AUTONOMOUS MISSION-PLANNING AND COLLABORATION BETWEEN MULTI-DOMAIN ROBOTS

by

Joel M. Lindsay

Submitted in partial fulfillment of the requirements
for the degree of Master of Applied Science

at

Dalhousie University
Halifax, Nova Scotia
December 2020

Table of Contents

List of Tables

List of Figures

vii

Abstract

Autonomous multi-robot path-planning and task-allocation to address unresponsive or even evasive targets on or under the water is studied. Challenges in robot localization and navigation in GPS-denied and communication constrained environments are addressed. The *mRobot* node developed autonomously path-plans and task-allocates to manage robotic collaboration in marine environments. The node is validated in simulation and controlled in-water tests with a team of heterogeneous marine robots (unmanned surface vehicle (USV), unmanned aerial vehicle (UAV), unmanned underwater vehicle (UUV)) to survey static floating targets (of known pose) in three-dimensions.

Then, a novel mutual-information incentivised Q-Learning algorithm is developed for UUVs to search for static underwater targets of uncertain pose. The planning considers the complex underwater environments where erroneous and false detections are expected. Simulation and controlled two-dimensional experiments show the algorithm performs notably better than alternative methods like greedy or Boustrophedon.

Lastly, a collaborative team of three UUVs is proposed to acoustically detect, track, and localize a mobile, evasive underwater target with uncertain pose. A novel algorithm combines predictive information measures with Q-Learning for trajectory planning. The algorithm adapts to conditions that impact detection with acoustic range-only measurements. Simulation results show superior performance of the 3-UUV system compared to the long baseline method.

Acknowledgements

First and foremost, I would like to thank Dr. Mae Seto and Dr. Robert Bauer for their expertise, direction and generous support. I am very proud to have completed this thesis under their supervision.

I would also like to acknowledge the members of the ISL research group for all of the discussions they shared with me. Specifically, I want to thank Jordan and Amy for their advice and friendship. I was very fortunate to work with such an excellent lab group.

Finally I would like to thank my parents for the support they have provided me throughout my studies. My achievements would not be possible without their encouragement.

Chapter 1

Introduction

1.1 Motivation and Thesis Contributions

Autonomous marine robots are now routinely considered for ocean monitoring, search and rescue, and naval applications. Developments in their autonomy minimizes human involvement in the difficult and harsh ocean environment [1] and expands their capabilities for more complex missions. An example of a naval application is addressing unresponsive or even evasive targets on or under the water, specifically when there is no prior information or situational awareness on their state, origin, or purpose.

If the target is a stationary non-responsive ship, below-water surveying could detect and identify objects attached to its wetted hull, and provide indications of its propulsive capabilities. Above-water situational awareness could yield information on the target's structure, purpose and, potentially, people and objects on-board. If maritime interdiction operations are considered then the information from such a multi-domain survey could determine if it is safe to attempt a boarding. If the target is an iceberg, then a multi-domain survey could characterize its structure towards a more informed approach to safely circumnavigate it. If the mission objective is to search for a downed aircraft or boating accident, then the information collected in such multi-domain surveys could inform the search on how to proceed [2].

Alternatively, if the targets are on the seabed and of uncertain location, then an area survey must be executed to detect, localize and classify them. Naval mine-countermeasures (NMCM) is a special case of the underwater target search, with a requirement that all mines be detected and localized to an acceptable certainty in the area survey. Previously, NMCM operations required ships and divers to complete this dangerous task. With advances in marine robotics, it is now possible to complete the surveys with unmanned underwater vehicles (UUVs) [3]. However, there are limitations that challenge UUVs in fulfilling NMCM requirements. To start, the acoustic sensors (sonar) used. Target detection by sonar is affected by environmental

conditions (physics of underwater sound propagation) like ambient noise, time-and spatially-varying sound velocities, multi-path, reflections, scattering, and bathymetry among others. Such conditions vary with location and time. Uncertain measurements mean that 100% sensor coverage of the search area does not ensure 100% detection [4]. The underwater target localization problem requires an adaptive method for resolving uncertain and false detection events to report a more accurate map of target placement. The other significant challenge for UUVs in NMCM surveys (and underwater applications in general) is self-localization and navigation without access to a global positioning system (GPS) or high bandwidth communication. Often, in GPS-denied conditions proprioceptive sensor fusion is relied upon for state-estimation. Combining internal sensors for localization or navigation without GPS influence or acoustic position support is termed inertial/dead reckoning (DR). An individual underwater vehicle relying on internal equipment can surface for GPS to update its state, at the cost of efficiency. The typical Inertial Navigation Systems (INS) can achieve a drift in position error of 2%-5% of the total distance travelled. Smaller position error is achievable with a higher accuracy (and more costly) INS [5]. However, the error is unbounded and unavoidable with INS due to uncertainty in the measurements, and environmental disturbances.

Lastly, if the target is a mobile underwater vehicle (marine mammal, submarine, UUV) entering a marine protected area or sea-port without authorization, then more advanced solutions are required to detect, localize and potentially track the target. The absolute position of the target can be difficult to obtain as the GPS is unavailable due to the rapid attenuation of high-frequency electromagnetic waves (EM) underwater [6]. To address this problem, acoustic long baseline (LBL) localization systems can be used. LBL systems use a network of stationary underwater beacons that range the target with acoustic measurements. Long baseline systems can provide accurate localization; however, LBL beacons are non-trivial to deploy and calibrate, and have very limited range [6]. They are better in small confined areas, but less so in open ones. As an alternate method to stationary underwater beacons, sensors on mobile UUVs may be an option to track the target.

This thesis makes contributions towards addressing these autonomous marine vehicle path-planning and robotic collaboration challenges —to obtain information and

situational awareness on marine targets depending on their pose uncertainty and mobility. The thesis objectives are, therefore, to:

1. create a node that autonomously plans missions and distribute tasks for collaborative multi-robot systems across multiple domains (under-, on- and above-water);

2. develop a light-weight autonomous planning algorithm for underwater search missions with increased efficiency and accuracy that is meant to be deployed on embedded platforms, and

3. build on objectives 1 and 2 to develop an autonomous planning method for a collaborative system of three UUVs to detect, localize, and track an unresponsive, mobile and evasive target.

1.2 Methodology

To achieve the objectives, this thesis explores the use of multi-robot systems to implement complex missions in marine environments. The problem of gaining awareness on targets is addressed through *in situ* autonomous planning and robotic collaboration.

The ultimate goal is to study the efficacy of an autonomous multi-robot system to detect, localize and track a non-responsive and mobile underwater target. The work is presented in three parts to systematically build and culminate in that goal. Each chapter builds on the previous by developing a solution which addresses non-responsive target(s) with progressively more pose uncertainty and mobility:

1. a single stationary target that is partially-submerged (floating) and its pose is known with good certainty prior to the mission;

2. multiple stationary underwater targets (of unknown quantity) whose pose is not known with good certainty, and

3. a single mobile and evasive underwater target whose location, velocity and heading are not known with good certainty. Evasive refers to the target's ability to change speed and heading.

Next, Chapter 2 provides a review of previous work on collaborative robot localization and navigation to address targets in the marine environment.

Chapter 3 details the development and validation of a node, *mRobot* (for multiple robot), which autonomously plans missions and distributes tasks for below-water, surface and above-water marine robots towards situational awareness on a partially-submerged target. Simulation and experimental validation of the systems' capability across multiple domains is described.

Chapter 4 describes an autonomous planning algorithm for underwater search missions for targets (e.g. seabed mines) whose pose are only known to the extent that they might exist in the area. The algorithm combines on-line reinforcement Q-Learning trajectory planning with information gain rewards. To quantify the merit of this proposed solution, its results are compared with the standard greedy and Boustrophedon approaches. A sensitivity study is provided which identifies the parameters that drive the planner's performance. Simulation and two-dimensional experimental validation is undertaken.

Chapter 5 describes the development of a system which autonomously plans and distributes missions for three collaborating UUVs to detect, localize, and track an evasive target. The solution builds upon the prior two chapters. The *mRobot* node is adapted to network the team of UUVs. The search channel formulation from [7] is adapted to a dynamic channel and predictive information measures, and the Q-Learning trajectory planner is expanded to policy learning for three UUVs. Simulation results show an improved performance and capability over an LBL system.

Lastly, Chapter 6 provides conclusions and thoughts on future work.

Chapter 2

Literature Review of Marine Multi-Robot Collaboration

This chapter provides a review of research and development pertinent to collaborative robotics in marine environments related to the three marine target types presented in Chapter 1 —namely a static target of known pose, multiple stationary targets of unknown pose, and a mobile target of unknown pose.

To start, relevant background in marine robot sensing is presented (section 2.1). Then, the impact of underwater communication limitations on multi-robot collaboration is outlined (section 2.2). Section 2.3 then defines robot localization and discusses standard acoustic localization techniques and their limitations, while section 2.3 concludes with multi-robot collaborative localization.

Robot navigation is addressed in section 2.4 with a review of multi-robot systems in marine applications. Firstly, the communication and navigation aid (CNA) autonomous vehicle is presented. Secondly, work related to surveying a partially-submerged target is reviewed. Thirdly, the challenges and current methods for marine target searching are presented. A discussion on reinforcement learning-based path-planners, as a proposed solution, is also provided. Lastly, mobile marine robots in target-tracking and localization applications are reviewed.

Prior to the review, the definitions of terms used in the thesis are described. The *domain* refers to whether the robot operates above or below-water. The environment for a particular domain includes other physical properties like temperature, pressure, obstacles, and fluid medium density. The term *marine environment* refers to both above and below water domains. Any combination of multiple robots tasked to a common goal or *mission* is referred to as a *team*. The act of working together to achieve the goal is robotic *cooperation*. The amount that a team's members interact among themselves to achieve this goal can range from *coordination* to *collaboration*. The distinction is based on how tightly coupled the team members' individual tasks are. In *coordination* the team may be centrally managed by a supervisor (such as

a robot, agent, or human operator) which assigns their allocated task so the robots do not need to interact much among themselves to achieve the mission. An example is two robots performing a side-by-side survey which is coordinated by a central supervisor. The execution of the two tasks are not dependent on one another and it may not matter that one finishes before the other. With collaboration, there may be tight cross-robot timing constraints where one robot waits on the result of another prior to planning and executing its task. One definition of collaboration is: *multiple assets working together directly and sharing resources to support a shared goal* [8], where *asset* refers to any robot or vehicle. A collaboration assembles a robot team with consideration for the mission's sensing requirements against each robot's individual capabilities (e.g. sensors, mobility, endurance, pose certainty, on-board computations, environment, communications) and what each team member contributes towards the common goal. The justification for collaboration is discussed next.

A motivation to use robot teams is to better address more complex goals. A single robot system is limited by one robot's mobility, computation capability and sensors. The capabilities of multi-robot systems applied to a common goal can be greater purely from the amplification of having more robots. The force multiplication from more robots contributes scaling and can offer a wider range of complementary sensors. A collaborating multi-robot solution can reduce the time to complete a task, add measures of redundancy, and provide more adaptability to address dynamic missions [9]. However, collaboration is contingent upon some level of inter-asset communications which can increase mission complexity, as environmental conditions that impact the communications channel become important. Given wireless communications between robots, how tightly coupled the assets are can manifest in the communications bandwidth needed between members to execute the common task. The amount of bandwidth between team members can vary, as in this thesis, due to the environment (communications channel) or a robot's specific tasking. Given the underwater communications challenge, there is still considerable value to use collaborative teams. Furthermore, more complex goals can also be addressed by making the teams' planning adaptive to the dynamic robot and environment.

Dynamic mission planning is an emerging requirement in real-world applications.

Adaptive responses are necessary to address unanticipated robot-wide failure, sub-system failure, loss of communication, environmental changes, or new information that impacts the robot's interactions with its environment. The potential of robots collaborating in dynamic mission-planning, and to provide additional levels of redundancy are other benefits that outweigh the complexity required in the development and deployment of multi-robot systems.

To discuss and review work related to multi-robot collaboration and the localization and navigation requirements, as it pertains to the three target types investigated in this thesis, relevant background on marine sensing and communications is provided next.

2.1   Background on Robotic Sensing

The taxonomy for sensors can depend on modality, function, whether it is proprioceptive or exteroceptive, interaction with the environment, and operating domain. Table 2.1 summarizes some marine robot sensors with this taxonomy. The table is drawn from [10], [11], [5], and [12]. The sensors shown will be touched on in this subsection.

On-board a robot, *vehicle* sensors are necessary for basic functions. Examples of mobile robot sub-systems include: propulsion, pose determination, attitude control, navigation, power distribution, off-board communications, buoyancy (marine robots), as well as faults and failure detection. Sensors in these sub-systems contribute to the robot's functioning. Of interest to this thesis are navigation sensors which will be discussed later.

In the broader picture, the robot is a platform to support the *payload* sensors. These sensors collect the data to fulfil the mission objectives. An example of a payload sensor is a camera on an unmanned aerial vehicle to collect imagery for surveillance. The robot platform can support the payload sensor by providing power, protection against the environment, attitude stability, lift/locomotion to bring the camera to where it is to perform its sensing, on-board processing, planning, decision-making, communications with the operator, and pose information to geo-reference the sensing. The selection of sensors must balance speed of operation, cost, error rate, robustness, and power consumption [12].

Table 2.1: Summary of sensor descriptions and their operating domain

	sensor		description	domain	
				below-water	above-water
exteroceptive	radar *(radio detecting and ranging)*		Uses radio frequency EM waves for detection of objects. Can be used to classify range, angle and velocity of object in the environment		✓
	lidar *(light detection and ranging)*		Similar principles to the radar - utlizes light instead of EM waves for measurements	✓	✓
	laser range finder		Estimates distance to objects in the environment using time of flight, phase-based, or triangulation methods		✓
	infrared (IR)		Fast and inexpensive proximity sensor, used to estimate range based on intenstity of returned signals		✓
	sonar *(sound navigation ranging)*	Doppler velocity log (DVL)	Acoustic measurments to determine ground speed relative to sea floor. Determines surge, sway, and heave velocity vecotes based on the doppler shifted returns of acoustic pulses off the seabed.	✓	✓
		multi-beam/ bathymetric	Multiple sonar beams that return depth measuremnts based on time of flight - used to scan surfaces. Often used for creating 3D images of seabed or underwater structures	✓	✓
		forward looking sonar	Forward facing multi-beam sonar, more suitable to navigation target detection and guidance	✓	✓
		side-scan	Multi-beam sonar that creates 2D images through intensity measurment returns	✓	
	camera	monocular	Single camera vision	lighting and depth dependent	✓
		stereo	Dual camera system that can determine rough depth estimates based on relationship between the cameras (similar to human vision)		✓
	GPS		Global Positioning System, using time-of flight singals from syncronized satelittes		✓
proprioceptive	IMU *(inertial measurment unit)*		An inertial navigation system (INS) can calculate postion, velocity, and attitude using IMU meaurments. In general an IMU will consist of three accelerometers and three gyroscopes to measure the 6-DOF motion. Accelerometers and gyroscopes measure linear and angular acceleration respectively	✓	✓
	pressure sensor		Depth measurement device (barometer or pressure sensor)	✓	
	compass		Provides global heading reference, generally done by measuring the magnetice field vecotrs. Less common is the gyrocompas that uses a spinning disk and the earths rotation (not affected by other metalic objects)	✓	✓

Whether any particular sensor is a *vehicle* or *payload* sensor depends on its use. If the camera were used for the robot's obstacle avoidance it would be a vehicle sensor. It is also common to classify robot sensors as proprioceptive or exteroceptive types. Proprioceptive sensors measure quantities internal to the robot (e.g. motor speed, acceleration, heading, battery status). They do not require that energy be put into the environment. Exteroceptive sensors measure quantities that are external to the robot like the environment, e.g. ambient lighting, ranges to targets, wind/current speed.

Robot state estimation is achieved with vehicle sensors. These sensors include proprioceptive types like GPS, inertial measurement units/INS and compasses. Inertial sensing requires external references to bound the otherwise unbounded error growth. Therefore, active exteroceptive sensors for ranging like free space optical, LIDAR, IR, radar, sonar, beacons are used. Passive exteroceptive sensors like cameras for vision-based approaches can also be used to bound error growth [12]. With the exception of GPS, radar and IR these sensors types are available both above and below water for marine robots.

Robotic payload sensors vary vastly across domains and are application specific. Sensing modalities like free space optical and electromagnetic (EM) waves are limited in range underwater. These waves are very high frequency and rapidly attenuate with range. At short ranges they are a good option as they have more bandwidth. As well, optical sensing requires a light source and so robots that work in deeper (greater than 20 ft) must carry sufficient lighting. Even at shallow depths, turbidity will compromise image quality for visible wavelength cameras. Given the high specific heat of water, infrared cameras are not effective in water.

Underwater, acoustics provides the best sensing and communications modality given the poor propagation of electromagnetic, optical and other energies in water. However, underwater acoustic propagation is not without its limitations and suffers from latency, frequency and range dependent losses, multi-path effects, and ambient noise [11]. More discussion of electromagnetic and acoustic frequencies in water is discussed in the next section.

Acoustic energy is used in vehicle state estimation (e.g. altimeters) and payload

sensing (e.g. sonar-based sea-floor mapping). Sonars can be downward-looking multi-beam where multiple acoustic beams are emitted and reflected by targets (like the seabed). Based on the two-way travel time of the reflected beam, the range between transmitter and reflector can be determined. Many such returns can be assembled to yield a three-dimensional image of the target (e.g. seabed, marine structure). Side scan and forward-looking sonars are similar in principle to multi-beam but project sideways and forward, respectively.

A robot's navigation and localization accuracy depends on its sensor suite and the state estimation algorithm which fuses them. Navigation and localization accuracy could be improved through collaboration with other sensors (e.g. UUV or a stationary beacon). As presented earlier in the chapter underwater sensing is best achieved with acoustics. Similarly, acoustic methods are best for underwater communications as well. The next sub-section provides background on acoustic communications as it pertains to multi-robot collaboration.

2.2   Underwater Communications

The bandwidth, range between vehicles, and network topology are some of the parameters that define the communication channel for a multi-robot system [11]. Depending on the operating domain a collaborative robot system will face different communication constraints. Radio frequency (RF) EM waves are effective for above-water/in-air communications. For example, as shown in [13] RF communication for a team of collaborative unmanned aerial vehicles (UAVs) was utilized to improve navigational accuracy over that of GPS. However, underwater RF electromagnetic waves do not propagate far because of high absorption in those frequencies. The absorption is two orders of magnitude greater in salt than in fresh water. The absorption coefficient (dB/km) in seawater at frequency f (kHz) can be written as the sum of chemical relaxation processes and absorption by Eq. 2.1 [14]:

$$\alpha = 0.106 \frac{f_1 f^2}{f_1^2 + f^2} e^{-(pH-8)/0.56}$$
$$+ 0.52(1 + \frac{T}{43})(\frac{S}{35}) \frac{f_1 f^2}{f_2^2 + f^2} e^{-D/6} \tag{2.1}$$
$$+ 4.9 \times 10^{-4} f^2 e^{-(T/27 + D/17)}$$

where D is depth in km, T is temperature in °C, S is salinity in ppt, and f_1 and f_2 are the boric acid and magnesium sulphate relaxation frequencies in kHz. The insight to be gained from Eq. 2.1 is that lower frequencies travel further in water. As the bandwidth is related to the carrier frequency this means that lower frequencies have less bandwidth [15]. Therefore, there is a trade-off between bandwidth and propagation distance for a given carrier frequency. The general result is that EM and free-space optical (FSO) waves are better than acoustic frequencies for short range communication underwater. Acoustic frequencies become necessary at deep depths and long ranges between transmitter and receiver; however, acoustic communications (and sensing) are limited as follows [11]:

1. Low sound speed – high latency between transmission and reception over a range of 100 meters or more results in outdated information, and signal scattering. Low sound speed contributes to Doppler shifts on underway UUVs with transmitter/receiver or wave action.

2. Losses – spreading, absorption, and scattering of acoustic signals underwater lead to unreliable communication and potentially significant packet loss.

3. Multi-path – reflections and refractions of acoustic waves creates multiple propagation paths between the transmitter and receiver, which leads to delays between the reception of the first and last instances of a message. Interference between signals can become a problem when that delay is large enough.

4. Ambient noise – ships and underwater equipment are noise sources and can create interference, which is more prominent over longer range missions and communications.

A real-world example that highlights the relative performance of underwater versus in-air communication is provided in [16]. Ludvigsen *et al.* conduct experiments

with a manned ship, UAV, UUV, and an unmanned surface vehicle (USV) in a network of heterogeneous vehicles conducting a bathymetric mapping survey. As expected, above water RF communications had good speed and reliability unaffected by range; however, the connection between the surface and underwater vehicle was intermittent and poor quality at certain stages. The underwater communication quality was the limiting factor in their system [16].

Overall, acoustic communications has developed to a point where underwater network development is feasible and complex missions can be accomplished. Transformational improvements to acoustic communication in the short term are unlikely – meaning that multi-robot systems operating in marine environments must combat and adapt to these issues through other means like collaboration.

In the literature there are a variety of recent studies that present cooperative strategies as solutions to unstable networks between mobile nodes (i.e. multi-robot systems). For example, Benavides *et al.* [17] contribute a discussion on recent advances in dealing with constrained communication environments in the application of robotic exploration. Their proposal to improve exploration is through collaboration between robots, using both survey and relay vehicles. Vehicle networks are becoming a standard approach to address constrained underwater communications and operations over large spatial environments. Similarly, a common trend in current research surrounds the use of USVs as acoustic-to-RF communication relays for heterogeneous robot teams operating above and under-water.

In summary, a brief overview of underwater communications has detailed challenges that collaborative multi-robot systems face especially in the underwater domain. The concept of heterogeneous robot systems as members in a communication network for above and below water was introduced. Given the background on underwater sensing and communications, it is now relevant to review how robotic localization is performed and the advantage of collaborative localization.

2.3 Single-and Multi-Robot Marine Localization

This section introduces commonly used acoustic localization techniques and their limitations. Then, collaborative localization is defined and recent research and development in the area is reviewed. The ultimate goal for this subsection is to illustrate

the benefit of mobile and collaborative localization methods in the marine environment.

The robotic localization problem aims to answer the question: where is a robot within its environment? Siegwart and Nourbkhsh describe robotic navigation as built upon four building blocks of perception, localization, cognition, and motion control [18]. By definition, localization and navigation overlap in some aspects; however, a distinction is made where localization is specific to how accurately a robot can estimate its pose within its environment. Navigation is the ability for the robot to transit between 2 points, avoid obstacles and find traversable paths. Accurate robot localization is critical to geo-reference perceived marine targets for follow-on actions.

To start, Paull *et al.* [5] offers an in-depth review on the state-of-the-art in underwater localization as of 2014. The authors break down localization (and navigation) techniques into the three main categories of inertial, acoustic, and geophysical. Inertial state-estimation was introduced in Sections 1.1 and 2.1 in the discussion on proprioceptive sensing. The focus here is on underwater acoustic (exteroceptive) methods. It is noted that the principles for underwater localization are the same as those for above water vehicles. The transducers are of course different. For example, unmanned aerial vehicles operating in GPS-denied environments face similar localization challenges. Localization with acoustic transponders and modems based on range and bearing measurements from a reference with good pose certainty has become common. Here, the time-of-flight (TOF) for a transmitted signal determines the range between a transmitter and receiver. Range-only measurements can bound the position error of the receiver vehicle, eliminating the need for underwater vehicles to surface for GPS. Paull *et al.* [5] summarize common localization methods using acoustic TOF measurements. Commonly used methods include ultra-short baseline (USBL), short baseline (SBL), long baseline (LBL), and a single fixed beacon. USBL navigation allows underwater vehicles to localize relative to a ship (which knows its position well from GPS), where range and bearing can be determined using TOF and the phase differences across the transceiver array. SBL triangulates with transceivers at either end of a surface ship. USBL is limited in its range and SBL's accuracy depends on the ship's length [5]. In LBL, acoustic triangulation determines the range

of a transmitting (interrogating) underwater target from a set of widely-spaced geo-referenced, fixed beacons. Another established method is a single fixed beacon system which is more economical and faster to deploy. However, range measurements from a single fixed beacon does not provide a unique solution for estimating a mobile target's pose. A singe beacon could provide some bound on the pose error of a UUV and is thus an improvement on dead-reckoning. LBL and single beacon systems use the two-way travel time (TWTT) of the acoustic signal between the transmitting target (e.g. UUV) and beacons to determine range. However, [5] present refinement on TWTT where synchronized clocks on the beacons and transmitting target can support one-way travel time (OWTT) ranging with LBL. However, OWTT range measurements will not work to detect and localize an unresponsive target. The challenges of LBL include cost, the complexity to deploy and recover multiple beacons, and the maintenance of their wireless network while deployed. Furthermore, the fixed locations of LBL beacons inherently limits the range that they are effective over. A visual summary of these techniques [5] is shown in Fig. 2.1. From [5], TWTT will be considered for this thesis. As well, this thesis considers the limitations of LBL which could be improved upon with the proposed methodologies.
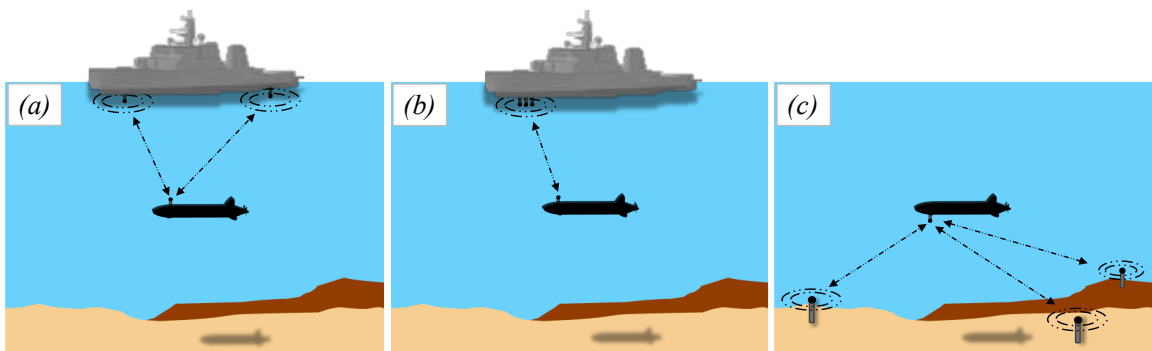
Figure 2.1: Traditional underwater localization or navigation using acoustic communication a) SBL, b) USBL and c) LBL; adapted from [5]

At the forefront of recent acoustic localization research is the use of intelligent mobile beacons to aid localization, specifically of a collaborative nature. In cooperative or collaborative localization (CL) the individual team members use on-board proprioceptive sensors for inertial localization that is combined with range updates from other team-members. Examples of underwater CL architectures include [9]

where the authors describe how a single UUV and two USVs can be used to improve the position estimation error of the underwater asset. Chitre [19] illustrates a CL example between two heterogeneous UUVs (one supporting the other with more advanced DR equipment), while other examples of homogeneous UUV teams is also not uncommon. Above water, teams of varying make-up can be used for collaborative localization including UAVs, USVs and terrestrial robots. For example, in [13] the authors explore localization and navigation techniques of a multi-UAV team where GPS is either denied or nominal. The trend seen in both above-and underwater CL is the use of a beacon vehicle that either has access to GPS or sophisticated localization sensors used to reduce the localization error of all team members within communication range. The shared information between the beacon and other vehicles is generally a pose estimate with a level of uncertainty. Measuring the propagation delay of messages sent between the vehicles is used to estimate range if synchronized clocks are available on both. However, if clock synchronization is not available then two-way propagation delay may be used to measure the range [20]. Acoustic mobile beacons like those integrated on a UUV, and especially those on a USV, compared to the aforementioned baseline systems are operational over much larger areas. These beacons may also be a cost-savings compared to say an USBL which requires ship support. Some examples are presented next.

An example of collaborative localization research is provided by Bahr *et al.* [9] who present a technique for underwater cooperative localization where team members share range measurements acoustically to perform mobile trilateration. In [9] every vehicle is equipped with proprioceptive sensors for dead-reckoning and exteroceptive sensors to measure range and bearing to other team-mates. The approach provided is flexible and decentralized as it does not require any vehicle to acquire all messages that are broadcast. The method is beneficial as it helps solve a major challenge in cooperative navigation and localization—the cross-correlation between shared position estimates of the vehicles in the cooperation. For example, when robot A uses the pose estimate of robot B to update its own position estimate (from a knowledge of the relative range and bearing between the two) then the position estimates of vehicles A and B are correlated [9]. The concept of mobile trilateration from [9] is explored in this thesis, except for the case where the vehicle being localized is non-responsive

and a collaborative team is attempting to track the target using active sonar TWTT measurements. Another example of collaborative localization uses a surface relay in the form of a USV.

Fallon *et al.* [21] built on [9] by introducing the concept of a single mobile surface craft supporting the localization and navigation of multiple UUV team members. This solution eliminates the need for UUVs to surface for GPS position fixes, and thus improves mission efficiency. This work published in 2010 claims to be the first to show on-line localization and navigation of a UUV using range measurements from a USV [21]. A discussion of scalability is offered for this method where it is suggested that one-way ranging allows for scalability to any number of UUVs within the USV broadcast range. However, a limited underwater broadcast range is itself a restriction on a system relying on a support USV. Ultimately, simulation and experimental results were provided to show an error reduction of approximately 80%, which is substantial. A single USV is used as a support vehicle for the collaborative system proposed towards surveying a floating target (Chapter 3).

Related to surveying a floating target, a unique localization problem arises from 'GPS shadows' near the floating structure. This GPS shadow is the deterioration of a GPS signal near, or under, floating structures [22]. Han *et al.* [22] propose a simultaneous localization and mapping (SLAM) solution for 3D reconstruction of a floating structure with a single USV, where landmarks, surface texture, colouring, and other environmental features are used to estimate the USV pose during its mapping survey. A number of researchers address solutions to localize in these GPS shadows, although a collaborative multi-robot solution has not been found by the present author.

In summary, a review of acoustic localization techniques was presented. The idea of collaborative localization was introduced and discussed, where it was shown how organized communication between robots can be utilized to improve localization and information sharing. Included was a short discussion on the relevance of the localization problem in relation to the 3D reconstruction of partially-submerged structures. The following section transitions to a review of navigation, dynamic planning, and collaborative robot systems towards addressing the three target types of interest in this thesis.

2.4   Marine Multi-Robot Navigation and Autonomous Planning

If localization answers the question of where does a robot exist within its environment, then navigation and planning answers the question: where should the robot go, and how does it get there [18]? Siegwart and Nourbakhsh [18] suggest that the functionalities for navigation are planning and reactive response, where the challenge involves executing a course of action to acquire a goal.

Navigation objectives within a collaborative system can vary with the robotic vehicle. For example, navigation objectives can include optimizing a trajectory for a energy-efficient survey or search mission or optimizing the way one asset aids its team-mates. The benefits and complexities of multi-robot systems have been previously discussed. These complexities are largely in the multi-robot navigation and planning phases. Developing an autonomous decision-making process to include task allocation, coordination and planning scales in complexity with the number of assets involved. The design of a multi-robot system must include consideration of the team make-up (vehicle characteristics and sensors), system control (central or decentralized), and the communication structure [23]. In summary, the multi-robot cooperative navigation problem can be broken into three components in [11]:

1. shared task view and goal representation;

2. decision-making process, and

3. behaviours enabling agents to perform tasks (not discussed).

Understanding the shared task is important for multi-robot navigation and improving team coordination. For example, shared task and goal representations is in relation to area coverage percentage, operating within a time limit, or searching based on probability distributions [11]. Another important consideration is data management and the use of navigation techniques to improve the sharing of information. Decision-making including task allocation and system hierarchies are also major considerations.

The following sub-sections review the use of multi-robot systems as well as their planning towards the challenges of the three target types of interest in this thesis. Firstly, the communication and navigation aid (CNA) vehicle is introduced as a collaborative asset to improve communications and information sharing between team

mates. Secondly, multi-robot systems research work related to multi-domain surveys on a floating target are presented. Thirdly, marine target search is reviewed —with an introduction to reinforcement learning-based navigation as a proposed solution. Lastly, a review is conducted on the state-of-the-art for autonomous target tracking.

2.4.1 Communication and Navigation Aid (CNA) Research

In multi-robot solutions the navigation and data flow are constrained by the communication channel bandwidth. In GPS-denied or nominal conditions (specifically underwater) cooperative/collaborative methodologies between robots has become a solution for completing complex navigational challenges that are not otherwise possible without some level of cooperation. The CNA vehicle, often referred to as a beacon vehicle, supports or aids other team-mates by providing information that can be used to improve their state estimation through collaborative localization. The CNA is a well researched topic and is an example of how collaborative navigation can improve a solution over that of a single robot system. The CNA is also crucial for more reliable communication networks between assets that are distributed above and below the water to improve mission efficiency, and allow for more timely and opportunistic task allocation. A CNA team-mate is central to the system developed in Chapter 3. A centralized architecture for dynamic planning with a lead UUV managing task allocation is also applied to the collaborative-UUV tracking system proposed in Chapter 5. A review of the CNA literature is provided next.

To start, Chitre [19] offers an optimized path-planning algorithm for a beacon UUV to minimize errors accumulated by other UUVs. It is assumed that the beacon, or CNA, knows its position well and transmits a signal periodically, which is used by other UUVs within range to bound their position error. It is assumed that the UUVs do not alter their depth so that the path-planning remains two-dimensional. The CNA path-planning can be performed off-line prior to a mission or on-line during the mission. On-line planning is a critical advancement as dynamic environments often require adaptivity that cannot be accounted for off-line. The algorithm optimizes the CNA path by deciding on a new heading at each time-step that minimizes the total UUVs' position errors [19]. In simulation, the author shows how the position error of one or two survey vehicles can be bounded by one CNA. The proposed algorithm

also considers look-ahead levels which account for future UUV positions (waypoints) for better CNA planning [19]. However, a significant trade-off between number of look-ahead levels and computational cost must be managed. The concept of the look-ahead level is applied to the problem of evasive target tracking (objective 3) in this thesis.

Unlike the previous work, Fallon *et al.* [21] introduced the concept of a single mobile surface craft to support localization and navigation of UUV team members. The USV path should be planned based on the dedication of the communication system to localization and the USV mobility compared to the UUV fleet. They conclude that the major considerations be the communication system's level of use, maintenance of an upper bound on the UUVs' pose errors, and transmissions occurring only when there is sufficient error reduction [21]. They also discuss observability as it is critical to state estimation for cooperative navigation (and localization) with range-only measurements. A definition of observability is as follows: *if a UUV receives a range measurement from another vehicle from the same relative direction then the linearised system is unobservable. However, if a beacon vehicle were to manoeuvre to achieve radial coverage of a survey UUV then the system would be fully observable.* Fallon *et al.* ultimately utilize two path-planning algorithms: a 45° zig-zag pattern behind a single UUV, and another that circles a single UUV continuously to achieve full system observability. As well, the authors offer insight and results suggesting that the effects of range on the range measurement itself is broadly independent. However, completing mobile versus stationary range measurements likely reduces the accuracy of the range function [21]. The solution was demonstrated in simulation and experimentally. A conclusion is that a USV as a CNA is beneficial for access to GPS and its higher speed and maneuverability better supports the UUV fleet compared to a submerged CNA. Others have built on this work with a more optimal planner.

Hu *et al.* [24] present a novel algorithm for cooperative navigation with a system of surface and underwater marine vehicles. Like previous work, their USV has GPS access and their UUV is enabled with an INS. They show an example of adaptive USV path-planning to ensure a set 'communications' range is maintained, where the USV uses line-of-sight (LOS) based proportional-integral-derivative (PID) control. LOS PID control maintains a set radius between the UUV and USV based on range

measurements between the two vehicles. However, their focus was to maintain communications between the UUV, USV, and a base station as opposed to minimizing UUV position error. Simulation results compared the UUV set point trajectory under dead-reckoning to that of cooperative navigation. Their system was still successful in bounding UUV position error [24] even though it was not the primary objective. However, this method was not explored for more than one underwater vehicle, nor verified experimentally where communication quality between the UUV and USV depends on more than just range. An example of yet a more advanced USV planner which aims to minimize the UUV pose estimate is described next.

Hudson and Seto [1][3] focused on algorithmic development for adaptive CNA path-planning. Specifically, their algorithm built on [19] by applying distance penalties to the path-planner to avoid losing communications while also avoiding collisions between team-mates. They add to the discussion of trade-off's in performance versus computational cost with look-ahead levels. Their results illustrated that a distance penalty is not required if the number of look-ahead levels is sufficient to avoid collision. The present work in Chapter 3 builds upon this algorithm to consider penalties for extending beyond the communication range with both above- and below-water team-mates, along with penalties for approaching the surface target too closely —to remain covert. The team-mate and target-stand-off constraints are implemented in the underwater target tracking method proposed in Chapter 5. Furthermore, the proposed algorithm utilizes collaborative planning with look-ahead levels that indirectly influence the assets to maintain a reliable communication range while maintaining a safe stand-off.

The CNA vehicle and its role within a collaborative marine robot system has been discussed. A CNA-based multi-robot navigation solution could be used to address the three target types in this thesis —it is only considered directly in Chapter 3 (floating target with known pose). A review of research related to gaining awareness on a floating marine target (Chapter 3) is presented next.

2.4.2   Multi-Robot Systems to Survey Floating Targets

Obtaining awareness on a partially-submerged target requires sensing in two-domains. A single vehicle cannot complete an exhaustive survey of the target in both domains.

A collaborative system of heterogeneous robots, which must communicate, is proposed to address this. Examples of related work is in collaborative robot systems for data collection in communication-constrained and other challenging environments. Some examples are presented next.

To start, Faria *et al.* [25] considers the coordination of UAVs and UUVs for oceanographic observation. They report the difficulty encountered with underwater communication and limited observability of the autonomous UUVs during run-time. Challenges were encountered assessing the UUVs' activities and states at any one instance in time, so adaptive mission-planning was difficult. Ultimately, the solution in [25] required a some level of operator influence during deliberation processes. Their work provides an example where a CNA vehicle could have been used to establish more reliable communication and information sharing with the UUVs.

Shkurti *et al.* [26] used a collection of marine robots (UUV, USV and UAV) to assess the practicality of using them for repeat visual monitoring and inspection of underwater features like coral reefs. Their robot system was guided in real-time with remotely-located marine biologists. The level of human interaction for their system was greater than what is proposed in this thesis – where the proposed mission-planning autonomously manages the missions to collect sensor measurements on targets.

Manjanna *et al.* [27] examined a system with a USV that gathered measurements in select underwater regions based on transmitted cues from distributed low-cost drifting sensor nodes. Their strategic sampling approach was more energetically favourable than the alternative method to exhaustively survey an area. However, their work focused solely on finding the surface targets rather than surveying them and thus is of less relevance to this thesis.

Eckstein *et al.* [28] deployed heterogeneous UUVs to cooperatively map unstructured underwater environments. Their UUVs could adaptively reconfigure to sense in very different environments like the sea bed and sea cliffs. Their contribution was a mission-planner which coordinated the team's behaviours and could receive and distribute waypoints, mission files and updated information from the central controller. The concepts of their mission-planner inspired the centralized mission-planner developed in this thesis. The work in Chapter 3 is also built on previous research which

established a communications network between under, on and well-above-the-water marine robots [29].

Kwon, Park and Kim [30] created three-dimensional models of a floating target with a wide-beam imaging sonar. These models were experimentally verified in a controlled environment. More advanced is the previously introduced work of Han *et al.* [22] who created a three-dimensional reconstruction of a semi-submersible offshore platform with a USV. Their USV is equipped with lidar and sonar sensors to survey a floating structure above and below the waterline. The above and below water data is combined to obtain a volumetric point cloud of the structure. Their method is validated with field trials —surveying the offshore platform. The authors required an autonomous system for three-dimensional reconstruction towards equipment/structure maintenance in marine environments. Similarly, Papadopoulos *et al.* [31] provides an example of a USV solution and a novel algorithm to combine above and below water maps. Experimental results show positive results for three-dimensional reconstruction of a semi-submerged jetty. The authors stated that, while effort was dedicated to the three-dimensional surface reconstruction of structural systems, multi-domain surveying of marine structures has been insufficiently studied. Specifically referenced is the BP oil spill in the Gulf of Mexico [31].

In general the floating target three-dimensional reconstruction problem is insufficiently addressed in robotics research. Solutions have used dive teams and above-water sensors like cameras and lidar on manned aerial vehicles, which is a largely inefficient [31]. Presently, there are no examples found that utilize heterogeneous robot teams to accomplish the above and below water map-merging presented in Chapter 3. Research addressing underwater target search with uncertain target poses is reviewed next.

2.4.3 Underwater Target Search Research

For autonomous robotic target search, related work is in marine search and rescue (SAR) and NMCM. The purpose is to identify the challenges in searching for marine targets. A discussion on reinforcement learning-based path-planners, as a proposed solution, is also provided.

A first example of target search research is provided by Dudek *et al.* [32] who

address the search for multiple passive targets. Their work focused primarily on finding drifting debris, sensors, or lost divers at sea, and their methods depend on some information prior to a search. The effectiveness of their methods were quantified with metrics like distance travelled, failure rate, and computation time. Results were obtained through simulation. As a development, the underwater target search methods proposed in this thesis are not dependent upon prior information and are focused on finding targets in environments where erroneous and false detections are expected —not usually considered by others. The effectiveness of the methods in this thesis use energy consumption as a metric to compare search efficacy.

In another example, Waharte and Trigoni [33] compare three heuristic methods for target search (victims in search and rescue): greedy, potential-based and partially observable Markov decision process (POMDP)-based. They reported that the third method yielded the best results; however, in large environments the POMDP proved too computationally expensive for optimal solutions. The operational environment drives the navigation method selection through its size and level of clutter. In maritime environments, the size of the search or survey area can be large; however, in smaller environments like harbours and approaches clutter starts to be a consideration in the search. The underwater planning methods proposed in this thesis prioritize computationally light-weight and energy-efficient methods.

In practice, when the environment spans large distances and has no prior map (environment knowledge), widely-used paths like the common spiral and lawnmower (Boustrophedon) could maximize potential to find a target as these are exhaustive searches. Specifically, Bernardini *et al.* [34] propose a solution to the search and tracking problem that does not rely on the target state's prior. Their method mitigates the high computational requirement that plagues many solutions. Their proposed algorithm utilizes both a spiral-out and a lawnmower pattern to maximize the likelihood of finding a single target. Through their methods they obtained promising results both in simulation and experimentation for large 100 kilometre square environments [34]. The results from [34] are promising for a single target search in a large environment, but the method is less adaptable to a multiple target search where the targets' pose uncertainty is high.

Methods for marine target searches do not always address environments that are

difficult to sense [7]. In NMCM survey planning it is important to consider the possibilities of false and missed detections. Additional planning and effort applied to resolving uncertain detections can ensure that safe follow-on actions with the target can be pursued (e.g. mine disposal). Using NMCM as their motivational use case, Baylog and Wettergren [7], [35], and [36] offer a method to improve UUV search planning in the difficult underwater environment. First, they define a search channel model of the underwater environment, and then pose the search problem with a game-theoretic approach to maximize information flow when a UUV search is conducted over a region of discrete cells. The underwater search performance depends on the UUV sensor's (sonar) ability to detect the target in its environment. This ability is captured by modelling the sonar signal-to-noise ratio (SNR). Their work develops a method to maximize information payout, which is a function of the SNR, in each cell. An important distinction of their work is the information payout for repeat searches in the same region. Multiple search optimization is an important consideration when searching in areas where repeat surveys of the same region might be required to resolve inconsistent or false detections (the target re-acquire phase in NMCM). Simulation results are presented to compare a greedy search (incentivized by the information gain measures) versus a classic Boustophedron to optimize the cell visitation sequence. Their results show that the greedy method is a clear improvement over the Boustrophedon in terms of information gained given the same UUV search effort, which is defined as the number of cell visits during the mission. Their work was simulation-based and used an infinitely powerful and energetic agent with no consideration for the robotics capabilities or dynamics. The work of Baylog and Wettergen is extended in this thesis. A trajectory planner is developed with a reinforcement Q-Learning approach as an alternate way (as opposed to greedy approach) to maximize information gain in an NMCM mission. Chapter 5 takes the work a step further to consider predictive information measures and planning for a dynamic target. The development of the information measure and search game optimization is detailed thoroughly in Chapter 4.

As stated, the underwater target search method proposed in this thesis (Chapter 4) implements reinforcement Q-Learning for trajectory planning. Reinforcement Learning (RL) is a subcategory of Machine Learning (ML), which is generally used to

solve sequential decision-making problems. RL is independent of environment models and no prior knowledge of the environment is required for the learning process [37]. RL approaches allow a robot to develop a policy for decision making through instant feedback from its own decisions. Specifically, Q-Learning is a value-based reinforcement learning algorithm that develops a policy to inform an agent on actions to take as it navigates towards a goal state [38]. For this application, as Q-Learning learns, it can adapt to the dynamic underwater environment (bathymetry, acoustic propagation). Examples of reinforcement learning applied to UUV navigation in this environment will be described next.

