International Journal of Production Research

ISSN: 0020-7543 (Print) 1366-588X (Online) Journal homepage: www.tandfonline.com/journals/tprs20

A hierarchical reinforcement learning approach for real-time berth allocation and quay crane scheduling

Seongbae Jo & Ilkyeong Moon

To cite this article: Seongbae Jo & Ilkyeong Moon (2025) A hierarchical reinforcement learning approach for real-time berth allocation and quay crane scheduling, International Journal of Production Research, 63:24, 10027-10052, DOI: 10.1080/00207543.2025.2542518

To link to this article: https://doi.org/10.1080/00207543.2025.2542518

Published online: 05 Aug 2025.

INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH
2025, VOL. 63, NO. 24, 10027–10052
https://doi.org/10.1080/00207543.2025.2542518

Taylor & Francis
Taylor & Francis Group

A hierarchical reinforcement learning approach for real-time berth allocation and quay crane scheduling

Seongbae Jo[a] and Ilkyeong Moon [a,b]

[a]Department of Industrial Engineering, Seoul National University, Seoul, Republic of Korea; [b]Institute of Engineering Research, Seoul National University, Seoul, Republic of Korea

ABSTRACT

Efficient and low-carbon berth allocation and quay crane scheduling are crucial for enhancing the competitiveness and sustainability of container ports. This study addresses the berth allocation and quay crane assignment and scheduling problem (BACASP) by incorporating carbon emission costs and uncertainties in vessel arrival times and quay crane processing times. A novel hierarchical reinforcement learning (HRL)-based scheduling framework is proposed, employing three cooperative agents to support real-time berth allocation and quay crane scheduling in dynamic environments. The upper-level agent determines whether to release waiting vessels, while two lower-level agents allocate berth locations and assign quay cranes. Numerical experiments demonstrate the effectiveness of the HRL framework compared to a mixed integer programming (MIP) approach with perfect information, highlighting its capability to achieve near-optimal solutions under sequentially observed information. The study also investigates the impact of uncertainty on operational and carbon emission costs, providing practical managerial insights for port operators. These findings underscore the potential of leveraging well-structured HRL frameworks to address complex and dynamic port operation problems.

ARTICLE HISTORY
Received 27 January 2025
Accepted 22 July 2025

KEYWORDS
Berth allocation and quay crane assignment and scheduling problem; hierarchical reinforcement learning; carbon emissions; real-time decision-making; maritime industry

1. Introduction

Nowadays, a considerable portion of global trade is transported via maritime transportation. The trade volume carried through maritime shipping increased by 2 percent in 2024 and is projected to grow steadily at an annual rate of 2.4 percent over the next five years (UNCTAD 2024). Furthermore, UNCTAD (2024) also anticipated that the growth rate of the cargo volume transported via containers would be even higher. As a result, the significance of efficient container port operations is being highlighted.

A berth is where a vessel stays while its containers are loaded or unloaded, and a quay crane (QC) is the equipment that performs these operations, as illustrated in Figure 1. Vessels are berthed parallel to the quay, and QCs move along the quay to perform container operations. In Figure 1, each rectangle enclosing a vessel indicates the space and time it occupies at the allocated berths. As shown in the bottom-right corner, the berthing period, indicated by the horizontal length, is calculated based on the container workload and the number of deployed QCs for each vessel. Given that QCs cannot move across each other, their schedules, represented by black arrows, are visualised without overlaps.

A port manager constructs a baseline schedule for the given planning horizon (e.g. 168 h) based on available information, such as each vessel's expected arrival times and expected workload. The baseline schedule includes the berth-vessel assignments and the scheduling of each QC for container operations, as shown in Figure 1. This schedule determines the dwelling times of vessels, which significantly impacts not only the productivity of the port but also berth operations at the subsequent ports where the vessels are scheduled to arrive. It also has a considerable impact on yard operations, including container storage and the scheduling of internal and external trucks. For example, based on the vessels' berthing location, the storage location of containers within the yard is determined, and the schedules for vehicles transporting the containers are constructed according to the vessels' berthing and departure times. Therefore, establishing an effective baseline schedule for berths and QCs is a critical decision in port operations.

However, unexpected weather changes, variations in workload, and other uncertainties make it difficult to

CONTACT Ilkyeong Moon ikmoon@snu.ac.kr Department of Industrial Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea Institute of Engineering Research, Seoul National University, Gwanak-ro 1, Gwanak-gu, Seoul 08826, Republic of Korea

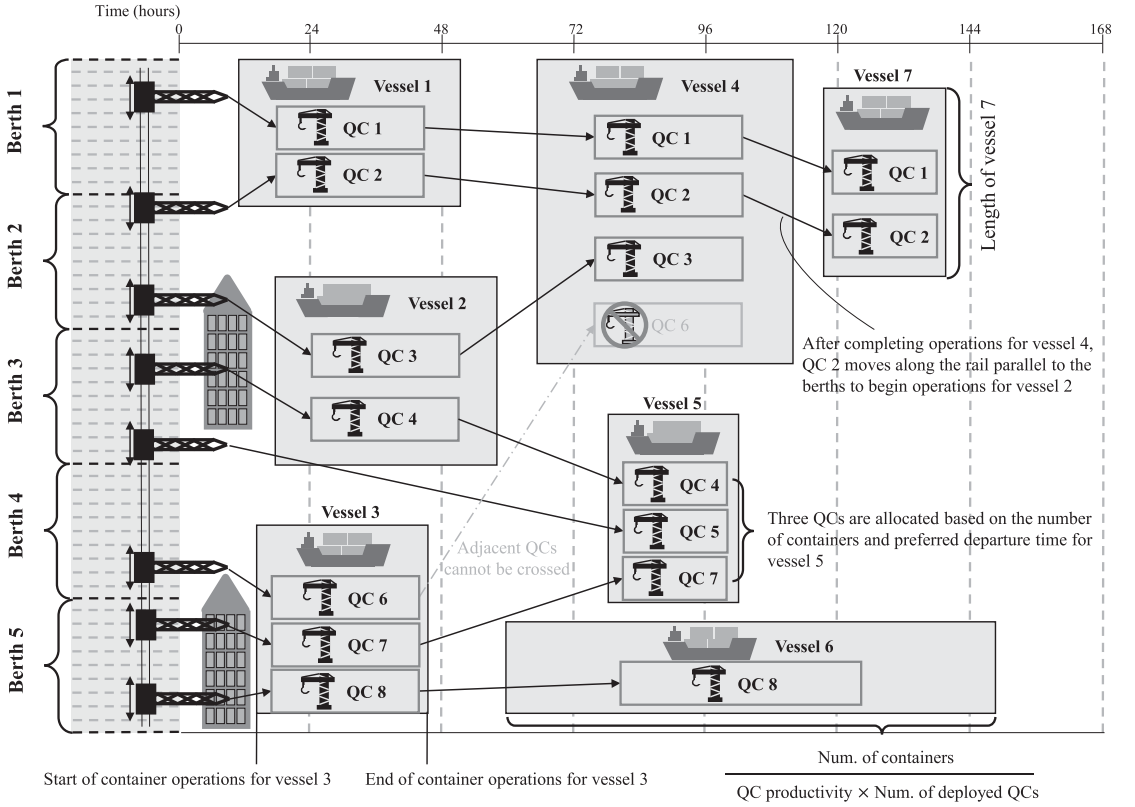


Figure 1. Continuous layout of a container terminal with five berths and eight QCs.

implement the baseline schedule as planned. As a result, the port manager must take corrective actions, such as adjusting the berthing location of delayed vessels or deploying more QCs than the initial plan. Figure 2 shows a real-world example of berth allocation and QC schedule adjustments at Busan Port in South Korea. In Figure 2, the vertical axis represents the berths, while the horizontal axis represents the time horizon. The light gray rectangles indicate vessels, and the assigned QCs for each vessel are shown in dark gray. For each QC, the number in the top-left corner indicates the start time of operations, while the number in the top-right corner represents the completion time. The time difference between the baseline schedule and the modified schedule is 4 hours. It can be observed that both the berthing locations and assigned QCs for Vessel 7 and Vessel 12 have been changed. For Vessel 10, the overall schedule has been delayed due to the departure delay of Vessel 9. Such schedule changes can occur as little as one hour before a vessel is berthed or even less. In this context, inadequate schedule adjustments can lead to increased vessel dwelling times or inefficient QC operations, which in turn increase the port's operating costs and carbon emissions from both vessels and QCs. If vessels depart the port later than their preferred departure times, penalty costs are incurred. Therefore, making appropriate real-time decisions regarding berth allocation and QC scheduling in response to uncertainty is an essential capability for port operators.

The problem of determining the berthing locations of vessels is known as the berth allocation problem (BAP), and the problem of determining which vessel each QC will handle in specific time slots is referred to as the QC scheduling problem (QCSP). As the decisions made in each problem significantly impact the outcomes of the other, their integrated optimisation problem, known as the berth allocation and QC assignment and scheduling problem (BACASP), has been widely studied in recent literature. Our study also aims to propose a new approach for solving the BACASP, taking into account several considerations outlined below.

Because the berth allocation and QC scheduling are the starting point for all container operations in ports, their appropriate implementation is essential for port efficiency and serves as a key factor of container terminals' competitiveness (Bierwirth and Meisel 2015; Yang, Wang, and Li 2012). However, as mentioned above, the inherent uncertainty in the maritime logistics industry makes it challenging (Rodrigues and Agra 2022; Tan and He 2021; Tasoglu and Yildiz 2019). Due to various types of uncertainty, schedules of berths and QCs often require real-time adjustments. Consequently, it is necessary to develop proactive or reactive approaches to determine

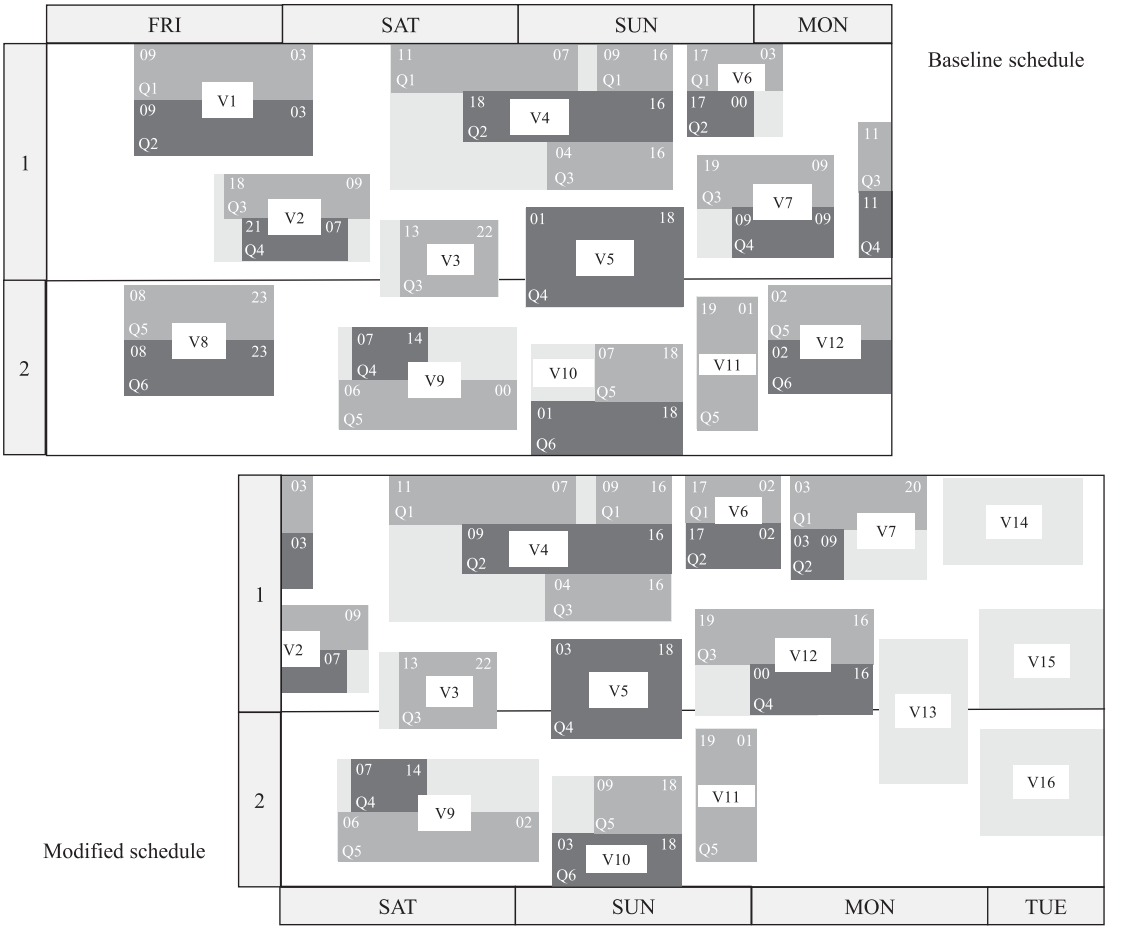


Figure 2. A real-world example of baseline and modified schedules.

schedules of berths and QCs that consider those uncertainties.

Another important issue in port operations is sustainability. Because the maritime logistics industry accounts for approximately 3 percent of global greenhouse gas (GHG) emissions (UNCTAD 2023), many international organisations and individual ports are striving to reduce GHG emissions from port operations and vessels. For example, the International Maritime Organization (IMO) adopted a long-term strategy for reducing GHG emissions from international shipping and related activities (International Maritime Organization 2018). To align with this international trend, port operators can replace the power sources of cargo-handling equipment with eco-friendly sources or implement incentive programs to encourage vessels to reduce GHG emissions. In addition, advancement in operational strategies to achieve both the maximisation of port productivity and the minimisation of GHG emissions is essential.

Although effectively managing uncertainty and reducing GHG emissions are both critical objectives in port operations, these two aspects have rarely been considered simultaneously in existing BACASP literature. Consequently, little attention has been given to analysing how uncertainty in port environments impacts operational decision-making, especially when GHG emissions are considered. To fill this research gap, we introduce a new dynamic problem that enables real-time decision-making in the BACASP context. Furthermore, we incorporate the carbon emission costs into the objective function, aiming to produce optimal decisions that minimise the total cost, encompassing both operational expenses and environmental impacts. To solve this challenging problem, we develop a novel reinforcement learning (RL) algorithm tailored to uncertain and dynamically changing port environments. We also validate its practical applicability through numerical experiments. Finally, leveraging our proposed approach, we provide valuable managerial insights to port operators responsible for operational decision-making, aiming to enhance ports' competitiveness in the evolving maritime logistics industry. More detailed explanations of our contributions to existing literature are discussed at the end of Section 2.

The remainder of this paper is organised as follows. Section 2 summarises previous studies related to this study and our contributions to the literature. Section 3 presents a problem description of the BACASP. We first explain the deterministic BACASP considering carbon emission costs and its mathematical model. Then, a real-time decision-making problem is described, in which several problem settings of the deterministic problem are modified. In Section 4, a novel RL-based approach to solve the real-time decision-making problem is proposed. Section 5 includes the results of numerical experiments, and concluding remarks are provided in Section 6.

2. Literature review

In this section, we review two distinct bodies of literature relevant to our study. The first focuses on studies addressing the BACASP, particularly under the considerations of uncertainty or carbon emissions. The second investigates RL approaches for tackling complex decision-making problems in manufacturing and logistics environments. Finally, contributions of this study to the existing literature will be provided.

Because the BAP and QCSP are NP-hard (Xu and Lee 2018; Zhu and Lim 2006), many studies have proposed efficient solution methods to address the BACASP, which integrate two computationally challenging optimisation problems. Among these, we briefly introduce studies that consider uncertainty or incorporate carbon emissions into their analysis. Although the berth allocation and QC assignment problem (BACAP) is a simpler version of the BACASP (Zheng et al. 2019), it is still a challenging problem, and therefore, it is included in the studies reviewed in this section.

