http://dx.doi.org/10.1016/j.trc.2024.104535


# Integrated planning model for two-story container ports

Lu Zhen [1], Zhiyuan Yang [1], Shuaian Wang [2*], Hongtao Hu [3], Ek Peng Chew [4], Tianyi Fan [1]

[1] *School of Management, Shanghai University, Shanghai, China*
[2] *Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*
[3] *School of Logistics Engineering, Shanghai Maritime University, Shanghai, China*
[4] *Department of Industrial and Systems Engineering, National University of Singapore, Singapore*

* Corresponding author: hans.wang@polyu.edu.hk

**Abstract**: This study introduces a novel two-stage stochastic programming model tailored for the unique infrastructure of two-story container ports, a design that significantly enhances land productivity and operational efficiency. Our model, addressing uncertainties in vessel arrival times and container loads, incorporates advanced features like innovative quay cranes, rooftop solar panels, and dense container storage. An efficient metaheuristic is designed for large-scale applications, demonstrating a 25.63% reduction in operational costs compared to traditional decision rules. We also distill key managerial insights, notably the advantage of adopting solar power technology with consistent energy output throughout the day, offering practical guidance for port operators contemplating vertical expansion strategies.

**Keywords**: Port operations, next-generation container ports, berth allocation, yard management.

## 1. Introduction

Despite the global economic slowdown caused by the COVID-19 pandemic and the Russia–Ukraine war, container throughput at some of the world's mega ports continues to increase at a slow but steady rate. As the largest container port in China, Shanghai Port's annual throughput is predicted to reach 48.2 million twenty-foot equivalent units (TEUs), a year-on-year growth of 2.5%; this implies that the port must handle over 100,000 TEUs per day. Faced with this operational burden, global port operators are exploring new port designs and advanced port technologies to help them increase their operational efficiency in the port area and decrease the environmental impact of their operations. In this regard, vertical expansion is a viable alternative to increase port efficiency, and it may reduce the cost of land reclamation/acquisition for port operators (Zaerpour et al., 2019). The two-story container port system may well be the future of port design in terms of vertical expansion. It was designed jointly by the Center for Next Generation Logistics (C4NGL) at the National University of Singapore, Shanghai Maritime University, and ZPMC, the world's largest manufacturer of port equipment. Figure 1 illustrates the novel design of a two-story container port. The advantages of this design are quite obvious. It can increase land productivity, and the two-story structure and solar panel roof can house refrigeration facilities. Another benefit of the two-story structure is multiple access points for quay cranes (QCs). ZPMC invented a specially designed QC for this port system that can handle container loading and

unloading on both floors of the port simultaneously. Port authorities (e.g., Port of Singapore Authority known for their innovative approaches to port operations) consider the design of this two-story container port to be realistic. It also won the grand prize in the Next Generation Container Port Challenge global contest, organized by the Port of Singapore Authority. The new port design can achieve the following goals: annual throughput of 20.125 million TEUs, berth-on-arrival (BOA) of 95.616%, and land productivity of 1,768 TEUs/hectare compared with 985 TEUs/hectare in the conventional design. In addition, the new design is pioneering in building a sustainable port system, as it uses solar energy as the energy source and a smart electric power grid for better management of energy generation and usage.

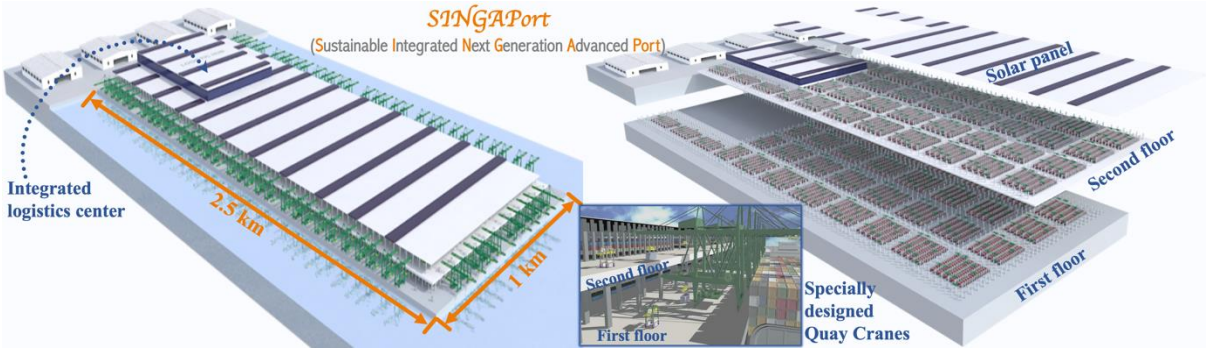


**Figure 1**: The design of a two-story container port

To operate the new port design at a high level of efficiency, port resources should be planned by keeping the new features of this port in mind. Figure 1 shows that the numbers of berths and yard blocks per unit of land area in the new port system are considerably higher than that of conventional ports. As mentioned, BOA and land productivity are the main metrics for evaluating a port's efficiency (Baird & Rother, 2013; Salido et al., 2011). For the new port system, the efficient planning of berth allocation and yard space reservation for visiting vessels by port operators is especially important. Berth allocation is linked to QC assignment because the latter decision affects the dwell time of vessels, which in turn affects the former decision. In this port system, QCs are specially designed; thus, QC assignment is more complex than in a conventional port. Moreover, the starting point of a port's resource planning is usually its long-term plan for berth allocation and yard space reservation (Cheimanoff et al., 2023), forming the basis for making other types of resource plans at different operational levels. To formulate long-term plans, it is essential to consider uncertainties, such as the random arrival times of vessels and fluctuations in their container volume in different periods (Hu et al., 2021). It is necessary to design model-driven decision tools to support port operators who wish to adopt the new two-story port in making the above long-term baseline plan. The goal of a port operator's decision should be to minimize the cost of the baseline plan and the expected cost of executing the plan in various future scenarios using random vessel arrival times and a random number of containers to be handled.

This paper conducts an exploratory study of this two-story container port and proposes an integrated optimization model to determine a long-term baseline plan from a tactical perspective for berth allocation and yard space reservation under uncertain vessel arrival times and number of containers to be loaded/unloaded. Unlike traditional single-layer port structures where containers are stacked only horizontally, double-layered docks utilize vertical space, allowing for stacking containers on two levels. The design shown as Appendix B significantly increases the container storage density, accommodating twice the number of containers in the same ground area compared to traditional single-layer yards. Because of the features of this novel port design, such as a two-story infrastructure, a triple hoist quay crane (a new type of tailored quay crane for the two-story container ports), solar panels on the port roof, and a high density of subblocks (i.e., the unit of yard space management), which are described in the Section 3, the decision models developed in the literature for conventional ports may not be applicable. The proposed model involves the above features of the port in its newly designed constraints and demonstrates a more complex structure than the related models for conventional ports. The proposed integrated optimization model is formulated as a two-stage stochastic programming model, in which uncertain arrival times and number of containers to be handled are taken into account in multiple scenarios in the execution of the baseline plan. Then, the model is linearized as a mixed-integer linear programming (MILP) model, which can be solved using commercial solvers in small-scale instances. A metaheuristic is also implemented to solve the model efficiently in large-scale instances. Furthermore, numerical experiments are conducted to validate the effectiveness of the proposed model and the efficiency of the algorithm. Sensitivity analyses are performed to derive useful managerial insights for port operators who may be interested in adopting the new port design or integrating some of its novel features into their current port systems.

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 elaborates the problem background for the new design. Sections 4 and 5 discuss the mathematical model and the suggested algorithm, respectively. Section 6 reports and discusses the experimental results. Last, Section 7 concludes the paper.

## 2. Related literature

The subject of this study belongs to the operation optimization models for next-generation container ports. This study is related to two streams of the literature. One deals with analytical studies on next-generation container ports and the other deals with optimization models and algorithms on port operations in related areas, such as berth allocation and yard management.

The first stream of the literature concerns analytical studies on next-generation container port systems, such as a high-density container tower port, a frame bridge-based automated container terminal (ACT) system, and an overhead grid system. Zaerpour et al. (2019) designed an innovative cylindrical container storage system for a high-density container tower port, which loads containers into the designated stacking area using a crane located in the center of the tower. Compared with the cube storage system stacked vertically in a traditional terminal, the cylindrical container storage system does not need to turn over the container, considerably decreasing the container handling time. Similar to Zaerpour et al. (2019), a few studies (Gue and Kim, 2007; de Koster et al., 2008; Zaerpour et al., 2015) have focused on a compact storage system, which is a rectangular cuboid system. Zhu et al. (2010) studied a new container terminal system for a frame bridge-based ACT system, in which shuttles are used to transfer containers between the seaside and the stacking area. Similarly, Zhen et al. (2012) introduced a new type of container terminal system that utilizes multi-story frame bridges and rail-mounted trolleys to transport containers between the quayside and the yardside areas. Several analytical models and performance measures have been developed to compare the new ACT system and the commonly used automated guided vehicle-based system. Unlike Zhen et al. (2012), who studied modeling methods, Hu et al. (2013) developed a Markov chain model to analyze the throughput of transfer platforms in the new container terminal system. Abou Kasm & Diabat (2020) designed a new QC technology: ship to shore multi-trolley portal gantry container crane which can access two bays simultaneously. Abou Kasm et al. (2022) explored a scheduling problem for a new type of QC, which has four trolleys for container loading and unloading operations on both sides of a ship. Furthermore, Zhou et al. (2016, 2017) introduced the hybrid goods retrieval and inventory distribution system, which is a fully automated multi-directional overhead handling structure and uses an underslung transfer unit to handle containers. The authors carried out simulations to validate its high yard storage capacity and handling productivity through comparison with three typical container port layouts. In addition to the above-mentioned innovations on automated equipment or systems applied to next-generation container ports, some studies have focused on terminal layouts to accommodate an increasing number of containers. Öztürkoğlu et al. (2012) introduced three non-traditional warehouse designs (chevron, leaf, and butterfly) in continuous space. The authors formulated a discrete model to demonstrate that the chevron design is best for many industrial applications. Furthermore, Kalmar (part of Cargotec Corporation) proposed a novel layout wherein an underground transportation system is used to separate transshipment movement from vessel operations (Gharehgozli et al., 2020). Higher capacity, lower horizontal expansion, and higher efficiency in handling containers have gradually become the industrial norms for next-generation container ports with increasing container throughput. The two-story

container port introduced in this study was developed for the future needs of Singapore; it can store more containers with higher land utilization, handle more containers using the novel triple hoist QC (described in the following section), and achieve sustainable development for green ports (Gharehgozli et al., 2020). The findings of this study yield valuable insights that can be effectively applied to enhance the operational efficiency of such next-generation container ports in the future.

The second stream of the literature concerns optimization models and algorithms for port operations in related areas, such as berth allocation and yard management. Some scholars have conducted integrated optimization studies that simultaneously consider decisions on the berth side, QC side, and yard side (Wang et al., 2018; Zhen et al., 2022). The column generation-based solution framework is effectively used to solve the proposed complex mixed-integer planning (MIP) model in large-scale instances. Similarly, Liu (2020) considered transshipment, import/export activities, and productivity losses caused by QC interference. The author proposed an iterative heuristic to solve the proposed model. The continuous or discrete berth allocation problem (CBAP/DBAP) is commonly linked to the QC assignment problem (QCAP). In this regard, Thanos et al. (2021) presented a MIP model to address CBAP with time-variant QCAP to minimize container transshipment distances within the terminal yard. The authors proposed a local search-based heuristic algorithm to accommodate fixed and flexible vessel departure time settings, which can solve the real-world instance with 60 QCs and 60 vessels in a quay length of 4,000 m. Cheimanoff et al. (2022) also studied a similar algorithm, which addresses continues quay topology and dynamic vessel arrivals and considers specific time-invariant QC assignment. Robenek et al. (2014) studied the above integrated optimization problem in the context of bulk ports and proposed a branch-and-price algorithm to solve the proposed MIP model; instances with 40 vessels were solved within approximately 4 hours. Uncertainty has been widely considered in research on the berth allocation problem (BAP) (Zhen, 2015; Tang et al., 2022). Some studies (Martin-Iradi et al., 2022; Guo et al., 2023) have further examined a multi-port BAP. The yard space allocation problem (YSAP) is also an important topic that has been widely studied in yard management (He et al., 2020). Hu et al. (2021) investigated a yard template planning problem that considers an uncertain number of transport containers. The authors developed an improved benders decomposition to solve the two-stage stochastic programming model. Their numerical experiment showed that it could solve large-scale instances with 36 container groups, 10 berths, and 160 subblocks under 10 scenarios. However, the computing time was more than 6 hours for solving the largest-scale instances, which is too time-consuming to cater to industry demand. Yu et al. (2021) applied the concept of flexible loading cluster in flexible yard management, where the numbers of bays, rows, and tiers can be changed. Yang et al. (2022a) formulated a two-stage robust approximation model in the form of a single MILP to optimize the yard template

under uncertain vessel arrival times. Furthermore, Jin et al. (2016) and Yang et al. (2022b) incorporated YSAP and the yard crane deployment problem (YCDP) into one MIP model. Most studies have addressed the yard space allocation problem based on the consignment strategy (Lee et al., 2006; Moorthy and Teo, 2006; Han et al., 2008). For a comprehensive understanding, Table 1 categorizes the existing studies by their research focus and methodologies employed.

**Table 1**: Overview of studies on the berth allocation and yard space allocation

| Reference | Problem | Uncertainty | | | Yard allocation strategy | | Solution methods |
|---|---|---|---|---|---|---|---|
| | | N | T | O | consignment | sharing & exclusive | |
| Cheimanoff et al. (2023) | DBAP&YSAP | | | | √ | | **E&H** |
| Guo et al. (2023) | DBAP | | √ | | | | **E&H** |
| Tang et al. (2022) | CBAP | √ | √ | √ | | | **H** |
| Yang et al. (2022a) | YSAP | | √ | | √ | | **E&H** |
| Yang et al. (2022b) | YSAP& YCDP | | | | √ | | **H** |
| Zhen et al. (2022) | DBAP&QCAP& YSAP | √ | √ | | √ | | **E&H** |
| Yang et al. (2022) | YAP&YCDP | √ | | | √ | | **H** |
| Cheimanoff et al. (2022) | CBAP&QCAP | | | | | | **H** |
| Hu et al. (2021) | YSAP | √ | | | √ | | **E** |
| Thanos et al. (2021) | CBAP&QCAP | | | | | | **H** |
| Yu et al. (2021) | YSAP | | | √ | | √ | **H** |
| He et al. (2020) | YSAP | √ | | | √ | | |
| Liu (2020) | DBAP&QCAP& YSAP | | | √ | √ | | **H** |
| Wang et al. (2018) | DBAP&QCAP& YSAP | | | | √ | | **E&H** |
| He and Tan (2018) | YSAP | √ | | | | √ | **H** |
| Jin et al. (2016) | YSAP &YCDP | | | | √ | | **H** |
| Zhen (2015) | CBAP | | √ | | | | **H** |
| Zhen (2014) | YSAP | √ | | | | √ | **H** |
| Robenek et al. (2014) | CBAP& QCAP | | | | | | **E&H** |
| This study | DBAP&QCAP& YSAP | √ | √ | | | √ | **H** |

***Note:*** **DBAP**: Discrete berth allocation problem; **CBAP**: Continuous berth allocation problem; **QCAP**: quay cranes assignment problem; **YSAP**: yard space allocation problem; **YCDP**: yard cranes deployment problem; **N**: uncertain number of incoming/storing containers or uncertain number of vessels; **T**: uncertain arrival times/operation times of vessels; **O**: other uncertain factor such as weather condition, equipment failure and retrieving sequence. **Yard allocation strategy**: the approach adopted by port operators in managing the assignment of yard space; **Consignment**: a dedicated approach in the allocation of yard subblocks and each subblock is reserved for a specific vessel's containers; **Sharing & exclusive**: an innovative approach introduces adaptability in yard allocation by blending two modes. **E**: exact algorithm (benders decomposition, branch and price, branch and cut et.al); **H**: heuristic (metaheuristic and swamp swarm intelligence algorithm); **E&H**: solution that combines exact and heuristic algorithm (column generation-based approach and other hybrid algorithm).