Frost and Lane [39] studied Q-Learning for UUV path-planning of marine archaeological missions. They use a coarse resolution map of the environment to learn an optimal policy off-line before deploying the UUV to search the environment. The authors provide results from a study where they vary the parameters of the Q-Learning algorithm to determine how to decrease convergence time. In the present work, the convergence time for the on-line planning algorithm is minimal. This thesis will similarly perform a parametric study of the Q-Learning parameters to assess their impact on the planned path, given a dynamic environment, and search outcomes. The implementation of the Q-Learning (or any other learning method) must consider what can be implemented on an embedded platform like a UUV.

Lamini et $al.$ [38] present a novel collaborative Q-Learning based planner to solve the classic path-planning problem to find the shortest path-to-goal for multiple agents. Their results report that their system reduces computation time over the classic A* algorithm. The motivation for utilizing Q-Learning is apparent in the lightweight nature illustrated by the performance in [38] and influenced the decision to use the method in this thesis.

Gautam et $al.$ [40] apply Q-Learning to UUV navigation in benthic ocean zones for efficient planning under ocean currents. The purpose of the work is to minimize the time taken to travel between destinations. A cost function is proposed which incorporates UUV interaction with ocean currents. Simulation results are provided which conclude that Q-Learning was better with respect to computational complexity and ease of environment simulation in comparison to the Genetic Algorithm and Flower pollination algorithm. The work is of interest to the present thesis as efficient

planning is one of the main goals of the proposed search algorithm.

In summary, Q-Learning has been applied to robot path-planning in many other applications. Information gain incentive planning is also not a new concept. However, while the SNR (and resulting information gain) has always varied in an environment, to the best of the authors knowledge, these elements have not been incorporated into a Q-Learning-based mission-planner for NMCM underwater target searches.

2.4.4   Dynamic Target Localization and Tracking Research

The final focus of this review addresses mobile marine robots in target tracking and localization applications (Chapter 5 and thesis objective 3). Towards localizing vehicles underwater, the state-of-the-art acoustic localization techniques including LBL and CL have been presented. As stated, the limitations of LBL is in its fixed range and inability to optimize the sensor placements to better localize a mobile target over time. Collaborative localization relies on some level of communications between assets to bound the survey vehicles' error in their state estimate, which is not suitable for tracking an unresponsive target. Unresponsive targets could be tracked and localized using range-only methods with both surface and underwater vehicles.

One topic of research focus in marine mobile target tracking is range-only single-beacon (ROSB) localization where the single-beacon is the mobile autonomous tracker. The tracker periodically measures its slant range from the mobile target using signal time-of-flight [6]. Triangulation of the mobile target is achieved with many such slant range measurements by the autonomous tracker. ROSB methods are generally broken into two streams: (i) determine the autonomous tracker's optimal path to achieve sufficient target pose observability and (ii) develop state-estimation algorithms to solve the complex multi-modal probability density functions that describe the target's evolving state. The target state is described by a multi-modal distribution because, with only partial observability of the target position, the solution is not unique. The interest in ROSB methods is driven by reduced cost and perhaps more importantly, reduced complexity compared to multi-asset systems like LBL [6]. A ROSB tracking example is illustrated in Fig. 2.2 where the tracker is circling the target while collecting slant range measurements towards localizing it.

In practice, Masmitja *et al.* [41] studied the optimal path for a mobile robot
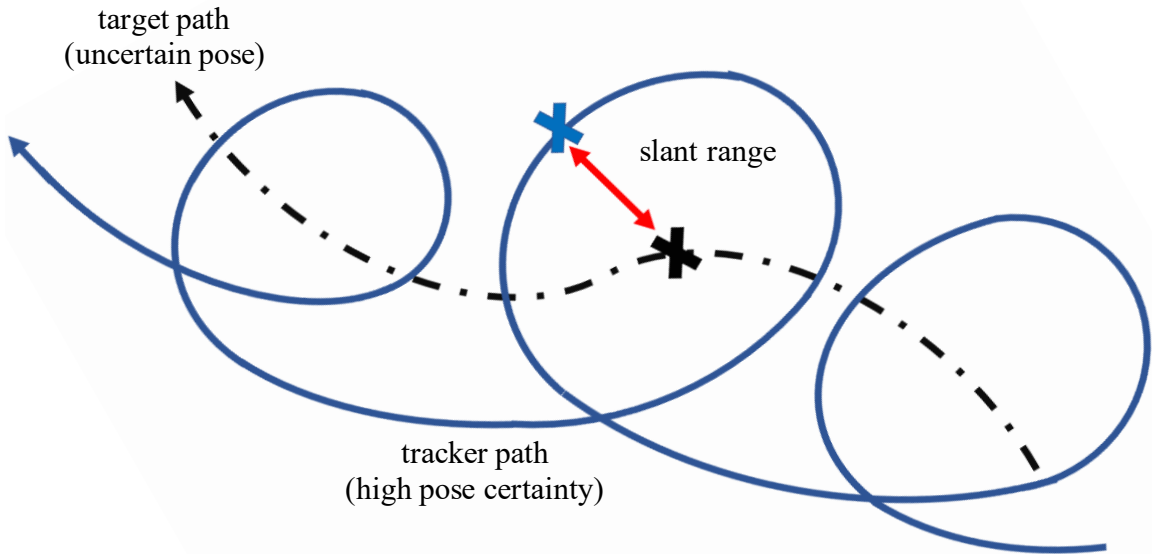
Figure 2.2: Illustrative ROSB tracking example of a tracker performing a circular path around the target to achieve observability of the target pose.

localizing a static target with the ROSB method. Not surprisingly, their simulation and in-water results with a wave-glider determined that a circular USV/target path was optimal. Crasta *et al.* [42] extended the work of [41] to localize a mobile target with an autonomous surface vehicle. Similarly, their optimal target trajectory was circular as it maximized the target pose observability. Subsequently, they developed a method to calculate the tracker's optimal trajectory radius given the relative target and tracker velocities. In practice, this solution is limited to targets that move quite slowly relative to the tracker. The reason is that the tracker must travel significantly further than the target to constantly circle the the target as it progresses on its path. Consequently, the tracker's power consumption and endurance become considerations. Another ROSB limitation is that it obviously requires that the initial target pose be known to the tracker.

Masmitja *et al.* [6] extended the previous work with a field study to evaluate the best filter for the tracker's estimate of the mobile target's state (pose) using ROSB. The tracker executes the circular trajectory used by Crasta *et al.* [42] at velocities of of 0 m/s (static) and 0.2 m/s. The results show that a particle filter provided the best state estimation performance compared to three other widely used state-estimation techniques: an extended Kalman filter (EKF), an unscented Kalman filter (UKF), and the maximum *aposteriori* estimation (MAP). A particle filter is used in

the target tracking method developed in this thesis; however, the work in this thesis aims to localize faster moving targets with larger initial target pose uncertainty.

An additional consideration for a USV tracker is obstacle avoidance. Agrawal and Dolan [43] identify the need for responsive autonomous planning (obstacle avoidance) to track surface targets in cluttered environments. In their research, Monte-Carlo (MC) simulations are used to predict possible paths for a target surface vessel in a discrete and limited search region with obstacles, based on the previous target path. An optimal path estimation algorithm determines the most likely target vessel position at each time step based on the MC simulation. Then, a USV path-planning algorithm based on an iterative A* heuristic method in 4-dimensional space is developed. The proposed algorithm looks to find the shortest path in time, while also maintaining a safe distance from the obstacles in the cluttered environment. The USV uses a lidar sensor to detect the target and obstacles. Their proposed obstacle avoidance behaviour follows the COLREGS rulebook [44] that specifies 38 rules which prevent collision between marine vehicles. Simulations and field results show that this proposed modified Monte-Carlo (M-MC) algorithm outperforms other methods like pure-pursuit path, constant bearing path, and the usual Monte-Carlo method. The research highlights the additional obstacle avoidance complexity of tracking in cluttered marine environments. Although the simulation and field trials are over a relatively small 400 m × 400 m environment, the M-MC method was found to be the most computationally expensive of the four algorithms considered. The environment considered for the tracking problem in this thesis is substantially larger. While not explored directly, the thesis' proposed dynamic tracking algorithm can implement basic obstacle avoidance by setting the planning reward function to zero in regions with known obstacles. Future work could consider testing in cluttered search environment.

An example of single-UUV (underwater) tracking is one that uses a reinforcement learning-based approach. Sun *et al.* [45] apply a modified Q-learning method to the single-UUV tracking problem. Their Q-Learning planner informs the UUV on a heading change to follow the exact path of the target (pure-pursuit path). Simulation results suggest that the proposed method is theoretically viable for such a task; however, the simulations assume instantaneous and reliable sensing of the target using

vision from an optical camera and side-scan sonar. This thesis considers the complexities of detecting a target acoustically, and also the case when the target has the potential to transit faster than the UUVs. Beyond single-beacon tracking methods, there are also examples of research with multi-UUV systems for underwater tracking.

More closely related to the work in Chapter 5, Salinas *et al.* [46] study multi-UUV systems for underwater target tracking in 3D space. Salinas *et al.* analytically determine the optimal tracker sensor placement. Specifically, the optimal tracker sensor placement is the best geometric configuration of a sensor network to maximize the range-related information available. A network of five or more beacons was considered. The authors assume an ideal range sensor with zero-mean additive white Gaussian noise; however, they do not consider range or environmental factors that can reduce the SNR and limit acoustic sensing and communication. Furthermore, the tracker paths to achieve the optimal tracker configurations are not considered (unlike the work in this thesis), such as their initial position, linear and angular speeds, and initial course angle.

In conclusion, target-tracking and underwater localization is receiving research attention. Methods that use both networks of stationary and mobile sensors have been introduced. However, to the best of the author's knowledge, a multi-UUV triangulation method that considers the limitations of acoustic communication and sensing has not been explored. Furthermore, adaptive planning that considers range and environmental effects on sonar SNR (and detection performance) has also not been incorporated into a mobile tracking system.

2.5   Summary

The goal of this review was to explore the challenges and current research in multi-robot collaboration. It covered relevant sensing, communication, localization and navigation – as these are the building blocks of multi-robot collaborative systems. In the author's case, the multi-robot problem includes complexities introduced by large marine environments with communication, sensing, and GPS-denied constraints.

The challenges of sensing and communicating in the marine environment was discussed. The concepts of collaborative localization and CNA navigation was presented

to illustrate the potential for multi-robot systems to overcome these difficult challenges. Current multi-robot systems were reviewed towards addressing the problem of surveying a floating target. A review was also conducted on navigation methods for underwater search and target tracking applications.

The next chapter presents the development of a multi-robot system that addresses the problem of situational awareness on a partially-submerged target. The objective is to present and validate a node for centralized autonomous mission-planning and task-allocation for a multi-robot system to address the first thesis objective.

Chapter 3

Collaboration of Heterogeneous Marine Robots Towards Multi-Domain Sensing on Partially-Submerged Targets

This chapter reports on a multi-robot system that collaboratively obtains above-water, surface and below-water information on a floating target (Fig. 3.1). This capability allows a ship to autonomously survey and obtain situational awareness on a floating unresponsive target from a safe stand-off prior to inspecting it more closely or navigating around it.
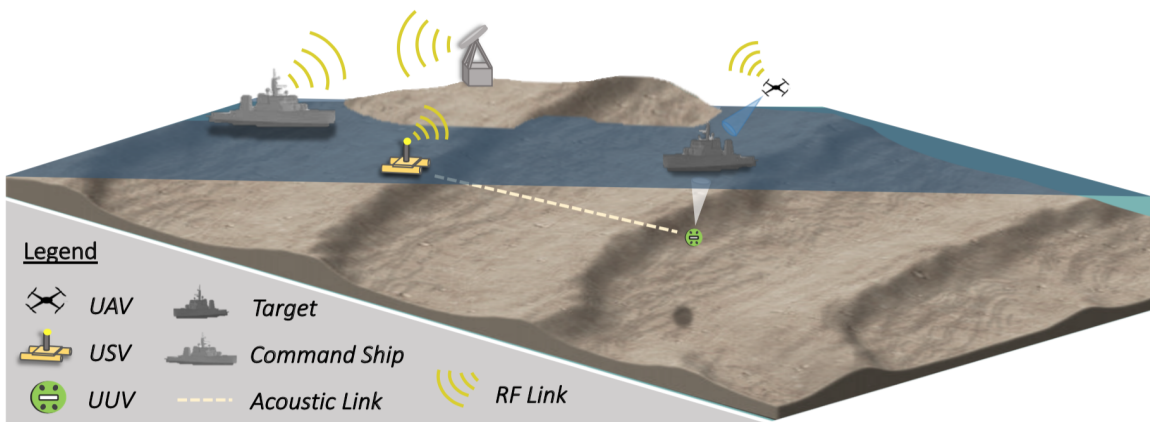
Figure 3.1: Depiction of multi-domain collaborative robotic survey on a partially-submerged target

The proposed solution is a collaborative system with an unmanned aerial vehicle (UAV), an unmanned underwater vehicle (UUV), and an unmanned surface vehicle (USV). The UAV captures visual imagery to create a three-dimensional model of the target's above-water geometry using photogrammetry. The UUV surveys the target's submerged hull with its integrated imaging and profiling bathymetric sonars. The USV hosts an intelligent mission-planning node which manages the robotic collaboration in a centralized architecture by autonomously planning and distributing the missions for the UUV and UAV. The intelligent node also plans an adaptive USV

trajectory to support the UUV and UAV. The resulting above- and below-water sensor data is fused at the water-plane to yield a three-dimensional representation of the floating unresponsive target.

The work presented in this chapter is an extension of the work published in [2]. The work presented here is focused on the present author's contribution of the *mRobot* node for managing robotic collaboration in multiple-domains, and the autonomous planning algorithms towards surveying a floating target. The simulation and experimental validation of the entire multi-robot system was a collaborative effort by the authors of [2]. The development and validation of the *mRobot* autonomous mission-planning node contributes to the foundation of the collaborative target tracking system presented later in Chapter 5.

The rest of this chapter is organized as follows. First, the *mRobot* node and the autonomous mission-planning algorithms are presented. Then the setup and results from simulations and in-water testing with a UUV, USV and UAV are described.

3.1 Collaborative Navigation and Multi-Robot Networking

The challenge in obtaining situational awareness on floating targets is that they simultaneously occupy the underwater and above-water domains. The proposed solution must have the ability to overcome the challenges associated with navigating, communicating, and collecting sensor data in both environments. For this scenario, the obstacles to overcome are as follows:

1. underwater localization over an extended range in GPS-denied environments

2. underwater acoustic communication limitations like bandwidth, latency, and drop-outs, and

3. to establish a constant, or at least reliable, communications network between all collaborating assets across the two (under and above-water) operating domains.

As described in Chapter 2, while acoustic methods are the best to transmit messages underwater, there are still limitations, which makes high bandwidth information sharing with a UUV at extended ranges often impractical. Furthermore, the UUV

INS-based state estimation error grows unbounded if no external information is available to aid in estimating its pose [47]. The collected underwater sensor data is of limited value if the UUV cannot accurately navigate to the target. The range of the system is, therefore, restricted by the communication range between the unmanned autonomous assets. In any collaborative network, the mission potential of the system deteriorates if any of the assets lose their ability to communicate and share information.

The proposed solution to network multi-domain robots is to designate the USV as a communications and navigation aid (CNA) vehicle, and to have it host the intelligent mission-planning node. A USV is ideal for the CNA role as it is located at the surface/underwater interface, which allows it to relay information between assets both above and below the water. The purpose of a CNA, which knows its pose to GPS accuracies, is to extend the operating range of the collaborative system from the command ship from which it deployed. The CNA also allows the UUV to remain underwater for the entire duration of the mission via collaborative localization [47].

3.2   Multi-Robot Autonomous Planning and Task Allocation (*mRobot*)

This section details the *mRobot* node and the autonomous mission planning algorithms which collects multi-domain sensor data on a floating target. Details of the simulation environment and preliminary results are presented at the end of this section.

The autonomous planning and task allocation tool, *mRobot*, is the ROS control node to manage multi-robot collaboration (Fig. 3.2). *mRobot* is responsible for networking and mission planning for the system of robots. The ability to network and plan a mission across two domains to sense and characterize the same object is the major contributions of the work in this chapter.

3.2.1   Robot Operating System (ROS)

The software developed to realize the new capabilities developed in this thesis, including the *mRobot* node, was coded in C++ to work in ROS. ROS is a middleware that manages robotic systems and includes software design and interfacing with hardware. It is a collection of tools and libraries designed to develop software specifically
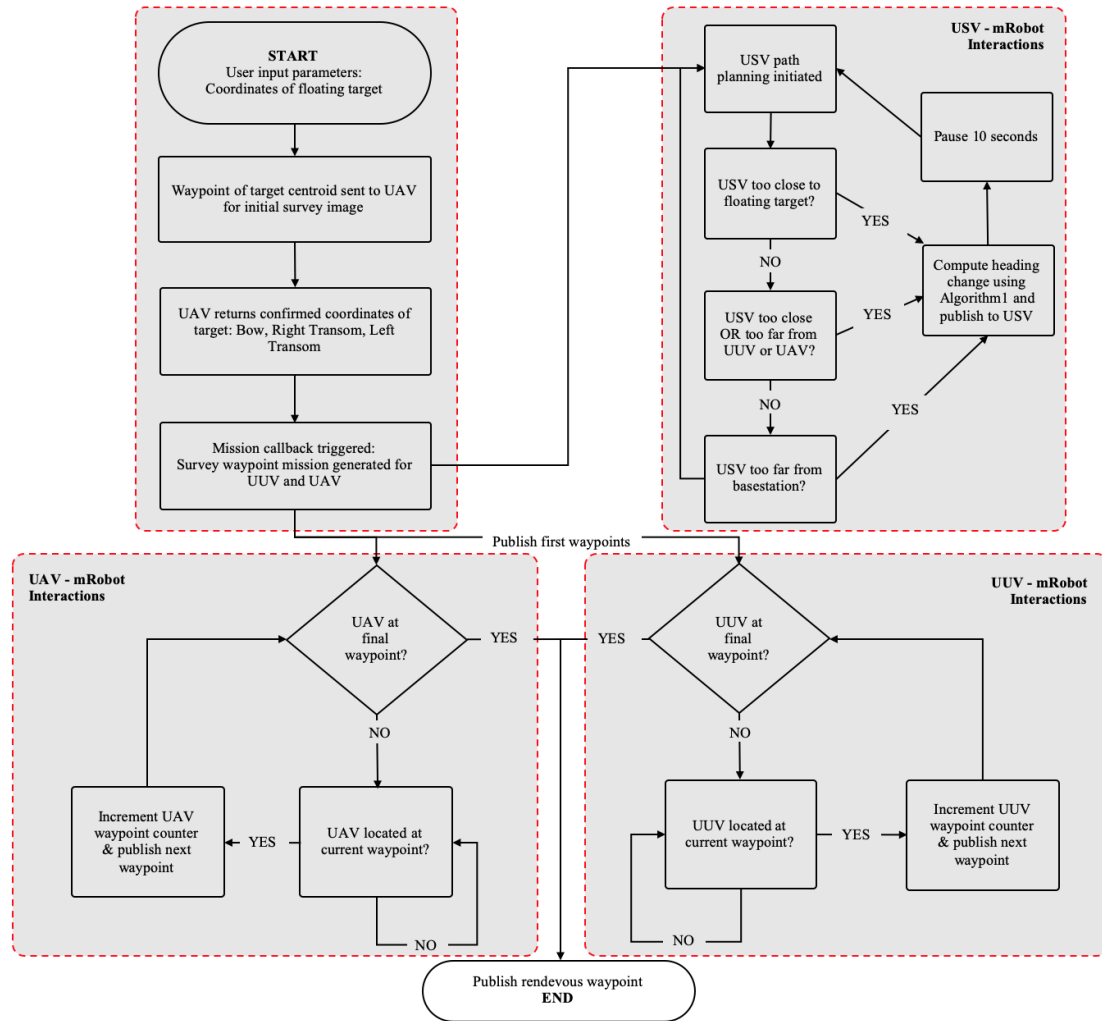
Figure 3.2: Schematic of *mRobot* components for the UAV, USV, and UUV.

for robot applications using a service-oriented architecture [48]. The author's work in simulations and interfacing with hardware for experimentation used ROS. There are several motivating factors that led to the use of ROS. The most critical are its modularity for code reuse and sharing, and its publish-and-subscribe architecture. The distributed and modular-nature of ROS means that it fosters a large community of contributors, third-party developers and forums that add considerable value on top of the standard ROS system. Contributions from community members has led to a collection of over 3000 packages that are available as open source resources [48]. These packages include everything from hardware drivers to high-level algorithms for autonomous systems. This level of resource interoperability makes ROS attractive for designing software at any level. The author has benefited from the use of ROS packages others have developed and will contribute packages developed in the thesis upon completion.

The ROS architecture facilitates near-seamless transitions from simulation to interfacing with hardware for experimental validation. The publish-and-subscribe architecture provides a framework for communication between robots, which is ideal for developing collaborative robotics systems where management of communication between assets (nodes) is critical.

Similar tools exist like MOOS-IvP which also has open source C++ modules for developing autonomous vehicle software. MOOS-IvP would also be a valid tool to have selected for the purposes of the author's thesis. The Intelligent Systems Laboratory currently has more efforts and resources for development using ROS for collaborative robotics.

3.2.2   UAV and UUV Path-Planning

mRobot follows the standard ROS publish-and-subscribe architecture, where messages between the three autonomous vehicles (UUV, UAV, and USV) are published through *mRobot*. It then uses *in situ* sensor measurements to continuously adapt to best support each robot team member to collect sensor data. Specifically, *mRobot* subscribes to the pose published by each robot and uses that information to manage the distribution of robot-specific tasks for the multi-domain survey mission. As shown in the top left block of Fig. 3.2, *mRobot* starts a mission by first publishing

a single waypoint to the UAV for an initial survey of a floating target. The purpose of the initial survey is to acquire the pose of the target in the form of three coordinate points describing the targets footprint and orientation. When the information from the initial survey is returned, a mission callback generates an array of waypoints for the UUV and UAV. As shown in the bottom left and right corners of Fig. 3.2, waypoints are published sequentially to the UUV and UAV every time the previous waypoint is acquired. *mRobot* determines that a waypoint is 'acquired' when the reported robot pose is within a pre-determined diameter of its respective waypoint location. At the end of the survey, *mRobot* publishes a final waypoint to gather the team at a pre-determined rendezvous point, and the mission is complete. At this point data transfer between each robot can be completed and then delivered to a base station (on the command ship) for post-processing.

The UAV and UUV survey missions are generated autonomously by *mRobot* based upon the size and orientation of the floating target and the requirements of the sensing equipment/techniques. Aerial photogrammetry requires a set of images where the target must be entirely within the camera field of view (FOV), and ultimately uses common features between images in the entire set to generate a point cloud. Therefore, the challenge in generating a mission plan for the UAV is to obtain enough images to capture the above-water target features. The challenge with underwater survey planning is to ensure complete (exhaustive) coverage of the wetted hull through multiple consecutive sonar scans. An example of the survey waypoints for the UAV and UUV in simulation is shown in Fig. 3.3, where the target would be at the center of the two paths.

The UAV path is layered in a descending rectangular spiral nominally around the target. The red triangles shown in Fig. 3.3 for the UAV path corresponds to waypoints where an image is taken. The algorithm to generate the waypoints was established empirically by assessing the photogrammetry reconstruction quality on objects both on land and in water, using different values of the path construction parameters. These parameters are the UAV altitude, number of layers in the spiral (different altitudes), number of waypoints per layer, and the length and width of the rectangular path. The UAV planning algorithm generates a unique path, defined by these parameters, for each mission given the camera FOV, the target's pose and
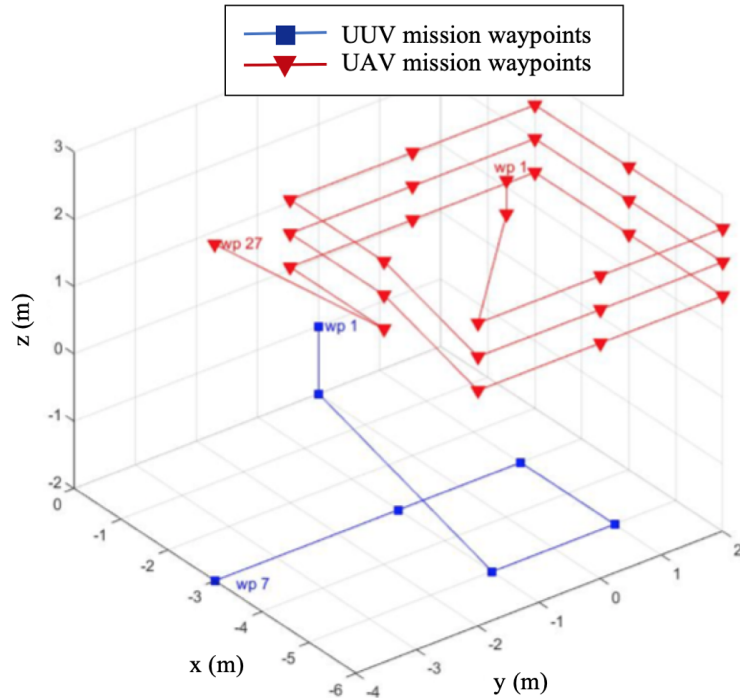
Figure 3.3: Sample waypoints for UUV and UAV missions in MATLAB simulations.

volume (length, width, and height as determined from an initial survey of the target) as inputs.

UUV path-planning is conducted using a similar empirically-developed algorithm. For the UUV, *mRobot* establishes a Boustrophedon survey path defined by the water plane footprint of the target, the sonar sensor swath width, and a sensor overlap percentage, which was tuned in simulation. Again, *mRobot* generates a unique path for each unique target footprint.

3.2.3 CNA Adaptive Path-Planning

An important aspect of this solution is that *mRobot* is hosted on the unmanned surface vehicle, which allows the collaborative system to autonomously and adaptively plan at an extended range from the command ship. The USV, acting as a CNA, requires that *mRobot* adaptively generates a path for the USV to best support the UUV and UAV. As shown in the top right corner of Fig. 3.2, *mRobot* uses pose information from the UUV and UAV to compute and publish a USV heading to minimize the UUV state-estimation error. The selected heading is constrained by communication

quality, which is based on the range to each vehicle and command ship, a safe stand-off distance from the target, and collision avoidance radii from other robots in the system. This aspect of *mRobot* was proven in simulation only. It was not a priority for the in-water testing given the size of the in-water test facility. Fig. 3.4 summarizes the adaptive CNA path-planning (Algorithm 1), which was modified from the works of [47] and [3] to include planning constraints related to the target and command ship. For every 10s-time step, Algorithm 1 cycles through possible heading options, which are defined by the maximum turn rate of the USV. The radial and tangential components of the UUV position error are used to calculate the cost of a heading choice in Eq. 5. Equation 6 calculates a sum if more than one UUV is deployed; however, this scenario is not explored in this work. Equations 7 through 10 are used to add distance penalties to the cost of a heading choice if it leads the CNA too close or too far from any asset (UUV, UAV, command ship station) in the system, or if it leads the CNA too close to the target. Equations 9 and 10 are the additional cost penalties for approaching too close to the target, or being too far from the command ship, which were added to the algorithm from [3] to better suit the application to this work. Ultimately, the most desirable heading option, with the least cost, is selected at each time step. Fig. 3.5 shows a top-down two-dimensional view of the robot paths for a MATLAB-simulated mission. The UAV and UUV paths are the same as shown previously in Fig. 3.3, while the blue path illustrates the adaptive path of the USV in its CNA role during the mission.

Algorithm 1 CNA Path Planning with Distance Penalities

Output: New target heading for CNA vehicle (δ_t^{CNA})

---

for i = 1, A, 1; increment through A possible δ_t^{CNA} options

$$
\begin{bmatrix} \psi_{t+1}^{CNA} \\ x_{t+1}^{CNA} \\ y \end{bmatrix} = \begin{bmatrix} \psi_t^{CNA} + \delta_t^{CNA} \\ x_t^{CNA} + \tau \cdot s^{CNA}\cos(\psi_{t+1}^{CNA}) \\ y_t^{CNA} + \tau \cdot s^{CNA}\sin(\psi_{t+1}^{CNA}) \end{bmatrix} \tag{1}
$$

    for j = 1, N, 1; increment through N survey UUVs and UAVs

$$
\theta_{t+1}^j = \angle (x_{t+1}^j - x_{t+1}^{CNA}) \tag{2}
$$

$$
\gamma_{t+1}^j = \theta_{t+1}^j - \theta_t^j \tag{3}
$$

$$
\varepsilon_{t+1}^{-j} = \sqrt{\frac{\left(\varepsilon_t^j \varepsilon_t^{-j}\right)^2}{\left(\varepsilon_t^j \cos\gamma_t^j\right)^2 + \left(\varepsilon_t^j \sin\gamma_t^j\right)^2} + \alpha\tau} \tag{4}
$$

$$
C(S_t, a_t)_j = \left(\varepsilon_{t+1}^j\right)^2 + \left(\varepsilon_{t+1}^{-j}\right)^2 \tag{5}
$$

    end for

$$
C(S_t, a_t) = \sum_j C(S_t, a_t)_j \tag{6}
$$

    if dist between CNA and any teammate < min_dist

$$
C(S_t, a_t) = 2\left(C(S_t, a_t) + 2(min\_dist - min(x_t^j - x_t^{CNA}))\right) \tag{7}
$$

    elseif dist between CNA and any teammate > max_dist

$$
C(S_t, a_t) = 2\left(C(S_t, a_t) + max(x_t^j - x_t^{CNA}) - comms\_range\right) \tag{8}
$$

    elseif dist between CNA and Target < min_standoff_dist

$$
C(S_t, a_t) = 2\left(C(S_t, a_t) + 2(min\_standoff - min(x_t^j - x_t^{CNA}))\right) \tag{9}
$$

    elseif dist between CNA and Command Ship < comms_range

$$
C(S_t, a_t) = 2(C(S_t, a_t) + max(ship\_x_t - x_t^{CNA}) - comms\_range) \tag{10}
$$

    end if

end for

Choose δ_t^{CNA} that returns lowest $C(S_t, a_t)$

Figure 3.4: Algorithm 1 for adaptive CNA path-planning in support of above- and below-water team-mates, with distance penalties.
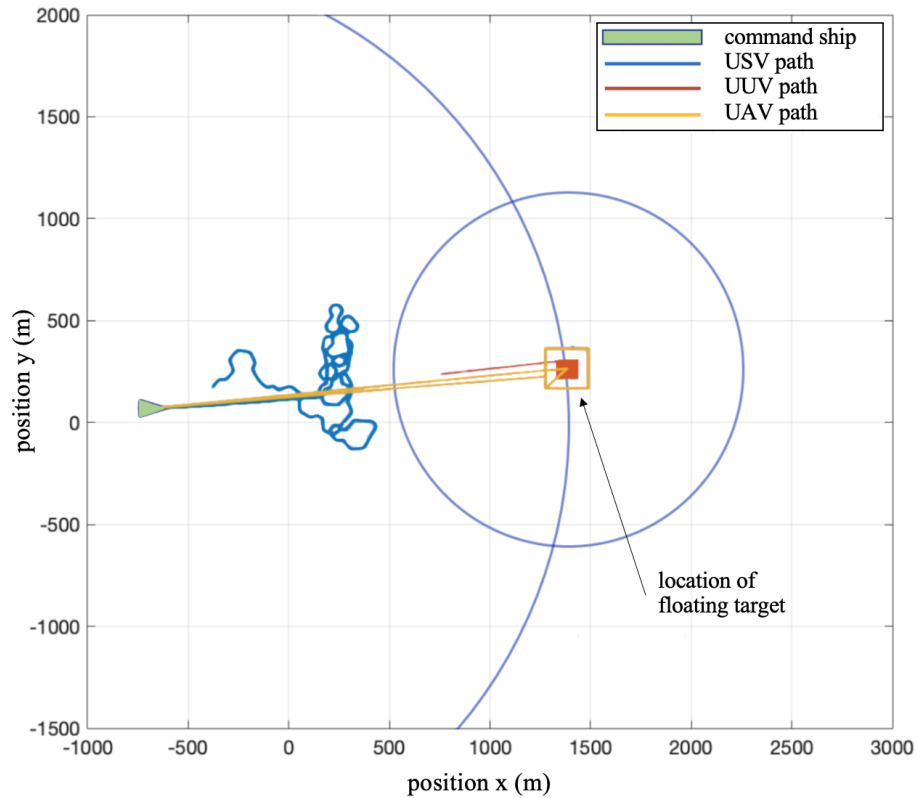
Figure 3.5: MATLAB simulation showing top down 2D view of the paths taken by the three autonomous vehicles during a survey mission. Circles in the image represent minimum and maximum distance constraints for the CNA during its adaptive planning.

3.2.4 Simulations

Computer simulations were utilized as a testbed to develop and validate the *mRobot* node, and to analyze concepts of operation for the multi-domain survey mission.

Simulations were performed with Gazebo [49], specifically within the underwater world environment from the UUV Simulator package [50]. The UUV was simulated with the IMOTUS in-house developed simulator with an upward looking bathymetric sonar, while the USV was simulated with the Clearpath Heron simulator [51]. Lastly, the UAV simulator was based on the Hector quad-rotor package with an optical camera [52]. The unresponsive targets surveyed during simulated missions were the Heron, a rectangular barge, and a large planning craft (Leeway Striker).

Computer simulations were carried out in both the open-ocean model environment, and in a model of Dalhousie University's Aquatron Pool tank (Fig. 3.6), where
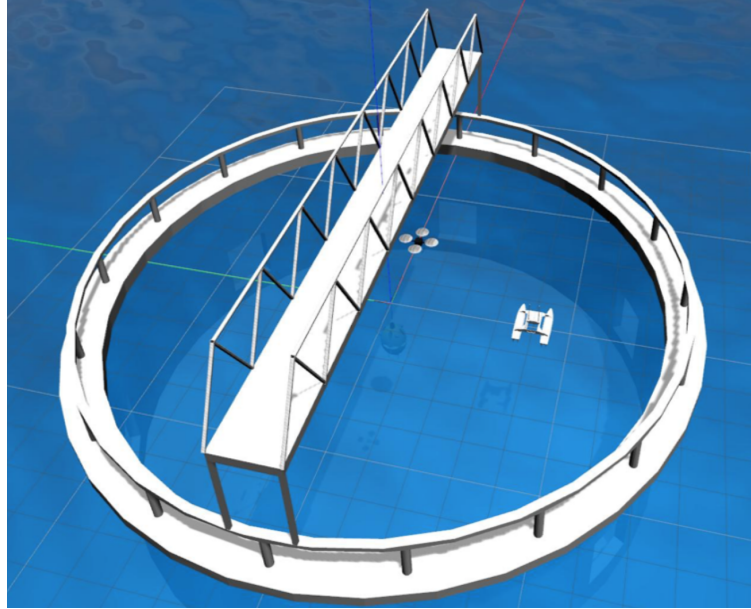
Figure 3.6: Dalhousie University Aquatron Pool Tank model used in simulation.

in-water testing was ultimately carried out. Simulations in both environments facilitated analysis of the system in both small- and large-scale environments.

The simulations were used to primarily test the path-planning for the marine robots. The major parameters considered in the planning for each robot are summarized in Table 3.1.

Table 3.1: Summary of multi-robot simulation parameters [2]

marine robot	simulation parameters varied
UUV	· survey distance and time · UUV path, position and depth · number of sonar passes · sonar range and radius
USV	· communication between UAV–UUV · detection distance from target · assistance to UUV localization
UAV	· survey distance and time · path, position and altitude · number and quality of images · camera specifications

For each simulated mission, the UUV sonar scans the underside of the target and above-water aerial images from the UAV were extracted and transformed into point clouds (Fig. 3.7). The targets and their respective sonar scans generated in simulation are provided in Fig. 3.8. Photogrammetry reconstructions were not successful in simulation as there was not enough unique texture on the targets within Gazebo to generate three-dimensional point clouds from the collected imagery. The photogrammetry was instead validated using a drone in a lab setting with smaller targets.

Gazebo provided an ideal environment to develop the proposed system, without the implementation complexity of including hardware-in-the-loop. However, the results from the simulation proved successful in providing preliminary validation of the multi-robot path-planning and task allocation node, and the overall order of operations for a multi-domain survey mission.
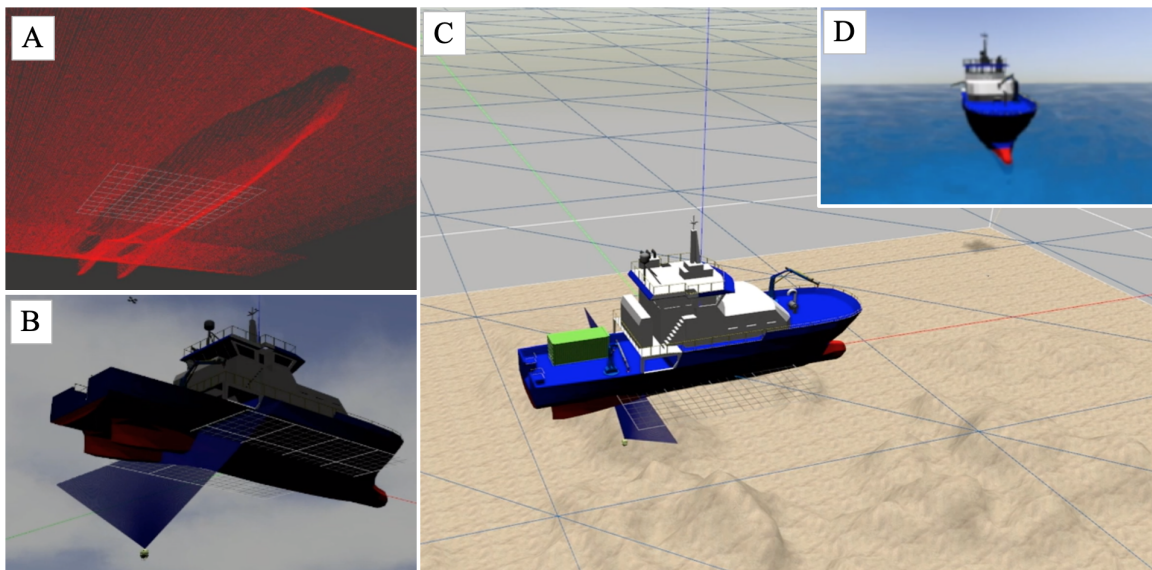
Figure 3.7: From simualtions (a) sonar output from IMOTUS UUV, (b) ship surveyed from below with the IMOTUS UUV integrated with a 256-beam sonar,(c) top isometric view of IMOTUS UUV surveying ship target, and (d) aerial imagery output from Pelican UAV [2].
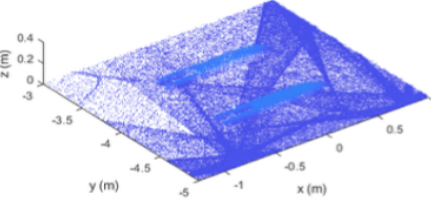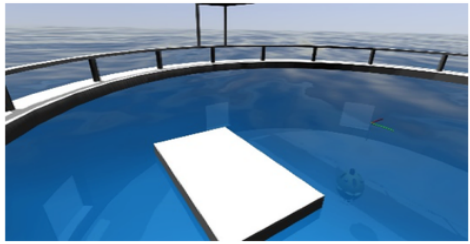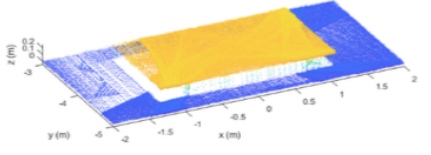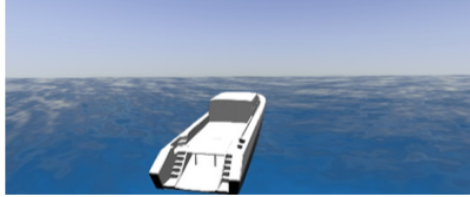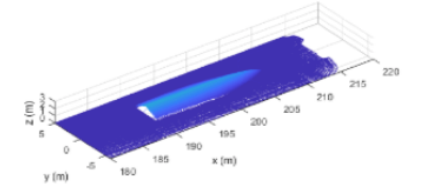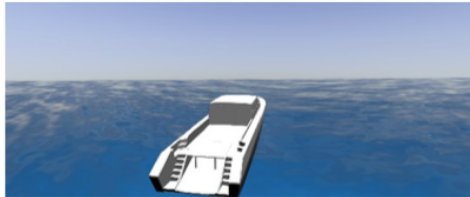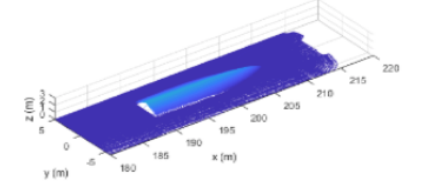
target	sonar
	target scan time: 3 min UAV mission target scan time: 9 min UUV mission
	target scan time: 3 min UAV mission target scan time: 9 min UUV mission
	target scan time: 6 min UAV mission target scan time: 22 min UUV mission
	target scan time: 10 min UAV mission Target scan time: 91 min UUV mission

Figure 3.8: Summary of data from simulations of surveys on a barge, USV, and large planning craft in both the simulated Aquatron and open ocean environments [2].

3.3 Experimental Setup

The novel approach presented in this chapter was implemented and tested in-water using a large indoor tank test facility with a UUV, USV and UAV. The objective of these tank tests was to assess the efficacy of acquiring and merging above- and

below-water imagery with real systems.

Above-water imagery of the target is captured with a regular visible wavelength video camera, which is merged with the below-water UUV bathymetric sonar imagery of the target. The merging of the above-water and below-water information is performed in post-mission processing. In the future, the USV would perform the merging autonomously. The algorithm developed by [2] for merging the imagery is not discussed here.

This section describes the robots used in experimentation, and the specifics of laboratory and in-water experimental setups. The individual robots are described next. All three marine robots in this research used the service-oriented ROS software framework for their middleware.

Unmanned underwater vehicle

A hover-capable UUV was desired as it is stable at very low, or hovering speeds. It requires an on-board sensor to insonify the target's wetted hull to identify hull features. The IMOTUS UUV (Fig. 3.9A) is a spherically-shaped robot that differentially drives thrusters to actuate all six degrees-of-freedom, to a top speed of 0.5 m/s. This hover-capable UUV was equipped with the Kongsberg M3 (500 kHz) and Flexview (1200 KHz) bathymetric sonars, which can insonify a target from a slant range of 150 m. For the in-water tests, these upward-looking sonars surveyed at a 4 m range from the hull. The IMOTUS navigation system uses an extended Kalman filter (EKF) for its pose estimation. The EKF performs sensor fusion of a pressure depth sensor, a Doppler velocity log (DVL) and a north-seeking fiber optic gyro (FOG), which provides the dead-reckoning used by IMOTUS for its underwater navigation solution [53].

Unmanned surface vehicle

The USV used in experimentation is a catamaran hull with 150 lb of reserve buoyancy, which was designed and constructed by the Intelligent Systems Laboratory (ISL) (Fig. 3.9B). It is propelled by a trolling motor that is mounted on the bow, which provides a top speed of 3 knots. The USV can be remotely controlled through a hand-held unit or autonomously through the *mRobot* ROS node. It is also equipped with an RF

radio that can relay messages to/from its underwater acoustic modem.

Unmanned aerial vehicle

The specific requirement of the UAV used in the experiments was to be splash-proof and positively buoyant in the event it went into the water. The ISL designed and built the prototype Pelican (Fig. 3.9C) UAV for the in-water testing phase of this project. The Pelican is a marinized quadrotor designed with a communication interface to allow *mRobot* to access the on-board position estimates and control the UAV. It was also necessary to have a marinized quadrotor that could tolerate the harsh elements of a marine environment (higher moisture, salt, winds, etc.). Its payload sensor is an optical camera mounted on a gimble. In autonomous mode, *mRobot* controls the UAV but the operator can override *mRobot*'s instructions with a remote controller. The UAV receives its mission and sends back imagery along a separate WIFI link.

Figure 3.9: (a) IMOTUS hover-capable unmanned underwater vehicle, (b) Intelligent Systems Laboratory's high buoyancy unmanned surface vehicle, approaching the camera, and (c) unmanned aerial vehicle, Pelican, designed to be marinized and networked through the motion capture system and controlled through *mRobot* [2].

3.3.1 Laboratory Testing

mRobot and the Pelican UAV were tested in the ISL flying space (3 m × 4 m × 4 m). The flight space is instrumented with a motion capture system [54] to provide ground truth localization to verify the UAV on-board estimates. The motion capture system fuses visual information from eight cameras, with the Motion Analysis (Rohnert Park, California) proprietary software, to track markers placed on robots in the space to determine their pose. The system provides pose accuracy to within fractions of a millimetre in translation and 1/10th of a degree in attitudes.

