural rules or mathematical laws. Therefore, in this work it is required that the model makes use of AIS data to infer the behaviour of ships in the model from real-world observations.

Recent work by Pham et al. (2025) documents the development of ShipNaviSim, a "data-driven maritime traffic simulator". Pham et al.'s work shows that it is possible to accurately simulate maritime traffic, specifically the movements and interactions of multiple ships in complex and/or high traffic areas, through imitation learning algorithms. Pham et al.'s work also shows that their model "indirectly captures captures the effect of ocean currents, wind, and other weather condition on the vessel movement" by accounting for drift (the difference between heading and COG) in the evaluation of their agents. It must be noted however, that by capturing these factors implicitly, the model cannot distinguish *why* drift occurs, but only *that* drift occurs.

Whether the same vessel would change its behaviour under different wind or current conditions depends on how those conditions are represented during training. If the model is trained on trajectories that span a range of environmental states or is explicitly conditioned on wind/current, it can likely learn to adapt its actions to environmental conditions, which allows these conditions to be varied when assessing infrastructure designs. However, if environmental effects are only captured indirectly (e.g. via drift, as in ShipNaviSim) the agent may not generalise reliably to deliberately altered weather and current conditions, which means that the influence of weather conditions cannot be studied using this model.

Other work by Basrur et al. (2021) shows that it is possible to use generative adversarial networks to stochastically model maritime traffic and ship travel times across a waterway.

Furthermore, reinforcement learning algorithms have been used to create multi-ship collision avoidance models, which account for safety, efficiency, and compliance with COLREGs, in order to aid ship crews with navigational decision making (Pan et al., 2025). Other work by Düz and van Iperen (2025) shows that it is possible to use reinforcement learning to predict the trajectories of ships based on their historical AIS data. The commonality between the works of Pham et al., Pan et al., Basrur et al., and Düz and van Iperen is that works make use of agents trained on real-world navigational data in order to simulate complex encounters between ships.

2.3. Safe Reinforcement Learning

The key problem in using data-driven traffic simulations based on a reinforcement learning approach for modelling interventions in an environment is that no real-world data is available before the intervention is realized. Therefore, real-world AIS data from the site of a proposed infrastructure cannot be used to study the effects of a new obstacle directly, although data from reference situations where similar projects have been undertaken may be able to partially infer the influence of the new infrastructure. Nonetheless, using only reference data means that local peculiarities may be lost. Therefore, the ability to balance between local observed behaviours and the ability to incorporate *new* obstacles into the model is a key requirement for the model, as the aim of the extended model is to be suitable to assess changes in the design of the waterway.

There are several established ways to add safety guarantees to reinforcement-learning-based control policies in continuous control settings, especially when an explicit kinematic model is available. One common class of approaches are optimization-based safety filters, which take the action proposed by a learned policy and adjust it as little as possible so that state constraints such as obstacle avoidance are still satisfied. In particular, Control Barrier Functions provide conditions under which a chosen safe set remains forward invariant, and in practice they can be enforced by solving a small quadratic program that projects the policy action onto the set of safe actions (Ames et al., 2019; Cheng et al., 2019). A closely related idea is to use model predictive control(MPC) as a safety certification

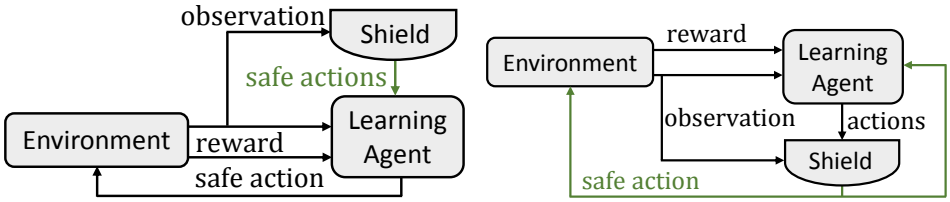


Figure 2.1: Preemptive (left) and Post-Posed (right) shielding (Alshiekh et al., 2017).

layer, where constraints are enforced by solving a constrained optimization problem over a short horizon while the learning-based controller acts as the nominal reference (Wabersich & Zeilinger, 2019). Finally, for low-dimensional kinematic models, Hamilton-Jacobi reachability can be used to compute the set of states that would eventually lead to a collision and to synthesize safety-preserving controls with worst-case guarantees under bounded disturbances, although this typically comes with a higher offline computational cost (Bansal et al., 2017; Fisac et al., 2018).

While these alternatives are well-founded, they generally require an explicit model and constraints that can be expressed in a way that is suitable for online optimization or reachability computations. In contrast, the main appeal of safe reinforcement learning (Alshiekh et al., 2017) in is its modularity: safety constraints can be enforced by an additional component layered on top of an existing machine-learning-based agent, which makes it possible to add safety during both training and deployment without having to redesign the underlying learning pipeline. This relative simplicity is the primary consideration for selecting this method.

Avoiding obstacles is a key task in safe navigation for ships. Therefore, adding a safety layer to the ship control agent can be a method to ensure safety (i.e., obstacle avoidance) in addition to imitating real-world behaviour (Alshiekh et al., 2017).

Safe reinforcement learning is “the process of learning policies which maximise the expected cumulative reward in problems where it is non-trivial to ensure compliance with the safety constraints in both the learning process and the implementation” (Odriozola-Olalde et al., 2023). In other words, it is a concept where an agent trained using reinforcement learning is bounded by safety constraints imposed on the agent. Crucially, safe reinforcement learning is a method that is useful for applications where it is not possible to have the agent reach an unsafe state, such as collision with an obstacle.

Safe reinforcement learning through *shielding* was first introduced by Alshiekh et al. (2017). This concept augments the traditional reinforcement learning setting by introducing a shield, which either limits the action space of the agent such that the agent can only select safe actions (pre-emptive shielding), or corrects the agent’s actions if an unsafe action is selected (post-posed shielding). These frameworks are illustrated in Figure 2.1.

The advantage of this approach is that the shield operates independently from the agent, and is able to correct an agent’s actions in both the learning and deployment phases. Especially, post-posed shielding has the advantage that “it works even if the learning algorithm is already in the execution phase and therefore follows a fixed policy” (Alshiekh et al., 2017). Furthermore, a shielding approach causes minimal interference with the agent’s actions, as the shield only interferes when a safety constraint is violated (Odriozola-Olalde et al., 2023).

Since shielded reinforcement learning works even when the agent is never able to explore an “unsafe” state, it can be used to filter out actions that would lead the ship agent to collide with an obstacle, without the agent having seen the obstacle during training. Furthermore, since shielding is a reactive method, the agent’s decisions are only modified with minimum intervention (Odriozola-Olalde et al., 2023). Hence, shielded reinforcement learning can be used to extend a data-driven maritime traffic model such that the ship agents emulate real-world behaviour, being only minimally corrected by a shield. Therefore, in this study we require a shield that “takes over” from the data-driven agent when the ship encounters a new obstacle in the waterway.

Shielded reinforcement learning approaches have also been shown to work for problems with continuous action spaces (Dalal et al., 2018; Dawood et al., 2024), and with multi-agent systems (Sheebaelhamd et al., 2021). These works propose that a minimum corrective action can be determined analytically by solving a quadratic program that minimizes the difference between the action selected by a policy and the actual action carried out, subject to a safety constraint.

To summarize, shielded reinforcement learning can be used to introduced new, and crucially, un-

seen constraints on the agents' action space. These constraints can, for example, be geographic (i.e., constraints on navigation imposed by changing waterway geometry), based on ship COG, or based on ship speed. Therefore, incorporating the principle of safe reinforcement learning allows the fusion of data-driven models with situations that have not been seen before.

2.4. Conclusions and Model Requirements

In conclusion, the current state of the art in data-driven maritime traffic modelling can be described as follows: the use of microscopic simulations allows for the individual movements and behaviour of ships in a waterway to be modelled, and allows us to capture external factors, as well as interaction effects (ship-ship and ship-obstacle). This in turn allows us to use these microscopic behaviours to model emergent patterns like traffic flow, capacity, density at the macroscopic level.

Generally, there are three modelling paradigms that capture these microscopic behaviours. Models can either be grounded in rules-based logic, mathematical law, or data-driven paradigms. The latter allows us to let the model learn real-world behaviours from AIS data, without the need to explicitly impose mathematical or logical rules on the model. The advantage of the data-driven approach is therefore that such models can also capture real-world behaviours that are not captured by logic or mathematical law. This allows for non-standard situations to be learned by a model without having to explicitly model them, meaning that theoretically such models should be able to capture more complex maritime manoeuvring behaviours with relatively fewer explicit model rules compared to rules-based or explicit mathematical models.

However, these models lack the ability to capture changing conditions, as there is inherently no data available for a new situation before that situation is realized. Therefore, it can be concluded that in order to use data-driven models as a design tool, they must be extended so that the model, in addition to imitating real-world behaviours, also accounts for new obstacles and interaction behaviour with these new obstacles. Since data-driven traffic models are generally based on reinforcement learning paradigms, it can be concluded that *safe* reinforcement learning is a suitable extension method that is useful for handling unseen situations since, in its post-posed form, this method allows for the minimal correction of the agent when it tries to perform an unsafe action, such as colliding with an obstacle. The advantage of this method, as established in Section 2.3, is that it is relatively simple to implement on top of existing RL maritime traffic simulators.

Based on these conclusions, the following theoretical requirements for a data-driven maritime traffic simulation that is useful as a design tool for maritime infrastructures can be formulated:

i. The model must be based on a microscopic simulation paradigm.
   - This requirement follows from the notion that this paradigm allows the model to capture both ship-ship and ship-obstacle interaction on a local level, which can then be used to determine macro-level traffic effects (see Section 2.1).
   - In turn, this requirement also means that the model should be based on a Multi-Agent Simulation paradigm, where there are multiple ships (agents) present in the simulation environment, which are allowed to interact and change their decisions based on each other's actions.

ii. The model should be an extension of a data-driven model that is based on AIS data.
   - This requirement follows from the idea that data-driven models are able to capture complex behavioural patterns without having to impose these patterns through extensive sets of rules-based logic or mathematical laws (see Section 2.2).

iii. The model must be able to include microscopic interactions with *modified* environmental conditions.
   - This requirement is inherent to the design of infrastructure, as the purpose of such models is to be used to assess the impact of new infrastructures. However, as established in Section 2.3, there is inherently no data available for situations that have not been realized yet.
   - To mitigate the lack of available data, a safety filter can be introduced to prevent the ship agents from taking actions that would lead to undesirable states under modified waterway

conditions. Such a filter allows the model to determine when a corrective action should be determined, and allows the agent to interact with previously unseen obstacles.

– If this approach is used, then the model must also include a corrective method that corrects an unsafe action to a safe action, whilst minimizing the impact of this correction. This allows for the model to capture real-world behaviours without imposing rules or mathematical law on the behaviour of ships from historic data, whilst falling back on more explicit modelling of interactions with new obstacles, limiting the deviation of the model's behaviour under new conditions.

Table 2.1 shows an overview of studies examined in this literature review that meet at least some of these requirements. This table also shows the method by which these studies control their simulation agents.

Table 2.1: Overview of selected related studies in the fields of data-driven maritime traffic behaviour models and safe reinforcement learning.

		Interaction		Methodology			
	AIS Data	Agent-Obstacle (Ship-Obstacle)	Agent-Agent (Ship-Ship)	Space Representation	Agent Control Type	Multi-Agent Simulation	Trajectory Prediction
Requirement	ii	i, iii	i	iii	i, iii	i	i
Basrur et al. (2021)	Y	N	N	zones	RL	N	N
Dalal et al. (2018)	N	N	N	continuous	Safe RL	Y	N
Dawood et al. (2024)	N	Y	N	continuous	Safe RL + MPC	N	Y
Düz and van Iperen (2025)	Y	N	N	continous	RL	N	Y
Koenighofer et al. (2024)	N	Y	Y	discrete cell	Safe RL	Y	Y
Pan et al. (2025)	N	N	Y	continuous	RL	Y	Y
Pham et al. (2025)	Y	N	Y	continuous	RL	Y	Y
Sheebaelhamd et al. (2021)	N	N	N	continuous	Safe RL	Y	Y
Shu (2019)	Y	Y	Y	continuous	MPC	N	Y
Design tool	Y	Y	Y	continuous	Safe RL	Y	Y

Key: Y = included in study, N = not included in study.
Abbreviations: RL: Reinforcement Learning, MPC: Model Predictive Control.

From this table, it can be concluded that the works by Dawood et al. (2024), Pham et al. (2025), can be used to form a theoretical basis for an extended data-driven maritime traffic model that can also be used as a design tool, since these models together meet the requirements, and complement each where they lack the required features. Specifically, the work by Pham et al. provides the basis for the AIS data processing, ship-ship interaction, and reinforcement learning training, while the work by Dawood et al. (2024) provides more general methods to implement safe reinforcement learning to detect unsafe actions (i.e., ship-object interaction) and calculates minimal corrections during the manoeuvring of an agent. Using these approaches together also has the advantages that both models are grounded in a continuous representation of space and that both models feature trajectory prediction.

Using these ideas, the work by Bellsolà Olba et al. (2019) provides a framework that transforms microscopic traffic behaviours into macroscopic traffic characteristics (see Section 2.1). Through this

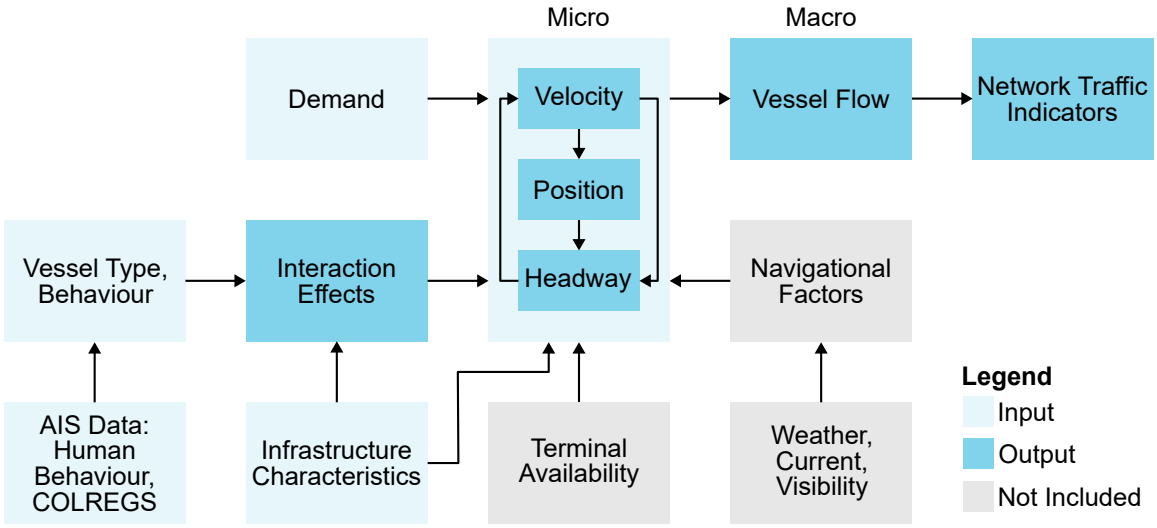


Figure 2.2: Conceptual Port Network Capacity Model, (adapted from Bellsolà Olba et al., 2017).

framework, the microscopic (local) effects of the proposed gate structure in order to determine changes in macroscopic traffic characteristics. In turn, these macroscopic traffic features are useful to determine the impact of the HSCGC on the efficiency and traffic capacity of the ports that lie beyond the HSCGC.

The combination of the works by Dawood et al. and Pham et al. can be used to shape the interaction effects and micro behaviour elements of this framework, such that this framework is adapted to the data-driven paradigm as shown in Figure 2.2. The primary adaptation to this framework from the formulation by Bellsolà Olba et al. is that the input for vessel types and behaviours is derived from AIS Data, and that this adaptation, in contrast to the work by Bellsolà Olba et al. does not explicitly include the effect of navigational rules (COLREGS), weather, and current, opting to infer behaviours caused by these factors from AIS data instead. Furthermore, this study does not consider constraints in waterway capacity caused by terminals inland, as the proposed location of the HSCGC is away from the inland ports.


SAL
SWL 320t
WIEBKE
SAL HEAVYLIFT

3

Data Analysis

Realistic maritime traffic simulation depends on an accurate representation of how vessels actually move and interact in a given area. Simplified assumptions about traffic flows or vessel behaviour risk overlooking important features such as route variability, encounter situations, and localised congestion. This chapter therefore focuses on analysing AIS data from Bolivar Roads in order to identify characteristic vessel movements, traffic patterns, and behavioural features that are essential for informing the modelling process. Specifically, this chapter aims to answer Research Question 2: *"What are the existing characteristic vessel movements, traffic patterns, and behavioural features observed from AIS data at the proposed site of the HSCGC that a maritime traffic model must be able to capture?"*.

This chapter aims to answer this question by analysing AIS data from Bolivar Roads, in order to gain an insight into what behaviours are present in the area of the proposed HSCGC. Studying existing local maritime traffic behaviours is necessary because these behaviours provide the baseline to which any model's performance should be compared. Additionally, studying these existing behaviours also provides an understanding of the existing conditions at the site of the proposed HSCGC, which means that the impact of a potential infrastructure intervention can be determined by comparing the existing and modified situations.

Therefore, first AIS data is collected and processed into trajectories, in order to determine the quality and characteristics of the available manoeuvring data. Next, this data is analysed and visualized using Python (v3.13) and Geopandas (v1.0.1). Specifically, the characteristics of the processed set of tracks are described in order to determine the spatio-temporal distribution and quality of the collected trajectories.

Next, a more detailed analysis into the behaviour captured in these trajectories is performed in order to determine the broad patterns in maritime traffic behaviour observed in Bolivar Roads. This is done by analysing the trajectory data itself: positions, speeds, course over ground (COG), and heading (HDG). Analysing these trajectories is necessary because this step reveals what kind of behaviours should be captured in a maritime traffic simulation model, and therefore what information is used to train the data-driven maritime traffic simulation model.

Finally, the kinematic properties of the trajectories are also analysed. These properties are used to determine in which parts of the study area turning, acceleration, and interaction behaviours are to be expected. These properties are studied in order to use them as metrics that provide a real-world baseline for model validation.

This chapter is structured as follows: firstly, the data selection and filtering process is described. After that, the data is explored to determine which movements, traffic patterns, and behavioural features are characteristic for the study area. These characteristic movements, patterns, and behavioural features are then summarized into a list of requirements in order to identify which patterns data-driven maritime traffic models should be able to capture in order to be used as a design tool.

3.1. Data Selection and Quality

In order to achieve insight into the characteristic vessel behaviours in Bolivar Roads, AIS data is collected from NOAA's MarineCadastre (n.d.-a), a publicly available database of AIS transmissions

in US coastal waters. The selected transmissions were made in 2024, within a bounding box around Bolivar Roads (longitudes between -94.85º and -94.68º, latitudes between 29.32º and 29.39º, 16.48 x 7.78 km).

A full AIS transmission from MarineCadastre consists of the following information:

- Transmission identification: Vessel's MMSI (Maritime Mobile Service Identity) number and time (UTC) of transmission.
- Movement information: Vessel's position (latitude, longitude), speed over ground (SOG), course over ground (COG), and heading (HDG),
- Vessel status information: Vessel status code, cargo code and draft.
- Static vessel information: Vessel's name, IMO (International Maritime Organization) number, call sign, vessel type code, vessel width and draft, and transceiver class code.

This analysis makes use of individual transmissions, which consist of MMSI, time, position, SOG, COG, and status code. The MMSI and status code are used to identify each unique ship and its movement type, whereas the timestamp, position, SOG, and COG are used to construct trajectories. Furthermore, the heading information from a transmission is also used if it is reported by the ship. Finally, the vessel type category is also recorded.

Including heading data is a requirement for Pham et al. (2025)'s ShipNaviSim, since accounting for heading allows their behaviour cloning (BC) method to implicitly capture external forces (wind, current) that influence the manoeuvring behaviour of a ship.

To analyse movement in the study area from AIS data, transmissions are filtered by AIS status code and speed over ground (SOG). Only transmissions with AIS status codes 0 (under way using engine), 3 (restricted manoeuvrability), 4 (constrained by her draught), 7 (engaged in fishing), 12 (tug towing ahead or aside), and 15 (undefined) are selected and SOG ≥0.2 kts are selected. The reason for including more status codes than simply code 0 is because a significant share of ship traffic consists of articulated tug and barge combinations (ATBs) (see Figure 3.1), which tend to self-report their status as "tug" rather than "under way using engine", even though they do not operate as a tug in the sense of assisting another ship. The use ATBs is characteristic for US inland shipping, and cannot be disregarded in this analysis. This means that although tugs contribute a sizeable part of the AIS data, whether a tug is sailing independently or assisting a ship is unknown, and therefore data produced by these vessels is less reliable than data produced by vessels sailing independently.

Filtering by these status codes makes sure that the manoeuvring data does not include ships that are moored, aground, or at anchor, and also removes transmissions made by non-ship objects like personal distress beacons or search and rescue aircraft. However, AIS status code is self-reported by the ship's crew, and not all ships properly update their AIS status codes (either immediately or at all) when they change status. Conversely, filtering by status code also means that ships may be moving, while their status code indicates that they are not. This means that the status code filter alone is not enough to distinguish between ships that are in transit and ships that are not. Therefore, filtering by speed therefore allows us to remove ships that are moored or at anchor which have not properly updated their status code, as these ships will still broadcast (very) low speeds through their AIS if their AIS is not disabled when moored or anchored.

From these selected transmissions tracks are constructed. This step is necessary because the transmissions received are not independent data points. Rather, sequential transmissions are related to each other, and show the progression of a ship's movement through the study area. Tracks are constructed from sequential transmissions made by the same ship, as identified by their MMSI number. A new track is started every time the ship leaves and re-enters the bounding box of the study area, or when there is a gap between two transmissions greater than 30 minutes. A new track is also started if a sharp reversal (≥160°) in COG is observed across five transmissions. In that case, the middle of these 5 transmissions is considered the start of the new track. These thresholds are set based on arbitrary practical considerations. These considerations are made in order to ensure that ships that are continuously present in the study area but produce distinct movements (e.g., ferries) are split into distinct trajectories.

Finally, tracks are only considered if their length is greater than 500 m, and the area of the bounding box of the track is greater than 10000 m^2. The goal of this filtering process is to remove noisy tracks

Figure 3.1: Articulated Tug and Barge combinations (ATBs) on the Houston Ship Channel seen from the San Jacinto Battleground Monument.

(e.g., produced by a ship at anchor), or tracks of ships that were not significantly moving, thereby making sure that the model training data consists only of ships that are actually moving.

The raw data observed in the study area in the year 2024 contains 12042357 transmissions produced by 5525 unique ships. After filtering and processing, 3361084 transmissions (30.40 % of total transmissions) produced by 4768 (86.30 % of total ships) remain in the dataset. These transmissions are distributed across 139268 tracks.

Table 3.1 shows the distribution of the transmissions that compose the dataset by ship type. This table shows that the majority of traffic in Bolivar Roads consists of tanker, tug, and cargo vessels, and that roughly 56.87% of tracks are produced by ships that report their heading through AIS.

Of the selected transmissions, 2501836 (68.34 % of selection, 20.78 % of all transmissions) have heading data available. This corresponds with 4355 ships (91.34 % of selection, 78.82 % of all transmissions) that report their heading data. A detailed distribution of these transmissions across the various vessel categories and status codes can be found in Appendix A. Table 3.1 also shows that between the different ship type categories, there is a sizeable difference in how many ships report their heading, ranging from none (0.00% of military vessels retained) to nearly all (98.99% of cargo vessels retained).

This means that these heading-reporting ships are part of the sample that is suitable for data-driven modelling through ShipNaviSim. Furthermore, Table 3.1 shows that the fleet of ships present in Bolivar Roads is heterogeneous in terms of ship type, which means that a data-driven maritime traffic simulation model applicable in this region should reflect this heterogeneity in ship types. Furthermore, a ShipNaviSim model trained on this dataset may not accurately capture all ship types, and that some ship types will be overrepresented in their contribution to the training dataset, compared to their real-world share of traffic.

Table 3.1: Ship type distribution per transmission in the raw and processed datasets.

	Raw Data	Processed Data (Full)	Processed Data (Heading available)	Share of Ships Retained
Points	12042357	3661084	2501836	
Tracks	N/A	139268	79199	
Cargo	7.00%	5.92%	7.52%	98.99%
Fishing	3.62%	1.92%	1.31%	20.77%
Military	0.09%	0.00%	0.00%	0.00%
Other	22.22%	13.95%	16.31%	40.00%
Passenger	14.86%	17.05%	1.33%	37.63%
Pilot Vessel	3.21%	4.34%	5.97%	88.57%
Pleasure Craft	1.55%	0.74%	0.78%	1.90%
Tanker	9.70%	13.91%	20.35%	99.89%
Tug	36.75%	41.70%	45.84%	73.19%
Unknown	1.01%	0.48%	0.59%	61.11%
Total	100.00%	100.00%	100.00%	

3.2. Track Statistics

This section describes the statistical and spatial properties of the constructed vessel tracks. The aim of this section is to assess the characteristics and representativeness of the processed AIS trajectories and to identify the dominant movement patterns observed in the study area. These statistics provide context for the subsequent analysis of vessel behaviour and traffic patterns at the proposed HSCGC site.

The geographic distribution of the tracks constructed using the process in the previous section is shown in Figure 3.2. This figure shows that the vast majority of vessel movements in Bolivar Roads are concentrated within the designated ship channels. This illustrates that ship traffic in the study area is highly structured and predominantly bound to these channels.

The average track consists of 26.3 transmissions and covers 32.3 minutes, with the median time between transmissions being 1 minute and 9 seconds. Tracks where heading information is available are slightly longer, with an average of 31.6 transmissions over 35.6 minutes. Table 3.2 presents the descriptive statistics for the number of points per track, track duration, and track distance.

Table 3.2: Descriptive statistics of the tracks in the processed dataset

	n	Mean	Median	St. Dev.	Min.	Max.
Points per track						
All tracks	139268	26.3	25	15.3	5	498
Tracks with heading available	79199	31.6	31	14.7	5	498
Track duration (min)						
All tracks	139268	32.3	30.7	18.0	3.1	806.8
Tracks with heading available	79199	36.0	35.6	17.8	3.1	580.0
Track distance (m)						
All tracks	139268	8464	7710	4467	500	37319
Tracks with heading available	79199	10216	10060	4476	500	37319

These statistics show that most tracks represent transits of the study area, as the geographic spread of the trajectories along the ship channels remains relatively constant, with median durations of a trajectory are approximately 30-35 minutes, and median distances range between 7.7 and 10.1 km.

Figure 3.3 shows the distribution of the number of points per track for both the full dataset and the

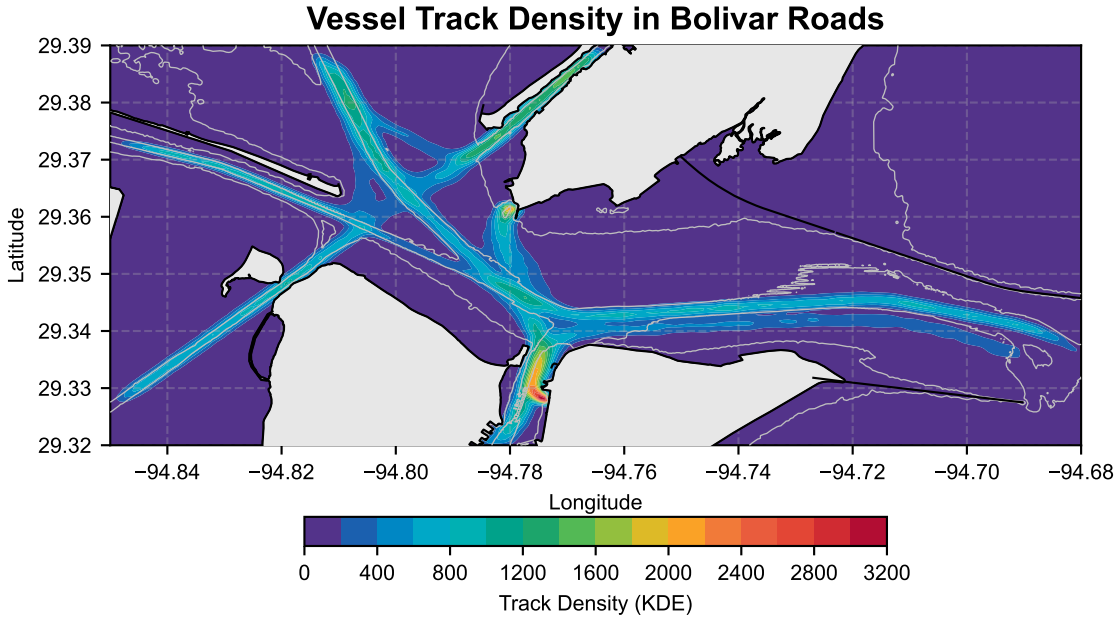


Figure 3.2: Geographic distribution of AIS transmissions in the filtered dataset.

subset where heading information is available. A distinct peak around 15 points is present in the full dataset but largely absent in the subset with heading information. This indicates that shorter tracks are more likely to lack heading data.

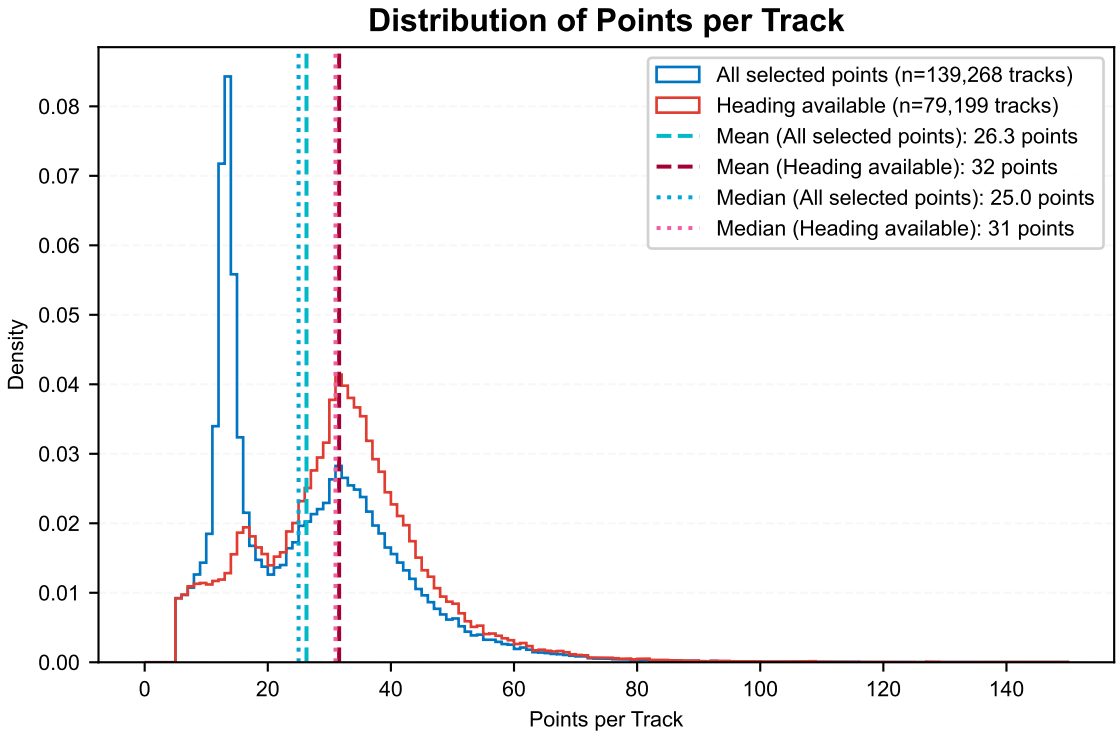


Figure 3.3: Distribution of points per track in the processed tracks.

Similarly, Figure 3.4 shows that a considerable number of tracks around 4500 m are missing heading information. This suggests that certain movement types are disproportionately excluded when analyses rely on heading.

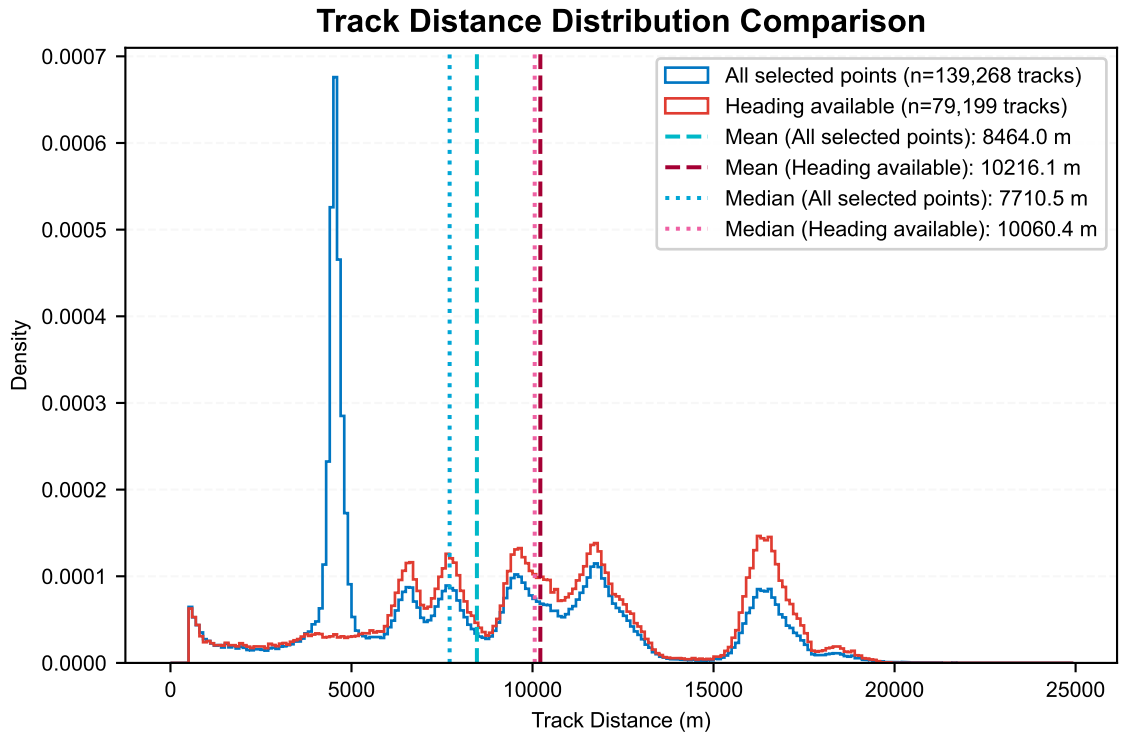


Figure 3.4: Distribution of track distances in the processed tracks.

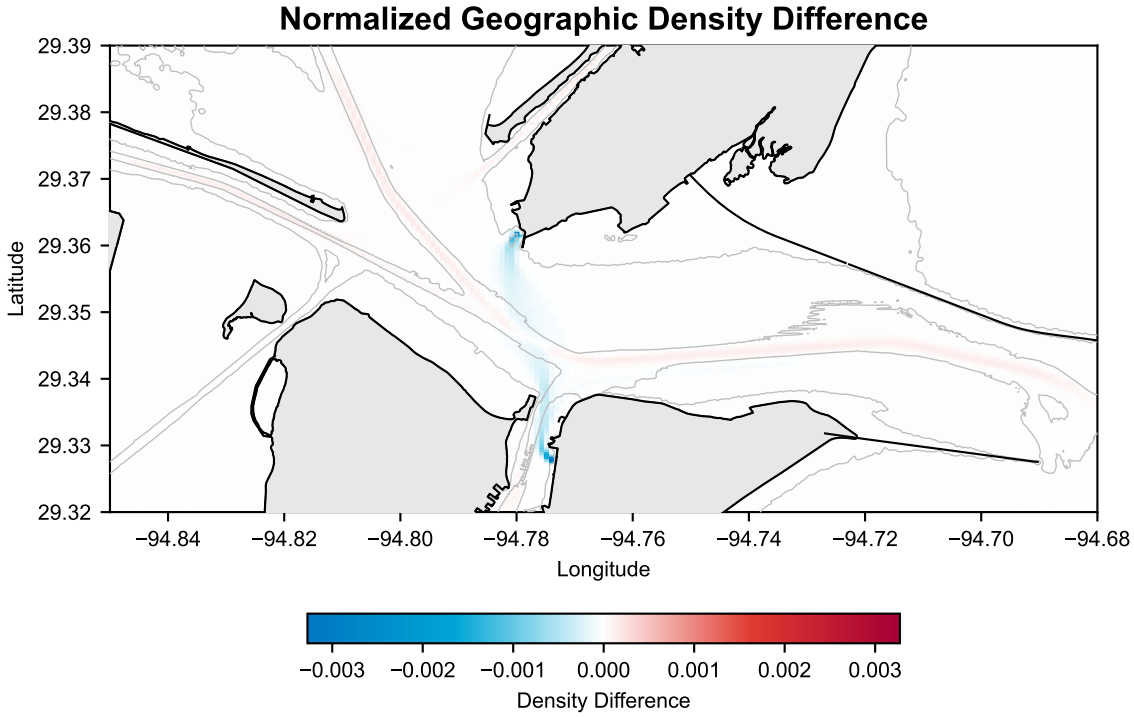


Figure 3.5: Normalized relative geographic density difference between the full dataset and dataset with heading available. Areas shaded in red show where transmissions with heading information are over-represented relative to the full dataset. Areas shaded in blue show where transmissions with heading information are under-represented.

The normalized geographic density difference shown in Figure 3.5 indicates that the main spatial discrepancy between the full dataset and the subset with heading information lies in a north-south

corridor between Galveston and Point Bolivar. This corridor corresponds to the Galveston-Point Bolivar ferry route (Texas Deartment of Transportation, n.d.). The under-representation of this area in the heading-available subset indicates that not all ferry traffic reports heading information.

This implies that movements by the ferries are under-represented in analyses that rely on heading data (see also the share of passenger ships reporting their heading in Table 3.1). Since the ferry traffic represents a recurring crossing flow in the area, this under-representation should be considered when interpreting heading-based behavioural analyses or model training.

Overall, the track statistics provide an overview of the temporal, spatial, and structural characteristics of vessel movements at the proposed site of the Houston Ship Channel Gate Complex. They show that traffic is strongly channel-constrained, that movements are typically of moderate duration and distance, and that certain traffic components, most notably the ferry crossings, are not fully captured when restricting the dataset to tracks with heading information. These observations provide important context for the subsequent analysis of characteristic vessel movements and traffic patterns.

3.3. Traffic Patterns

In order to identify the dominant traffic flows in the study area, the origins and destinations of the constructed tracks are analysed. For this purpose, origins are defined as the first point in a track, and destinations are defined as the last point in a track. By examining the spatial distribution of these points, clusters of origin and destination (OD) locations can be identified. The locations of these clusters are shown in Figure 3.6.

The analysis of OD patterns provides insight into how vessels move through the system as a whole. While the track statistics in the previous section describe the characteristics of individual movements, the OD structure reveals the larger-scale traffic patterns and connectivity between different parts of the study area. This is relevant for understanding which traffic flows dominate the area and how different waterways interact.

From Figure 3.6, it is evident that most traffic in Bolivar Roads enters and leaves the area via one of the ship channels. This conclusion follows from the alignment of the origin and destination clusters with the Houston Ship Channel, the Galveston Ship Channel, and the Texas City Ship Channel. The concentration of OD points at these locations confirms that these channels form the primary access routes to and from the ports in the region.

These clusters can be used to divide the study area into sections, as shown in Figure 3.6. By determining in which section the start and end points of a track lie, the share of trips between each pair of zones can be calculated. This allows for the construction of an OD matrix, which quantifies the relative importance of different traffic flows. The resulting OD-matrix is visualized in Figure 3.7, and can be found in full in Appendix B.

The OD diagram shows that the primary traffic flows through Bolivar Roads are from the Gulf of Mexico to the ports of Houston, Galveston, and Texas City. In addition, a substantial share of the traffic in the study area consists of ferry movements between the Port of Galveston and Point Bolivar. The OD representation thus highlights both the dominant inbound and outbound flows and the recurring cross-channel ferry traffic identified in the previous section.

Furthermore, Figure 3.7 indicates the presence of several intersecting, diverging, and merging traffic flows east of Galveston, particularly in the area where the Gulf Intracoastal Waterway (GIWW) meets the Houston and Texas City Ship Channels. These intersecting flows imply a structurally complex traffic system, in which vessels navigating to different destinations share constrained waterways and interact at junctions.

In relation to the research question, the OD analysis identifies the main traffic corridors, the relative magnitude of different flows, and the locations where flows intersect. These features constitute key characteristic traffic patterns observed in the AIS data, which means that a maritime traffic model of the proposed HSCGC site should therefore be capable of representing these dominant inbound and outbound channel traffic flows, recurrent cross-channel ferry movements, and interactions arising from merging and crossing flows in confined waterways.

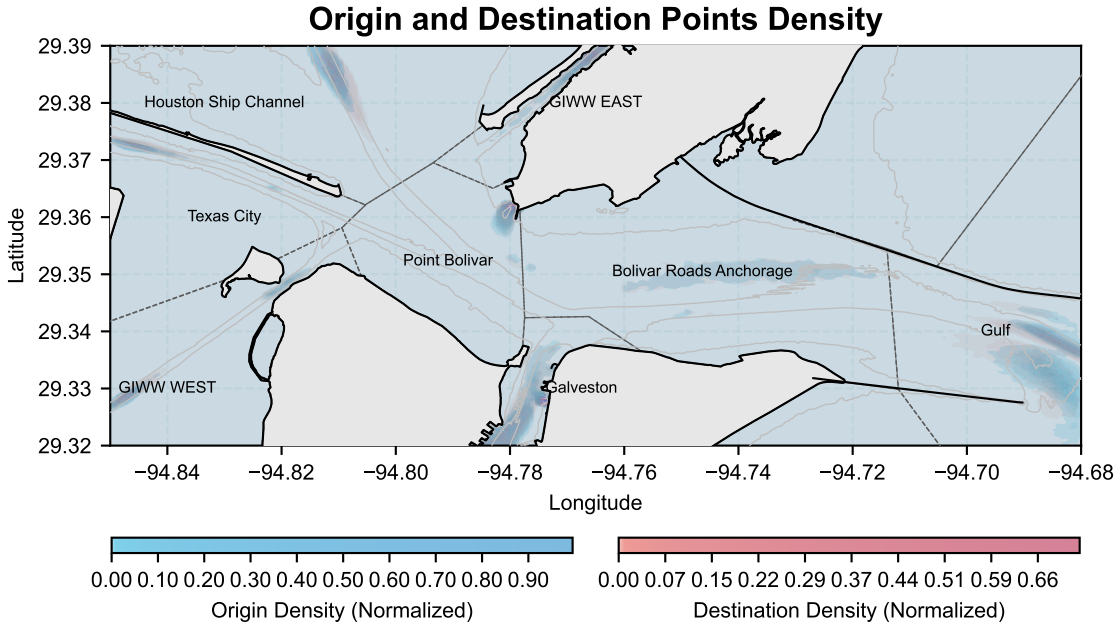


Figure 3.6: Heatmap (kernel density estimation) showing the positions of origin (blue) and destination (red) clusters.

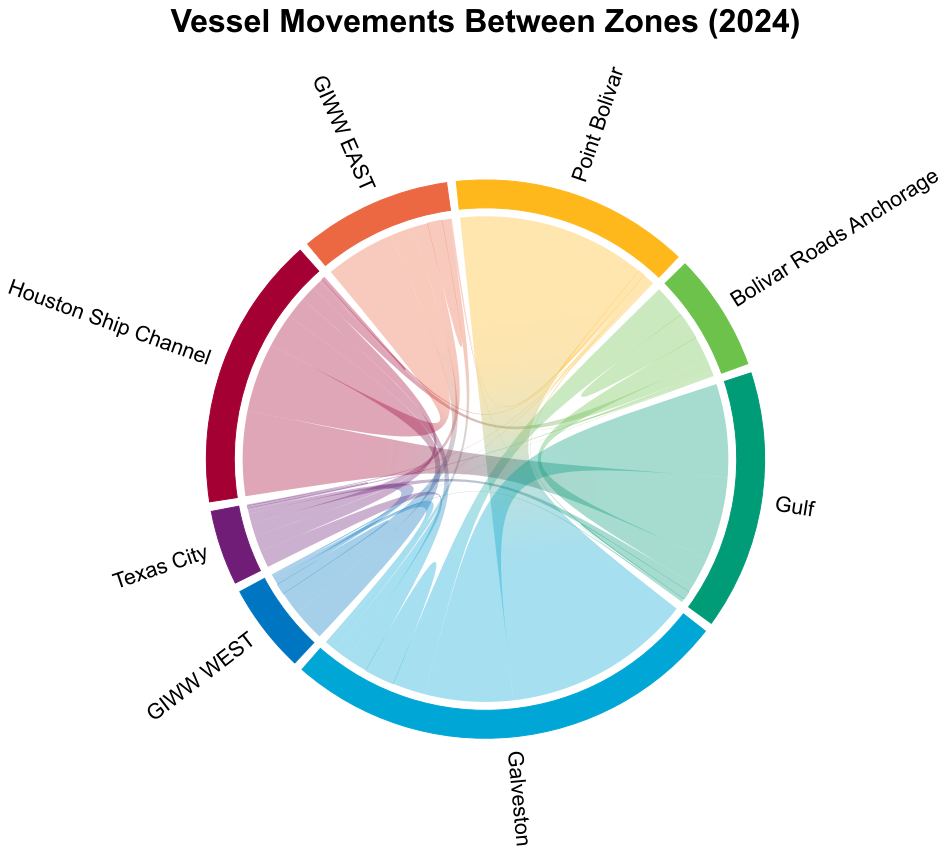


Figure 3.7: Diagram illustrating the share of origins and destinations for each zone as defined in Figure 3.6, and the size of the flows between these zones.

In addition to temporal variation in vessel speeds, the overall distribution of speeds and directions provides further insight into characteristic movement behaviour in the study area. By examining the

joint distribution of speed over ground (SOG) and course over ground (COG), dominant movement regimes and recurring traffic streams can be identified.

Figure 3.8 shows the distribution of SOG in the study area. This figure indicates that the speed distribution in the subset with heading information closely follows that of the full dataset, suggesting that restricting the dataset to transmissions with heading does not substantially distort the overall speed characteristics. Two frequently observed speed peaks are visible, at approximately 5 and 10 knots. The presence of these distinct modes suggests that vessel movements in the area are not uniformly distributed across speeds, but instead cluster around typical operational regimes.