While there has been considerable research in traditional single-layer container port operations, it is observed that analytical studies specifically addressing the unique challenges and integrated decision models of two-story container port systems are relatively scarce. Through a comprehensive review of existing studies, this paper identifies and addresses a critical gap in the literature concerning two-story container port systems. Our model innovatively addresses the unique operational challenges of these ports, focusing on load distribution strategies for quay cranes engineered for vertical operations and the

integration of solar power into operational planning. This approach significantly diverges from existing literature that predominantly concentrates on single-layer port operations, by introducing a novel consideration of QC profile allocation for two-layer cranes. Moreover, this study extends its contribution by conducting sensitivity analysis experiments specifically designed for the attributes of two-story ports. These experiments offer tailored management insights and analytical perspectives that are absent in current literature, thereby furnishing a nuanced understanding of how dual-layer ports can adapt to various operational scenarios.

In essence, this research not only enriches the theoretical framework of port operation optimization but also delivers practical insights for the efficient management and strategic planning of two-story container ports. By addressing these novel and complex challenges, the study paves the way for significant advancements in the design and operation of future port infrastructure.

## 3. Problem background

Irrespective of the type of port, berth allocation is one of the most important decisions in a port because for port operators, the berth allocation plan is the starting point to make other types of plans for the port, for example yardside resources and vehicle-related plans. In addition, the baseline plan for berth allocation determines the expected departure time of vessels, which in turn influences the planning of voyage and service schedules of vessels by vessel owners (i.e., liners) (Dai et al., 2023). Therefore, this study focuses on the berth allocation decision for a novel type of terminal system—a two-story container port. As berth allocation is a mature research topic in the field of port operations (Stahlbock and Voß, 2008; Bierwirth and Meisel, 2010; Rodrigues and Agra, 2022), this section mainly elaborates the features of this new type of port and their influence on the related decisions in our problem.

In berth allocation, we need information about the dwell time of vessels, which is determined by QC assignment. Thus, usually, berth allocation is jointly decided with QC assignment. Many recent related studies have adopted the concept of QC profile in the berth allocation and QC assignment problems (Giallombardo et al., 2010; Liu et al., 2016; Wang et al., 2018; Xie et al., 2019); this concept reduces the solution space and improves decision efficiency. This study extends the concept of QC profile to the context of a two-story container port. Figure 2 illustrates an example of berth allocation and the extended QC profile in a two-story container port.

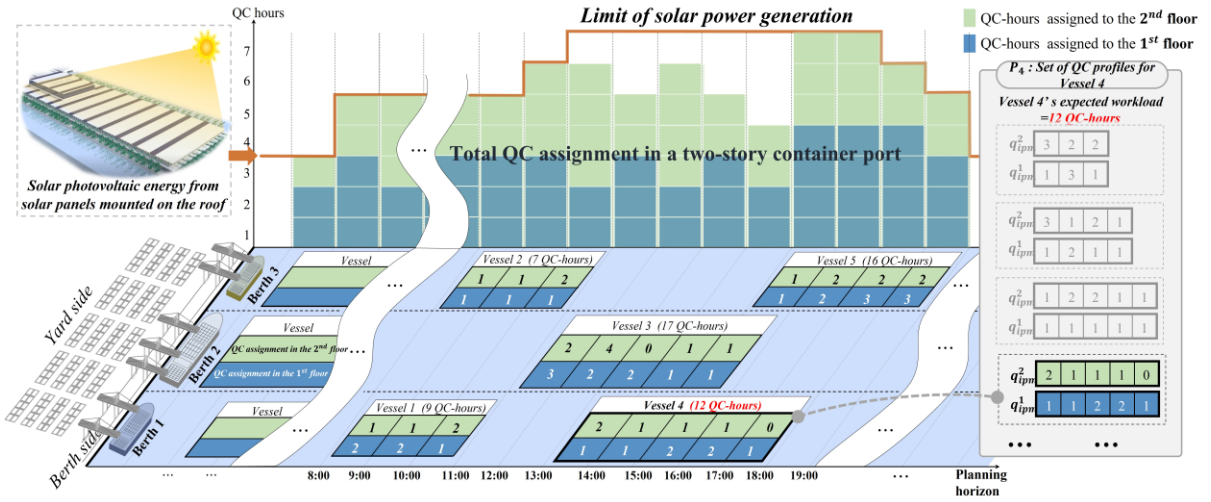


**Figure 2**: Berth allocation and QC profile assignment in a two-story container port

Figure 2 shows that the workload of loading and unloading containers for Vessel 4 is 12 QC hours; here, we assume that time is discretized into time steps and that one time step is 1 hour. In a traditional port, the QC profile distributes the workload among the hours of the vessel's total dwell time and determines the number of QCs that will serve the vessel in each hour. However, in a two-story port, the QC profile must specify the number of QCs that will serve on the first and second floors in each hour. In the example shown in Figure 2, one QC is assigned to the first floor and two QCs are assigned to the second floor to handle Vessel 4's workload in the first hour of its dwell time. The distribution of the workload of the QCs on both floors is specified for each hour because the gap between the total workload of the QCs on the first floor and those on the second floor should not exceed a pre-set value, which guarantees that the workload is balanced between the two floors of the port.

One thing must be clarified here. In the example given above, this does not mean that only one QC (or two QCs) will be dedicated to operations on the first (or the second) floor during the entire hour. Also note that the triple hoist QC, which is a novel type of QC, is specially designed for two-story container ports. The triple hoist QC can handle the unloading/loading operations of both floors simultaneously. Figure 3 shows that the triple hoist QC has a tandem lift. Specially designed components (e.g., front trolley, rear trolley, portal trolley, and lashing platform) allow the QC to serve both floors simultaneously. Therefore, the above-mentioned arrangement in which Vessel 4's workload is handled by one QC on the first floor and two QCs on the second floor could be executed in practice as follows. Three QCs serve Vessel 4 simultaneously, and each QC spends 33.3% and 66.7% of its operating time on the first floor and the second floor, respectively. In Figure 3, other optional combinations for the division of the operating time of each QC are equivalent to the above-mentioned case. Note that the decision to specify the distribution of the QCs' workload between the two floors is

not necessarily considered in this problem. This operational-level decision could be made according to the work schedule in reality.

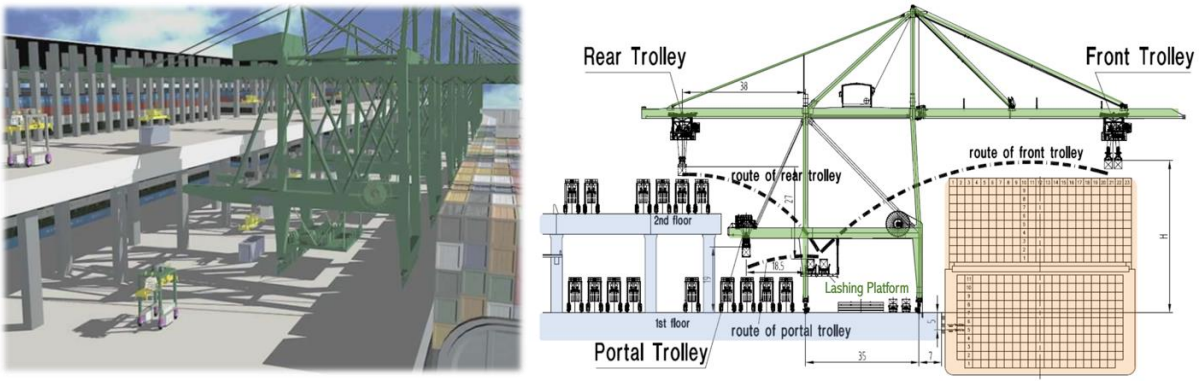


**Figure 3**: A triple hoist QC specially designed for a two-story container port

Figure 1 shows another feature of the new container port type—the solar panels on the roof of the port. The design philosophy here is to build a zero-emission port. This feature also influences the plan for port resource allocation, especially QC assignment. The number of QCs used in each hour reflects the total workload fulfilled, which is approximately proportional to the electricity used for the entire port's operations. As mentioned, the solar panels generate electricity for port operations and there is intraday variation in power generation capacity. Thus, in this decision problem, the number of QCs used in each hour should not be greater than a time-dependent function that is related to the intraday variation in the power generation capacity of the solar panels.

Within the distinctive framework of a two-story container port, efficient yard space management becomes paramount, given the substantial increase in available storage area. This novel context necessitates a sophisticated approach to yard allocation, prompting our adoption of a dual-mode strategy, specifically exclusive and sharing modes. The dual-mode allocation system is a direct response to the enhanced storage capacity of two-story ports, addressing the need to avoid both congestion and underuse. By dynamically allocating yard space, this system mitigates the risks associated with unpredictable container flows. The exclusive mode safeguards space for specific vessels, enhancing predictability for port operators and shipping companies alike. Conversely, the sharing mode offers operational agility, accommodating excess containers beyond the reserved exclusive subblocks, thus optimizing space utilization and maintaining smooth operations. Usually, a subblock is the unit area for yard management in ports; the capacity of a subblock is normally 240 TEUs. In this study, yard management mainly refers to the allocation of subblocks to vessels. For example, some subblocks are allocated to vessel $i$. In the period before the arrival of vessel $i$ at the port, containers are unloaded from other vessels or transported by trucks from hinterland to the port; these containers will be loaded onto vessel $i$ later. Then, these containers are stored at the subblocks that have been allocated to vessel

$i$. The benefit of this arrangement is that all of the containers stored at the subblocks can be loaded onto vessel $i$ when it arrives at the port, and there is no need to reshuffle the containers, which reduces vessel $i$'s dwell time and increases the port's operational efficiency. Because the number of containers to be loaded onto a vessel during one cycle time (usually 1 week) is unknown and changes weekly, some subblocks are reserved for the vessel in the exclusive mode when making the baseline plan; here, the exclusive mode means that these subblocks cannot be used by other vessels. However, if the reserved subblocks are not sufficient in a given week for a vessel, the vessel could use the subblocks in the sharing mode. The exclusive reservation of subblocks to vessels is made in the baseline plan and is a long-term decision, which forms the basis for the contract between the port and the shipping liners that operate the vessels. The unit cost of using a subblock in the exclusive mode should be lower than the unit cost in the sharing mode. If too many subblocks are reserved exclusively for a vessel, then many of the subblocks may end up as surplus in some weeks. In contrast, if too few subblocks are reserved for a vessel, then more subblocks may have to be shared in some weeks, which has a higher unit cost. Thus, it is important to determine the appropriate number of subblocks to be reserved exclusively for each vessel in this problem. Figure 4 shows that the port operator usually reserves the subblocks (which are close to a vessel's allocated berth) to a vessel in such a way that it decreases the transportation distance of the trucks. The subblocks that are further away from the berths but in the center of the yard could be used in the sharing mode, as Figure 4 shows.

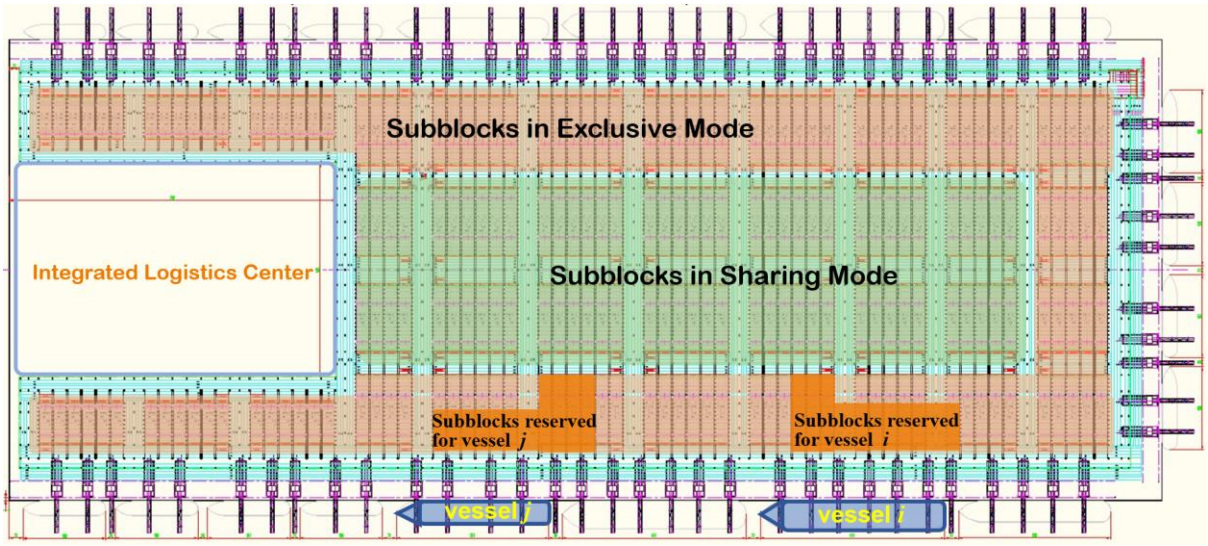


**Figure 4**: Dual-mode (exclusive and shared) yard management for a two-story container port

Based on the above-mentioned problem background of the new port design, the decision problem examined in this study is summarized as follows. The main decision is to make a baseline plan for berth allocation and yard space reservation for vessels. The baseline plan includes the berth to be allocated to

each vessel and the number of subblocks to be reserved exclusively for each vessel. QC profile assignment, which affects vessel dwell time, is also determined with berth allocation. In addition, the baseline plan for berth allocation determines the planned service time slot for each vessel. These decisions are made in the first stage of the formulated two-stage stochastic programming model. The objective of the first-stage decision is to minimize the deviation in the vessels' planned service time slots from their expected time windows and the number of exclusively reserved subblocks. This objective stems from a cost-minimization strategy, where the port operator, wary of the uncertainties inherent in maritime logistics, opts for a flexible approach. The allocation of the reserved quantity for exclusive subblocks is also strategically planned by the port operator. This allocation directly influences the human resource maintenance costs and the operational and maintenance costs of various equipment for the port operator. Therefore, in the first phase of planning, it is imperative for the port operator to consider how to fulfill the container loading and unloading service contracts for different shipping lines with the least possible yard space, including the corresponding operational equipment, under the presence of uncertainties. This approach not only optimizes resource utilization but also aligns with the overarching goal of enhancing operational efficiency while minimizing costs in a dynamic and unpredictable maritime environment.