The most commonly considered sources of uncertainty are the arrival times of vessels and the processing times of QCs. Xiang, Liu, and Miao (2018) considered uncertainties in vessel arrival times and QC processing times simultaneously in the BACASP. The arrival of unscheduled vessels and breakdown of QCs are also considered, and they utilised a rolling horizon optimisation algorithm to handle the computational complexity of the BACASP. In Tasoglu and Yildiz (2019), only QC processing times are considered as the source of uncertainty. They utilised the simulation model to handle complex constraints for QC movements and uncertainty in QC handling times. More recently, the main research stream in addressing the BACAP and BACASP under uncertain vessel arrival times or QC processing times has focussed on proposing mixed integer program (MIP)-based problem formulations and on developing efficient algorithms to solve them. Although uncertainty was not considered in Agra and Oliveira (2018), they proposed a new formulation for the deterministic BACASP, which can serve as a good starting point for developing MIP models that incorporate uncertainty. Furthermore, a rolling horizon heuristic algorithm was employed to solve large-scale instances. Based on this deterministic model, Rodrigues and Agra (2021) proposed a new robust formulation and decomposition algorithm that can address uncertainty in vessel arrival times. In Chargui et al. (2023), a robust BACASP model and a decomposition algorithm for solving it were also proposed. The key difference from the Rodrigues and Agra (2021) lies in the consideration of both uncertainties in vessel arrival times and QC processing times, as well as in the incorporation of energy price variations. Similarly, Zhen, Zhuge, Wang et al. (2022) developed a two-stage stochastic programming model that integrates berth and yard space allocation under uncertain vessel arrival times and handling workloads. They also proposed a new decomposition algorithm to solve the model efficiently. While their work focuses on yard allocation rather than QC assignment, both the modelling and solution approaches are closely aligned with the BACAP and BACASP. Zhen, Sun et al. (2021) and Ji, Huang, and Samson (2022) also formulated the BACASP as a scenario-based stochastic model to incorporate uncertainty in vessel arrival times. The former developed a column generation-based solution method, and the latter employed an enhanced non-dominated sorting genetic algorithm II to solve the proposed problems. In Xiang and Liu (2021), an almost robust model for the BACAP and a decomposition method are proposed to address uncertainty in vessel arrival times. In C. Wang, Liu et al. (2024), a distributionally robust optimisation (DRO) model was proposed to overcome the drawbacks of stochastic programming and robust optimisation methods for the BACAP under uncertain vessel arrival times. Following this line of research, C. Wang, Wang et al. (2025) developed a two-stage DRO approach, which demonstrated improved computational performance compared to the previous study.

However, many of the solution methods proposed in the aforementioned studies are generally effective in uncertain but static port environments. They are limited in their ability to support real-time decision-making in highly dynamic port environments. For instance, scenario-based two-stage stochastic models, in which baseline schedules for berths and QCs are determined in the first stage and modified in the second, assume that all uncertainties in the planning horizon are realised simultaneously rather than sequentially. Other two-stage models, where berthing positions of vessels are decided in the first stage and QC schedules are determined in the second stage, also have limitations in that the berthing positions cannot be modified even when changes in vessel

arrival times are observed. Furthermore, many papers have adopted rolling horizon heuristic algorithms to address large-scale instances, but these approaches result in a smaller number of vessels or shorter time periods being considered at a single decision point. Therefore, as demonstrated in Zhen, He et al. (2024), which proposed a multi-stage stochastic integer programming model for berth planning under uncertain vessel arrival and service times, it is important to develop practical decision-making frameworks that account for the sequential realisation of uncertainty throughout the planning horizon.

Carbon emissions, which have become increasingly important in the maritime logistics industry, have also been considered in numerous studies addressing optimisation problems in port operations (Jauhar et al. 2023; Jiang et al. 2024; Karakas, Kirmizi, and Kocaoglu 2021; Peng, Dong et al. 2021; Peng, Wang et al. 2016; Venturini et al. 2017). Among them, several studies addressed the QC assignment problem (QCAP) or QCSP with considerations of carbon emissions by QCs (Kenan, Jebali, and Diabat 2022; Liu and Ge 2018). Carbon emissions are also considered in the BACAP literature, where QC operations are the primary sources of emissions. Wang, Wang, and Meng (2018) incorporated several carbon emission taxation policies into the BACAP and investigated their impact on the problem. Similarly, T. Wang et al. (2020) proposed a bi-objective model considering carbon emission taxation and developed an efficient algorithm to solve it. They also evaluated the trade-off between service efficiency and carbon emissions. In Yu et al. (2023), the BACAP was formulated in conjunction with the vessel speed optimisation problem, taking into account the adoption of green technologies aimed at reducing vessel emissions. In addition, Wang, Hu, and Zhen (2024) addressed the BACAP considering the assignment of onshore power supply, a technology for reducing carbon emissions of ports and vessels.

However, uncertainty in port operations was not incorporated into those models. To the best of our knowledge, only a few studies, such as Chargui et al. (2023), Chargui, Zouadi, and Sreedharan (2023), and Zhen, Sun et al. (2021), have simultaneously considered both carbon emissions and uncertainty in the context of the BACAP or BACASP. This highlights the need for further development of mathematical models that integrate these two critical factors within the context of the BACASP. It enables exploring how uncertainty affects low-carbon berth allocation and QC scheduling, providing valuable insights into more sustainable and optimised port operations. One point to consider is that, because carbon taxes have not yet been directly imposed on ports, it remains unclear which emission sources they are responsible for. In this study, because berthing and waiting times of vessels are also determined by port operators, we consider the carbon emission costs from vessels in the berths and roadstead as part of the ports' emission costs, as in Kenan, Jebali, and Diabat (2022).

Incorporating uncertainty, particularly in vessel arrival times and QC processing times, along with carbon emissions resulting from port operations, makes the BACASP an even more challenging problem in terms of computational complexity. Furthermore, as in T. Wang et al. (2020), constraints for complex QC movements and their interference effects must be included to measure the amount of carbon emissions from QCs accurately. To address these challenges, we propose a novel RL-based scheduling framework capable of providing real-time decision-making in dynamic and complex port environments. It can leverage historical data to train the agent to optimise the objective over the entire planning horizon while accounting for the sequential realisation of uncertainty. According to Filom, Amiri, and Razavi (2022), several studies have utilised RL algorithms to address optimisation problems in port operations. However, to the best of our knowledge, they have not been explored in the BACASP literature, where multiple decisions must be made jointly. In contrast, RL using multiple collaborative agents has been frequently applied to complex decision-making problems in the logistics and production domains. We briefly introduce studies related to the solution method proposed in our study in the next paragraph.

To handle complex real-time decision-making problems, several studies have employed multiple cooperative agents. In H. Wang et al. (2021), a dual Q-learning method was proposed to address assembly job shop scheduling problems. Their top-level Q-learning agent chooses one of the jobs in a global job buffer; then the job is automatically assigned to a machine with the minimum loading. Next, the bottom-level agent determines the priority of jobs assigned to each machine. Ma et al. (2021) also employed two RL agents for dynamic pickup and delivery problems: the upper-level agent decides whether to release accumulated orders for processing, while the lower-level agent determines the sequence in which each vehicle processes these released orders. In Liu, Piplani, and Toro (2022), two cooperative agents were trained to allocate arriving jobs to machines and to determine the sequence of jobs to be processed by each machine in flexible job shop scheduling problems (FJSPs). Additionally, Zhang et al. (2024) developed a collaborative agent RL for FJSPs, where the job agent selects one of the candidate operations, and the machine agent determines which machine will execute the selected operation. In Lei

et al. (2023), three RL agents were trained within a hierarchical structure. First, the upper-level agent determines whether to release the cached jobs to the lower-level agents. Next, the first lower-level agent, referred to as the job agent, selects one of the released jobs, while the second lower-level agent, referred to as the machine agent, assigns each job to one of the compatible machines. The BACASP requires more complex decision-making structures compared to the problems addressed in the previous studies explained above. Therefore, we construct a well-designed hierarchical RL (HRL)-based scheduling framework in which three agents cooperate to make major decisions, while several detailed decisions are made based on predefined rules. Its detailed explanation will be presented in Section 4.

Finally, our contributions to the existing literature can be summarised as follows:

- We introduce a new real-time BACASP, in which schedules of berths and QCs are sequentially determined based on realised information to manage uncertainty in port environments. By addressing this problem, port operators can determine feasible schedules at each period that minimise the expected total cost incurred over the entire planning horizon. Furthermore, we incorporate carbon emission costs into the proposed problem to explore the relationship between uncertainty and carbon emission costs, which are critical factors in port operations.
- We develop a novel HRL-based scheduling framework, which consists of three cooperative RL agents, capable of supporting real-time berth allocation and QC scheduling. The effectiveness of the proposed framework is demonstrated through comparisons with an exact algorithm. Additionally, its practical applicability is verified in the aspect of consistency and computation time.
- Managerial insights for practitioners involved in port operations are provided through numerical experiments. These insights enable ports to enhance their sustainability and competitiveness by utilising the proposed HRL-based scheduling framework.

3. Problem description and mathematical model

3.1. Deterministic BACASP

In this subsection, we present a problem description for the deterministic BACASP in a static manner. In the deterministic BACASP, the arrival times of all vessels and completion times of all QCs are known to a decision-maker a priori, and they remain unchanged. Based on the known information, berthing locations of waiting vessels and a set of specific QCs to serve each vessel are determined in an integrated manner. We use a continuous layout, where berths are divided into smaller units called berth sections, and a single vessel can be moored across multiple consecutive berth sections, as shown in Figure 1. QC schedules are time-invariant, meaning that QCs assigned to a vessel cannot be reassigned to another vessel until the assigned vessel departs, and an idle QC cannot be assigned to a vessel currently being served by other QCs. All QCs move along a straight line and each QC operates within a specified range without intersecting with one another as described in Chung and Chan (2013). Furthermore, we assume that there is QC interference. It means that as the number of QCs working on a single vessel increases, the productivity of each QC decreases due to interference among QCs.

The objective is to minimise the total costs that include operational costs and carbon emission costs. The operational costs include the earliness income and tardiness penalty for vessel departure times, as well as the QC operating costs. The carbon emission costs occur from three sources of emissions: waiting vessels, berthing vessels, and operating QCs. We assume that the carbon emission costs are proportional to vessels' waiting time, berthing time, and QC operation time. This assumption aligns with the scenario in which a carbon tax is imposed on the carbon emissions generated by vessels and QCs consuming their fuel. Based on Agra and Oliveira (2018), Rodrigues and Agra (2021), and T. Wang et al. (2020), we construct a mathematical model of the deterministic problem described above. We provide it in Appendix 1.

3.2. Real-time decision-making problem

Due to highly uncertain arrival times of vessels and completion times of QC operations, schedules calculated from the deterministic model cannot be directly implemented in real-world situations. Therefore, a dynamic scheduling framework that can sequentially generate the best schedule under the realised information must be developed. To achieve this, we modify the problem setting from the deterministic BACASP as follows. We first assume that uncertainty exists in vessels' arrival time and completion time of QC operations. More specifically, the arrival time of vessel k, $\overline{A}_k$, and the departure delay of vessel k caused by the delay in QC operations, $\overline{D}_k$, are no longer known in advance. Therefore, the decision-maker utilises only the realised observations, including information related to vessels that have already arrived and QCs that have already completed their tasks. Based on the observations, the decision-maker selects one of the

waiting vessels in the roadstead, determines its berthing location, and assigns idle QCs to perform its container operations.

Because the allocation of berth sections and QCs at each decision point significantly impacts subsequent decisions, the decision-maker has to minimise not only the immediate costs but also the total costs incurred over the entire planning horizon. To this end, we propose a sequential decision-making framework based on reinforcement learning (RL). In the proposed framework, the decision-maker dynamically performs berth allocation and QC scheduling in real time, without relying on a predefined baseline schedule. In each time period, the decision-maker determines whether to allocate new berths and assign new QCs to a single vessel. A more detailed explanation of the sequential decision-making process in the real-time BACASP is provided in Section 4.

4. RL-based decision-making framework for the BACASP

To address the real-time decision-making problem explained in Section 3.2, we propose a new RL-based scheduling framework in this section. Recently, a vast amount of data has been accumulated in the maritime logistics industry, and many studies have been conducted to utilise this historical data in addressing inefficiencies caused by uncertainty. In the context of the BACASP, we also recognise the potential of utilising historical data to effectively address issues stemming from uncertainty. Due to the inherent unpredictability of vessel arrival times and QC processing times, the need for real-time berth allocation and QC scheduling is becoming increasingly prevalent. Once we train the RL agents in a virtual port environment simulated using historical data, the RL agents can generate decisions within a significantly short amount of time based on realised information. Moreover, even when incorporating complex constraints, such as those related to QC movement, the RL agents can still be trained within a reasonable computation time. Therefore, an RL-based decision-making framework can support real-time decision-making aimed at minimising the expected cost in highly uncertain and dynamic real-world port environments. These considerations motivated us to develop a new RL-based solution approach for the BACASP.

4.1. HRL-based scheduling framework

In the BACASP, multiple decisions must be made in an integrated manner. However, making multiple decisions is quite difficult for a single agent. Therefore, we employ multiple agents, consisting of one upper-level agent and two lower-level agents, that take actions in a hierarchical structure. As described in Section 2, such decision-making structures are commonly found in previous studies that utilised RL to address complex real-time decision-making problems in the logistics and production environments. While not employing multiple RL agents, similar hierarchical decision-making structures have been applied to complex optimisation problems in the maritime logistics domain (Yu et al. 2023; Zhen, Zhuge, Zhang et al. 2024). For example, Zhen, Zhuge, Zhang et al. (2024) developed a two-level optimisation model, where the upper-level model determines the layout of emission control areas (ECAs), and the lower-level model solves the vessel routing problem given the ECA width and sulfur content limits. These prior studies motivate the design of our hierarchical decision-making framework based on RL.

Figure 3 shows the hierarchical structure of our three RL agents. The port environment is a discrete-time simulation framework constructed with berth sections, QCs, and vessels as the main entities of interaction. It maintains and updates the current state of each entity, which includes the state representations used in the MDP formulations of the RL agents described in Sections 4.2 and 4.3. For instance, the vessel entity is characterised by various attributes such as its arrival time, workload, berthing time, and the assigned QCs and berth sections. The port environment updates these attributes in each time period considering the actions taken by RL agents. Additionally, it computes costs based on the updated attributes and uses them to generate rewards for the RL agents.