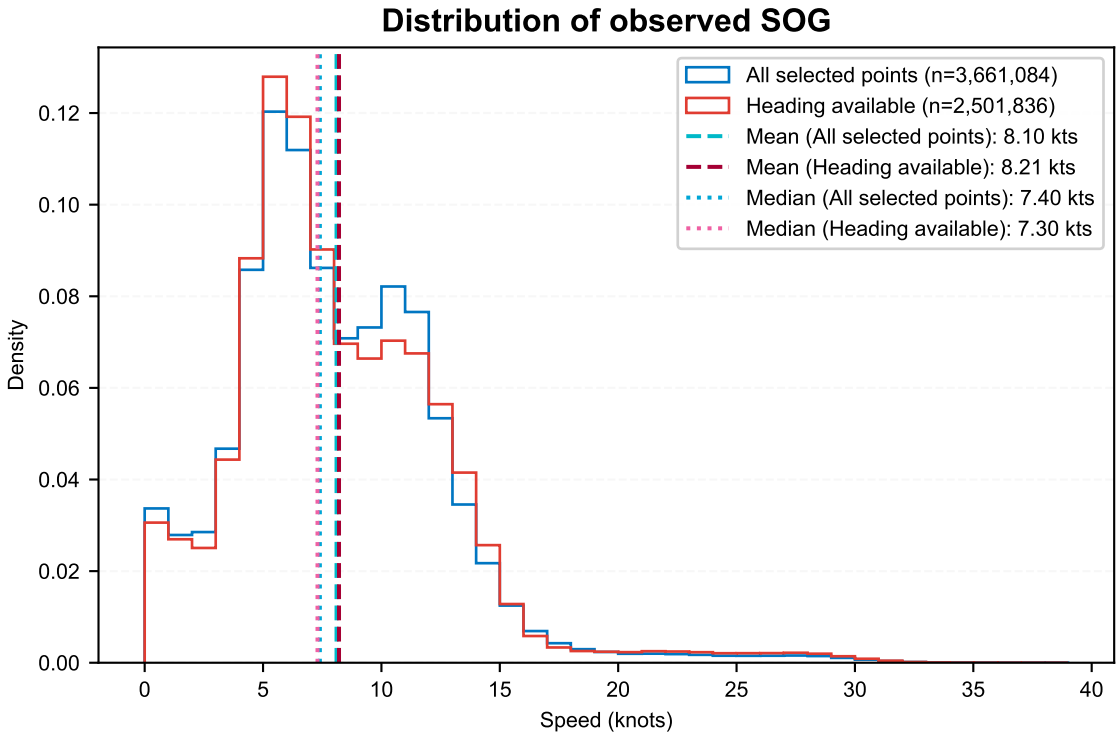


Figure 3.8: Distribution of SOG in the dataset.

Figure 3.9 presents a rose diagram of ship speeds and COG in Bolivar Roads. This figure shows a dominant East-West traffic stream, with vessels frequently transiting the area at speeds between 10 and 14 knots. In addition, a slower-moving stream of traffic is visible crossing the area from South-West to North-East, generally at speeds between 4 and 8 knots. These directional patterns correspond to the main ship channel traffic and the cross-channel ferry movements identified in previous sections.

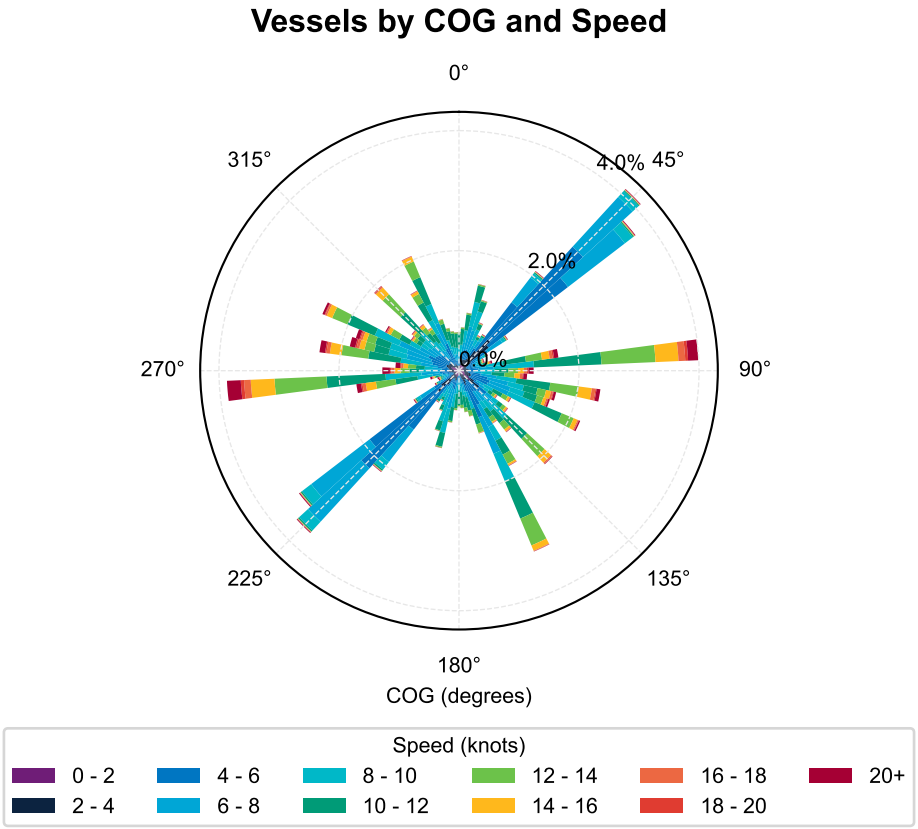


Figure 3.9: Wind rose plot of speed observations against COG

Figure 3.10 shows the geographic variability in average speed and COG using a vector field representation of the mean direction and speed of vessels transiting the study area. This visualization highlights how both direction and speed vary spatially, reflecting channel alignment, crossing flows, and localized manoeuvring behaviour near intersections of waterways.

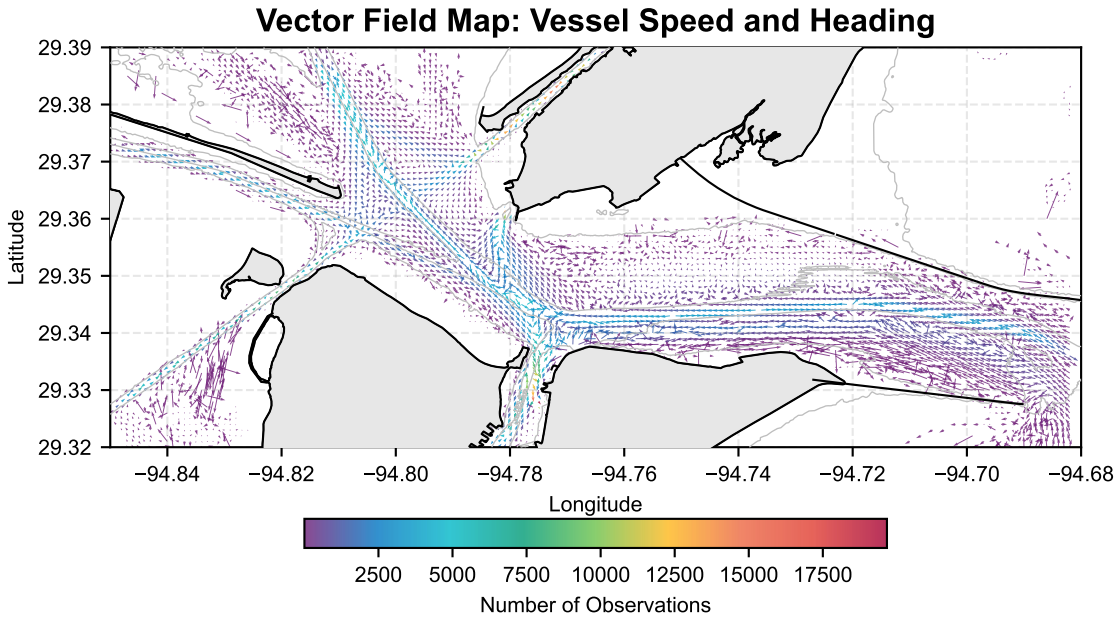


Figure 3.10: Vector field showing average speed and direction of ship traffic in Bolivar Roads. Arrow direction indicates COG, arrow length indicates speed.

Together, these distributions show that vessel movements in the study area are characterized by distinct operational speed regimes, strongly directional east-west channel traffic, and a recurring slower cross-channel flow. This means that, these features represent characteristic behavioural patterns observed in the AIS data. A maritime traffic model of the proposed HSCGC site should therefore be capable of reproducing both the directional structure of traffic flows and the presence of multiple typical speed regimes associated with different movement types.

In addition to spatial traffic structure, temporal variations in traffic intensity and vessel speed provide insight into characteristic behavioural patterns in the study area. By analysing how the number of active vessels and their speeds vary over the course of the day, week, and year, it can be assessed whether traffic conditions are relatively stable or exhibit systematic temporal fluctuations. Such patterns are relevant for understanding typical operating conditions in the area.

Figure 3.11 shows the daily pattern in the average number of active ships in the study area, compared to averages per weekday and per month. The average daily pattern does not vary substantially throughout the week, indicating that traffic intensity is relatively stable across weekdays. In the monthly patterns, there is somewhat more variation in average traffic levels. However, the monthly averages all lie within one standard deviation of the overall mean (see Appendix C), suggesting that no single month deviates significantly from typical traffic conditions.

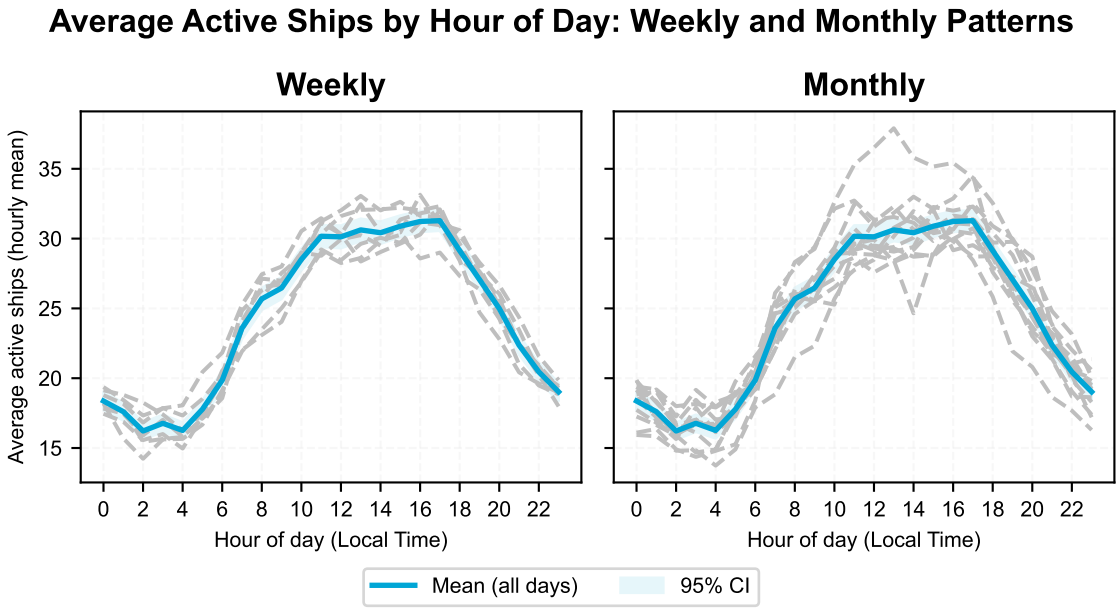


Figure 3.11: Daily average active ships in the study area compared to weekly and monthly average patterns.

These results indicate that, while minor seasonal differences exist, the overall traffic intensity in the study area remains relatively consistent over time. This suggests that the characteristic traffic patterns identified in previous sections are not dominated by isolated temporal peaks but instead reflect persistent operating conditions.

Figure 3.12 shows the distribution of vessel speeds (SOG) by hour of the day. Within the study area, vessel speeds generally vary between 5 and 12 knots. This range reflects typical operating speeds in a confined waterway system with channel constraints and interacting traffic flows.

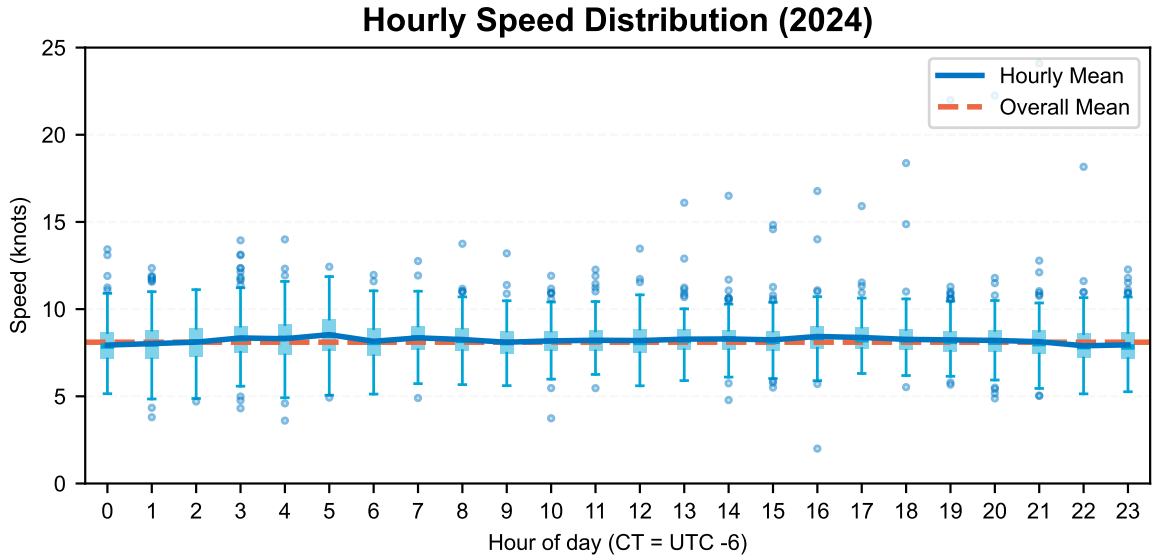


Figure 3.12: Average speed distribution by hour.

Based on the 3661141 AIS transmissions, the overall average vessel speed in the dataset is 8.10 knots, with a standard deviation of 4.63 knots. However, vessel speeds are not constant throughout the day. The average speed varies by hour, and this variation is unlikely to be due to random fluctuations alone. An analysis of variance confirms a statistically significant difference between hourly averages ($F = 157.5$, $p \leq 0.05$). Although this result is descriptive and not further modelled in this study, it indicates that a single daily average speed does not adequately represent vessel behaviour at different times of day.

The observed hourly variation suggests that operational conditions, such as traffic density and vessel interactions, may influence speed selection. Therefore, this implies that temporal variability in both traffic intensity and vessel speed forms part of the characteristic behavioural patterns observed at the proposed HSCGC site. A maritime traffic model intended to represent these movements should therefore allow for variation in traffic intensity over time and for speed behaviour that reflects interaction effects rather than assuming constant vessel speeds.

3.4. Manoeuvring Behaviour

To understand the characteristic behavioural features of vessel traffic in the study area, it is necessary to examine not only where vessels travel, but also how they manoeuvre and interact with each other and with their environment. Variations in speed, turning behaviour, encounters, and drift provide insight into how vessels respond to navigational constraints, traffic density, and environmental conditions. These aspects complement the previously identified spatial and OD patterns by describing how vessels operate within the system.

The acceleration of ships is a useful indicator of manoeuvring behaviour, as it captures where vessels change their speeds. Such changes often reflect local navigational constraints, including traffic density, intersections, bottlenecks, or proximity to ports. By analysing acceleration patterns, zones can be identified where ships systematically alter their behaviour. Figure 3.13 shows the spatial variation in mean absolute acceleration in Bolivar Roads, determined from the AIS data using the central difference in SOG between observations. The figure indicates that speeds are relatively constant within the main ship channels and the GIWW, while higher acceleration levels occur where vessels enter or exit the ship channels and near the port of Galveston. In addition, increased acceleration and deceleration are observed near intersections, which may reflect interaction with other vessels. These patterns suggest that manoeuvring behaviour is not uniformly distributed, but concentrated in specific areas associated with merging, diverging, or crossing flows.

The rate of turn (ROT), derived from AIS data using a central difference in COG, provides further insight into course-altering behaviour. ROT indicates where vessels alter their courses and at what speeds these alterations occur. Course changes may result from channel-following behaviour, port

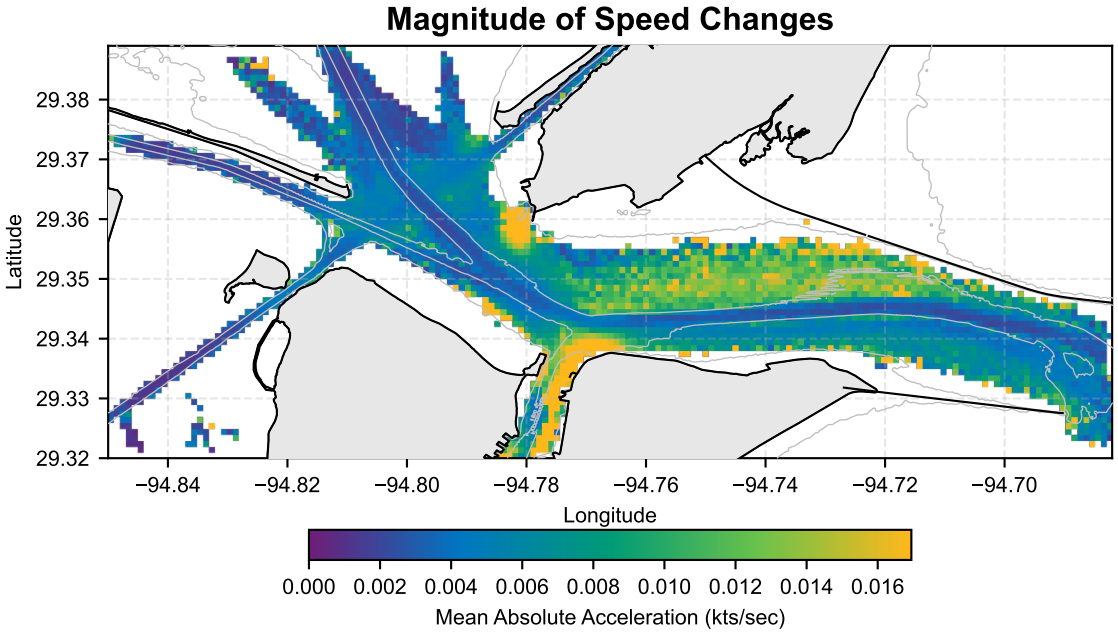


Figure 3.13: The spatial variation in mean absolute acceleration in Bolivar Roads.

approaches, avoidance manoeuvres, or interaction with other vessels. Figure 3.14 shows the relationship between ROT and SOG. From this figure it is evident that vessels operating at lower speeds (5-7 knots) tend to execute slightly sharper turns than vessels operating at higher speeds (10-12 knots). In addition, there is greater variability in ROT at lower speeds. This suggests that more complex manoeuvres are generally associated with lower-speed operations, such as port approaches or crossing traffic.

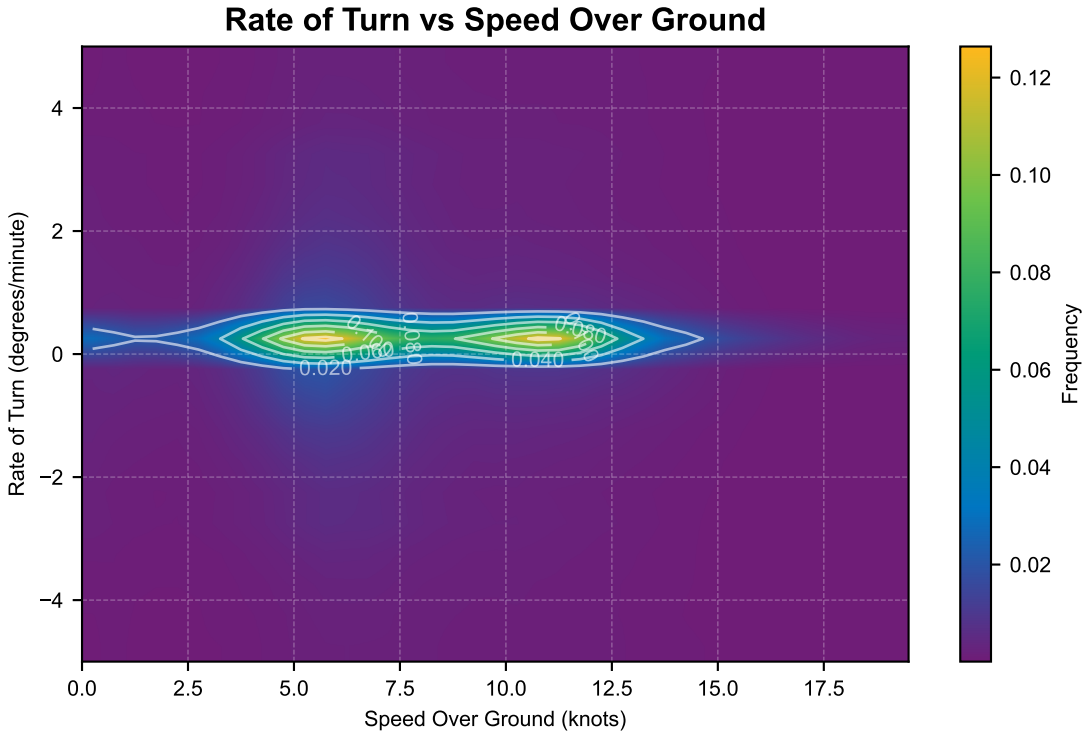


Figure 3.14: The relationship between ROT and SOG observed in Bolivar Roads.

Figure 3.15 highlights the spatial distribution of turning behaviour. As expected, higher rates of turn are observed near intersections and in the Bolivar Roads anchorage area, whereas turning behaviour within the main ship channels is more limited and structured. This indicates that manoeuvring complexity is concentrated in specific zones where traffic streams intersect or where vessels change operational states (e.g., entering anchorage areas or ports).

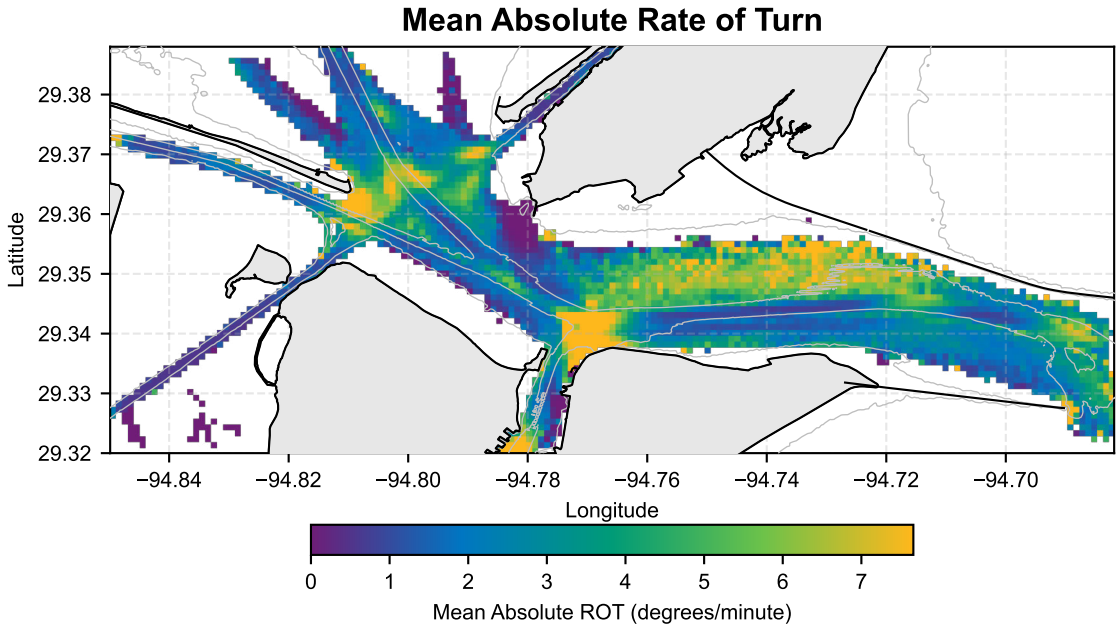


Figure 3.15: The spatial variation in ROT in Bolivar Roads.

Ship-ship encounters provide another indicator of interaction behaviour. Figure 3.16 shows the spatial density of encounters in Bolivar Roads, where an encounter is defined as two ships within 500 m of each other. Encounters were identified by linearly interpolating tracks to 10-second intervals and determining whether two vessels at the same interpolated time stamp were within 500 m. The figure shows that most encounters occur within the ship channels and at their intersections. A particularly high concentration of encounters is observed near the ferry terminals in Galveston and Port Bolivar. This is consistent with the near-continuous ferry operations, with up to five vessels in service simultaneously (Texas Deartment of Transportation, n.d.). These patterns indicate that interaction intensity is not evenly distributed, but concentrated in structurally constrained and operationally active areas.

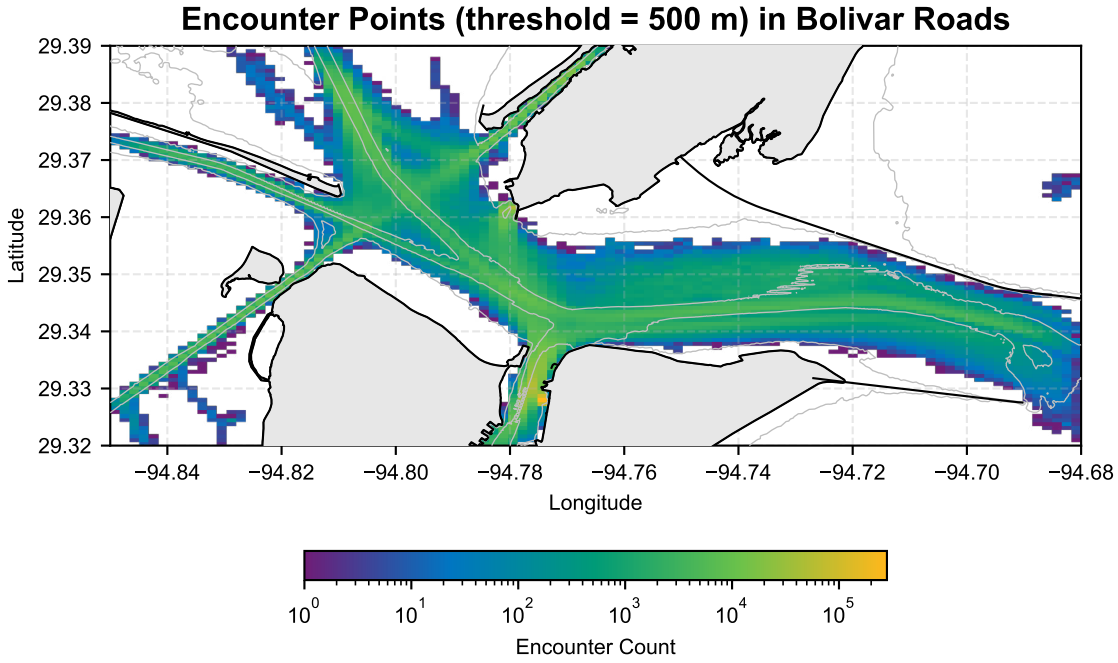


Figure 3.16: Map showing the density of encounters (ships within 500 m of each other) in Bolivar Roads.

Finally, Figure 3.17 shows the mean absolute drift angle in Bolivar Roads. Drift represents the difference between a vessel's heading and its movement over ground, reflecting the combined effects of environmental forces and steering corrections. Analysing drift helps identify locations where vessels must compensate for tidal currents or other environmental influences. The map shows higher drift values along the GIWW and on the ferry route between Galveston and Port Bolivar. Both routes have a north-south component, making vessels more susceptible to the predominantly east-west tidal flow. In contrast, vessels following the east-west alignment of the main ship channels experience lower average drift. Figure 3.18 further shows that drift variability is generally higher outside the main Houston Ship Channel, indicating that environmental or operational variability is greater in those areas.

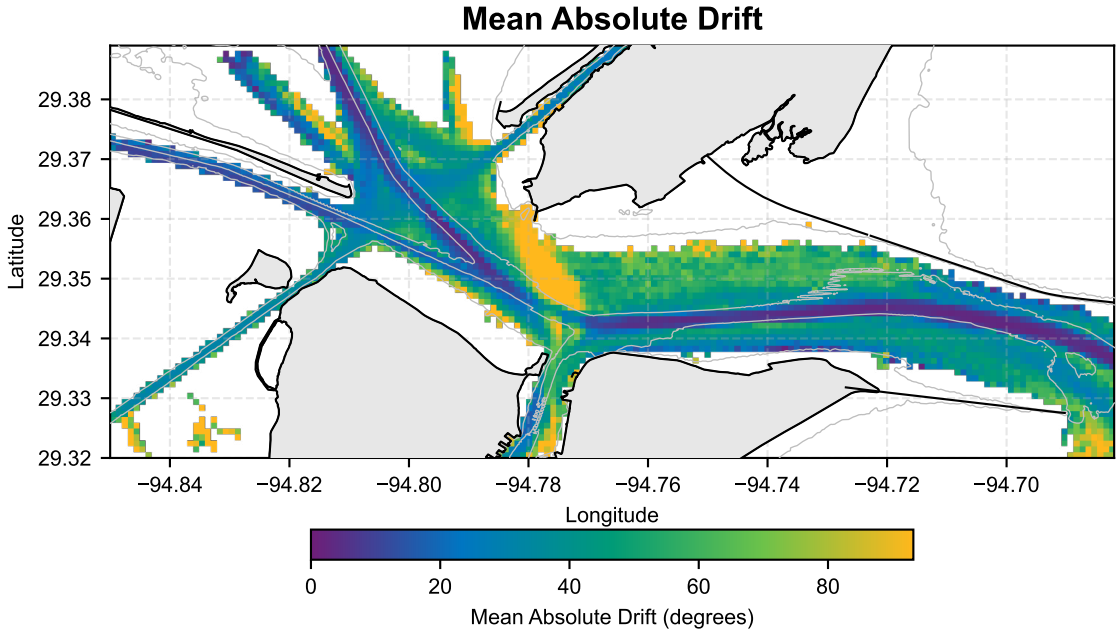


Figure 3.17: The mean drift in Bolivar Roads.

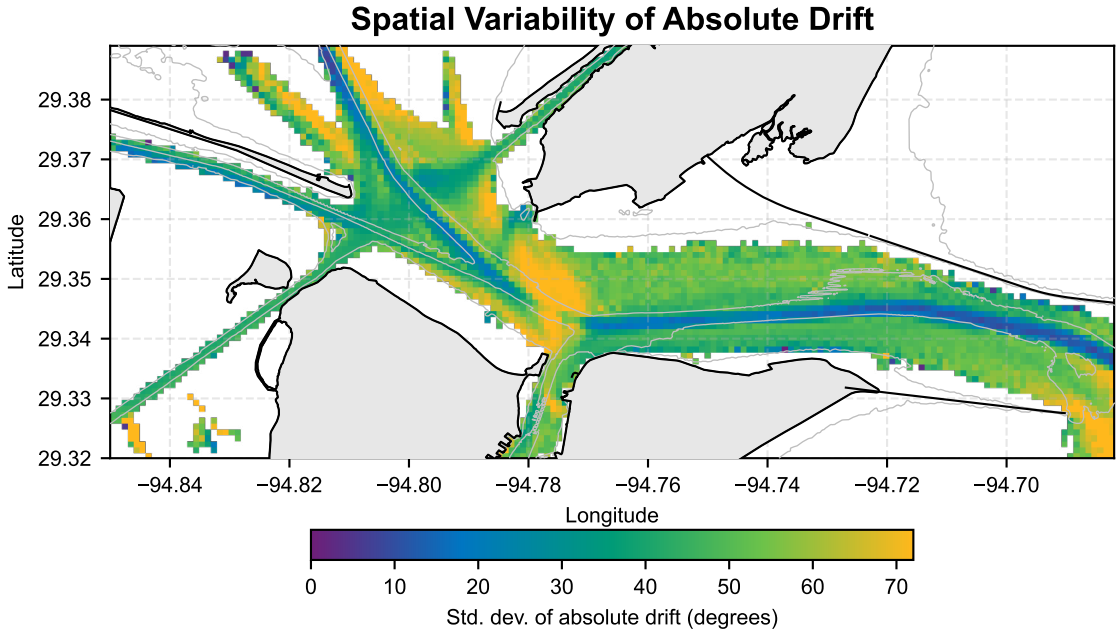


Figure 3.18: The spatial variation in drift in Bolivar Roads.

Together, the acceleration, turning, encounter, and drift patterns show that vessel behaviour in Bolivar Roads varies systematically across space and operational context. Manoeuvring complexity increases near intersections, ports, ferry terminals, and anchorage areas, while movements within the main ship channels are more stable and structured. In relation to the research question, these findings describe characteristic behavioural features observed in the AIS data, including speed adjustments, course alterations, interaction hotspots, and environmental compensation. These behavioural patterns form an important component of the existing traffic dynamics that a maritime traffic model of the proposed HSCGC site should be able to represent.

3.5. Conclusions and Model Requirements

This chapter identified from AIS data which existing characteristic vessel movements, traffic patterns, and behavioural features observed from AIS data at the proposed site of the HSCGC are observed that a maritime traffic model must be able to capture. The analysis reveals that the waters around the proposed site for the HSCGC consist of a highly structured, channel-constrained system. Most movements align with the Houston, Galveston, and Texas City Ship Channels, with typical transits through the study area lasting about 30-35 minutes and covering 7-10 km. A persistent north-south ferry between Galveston and Point Bolivar regularly crosses the dominant east-west traffic, and intersections east of Galveston and near the GIWW junctions produce locally complex situations.

The OD-patterns show that most traffic links the Gulf of Mexico with the ports of Houston, Galveston, and Texas City, while smaller, frequent ferries and ATBs routinely share confined space with deep-draft transits. These interactions are structurally embedded in the system rather than incidental. Speed and directional analysis reveals two common regimes near 5 and 10 knots: higher speeds correspond to ship channel transits by predominantly ocean-going ships, whilst lower speeds correspond to port approaches, crossings, and manoeuvring, and lower-speed barge traffic. Overall traffic intensity is relatively stable on a daily or monthly timescale, but statistically significant hourly variations in average speed indicate time-dependent operational conditions and interaction effects.

Manoeuvring complexity is spatially heterogeneous: acceleration and deceleration concentrate near channel entrances, approaches, and intersections. Turning is more frequent at lower speeds near junctions and anchorages. Encounter densities peak inside channels and at intersections, especially near ferry terminals. Vessels on north-south routes (specifically, the GIWW and Galveston-Port Bolivar ferry corridor) are more affected by east-west tidal drift than those aligned with the main channels, showing how environmental forces, infrastructure layout, and traffic interactions jointly shape behaviour.

Based on these findings, the maritime traffic model should be able to capture the following:

- Elements representative for describing and validating the traffic system:
  - The heterogeneity of vessel types (e.g., tankers, tugs, cargo vessels, ferries) and their differing operational characteristics and behaviours (Section 3.1).
  - The strong channel-constrained structure of vessel movements, including route-following behaviour within the Houston, Galveston, and Texas City Ship Channels (Section 3.2).
  - The relative magnitude of dominant OD flows between the Gulf of Mexico and the ports of Houston, Galveston, and Texas City. (Section 3.3)
  - The recurring cross-channel ferry movements between Galveston and Point Bolivar, including their intersection with east-west channel traffic (Section 3.3).
  - The presence of distinct operational speed regimes (e.g., clustered around approximately 5 and 10 knots) (Section 3.3).
  - The spatial variability in average direction and speed, reflecting channel alignment and crossing flows (Section 3.3).
  - The observed temporal stability in overall traffic intensity, combined with hourly variation in average vessel speed (Section 3.3).
- Elements representative for safety, interaction, and efficiency assessment:
  - Spatially varying manoeuvring complexity, particularly in areas with merging, diverging, and crossing traffic flows (Section 3.3).
  - Interaction effects that influence vessel speed selection, rather than assuming constant speeds (Section 3.3).
  - Speed adjustments (acceleration and deceleration) near channel entrances, port approaches, and intersections (Section 3.4).
  - Increased turning behaviour at lower speeds and near junctions, anchorage areas, and port approaches (Section 3.4).
  - Concentrations of ship-ship encounters within channels and especially at intersections and ferry terminals (Section 3.4).

In conclusion, the AIS analysis shows that traffic at the proposed HSCGC site is structured, heterogeneous, and interaction-driven. The identified movement patterns, speed regimes, interaction hotspots, and environmental effects form the empirical baseline against which a maritime traffic simulation model can be validated and upon which infrastructure impact assessments can be based. A model that does not reproduce these characteristic features would risk misrepresenting both the current operational reality and the potential consequences of modifications to the waterway system.


PILOT
TEXAS

4

Interviews

To define which vessel manoeuvring behaviours should be represented in ship-traffic models, input from experienced navigators is required in addition to analysis of traffic data. AIS data describes observed motions, but provides limited information on decision-making, prioritisation of actions, and the relevance of rare but safety-critical manoeuvres. Therefore, expert judgement is needed to identify which behaviours a model must capture and how these should be prioritised.

This chapter addresses Research Question 3: *"What are the characteristic vessel movements, traffic patterns, and behavioural features as observed under current conditions and predicted under future conditions by expert navigators that a maritime traffic simulation model must be able to capture?"* To answer this question, semi-structured interviews with local navigators were conducted to explore typical manoeuvres and behaviours seen in Bolivar Roads, as well as how introducing a gate complex in this area would change these manoeuvres and behaviours.

This information is necessary for two reasons: Firstly, local experts are able to identify navigational behaviours that may have been overlooked in the data analysis. Secondly, local experts are able to qualitatively identify how they would expect their behaviours and procedures to change compared to the current situation. These insights can be used to qualitatively assess the performance of maritime traffic simulation models when new obstacles are introduced, and are therefore relevant to determine to what extent data-driven maritime traffic simulations capture interactions with *new* maritime infrastructures.

For this interview, two experts associated with navigational organizations that serve Bolivar Roads were interviewed in separate sessions that lasted roughly one hour each. While the sample size of this interview is very small in terms of the number of respondents, it must be noted that the interviewed experts acted in their capacity of representatives of organizations that are heavily involved in maritime navigation, and were asked to not only share their own perspectives, but the general opinions of their organizations, and by extension their colleagues too. During the interview, notes[1] were taken by the interviewers. The respondents agreed to participate in this study voluntarily, and had the opportunity to review, and if desired correct, the interview notes shortly after their participation. These notes were further anonymized in order to remove identifying information, in order to prevent traceability to any individual participant. Interviews were held in-person by the author, assisted by a colleague from MARIN USA.

Interviews were held in a semi-structured fashion. This means that while the interview was structured around a set of themes, the conversation was allowed to flow naturally. The semi-structured interviews were guided by a list of questions (Appendix D), and a hydrographic chart (Figure 4.1) of the study area was available to the respondents to allow them to point out zones and routes of interest. These questions are formulated to specifically seek out insights into strategic planning of passages through Bolivar Roads, how interactions with other traffic are handled, and how respondents would anticipate their behaviour around a new gate structure would change. These themes are selected in order to correlate the findings from the interviews with the AIS data analysis in Chapter 3, as well as identify expected behaviours and navigational decisions around new obstacles (especially, the HSCGC) for which historic data is not available.

[1] The interview notes are not published in this work in order to protect the participants' privacy. Anonymized versions of these notes are available from the author or thesis committee upon request.

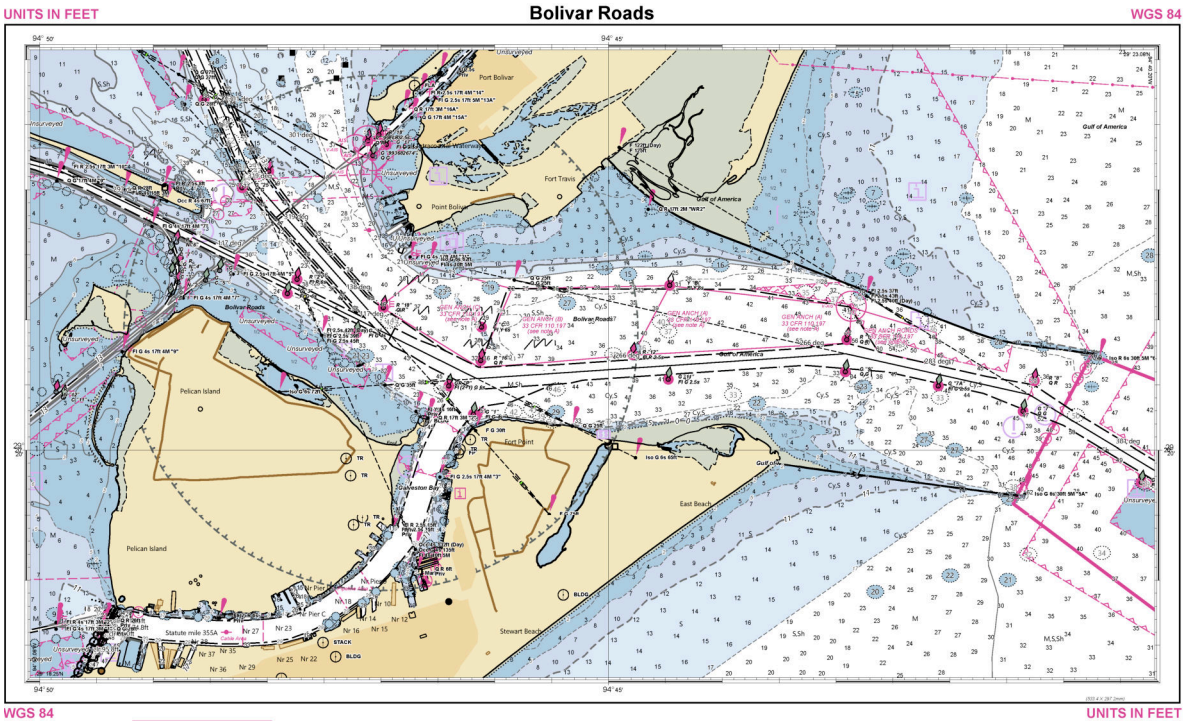


Figure 4.1: Hydrographic chart of Bolivar Roads. Generated using NOAA ENC Viewer (NOAA, 2025). A full-resolution version of this chart is available in Appendix E.

The interview notes were processed using a focused qualitative selection approach. After a full review of the notes, statements directly relevant to the design of data-driven maritime traffic simulations were systematically identified and extracted. The selection of these statements was guided by thematic relevance to the three key themes, particularly insights concerning practical decisions, operational constraints, and navigation around a potential new obstacle. This targeted filtering allows key insights from these interviews to be translated into concrete modelling requirements.

The following sections present the responses of the navigators, and their implications for data-driven maritime traffic simulations from a practical perspective. After that, the findings of the semi-structured interviews are used to formulate additional model requirements, extending the requirements found in Chapters 2 and 3.

4.1. Operational Context

The chart shown in Figure 4.1 depicts clearly marked ship channels, with sections of deeper water available extending beyond the ship channels towards the Gulf. Between the Bolivar Peninsula and Galveston Island, north of the Houston Ship Channel, the area includes a series of designated anchorages providing holding ground for vessels awaiting transit, orders, or refuelling. There is no Traffic Separation Scheme (TSS) within this particular section of the waterway. However, TSS arrangements are located further offshore and are not shown on this map. Vessel Traffic Service (VTS) coverage is available for information purposes only, and vessels are therefore expected to self-coordinate their movements and manoeuvres. Pilotage is compulsory for all vessels, except for US-flagged ships whose masters hold a valid pilotage certificate for their intended destination port.

4.2. Present-day Navigation and Interaction

First, respondents to the semi-structured interviews were asked to describe their perspective on navigating Bolivar Roads under present-day conditions. Respondents were also asked to describe the general principles and strategies they adhere to when manoeuvring, as well as the typical traffic conditions they experience during a passage of Bolivar Roads.

The respondents observed that vessels generally keep to the starboard side of Ship Channels. Furthermore, they also observed that ships with deep drafts (over 18 ft, approximately 5.5 m) are

constrained to the ship channels due to their draft. Ships with smaller drafts generally stay just outside the ship channels, as they are not constrained by their draft. If these ships are following the ship channels, they are generally able to and do navigate outside the marked ship channels if necessary to avoid ships that are constrained to the ship channels.

The respondents also emphasized that the Texas City Ship Channel is too narrow for two-way ship traffic. Therefore, ships bound for Texas City that are constrained to the channel by their draft must coordinate with each other before entering the Texas City Ship Channel. However, due to their smaller draft, barge traffic bound for Texas City is generally able to transit and encounter ships in this area.

Furthermore, respondents observed overtaking behaviour in the deep-water section South of the Houston Ship Channel in Bolivar Roads, as well as north of the channel further east (see Figure 4.2) In addition to that, merging and diverging behaviour is especially prevalent in the entry to the Port of Galveston, and the split between the Houston and Texas City Ship Channels. Respondents noted that these areas are especially prone to traffic congestion, leading to ships slowing down and altering course in order to avoid each other. In addition to that, respondents also noted that due to the generally limited space for navigation bottlenecks may appear from time to time.

The respondents noted that crossing ships sometimes meet starboard to starboard (instead of the standard port to port meeting prescribed by the COLREGs) in turning zones near the entrance to the Port of Galveston. Such encounters are usually coordinated between ships over radio communications. Such behaviour is also seen in the intersection between the GIWW and Houston and Texas City Ship Channels. Furthermore, the respondents noted that especially for ships to and from Galveston, these ships tend to accelerate and decelerate for the turn from the port to the Houston Ship Channel and vice versa. These ships might not be up to full speed when performing these turns.

Finally, the respondents noted that the behaviours of recreational ships and fishing vessels in the area are generally less predictable than the behaviours of commercial vessels. Therefore, commercial vessels generally stick to the ship channels, hoping that these smaller ships stay out of the way. The respondents noted that collisions are rare, but near-misses do happen from time to time.

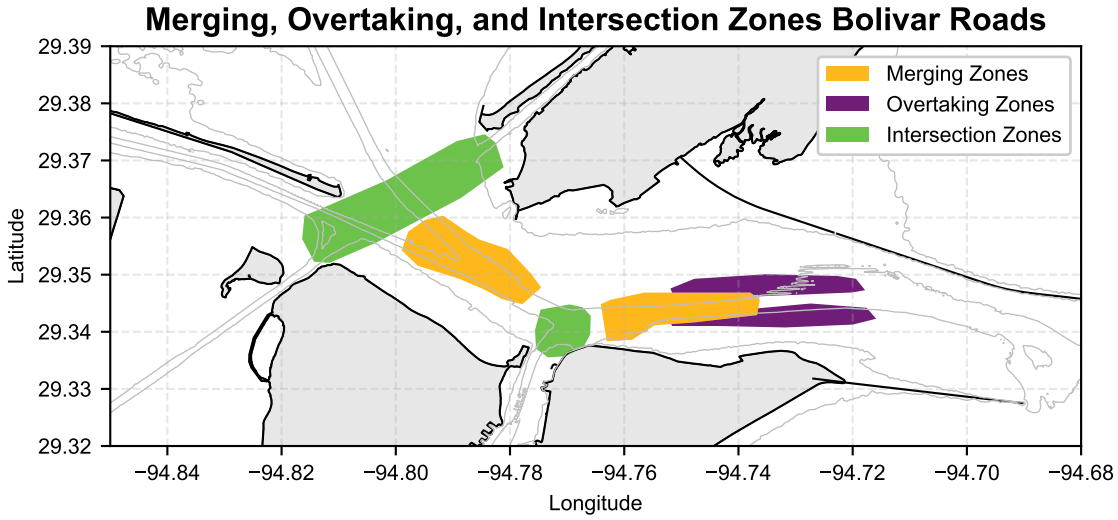


Figure 4.2: Zones where merging, overtaking, and intersecting behaviour is typically observed.