3.3.2 Aquatron Pool Tank Testing

The Aquatron Pool Tank at Dalhousie University (Fig. 3.10) was used to perform the integrated in-water tests with the UUV, USV and UAV. The tank is 15 m in diameter × 4 m in depth and filled with sea water. A GPS signal is not available within the tank facility, nor is it feasible to cleanly access magnetic north for the compass of either the UAV or USV. Consequently, the above mentioned motion capture system was installed around the Aquatron Pool Tank to provide pose information to the above-water robots.
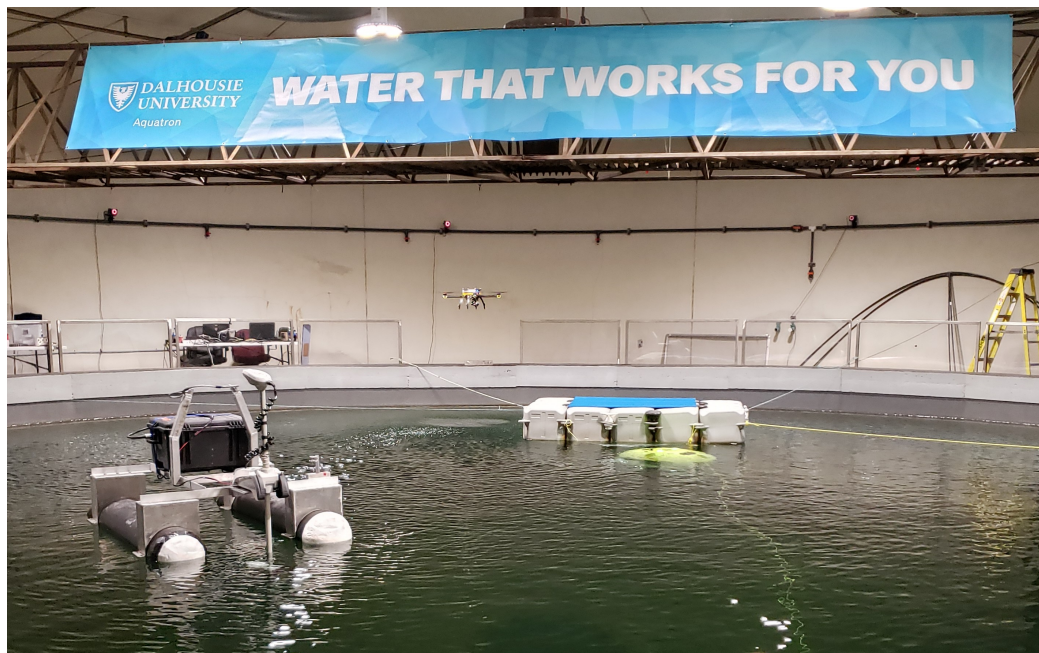
Figure 3.10: All three marine robots collaborating in the trial. The USV is left in the foreground. The surfaced UUV is right in the foreground (green surface with a tether). The barge is behind both. The UAV is left of the barge. On the wall, the red LED rings are three of the eight motion capture cameras installed at the Aquatron Pool Tank [2].

During the in-water validation experiments, *mRobot* would conduct the mission as described previously in Fig. 3.2. The result of the experimental procedure is above-water image data from the UAV and underwater sonar data from the UUV. The two sets of data are then merged together to form a complete three-dimensional model of the floating target.

Robot Networking

Given the relatively small Aquatron Pool Tank, WIFI was used between the UAV and *mRobot* for data transmission. IMOTUS communicated with *mRobot* through a tether integrated to its topside vehicle control computer. This setup facilitated the transmission of incoming sonar images for development purposes. Given the shallow depth of the tank and its concrete walls and floor, acoustic communications was not implemented. Underwater communications were emulated by sending the missions and messages using the user datagram protocol (UDP), common in underwater communications, through the tether.

Floating Targets

Two floating targets were used. The first was a barge of 2.4 m long × 1.4 m wide × 0.4 m deep (Fig. 3.11). The barge was an ideal first target to test as it exhibited identical geometry and texture above and below the water. The barge's symmetry provided a baseline to analyze and compare the data quality collected across the two domains, as the point clouds can be compared for the same symmetry and likeness. To make the barge more responsive to the 500 kHz and 1200 kHz sonars, part of the barge underside had a metal roofing panel attached to it. The panel has structural ridges that provided larger scale features then what was inherently on the barge. An identical roofing panel was sometimes placed on the top-side of the barge to test the UAV optical camera photogrammetry against the same surface. The second target was the unmanned surface vehicle (Fig. 3.12A) which is a twin-hull catamaran. Its expression at the water plane were the two catamaran pontoons, the trolling motor and the aluminum frame holding the pontoons together. The USV was an example of a more complex target to survey by comparison to the barge.

3.4 Results and Discussion

In-water experimentation allowed for real-time tests of the networking and collaboration of the three autonomous robots as a solution to collect multi-domain measurements on both a barge and catamaran hullform USV target.

From the experimental results, the missions generated by *mRobot* for the UUV

and UAV achieved their goals and collected data that was successfully merged, and at a high enough resolution to distinguish features, on several scales, on the barge and USV targets. The results are qualitative in nature, as the goal was to recreate visual representations of unresponsive targets. The visual representations are used to assess the target state and gain situation awareness.

The required resolution of the reconstruction is mission and environment dependent. The quality of the resulting three-dimensional reconstruction is limited by the resolution of the sonar (which the optical camera exceeds), which is dependent on its operating frequencies and range from the sonar to the target. Ultimately, the resolution of the three-dimensional target reconstruction is the size of the smallest feature that can be detected. Depending on the target being surveyed, the required resolution might be on the order of several centimetres to identify small objects attached to the wetted hull, or on the deck of a non-responsive ship. In the instance of surveying an iceberg, the required resolution might be on the order of a few meters, where only a general idea of the mass distribution is sufficient.

The results of surveying the first target, the barge, are presented in Fig. 3.11 showing (a) the target, (b) the insonification of the underside with the Kongsberg Flexview (1500 kHz) bathymetric sonar, and (c) an isometric view of the merged optical camera photogrammetry and sonar point clouds. Notice the sonar image can discriminate specific features of the barge, beyond its gross dimensions (e.g. the 15 cubes that make up the barge and their boundaries). A roofing panel was tied to the barge underside to give it different features (and better response at higher sonar insonification frequencies). The stiffeners in the roofing panel, which are 5 cm wide, can be clearly seen. The topside of the barge is captured well with the photogrammetry-based point cloud where texture on the topside of the barge is visible, which clearly surpasses the resolution of the sonar imaging at 1200 kHz. The process of merging the above-and-below-water data was easier with the 1200 kHz sonar that produced higher resolution images (≈ 1 mm). The photogrammetry resolution was good and exceeded that achieved by the sonar's. As shown in Fig. 3.11, dimensions can be extracted for features of interest on the top or bottom side of the floating target.
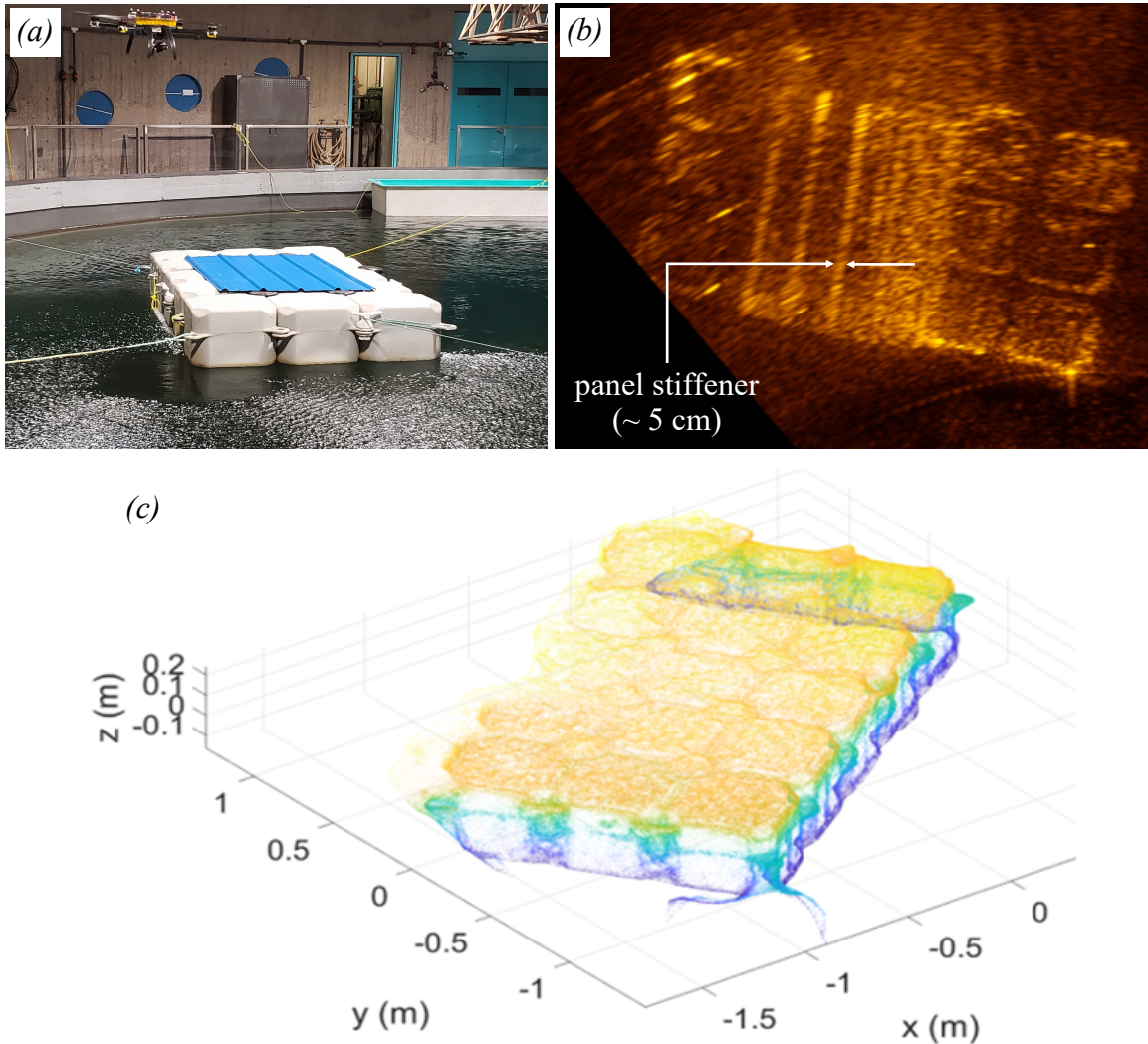
Figure 3.11: (a) UAV underway surveying barge target from the air, (b) Flexview sonar imaging of the barge underside (underwater) from the IMOTUS UUV (notice that the roofing panel stiffeners running top to bottom are captured), and (c) optical camera photogrammetry reconstruction merged with bottom-side sonar (isometric view) [2].

The results of the multi-domain USV survey (Fig. 3.12) show an example of the details that can be captured by the proposed robotic system. The sonar image reveals the submerged portion (wetted hull) of the two pontoons, trolling motor, and even the brace members that fasten the pontoons to the hull of the USV – features that are 5 cm and smaller. The features and texture on the topside of the USV are also captured well. The 4 cm diameter circular shaft extrusion (from the trolling motor) on the topside of the USV was captured also. The teal streaks in the photogrammetry

model that run at a 45° angle are due to the photogrammetry capturing features from rope segments anchoring the USV.

The UAV photogrammetry images were easily transmitted with in-air radio to *mRobot*. However, the extracted sonar images and resulting point clouds could not have easily been transmitted through underwater acoustic modems - especially at greater ranges between UUV and USV. Image compression and encoding techniques will help overcome this challenge [55].
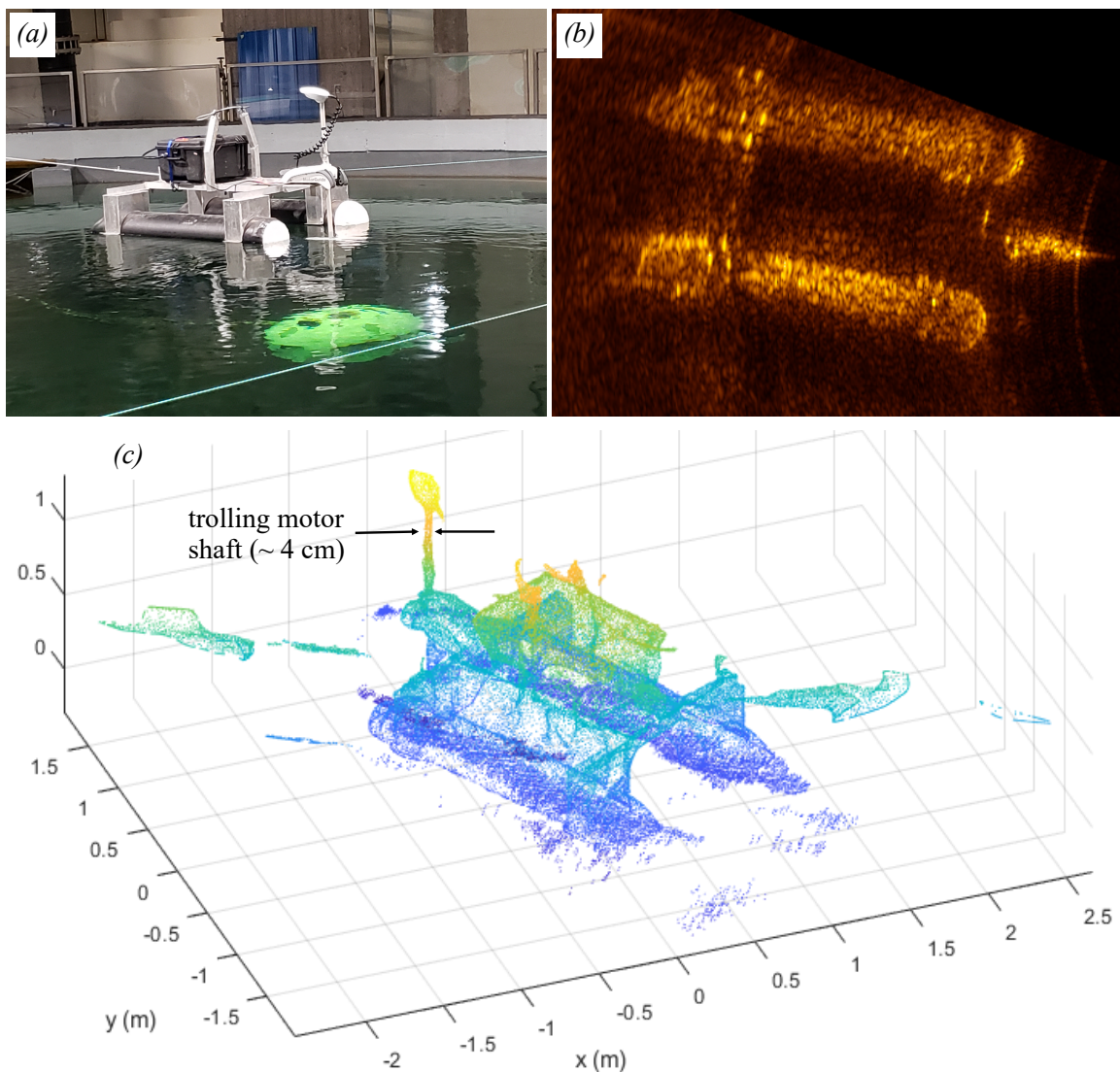
Figure 3.12: (a) USV target, (b) Flexview sonar imaging of the USV underside (underwater) from the IMOTUS UUV, and (c) optical camera photogrammetry reconstruction merged with bottom-side sonar (isometric view).

3.5 Summary

Simulations and controlled in-water experimentation provided preliminary validation of the planning algorithms for a UUV, USV and UAV working collaboratively to survey a floating target. *mRobot* is a stand-alone autonomous mission-planner that creates missions for multi-domain robots towards a common goal. The aforementioned components have been developed, tested and validated.

The results showed that the proposed solution can provide three-dimensional reconstructions with a resolution of 5 cm (and smaller), which was shown from experimentation with a barge and USV target. The reported resolution is adequate considering that the goal of this work is to provide the user with information on the state of the target.

Future work for the system presented in this chapter will take the heterogeneous collaborating robots into a larger in-water environment to test the underwater acoustic communications as well as study how to represent the in-situ sonar information so that it can be transmitted to the USV. A larger environment will also allow for experimental validation of USV path-planning in support of its team-mates, as well as quantifying the benefit, in-terms of reconstruction quality, of its role in the system. With regards to the merging of above- and below-water models, more complex and larger targets will be trialed.

In the next chapter, the *mRobot* mission planner is applied towards gaining information on marine targets with uncertain poses. A dynamic planning method is developed for underwater search missions with the goal to locate underwater static targets of uncertain poses.

Chapter 4

Autonomous Underwater Search for Static Targets

The objective of a UUV in an exploration mission is to maximize its information gathering about the environment it explores. For example, in a mine countermeasures (MCM) survey the goal is to detect and localize targets, whose states are not known prior to the mission (Fig. 4.1). The desired outcome of the search mission is an occupancy grid map of the search area indicating the location of all targets present as illustrated in Fig. 4.2.
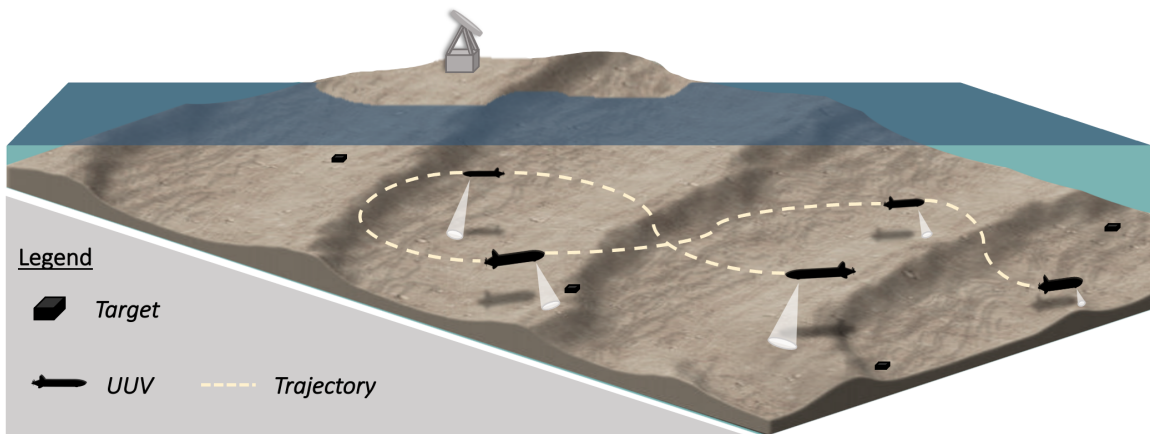
Figure 4.1: Depiction of underwater search for hidden targets (NMCM).

The work presented in this chapter addresses the efficacy of reinforcement Q-Learning-based planning to localize targets in an area where the UUV's sonar signal-to-noise-ratio (SNR) may vary, and consequently, its probability of detection and likelihood of false alarm.

The contribution of the work presented in this chapter is an adaptive mission / path-planner which addresses the challenges of autonomous NMCM surveys in the unstructured underwater environment where erroneous detections are expected, and energy-efficient missions and timely planning is desirable. Specifically, the contributions of this work are as follows:
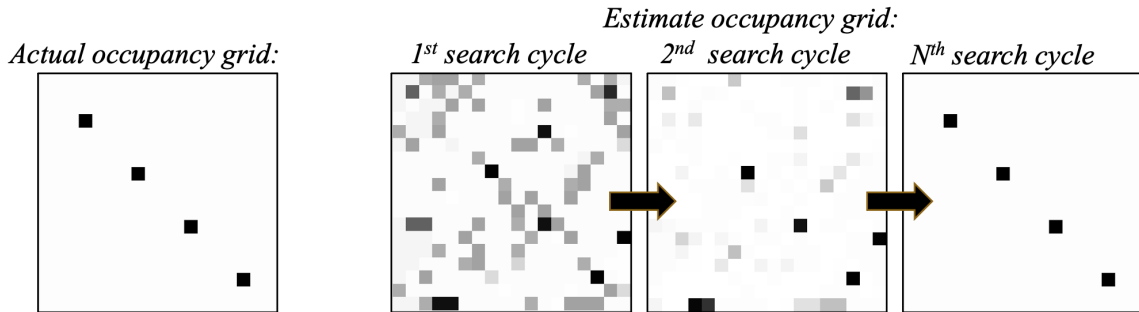
Figure 4.2: Example MCM search mission showing estimate occupancy grid (18 × 18 cells) development over the duration of a mission with multiple search cycles (replans).

1. Developments to the work of [7], [35], and [36], by applying Q-Learning to optimize the search cell visitation sequence. To quantify the merit of the proposed solution, its results are compared with the standard greedy and boustrophedon approaches.

2. A sensitivity study of the proposed algorithm which identifies the parameters that drive the proposed planner's performance. With a machine learning approach, parameter sensitivity is not always apparent. The sensitivity study results are summarized visually to illustrate the parameters' impact on the path-planning policy.

3. Simulation and two-dimensional experimental validation of the proposed algorithm.

The rest of this chapter is organized as follows. The first section details the work from [7], [35] and [36] which developed a method to measure and maximize mutual information gain in uncertain search environments with varying probabilities of detection and false alarm. Next, the proposed autonomous mission-planning algorithm is developed which combines information theoretic rewards with Q-Learning optimization. Then a brief description of the simulation development and experimental validation environments follows. Finally, a sensitivity study is detailed, which is followed by NMCM results from four different search environments which compare Q-Learning to two classic UUV search planning methods (greedy, and nominal/boustrophedon).

4.1 Optimized Mutual Information Gain Measure for UUV Search Planning

The purpose of this section is to define the search channel, search game, and the measure of information gained from UUV searcher agents in a target search mission. The work is this section is condensed and summarized from [7], [35], and [36].

4.1.1 Search Game and Cell Search Channel

The information flow while a UUV explores its underwater search area is not dissimilar to a communication channel and the message transfer between a transmitter and receiver [7]. The message to communicate is the location of the targets —which are initially unknown. Although, there may be prior knowledge on the likelihood of target positions, either relative to each other, or within the search area itself. The search channel definition is based on Shannon's model of the communications channel [56]. Firstly, the channel presumes a search area which is partitioned into L independent cells (or subregions as in Fig. 4.2). Targets are placed in a subset of the cells. The intended message, $\mathbf{M} = \langle M_1, M_2, ..., M_l \rangle$, is the binary occupancy status of the set of cells, and the UUV search is conducted to decode the intended message as the estimate, $\hat{M}$. The search plan is then defined as the cell visitation sequence executed by the searcher. Replanning during the search based upon the cell detection outcome sequence $D_l = \langle d_1, d_2, ..., d_l \rangle$, and the aggregate detection outcome sequence $D = \langle D_1, D_2, ..., D_l \rangle$ make up the decoding framework of the channel. The binary cell search channel, and the search channel information flow are shown in Fig. 4.3, respectively.
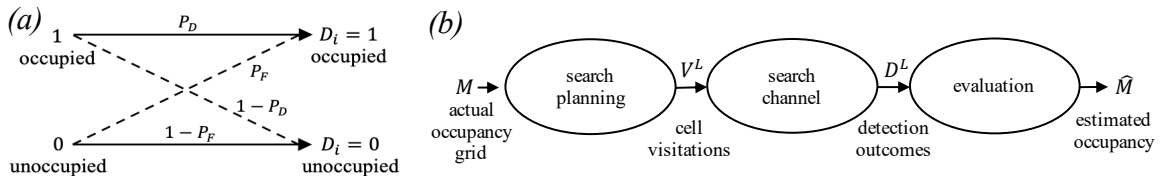

Figure 4.3: (a) binary cell search channel and (b) the search channel flow from [7].

The execution of a search plan is expressed mathematically as a sequential game problem. In the context of search theory, a target search mission is a two-player game

such that player one is the searcher, and player two, the target of the search, wants to stay hidden [36]. A search game assumes independent target placement across the set of cells, and it is assumed a cell is occupied if there are any targets within its boundaries, and if not, it is unoccupied. The status of a cell is described as:

$$M_l = \begin{cases} 1, & \exists x | x \in C_l \\ 0, & otherwise. \end{cases} \tag{4.1}$$

where C_l is a cell in the discrete search region.

To quantify the amount of information passed through the channel during the search, the probability space is first defined. The valid detection of a target within cell l occurs with the probability P_{D_l}, and false detections occurs with likelihood P_{F_l}. These search channel properties are determined from the receiver operator characteristic (ROC) analysis commonly applied to underwater sonar applications [57]. The relationship between the P_{D_l} and P_{F_l} is from the sensor's ROC curve. As in [7], the work here employs a theoretical ROC analysis that employs a Gauss-Gauss channel model for energy detection. In this model the null hypothesis $\mathcal{H}_{0_l}$ maps to a signal model where the energy detected by the sonar is noise alone (no target present). The alternative hypothesis $\mathcal{H}_{1_l}$ rejects the null hypothesis if the detected energy level maps to a model of acoustic signal and noise, and the received signal exceeds the prescribed detection threshold. The final decision is made by applying the true detection and false alarm likelihoods in the region (or cell) to determine whether the signal suggests a target is present or if the signal is noise-only. The signal and noise within each cell are modelled as zero-mean white Gaussian processes.

Of specific interest to the work here is the impact of variable probabilities of detection and false alarm on search performance due to the environment [7]. When detecting targets on the ocean floor the topography and texture can change significantly, as perceived at sonar frequencies, throughout the search area, which in turn affects the search performance. As previously identified, other factors like multipath and time and spatially varying sound velocities can impact search performance. From cell-to-cell, these variable environmental conditions are captured in the acoustic signal-to-noise-ratio (SNR_l). Towards calculating the probability of false alarm within a cell, let $\breve{x}$ be the position where the false alarm occurs, and let γ_l be the detection threshold in the log likelihood ratio space. Following the work of [58], the

probability of false alarm is then calculated over the cell by:

$$P_{F_l}(\gamma_l) = \int_{C_l} \mathbf{Pr}(d = 1|\breve{x}; \gamma_l) f(\breve{x}|x \notin C_l) d\breve{x}$$
$$\approx Q_{\chi_N^2}(\grave{\gamma}_l) \tag{4.2}$$

where $\grave{\gamma}_l = \gamma_l/\sigma_l$ is the normalized detection threshold, and

$$Q_{\chi_\nu^2}(x) = \int_x^\infty \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} t^{\frac{\nu}{2}-1} exp(-\frac{1}{2}t) dt \tag{4.3}$$

is the integral of the chi-squared distribution with ν degrees-of-freedom (DOF). In this target search game, the binary search channel has $\nu = 4$ DOF (or possible outcomes) in a detection event: true positive, true negative, false positive, and false negative. The probability of valid detection within a cell can be calculated by:

$$P_{D_l}(\gamma_l) = \int_{C_l} \mathbf{Pr}(d|x; \gamma_l) f(x|x \in C_l) dx$$
$$= Q_{\chi_N^2}\left(\frac{Q_{\chi_N^2}^{-1}(P_{F_l})}{SNR_l + 1}\right). \tag{4.4}$$

Equation 4.4 is formulated for a ROC curve analysis without directly solving for the detection threshold [7]. Figure 4.4 provides an example of four ROC curves from regions of increasing SNR. Potential operating points are highlighted Fig. 4.4. The information flow through the channel can be maximized by careful selection of $P_{D_l}(\gamma_l)$ and $P_{F_l}(\gamma_l)$, [36].

4.1.2 Development of the Information Payout

From information theory, the mutual information gain is defined by the reduction of Shannon entropy, or a reduction of uncertainty, about one random variable while observing another random variable [59]. In this target search application the reduction of entropy is the reduction of uncertainty about a cell's occupancy after visiting and observing the region. Entropy and information are measured in units of Shannons or bits.

If the search is a single pass over a cell with a detection outcome, d_1, the mutual information gain can be calculated by:
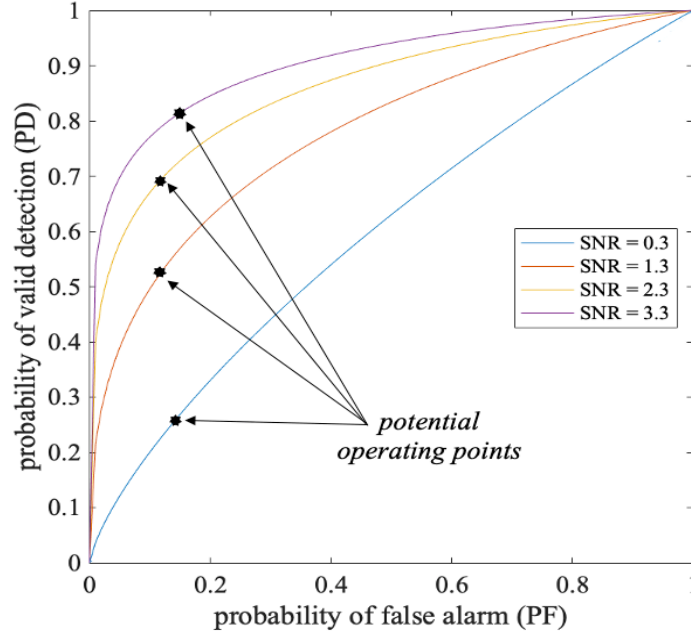
Figure 4.4: ROC curves for four increasing levels of SNR. The operating point determines the properties of the cell through selection of the relationship between P_{D_l} and P_{F_l}.

$$I(M_l; d_1) = H(M_l) - H(M_l|d_1) \tag{4.5}$$

such that $H(M_l)$ and $H(Ml|d1)$ are, respectively, the the prior and the posterior entropy of the cell's occupancy. The prior uncertainty of the cell's occupancy, or the entropy of the message, is given by:

$$H(M_l) = [P_{M_l} \log_2(P_{M_l}) + P^c_{M_l} \log_2(P^c_{M_l})] \tag{4.6}$$

where P_{M_l} is the prior probability that the cell is occupied, and $P^c_{M_l}$ is the complementary probability that the cell is not occupied. Note that log_2 is used as it follows Shannon's convention for measuring information [56] in the binary channel. This convention leads to a normalization where $I(M_l; D_l) \in [0, 1]$.

In a similar form, the message posterior entropy is calculated with:

$$H(M_l|d_1) = P_{d_1} H(M_l|d_1 = 1) + (1 - P_{d_1})H(M_l|d_1 = 0) \tag{4.7}$$

where P_{d_1} is the marginal probability of detection, which is given by:

$$P_{d_1} = P_d P_{M_l} + P_f P_{M_l}^c. \tag{4.8}$$

Given a positive detection event, the posterior probability that the cell is occupied is given as:

$$P_{M_l|d_1=1} = \frac{P_D P_{M_l}}{P_{d_1}} \tag{4.9}$$

and alternatively, if the detection event is negative, then the posterior probability of occupancy is given by:

$$P_{M_l|d_1=0} = \frac{(1 - P_D)P_{M_l}}{1 - P_{d_1}}. \tag{4.10}$$

The two posterior entropy components can be computed by:

$$H(M_l|d_1 = j) = -P_{M_l|d_{1=j}} \log_2(P_{M_l|d_{1=j}}) - (1 - P_{M_l|d_{1=j}}) \log_2(1 - P_{M_l|d_{1=j}}) \tag{4.11}$$

where $j = 1$ for the detection event and $j = 0$ for a non-detection event.

When considering the entire set of cells in a search environment, the information collected upon completion of a search takes the form:

$$I(\mathbf{M}; \mathbf{D}) = H(\mathbf{M}) - H(\mathbf{M}|\mathbf{D}) \tag{4.12}$$

where $H(\mathbf{M})$ is now the total uncertainty of the source message (occupancy grid) and $H(\mathbf{M}|\mathbf{D})$ is the uncertainty of the message given the detection outcomes of the executed search plan. From the assumption of independent cells, Eq. 4.12 can written in the following form:

$$I(\mathbf{M}; \mathbf{D}) = \sum_{l=1}^{L} [H(M_l) - H(M_l|D_l(n_l))]. \tag{4.13}$$

The message component posterior entropy of Eq. 4.13 is calculated over the set of all detection events that can occur in n-passes (or n-searches) of the the cell, by:

$$H(M_l|D_l) = -\sum_{\{D_l\}} \mathbf{Pr}(D_l) \cdot \left[\sum_{\{M_l\}} P_{M_l|D_l} \log_2(P_{M_l|D_l}) \right] \tag{4.14}$$

where $P_{M_l|D_l}$ is the cell's posterior occupancy probability. $P_{M_l|D_l}$ can be calculated using Bayes' rule:

$$P(M_l|D_l) = \frac{\mathbf{Pr}(D_l, x \in C_l)}{\mathbf{Pr}(D_l, x \in C_l) + \mathbf{Pr}(D_l, x \notin C_l)}. \tag{4.15}$$

Assuming that detection events are independent, the probability of a sequence of detection events given that the cell is occupied can be calculated as follows:

$$\mathbf{Pr}(D_l, x \in C_l) = \left[\prod_{i=1}^{n_l} \mathbf{Pr}(d_i, x \in C_l) \right] P_{M_l} \tag{4.16}$$

where the probability $\mathbf{Pr}(d_i, x \in C_l)$ is P_{D_i} in the event of a detection ($d_i = 1$), and by $1 - P_{D_i}$ in the event of a non-detection ($d_i = 0$). If the cell is not occupied, then the probability of a given set of n detections in a cell follows:

$$\mathbf{Pr}(D_l, x \notin C_l) = \left[\prod_{i=1}^{n_l} \mathbf{Pr}(d_i, x \notin C_l) \right] P_{M_l}^c \tag{4.17}$$

where $\mathbf{Pr}(d_i, x \notin C_l)$ is given by P_{F_i} if $d_i = 1$ and by $1 - P_{D_i}$ if $d_i = 0$.

Wait Time Information Measures

The complexity of the calculations in Eqs. 4.16 and 4.17 is determining the marginal probabilities of every different possible detection outcome sequence in an n-pass search of a cell. If the probabilities of detection and false alarm within the cell are held constant for each visit to the cell, then the information obtained during that search can be computed as a function of wait time probabilities [36]. Wait time event probabilities indicate the number, n, of passes (effort) that must occur before one can expect detection events to happen. In other words, n denotes the number of search passes that must be attempted to achieve r detections. For independent trials (or cell searches), the rth detection event happens during the nth search pass ($W_r = n$) with the wait time probability given by the negative binomial distribution [60]:

$$\mathbf{Pr}(W_r = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r} \tag{4.18}$$

where p is a fixed probability of occurrence P_{d_1}. Wait time probabilities are valuable since in an n-pass search of a cell there are 2^n different detection sequences possible, while there are only $2n$ wait time outcomes. Therefore, using wait time measures greatly reduce computation effort when determining information measures.

However, selecting occurrence probabilities (P_D and P_F) that remain fixed over the duration of a mission is a simplification that does not allow for a full optimization of the information measure. In the more complex scenario where detection and false alarm probabilities are allowed to vary over each search pass, wait time probabilities cannot be calculated using Eq. 4.18. Instead, Baylog and Wettergren [35] provide a wait time probability recursion developed from the Bernoulli trial event associated with each search pass ($i = 1 : n$). Firstly, let $P_{\delta_l}(n, r) = \mathbf{Pr}(W_r = n)$. Then, for the first detection, $r = 1$,

$$P_\delta(n, r) = p(n)[1 - P_\Sigma(n - 1, r)] \qquad (4.19)$$

and for each subsequent occurrence, $r > 1$,

$$P_\delta(n, r) = p(n)[P_\Sigma(n - 1, r - 1) - P_\Sigma(n - 1, r)]. \qquad (4.20)$$

The above recursions are again conditioned on the occupancy of the cell such that $p(n) = P_{D_l}$ when occupied and $p(n) = P_{F_l}$ when unoccupied.

The cumulative probability of detection (CPD) $P_{\Sigma_l}(n, r)$ denotes the marginal wait time probability of making r detections during n-searches of a cell with:

$$P_{\Sigma_l}(n, r) = P_{\Sigma_l}^d(n, r)P_{M_l} + P_{\Sigma_l}^f(n, r)P_{M_l}^c \qquad (4.21)$$

where $P_\Sigma^d(n, r)$ is the probability of making r valid detections (true positives) with:

$$P_\Sigma^d(n, r) = \mathbf{Pr}(W_r \leq n | x \in C_l) = \sum_{i=1}^{n} P_{\delta_l}^d(i, r) \qquad (4.22)$$

and $P_\Sigma^f(n, r)$ is the probability of making r invalid detections (false positives) during the same n-pass search with:

$$P_\Sigma^f(n, r) = \mathbf{Pr}(W_r \leq n | x \notin C_l) = \sum_{i=1}^{n} P_{\delta_l}^f(i, r). \qquad (4.23)$$

Finally, with the $n \times r$ set of wait time probabilities $P_\Sigma^d(n, r)_{r=1}^{n}$ and $P_\Sigma^f(n, r)_{r=1}^{n}$, the channel information can be calculated as:

$$I(M_l; D_l) = I(M_l; W_{1_l}) + \sum_{r=2}^{n} \mathbf{Pr}(W_{r_l-1} \leq n)I(M_l; W_{r_l}|W_{r_l-1} \leq n) \qquad (4.24)$$

where the first term $I(M_l; W_{1_l})$ is found with Eq. 4.5. The prior probability of cell occupancy (Eq. 4.15), conditioned on the rth wait time event, now becomes:

$$P_{M_l|W_{r_l}} = \mathbf{Pr}(x \in C_l | W_{r_l} \le n) = \frac{P_{\Sigma_l}^d(n,r) P_{M_l}}{P_{\Sigma_l}(n,r)}. \tag{4.25}$$

The r-criterion mutual information component of 4.24 is calculated as:

$$I(M_l; W_{r_l}|W_{r_l-1} \le n) = H(M_l|W_{r_l-1} \le n) - H(M_l|W_{r_l}, W_{r_l-1} \le n). \tag{4.26}$$

Calculation of the r-criterion component in Eq. 4.24 involves four possible outcomes: two occupancy outcomes and two-wait time outcomes. As $(W_{r_l} \le n, W_{r_l-1}) = (W_{r_l} \le n)$ the latter term of Eq. 4.24 can be calculated by:

$$
\begin{aligned}
H(M_l|W_{r_l}, W_{r_l-1} \le n) &= P_{W_l|W_l^-} H(M_l|W_{r_l} \le n) \\
&\quad + (1 - P_{W_l|W_l^-}) H(M_l|W_{r_l} > n, W_{r_l-1} \le n).
\end{aligned}
\tag{4.27}
$$

Let $P_{W_l|W_l^-} = Pr(W_{r_l} \le n | W_{r_l} \le n)$ be the marginal probability of the rth wait time given the preceding $(r-1)$th event such that:

$$P_{W_l|W_l^-} = \frac{P_{\Sigma_l}(n,r)}{P_{\Sigma_l}(n,r-1)}. \tag{4.28}$$

The right side of Eq. 4.27 combines weighted binary entropy terms, where for the last term the required probability can be found by:

$$P_{M_l|W_l>n, W_{r_l-1}\le n} = \frac{[P_{\Sigma_l}^d(n,r-1) - P_{\Sigma_l}^d(n,r)] P_{M_l}}{P_{\Sigma_l}(n,r-1) - P_{\Sigma_l}(n,r)}. \tag{4.29}$$

Maximization of the Information Payout

Having presented the method for measuring information gain, the next step is the optimization of the payout for an n-pass search of each cell. In [36], Baylog and Wettergren offer mathematical proof that the information gain through the search channel can be optimized as there is local concavity of the information measure over the subspace of ROC operating points, and the measure over the subspace of search passes is submodular. The algorithm to maximize information flow through the search channel is as follows [36]:

1. For $n = 1, ..., N$ and $l = 1, ..., L$, determine the optimal value for the probability of false alarm P_{F_l} and its corresponding information gain as:

$$I^*(n, l) = max_{P_{F_l}} I(M_l; D_l(n)). \qquad (4.30)$$

2. Construct the information payout matrix by calculating the information gain with increasing pass count:

$$\pi(n, l) = I^*(n, l) - I^*(n - 1, l). \qquad (4.31)$$

3. Determine a cell visitation sequence to maximize information gain based on the payout matrix.

In this thesis, the Dlib optimization [61] C++ package is used to find I^* in Eq. 4.30. The Dlib optimizer of choice performs a constrained maximization of the non-linear information gain function.

As an example, a hypothetical search region is used which has 100 equally-sized cells numbered 1 to 100. This sequence of cells is assigned increasing SNR values from 0.3 to 3, where cell #1 has the lowest value and cell #100 has the highest. Following steps #1 and #2 of the above algorithm the information payout per visitation of each cell is plotted in Fig. 4.5.

Figure 4.5 illustrates the properties of the information measure. Each line represents the incremental information gained as additional visitations are added for each cell. The important properties to notice is that all payouts are monotonically decreasing and non-negative [7]. However, the cell that offers the largest information gain varies with increasing cell visitation count. Another visual example of the payout matrix calculated from Eq. 4.31 is shown in Table 4.1 where the information payout of 6 cells is shown for 6 passes. Note that the cells referenced within the table are not related to the cells in Fig. 4.5.
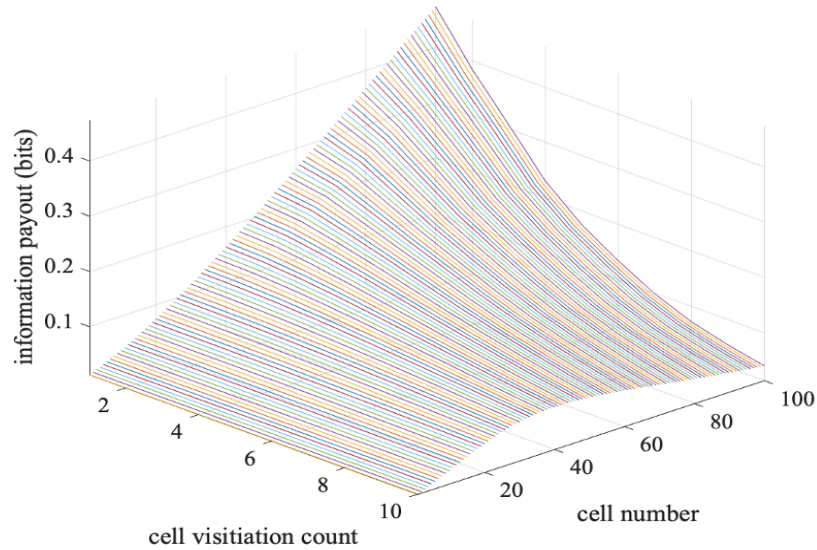
Figure 4.5: Information payout in bits versus cell visitation count for 100 cells with increasing SNR values applied. Adapted from [7].

Table 4.1: Information payout matrix for six random cells over six visitations illustrating the monotonically decreasing nature of the measure

Pass	cell 1	cell 2	cell 3	cell 4	cell 5	... cell N
1	0.0230	0.0420	0.0420	0.0220	0.0220	0.0550
2	0.0180	0.0270	0.0270	0.0180	0.0180	0.0180
3	0.0140	0.0160	0.0160	0.0140	0.0140	0.0080
4	0.0110	0.0100	0.0100	0.0110	0.0110	0.0020
5	0.0070	0.0060	0.0060	0.0070	0.0070	0.0004
6	0.0033	0.0003	0.0003	0.0033	0.0033	0.0001

In step #3 of the algorithm presented for maximizing information flow, Baylog and Wettergren [7] suggest a greedy algorithm to determine the cell visitation sequence (trajectory) of a UUV (or UUVs) conducting the search. Their greedy algorithm selects the cell with the highest information gain and set that as the next cell to be searched. The restriction set on the search game is limiting the searcher effort by the number of cell visitations allowed during the mission, for example total search effort is fixed at $4L = 400$ cell visits. The greedy search shifts search effort towards the higher SNR cells and away from those with the least information gain.

Within this underwater search game formulation, another important parameter to define is a replanning cycle. In a replanning cycle, the information payout matrix is recalculated with updated beliefs of target placement P_{M_l} based on the detection

outcomes collected during the search. After a re-plan the information gain available throughout the search area can change drastically from the initial payout matrix due to the probabilistic nature of the search. For example, if no prior knowledge of object placement is available at the start of the mission, the prior belief in each cell will be $P_{M_l} = 0.5$, which indicates no information is known about the occupancy of cell l. If several positive detections have occurred in a specific cell then the outcome will be a $P_{M_l} > 0.5$. A cell with multiple non-detection events will have a resulting $P_{M_l} < 0.5$. For both cases the information payout for future visits to that cell will be less. However, missed and false detections will also affect the beliefs of target placement. Therefore, a replanning cycle allows the searcher to refocus its efforts towards cells with lower certainty and higher information gain to resolve uncertain mixed detection outcomes.