The baseline plan is made for a relatively long planning horizon. However, the arrival time of vessels and the number of containers to be loaded/unloaded are not the same in different weeks of the planning horizon; they are uncertain in this problem. Therefore, the second-stage subproblem minimizes the expected value of actual operational costs in different scenarios; the actual cost in each scenario includes the penalty for delay in the actual departure time from the planned time, the yard operating cost related to the actual number of containers handled, and the extra cost of using subblocks in the sharing mode in some weeks. Note that the berth allocated to a vessel in the baseline plan is kept unchanged in the actual scenarios, which is also in keeping with realistic port operations. In reality, shipping liners usually enter into a contract with a port operator, which specifies the dedicated berths and yard space for their vessels. Except the decisions on berth allocation and yard space reservation, the remaining decision variables in the first stage are similar to those in the second stage. It is important to note that the present study does not delve into the allocation decisions of yard space. This omission is partly due to the similarity in yard allocation decisions between double-layered container terminals and traditional terminals, a topic which has already been extensively explored in existing literature (e.g., Hu et al., 2021; Wang et al., 2023). Furthermore, from a resource perspective, the yard resources in double-layered container terminals are significantly expanded due to the dual-layer platform, as illustrated in Appendix B. This expansion inherently leads to a situation where the decision-making regarding the

new type of QCs load becomes more critical in constraining the operational efficiency and service costs of the port than yard space allocation. Consequently, our study focuses more on the allocation of double-layered QC profiles in this novel port system, temporarily sidelining the specific allocation of yard space for each vessel.

The constraints of the proposed two-stage model take into account the aforementioned features of the new port design, such as the balancing of the workload between the two floors of the port, the new QC profile concept for the novel type of QC that handles container operations on both floors, electric power generated by the solar panels on the port roof, and dual-mode (exclusive and shared) yard management.

## 4. Mathematical model

The integrated planning problem for the two-story port is formulated as a two-stage stochastic programming model. The model aims to minimize the total costs in deciding the baseline plan for berth and yard space allocation by considering future uncertainties, which are featured using a set of scenarios in the second-stage subproblem. The main decision of the problem is the baseline plan determined in the first stage of the model; auxiliary decisions for actual operations in each future scenario are determined in the second stage of the model. The next section proposes the two-stage stochastic programming model and the linearization of the proposed mathematical model.

### 4.1 Notations

Before formulating the mathematical model, the indices, sets, parameters and decision variables are defined first. For the easy of understanding, we use capital letters, Latin letters and Greek letters to denote sets, parameters and decision variables, respectively.

**Indices and sets:**

$V$ set of arriving vessels, indexed by $i, j$.

$B$ set of all berths, indexed by $b$.

$K$ set of all subblocks in the two-story container terminal, indexed by $k$.

$P_i$ set of QC-profiles for vessel $i$ according to the expected workload, indexed by $p$.

$\tilde{P}_{i\omega}$ set of QC-profiles for vessel $i$ according to the actual workload in scenario $\omega$.

$T$ set of time steps, $T = \{1, \dots, H\}$, indexed by $t$.

$\Omega$ set of scenarios, indexed by $\omega$.

$e, e'$ dummy vessel denoting the first and the last one in a sequence of vessels that moors at a berth.

**Parameters:**

$h_{ip}$ handling time of vessel $i$ by using QC-profile $p$ with unit of time step.

$q_{ipm}^1$ workloads fulfilled by vessel $i$' s QC-profile $p$ for the first floor at the $m^{th}$ time step; $m = 1,2, \dots, h_{ip}$.

$q_{ipm}^2$ workloads fulfilled by vessel $i$' s QC-profile $p$ for the second floor at the $m^{th}$ time step.

$d^{LE}$ unit cost for handling a container loaded from subblocks in the exclusive mode.

$d^{LS}$ unit cost for handling a container loaded from subblocks in the sharing mode.

$d^U$ unit cost for handling a container unloaded to subblocks.

$\tilde{a}_{i\omega}$ actual arrival time of vessel $i$ in scenario $\omega$.

$\tilde{l}_{i\omega}$ actual number of loading containers for vessel $i$ in scenario $\omega$.

$\tilde{u}_{i\omega}$ actual number of unloading containers for vessel $i$ in scenario $\omega$.

$[a_i^f, b_i^f]$ feasible service time when making baseline plan for vessel $i$.

$[a_i^e, b_i^e]$ expected service time when making baseline plan for vessel $i$.

$c_i^P$ unit penalty cost caused by the deviation of vessel $i'$s planned service time window from its expected service time, during the stage of planning the baseline plan.

$c_i^D$ unit penalty cost caused by the delay of vessel $i$'s actual departure time from its planned departure time determined in the baseline plan.

$c^E$ unit cost of occupying a subblock in the exclusive mode for a planning horizon.

$c^S$ unit cost of occupying a subblock in the sharing mode for a planning horizon.

$y_i$ minimum number of subblocks that should be reserved for vessel $i$ in the exclusive mode.

$f_t$ maximum workload that can be supported by the port's solar panel at time step $t$.

$x^{Gap}$ maximum gap between the two floors' workloads.

$p_\omega$ probability of $\omega$, $\sum_{\omega\in\Omega} p_\omega = 1$.

$z$ capacity of a subblock in terms of TEUs, which equals to 240 (5 tiers $\times$ 6 lanes $\times$ 8 slots, Zhen et al., 2011, the details are shown in Appendix C).

$Q$ maximum number of available QCs.

$M$ a sufficiently large positive number.

**Decision variables:**

$\alpha_{ib}$ binary; equals one if vessel $i$ is allocated to berth b in the baseline plan, otherwise zero.

$\beta_{bij}$ binary; equals one if vessel $j$ moors at berth $b$ immediately after $i$, otherwise zero.

$\gamma_{ip}$ binary; equals one if vessel $i$ is served by QC-profile $p$ in baseline plan, otherwise zero.

$\eta_{ipt}$ binary; equals one if vessel $i$ is served by QC-profile $p$ and begins handling at time step $t$ in baseline plan, otherwise zero.

$\rho_t^1, \rho_t^2$ workload that provided by used QCs for first floor and second floor at time step $t$.

$\mu_i^s, \mu_i^e$ start and end time steps of handling for vessel $i$ in the baseline plan, it should be noted that both $\mu_i^s$ and $\mu_i^e$ indicate the commencement of the time step.

$\sigma_i$ number of subblocks that are reserved for vessel $i$ in the exclusive mode.

### 4.2 Objective of the model

The objective of the model is to minimize the total cost that contains three parts:

(*i*) The cost related to the deviation of vessels' planned service time slot from their expected time windows. For vessel $i$, the planned service time slot is denoted by $[\mu_i^s,\ \mu_i^e]$, the expected time window is denoted by $[a_i^e,\ b_i^e]$. Then the cost is calculated as $\sum_{i\in V} c_i^p[(a_i^e - \mu_i^s)^+ + (\mu_i^e - b_i^e)^+]$, here $c_i^p$ is the unit penalty cost caused by the above mentioned deviation.

(*ii*) The cost related to the number of subblocks reserved for ships, i.e., $c^E \sum_{i\in V} \sigma_i$; here, $\sigma_i$ is number of subblocks reserved for ship $i$ in the exclusive mode, $c^E$ is fixed cost of occupying a subblock in the exclusive mode for a planning horizon.

(*iii*) The expected cost of actual operations in future scenarios, i.e., $\sum_{\omega\in\Omega} p_\omega \Theta(\omega, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha})$; here $p_\omega$ is the probability of scenario $\omega$, and the cost of actual operations in scenario $\omega$ is denoted by $\Theta(\omega, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha})$, which is the objective of the second-stage subproblem and is elaborated later.

Based on the above three types of costs, the objective of the model is defined as follows:

$$[\boldsymbol{\mathcal{M}0}]\ \text{Minimize}\ \sum_{i\in V}\{c_i^p[(a_i^e - \mu_i^s)^+ + (\mu_i^e - b_i^e)^+] + c^E\sigma_i\} + \sum_{\omega\in\Omega} p_\omega \Theta(\omega, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}). \quad (1)$$

### 4.3 Constraints in the first stage of the model

As aforementioned at the beginning of this section, the core decision of the problem is the baseline plan of the berth and yard space allocation for vessels. The constraints in the first stage of the model are mainly related to the details of these decisions.

*4.3.1 Berth allocation in the baseline plan*

When allocating vessels to berths, the vessels allocated to each berth form a sequence of vessels. The following constraints are formulated for the berth allocation.

$$\sum_{b\in B} \alpha_{ib} = 1 \qquad \forall i \in V \quad (2)$$

$$\sum_{j\in V\cup\{e'\}} \beta_{bej} = \sum_{i\in V\cup\{e\}} \beta_{bie'} = 1 \qquad \forall b \in B \quad (3)$$

$$\sum_{j\in V\cup\{e'\}} \beta_{bij} = \sum_{j\in V\cup\{e\}} \beta_{bji} = \alpha_{ib} \qquad \forall i \in V, b \in B\,. \quad (4)$$

Constraints (2) ensure that each vessel is allocated to one berth. Constraints (3) are about a dummy vessel ($e$ or $e'$) that is in the beginning or the end of the sequence of vessels mooring at a berth. Constraints (4) guarantee that a vessel is at one certain position of the sequence of vessels mooring at a

berth if the vessel is allocated to the berth.

*4.3.2 QC arrangement in the baseline plan*

The QC arrangement affects each vessel's dwell time at the port, which further influences the berth allocation decision. For the purpose of reducing the solution space of the model and improving the decision efficiency, this study adopts the concept of "QC-profile" and allocates the QC-profiles to vessels. These profiles, representing the allocation of quay cranes to vessels and the subsequent workload distribution, are meticulously calculated to optimize handling time and operational throughput. Recognizing the technical depth of this process, we have detailed the generation methodology within the Appendix A. The constraints related the QC-profile allocation are listed as follows.

$$\sum_{p\in P_i}\gamma_{ip}=1 \qquad \forall i\in V \qquad (5)$$

$$\sum_{p\in P_i}\sum_{t\in T}\eta_{ipt}=1 \qquad \forall i\in V \qquad (6)$$

$$\sum_{p\in P_i}\sum_{t\in T}t\eta_{ipt}=\mu_i^s \qquad \forall i\in V \qquad (7)$$

$$\eta_{ipt}\le\gamma_{ip} \qquad \forall i\in V, p\in P_i, t\in T \qquad (8)$$

$$\mu_i^s+\sum_{p\in P_i}h_{ip}\gamma_{ip}-1=\mu_i^e \qquad \forall i\in V \qquad (9)$$

$$\mu_i^e+1\le\mu_j^s+M(1-\beta_{bij}) \qquad \forall i,j\in V, i\ne j, b\in B \qquad (10)$$

$$\mu_i^s\ge a_i^f \qquad \forall i\in V \qquad (11)$$

$$\mu_i^e\le b_i^f \qquad \forall i\in V \qquad (12)$$

Constraints (5) ensure one QC-profile is allocated to each ship. Constraints (6)–(8) are formulated for specifying the start time step of handling for each vessel in the baseline plan. Constraints (9) connect the start and end time steps of handling for each vessel. Constraints (10) ensure the end time of handling for a vessel $i$ is no later than the start time of handling for another vessel $j$ that immediately follows vessel $i$ in the sequence of vessels mooring at the same berth. Constraints (11) and (12) guarantee the planned service time slot for a vessel is completely inside its feasible time window.

*4.3.3 Workload balance between the two floors of a port*

As the "two-story" and "solar energy driven" are the main features of the port, the workload for container loading and unloading should be balanced between the two floors of the port, and the total workload should also be not exceed a certain limit that is related to electric power of the port roof's solar panel. These constraints are as follows.

$$\rho_t^1=\sum_{i\in V}\sum_{p\in P_i}\sum_{m=\max\{1,t-h_{ip}+1\}}^{t}\eta_{ipm}q_{ip(t-m+1)}^1 \qquad \forall t\in T \qquad (13)$$

$$\rho_t^2=\sum_{i\in V}\sum_{p\in P_i}\sum_{m=\max\{1,t-h_{ip}+1\}}^{t}\eta_{ipm}q_{ip(t-m+1)}^2 \qquad \forall t\in T \qquad (14)$$

$$\rho_t^1+\rho_t^2\le\min\{f_t,Q\} \qquad \forall t\in T \qquad (15)$$

$$|\rho_t^2 - \rho_t^1| \le x^{Gap}. \qquad \forall t \in T \qquad (16)$$

Constrains (13) and (14) calculate the QC workload occurred in the first floor and the second floor, respectively, at the time step $t$. Constraints (15) ensure the total workload (the number of required QCs) of the two floors at time step $t$ cannot exceed $f_t$ (i.e., the maximum workloads that can be supported by the port's solar panel at time step $t$) and $Q$ (i.e., the number of QCs). Constraints (16) guarantee the workload gap between two floors is no greater than a threshold, which ensures the workload balance between the two floors.

*4.3.4 Yard space reservation in baseline plan*

Besides the berth allocation, the yard space reservation is another important decision in the baseline plan, and mainly aims to reserve a proper number of subblocks to vessels in the exclusive mode. The related constraints are formulated as follows.

$$\sigma_i \ge y_i \qquad \forall i \in V \qquad (17)$$

$$\sum_{i \in I} \sigma_i \le |K|. \qquad (18)$$

Constraint (17) guarantees that the allocation of subblocks to a vessel meets or exceeds the minimum reservation requirement, a critical component of yard template planning amidst the uncertainties of the maritime market, as discussed in literature such as Zhen (2014). This minimum reservation threshold is typically derived through contractual negotiations between port authorities and shipping liners, informed by demand projections and historical data analysis. Such a strategy ensures the port's ability to efficiently manage container loading and unloading operations for vessels with substantial cargo handling needs, even under fluctuating market conditions. Constraint (18) states the total number of reserved subblocks should not exceed the total available number of subblocks in the port.