4.3. Introduction of a Gate

In addition to being asked to describe present-day navigation, respondents were also asked to consider a hypothetical gate complex in Bolivar Roads, and to describe how, in their opinion, traffic patterns and manoeuvring behaviour would change. This hypothetical is relevant to this research because this way, an indication of possible future behaviours around new obstacles, and therefore a (qualitative) description of how a maritime traffic simulation model should reflect these behaviours is obtained.

Since the gate complex is hypothetical, respondents' answers reflect expert judgement about expected behavioural adaptations rather than observed behaviour. The results of this section should

therefore be interpreted as qualitative indications of plausible operational constraints and safety-critical interactions that could be observed once a gate is realized. However, in order to quantify the expected effects of introducing a gate complex, a more extensive approach to determining the operational constraints and safety-critical interactions (for example, through human-in-the-loop simulations) should be taken. The latter is left for future research.

Respondents observed that upon approaching the gate inbound to the ports, they would have to slow down their ships in order to safely navigate through the gate. They also observed that this would be difficult to achieve on a flood tide, which means that inbound ships may be less able to slow down. Current and wind directions would push these ships towards the northern edge of the gate channel. Outbound ships' manoeuvring behaviour would likely be less affected by the gates.

Furthermore, respondents expressed concern about ship-ship encounters within the gate and the approaches to it, especially in single-channel designs. In their general opinion, head-on encounters within the gate complex would be undesirable, and also require ships to slow down and remain on the starboard side of the passage. When not encountering another ship, a passage along the centreline of the gate would be preferable. Moreover, the introduction of a gate would require ships to leave larger headways between them, which would reduce the space available in Bolivar Roads for overtaking manoeuvres.

Additionally, respondents noted that the introduction of the gate would also affect the traffic demand patterns in Bolivar Roads. This is because the gate would split the anchorages north of the Houston Ship Channel, which would render parts of this area unusable as an anchorage. This would reduce the traffic demand to and from this area.

4.4. Conclusions and Model Requirements

Based on the responses to the semi-structured interviews, a set of additional modelling requirements can be formulated for data-driven maritime traffic simulation in Bolivar Roads.

Firstly, while the general tendency for vessels to follow the ship channels is consistent with the AIS-based observations in Chapter 3, the interviews indicate that the model must explicitly represent two distinct categories of traffic: deep-draft vessels that are constrained to the marked channels and smaller-draft vessels that can and do deviate outside the channel when needed. This distinction affects both lateral position choices and interaction possibilities, and is therefore necessary to reproduce observed meeting and avoidance behaviour. In addition, the model must be able to reproduce overtaking manoeuvres in the deep-water sections identified in Figure 4.2, including the fact that available space for overtaking is operationally limited and can vary with local congestion.

Secondly, the interviews emphasize that vessel interactions in Bolivar Roads are not limited to simple head-on meetings. The model must be able to represent intersecting, merging, and diverging behaviour at the intersections between the GIWW, Houston Ship Channel, and Texas City Ship Channel, including the possibility of both port-to-port and coordinated starboard-to-starboard meetings in specific turning zones. To be credible in these areas, the model must also allow for local speed adaptation (acceleration and deceleration) associated with turning and with congestion-driven conflict avoidance.

Thirdly, the respondents highlight that not all traffic follows the same strict conventions that channel-bound ships must follow. Therefore, the model should include the presence of traffic with less predictable behaviour (notably recreational and fishing vessels), because this variability increases uncertainty in decision making and can trigger avoidance manoeuvres and near-miss situations even when collisions are rare.

Finally, the interviews provide explicit expectations for how behaviour and demand would change when a new obstacle such as the HSCGC is introduced. The model must be able to represent inbound speed reductions when approaching and transiting the gate, including reduced ability to slow down on a flood tide, systematic lateral bias (slight Northward movement) induced by current and prevailing wind effects, and constrained interaction manoeuvres in the gate and its approaches, where head-on encounters are undesirable and a centreline passage is preferred when unopposed. In addition to these behavioural effects, the model must support scenario-specific demand patterns. This is necessary because a gate fragments the anchorages North of the Houston Ship Channel, it should be possible to specify OD-patterns that differ from those inferred from historic AIS data in order to represent the resulting shift in anchorage use and associated traffic demand.

In conclusion, the interviews make explicit which behaviours and constraints a maritime traffic simulation model must represent to be credible for Bolivar Roads and for assessing a new obstacle such as the HSCGC. In addition to the channel-following structure seen in AIS, the model must capture draft-dependent manoeuvring options, overtaking and congestion in limited space, interaction-rich junction behaviour (including coordinated non-standard meetings and local speed adaptation), and uncertainty introduced by small-vessel traffic. The interviews further indicate that a gate affects both manoeuvring (speed reduction, lateral bias, and constrained encounters) and demand through anchorage fragmentation, so scenario-specific OD-patterns must be configurable rather than inferred solely from historic AIS. A model that cannot represent these requirements would risk reproducing tracks without capturing the operational logic that drives safety-critical interactions and infrastructure-induced change.

5

Model Requirements

This thesis aims to contribute to the development of a maritime traffic simulation model suitable as a design and assessment tool for maritime infrastructure changes in Bolivar Roads, specifically in the context of introducing a new obstacle such as the Houston Ship Channel Gate Complex (HSCGC). As such, the model must reproduce key features of present-day traffic and vessel behaviour derived from AIS data and expert knowledge, and remain applicable under modified environmental and infrastructural conditions for which historical data are inherently incomplete.

Across the literature review, AIS data analysis, and semi-structured interviews, a consistent theme emerges: a data-driven maritime traffic simulation capable of suitable for the design of maritime infrastructures requires a microscopic, interaction-driven representation of vessel motion, with the ability to reproduce observed traffic structure and manoeuvring behaviour, while also enabling the modelling of previously unseen scenarios (new obstacles, altered interactions, and altered demand).

This chapter consolidates the modelling requirements derived from the literature review (Chapter 2), AIS-based data analysis (Chapter 3), and expert interviews (Chapter 4), which are then used to formulate, test, and validate a data-driven maritime traffic simulation model in the next chapters. These requirements are presented as an indexed list, which are used for reference in the rest of this document.

5.1. Theoretical Foundations

Based on the findings of Chapter 2, the model should be grounded in the following theoretical paradigms:

R 1 Microscopic simulation.
The model must use a microscopic simulation paradigm, representing individual vessels and their movement decisions over time. This is required to capture local ship behaviour and ship-ship and ship-obstacle interactions, and to translate these local dynamics into emergent macroscopic outcomes such as traffic flow, capacity, density, and delay (Section 2.1).

R 2 Multi-agent representation.
The model must be a multi-agent simulation, with multiple vessel agents (in the sense that each agent controls one ship) simultaneously present, able to adapt their decisions based on the actions of other vessels and on the environment (Section 2.1).

R 3 Data-driven behavioural basis (AIS-grounded).
The model should be an extension of a data-driven maritime traffic model grounded in AIS data, in order to reproduce complex and potentially non-standard real-world behaviours without requiring an exhaustive set of explicit handcrafted rules or complete mathematical laws for manoeuvring (Section 2.2).

5.2. Generalizability to Design Scenarios

Since the focus of this study lies on the applicability of data-driven maritime traffic simulation models in the context of maritime infrastructure design, the following requirements address the need to have a model that generalizes to design scenarios:

R 4 Ability to simulate modified infrastructure and obstacles.
The model must be able to include modified environmental conditions, including new obstacles (e.g., the HSCGC) and altered navigable space, and must allow vessel agents to interact with these modifications (Section 2.3).

R 5 Safety mechanism for unseen situations (safe RL / safety filtering).
Because historical AIS data cannot contain behaviour for infrastructure that does not yet exist, the model must include a method to prevent unsafe outcomes (e.g., collisions with new obstacles) when agents encounter unseen conditions. Therefore, a suitable approach is a safety filter consistent with safe reinforcement learning concepts, capable of intervening when an intended action would lead to an unsafe state (Section 2.3).

R 6 Minimal-impact corrective control.
If a safety filter is used, the model must include a corrective method that modifies unsafe actions into safe actions while minimizing behavioural distortion. The intent is to preserve data-driven realism where possible, and only impose explicit corrections where necessary to maintain safety under new conditions (Section 2.3).

5.3. Baseline Representation

To be valid for Bolivar Roads, the model should reproduce the empirical baseline observed in AIS data (2024) and use these features as validation targets:

R 7 Heterogeneous vessel fleet.
The model should represent heterogeneity in vessel types (e.g., tankers, cargo vessels, tugs/barges, ferries) and reflect that these types have different operational characteristics, including speed regimes and manoeuvring patterns (Section 3.1), and draft constraints (Section 4.2).

R 8 Channel-constrained structure and route-following.
The model should reproduce the strongly channel-constrained structure of traffic, including route-following behaviour aligned with the Houston, Galveston, and Texas City Ship Channels (Section 3.1).

R 9 Dominant OD flows (macro structure).
The model should reproduce the relative magnitude of dominant origin-destination flows between the Gulf of Mexico and the ports of Houston, Galveston, and Texas City, since these flows structure the system and determine where interactions concentrate (Section 3.3). Furthermore, the model should represent the recurring Galveston-Point Bolivar ferry corridor, including its persistent crossing of the dominant East-West channel traffic and the interaction implications at crossing locations (Section 3.3).

R 10 Distinct speed regimes.
The model should reproduce the presence of distinct operational speed regimes (notably clustering around 5 and 10 knots), reflecting differences between deep-draft transits versus lower-speed approaches, crossings, and manoeuvring traffic,including barge traffic (Section 3.3).

R 11 Spatial variability in direction and speed.
The model should capture spatial patterns in average direction and speed driven by channel alignment, crossings, and local manoeuvring zones (Section 3.3).

R 12 Temporal structure: stable intensity, variable speeds.
The model should reflect the observed pattern that overall traffic intensity is relatively stable at daily/monthly scales, while hourly variability in average speeds exists and likely reflects time-dependent operational conditions and interaction effects (Section 3.3).

5.4. Interaction, Manoeuvring, and Safety

Beyond reproducing tracks, the model must capture behavioural mechanisms that drive interaction outcomes in constrained waterways:

R 13 Spatially heterogeneous manoeuvring complexity.
The model should represent that manoeuvring complexity is not uniform: acceleration/deceleration

and turning concentrate near channel entrances, port approaches, intersections, junctions, anchorages, and ferry terminals (Section 3.4).

R 14 Speed adaptation driven by interactions and context.
The model must allow vessels to adapt speed (including acceleration and deceleration) due to congestion and conflict avoidance, rather than assuming constant speed profiles (Section 3.4).

R 15 Turning behaviour linked to low speeds and junction areas.
The model should reproduce increased turning behaviour at lower speeds and near junctions, approaches, and anchorages, consistent with AIS-derived manoeuvring indicators (Section 3.4).

R 16 Ship-ship encounter hotspots.
The model should reproduce concentrations of ship-ship encounters inside channels and especially at intersections and ferry terminals, since these locations are critical for safety and capacity assessment (Section 3.4, Section 4.2).

R 17 Complex junction interactions.
The model must represent interacting flows at key intersections (GIWW, Houston Ship Channel, Texas City Ship Channel), including merging, diverging, and crossing behaviours rather than only head-on meetings (Section 3.4, Section 4.2).

R 18 Meeting conventions including non-standard coordination.
To be credible in turning zones, the model must allow for both standard and coordinated meeting behaviours, including the possibility of port-to-port and coordinated starboard-to-starboard meetings where operationally relevant (Section 4.2).

R 19 Overtaking in limited deep-water sections.
The model must be able to reproduce overtaking manoeuvres in deep-water sections where overtaking is feasible, while also representing that overtaking space is operationally limited and can vary with congestion (Section 4.1, Section 4.2).

R 20 Less predictable small-vessel behaviour.
The model should include traffic with less predictable behaviour (notably recreational and fishing vessels), because this increases uncertainty and can induce avoidance manoeuvres and near-miss dynamics even when collisions are rare (Section 4.2).

5.5. HSCGC Design Scenarios

To function as a design tool, the model must represent behavioural and demand changes induced by introducing a gate:

R 21 Gate approach and transit speed reduction.
The model must be able to represent inbound speed reductions when approaching and transiting the gate, including the operational reality that the ability to slow down can be reduced during flood tide (Section 4.3).

R 22 Environmental-force-induced lateral bias.
The model must be able to represent systematic lateral bias (e.g., slight northward movement) induced by current and prevailing wind effects, particularly for gate approaches and other relevant constrained segments (Section 4.3).

R 23 Constrained interaction logic near the gate.
The model must represent constrained manoeuvring and interaction near/within the gate and its approaches, including head-on encounters being undesirable, and a preference for centreline passage when unopposed (Section 4.3).

R 24 Configurable scenario-specific OD demand.
The model must support scenario-specific OD-patterns that can differ from those inferred from historic AIS, to represent demand shifts caused by infrastructure changes (Section 4.3).

Collectively, these requirements imply that model validity is not demonstrated solely by reproducing trajectories in aggregate, but by reproducing traffic structure, speed and manoeuvring regimes, and interaction hotspots and behaviours. A model that fails to reproduce these characteristic features risks misrepresenting both current operational reality and the safety and capacity consequences of infrastructure modifications such as the HSCGC.


LEG

6

Model Descriptions

The previous chapters lead to a set of requirements (presented in Chapter 5) for a data-driven maritime traffic model that is realistic enough to reproduce observed behaviour, but also flexible enough to test design changes. In this chapter, these requirements are translated into a mathematical model. By formulating the processes and principles of a data-driven maritime traffic simulation model (ShipNaviSim and extensions), this chapter is the first step in answering Research Question 4: *"To what extent does the selected data-driven maritime traffic model and its extensions meet the performance criteria for both historical realism and situational adaptability?"*, and Research Question 5: *"Based on their performance, which developed model extensions demonstrate the highest efficacy in capturing critical vessel behaviours and should be prioritized for further investigation in order to apply data-driven maritime traffic simulations as design tools for maritime infrastructure design?"*. The model and extensions developed in this chapter are then tested in Chapter 7.

This chapter is structured as follows: first, the basic principles that underlie the data-driven maritime ship traffic model are introduced. After that, the working of ShipNaviSim is briefly covered, and three extensions to ShipNaviSim are formulated: adding a multiple goals to the agent, correcting actions to safe actions with a safe reinforcement learning approach, and the combination of these three elements. The model and its extensions are formulated by loosely using the framework of the Overview, Design, and Detail (ODD) protocol as formulated by Grimm et al. (2006).

6.1. Basic Principles

This section will first provide an overview of the model's purpose, variables, and processes. After that, this section will provide the design principles and details of the model tested in this study, as well as extensions to that model.

6.1.1. Purpose

The purpose of the model is to understand how high-level traffic patterns (average speed, traffic flow, capacity) are affected by (new) obstacles in the study area. As established in Chapter 2, the model should capture ship-ship and ship-obstacle interaction to simulate microscopic manoeuvring behaviour, from which these macroscopic metrics emerge. These macroscopic metrics (aggregate speeds, traffic flows, and capacities) can then be used to assess the network-level impacts of local infrastructure interventions. Section 6.6 elaborates on these macroscopic metrics.

6.1.2. State Variables and Scales

The simulation describes the movements of individual ships (conform Requirement 1 and 2) within in a 2-dimensional plane. Latitudes and longitudes are transformed to northing (y) and easting (x) respectively, relative to the south-western corner of the study area. This allows the for spatial dimensions of the model to be represented in meters. Furthermore, space is considered to be continuous between $[0, x_{\max}]$ and $[0, y_{\max}]$.

The state of the simulation environment is described by the set of vectors $S_t = \{\mathbf{s}_t^1, \dots, \mathbf{s}_t^k\}$, where $k \in K$, the set of active ships. A single ship's state $\mathbf{s}_t^k = \langle x_t^k, y_t^k, v_t^k, h_t^k \rangle$ is a vector that corresponds to that ship's position, speed, and heading. Additionally, a ship is also issued a set of goals $G_k =$

$\{\mathbf{g}_k^1, \ldots \mathbf{g}_k^n\}$, with $\mathbf{g}_k^i = \langle x_g^i, y_g^i, dist_g^i \rangle$, the position (x_g^i, y_g^i) of the goal and the $(dist_g^i)$ range within which a goal is considered to be reached.

Furthermore, in the simulation environment a static set of obstacles for the ships is described by a set of polygons $p \in P$. Each polygon is closed and described by a series of points $P = (p_1, p_2, \ldots, p_i)$, with $p_i = (x_i, y_i)$, the coordinates of a polygon vertex in the simulation environment.

Finally, in order to maintain consistency with Pham et al. (2025)'s ShipNaviSim, and in order to maintain a balance between the temporal resolution of AIS data (minutes) and human reaction times (seconds), the model is formulated with a timestep dt = 10 s. The use of ShipNaviSim as a model basis also ensures that the model is grounded in AIS data as established by Requirement 3.

6.1.3. Processes

The model follows a closed-loop structure as shown in Figure 6.1. After initialization of the simulation, ships are added to the simulation environment. For each ship that is active, an observation is constructed from the simulation state S_t, which is used to determine the actions $(\mathbf{a}_t^k)$ of each active ship at that timestep. These actions are then used to update each ship's state. After that if a ship has reached an intermediate goal the goal position of the ship is updated, and if the ship has reached its final goal it is removed from the simulation. Then, ships that enter the simulation at $t + 1$ are added, and the process repeats itself until $t = t_{end}$

The construction of observations from the environment state, determining actions, updating ship states, and updating ship goals all happen in parallel for each ship. The rationale for this is that, in reality, each ship's navigator makes decisions independently. Navigation is not controlled by a central authority or the result of sequential decision-making. Each real-world decision is based solely on the navigator's local perception of the environment. Therefore, a parallel structure for these elements ensures that the modelled behaviour more closely mirrors real-world maritime navigational decision-making.

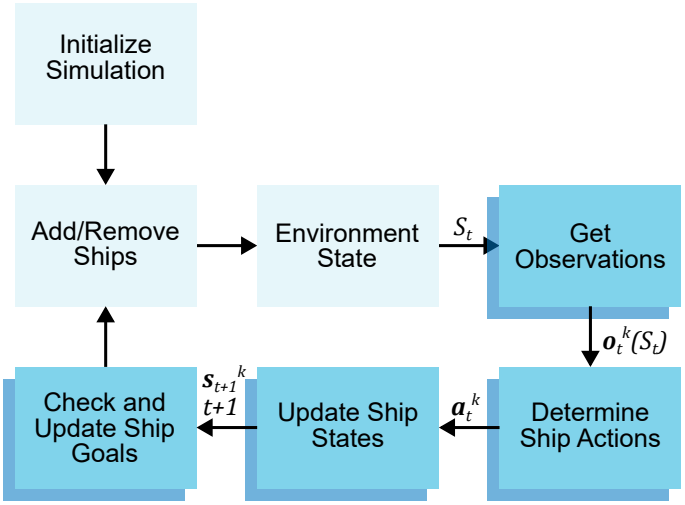


Figure 6.1: Overview of the simulation processes.

6.1.4. Initialization and Input

The simulation is initialized with a precomputed list of ships. For each ship, the following information is recorded in the initialization.

- The timestep t_{init}^k at which the ship must be inserted into the simulation
- The initial state at that timestep $\mathbf{s}_{t_{init}^k}^k$
- The set of goals G_k for the ship. This set contains at least one goal.

The initial conditions for each ship should be generated based on the origin-destination flows and the observed conditions at the starting points of tracks as shown in Section 3.3. Alternatively, synthetic OD-data can be used to generate the initial conditions for each ship, allowing model flexibility when

testing infrastructures under changing demand patterns, ensuring compliance with Requirement 9 and 24.

The geometries of the obstacles in Bolivar Roads are the main input for the different simulation runs. Since these obstacles are considered static, the simulation does not have dynamic inputs that are imposed on the state variables.

6.1.5. Adding and removing ships

Ships are added to the simulation when the simulation reaches the timestep $t = t^k_{init}$. Ships are removed from the simulation when they have reached the last goal in their set of goals G_k, when they leave the bounds of the simulation, or if they have been in the simulation environment for over 1000 timesteps, such that the simulation does not perpetually simulate ships that have gotten stuck.

6.2. ShipNaviSim

ShipNaviSim (Pham et al., 2025) is chosen as the baseline model for this study (see Section 2.4). When using ShipNaviSim, the action of each ship is determined by inferring an action using a policy (π_θ) trained on AIS data. This action is inferred from the by the following function:

$$\mathbf{a}^k_t = \pi_\theta(\mathbf{o}^k_t(S_t)) \tag{6.1}$$

Pure ShipNaviSim then directly executes the action $\mathbf{a}^k_t$, and updates the state of each ship and the environment accordingly. By inferring actions from real-world data, ShipNaviSim is used as a tool to theoretically capture the behaviours required by Requirements 8 - 20.

The policy π_θ is trained deterministically to minimize the difference between an action produced under the policy π_θ and the action from an expert demonstration ($\mathbf{a}^*$), such that the objective function in the training process is given by:

$$\min \sum_{\{\mathbf{o}^*, \mathbf{a}^*\}} (\pi_\theta(\mathbf{o}^k_t(S_t)) - \mathbf{a}^*)^2 \tag{6.2}$$

However, using ShipNaviSim without modification requires the assumption that the fleet of ships in the model is homogeneous, meaning it is assumed that all ships in the model have the same physical characteristics (length, width, draft) and manoeuvring behaviour. Consequently, the use of ShipNaviSim as a base model means that the requirement that ship type-specific behaviours and manoeuvring constraints be captured in the model cannot be met without significant modifications to ShipNaviSim, which is contrary to Requirement 7. However, modifying the ShipNaviSim project to account for this is left outside the scope of this study.

6.2.1. Observation Construction

The observation $\mathbf{o}^k_t$ from the perspective of the ship k of the state S_t of the environment is the combination of:

- The current and past H own-ship states ($\mathbf{s}^{k=k}_t, \mathbf{s}^{k=k}_{t-1}, \ldots \mathbf{s}^{k=k}_{t-5}$).
- The current and past H states of the n nearest ships at timestep t.
- The goal position of the ship.
- The distances to the nearest n ships.

Any information that is missing or unavailable is padded with a value of 0 in order to maintain consistent observation size. The structure of the observation aligns with the structure of observations used in ShipNaviSim.

6.2.2. Action Space

The action taken by the agent is in delta action space. This means that the agent determines its own change in position and heading between two time steps. The use of delta action space also means that the action space of the agent is continuous. An action taken at time t contains a change in position (dx, dy), and a change in heading (dh). Hence, an action is described by the following vector:

$$\mathbf{a}_t^k = \langle dx_t^k, dy_t^k, dh_t^k \rangle$$

This way, the state of each ship is updated every timestep, such that:

$$\mathbf{s}_{t+1}^k = f(\mathbf{s}_t^k, \mathbf{a}_t^k) = \langle x_t^k + dx,\ y_t^k + dy,\ v_{t+1}^k,\ h_t^k + dh \rangle \tag{6.3}$$

With the speed (v) of the ship as a function of the ship's change in position across a single timestep (dx, dy), such that:

$$v_{t+1}^k = \frac{\sqrt{{dx_t^k}^2 + {dy_t^k}^2}}{dt} \tag{6.4}$$

ShipNaviSim outputs each dimension of an action between $[-1, 1]$ using a $\tanh$ activation function, which are then scaled to the action spaces. ShipNaviSim specifies these as functions of the world bounds (i.e., the maximum x and y coordinates of the simulation) as follows:

$$dx \in \left[-\frac{1}{20}x_{\max}, \frac{1}{20}x_{\max}\right]$$
$$dy \in \left[-\frac{1}{20}y_{\max}, \frac{1}{20}y_{\max}\right]$$
$$dh \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

6.3. Extension: Multi-Goal ShipNaviSim

To encourage the ship control agent to adhere more closely to designated navigation channels (as specified by Requirement 8), ShipNaviSim is extended with a multi-goal control formulation. Rather than navigating toward a single terminal destination, the ship agent is initialized with an ordered set of intermediate goals G_k, such that $|G_k| > 1$, of which only one is active at any given time. Intermediate goals are given as input to the simulation during the initialization of the list of ships.

The motivation for this extension is that real-world maritime navigation is inherently waypoint-based: vessels follow planned routes of successive waypoints rather than continuously optimizing toward a distant terminal objective. Representing navigation as a sequence of local goals allows the control agent to operate on short-horizon objectives, improving path adherence and stability while remaining consistent with the available trajectory information.

During each simulation step, the agent's position is evaluated relative to the active goal. Once the ship the specified range of the goal, the active goal is advanced to the subsequent element in G_k. This process continues sequentially until the final goal is reached, at which point the ship is removed from the simulation.

It must be noted that ShipNaviSim does not include the ability to train an agent in a multi-waypoint setting. Given that the model structure of ShipNaviSim includes goal and position information (Pham et al., 2025), it is possible that the model learns to correlate behaviours to the distance-to-goal, which may result in different behaviours from ships nearer to goals and ships that are en-route. Therefore, it is possible that a ship controlled by ShipNaviSim will deviate from en-route behaviours near intermediate goals, even though this may not be representative of actual transits near waypoints.

6.4. Extension: Obstacle Avoidance with a Safety Layer

Furthermore, in order to ensure that the ship control agent avoids *both* new and existing obstacles, a safety layer using the framework of safe reinforcement learning is introduced. This approach uses raycasting to detect obstacles, followed by a decision and correction mechanism to determine whether the action provided by ShipNaviSim should be altered. This framework is illustrated in Figure 6.2.

By introducing and applying the correction mechanism, Requirement 4 and 5 are satisfied. Furthermore, through tuning of the parameters of this safety layer, Requirements 21 - 23 should be covered. Nonetheless, it must be noted that the formulation presented here does not involve ship-ship interaction in the safety layer, which means that Requirement 23 is only partially covered.

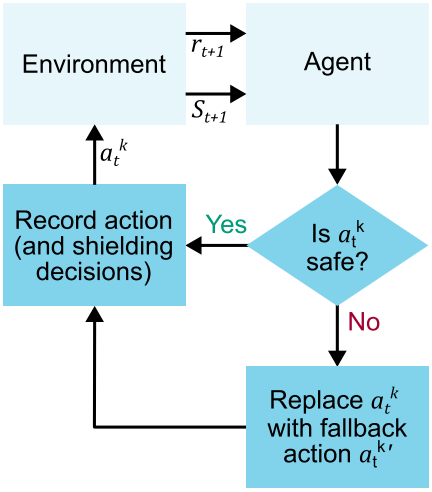


Figure 6.2: Decision mechanism for obstacle avoidance with a safety layer using the safe reinforcement learning paradigm.

6.4.1. Raycasting

The agent performs obstacle detection is done by means of raycasting. By casting rays originating from the ship's position along regular intervals, the agent can check whether there are obstacles along the ray within a set threshold $dist_{safe}$. This approach is analogous to the approach by Dawood et al. (2024), who used lidar data (a real-world form of distance detection through raycasting) to determine the distance between their agent and an obstacle. This way, ship domain is approximated. The distance between a ship and an obstacle can then be sampled by measuring along the rays cast from the ship as shown in Figure 6.3.

The advantage of this approach is that by sampling the distance between the ship and any obstacles, the ship can be aware of its proximity to obstacles. The rays $m \in M$ are spaced at regular intervals of φ degrees. The rays that are cast from the ship are defined relative to the ship's heading.

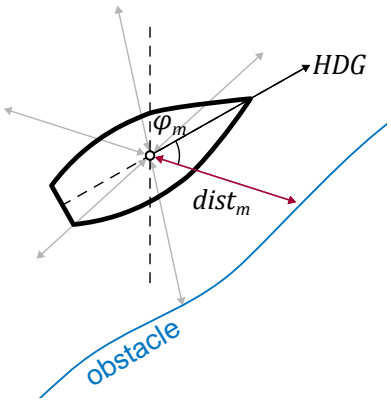


Figure 6.3: Definition of distance along a ray cast from the ship's position.

6.4.2. Safety Shield

After an initial action is determined, the safety shield checks whether this action is safe. For this safety check, the action determined by the ship is linearly extrapolated for 10 timesteps. An action is

considered safe when:

1. The vector $\mathbf{s}_t^k + 10\mathbf{a}_t^k$ does not cross any obstacle boundaries.
2. At each intermediate position $\mathbf{s}_t^k + n\mathbf{a}_t^k$, $n \in \{1, \dots, 10\}$, none of the rays along the ship find an obstacle within $dist_{safe}$ m.

If the action inferred from the ShipNaviSim policy π_θ is safe, it is passed directly to the simulation. However, if the action is unsafe, the correction mechanism is applied.

6.4.3. Correction of Unsafe Actions

The correction mechanism aims to find a minimal safe corrective action $\mathbf{a}_t^{k\prime}$ for the initial action $\mathbf{a}_t^k$ as determined by inference using the ShipNaviSim policy (π_θ).

The idea minimal safe corrective action is the action that deviates minimally from the original action $\mathbf{a}_t^k$, while guaranteeing to be safe. Therefore, the minimal corrective action is generated by solving an optimization problem (Alshiekh et al., 2017). By determining a minimal safe corrective action, the agent stays as close to the original behaviour learned in the training of the policy π_θ as is reasonably possible, as required by Requirement 6.

In order to determine the corrected safe action $\mathbf{a}_t^{k\prime}$, a Model Predictive Control (MPC) problem is activated once an action produced by the BC agent is determined to be unsafe. Formally, this model should guarantee that unsafe actions are never selected. This model at timestep t with a prediction horizon N timesteps in the future can be formulated as (Dawood et al., 2024):

$$\min \quad Z = ||\mathbf{a}_t^{k\prime} - \mathbf{a}_t^k||_{R_0}^2 + \sum_{i=1}^{N} ||\mathbf{a}_{t+i}^k{}'||_R^2 \tag{6.5}$$

Subject to: (6.6)

$$\mathbf{s}_{t+1}^k = f(\mathbf{s}_t^k, \mathbf{a}_t^k) \qquad \forall t \in \{t, t+N\} \tag{6.7}$$

$$\mathbf{s}_t \in \mathcal{S}_{safe} \qquad \forall t \in \{t, t+N\} \tag{6.8}$$

$$\mathbf{a}_t^{k\prime} \in \mathcal{A}_{safe} \qquad \forall t \in \{t, t+N\} \tag{6.9}$$

$$dist(\mathbf{s}_t^k, p) \geq dist_{safe} \qquad \forall p \in P, t \in \{t, t+N\} \tag{6.10}$$

The weighted euclidean norm is represented by $||\cdot||_R^2$, which is defined as $||\mathbf{x}||_R^2 = \mathbf{x}^\top R\mathbf{x}$, where R is positive definite matrix that contains the weights.

The objective function (6.5) of this problem aims to minimize the weighted (by R_0 magnitude Z of the difference between corrective action decided by this problem ($\mathbf{a}_t^{k\prime}$) and the original action produced using the BC policy ($\mathbf{a}_t^k$), plus the weighted (by R) magnitude of future corrective actions ($\mathbf{a}_{t+i}^k{}'$). Constraint 6.7 ensures that the actions and states visited during the lookahead time t are consistent with the model of the ship's dynamics ($f(\mathbf{s}, \mathbf{a})$), as defined in Equation 6.3. Constraints 6.8 and 6.9, ensure that the model can only select states and actions that are within the sets of safe states ($\mathcal{S}_{safe}$) and actions ($\mathcal{A}_{safe}$), respectively. Finally, constraint 6.10 ensures that at any timestep, the distance between the ship and an obstacle is greater than the minimum threshold distance $dist_{safe}$.

The non-linear nature of this problem as formulated by Dawood et al. (2024) makes the problem computationally hard to solve, as formally guaranteeing an action is within the safe set of actions and reaches a safe set of states would require an infinite amount of actions to be checked, since space is considered continuous in this model.

Raycasting provides a method to simplify the problem. By considering distance between a ship and the obstacles along the rays as soft constraints, a linearized version of the problem can be formulated as follows:

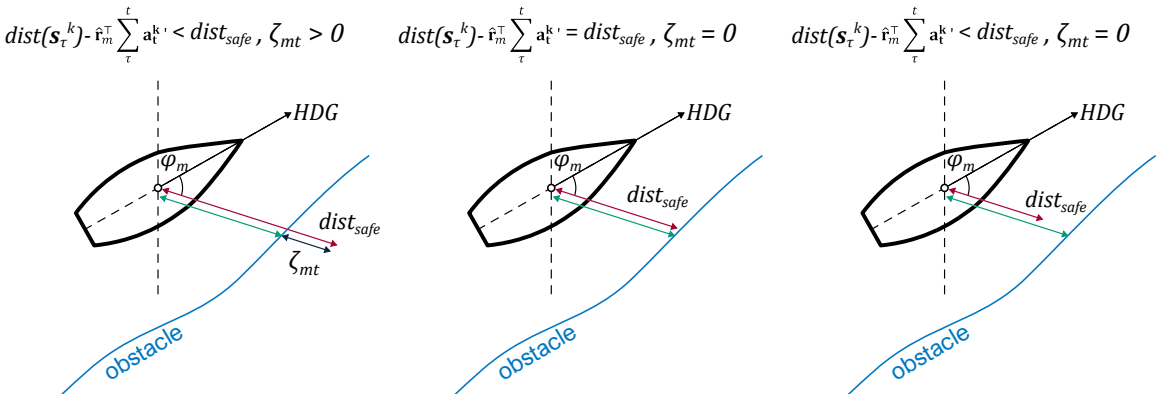


Figure 6.4: The three cases for the distance between a ship and an obstacle along a ray: distance is shorter than d_{safe}, distance is exactly d_{safe}, or distance is greater than d_{safe}

$$\min \quad Z = ||\mathbf{a}_\tau^{k\prime} - \mathbf{a}_\tau^k||_{R_0}^2 + \sum_{t=\tau+1}^{\tau+N} ||\mathbf{a}_t^{k\prime}||_R^2 + \sum_\tau^{\tau+N} \sum_m^M \lambda_{mt}\zeta_{mt} \tag{6.11}$$

Subject to: (6.12)

$$\mathbf{s}_{t+1}^k = f\left(\mathbf{s}_t^k, \mathbf{a}_t^{k\prime}\right) \qquad \forall t \in T \tag{6.13}$$

$$dx^{\min} < dx_t^k < dx^{\max} \qquad \forall t \in T \tag{6.14}$$

$$dy^{\min} < dy_t^k < dy^{\max} \qquad \forall t \in T \tag{6.15}$$

$$dh^{\min} < dh_t^k < dh^{\max} \qquad \forall t \in T \tag{6.16}$$

$$dist_m(\mathbf{s}_\tau^k) - \hat{\mathbf{r}}_m^\top \sum_\tau^t \mathbf{a}_t^{k\prime} + \zeta_{mt} \geq dist_{safe} \qquad \forall m \in M, t \in \{\tau, \dots, \tau+N\} \tag{6.17}$$

$$\zeta_m \geq 0 \qquad \forall m \in M \tag{6.18}$$

The objective function of this problem (6.11) minimizes the weighted (by R_0) norm of the difference between the original action $\mathbf{a}_t^k$ and the corrective action $\mathbf{a}_t^{k\prime}$, as well as the control effort over future steps ($\sum_{t=\tau+1}^{\tau+N} ||\mathbf{a}_t^{k\prime}||_R^2$) and the penalties incurred by using slack on the rays radiating from the ship's centre. Constraint 6.13 ensures that the state and actions are continuous with each other across sequential timesteps. In this constraint, $f(\mathbf{s}_t^k, \mathbf{a}_t^{k\prime})$ is defined as in equation 6.3. Next, constraints 6.14, 6.15, and 6.16 ensure that the elements of the corrective action do not exceed physical limits of the action space specified in Section 6.2.

Constraint 6.17 ensures that if the distance along a ray m from the ship to an obstacle after the action $\mathbf{a}_t^{k\prime}$ is smaller than $dist_{safe}$, a slack penalty ζ_m (weighted by λ_{mt}) is applied in the objective function. This constraint is based on the assumption that the heading of the ship rotates very little (i.e., using a small angle approximation) between timesteps, and that the edge of the obstacle that meets the ray is locally smooth. The full derivation of this constraint can be found in Appendix F.

The effect of this constraint is that at each timestep a penalty is applied for selecting an action that leads to the ship getting close to an obstacle. By taking the sum over $\{\tau, t\}$, the model accounts for the cumulative actions at timesteps $\tau, \tau+1, \tau+2$, etc. This way, this constraint remains linearized around $\mathbf{s}_t^k$. This way, the continuous nature of space in the model is addressed, and the model will always produce at least one action.

Finally, constraint 6.18 ensures that this slack penalty is non-negative, which implies if that the distance between the ship and the obstacle along ray m after taking the action $\mathbf{a}_t^{k\prime}$ is greater than or equal to $dist_{safe}$, then $\zeta_m = 0$. This avoids pushing the ship further away than the ray is long, which would cause the second term of the objective function to dominate over the first term. This idea is illustrated in Figure 6.4.

Since this problem is quadratic and can be solved quickly, using this version of the problem avoids a bottleneck in the performance of the model at runtime. The solution is guaranteed to be unique if

R is set to be a positive-semidefinite matrix. Combined with the safety shield, this problem approach filters unsafe actions produced by the ShipNaviSim agent, and thereby prevent ships from running into obstacles. When the optimization is complete, the first action in the sequence found by the optimization ($\mathbf{a}_\tau^{k\prime}$) is used as a replacement for the original action $\mathbf{a}_t^k$.

6.5. Extension: Multi-Goal with Safety layer

Finally, the combination of multi-goal ShipNaviSim with a safety layer is considered. This model integrates the route-level guidance provided by the multi-goal formulation (Section 6.3) with the local constraint enforcement of the safety layer (Section 6.4). Since these components operate at different levels of the navigation process, their combined effect is not trivial and is therefore evaluated explicitly. The resulting model reflects a layered navigation structure in which waypoint-based route guidance is complemented by local safety constraints, enabling the agent to follow designated channels while respecting operational safety requirements.

6.6. Model Outputs

The outputs from the model can be categorized into two categories: microscopic outputs and macroscopic outputs. This section first describes first the microscopic outputs, followed by the macroscopic outputs.

6.6.1. Trajectories and Control Decisions

In order to assess the realism of the trajectories determined by model agents, the model outputs the states, proposed actions, executed actions, and magnitude of any corrective action (if applicable) for each ship at each timestep. This informations allows for the comparison between the model-generated trajectories and historical data, from which it can be determined whether the models are able to accurately recreate historical manoeuvring behaviours. In addition to that the magnitude of the difference between the proposed and corrective action gives insight into what the influence of the safety layer is on the decisions of the model agent, as this value quantifies the deviation between the proposed and corrective actions. Greater control magnitudes indicate that the action as determined by the safety layer deviates more from what the ShipNaviSim agent determined.

6.6.2. Traffic Metrics

Since the purpose of the model is to determine aggregate metrics that emerge from microscopic behaviours, the macroscopic output of the model is determined by aggregating the states of the ships over time and space. These macroscopic metrics can then be used to determine the network-level effects of introducing new infrastructure on maritime traffic patterns.

For the sake of aggregation, gates and zones are defined. These gates and zones are necessary to define location-specific metrics, as aggregating traffic behaviour across the whole simulation environment would not reveal where certain changes in e.g., traffic flow and capacity will take place. A gate (L) is defined as the line segment between two points. Gates are placed at the entries and exits of zones. A zone (Ω) is a control volume that describes a section of a waterway, and is defined as a polygon enclosing a region of the waterway (Wang et al., 2025).

Space-mean speed per zone

In zones, the instantaneous space-mean speed of traffic in that zone at timestep t can be computed as:

$$u_{\Omega t} = \frac{N_{\Omega t}}{\sum_{i=1}^{N_{\Omega t}} \frac{1}{v_t^k}} \tag{6.19}$$

This metric captures the average speed of all ships occupying the zone at a given instant, weighting slower-moving ships more strongly because they spend more time within the control volume. The harmonic mean formulation arises directly from the definition of space-mean speed as total distance travelled divided by total travel time, and is therefore the appropriate aggregation for spatially distributed traffic states.

Time-aggregated space-mean speed
By weighting the speeds by timestep dt, the time-weighted harmonic space-mean speed of traffic in a zone can be calculated as:

$$u_{\Omega,agg} = \frac{\sum_t^{\Delta T} N_{\Omega t} dt}{\sum_t^{\Delta T} \sum_{i=1}^{N_{\Omega t}} \frac{dt}{v_t^k}} \tag{6.20}$$

This aggregation accounts for temporal variability in traffic conditions by weighting each instantaneous observation by the total ship-time spent in the zone. As a result, periods with higher occupancy or congestion contribute proportionally more to the aggregate speed estimate, preventing bias toward short-lived free-flow conditions.

For constant timesteps, this expression reduces to:

$$u_{\Omega,agg} = \frac{\sum_t^{\Delta T} N_{\Omega t}}{\sum_t^{\Delta T} \sum_{i=1}^{N_{\Omega t}} \frac{1}{v_t^k}} \tag{6.21}$$

This reduced form preserves the physical interpretation of space-mean speed while allowing efficient computation over long simulation horizons.

Density in a zone
The density of traffic in a zone ($\rho_{\Omega t}$) at timestep t is a function of the number of ships in that zone ($N_{\Omega t}$) and the area of that zone ($A_{\Omega t}$):

$$\rho_{\Omega t} = \frac{N_{\Omega t}}{A_{\Omega t}} \tag{6.22}$$

Density provides a measure of how crowded a waterway segment is at a given time and serves as a key macroscopic state variable linking individual ship movements to aggregate flow behaviour. The choice of zone boundaries, and hence $A_{\Omega t}$, is a modelling decision that reflects the desired spatial resolution and analytical focus. Expressing density per unit area is appropriate for two-dimensional waterways and allows comparison across zones of different sizes.

Flow across a gate
The flow of traffic across a gate (q_L) can be calculated by dividing the number of gate crossings (N_{cross}) across a time window ΔT:

$$q_L = \frac{N_{cross}}{\Delta T} \tag{6.23}$$

Gate-based flow directly measures throughput at strategic locations such as channel entrances, exits, or bottlenecks. This formulation is independent of zone geometry and provides a robust, observable quantity that aligns closely with real-world traffic counting practices.

Flow within a zone
Furthermore, the flow of traffic in a zone (q_Ω) can be quantified by the fundamental relation of traffic flow:

$$q_\Omega = \rho_\Omega u_\Omega \tag{6.24}$$

This relation links density and space-mean speed to flow, enabling internal zone performance to be inferred without relying on discrete crossing events. It provides a consistent macroscopic description of traffic dynamics and allows comparison with classical fundamental diagrams.

Capacity
Finally, the capacity of a region can be determined by determining the empirical maximum flow observed when traffic demand exceeds supply during the simulation:

$$q_{\Omega,\max} = \max_{\Delta T}, q_{\Omega}(\Delta T) \tag{6.25}$$

Capacity is defined operationally as the highest sustainable flow produced under simulated conditions, reflecting the combined effects of geometry, interactions, and behavioural constraints. Given that waterways may not yet be at capacity under present-day traffic patterns, to effectively estimate capacity the amount of traffic should be increased such that the simulation starts exhibiting signs of oversaturation. Specifically, the amount of ships that is simulated should be increased until “the waterway is so crowded that overtaking is almost impossible, and vessels form into groups with almost equal speed” (Fujii and Tanaka, 1971, p. 543). Estimating capacity empirically avoids imposing a priori assumptions and ensures consistency with the modelled dynamics.

6.7. Conclusions

With the processes and principles of ShipNaviSim and its three extensions now formulated, this chapter establishes the modelling basis needed to address Research Question 4 and 5. In relation to Research Question 4, the mathematical model specifies how historical realism is pursued through AIS-driven action inference and trajectory-level outputs, while situational adaptability is enabled by making demand patterns and obstacle configurations explicit model inputs, and by introducing a safety layer that can enforce obstacle avoidance when the simulated environment deviates from the training context. In relation to Research Question 5, the extensions are formulated such that their contribution can be assessed separately: the multi-goal formulation is introduced to improve route adherence and thereby strengthen the reproduction of channel-following behaviour, whereas the safety layer is introduced to maintain safe behaviour around both existing and newly introduced obstacles while deviating minimally from the original policy actions. The combined model integrates these elements into a layered navigation structure in which waypoint-based guidance is complemented by local safety constraints.

In the next chapter, these formulations are evaluated against empirical data and test cases to determine to what extent the baseline model and each extension meet the performance criteria, and which extensions provide the most effective improvement in capturing critical vessel behaviours for application in maritime infrastructure design.


PILOT
GALVESTON
NEPTUNE

7

Experiments

Using the models formulated in Chapter 6, this chapter presents the experimental evaluation of the developed data-driven maritime traffic model and its extensions. The experiments are intended to assess both the baseline model and the added value of the extensions, and to provide a basis for judging their practical usefulness and their potential application in the design of maritime infrastructures.

Therefore, this chapter aims to answer the questions: *"To what extent do data-driven maritime traffic models and extensions to these models possess the ability to capture the characteristic vessel movements, traffic patterns, and behavioural features?"* (Research Question 4), and *"What improvements can be made to apply these data-driven models as a design tool for maritime infrastructures?"* (Research Question 5). These questions are answered by means of assessing model performance through experimental model runs.

This chapter is structured as follows: First, the metrics used to assess the models are introduced. After that, the model testing experiment method is described, followed by the results of the tests. These results are then used to determine how well data-driven maritime traffic models and their extensions capture vessel behaviour, and whether the extensions proposed in Chapter 6 improve ability of data-driven maritime traffic models to be used as a design tool for maritime infrastructures.

7.1. Model Assessment Metrics

Firstly, Goal-Conditioned Average Displacement Error (GC-ADE) is used to compare the trajectories of agents to the trajectories of historical observations (Pham et al., 2025). This metric aims to assess whether the model is able to replicate the observed channel-constrained structure (Requirement 8).

Given an agent i and a historical trajectory j, with trajectory lengths T_i, T_j, respectively, the GC-ADE is the point-wise comparison between the trajectories of the agent and its corresponding historical track given by:

$$\frac{1}{\min(T_i, T_j)} \sqrt{\sum_{t=1}^{\min(T_i,T_j)} (x_t^i - x_t^j)^2 + (y_t^i - y_t^j)^2} \tag{7.1}$$

Furthermore, the kinematic metrics drift, curvature, and acceleration are determined from the trajectories produced by the agents. By comparing these metrics to the kinematic behaviour observed in real-world data, the extent to which the models imitate the behaviour and manoeuvrability of real-world ships is assessed (Pham et al., 2025). This way, the performance of the agents with respect to Requirements 10, 11 13, and 14 is assessed. Together with GC-ADE, these metrics are used to assess how well an agent is able to reproduce historical trajectories.

In addition to that, the model extension described in Section 6.4 introduces ship-obstacle interaction using a safety layer over ShipNaviSim. In order to illustrate the influence of this safety layer (as required by Requirement 5, the magnitude of the actions ($||\mathbf{a}_t^k||$) produced by different models and the magnitude of the corrections ($||\mathbf{a}_t^k - \mathbf{a}_t^{k\prime}||$ that are applied by the safety layer are also calculated.

This way, the degree to which the safety filter performs only minimal intervention (as required by Requirement 6) is quantified. Finally, the number of entries into obstacles is also recorded to determine whether the safety layer is indeed able to prevent the agent from choosing actions that lead to the agent entering into an obstacle, again addressing Requirement 5.