With fixed search effort and replanning conducted after every $1L$ cells searched, Baylog and Wettergren [36] show their greedy approach is an improvement over the classic boustrophedon search. The authors also conclude that the greedy search is locally optimal for the restrictions of the search game, and for a UUV where effort is measured in cell visits.

In the next section, a reinforcement learning approach is proposed to plan the cell visitation sequence (or searcher UUV trajectory) to maximize information gain. The total search effort is limited by UUV energy usage as opposed to the number of cells visited.

4.2 Information Motivated Q-Learning Autonomous Trajectory Planning

This section presents a novel planner which uses a machine learning approach that exploits any *apriori SNR* variances and regions of higher information gain that arise during the mission.

Reinforcement learning approaches, like Q-Learning, enable an agent (e.g. a UUV) to learn through interactions with its environment and feedback on its decisions. It does not require a world model. Q-Learning develops a policy to inform the agent on actions to reach a goal state, while also seeking to maximize reward collection [38].

Q-Learning was a preferred choice for several reasons. It is flexible in implementation and computationally light-weight so it can run on an embedded UUV platform for *in situ* planning [38]. For the same number of state-action pairs the computational load of Q-Learning does not increase with more complex reward functions, since the outcome only depends on feedback from the environment. For the present work, the optimized cell information payout is the reward function. The reward includes the mutual information gain and exclusion zones (obstacle/blackout avoidance areas). States with $SNR < 1$ or known obstacles/occupancy can be avoided by pre-assigning a reward of 0. As a tool, Q-Learning parameters are intuitive and easily tuned to maximize performance for the dynamic underwater environment. As a machine learning method, it analyzes many possible state-action pairs in (near) real-time. The potential for solutions given large numbers of inputs and complex stochastic environments is yet another justification to choose Q-Learning. Ultimately, the proposed solution seeks to concentrate search efforts where the information gain is larger, while also continuously responding to detection outcomes throughout the mission.

Q-Learning realizes the Bellman equation for optimality based on dynamic programming. A form of this is shown in Eq. 4.32:

$$Q^{new}(s_t, a_t) \leftarrow (1 - \alpha) \cdot Q^{old}(s_t, a_t) + \alpha \cdot (r_t + \gamma \cdot max_a Q(s_{t+1}, a)) \qquad (4.32)$$

where s_t is the current agent state, α is the learning rate, r_t is the reward, γ is the discount factor, and a_t is any action the agent can take. Specifically, an allowed action is the transition to any neighbouring cell. The resulting Q-values and Q-table (not shown) define a policy for the quality of an action given a state where a higher quality Q-value indicates a better action.

Table 4.2 summarizes the standard Q-Learning parameters $-$ discount factor γ, learning rate α, and exploration vs exploitation ϵ. Reward gain is added by the author to weight the importance of reaching the goal cell. The Q-learning algorithm in its entirety is shown in Algorithm 2 in Fig. 4.6.

Table 4.2: Q-Learning Parameters

parameter		description
reward gain		A reward is the value received for a specific action from a given state. The reward gain is the value multiplied by the reward of reaching the goal state. It can be defined as the importance of the goal state.
discount factor	γ	Balances the importance for intermediate versus future rewards. A large γ influences the agent to desire future rewards
learning rate	α	Defines the rate at which new values are accepted when updating $q-$values
exploration vs exploitation	ϵ	The agent interacts with its environment in 2 ways; using prior knowledge (exploitation) or choosing an action randomly (exploration). The ϵ value specifies the probability of exploration vs exploitation.

Algorithm 2 Q-Learning Algorithm with Information Gain Rewards

---

Output: Q-table, Q(s,a), with optimized q-value for each state-action pair

---

Require:

   $States\!: S = \{1, ..., n_s\}$

   $Actions\!: A = \{1, ..., n_a\}$

   $Reward\ matrix\!: R : S \times A \quad (Information\ payout\ matrix)$

Begin:

   $initialize\ Q : S \times A\ randomly$

   while $Q\ is\ not\ converged$

      $start\ in\ state\ s\ \in S$

      while $s\ is\ not\ the\ goal\ state$

         $choose\ a\ based\ upon\ greedy - exploration\ strategy$

            $a \leftarrow random\ a\ \in A_s\ ...\ (exploration)$

            OR

            $a \leftarrow argmax_a Q(s, a)\ ...\ (exploitation)$

            $r \leftarrow R(s, a)$

            $s' \leftarrow s + a$

            $Q(s', a) \leftarrow (1 - \alpha) \cdot Q(s', a) + \alpha \cdot (r + \gamma \cdot max_{a'} Q(s', a'))$

            $s \leftarrow s'$

   return Q

Figure 4.6: Q-Learning algorithm with information gain rewards.

The selection of goal cell for the Q-Learning algorithm is similar to the greedy method described in the previous section. The region in the environment with the highest information gain potential is selected as the goal, and then Q-Learning determines the policy to reach that state. After reaching the goal state, the pass count of the searched cells is incremented and the reward matrix is updated with the information payout values for the next visit to each cell, respectively.

The *mRobot* node developed in Chapter 3 is modified here to subscribe to the state of the searcher UUV and distribute a trajectory to the UUV in real-time. A summary of the proposed underwater search planner is presented in Fig. 4.7, with the the author's contribution highlighted in the blue dotted box.
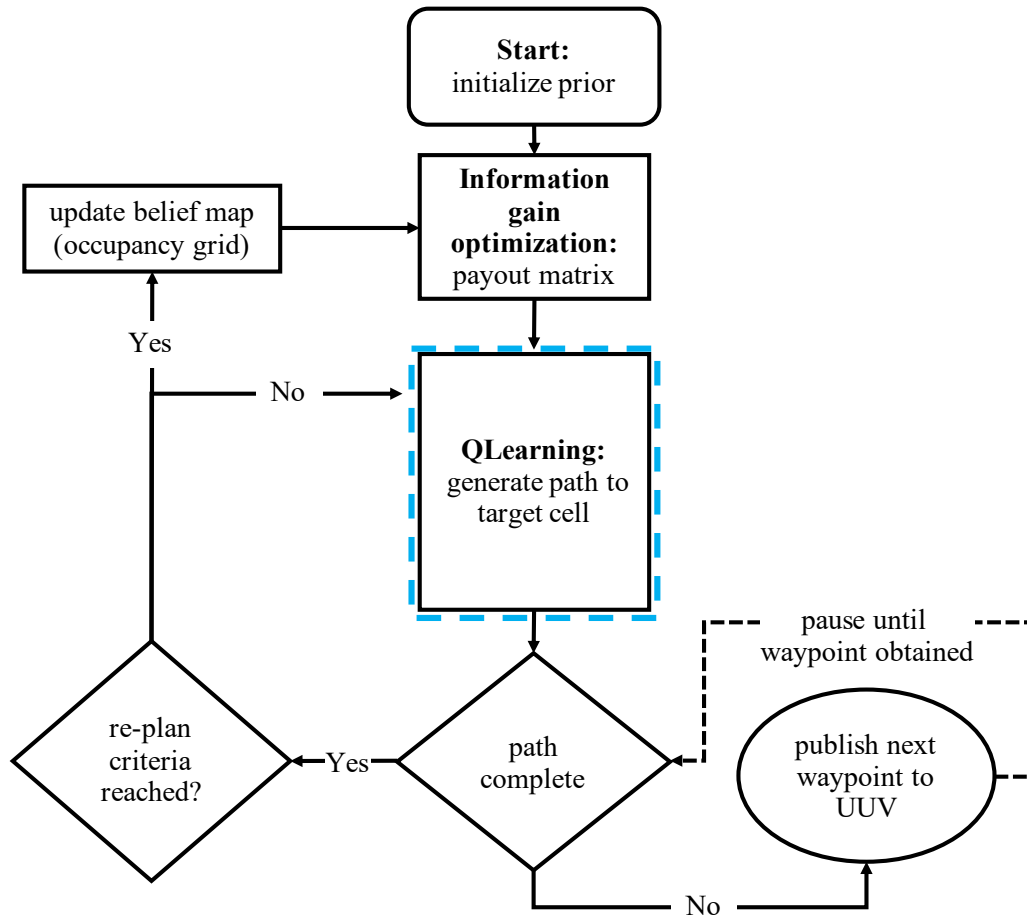

Figure 4.7: Underwater search planning flow diagram implemented within the *mRobot* ROS node [4].

4.3   Simulation and Experimental Configuration

A Gazebo virtual world was developed to evaluate the path-planner's effectiveness in a controlled underwater environment through simulations. The ECA A9 autonomous underwater vehicle (AUV) and the UUV Gazebo simulator tools were used [27]. April-Tags uniquely labelled the targets in the UUV sensor's (camera) FOV. This abstracts out target features and eliminates the feature recognition problem as it is peripheral to the path-planning objectives. To incorporate AprilTag, the AprilTag ROS package was integrated into the Gazebo underwater world [62]. Initial experimental validation of the adaptive path-planner was conducted in an indoor setting with a quadcopter representing the UUV. The environments for experimentation and simulation are shown in Fig. 4.8(a) and (b), respectively [4].
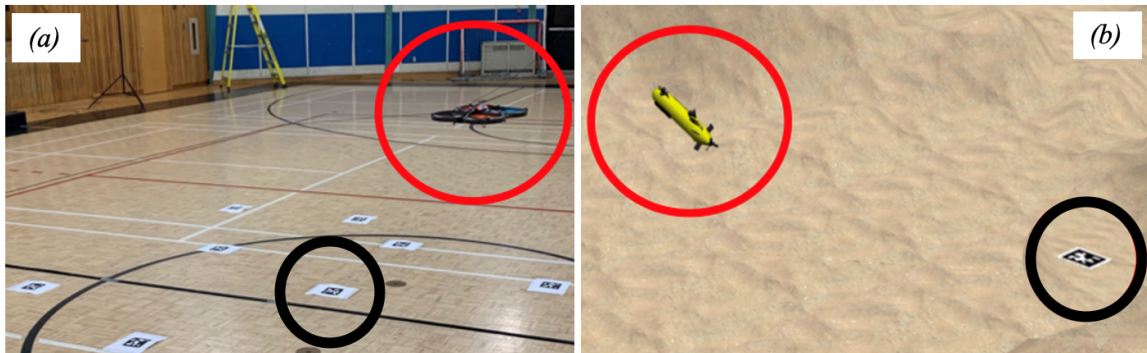

Figure 4.8: Robots (circled red) search for AprilTag targets [62] (circled black) in a: (a) structured gym environment with the Parrot UAV (experiments) and (b) Gazebo underwater world (simulations) with the ECA A9 AUV [27]

The simulation environment mirrors the square search region in Fig. 4.9 used in experiments (Fig. 4.8a), with the search cells scaled up to 40 m × 40 m. Note that the pattern of strong and poor sensing regions (white and gray cells, respectively) was varied to represent different environmental conditions. Search areas with variable environmental conditions were simulated by applying SNR values to each cell in either a homogeneous or variable distribution [36]. The SNR is a property of the sonar's interaction with the environment, and varying the SNR value in a region is a method used here to capture the varying conditions that affect the search performance.
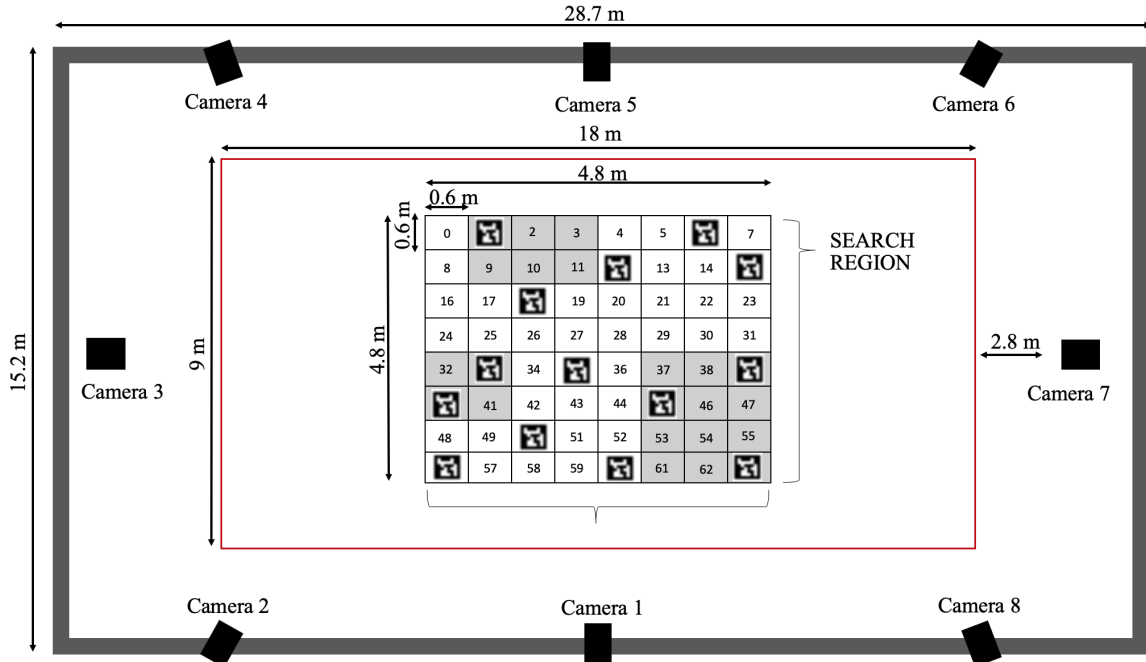
Figure 4.9: Experimental setup floor plan showing the dimensions of the search region and the 8 motion capture cameras. AprilTag targets are randomly distributed in the 64-cell search region, where the white and gray cells represent regions of differing environmental conditions and consequently, variable probabilities of detection and false alarm [4].

4.4 Results and Discussion

This section provides results from sensitivity studies on the proposed algorithm, and simulated NMCM search mission results.

4.4.1 Parameter Sensitivity Study

To begin, the influence of the discount factor γ and reward gain on the UUV trajectory is graphically demonstrated in Fig. 4.10 and 4.11. α and ϵ are not explored here as they only affect the algorithm convergence rate, as the reward values in each state are deterministic [63].

Figure 4.10 shows an example of a search region with a random SNR distribution of values from $0 - 3$ over the cells. Note the 'start' and 'goal' points of the trajectory are indicated. This map example will be used to highlight the influence of reward gain and γ.

Figure 4.11 shows the environment of Fig. 4.10 under the case of low reward gain, 2.0 (Fig. 4.11a) and high reward gain, 24.0 (Fig. 4.11b) with $\gamma \in \{0.1, 0.3, 0.7, 0.9\}$. Note that as γ increases, the UUV path becomes more direct and thus more fuel efficient; however, the increase in information gain is smaller. The path is even more direct with increased reward gain (Fig. 4.11(b)).
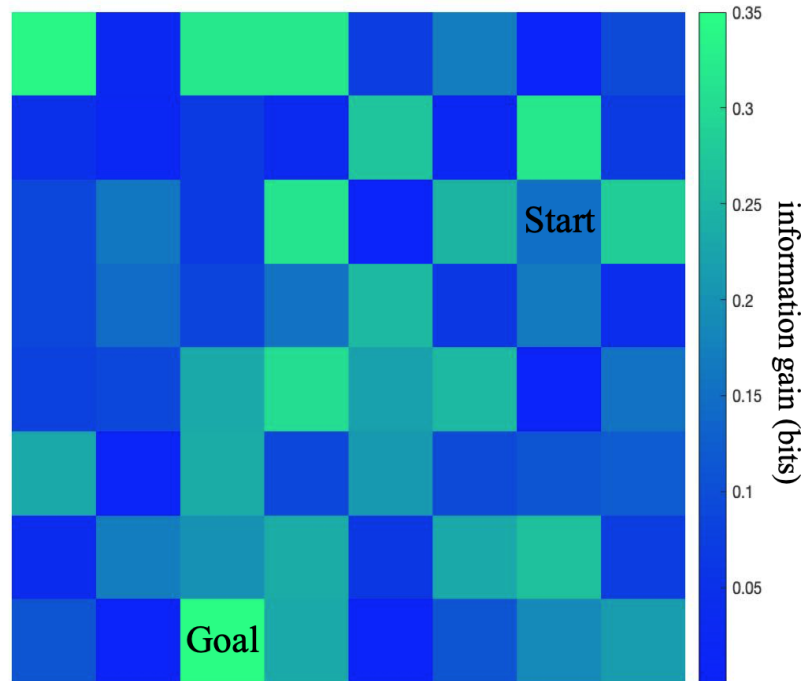

Figure 4.10: Simulated 64-cell search region with each cell assigned a randomly distributed SNR value from $0 \rightarrow 3$. The colour map and scale show the optimized information gain payout for each cell [7]. The start and goal cells have been identified.
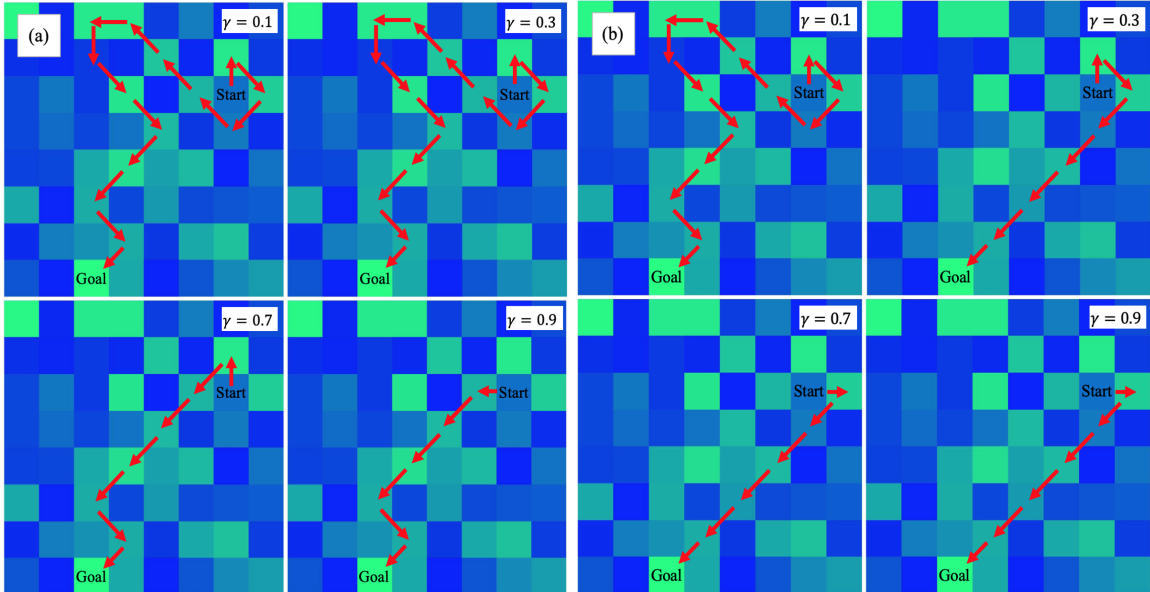
Figure 4.11: Q-Learning path for an agent navigating through the environment in Fig. 4.10 with (a) low reward gain (2) and increasing values of discount factor and (b) high reward gain (24) and increasing values of discount factor. The path of the agent for each simulation is shown with red arrows. These results illustrate the impact of an increasing reward gain and discount factor to draw the agent towards the goal without wandering from the shortest possible path.

A comprehensive sensitivity study is now detailed using results from full simulated search missions as described in section 4.3. Figure 4.12 shows the process for the parameter sensitivity studies in (ROS) simulations for three environments:

1. Homogeneous SNR 3 environment: $SNR = 3$ applied to each cell (Fig. 4.16a)

2. Homogeneous SNR 1 environment: $SNR = 1$ applied to each cell (Fig. 4.17a)

3. Variable SNR gradient environment: SNR increases from 0.3 to 3 in a gradient across the environment(Fig. 4.16b)
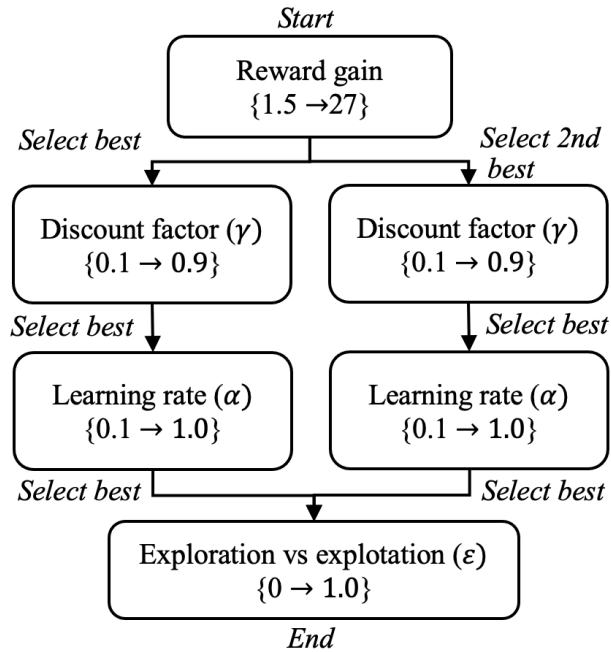
Figure 4.12: Parameter study flow diagram [4].

For any change in parameter value, 100 simulations were performed, and the results were averaged. To start, the discount factor and learning rate were both set to mid-range values of 0.5 and the reward gain was studied over a range of 1.5 to 27. The reward gain value that produced the highest average information gain over a search mission was selected. A second reward gain value was also selected to ensure each study was more comprehensive. The same process was continued until each variable was studied, and the best set of parameters was selected. The selected range in reward gain encompasses low to high values of the parameter. As shown previously in Fig. 4.11, a reward gain of 24 is sufficiently large to draw the agent to the goal node on a direct path for most values of discount factor γ. Larger values of reward gain do not result in a statistically significant difference in the path or total information gain. The discount factor γ is traditionally studied in a range from 0.1 to 0.99, which is a range from short-sidedness (immediate desire for rewards) to striving for long-term high rewards. Discount factor $\gamma \geq 1$ can lead to divergence in Q-values [64].

For this study, the Q-Learning search was allotted the amount of energy required by the Boustrophedon path to exhaustively search the 64-cell region three times. Two replanning opportunities were allowed. For brevity, the results of this analysis are

not all shown. The select example of the homogeneous SNR 3 environment is shown. The SNR 3 environment is a good representative example as it provided the results showing the greatest sensitivity to changing parameters. The increased sensitivity occurred as it is the search region with the highest average SNR across all cells, and therefore the highest total information gain potential.

Figure 4.13 and Table 4.3 shows the first results from the study that examined increasing values of reward gain.
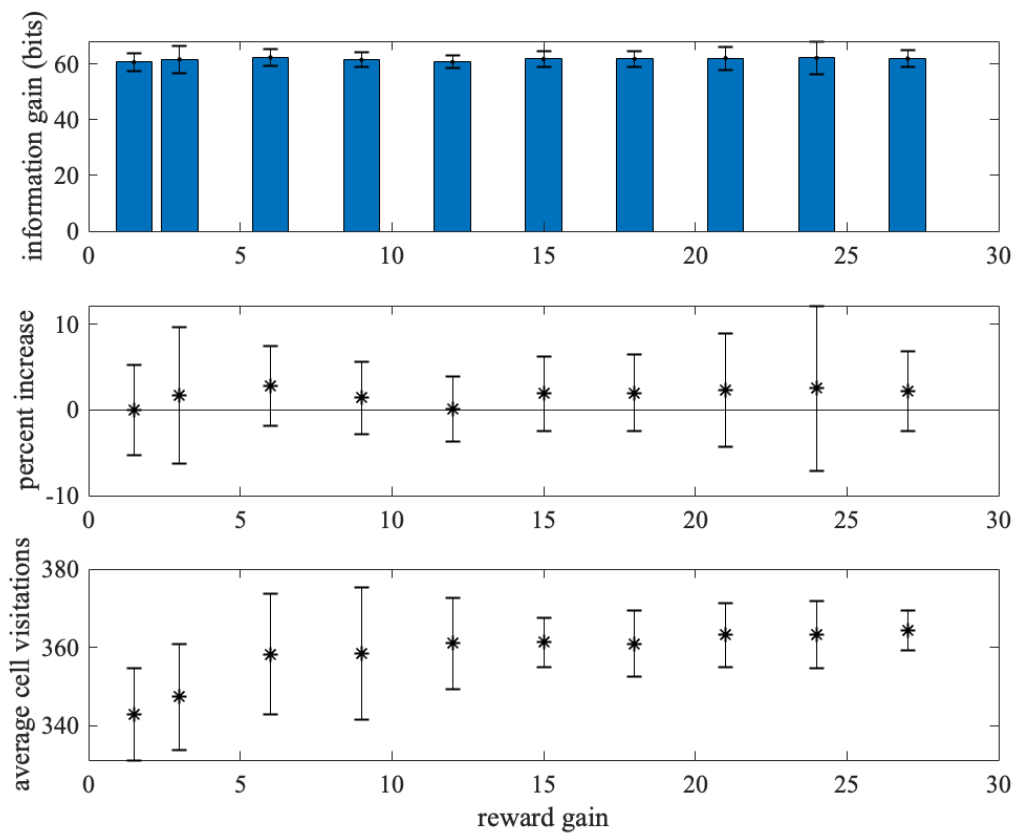

Figure 4.13: Reward gain parameter study for homogenous SNR 3 environment, with error bars representing a 95% confidence interval. Maximum information gain average (bits) is found with a reward gain of 6 and 24, respectively. The results show no significant parameter sensitivity for total information gain. The average cell visitation does increase with increasing reward gain.

Table 4.3: Summary of results for the reward gain sensitivity study in an SNR 3 environment (Fig. 4.13)

reward gain	information gain (bits)				cell visitations	
	mean	*max*	*min*	*std. deviation*	*mean*	*std. deviation*
1.5	60.7	62.5	57.3	1.6	343.0	5.9
3	61.7	66.1	59.1	2.4	347.4	6.8
6	62.4	65.2	61.0	1.4	358.3	7.7
9	61.5	63.0	59.4	1.3	358.4	8.4
12	60.7	63.1	59.2	1.2	361.0	5.8
15	61.8	64.9	60.2	1.4	361.3	3.2
18	61.9	64.4	59.2	1.4	360.9	4.2
21	62.1	64.3	57.7	2.1	363.2	4.1
24	62.2	67.6	58.0	3.0	363.3	4.3
27	62.0	63.9	59.8	1.4	364.4	2.5

set mean = 61.69, standard deviation = 1.81

The findings from Fig. 4.13 and Table 4.3 are that changing the reward gain had no statistically significant impact on the total information gained during the search missions. However, there does appear to be a trend of increasing average cell visitations with increasing reward gain —which is expected. As seen in Fig. 4.11 higher values of reward gain lead to more direct paths to the goal, albeit with less total information gained. For a UUV under-way in a horizontal plane, the majority of the energy consumption is due to surge, yet significant energy is still required for yaw control [65]. Heading changes between waypoints also increases the distance travelled relative to the Euclidean distance between those two points. Therefore, navigating with mostly constant headings is more efficient, leading to the expectation that a higher reward gain would lead to more cell visitations for the same given total energy. Ultimately, a reward gain of 6 and 24 were chosen for the next study as they provided they highest average information gain.

Figure 4.14 and Table 4.4 provide results from the discount factor γ parameter study for a reward gain = 6. Figure 4.15 and Table 4.5 provide results from the discount factor γ parameter study with a constant value of reward gain = 24.
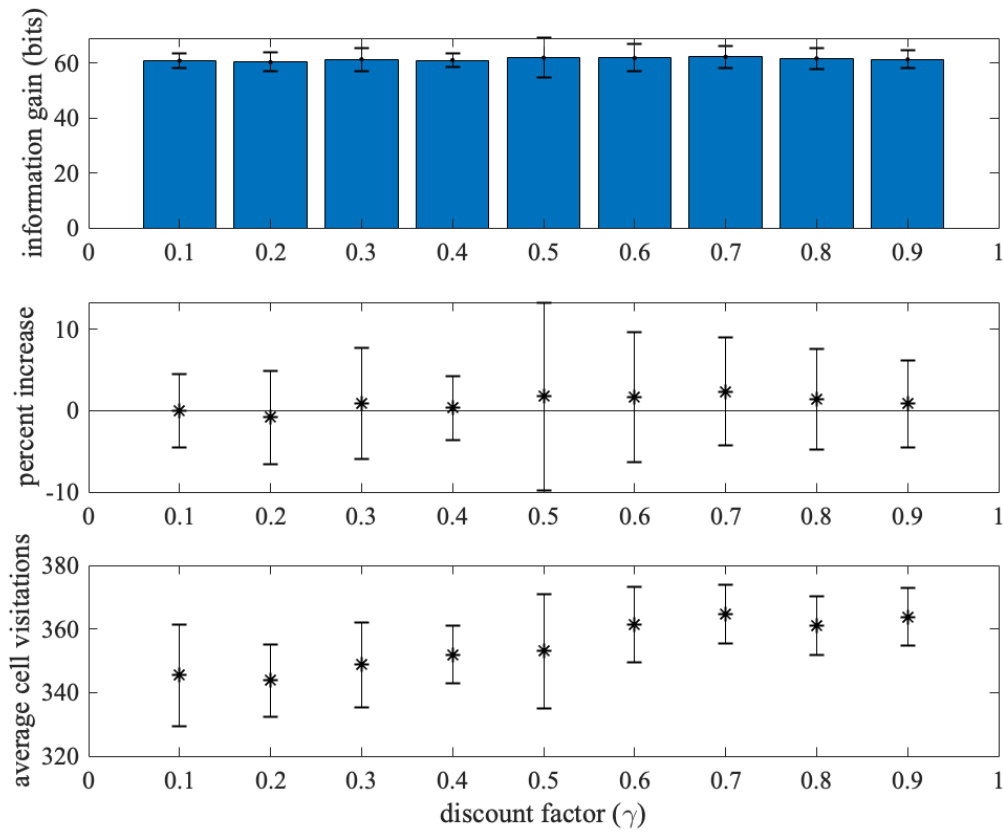
Figure 4.14: Discount factor, γ, parameter study for homogenous SNR 3 environment and a reward gain of 6, with error bars representing a 95% confidence interval. The result is no significant sensitivity to changes in the γ parameter.

Table 4.4: Summary of results for γ sensitivity study in SNR 3 environment, with reward gain held to a constant value of 6 (Fig. 4.14)

discount factor	information gain (bits)				cell visitations	
	mean	*max*	*min*	*std. deviation*	*mean*	*std. deviation*
0.1	60.9	63.0	58.7	1.4	345.4	8.0
0.2	60.3	62.4	57.4	1.7	343.8	5.7
0.3	61.4	65.4	58.0	2.1	348.7	6.7
0.4	61.0	62.8	59.1	1.2	351.9	4.5
0.5	61.9	68.0	57.1	3.6	353.0	9.0
0.6	61.9	65.3	58.2	2.5	361.4	5.9
0.7	62.3	64.8	57.8	2.1	364.7	4.6
0.8	61.7	64.3	59.0	1.9	361.2	4.6
0.9	61.4	63.0	58.5	1.6	363.8	4.6

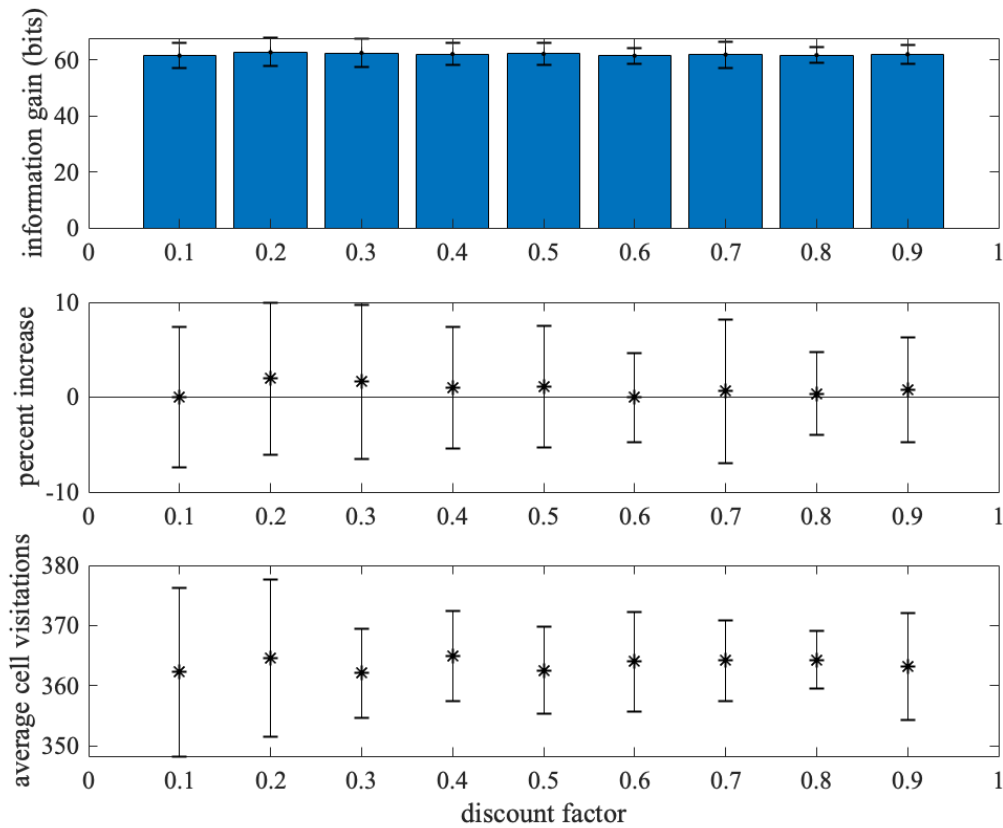*set mean = 61.42, standard deviation = 2.09*

Figure 4.15: Discount factor, γ, parameter study for homogeneous SNR 3 environment and a reward gain of 24, with error bars representing a 95% confidence interval. Again, showing the expected result of no significant sensitivity to changes in the parameter.

Table 4.5: Summary of results for discount factor sensitivity study in SNR 3 environment with reward gain equal to 24 (Fig. 4.15)

discount factor	information gain (bits)				cell visitations	
	mean	max	min	std. deviation	mean	std. deviation
0.1	61.5	65.7	58.4	2.3	362.3	7.0
0.2	62.7	67.6	58.5	2.5	364.6	6.5
0.3	62.5	66.1	57.6	2.5	362.1	3.7
0.4	62.1	64.8	59.2	2.0	364.9	3.7
0.5	62.1	65.5	59.1	2.0	362.6	3.6
0.6	61.5	63.7	59.0	1.4	364.0	4.2
0.7	61.9	66.3	58.7	2.3	364.2	3.4
0.8	61.7	63.3	59.1	1.3	364.3	2.4
0.9	62.0	65.2	59.6	1.7	363.2	4.5

set mean = 61.98, standard deviation = 2.00

The results from the two discount factor γ studies provide similar insight, and showed little statistical significance in the total information gain. For the study with reward gain = 6, a similar trend of increasing average cell visitations is seen with increasing γ (Fig. 4.14). This result follows the same reasoning that direct paths are more energy efficient. In the second γ study, no trend is seen for either information gain or number of cells visited. This result should again be expected as a high reward gain of 24 already drives the agent towards the goal at low γ values. A reward gain = 6 and discount factor γ = 0.5 were selected as the best set of parameters moving forward, as the highest average information gain was seen with those values.

The results from the α and ϵ studies are also not shown for brevity reasons. The expected result was seen from those studies: for a deterministic reward map, α and ϵ will only affect the convergence rate, and not the policy.

In summary, it was found that the reward gain and the discount factor γ are the only parameters that had any statistically significant effect on the UUV trajectory or the search outcome. While the trajectory varies, the total information gained during the search mission is not significantly affected from varying the Q-Learning parameters. With low values of reward gain and discount factor γ, the agent prioritizes immediate rewards with a slight sacrifice to efficiency, while with high values the agent is drawn to the maximum reward cells without much deviation from the shortest path. The two opposing strategies converge to a similar information gain total over the mission duration.

These results provide confidence that the proposed solution is suitable and insensitive for a wide range of parameters. However, if the constraints of the mission were different (shorter duration or less replanning cycles) the parameters could be easily tuned to fit that scenario depending on the specific needs.

4.4.2 Search Mission results

The Q-Learning information gain as a function of time was compared with the nominal boustrophedon (lawnmower) and greedy (locally-optimal) planning methods (Figs. 4.16 and 4.17). The comparison is over four different 64-cell environments for SNR = 3.0 (Fig. 4.16a), an SNR that decreases in a gradient from left to right (Fig. 4.16d), SNR = 1.0 (Fig. 4.17a), and an SNR that varies in distinct regions (Fig.

4.17d). Both the experiments and simulations used these environments.

Again, two replanning opportunities were allowed for each method to provide equal information gain potential. Each search method was allotted the energy to complete the boustrophedon search. Thus time axes in Figs. 4.16 and 4.17 were normalized by the boustrophedon search duration.

These results demonstrate the improved capability of Q-Learning to exploit SNR heterogeneity and/or prior belief of cell occupancy throughout the environment. With the exception of a small difference between the simulation (Fig. 4.16c) and experiments (Fig. 4.16b) for the $SNR = 3.0$ case, and a similar difference in the $SNR = 1.0$ case (Fig. 4.17c and Fig. 4.17b), there is agreement in the relative performance of the 3 approaches between experimental measurements and simulation predictions.

Not unexpectedly, the greedy planner outperforms (or matches) the boustrophedon for most cases. The exception to that is seen in the $SNR = 1$ case where the greedy planner performs the worst. The greedy planner works well when there are areas of high information gain to exploit (gradient and zoned environments). In a homogeneous search area of low SNR the greedy planner becomes ineffective relative to the other two methods, as it makes energy inefficient decisions with low payout. For every result the greedy method utilizes all available energy first due to its local approach that does not consider energy efficiency in its cell selections. The duration of the greedy search in the experiment is even shorter due to the difference in holonomic constraints between the UAV and UUV, which had more of an impact on the greedy search than the other methods. Conversely, the boutrophedon approach takes the longest to complete the search being the most energy efficient method, as it utilizes the minimum the number of turns to cover the area.

Generally, Q-Learning outperforms both of them with slightly better performance in variable SNR environments over homogeneous ones. Again, the greedy search is more inefficient as it chooses the best cells versus considering a more optimal trajectory like Q-Learning.

The experiments at this stage are for initial validations to show the algorithm

works and is correctly implemented in two-dimensions. The experiments were performed with an unmanned aerial vehicle at constant altitude which is a near two-dimensional system albeit with more degrees-of-freedom than some UUVs. This approach is part of the process that the Lab uses to mitigate risk prior to implementing and deploying on expensive systems. Later phases of this research will consider the holonomic constraints of UUVs.
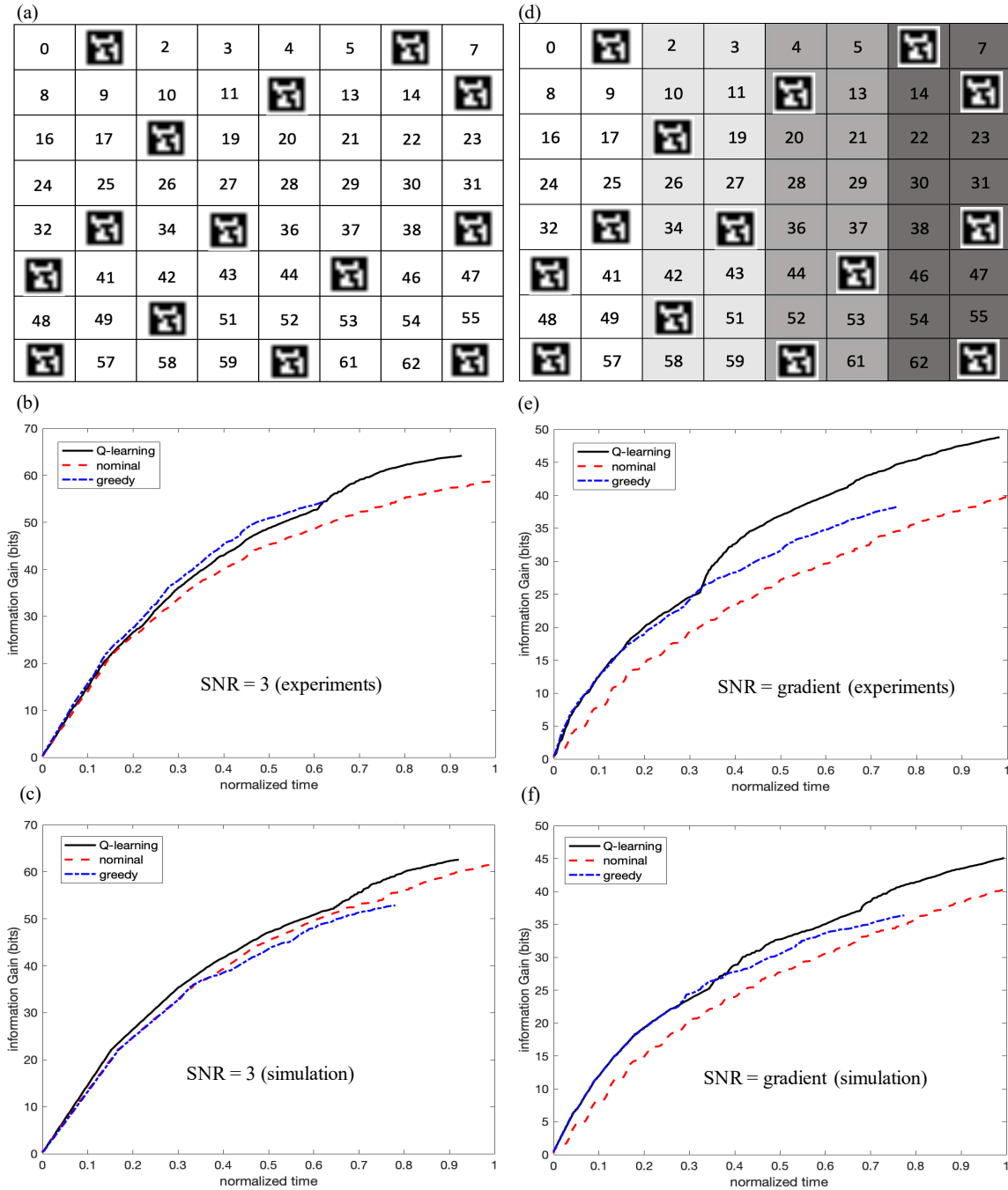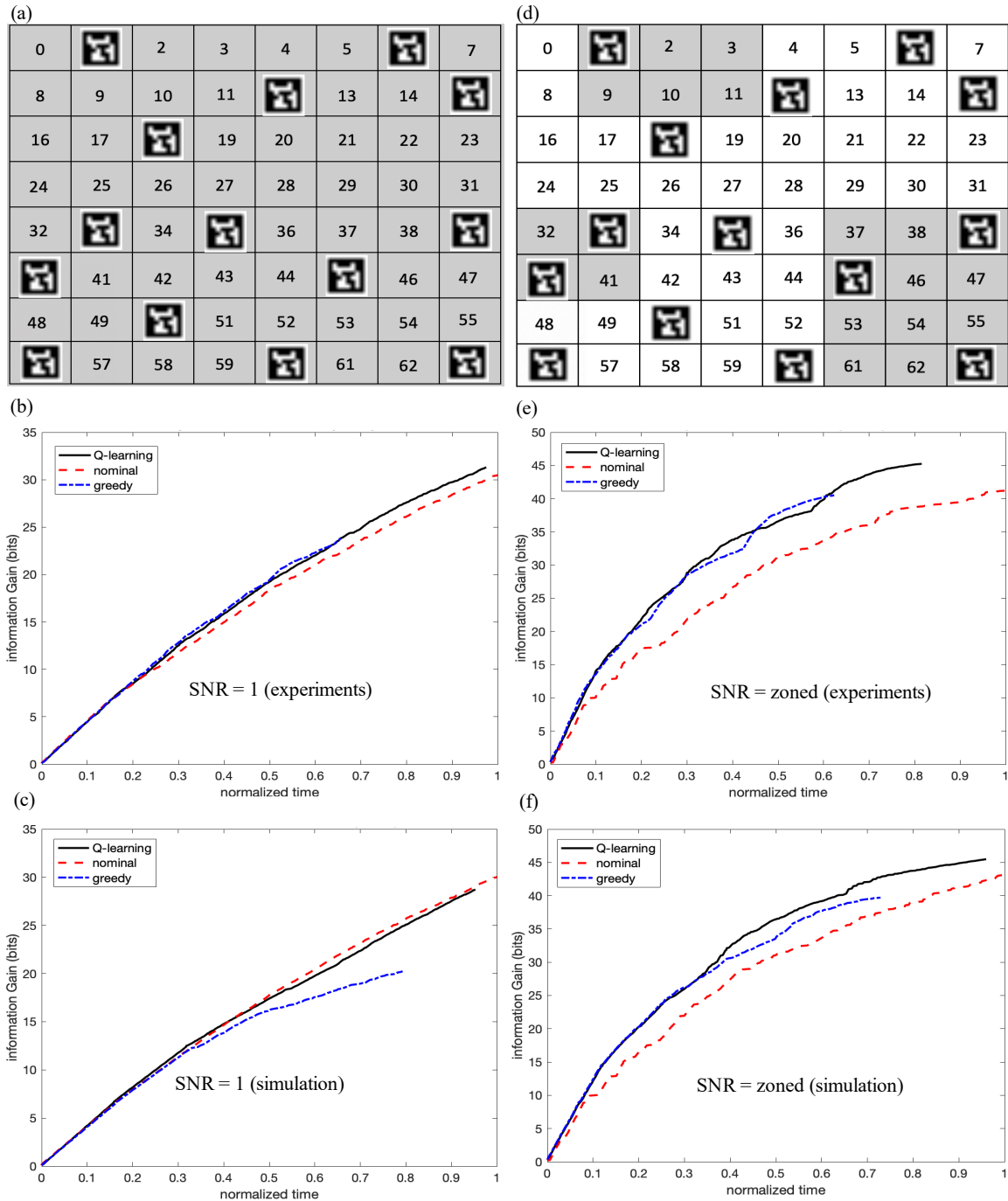
Figure 4.16: Performance comparison of the Q-Learning based planner against the greedy and nominal boustrophedon algorithms for a 64-cell environment. (a) A homogeneous environment with a high SNR of 3.0 (high probability of detection). (b) Experimental measurements from model-scale runs with a quadcopter and cell size of 0.6 m × 0.6 m.(c) Full-scale simulation with the ECA A9 AUV in Gazebo with cell size 40 m × 40 m. (d, e, f) provide the same flow of results for a search region with the SNR values decreasing in a gradient from left to right.) [4]

Figure 4.17: Performance comparison of the Q-Learning based planner against the greedy and nominal boustrophedon algorithms for a 64-cell environment. (a) A homogeneous environment with a high SNR of 1.0 (low probability of detection). (b) Experimental measurements from model-scale runs with a quadcopter and cell size of 0.6 m × 0.6 m.(c) Full-scale simulation with the ECA A9 AUV in Gazebo with cell size 40 m × 40 m. (d, e, f) provide the same flow of results for a search region with the SNR values varying in zoned regions (SNR of 1.0 & 3.0).)