*4.3.5 Decisions in the baseline plan*

The core of this decision problem is to determine the baseline plan, which is reflected by the decision variables in the first stage of the model. The decision variables as well as their domains are defined in Constraints (19)–(25).

$$\alpha_{ib} \in \{0,1\} \qquad \forall i \in V, b \in B \qquad (19)$$

$$\beta_{bij} \in \{0,1\} \qquad \forall i, j \in V, i \ne j, b \in B \qquad (20)$$

$$\mu_i^e, \mu_i^s \ge 0 \qquad \forall i \in V \qquad (21)$$

$$\gamma_{ip} \in \{0,1\} \qquad \forall i \in V, p \in P_i \qquad (22)$$

$$\eta_{ipt} \in \{0,1\} \qquad \forall i \in V, p \in P_i, t \in T \qquad (23)$$

$$\rho_t^1, \rho_t^2 \ge 0 \qquad \forall t \in T \qquad (24)$$

$$\sigma_i \in Z^+ \qquad \forall i \in V. \qquad (25)$$

### 4.4 Objective of the second-stage model

The second-stage model is formulated for the actual realizations in future scenarios. The objective of the second stage model, i.e.,$\Theta(\omega, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha})$, which is contained in the original model's objective as aforementioned at the beginning of this section. The objective of the second stage model is to minimize the total cost of the actual operations in each scenario, indexed by $\omega$. Some of the decision variables' values (i.e., $\mu_i^e, \sigma_i, \alpha_{ib}$) are the parameters of the second stage model; $\alpha_{ib}$, $\mu_i^e$ and $\sigma_i$ are about berth allocation, end time of service, and yard space reservation in the baseline plan, respectively. Thus the objective of the second stage model is denoted by $\Theta(\omega, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha})$. In the context of mathematical modeling, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\alpha}$ represent the value vectors of variables $\mu_i^e$, $\sigma_i$ and $\alpha_{ib}$, respectively. The objective of the second stage contains three parts:

(*i*) The cost related to the delay of a vessel's actual departure time ($\tilde{\mu}_{i\omega}^e$) from its expected departure time ($\mu_i^e$). The $\mu_i^e$ is determined in the baseline plan and is informed to the liner who operates the vessel; and the voyage following the visit of this port is also arranged according to the $\mu_i^e$. If there is a delay for the departure from the port, the remainder voyage of the vessel will be influenced. Suppose $c_i^D$ is the unit penalty cost caused by the delay of vessel $i$'s actual departure time from its planned departure time determined in the baseline plan. The above type of cost is calculated as: $\sum_{i \in V} c_i^D (\tilde{\mu}_{i\omega}^e - \mu_i^e)^+$.

(*ii*) The cost related to yard operations for loading and unloading containers. Suppose $d^{LE}$ is the unit cost for handling a container loaded from subblocks that have been reserved to vessels in the exclusive mode; and $d^{LS}$ is the unit cost for handling a container loaded from subblocks that belong to the sharing mode. Here $d^{LE}$ is usually lower than $d^{LS}$ because the plan of the supporting resources in subblocks of sharing mode is usually made in temporary manner while the plan for the exclusive mode is usually on a regular basis; in addition, the exclusively reserved subblocks are closer to their vessel than the subblocks of sharing mode, the trucks' travel distance for the former mode is also shorter than the distance for the latter mode. Thus we differentiate the above two types of unit cost $d^{LE}$ and $d^{LS}$. As defined in subsection on notations, $\tilde{l}_{i\omega}$ is actual number of loading containers for vessel $i$ in scenario $\omega$, $z \times \sigma_i$ equals the maximum capacity (in TEUs) of the subblocks that have been exclusively reserved to vessel $i$; then for vessel $i$, $\min\{\tilde{l}_{i\omega}, z\sigma_i\}$ and $(\tilde{l}_{i\omega} - z\sigma_i)^+$ are the number of containers loaded from the subblocks in the exclusive model and the sharing mode, respectively. In addition, $d^U$ is defined as the unit cost for handling a container unloaded to subblocks. Here $d^U$ is not distinguished between the exclusive and the sharing modes because it is difficult to predict the destination of one container unloaded from a vessel is a subblock that is in either exclusive mode or

sharing model for the vessel. For the easy of model formulation, we do not distinguish the $d^U$ between two modes. As defined in subsection on notations, $\tilde{u}_{i\omega}$ is actual number of unloading containers for vessel $i$ in scenario $\omega$. Then this type of cost is calculated as: $\sum_{i\in V}\left[d^{LE}\min\{\tilde{l}_{i\omega}, z\sigma_i\} + d^{LS}(\tilde{l}_{i\omega} - z\sigma_i)^+ + d^U\tilde{u}_{i\omega}\right]$.

(*iii*) The cost related to usage of the subblocks in sharing mode. As the cost related to the reservation of subblocks in the exclusive mode for vessels has been taken account in the first-stage objective. Different from the reservation of subblocks in the exclusive mode which is a long-term decision, the usage of subblocks in sharing mode depends on the actual number of containers loaded onto each vessel, which is uncertain. For vessel $i$, besides $\sigma_i$ subblocks exclusively reserved for it, $\left\lceil\left(\tilde{l}_{i\omega} - z\sigma_i\right)^+/z\right\rceil$ subblocks are used in the sharing mode. Then this type of cost is calculated as: $c^S\left\lceil\left(\tilde{l}_{i\omega} - z\sigma_i\right)^+/z\right\rceil$.

Based on the above three types of costs, the objective of the second-stage model is defined as follows:

$$\Theta(\omega, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}) = \text{Minimize } \sum_{i\in V}\left\{c_i^D(\tilde{\mu}_{i\omega}^e - \mu_i^e)^+ + d^{LE}\min\{\tilde{l}_{i\omega}, z\sigma_i\} + d^{LS}(\tilde{l}_{i\omega} - z\sigma_i)^+ + d^U\tilde{u}_{i\omega} + c^S\left\lceil\left(\tilde{l}_{i\omega} - z\sigma_i\right)^+/z\right\rceil\right\}. \quad (26)$$

### 4.5 Constraints in the second stage of the model

The constraints in the second stage are similar as the ones in the first stage because both of them need to decide the berthing time and QC-profile assignment for vessels. The difference between them mainly lies in that the decisions in the second stage are made according to the vessels' actual arrival time, actual workload for loading and unloading containers. Thus the decision variables in the second stage are scenario-dependent, and the subscripts of the variables contain $\omega$ by comparing with the ones in the first stage. The constraints for the second stage are listed as follows. The explanations on Constraints (27)–(39) are similar as the ones on Constraints (3)–(16), respectively. Constraint (40) states the number of extra used subblocks in sharing mode should not exceed the number of remainder available subblocks that can be used as sharing mode, i.e., the total number of all subblocks in the port minus the number of subblocks that have been exclusively reserved in advance. In our model, Constraint (40) simplifies yard management by not accounting for initial containers in shared subblocks, focusing instead on strategic allocation of space for excess containers. It is noted that the following model presented the decision-making process is under a determined scenario $\omega'$.

$$\sum_{j\in V\cup\{e'\}}\tilde{\beta}_{bej\omega} = \sum_{i\in V\cup\{e\}}\tilde{\beta}_{bie'\omega} = 1 \qquad \forall b \in B, w = \omega' \quad (27)$$

$$\sum_{j\in V\cup\{e'\}}\tilde{\beta}_{bij\omega} = \sum_{j\in V\cup\{e\}}\tilde{\beta}_{bji\omega} = \alpha_{ib} \qquad \forall i \in V, b \in B, w = \omega' \quad (28)$$

$$\sum_{p\in\bar{P}_{i\omega}}\tilde{\gamma}_{ip\omega} = 1 \qquad \forall i \in V, w = \omega' \quad (29)$$

$$\sum_{p\in\tilde{P}_{i\omega}}\sum_{t\in T}\tilde{\eta}_{ipt\omega}=1 \qquad \forall i\in V, w=\omega' \qquad (30)$$

$$\sum_{p\in\tilde{P}_{i\omega}}\sum_{t\in T}t\tilde{\eta}_{ipt\omega}=\tilde{\mu}^{s}_{i\omega} \qquad \forall i\in V, w=\omega' \qquad (31)$$

$$\tilde{\eta}_{ipt\omega}\le\tilde{\gamma}_{ip\omega} \qquad \forall i\in V, p\in\tilde{P}_{i\omega}, t\in T, w=\omega' \qquad (32)$$

$$\tilde{\mu}^{s}_{i\omega}+\sum_{p\in\tilde{P}_{i\omega}}h_{ip}\tilde{\gamma}_{ip\omega}-1=\tilde{\mu}^{e}_{i\omega} \qquad \forall i\in V, w=\omega' \qquad (33)$$

$$\tilde{\mu}^{e}_{i\omega}+1\le\tilde{\mu}^{s}_{j\omega}+M(1-\tilde{\beta}_{bij\omega}) \qquad \forall i,j\in V, i\neq j, b\in B, w=\omega' \qquad (34)$$

$$\tilde{\mu}^{s}_{i\omega}\ge max\{\tilde{a}_{i\omega}, a^{f}_{i}\} \qquad \forall i\in V, w=\omega' \qquad (35)$$

$$\tilde{\rho}^{1}_{t\omega}=\sum_{i\in V}\sum_{p\in\tilde{P}_{i\omega}}\sum^{t}_{m=\max\{1,t-h_{ip}+1\}}\tilde{\eta}_{ipm\omega}q^{1}_{ip(t-m+1)} \qquad \forall t\in T, w=\omega' \qquad (36)$$

$$\tilde{\rho}^{2}_{t\omega}=\sum_{i\in V}\sum_{p\in\tilde{P}_{i\omega}}\sum^{t}_{m=\max\{1,t-h_{ip}+1\}}\tilde{\eta}_{ipm\omega}q^{2}_{ip(t-m+1)} \qquad \forall t\in T, w=\omega' \qquad (37)$$

$$\tilde{\rho}^{1}_{t\omega}+\tilde{\rho}^{2}_{t\omega}\le\min\{f_t, Q\} \qquad \forall t\in T, w=\omega' \qquad (38)$$

$$|\tilde{\rho}^{2}_{t\omega}-\tilde{\rho}^{1}_{t\omega}|\le x^{Gap} \qquad \forall t\in T, w=\omega' \qquad (39)$$

$$\sum_{i\in I}\left[\left(\tilde{l}_{i\omega}-z\sigma_i\right)^{+}/z\right]\le|K|-\sum_{i\in I}\sigma_i. \qquad w=\omega' \qquad (40)$$

The decision variables in the second stage as well as their domains are defined as follows. These variables specify the vessels' actual berthing time, actual departure time, the actual position in the sequence of vessels moored at the same berth, actual usage of QC-profiles in one certain future scenario.

$$\tilde{\beta}_{bij\omega}\in\{0,1\} \qquad \forall i,j\in V, i\neq j, b\in B, w=\omega' \qquad (41)$$

$$\tilde{\mu}^{s}_{i\omega},\tilde{\mu}^{e}_{i\omega}\ge 0 \qquad \forall i\in V, w=\omega' \qquad (42)$$

$$\tilde{\gamma}_{ip\omega}\in\{0,1\} \qquad \forall i\in V, b\in B, w=\omega' \qquad (43)$$

$$\tilde{\eta}_{ipt\omega}\in\{0,1\} \qquad \forall i\in V, p\in\tilde{P}_{i\omega}, t\in T, w=\omega' \qquad (44)$$

$$\tilde{\rho}^{1}_{t\omega},\tilde{\rho}^{2}_{t\omega}\ge 0 \qquad \forall t\in T, w=\omega'. \qquad (45)$$

### 4.6 Linearization of the model

As Constratins (16) and (39) contain the form of absolute value, we linearize them by defining some auxilary variables $\zeta_t$, $\zeta'_t$, $\xi_{t\omega}$ and $\xi'_{t\omega}$. Then the following Constraints (46)–(48) replace original Constraints (16); and Constraints (49)–(51) replace original Constraints (39).

$$\zeta_t-\zeta'_t=\rho^{2}_{t}-\rho^{1}_{t} \qquad \forall t\in T \qquad (46)$$

$$\zeta_t+\zeta'_t\le x^{Gap} \qquad \forall t\in T \qquad (47)$$

$$\zeta_t,\zeta'_t\ge 0 \qquad \forall t\in T \qquad (48)$$

$$\xi_{t\omega}-\xi'_{t\omega}=\tilde{\rho}^{2}_{t\omega}-\tilde{\rho}^{1}_{t\omega} \qquad \forall t\in T, w=\omega' \qquad (49)$$

$$\xi_{t\omega}+\xi'_{t\omega}\le x^{Gap} \qquad \forall t\in T, w=\omega' \qquad (50)$$

$$\xi_{t\omega},\xi'_{t\omega}\ge 0 \qquad \forall t\in T, w=\omega'. \qquad (51)$$

Objective (26) and Constraint (40) in the second stage contains some non-linear forms. Some

auxiliary variables and constraints are added for linearize the objective. First, we define a new auxiliary variable $\tilde{\varrho}_{i\omega}$ to replace the form $\left\lceil \left(\tilde{l}_{i\omega} - z\sigma_i\right)^+ / z \right\rceil$, i.e., the number of extra used subblocks in sharing mode. The following constraints are also newly defined and the Constraint (53) replaces the original Constraint (40).

$$\tilde{\varrho}_{i\omega} - 1 \le \left(\tilde{l}_{i\omega} - z\sigma_i\right)^+ / z \le \tilde{\varrho}_{i\omega} \qquad \forall i \in V, w = \omega' \qquad (52)$$

$$\sum_{i\in I} \tilde{\varrho}_{i\omega} \le |K| - \sum_{i\in I} \sigma_i \qquad w = \omega' \qquad (53)$$

$$\tilde{\varrho}_{i\omega} \ge 0 \qquad \forall i \in V, w = \omega'. \qquad (54)$$

Another auxiliary binary variable $\pi_{i\omega}$ is defined to denote whether or not the subblocks in the sharing mode are needed. The following constraints should also be added.

$$\tilde{l}_{i\omega} - z\sigma_i \le M\pi_{i\omega} \qquad \forall i \in V, w = \omega' \qquad (55)$$

$$\tilde{l}_{i\omega} - z\sigma_i \ge M(\pi_{i\omega} - 1) \qquad \forall i \in V, w = \omega' \qquad (56)$$

$$\pi_{i\omega} \in \{0,1\} \qquad \forall i \in V, w = \omega'. \qquad (57)$$

After adding the above newly defined variables and constraints, Objective (26) turns to:

$$\sum_{i\in V}\left\{c_i^D(\tilde{\mu}_{i\omega}^e - \mu_i^e)^+ + d^{LE}\left[z\sigma_i\pi_{i\omega} + \tilde{l}_{i\omega}(1-\pi_{i\omega})\right] + d^{LS}\left(\tilde{l}_{i\omega}\pi_{i\omega} - z\sigma_i\pi_{i\omega}\right) + d^U\tilde{u}_{i\omega} + c^S\tilde{\varrho}_{i\omega}\right\} \qquad (58)$$

For the forms "$(\cdot)^+$" contained in Objective (1) and Objective (58), some more auxiliary variables and constraints are added to linearize them.

$$\psi_i \ge a_i^e - \mu_i^s \qquad \forall i \in V \qquad (59)$$

$$\chi_i \ge \mu_i^e - b_i^e \qquad \forall i \in V \qquad (60)$$

$$\tilde{\psi}_{i\omega} \ge \tilde{\mu}_{i\omega}^e - \mu_i^e \qquad \forall i \in V, w = \omega' \qquad (61)$$

$$\psi_i, \chi_i, \tilde{\psi}_{i\omega} \ge 0 \qquad \forall i \in V, w = \omega'. \qquad (62)$$

For the product of two variables "$\sigma_i\pi_{i\omega}$" in Objective (58), some more auxiliary variables and constraints are added to linearize it.

$$\tau_{i\omega} \le M\pi_{i\omega} \qquad \forall i \in V, w = \omega' \qquad (63)$$

$$\tau_{i\omega} \le \sigma_i \qquad \forall i \in V, w = \omega' \qquad (64)$$

$$\tau_{i\omega} \ge \sigma_i + M(\pi_{i\omega} - 1) \qquad \forall i \in V, w = \omega' \qquad (65)$$

$$\tau_{i\omega} \ge 0 \qquad \forall i \in V, w = \omega'. \qquad (66)$$

By summarizing the above model formulation and linearization, a linearized model that is equivalent to the original model $\mathcal{M}0$ is formulated as follows. This model can be solved by commercial solvers.

$$[\boldsymbol{\mathcal{M}0}]\ \text{Minimize} \sum_{i\in V}\left\{c_i^p(\psi_i + \chi_i) + c^E\sigma_i + \sum_{\omega\in\Omega} p_\omega\left\{c_i^D\tilde{\psi}_{i\omega} + (d^{LE} - d^{LS})z\tau_{i\omega} + (d^{LS} - d^{LE})\tilde{l}_{i\omega}\pi_{i\omega} + d^{LE}\tilde{l}_{i\omega} + d^U\tilde{u}_{i\omega} + c^S\tilde{\varrho}_{i\omega}\right\}\right\} \qquad (67)$$

subject to: Constraints (2)–(15), (17)–(25), (27)–(38), (41)–(66).