The cooperation among the three RL agents in the port environment can be explained as follows. First, the upper-level agent decides whether to release one of the vessels waiting for berthing. Suppose the upper-level agent does not release any vessels, as described by arrow 1. In that case, the two lower-level agents, the QC scheduler and the berth scheduler, do nothing, and only the port operations that are currently in progress continue to be executed. Otherwise, when the upper-level agent decides to release one of the waiting vessels, a vessel to be berthed is selected based on a predefined rule. We employ the first-in-first-out (FIFO) rule in this study. When preferred berthing locations are not considered, releasing a later-arriving vessel before an earlier one may be perceived as unfair and may cause dissatisfaction for the earlier vessel. For such a decision to be justifiable from the perspective of port operational efficiency, the later-arriving vessel must fit into the available berth space, while the earlier-arriving vessel does not. Nonetheless, our preliminary numerical experiments indicated that

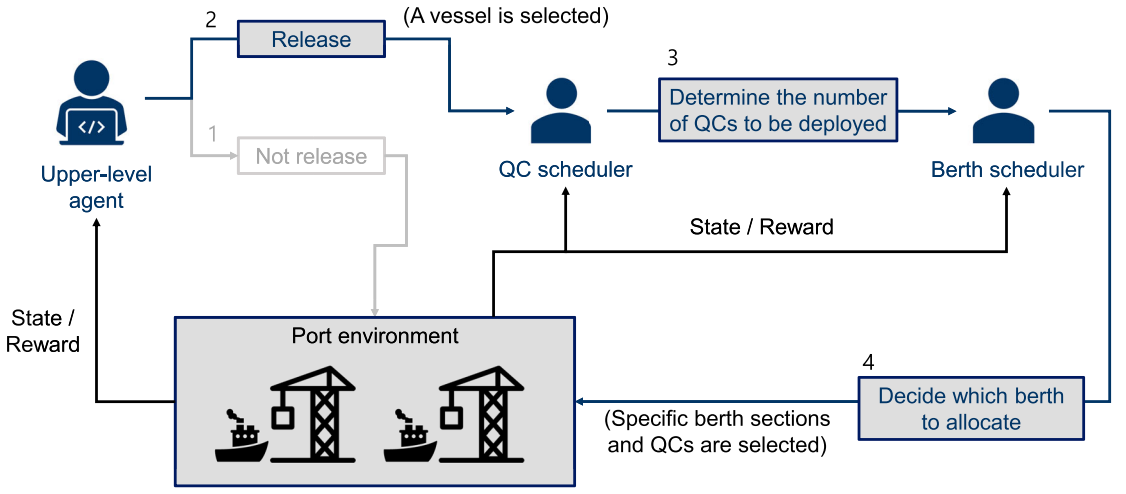


Figure 3. HRL-based scheduling framework.

such situations occurred only rarely. Therefore, we conclude that the FIFO rule remains a reasonable and fair policy for both vessels and port operators. After the vessel to be released is selected, the QC scheduler determines how many QCs to assign to it, as shown in arrow 3. The berth scheduler selects the berthing location of the vessel as described in arrow 4. Because the number of berth sections considered in the BACASP is usually in the range of dozens, it is challenging to match each berth section to an action of the berth scheduler. To address this issue, we define its action as selecting one of the berths rather than a berth section. Once a berth is selected by the berth scheduler, the first feasible berth section within the selected berth is set as the starting berthing location of the released vessel. Finally, specific QCs that can move to the determined berthing location are deployed in ascending order of their identifiers (IDs). All these processes are implemented by updating the attributes of the vessel, QC, and berth section entities within the port environment.

Note that the decisions made by each agent are written inside boxes, while the operations automatically executed according to predefined rules are written inside parentheses in Figure 3. These operations could also be executed by the QC scheduler or berth scheduler; however, because it complicates their action space, we use the predefined rules instead to enhance the performance of the proposed framework. By repeating the explained process in each time period until all vessels are handled within the planning horizon, the complex decision-making in the BACASP can be accomplished through the cooperation of three agents. We formulate this as a Markov decision process (MDP) for each agent, and these are explained in the following subsection.

4.2. Upper-level agent

4.2.1. MDP formulation

According to the workflow described in the previous section, the MDP of the upper-level agent can be formulated as follows. The state of the upper-level agent, s_t^u, includes the proportion of empty berth sections, the proportion of idle QCs, the number of waiting vessels, and properties of the earliest-arrived vessel in the roadstead. These properties consist of its length, the number of containers to be processed, and the remaining time until its requested departure time. We selectively use these abstracted information of the port environment as state variables to ensure a fixed state dimension while improving the performance of the training algorithm. The action a_t^u is determining whether to release one of the waiting vessels. The notation $a_t^u = 1$ corresponds to releasing, while $a_t^u = 0$ indicates not releasing. The reward r_t^u is the negative value of the sum of the carbon emission costs and operating costs incurred during the current period by all vessels and QCs staying in the port. Using this reward function, the upper-level agent can be trained to achieve our ultimate goal: minimising the sum of operational and carbon emission costs incurred over the entire planning horizon.

4.2.2. Training algorithm

To train the upper-level agent, we use the double deep Q-network (DDQN) algorithm proposed by Van Hasselt, Guez, and Silver (2016). Q-function is parameterised using the multilayer perceptron (MLP), in which ϕ^u and $\bar{\phi}^u$ are parameters of the Q-network and the target Q-network, respectively. At each time period t, we store the transition data $(s_t^u, a_t^u, r_t^u, s_{t+1}^u)$ in the replay buffer D.

Note that even if $a_t^u = 1$, when no feasible actions for the lower-level agents exist, it is masked by $a_t^u = 0$. In the kth update, the transition data $(s, a, r, s') \sim U(D)$ are uniformly sampled, and the Q-network is updated via the stochastic gradient descent (SGD) with the following loss function:

$$L(\phi_k^u) = \mathbb{E}_{(s,a,r,s')\sim U(D)}\left[\left(r + \gamma\, Q(s', \arg\max_{a'} + Q(s', a'; \phi_k^u); \bar{\phi}_k^u) - Q(s, a; \phi_k^u)\right)\right] \quad (1)$$

where γ is the discount factor. The target network is updated every E_u training episodes by copying the parameters of the original Q-network.

4.3. Lower-level agents

4.3.1. MDP formulation of the QC scheduler

When the upper-level agent releases one of the waiting vessels, the QC scheduler takes an action and receives a reward. Its state s_t^q contains the number of containers to be processed in the released vessel and the time remaining until the requested departure time of the vessel. Furthermore, the state variables of the upper-level agent are also included. The action a_t^q is determining the number of QCs to be deployed for the selected vessel. Because the maximum number of QCs that can be assigned to a single vessel simultaneously is $\mathcal{N}^{QC}$, we let $a_t^q \in \{1, \ldots, \mathcal{N}^{QC}\}$. The reward r_t^q is the negative value of the estimated sum of carbon emission costs and operating costs incurred by the released vessel and QCs working on it. We used the estimated costs, assuming no uncertainty in $\overline{D}_k$, because the actual carbon emissions can be calculated after the vessel leaves the port and the QCs complete their work.

4.3.2. MDP formulation of the berth scheduler

After the QC scheduler takes an action, the berth scheduler also takes its action and gets a reward. The state of the berth scheduler s_t^b consists of the number of idle QCs in each berth section and the number of QCs that will be deployed. Its action a_t^b is to determine the berth to be assigned to the released vessel. For the berth scheduler, because which vessel to be berthed and the number of QCs to be deployed have already been determined by the upper-level agent and the QC scheduler, its action does not affect the immediate costs. However, its action ultimately affects which specific QCs will be deployed. Therefore, we design the reward function for the berth scheduler to leave as many feasible QCs as possible for the next vessel rather than directly minimising costs. To achieve this, we use a binary matrix F representing the feasibility of each QC being deployed to each berth section, where $F_{i,j} = 1$ if QC i can be deployed to berth section j. We define the reward r_t^b as the difference between the total sums of the elements in the matrix F before and after the action of the berth scheduler.

4.3.3. Training algorithm for the lower-level agents

Unlike the upper-level agent, the lower-level agents take their actions and receive rewards in their respective environments that are more sensitive to the actions of other agents. Therefore, we employ the proximal policy optimisation (PPO) proposed by Schulman et al. (2017). It is one of the most widely used policy gradient algorithms, ensuring more stable learning for our lower-level agents in their non-stationary environments. Lei et al. (2023) proposed the multi-PPO algorithm, where two lower-level agents share a single state value function. However, in our numerical experiments, we observed that this approach is not effective for our problem. Therefore, we allow the two lower-level agents to have independent value functions. Policy π and value function v of each lower-level agent are also parameterised using the MLPs. We denote θ^q and ϕ^q as the parameters of the policy and value networks for the QC scheduler, and θ^b and ϕ^b as those for the berth scheduler.

Because each lower-level agent has a discrete action space, the number of output nodes in the policy network is equal to the size of each agent's action space, and each node outputs the logit of the corresponding action. Then, the logits of infeasible actions are masked by $-\infty$, and action probabilities are calculated using the softmax function. Unlike the upper-level agent, the lower-level agents do not use the transition data for their training if either of them has no feasible actions. As a result, the number of transition data used in a single episode is equal to the number of vessels arriving within the planning horizon. In this case, the amount of batch data for training the agents using the PPO algorithm is insufficient. To address this issue, we gather data from a considerable number of E_l episodes and perform a batch update. Finally, the policy and value networks of the QC scheduler and berth scheduler are updated using individual loss functions. The loss functions for the QC scheduler are computed using Equations (2) and (3) with ϵ, γ, and λ as the hyperparameters of the PPO algorithm.

$$L_{policy}(\theta^q) = \hat{\mathbb{E}}_t\left[\min(p_t^q(\theta^q)\hat{A}_t^q, \text{clip}(p_t^q(\theta^q), 1-\epsilon, 1+\epsilon)\hat{A}_t^q\right] \quad (2)$$

$$L_{value}(\phi^q) = \hat{\mathbb{E}}_t\left[\left(\sum_{t'=t}^{T} \gamma^{t'} r_{t'}^q - v_{\phi^q}(s_t^q)\right)^2\right], \quad (3)$$

where $\hat{A}_t^q$ and $p_t^q(\theta^q)$ represent its advantage function and probability ratio, respectively. We present them using Equations (4) to (6). Here, θ_{old}^q indicates the parameters of the policy network before the update.

$$\hat{A}_t^q = \delta_t^q + (\gamma\,\lambda)\delta_{t+1}^q + \cdots + (\gamma\,\lambda)^{T-t+1}\delta_{T-1}^q \tag{4}$$

$$\delta_t^q = r_t^q + \gamma\, v_{\phi^q}(s_{t+1}^q) - v_{\phi^q}(s_t^q) \tag{5}$$

$$p_t^q(\theta^q) = \frac{\pi_{\theta^q}(a_t^q \mid s_t^q)}{\pi_{\theta_{old}^q}(a_t^q \mid s_t^q)} \tag{6}$$

The loss functions for the berth scheduler can be calculated using Equations (7) to (11). We omit the detailed explanation for them because they share the same structure as those of the QC scheduler.

$$L_{policy}(\theta^b) = \hat{\mathbb{E}}_t\Big[\min(p_t^b(\theta^b)\hat{A}_t^b, \mathrm{clip}(p_t^b(\theta^q), 1-\epsilon, 1+\epsilon)\hat{A}_t^b\Big] \tag{7}$$

$$L_{value}(\phi^b) = \hat{\mathbb{E}}_t\left[\left(\sum_{t'=t}^{T}\gamma^{\,t'} r_{t'}^b - v_{\phi^b}(s_t^b)\right)^2\right] \tag{8}$$

$$\hat{A}_t^b = \delta_t^b + (\gamma\,\lambda)\delta_{t+1}^b + \cdots + (\gamma\,\lambda)^{T-t+1}\delta_{T-1}^b \tag{9}$$

$$\delta_t^b = r_t^b + \gamma\, v_{\phi^b}(s_{t+1}^b) - v_{\phi^b}(s_t^b) \tag{10}$$

$$p_t^b(\theta^b) = \frac{\pi_{\theta^b}(a_t^b|s_t^b)}{\pi_{\theta_{old}^b}(a_t^b \mid s_t^b)} \tag{11}$$

The overall process in which the upper-level agent, the QC scheduler, and the berth scheduler are trained using their respective training algorithms is presented in Algorithm 1.

5. Numerical experiments

Using the proposed HRL-based scheduling framework, we conduct numerical experiments to answer the following research questions:

(i) Can the proposed framework generate effective berth and QC schedules with limited observations in the port environment?
(ii) Is the utilised training algorithm practically applicable to support real-time decision-making in real-world port environments?
(iii) How does the level of uncertainty affect decision-making in the BACASP?

In Section 5.1, we aim to verify the effectiveness of the proposed HRL algorithm by comparing it with the MIP approach. Furthermore, we conduct ablation studies to evaluate the effectiveness of several techniques used in the HRL algorithm, as well as the effectiveness of the upper-level and lower-level agents. In Section 5.2, we repeatedly train the agents using the HRL algorithm to verify whether it yields consistent outcomes across multiple seeds. Computation time for training agents and their implementation is also measured. In Section 5.3, we analyse the impact of uncertainty on the HRL-based scheduling framework. Several system metrics and the behaviour of the HRL agents are evaluated under different levels of uncertainty for each uncertainty source. Finally, we provide managerial insights based on the results of the numerical experiments in Section 5.4.

All experiments were conducted using AMD Ryzen 5 7600X CPU and 32 GB of RAM. We implemented the HRL algorithm with Python 3.10 and Pytorch 2.4.0, and the MIP approach was implemented using FICO Xpress 9.2. We used one of the container terminals at Busan Port as the physical model of our port environment, which includes 15 QCs and five berths. Five berths correspond to 50 berth sections, and one time period corresponds to one hour. Historical real-world data on each vessel's arrival time, requested departure time, length, and the number of containers to be processed were provided by the same container terminal at Busan Port.

We constrained the HRL agents to take actions only at discrete time intervals of one hour. In other words, the HRL algorithm was implemented within the port environment modelled as a discrete-time simulation. Consequently, both the MIP approach and the HRL algorithm make decisions under identical temporal conditions, and their objective function values are directly comparable. It is worth noting that, although the HRL algorithm was implemented in a discrete-time setting during both the training and evaluation phases, it is inherently capable of supporting real-time decision-making in practice. This capability will be demonstrated in Section 5.2.

To provide the HRL agents with diverse training episodes, we generate the actual vessel arrival times in the port environment by adding noise to the historical vessel arrival times in the data. Additionally, we generate the actual vessel departure time by adding noise to the estimated departure time calculated based on the number of QCs to be deployed, determined by the QC scheduler. This is an alternative method used due to a lack of data to generate noise for QC processing times. For these processes, we first calculate the proportions of vessels whose actual arrival and departure times differ from their estimated times in the historical data. Next, we add each type of noise to the arrival or departure times of vessels based on the calculated proportions.

Algorithm 1 HRL algorithm for the BACASP

Initialize the network parameters ϕ^u and $\bar{\phi}^u$, and replay memory D for the upper level agent
Initialize the network parameters θ^q, ϕ^q, θ^b, and ϕ^b for the lower-level agents
Set $\theta^q_{old} \leftarrow \theta^q$, $\theta^b_{old} \leftarrow \theta^b$
for episode $= 1, \ldots, C$ **do**
    set $t = 0$
    while all vessels are served **do**
        Obtain s^u_t
        Determine whether to release a vessel or not, where
        $a^u_t = \arg\max_a Q(s^u_t, a; \phi^u)$
        if $a^u_t = 1$ **then**
            One of the waiting vessels is selected based on the predefined rule
            Obtain s^q_t
            Determine the number of QCs to be deployed for the selected vessel,
            where $a^q_t = \pi_{\theta^q_{old}}(s^q_t)$
            Obtain s^b_t
            Determine which berth to allocate the selected vessel, where
            $a^b_t = \pi_{\theta^b_{old}}(s^b_t)$
            if either of the two lower-level agents has no feasible action **then**
                The action of the upper-level agent is masked by $a^u_t = 0$
                Proceed with the port operations without new berth and QC
                allocation
                Get reward r^u_t
            else
                Specific berth sections and QCs are deployed based on the predefined
                rules
                Get rewards r^u_t, r^q_t, and r^b_t
                Store the transition data (s^q_t, a^q_t, r^q_t) in the QC scheduler's rollout
                buffer
                Store the transition data (s^b_t, a^b_t, r^b_t) in the berth scheduler's rollout
                buffer
            end if
        else
            Proceed with the port operations without new berth and QC allocation
            Get reward r^u_t
        end if
        Compute s^u_{t+1}
        Store transition data $(s^u_t, a^u_t, r^u_t, s^u_{t+1})$ in the upper-level agent's replay
        memory D
        Update ϕ^u to minimize the loss function presented in Equation (1)
        Set $t = t + 1$
    end while
    Set $\bar{\phi}^u \leftarrow \phi^u$ every E_u episodes
    Update ϕ^q, θ^q, ϕ^b, and θ^b every E_l episodes to minimize the loss functions
    presented in Equations (2), (3), (7) and (8), respectively
    Set $\theta^q_{old} \leftarrow \theta^q$ and $\theta^b_{old} \leftarrow \theta^b$ every E_l episodes
    Clear the lower-level agents' rollout buffers every E_l episodes
end for

Table 1. Cost coefficients.