4.5 Summary

The development of the information measure for the underwater search channel was defined. A novel Q-Learning approach to the underwater search problem with information gain rewards was developed. The proposed algorithm outperformed the classic boustrophedon and greedy approaches in NMCM search missions, as shown in both simulations and two-dimensional experimentation.

The next chapter pushes the problem further to consider a more capable and intelligent underwater target. A collaborative robot system with 3 UUVs is proposed to detect, track and localize an evasive underwater target. The solution builds upon the information incentivized Q-Learning algorithm and the multi-robot task allocation *mRobot* ROS node.

Chapter 5

Collaboration of Marine Robots Towards Dynamic Target Localization and Tracking

A mobile underwater vehicle that is unresponsive and evasive represents the most capable and intelligent of the three targets considered in this thesis. The proposed solution uses a collaborative system of three UUVs to detect, track, and localize the target (Fig. 5.1).
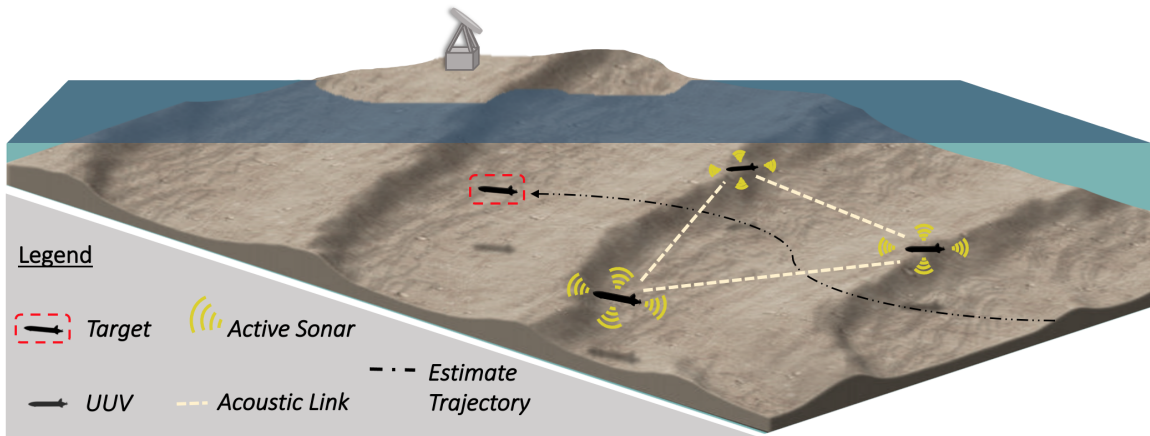

Figure 5.1: Depiction of underwater collaborative tracking and localization of an evasive target

First, an adaptive planner that combines Q-Learning with dynamic and predictive information measures is developed for collaborative trajectory planning − which is integrated with the *mRobot* node. Then, the simulation configuration, in particular the acoustic environment, is presented. Lastly, results are provided comparing the collaborative system against the standard LBL solution.

5.1 Autonomous Multi-Robot Target Tracking Algorithm

Estimating the current position, velocity, and heading of a non-responsive target is a non-trivial problem. As presented in the literature, one solution is LBL localization

that uses stationary acoustic beacons to range the target with two-way-travel-time (TWTT) measurements. With this type of active ranging, the beacons transmit an acoustic signal and measures the two-way-travel-time (TWTT) to determine the distance to a target. An unambiguous solution is not possible unless three or more measurements are available, which is the principle of trilateration [66]. The system developed here uses trilateration and a particle filter for target state-estimation (the state being its time-varying pose).

5.1.1 Target State Estimation

The filtering of available information to best estimate the vehicle state is the basis of any navigation algorithm. The goal of recursive state estimation is a refined estimate of the system state x, given its: prior state x_{t-1}, an input, control or action u, and measurement z [11]. In this work the target input or action is also based on the estimate of its velocity and heading. As well, a particle filter is used to fuse the measurements and estimate the state of the target.

Particle filtering (PF) is a sequential state estimation technique for stochastic non-linear and/or non-Gaussian state-space models. A particle filter provides a strong alternative to the commonly used extended Kalman filter. In particle filtering, continuous distributions are approximated by discrete random measures, which are comprised of particles (hypotheses in state space) and their associated weights [67]. Each particle represents a potential solution to the target state, and the particle weight is a measure of its importance, or likelihood. In short, the particles are sampled from a prior distribution and then updated in future time steps based on the estimated motion model of the target. When a measurement becomes available the particles are re-sampled based upon their prior weights and their likelihood given the most recent measurement, using Bayes' rule. Then, the predicted target position (state) can be given by the weighted mean of all the particles. The three more commonly used ways are [68]:

1. weighted mean is chosen;

2. best particle (the one with the largest likelihood) is chosen, and

3. mean of the best group of particles is chosen.

The first option is applied in this thesis. The entire PF algorithm is not shown here as there are many references for this.

In this work, the target velocity $\hat{v}$ and heading $\hat{h}$ are estimated with running averages based on the history of the target state estimate and a UUV kinematic model in the horizontal plane by:

$$\hat{x}_t = \hat{x_{t-1}} + (\hat{v} \cdot ts \cdot cos(\hat{h})) \tag{5.1}$$

and

$$\hat{y}_t = \hat{y_{t-1}} + (\hat{v} \cdot ts \cdot sin(\hat{h})) \tag{5.2}$$

where ts is the time in between measurements.

To illustrate the principle of a working particle filter, a MATLAB simulation was developed for a stationary LBL system tracking a point target. The target is translating in the horizontal plane with a velocity of 2 m/s and a heading that changes randomly by $\pm[0,1]$ degrees per second. The measurement input into the particle filter is the triliteration solution of three static beacons using ideal range sensors with additive zero-mean Gaussian noise with 1 m variance. Acoustic delay is considered for communication and ranging based on a sound speed in water of 1500 m/s. Figure 5.2 shows the target state estimate in four different time instances of the described simulation. The particles updated according to the estimated target motion model, given time in between measurements, are shown in magenta. The re-sampled particles (given new measurement information) are shown in black, and the measurement is shown in red. A PF with 1000 particles is implemented in this work.
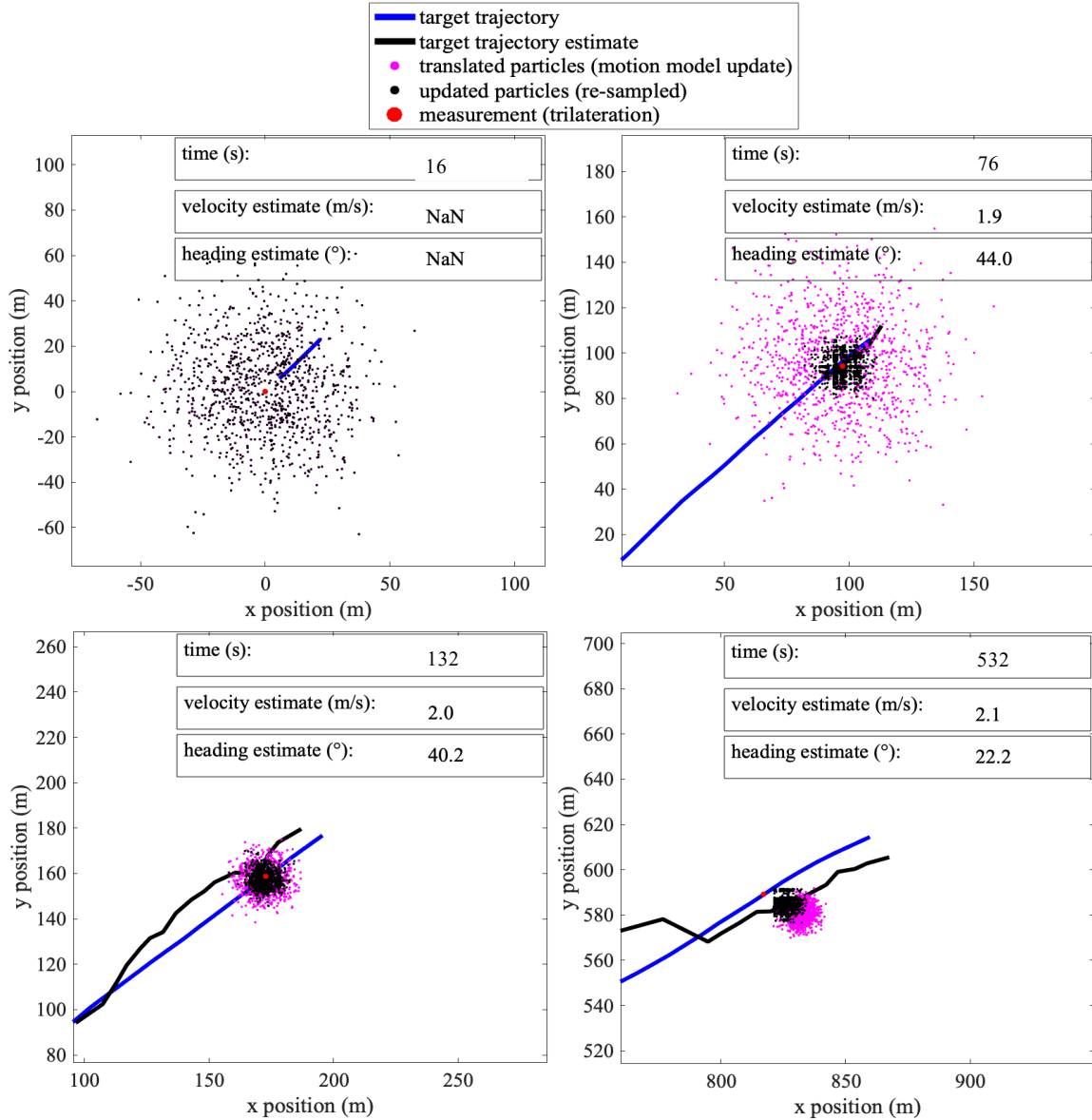
Figure 5.2: Particle filter state-estimation applied to target tracking simulation at four different instances in time. The first measurement becomes available at $t = 16$ s, and the particles are then drawn from a prior distribution surrounding the measurement. The particle distribution becomes more certain (less distributed) and the velocity and heading estimates improve with each measurement.

The plots in Fig. 5.2 illustrate in this example how the distribution of the particles can change over time and become less distributed (or more certain) with more measurements. The velocity and heading estimates are provided to illustrate how the estimate of those parameters can adapt over time to respond to changes in the target's trajectory.

In underwater applications the performance of a recursive state estimator relies on the ability of the system to acoustically detect the target. Similar to the sonar imaging techniques used to detect targets on the sea floor, the SNR quantifies a sonar's ability to range a target given an acoustic ambient.

5.1.2   Sonar Equation

The results and method presented in this chapter use the sonar equation to estimate the acoustic channel for both communications and active sonar ranging. For active acoustic ranging of a target the SNR can be approximated using the sonar equation in the form of Eq. 5.3. One-way-travel communication SNR can be approximated using the sonar equation in the form of Eq. 5.4 [69]

$$SNR = SL + TS - 2(TL) - NL \qquad (5.3)$$

$$SNR = SL - TL - NL \qquad (5.4)$$

where SNR is the received signal-to-noise ratio in dB, SL is the source level, TS is the target strength, TL is the transmission loss, and NL is the ambient noise level.

The source level is the ratio of the transmitted intensity 1 m from the source to a reference intensity and can be calculated by [69]:

$$SL = 10log_{10}P + 170.8 \qquad (5.5)$$

where P is the transmitted power in Watts.

The target strength is the ratio of the reflected signal intensity at 1 m from a target to the incident intensity, in dB [69]. The target strength TS can be calculated by:

$$TS = 10log_{10}\frac{\sigma}{4\pi} \qquad (5.6)$$

where σ is the effective cross-sectional area based on the target aspect angle.

The final term TL is the attenuation of the acoustic signal as it propagates through the underwater channel. There are two components to transmission loss. The larger of the two is geometrical spreading of the wavefront. In an infinite-sized medium

the wavefront will spread spherically. In reality, the wavefront will spread spherically until it reaches the water-surface or ocean-floor (e.g. in shallow water), which will cause the wavefront to then spread cylindrically. The transmission loss for spherical and cylindrical spreading can be calculated using Eqs. 5.7 and 5.8, respectively. Equations 5.7 to 5.13 are summarized from [14].

$$TL = 20log_{10}R + (\alpha_{vis} + \alpha_B + \alpha_M)R \tag{5.7}$$

$$TL = 10log_{10}R + (\alpha_{vis} + \alpha_B + \alpha_M)R \tag{5.8}$$

The second, and lesser contribution to transmission loss, is absorption of the sound as it propagates. The absorption loss has three components —viscous, boric acid relaxation, and magnesium sulphate relaxation [14]. All absorption components are modelled with a linear dependence on range αR, where α is the absorption constant (dB/km) and R is range (km). The viscous absorption coefficient α_{vis} is calculated per Eq. 5.9, which is a function of frequency f, temperature in Celsius T, and depth D.

$$\alpha_{vis} = 4.9 \times 10^{-4} f^2 e^{-(T/27+D/17)} \tag{5.9}$$

The absorption second mechanism is the relaxation process of boric acid, whose co-efficient α_B, can be described by:

$$\alpha_B = 0.106 \frac{f_1 f^2}{f_1^2 + f^2} e^{-(pH-8)/0.56}. \tag{5.10}$$

The final absorption mechanism is the magnesium sulphate relaxation process α_M, which is described by:

$$\alpha_M = 0.52(1 + \frac{T}{43})(\frac{S}{35}) \frac{f_1 f^2}{f_2^2 + f^2} e^{-D/6}. \tag{5.11}$$

The terms f_1 and f_2 can be found with Eqs. 5.12 and 5.13:

$$f_1 = 0.78\sqrt{S/35}e^{T/26} \tag{5.12}$$

$$f_2 = 42e^{T/17} \tag{5.13}$$

where S is salinity in parts per thousand (ppt). The acoustic modem parameters used are from the Woods Hole Oceanographic Institute (WHOI) micromodem with a carrier frequency $f_c = 24$ kHz and $P = 100$ W [70].

In summary, the sonar equation gives an estimate of the SNR for acoustic ranging given the position of the target relative to the sensing UUV. It is also applied to estimate the SNR for communications between collaborating UUVs. In continuation, the SNR impacts the information measure when ranging a target —similar to that in Section 4.1.

5.1.3 Dynamic Information Motivated Q-Learning Trajectory Planner

The autonomous planner developed here builds on the information measure and channel definition presented in Section 4.1. The search game discussed previously was the two-player game where the second player is the target whose objective is to stay hidden. When the second player can move in a manner that is uncertain to the searcher it is called a rendezvous game. The information measures in a rendezvous game are solved with a similar mathematical procedure, such that Eq. 4.24 still applies [7].

The proposal here is for the search channel defined in Section 4.1 to be adapted to the underwater target-tracking problem. That is, the search channel becomes the acoustic channel to range a target in the water-column rather than to detect it on the sea floor. The environment is still partitioned into discrete cells.

To make decisions on where the each of the three UUVs should navigate, a prediction of the target's state into the future is required. First, the look-ahead level L_n is defined as the estimated state of the target at a future time step ts_n. The time step ts is defined as the time it takes for a UUV to travel the distance across a cell. For continuity, the cell sizes here are kept the same (40 m × 40 m) as in Chapter 4. Each look-ahead level is used to predict the SNR in each cell at that time, per Eq. 5.3. Following Eq. 4.24 a predictive information payout map is computed for each look-ahead level. Multiple-pass search calculations ($n > 1$) are no longer considered as the information payout per cell changes with each time step. Increasing the look-ahead level inherently increases the computations. However, too few can lead to the planner being near-sighted, which is making short-term decisions that could negatively affect future performance.

The payout map's size are parametrized by length and width so that computations can be limited to a focused region around the target. For the results in this chapter the map size was set to 3 km $\times$ 3 km. A map of this size is approximately equal to the region around the target where the $SNR > 0$ (and information gain > 0), which will be discussed further in section 5.2.1. As the payout maps no longer cover the whole search area, they are centred around the estimated location of the target in the final look-ahead level. To avoid the collaborating UUVs from approaching too close to the target, a constraint is applied which sets the information payout to zero within a set radius from the target.

An example of the proposed predictive information payout maps are shown in Fig. 5.3, where $L = N$ predictive payout maps are shown for a target moving along the x-axis with a constant position paralleling the y-axis at 1500 m.
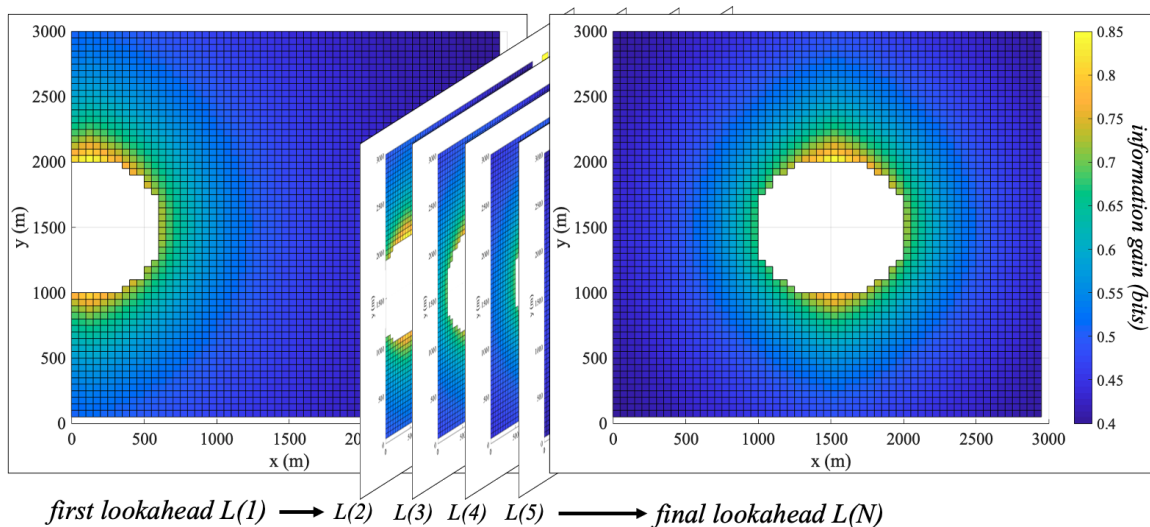
Figure 5.3: Predictive payout maps for $L = N$ look-ahead levels. The estimate of the target position at each look-ahead is at the center of the white region that is the stand-off distance to the target. The heat-maps shows the predicted levels of information gain (bits) for each cell (per look-ahead level), which decreases as a function of range from the target. Maximum information gain is found next to the broadside of the target where TS is maximum.

The important features to note in Fig. 5.3 is that the information payout (and SNR) is maximum next to the broadside of the target where the target strength level TS is maximum. As expected, the information gain is also a function of range to the target.

Q-Learning and Multi-Robot Goal State Selection

Similar to Algorithm 2 (Fig. 4.6), Q-Learning is applied here to generate path-planning policies with information payout as the reward function. The modified Q-Learning algorithm, denoted 'Algorithm 3' in Fig. 5.5, plans trajectories for the three collaborating UUVs to track the target. The major difference here is the information payout changes at each time-step, or each time a new action is selected, during the learning process. The reward matrix is updated until a terminal look-ahead is reached, which is used for the remainder of the learning process. The number of look-ahead levels is therefore a tunable parameter. Note that if a UUV is located outside of the focused payout map, then the starting state s is set to the nearest cell. The learning environment here is no longer deterministic as the rewards in each cell change after each action. The reward gain = 6 and discount factor $\gamma = 0.5$ were carried over from the previous sensitivity study in Chapter 4. The learning rate was set to a constant $\alpha = 0.1$ and $\epsilon = 0.4$ which are values often used in practice [71]. In most applications $\alpha = 0.1$ is a sufficiently small step-size that does not significantly increase the time required for the learning process. An ϵ-greedy approach with $\epsilon = 0.4$ is practical as it balances exploration and exploitation by choosing between two methods randomly. Future work could conduct a sensitivity study to determine optimal parameters. Expectation is that the results would not vary in a statistically significantly manner.

The problem of goal state selection is expanded to the three-vehicle problem. The purpose is to position the vehicles in the best possible configuration in the the final look-ahead L_N. A simple selection process is used that balances the desire for high rewards while maintaining space between the three UUVs to achieve the best possible trilateration solution. The result of the selection process is a triangular configuration of goal states that is illustrated in Fig 5.4. Note that the solution is for a target travelling parallel to the x-axis.
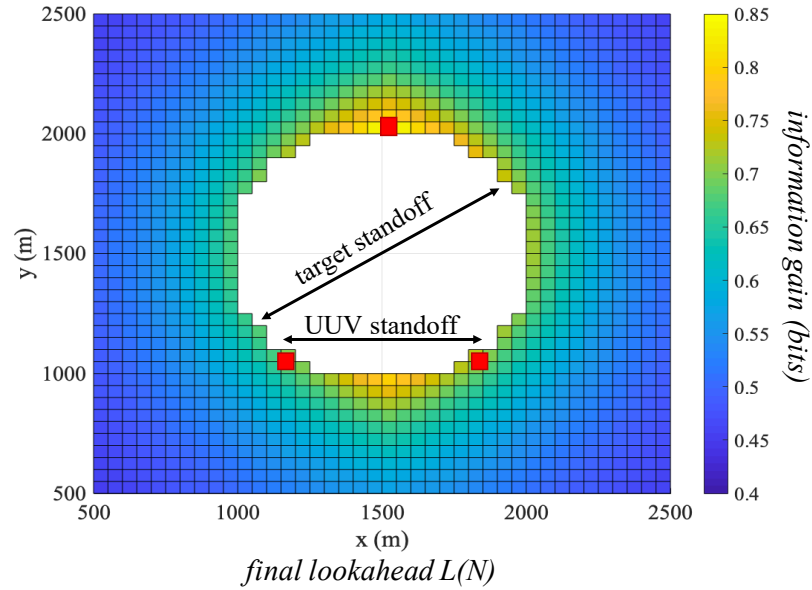
final lookahead L(N)

Figure 5.4: Outcome of goal selection process where the 3 goal states maximize information payout with the imposed stand-off constraints. Goal states are shown in red, and the related stand-off parameters are labelled.

Algorithm 3 Q-Learning Algorithm with Predictive Information Gain Reward Map

---

Output: Q-table, Q(s,a), with optimized q-value for each state-action pair

---

Require:
    $States: S = \{1, \dots, n_s\}$
    $Actions: A = \{1, \dots, n_a\}$
    $Reward\ matrix: R : S \times A \times L \quad (Information\ payout\ at\ each\ lookahead)$

Begin:
    for j = 1: N (increment thru *N=3* UUVs)
        $initialize\ Q : S \times A\ randomly$
        while Q *is not converged*
            $start\ in\ state\ s\ \in S$
            $l = L_1$
            while *s is not the goal state*
                $choose\ a\ based\ upon\ greedy - exploration\ strategy$
                    $a \leftarrow random\ a\ \in A_s \dots (exploration)$
                    OR
                    $a \leftarrow argmax_a Q(s,a) \dots (exploitation)$
                    $r \leftarrow R(s,a,l)$
                    $s' \leftarrow s + a$
                    $Q(s',a) \leftarrow (1-\alpha) \cdot Q(s',a) + \alpha \cdot (r + \gamma \cdot max_{a'} Q(s',a'))$
                    $s \leftarrow s'$
                    $l \leftarrow l + 1$
    return Q

Figure 5.5: Proposed *Q*-Learning algorithm with predictive information gain rewards based on target state look-ahead levels.

Given the proposed algorithm, the last decision for *mRobot* to manage is when a re-plan (new trajectories) should be conducted. Ultimately, three criteria were chosen that trigger a re-plan:

1. if the velocity-estimate change (% difference) between a planning instance is greater than a threshold;

2. if the heading-estimate change (% difference) between a planning instance is greater than a threshold, and

3. if a time limit is reached.

A more optimal solution to determine when a re-plan is warranted can be explored in future work.

The entire autonomous planning and task allocation system, as it is implemented within the *mRobot* ROS node, is summarized by Fig 5.6.
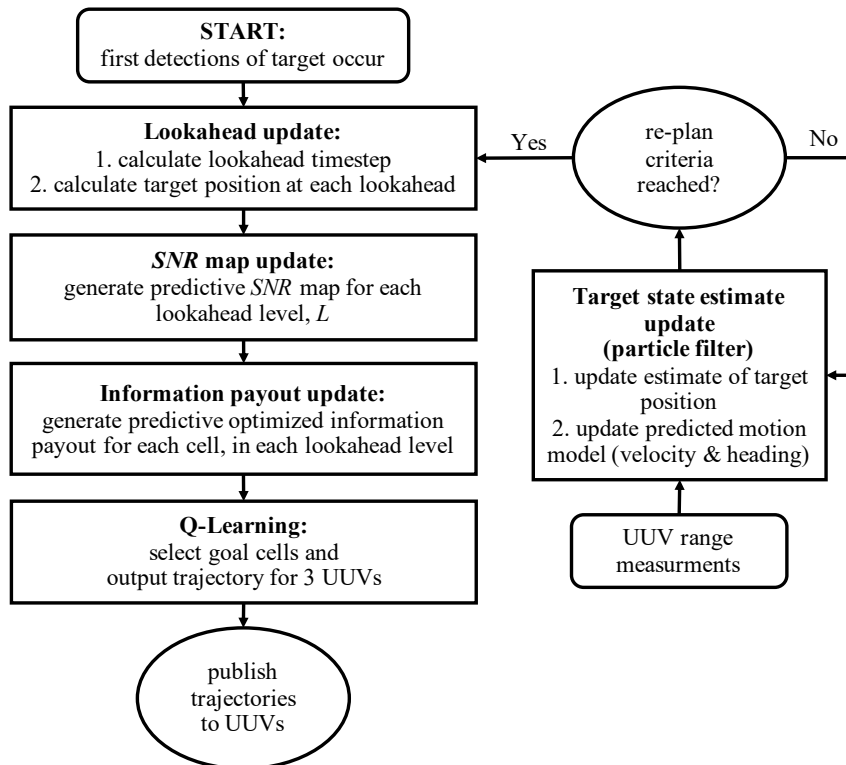
Figure 5.6: Autonomous target state-estimate and tracking algorithm implemented in the *mRobot* ROS node for multi-UUV collaboration.

There are several important parameters of the proposed solution that have been introduced. Certain parameters were tuned during development, and from previous studies. A complete sensitivity study was not completed. A list of the parameters and the values used to obtain the results in this work are listed in Table 5.1.

Table 5.1: Parameters of proposed autonomous planning algorithm

reward function	max lookahead levels L_N	7
	payout map dimensions	$3\ km \times 3\ km$
Q-Learning	discount factor γ	0.5
	reward gain	6
	ϵ	0.4
	α	0.1
replanning criteria	velocity change threshold (% difference)	20
	heading change threshold (% difference)	20
	time limit (s)	$L_{N-2} \times ts$

5.2   Simulated Environment

The ECA A9 AUV and the UUV Gazebo simulator tools were used again to evaluate the proposed path-planner's effectiveness in a controlled underwater environment through simulations [27]. Each of the three collaborating assets and the target were simulated by the ECA AUV simulator package. The following section describes the simulated search region and acoustic environment.

5.2.1   Acoustic Environment

In this chapter, a few simplifying assumptions are made to the acoustic environment. Firstly, the search area of interest in this chapter is the Bedford Basin (an inland body of water near Halifax, Canada). The basin has a maximum depth of 71 m and its bathymetry is shaped like a bowl with its deepest point close to the geometric center [72]. However, in simulations the acoustic channel is simplified to a constant depth of 40 m. Secondly, the target of interest is a Victoria Class submarine which

has a 70.86 m length, 7.6 m beam, and 5.5 m draught [73]. The target is assumed to be a rectangular prism for the purpose of reducing the complexity when computing the target cross-section σ. Thirdly, α is assumed constant with all absorption model parameters fixed at $T = 10^{\circ}$C, $S = 35$ ppt, and $pH = 8$ [14]. The range in temperature of a body of water can be large, where 10°C resides near the accepted yearly average of the Atlantic Ocean. The range of salinity and pH are less significant, and the values selected here match the results found in [74] for a similar 70 m depth environment. Furthermore, if the SNR for one-way-travel communication is above 1 then it is assumed that acoustic packets will be transmitted and received reliably.

Figure 5.7 plots the SNR in dB versus range using assumptions detailed in this subsection. The source level of a target submarine can vary depending on its relative aspect to the UUV. A typical value for the broadside of the submarine is 25 dB [57], which was used for generating the plots in Fig. 5.7.
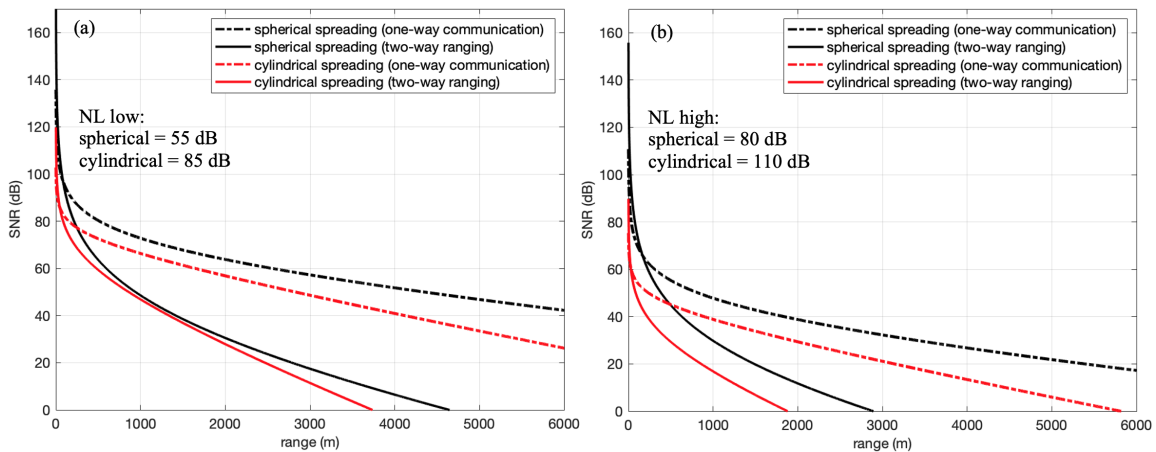
Figure 5.7: SNR versus range for both two-way acoustic ranging and one-way communications for an acoustic environment with: (a) lower NL (85 dB) and (b) higher NL (110 dB). Spherical and cylindrical spreading are modelled for comparison.

The respective noise levels in Fig. 5.7 were selected based on the communication range provided by cylindrical wavefront model. In the low noise case (85 dB) the communications go to the horizon at sea level ($SNR > 1$ at 6 km), while they do not for the higher noise level of 110 dB. In continuation, the noise level also impacts the target detection range. These respective noise levels will be applied to test the proposed system's ability to track a target in relatively high and low noise environments. The noise levels are higher for the cylindrical spreading model to

capture the additive noise from reverberation in a shallow environment. Environments with both types of spreading and ambient noise levels are studied in this chapter, while the spherical model is the baseline case. The ambient noise levels in an actual ocean environment would vary spatially and with time.

Lastly, simulated UUV acoustic sensors are implemented to measure the TWTT to the target based on a sound speed of 1500 m/s. The TWTT measurements are subject to additive zero-mean Gaussian noise with 1 m variance. The detection outcome is determined by applying the probabilities P_{D_l} and P_{F_l} from the optimization process of Eq. 4.30. Delays due to time-of-flight for the range measurements and communications between assets are modelled.

5.2.2 3-UUV Configurations for Target Tracking in Confined Spaces

The geometry of the Bedford Basin model used in simulation is shown in Fig. 5.8.
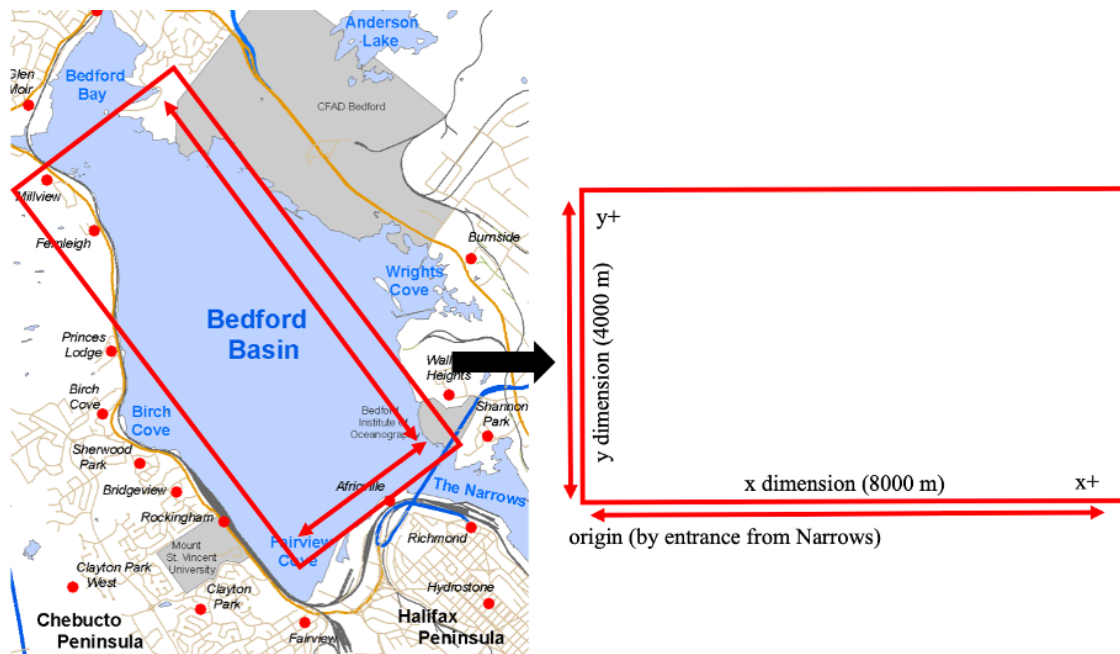
Figure 5.8: Simulated environment based on the Bedford Basin (Halifax, Canada). The coordinate system is defined with the x-dimension on the long side of the environment and the origin at the entrance to the basin from the Narrows.

Parameters of simulated environment were varied to test and compare the 3-UUV collaborative system with the standard stationary LBL solution. The placement of

the three LBL beacons (at the start positions of the UUVs) was varied, along with the target velocity, target trajectory, and ambient noise level NL.

The two 3-UUV configurations (yellow, green) and the three target trajectories considered in simulations are shown in Fig. 5.9. Tables 5.2 and 5.3 describe the configurations and target trajectories. Note that communication between assets is possible (Fig. 5.7) for the initial configurations. However, the target cannot be reliably detected at all points on the trajectory lines. The *mRobot* node is located on the lead UUV, labelled UUV 1, which subscribes to the state and measurements of the other UUVs to conduct the mission-planning.
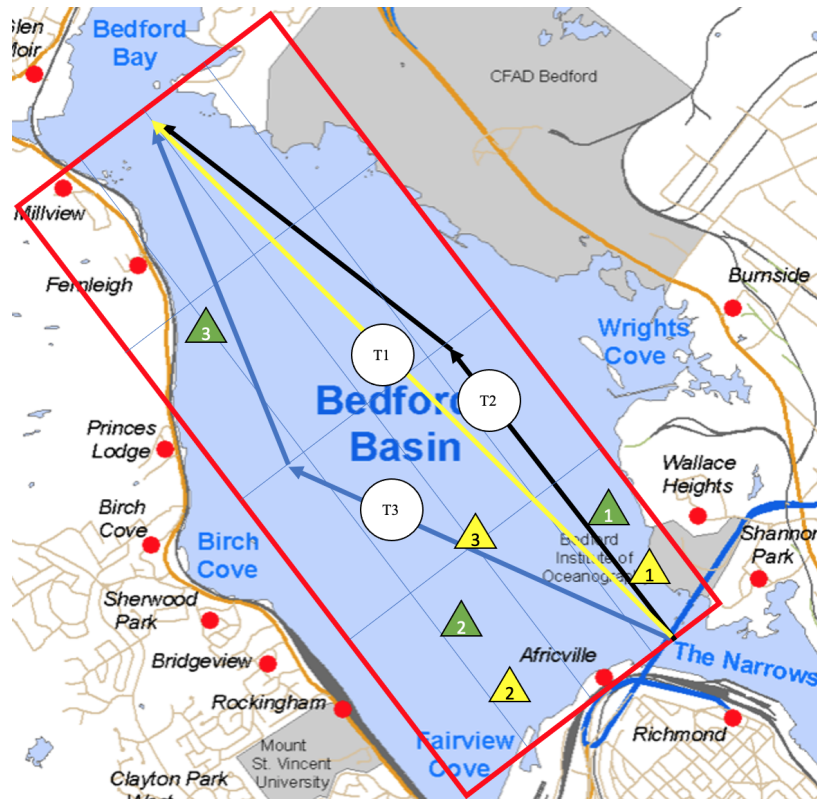
Figure 5.9: Bedford Basin showing the environment boundary and locations of LBL beacons / initial UUVs' positions. Yellow triangles represent configuration #1 that provide a baseline set-up where the beacons and UUVs are within range of the target at the start of the mission. The green triangles show configuration #2 which disperses the beacons and UUVs to provide greater coverage of the environment. The numbers within each triangle identify the number for the different beacons/UUVs. The three target trajectories for the simulations are labelled T1, T2, T3. The grid-lines within the rectangle are 1 km (y axis) and 2 km (x axis) spacing lines for reference.

Table 5.2: Coordinates of LBL beacon and UUV start positions for the two configurations as shown in 5.9.

configuration #	beacon/ UUV #1 initial pose (x, y) (m)	beacon/ UUV #2 initial pose (x, y) (m)	beacon/ UUV #3 initial pose (x, y) (m)
1 *(yellow)*	(700, 700)	(700, 2700)	(2000, 2000)
2 *(green)*	(1500, 700)	(1500, 2700)	(5700, 3000)

2

Table 5.3: Target trajectories studied in simulation

trajectory	initial heading (°) *(measured from positive x-axis)*	heading change (°) *(at x = 4000 m)*
T1	9	0
T2	0	port 17
T3	32	starboard 34

5.3   Results and Discussion

Trials matrix 1 (Table 5.4) details the baseline simulations where the beacons are placed close to the basin entrance from the Narrows such that the target is more likely to be detected early (yellow, configuration 1). The spherical spreading acoustic model is used for configuration 1. Trials matrix 2 (Table 5.5) represents a more realistic and challenging set-up with a cylindrical spreading acoustic model. The distribution of beacons in configuration 2 are representative of a set-up more suitable for the LBL system (green, configuration 2).

The average root-mean-square error (RMSE) is reported to compare the collaborative and LBL tracking solutions in each trials matrix. The RMSE measures the difference between the estimated and actual target state at one second intervals, which is averaged over the duration of the mission. Note, the LBL solution uses the same target state-estimation algorithm and sensor model as the 3-UUV collaborative system. The combined set of trials are not an exhaustive set of simulations. Representative examples from both trials are presented and discussed next. All the simulation trials results are in Appendix A.

Table 5.4: Trials matrix 1 describes the simulation parameters for trials 0-18. The average RMSE is reported for the collaborative UUV and LBL solution for comparison. Superior performance is reported for the 3-UUV system for all trials. Error for both solutions increases with increasing ambient noise and target velocity.

trial #	target		UUVs		acoustic environment		avg. error (RMSE)		configuration
	velocity (kn)	trajectory #	initial position (conf. #)	velocity (kn)	one-way SNR at 6km (spherical wavefront)	NL (dB)	LBL solution (m)	UUV solution (m)	
0 (ref.)	2	T1	1	0	42.2	55	58.2	61.9	
1	2	T1	1	2	42.2	55	56.7	8.5	
2	2	T2	1	2	42.2	55	65.6	9.6	
3	2	T3	1	2	42.2	55	71.7	9.8	
4	3	T1	1	2	42.2	55	101.5	14.8	
5	3	T2	1	2	42.2	55	113.4	14.8	
6	3	T3	1	2	42.2	55	111.5	16.4	
7	4	T1	1	2	42.2	55	143.7	42.8	
8	4	T2	1	2	42.2	55	153.9	40.7	
9	4	T3	1	2	42.2	55	121.6	40.6	
10	2	T1	1	2	17.2	80	86.3	43.4	
11	2	T2	1	2	17.2	80	83.9	38.1	
12	2	T3	1	2	17.2	80	95.6	42.2	
13	3	T1	1	2	17.2	80	120.4	63.4	
14	3	T2	1	2	17.2	80	143.0	77.3	
15	3	T3	1	2	17.2	80	129.4	76.2	
16	4	T1	1	2	17.2	80	224.2	89.2	
17	4	T2	1	2	17.2	80	148.1	101.7	
18	4	T3	1	2	17.2	80	252.8	121.5	

Table 5.5: Trials matrix 2 describes the simulation parameters for trials 19-37. The average RMSE is reported for the collaborative UUV and LBL solution for comparison. Significantly lower average error is reported by the 3-UUV system for all trials. The cylindrical spreading model with increased acoustic noise lead to higher errors reported for both solutions. The trends of increasing error with increasing noise and target velocity are preserved.

| trial # | target | | UUVs | | acoustic environment | | avg. error (RMSE) | | configuration |
	velocity (kn)	trajectory #	initial position (conf. #)	velocity (kn)	one-way SNR at 6km (cylindrical wavefront)	NL (dB)	LBL solution (m)	UUV solution (m)	
19 (ref.)	2	T1	2	0	26.1	85	93.4	95.3	
20	2	T1	2	2	26.1	85	101.6	27.3	
21	2	T2	2	2	26.1	85	85.8	19.0	
22	2	T3	2	2	26.1	85	75.2	18.0	
23	3	T1	2	2	26.1	85	122.6	25.7	
24	3	T2	2	2	26.1	85	120.5	33.7	
25	3	T3	2	2	26.1	85	122.9	44.6	
26	4	T1	2	2	26.1	85	143.6	57.2	
27	4	T2	2	2	26.1	85	118.3	75.9	
28	4	T3	2	2	26.1	85	150.8	92.5	
29	2	T1	2	2	-1.3	110	125.7	30.2	
30	2	T2	2	2	-1.3	110	126.2	31.6	
31	2	T3	2	2	-1.3	110	164.4	31.1	
32	3	T1	2	2	-1.3	110	177.4	54.7	
33	3	T2	2	2	-1.3	110	223.5	56.2	
34	3	T3	2	2	-1.3	110	177.4	64.5	
35	4	T1	2	2	-1.3	110	257.8	134.3	
36	4	T2	2	2	-1.3	110	303.4	195.5	
37	4	T3	2	2	-1.3	110	241.4	97.1	

Results from configuration 1 trials are discussed first. The results from Table 5.4 (configuration 1) show the proposed 3-UUV collaborative system is a notably better solution for the simulation parameters in trials 1-18. Not unexpectedly, the RMSE for both increases with increasing ambient noise NL and target velocity.