## 5. Solution approach

It is difficult for commercial solvers (e.g., CPLEX) to solve the linearized model $\mathcal{M}0$ even in small-scale instances, and designing exact algorithms to solve it is also difficult because of the complex structure of the model. Thus, a metaheuristic solution approach is designed to solve $\mathcal{M}0$ in large-scale instances in a reasonable time. The main idea of the solution approach is to encode the solution (i.e., a baseline plan and a recovery plan) into the problem with a sequence of vessels and using a metaheuristic to improve the sequence to obtain a near-optimal solution. The following subsections address how to decode a sequence of vessels into a feasible solution and how to improve the sequence using a metaheuristic.

### 5.1 Decoding a sequence of vessels into a feasible solution

The solution for the original two-stage programming model, $\mathcal{M}0$, contains two parts: a baseline plan and a recovery plan, which are usually determined in the first and second stages, respectively. Then, the process of decoding a sequence of vessels into a feasible solution contains three phases: (1) decoding a sequence into a baseline plan, (2) obtaining a recovery plan based on the baseline plan, and (3) refining the obtained baseline and recovery plans.

The first phase decodes a sequence into a baseline plan. The baseline plan is mainly determined using the decision variables in the original first-stage problem, which is difficult to solve using CPLEX directly when the instance scale is large. Thus, a sequential method is proposed to obtain a baseline plan based on the sequence of vessels. The main idea of the sequential method is explained as follows.

Suppose that the sequence of vessels is $seq = \{v_{(1)}, v_{(2)}, \cdots, v_{(i)}, \cdots, v_{(|V|)}\}$, and the step length of the sequential method is set at $s$, which means that the arrangement of $s$ vessels in the baseline plan is determined in one step of the sequential method. Note that the sequential method contains $|V| - s + 1$ steps.

In the first step, we solve model $\mathcal{M}1$ to obtain the values of the decision variables related to vessels $v_{(1)}, v_{(2)}, \cdots, v_{(s)}$. Model $\mathcal{M}1$ is discussed later. When the model is solved, the arrangement for the first vessel, $v_{(1)}$, is determined in the baseline plan, which means that the solved values of the decision variables related to vessel $v_{(1)}$ are the known parameters in $\mathcal{M}1$ solved in the next step.

In the second step, we solve model $\mathcal{M}1$ to obtain the values of the decision variables related to vessels $v_{(2)}, v_{(3)}, \cdots, v_{(s+1)}$. As mentioned above, the related variables of vessel $v_{(1)}$ are set at the values solved in the previous step; they become the parameters in this step. When the model is solved, the arrangement for the second vessel, $v_{(2)}$, is determined in the baseline plan, and the solved values of the variables related to vessels $v_{(2)}$ and $v_{(1)}$ become parameters in $\mathcal{M}1$ solved in the next step.

Similarly, in the $n^{\text{th}}$ step, model $\mathcal{M}1$ is solved to obtain the values of the decision variables related to vessels $v_{(n)}, v_{(n+1)}, \cdots, v_{(s+n-1)}$. The variables related to vessels $v_{(1)}, v_{(2)}, \cdots, v_{(n-1)}$ are set at the values solved in the previous steps; they become the parameters in this step. When the model is solved, the arrangement for the second vessel, $v_{(n)}$, is determined in the baseline plan, and the solved values of the decision variables related to vessels $v_{(1)}, v_{(2)}, \cdots, v_{(n)}$ become the parameters in $\mathcal{M}1$ solved in the next step.

In the last step, i.e., the $(|V|-s+1)^{\text{th}}$ step, $\mathcal{M}1$ is solved to obtain the values of the decision variables related to vessels $v_{(|V|-s+1)}, \cdots, v_{(|V|)}$. When the model is solved, the arrangements for the remaining vessels $v_{(|V|-s+1)}, \cdots, v_{(|V|)}$ are determined in the baseline plan. Next, all of the arrangements for the vessels in the baseline plan are determined.

Model $\mathcal{M}1$ is as follows.

$$[\boldsymbol{\mathcal{M}1}] \quad \text{Minimize } \sum_{i\in V}\{(a_i^e-\mu_i^s)^+ + (\mu_i^e-b_i^e)^+\} \tag{68}$$

subject to Constraints (2)–(16), (19)–(24).

Model $\mathcal{M}1$ differs from the first-stage part of the original model $\mathcal{M}0$ in that it ignores the yard assignment decision, which is determined by solving another model, $\mathcal{M}2$.

$$[\boldsymbol{\mathcal{M}2}] \quad \text{Minimize } \sum_{i\in V} c^E\sigma_i + \sum_{\omega\in\Omega} p_\omega\{c^S\left\lceil\frac{(\tilde{l}_{i\omega}-z\sigma_i)^+}{z}\right\rceil\} \tag{69}$$

subject to Constraints (17)-(18), (25), (40), (52)–(54).

For the baseline plan, the decisions on berth and yard assignment are made separately to expedite the solving process and enlarge the scale of the solvable problem instances.

Based on the baseline plan obtained above, the second phase of the decoding process involves obtaining a recovery plan for a given baseline plan. The recovery plan is a set of solutions, each of which is obtained using CPLEX to solve a second-stage subproblem oriented to one scenario. Because the first-stage variables are fixed, the solution for the related second-stage variables of one scenario could be solved in a relatively short time.

The last phase of the decoding process involves tuning the sequence of vessels iteratively to improve the solution until no further improvement can be achieved. This phase can be regarded as a type of local refinement. The main idea for designing this local refinement is that the vessels in the earlier part of the sequence have a higher probability (priority) of achieving better performance than the vessels in the later part of the sequence. Here, the performance of each vessel is evaluated using the deviation cost ($Cost_i$), which is calculated based on the deviations in each vessel's baseline and recovery plans, i.e., $Cost_i = c_i^p[(a_i^e-\mu_i^s)^+ + (\mu_i^e-b_i^e)^+] + \sum_{\omega\in\Omega} p_\omega\, c_i^D(\tilde{\mu}_{i\omega}^e-\mu_i^e)^+$. Based on this main idea, the detailed procedure is explained as follows.

A feasible solution is obtained according to the first and second phases. Then, we calculate the deviation cost ($Cost_i$) of each vessel according to the solution (the baseline and recovery plans). We sort the vessels again in decreasing order of $Cost_i$, which implies that we move the vessels with greater deviation to the earlier part of the sequence to increase the probability that their deviation cost is reduced in the next iteration. Then, the first and second phases (i.e., decoding a sequence to obtain a baseline plan and a recovery plan) are executed again; when a new solution (i.e., a new baseline plan and a new recovery plan) is obtained, the deviation cost ($Cost_i$) of each vessel is also evaluated again. This procedure is repeated until no further improvement can be achieved.

These three phases constitute the process of decoding a sequence of vessels into a solution. The third phase (i.e., local refinement) can be considered a loop framework. The loop contains the first and second phases; both phases are executed iteratively until the termination condition of the loop is reached.

### 5.2 Meta-heuristic for improving the process of decoding a vessel sequence to a feasible solution

The process elaborated in Section 5.1 can derive a local optimal solution iteratively. To obtain a global optimal solution, the critical shaking neighborhood search (CSNS) metaheuristic is adopted to search a better sequence of vessels in the solution space. The CSNS was created by Lim and Xu (2006) for use in a yard allocation problem; others, such as Zhen et al. (2011), have used the CSNS for berth allocation. Studies (Lim and Xu, 2006; Zhen et al., 2011) have validated the good algorithmic performance of the CSNS in port-related problems. The main idea of the CSNS is to randomly shake the priority of the critical elements in the sequence to help sequences avoid the local optimum. Here, we regard the vessels with a high schedule deviation cost ($Cost_i$) as critical elements in the iteration process. We define $NC$ as the number of exchanged vessels, that is, we select the top $NC$ vessels with the highest $Cost_i$ and then randomly shake the $NC$ vessel to other places in the sequence to obtain a new sequence. We create a tabu list to store each sequence to avoid generating the same sequence. If one or more sequences have been generated in a previous iteration, we ignore it and restart the neighborhood search. The above CSNS metaheuristic terminates if the best solution obtained cannot be further improved. Specifically, we monitor the progression of the solution quality across successive iterations. If the algorithm, fails to discover a superior solution (i.e., it cannot locate a solution with a better evaluation based on our objective function) after more than 10 iterations, we interpret this scenario as having reached a satisfactory approximate solution. This indicates that further iterations are unlikely to yield substantial improvements, signaling an appropriate juncture for termination to avoid unnecessary computational expenditure.

Last, the initialization of the sequence of vessels in the metaheuristic is important for algorithmic

performance. In this study, we first generate an initial sequence of vessels based on the following practical tactics: (1) the most intuitive tactic is to generate the sequence in decreasing order of the unit penalty cost ($c_i^p$) of each vessel, which weighs the importance of the vessels. (2) Another factor that should be considered is the number of containers in each vessel, which reflects the port resources that need to be assigned. Vessels with more containers should be assigned more QCs and longer time slots for loading and unloading. Considering the above two tactics, this study uses tactics (1) and (2) as primary and secondary criteria, respectively, to generate the initial sequence of vessels.

## 6. Computational experiments

Numerical experiments are conducted to validate the effectiveness of the proposed model and the efficiency of the metaheuristic solution approach for solving large-scale instances. All of the experiments are performed on a workstation equipped with two Intel Xeon E5-2643 v4 CPU running at 3.40GHz with 128 GB of memory under Windows 7. All of the algorithms are programmed in C# (VS 2019), and the related sub-models are solved using CPLEX 12.6.1.

### 6.1 Instance generation

The planning horizon considered in the experiments is one week and each time step equals four hours. In this way, there are 42 time steps for one planning horizon (i.e., $H = 42$). The instances are generated with six different scales, and the parameters of six instance groups (ISGs) are listed in Table 2. It should be noted that although the two-story port is much more complex than the traditional ports, this study's largest scale instance group (i.e., ISG 6) for the two-story port contains 100 vessels, 14 berths, 45 QCs, and 360 subblocks, which may be larger than the problem scales in all the literature related to berth allocation, berth and QC allocation, or berth, QC and yard space allocation problems.

**Table 2**: Problem scales of six instance groups

| Scales of instance groups | Num. of vessels ($\|\boldsymbol{V}\|$) | Num. of berths ($\|\boldsymbol{B}\|$) | Num. of QCs ($\|\boldsymbol{Q}\|$) | Num. of subblocks ($\|\boldsymbol{K}\|$) |
|---|---|---|---|---|
| ISG 1 | 15 | 2 | 7 | 60 |
| ISG 2 | 20 | 3 | 10 | 80 |
| ISG 3 | 45 | 6 | 20 | 180 |
| ISG 4 | 60 | 8 | 30 | 240 |
| ISG 5 | 75 | 10 | 35 | 300 |
| ISG 6 | 100 | 14 | 45 | 360 |

According to the usual setting of berth allocation related computational experiment, we assume there are three classes of vessels, i.e., feeder, medium, jumbo; each of which occupies one third of the all the vessels (Meisel and Bierwirth, 2009; Zhen et al., 2011). For each vessel class, the parameters used for generating the QC-profiles are not identical. Table 2 illustrates the detailed parameters on QC-profile generation for the first floor and the second floor of the port, which refers to the parameter settings in related studies (Wang et al., 2018; Zhen et al., 2022). The average expected workload the average actual workload for the two-story port are shown in the last two rows of Table 3. These average values will be used to estimate the berth and QC utilization rates in the instances for validating the experimental setting, which is later shown in Table 4.

**Table 3**: Parameters on QC-profile generation and workloads for three different vessel classes

| | **Vessel classes** | **Feeder** | **Medium** | **Jumbo** |
|---|---|---|---|---|
| | Percentage | 1/3 | 1/3 | 1/3 |
| *QC-profiles of the first or the second floor* | Range of used QCs | [0,1] | [1,2] | [2,3] |
| | Range of handling time (time_step) | [2,4] | [3,5] | [4,6] |
| | Average handling time (time_step) | 3 | 4 | 5 |
| | Range of expected workload (QC×time_step) | [1,3] | [3,7] | [8,10] |
| | Range of actual workload in scenarios (QC×time_step) | [1,4] | [2,8] | [7,11] |
| Average expected workload of two floors (QC×time_step) | | 4 | 10 | 18 |
| Average actual workload of two floors in scenarios (QC×time_step) | | 5 | 10 | 18 |

According to the data in Table 3, the average handling time for all of the vessel could be calculated as $(3+4+5)/3=4$, and the average actual workload for all of the vessels as $(4+10+18)/3=10.7$ and $(5+10+18)/3=11$ in scenarios. Based the QC-profile settings, it can be concluded that the average duration of a vessel in a berth is four time steps, and a vessel uses QC resources for at least 10.7 QC $\times$ time steps, on average. Therefore, the berth utilization rate and the QC utilization rate could be calculated as shown in the Table 4 when all incoming vessels are served. It can be seen that the berth utilization rate and QC utilization rate for all instance group ranges in 63%～78%, which closely reflects reality (Wang et al., 2018).

The yard utilization is also estimated for the six instance groups. Here, the utilization of the yard resources is defined as the ratio of the number of used subblocks to the total number of subblocks. In the numerical experiments, we set the total number of loaded and unloaded containers are range in [240,600], [720,1800] and [1800,2400] for three vessel classes, respectively. When measure the yard utilization, the proportion of the loading and unloading containers are set as half to half. Then the average number of used subblocks of each feeder vessel could be calculates as $[(240\times0.5+600\times0.5)/2]/240=0.875$. The average number of used subblocks of medium and jumbo vessels

are the same as the above calculation procedure. Therefore, the average number of used subblocks for all vessels could be calculated as $(0.875 + 2.5 + 4.375)/3 = 2.58$. Based on above setting, the yard utilization of different ISGs is showed in Table 5. It can be seen that the yard utilization rate for all instance group ranges in 65%～72%, which also closely reflects reality (Feng et al., 2022).