Parameter	Value ($/h)
w	800
τ	300
c^W	5,400
c^B	5,400
c^Q	3,300
p^Q	100

We assume that a single QC can process 25 containers per hour. The maximum number of QCs that can work on a single vessel simultaneously is set to four, following the settings in Agra and Oliveira (2018) and Rodrigues and Agra (2021). This setting is also consistent with historical QC scheduling data from Busan Port. Furthermore, we assume that five QCs can move between berth sections 0 and 19, another five QCs can move between berth sections 10 and 39, and the remaining five QCs can move between berth sections 30 and 49. We set the value of the QC interference parameter $\tilde{\alpha}$ to 0.9, referenced from T. Wang et al. (2020) and Wang, Hu, and Zhen (2024). This value implies that if three QCs are deployed, they can handle approximately 2.67 times the amount of work that a single QC can handle. We assume that the QCs are electrically powered, and adopt the QC operating cost and energy consumption rate provided in Kenan, Jebali, and Diabat (2022), T. Wang et al. (2020), and Wang, Wang, and Meng (2018). More specifically, the QC operating cost of a single QC is $100 per hour. Each QC consumes 200 kWh of electricity during one hour of operation, and the carbon emission factor for electric-powered QCs is 1.1 kg/kWh. Given that a carbon tax of $15 is imposed per kilogram of carbon emissions, the resulting carbon emission cost from operating a single QC for one hour is $3,300.

Following Kenan, Jebali, and Diabat (2022), the penalty cost for delayed vessels is set to $800 per hour, while the earliness income for early departures is set to $300 per hour. It is assumed that vessels use auxiliary engines powered by liquefied natural gas (LNG) while they are either berthed or waiting. In this context, it is assumed that each vessel consumes 2,000 kWh of electricity per hour, based on the data provided by Stolz et al. (2021). The carbon emission factor for LNG is 0.18 kg/kWh. Accordingly, the carbon emission cost incurred by a waiting or berthed vessel is calculated to be $5,400 per hour. All coefficients related to operational and carbon emission costs are summarised in Table 1, with detailed definitions provided in Appendix 1. Note that all costs in this section will be represented as negative values because the upper-level agent and the QC scheduler receive the negative values of the costs as their rewards.

5.1. Performance evaluation of the proposed HRL-based framework

Because the complicated decision-making in the BACASP is achieved through the sequential cooperation of multiple agents in our scheduling framework, its optimality cannot be guaranteed. Therefore, we evaluate its performance by comparing it with the results obtained from an exact algorithm, namely the MIP approach, under a perfect information setting. In the MIP approach, we assume that all uncertainties are perfectly known in advance. Based on this assumption, we solve the deterministic BACASP and obtain lower bounds on the total costs achieved by the HRL algorithm. Note that the MIP approach is employed not to demonstrate the superiority of the proposed HRL algorithm over traditional methods, but rather to serve as an exact algorithm for obtaining the optimal benchmark.

The procedure of the performance evaluation can be explained as follows. First, we train the HRL agents in the port environment simulated using one week of historical vessel data. Note that while the time horizon of a single episode corresponds to one week, the HRL agents are trained using numerous episodes, each with different realizations of uncertainty. As mentioned above, uncertainties in vessel arrival times and QC processing times are incorporated into each episode by adding normally distributed noises to a subset of vessels. Next, we generate a single episode and incorporate the realised uncertainty into the parameters of the deterministic MIP model presented in Appendix 1. We solve it to obtain the optimal objective value, which serves as a lower bound for the total cost under the selected episode. However, one drawback of the MIP approach is that it requires an excessive amount of computation time to find the optimal solution for large-scale instances. Although recent advancements in MIP approaches enable decision-making with shorter time intervals and longer planning horizons, it remains challenging to obtain optimal solutions using exact algorithms in the context of the BACASP. To address this issue, we divide the generated episode into smaller segments so that the commercial solver can find the optimal solution for each segment in a reasonable computation time. A more detailed explanation of how we divide the episode is provided in Appendix 2. In the evaluation, all HRL agents use deterministic policies based on their trained policy networks.

Table 2 shows the total costs in a single episode obtained from the MIP approach and several variations of the HRL algorithm. A total of 20 instances are used to validate the effectiveness of the HRL algorithm under diverse scenarios. Instances 1 through 5 are constructed based on real-world historical data and are intended to

Table 2. Comparison between the MIP approach and HRL algorithms.

	MIP-PI		HRL-QF-IR			HRL-QF-SR			HRL-BF		
Instance	obj_L[a]	obj_F[b]	*obj*	gap_L[c]	gap_F[d]	*obj*	gap_L[c]	gap_F[d]	*obj*	gap_L[c]	gap_F[d]
1	8.29	8.29	8.30	0.12%	0.12%	9.52	14.84%	14.84%	10.74	29.55%	29.55%
2	6.37	6.37	6.58	3.30%	3.30%	6.66	4.55%	4.55%	9.27	45.53%	45.53%
3	7.66	7.66	7.66	0.00%	0.00%	8.00	4.44%	4.44%	10.61	38.51%	38.51%
4	3.93	3.93	4.08	3.82%	3.82%	4.07	3.56%	3.56%	4.57	16.28%	16.28%
5	5.71	5.71	5.87	0.03%	0.03%	5.80	1.58%	1.58%	6.21	8.76%	8.76%
6	7.67	7.77	7.91	3.13%	1.80%	8.66	12.91%	11.45%	10.55	37.55%	35.78%
7	5.55	5.61	5.67	2.12%	1.07%	5.92	6.67%	5.53%	8.35	50.45%	48.84%
8	7.44	7.52	7.46	0.27%	−0.80%	7.91	6.32%	5.19%	10.11	35.89%	34.41%
9	4.04	4.07	4.23	4.70%	3.93%	4.26	5.45%	4.67%	4.21	4.21%	3.33%
10	6.14	6.14	6.25	1.79%	1.79%	6.15	0.16%	0.16%	6.72	9.45%	9.45%
11	8.23	8.35	8.46	2.79%	1.32%	9.97	21.14%	19.40%	10.83	31.59%	29.70%
12	6.23	6.23	6.38	2.41%	2.41%	6.88	10.43%	10.43%	9.17	47.19%	47.19%
13	7.48	7.48	7.48	0.00%	0.00%	8.38	12.32%	12.32%	10.41	39.17%	39.17%
14	4.05	4.07	4.16	2.72%	2.21%	4.16	2.72%	2.21%	4.40	8.64%	8.11%
15	5.81	5.81	5.96	2.58%	2.58%	6.13	5.51%	5.51%	6.52	12.20%	12.20%
16	7.72	7.78	8.00	3.63%	2.83%	9.33	20.85%	19.92%	10.60	37.31%	36.25%
17	5.59	5.65	5.81	3.94%	2.83%	6.16	10.20%	9.03%	8.59	53.67%	52.03%
18	7.40	7.52	7.46	0.81%	−0.80%	8.50	14.86%	13.03%	10.50	41.89%	39.63%
19	3.99	4.01	4.16	4.26%	3.74%	4.16	4.26%	3.74%	4.30	7.77%	7.23%
20	6.18	6.24	6.31	2.10%	1.12%	6.42	3.88%	2.88%	7.05	14.08%	12.98%
Average gap				2.23%	1.67%		8.33%	7.72%		28.48%	27.75%

[a] obj_L: the best lower bound obtained within 30,000 s.
[b] obj_F: the objective value of the best feasible solution obtained within 30,000 s.
[c] gap_L: (total cost under the HRL algorithm − obj_L) × 100/obj_L.
[d] gap_F: (total cost under the HRL algorithm − obj_F) × 100/obj_F.

represent typical operating conditions. From these, we generate Instances 6 through 10 by increasing the levels of uncertainties in vessel arrival times and QC processing times. This implies that the HRL agents make decisions under vessel arrival and departure times that significantly differ from those encountered during the training phase. Instances 11 through 15 are constructed by reducing the intervals between vessel arrivals in Instances 1 through 5, creating more congested scenarios with a higher number of waiting vessels and more vessels berthed simultaneously. Instances 16 through 20 are generated by increasing the level of uncertainty based on these congested scenarios. Each instance consists of 5 vessels, and their lengths range from 4 to 8 berth sections. Each vessel's workload ranges from 124 to 1,589 containers, with an average of 573 containers. The noises added to vessel arrival times are drawn from $\mathcal{N}(0, 0.395^2)$, while the noises for vessel departure times are drawn from $\mathcal{N}(0, 0.955^2)$. The generated noises are then rounded up if positive and rounded down if negative before being added to the original times. In Instances 1 through 5 and 11 through 15, the uncertainty level, defined as the probability that a vessel's arrival or departure time is changed, is set to 0.395 for arrival times and 0.300 for departure times. In Instances 6 through 10 and 16 through 20, these levels are modified to 0.9 to simulate highly uncertain conditions. All vessel characteristics, distributions, and probabilities were extracted and estimated based on one year of historical vessel arrival and departure data from Busan Port.

MIP-PI indicates the MIP approach with perfect information. Because there exist instances for which the optimal solution could not be obtained even after a long computation time, we report both the best lower bound and the best feasible solution obtained from the MIP approach using Xpress, within a time limit of 30,000 s. The results of the HRL algorithms are then compared with these two values, and the corresponding gaps are calculated and reported. HRL-QF-IR refers to the HRL algorithm we ultimately proposed in Section 4, where the QC scheduler and the berth scheduler have independent reward functions. HRL-QF-SR indicates the same algorithm; however, the berth scheduler and QC scheduler share the same reward. Hence, the reward functions of both lower-level agents are set to the estimated costs in HRL-QF-SR. Lastly, HRL-BF refers to the HRL algorithm where the berth scheduler first assigns a berth, and then the QC scheduler determines QCs to be deployed accordingly. The reward function is the same as that used in HRL-QF-IR.

From Table 2, we could ascertain that the proposed HRL algorithm gives sufficiently effective berth and QC schedules. It achieves an average gap of 1.67% compared to the best feasible solutions and 2.23% compared to the best lower bounds, which are remarkable results given that decisions are made in real time based on limited but certain observations. This means that although the HRL agents address uncertainty through sequential cooperation, they can still make decisions that are not significantly different from the integrated decision-making

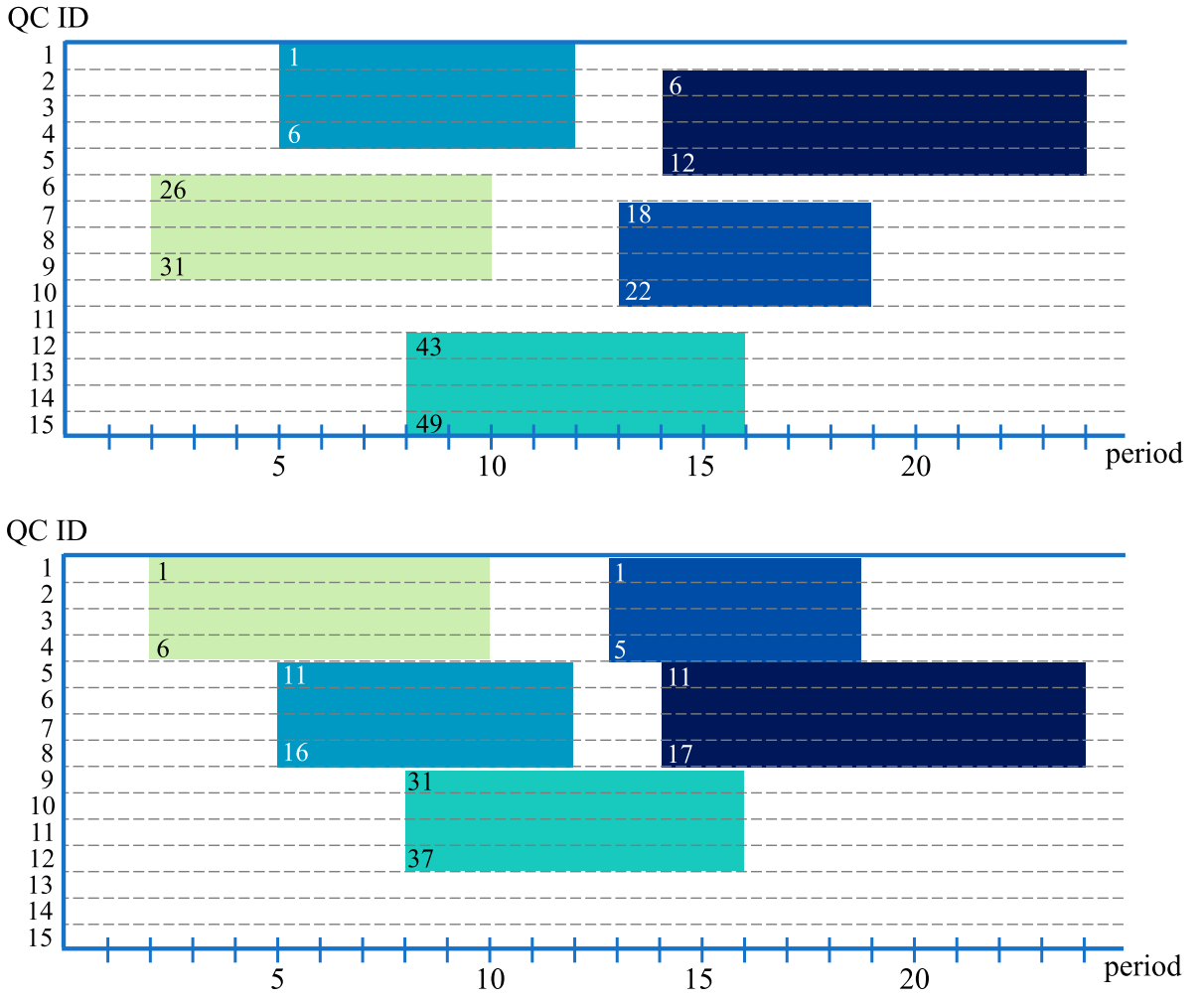


Figure 4. Illustrative example of berth allocation and QC scheduling decisions obtained from MIP-PI and HRL-QF-IR. (a) Berth and QC schedules obtained from MIP-PI. (b) Berth and QC schedules obtained from HRL-QF-IR.

under fully known uncertainties. Even in instances with higher levels of uncertainty than those in the training data or in more congested vessel traffic scenarios, the performance of the HRL algorithm does not deteriorate. Furthermore, the berth allocation and QC schedule determined by MIP-PI and HRL-QF-IR are illustrated in Figure 4. The vertical axis represents QC IDs, and the horizontal axis denotes time periods. Each coloured rectangle corresponds to a vessel and is positioned over the QCs assigned to it. The numbers at the top-left and bottom-left corners of each rectangle indicate the starting and ending berth sections allocated to the vessel, respectively. Although HRL-QF-IR determined different berthing locations for the vessels compared to MIP-PI, it adopted the same berthing times and number of assigned QCs as the optimal solution, thereby achieving the minimum total cost.