Figure 5.10 plots the trajectories of all four vehicles from six trials in the 55 dB noise environment (trials 1-3, 7-9), which demonstrate the collaborative 3-UUV system response to the target. Being able to track the target and adapt to evasive manoeuvres was one of the main objectives. As shown in Fig. 5.10, the 3-UUV collaborative system was able to track and maintain a stand-off from the target when it travelled at the same velocity (trials 1-3). When the target travelled twice as fast (trials 7-9), the UUVs were still able to track the target and adapt to the heading changes at 4000 m. However, the UUVs eventually fall out-of-range. The trajectory plots for the other trials in Table 5.4 are not shown as they do not differ significantly from Fig. 5.10.

Figure 5.11 plots the target state estimate-error versus time for the six trials shown in Fig. 5.10. With the slower moving target (2 knots), the collaborative UUV system was able to track and localize the target with low error for the mission duration (trials 1-3 in Fig. 5.11). The LBL solution eventually falls out of detection range and the estimate error grows unbounded after 4000 m. Configuration 1 was at a disadvantage for both the LBL and 3-UUV collaborative solution, as the error generally increases with time. Specifically at high target velocity (4 knots), the 3-UUV collaborative system is chasing the target early and is not able to provide a consistent estimate for the full duration (trials 7-9). However, the advantage provided by the collaborative system is clear.

Figure. 5.12 provides the estimate-error plots for the same six trials in the 80 dB noise environment (trials 10-12, 16-18). In these high noise trials, both solutions provide a higher average error —as expected. The collaborative UUV system bounds and maintains a relatively low error for the duration of trials 10-12, and again shows improvement over the LBL system for the faster moving target in trials 16-18. In the later three trials, the average RMSE for the collaborative UUV solution was 89.2 m, 101.7 m, and 121.5 m, which is quite high relative to the other fifteen trials. This result can be explained by the presence of higher ambient noise and target velocity.

Under such conditions the probability of making a valid detection is lower, and the UUVs have less time to respond to the faster moving target. For both the high and low noise cases, the different target trajectories do not significantly affect the 3-UUV system. However, in the high noise trials the LBL solution is not able to accurately update its estimate of the target after it makes a heading change —a function of the reduced detection range and the beacon configuration. This result is less significant at lower target velocities when more measurements are possible for the LBL solution due to more favourable sensing conditions and a longer measurement period.
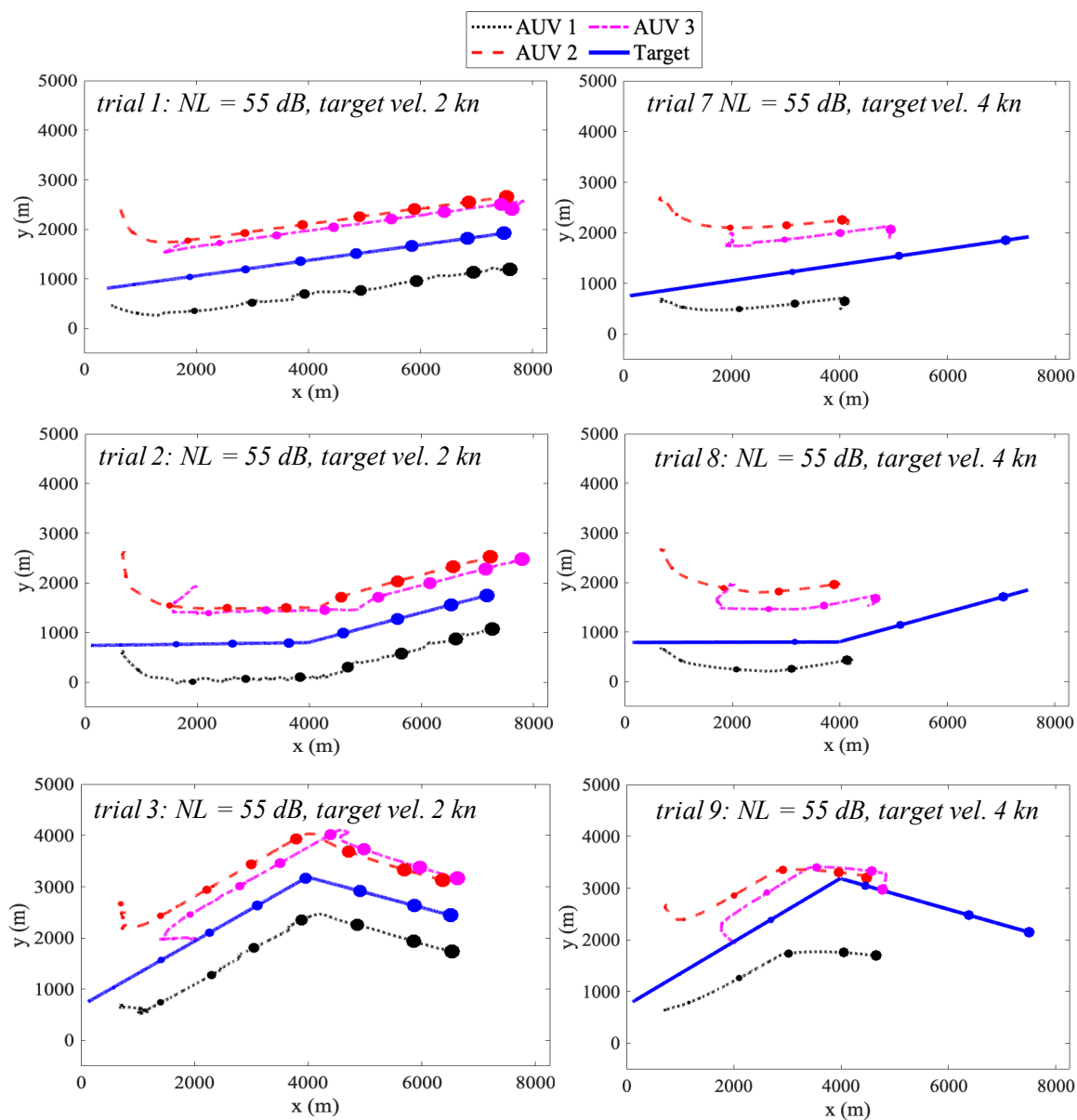
Figure 5.10: Trajectories of the collaborating UUVs adapting to the target over the tracking mission duration. Six trials are shown from configuration 1, where trials 1-3 are with a target velocity of 2 kn (or 1 m/s) and 7-9 are with a target trajectory of 4 kn (or 2 m/s). The increasing dot size on the trajectory lines provides a reference for the passage of time. The UUV system tracks the target successfully for the full duration of trials 1-3. In trials 7-9 the faster moving target does outrun the UUVs; however, the UUVs are still shown to track and adapt to changes in the target trajectory.
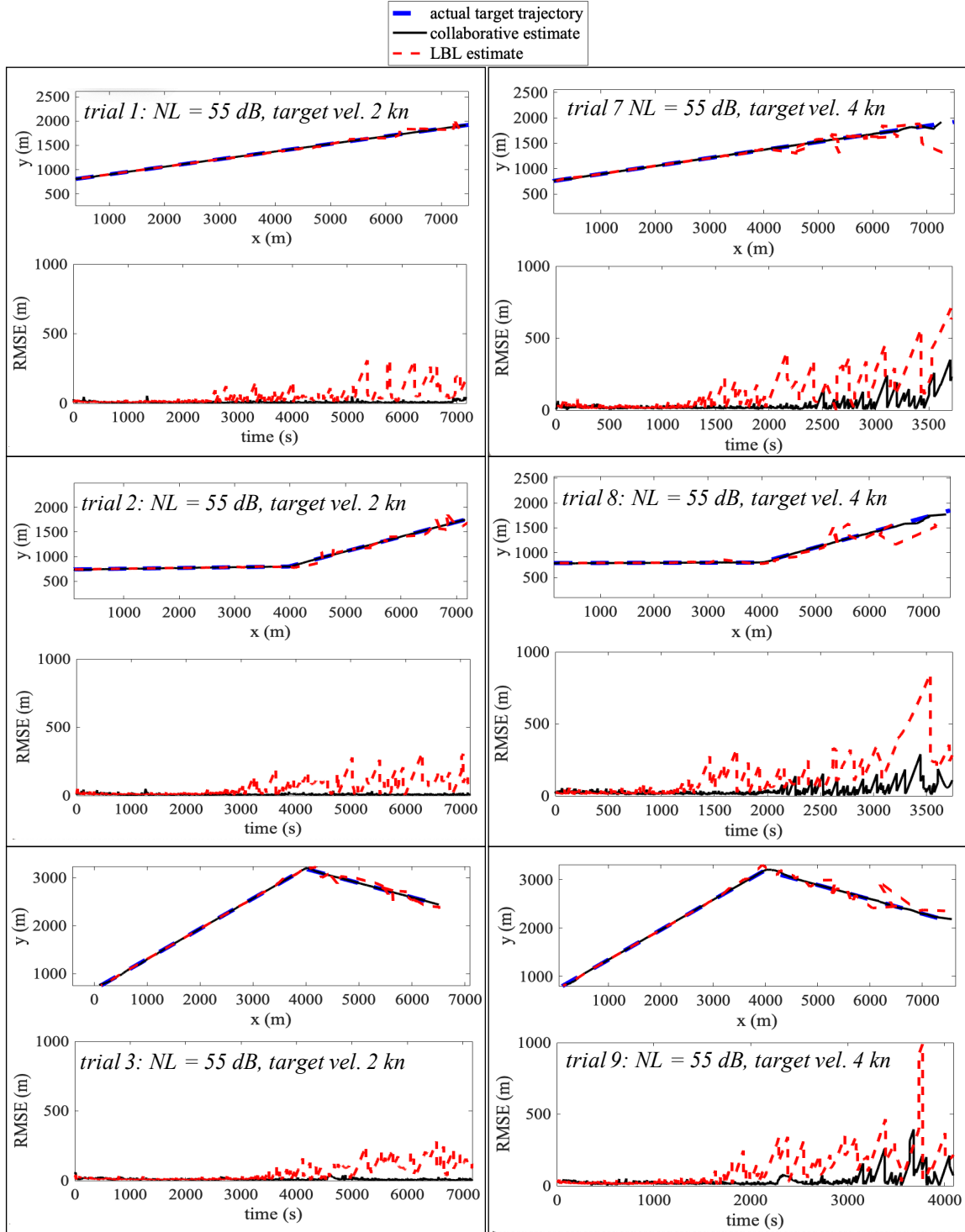
Figure 5.11: Comparison of 3-UUV versus LBL target state estimate tracking solutions over the mission duration, and the RMSE versus time for trials 1-3 and 7-9 in configuration 1 (trials detailed in Fig. 5.10). The 3-UUV system provides a lower average state estimate than the LBL for all trials —as the LBL solution falls out of range at approximately 4000 m.
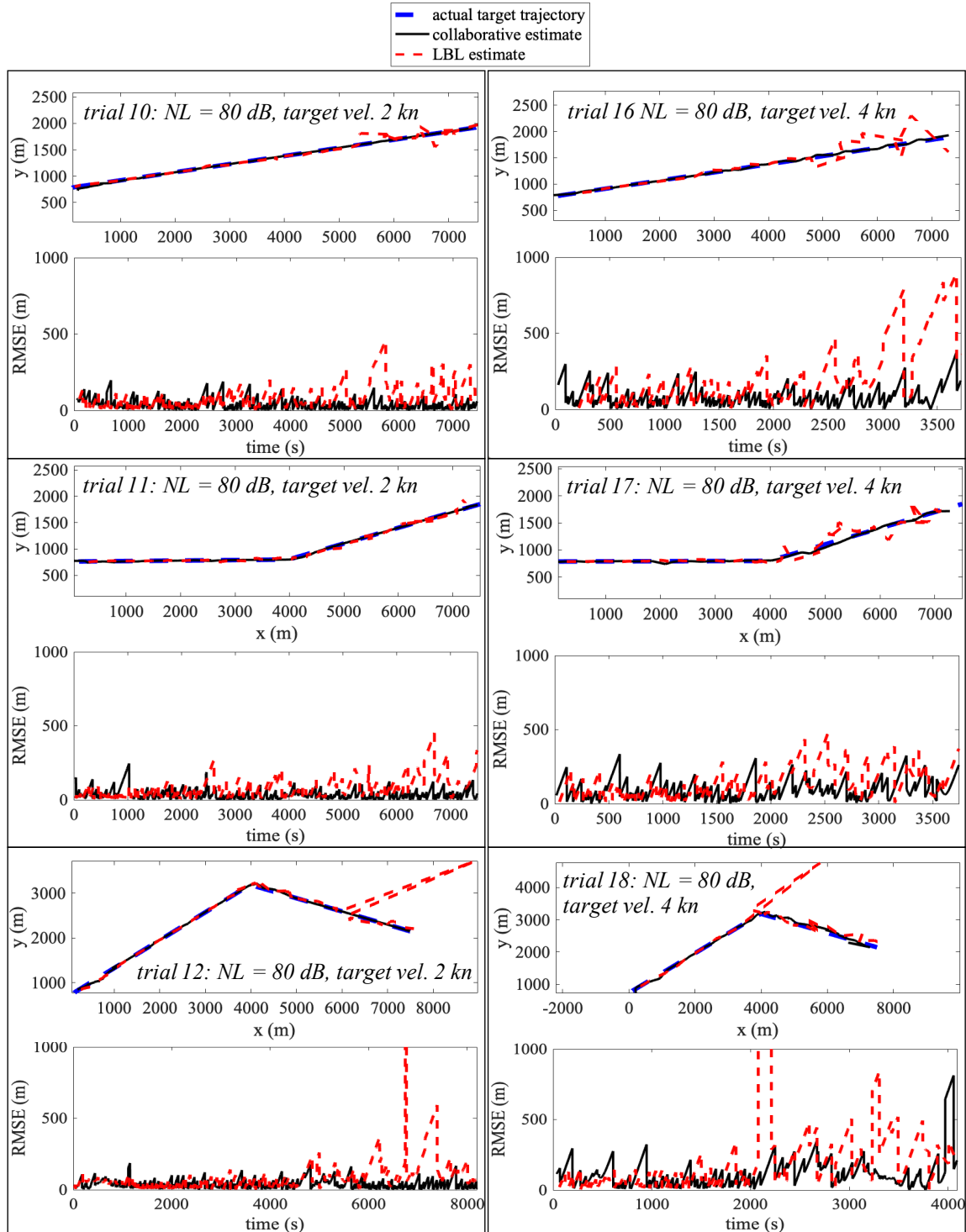
Figure 5.12: Comparison of 3-UUV versus LBL target state estimate tracking solutions over the mission duration, and the RMSE versus time for trials 10-12 and 16-18. These trials have the same parameters as the trials in 5.11 with the exception of increased noise (80 dB). Higher error is provided by both solutions in the higher noise conditions, where the 3-UUV system again provides an improvement over the LBL for all trials.

Results are now discussed for the configuration 2 trials, corresponding to the green triangles in Fig. 5.9. The average RMSE reported in Table 5.5 present a similar trend to the first set of trials when comparing the LBL and 3-UUV collaborative solutions. The total average error is larger for both systems due to the more difficult acoustic environment: cylindrical wavefront spreading and added noise due to reverberation. The 3-UUV collaborative solution again provides a better solution for all trials.

Towards tracking the target, the trajectories in Fig. 5.13 illustrate the advantage of the UUVs starting further downstream from the basin entrance in configuration 2. With a lower target velocity (2 knots in trials 20-22), the UUVs navigate closer to the entrance of the basin to improve the detection probability, and then continue to track the target for the duration of the mission. With a higher target velocity (4 knots in trials 26-28), the UUVs can still fully adapt to the evasive manoeuvres of the target and maintain a good stand-off. The target does start to outrun the lagging two UUVs in the last kilometre of the missions. Again, the trajectory plots for the complete set of trials in configuration 2 are not provided as they do not provide any further insight than in Fig. 5.13.

The sample of low ambient noise (85 dB) trials in Fig. 5.14 provide insight into the challenges brought on by the more difficult acoustic environment, and dispersed beacon configuration. In the trials with low target velocity (2 knots in trials 20-22), the LBL solution is unable to develop a strong estimate of the target pose prior to approximately 2000 s, when the target comes into range. The 3-UUV collaborative solution is challenged in a similar manner; however, an improvement is seen as the error is minimized approximately 500 s earlier than the LBL solution. Note, these times associate with a target location of approximately 1500 m and 2000 m in the x-axis for the 3-UUV and LBL solutions, respectively. The UUVs' abilities to mobilize and head towards the target early in the mission provides the advantage seen here. Consequently, the UUVs can also maintain and bound the estimate error for the remainder of the missions. The LBL solution is only able to provide a window from approximately $2000 - 4500$ m where the error is bounded and equivalent to the 3-UUV collaborative solution. Outside of that window, the target appears to leave the region covered by the stationary beacons. The results are similar when the target is in transit at 4 knots, with the exception that the error is larger for both tracking

methods at the start and end of the trials (trials 26-28). The increased error observed is of course expected, as the detection window is shorter and the delay in the TWTT measurements and communications is more significant when the target travels with higher velocity. However, the 3-UUV collaborative solution maintained a relatively small bounded error through to the end of each trial. The target trajectory again does not appear to play a significant role in either solution, with the exception of trial 28 where there is a bump in estimate error at 4000 m (2400 s) when the target changes its heading. Both solutions are able to recover and detect the large heading change in trajectory T3. The LBL solution falls out-of-range shortly after. In general, the LBL system provides a better estimate in trials where the target follows trajectory T3, relative to the other two trajectories. This result can be explained by the asymmetrical configuration of the beacons, which is more ideal when the target navigates closer to centroid of the triangle of beacons (as in T3) prior to making a heading change.

Continuing with results from configuration 2, a sample of the high noise (110 dB) environment trials are shown in Fig. 5.15. The higher ambient noise and reverberation in the acoustic environment increases the average error for both solutions, which again affects the LBL solution more. As in the other trials, the UUVs navigate to reduce their respective ranges to the target where the sonar SNR is larger and detections can be reliable, even in a high-noise environment. However, an increase in noise (and a decreased SNR) directly reduces the detection range of a stationary beacon, and the LBL system performance decreases in a proportional manner. Observing the low velocity trials (29-31) first, the mobile UUVs can bound and reduce the estimate error after approximately 2000 s, while the LBL solution is poor throughout. The high velocity trials (35-37) approach the failure point for both systems; however, the 3-UUV collaborative system was able to provide short durations with low-error. Here, the target trajectory does have a more significant impact on the results. The most significant impact can be observed in comparing trials 36 and 37 (T2 and T3). In trial 37, the target heads towards the 3rd beacon/UUV and both systems see an improved average estimate. In trial 36, the target maintains a larger stand-off from the 3rd beacon which increases the error for the LBL solution. The error is also higher for the collaborative solution in trial 36 as the system has less time to adapt to the target

and improve the measurements.
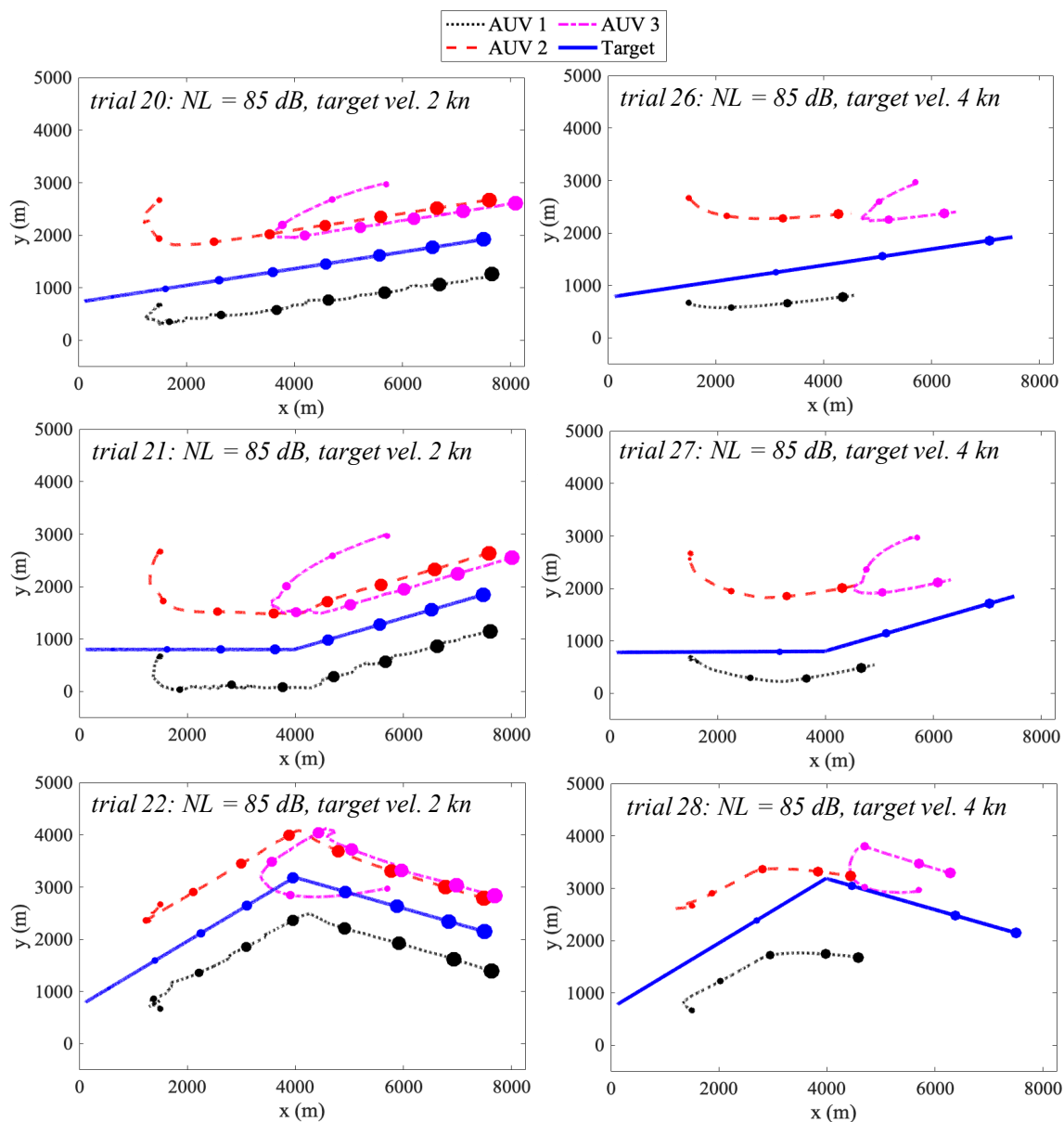
Figure 5.13: Trajectories of the collaborating UUVs adapting to the target over the mission duration. Six different trials are shown from configuration 2, where trials 20-22 have a target velocity of 2 m/s (or 1 kn) and 26-28 are with a target trajectory of 4 m/s (or 2 kn). The increasing dot size on the trajectory lines provides a reference for the passage of time. The UUVs are able to track and adapt to the target throughout the six trials. Specifically, in trials 26-28 the UUVs are still able to track the target travelling at a higher velocity.
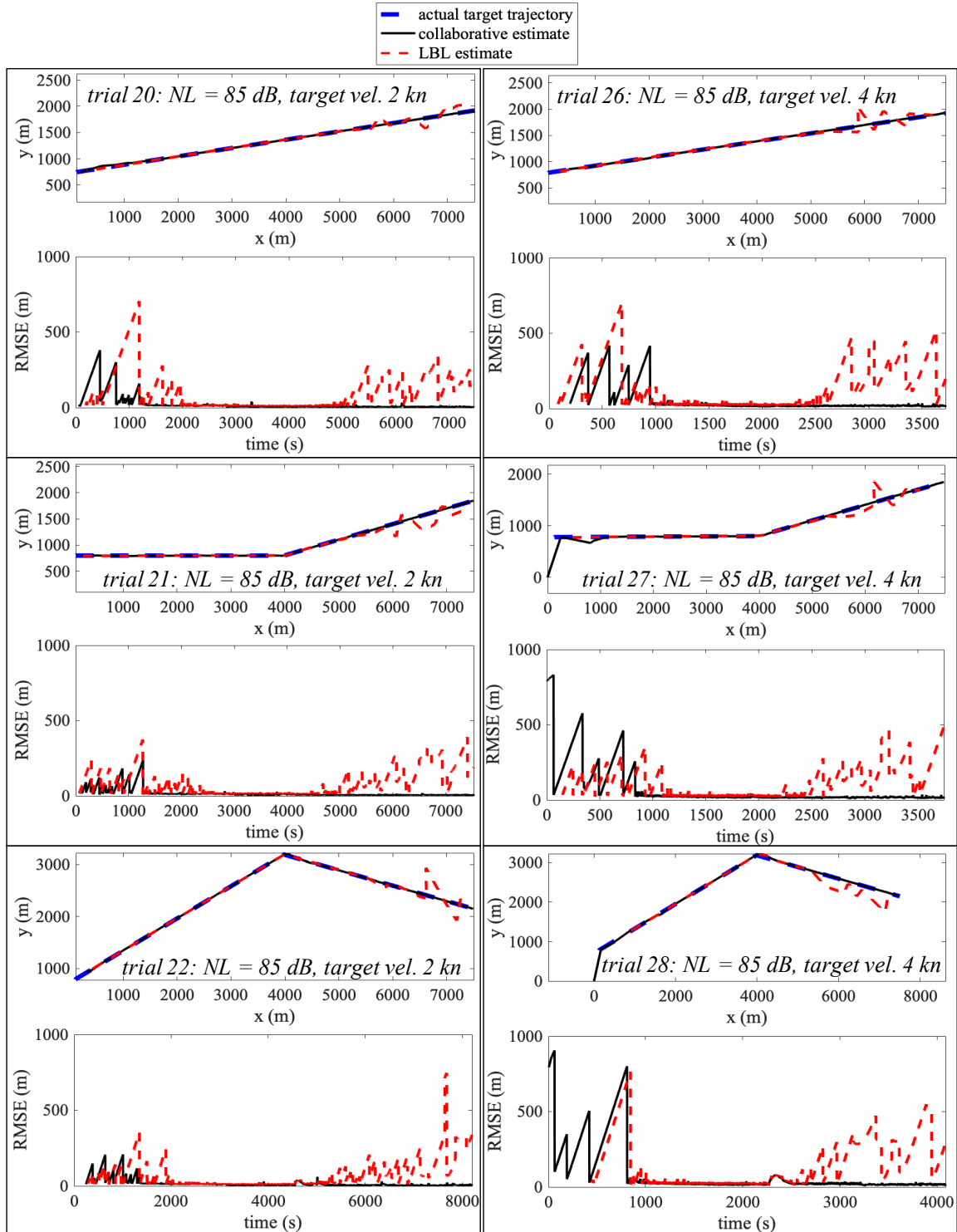
Figure 5.14: Comparison of 3-UUV versus LBL target state estimate tracking solutions over the mission duration, and the RMSE versus time for trials 20-22 and 26-28 (trials detailed in Fig. 5.13). The 3-UUV system accurately tracks the target for all trials after an initial duration when the UUVs converge to the target. The LBL solution provides an accurate estimate of the target pose for a shorter period of time in all trials.
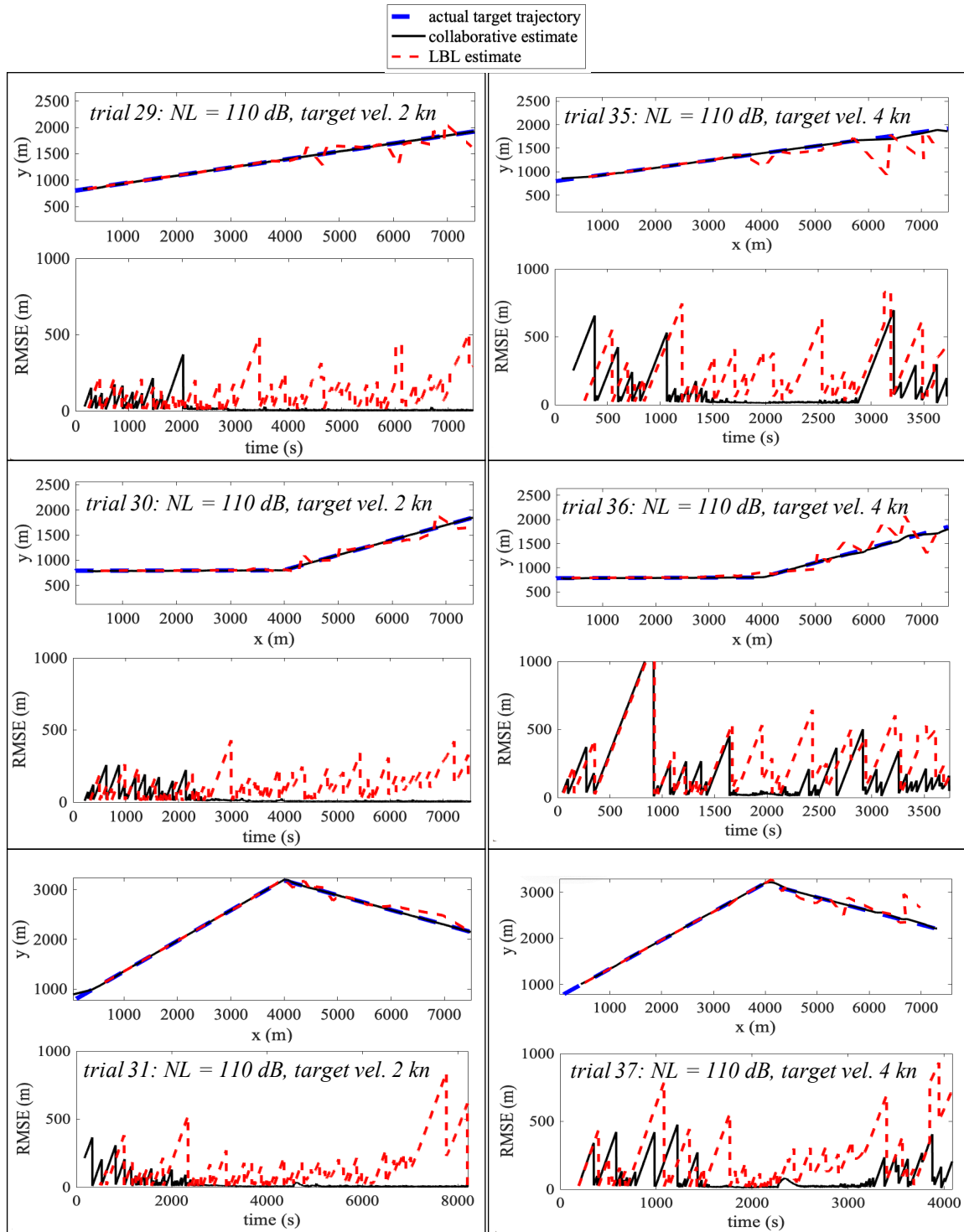
Figure 5.15: Comparison of 3-UUV versus LBL target state estimate tracking solutions over the mission duration, and the RMSE versus time for trials 29-31 and 35-37. These 6 trials have the same parameters as the trials in 5.14 with the exception of increased noise (110 dB). The significance of these results is again the superior performance of the 3-UUV system. Trials 35-37 challenge both solutions, where the LBL system is ineffective. The UUVs are able to provide short windows of accurate estimates in trials 35-37.

5.4 Summary

An autonomous trajectory planner was developed towards underwater evasive target-tracking and localization. The information measure was adapted to the acoustic channel for range only measurements of a mobile target. Predictive look-ahead information payout maps were developed as the reward for Q-Learning. Particle filters performed the target state (pose) estimation and thus a prediction for the target state with increasing look-ahead levels.

The simulated environment captured the acoustic propagation conditions for communication and sensing based on the the sonar equation with assumed levels for the ambient noise. Results from 38 simulated trials cases were presented with variable target velocity, target trajectory, acoustic noise, and sensor configuration. The results were used to compare the tracking solution for the proposed 3-UUV system and the standard LBL method.

The proposed 3-UUV collaborative system can track the target through a simulated Bedford Basin environment and provide better tracking than the LBL solution for all trials cases. The 3-UUV system was able to navigate in order to increase the target detection likelihood and ultimately provide superior performance. In trials with increasing target velocity, the collaborative UUV system was able to react and extend the window where target detections were reliable. As expected, the ability of the mobile UUVs to optimize their position to sense the target provided a significant advantage. Higher ambient noise and target velocities decreased the accuracy of both solutions, where the final trials, 35-37, significantly challenged both systems.

Chapter 6

Conclusions and Future Work

This thesis studied autonomous multi-robot path-planning and task-allocation towards gaining information on three types of marine targets with increasing uncertainty and mobility. Firstly, a collaborative robot system was proposed as a method towards surveying and creating 3D representations of stationary and partially-submerged targets. Results from simulation and controlled in-water testing were reported, which demonstrated the ability of the $mRobot$ ROS node to plan and distribute tasks to a network of above-and-below-water robots. The contribution was the development and validation of the $mRobot$ node. Secondly, towards searching for underwater targets, an autonomous algorithm was developed that combines optimized multi-pass information measures with reinforcement Q-Learning as a method for trajectory planning. The $mRobot$ node was used to manage the planning and distribution of waypoints to a UUV during search missions. A sensitivity study was completed, which illustrated that the value of reward gain and discount factor γ can play a significant role in changing the trajectory of the UUV. In simulated NMCM target search missions, the Q-Learning based planner showed improved information gain performance compared to the boustrophedon and greedy approaches. This improvement was observed for both homogeneous and variable SNR environments. Lastly, a collaborative UUV system was developed as a method to detect, track, and localize unresponsive targets. The proposed system built on the previous chapters by adapting the information measure into predictive look-ahead information payout maps. The predictive information measures were used as the reward function for Q-Learning with three UUVs. The $mRobot$ ROS node was used again as a tool to manage the planning and task-allocation. As expected, increasing acoustic noise and target velocity reduced the effectiveness of the 3-UUV system. However, the collaborating UUVs outperformed the LBL solution for all of the trials considered in controlled simulations.

In conclusion, the robotic path-planning and task-allocation requirements varied

significantly given static targets of known pose, static targets of unknown pose, or mobile targets of unknown pose. Firstly, when surveying a floating target the main challenge was communication and planning in multiple-domains. The planning was aimed at overcoming constrained communication with the UUV. However, as the pose of the static target was known prior to the mission, neither complex path-planning nor tightly-coupled task-allocation was required to complete the survey. Secondly, searching for underwater static targets of uncertain pose required planning that could adapt to the continuous flow of information obtained during the search. In continuation, dynamic planning that considered variability of environmental conditions and uncertain detection outcomes was crucial to maximizing search performance underwater. Lastly, tracking a mobile target of unknown pose required a combination of dynamic state-estimation and tightly-coupled collaboration between the three UUVs. Tightly-coupled task-allocation was essential to ensure each UUV maintained a stand-off with both the target and its team-mates to minimize the target state estimate error.

There are several considerations for future work in this thesis. Firstly considering Chapter 3, the project will take the heterogeneous collaborating robots into a larger in-water environment to test the underwater acoustic communications. A larger environment will also allow for experimental validation of the USV path-planning in support of its team-mates, as well as quantifying the benefit of its role in the system. With regards to the merging of above- and below-water models, more complex and larger targets will be trialled. Secondly, immediate future work is conducting a more exhaustive set of simulations to better understand the capability and limitations of the collaborative UUV system for underwater target tracking. Different target sizes, velocities, and trajectories will be explored. Thirdly, collaborative localization was introduced as a method for improving and bounding the localization error of UUVs; however, the state-estimate error of the underwater robots was not considered in this thesis. Accounting for UUV self-localization error and implementing either a CL or CNA method is an important consideration as future work for both Chapters 4 and 5. Lastly, controlled in-water testing would take a large step towards validating the the proposed target tracking method.

Bibliography

[1] Jonathan Hudson. *Adaptive path planning for an autonomous marine vehicle.* PhD thesis, Dalhousie University, 2012.

[2] J. Ross, J. Lindsay, E. Gregson, A. Moore, J. Patel, and M. Seto. Collaboration of multi-domain marine robots towards above and below-water characterization of floating targets. In *2019 IEEE International Symposium on Robotic and Sensors Environments (ROSE)*, pages 1–7, 2019.

[3] J. Hudson and M. L. Seto. Underway path-planning for an unmanned surface vehicle performing cooperative navigation for uuvs at varying depths. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2298–2305, 2014.

[4] J. Lindsay, M. Seto, and R. Bauer. Reinforcement q-learning applied to underwater search planning towards maximizing information gain in environments with variable target detection probabilities. *2020 Symposium on Autonomous Underwater Vehciles (AUV 2020)*, 2020.

[5] L. Paull, S. Saeedi, M. Seto, and H. Li. Auv navigation and localization: A review. *IEEE Journal of Oceanic Engineering*, 39(1):131–149, 2014.

[6] I. Masmitja, S. Gomariz, J. Del-Rio, B. Kieft, T. O'Reilly, P. Bouvet, and J. Aguzzi. Range-only single-beacon tracking of underwater targets from an autonomous vehicle: From theory to practice. *IEEE Access*, 7:86946–86963, 2019.

[7] J. G. Baylog and T. A. Wettergren. A search game for optimizing information collection in uuv mission planning. In *OCEANS 2015 - MTS/IEEE Washington*, pages 1–8, 2015.

[8] Y. Huang, Y. Zhang, and H. Xiao. Multi-robot system task allocation mechanism for smart factory. In *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, pages 587–591, 2019.

[9] A. Bahr, M. R. Walter, and J. J. Leonard. Consistent cooperative localization. In *2009 IEEE International Conference on Robotics and Automation*, pages 3415–3422, 2009.

[10] Z. Liu, Y Zhang, X. Yu, and C. Yuan. Unmanned surface vehicles: An overview of developments and challenges. *Annual Review of Control*, vol. 41, pp.:71–93, 2015.

[11] S. Saeedi M. Seto, L. Paull. *Marine Robot Autonomy.* New York: Springer,, 2013.

[12] G. Dudek and M. Jenkin. *Computational Principles of Mobile Robotics, 1st edition.* Press Syndicate of the University of Cambridge, 2000.

[13] A. R. Vetrella and G. Fasano. Cooperative uav navigation under nominal gps coverage and in gps-challenging environments. In *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, pages 1–5, 2016.

[14] M.A. Ainslie and J.G. McColm. A simplified formula for viscous and chemical absorption in sea water. *Journal of the Acoustical Society of America*, 103(3):1671–1672, 1998.

[15] F. Mosca, G. Matte, V. Mignard, and M. Rioblanc. Low-frequency source for very long-range underwater communication. In *2013 OCEANS - San Diego*, pages 1–5, 2013.

[16] M. Ludvigsen, S. M. Albrektsen, K. Cisek, T. A. Johansen, P. Norgren, R. Skjetne, A. Zolich, P. Sousa Dias, S. Ferreira, J. B. de Sousa, T. O. Fossum, Ø. Sture, T. Røbekk Krogstad, Ø. Midtgaard, V. Hovstein, and E. Vågsholm. Network of heterogeneous autonomous vehicles for marine research and management. In *OCEANS 2016 MTS/IEEE Monterey*, pages 1–7, 2016.

[17] F. Benavides, P. Monzón, C. P. Carvalho Chanel, and E. Grampín. Multi-robot cooperative systems for exploration: Advances in dealing with constrained communication environments. In *2016 XIII Latin American Robotics Symposium and IV Brazilian Robotics Symposium (LARS/SBR)*, pages 181–186, 2016.

[18] R. Siegwart and I. Nourbakhsh. *Introduction to Autonomous Mobile Robotics, 2nd ed.* Cambridge, Massachusetts: The MIT Press, 2011.

[19] M. Chitre. Path planning for cooperative underwater range-only navigation using a single beacon. In *2010 International Conference on Autonomous and Intelligent Systems, AIS 2010*, pages 1–6, 2010.

[20] G. Rui and M. Chitre. Cooperative positioning using range-only measurements between two auvs. In *OCEANS'10 IEEE SYDNEY*, pages 1–6, 2010.

[21] m. Fallon, G. Papadopoulos, J. Leonard, and N. Patrikalakis. Cooperative auv navigation using a single maneuvering surface craft. *I. J. Robotic Res.*, 29:1461–1474, 10 2010.

[22] J. Han, M. Kang, J. Wang, and J. Kim. Three-dimensional reconstruction of a semi-submersible offshore platform with an unmanned surface vehicle. In *OCEANS 2016 - Shanghai*, pages 1–6, 2016.

[23] A. Gautam and S. Mohan. A review of research in multi-robot systems. In *2012 IEEE 7th International Conference on Industrial and Information Systems (ICIIS)*, pages 1–5, 2012.

[24] C. Hu, L. Fu, and Y. Yang. Cooperative navigation and control for surface-underwater autonomous marine vehicles. In *2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pages 589–592, 2017.

[25] M. Faria, J. Pinto, F. Py, J. Fortuna, H. Dias, R. Martins, F. Leira, T. A. Johansen, J. Sousa, and K. Rajan. Coordinating uavs and auvs for oceanographic field experiments: Challenges and lessons learned. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6606–6611, 2014.

[26] F. Shkurti, A. Xu, M. Meghjani, J. C. Gamboa Higuera, Y. Girdhar, P. Giguère, B. B. Dey, J. Li, A. Kalmbach, C. Prahacs, K. Turgeon, I. Rekleitis, and G. Dudek. Multi-domain monitoring of marine environments using a heterogeneous robot team. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1747–1753, 2012.

[27] S. Manjanna, J. Hansen, A. Q. Li, I. Rekleitis, and G. Dudek. Collaborative sampling using heterogeneous marine robots driven by visual cues. In *2017 14th Conference on Computer and Robot Vision (CRV)*, pages 87–94, 2017.

[28] S. Eckstein, T. Glotzbach, and C. Ament. Design of a software structure and a mission handler for cooperative marine robots. In *OCEANS 2015 - Genova*, pages 1–6, 2015.

[29] DRDC. Exercise unmanned warrior: an international exercise using autonomous tech to detect underwater mines. `https://www.canada.ca/en/defence-research-development/news/articles/exercise-unmanned-warrior-an-international-exercise-using-autonomous-tech-to-detect-underwater-mines.html'`, 2016. Accessed: 2019-03-01.

[30] S. Kwon, J. Park, and J. Kim. 3d reconstruction of underwater objects using a wide-beam imaging sonar. In *2017 IEEE Underwater Technology (UT)*, pages 1–4, 2017.

[31] G. Papadopoulos, H. Kurniawati, A. S. Bin Mohd Shariff, L. J. Wong, and N. M. Patrikalakis. 3d-surface reconstruction for partially submerged marine structures using an autonomous surface vehicle. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3551–3557, 2011.

[32] M. Meghjani, S. Manjanna, and G. Dudek. Multi-target search strategies. In *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 328–333, 2016.

[33] S. Waharte and N. Trigoni. Supporting search and rescue operations with uavs. In *2010 International Conference on Emerging Security Technologies*, pages 142–147, 2010.

[34] S. Bernardini, D. Long, M. Fox, and J. Bookless. Autonomous search and tracking via temporal planning. In *ICAPS 2013 - Proceedings of the 23rd International Conference on Automated Planning and Scheduling*, 06 2013.

[35] J. G. Baylog and T. A. Wettergren. Extended search games for uuv mission planning. In *OCEANS 2017 - Anchorage*, pages 1–9, 2017.

[36] J. G. Baylog and T. A. Wettergren. A roc-based approach for developing optimal strategies in uuv search planning. *IEEE Journal of Oceanic Engineering*, 43(4):843–855, 2018.

[37] C. Yan and X. Xiang. A path planning algorithm for uav based on improved q-learning. In *2018 2nd International Conference on Robotics and Automation Sciences (ICRAS)*, pages 1–5, 2018.

[38] C. Lamini, Y. Fathi, and S. Benhlima. Collaborative q-learning path planning for autonomous robots based on holonic multi-agent system. In *2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA)*, pages 1–6, 2015.

[39] G. Frost and D. M. Lane. Evaluation of q-learning for search and inspect missions using underwater vehicles. In *2014 Oceans - St. John's*, pages 1–6, 2014.

[40] U. Gautam, R. Malmathanraj, and C. Srivastav. Simulation for path planning of autonomous underwater vehicle using flower pollination algorithm, genetic algorithm and q-learning. In *2015 International Conference on Cognitive Computing and Information Processing(CCIP)*, pages 1–5, 2015.