**Table 4**: Berth and QC utilization rates of the six instances group in the experiments

| Instance groups | **Berth utilization** | | | **QC utilization** | | |
|---|---|---|---|---|---|---|
| | Vessel usage ($\|V\| \times 4$) | Port resource ($\|B\| \times H$) | **Utilization Rate** (%) | Vessel usage ($\|V\| \times 10.7$) | Port resource ($\|Q\| \times 30.8$) | **Utilization Rate** (%) |
| ISG 1 | 60 | 84 | 71.43% | 161 | 216 | 74.54% |
| ISG 2 | 80 | 126 | 63.49% | 214 | 308 | 69.48% |
| ISG 3 | 180 | 252 | 71.43% | 482 | 616 | 78.25% |
| ISG 4 | 240 | 336 | 71.43% | 642 | 924 | 69.48% |
| ISG 5 | 300 | 420 | 71.43% | 803 | 1078 | 74.49% |
| ISG 6 | 400 | 588 | 68.03% | 1070 | 1386 | 77.20% |

***Notes***: For the calculation of port resource of QC, the power generated by solar panel could only support QCs to work simultaneously during 8:00-16:00 and otherwise there are 60% QC that can be used for work. Thus, the total available QC resources in a planning horizon is calculated as $(0.6 \times 4 + 1 \times 2) \times 7 \times |Q| = |Q| \times 30.8$.

**Table 5**: Yard utilization rates of the six instances group in the experiments

| Instance groups | **Yard utilization** | | |
|---|---|---|---|
| | Subblock usage ($\|V\| \times 2.58$) | Port resource ($K$) | Utilization Rate (%) |
| ISG 1 | 38.7 | 60 | 64.50% |
| ISG 2 | 51.6 | 80 | 64.50% |
| ISG 3 | 116.1 | 180 | 64.50% |
| ISG 4 | 154.8 | 240 | 64.50% |
| ISG 5 | 193.5 | 300 | 64.50% |
| ISG 6 | 258.0 | 360 | 71.67% |

Some other parameters are set as follows. The coefficient $c_i^P$ for the penalty cost of each vessel in the first stage problem is randomly generated in the range of [2,6], [6,10] and [10,14] for the above three classes, respectively (Meisel and Bierwirth, 2009). And we assume the coefficient $c_i^D$ for the penalty cost of each vessel in the second stage problem is the same as $c_i^P$. The $c^E$ and $c^S$ for the cost of occupying the subblocks in exclusive mode and sharing mode are set to are set to 3 and 5, which refers to the parameters setting in Zhen (2014). The unit cost $d^{LE}$ and $d^{LS}$ for handling a container loaded from subblocks in the exclusive mode and sharing mode are calculated as $c^T \cdot \bar{d}^E$ and $c^T \cdot \bar{d}^S$, respectively. Here, $\bar{d}^E$ and $\bar{d}^S$ indicate the average distance between berth-side and the subblocks in exclusive mode and sharing mode, respectively. $c^T$ is the unit cost of transportation, which is set to $5 \times 10^{-6}$ to ensure the berth-side cost and yard-side cost are of the same order of magnitude in the

objective function during our numerical experiments. For the feasible time windows $[a_i^f, b_i^f]$, they are randomly distributed along the planning horizon; and the length of each time window is about the four times of its average handle time. For generating the expected time window $[a_i^e, b_i^e]$, it is randomly generated in a range of $[a_i^f, b_i^f]$ and the length of its time window is set to the same as long as its average handle time. In assessing the real-time arrival of vessels across various scenarios, the actual arrival times, denoted as $\tilde{a}_{i\omega}$, are stochastically generated within the interval $[a_i^f - r_\omega, a_i^f + r_\omega]$, where $r_\omega$ fluctuates between 1 and 3, contingent upon the specific scenario (Zhen et al., 2022). Regarding $y_i$, that is, the minimal quota of reserved sub-blocks in an exclusive modality, the intervals are delineated as [1,3], [2,4], and [3,5] for the tripartite vessel classifications, correspondingly. Pertaining to $\tilde{l}_{i\omega}$ and $\tilde{u}_{i\omega}$, indicative of the concrete quantities of containers slated for loading and unloading, these figures are predicated upon the quintessential operational efficiency of a quay crane, benchmarked at 30 containers hourly, in conjunction with the workload metrics explicated in Table 3. Subsequently, we postulate a random variable $\epsilon$ within the range [0.4,0.6] to represent the fraction of containers designated for loading, with the residual $1 - \epsilon$ earmarked for unloading operations amongst the total consignment of containers requiring handling. This methodology underpins the derivation of input parameters $\tilde{l}_{i\omega}$ and $\tilde{u}_{i\omega}$ (Wang et al., 2018). For $f_t$, i.e., the maximum workload that can be supported by the ports' solar panel at time step $t$, we assume that the solar panel equipped in the two-story port could generate sufficient power providing the energy demand of all QCs during 8:00 to 16:00 and store extra energy for other time in one day, which has been already implemented in realistic ports (e.g., Tianjin Port and Qingdao Port) with the solar panel related applications. Thus, we assume a 60% proportion of QCs that could for work during 0:00 to 8:00 and 16:00 to 24:00 in one day due to the lack of support of the solar power. It should be noted that we later investigate the influence of $f_t$ on results through some sensitivity analysis.

### 6.2 Solution quality

The first series of experiments are conducted to validate the quality of solutions solved by our proposed meta-heuristic solution approach. As the CPLEX can only solve some of the small-scale instances optimally, the comparative experiments between our approach and the CPLEX are performed in small-scale instances for investigating the solution quality of our approach. Table 6 illustrates the comparative results in the aspects of objective value and the solution time. As shown in Table 6, the CPLEX can solve the problem only for the first two small-scale instance groups (i.e., ISG1 and ISG2) with five scenarios during three hours (10,800 seconds). All the instances in the ISG3 with five scenarios

cannot be solved by the CPLEX within three hours, while our proposed meta-heuristic can solve them within one hour. Besides, the gap of the solved solutions' objectives between the CPLEX and the proposed algorithm is around 0.02% under the small-scale instances, which validate the effectiveness of the proposed algorithm.

We also propose a lower bound for the original problem $\mathcal{M}0$ to evaluate the performance of our proposed algorithm in large-scale instances, for which the CPLEX cannot obtain the optimal results but a lower bound can be provided as a benchmark. The model for obtaining the lower bound is: Min $\sum_{\omega\in\Omega} p_\omega \{c^E\sigma_i + c_i^p(\psi_{i\omega}^+ + \psi_{i\omega}^-) + (d^{LE} - d^{LS})z\tau_{i\omega} + (d^{LS} - d^{LE})\tilde{l}_{i\omega}\pi_{i\omega} + d^{LE}\tilde{l}_{i\omega} + d^U\tilde{u}_{i\omega} + c^S\tilde{\varrho}_{i\omega}\}$ , subject to: Constraints (2)–(4), (11)–(12), (17)–(21), (25), (52)–(57), (59)–(60), and $\mu_i^s + \min_{p\in P_i} h_{ip} - 1 = \mu_i^e$ for $i \in V$. By comparing with the original model, the above lower bound model relaxes the QC assignment decision and decouples the first and the second stages' decisions. For relaxing the QC assignment, we assume all vessels' dwell time are their shortest handling time according to their QC-profiles. For decoupling the two stages, we adopt the model formulation based on the "perfect information", which is a common concept in the fields related to the stochastic programming and is elaborated later (the second paragraph in Section 6.3). For some large-scale instances, the lower bound model can be solved in a fast way. Table 6 also demonstrates the gap between the lower bound and the objective value of the solution solved by our algorithm; the average gap value is about 1.24%, which implies the lower bound is very tight to the original problem's result and it can act as a good benchmark for large-scale experiments on solution quality.

For some large-scale instances, Table 7 reports the comparative results (i.e., Gap1) between the our proposed solution approach and the lower bound, which is a benchmark and evaluated in the small-scale experiments. The average value of Gap1 is 2.34%. Recall that the Gap1 in small-scale experiments is about 1.24% and the optimality gap the solved results is 0.02%; it implies the lower bound's optimality gap is approximately 1.22%. Then the results in the large-scale experiments, which deviate from the lower bound by 2.34% on average, could demonstrate the actual optimality gap of the results in the large-scale experiments should be quite low. The solution quality in the large-sale experiments could be validated according to the above results.

In realistic port operations management environment, port operators may use some straightforward decision rule to arrange the port resources rather than using an MILP model. The first-come-first-service (FCFS) rule is a widely used decision rule in berth allocation related port management (Zhen et al., 2011; Wang et al., 2018). For some large-scale instances, Table 7 reports the comparative results (i.e., Gap2) between the our proposed solution approach and the FCFS rule, which represents a decision way

in reality. The average value of Gap2 is 25.63%, which implies our proposed methodology (i.e., the decision model as well as the meta-heuristic solution approach) can help the port operator to save total operational cost by about 25% when comparing with the FCFS rule used in reality.

Last but not the least, Table 7 also demonstrates that the proposed meta-heuristic can solve the largest scale instance group, i.e., ISG6, which contains 100 vessels, 14 berths, 45 QCs, and 360 subblocks. This problem scale is larger than the existing literature related to berth, QC and yard space allocation problems in traditional ports. Because the two-story port is more complex than the traditional ports, the complexity and challenging degree of the ISG6 in this study is much higher than not only the literature in academia, but also the problem scale in realistic ports around the world. All of the above validates the solution quality and the applicability for handling the realistic port operations context.

**Table 6**: Comparison with the CPLEX's optimal results in small-scale instances

| Instance ID | $LB_{model}$ | CPLEX | | Meta-heuristic algorithm | | $Gap1$ | $Gap2$ |
|---|---|---|---|---|---|---|---|
| | Obj | Obj | Time (s) | Obj | Time (s) | | |
| ISG1-5-1 | 191.97 | 193.57 | 713 | 193.57 | 36 | 0.83% | 0.00% |
| ISG1-5-2 | 188.54 | 194.95 | 697 | 194.95 | 137 | 3.40% | 0.00% |
| ISG1-5-3 | 191.74 | 194.54 | 838 | 194.54 | 264 | 1.46% | 0.00% |
| ISG1-5-4 | 208.68 | 210.09 | 1107 | 210.09 | 29 | 0.68% | 0.00% |
| ISG1-5-5 | 191.28 | 195.28 | 2640 | 195.28 | 112 | 2.09% | 0.00% |
| ISG2-5-1 | 266.19 | 267.19 | 2441 | 267.19 | 40 | 0.38% | 0.00% |
| ISG2-5-2 | 255.79 | 258.4 | 2265 | 258.40 | 95 | 1.02% | 0.00% |
| ISG2-5-3 | 267.98 | 269.18 | 3074 | 269.18 | 35 | 0.45% | 0.00% |
| ISG2-5-4 | 252.51 | 254.71 | 2005 | 254.72 | 133 | 1.03% | 0.16% |
| ISG2-5-5 | 258.48 | 263.08 | 5493 | 263.08 | 27 | 1.78% | 0.00% |
| ISG3-5-1 | 675.15 | —— | —— | 685.16 | 2404 | 1.48% | —— |
| ISG3-5-2 | 678.47 | —— | —— | 691.27 | 2713 | 1.89% | —— |
| ISG3-5-3 | 688.26 | —— | —— | 693.28 | 2158 | 0.73% | —— |
| ISG3-5-4 | 705.48 | —— | —— | 710.00 | 2020 | 0.64% | —— |
| ISG3-5-5 | 688.79 | —— | —— | 694.40 | 2619 | 0.81% | —— |
| Average | | | | | | 1.24% | 0.02% |

***Notes***: (1) The instance ID "ISG *a*-*b*-*c*" means the problem scale is ISG *a*, the number of scenarios is *b*, and the index of instance is *c*. (2) A dash means the computation time for obtaining a feasible solution exceeds 10,800 second; (3) "$Gap1$" represents the gap between the objective values of solutions obtained by lower-bound model and the meta-heuristic algorithm; (4) "$Gap2$" represents the optimality gap between the solutions obtained by CPLEX solver and the meta-heuristic algorithm.

**Table 7**: Performance evaluation of the metaheuristic algorithm in large-scale instances

| Instance ID | $LB_{model}$ Obj | Meta-heuristic algorithm Obj | Meta-heuristic algorithm Time (s) | FCFS Obj | $Gap1$ | $Gap2$ |
|---|---|---|---|---|---|---|
| ISG4-5-1 | 896.09 | 909.68 | 3795 | 1185.28 | 1.52% | 30.30% |
| ISG4-5-2 | 895.91 | 905.34 | 3669 | 1143.34 | 1.05% | 26.29% |
| ISG4-5-3 | 885.54 | 907.76 | 3047 | 1243.36 | 2.51% | 36.97% |
| ISG4-5-4 | 916.80 | 936.02 | 4833 | 1138.82 | 2.10% | 21.67% |
| ISG4-5-5 | 927.54 | 936.96 | 4440 | 1113.96 | 1.02% | 18.89% |
| ISG5-5-1 | 1113.16 | 1154.62 | 5193 | 1375.02 | 3.72% | 19.09% |
| ISG5-5-2 | 1085.43 | 1122.65 | 5799 | 1459.45 | 3.43% | 30.00% |
| ISG5-5-3 | 1158.04 | 1213.45 | 4661 | 1374.65 | 4.78% | 13.28% |
| ISG5-5-4 | 1107.65 | 1123.47 | 3468 | 1420.47 | 1.43% | 26.44% |
| ISG5-5-5 | 1069.18 | 1086.39 | 5324 | 1449.19 | 1.61% | 33.40% |
| ISG6-5-1 | 1390.55 | 1419.56 | 11684 | 1813.76 | 2.09% | 27.77% |
| ISG6-5-2 | 1453.18 | 1483.81 | 14347 | 1934.61 | 2.11% | 30.38% |
| ISG6-5-3 | 1453.50 | 1476.34 | 12164 | 1730.94 | 1.57% | 17.25% |
| ISG6-5-4 | 1408.83 | 1456.50 | 14676 | 1813.90 | 3.38% | 24.54% |
| ISG6-5-5 | 1424.04 | 1462.68 | 12054 | 1929.08 | 2.71% | 31.89% |
| Average | | | | | 2.34% | 25.63% |

***Notes***: (1) "$Gap1$" represents the objective value's gap between the solutions obtained by lower-bound model and the meta-heuristic algorithm; (2) "$Gap2$" represents the objective value's gap between the solutions obtained by FCFS decision rule and the meta-heuristic algorithm.