In addition to these quantitative metrics, the behaviour of ships is also assessed using face validation using a visualisation module for the models. Face validation was carried out by visually checking the model outputs to see whether the simulated ship behaviour looked realistic. In particular, attention is given to the consistency of ship headings (especially whether they align with the direction of travel, and whether headings change in small, incremental steps), the realism of ship speeds (do model ship speeds match historical ship speeds?), and whether vessels manoeuvre sensibly when encountering obstacles (do ships anticipate and move around obstacles with consistent heading and directions?). These checks help to confirm that the simulated trajectories and navigation patterns are in line with what would be expected in real-world conditions.

7.2. Set-Up

The experiments are based on a single test case in which one month of real-world vessel traffic data (November 2024) is reconstructed. Using observed traffic allows the models to be grounded in realistic operating conditions and makes it possible to directly compare simulated trajectories with historical vessel tracks. November 2024 was selected arbitrarily from the months that show traffic count patterns that align with the annual mean (see Section 3.3, Appendix C). The simulation is limited to a single month of tracks due to computational limits, under the assumption that the daily and monthly traffic patterns do not vary significantly from each other (see Section 3.3. In total, 6601 tracks were simulated.

This way, the performance of the data-driven maritime traffic simulation models is tested with respect to real-world conditions observed in Bolviar Roads, in order to determine whether the models presented in Chapter 6 meet Requirements 8 - 20, establishing how well the models are able to represent real-world behaviours observed from AIS data.

The simulation environment of Bolivar Roads is derived from a local flat-earth approximation in the South-Western corner of the study area (latitude, longitude = 29.32°, -94.85°, using $R_{Earth} = 6371$ km). The environment is bounded by the study area (longitiudes between -94.85° and -94.68°, latitudes between 29.32° and 29.39°).

Existing coastline geometries are used to assess the obstacle-avoidance capabilities of the safety layer. For this purpose, a simplified coastline is adopted as obstacle geometry. This geometry, shown in Figure 7.1, is derived from bathymetric data (NOAA, n.d.-b) using the QGIS (v3.42) implementation of the Visvalingam algorithm.

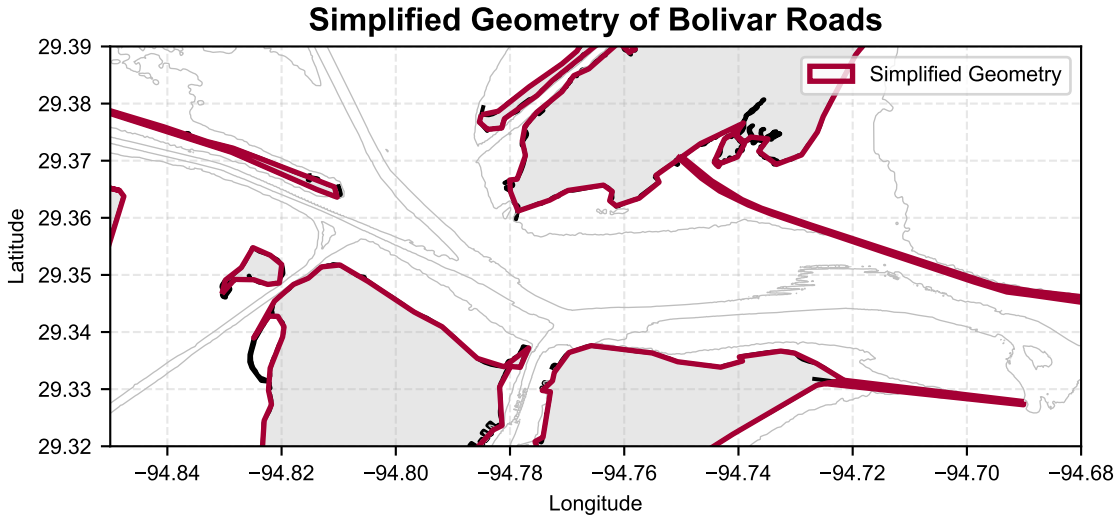


Figure 7.1: Simplified coastline of Bolivar Roads used in model testing.

All ShipNaviSim models are trained using a history length of 5 and awareness of the 10 nearest ships. Pham et al. (2025) showed that this configuration yielded the best results in their study. While

other hyperparameter combinations may lead to improved performance, a systematic hyperparameter search is beyond the scope of this experiment. The objective of this study is not to optimise predictive performance, but to evaluate the applicability and qualitative behaviour of the proposed reinforcement learning model within a design context. As such, a systematic hyperparameter search is considered outside the scope of this study.

During initial testing, it was found that training a model with the publicly available ShipNaviSim repository on the full dataset described in Chapter 3 resulted in numerical instabilities after 166 training epochs (training cycles), making the trained model unsuitable for a reliable evaluation. The precise source of these instabilities was not investigated further, as a detailed analysis of reinforcement learning training dynamics falls outside the scope of this study. Due to these numeric instabilities, a reduced training set was constructed for use in these experiments. This reduced training set was constructed by randomly selecting 10000 tracks from the full 2024 AIS dataset described Chapter 3.

Random selection of tracks is used as a pragmatic mitigation strategy to obtain a stable experimental baseline. Random sampling preserves a representative subset of the original data distribution while limiting the influence of extreme or rare trajectories that may disproportionately affect numerical stability. Based on observed action magnitudes during preliminary tests and a face-validation of the resulting trajectories, this reduced configuration resulted in stable training behaviour and was therefore selected for further analysis.

Furthermore, the model was trained on a random sample the full 2024 dataset, including November. Therefore, the evaluation of the model on the AIS tracks observed from November does not constitute a strictly out-of-sample temporal test. This may lead to optimistic performance estimates due to temporal dependencies in AIS traffic patterns. The chosen data split represents a pragmatic compromise between methodological rigour and project time limitations.

Four model variants are evaluated and compared against historical vessel tracks:

1. a ShipNaviSim model trained on the full dataset for 166 epochs.
2. a ShipNaviSim model trained on the reduced dataset for 300 epochs.
3. the same reduced-dataset model combined with a model predictive control (MPC) safety layer.
4. the reduced-dataset model with both MPC and additional intermediate goals.

As described previously, model 1 was trained on the full dataset. However, training was terminated prematurely after 166 epochs due to numerical instability, preventing the model from reaching full convergence. To address this, model 2 was trained on a reduced dataset, which mitigated the numerical instability and allowed for stable training throughout. Upon face validation, model 2 demonstrated qualitatively superior performance compared to model 1. Specifically, model 2 demonstrated action magnitudes and ship headings that are more aligned with real-world observations, suggesting that the reduced dataset, despite containing fewer samples, yielded a more robust representation of real-world maritime traffic behaviour in Bolivar Roads. Therefore, models 3 and 4 are developed as extensions of model 2, incorporating the safety shield and multi-goal extensions described in Chapter 6. This progression ensures that the extended models build upon a more stable and promising foundation identified during the initial evaluation phase.

For the safety layer, a prediction horizon of $N = 10$ steps and $M = 12$ rays are used, with a safe distance of $dist_{safe} = 150$ m (roughly half the width of the Ship Channels). The weighting matrices are set to $R_0 = I$ and $R = 0.1I$. The ray sensitivity parameters are defined as $\lambda_{mt} = 100(1-0.7\sin^2(\varphi_m))$, such that λ_{mt} depends on the angle φ_m of the corresponding ray.

For the model with additional intermediate goals, the intermediate goals are constructed by sampling positions along the corresponding historical trajectory at fixed intervals of n points. This choice provides a piecewise representation of the intended navigation path while avoiding continuous path-following constraints. This goal-construction approach is chosen as a pragmatic trade-off that preserves the geometric structure of the historical trajectory while limiting the number of goal transitions, thereby avoiding overly frequent goal switching. The consequence of this track-based goal-setting method is that it must be assumed that historical trajectories remain representative of the operating environment. Since real-world navigation waypoints may evolve as a result of infrastructural changes to waterways, the set of goals corresponding to a movement through new infrastructures should be updated accordingly to reflect new navigation conditions.

For the multi-goal experiments, intermediate goals are placed along each historical trajectory at intervals of $n = 50$ timesteps, which provides a pragmatic balance between encouraging the model to preserve the geometric structure of the historical trajectories while limiting the number of goal transitions, thereby avoiding overly frequent goal switching and avoiding forcing the agent to *exactly* follow historical trajectories.

Each simulation is initialized using the initial conditions of a historical vessel track and the simplified 0 m depth contour of the Bolivar Roads area as obstacle geometry (see Figure 7.1). During the simulations, detailed model logs are recorded and subsequently used to compute the evaluation metrics described in Section 7.1.

The models are implemented in Python (v3.11), with the MPC-based safety layer formulated and implemented using the Gurobi (v11.0.1) optimization solver. The model visualisation module is implemented in Pygame (v2.6.1). All experiments are prepared to run at the Delft High Performance Computing Centre (DHPC) (2024). The model source code is made available on GitHub to support reproducibility.[1] This repository is built upon a fork of the original ShipNaviSim repository[2] as published by Pham et al. (2025). The models are assessed using the methods and metrics described in the previous section.

7.3. Performance Under Existing Environmental Conditions

This section presents the results of the simulation experiments. A complete overview of the descriptive statistics for all evaluated metrics is provided in Appendix G.

Figure 7.2 provides a qualitative comparison between the models studied by visualising the trajectories of the same vessel under identical initial conditions and goals for the different simulation agents. This figure highlights the differences between historical and simulated tracks in terms of position, drift behaviour, and interactions with obstacle regions, thereby illustrating the effects of the different modelling choices on the resulting vessel motion.

From this figure, it is evident that the models with only a single goal appear to seek out the shortest route towards that goal, both with and without the safety layer enabled. Furthermore, this figure shows that although all models tested exhibit the ability to move towards a goal, they do not produce the expected headings that align the COG of the produced trajectory. Specifically, the ships tend to move along a track in reverse or nearly perfectly perpendicular to their heading.

The Goal-Conditioned Average Displacement Error (GC-ADE) for the four model variants is shown in Table 7.1. For models 1, 2, and 3, the GC-ADE values are relatively high and are accompanied by a large standard deviation. Although Model 4 shows an improvement compared to the other variants, the resulting GC-ADE remains larger than the characteristic length of many tanker vessels (200-300 m), or the width of the Houston Ship Channel (approximately 200 m), indicating that the models produce substantial deviations from the historical trajectories.

Table 7.1: Mean and standard deviation of GC-ADE for the tested models.

	GC-ADE (m)	
Model	Mean	Std. Dev
Historic data	0.00	0.00
1: Full Set - 166 Epochs	1323.13	463.29
2: Limited Set - 300 Epochs	1053.15	463.71
3: Limited Set - 300 Epochs + Safety	917.85	497.82
4: Limited Set - 300 Epochs + Safety + Extra Goals	397.07	234.63

The ranges of the kinematic metrics are summarised in Table 7.2. This table shows that all model variants exhibit a wider range of acceleration values than is observed in the historical data. For drift angle and trajectory curvature, the overall ranges produced by the models are comparable to those observed in reality, which means that the models do properly capture the full range of these kinematic metrics as observed in real-world data. However, the distributions of these metrics as produced by the models differs markedly from the distributions of historical data. As shown in Figure 7.3, the

[1] https://github.com/TUDelft-CITG/traffic-behaviour-cloning
[2] https://github.com/quanganh1999/ShipNaviSim

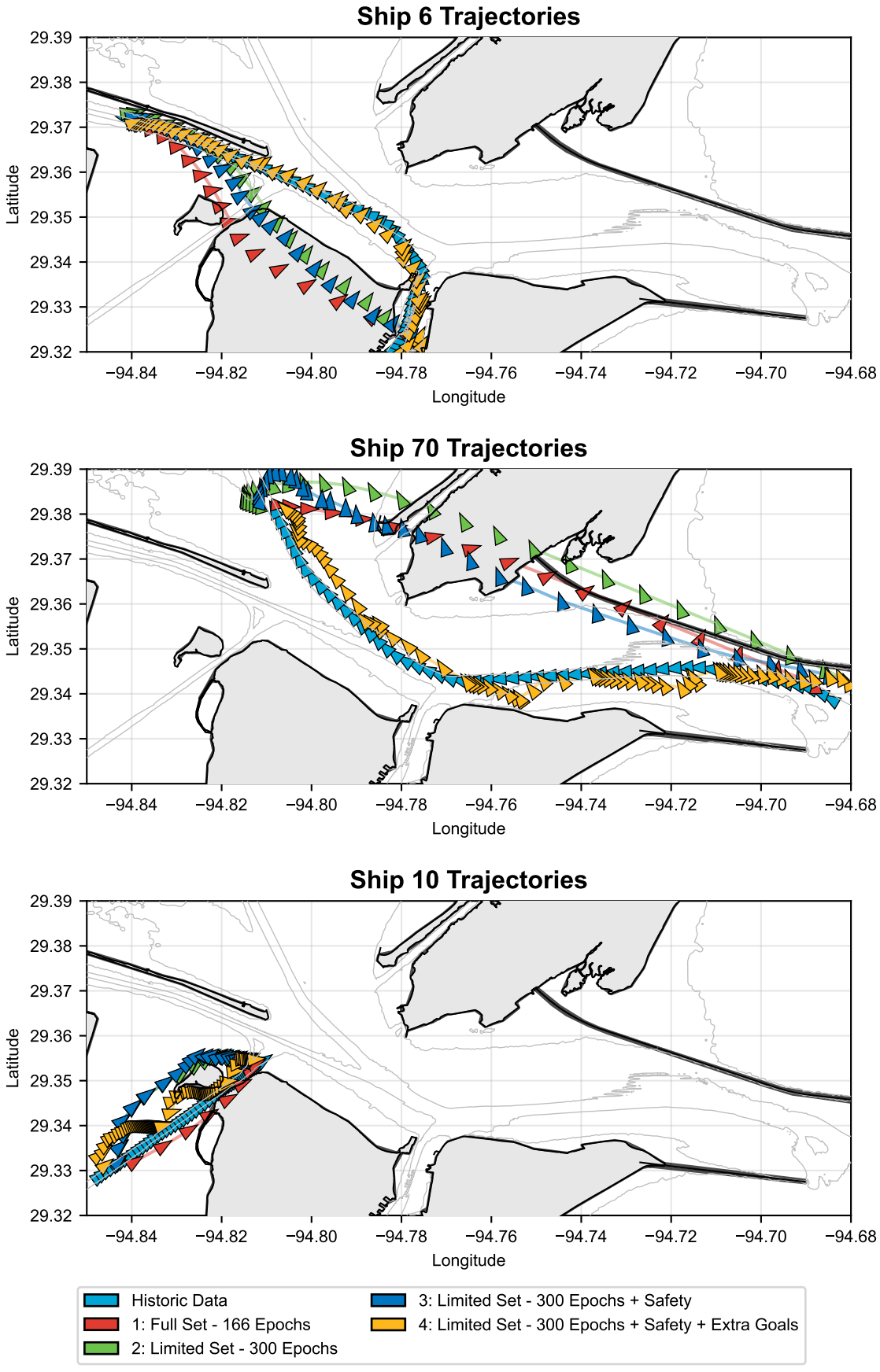


Figure 7.2: Examples showing the trajectories of a single ship in multiple simulations.

simulated trajectories display substantially higher variability in both drift and track curvature than the historical tracks, which indicates that that the models generate more erratic manoeuvring behaviour than observed in real-world data.

Table 7.2: Ranges of kinematic metrics.

	Acceleration		Drift (°)		Curvature (°/s)	
Model	Min	Max	Min	Max	Min	Max
Historic data	-1.39	1.84	-179.99	180.00	-18.00	18.00
1: Full Set - 166 Epochs	-8.40	7.94	-179.99	179.99	-17.77	17.64
2: Limited Set - 300 Epochs	-7.65	7.51	-179.96	179.98	-17.94	17.98
3: Limited Set - 300 Epochs + Safety	-7.59	7.54	-180.00	180.00	-18.00	18.00
4: Limited Set - 300 Epochs + Safety + Extra Goals	-7.71	8.27	-180.00	180.00	-18.00	18.00

Figure 7.3 further shows that the magnitude of the actions produced by all models is larger than that inferred from the historical data. Consequently, the simulated vessels tend to move faster than their real-world counterparts, which contributes to the increased GC-ADE observed when comparing simulated and historical trajectories.

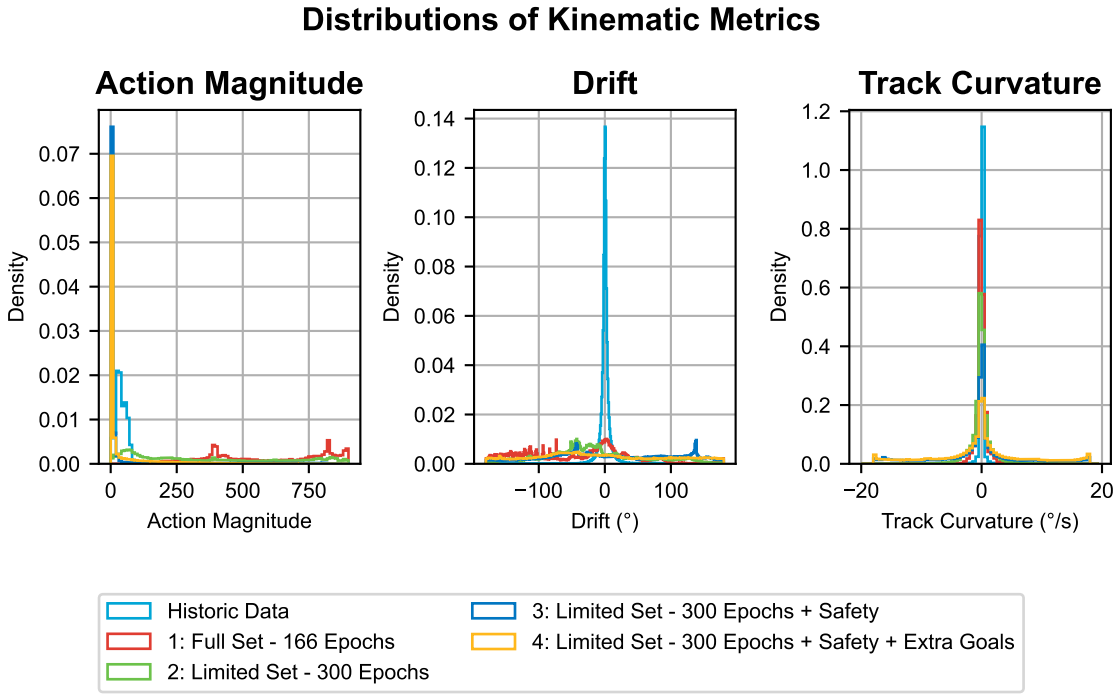


Figure 7.3: Distributions of action magnitude, drift, and track curvature.

The effect of the safety layer is quantified in Table 7.3. This table shows that models without the safety layer exhibit a higher number of entries into obstacle regions, whereas the inclusion of the safety layer significantly reduces such violations. In addition, the table shows that the magnitude of the applied control actions is generally smaller when the safety layer is active, indicating a smoothing effect on the vessel manoeuvres. This means that the models that include the safety layer do exhibit the ability to alter the ShipNaviSim agent's behaviour to account for obstacles the ship encounters. Figure 7.4 shows where entries are prevalent across all simulations. This figure shows that the obstacle detection can be improved in general, as there are no geographic distributions of obstacle entries appears to be similar across simulations 1, 2, and 3. The obstacle entries from simulation 4 are more concentrated around sharp curves in the obstacles, suggesting that the multi-goal routing does also play a part in avoiding the banks of a channel. A dedicated overview of the observed entries per simulation can be found in Appendix H.

Table 7.3: Effects of the safety layer: comparison between obstacles entered and control magnitude.

	Obstacles Entered	Control Magnitude	
Model		Mean	St. Dev.
Historic data	0	0.00	0.00
1: Full Set - 166 Epochs	6334	0.00	0.00
2: Limited Set - 300 Epochs	7340	0.00	0.00
3: Limited Set - 300 Epochs + Safety	5889	92.78	73.16
4: Limited Set - 300 Epochs + Safety + Extra Goals	1174	109.59	82.73

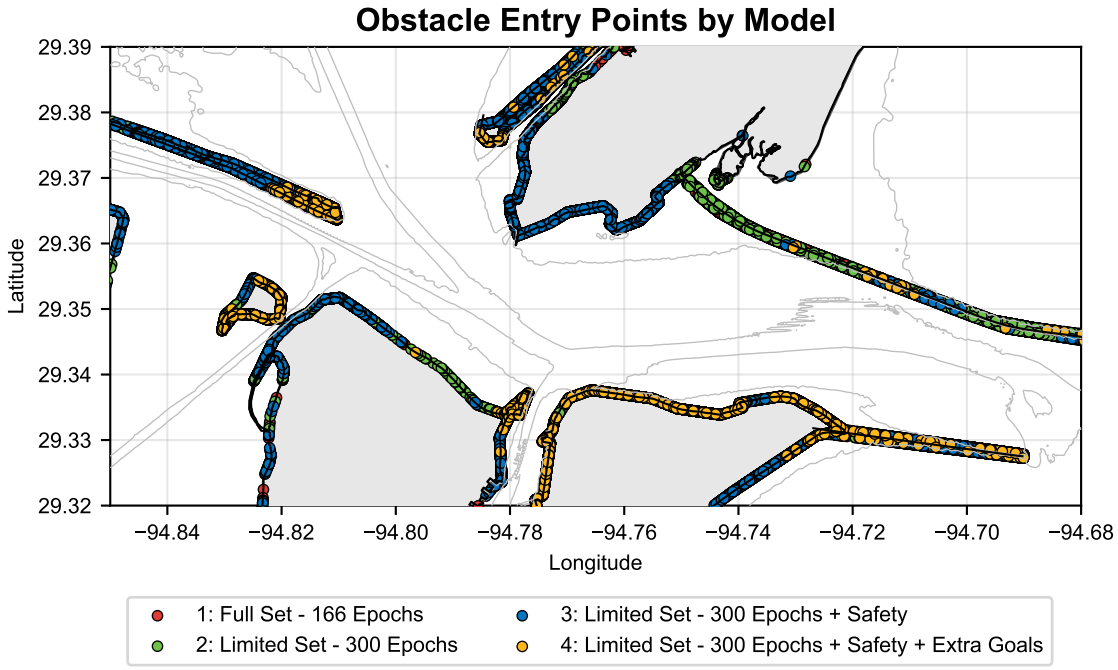


Figure 7.4: Points where ships entered the coastline by simulation. Individual maps per simulation can be found in Appendix H.

Overall, these results show that the tested models still differ considerably from what is seen in the historical data. The baseline versions (models 1 and 2) often produce trajectories that deviate strongly from real vessel tracks, both in terms of kinematics and position. However, adding obstacle awareness through the safety layer reduces the number of crossings into restricted areas and leads to more realistic behaviour near the coastline. In addition, introducing intermediate goals helps the simulated vessels follow routes that are closer to the historical tracks, suggesting that providing extra navigational guidance to the models improves how well the models reproduce observed movement patterns.

7.4. Post-Hoc Safety Shield Parameter Sensitivity Analysis

The results in the previous section are obtained with a single baseline setting of the safety shield parameters, yet that choice can strongly shape what the models appear to do. Because the shield overlays the ShipNaviSim controller and can modify its actions, its tuning may trade tracking accuracy against constraint satisfaction and control effort, and could even be responsible for the poor performance observed for the main model if it intervenes too aggressively or in unintended ways. This sensitivity analysis is therefore needed to show whether the main findings from the previous section are robust to plausible parameter changes, to identify which parameters actually matter for the KPIs, and to distinguish limitations of the underlying model from effects introduced by the safety shield layer.

Therefore, this sensitivity analysis begins by selecting vessel tracks that represent a distinct and diverse set of manoeuvres through Bolivar Roads, ensuring that the results are evaluated across the

most operationally relevant scenarios. Key model parameters ($\lambda_{mt}, M, N, dist_{safe}$) are then varied one at a time from the baseline configuration described in Section 7.2 to isolate the influence of each parameter on system behaviour. Performance is assessed using a set of KPIs: GC-ADE on each of the selected tracks, obstacle entries by simulation model, and control magnitude, in order to capture accuracy, safety, and control effort, respectively. Finally, combinations of parameters identified as providing the best individual performance are tested together to investigate potential second-order interactions and coupled effects that may not be evident when parameters are varied independently.

7.4.1. Test Cases

For this sensitivity analysis, a set of 12 historic tracks were selected as test-cases. Tracks were selected through visual inspection such that one track from each of the major flows identified in Figure 3.7 is included in the test-case set. This approach allows this sensitivity analysis to focus on the most important flows through Bolivar Roads. The selected tracks are shown in Figure 7.5.

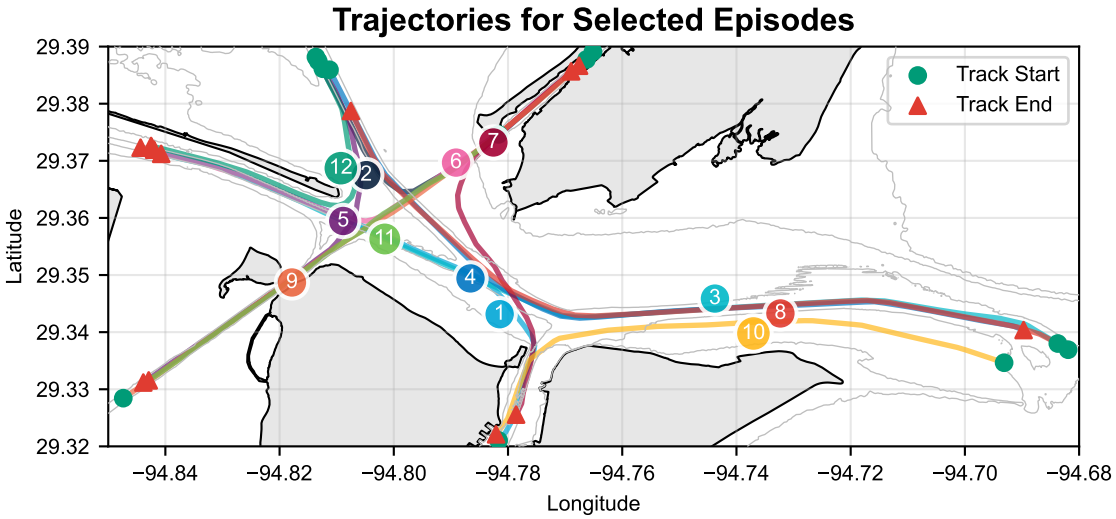


Figure 7.5: Tracks selected for this sensitivity analysis.

The simulations are configured to isolate the effects of parameter variation by staggering vessel tracks in time, with 1000 time steps allocated per track, ensuring that each track is simulated independently and that ship-ship interactions do not influence the results. The same geometric configuration used in the main simulations is retained to maintain consistency and ensure that observed differences are attributable solely to changes in model parameters rather than environmental or layout effects. All simulations are executed at the Delft High Performance Computing Centre (DHPC) (2024) while systematically varying the parameters $\lambda_{mt}, M, N,$ and $dist_{safe}$, providing a controlled and computationally efficient framework for evaluating their impact on system performance. The model outputs are then used to determine the GC-ADE and obstacle entry scores, which are in turn used to assess the quality of each set of parameters.

The base combination of parameters for the sensitivity is the same as described as in Section 7.2: a prediction horizon of $N = 10$ steps and $M = 12$ rays are used, with a safe distance of $dist_{safe} = 150$ m. The weighting matrices are set to $R_0 = I$ and $R = 0.1I$. The ray sensitivity parameters are defined as $\lambda_{mt} = 100(1-0.7\sin^2(\varphi_m))$, and intermediate goals are placed every $n = 50$ steps of the historical track corresponding to the test case. Parameters are then individually varied from this base case as follows:

- λ_{mt}: Different shapes of the function defining this parameter are tested in order to determine whether the model is sensitive to the shape of the approximated ship domain. The following functions are tested:
  - $\lambda_{mt} = 100(1 - 0.7\sin^2(\varphi_m))$ (Base Case).
  - $\lambda_{mt} = 100(0.5 - 0.5\sin^2(\varphi_m)$

  - $\lambda_{mt} = 100(0.6 + 0.4\cos(\varphi_m))$
  - $\lambda_{mt} = 100(0.5 + 0.5\cos(\varphi_m))$
- N: varied between $N = 0$ (no lookahead), $N = 5$, and $N = 10$ (Base Case).
- M: varied between $M = 8$, $M = 12$ (Base Case), $M = 24$,and $M = 36$. These values ensure that the agent always has rays at $\varphi_m = 0°, 90°, 180°$, and $270°$.
- $dist_{safe}$: is varied in 50 m intervals from 50 m to 300 m. Base Case: $dist_{safe} = 150$ m.

7.4.2. Findings

After running the model on the test cases with the varied parameters, the following observations can be made: Firstly, tracks that do not pass within $dist_{safe}$ of any obstacles, the safety shield does not intervene and therefore does not affect the performance of the base ShipNaviSim agent (see Figure 7.6). In these cases, trajectories are effectively left unchanged: they are not influenced by the shield in any way, which also implies that the shield does not exhibit a stabilizing effect on the ShipNaviSim agent when obstacles are absent from the relevant operating region.

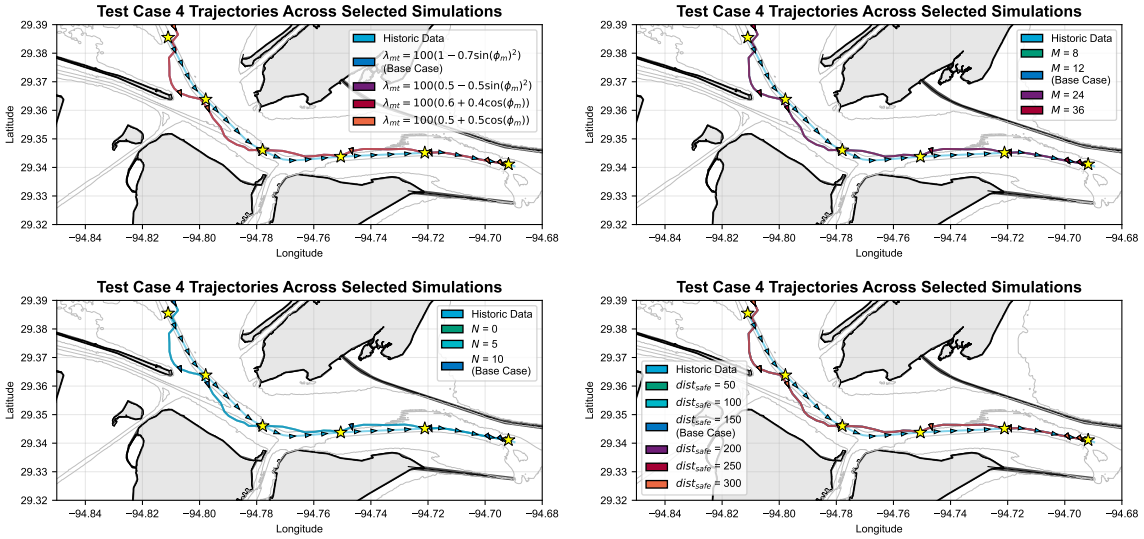


Figure 7.6: Comparison of tracks produced by different simulations for test case 4, which does not pass within $dist_{safe}$ of an obstacle. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

Furthermore, when the safety shield does intervene, its impact is most pronounced for tracks in which intervention happens early along the trajectory. As illustrated in Figure 7.7, early intervention tends to have a knock-on effect on the rest of the trajectory. In contrast, interventions that occur when a ship encounters an obstacle later in its trajectory tend to yield smaller observable differences in overall performance, which shows that the safety filter only has an effect when the agent is near enough to an obstacle.

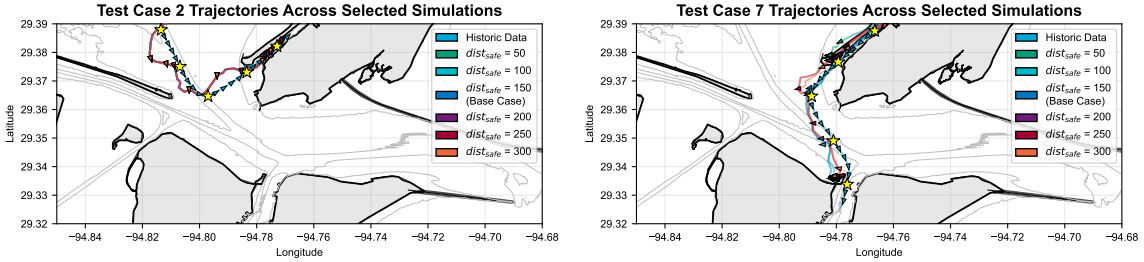


Figure 7.7: Comparison of safety shield intervention early (left, test case 2) and late in the track (right, test case 7).

These examples illustrate how functional form used for λ_{mt} affects the performance of the obstacle avoidance layer. Using a cosine-shaped function for λ_{mt} results in more obstacle entries than using a sine-shaped function (see Table 7.4), suggesting that the sine-shaped definition is more effective at detecting (and therefore avoiding) obstacles. In addition, reducing the number of lookahead steps increases the number of obstacle entries (see Table 7.4), which indicates that extrapolating the ship's motion forward in time is important for filtering whether a decision proposed by the ShipNaviSim agent remains effective once its near-future consequences are taken into account.

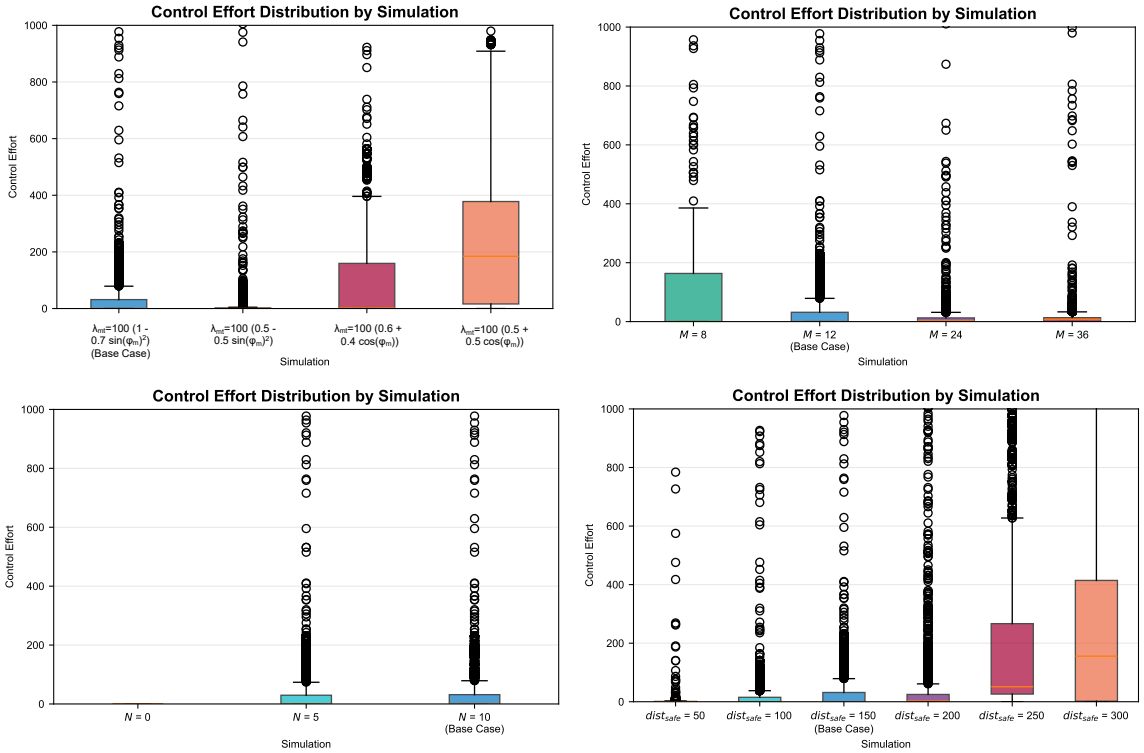


Figure 7.8: Comparison of control effort distributions entry points produced under different parameter variations. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

These differences are also reflected in the control behaviour. Cosine-shaped functions for λ_{mt} exhibit a less consistent control effort distribution than sine-shaped functions (see Figure 7.8), implying that the control decisions induced by the sine-shaped definition of λ_{mt} are more stable. Similarly, stability tends to increase with the number of rays used by the safety shield, particularly in the presence of non-convex obstacles (see Figure 7.10). This aligns with the expectation described by Dawood et al. (2024) that choosing ray sensitivities λ_{mt} statically and equally can lead to agents getting stuck in such complex geometries, and that increasing ray coverage can mitigate this effect.

Across the parameter variations considered, the locations where ships enter obstacles do not appear to shift meaningfully. The entry points remain visually consistent (see Figure 7.9), indicating that the parameters varied in this analysis do not substantially affect where failures of the safety shield occur, even if they do affect how often they occur or how the agent behaves near obstacles.

Table 7.4: Number of obstacle entries observed per simulation

Variation	Obstacle Entries	Variation	Obstacle Entries
$\lambda_{mt} = 100(1 - 0.7sin(\varphi_m)^2)$	2	$N = 0$	16
$\lambda_{mt} = 100(0.5 - 0.5sin(\varphi_m)^2)$	2	$N = 5$	2
$\lambda_{mt} = 100(0.6 + 0.4cos(\varphi_m))$	4	$N = 10$	2
$\lambda_{mt} = 100(0.5 + 0.5cos(\varphi_m))$	5	$dist_{safe} = 50$ m	2
$M = 8$	4	$dist_{safe} = 100$ m	2
$M = 12$	2	$dist_{safe} = 150$ m	2
$M = 24$	1	$dist_{safe} = 200$ m	5
$M = 36$	0	$dist_{safe} = 250$ m	4
		$dist_{safe} = 300$ m	4

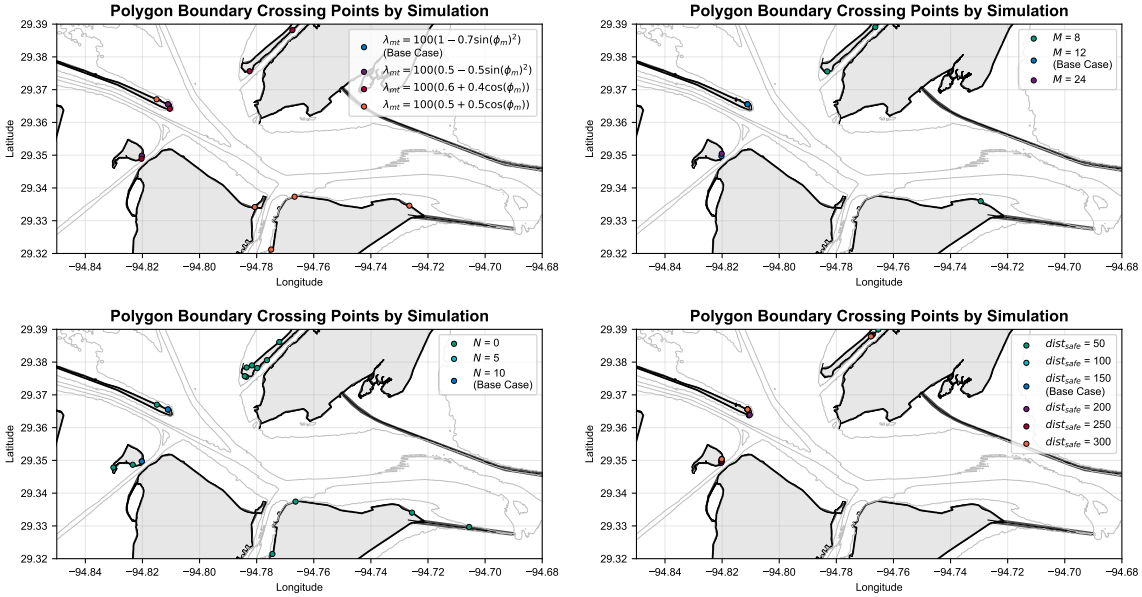


Figure 7.9: Comparison of obstacle entry points produced under different parameter variations. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

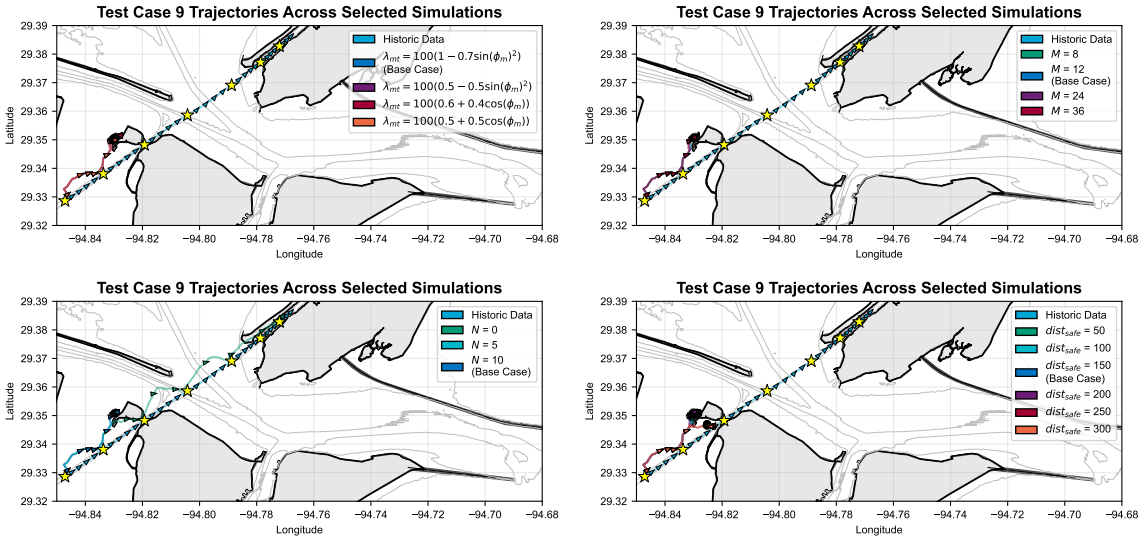


Figure 7.10: Example of ships getting stuck in test case 9. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

7.5. Performance on New Situations

In addition to considering how well the different models are able to recreate traffic patterns under initial conditions, the ability of these models to account for *new* situations should also be considered, as described by Requirements 21 - 24 . Given that model 4 (the reduced-dataset model with both MPC and additional intermediate goals) most closely matched the real-world behaviours observed in the historic dataset, this model is chosen to test the ability of a safety layer to account for previously unseen obstacles.

For this experiment, new obstacles are added to the simulation, in addition to the coastlines of Bolivar Roads as shown in Figure 7.1. For the first experiment, a rectangular obstacle is added east of the port of Galveston, overlaid on a part of the Houston Ship Channel. In the second experiment, a simplified gate structure is introduced, adapted from the designs by Konstapel (2026). These obstacles are introduced in the context of the existing conditions in Bolivar Roads, in order to demonstrate how the traffic patterns under the introduction of new obstacles compare to the traffic patterns observed under real-world conditions. The rectangular obstacle is chosen as a situation to be tested because of its geometric simplicity, and since it resembles simple no-go zones like the zones that would be defined in traffic separation schemes (TSS) or precautionary areas. The gate obstacle is chosen as it is a simple representation that incorporates at least some of the features desired in the design of the Houston Ship Channel Gate Complex (Burkley et al., 2022; Konstapel, 2026).

Other simulation parameters, including goal-setting logic are left unchanged. The consequence of this is that some tracks *may* have goals set inside the newly introduced obstacles. This is the consequence of using a regular interval in goal-setting along a historic track. The strategic route choice modelling, i.e., the placement of intermediate goals along a desired route, and how these would change when a new obstacle is introduced, is not studied in this work, and is left for future research.

Table 7.5 shows the number of obstacles entered and control magnitude for each of these scenarios. From this table, it can be concluded that the average magnitude of the interventions of the safety layer increases when new obstacles are present. Furthermore, this table shows that the scenarios in which new obstacles are present the number of obstacle entrances increases. Figure 7.11 shows that this can largely be attributed to the introduction of these obstacles, since the pattern of obstacle entries on the existing coastline does not have noteworthy differences when compared to the baseline scenario.

A noteworthy pattern in these entries of new obstacles is the higher density of entries along obstacle edges that ships would encounter head-on. Both the rectangular and gate obstacles show a higher density of obstacle entries along their north-south axis than their east-west axis, suggesting that ships that transit these obstacles in an east-west movement enter these obstacles more frequently. This is likely a result of their head-on movement towards the new obstacle. This idea aligns with the general

direction of motion shown in Figure 7.12. Furthermore, this also suggests that the safety shield is less useful for detecting new obstacles that ships meet head-on than for ships

Additionally, Figure 7.12 shows that the trajectories along these new obstacles do show slowdowns compared to the baseline scenario. The spread of motion around the rectangular obstacle and gate obstacles also suggests that the models are partially able to manoeuvre around the obstacle. However, these figures do also show that ships enter and manoeuvre within the obstacles. This is the result of obstacles and obstacle detection being defined based on polygons, without regard for whether a ship is inside or outside an obstacle polygon.

Table 7.5: Effects of the safety layer: comparison between obstacles entered and control magnitude.

	Obstacles Entered	Control Magnitude	
Scenario		Mean	St. Dev.
Baseline	1174	109.59	82.73
Rectangular Obstacle	2024	111.10	72.46
Gate Obstacle	1534	151.56	77.97

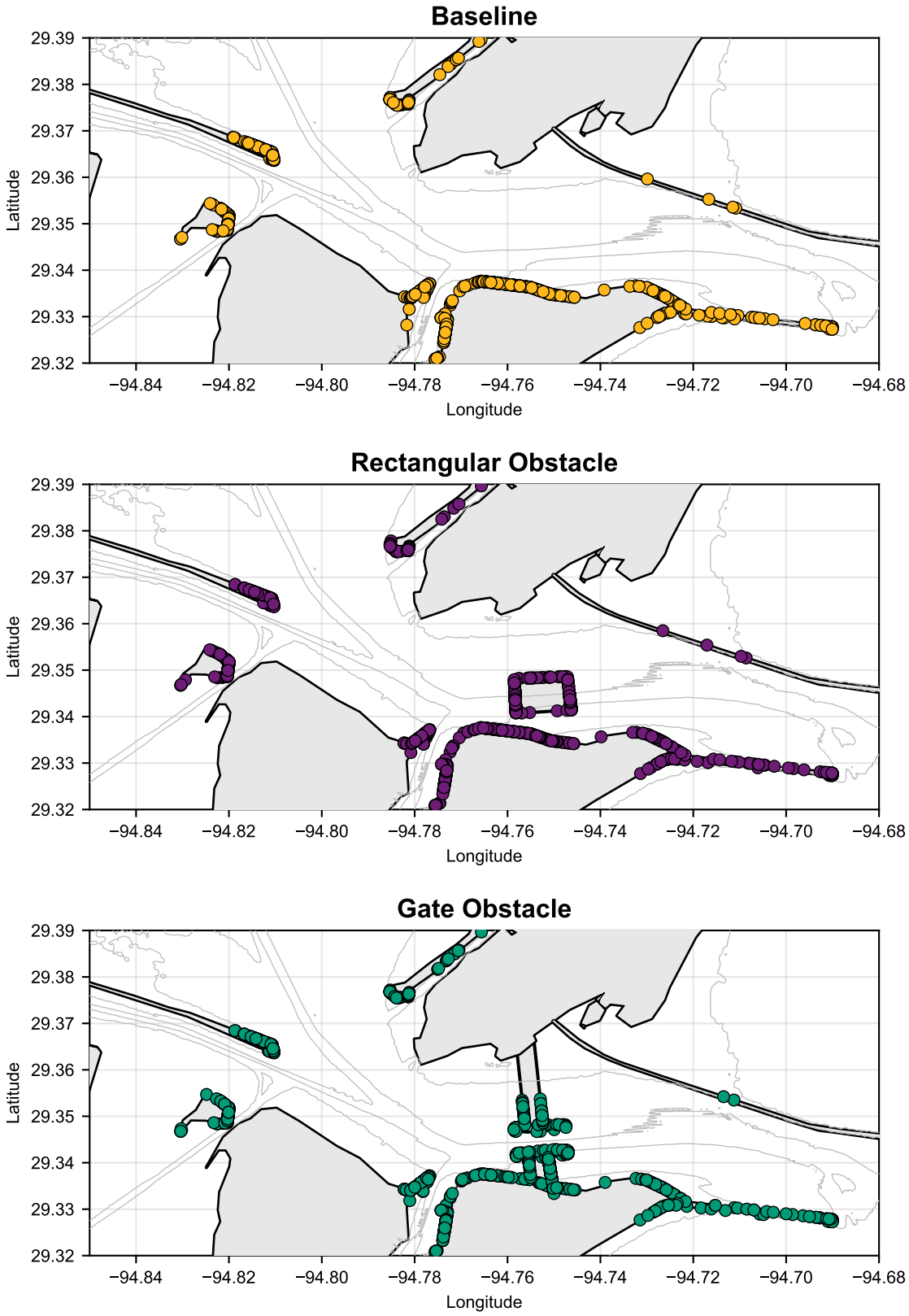


Figure 7.11: Points where ships crossed the coastline by scenario.