[41] Ivan Masmitja, Spartacus Gomariz, Joaquin del Rio, Brian Kieft, Tom O'Reilly, Pierre-Jean Bouvet, and Jacopo Aguzzi. Optimal path shape for range-only underwater target localization using a wave glider. *The International Journal of Robotics Research*, 37:027836491880235, 10 2018.

[42] N. Crasta, D. Moreno-Salinas, A.M. Pascoal, and J. Aranda. Multiple autonomous surface vehicle motion planning for cooperative range-based underwater target localization. *Annual Reviews in Control*, 46:326 – 342, 2018.

[43] P. Agrawal and J. M. Dolan. Colregs-compliant target following for an unmanned surface vehicle in dynamic environments. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1065–1070, 2015.

[44] M. R. Benjamin and J. A. Curcio. Colregs-based navigation of autonomous marine vehicles. In *2004 IEEE/OES Autonomous Underwater Vehicles (IEEE Cat. No.04CH37578)*, pages 32–39, 2004.

[45] Tingting Sun, Bo He, Rui Nian, and Tianhong Yan. Target following for an autonomous underwater vehicle using regularized elm-based reinforcement learning. In *OCEANS 2015 - MTS/IEEE Washington*, pages 1–5, 2015.

[46] D. Moreno-Salinas, A. Pascoal, and J. Aranda. Optimal sensor placement for acoustic underwater target positioning with range-only measurements. *IEEE Journal of Oceanic Engineering*, 41(3):620–643, 2016.

[47] G. Rui and M. Chitre. Cooperative multi-auv localization using distributed extended information filter. In *2016 IEEE/OES Autonomous Underwater Vehicles (AUV)*, pages 206–212, 2016.

[48] T. Pastore, G. Galdorisi, and A. Jones. Command and control (c2) to enable multi-domain teaming of unmanned vehicles (uxvs). In *OCEANS 2017 - Anchorage*, pages 1–7, 2017.

[49] Open Source Robotics Foundation. Gazebo. `http://gazebosim.org/`', 2012. Accessed: 2019-03-04.

[50] Musa Manhães, Sebastian A. Scherer, Martin Voss, Luiz Ricardo Douat, and Thomas Rauschenbach. UUV simulator: A gazebo-based package for underwater intervention and multi-robot simulation. In *OCEANS 2016 MTS/IEEE Monterey*. IEEE, Sep 2016.

[51] Clearpath Robotics Incorporated. Simulator package for heron usv. `https://github.com/heron/heron_simulator`, 2019.

[52] J. Meyer, A. Sendobry, S. Kohlbrecher, U. Klingauf, and O. vonStryk. Comprehensive simulation of quadrotor uavs using ros and gazebo. *Simulation, Modeling, and Programming for Autonomous Robots. SIMPAR 2012*, pages 400–411, 2012.

[53] J. Nichols, D. Leslie, S. Soylu, T. Crees, A. Woodroffe, and S. McLean. Imotus: An autonomous underwater resident vehicle for vertical profiling. In *OCEANS 2018 MTS/IEEE Charleston*, pages 1–5, 2018.

[54] Motion Analysis. Cortex user manual, 2018.

[55] A. Danckaers and M.L. Seto. Transmission of images by unmanned underwater vehicles. *J. Autonomous Robots, accepted for publication April 19, 2019*, page 22, 2019.

[56] C.E. Shannon. A mathematical theory of communication. *Bell Syst.Tech. J*, 27:379–423, 1948.

[57] R. Urick. *Principles of Underwater Sound*. New York, NY, USA: Mcgraw Hill, 1983.

[58] S. Kay. Fundamentals of statistical signal processing. *Detection Theory*, 1998.

[59] T. M. Cover and J. A Thomas. *Elements of Information Theory*. Hoboken, NJ, USA: Wiley-Interscience, 2006.

[60] G.Grimmett and D.Stirzaker. *ProbabilityandRandomProcesses*. U.K.: Oxford Univ. Press, 2001.

[61] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.

[62] University of Michigan The APRIL Robotics Laboratory. Apriltag: a visual fiducial system popular for robotics research. `https://april.eecs.umich.edu/software/apriltag`, 2010.

[63] A. Konar, I. Goswami Chakraborty, S. J. Singh, L. C. Jain, and A. K. Nagar. A deterministic improved q-learning for path planning of a mobile robot. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43(5):1141–1153, 2013.

[64] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 3 edition, 2010.

[65] N. Yang, D. Chang, M. R. Amini, M. Johnson-Robersor, and J. Sun. Energy management for autonomous underwater vehicles using economic model predictive control. In *2019 American Control Conference (ACC)*, pages 2639–2644, 2019.

[66] M. Moradi Zaniani, A. M. Shahar, and I. Abdul Azid. Trilateration target estimation improvement using new error correction algorithm. In *2010 18th Iranian Conference on Electrical Engineering*, pages 489–494, 2010.

[67] E. E. Tsakonas, N. D. Sidiropoulos, and A. Swami. Optimal particle filters for tracking a time-varying harmonic or chirp signal. *IEEE Transactions on Signal Processing*, 56(10):4598–4610, 2008.

[68] Georgios Papadopoulos. Underwater vehicle localization using range measurements. Master's thesis, Massachusetts Institute of Technology (MIT), 2010.

[69] The MathWorks, Inc. Sonar equation. `https://www.mathworks.com/help/phased/ug/sonar-equation.html#mw_f7e25a3e-7e33-4ed6-9b1f-87572479fcc0`.

[70] WHOI Acoustic Communications Group. Micromodem. https://acomms.whoi.edu/micro-modem/, 2020.

[71] R. Sutton and A. Barto. *Reinforcment Learning: An Introduction*. The MIT Press Cambridge, Massachusetts, 1998.

[72] S. Blouin. Bedford basin sound speed profiles - statistical analysis and low-order approximations. Technical report, DRDC, 2015.

[73] Government of Canada. HMCS victoria. `https://www.canada.ca/en/navy/services/history/ships-histories/victoria.html`, 2020. Accessed: 2020-11-07.

[74] G. Pawlak, M. McManus, L. Tuthill, J. Sevadjian, M. Ericksen, and A. Rocheleau. Real-time ocean water quality monitoring for the south shore of oahu. In *OCEANS'11 MTS/IEEE KONA*, pages 1–8, 2011.

Appendix A

Complete Results for Target Tracking and Localization Simulations

This appendix provides the complete set of target tracking simulations in Chapter 5 (Tables 5.4 and 5.5). The figures here are labelled by their trial number.

A.1 Configuration #1 Results
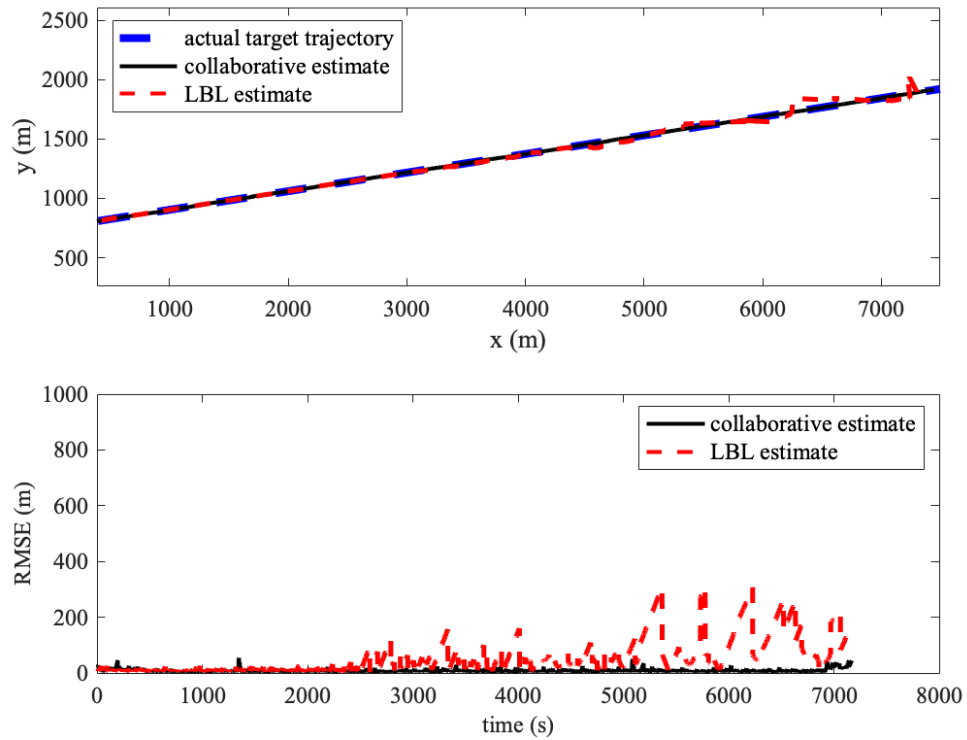
Figure A.1: Trial #1
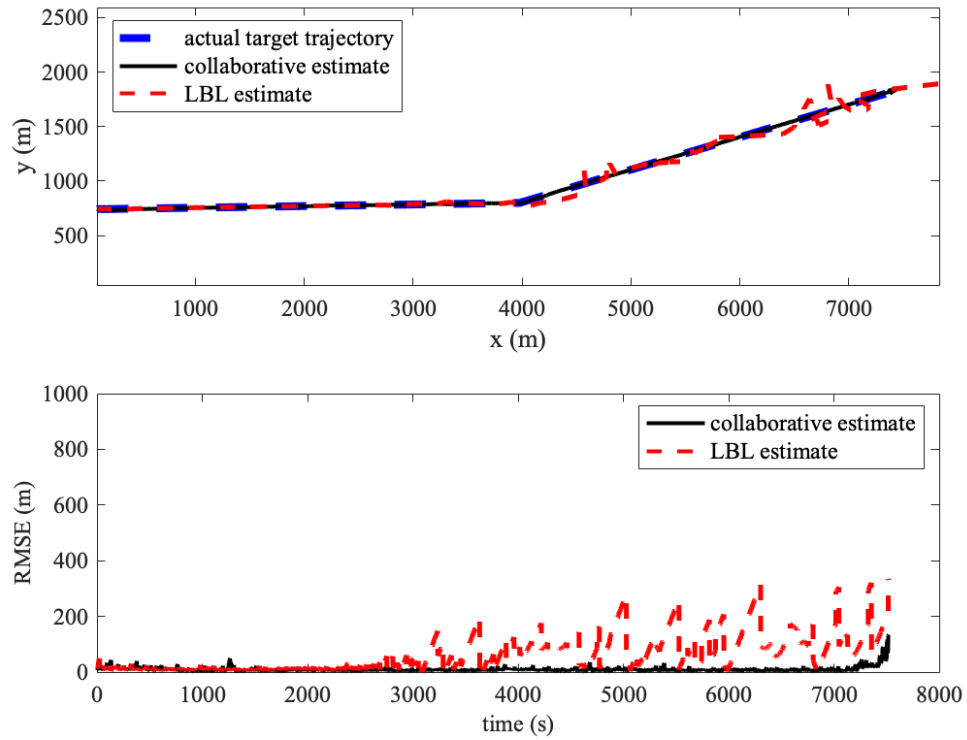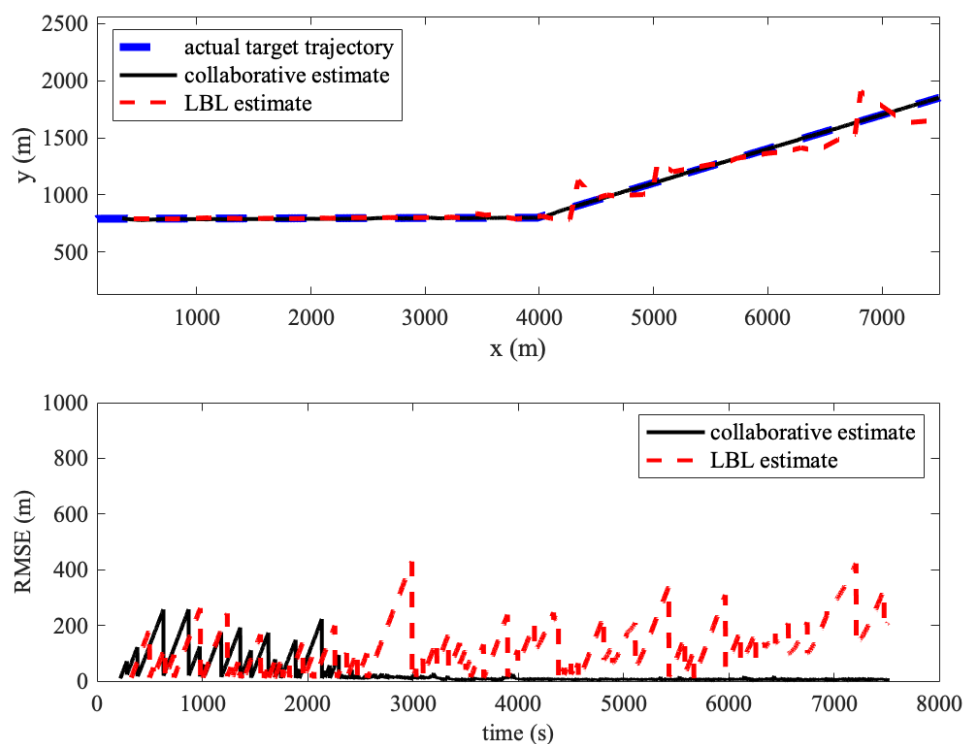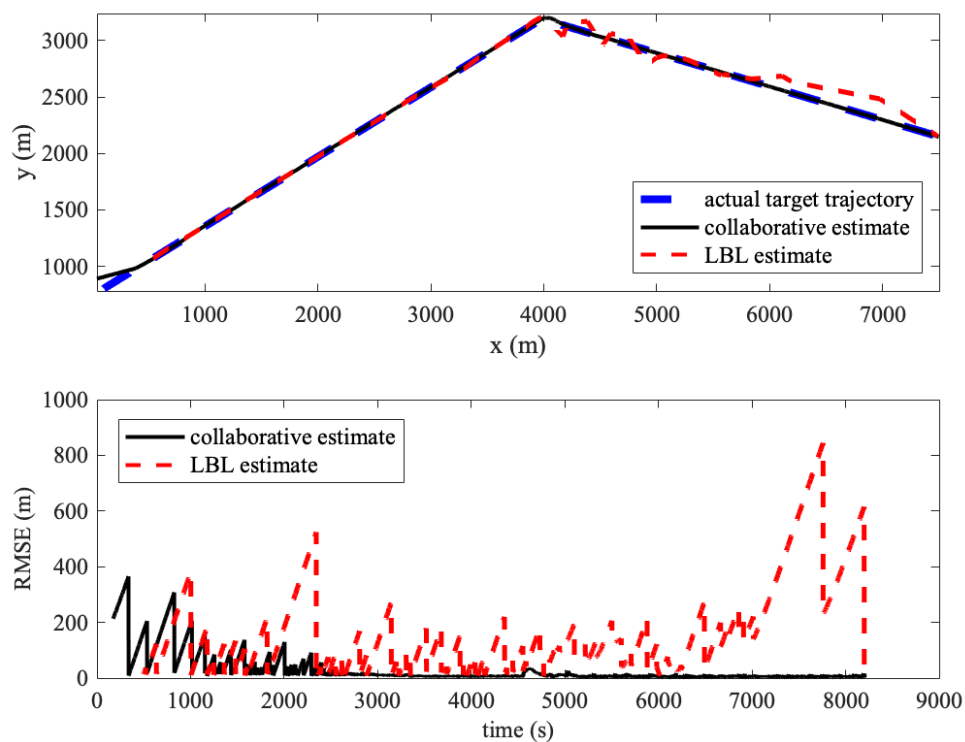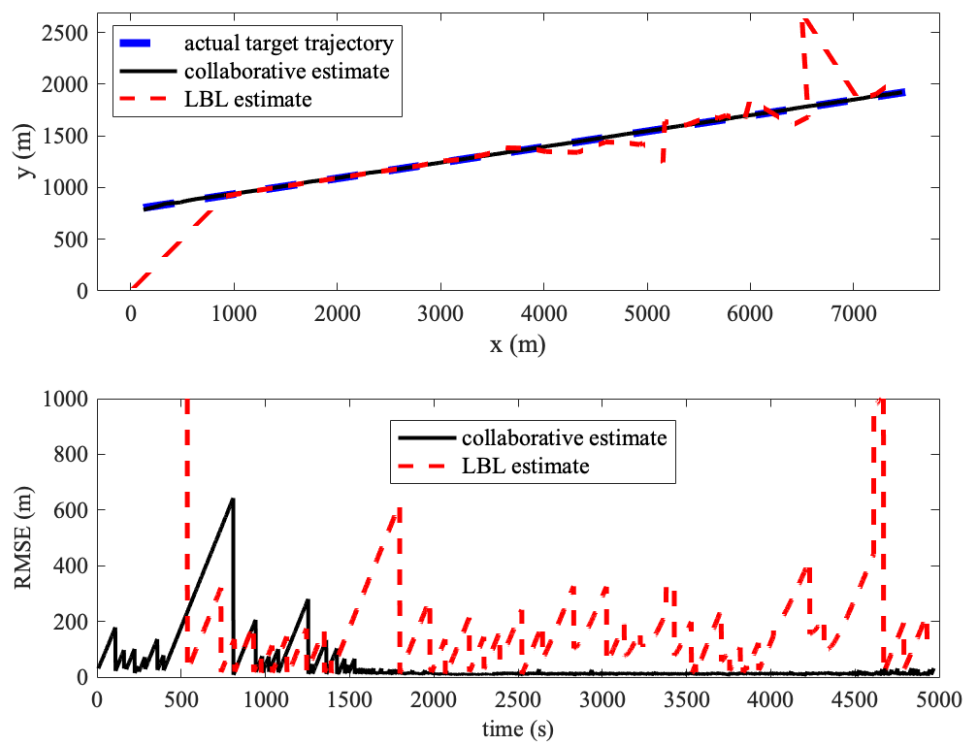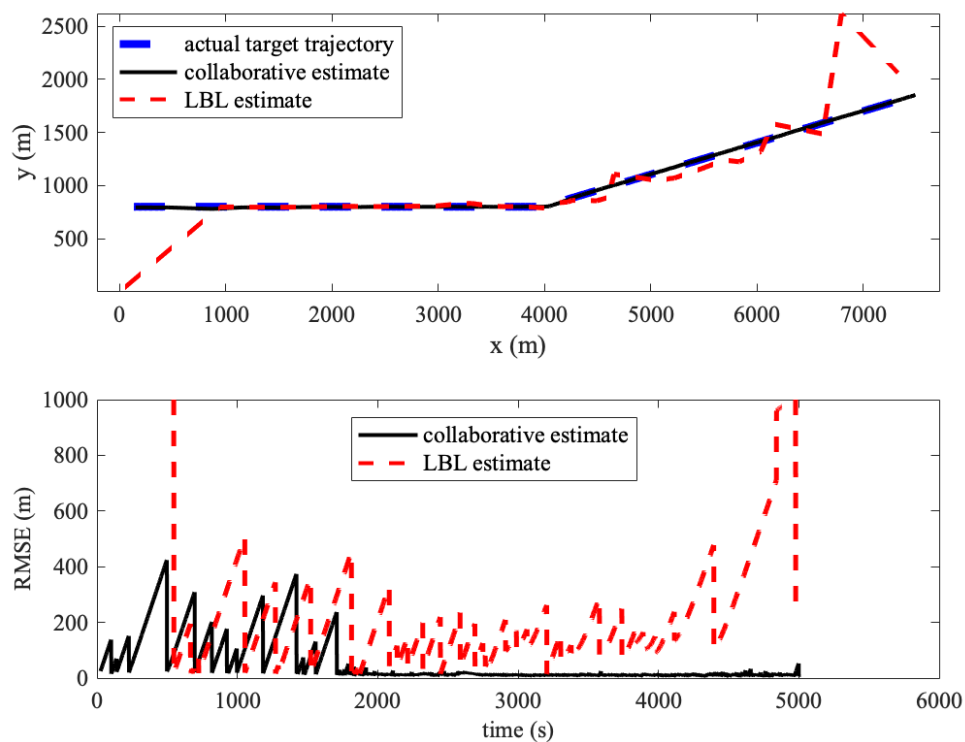
Figure A.2: Trial #2
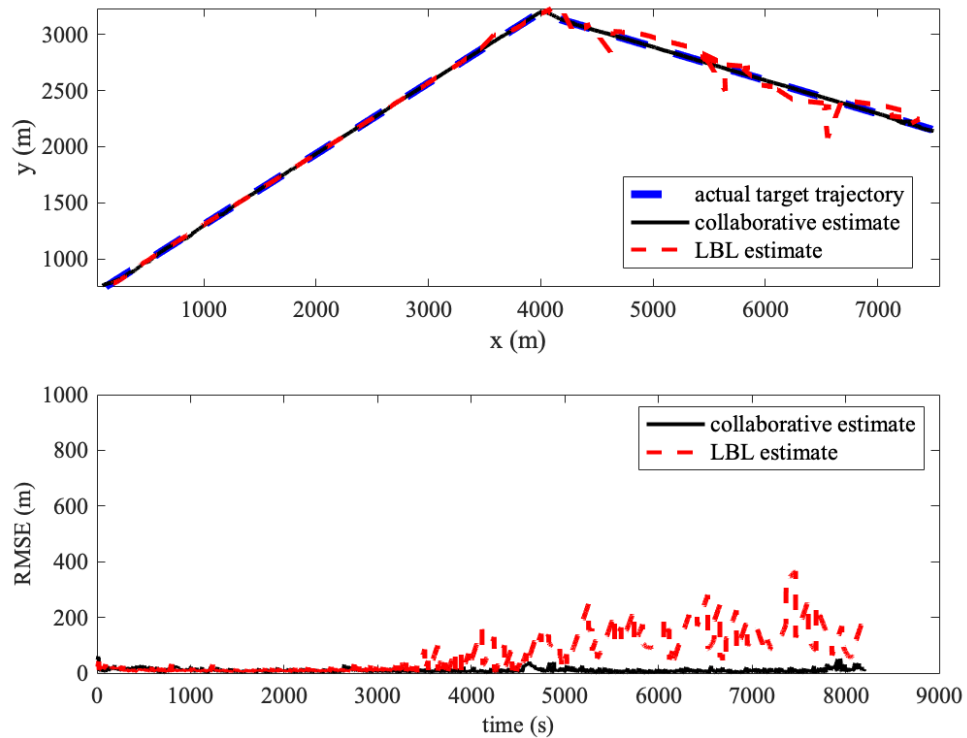
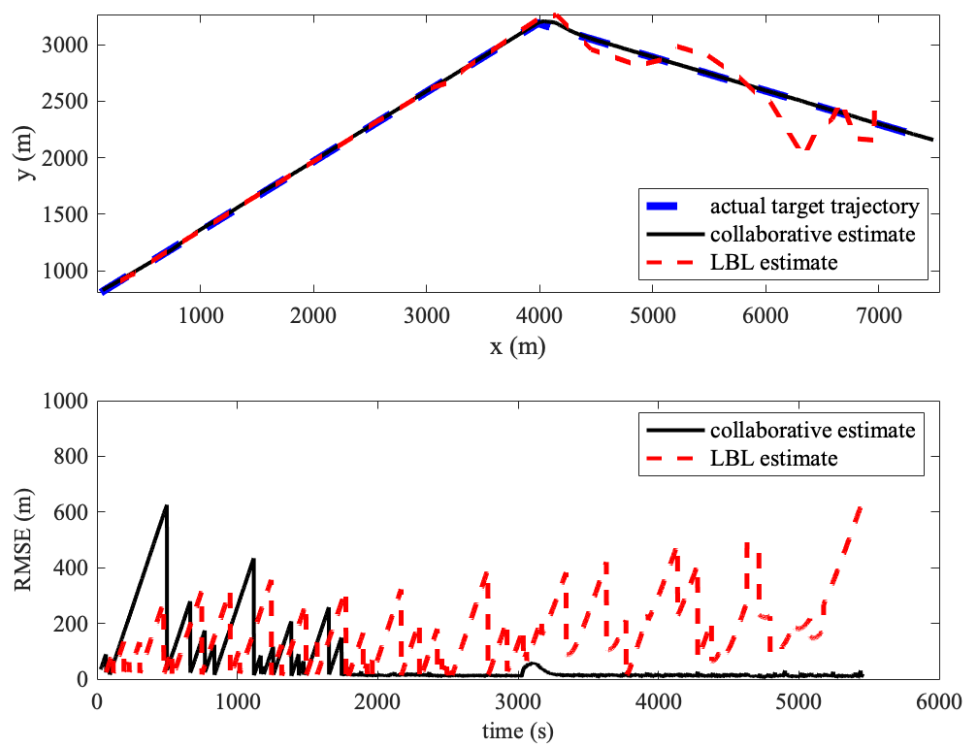Figure A.3: Trial #3
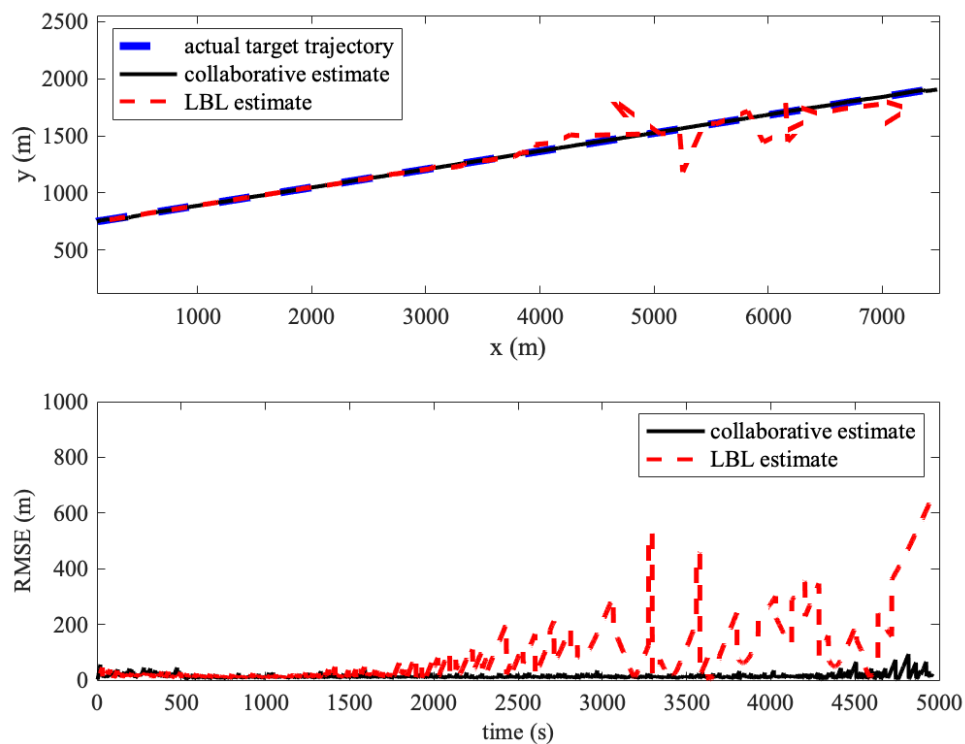
Figure A.4: Trial #4
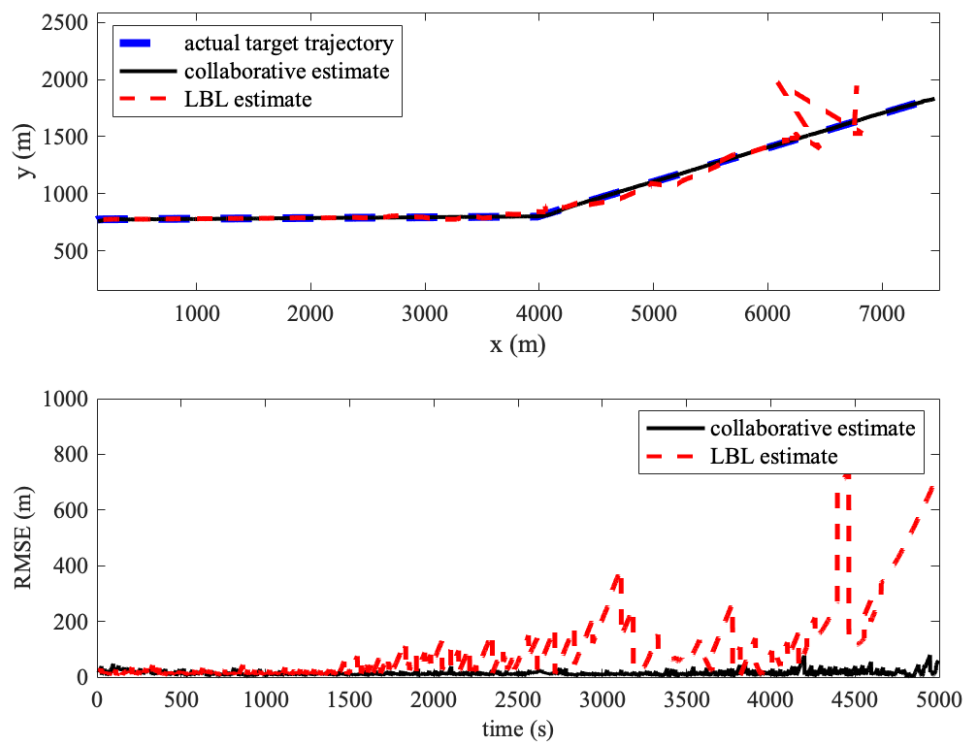
Figure A.5: Trial #5
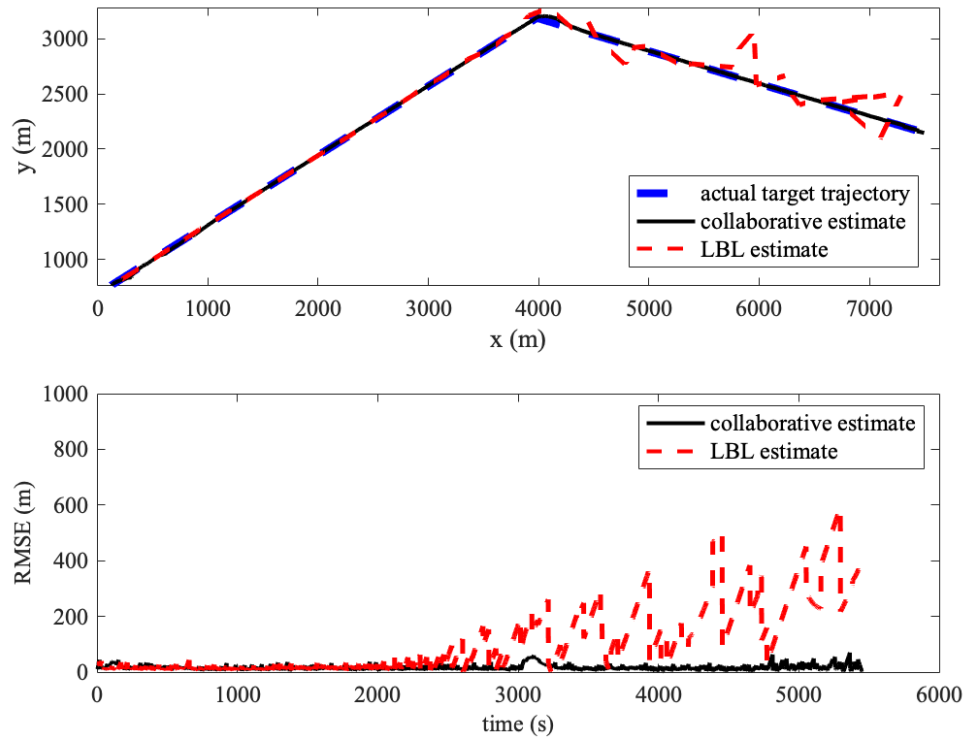
Figure A.6: Trial #6
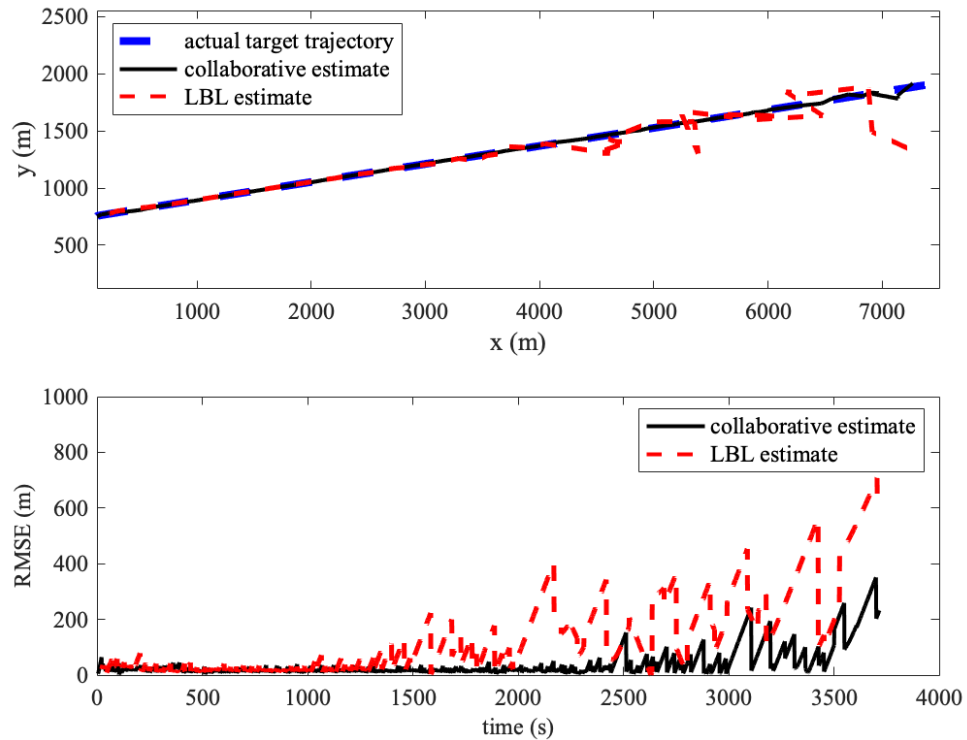
Figure A.7: Trial #7
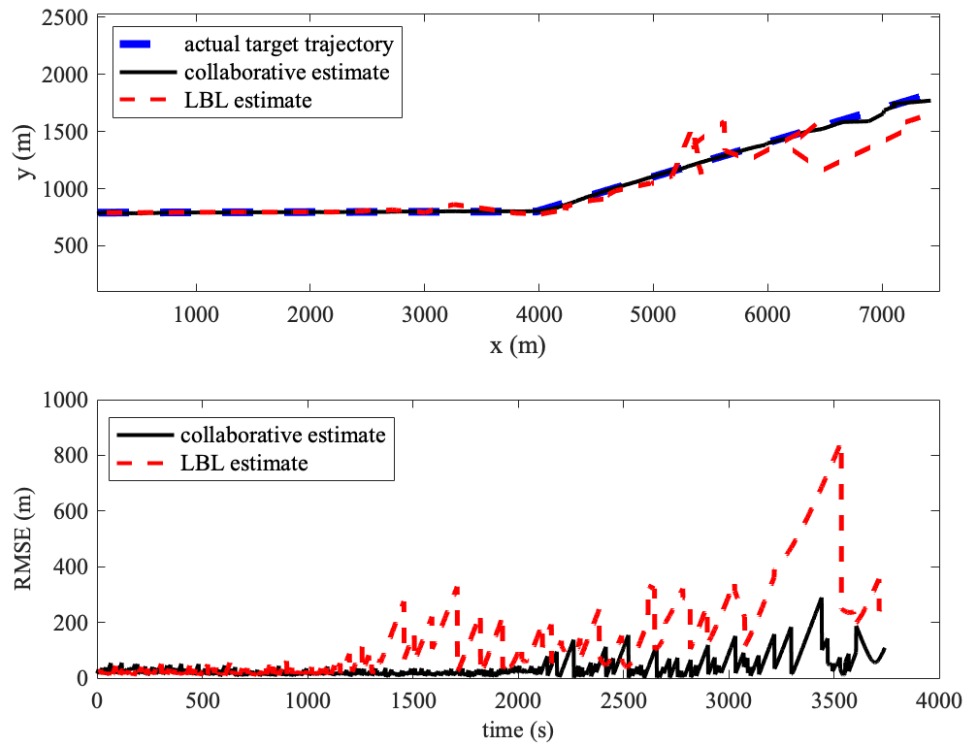
Figure A.8: Trial #8
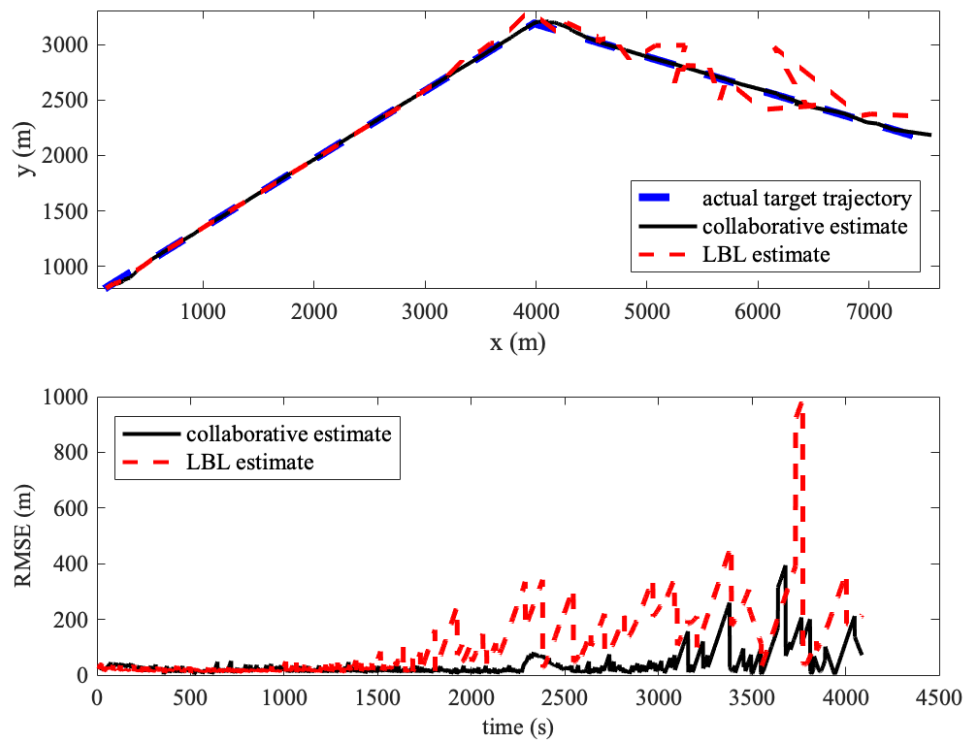
Figure A.9: Trial #9
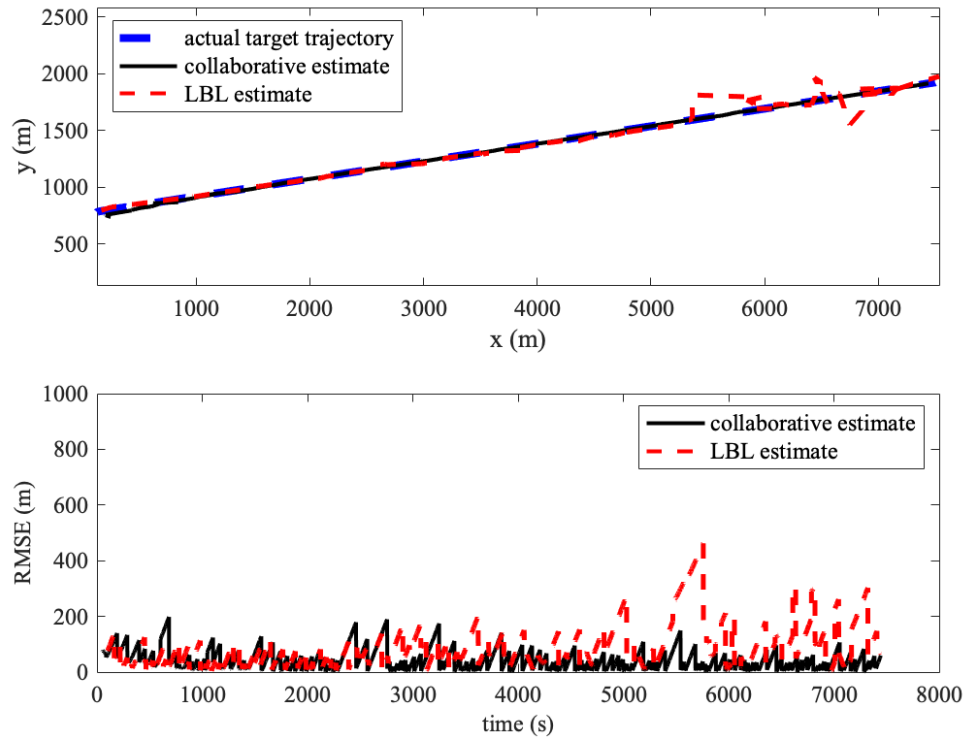
Figure A.10: Trial #10
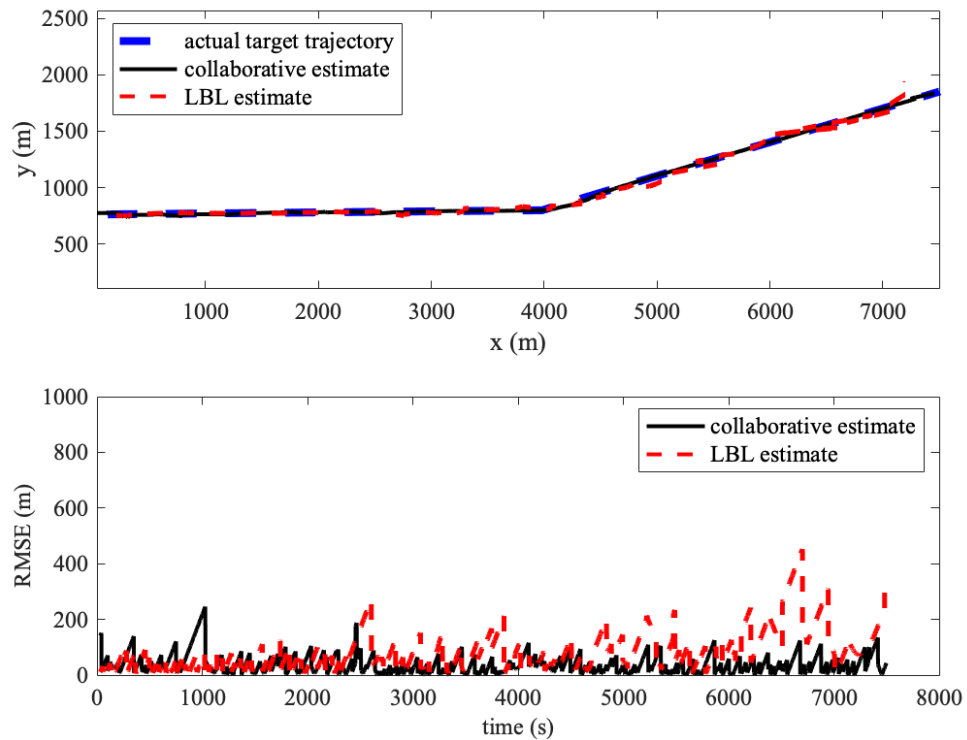
Figure A.11: Trial #11
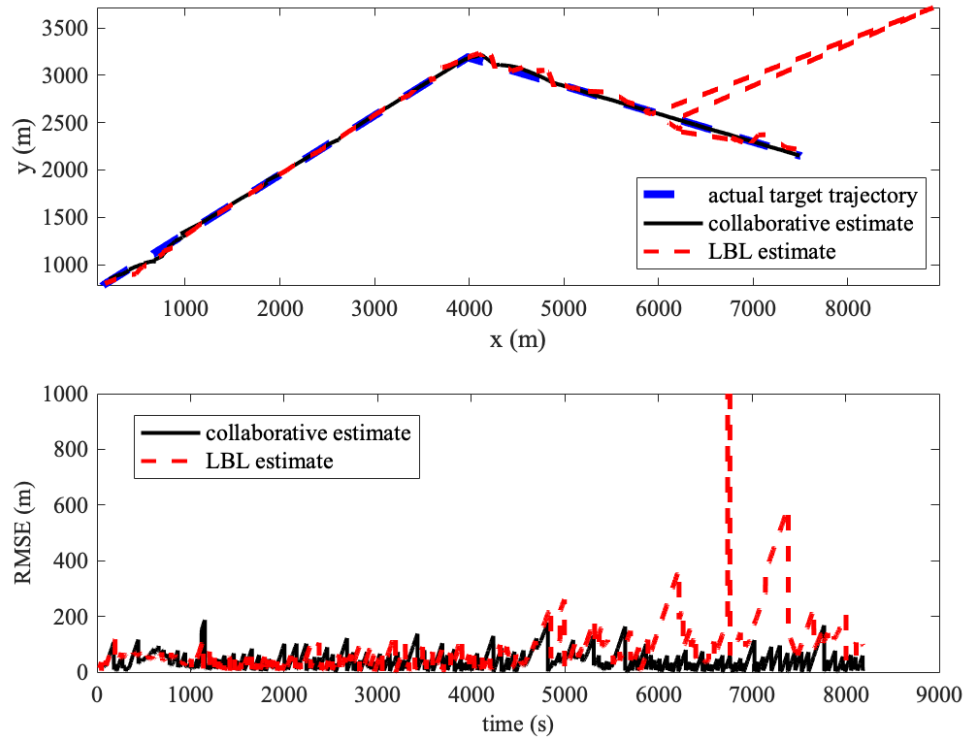
Figure A.12: Trial #12
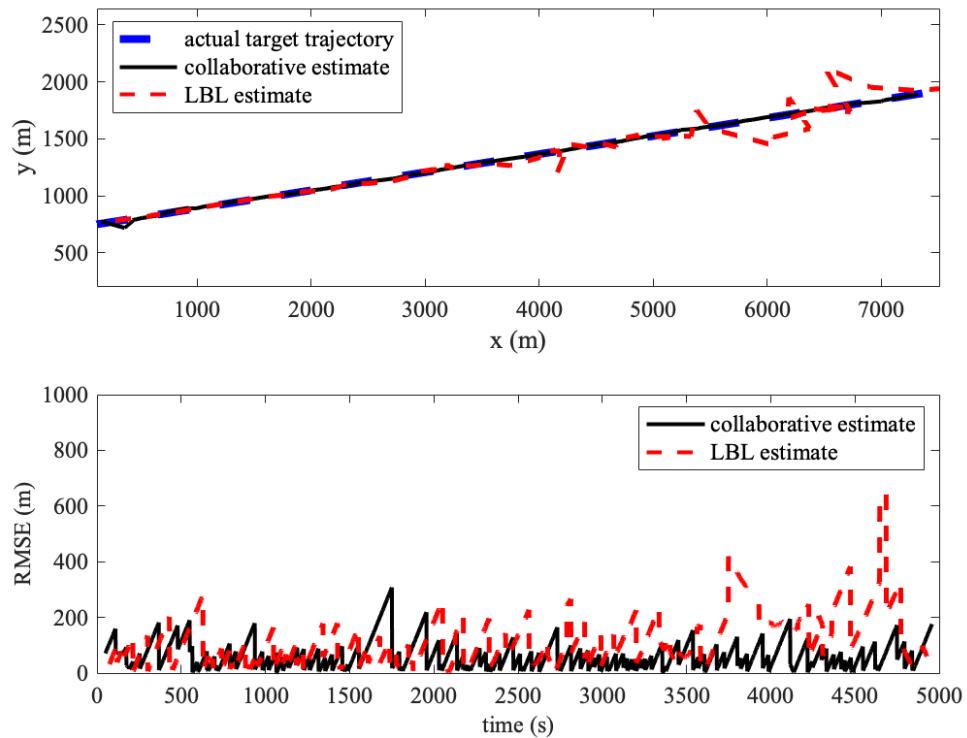
Figure A.13: Trial #13
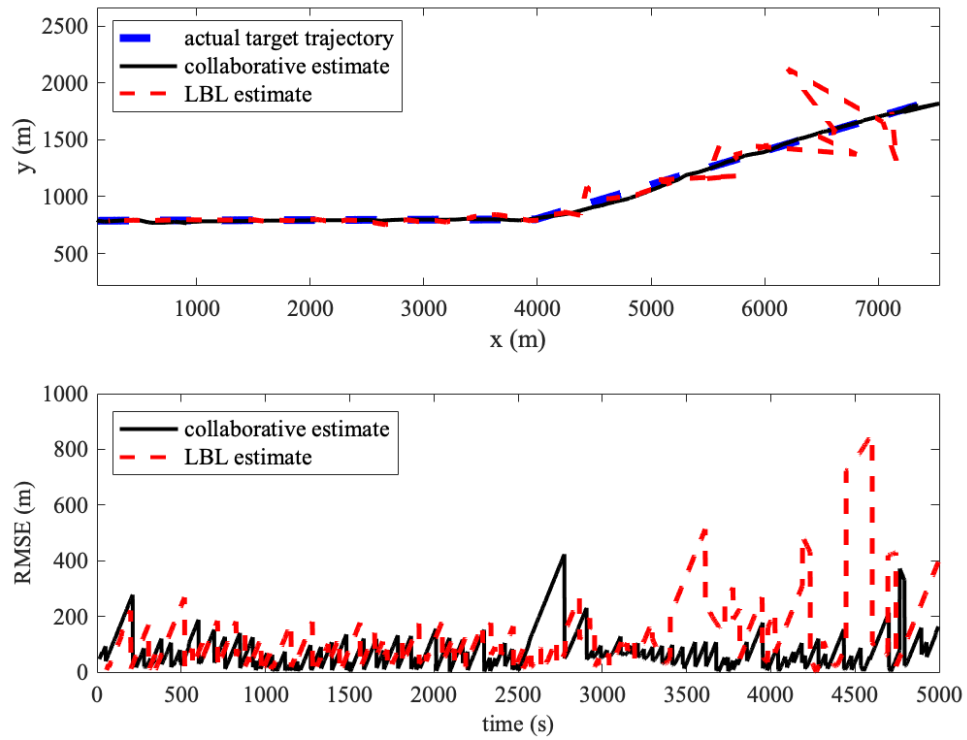
Figure A.14: Trial #14
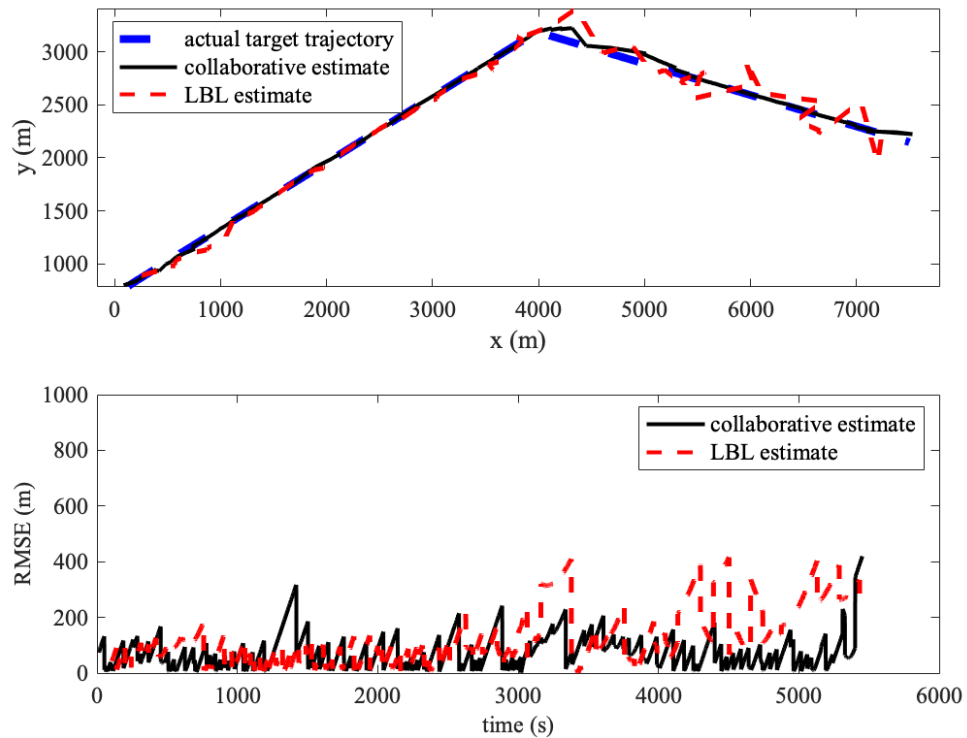
Figure A.15: Trial #15
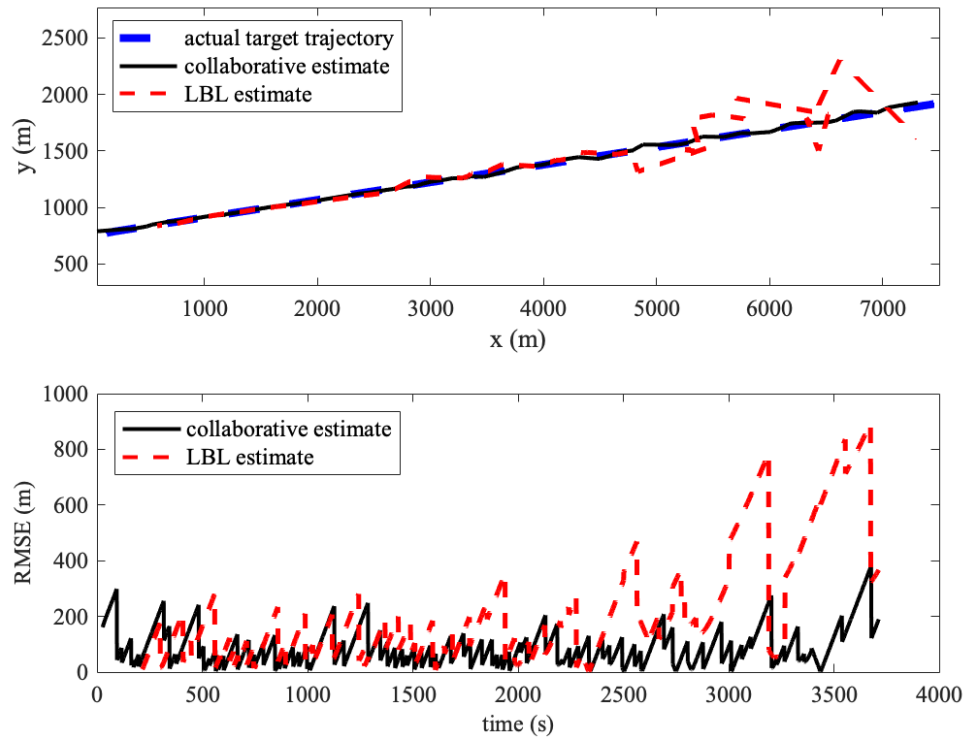
Figure A.16: Trial #16
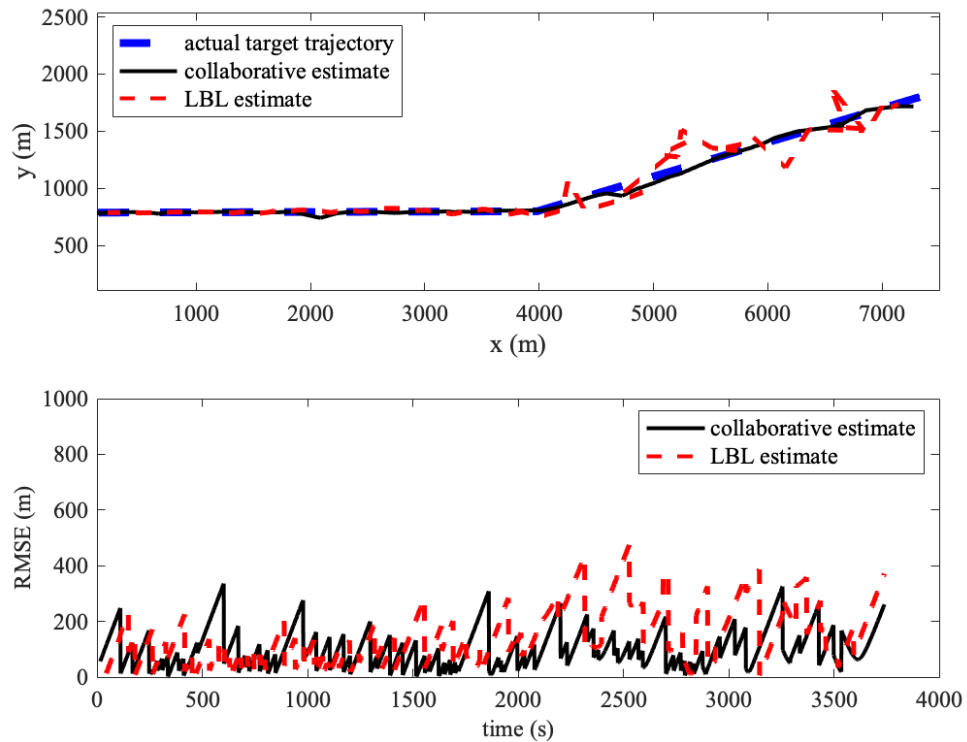
Figure A.17: Trial #17
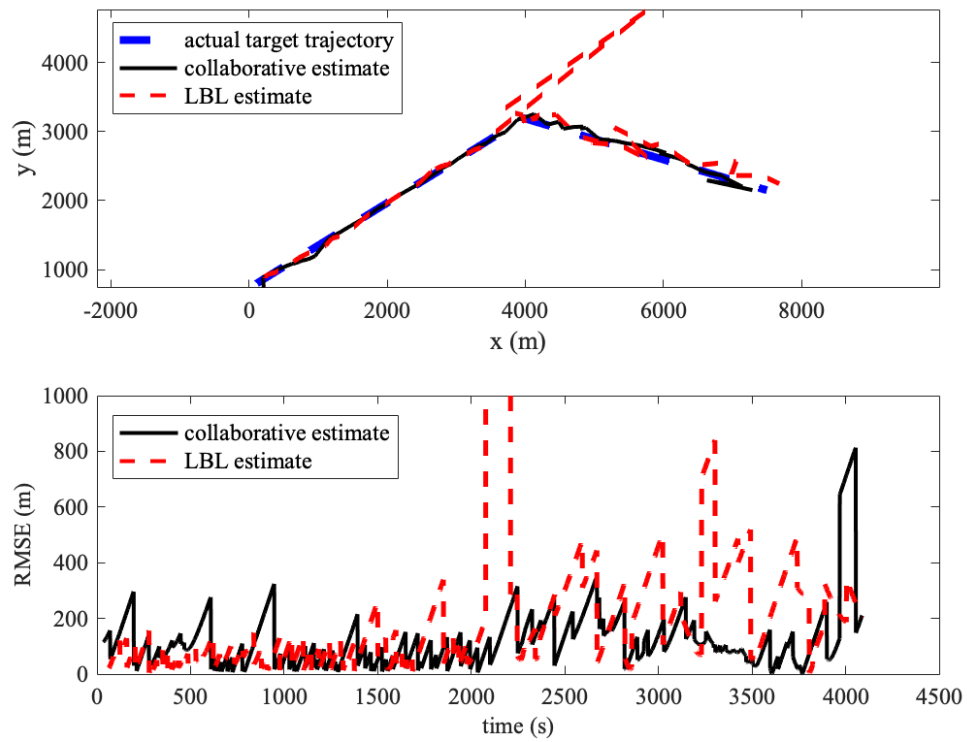
Figure A.18: Trial #18