### 6.3 Benefit of stochastic programming

This study adopts the stochastic programming methodology to formulate a two-stage model. Experiments are conducted to investigate the benefit of the stochastic programming. More specifically, the value of stochastic programming and the value of perfect information are calculated on the basis of the three methods. The first method is to use the proposed algorithm to solve the model $\mathcal{M}0$; and the solved solution's objective value is denoted by $Z_0$. The second method is to construct a plan exclusively around expected data, sidestepping the stochastic elements. Post formulation, this plan is applied in actual scenarios, undergoing evaluation via the identical objective function inherent in model $\mathcal{M}0$. The resultant objective value is cataloged as $Z_1$. In consideration of the potential for Method 2 to yield plans that may be infeasible in practical scenarios, we have implemented specific measures as detailed in Appendix D. Theoretically, this method mirrors a deconstructed version of our proposed model $\mathcal{M}0$, treating the stages as isolated events rather than intertwined components. A salient point emerges here: the solution borne from the second method qualifies as a feasible solution for the original model $\mathcal{M}0$, positioning $Z_0$ as equal to or less than $Z_1$. The gap between them, i.e., $Z_1 - Z_0$, reflects the value of stochastic programming. More specifically, this experiment uses the percentage value $(Z_1 - Z_0)/Z_0$

to reflect the above mentioned value of stochastic programming; the values are listed in the column "$Val_{Stochas}$" of Table 8. These values offer a clear, comparative insight into the stochastic programming's tangible impact on operational planning and decision-making.

Besides the value of stochastic programming, the value of perfect information can also be calculated through the third method, which is to craft tailored plans for each discrete scenario. This method entails making the most judicious strategic decisions within each individual scenario, assuming complete foreknowledge of the outcomes. The expected value of the objectives for all the scenarios' plans is denoted by $Z_2$. While $Z_2$ is derived from an idealized context where perfect information allows for optimal decisions in every scenario, we recognize that such a situation is impossible in real-world circumstances. Consequently, $Z_2$ serves as a theoretical lower bound rather than a feasible solution for our original model $\mathcal{M}0$. And the gap between them, i.e., $Z_0 - Z_2$, reflects the value of perfect information. This experiment uses the percentage value $(Z_0 - Z_2)/Z_0$ to reflect the above mentioned value of perfect information; the values are listed in the column "$Val_{Info}$" of Table 8.

According to the columns of "$Val_{Stochas}$" and "$Val_{Info}$" in Table 8, we can see that the values of stochastic programming and the perfect information increase along with the randomness increasing, which is reflected by the growing trends of probability $p_e$ and the number of varying vessels $V_{var}$. It implies that when then randomness in the operational environment is severe, the port operator should adopt the stochastic programming methodology to make a robust baseline plan or spend more effort on predicting the perfect information in the future.

**Table 8**: Results of numerical investigation with varied arrival time

| Instance | $S$ | $p_e$ | $V_{var}$ | $M_{var}$ | $Z_0$ | $Z_1$ | $Z_2$ | $Val_{Stochas}$ | $Val_{Info}$ |
|---|---|---|---|---|---|---|---|---|---|
| ISG1 | 5 | 0.10 | 1 | 2 | 247.11 | 252.68 | 219.54 | 2.25% | 11.16% |
| | 5 | 0.12 | 1 | 2 | 248.42 | 254.2 | 220.08 | 2.33% | 11.41% |
| | 5 | 0.15 | 1 | 2 | 250.33 | 256.57 | 220.92 | 2.49% | 11.75% |
| | 5 | 0.10 | 2 | 2 | 253.51 | 259.88 | 222.32 | 2.51% | 12.30% |
| | 5 | 0.12 | 2 | 2 | 256.10 | 262.84 | 223.44 | 2.63% | 12.75% |
| | 5 | 0.15 | 2 | 2 | 259.93 | 267.37 | 225.12 | 2.86% | 13.39% |
| | 5 | 0.10 | 3 | 2 | 258.21 | 268.67 | 225.12 | 4.05% | 12.82% |
| | 5 | 0.12 | 3 | 2 | 261.74 | 272.20 | 226.81 | 4.00% | 13.35% |
| | 5 | 0.15 | 3 | 2 | 264.73 | 279.07 | 229.32 | 5.42% | 13.38% |

***Notes*:** (1) "$S$" represents the number of scenarios with vessels' arrival times deviating from estimated ones, where the estimated arrival times are equal to the expected earliest service times; the scenario $s_0$ with all vessels arriving punctually is not included. (2) "$V_{var}$" represents the number of varied vessels (i.e., vessels with arrival times deviating from estimated ones) in each scenario, for example, $V_{var} = 1$ means that one vessel's actual arrival time deviates from its scheduled time. (3) "$p_e$" is the probability of each scenario with varied vessels, so $S \cdot p_e$<1, and $p_0 = 1 - S \cdot p_e$. $p_0$ is the probability of above mentioned the scenario $s_0$. (4) "$M_{var}$" is the average deviation of the actual arrival times from the estimated ones in all scenarios, for example, $M_{var} = 2$ means that this deviation could be plus or minus 2 time steps. Specifically, if a vessel is expected at $T = 5$, it could arrive either 2 time steps earlier (at T=3) or 2 time steps later (at T=7) in various scenarios. (5)$Val_{Stocha} = (Z_1 - Z_0)/Z_0$, $Val_{Info} = (Z_0 - Z_2)/Z_0$. $Z_1$ and $Z_1$ are calculated by the CPLEX solver.

It is noted that the experiments in Table 8 focus on the random arrival time. Then two more series of experiments are conducted on the context of random workload (number of handling containers), random arrival time and workload; the results of the two series of experiments are shown in Table 9 and Table 10, respectively. The above results show that the influence of the random arrival time is more significant than the influence of the random workload. The reason may lie in the subblocks of sharing mode, which could alleviate the negative influence of the random number of loading/unloading containers. The random arrival time of vessels could affect the following vessels' turnaround intervals, and further incur the penalty cost of deviation from their expected turnaround intervals.

**Table 9**: Results of numerical investigation with varied workload (number of handling containers)

| Instance | $S$ | $p_e$ | $V_{var}$ | $M_{var}$ | $Z_0$ | $Z_1$ | $Z_2$ | $Val_{stochas}$ | $Val_{Info}$ |
|---|---|---|---|---|---|---|---|---|---|
| | 5 | 0.10 | 1 | 2 | 240.91 | 245.08 | 216.92 | 1.73% | 9.96% |
| | 5 | 0.12 | 1 | 2 | 240.99 | 245.11 | 216.97 | 1.71% | 9.97% |
| | 5 | 0.15 | 1 | 2 | 241.03 | 245.19 | 217.03 | 1.73% | 9.96% |
| | 5 | 0.10 | 2 | 2 | 241.62 | 245.75 | 217.43 | 1.71% | 10.01% |
| ISG1 | 5 | 0.12 | 2 | 2 | 241.80 | 245.97 | 217.57 | 1.72% | 10.02% |
| | 5 | 0.15 | 2 | 2 | 242.09 | 246.25 | 217.78 | 1.72% | 10.04% |
| | 5 | 0.10 | 3 | 2 | 243.34 | 247.50 | 218.53 | 1.71% | 10.20% |
| | 5 | 0.12 | 3 | 2 | 243.87 | 248.03 | 218.89 | 1.71% | 10.24% |
| | 5 | 0.15 | 3 | 2 | 244.66 | 248.83 | 219.43 | 1.70% | 10.31% |

***Notes***: (1) "$S$" represents the number of scenarios with vessels' handling containers deviating from estimated ones, which need to assign the extra workload of QCs to finish the unloading or loading tasks. (2) "$V_{var}$" represents the number of varied vessels (i.e., vessels with the number of handling containers deviating from estimated ones) in each scenario, for example, $V_{var} = 1$ means that one vessel's actual workload deviates from its scheduled workload. (3) "$M_{var}$" is the average deviation of the actual assigned workload to the estimated ones in all scenarios, for example, $M_{var} = 2$ means that that this deviation could be plus or minus 2 QC time steps resulting in a fluctuation of about [240-480] TEUs in different scenarios.

**Table 10**: Results of numerical investigation with varied arrival time and workload

| Instance | $S$ | $p_e$ | $V_{var}^{ar}$ | $V_{var}^{w}$ | $M_{var}^{ar}$ | $M_{var}^{w}$ | $Z_0$ | $Z_1$ | $Z_2$ | $Val_{stochc}$ | $Val_{Info}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 0.10 | 1 | 1 | 2 | 2 | 247.31 | 252.88 | 219.73 | 2.25% | 11.15% |
| | 5 | 0.12 | 1 | 1 | 2 | 2 | 248.67 | 254.47 | 220.33 | 2.33% | 11.40% |
| | 5 | 0.15 | 1 | 1 | 2 | 2 | 250.63 | 256.89 | 221.23 | 2.50% | 11.73% |
| | 5 | 0.10 | 2 | 2 | 2 | 2 | 254.42 | 260.78 | 223.03 | 2.50% | 12.34% |
| ISG1 | 5 | 0.12 | 2 | 2 | 2 | 2 | 257.16 | 263.97 | 224.29 | 2.65% | 12.78% |
| | 5 | 0.15 | 2 | 2 | 2 | 2 | 261.29 | 268.75 | 226.18 | 2.86% | 13.44% |
| | 5 | 0.10 | 3 | 3 | 2 | 2 | 260.84 | 270.30 | 226.93 | 3.63% | 13.00% |
| | 5 | 0.12 | 3 | 3 | 2 | 2 | 264.87 | 275.39 | 228.97 | 3.97% | 13.55% |
| | 5 | 0.15 | 3 | 3 | 2 | 2 | 268.66 | 283.03 | 232.03 | 5.35% | 13.63% |

***Notes***: (1) "$S$" represents the number of scenarios with vessels' handling containers and arrival time deviating from estimated ones. (2) "$V_{var}^{ar}$" represents the number of vessels with arrival times deviating from estimated ones in each scenario. (3) "$V_{var}^{w}$" represents the number of vessels with the number of handling containers deviating from estimated ones in each scenario. (4) "$M_{var}^{ar}$" is the average deviation of the actual arrival times from the estimated ones in all scenarios. (5) "$M_{var}^{w}$" is the average deviation of the actual assigned workload to the estimated ones in all scenarios.

### 6.4 Sensitivity experiments

In our study, the two-story container port design, envisaged as a future-oriented green port initiative, significantly diverges in spatial layout and operational equipment from traditional terminals. Firstly, this terminal integrates solar panels on its rooftop to generate power, supporting the operational equipment. However, the fluctuating power generation throughout the day underlines the impact of energy supply curves on equipment load distribution, assuming a fixed energy capacity. Secondly, the custom-designed triple hoist QCs, unique to the dual-level terminal, bear crucial influence on the overall operational costs, warranting analysis for optimal crane operational power. Lastly, within the high-density dual-level port environment, the costs associated with exclusive mode subblocks and sharing mode subblocks significantly influence the arrangement of yard quantities, thereby impacting the terminal's overall operational costs.

The subsequent discussion revolves around these three pivotal features of a dual-level terminal, delving into the relevant sensitivity analysis experiments. We aim to provide valuable managerial insights for operators of such terminals, assisting them in effectively optimizing their terminal operations in the future.

*(1) Influence of the port's solar power capacity*

One of the main features of the new port design is that it is driven by solar energy. In the model, the parameter $f_t$ denotes the maximum workload that can be supported by the port's solar panels at time step $t$. Figure 5 illustrates different curves of the port's solar power capacity. Here, some experiments are conducted to investigate the influence of solar power capacity on operating costs.

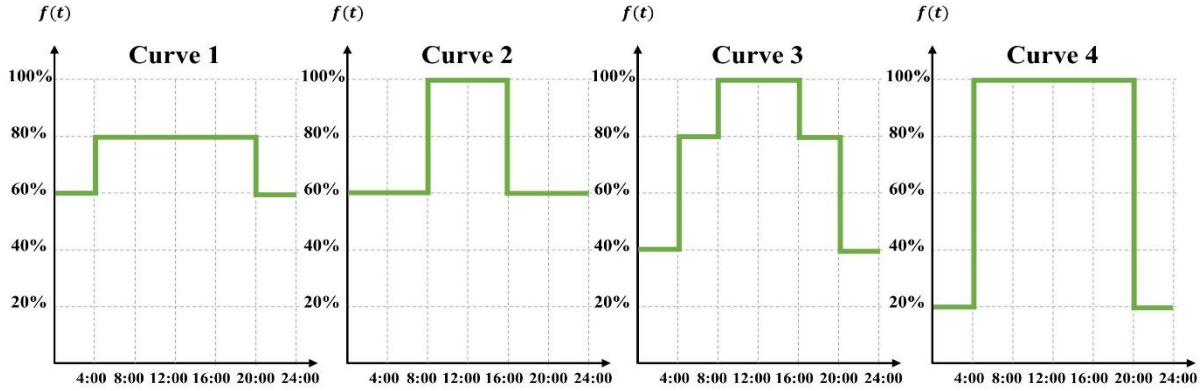


**Figure 5**: Four types of solar power capacity curves

In our analysis, as represented in Figure 5, the vertical axis indicates the percentage of solar power generated relative to the maximum power capacity, demonstrating the variability of power output over time (depicted on the x-axis). In this analysis, we indeed assumed that that once solar energy is stored, it can be supplied across different time periods. This assumption is reflected in the consistent energy

supply area under all four curves in Figure 5. The critical differential between these curves lies in the extremity of their peaks and troughs. Curve 1 exhibits the most modest variance, characterized by its minimal deviations, while Curve 4 presents the most substantial fluctuation in energy output.

Our subsequent findings, illustrated in Figure 6, indicate a direct correlation between these variances in energy supply and operational costs. Specifically, we observed that a more uniform solar power output—evidenced by the lesser disparity between its highest and lowest points—results in decreased operational expenditures. This outcome underscores a significant operational insight for port logistics: there is a tangible economic advantage in utilizing solar power technologies capable of delivering a steadier stream of energy throughout the day. By minimizing the peaks and troughs in the power generation curve, ports can achieve a more efficient energy utilization rate, consequently lowering the overall operating costs.

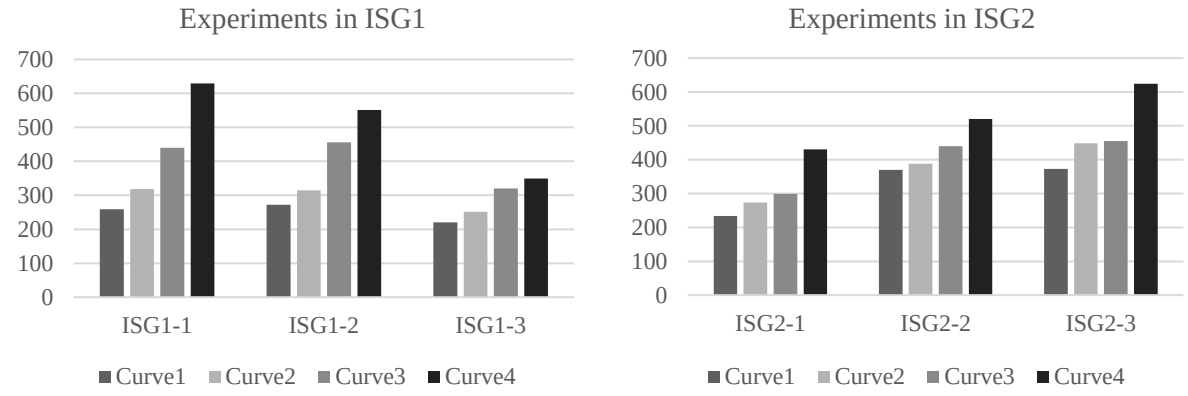


**Figure 6**: Sensitivity analysis of the port's solar power capacity

*(2) Influence of QC handling efficiency*

The handling efficiency of QC, represented by 'C' in our study, is integral to port operations, directly affecting overall performance and operating costs. While a more substantial handling rate (measured in TEU/hour) potentially lowers these costs, reflected in the model's objective value, it also necessitates more substantial investments in QC design and potentially leads to increased energy consumption and emissions. Our research indicates a diminishing return on investment and environmental impact when 'C' exceeds a threshold. Specifically, upon reaching a QC handling efficiency of 35 TEU/hour, further enhancements do not proportionately contribute to operational performance. This plateau suggests that beyond this point, the QC's heightened efficiency cannot significantly offset impacts arising from uncertainties such as vessel delays or miscalculations in container volume.