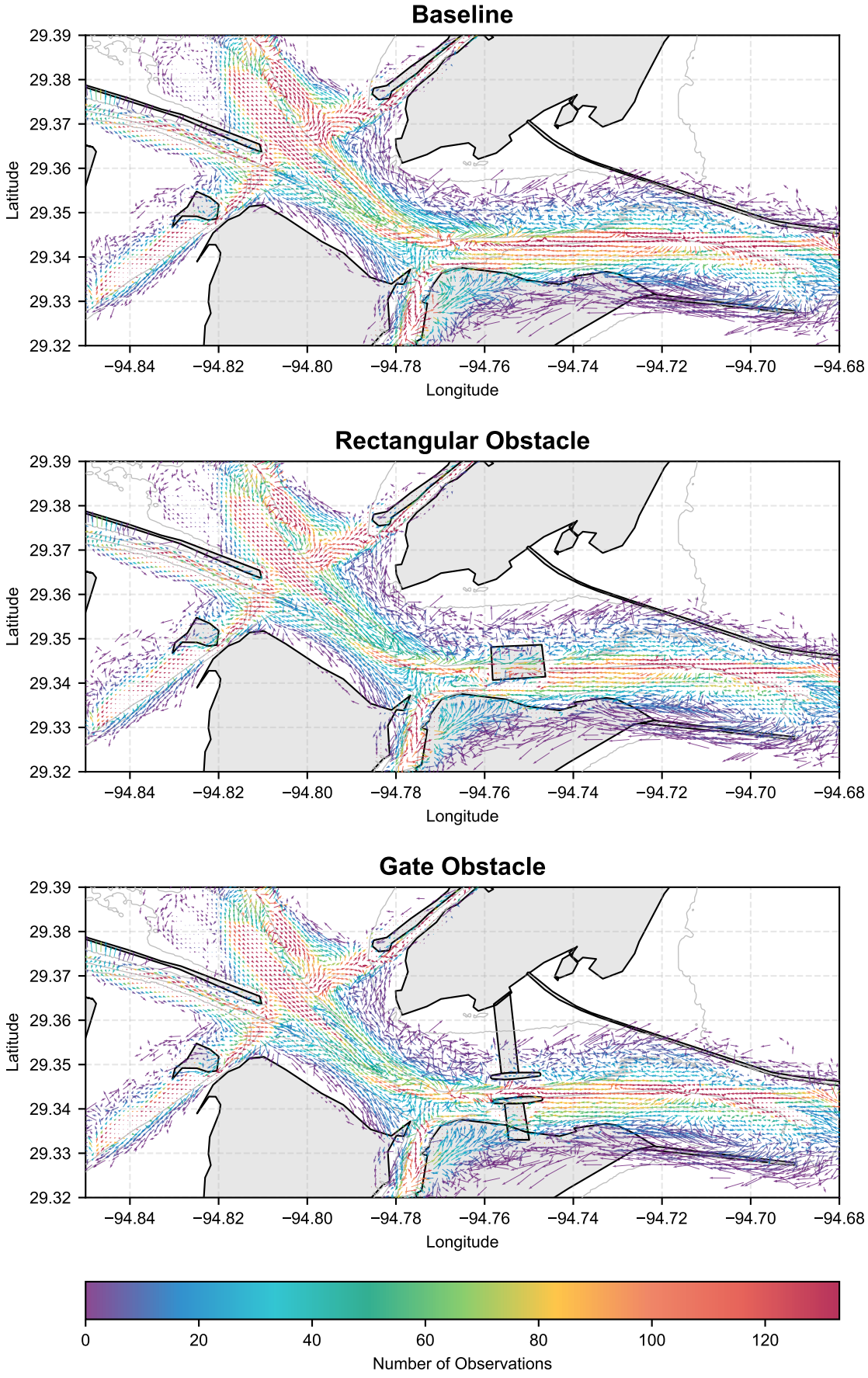


Figure 7.12: Vector field of trajectories produced by the reduced-dataset model with both MPC and additional intermediate goals without new obstacles, with a rectangular obstacle, and a gate obstacle. Arrow direction indicates COG, arrow length indicates speed.

7.6. Conclusions

Overall, the experimental results show that the evaluated data-driven maritime traffic models capture some of the characteristic structure of vessel movement in Bolivar Roads, but that their ability to reproduce observed behaviour remains limited under the tested configurations. In relation to Research Question 4, all model variants exhibit goal-seeking behaviour and reproduce the full range of drift angles and trajectory curvatures found in the AIS data, yet they do not consistently reproduce the distribution of these kinematic features: simulated trajectories show substantially higher variability in drift and curvature, wider acceleration ranges, and action magnitudes that imply higher vessel speeds than observed historically. This mismatch is also reflected in the trajectory-level performance, where models with a single goal yield high GC-ADE values (917-1323 m) and face validation reveals unrealistic heading-COG alignment, indicating that the models often fail to capture key behavioural features required by Requirements 8 - 20, even when the overall behavioural patterns expected in Bolivar Roads are broadly represented.

With respect to Research Question 5, the experiments indicate that the introduction of an MPC-based safety layer and the addition of intermediate goals are effective steps in moving towards a design-oriented tool. However, these steps need further detailing and calibration to truly capture the expected behaviours required by Requirements 8 - 20 and 21 - 23. The safety layer reduces obstacle violations and smooths control, while intermediate goals substantially improve route adherence (GC-ADE reduced to 397 m) and concentrate remaining obstacle entries around sharper geometric features. However, additional tests with newly introduced obstacles show that, although the safety layer increases intervention effort and induces partial slowdowns and avoidance behaviour, obstacle entries also increase and tend to occur at head-on encounters, highlighting that robust constraint satisfaction under novel infrastructure geometries remains a key limitation.


639
1109029
SUZUKI
DO NOT SIT ON RAIL
NO SMOKING
VAPOR AND
E-CIGS
FIRE

8

Discussion

While the results presented in Chapter 7 show that data-driven maritime traffic models with safety extensions can be used as a basis for infrastructure impact assessment, they also highlight that the approaches presented in this work are not yet fully mature. Although a list of requirements that should be met by a design-capable maritime traffic simulation model was formulated in Chapter 5, several assumptions and simplifications were made during the modelling, experiments, and evaluation of the models developed in this work. In addition to that, the experiments using the models have shown that the models proposed in this work are not yet suitable to be used as a design tool in a complicated setting such as Bolivar Roads. Therefore, it is necessary to reflect on the limitations of this study before drawing broader conclusions about the results presented in this work.

Furthermore, since this work aims to contribute to the development of data-driven maritime traffic simulation models, and this work demonstrates that there are still many developments to be made, it is also necessary to consider the position of this work in relation to other studies and practical uses of such simulation models.

Therefore, this chapter explains the limitations of this study and the position of this study relative to other works in the field data-driven maritime traffic simulation models. The first section will focus on the limitations of this work, and the second section will focus on the relation of this work to other research and practical applications of maritime traffic simulation models.

8.1. Limitations

This section identifies the limitations of this study. Firstly the limitations that are a result of methodological choices will be elaborated upon. Secondly, the limitations that arise from the use of ShipNaviSim as the basis for the model are explained. Next, the limitations of the model evaluation process are identified, and finally the usability of the results of this study are addressed.

8.1.1. Methodological Choices

To start, this study deliberately narrowed its focus to data-driven maritime traffic simulation models applied within a design context. Maritime traffic simulation serves a broad range of purposes, from traffic management and operator training to infrastructure planning. Each of these uses carries a distinct set of requirements regarding model fidelity and validation criteria. Restricting the scope to the design phase of maritime infrastructures ensures that models are evaluated against criteria directly relevant to infrastructure designers, such as the ability to generate realistic traffic scenarios under hypothetical layout configurations. Furthermore, while rule-based and mathematical simulation approaches have been studied extensively in design contexts, the suitability of data-driven methods for such applications has received comparatively little systematic attention. This study therefore targets a genuine gap in the literature. Future work could extend these findings to other domains, such as traffic management or human-in-the-loop simulation, where requirements differ substantially from those of the design phase.

The model also does not explicitly account for the effect of external factors such as wind and currents on ship behaviour. Although these effects are implicitly present in the AIS data used for training, this means that the model cannot be used to systematically study how different environmental

conditions influence traffic patterns, for example under extreme weather such as hurricanes or to isolate the impact of tidal currents. This is the result of a scoping choice, choosing not to explicitly model these factors but preferring to let the RL model learn these effects implicitly. The consequence of this choice is that the models formulated in this work do not have the capability to account for these external (weather and current) effects, limiting the applicability of the model as a design tool to only "average" conditions. Explicitly accounting for these effects could improve the versatility in situations a data-driven maritime traffic simulation model is able to represent.

Furthermore, this study does not involve a quantitative assessment of the model performance on novel situations. This limits the ability of this study to quantitatively assess the model when new infrastructure is introduced. While methods to quantitatively validate the model's ability to simulate new infrastructure conditions exist, such as human-in-the-loop simulation like in Burkley et al. (2022), such resources were not available for this project. Therefore, this project has relied solely on the qualitative judgements of model validity under new situations derived from the interviews in Chapter 4.

This study has also left (hyper)parameter tuning for future research. ShipNaviSim parameter setting was assumed based on literature (Pham et al., 2025), in order to focus on the data processing and methodical contributions (safety shield) of this work. Similarly, the sensitivity of the model to route choice and waypoint setting was also not studied for the same reasons. While a brief post-hoc sensitivity analysis was carried out to determine the sensitivity of the safety shield to parameter settings, a full-scale calibration of the safety shield based on real-world or synthetic reference data has not been performed. Therefore, the parameter setting of the safety shield should also be subject of further research. This parameter setting could also be extended in order to attempt to capture hydrodynamic effects caused by the presence of a structure, such as suction or bank effects, which are not included in the model presented in this work. While capturing such effects is outside the scope of the present study, these effects can become important in close encounters between ships and obstacles and may affect both safety and manoeuvring behaviour.

Adding to this idea, the way ships interact with new infrastructure has also not been studied in detail. The current safety and interaction settings are based mainly on the interviews with experts, but there may be other relevant factors that were not identified, which could influence how vessels behave around new structures. These factors should, together with the aforementioned parameter setting to capture hydrodynamic effects, be subject of further research.

Finally, this study focuses on a microscopic view of a limited set of waterways. While this allows for a detailed analysis of local ship behaviour and interactions, the broader, network-wide effects of an infrastructure intervention are not considered. In reality, changes at one location, such as the introduction of a gate, may propagate through the wider waterway network and affect traffic patterns at a much larger scale.

8.1.2. Use of ShipNaviSim

The main limitation of using ShipNaviSim is that the software assumes that all ships are of homogeneous type, which is contrary to the finding that in a setting like Bolivar Roads it is desirable to represent many types and characteristics of ships, as a ship's type and characteristics greatly influence its manoeuvring abilities. Implementing this would require a significant rework of the ShipNaviSim source code, which would allow a more diverse set of agents to be represented in the simulation. Therefore, implementing a multi-class approach to formulating data-driven maritime traffic simulation models would be more appropriate.

Second, not all ships in the AIS data report their heading. Ships without heading information had to be left out of the training, which means that part of the real traffic and some ship-ship interactions are missing from the model. As a result, the learned behaviour does not fully reflect everything that happens in practice.

On the computational side, ShipNaviSim became numerically unstable when trained on the full dataset, which limited the experiments that could be carried out. This means that a full evaluation of ShipNaviSim on all available data was not within reach. The causes of these numeric instabilities have not been investigated within the scope of this research in order to keep the research aligned with the research goals, as researching the numerical stability of a small part of the model veers more into the field of computational science. Nonetheless, with other (cleaner) datasets a full training run might be feasible, and this should be investigated in future research.

Finally, another limitation of using ShipNaviSim is that ShipNaviSim trains agents with a single goal, defined as one geographic target location that the ship has to navigate to. In constrained waterways,

however, ships typically follow specific routes that can be described by a sequence of intermediate waypoints rather than by a single end point. Training agents to navigate along multiple goals or waypoint sequences could therefore better represent route choice and channel-following behaviour, and may lead to more realistic performance in complex and intersecting waterways.

8.1.3. Model Evaluation

First, the evaluation does not fully separate the training and test samples, meaning that some of the tracks used for evaluation were likely also present in the training data. This is problematic because it can lead to overly optimistic performance estimates: models may appear to perform well not only because they have learned general behavioural patterns, but also because they are evaluated on scenarios that are (partly) familiar. As a result, the reported metrics provide stronger evidence of in-distribution performance than of robustness to genuinely unseen traffic situations. This matters in particular for the intended use case, where models are expected to cope with variation in vessel behaviour, traffic density, and local conditions. Therefore, a more appropriate evaluation would explicitly enforce an out-of-training-sample split (e.g., by holding out entire tracks, time periods, or geographic sub-areas) to better quantify generalization and reduce the risk of inadvertent information leakage.

Second, the evaluation is largely aggregate, with little analysis of how the models behave on individual test cases. While overall averages are useful for comparing methods, they can hide important case-by-case variability: a model can score well on average while consistently underperforming in specific situations, such as close encounters, overtaking manoeuvres, merges, or other high-interaction scenarios. Because the evaluation does not systematically distinguish between different classes of test cases beyond the post-hoc safety analysis, it is difficult to identify where errors occur, whether failures are concentrated in particular manoeuvre types, and how consistent the models are across conditions. As a result, the evaluation offers limited diagnostic insight for improving the models and limited evidence about reliability in rare but operationally important scenarios. Therefore, a more informative evaluation would include stratified or scenario-based reporting (e.g., by interaction intensity, geometry, traffic density, or vessel context) alongside targeted qualitative inspection of representative success and failure cases.

8.1.4. Quality of Developed Model

Finally, the quality of the developed model is not yet sufficient for direct use as a design tool. Although the model is able to reproduce broad movement patterns, such as general route-following and overall direction of travel, its behaviour remains noticeably erratic at a finer scale. In particular, the kinematic details of ship motion are not consistently captured: the simulated trajectories show unrealistic drift behaviour and do not reproduce acceleration (and deceleration) dynamics in a stable, physically plausible way. This is a substantial limitation because design-oriented applications typically depend on credible short-horizon motion and interaction dynamics, not only on coarse trajectory shape. As a result, the model in its current form is better viewed as a proof of concept for learning high-level behavioural structure than as a validated representation of vessel manoeuvring performance, and further work is needed to improve kinematic fidelity and smoothness before it can support design decisions.

8.2. Position of this Work

In addition to identifying the limitations of this work, it is also necessary to consider the position of this research in relation to other works and in relation to practice. This way, the position and use-cases of the findings presented in this work are placed in the larger context in which this research takes place. Therefore, this section presents the position of this work in relation to other works in research and in relation to practical applications of data-driven maritime traffic simulation models.

8.2.1. In Relation to Other Research

This thesis is positioned at the intersection of data-driven vessel behaviour modelling and maritime infrastructure assessment. First, it provides a concrete instantiation of the conceptual framework proposed by Bellsolà Olba et al. (2019) by operationalising the microsimulation component. In doing so, the work moves from a high-level methodological framing to an implementable modelling workflow, and demonstrates what is practically involved in constructing and evaluating a microscopic simulation of vessel movement from observational data.

Second, the work builds directly on ShipNaviSim as introduced by Pham et al. (2025). Rather than

treating ShipNaviSim as an end product, this thesis explores how it could be extended and adapted towards a maritime infrastructure model, i.e., a model that not only reproduces individual vessel motions but can ultimately support system-level questions such as throughput and operational performance in constrained waterways.

At the same time, the present study highlights that a meaningful link between microsimulation outcomes and infrastructure-relevant conclusions depends on a principled definition of safety. In that respect, the approaches developed by Burkley et al. (2022) and Konstapel (2026) are complementary: they provide methods to quantify and establish "safety" in a way that can serve as input to a dedicated safety layer. With such a safety layer in place, the microsimulation developed here can be used as a mechanism to propagate vessel-level interactions and behavioural changes into macro-level effects.

Therefore, a logical next step is to refine the developed modelling approach and to connect it to a macro-level network study, in which capacity is established from the type of model developed in this thesis. Such a translation would not only advance the model family to which this thesis contributes, but would also support future research on the HSCGC by enabling network-level analysis grounded in microscopic, data-driven vessel behaviour.

8.2.2. In Practice

The practical relevance of data-driven maritime microsimulation extends beyond its use as an infrastructure assessment tool. In an operational context, microsimulation can support traffic management by providing a structured way to explore the consequences of alternative traffic measures, operational rules, or control strategies under realistic traffic demand and interaction patterns. Rather than relying only on aggregate historical indicators, a calibrated microsimulation can be used to examine how local interventions may affect encounter rates, delays, and emergent traffic dynamics, and can help identify which scenarios merit closer attention by operators and policymakers.

In addition, microsimulation provides a natural basis for more interactive and training-oriented applications. A particularly promising direction is the "gamification" of microsimulation by coupling data-driven maritime traffic simulation models to full mission bridge simulators. Such an integration would enable human operators to interact with a realistic surrounding traffic field generated by the model, while the simulated traffic in turn responds to the evolving situation. This creates a parallel with infrastructure design practice: in the same way that simulators can be used to train crews around new port layouts or route changes, a coupled microsimulation-bridge-simulator setup could support training and evaluation around new obstacles, novel traffic arrangements, or revised operational procedures. In that sense, the modelling approach developed in this thesis can be viewed not only as an analytical instrument, but also as a potential foundation for decision support and scenario-based training in complex waterways.

9

Conclusions

Using the case of the construction of a Gate Complex in the Houston Ship Channel through Bolivar Roads, this study aims to contribute to the development of data-driven maritime traffic simulation models as a design tool for maritime infrastructures. This chapter first summarizes the findings of the previous chapters, and aims to answer the main research question posed in Chapter 1. Additionally, this chapter also formulates recommendations for future research.

9.1. Answers to Resarch Questions

Research Question 1: "*What is the current state of the art in data-driven ship traffic models and their potential adaptations for use in maritime infrastructure design?*"

Data-driven maritime traffic simulation modelling increasingly relies on microscopic simulation, because representing vessel behaviour at the individual level makes it possible to capture ship-ship and ship-obstacle interactions and to recover macroscopic traffic phenomena (e.g., flow, density, capacity) as emergent outcomes.

The literature shows three dominant paradigms for modelling microscopic behaviour: rules-based logic, mathematical law, and data-driven learning. Data-driven approaches, typically learned from AIS observations, can reproduce real-world manoeuvring patterns without requiring that these patterns are explicitly formalized in rules or equations. This makes them well-suited to capturing complex and context-dependent behaviour, including non-standard manoeuvres that may be difficult to prescribe a priori.

Purely data-driven models generalize poorly to modified environments, because no AIS data exist for conditions that have not yet occurred. Therefore, to function as design tools, data-driven models must be extended to safely handle previously unseen obstacles and interaction situations. Since many data-driven simulators are reinforcement-learning-based, *safe* reinforcement learning is a suitable extension direction; specifically, post-posed safety filtering can minimally correct unsafe actions (e.g., collision-inducing actions) while preserving the learned policy as much as possible. As established in Section 2.3, this strategy is also practically attractive because it can be added as a layer on top of existing RL maritime traffic simulators.

Research Question 2: "*What are the existing characteristic vessel movements, traffic patterns, and behavioural features observed from AIS data at the proposed site of the HSCGC that a maritime traffic model must be able to capture?*"

To define an empirical baseline for both model validation and infrastructure-impact assessment, AIS data from 2024 were processed into trajectories and analysed for spatial structure, OD flows, temporal variability, kinematic regimes, and manoeuvring behaviour. Traffic at the proposed HSCGC site is highly structured and channel-constrained: most movements align with the Houston, Galveston, and Texas City Ship Channels. Typical transits through the study area last approximately 30-35 minutes and cover roughly 7-10 km. OD flows are dominated by movements between the Gulf of Mexico and the ports of Houston, Galveston, and Texas City, while smaller but frequent flows (notably

ferries and tug-barge combinations) co-occupy constrained space with deep-draft transits. A persistent north-south Galveston-Point Bolivar ferry corridor crosses the dominant east-west channel traffic, generating recurring interaction hot-spots. Speed and direction analyses show two common operating regimes around 5 and 10 knots, and while overall traffic intensity is relatively stable across longer timescales, hourly variation in average speed indicates time-dependent operational conditions and interaction effects. Manoeuvring complexity is spatially heterogeneous: acceleration and deceleration concentrates near entrances and intersections, turning increases at lower speeds near junctions and anchorages, and encounter densities peak inside channels and at intersections (especially near ferry terminals). Environmental forcing is observable: North-south routes are more affected by east-west tidal drift than channel-aligned transits.

A maritime traffic model intended for this site must reproduce not only route structure and OD magnitudes, but also the interaction-driven kinematic and manoeuvring patterns that shape safety and efficiency under current conditions. The AIS-derived features therefore define concrete targets for calibration and validation, and establish what constitutes a credible "baseline" simulation against which HSCGC-induced changes can be assessed.

Therefore, the AIS analysis shows that the HSCGC site is best described as a structured but interaction-rich system in which heterogeneity, crossings, and localized manoeuvring hotspots are central. These characteristics provide the empirical benchmark for assessing whether a simulation model is realistic enough to support infrastructure design decisions.

Research Question 3: "*What are the characteristic vessel movements, traffic patterns, and behavioural features as observed under current conditions and predicted under future conditions by expert navigators that a maritime traffic simulation model must be able to capture?*"

Semi-structured interviews with expert navigators were used to complement AIS observations by eliciting operational constraints, encounter-handling logic, and expected behavioural and demand changes under an HSCGC scenario. The interviews confirm the channel-following character of Bolivar Roads, but stress that a credible model must distinguish between deep-draft vessels that are constrained to the marked channels and smaller-draft vessels that can deviate outside the channel when needed. Experts also emphasize that overtaking occurs only in limited deep-water sections (see Figure 4.2) and that overtaking feasibility depends on local congestion. Interaction behaviour is multi-modal: in addition to head-on meetings, the model must represent intersecting, merging, and diverging encounters at the GIWW, Houston Ship Channel, and Texas City Ship Channel junctions, including both port-to-port and coordinated starboard-to-starboard meetings in specific turning zones, with associated local speed adaptation. Finally, experts highlight that less predictable traffic (notably recreational and fishing vessels) increases uncertainty and can trigger avoidance manoeuvres even when collisions remain rare. Under future conditions with the HSCGC, experts expect systematic changes: inbound speed reductions near the gate (with reduced ability to slow on flood tide), a slight northward lateral bias induced by current and wind, and constrained encounters in the gate approaches (preference for centreline passage when unopposed; avoidance of head-on encounters). They also expect demand changes because a gate fragments anchorages north of the Houston Ship Channel, implying that scenario OD-patterns must be configurable.

The interviews imply that reproducing trajectory geometry alone is insufficient: the model must capture the decision constraints and encounter logic that generate safety-critical behaviour, and it must support explicit scenario specification for both geometry (obstacles) and demand (OD changes). These requirements are especially important in high-conflict junctions and in constrained gate approaches where infrastructure-induced changes are expected to be most pronounced.

In conclusion, the expert perspective translates into concrete modelling requirements: draft-dependent manoeuvring envelopes, operationally limited overtaking, junction-specific encounter handling (including non-standard coordinated meetings), uncertainty due to small-vessel traffic, and explicit scenario control over both behaviour (via obstacle interactions) and demand (via configurable OD-patterns).

Research Question 4: "*To what extent does the selected data-driven maritime traffic model and its extensions meet the performance criteria for both historical realism and situational adaptability?*"

ShipNaviSim and its extensions were evaluated with two objectives: historical realism, pursued via AIS-driven action inference and trajectory-level outputs, and situational adaptability, pursued by mak-

ing obstacle configurations and demand patterns explicit inputs and by introducing a safety layer for obstacle avoidance under modified conditions.

Across configurations, the evaluated models exhibit goal-seeking behaviour but do not consistently reproduce the observed distributions of kinematic behaviour. Simulated trajectories show substantially higher variability in drift and curvature, wider acceleration ranges, and action magnitudes implying higher speeds than historically observed. Trajectory-level metrics reflect this mismatch: models with a single goal yield high GC-ADE values (917 - 1323 m), and face validation indicates unrealistic heading-COG alignment. Consequently, under the tested configurations, the models do not reliably satisfy the behavioural requirements needed for a credible representation of Bolivar Roads (Requirements 8 - 20).

The results indicate that the current model family achieves partial historical realism at a qualitative level (plausible transits and motion classes) but remains limited in quantitative fidelity (speed, curvature, drift, acceleration statistics). Situational adaptability is enabled in principle through explicit scenario inputs and safety filtering, but the empirical evidence suggests that achieving robust, safety-critical behaviour under modified conditions requires further development before the model can serve as a dependable design tool.

In conclusion, the selected data-driven model and extensions provide a workable foundation for scenario-based simulation, but the present performance does not yet meet the combined criterion of statistical realism and reliable adaptability required for infrastructure design applications.

Research Question 5: "*Based on their performance, which developed model extensions demonstrate the highest efficacy in capturing critical vessel behaviours and should be prioritized for further investigation in order to apply data-driven maritime traffic simulations as design tools for maritime infrastructure design?*"

The extensions are designed to target complementary limitations: intermediate goals to improve route adherence (channel-following realism), and an MPC-based safety layer to enforce safe behaviour around both existing and newly introduced obstacles while deviating minimally from the learned policy. Intermediate goals provide the clearest improvement in channel-following performance, reducing GC-ADE to 397 m and concentrating trajectories along expected corridors. The MPC-based safety layer reduces obstacle violations and smooths control, which is essential for using the simulator under modified infrastructure configurations. In combined form, these extensions implement a layered navigation structure in which waypoint-based guidance supports global route structure while local safety constraints prevent unsafe actions.

Based on measured improvements and relevance to infrastructure design use, intermediate goals and the MPC-based safety layer should be prioritized for further investigation. However, additional tests with newly introduced obstacles show that obstacle entries can still increase and tend to occur during head-on encounters, even when safety interventions intensify. This indicates that robust constraint satisfaction under novel geometries and high-conflict interactions remains the main limitation. Future work should therefore focus on calibrating and stress-testing the safety layer in encounter-dense settings, improving the coupling between policy actions and safety corrections, and evaluating performance against the full behaviour requirement set (Requirements 8 - 23).

In conclusion, intermediate goals (for route realism) and an MPC-based safety layer (for safe scenario execution) demonstrate the highest efficacy and constitute the most promising path toward applying data-driven maritime traffic simulation as an infrastructure design tool, provided that robustness in head-on and high-density encounters is improved.

9.2. Answer to Main Research Question

Based on this information, the answer to the main research question: "*What are the requirements and characteristics of a data-driven maritime traffic simulation model that is useful to assess the impact of prospective maritime infrastructures on maritime traffic patterns in the context of designing the Houston Ship Channel Gate Complex?*" is as follows:

A data-driven maritime traffic simulation model is useful for assessing the impact of prospective maritime infrastructures in the context of the Houston Ship Channel Gate Complex (HSCGC) if it can reproduce the empirically observed structure and interaction dynamics of the existing traffic system with sufficient realism to serve as a credible baseline, and remain behaviourally plausible and

safe under *modified* waterway configurations and demand patterns that have no direct historical AIS equivalent.

First, such a model must be grounded in a microscopic and multi-agent simulation paradigm. Microscopic representation is necessary because infrastructure impacts emerge through local interaction mechanisms: ship-ship encounters, ship-obstacle avoidance, and conflict resolution in constrained geometry, which then aggregate into macroscopic outcomes such as throughput, congestion, speed regimes, and encounter exposure. The multi-agent character is essential because traffic patterns in Bolivar Roads are interaction-driven and structurally embedded (rather than incidental), meaning that individual decisions must be responsive to the behaviour of other vessels.

Second, the model must be data-driven in the sense that its nominal behavioural policy is learned from AIS data, enabling it to capture complex operational patterns without requiring exhaustive rules or explicit mathematical law. For Bolivar Roads, this includes reproducing the strong channel-constrained structure of movement, the relative magnitudes of dominant OD flows between the Gulf of Mexico and the ports of Houston, Galveston, and Texas City, and recurring crossing flows such as the Galveston-Point Bolivar ferry corridor. It must further reproduce characteristic kinematic regimes (notably the observed speed clustering around approximately 5 and 10 knots), spatial heterogeneity in manoeuvring demand (acceleration/deceleration and turning near entrances, junctions, and anchorages), and concentrated encounter hotspots inside channels and at intersections and ferry terminals. These AIS-derived characteristics define the minimum set of features required for model calibration and validation. Without them, infrastructure impact results would not be interpretable as deviations from current operational reality.

Third, the model must explicitly represent operational heterogeneity and encounter logic that govern safety-critical behaviour in the study area. Expert navigators indicate that credibility requires at least a distinction between deep-draft channel-bound traffic and smaller-draft vessels that can deviate outside the channel, because this difference determines feasible lateral position choices and avoidance options. The model must also permit overtaking where it is operationally feasible (and limited), and represent interaction modalities beyond head-on meetings, including merging, diverging, and crossing encounters at the GIWW, Houston Ship Channel, and Texas City Ship Channel junctions, potentially including coordinated non-standard meetings in specific turning zones. In addition, the model should include the influence of less predictable small craft (e.g., recreational and fishing vessels), as this behavioural variability increases uncertainty and can induce avoidance manoeuvres even when collisions are rare. Collectively, these elements ensure that the simulation reproduces not only where vessels go, but also why they make safety-relevant decisions in encounter-dense locations.

Fourth, and most critically for infrastructure design, the model must be situationally adaptable: it must accept obstacle configurations and demand patterns as explicit scenario inputs, rather than treating the environment as fixed and implicitly defined by historic AIS. This is necessary because the HSCGC introduces new geometry and constraints (gate approaches, constrained passages) and can induce demand shifts (e.g., through anchorage fragmentation north of the Houston Ship Channel) that cannot be inferred from pre-construction AIS alone. Therefore, the model must support scenario-specific OD-patterns and allow behavioural changes expected under the gate scenario to emerge in a plausible manner, including inbound speed reductions near the gate, constrained encounter handling in approaches (preference for centreline passage when unopposed; reduced tolerance for head-on encounters), and systematic lateral bias due to current and wind.

Finally, because purely data-driven policies cannot be expected to generalize safely to unseen infrastructure layouts, the model must include a mechanism for safe operation under novel conditions. A suitable characteristic is a safety layer or safety filter (e.g., in a safe reinforcement learning framework) that prevents unsafe actions such as obstacle collisions while minimally deviating from the learned AIS-driven behaviour. This layered structure, combining data-driven nominal control with explicit safety constraints and corrective action selection, is what allows the model to be both historically grounded and practically usable for evaluating prospective infrastructure.

9.3. Recommendations

Based on the conclusions presented in this chapter and the limitations presented in Chapter 8, the following recommendations for further research are made:

In order to improve model fidelity, a first priority is to increase the model's ability to represent the main drivers of behaviour variation that matter for design assessment. In its current form, environ-

mental forcing (wind and currents) is only implicitly present through AIS training data, which limits systematic testing under non-average or extreme conditions (e.g., strong tidal states or hurricanes). Future work should therefore make these exogenous conditions explicit scenario inputs and evaluate whether doing so reduces the observed instability in drift and improves speed and acceleration realism.

Second, the simulator should move beyond homogeneous ship representations. Both AIS analysis and expert interviews indicate that vessel heterogeneity is central at Bolivar Roads: deep-draft vessels are channel-bound, while smaller vessels can deviate and introduce additional uncertainty. A multi-class (or conditional) agent formulation, where manoeuvring envelopes, feasible lateral positioning, and interaction logic depend on vessel type, would better reflect operational reality and would also make scenario-based assessment more credible.

Third, waypoint- or route-sequence navigation should be treated as the default for constrained waterways. The intermediate-goal extension produced the clearest improvement in channel-following, and extending this into robust multi-goal routing (with sensitivity analysis for waypoint choice and explicit scenario control over OD-patterns and routes) is a direct path toward a model that can represent both current traffic structure and plausible re-routing under an HSCGC scenario.

Fourth, the safety layer should be strengthened and calibrated with the explicit objective of robust performance in encounter-dense situations. Although the MPC-based safety layer reduces obstacle violations in general, the remaining failure modes, particularly in head-on and high-conflict encounters under novel obstacle layouts, are exactly the cases that determine whether a model is usable for infrastructure design. Calibration against reference encounter cases (real or synthetic), systematic stress testing by encounter type, and improved coupling between the learned policy and safety corrections are therefore recommended.

Fifth, the model's kinematic fidelity and computational robustness require further development before design use. The current erratic fine-scale motion (unrealistic drift, unstable acceleration dynamics) and the observed numerical instability during full-dataset training limit both credibility and evaluability. Introducing stronger physical consistency constraints, performing systematic (hyper)parameter tuning, and addressing the sources of numerical instability are necessary steps to progress from proof-of-concept behaviour to design-capable simulation.

On the evaluation side, future studies should enforce strict out-of-sample testing (e.g., holding out entire tracks, time periods, or geographic sub-areas) to quantify generalization and avoid overly optimistic estimates caused by information leakage. In addition, evaluation should become more diagnostic: rather than relying mainly on aggregate scores, performance should be reported by scenario class (head-on, crossing, merging, overtaking), interaction intensity, and traffic density, supported by targeted inspection of representative successes and failures in the junctions and ferry corridor where operational complexity concentrates. Where the goal is to assess modified infrastructure conditions, expert-in-the-loop validation (for example via bridge-simulator sessions or structured expert review of encounter replays) is recommended to establish acceptance criteria for "design-capable" behaviour under HSCGC scenarios.

Finally, to answer HSCGC-oriented design questions, the modelling workflow should explicitly separate baseline calibration from scenario intervention, so that impacts are interpretable as deviations from a credible "current-conditions" reference. In parallel, the microscopic simulator should be connected to macro-level performance assessment, because network-wide effects and capacity constraints are ultimately the quantities of interest for infrastructure planning. A promising practical direction is to couple the microsimulation to full-mission bridge simulators, both as a validation mechanism and as a pathway toward decision support and scenario-based training around gate operations and traffic measures.

In summary, a data-driven maritime traffic simulation model is useful for HSCGC design assessment if it is a validated microscopic multi-agent AIS-driven simulator that reproduces the site-specific structure, kinematic regimes, and interaction hot-spots of Bolivar Roads, explicitly represents vessel heterogeneity and encounter logic identified by experts, and is extended with scenario configurability (obstacles and OD-demand) and a robust safety mechanism that enables credible behaviour under modified, previously unseen infrastructure geometries.

This thesis shows that data-driven maritime traffic microsimulation, augmented with safety extensions, is a promising basis for infrastructure-impact assessment. At the same time, the current model family is not yet sufficiently mature to function as a dependable design tool in a complex, interaction-rich setting such as Bolivar Roads. These recommendations therefore focus on the most direct steps

needed to improve behavioural realism and robustness, validate performance under genuinely new conditions, and connect microscopic simulations to the infrastructure design questions that are relevant to the HSCGC. This way, this study is a step towards making the Houston Ship Channel Gate Complex a piece of infrastructure that efficiently helps to create a safer Texas.

References

Alshiekh, M., Bloem, R., Ehlers, R., Könighofer, B., Niekum, S., & Topcu, U. (2017, September 3). Safe reinforcement learning via shielding. https://doi.org/10.48550/arXiv.1708.08611

Ames, A. D., Coogan, S., Egerstedt, M., Notomista, G., Sreenath, K., & Tabuada, P. (2019, March 27). Control barrier functions: Theory and applications. https://doi.org/10.48550/arXiv.1903.11199

Bansal, S., Chen, M., Herbert, S., & Tomlin, C. J. (2017, September 21). Hamilton-jacobi reachability: A brief overview and recent advances. https://doi.org/10.48550/arXiv.1709.07523

Basrur, C., Singh, A. J., Sinha, A., & Kumar, A. (2021). Ship-GAN: Generative modeling based maritime traffic simulator. *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 1755–1757.

Bellsolà Olba, X., Daamen, W., Vellinga, T., & Hoogendoorn, S. P. (2017). Network capacity estimation of vessel traffic: An approach for port planning. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *143*(5), 04017019. https://doi.org/10.1061/(ASCE)WW.1943-5460.0000400

Bellsolà Olba, X., Daamen, W., Vellinga, T., & Hoogendoorn, S. P. (2019). A method to estimate the capacity of an intersection of waterways in ports. *Transportmetrica A: Transport Science*, *15*(2), 1848–1866. https://doi.org/10.1080/23249935.2019.1652866

Burkley, G. B., Webb, D., & Pierce, J. J. (2022, November 30). *Ship simulation transit scenario modeling & ship pilot study of the maritime implications for commercial ship transits of the proposed houston ship channel gate complex* (Version 7.2). Locus LLC. Arnold, MD, USA.

Buşoniu, L., Babuška, R., & De Schutter, B. (2010). Multi-agent reinforcement learning: An overview [Series Title: Studies in Computational Intelligence]. In J. Kacprzyk, D. Srinivasan, & L. C. Jain (Eds.), *Innovations in multi-agent systems and applications - 1* (pp. 183–221, Vol. 310). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14435-6_7

Cheng, R., Orosz, G., Murray, R. M., & Burdick, J. W. (2019, March 21). End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. https://doi.org/10.48550/arXiv.1903.08792

Daguano, R. F., Yoshioka, L. R., Netto, M. L., Marte, C. L., Isler, C. A., Santos, M. M. D., & Justo, J. F. (2023). Automatic calibration of microscopic traffic simulation models using artificial neural networks. *Sensors (Basel, Switzerland)*, *23*(21), 8798. https://doi.org/10.3390/s23218798

Dalal, G., Dvijotham, K., Vecerik, M., Hester, T., Paduraru, C., & Tassa, Y. (2018, January 26). Safe exploration in continuous action spaces. https://doi.org/10.48550/arXiv.1801.08757

Dawood, M., Shokry, A., & Bennewitz, M. (2024, December 5). A dynamic safety shield for safe and efficient reinforcement learning of navigation tasks [version: 1]. https://doi.org/10.48550/arXiv.2412.04153

Delft High Performance Computing Centre (DHPC). (2024). *DelftBlue supercomputer (phase 2)*. https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2

Desyani, R. (2019). Comparison of paradigms in nautical traffic models. Retrieved June 24, 2025, from https://repository.tudelft.nl/record/uuid:76029292-c011-4bba-a313-2a28de3be380

Düz, B., & van Iperen, E. (2025). Ship trajectory prediction using encoder–decoder-based deep learning models [_eprint: https://doi.org/10.1080/17489725.2024.2306339]. *Journal of Location Based Services*, *19*(2), 146–166. https://doi.org/10.1080/17489725.2024.2306339

Fisac, J. F., Akametalu, A. K., Zeilinger, M. N., Kaynama, S., Gillula, J., & Tomlin, C. J. (2018, February 14). A general safety framework for learning-based control in uncertain robotic systems. https://doi.org/10.48550/arXiv.1705.01292

Fujii, Y., & Tanaka, K. (1971). Traffic capacity. *Journal of Navigation*, *24*(4), 543–552. https://doi.org/10.1017/S0373463300022384

Goodwin, E. M. (1975). A statistical study of ship domains. *Journal of Navigation*, *28*(3), 328–344. https://doi.org/10.1017/S0373463300041230

Greater Houston Port Bureau. (2023a). Coastal spine must balance storm protection with safe navigation and long-term economic vitality of the port of houston. Retrieved July 11, 2025, from

https://assets.noviams.com/novi-file-uploads/ghpb/images/Coastal_Barrier/EPCO_IkeDike_Brochure_Final_pdf_11_28_rev-13af9cd3.pdf

Greater Houston Port Bureau. (2023b). *Proposed houston ship channel gate complex - greater houston port bureau*. Retrieved July 11, 2025, from https://www.txgulf.org/houston-ship-channel-gate-complex

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., … DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, *198*(1), 115–126. https://doi.org/10.1016/j.ecolmodel.2006.04.023

Koenighofer, B., Bloem, R., Jansen, N., Bochum, R.-U., Junges, S., & Pranger, S. (2024). Shields for safe reinforcement learning.

Konstapel, S. (2026). *Quantitative assessment of navigational safety in storm surge barrier design: A systematic maneuvering-based approach for early evaluation of alternative configurations* [Master Thesis]. Delft University of Technology. https://resolver.tudelft.nl/uuid:f1788d5a-a049-49e5-88b2-30e359965576

Li, R., Li, T., Bu, R., Zheng, Q., & Chen, C. L. P. (2013). Active disturbance rejection with sliding mode control based course and path following for underactuated ships. *Mathematical Problems in Engineering*, *2013*, 1–9. https://doi.org/10.1155/2013/743716

Mathioudakis, M., Papandreou, C., Stouraitis, T., Margari, V., Nikitakis, A., Paschalakis, S., Kyriakopoulos, K., & Spyrou, K. J. (2025, January 23). Towards real-world validation of a physics-based ship motion prediction model. https://doi.org/10.48550/arXiv.2501.13804

NOAA. (n.d.-a). AccessAIS. Retrieved September 3, 2025, from https://marinecadastre.gov/accessais/

NOAA. (n.d.-b). *Bathymetric data viewer*. Retrieved August 28, 2025, from https://www.ncei.noaa.gov/maps/bathymetry/

NOAA. (2025, November 17). *NOAA ENC® - electronic navigational charts*. Retrieved November 17, 2025, from https://nauticalcharts.noaa.gov/charts/noaa-enc.html

Odriozola-Olalde, H., Zamalloa, M., & Arana-Arexolaleiba, N. (2023). Shielded reinforcement learning: A review of reactive methods for safe learning. *2023 IEEE/SICE International Symposium on System Integration (SII)*, 1–8. https://doi.org/10.1109/SII55687.2023.10039301

OpenStreetMap. (2025). Retrieved September 11, 2025, from openstreetmap.org

Pan, R., Zhang, W., Wang, S., & Kang, S. (2025). Deep reinforcement learning model for multi-ship collision avoidance decision making design implementation and performance analysis. *Scientific Reports*, *15*, 21250. https://doi.org/10.1038/s41598-025-05636-3

Papandreou, C., Mathioudakis, M., Stouraitis, T., Iatropoulos, P., Nikitakis, A., Paschalakis, S., & Kyriakopoulos, K. (2025, March 3). Interpretable data-driven ship dynamics model: Enhancing physics-based motion prediction with parameter optimization. https://doi.org/10.48550/arXiv.2502.18696

Pham, Q. A., Brahmanage, J. C., & Kumar, A. (2025). ShipNaviSim: Data-driven simulation for real-world maritime navigation. https://www.ifaamas.org/Proceedings/aamas2025/pdfs/p1641.pdf

Sheebaelhamd, Z., Zisis, K., Nisioti, A., Gkouletsos, D., Pavllo, D., & Kohler, J. (2021, August 11). Safe deep reinforcement learning for multi-agent systems with continuous action spaces. https://doi.org/10.48550/arXiv.2108.03952

Shu, Y. (2019). *Vessel route choice model and operational model based on optimal control* [Dissertation (TU Delft)]. TRAIL Research School. https://doi.org/10.4233/uuid:5e700fc1-7620-4ab0-9b72-859e2db7926b

Texas Deartment of Transportation. (n.d.). *Galveston-port bolivar ferry*. Retrieved December 28, 2025, from https://www.txdot.gov/discover/ferry-boat-schedules/galveston-port-bolivar-ferry.html

Wabersich, K. P., & Zeilinger, M. N. (2019, April 8). Linear model predictive safety certification for learning-based control. https://doi.org/10.48550/arXiv.1803.08552

Wang, Y., Wen Liu, R., Liu, J., Yang, L., & Liu, Y. (2025). AIS data-driven maritime traffic flow prediction and density visualization using multitime scale temporal feature fusion network. *IEEE Sensors Journal*, *25*(7), 11357–11365. https://doi.org/10.1109/JSEN.2024.3525094

Yip, T. L. (2013). A marine traffic flow model. *TransNav, International Journal on Marine Navigation and Safety od Sea Transportation*, *7*(1), 109–113. https://doi.org/10.12716/1001.07.01.14

Zhou, Y. (2022). *Ship behavior in ports and waterways: An empirical perspective* [Doctoral dissertation, Delft University of Technology]. https://doi.org/10.4233/UUID:DAB6C5BB-5DAA-496A-A912-8C3DFAFF0EBE

Zhou, Y., Daamen, W., Vellinga, T., & Hoogendoorn, S. (2019). Review of maritime traffic models from vessel behavior modeling perspective. *Transportation Research Part C: Emerging Technologies*, *105*, 323–345. https://doi.org/10.1016/j.trc.2019.06.004

A

Data Filtering Statistics

This table presents the full data filtering statistics as described in Section 3.1.

Table A.1: Distribution of transmissions throughout filtering process by vessel type and status code.

	Raw Data		Filtered Data			Processed Data		
	Points	Share (%)	Points	% Retained	% of original	Points	% Retained	% of original
AIS Transimissions								
AIS Transmissions	12042357	100.00%	3661084	30.40%	30.40%	2501836	68.34%	20.78%
Unique Ships	5525	100.00%	4768	86.30%	86.30%	4355	91.34%	78.82%
Tracks	-	-	139268	-	-	79199	56.87%	-
Transmissions by Vessel Type								
Cargo	843157	7.00%	216732	25.70%	1.80%	188013	86.75%	1.56%
Fishing	435473	3.62%	70164	16.11%	0.58%	32752	46.68%	0.27%
Military	10649	0.09%	0	0.00%	0.00%	0	0.00%	0.00%
Other	2675638	22.22%	510627	19.08%	4.24%	408135	79.93%	3.39%
Passenger	1789360	14.86%	624292	34.89%	5.18%	33196	5.32%	0.28%

Continued on next page...