A.2   Configuration #2 Results
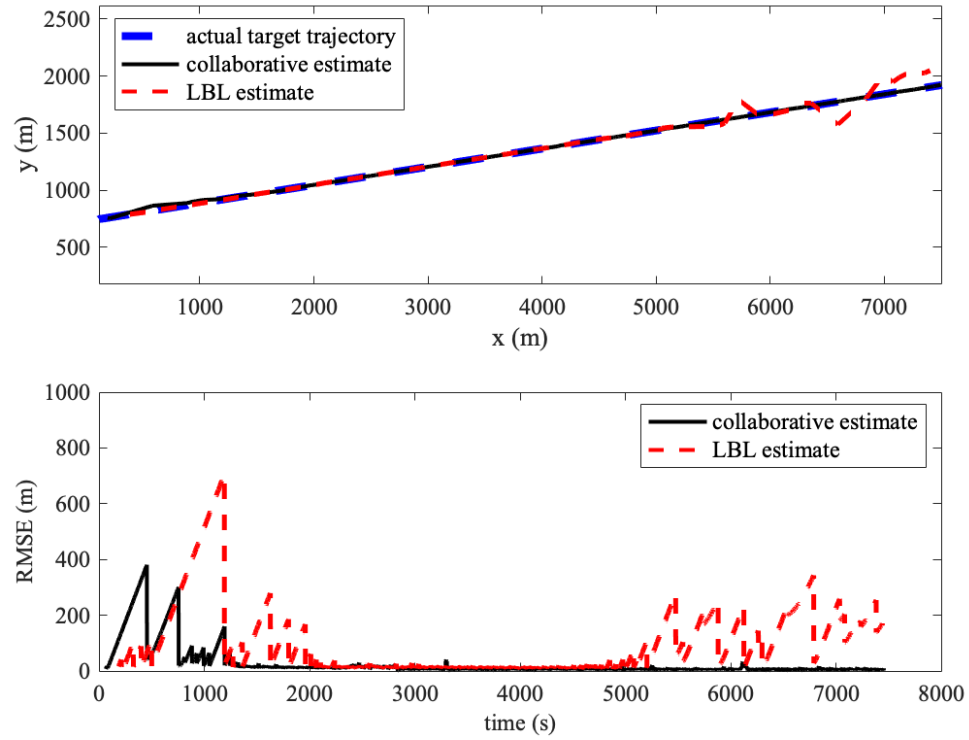
Figure A.19: Trial #20
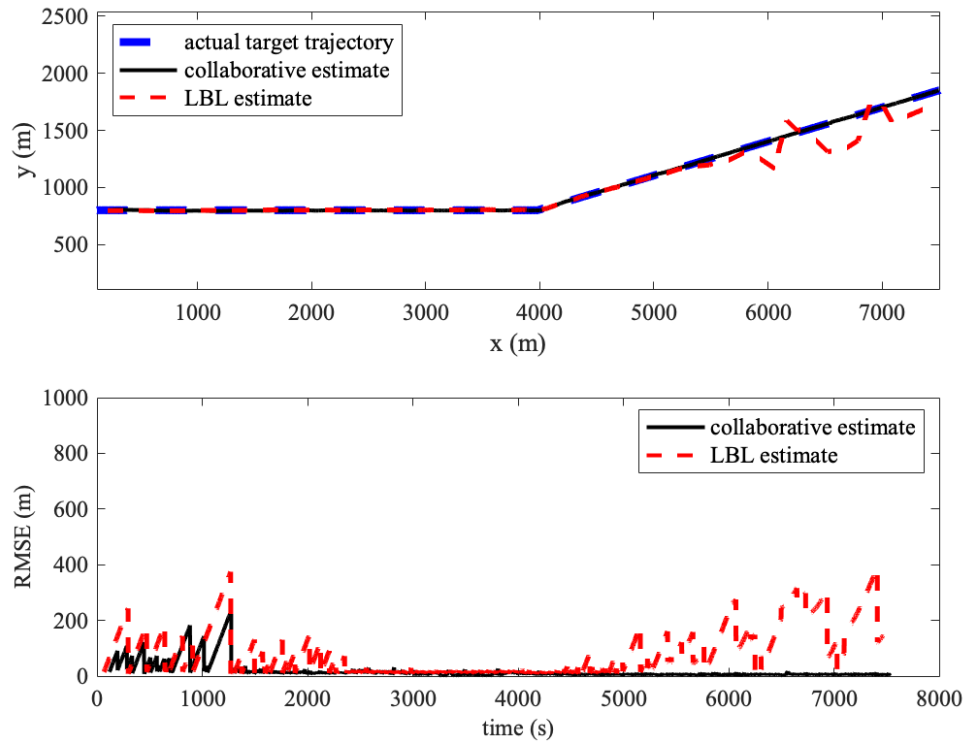
Figure A.20: Trial #21
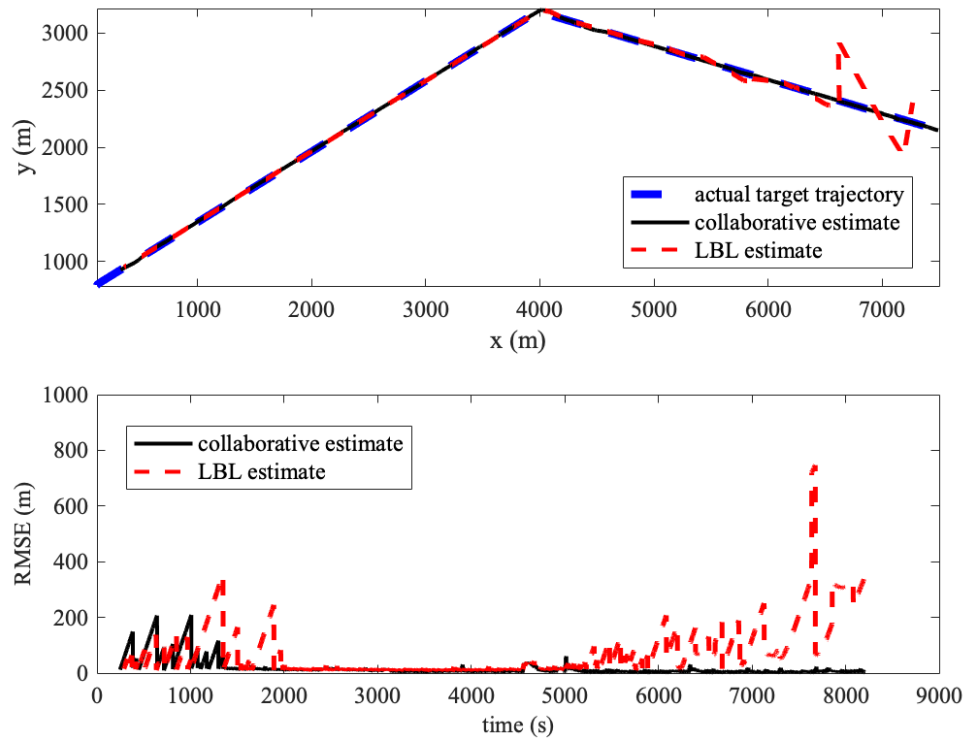
Figure A.21: Trial #22
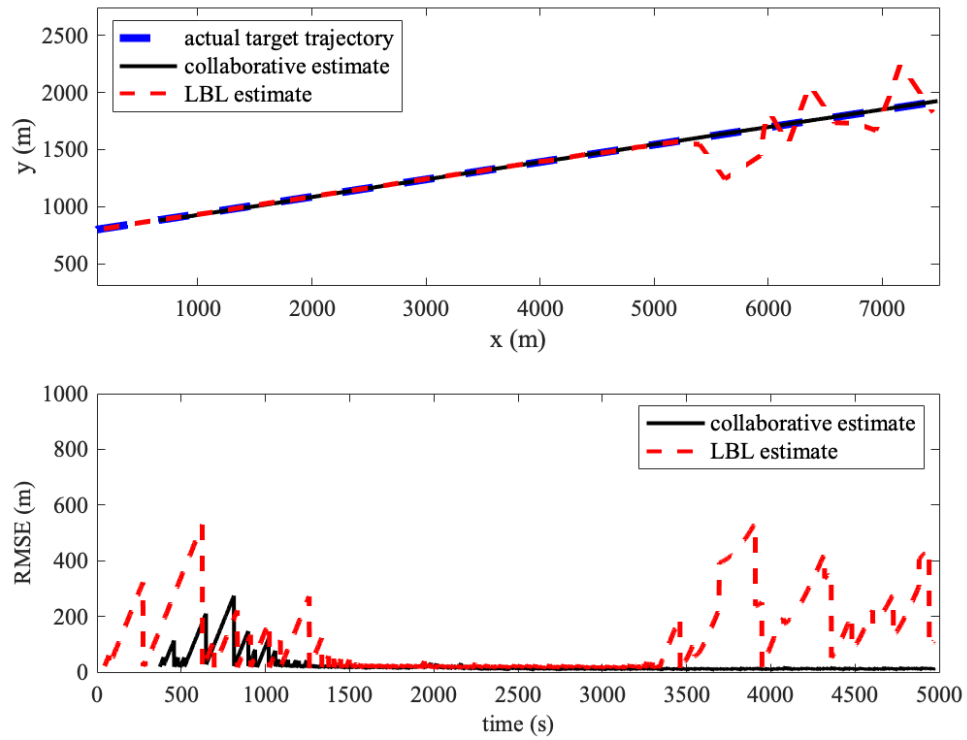
Figure A.22: Trial #23
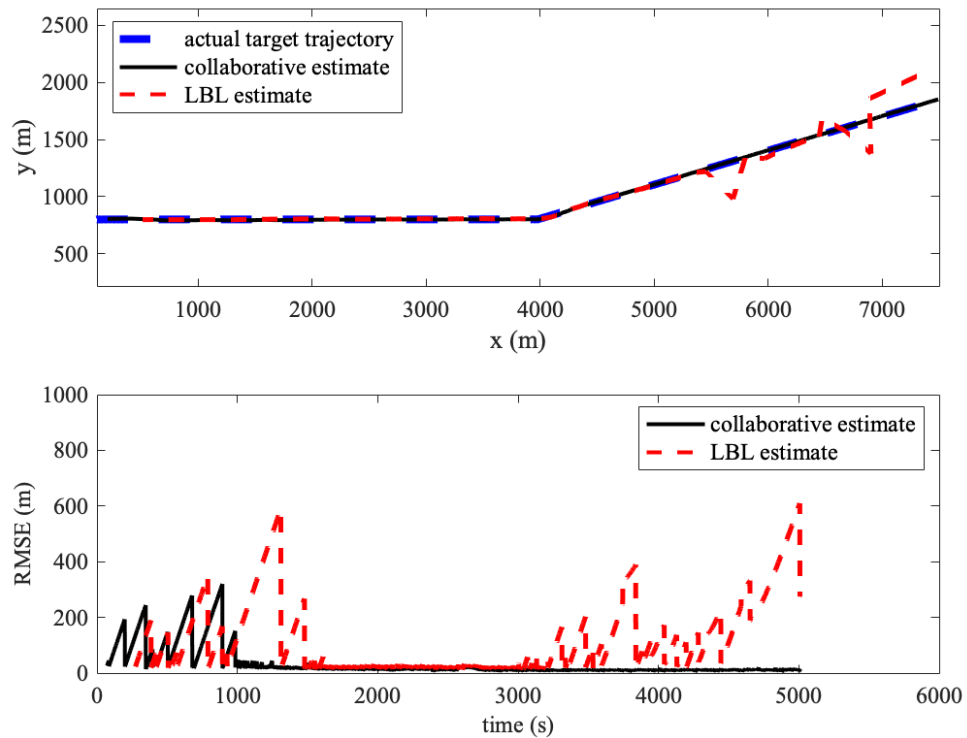
Figure A.23: Trial #24
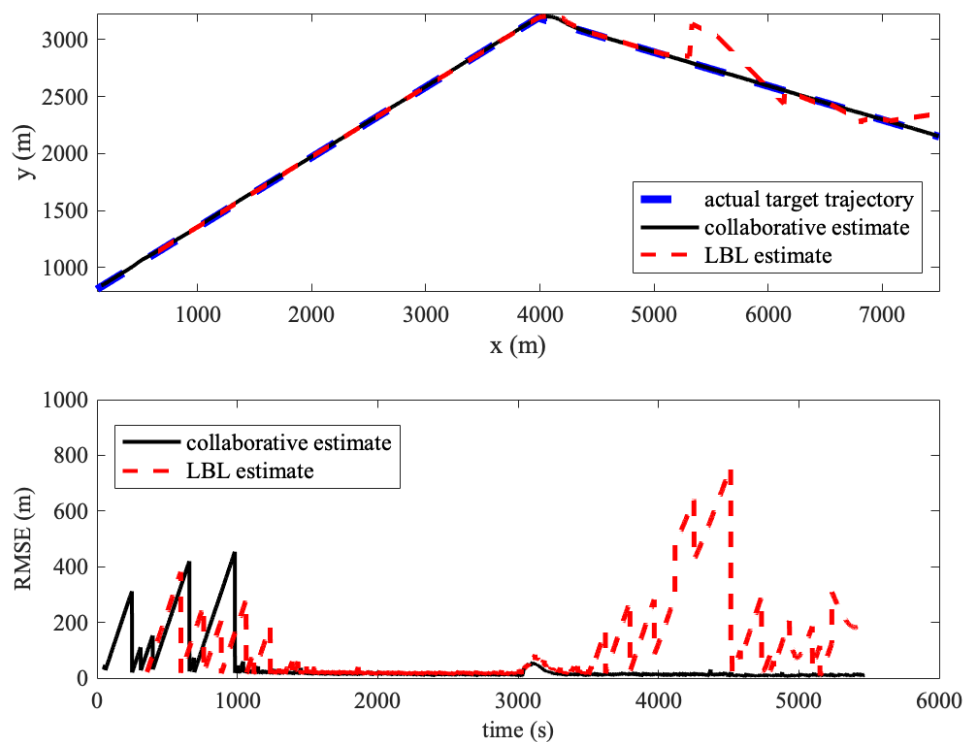
Figure A.24: Trial #25
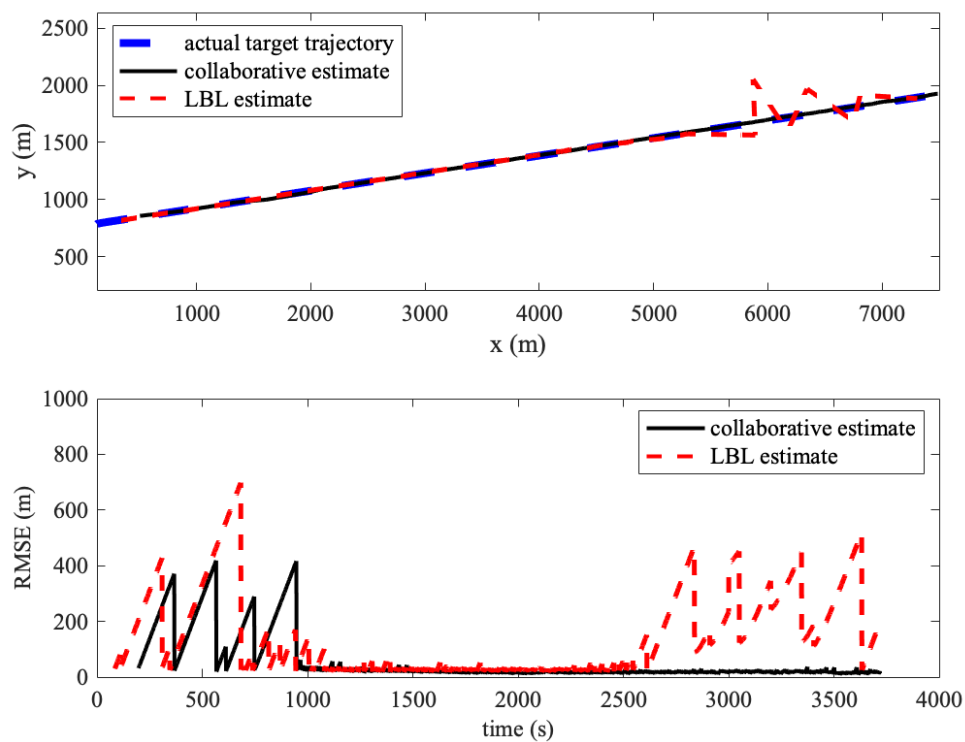
Figure A.25: Trial #26
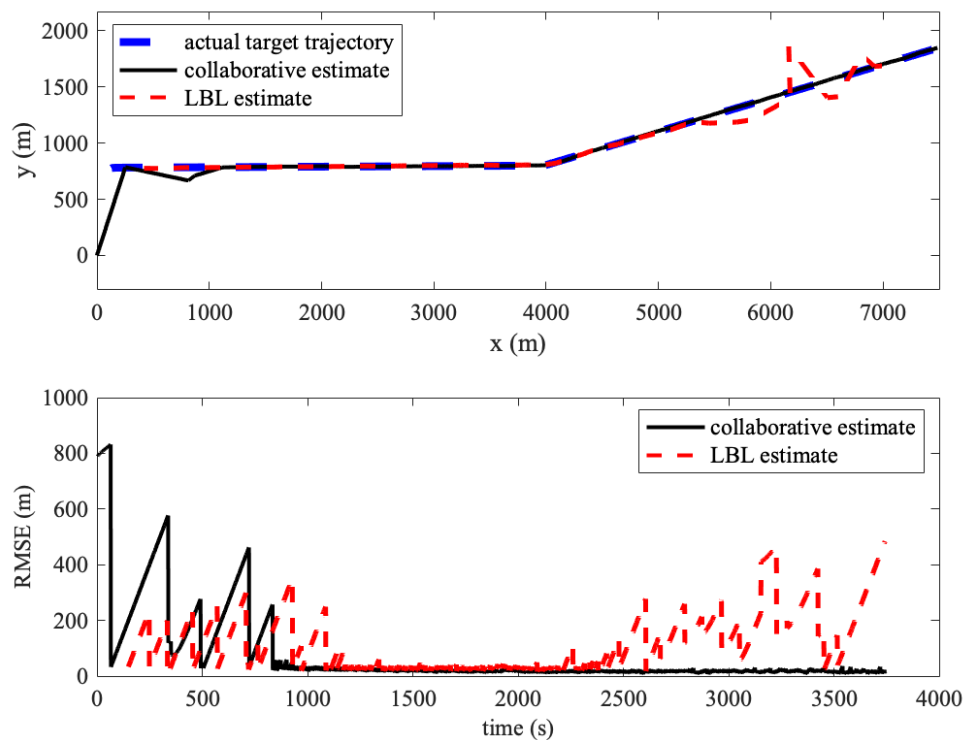
Figure A.26: Trial #27
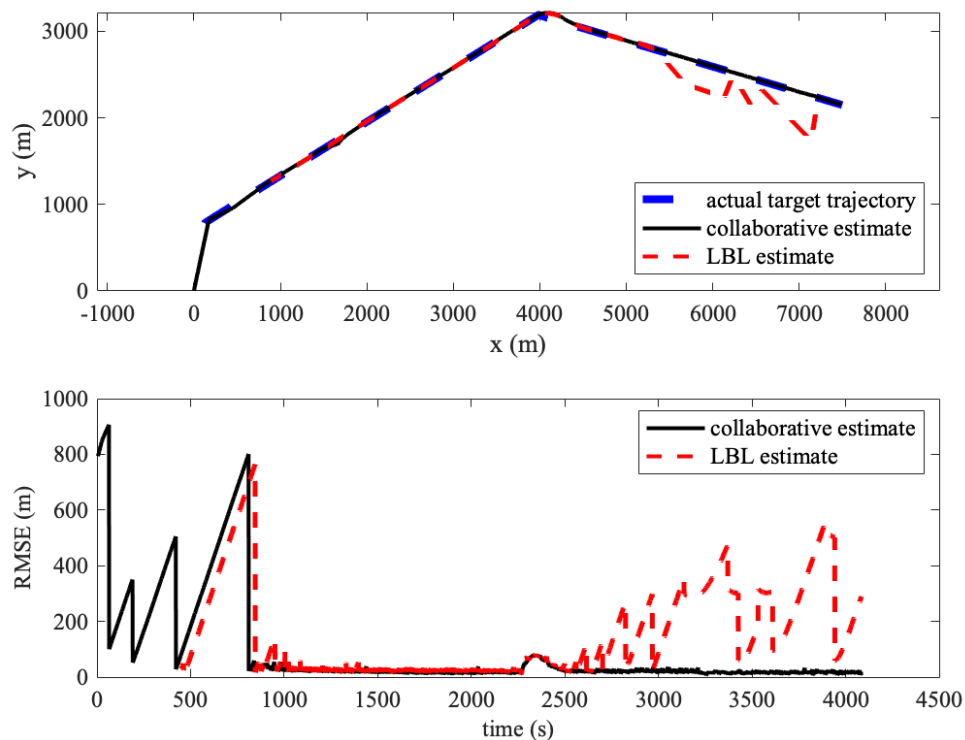
Figure A.27: Trial #28
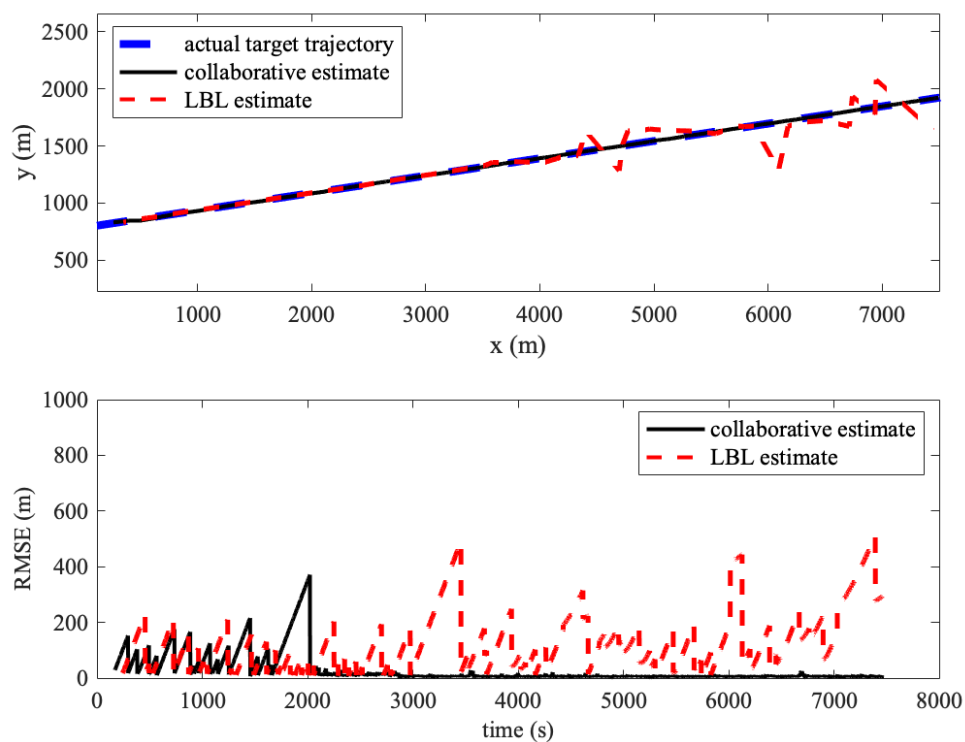
Figure A.28: Trial #29

Figure A.29: Trial #30

Figure A.30: Trial #31

Figure A.31: Trial #32
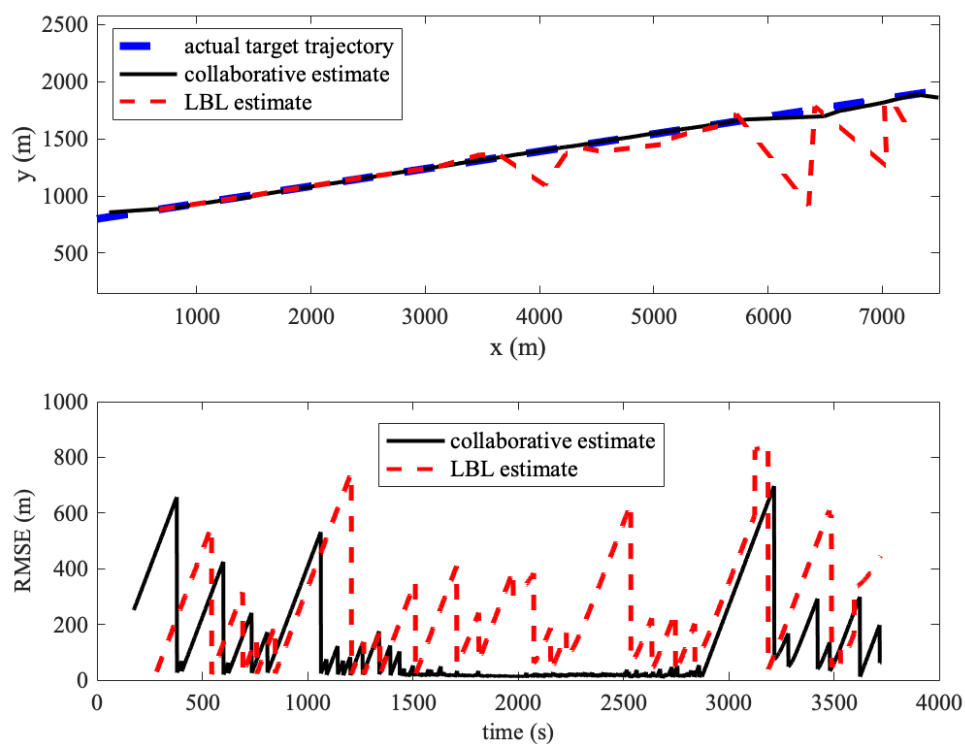
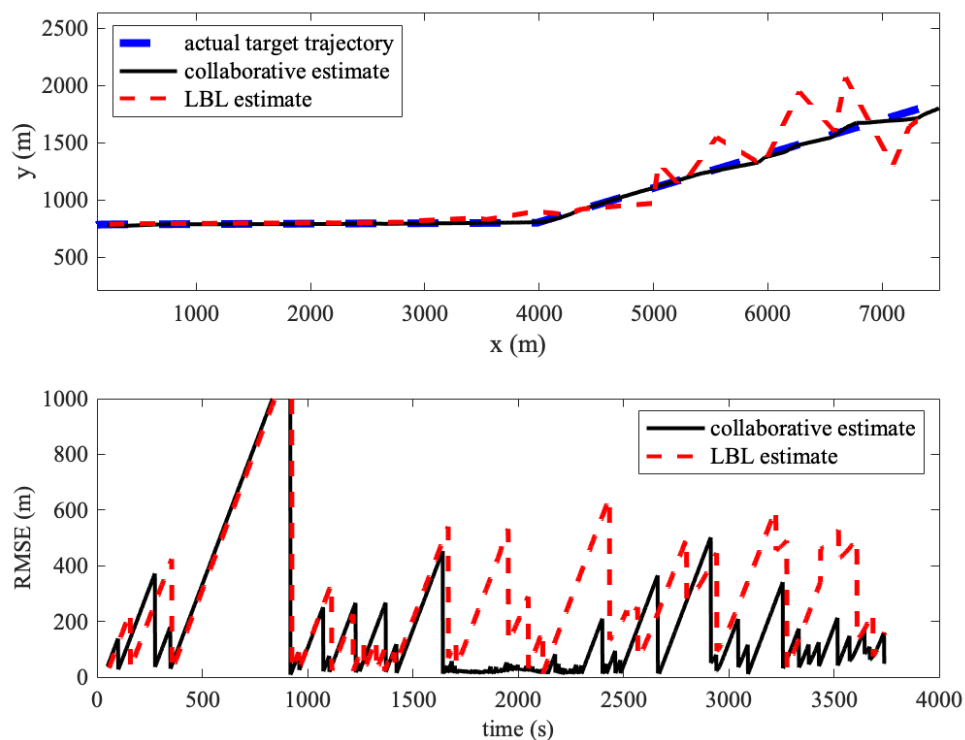Figure A.32: Trial #33

Figure A.33: Trial #34

Figure A.34: Trial #35
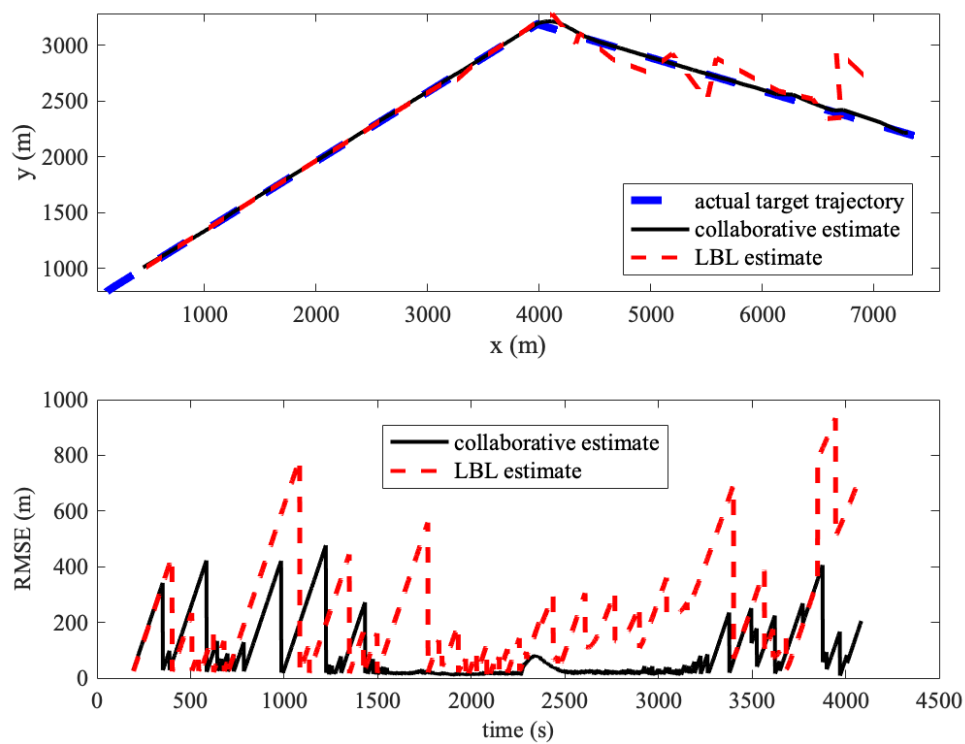

Figure A.35: Trial #36

Figure A.36: Trial #37