By comparing HRL-BF and HRL-QF-IR, we can verify that the order of decision-making of the two lower-level agents is a crucial factor in the HRL-based scheduling framework. This is because the number of QCs assigned to vessels and the resulting departure times are more critical to operational and carbon emission costs than are the locations where the vessels are berthed. In fact, we frequently observed cases where a vessel was handled inefficiently due to the limited number of QCs that could move to the assigned berth sections when we used HRL-BF. Therefore, determining the number of QCs for the selected vessel first, followed by deciding on the feasible berthing location, is an appropriate strategy for minimising the total cost. In addition, we examined the validity of the reward function for the berth scheduler through a comparison between HRL-QF-SR and HRL-QF-IR. In most instances, HRL-QF-IR outperformed HRL-QF-SR. This implies that the reward function explained in Section 4.3.2 successfully guided the berth scheduler to allocate a berth in a way that leaves as many QCs as possible for the next arriving vessel.

We also conduct an ablation study on the agents at each hierarchical level. Table 3 shows the result, where

Table 3. Ablation study for the HRL agents.

	Upper-level agent + Lower-level agents			Random release + Lower level agents			Upper-level agent + Random assignment		
Case	Mean	Var	Gap	Mean	Var	Gap	Mean	Var	Gap
1	8.32	0.13	–	8.76	0.25	5.29%	10.23	1.18	22.96%
2	6.50	0.12	–	6.96	0.23	7.08%	7.86	0.62	20.92%
3	7.65	0.12	–	7.99	0.18	4.44%	9.48	0.89	23.92%
4	4.08	0.09	–	4.44	0.15	8.82%	4.66	0.27	14.22%
5	5.87	0.10	–	6.22	0.17	5.63%	7.03	0.67	19.61%
Average		0.11	–		0.20	6.25%		0.73	20.33%

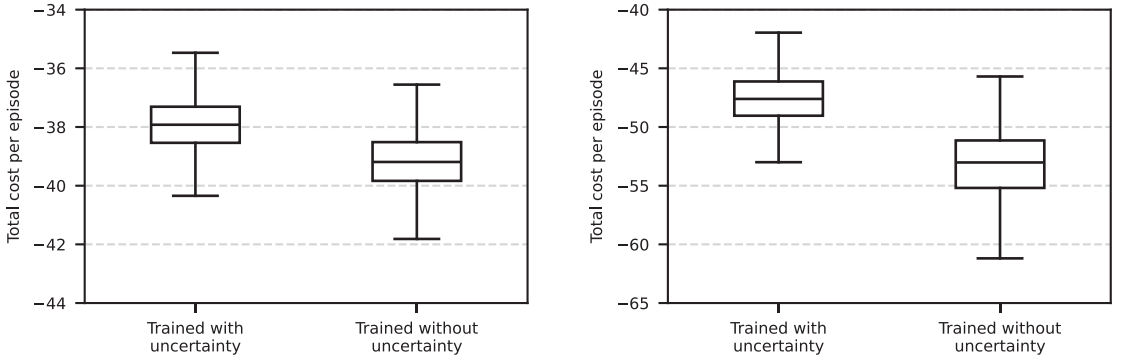
each case indicates a set of 1,000 test episodes generated by adding noise to the vessel arrival and departure times in each instance of Table 2. Random release refers to releasing a vessel with a probability of $\frac{1}{2}$ without using the upper-level agent, and random assignment refers to randomly selecting the number of QCs and berthing location instead of using the lower-level agents. For each scheduling framework, the mean and variance of the total cost over 1,000 episodes are presented. Values in the third column of each framework indicate the gap in total cost compared to that of using all three agents. By comparing the gap, we can ascertain that the agents at both hierarchical levels are effective in minimising total costs. In particular, the lower-level agents play a crucial role in enhancing the performance of the HRL algorithm. Furthermore, a comparison of the variance shows that the cooperation of all three agents can effectively address the uncertainty, resulting in consistent performance.

Lastly, in order to assess the impact of considering uncertainty in berth allocation and QC scheduling, we compare the decisions made by two HRL agents: one trained on deterministic episodes and the other trained on episodes incorporating uncertainty. Figure 5 presents box plots of the total cost achieved by these two HRL agents. Figure 5(a) shows the total cost over 1,000 evaluation episodes with the same vessel arrival rate as the historical data, while Figure 5(b) shows the total cost over 1,000 evaluation episodes with a higher vessel arrival rate, representing more congested scenarios. The HRL agent trained on episodes incorporating uncertainty achieves a lower total cost compared to the agent trained on deterministic episodes, and this difference becomes more pronounced in congested scenarios. The average total cost increases by approximately 3% in the normal scenarios and by about 12% in the congested scenarios. These results indicate that uncertainty has a significant impact on optimal berth allocation and QC scheduling, and that ignoring uncertainty in decision-making can lead to substantially higher costs, especially in more congested scenarios.

5.2. Validation of the practical applicability of the HRL algorithm

In a typical port operation, baseline schedules for berths and QCs are determined over multiple days (e.g. one week) based on the estimated vessel arrival times. When unexpected changes occur in vessel arrival or departure times, the baseline schedule is adjusted by a port manager. Our HRL-based scheduling framework can accommodate such practical situations in the following manner: Prior to the start of port operations, HRL agents are trained using training episodes generated based on one week of data, which includes information of vessels scheduled to arrive within the planning horizon. Due to the limited generalisation capability of the HRL agents, this training approach is employed instead of utilising training episodes with diverse vessel data configurations. Its practical feasibility will be verified later. Next, the trained agents at each level make decisions based on realised vessel arrival and departure times. Because they make feasible decisions using already observed information, modifications to berth and QC schedules are not required. In addition, because the training episodes incorporate uncertainties in vessel arrival times and QC processing times, the trained agents, as demonstrated in Section 5.1, can effectively address changes in vessel arrival times or QC processing times within the one-week planning horizon.

However, several additional aspects need to be verified for the practical implementation of the proposed algorithm. The first aspect is the computation time required for training and implementation. Table 4 shows the training and implementation times for the HRL algorithm at each random seed. Because we determined that 30,000 episodes are sufficient to converge the HRL agents, we measured the computation time required for this number of training episodes. As presented in the table, training the three HRL agents with 30,000 episodes, each having a one-week time horizon, required less than 6,000 s in all training sessions. It also took less than 0.1 s to execute a single episode once the HRL agents



(a) Normal scenarios (b) Congested scenarios

Figure 5. Box plots of total cost by HRL agents trained with and without uncertainty. (a) Normal scenarios. (b) Congested scenarios.

Table 4. Computation times of the HRL algorithm.

Seed	Training time for 30,000 episodes	Implementation time for a single episode
1	5,797 s	< 0.1 s
2	5,710 s	< 0.1 s
3	5,738 s	< 0.1 s
4	5.694 s	< 0.1 s
5	5,710 s	< 0.1 s

were trained. Therefore, we can conclude that the proposed framework is practical in terms of the computation time.

In addition, if the agents produce different results each time they are trained, the decisions made by a trained HRL agent cannot be trusted in real-world scenarios. Therefore, it is necessary to verify whether it produces consistent results across different training runs. To achieve this, we compare the learning curves and total cost distributions of the HRL agents trained on five different random seeds. Learning curves of each training run are presented in Figure 6, where each point represents the average total cost over the most recent 100 training episodes. First, we observed that the initial weights of policy networks significantly impact the performance of the HRL algorithm in initial training episodes. Although there are cases where the learning curve temporarily drops, as observed in seed 1, the curves eventually converge to nearly identical total costs across all training runs.

Furthermore, Figure 7 depicts distributions of the total costs in the last 1,000 training episodes across five random seeds. As in the learning curves, we could observe that all trained agents hold policy networks that yield almost the same total costs. In conclusion, the proposed HRL algorithm produces consistent results across multiple training runs, making it appropriate for practical application in real-world port operations.

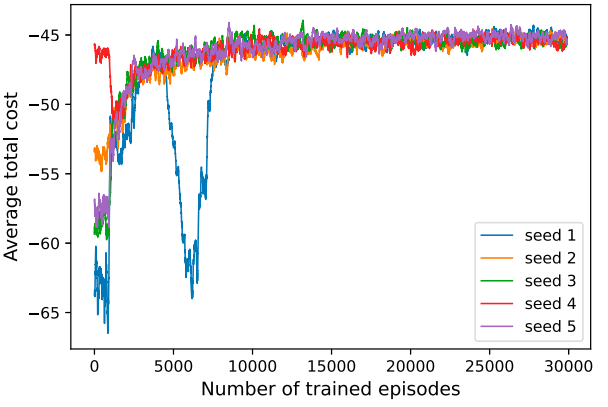


Figure 6. Learning curves of the upper-level agent trained on five random seeds.

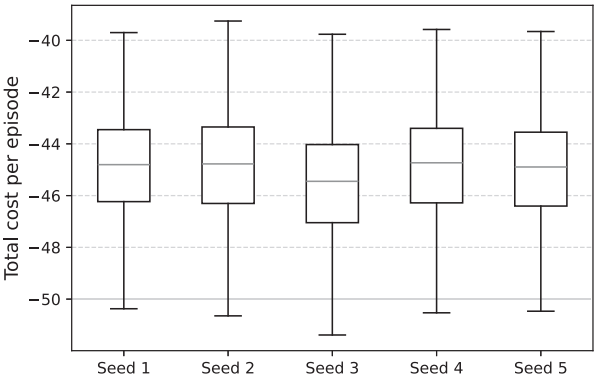


Figure 7. Distributions of total costs across five random seeds.

5.3. Analysis of the impact of uncertainty on the BACASP

In this subsection, we aim to analyse how the system metrics, including the operational and carbon emission costs, vary with uncertainty levels. In addition, we examine how the HRL agents respond to uncertainty. Note that the level of uncertainty is defined as the proportion of vessels

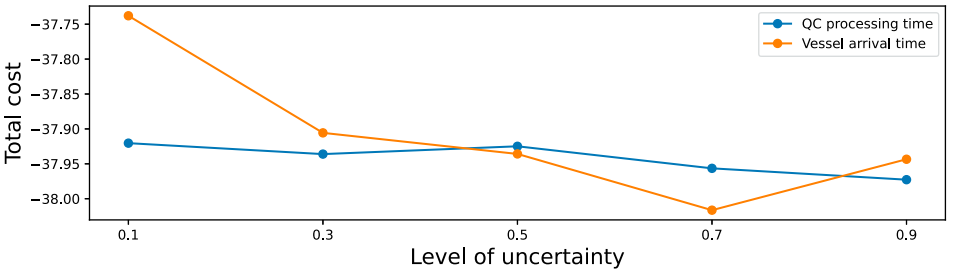


Figure 8. Total costs by uncertainty level.

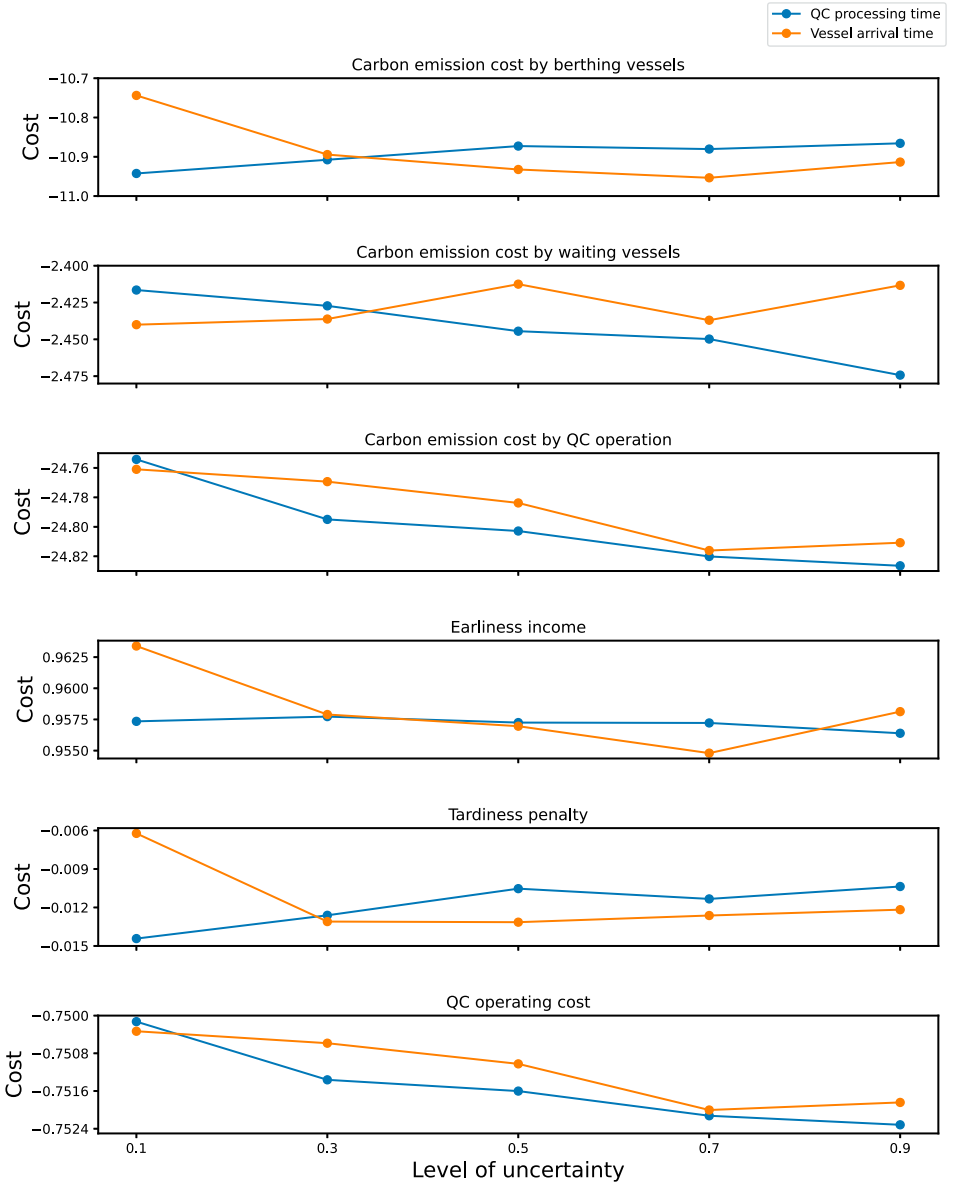
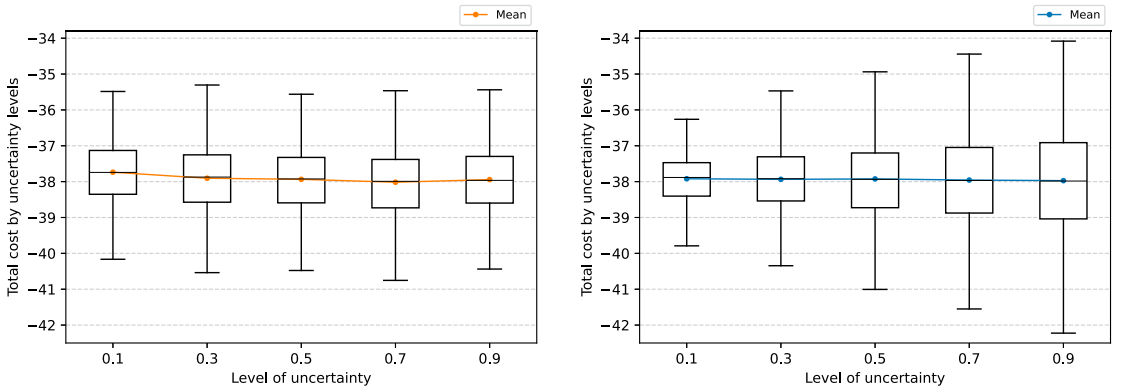


Figure 9. Operational and carbon emission costs by uncertainty level.

whose actual arrival or departure times differ from their estimated times. In all experiments in this subsection, the uncertainty level of only one source of uncertainty is controlled at a time.