		Raw Data		Filtered Data			Processed Data		
		Points	Share (%)	Points	% Retained	% of original	Points	% Retained	% of original
	Pilot Vessel	386896	3.21%	159037	41.11%	1.32%	149352	93.91%	1.24%
	Pleasure Craft	186299	1.55%	27034	14.51%	0.22%	19571	72.39%	0.16%
	Tanker	1167708	9.70%	509215	43.61%	4.23%	509054	99.97%	4.23%
	Tug	4425625	36.75%	1526539	34.49%	12.68%	1146966	75.14%	9.52%
	Unknown	121552	1.01%	17444	14.35%	0.14%	14797	84.83%	0.12%
	TOTAL	12042357	100.00%	3661084		30.40%	2501836		20.78%
Transmissions by Status Code									
0	Under way using engine	6100149	50.66%	2459630	40.32%	20.42%	1516617	61.66%	12.59%
1	At anchor	888012	7.37%	0	0.00%	0.00%	0	0.00%	0.00%
2	Not under command	11283	0.09%	0	0.00%	0.00%	0	0.00%	0.00%
3	Restricted manoeuverability	189707	1.58%	141166	74.41%	1.17%	131676	93.28%	1.09%
4	Constrained by her draught	3672	0.03%	1393	37.94%	0.01%	1100	78.97%	0.01%
5	Moored	1969973	16.36%	0	0.00%	0.00%	0	0.00%	0.00%
6	Aground	429	0.00%	0	0.00%	0.00%	0	0.00%	0.00%
7	Engaged in Fishing	6383	0.05%	4375	68.54%	0.04%	1998	45.67%	0.02%
8	Under way sailing	26881	0.22%	0	0.00%	0.00%	0	0.00%	0.00%
9	Reserved for future amendment of Navigational Status for HSC	252214	2.09%	0	0.00%	0.00%	0	0.00%	0.00%
10	Reserved for future amendment of Navigational Status for WIG	3125	0.03%	0	0.00%	0.00%	0	0.00%	0.00%
11	Tug Towing Astern	41798	0.35%	0	0.00%	0.00%	0	0.00%	0.00%
12	Tug Towing Ahead or Aside	1652183	13.72%	974931	59.01%	8.10%	822279	84.34%	6.83%
13	Reserved for future use	1666	0.01%	0	0.00%	0.00%	0	0.00%	0.00%
14	AIS-SART is active	109	0.00%	0	0.00%	0.00%	0	0.00%	0.00%
15	Not defined (default)	417475	3.47%	79589	19.06%	0.66%	28166	35.39%	0.23%
N/A	Unknown	477298	3.96%	0	0.00%	0.00%	0	0.00%	0.00%
	TOTAL	12042357	100.00%	3661084		30.40%	2501836		20.78%

B

Origin-Destination Matrix for Bolivar Roads (2024)

This appendix provides the origin-destination matrix of movements between zones as shown in Figure 3.7.

Table B.1: Origin–Destination matrix of all passages included in the filtered dataset through the study area in 2024.

	Bolivar Roads Anchorage	Point Bolivar	GIWW EAST	Houston Ship Channel	Texas City	GIWW WEST	Galveston	Gulf
Bolivar Roads Anchorage	2590	194	50	1140	225	75	3216	2376
Point Bolivar	229	614	324	169	119	103	18314	222
GIWW EAST	47	70	1382	6760	1463	2220	779	63
Houston Ship Channel	909	315	6693	246	1452	3251	2066	8232
Texas City	187	143	1230	1556	184	385	1674	936
GIWW WEST	73	125	2297	3087	488	1004	306	130
Galveston	3183	18377	770	2095	1497	336	2876	8542
Gulf	2772	201	61	8012	883	113	9048	781

C

Description of Outliers in Monthly Vessel Count Patterns

As shown in Figure 3.11, the average vessel counts in Bolivar Roads during the months of January and March appear to deviate from the overall pattern. Figure C.1 highlights the deviation of the average traffic count over the day during these months. This figure shows that although March sees a noticeable peak in traffic, and January a noticeable dip, the averages over these months still lie within 1 standard deviation of the overall mean.

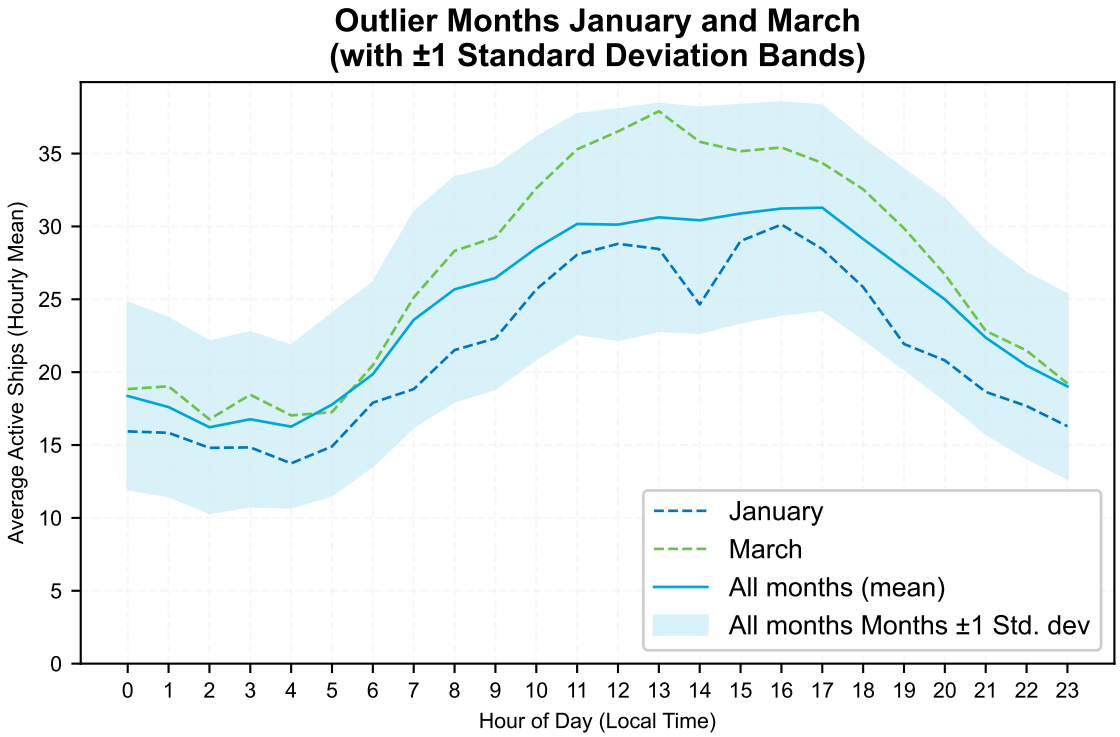


Figure C.1: Deviation of monthly average traffic volume during the months January and March

D

List of Interview Questions

The following is a list of interview questions that were used as guidance during the semi-structured interviews as described in Chapter 4.

Strategic Navigational Decisions

- How do you plan and execute a passage through Bolivar Roads?
- Which information do you need before and during the passage?
- Which routes do you choose, what information influences your choice?
- Can you describe a situation where you had to change your plan in a significant way?
- How do you verify that the route you chose remains viable during the passage?

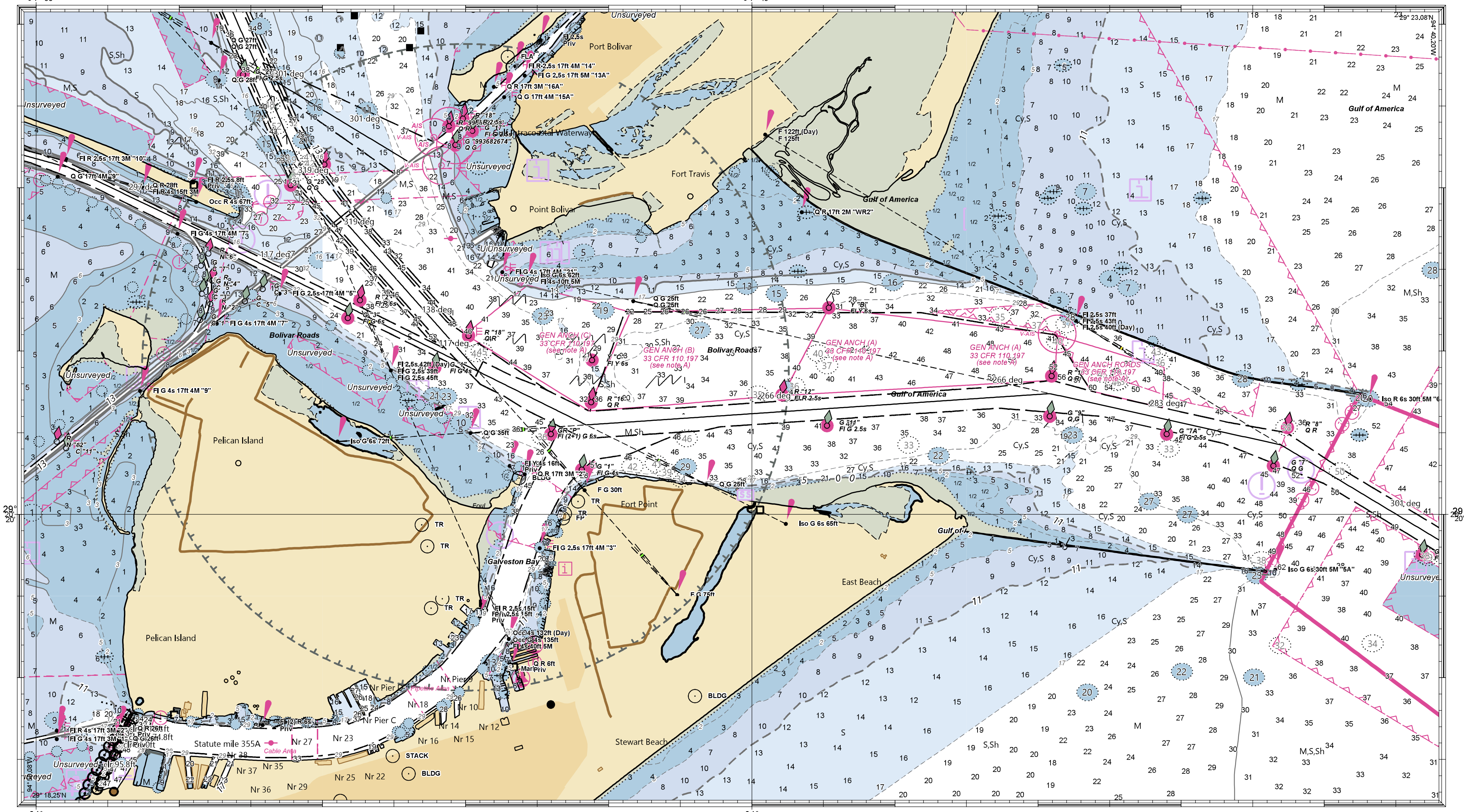
Handling Traffic

- Which flows of ship traffic do you pay attention to when coming from a certain area?
- Which patterns are most predictable, and which patterns are more uncertain?
- Where are the current chokepoints in the area?
- How do you handle oncoming and crossing traffic?
- How do you communicate with other ships you encounter during a passage?
- What is the role of VTS during your passages? What influence does their advice have on your movements?
- How do you prioritize vessels when multiple conflicts arise simultaneously?
- How do you handle situations where other vessels do not follow COLREGS or local navigational rules?
- How do you determine when to take early and bold action versus small course corrections?

Handling obstacles

- How would you handle an obstacle such as the HSCGC?
- In case the HSCGC is realized, how would your route choices and navigational decisions change?
- What should a pilot or captain pay extra attention to when navigating through such a structure?
- Can you sketch on this map how you would handle the HSCGC?
- How do you reassess risk when passing close to fixed structures or ongoing construction zones?
- Which hydrodynamic effects are you most vigilant for when navigating near breakwaters, jetties, or narrow channel edges?

Bolivar Roads



(533.4 X 297.2mm)

To ensure that this chart was printed at the proper scale, the line above should measure six inches (152 millimeters).

WGS 84

Generation Date: 11/17/2025

Page Size: A2 - 420 x 594 mm (Landscape)

UNITS IN FEET

CAUTION
AUTOMATED CHART GENERATION

This NOAA Custom Chart has been automatically rendered from NOAA Electronic Navigational Chart (NOAA ENC®) data. Mariners using this NOAA Custom Chart are advised that this is a static reproduction of the NOAA ENC®. This NOAA Custom Chart has not been individually quality checked or adjusted for optimal use for navigation. The portrayal may be at a different scale from that of the original NOAA ENC®. Mariners are advised to use caution when using this NOAA Custom Chart for navigation and are encouraged to use the latest NOAA ENC® to access the most up-to-date information. Mariners must also comply with all applicable regulatory requirements.

HEIGHTS

Heights of fixed aids to navigation and vertical clearances of overhead obstructions will be shown in feet if the units are set to feet or fathoms. If units are set to meters, heights will be shown in meters. Land elevation values are shown in meters only.

CAUTION
CHART UPDATES

This NOAA Custom Chart contains up-to-date information only as of the time of creation, and will become outdated. Mariners are advised to visit https://distribution.charts.noaa.gov/navigation-updates/ to check for critical and routine updates, and to render a new NOAA Custom Chart when the ENC data used to make the chart is updated. Notices to Mariners are not issued for corrections to this NOAA Custom Chart.

WARNING

The prudent mariner will not rely solely on any single aid to navigation, particularly on floating aids. See U.S. Coast Guard Light List and U.S. Coast Pilot for details.

Projection Information
WGS 1984 World Mercator (Calculated)
SCALE 1:30000 at Lat. 29°20.66'
GCS WGS 1984

Additional information can be obtained at www.nauticalcharts.noaa.gov.

Nautical Miles: 0 0.35 0.7 1.05 1.4

Meters: 0 740 1,480 2,220 2,960

AUTHORITIES

Hydrography and topography by the National Ocean Service, Coast Survey, with additional data from the Corps of Engineers, Geological Survey, U.S. Coast Guard and National Geospatial-Intelligence Agency.

POLLUTION REPORTS

Report all spills of oil and hazardous substances to the National Response Center via 1-800-424-8802 (toll free), or to the nearest U.S. Coast Guard facility if telephone communication is impossible (33 CFR 153).

WATER LEVELS, CURRENTS, AND TIDES

Real-time water levels, tide predictions, and tidal current predictions are available on the internet from NOAA's Center for Operational Oceanographic Products and Services (CO-OPS) at https://tidesandcurrents.noaa.gov/water_level_info.html and https://tidesandcurrents.noaa.gov/currents_info.html .

FATHOMS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
FEET	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102

METERS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31

NOAA encourages users to submit inquiries, discrepancies, or comments about this chart via NOAA's ASSIST tool at https://www.nauticalcharts.noaa.gov/customer-service/assist/ .

F

Derivation of Constraint 6.17

This appendix provides the derivation of the ray distance constraint as defined in Equation 6.17.

Let:

- $\mathbf{s}_t^k$ be the state of the ship, which contains the position (x_t^k, y_t^k) of the ship.
- P be the set of obstacles.
- $\mathbf{a}_t^k$ be the change in position of the ship between t and $t+1$.
- $\hat{\mathbf{r}}$ be the unit vector from the position of the ship along the direction of the ray such that $l_{\mathbf{s}_t^k} = \mathbf{s}_t^k + \hat{\mathbf{r}}q$
- $dist(\mathbf{s}_t^k)$ be the distance along the ray $m_{\mathbf{s}_t^k}$ from the position of the ship to the first intersection with the shoreline such that $dist(\mathbf{s}_t^k) = \min\{q \geq 0 | m_{\mathbf{s}_t^k}(q) \in P\}$

Using the definition of the $dist(\mathbf{s}_t^k)$ such that it is the point with the smallest q where the ray m intersects the set of obstacles P, Suppose that the ship moves by a small vector $\mathbf{a}_t^k$. Now, we are interested in the change in distance along the ray, such that:

$$dist(\mathbf{s}_{t+1}^k) = dist(\mathbf{s}_t^k + \mathbf{a}_t^k)$$

If we decompose the displacement $\mathbf{a}_t^k$ into its components in $\hat{\mathbf{r}}$ and $\hat{\mathbf{r}}_\perp$ (see Figure F.1), we can write the displacement as the sum of the components parallel to the ray and perpendicular to the ray:

$$\mathbf{a}_t^k = (\hat{\mathbf{r}}^\top \mathbf{a}_t^k)\hat{\mathbf{r}} + (\hat{\mathbf{r}}_\perp^\top \mathbf{a}_t^k)\hat{\mathbf{r}}_\perp$$

Now, assuming that the motion perpendicular to the ray has zero first-order effect, it can be assumed that moving along the ray towards the obstacle reduces the distance to the obstacle along this ray by as follows:

$$dist(\mathbf{s}_t^k + (\hat{\mathbf{r}}^\top)\mathbf{a}_t^k\hat{\mathbf{r}}) = dist(\mathbf{s}_t^k) - \hat{\mathbf{r}}^\top \mathbf{a}_t^k$$

Assuming that the effect of the component perpendicular to the ray has zero effect (see Figure F.1):

$$dist(\mathbf{s}_t^k + (\hat{\mathbf{r}}_\perp^\top \mathbf{a}_t^k)\hat{\mathbf{r}}_\perp) \approx dist(\mathbf{s}_t^k)$$

Then, by superposition:

$$dist(\mathbf{s}_t^k + \mathbf{a}_t^k) \approx dist(\mathbf{s}_t^k) - \hat{\mathbf{r}}^\top \mathbf{a}_t^k$$

Hence, a linear constraint can be written in the following form:

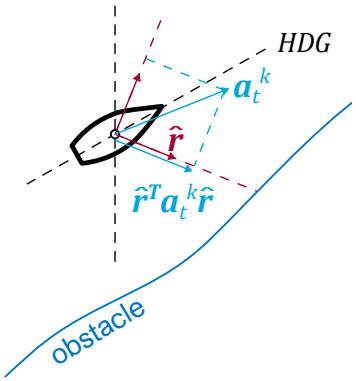


Figure F.1: Projection of the action $\mathbf{a}_t^k$ on a ray along unit vector $\hat{\mathbf{r}}$.

$$dist(\mathbf{s}_t^k) - \hat{\mathbf{r}}^\top \mathbf{a}_t^k \geq dist_{safe}$$

And by introducing slack penalty ζ_{mt}, this constraint can be written as:

$$dist(\mathbf{s}_t^k) - \hat{\mathbf{r}}^\top \mathbf{a}_t^k + \zeta_{mt} \geq dist_{safe}$$

G

Comparison of Model Results

This table presents the full set of descriptive statistics as described in Section 7.3.

Table G.1: Descriptive statistics of simulation results.

	Historic data	1: Full Set - 166 Epochs	2: Limited Set - 300 Epochs	3: Limited Set - 300 Epochs + Safety	4: Limited Set - 300 Epochs + Safety + Extra Goals
Executed dx (m)					
Min	-180.77	-824.08	-822.05	-820.75	-800.75
Max	179.60	824.08	805.14	809.90	798.22
Median	2.83	-98.80	41.98	0.04	0.00
Mean	-0.06	2.38	8.32	1.27	-1.99
St. Dev.	41.15	628.97	409.62	148.85	94.65
Executed dy (m)					
Min	-157.31	-389.18	-389.18	-389.18	-389.18
Max	198.16	389.18	389.18	389.18	389.18

Continued on next page...

	Historic data	1: Full Set - 166 Epochs	2: Limited Set - 300 Epochs	3: Limited Set - 300 Epochs + Safety	4: Limited Set - 300 Epochs + Safety + Extra Goals
Median	-0.15	-46.33	-11.50	0.04	0.00
Mean	-0.34	-3.80	2.28	0.74	-0.47
St. Dev.	24.10	291.35	219.26	77.76	64.25
Executed dh (°)					
Min	-30.71	-8.61	-15.93	-16.39	-14.43
Max	28.87	14.18	11.50	12.53	13.43
Median	0.00	2.47	0.11	-0.27	-0.09
Mean	0.00	2.72	-0.07	-0.13	0.01
St. Dev.	1.45	4.01	2.97	1.36	1.19
Action Magnitude $\|\|\mathbf{a}_t^k\|\|$					
Min	0.01	27.16	0.31	0.00	0.00
Max	238.78	911.40	906.15	906.71	883.22
Median	37.49	734.44	300.33	2.47	3.15
Mean	41.79	651.19	369.62	55.19	44.96
St. Dev.	23.02	237.64	281.64	158.62	105.22
Correction Magnitude $\|\|\mathbf{a}_t^k - \mathbf{a}_t^{k\prime}\|\|$					
Min	0.00	0.00	0.00	0.00	0.00
Max	0.00	0.00	0.00	334.63	328.24
Median	0.00	0.00	0.00	63.89	97.77
Mean	0.00	0.00	0.00	92.78	109.59
St. Dev.	0.00	0.00	0.00	73.16	82.73
Drift (°)					
Min	-179.99	-179.99	-179.96	-180.00	-180.00
Max	180.00	179.99	179.98	180.00	180.00
Median	0.37	-23.11	-24.41	2.88	-18.00

Continued on next page...

	Historic data	1: Full Set - 166 Epochs	2: Limited Set - 300 Epochs	3: Limited Set - 300 Epochs + Safety	4: Limited Set - 300 Epochs + Safety + Extra Goals
Mean	0.95	318.15	333.09	7.43	319.39
St. Dev. (circular)	16.63	89.07	76.45	115.05	110.71
Track Curvature (°/s)					
Min	-18.00	-17.77	-17.94	-18.00	-18.00
Max	18.00	17.64	17.98	18.00	18.00
Median	0.00	0.00	-0.06	0.00	0.00
Mean	360.00	0.01	359.88	359.60	359.85
St. Dev.	8.02	34.15	25.09	11.74	11.01
Acceleration (m/s^2)					
Min	-1.39	-8.40	-7.65	-7.59	-7.71
Max	1.84	7.94	7.51	7.54	8.27
Median	0.00	-0.27	-0.14	0.00	0.00
Mean	0.00	-0.42	-0.22	-0.03	-0.02
St. Dev.	0.02	1.03	0.87	0.39	0.58
GC-ADE (m)					
Min	0.00	116.46	72.32	15.37	18.17
Max	0.00	2343.39	2199.76	2219.78	1365.89
Median	0.00	1324.93	1051.52	919.88	328.61
Mean	0.00	1323.13	1053.15	917.85	397.07
St. Dev.	0.00	463.29	463.71	497.82	234.63
Obstacle Entries					
Count	0	6334	7340	5889	1174

H

Obstacle Entry Points by Simulation

This appendix contains the figures showing entries into obstacles per model as described in Section 7.3.

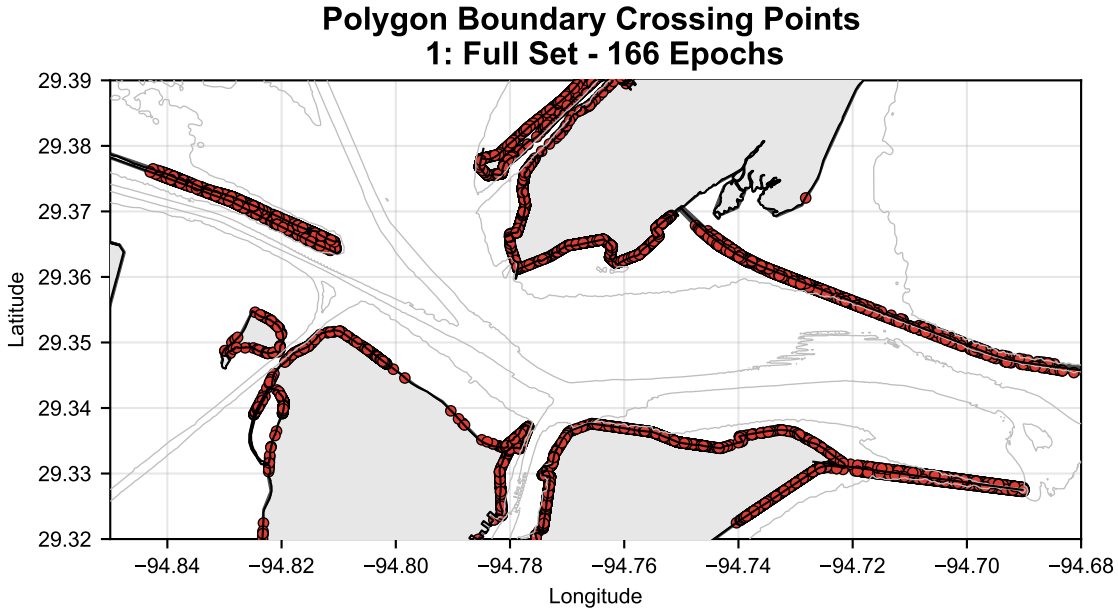


Figure H.1: Entry points observed using model 1: Full Set - 166 Epochs

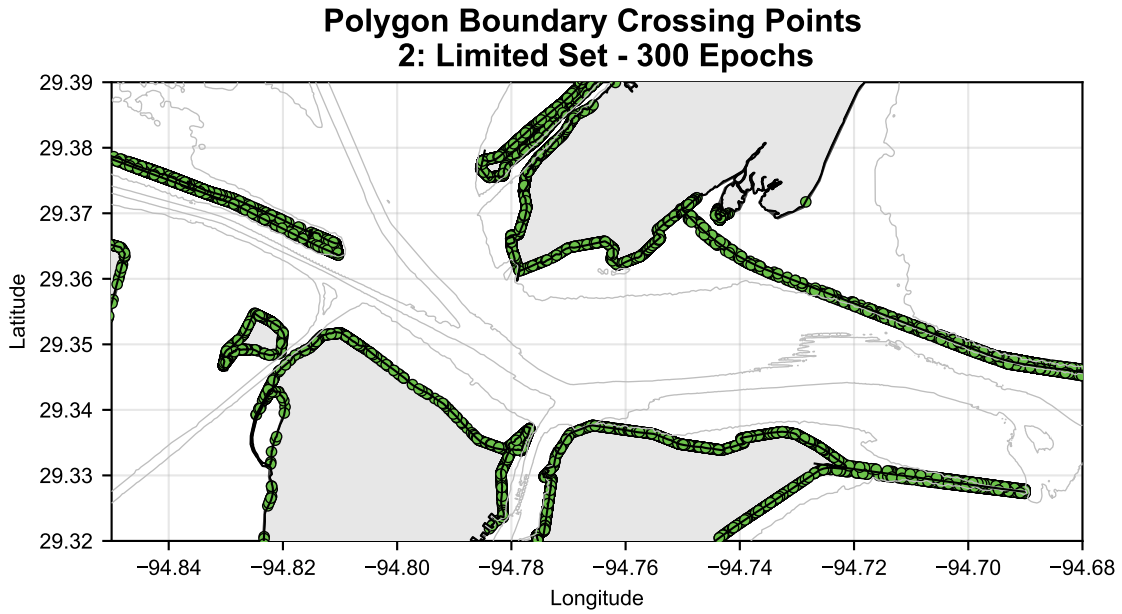


Figure H.2: Entry points observed using model 2: Limited Set - 300 Epochs

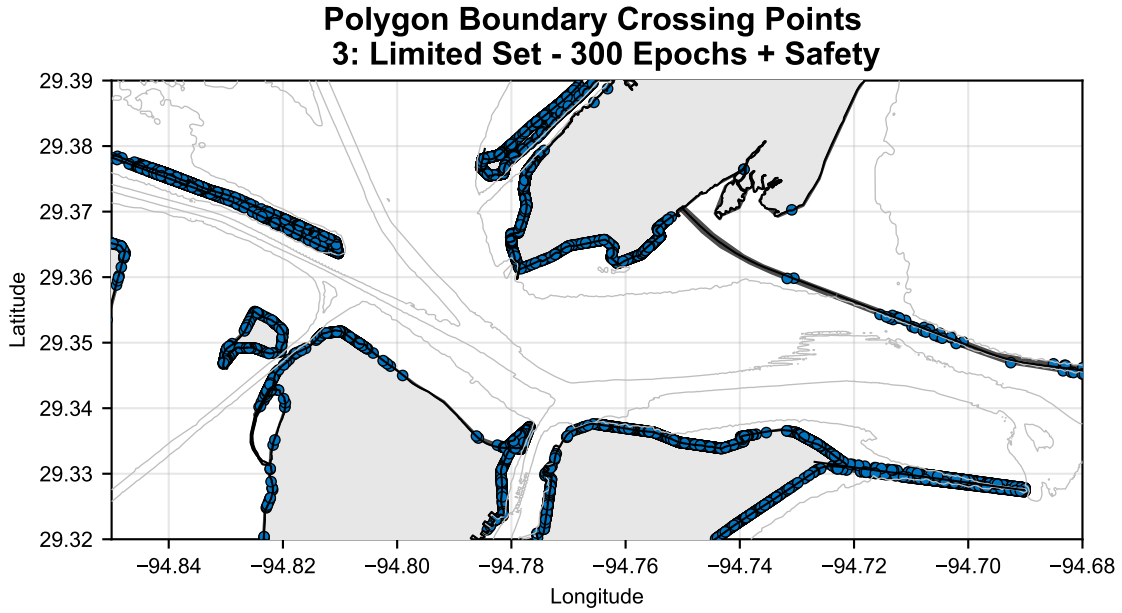


Figure H.3: Entry points observed using model 3: Limited Set - 300 Epochs + Safety

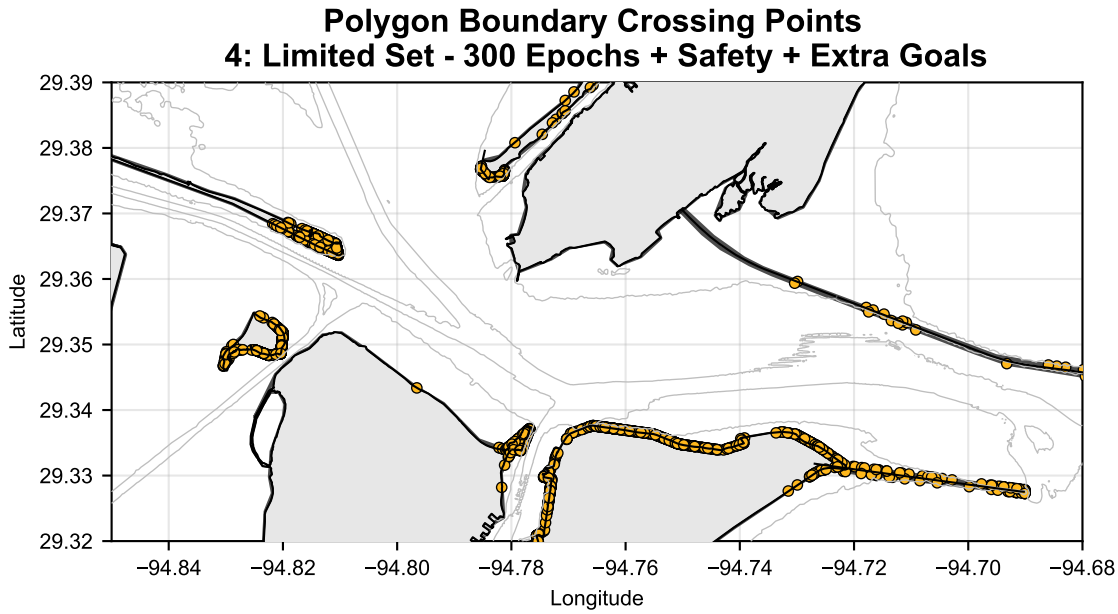


Figure H.4: Entry points observed using model 4: Limited Set - 300 Epochs + Safety + Extra Goals

Post-Hoc Safety Shield Parameter Sensitivity Analysis Results

This appendix provides the full set of results of the post-hoc safety shield parameter sensitivity analysis presented in Section 7.4. This appendix first presents tables of GC-ADE on each track, followed, track-level visualizations of the 12 test-cases presented in Figure 7.5.

Case-by-case GC-ADE Scores

Table I.1: Left: GC-ADE on the test cases for selected values of λ_{mt}. Right: GC-ADE on the test cases for selected values of M.

	$\lambda_{mt} = 100 \cdot ...$			
Test Case	$(1 - 0.7\sin(\varphi_m)^2)$	$(0.5 - 0.5\sin(\varphi_m)^2)$	$(0.6 + 0.4\cos(\varphi_m))$	$(0.5 + 0.5\cos(\varphi_m))$
1	621.29	626.88	593.89	287.04
2	200.48	196.68	115.99	112.62
3	650.48	650.46	652.44	647.93
4	729.97	729.97	729.97	729.97
5	101.73	101.56	310.59	354.10
6	348.96	222.20	403.22	438.47
7	176.35	173.93	208.74	500.80
8	506.45	506.45	506.92	530.53
9	226.28	228.00	223.44	231.10
10	638.82	630.55	664.86	553.44
11	483.60	469.31	478.92	511.55
12	71.59	155.34	158.45	163.41

	M			
Test Case	8	12	24	36
1	605.43	621.29	606.35	622.95
2	113.83	200.48	124.91	115.34
3	651.62	650.48	652.22	653.22
4	729.97	729.97	729.97	729.97
5	282.07	101.73	84.45	161.54
6	373.77	348.96	225.69	226.38
7	208.59	176.35	168.27	158.92
8	506.96	506.45	506.45	506.45
9	226.79	226.28	228.59	228.05
10	613.85	638.82	604.00	594.69
11	469.45	483.60	515.21	522.80
12	156.25	71.59	148.53	94.31

Table I.2: Left: GC-ADE on the test cases for selected values of N. Right: GC-ADE on the test cases for selected values of $dist_{safe}$.

	N		
Test Case	0	5	10
1	284.18	621.29	621.29
2	196.18	200.48	200.48
3	632.83	650.48	650.48
4	729.97	729.97	729.97
5	302.14	101.73	101.73
6	413.12	345.90	348.96
7	390.97	171.93	176.35
8	530.53	506.44	506.45
9	240.49	226.28	226.28
10	690.09	631.61	638.82
11	466.89	483.60	483.60
12	170.10	71.59	71.59

	$dist_{safe}$					
Test Case	50 m	100 m	150 m	200 m	250 m	300 m
1	224.61	224.00	621.29	619.40	615.52	620.06
2	191.44	191.05	200.48	199.12	202.24	196.39
3	647.68	650.19	650.48	650.95	665.72	652.16
4	729.97	729.97	729.97	729.97	729.97	729.97
5	308.03	102.65	101.73	101.82	292.60	102.91
6	450.95	475.46	348.96	410.19	449.29	463.84
7	204.04	179.80	176.35	202.89	159.49	167.45
8	506.91	506.44	506.45	530.41	474.58	474.48
9	223.23	224.71	226.28	227.72	220.97	222.34
10	666.33	657.82	638.82	642.11	600.10	600.55
11	489.88	494.89	483.60	122.47	132.03	131.74
12	69.40	69.90	71.59	154.93	72.61	71.75

Case-by-case Comparison of Tracks

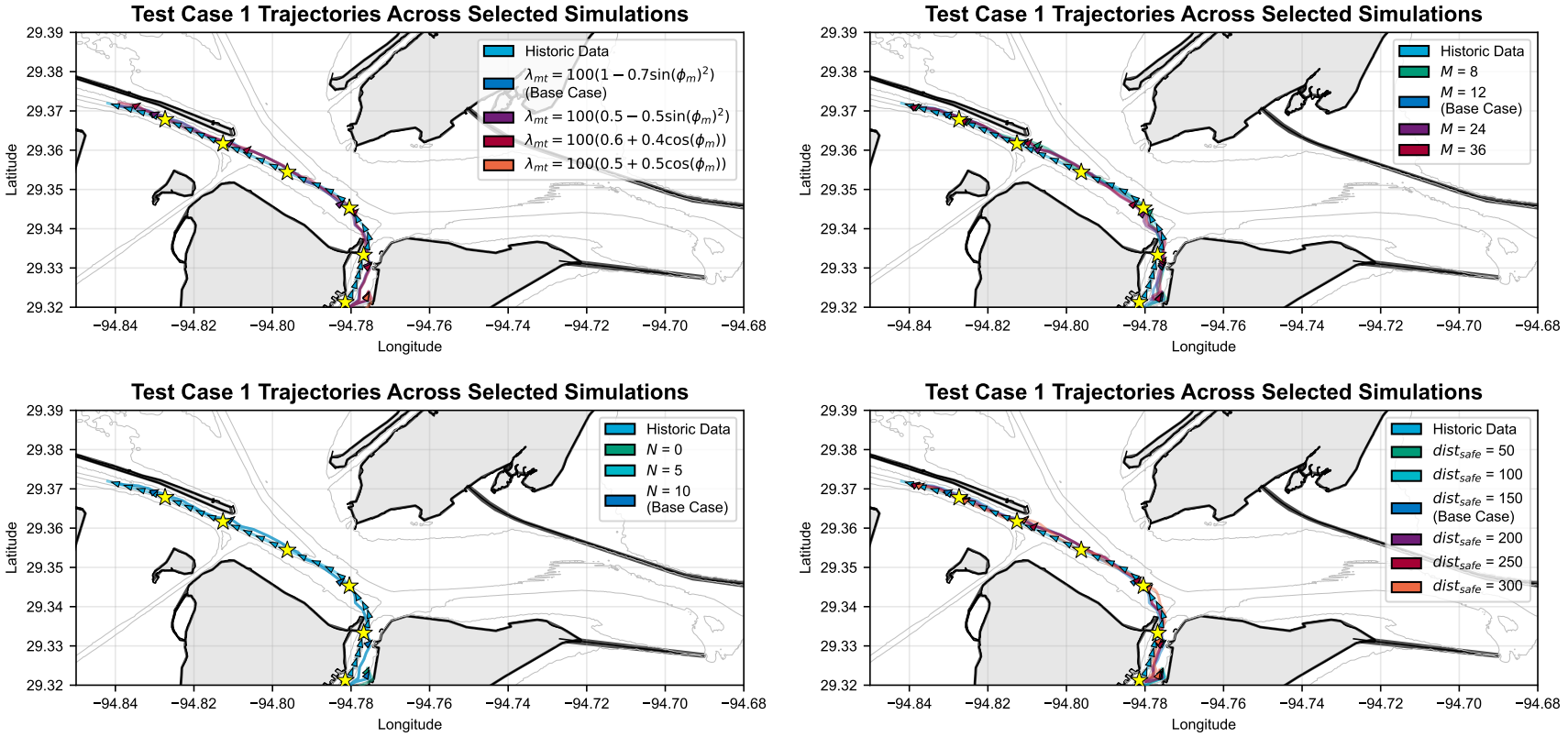


Figure I.1: Comparison of tracks produced by different simulations for test case 1. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

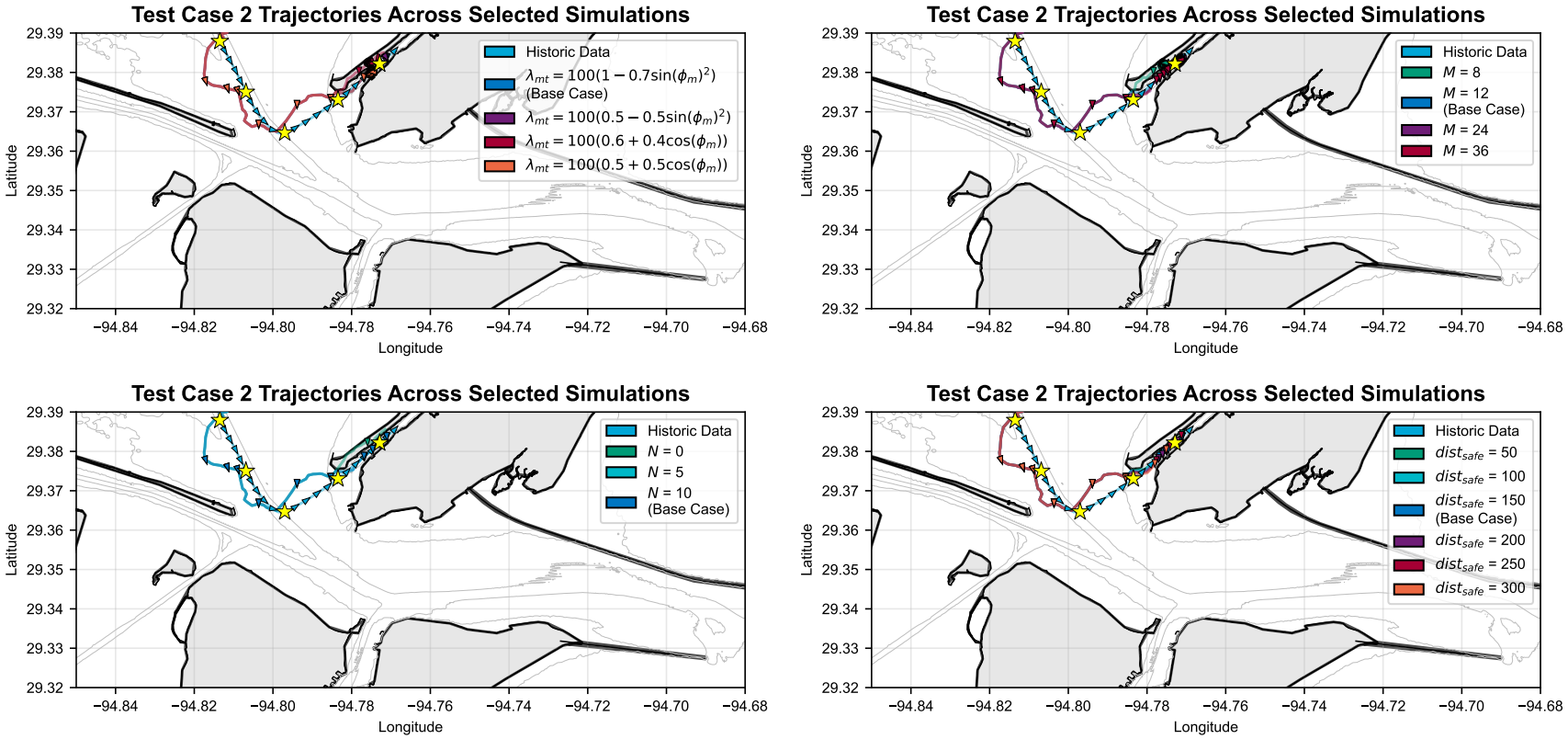


Figure I.2: Comparison of tracks produced by different simulations for test case 2. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

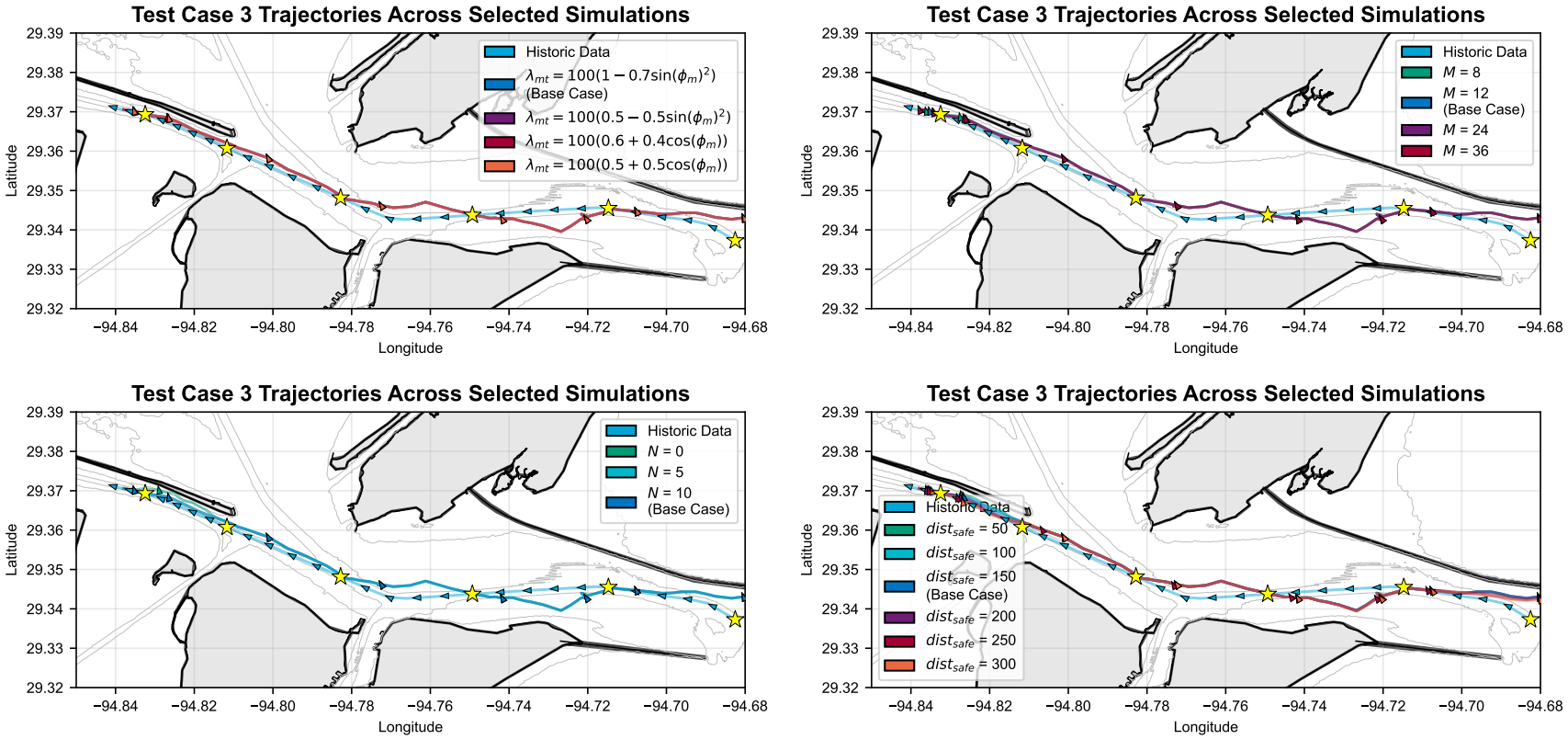


Figure I.3: Comparison of tracks produced by different simulations for test case 3. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