Figure 8 shows the impact of uncertainty levels on the total cost, and Figure 9 illustrates the impact of uncertainty levels on each cost component. Each point in Figures 8 and 9 indicates the average cost in 1,000 evaluation



(a) Vessel arrival time (b) QC processing time

Figure 10. Box plots of total cost by uncertainty levels. (a) Vessel arrival time. (b) QC processing time.

episodes. The orange lines represent costs across the proportion of vessels whose actual arrival times differ from their scheduled arrival times, while the blue lines represent costs across the proportion of vessels whose actual departure times differ from their scheduled departure times due to changes in QC processing times.

As shown in Figure 8, the total cost tends to increase as more vessels arrive or depart at times different from their scheduled arrival or departure times. A noteworthy observation is that uncertainty in vessel arrival times significantly impacts the mean total cost more than uncertainty in QC processing times. We analysed that this is due to the difference in the trends of the carbon emission cost by berthing vessels, which exhibits the largest variation among the cost components. As illustrated at the top of Figure 9, the emission cost from berthing vessels increases with higher uncertainty levels of vessel arrival times, whereas the opposite trend is observed in QC processing times. While other cost components also display their own trends, the magnitudes are smaller compared to the emission cost incurred by berthing vessels, resulting in limited contributions to the variation in the total cost.

In Figure 10, box plots of total cost distributions for 1,000 evaluation episodes across different uncertainty levels of each source are presented. Figure 10(a,b) show the correlation between the total cost distribution and the uncertainty levels of vessel arrival times and QC processing times, respectively. When the uncertainty in vessel arrival times increases, the variance remains nearly unchanged. However, it gradually increases as the uncertainty in QC processing times grows. From this result, we can conclude that, unlike the mean of the total cost, its variance is more significantly influenced by the uncertainty in QC processing times.

To understand how the behaviour of the upper-level agent changes with different levels of uncertainty, we

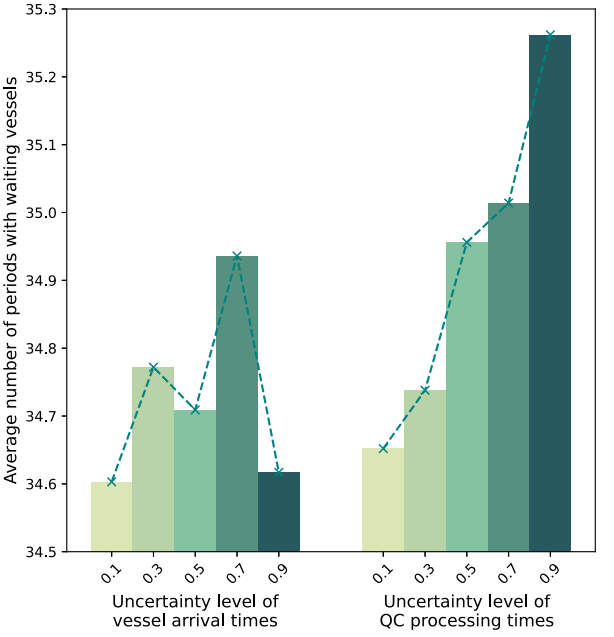


Figure 11. Average number of time periods with one or more waiting vessels for each uncertainty level.

analyse the number of time periods during which one or more vessels are waiting. Figure 11 illustrates the average number of time periods with waiting vessels per episode over 1,000 evaluation episodes. We observe a slight increase in the average number of periods during which vessels are waiting as the uncertainty of QC processing times rises. This means that as the uncertainty level of QC processing times increases, the upper-level agent postpones releasing vessels more frequently. The increase in the uncertainty of QC processing times makes QC scheduling more challenging, and postponing the release of vessels can be an effective action in such situations. Therefore, the upper-level agent is appropriately responding to the uncertainty of QC processing times.

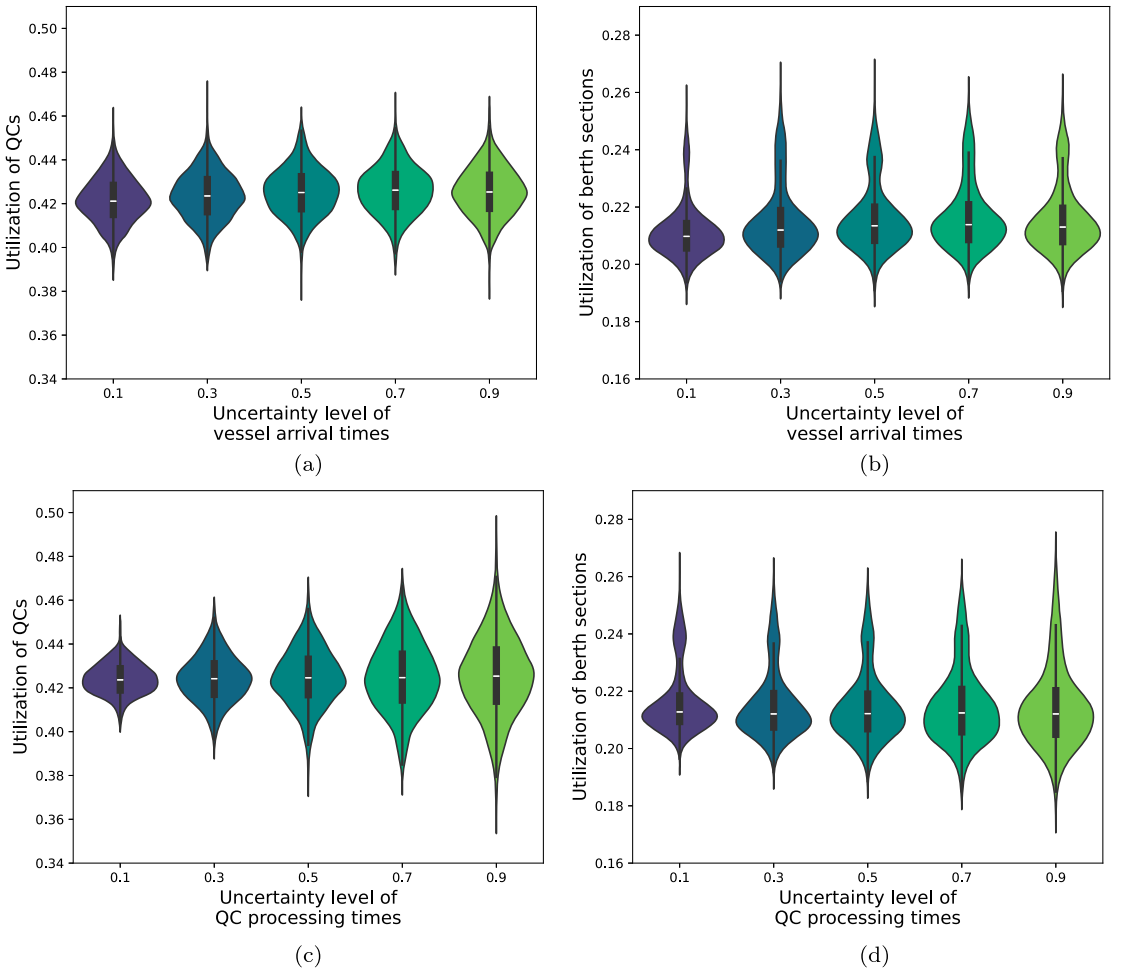


Figure 12. Average utilisation of QCs and berth sections by uncertainty levels.

Next, Figure 12 depicts violin plots for the average utilisation of QCs and berth sections per episode over 1,000 evaluation episodes. As illustrated in Figure 12(a,b), the uncertainty of vessel arrival times has a negligible effect on the distributions of the average utilisation of QCs and berth sections. In contrast, the uncertainty of QC processing times considerably affects them, as shown in Figure 12(c,d). As the uncertainty of QC processing time increases, the variation in the average utilisation increases, with a greater magnitude of the change observed in the utilisation of QCs. However, the increased uncertainty does not result in a significant increase in the utilisation of QCs or berth sections that could cause deterioration in the performance of the HRL agents.

Lastly, we analyse how the QC scheduler responds to the uncertainty. Figures 13 and 14 illustrates the frequency of actions taken by the QC scheduler. First, we can identify that the QC scheduler assigns the maximum

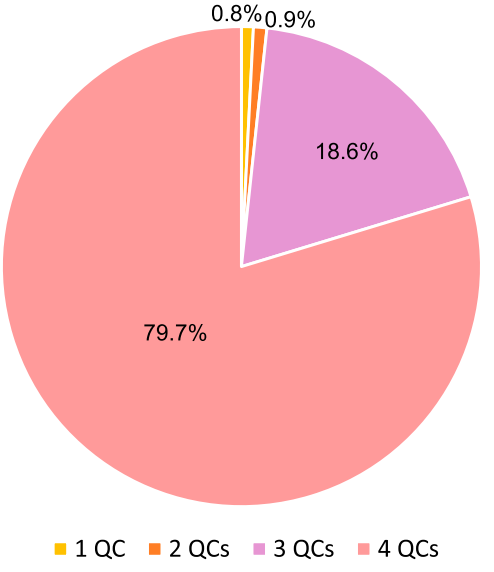


Figure 13. Execution rate of actions by the QC scheduler across all uncertainty levels.

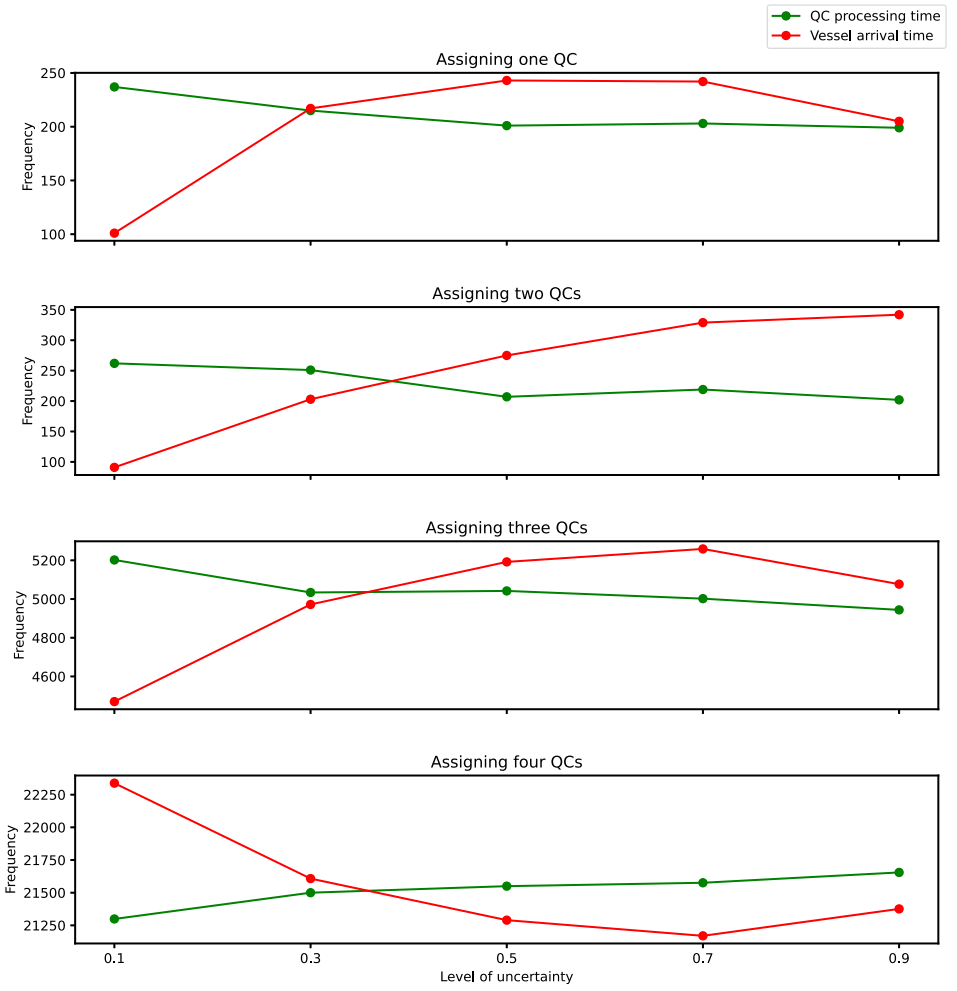


Figure 14. Frequency of actions taken by the QC scheduler under different uncertainty levels.

number of four QCs to approximately 80 percent of vessels, as shown in Figure 13. However, its policy slightly changes as the uncertainty level increases, and the tendency of these changes differs depending on the source of uncertainty, as illustrated in Figure 14. The QC scheduler tends to assign fewer QCs on average when the uncertainty of vessel arrival times increases, whereas it assigns four QCs more frequently as the uncertainty of QC processing times increases. Although the degree of change is quite small, the QC scheduler gradually adjusts its behaviour to respond to increasing uncertainty. In particular, we infer that this represents a cooperative strategy to address increased uncertainty in QC processing times, where the upper-level agent delays the release of vessels more frequently, allowing the QC scheduler to assign more QCs to a single vessel.

5.4. Managerial insights

Based on the numerical experiments explained above, we offer several managerial insights for practitioners engaged in real-time port operations.

- The proposed HRL-based scheduling framework demonstrated remarkable performance in real-time decision-making based on observed information within highly dynamic and uncertain port environments. Its consistency over multiple training iterations was guaranteed, ensuring it can be applied to real-world port environments without repetitive training. To train the HRL agents for a one-week planning horizon, approximately 30,000 training episodes are required to achieve their convergence. The training time for 30,000 episodes and implementation time

after the training were verified to be sufficiently short for practical application. Lastly, we recommend constructing training episodes by adding noise to the estimated arrival and departure times of vessels expected to arrive within the planning horizon.

- When the proposed HRL-based scheduling framework is utilised, if the carbon tax is imposed on port operations, the largest proportion of the total cost comes from carbon emission costs generated by QC operations and berthing vessels. Consequently, strategies to reduce carbon emissions from QCs and vessels, such as the use of eco-friendly fuels for them, are required not only for sustainable port operations but also for minimising port costs. Furthermore, an increase in uncertainty of vessel arrival times or QC processing times leads to an increase in the total cost. In particular, according to Figure 9, the cost increase caused by the uncertainty of vessel arrival times is primarily due to the rise in carbon emission costs from berthing vessels. Therefore, if such situations are expected, ports should implement policies such as encouraging berthing vessels to use less power.
- Uncertainties of vessel arrival times and QC processing times were found to have different types of impacts on system metrics resulting from the proposed HRL-based scheduling framework. The average total cost was more significantly affected by the uncertainty of vessel arrival times, whereas the variance of total costs was more heavily influenced by the uncertainty of QC processing times. As a result, reducing both types of uncertainty is recommended not only to lower costs but also to ensure stable and predictable costs when utilising our framework. In addition, while the increase in uncertainty of QC processing times was found to raise the variance of the utilisation of QCs and berths, it did not lead to an unmanageable increase in their maximum utilisation. Therefore, if our framework can be applied, constructing additional berths or QCs in the port to mitigate the inefficiencies caused by uncertainty is not necessary.

6. Conclusions

The performance of berth allocation and QC scheduling, which are fundamental components of port operations, is a critical determinant of port competitiveness. Recently, effective management of uncertain factors such as vessel arrival times and QC operation times has been focussed. Additionally, reducing carbon emissions in line with international trends is also required. However, the BACASP literature lacks studies that simultaneously incorporate uncertainty and carbon emissions into their models. To fill this research gap, we proposed a novel HRL-based scheduling framework to support real-time berth allocation and QC scheduling with consideration of carbon emissions.