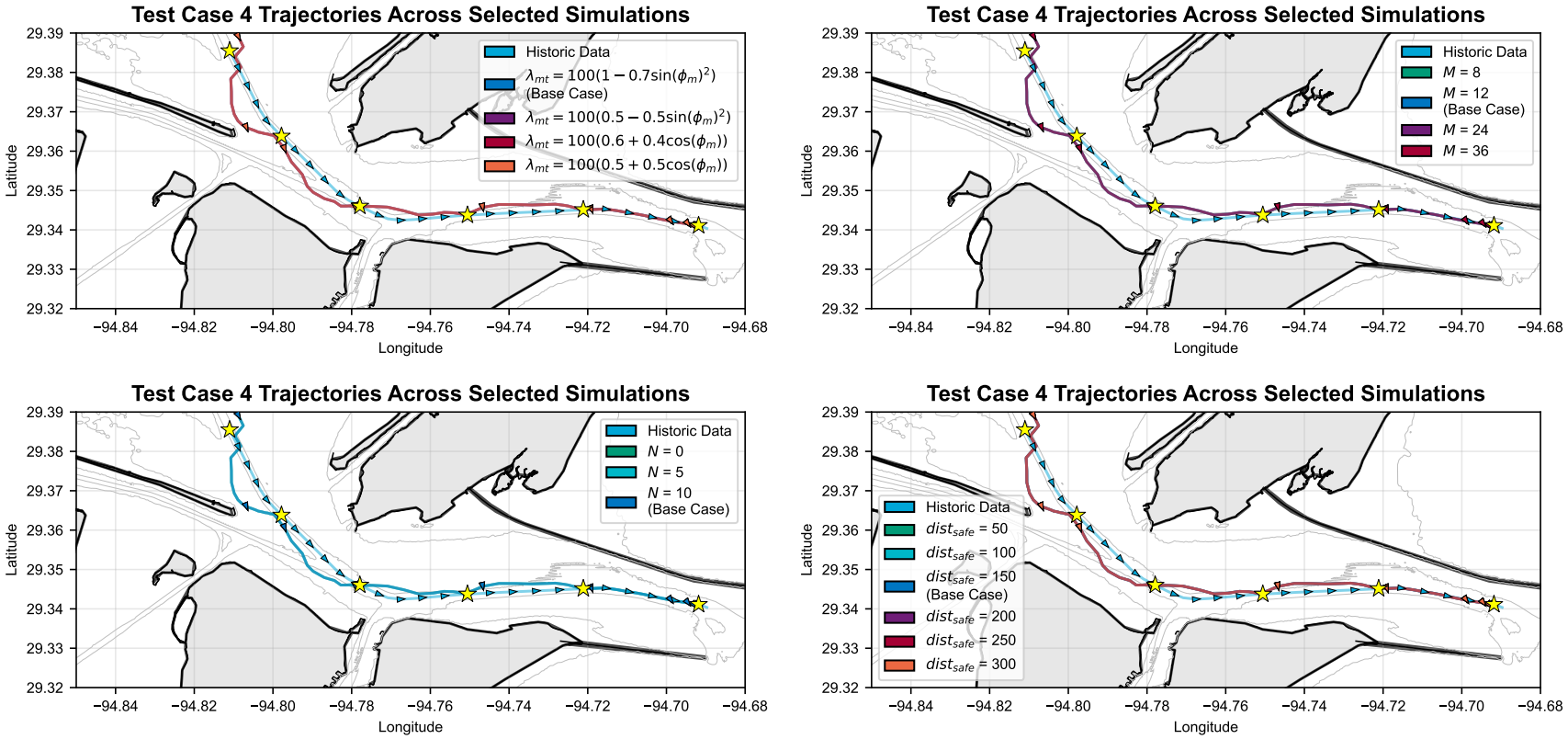


Figure I.4: Comparison of tracks produced by different simulations for test case 4. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

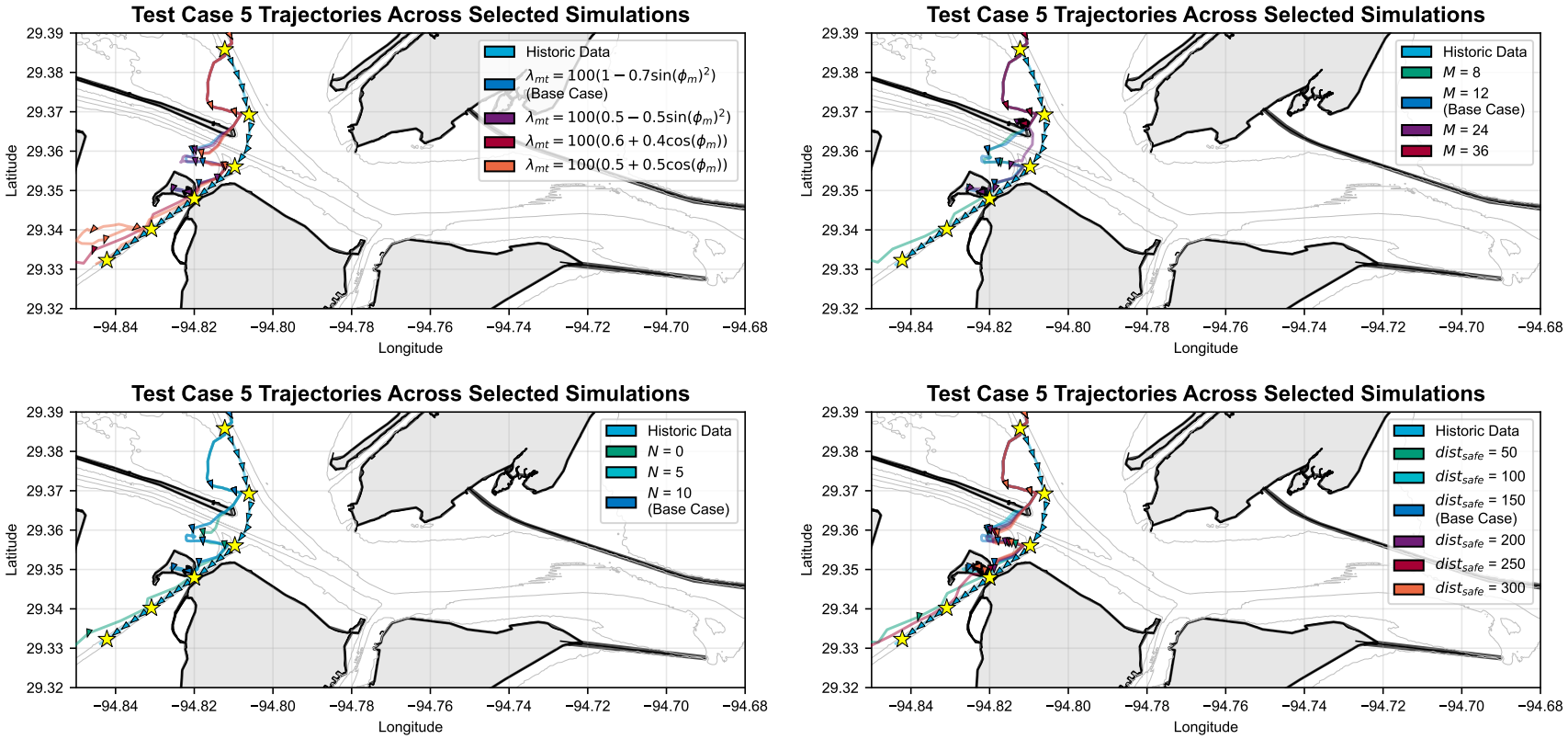


Figure I.5: Comparison of tracks produced by different simulations for test case 5. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

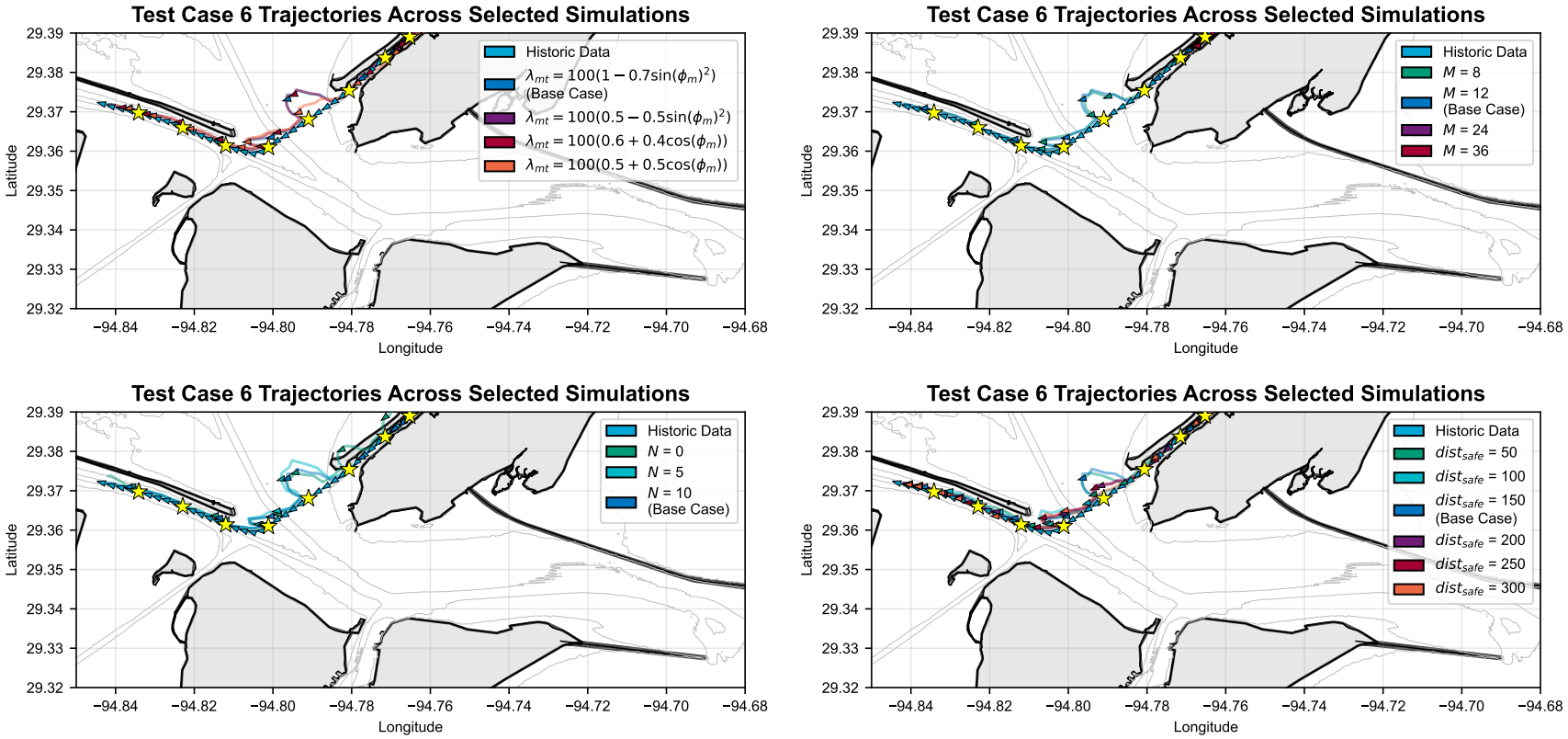


Figure I.6: Comparison of tracks produced by different simulations for test case 6. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

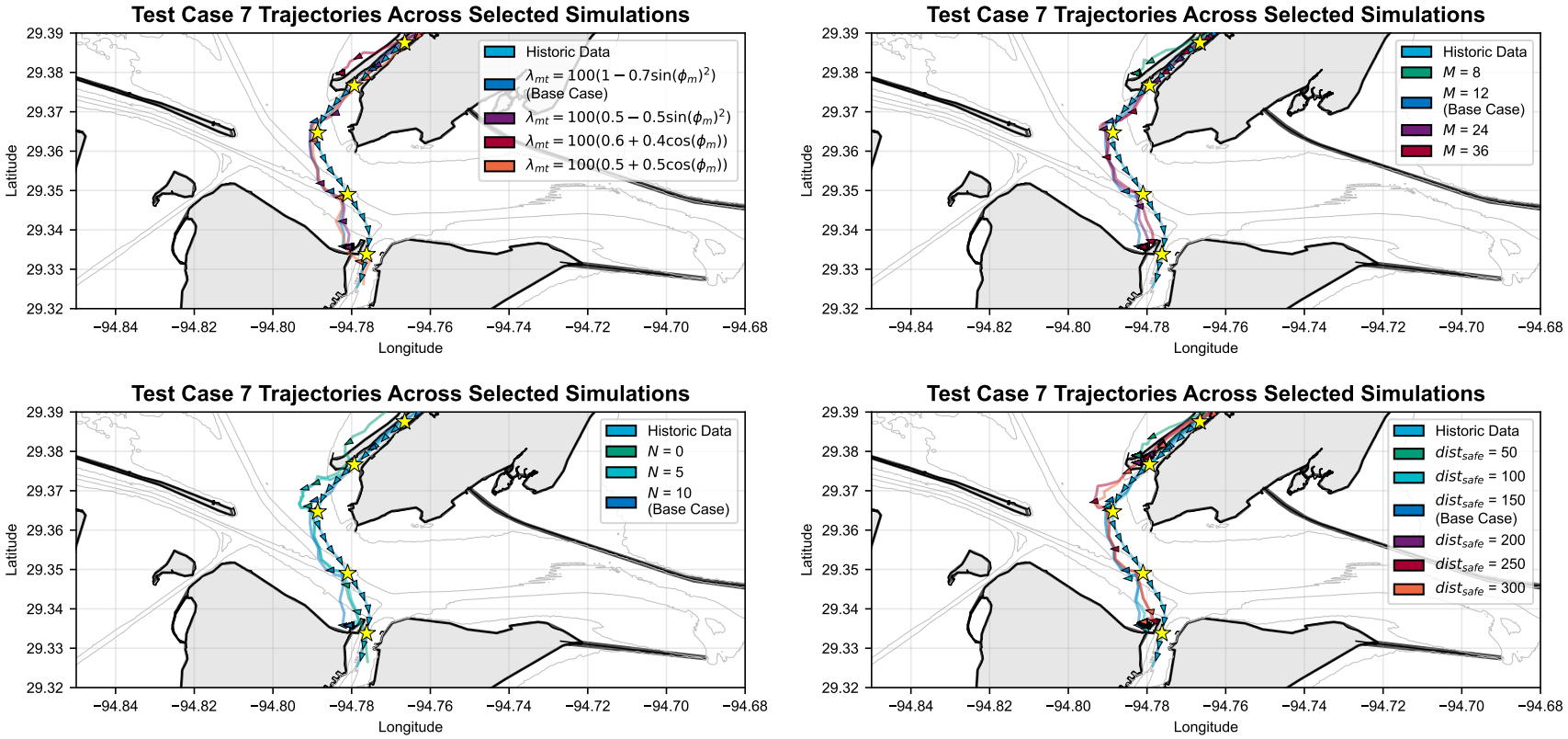


Figure I.7: Comparison of tracks produced by different simulations for test case 7. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

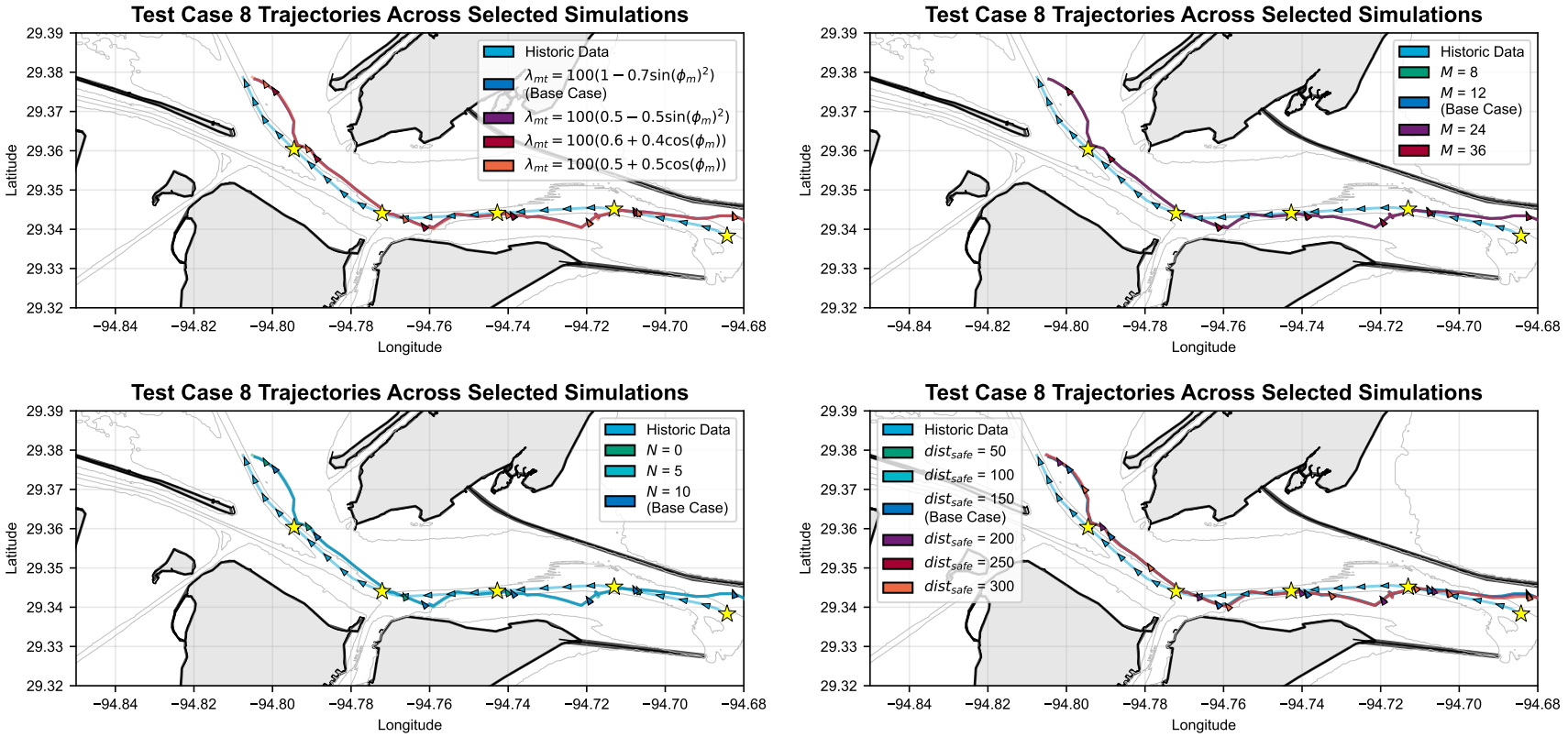


Figure I.8: Comparison of tracks produced by different simulations for test case 8. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

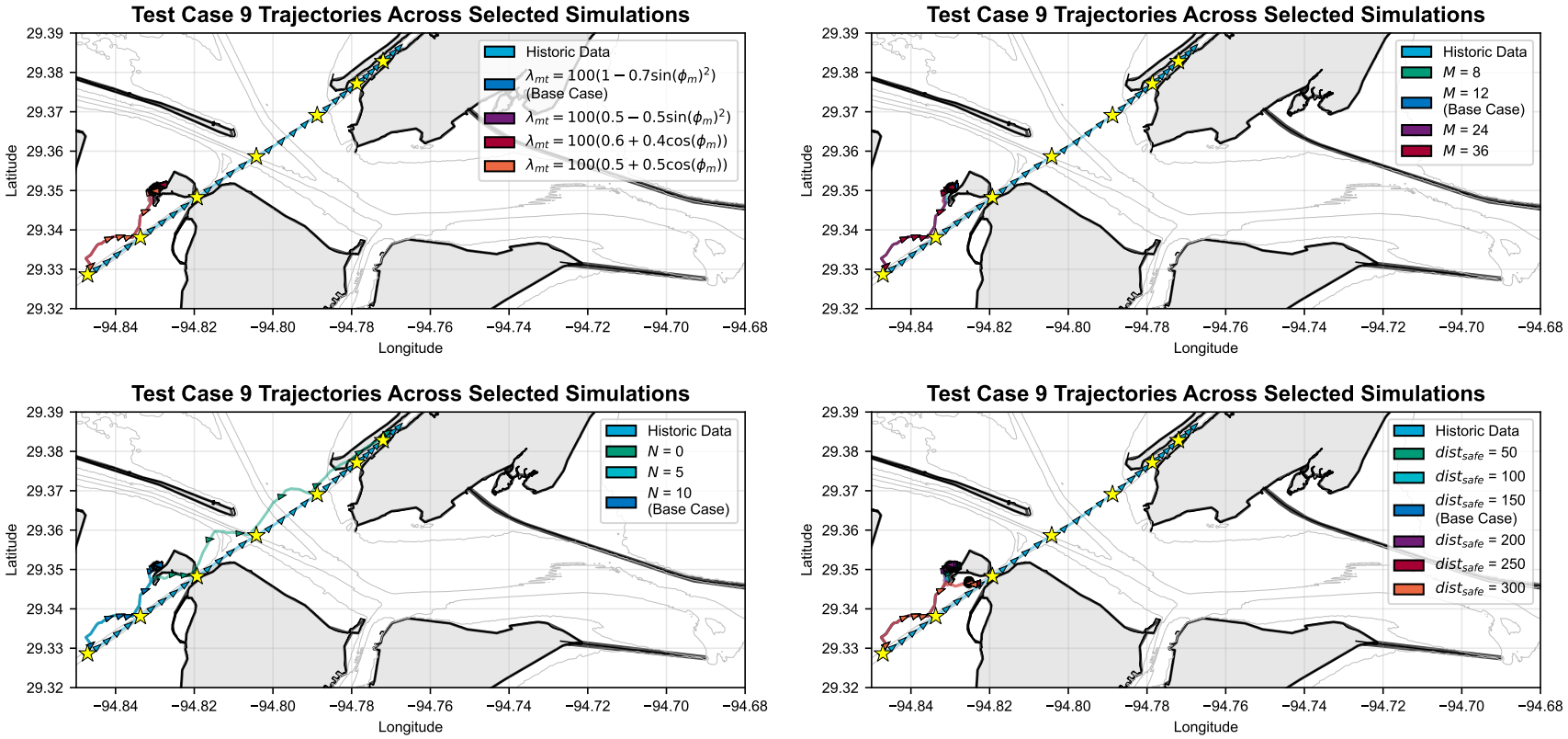


Figure I.9: Comparison of tracks produced by different simulations for test case 9. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

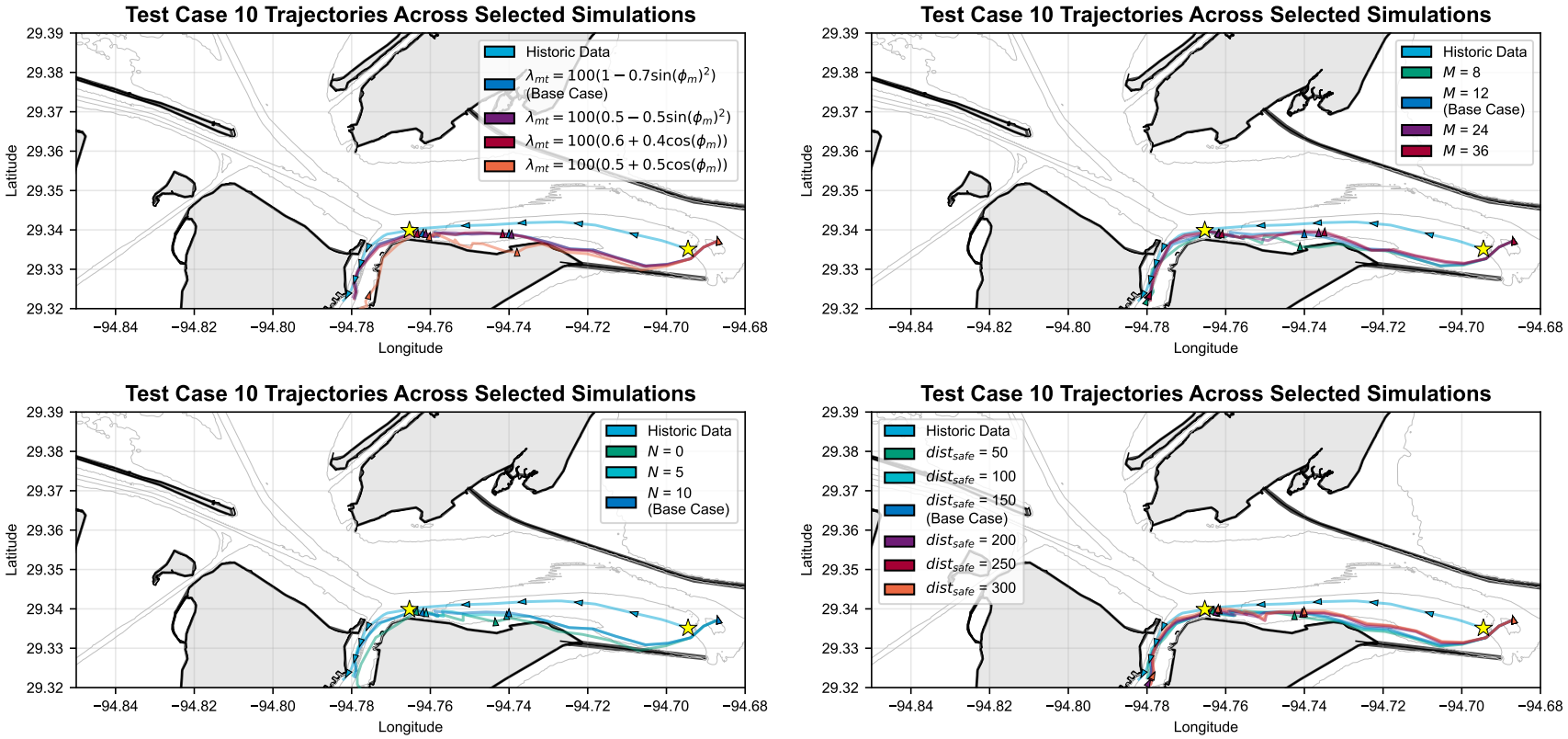


Figure I.10: Comparison of tracks produced by different simulations for test case 10. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

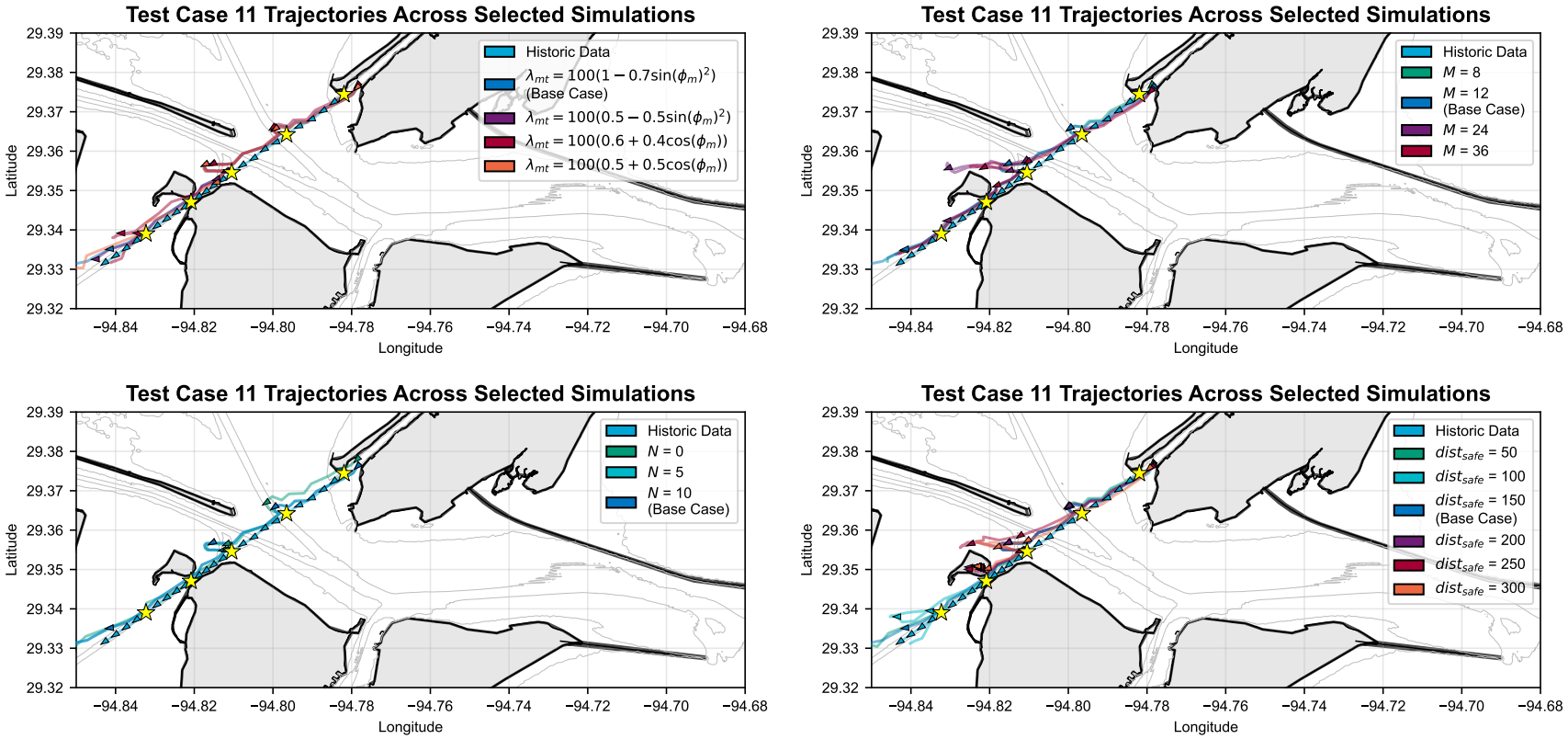


Figure I.11: Comparison of tracks produced by different simulations for test case 11. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

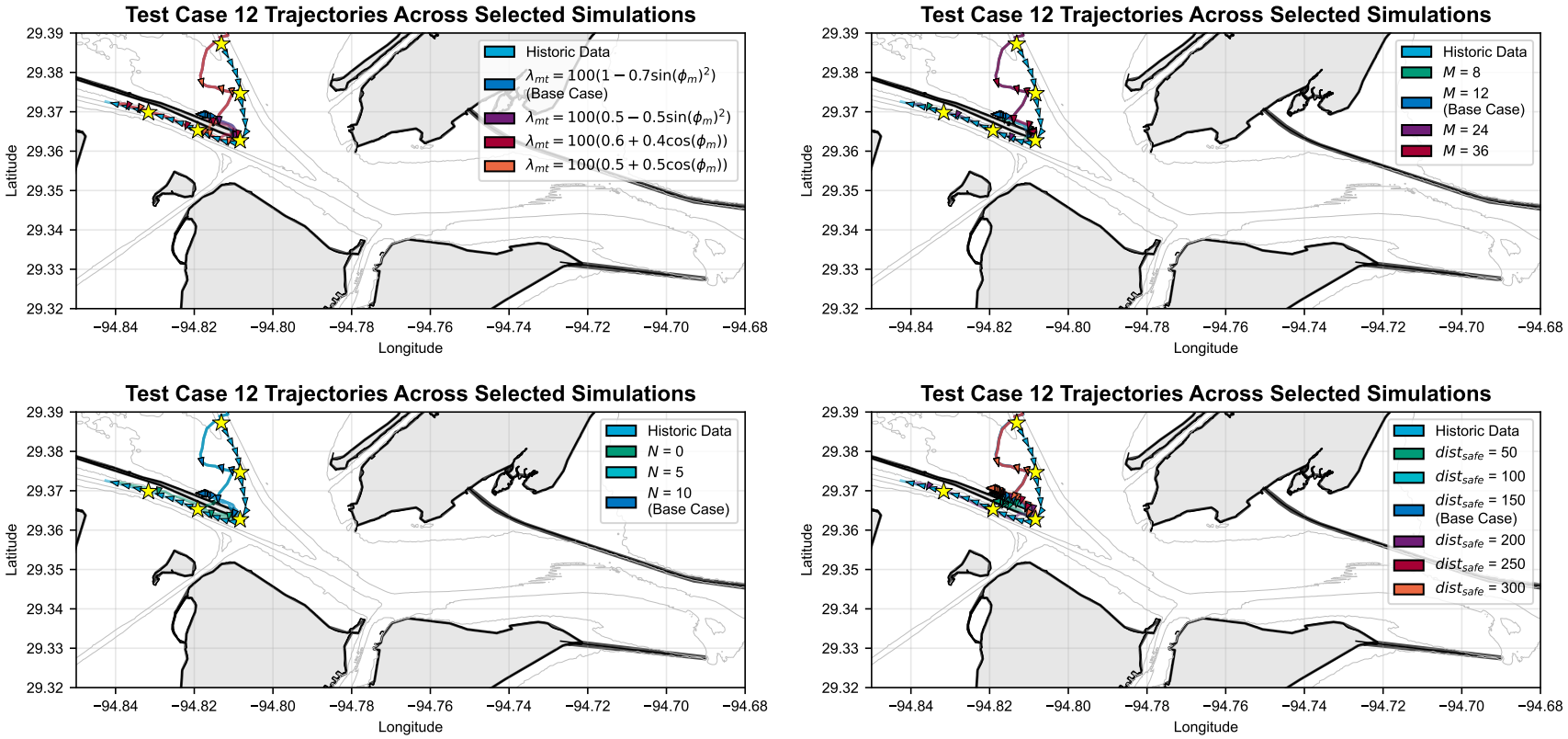


Figure I.12: Comparison of tracks produced by different simulations for test case 12. Top left: variation in λ_{mt}. Top right: variation in M. Bottom left: variation in N. Bottom right: variation in $dist_{safe}$.

How to Train Your Ship Traffic Model: Lessons from Developing Data-Driven Microscopic Ship Traffic Models as a Design Tool for the Houston Ship Channel Gate Complex

Jasper van den Broek
Delft University of Technology, Faculty of Civil Engineering and Geosciences

March 19th, 2026

This paper evaluates whether data-driven microscopic vessel-traffic simulators can support maritime infrastructure design using the proposed Houston Ship Channel Gate Complex as a case study. This paper combines AIS trajectory analysis and semi-structured interviews with local navigators to derive behavioural requirements and implements a continuous-space agent model based on ShipNaviSim. Two extensions are tested: a multi-goal (waypoint) mechanism to encourage channel following and a post-posed MPC safety shield that uses raycasting for obstacle detection and computes minimal corrective actions. Reconstructing one month of traffic, four model variants are compared using trajectory/kinematic metrics (GC-ADE, drift, curvature, acceleration), obstacle incursions and corrective-action magnitudes. Results show the behaviour-cloned baseline reproduces broad patterns but has large positional and kinematic errors; adding intermediate goals and the safety layer substantially improves route conformity and reduces obstacle violations, though head-on encounters and fine kinematic fidelity remain problematic. We conclude that combining AIS-trained agents with route guidance and safety-constrained correction is a promising path for design-capable simulators and recommend further work on multi-class agents, explicit environmental forcing, learning from safety interventions, and route-aware training.

I. Introduction

As part of the Texas coastal protection plan, constructing a storm surge barrier Houston Ship Channel has been proposed by the United States Army Corps of Civil Engineers (USACE) (Greater Houston Port Bureau, 2023b). This barrier, known as the Houston Ship Channel Gate Complex (HSCGC) should protect both the city and the critical industry in the ports of Houston, Galveston, and Texas City. However, recent studies have shown that the barrier is perceived as undesirable from a marine navigation point of view (Burkley et al., 2022), not only in terms of navigational safety, but also in terms of the gate introducing a "chokepoint for ship traffic in the area" which impacts the "spacing, timing, and navigation of ships for miles on either side of the HSCGC" (Burkley et al., 2022, p. 19).

The construction of the HSCGC would affect the traffic capacity and logistics of the Houston Ship Channel (Greater Houston Port Bureau, 2023a). Therefore, it is necessary to assess the impact of the HSCGC on maritime traffic patterns. To assess these impacts in a meaningful way, a modelling framework is needed that describes how individual vessels will respond to the new geometry and operational constraints introduced by the HSCGC. In this sense, the HSCGC serves as real-world example of a more general problem: new or modified maritime infrastructure changes the navigational environment in ways for which no historical traffic data exists.

Current research shows microscopic models can capture the behaviours of individual ships, which can be used to determine emergent traffic characteristics (Bellsolà Olba et al., 2019; Li et al., 2013; Mathioudakis et al., 2025; Papandreou et al., 2025; Shu, 2019; Zhou, 2022). Most models rely on either deterministic rule-following logic or mathematical law to determine the motions and interactions of each ship (Desyani, 2019). Recently, attention has shifted using machine learning based approaches instead,

making use of ship's self-reported AIS (Automatic Identification System) transponder data (Basrur et al., 2021; Pan et al., 2025; Pham et al., 2025). These models provide a useful basis for analysing existing situations, but lack the ability to be employed as a design tool due to the lack of available data for novel situations. This makes such data-driven models unable to represent traffic responses to modified waterways, limiting their usability as a design tool.

A. Contributions

In order to address these limitations, this paper aims to use the case of the HSCGC to contribute to the development of data-driven maritime traffic simulations as a design tool for maritime infrastructures through the following: *(i)* Collecting and analysing AIS data from the proposed site of the HSCGC (Bolivar Roads) to determine characteristic vessel movements, traffic patterns, and behavioural features that the model must be able to capture. *(ii)* Performing semi-structured interviews with expert navigators familiar with the site to extend these features with characteristic features not observed from AIS data. *(iii)* Assessing and extending existing data-driven maritime traffic simulator (ShipNaviSim) to establish its usability as a design tool for maritime infrastructures. *(iv)* Proposing directions for future development of maritime traffic simulators as a tool for maritime infrastructure design.

B. Methodology

This work adopts a mixed-methods approach that integrates quantitative AIS-derived trajectory analysis, qualitative expert interviews, and simulation development and evaluation: first, one year (2024) of AIS data for Bolivar Roads is collected from public databases, cleaned and segmented into vessel tracks, and analysed to extract characteristic manoeuvres, speed/heading distributions, traffic density, and interaction patterns using descriptive statistics. Second, semi-structured interviews with expert navigators who operate in Bolivar Roads are conducted to identify behavioural features, local rules, and edge-case scenarios not visible in AIS. Third, this work assesses ShipNaviSim by calibrating its parameters using AIS-derived machine learning, implementing targeted extensions informed by findings from the data analysis and interviews, and validate the models quantitatively and qualitatively. Finally, this work synthesises results to identify limitations, usability gaps, directions for future simulator development.

C. Document Structure

The rest of this paper is organized as follows: section II introduces related works. section III presents the results of the data analysis and semi-structured interviews. section IV summarizes the model requirements, and section V presents and evaluates the performance of ShipNaviSim and extensions to the model. Finally, section VI summarizes the conclusions and limitations of this study, and proposes steps for future research.

II. Related Work

The works by Shu (2019) and Zhou et al. (2019) found that in constrained waterways, vessel speed and manoeuvring behaviour is influenced by the following factors: size and type of the vessel, the geometry of the waterway, and the direction of navigation, and that external influences such as wind, visibility, and current, as well as head-on and overtaking encounters are non-negligible.

Bellsolà Olba et al. (2017, 2019)'s work showed that in systems of ports and waterways, network traffic indicators can be constructed from microscopic variables, which aligns with Yip (2013)'s idea that macroscopic traffic flow and capacity are the result of individual ship movements.

By constructing such microscopic models based on AIS data, these models are grounded in a commonality with other models that also use historical AIS data (Desyani, 2019), while also accounting for complex behaviours that are not properly captured by behavioural rules or mathematical laws. One such model is ShipNaviSim which was developed by Pham et al. (2025). Pham et al.'s work shows that it is possible to accurately simulate maritime traffic, specifically the movements and interactions of multiple

ships in complex and/or high traffic areas, through imitation learning algorithms. Other models that exploit reinforcement learning in order to develop ship traffic models were developed by Basrur et al. (2021), Düz and van Iperen (2025), and Pan et al. (2025).

The key problem in using data-driven traffic simulations based on a reinforcement learning approach for modelling interventions in an environment is that no real-world data is available before the intervention is realized. The safe reinforcement learning framework first proposed by Alshiekh et al. (2017) and applied by others like Dalal et al. (2018), Dawood et al. (2024), and Sheebaelhamd et al. (2021) provides a solution to handling previously unseen situations by determining a minimum corrective action analytically by solving a quadratic program that minimizes the difference between the action selected by a policy and the actual action carried out, subject to a safety constraint.

Therefore, in order to apply a data-driven maritime traffic model as a design tool the model must be based on a microscopic simulation paradigm and use real-world AIS data in order to inform agent behaviour. Additionally, extensions of such models should use the safe reinforcement learning paradigm to account for novel maritime infrastructures. Table 1 shows an overview of reviewed models, along with how they meet these criteria.

Table 1. Overview of selected related studies in the fields of data-driven maritime traffic behaviour models and safe reinforcement learning.

Study	AIS Data	Interaction		Methodology			
		Agent-Obstacle	Agent-Agent	Space Representation	Agent Control Type	Multi-Agent Simulation	Trajectory Prediction
Basrur et al. (2021)	Y	N	N	zones	RL	N	N
Dalal et al. (2018)	N	N	N	continuous	Safe RL	Y	N
Dawood et al. (2024)	N	Y	N	continuous	Safe RL + MPC	N	Y
Düz and van Iperen (2025)	Y	N	N	continous	RL	N	Y
Koenighofer et al. (2024)	N	Y	Y	discrete cell	Safe RL	Y	Y
Pan et al. (2025)	N	N	Y	continuous	RL	Y	Y
Pham et al. (2025)	Y	N	Y	continuous	RL	Y	Y
Sheebaelhamd et al. (2021)	N	N	N	continuous	Safe RL	Y	Y
Shu (2019)	Y	Y	Y	continuous	MPC	N	Y
Design tool	Y	Y	Y	continuous	Safe RL	Y	Y

Key: Y = included in study, N = not included in study.

Abbreviations: RL: Reinforcement Learning, MPC: Model Predictive Control.

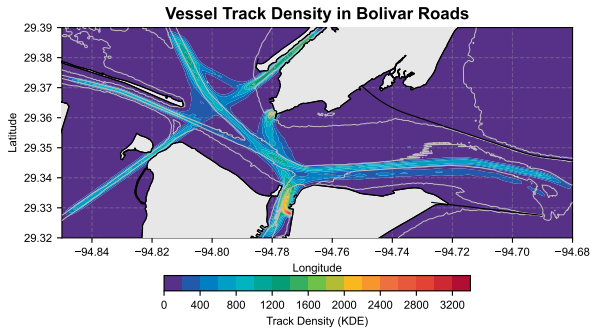


Figure 1. Geographic distribution of AIS transmissions in the filtered dataset (kernel density estimate).

III. Data Analysis and Interviews

A. AIS Data Analysis

AIS transmissions for 2024 within a bounding box around Bolivar Roads (longitudes -94.85° to -94.68°, latitudes 29.32° to 29.39°) were obtained from NOAA (n.d.).

Preprocessing and track construction. Transmissions were filtered by AIS navigational status (kept: 0, 3, 4, 7, 12, 15) and speed-over-ground (SOG $\geq$ 0.2 kts) to remove moored/anchored broadcasts and non-vessel transponders. Tracks were assembled by MMSI; a new track was started when a vessel left and re-entered the bounding box, when the inter-report gap exceeded 30 min, or when a sharp COG reversal ($\geq$ 160° across five reports) occurred (the midpoint of those five reports defines the new track). Tracks shorter than 500 m or with a bounding-box area $\leq$ 10000 m^2 were discarded.

Dataset summary. After processing there are 139,268 tracks (see Table Table 2). The modal vessel classes are tugs (45.84%), tankers (20.35%) and cargo (7.52%). Of selected transmissions, 2,501,836 (68.34% of selection; 20.78% of all raw transmissions) include heading; 4355 vessels (91.34% of selection; 78.82% of all ships) report heading. The mean track contains 26.3 reports and spans 32.3 min (tracks with heading: 31.6 reports, 36.0 min).

Table 2. Descriptive statistics of the tracks in the processed dataset

	n	Mean	Median	St. Dev.	Min.	Max.
Points per track						
All tracks	139268	26.3	25	15.3	5	498
Tracks with heading available	79199	31.6	31	14.7	5	498
Track duration (min)						
All tracks	139268	32.3	30.7	18.0	3.1	806.8
Tracks with heading available	79199	36.0	35.6	17.8	3.1	580.0
Track distance (m)						
All tracks	139268	8464	7710	4467	500	37319
Tracks with heading available	79199	10216	10060	4476	500	37319

Figure 1 shows that traffic is concentrated in the designated ship channels. Origin–destination clustering (Figures 2–3) reveals primary flows from the Gulf to Houston, Galveston and Texas City, and a substantial ferry flow between Galveston and Port Bolivar with frequent encounters at terminals.

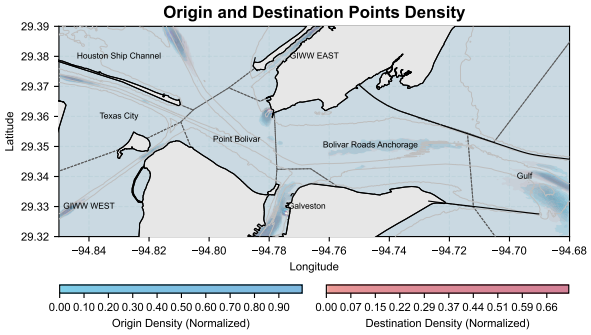


Figure 2. Heatmap (kernel density estimation) showing the positions of origin (blue) and destination (red) clusters.

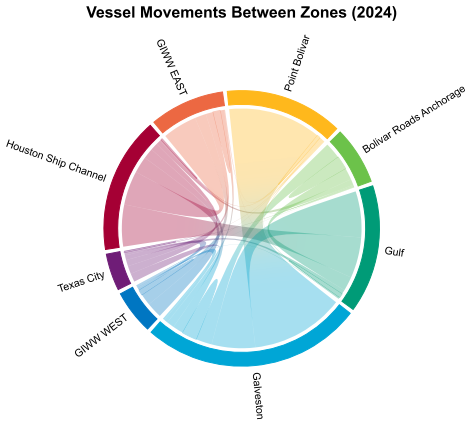


Figure 3. Diagram illustrating the share of origins and destinations for each zone as defined in Figure 2, and the size of the flows between these zones.

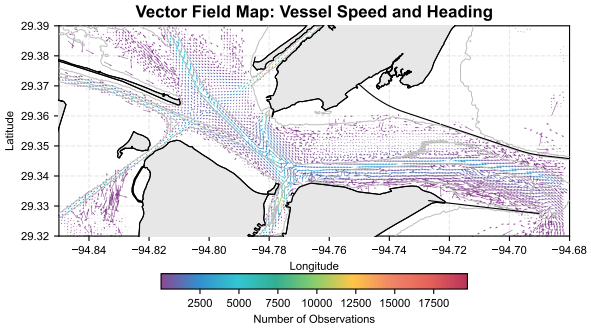


Figure 4. Vector field showing average speed and direction of ship traffic in Bolivar Roads.

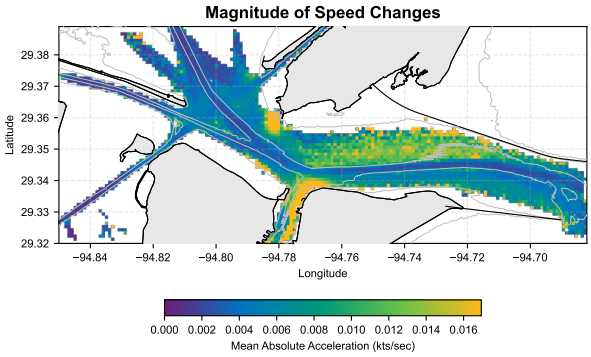


Figure 5. The spatial variation in mean absolute acceleration in Bolivar Roads.

A vector-field of mean COG and speed (Figure 4) confirms route-consistent headings and speeds within channels. Mean absolute acceleration (Figure 5) highlights acceleration/deceleration at channel entries/exits, ports and intersections. Rate-of-turn (ROT) versus SOG (Figure 6) shows greater ROT variability at low speeds; the spatial ROT map (Figure 7) concentrates turns near intersections and the Bolivar Roads anchorage. Encounter density (ships within 500 m; Figure 8) peaks in channels, intersections and ferry terminals. Drift (heading vs. movement over ground; Figures 9–10) is elevated and more variable on north-south routes (GIWW, ferry) consistent with cross-channel tidal influence, while the Houston Ship Channel shows lower drift.