Given the complexity of decision-making in the BACASP, three cooperative agents were employed: the upper-level agent determines whether to release waiting vessels in the roadstead, the QC scheduler decides the number of QCs to be deployed for each vessel, and the berth scheduler allocates berth sections for each vessel. Comparisons between the proposed HRL algorithm and the MIP approach with perfect information demonstrated the effectiveness of the proposed algorithm. In addition, we ascertained that the computation time for training and implementation is sufficiently short, indicating that the HRL-based scheduling framework is appropriate for real-time decision-making in port environments. The consistency of the HRL algorithm was also verified by comparing results from repetitive training of the agents. Finally, based on numerical experiments, managerial insights for port operators were provided to enhance sustainability and competitiveness. Through this study, we demonstrated the potential of utilising a well-structured RL framework to address complex port operation problems, including the BACASP.

Acknowledgments

The authors are grateful for the valuable comments from the associate editor and anonymous reviewers.

Data availability statement

The data used in this study are available from the first author upon reasonable request.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (Nos. RS-2023-00218913 and RS-2024-00410619).

Notes on contributors

Seongbae Jo received the B.S. and M.S. degrees in industrial engineering from Seoul National University, Korea, in 2021 and 2023, respectively. He is currently working toward the Ph.D. degree in industrial engineering from Seoul National University, Korea. His research interests include reinforcement learning and optimisation with applications in maritime logistics, supply chain management, and real-time decision-making problems.

Ilkyeong Moon received the B.S. and M.S. degrees in industrial engineering from Seoul National University, Korea, in 1984 and 1986, respectively, and the Ph.D. degree in operations research from Columbia University, USA, in 1991. He is currently a professor of industrial engineering with Seoul National University. He published 190 papers in international journals. His research interests include supply chain management, logistics, and inventory management. He has been an editor-in-chief for European Journal of Industrial Engineering. He was the president of the Korean Institute of Industrial Engineers (KIIE) from 2019 to 2020. He is a fellow of the Korean Academy of Science and Technology which is the most prestigious engineering and science society in Korea. He is a fellow of International Federation for Production Research (IFPR).

ORCID

Ilkyeong Moon http://orcid.org/0000-0002-7072-1351

References

Agra, Agostinho, and Maryse Oliveira. 2018. "MIP Approaches for the Integrated Berth Allocation and Quay Crane Assignment and Scheduling Problem." *European Journal of Operational Research* 264 (1): 138–148. https://doi.org/10.1016/j.ejor.2017.05.040.

Bierwirth, Christian, and Frank Meisel. 2015. "A Follow-up Survey of Berth Allocation and Quay Crane Scheduling Problems in Container Terminals." *European Journal of Operational Research* 244 (3): 675–689. https://doi.org/10.1016/j.ejor.2014.12.030.

Chargui, Kaoutar, Tarik Zouadi, and V. Raja Sreedharan. 2023. "Berth and Quay Crane Allocation and Scheduling Problem with Renewable Energy Uncertainty: A Robust Exact Decomposition." *Computers & Operations Research* 156: 106251. https://doi.org/10.1016/j.cor.2023.106251.

Chargui, Kaoutar, Tarik Zouadi, V. Raja Sreedharan, Abdellah El Fallahi, and Mohamed Reghioui. 2023. "A Novel Robust Exact Decomposition Algorithm for Berth and Quay Crane Allocation and Scheduling Problem considering Uncertainty and Energy Efficiency." *Omega* 118: 102868. https://doi.org/10.1016/j.omega.2023.102868.

Chung, Sai Ho, and Felix T. S. Chan. 2013. "A Workload Balancing Genetic Algorithm for the Quay Crane Scheduling Problem." *International Journal of Production Research* 51 (16): 4820–4834. https://doi.org/10.1080/00207543.2013.774489.

Filom, Siyavash, Amir M. Amiri, and Saiedeh Razavi. 2022. "Applications of Machine Learning Methods in Port Operations–A Systematic Literature Review." *Transportation Research Part E: Logistics and Transportation Review* 161: 102722. https://doi.org/10.1016/j.tre.2022.102722.

International Maritime Organization. 2018. "UN Body Adopts Climate Change Strategy for Shipping." https://www.imo.org/en/MediaCentre/PressBriefings/Pages/06GHGinitialstrategy.aspx.

Jauhar, Sunil Kumar, Saurabh Pratap, Sachin Kamble, Shivam Gupta, and Amine Belhadi. 2023. "A Prescriptive Analytics Approach to Solve the Continuous Berth Allocation and Yard Assignment Problem Using Integrated Carbon Emissions Policies." *Annals of Operations Research* 1–32. https://doi.org/10.1007/s10479-023-05493-1.

Ji, Bin, Han Huang, and S. Yu Samson. 2022. "An Enhanced NSGA-II for Solving Berth Allocation and Quay Crane Assignment Problem with Stochastic Arrival times." *IEEE Transactions on Intelligent Transportation Systems* 24 (1): 459–473. https://doi.org/10.1109/TITS.2022.3213834.

Jiang, Xing, Ming Zhong, Jiahui Shi, and Weifeng Li. 2024. "Optimization of Integrated Scheduling of Restricted Channels, Berths, and Yards in Bulk Cargo Ports considering Carbon Emissions." *Expert Systems with Applications* 255: 124604. https://doi.org/10.1016/j.eswa.2024.124604.

Karakas, Serkan, Mehmet Kirmizi, and Batuhan Kocaoglu. 2021. "Yard Block Assignment, Internal Truck Operations, and Berth Allocation in Container Terminals: Introducing Carbon-Footprint Minimisation Objectives." *Maritime Economics & Logistics* 23: 750–771. https://doi.org/10.1057/s41278-021-00186-7.

Kenan, Nabil, Aida Jebali, and Ali Diabat. 2022. "The Integrated Quay Crane Assignment and Scheduling Problems with Carbon Emissions Considerations." *Computers & Industrial Engineering* 165: 107734. https://doi.org/10.1016/j.cie.2021.107734.

Lei, Kun, Peng Guo, Yi Wang, Jian Zhang, Xiangyin Meng, and Linmao Qian. 2023. "Large-Scale Dynamic Scheduling for Flexible Job-Shop with Random Arrivals of New Jobs by Hierarchical Reinforcement Learning." *IEEE Transactions on Industrial Informatics* 20 (1): 1007–1018. https://doi.org/10.1109/TII.2023.3272661.

Liu, Ding, and Ying-En Ge. 2018. "Modeling Assignment of Quay Cranes Using Queueing Theory for Minimizing CO2 Emission at a Container Terminal." *Transportation Research Part D: Transport and Environment* 61: 140–151. https://doi.org/10.1016/j.trd.2017.06.006.

Liu, Renke, Rajesh Piplani, and Carlos Toro. 2022. "Deep Reinforcement Learning for Dynamic Scheduling of a Flexible Job Shop." *International Journal of Production Research* 60 (13): 4049–4069. https://doi.org/10.1080/00207543.2022.2058432.

Ma, Yi, Xiaotian Hao, Jianye Hao, Jiawen Lu, Xing Liu, Tong Xialiang, Mingxuan Yuan, Zhigang Li, Jie Tang, and Zhaopeng Meng. 2021. "A Hierarchical Reinforcement Learning Based Optimization Framework for Large-Scale Dynamic Pickup and Delivery Problems." *Advances in Neural Information Processing Systems* 34: 23609–23620.

Peng, Yun, Meng Dong, Xiangda Li, Huakun Liu, and Wenyuan Wang. 2021. "Cooperative Optimization of Shore Power Allocation and Berth Allocation: A Balance between Cost and Environmental Benefit." *Journal of Cleaner Production* 279: 123816. https://doi.org/10.1016/j.jclepro.2020.123816.

Peng, Yun, Wenyuan Wang, Xiangqun Song, and Qi Zhang. 2016. "Optimal Allocation of Resources for Yard Crane Network Management to Minimize Carbon Dioxide Emissions." *Journal of Cleaner Production* 131: 649–658. https://doi.org/10.1016/j.jclepro.2016.04.120.

Rodrigues, Filipe, and Agostinho Agra. 2021. "An Exact Robust Approach for the Integrated Berth Allocation and Quay Crane Scheduling Problem under Uncertain Arrival times." *European Journal of Operational Research* 295 (2): 499–516. https://doi.org/10.1016/j.ejor.2021.03.016.

Rodrigues, Filipe, and Agostinho Agra. 2022. "Berth Allocation and Quay Crane Assignment/scheduling Problem under

Uncertainty: A Survey." *European Journal of Operational Research* 303 (2): 501–524. https://doi.org/10.1016/j.ejor.2021.12.040.

Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. "Proximal Policy Optimization Algorithms." Preprint arXiv:1707.06347.

Stolz, Boris, Maximilian Held, Gil Georges, and Konstantinos Boulouchos. 2021. "The CO2 Reduction Potential of Shore-Side Electricity in Europe." *Applied Energy* 285: 116425. https://doi.org/10.1016/j.apenergy.2020.116425.

Tan, Caimao, and Junliang He. 2021. "Integrated Proactive and Reactive Strategies for Sustainable Berth Allocation and Quay Crane Assignment under Uncertainty." *Annals of Operations Research*349: 879–910. https://doi.org/10.1007/s10479-020-03891-3.

Tasoglu, Gokcecicek, and Gokalp Yildiz. 2019. "Simulated Annealing Based Simulation Optimization Method for Solving Integrated Berth Allocation and Quay Crane Scheduling Problems." *Simulation Modelling Practice and Theory* 97: 101948. https://doi.org/10.1016/j.simpat.2019.101948.

UNCTAD. 2023. "Review of Maritime Transport." Technical Report. Geneva: United Nations.

UNCTAD. 2024. "Review of Maritime Transport." Technical Report. Geneva: United Nations.

Van Hasselt, Hado, Arthur Guez, and David Silver. 2016. "Deep Reinforcement Learning with Double Q-Learning." In *Proceedings of the AAAI Conference on Artificial Intelligence*, 30: 2094–2100. https://doi.org/10.1609/aaai.v30i1.10295.

Venturini, Giada, Çağatay Iris, Christos A Kontovas, and Allan Larsen. 2017. "The Multi-port Berth Allocation Problem with Speed Optimization and Emission Considerations." *Transportation Research Part D: Transport and Environment* 54: 142–159. https://doi.org/10.1016/j.trd.2017.05.002.

Wang, Chong, Kaiyuan Liu, Canrong Zhang, and Lixin Miao. 2024. "Distributionally Robust Chance-Constrained Optimization for the Integrated Berth Allocation and Quay Crane Assignment Problem." *Transportation Research Part B: Methodological* 182: 102923. https://doi.org/10.1016/j.trb.2024.102923.

Wang, Chong, Qi Wang, Xi Xiang, Canrong Zhang, and Lixin Miao. 2025. "Optimizing Integrated Berth Allocation and Quay Crane Assignment: A Distributionally Robust Approach." *European Journal of Operational Research* 320 (3): 593–615. https://doi.org/10.1016/j.ejor.2024.08.001.

Wang, Haoxiang, Bhaba R. Sarker, Jing Li, and Jian Li. 2021. "Adaptive Scheduling for Assembly Job Shop with Uncertain Assembly times Based on Dual Q-learning." *International Journal of Production Research* 59 (19): 5867–5883. https://doi.org/10.1080/00207543.2020.1794075.

Wang, Tingsong, Xinchang Wang, and Qiang Meng. 2018. "Joint Berth Allocation and Quay Crane Assignment under Different Carbon Taxation Policies." *Transportation Research Part B: Methodological* 117: 18–36. https://doi.org/10.1016/j.trb.2018.08.012.

Wang, Tingsong, Yuquan Du, Debin Fang, and Zhi-Chun Li. 2020. "Berth Allocation and Quay Crane Assignment for the Trade-off between Service Efficiency and Operating Cost considering Carbon Emission Taxation." *Transportation Science* 54 (5): 1307–1331. https://doi.org/10.1287/trsc.2019.0946.

Wang, Zhu, Hongtao Hu, and Lu Zhen. 2024. "Berth and Quay Cranes Allocation Problem with on-shore Power Supply Assignment in Container Terminals." *Computers & Industrial Engineering* 188: 109910. https://doi.org/10.1016/j.cie.2024.109910.

Xiang, Xi, and Changchun Liu. 2021. "An Almost Robust Optimization Model for Integrated Berth Allocation and Quay Crane Assignment Problem." *Omega* 104: 102455. https://doi.org/10.1016/j.omega.2021.102455.

Xiang, Xi, Changchun Liu, and Lixin Miao. 2018. "Reactive Strategy for Discrete Berth Allocation and Quay Crane Assignment Problems under Uncertainty." *Computers & Industrial Engineering* 126: 196–216. https://doi.org/10.1016/j.cie.2018.09.033.

Xu, Zhou, and Chung-Yee Lee. 2018. "New Lower Bound and Exact Method for the Continuous Berth Allocation Problem." *Operations Research* 66 (3): 778–798. https://doi.org/10.1287/opre.2017.1687.

Yang, Chunxia, Xiaojun Wang, and Zhenfeng Li. 2012. "An Optimization Approach for Coupling Problem of Berth Allocation and Quay Crane Assignment in Container Terminal." *Computers & Industrial Engineering* 63 (1): 243–253. https://doi.org/10.1016/j.cie.2012.03.004.

Yu, Jingjing, Guolei Tang, Stefan Voß, and Xiangqun Song. 2023. "Berth Allocation and Quay Crane Assignment considering the Adoption of Different Green Technologies." *Transportation Research Part E: Logistics and Transportation Review* 176: 103185. https://doi.org/10.1016/j.tre.2023.103185.

Zhang, Wenquan, Fei Zhao, Yong Li, Chao Du, Xiaobing Feng, and Xuesong Mei. 2024. "A Novel Collaborative Agent Reinforcement Learning Framework Based on an Attention Mechanism and Disjunctive Graph Embedding for Flexible Job Shop Scheduling Problem." *Journal of Manufacturing Systems* 74: 329–345. https://doi.org/10.1016/j.jmsy.2024.03.012.

Zhen, Lu, Xueting He, Dan Zhuge, and Shuaian Wang. 2024. "Primal Decomposition for Berth Planning under Uncertainty." *Transportation Research Part B: Methodological* 183: 102929. https://doi.org/10.1016/j.trb.2024.102929.

Zhen, Lu, Qian Sun, Wei Zhang, Kai Wang, and Wen Yi. 2021. "Column Generation for Low Carbon Berth Allocation under Uncertainty." *Journal of the Operational Research Society* 72 (10): 2225–2240. https://doi.org/10.1080/01605682.2020.1776168.

Zhen, Lu, Dan Zhuge, Shuaian Wang, and Kai Wang. 2022. "Integrated Berth and Yard Space Allocation under Uncertainty." *Transportation Research Part B: Methodological* 162: 1–27. https://doi.org/10.1016/j.trb.2022.05.011.

Zhen, Lu, Dan Zhuge, Shuanglu Zhang, Shuaian Wang, and Harilaos N. Psaraftis. 2024. "Optimizing Sulfur Emission Control Areas for Shipping." *Transportation Science* 58 (3): 614–638. https://doi.org/10.1287/trsc.2023.0278.

Zheng, Feifeng, Ying Li, Feng Chu, Ming Liu, and Yinfeng Xu. 2019. "Integrated Berth Allocation and Quay Crane Assignment with Maintenance Activities." *International Journal of Production Research*57 (11): 3478–3503. https://doi.org/10.1080/00207543.2018.1539265.

Zhu, Yi, and Andrew Lim. 2006. "Crane Scheduling with Non-crossing Constraint." *Journal of the Operational Research Society* 57 (12): 1464–1471. https://doi.org/10.1057/palgrave.jors.2602110.

Appendices

Appendix 1. Deterministic BACASP model

We first present notations used in the mathematical model as follows.

Indices and sets

$\mathcal{B}$	Set of berth sections, $n \in \mathcal{B} = \{1, \ldots, J\}$
$\mathcal{Q}$	Set of QCs, $g \in \mathcal{Q} = \{1, \ldots, Q\}$
$\mathcal{V}$	Set of vessels, $k \in \mathcal{V} = \{1, \ldots, V\}$
$\mathcal{T}$	Set of time periods, $j \in \mathcal{T} = \{1, \ldots, M\}$
$\mathcal{N}^{QC}$	Set of the number of QCs that can be assigned to a single vessel simultaneously, $q \in \mathcal{N}^{QC} = \{1, \ldots, N^{QC}\}$

Parameters

$\overline{A}_k$	Arrival time of vessel k
D_k	Requested departure time of vessel k
$\overline{D}_k$	Departure delay of vessel k caused by the delay in QC operations
H_k	Length of vessel k, expressed in terms of the number of berth sections
W_k	Number of containers to be processed in vessel k
p	Processing rate of each QC
S_g	Starting berth section where QC g can operate
E_g	Ending berth section where QC g can operate
$\tilde{\alpha}$	QC interference exponent
w	Penalty cost incurred for each unit of tardiness in the departure time of a vessel
τ	Reward granted for each unit of earliness in the departure time of a vessel
c^W	Carbon emission cost incurred for each unit of time a vessel is waiting
c^B	Carbon emission cost incurred for each unit of time a vessel is berthing
c^Q	Carbon emission cost incurred for each unit of time a QC operates
p^Q	Operating cost incurred for each unit of time a QC operates

Decision variables

x_{kl}	Binary variable taking value one if vessel l berths after vessel k had departed
y_{kl}	Binary variable taking value one if vessel l berths below the berthing position of vessel k
b_k	Integer variable that indicates the first berthing position of vessel k
t_k	Integer variable that indicates the berthing time of vessel k
c_k	Integer variable that indicates the departure time of vessel k
ρ_k	Integer variable that indicates tardiness in the departure time of vessel k
e_k	Integer variable that indicates earliness in the departure time of vessel k
z_{gkj}	Binary variable taking value one if crane g is assigned to vessel k in period j
π_{kn}	Binary variable taking value one if the first berthing position of vessel k is n
σ_{kn}	Binary variable taking value one if the berth section n is assigned to vessel k
α_{kj}	Binary variable taking value one if the berthing time of vessel k is j
β_{kj}	Binary variable taking value one if vessel k is berthing at time period j
γ_{kj}	Binary variable taking value one if the departure time of vessel k is $j+1$
η_{qkj}	Binary variable taking value one if q QCs are assigned to vessel k at time period j
l_{gk}	Auxiliary variable to linearise constraints for the time-invariant QC scheduling
ζ_k	Auxiliary variable used to capture delays in QC processing time through workload adjustment

With notations described above, we propose the following integer program for the deterministic BACASP considering carbon emissions.

$$\min \quad \sum_{k\in\mathcal{V}} \left\{c^W(t_k - \overline{A}_k) + c^B(c_k - t_k) + w\rho_k - \tau e_k\right\} + \sum_{j\in\mathcal{T}}\sum_{k\in\mathcal{V}}\sum_{g\in\mathcal{Q}}(c^Q + p^Q)z_{gkj} \tag{A1}$$

$$\text{s.t.} \quad x_{lk} + x_{kl} + y_{lk} + y_{kl} \geq 1, \quad k,l \in \mathcal{V},\ k < l \tag{A2}$$

$$x_{lk} + x_{kl} \leq 1, \quad k,l \in \mathcal{V},\ k < l \tag{A3}$$

$$y_{lk} + y_{kl} \leq 1, \quad k,l \in \mathcal{V},\ k < l \tag{A4}$$

$$b_k \leq J - H_k, \quad k \in \mathcal{V} \tag{A5}$$

$$t_k \geq \overline{A}_k, \quad k \in \mathcal{V} \tag{A6}$$

$$\sum_{k\in\mathcal{V}} z_{gkj} \leq 1, \quad j \in \mathcal{T},\ g \in \mathcal{Q} \tag{A7}$$

$$t_k \leq jz_{gkj} + (1 - z_{gkj})M, \quad j \in \mathcal{T},\ k \in \mathcal{V},\ g \in \mathcal{Q} \tag{A8}$$

$$c_k \geq (j+1)z_{gkj}, \quad j \in \mathcal{T},\ k \in \mathcal{V},\ g \in \mathcal{Q} \tag{A9}$$

$$\sum_{g\in\mathcal{Q}} z_{gkj} = \sum_{q\in\mathcal{N}^{QC}} q\eta_{qkj}, \quad j \in \mathcal{T},\ k \in \mathcal{V} \tag{A10}$$

$$\zeta_k \geq \sum_{q\in\mathcal{N}^{QC}} q^{\tilde{\alpha}}\eta_{qkj}, \quad j \in \mathcal{T},\ k \in \mathcal{V} \tag{A11}$$

$$\sum_{j\in\mathcal{T}}\sum_{q\in\mathcal{N}^{QC}} \eta_{qkj}q^{\tilde{\alpha}}p \geq W_k + \zeta_k p\overline{D}_k, \quad k \in \mathcal{V} \tag{A12}$$

$$\sum_{k\in\mathcal{V}}\sum_{q\in\mathcal{N}^{QC}} \eta_{qkj} \leq Q, \quad j \in \mathcal{T} \tag{A13}$$

$$\sum_{q\in\mathcal{N}^{QC}} \eta_{qkj} \leq 1, \quad j \in \mathcal{T},\ k \in \mathcal{V} \tag{A14}$$

$$b_k + H_k \leq E_g z_{gkj} + (1 - z_{gkj})J, \quad j \in \mathcal{T},\ k \in \mathcal{V},\ g \in \mathcal{Q} \tag{A15}$$

$$b_k \geq S_g z_{gkj}, \quad j \in \mathcal{T},\ k \in \mathcal{V},\ g \in \mathcal{Q} \tag{A16}$$

$$z_{gkj} + z_{g'lj} \leq 2 - y_{kl}, \quad j \in \mathcal{T},\ k,l \in \mathcal{V},\ g,g' \in \mathcal{Q},\ g' < g \tag{A17}$$

$$\sum_{g\in\mathcal{Q}} z_{gkj} \le N^{QC}, \quad j\in\mathcal{T},\ k\in\mathcal{V} \tag{A18}$$

$$\sum_{j\in\mathcal{T}} z_{gkj} \le M l_{gk}, \quad k\in\mathcal{V},\ g\in\mathcal{Q} \tag{A19}$$

$$\sum_{j\in\mathcal{T}} z_{gkj} \le c_k - t_k, \quad k\in\mathcal{V},\ g\in\mathcal{Q} \tag{A20}$$

$$\sum_{j\in\mathcal{T}} z_{gkj} \ge c_k - t_k - M(1-l_{gk}), \quad k\in\mathcal{V},\ g\in\mathcal{Q} \tag{A21}$$

$$\rho_k - e_k \ge c_k - D_k, \quad k\in\mathcal{V} \tag{A22}$$

$$b_k = \sum_{n\in\mathcal{B}} n\pi_{kn}, \quad k\in\mathcal{V} \tag{A23}$$

$$\sum_{n\in\mathcal{B}} \sigma_{kn} = H_k, \quad k\in\mathcal{V} \tag{A24}$$

$$\sum_{n\in\mathcal{B}} \pi_{kn} = 1, \quad k\in\mathcal{V} \tag{A25}$$

$$\pi_{kn} \ge \sigma_{kn} - \sigma_{k,n-1}, \quad k\in\mathcal{V},\ n\in\mathcal{B},\ n>1 \tag{A26}$$

$$\pi_{k1} \ge \sigma_{k1}, \quad k\in\mathcal{V} \tag{A27}$$

$$\pi_{kn} \le \sigma_{kn}, \quad k\in\mathcal{V},\ n\in\mathcal{B} \tag{A28}$$

$$\pi_{kn} \le 1 - \sigma_{k,n-1}, \quad k\in\mathcal{V},\ n\in\mathcal{B},\ n>1 \tag{A29}$$

$$y_{kl} + \sum_{m=\max\{n-H_l+1,0\}}^{J} \pi_{km} + \pi_{ln} \le 2, \quad k,l\in\mathcal{V},\ k\ne l,\ n\in\mathcal{B} \tag{A30}$$

$$z_{gkj} \le \sum_{n=S_g}^{E_g-H_k} \pi_{kn}, \quad j\in\mathcal{T},\ k\in\mathcal{V},\ g\in\mathcal{Q} \tag{A31}$$

$$t_k = \sum_{j\in\mathcal{T}} j\alpha_{kj}, \quad k\in\mathcal{V} \tag{A32}$$

$$c_k \ge (j+1)\beta_{kj}, \quad j\in\mathcal{T},\ k\in\mathcal{V} \tag{A33}$$

$$z_{gkj} \le \beta_{kj}, \quad j\in\mathcal{T},\ k\in\mathcal{V},\ g\in\mathcal{Q} \tag{A34}$$

$$\sum_{j\in\mathcal{T}} \alpha_{kj} = 1, \quad k\in\mathcal{V} \tag{A35}$$

$$\alpha_{kj} \ge \beta_{kj} - \beta_{k,j-1}, \quad j\in\mathcal{T},\ j>1,\ k\in\mathcal{V} \tag{A36}$$

$$\alpha_{k1} \ge \beta_{k1}, \quad k\in\mathcal{V} \tag{A37}$$

$$\alpha_{kj} \le \beta_{kj}, \quad j\in\mathcal{T},\ k\in\mathcal{V} \tag{A38}$$

$$\alpha_{kj} \le 1 - \beta_{k,j-1}, \quad j\in\mathcal{T},\ j>1,\ k\in\mathcal{V} \tag{A39}$$

$$x_{kl} + \beta_{ki} + \beta_{lj} \le 2, \quad i,j\in\mathcal{T},\ i\ge j-1,\ k,l\in\mathcal{V},\ k\ne l \tag{A40}$$

$$\gamma_{kj} \ge \beta_{kj} - \beta_{k,j+1}, \quad j\in\mathcal{T},\ j<M,\ k\in\mathcal{V} \tag{A41}$$

$$\gamma_{k1} \ge \beta_{kM}, \quad k\in\mathcal{V} \tag{A42}$$

$$\gamma_{kj} \le \beta_{kj}, \quad j\in\mathcal{T},\ k\in\mathcal{V} \tag{A43}$$

$$\gamma_{kj} \le 1 - \beta_{k,j+1}, \quad j\in\mathcal{T},\ j<M,\ k\in\mathcal{V} \tag{A44}$$

$$\sum_{j\in\mathcal{T}} \gamma_{kj} = 1, \quad k\in\mathcal{V} \tag{A45}$$

$$c_k = \sum_{j\in\mathcal{T}} (j+1)\gamma_{kj}, \quad k\in\mathcal{V} \tag{A46}$$

$$c_k \ge t_k + \sum_{j\in\mathcal{T}} \beta_{kj}, \quad k\in\mathcal{V},\ g\in\mathcal{Q} \tag{A47}$$

$$x_{kl}, y_{kl} \in \{0,1\}, \quad k,l\in\mathcal{V},\ k\ne l \tag{A48}$$

$$z_{gkj} \in \{0,1\}, \quad j\in\mathcal{T},\ k\in\mathcal{V},\ g\in\mathcal{Q} \tag{A49}$$

$$b_k, t_k, c_k \in \mathbb{Z}^+, \quad k\in\mathcal{V} \tag{A50}$$

$$\pi_{kn}, \sigma_{kn} \in \{0,1\}, \quad k\in\mathcal{V},\ n\in\mathcal{B} \tag{A51}$$

$$\alpha_{k,j}, \beta_{k,j}, \gamma_{k,j} \in \{0,1\}, \quad j\in\mathcal{T},\ k\in\mathcal{V} \tag{A52}$$

$$\rho_k, e_k, \zeta_k \in \mathbb{Z}_0^+, \quad k\in\mathcal{V} \tag{A53}$$

$$\eta_{qkj} \in \{0,1\}, \quad j\in\mathcal{T},\ k\in\mathcal{V},\ q\in\mathcal{N}^{QC} \tag{A54}$$

$$l_{gk} \in \{0,1\}, \quad k\in\mathcal{V},\ g\in\mathcal{Q} \tag{A55}$$

The objective function (A1) includes the operational costs and the carbon emission costs under the unitary carbon emission taxation rate. Constraints (A10) to (A14) are formulated to represent the QC interference. In particular, (A12) and (A11) are used to represent vessel departure delay by adjusting the workload. Constraints (A19) to (A21) ensure that QC schedules remain time-invariant. Constraint (A22) indicates that the tardiness and earliness of a vessel's departure time are determined based on the actual departure time and the requested departure time of the vessel. The remaining constraints pertain to the general continuous-layout BACASP formulation proposed by previous studies. We omit a more detailed explanation for them because it can be found in Agra and Oliveira (2018) and Rodrigues and Agra (2021).

Appendix 2. Data segmentation for the MIP approach

As described in Section 5.1, we divide an episode, which is composed of data on vessels arriving at the port over a one-week period, into smaller segments to solve the MIP model in a reasonable computation time. Each segment consists of five vessels, while the number of QCs and berth sections remains the same as in the original episode. Figure A1 shows the outline of the data segmentation for the performance evaluation.

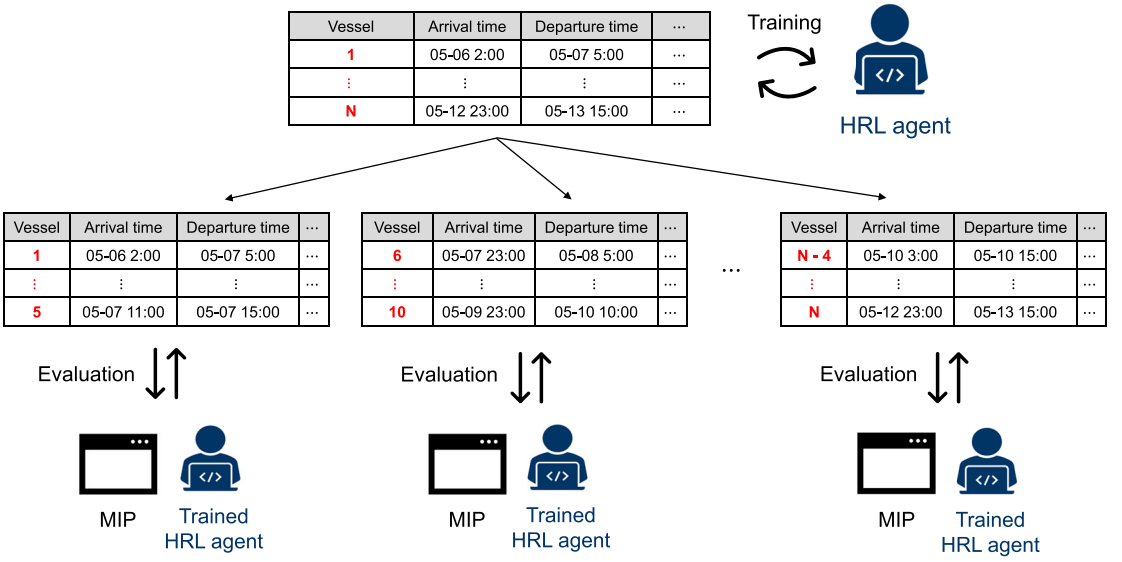


Figure A1. Instance generation for the performance evaluation of the HRL algorithm.