The AIS-derived patterns indicate that a traffic model must: *(i)* Represent diverse vessel classes and class-specific manoeuvring behaviour. *(ii)* Reproduce route-following with consistent within-channel speeds/headings and speed transitions at channel/port boundaries. *(iii)* Capture interaction dynamics

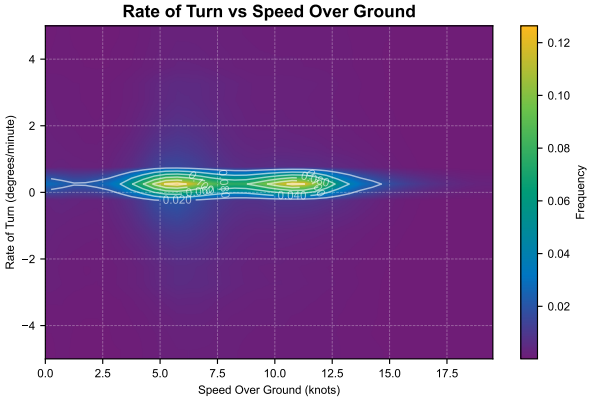


Figure 6. The relationship between ROT and SOG observed in Bolivar Roads.

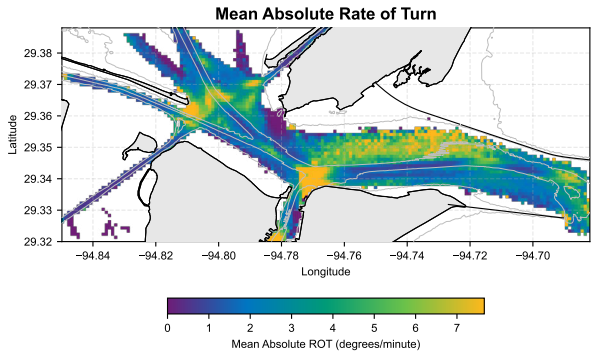


Figure 7. The spatial variation in ROT in Bolivar Roads.

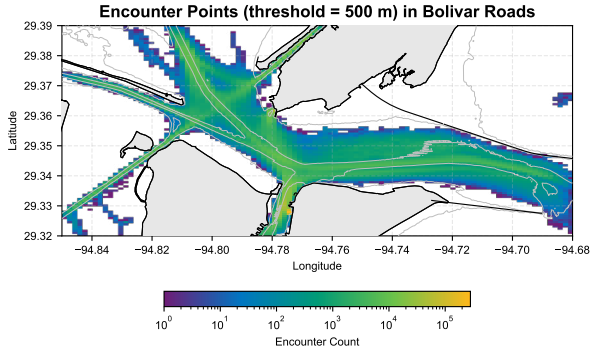


Figure 8. Map showing the density of encounters (ships within 500 m of each other) in Bolivar Roads

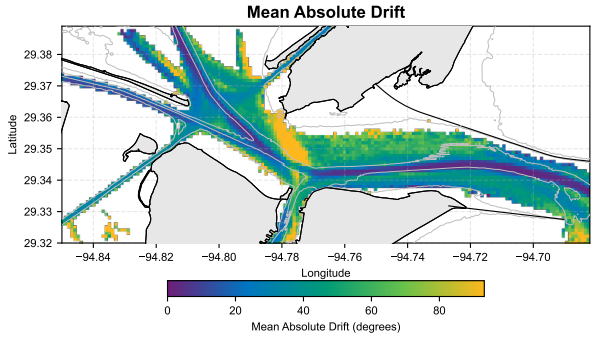


Figure 9. The mean drift in Bolivar Roads

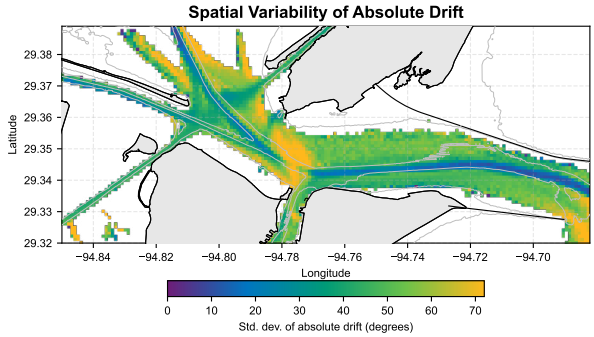


Figure 10. The spatial variation in drift in Bolivar Roads.

concentrated in channels and at intersections (high encounter density). *(iv)* And include route-specific environmental effects (drift) and increased ROT/acceleration variability near intersections and anchorages.

B. Interviews

Analysis of the semi-structured interviews with navigational experts corroborates several AIS-derived patterns while adding nuance critical for modelling. Respondents confirmed that most vessels follow the designated ship channels but emphasised the need to distinguish between draft-constrained ships that must remain in deep water and vessels that may operate outside the channels. These two categories display different channel-following and turning behaviour and therefore require separate manoeuvre representations. Experts also reported frequent overtaking in the deep-water segments highlighted in Figure 11, indicating the simulator must support overtaking logic in those zones.

Respondents highlighted the complex interaction dynamics at the ship channel junctions, where intersecting, merging and diverging flows produce both port-to-port and starboard-to-starboard meetings; they further noted systematic speed reductions when traffic volumes are high. This implies a need for the model to encode meeting-side conventions, conditional speed adjustments, and conflict-resolution behaviours at intersections. Respondents drew attention to high-variability traffic classes, particularly recreational and fishing vessels, that increase navigational uncertainty.

When considering the introduction of an obstacle such as the HSCGC, navigators expect local slowdowns, constrained interaction manoeuvres within the gate complex, and a systematic slight northward displacement of traffic driven by currents and prevailing winds, together with a tendency for vessels to favour paths along the structure's centreline. Finally, respondents anticipated that fragmentation of the Bolivar Roads anchorages would shift demand patterns; for design use the model must therefore allow user-specified OD distributions rather than relying solely on AIS-derived empirics.

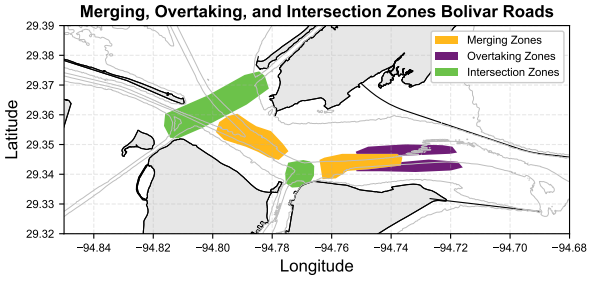


Figure 11. Zones where merging, overtaking, and intersecting behaviour is typically observed.

IV. Model Requirements

This thesis requires a microscopic, multi-agent maritime traffic simulation for Bolivar Roads that is grounded in AIS-derived behaviour yet remains applicable to design scenarios where historical data are incomplete (e.g., introducing the Houston Ship Channel Gate Complex, HSCGC). The model must capture interaction-driven vessel motion such that realistic local decisions and encounters yield credible system-level outcomes (flow, capacity, density, delay).

A. Theoretical and modelling basis.

The simulation must represent individual vessels over time in a multi-agent setting, where agents adapt to other vessels and to the environment. Behaviour should be AIS-grounded and data-driven to reproduce complex manoeuvring without relying solely on handcrafted rules.

B. Generalizability to infrastructure change.
The model must support modified waterway geometry and new obstacles, and enable vessel–obstacle interaction. Because new infrastructure introduces unseen situations, the model must include a safety mechanism (e.g., safety filtering consistent with safe reinforcement learning) that prevents unsafe outcomes such as collisions, and applies minimal-impact corrective control to preserve behavioural realism where possible.

C. Baseline validity targets (AIS 2024).
For present-day Bolivar Roads, the model should reproduce: *(i)* a heterogeneous fleet with type-dependent operations (including draft-related constraints); *(ii)* channel-constrained route-following aligned with the Houston, Galveston, and Texas City Ship Channels; *(iii)* dominant OD flows, including the Galveston–Point Bolivar ferry crossing and its conflicts with channel traffic; *(iv)* distinct operational speed regimes and spatial structure in speed and heading; and *(v)* relatively stable traffic intensity at daily/monthly scales alongside hourly variability in average speeds.

D. Interaction, manoeuvring, and safety behaviour.
The model should capture spatially heterogeneous manoeuvring complexity, with turning and speed changes concentrated near entrances, junctions, approaches, anchorages, and ferry terminals. Vessels must adapt speed due to congestion and conflict avoidance, reproduce increased turning at low speeds near junctions, and generate realistic encounter hotspots. The model must represent merging, diverging, and crossing interactions; allow both standard and coordinated meeting conventions where operationally relevant; enable overtaking where feasible but constrained; and include less predictable small craft that increase uncertainty and induce avoidance dynamics.

E. HSCGC design-scenario capabilities.
For gate scenarios, the model must represent: approach and transit speed reductions (including reduced ability to slow under flood tide); environmentally induced lateral bias due to wind and current; constrained interaction logic near the gate (discouraging head-on encounters and favouring centreline passage when unopposed); and configurable OD demand to reflect scenario-driven traffic shifts.

Overall, validity should be demonstrated not only by aggregate trajectory similarity, but by reproducing traffic structure, speed and manoeuvring regimes, and interaction locations and behaviours, enabling credible assessment of safety and capacity impacts under infrastructure change.

V. Modelling

A. Model Formulation

This section translates the empirical and interview-derived requirements into a microscopic, continuous-space, data-driven ship-traffic model that is both realistic (able to reproduce observed manoeuvring) and flexible (able to test design changes such as the HSCGC). To reconcile AIS-trained behaviour with novel design scenarios, the architecture combines a behaviour-cloning policy trained on AIS (ShipNaviSim) with two targeted extensions: a multi-goal mechanism to enforce route-following and an explicit safety layer that detects and corrects unsafe actions when ships encounter novel obstacles.

1. State, scale and initialization.
The simulation evolves in a flat-earth 2D metric frame (south-western corner at lat/lon 29.32°, -94.85°) with timestep dt = 10 s. The environment state is $S_t = \{\mathbf{s}_t^1, \ldots, \mathbf{s}_t^k\}$ and a ship state $\mathbf{s}_t^k = \langle x_t^k, y_t^k, v_t^k, h_t^k \rangle$; each agent has a set of goals $G_k = \{\mathbf{g}_k^1, \ldots\}$ with $\mathbf{g}_k^i = \langle x_g^i, y_g^i, dist_g^i \rangle$. Static obstacles are polygons $p \in P$. Ships are initialized from precomputed OD-driven conditions drawn from AIS-derived distributions;

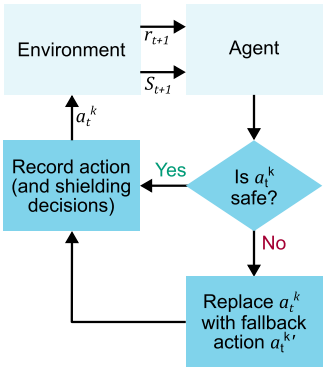


Figure 12. Working mechanism of the safety shield. Adapted from Alshiekh et al. (2017).

ships are spawned/removed according to t^k_{init} and goal completion, exiting simulation bounds, or timeout after 1000 timesteps.

2. ShipNaviSim

Actions are produced by a learned policy $\mathbf{a}^k_t = \pi_\theta(\mathbf{o}^k_t(S_t))$. Observations $\mathbf{o}^k_t$ include recent own-ship history, history of the n nearest ships, goal position and inter-vessel distances; missing entries are zero-padded. The agent outputs delta actions $\mathbf{a}^k_t = \langle dx^k_t, dy^k_t, dh^k_t \rangle$ (continuous, bounded), and states update via the kinematic mapping $\mathbf{s}^k_{t+1} = f(\mathbf{s}^k_t, \mathbf{a}^k_t) = \langle x^k_t + dx,\ y^k_t + dy,\ v^k_{t+1},\ h^k_t + dh \rangle$, with

$$v^k_{t+1} = \frac{\sqrt{dx^{k\,2}_t + dy^{k\,2}_t}}{dt} \tag{1}$$

3. Multi-goal extension

To encourage channel-conformant routing, agents receive intermediate goals sampled along historic tracks (every 50 points) so that agents sequentially pursue multiple goals G_k. Reaching the active goal advances to the next target.

4. Obstacle Avoidance through Safety Layer

A safety shield (Figure 12) employs raycasting (rays $m \in M$ at spacing φ relative to heading) to sample distances $dist_m$ to obstacles and flags unsafe candidate actions by linearly extrapolating the proposed action for 10 timesteps. Unsafe actions trigger a corrective optimisation that searches for a minimal-deviation safe action $\mathbf{a}^{k\prime}_t$. Formally the correction minimises deviation from the original action plus control effort (and ray slack penalties) subject to ship dynamics, action bounds and ray-based distance constraints. The original non-linear constrained formulation is relaxed and linearised via ray-based approximations to yield a quadratic program (QP) that can be solved efficiently at runtime; the first approved action from the QP sequence replaces the unsafe proposal. This problem is derived from the work of Dawood et al. (2024), and can be formulated as follows:

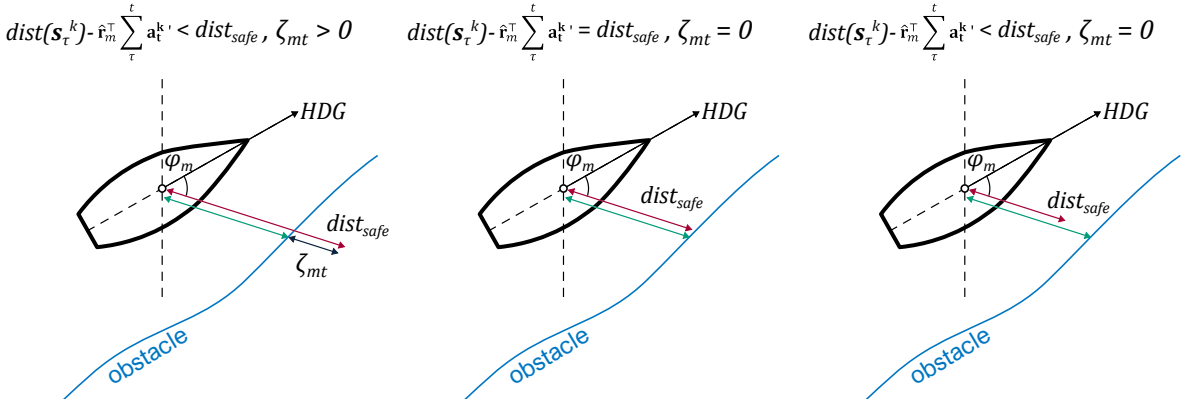


Figure 13. The three cases for the distance between a ship and an obstacle along a ray: distance is shorter than d_{safe}, distance is exactly d_{safe}, or distance is greater than d_{safe}

$$\min \quad Z = ||\mathbf{a}_\tau^{k\prime} - \mathbf{a}_\tau^k||_{R_0}^2 + \sum_{t=\tau+1}^{\tau+N} ||\mathbf{a}_t^{k\prime}||_R^2 + \sum_{\tau}^{\tau+N} \sum_{m}^{M} \lambda_{mt} \zeta_{mt} \tag{2}$$

Subject to: (3)

$$\mathbf{s}_{t+1}^k = f\left(\mathbf{s}_t^k, \mathbf{a}_t^{k\prime}\right) \qquad \forall t \in T \tag{4}$$

$$dx^{\min} \leq dx_t^k \leq dx^{\max} \qquad \forall t \in T \tag{5}$$

$$dy^{\min} \leq dy_t^k \leq dy^{\max} \qquad \forall t \in T \tag{6}$$

$$dh^{\min} \leq dh_t^k \leq dh^{\max} \qquad \forall t \in T \tag{7}$$

$$dist_m(\mathbf{s}_\tau^k) - \hat{\mathbf{r}}_m^\top \sum_{\tau}^{t} \mathbf{a}_t^{k\prime} + \zeta_{mt} \geq dist_{safe} \qquad \forall m \in M, t \in \{\tau, \ldots, \tau + N\} \tag{8}$$

$$\zeta_m \geq 0 \qquad \forall m \in M \tag{9}$$

The objective function minimizes the norm of the difference between the original action $\mathbf{a}_t^k$ and the corrective action $\mathbf{a}_t^{k\prime}$, as well as control effort ($\sum_{t=\tau+1}^{\tau+N} ||\mathbf{a}_t^{k\prime}||_R^2$) and the penalties incurred by using slack on the rays radiating from the ship's centre. Constraint 4 ensures that the state and actions are continuous with each other across sequential timesteps. Next, constraints 5, 6, and 7 ensure that the elements of the corrective action do not exceed physical limits imposed on the simulation.

Constraint 8 ensures that if the distance along a ray m from the ship to an obstacle after the action $\mathbf{a}_t^{k\prime}$ is smaller than $dist_{safe}$, a slack penalty ζ_m is applied in the objective function. This constraint is based on the assumption that the heading of the ship rotates very little (i.e., using a small angle approximation) between timesteps, and that the edge of the obstacle that meets the ray is locally smooth.

The effect of this constraint is that at each timestep a penalty is applied for selecting an action that leads to the ship getting close to an obstacle. By taking the sum over $\{\tau, t\}$, the model accounts for the cumulative actions at timesteps $\tau, \tau + 1, \tau + 2$, etc. This way, this constraint remains linearized around $\mathbf{s}_t^k$.

Finally, constraint 9 ensures that this slack penalty is non-negative, which implies if that the distance between the ship and the obstacle along ray m after taking the action $\mathbf{a}_t^{k\prime}$ is greater than or equal to $dist_{safe}$, then $\zeta_m = 0$. This avoids pushing the ship further away than the ray is long, which would cause the second term of the objective function to dominate over the first term. This idea is illustrated in Figure 13

The final modelling variants tested are *(i)* baseline ShipNaviSim, *(ii)* ShipNaviSim with Safety Layer, and *(iii)* Multi-Goal plus Safety Layer. Decisions and state updates are executed in parallel per agent to reflect independent navigator decision-making.

B. Model Performance

Using these models from, it is possible to assess whether data-driven agents (ShipNaviSim) and two extensions (Multi-Goal routing; MPC safety layer) can reproduce characteristic vessel movements and support design-oriented scenario testing. Performance is assessed with trajectory and kinematic metrics (Goal-Conditioned Average Displacement Error, GC-ADE; drift, curvature, acceleration), safety diagnostics (entries into obstacle regions, corrective-action magnitudes $||\mathbf{a}_t^k - \mathbf{a}_t^{k\prime}||$), and face validation of visualised behaviour.

C. Experiment Setup

Experiments reconstruct one month of traffic (November 2024) using simplified coastline geometry as obstacle input (simplified 0 m depth contour). Due to numeric instability on the full dataset, most experiments use a reduced training set of 10000 randomly sampled tracks. Four model variants are compared: (1) full-set model (166 epochs), (2) reduced-set model (300 epochs), (3) reduced + MPC safety layer, and (4) reduced + MPC + intermediate goals (multi-goal). The following safety layer settings were used: horizon $N = 10$, 12 rays, $dist_{safe} = 150$ m, $R_0 = I$, $R = 0.1I$. In models with extra goals, intermediate goals are placed every 50 timesteps along historic tracks. Models are logged for microscopic outputs.

D. Results: existing situations.

GC-ADE (Table 3) shows large deviations from historical tracks for models 1–3 (means $\approx$ 1.3 km, 1.05 km and 0.92 km, respectively), while model 4 (reduced + safety + extra goals) substantially improves GC-ADE to $\approx$ 397 m ($\sigma \approx 235$ m), but still remains large relative to vessel lengths. Kinematic ranges indicate all models produce larger acceleration magnitudes than observed historically; drift and curvature ranges are comparable in extent but the simulated distributions exhibit substantially higher variability (Figure 14), consistent with more erratic manoeuvring and larger action magnitudes than the historic data. Visual inspection (Figure 15) confirms single-goal agents tend to seek shortest-path solutions and fail to reproduce channel-aligned headings; adding intermediate goals produces trajectories that better follow historic routes.

Table 3. Mean and standard deviation of GC-ADE for the tested models.

Model	Mean GC-ADE (m)	Std. Dev (m)
Historic data	0.00	0.00
1: Full Set - 166 Epochs	1323.13	463.29
2: Limited Set - 300 Epochs	1053.15	463.71
3: Limited Set - 300 Epochs + Safety	917.85	497.82
4: Limited Set - 300 Epochs + Safety + Extra Goals	397.07	234.63

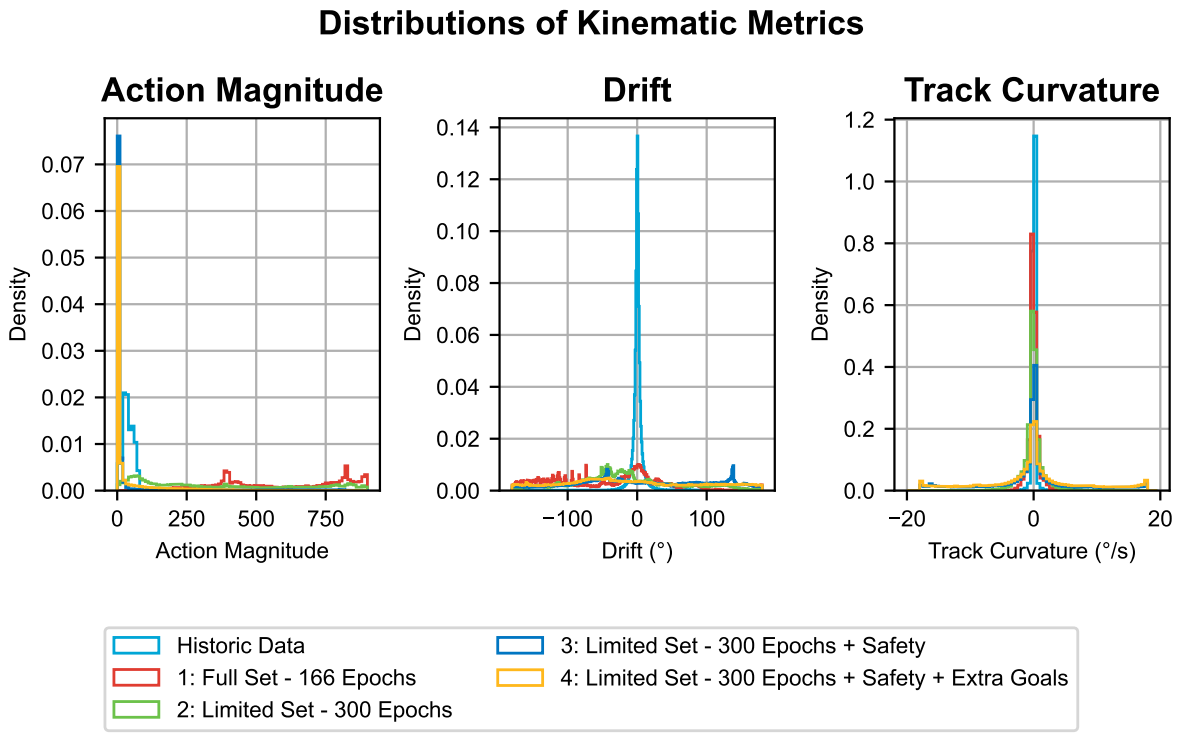


Figure 14. Distributions of action magnitude, drift, and track curvature.

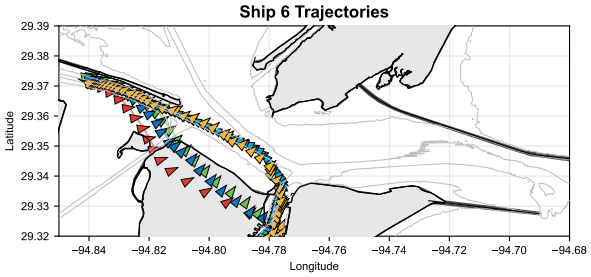


Figure 15. Example showing the trajectories of a single ship in multiple simulations.

The MPC safety layer markedly reduces obstacle violations: obstacle entries fall from several thousand for non-safe models to much lower counts when safety is enabled (Table 4). In the reduced + safety + goals configuration (model 4) observed obstacle entries drop to 1174 and the mean applied corrective-action magnitude rises (mean ≈ 109.6), indicating active intervention but also a smoothing effect on control (reduced extreme control variations). Spatial patterns of crossings (Figure 16) show where violations concentrate.

Table 4. Effects of the safety layer: comparison between obstacles entered and control magnitude.

	Obstacles Entered	Control Magnitude	
Model		Mean	St. Dev.
Historic data	0	0.00	0.00
1: Full Set - 166 Epochs	6334	0.00	0.00
2: Limited Set - 300 Epochs	7340	0.00	0.00
3: Limited Set - 300 Epochs + Safety	5889	92.78	73.16
4: Limited Set - 300 Epochs + Safety + Extra Goals	1174	109.59	82.73

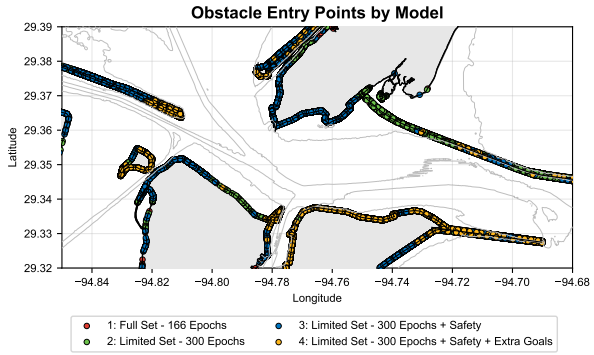


Figure 16. Points where ships entered the coastline by simulation

E. Safety Shield Parameter Sensitivity

The results in the previous were obtained for a single baseline configuration of the post-hoc safety shield. Since the shield operates on top of the ShipNaviSim controller and can override proposed actions, its tuning can directly trade off *(i)* tracking accuracy, *(ii)* constraint satisfaction (obstacle avoidance), and *(iii)* control effort. A sensitivity analysis is therefore required to *(a)* assess whether the main conclusions are robust to plausible parameter changes, *(b)* identify which shield parameters materially affect the key performance indicators (KPIs), and *(c)* separate limitations of the underlying controller from effects introduced by the shielding layer.

We vary the key shield parameters (ray sensitivity profile ζ_{mt}, prediction horizon N, number of rays M, and safe distance $dist_{safe}$) one at a time relative to the baseline and evaluate performance on a diverse set of vessel trajectories through Bolivar Roads. Performance is quantified using *(i)* GC-ADE per trajectory, *(ii)* the number of obstacle entries, and *(iii)* the control magnitude, capturing accuracy, safety, and actuation effort, respectively. Finally, parameter combinations suggested by the one-at-a-time study are jointly tested to probe potential coupled effects.

F. Test cases and experimental protocol

Twelve historic vessel tracks were selected as test cases. Tracks were chosen (via visual inspection) to cover the major traffic flows identified in Figure 3, such that each flow is represented at least once. The resulting selection is shown in Figure 17.

To isolate the effects of parameter variation, tracks are staggered in time and simulated independently (1000 time steps per track), preventing ship-ship interactions from confounding the results. The geometric configuration is kept identical to the main experiments, so that differences can be attributed to shield parameters rather than changes in the environment. All simulations were executed on the Delft High Performance Computing Centre (DHPC) (2024) system. Model outputs are used to compute GC-ADE

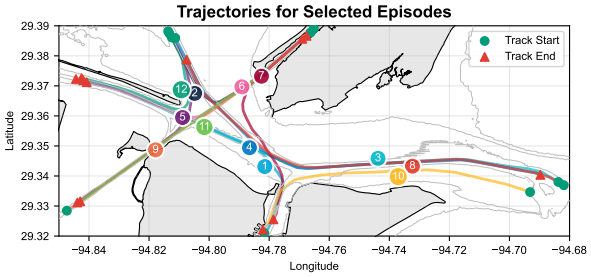


Figure 17. Tracks selected for the sensitivity analysis, covering the major flows through Bolivar Roads.

and obstacle-entry counts.

Unless stated otherwise, the baseline configuration matches subsection V.C: prediction horizon $N = 10$, $M = 12$ rays, and $dist_{safe} = 150\,\text{m}$. The weighting matrices are $R_0 = I$ and $R = 0.1I$. Ray sensitivities are

$$\zeta_{mt} = 100\left(1 - 0.7\sin^2(\varphi_m)\right), \tag{10}$$

and intermediate goals are placed every $n = 50$ steps along the historical track.

One-at-a-time parameter variations. Starting from the baseline, we consider:

- **Ray sensitivity profile ζ_{mt}:** to test sensitivity to the implied ship-domain shape, we evaluate:

$$\begin{aligned}
\zeta_{mt} &= 100\left(1 - 0.7\sin^2(\varphi_m)\right) \quad \text{(baseline)},\\
\zeta_{mt} &= 100\left(0.5 - 0.5\sin^2(\varphi_m)\right),\\
\zeta_{mt} &= 100\left(0.6 + 0.4\cos(\varphi_m)\right),\\
\zeta_{mt} &= 100\left(0.5 + 0.5\cos(\varphi_m)\right).
\end{aligned}$$

- **Prediction horizon N:** $N \in \{0, 5, 10\}$ (with $N = 10$ as baseline).
- **Number of rays M:** $M \in \{8, 12, 24, 36\}$ (with $M = 12$ as baseline). Values are chosen so that rays include $\varphi_m \in \{0°, 90°, 180°, 270°\}$.
- **Safe distance $dist_{safe}$:** $dist_{safe} \in \{50, 100, 150, 200, 250, 300\}$ m (with 150 m as baseline).

Across all variations, the shield only affects behaviour when the simulated trajectory enters the shield's relevant operating region (i.e., when obstacles fall within the influence implied by $dist_{safe}$ and the ray model). For tracks that do not pass within $dist_{safe}$ of any obstacle, the shield never intervenes, and the resulting trajectory matches the unshielded ShipNaviSim behaviour. This indicates that the post-hoc shield does not act as a general stabilizer in obstacle-free regimes; its influence is localized to situations where obstacles are near enough to trigger filtering.

When the shield does intervene, its impact is strongest when intervention occurs early in the trajectory: early deviations can propagate and yield a visibly different downstream path, whereas late interventions typically affect only a limited segment (see Figure 18). This is consistent with the shield operating as a local filter whose effects accumulate if activated early.

The ray sensitivity profile ζ_{mt} materially affects obstacle avoidance. In particular, cosine-shaped profiles yield more obstacle entries than sine-shaped profiles (see Table 5), suggesting that the sine-based definitions more effectively represent a conservative ship-domain approximation for this environment. The prediction horizon also matters: reducing lookahead increases obstacle entries substantially, with $N = 0$ producing many more failures than $N = 5$ and $N = 10$. This indicates that multi-step extrapolation is important for rejecting actions that appear safe instantaneously but lead to near-future violations.

Changes that improve obstacle avoidance also affect control behaviour. Cosine-shaped ζ_{mt} profiles produce a less consistent control-effort distribution than sine-shaped profiles, suggesting reduced stability.

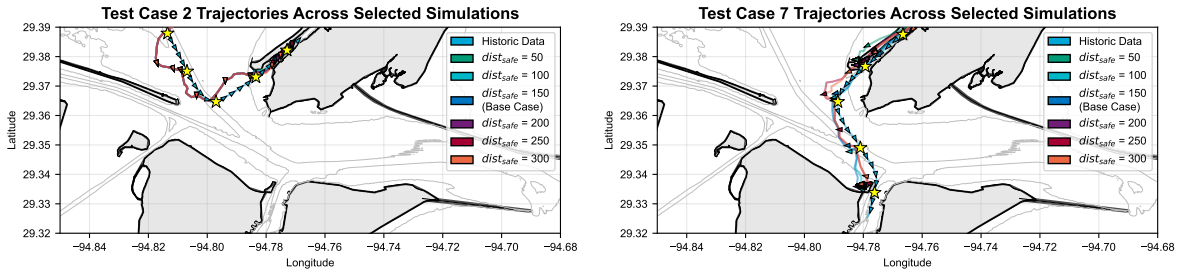


Figure 18. Example of early (left, test case 2) versus late (right, test case 7) shield intervention under different $dist_{safe}$ settings.

Table 5. Total obstacle entries across all test cases under one-at-a-time parameter variations (lower is better).

Variation	Obstacle entries	Variation	Obstacle entries
$\zeta_{mt} = 100(1 - 0.7\sin^2(\varphi_m))$ (baseline)	2	$N = 0$	16
$\zeta_{mt} = 100(0.5 - 0.5\sin^2(\varphi_m))$	2	$N = 5$	2
$\zeta_{mt} = 100(0.6 + 0.4\cos(\varphi_m))$	4	$N = 10$ (baseline)	2
$\zeta_{mt} = 100(0.5 + 0.5\cos(\varphi_m))$	5	$dist_{safe} = 50\,\text{m}$	2
$M = 8$	4	$dist_{safe} = 100\,\text{m}$	2
$M = 12$ (baseline)	2	$dist_{safe} = 150\,\text{m}$ (baseline)	2
$M = 24$	1	$dist_{safe} = 200\,\text{m}$	5
$M = 36$	0	$dist_{safe} = 250\,\text{m}$	4
		$dist_{safe} = 300\,\text{m}$	4

Increasing the number of rays M tends to improve stability, particularly near non-convex obstacles where insufficient ray coverage can cause the agent to get stuck. This aligns with the failure mode described by Dawood et al. (2024), where static and insufficient ray-based sensitivity choices can lead to trapping in complex geometries.

Across the variations considered, the geographic locations of obstacle entries do not shift meaningfully: entry points remain visually consistent. This suggests that the parameters tested primarily change the frequency of failures (and local behaviour near obstacles), rather than moving the system into qualitatively different failure locations.

G. Results: New Situations

Model 4 (best performer on historic reconstruction) is used to test unseen obstacles (rectangular obstacle overlaying the Houston Ship Channel; a simplified gate structure). Introducing new obstacles increases obstacle entries and the average magnitude of safety interventions (see Table 6), while control magnitudes increase notably for the gate scenario (mean ≈ 151.6). Trajectories around new obstacles show partial avoidance and local slowdowns (Figure 19), but obstacle entries are denser along North–South obstacle axes (head-on encounters), suggesting the raycasting + shield is less effective for obstacles encountered directly in the transit direction.

Table 6. Effects of the safety layer: comparison between obstacles entered and control magnitude.

	Obstacles Entered	Control Magnitude	
Scenario		Mean	St. Dev.
Baseline	1174	109.59	82.73
Rectangular Obstacle	2024	111.10	72.46
Gate Obstacle	1534	151.56	77.97

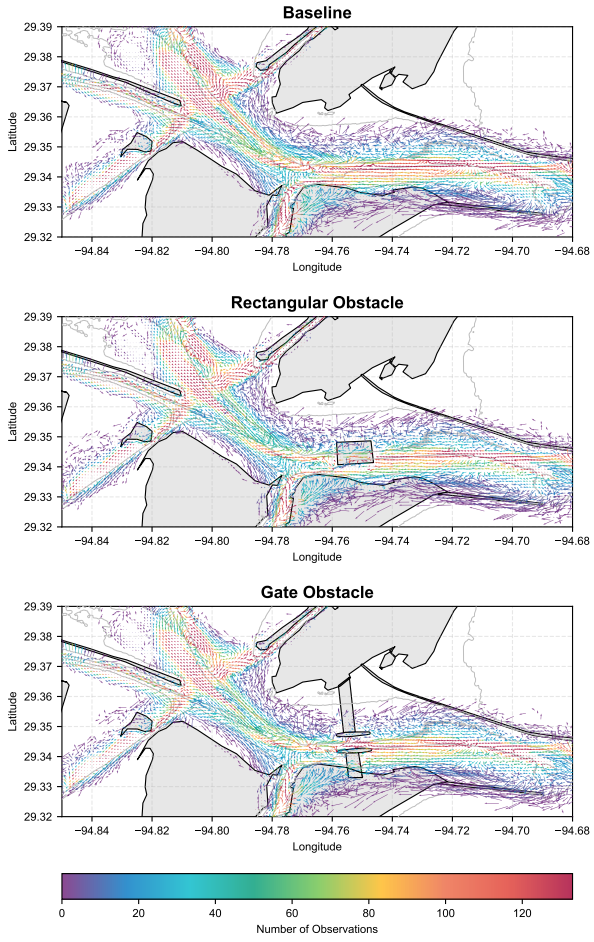


Figure 19. Vector field of trajectories produced by the reduced-dataset model with both MPC and additional intermediate goals without new obstacles, with a rectangular obstacle, and a gate obstacle. Arrow direction indicates COG, arrow length indicates speed.

The tested data-driven agents retain only a limited ability to reproduce precise historical trajectories and kinematic fidelity: GC-ADE and action/drift magnitudes indicate sizeable discrepancies. Nevertheless, the proposed extensions yield clear improvements: multi-goal routing reduces path deviation and encourages channel-conformant routes, and the MPC safety layer substantially reduces obstacle incursions and produces smoother executed controls. The safety layer also enables interaction with previously unseen obstacles, though its efficacy is reduced for head-on encounters and detailed obstacle interactions require further refinement.

These experiments demonstrate that behaviour-cloned ShipNaviSim agents augmented with intermediate goals and a light-weight safety correction provide a pragmatic step toward a design-usable simulator: they improve route conformity and operational safety under novel geometries, but additional

model developments are necessary to attain close alignment with real-world kinematics (larger/more representative training sets, explicit multi-class vessel dynamics, improved perception/avoidance for head-on scenarios, and enhanced action-space regularisation). The next steps focus on these improvements and on applying the best variant to design case studies.

VI. Conclusion

This study used the proposed Houston Ship Channel Gate Complex in the Bolivar Roads strait to evaluate whether data-driven microscopic traffic simulations can serve as a design tool for maritime infrastructure. The main findings are summarized below.

A. Conclusions

Data-driven microscopic simulation is an appropriate modelling paradigm for design-oriented assessment because it captures individual ship movements, local ship-ship and ship-obstacle interactions, and the macroscopic metrics that emerge from them (e.g., space-mean speed, flow, capacity). AIS analysis and expert interviews together define concrete modelling requirements for Bolivar Roads: route-following behaviour with route-specific speed and drift profiles; distinct handling of draft-constrained versus unconstrained vessels; explicit overtaking in deep-water segments; detailed interaction dynamics at channel intersections (including non-standard meeting conventions); and the ability to specify alternative origin–destination patterns for design scenarios.

Behaviour-cloned agents (ShipNaviSim) trained on AIS reproduce some broad movement patterns but show sizeable discrepancies from historical tracks in position and kinematics (GC-ADE, action magnitudes, drift and acceleration variability). Thus, the baseline data-driven model alone is not yet sufficient for reliable design assessment. Augmenting a behaviour-cloned agent with *(i)* intermediate goals (to encourage waypoint/route following) and *(ii)* a post-posed MPC safety shield substantially improves practical behaviour: intermediate goals reduce path deviation and encourage channel-aligned routes, while the safety layer reduces obstacle incursions and smooths executed controls. The combined variant also enables interaction with previously unseen obstacles, though head-on obstacle encounters and detailed obstacle interactions remain imperfect.

Taken together, a viable path toward design-usable data-driven maritime simulation is to combine AIS-trained behaviour with safety-aware correction and route-guidance mechanisms; this hybrid approach preserves learned, realistic manoeuvring where available and imposes minimal, safety-preserving corrections in novel situations.

B. Limitations

The study and its experiments have several important limitations that constrain generality:

- **Single generic agent.** ShipNaviSim in this work uses a single agent to represent all vessel classes, averaging away class-dependent dynamics (size, draft, manoeuvrability).
- **Incomplete observation set.** Not all AIS records report heading. Tracks without heading were excluded from training, reducing representativeness of the learned policy.
- **Training and numerical issues.** ShipNaviSim showed numerical instability on the full dataset, forcing use of a reduced subset (10000 tracks) and limiting robustness and generality. Hyperparameter tuning was not exhaustive.
- **Static, implicit environment forcing.** Environmental effects (wind, currents, tides) are only implicitly present in AIS training data. The model does not have the ability to explicitly simulate or vary these forces.
- **Post-posed, non-learning safety layer.** The MPC shield corrects unsafe actions at runtime but the agent does not learn from corrections; the potential benefit of integrating corrections into training was not exploited.
- **Perception and head-on sensitivity.** The raycasting + shield approach is less effective for some head-on encounters and for obstacles aligned with main transit directions.

- **Simplified hydrodynamics and local effects.** Hydrodynamic interactions (suction, bank effects) and structure-induced flow changes were not modelled.
- **Local scope.** Experiments reconstruct a single month and a local area. Network-level propagation of demand changes was not evaluated.

These limitations partly explain the remaining gap between simulated and historical kinematics and indicate where future work should focus.

C. Recommendations

To improve the applicability of data-driven maritime traffic simulation for design applications, the following steps for future research can be recommended:

1) **Multi-class agents:** Train separate agents per vessel class (tug, tanker, cargo, ferry, recreational/fishing) to capture class-specific manoeuvring and interaction behaviour.
2) **Data completeness and augmentation:** Increase the training set (and balance classes), recover or impute missing headings where possible, and include longer/higher-quality historic windows to stabilise training.
3) **Integrate learning from safety interventions:** Allow the agent to incorporate MPC corrections so that safety interventions improve future policy behaviour rather than act only post-hoc.
4) **Route-aware training:** Train agents to follow waypoint or multi-goal sequences (rather than single-point goals) so route-following and channel adherence emerge naturally.
5) **Explicit environment modelling:** Include explicit environmental effects and structure-induced dynamics to assess the effects of these external influences on ship behaviour.
6) **Perception and avoidance improvements:** Refine obstacle detection and avoidance for head-on cases.
7) **Hydrodynamic coupling:** Where close encounters and structure effects matter, couple simplified hydrodynamic models or empirically derived lateral biases into the agent dynamics.
8) **Robust training and calibration:** Address numerical stability, perform systematic hyperparameter optimisation, and validate across larger multi-month datasets and multiple seasons.
9) **Network-level assessment:** Combine microscopic outputs into network-scale analyses to evaluate how local infrastructure changes propagate through the wider waterway system.

With these improvements, a hybrid modelling approach (AIS-trained agents + goal/route conditioning + safety-aware correction + explicit environmental forcing) can become a practical, reproducible foundation for assessing how prospective infrastructure (such as an HSC gate complex) alters vessel behaviour, operational risk, and capacity.

References

Alshiekh, M., Bloem, R., Ehlers, R., Könighofer, B., Niekum, S., & Topcu, U. (2017, September 3). Safe reinforcement learning via shielding. https://doi.org/10.48550/arXiv.1708.08611

Basrur, C., Singh, A. J., Sinha, A., & Kumar, A. (2021). Ship-GAN: Generative modeling based maritime traffic simulator. *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 1755–1757.

Bellsolà Olba, X., Daamen, W., Vellinga, T., & Hoogendoorn, S. P. (2017). Network capacity estimation of vessel traffic: An approach for port planning. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *143*(5), 04017019. https://doi.org/10.1061/(ASCE)WW.1943-5460.0000400

Bellsolà Olba, X., Daamen, W., Vellinga, T., & Hoogendoorn, S. P. (2019). A method to estimate the capacity of an intersection of waterways in ports. *Transportmetrica A: Transport Science*, *15*(2), 1848–1866. https://doi.org/10.1080/23249935.2019.1652866

Burkley, G. B., Webb, D., & Pierce, J. J. (2022, November 30). *Ship simulation transit scenario modeling & ship pilot study of the maritime implications for commercial ship transits of the proposed houston ship channel gate complex* (Version 7.2). Locus LLC. Arnold, MD, USA.

Dalal, G., Dvijotham, K., Vecerik, M., Hester, T., Paduraru, C., & Tassa, Y. (2018, January 26). Safe exploration in continuous action spaces. https://doi.org/10.48550/arXiv.1801.08757

Dawood, M., Shokry, A., & Bennewitz, M. (2024, December 5). A dynamic safety shield for safe and efficient reinforcement learning of navigation tasks [version: 1]. https://doi.org/10.48550/arXiv.2412.04153

Delft High Performance Computing Centre (DHPC). (2024). *DelftBlue supercomputer (phase 2)*. https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2

Desyani, R. (2019). Comparison of paradigms in nautical traffic models. Retrieved June 24, 2025, from https://repository.tudelft.nl/record/uuid:76029292-c011-4bba-a313-2a28de3be380

Düz, B., & van Iperen, E. (2025). Ship trajectory prediction using encoder–decoder-based deep learning models [_eprint: https://doi.org/10.1080/17489725.2024.2306339]. *Journal of Location Based Services*, *19*(2), 146–166. https://doi.org/10.1080/17489725.2024.2306339

Greater Houston Port Bureau. (2023a). Coastal spine must balance storm protection with safe navigation and long-term economic vitality of the port of houston. Retrieved July 11, 2025, from https://assets.noviams.com/novi-file-uploads/ghpb/images/Coastal_Barrier/EPCO_IkeDike_Brochure_Final_pdf_11_28_rev-13af9cd3.pdf

Greater Houston Port Bureau. (2023b). *Proposed houston ship channel gate complex - greater houston port bureau*. Retrieved July 11, 2025, from https://www.txgulf.org/houston-ship-channel-gate-complex

Koenighofer, B., Bloem, R., Jansen, N., Bochum, R.-U., Junges, S., & Pranger, S. (2024). Shields for safe reinforcement learning.

Li, R., Li, T., Bu, R., Zheng, Q., & Chen, C. L. P. (2013). Active disturbance rejection with sliding mode control based course and path following for underactuated ships. *Mathematical Problems in Engineering*, *2013*, 1–9. https://doi.org/10.1155/2013/743716

Mathioudakis, M., Papandreou, C., Stouraitis, T., Margari, V., Nikitakis, A., Paschalakis, S., Kyriakopoulos, K., & Spyrou, K. J. (2025, January 23). Towards real-world validation of a physics-based ship motion prediction model. https://doi.org/10.48550/arXiv.2501.13804

NOAA. (n.d.). AccessAIS. Retrieved September 3, 2025, from https://marinecadastre.gov/accessais/

Pan, R., Zhang, W., Wang, S., & Kang, S. (2025). Deep reinforcement learning model for multi-ship collision avoidance decision making design implementation and performance analysis. *Scientific Reports*, *15*, 21250. https://doi.org/10.1038/s41598-025-05636-3

Papandreou, C., Mathioudakis, M., Stouraitis, T., Iatropoulos, P., Nikitakis, A., Paschalakis, S., & Kyriakopoulos, K. (2025, March 3). Interpretable data-driven ship dynamics model: Enhancing physics-based motion prediction with parameter optimization. https://doi.org/10.48550/arXiv.2502.18696

Pham, Q. A., Brahmanage, J. C., & Kumar, A. (2025). ShipNaviSim: Data-driven simulation for real-world maritime navigation. https://www.ifaamas.org/Proceedings/aamas2025/pdfs/p1641.pdf

Sheebaelhamd, Z., Zisis, K., Nisioti, A., Gkouletsos, D., Pavllo, D., & Kohler, J. (2021, August 11). Safe deep reinforcement learning for multi-agent systems with continuous action spaces. https://doi.org/10.48550/arXiv.2108.03952

Shu, Y. (2019). *Vessel route choice model and operational model based on optimal control* [Dissertation (TU Delft)]. TRAIL Research School. https://doi.org/10.4233/uuid:5e700fc1-7620-4ab0-9b72-859e2db7926b

Yip, T. L. (2013). A marine traffic flow model. *TransNav, International Journal on Marine Navigation and Safety od Sea Transportation*, *7*(1), 109–113. https://doi.org/10.12716/1001.07.01.14

Zhou, Y. (2022). *Ship behavior in ports and waterways: An empirical perspective* [Doctoral dissertation, Delft University of Technology]. https://doi.org/10.4233/UUID:DAB6C5BB-5DAA-496A-A912-8C3DFAFF0EBE

Zhou, Y., Daamen, W., Vellinga, T., & Hoogendoorn, S. (2019). Review of maritime traffic models from vessel behavior modeling perspective. *Transportation Research Part C: Emerging Technologies*, *105*, 323–345. https://doi.org/10.1016/j.trc.2019.06.004