Consequently, based on comprehensive experiments conducted on ISG1 and ISG2, as visualized in Figure 6, we advocate a QC handling rate setting of 30 TEU/hour as the optimal balance between performance enhancement and cost (including environmental considerations). The vertical axis in

Figure 6 represents the model's objective value, underscoring the balance achieved at this proposed QC efficiency level.

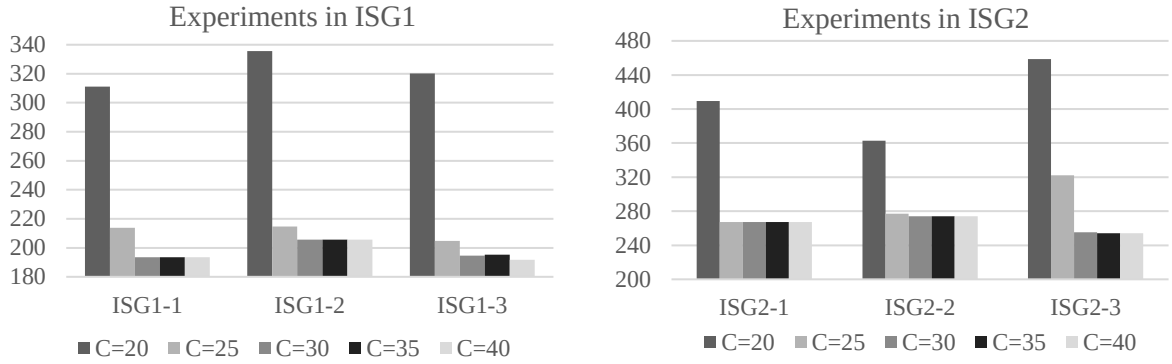


**Figure 7**: Sensitivity analysis of the handling efficiency of QCs

*(3) Influence of unit costs for occupying a subblock in different modes*

In the proposed model, two types of unit costs are defined, i.e., $c^E$ and $c^S$, which denote the unit cost of occupying a subblock in a planning horizon in the exclusive mode and the sharing mode, respectively. These costs, therefore, influence the decision-making process of shipping liners regarding how they allocate their containers - either within exclusive or shared spaces. The distinction in pricing reflects the premium nature of exclusive spaces that guarantee availability versus shared spaces that are subject to availability and potential delays. Sensitivity analyses are also conducted to investigate the influence of $c^E$ and $c^S$ on the final performance, i.e., the objective value of the model. The higher the two types of unit costs, the higher the objective value; this result is intuitive. To obtain more non-intuitive insights, sensitivity analyses should be conducted keeping a certain metric unchanged (when $d^{LE}$ and $d^{LS}$ vary). Here, this metric is calibrated as $[N^E(Y \cdot z \cdot d^{LE} + c^E) + N^S(c^S + Y \cdot z \cdot d^{LS})]/(N^E + N^S)$. In this sensitivity analysis, the value of the metric is a constant to make a fair comparison between settings using different combinations of unit costs $d^{LE}$ and $d^{LS}$. $N^E$ and $N^S$ denote the number of subblocks in the exclusive mode and the sharing mode, respectively. In addition, $d^{LE}$ and $d^{LS}$ (i.e., the unit costs of handling a container loaded from the subblocks in the exclusive mode and the sharing mode, respectively) are considered. Furthermore, $Y = 64.50\%$, i.e., yard utilization, according to Table 5, and $z = 240$ TEU, i.e., the capacity of a subblock in terms of TEUs, i.e., 5 tiers $\times$ 6 lanes $\times$ 8 slots.

By keeping this metric unchanged, we generate five combinations of $c^E$ and $c^S$, as shown in the five columns (cases) in Table 11.

**Table 11**: Different settings of unit cost for occupying subblocks in the two modes

| | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 |
|---|---|---|---|---|---|
| $c^E$ | 2.6 | 2.8 | 3 | 3.2 | 3.4 |
| $c^S$ | 6.1 | 5.6 | 5 | 4.4 | 3.8 |

Figure 8 shows the yard cost dynamics under various pricing strategies for exclusive ($c^E$) and shared ($c^S$) subblocks, spanning five cases along the x-axis. The depicted trends reveal a more pronounced positive correlation between yard costs and exclusive subblock pricing ($c^E$), with less significant fluctuations resulting from changes in shared subblock pricing ($c^S$). Moreover, the analysis suggests a strategic advantage in encouraging shipping liners' preference for exclusive yard space. By expanding the cost gap between shared and exclusive yards—specifically, by raising $c^S$ while reducing $c^E$—shipping liners are incentivized to reserve more exclusive subblocks. When exclusive subblocks are more cost-efficient than shared ones, shipping liners are inclined to reserve them preferentially, allowing for more streamlined operations and greater control over their container handling, thereby optimizing operational expenditures and efficiency. This strategy not only lowers uncertainty-related costs but also reduces terminal operators' variable costs associated with shared subblock reservations. Consequently, this approach presents a win-win scenario: while exclusive yards handle regular container flows, the existence of shared subblocks remains crucial for accommodating unexpected surges, safeguarding efficient operations. Therefore, the application of subblock pricing strategies, favoring exclusive subblock reservations, emerges as a pivotal factor in optimizing operational costs and handling uncertainties in container volumes.

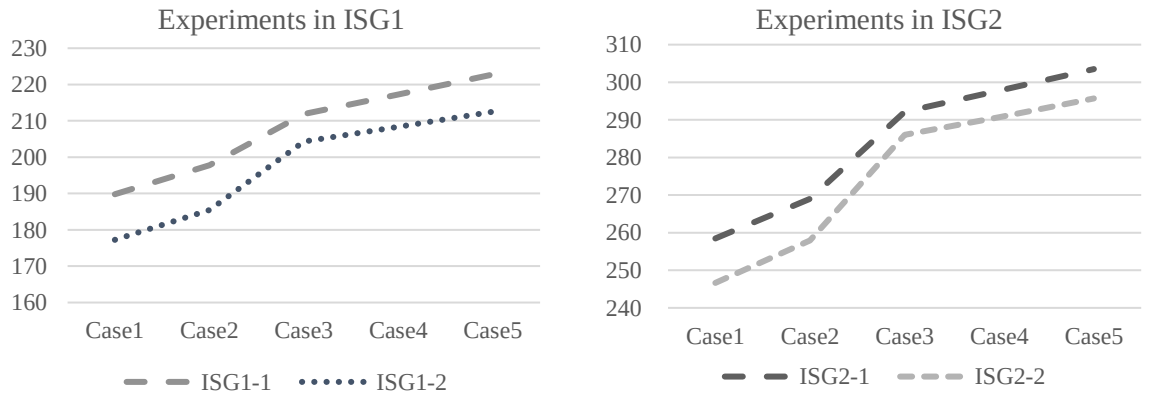


**Figure 8**: Sensitivity analysis of unit costs of occupying a subblock in different modes

### 6.5 Managerial Insights

This study unveils several key managerial insights, offering practical strategic guidance for port operators:

**Port Information Management:** Effective information management is the key. Our analysis in Section 6.3 shows that precise knowledge about ship arrivals and container volumes can lead to up to 13.38% cost savings. Investing in more sophisticated forecasting methods will enable port operators to anticipate container loads and vessel arrival times more accurately, reducing costly deviations from planned schedules.

**QC Scheduling Management:** The dynamic nature of solar power utilization and the unique structural design of two-story ports demand a thoughtful approach to QC deployment, as explored in Section 6.4. While increasing the number of QCs can enhance operational efficiency, our findings suggest a point of diminishing returns. Port authorities are advised to find an optimal balance, avoiding overinvestment in QCs. This involves considering the vertical dimension in QC operations to improve handling times and efficiency, necessitating a reevaluation of conventional QC scheduling strategies.

**Yard Operation Management:** Our sensitivity analysis (Section 6.4) highlights the economic implications of yard block allocation. Adjusting pricing strategies to favor dedicated over shared yard blocks can significantly reduce the uncertainties associated with fluctuating container numbers. This strategy not only minimizes unexpected costs for shipping companies but also benefits port operators by lowering variable expenses, such as maintenance and personnel. However, maintaining a proportion of shared yard space is crucial for accommodating sporadic increases in container volumes.

In conclusion, our study emphasizes that port operators need to adopt more nuanced management strategies in resource allocation, cost control, and uncertainty mitigation to enhance overall operational efficiency and economic benefits at the ports.

## 7. Conclusions

This study introduces a new two-story container port system, which is the future of port system design in terms of vertical expansion and can increase land productivity for storage and operational efficiency. An integrated decision model is formulated to determine the long-term baseline plan for berth allocation and yard space reservation under uncertain vessel arrival times and number of containers to be loaded/unloaded. To take these uncertainties into account, the integrated decision model is formulated as a two-stage stochastic programming model. The proposed model is nonlinear to accommodate the features of the new port design, such as a two-story infrastructure, a new type of QC, solar panels on the port roof, and a high density of yard space for storing containers. This study further linearizes the model as a MILP model, which could be solved using commercial solvers in small-scale instances. To solve the model efficiently in large-scale instances, a metaheuristic is also designed. In addition, this study conducts computational experiments to validate the effectiveness of the proposed model and the

efficiency of the algorithm. Managerial insights are obtained from the results of sensitivity analyses. These insights may be useful for port operators who are interested in adopting the new port design or integrating some of its novel features into their current port systems.

This study has the following limitations. An immediate direction for future research involves integrating yard reservation numbers with berth and crane scheduling to better reflect the interconnected nature of port operations. Designing exact algorithms for the decision models of two-story port systems is also an interesting and challenging research direction. Besides, the factor in tidal changes and traffic flow in the port's navigation channels if the new port design were to be implemented in a tidal port could be further considered in future studies. All of the above provide opportunities for further exploration by future studies.

## Acknowledgments

Professor Loo Hay Lee in National University of Singapore is one of leaders of the team that invents this new port design. The authors dedicate this article to commemorate Prof. Lee who passed away one year ago.

The authors would like to thank the editor-in-chief, the associate editor, and anonymous reviewers for their valuable comments and suggestions, which have greatly improved the quality of this paper.

## References

Abou Kasm, O., & Diabat, A. (2020) Next-generation quay crane scheduling. *Transportation Research Part C: Emerging Technologies* 114: 694–715.

Abou Kasm, O., Diabat, A., Chow, J. (2023) Simultaneous operation of next-generation and traditional quay cranes at container terminals. *European Journal of Operational Research* 308(3): 1110–1125.

Baird, A. J., & Rother, D. (2013) Technical and economic evaluation of the floating container storage and transhipment terminal (FCSTT). *Transportation Research Part C: Emerging Technologies* 30: 178–192.

Bierwirth, C., Meisel, F. (2010) A survey of berth allocation and quay crane scheduling problems in container terminals. *European Journal of Operational Research* 202(3): 615–627.

Cheimanoff, N., Féniès, P., Kitri, M., Tchernev, N. (2023) Exact and metaheuristic approaches to solve the integrated production scheduling, berth allocation and storage yard allocation problem. *Computers & Operations Resear*ch 153: 106174.

Cheimanoff, N., Fontane, F., Kitri, M., Tchernev, N. (2022) Exact and heuristic methods for the

integrated berth allocation and specific time-invariant quay crane assignment problems. *Computers & Operations Research* 141: 105695.

Dai, H., Ma, J., Yang, Y., Sun, J., & Dai, Y. (2023) A bi-layer model for berth allocation problem based on proactive-reactive strategy. *Computers & Industrial Engineering* 179: 109200.

de Koster, R., Le-Duc, T., Yu, Y. (2008) Optimal storage rack design for a 3-dimensional compact AS/RS. *International Journal of Production Research* 46(6): 1495–1514.

Feng, Y., Song, D. P., & Li, D. (2022) Smart stacking for import containers using customer information at automated container terminals. *European Journal of Operational Research* 301(2): 502–522.

Gharehgozli, A., Zaerpour, N., de Koster, R. (2020) Container terminal layout design: transition and future. *Maritime Economics & Logistics* 22: 610–639.

Giallombardo, G., Moccia, L., Salani, M. and Vacca, I. (2010) Modeling and solving the tactical berth allocation problem. *Transportation Research Part B* 44: 232–245.

Gue, K., Kim, B. (2007) Puzzle-based storage systems. *Naval Research Logistics* 54(5): 556–567.

Guo, L., Zheng, J., Liang, J., Wang, S. (2023) Column generation for the multi-port berth allocation problem with port cooperation stability. *Transportation Research Part B* 171: 3–28.

Han, Y., Lee, L., Chew, E., Tan, K. (2008) A yard storage strategy for minimizing traffic congestion in a marine container transshipment hub. *OR Spectrum* 30: 697–720.

He, J., Tan, C. (2019) Modelling a resilient yard template under storage demand fluctuations in a container terminal. *Engineering Optimization* 51(9): 1547–1566.

He, J., Tan, C., Yan, W., Huang, W., Liu, M., Yu, H. (2020) Two-stage stochastic programming model for generating container yard template under uncertainty and traffic congestion. *Advanced Engineering Informatics* 43: 101032.

Hu, H., Lee, B., Huang, Y., Lee, L., Chew, E. (2013) Performance analysis on transfer platforms in frame bridge based automated container terminals. *Mathematical Problems in Engineering* 9: 831–842.

Hu, H., Mo, J., & Zhen, L. (2021) Improved Benders decomposition for stochastic yard template planning in container terminals. *Transportation Research Part C: Emerging Technologies* 132: 103365.

Jin, J., Lee, D., Cao, J. (2016) Storage yard management in maritime container terminals. *Transportation Science* 50(4): 1300–1313.

Lee, L., Chew, E., Tan, K., Han, Y. (2006) An optimization model for storage yard management in transshipment hubs. *OR Spectrum* 28: 539–561.

Lim, A., Xu, Z. (2006) A critical-shaking neighborhood search for the yard allocation problem.

*European Journal of Operational Research* 174: 1247–1259.

Liu, C. (2020) Iterative heuristic for simultaneous allocations of berths, quay cranes, and yards under practical situations. *Transportation Research Part E* 133: 101814.

Liu, M., Lee, C., Zhang, Z., Chu, C. (2016) Bi-objective optimization for the container terminal integrated planning. *Transportation Research Part B* 93: 720–749.

Martin-Iradi, B., Pacino, D., Ropke, S. (2022) The multiport berth allocation problem with speed optimization: exact methods and a cooperative game analysis. *Transportation Science* 56(4): 972–999.

Moorthy, R., Teo, C. (2006) Berth management in container terminal: the template design problem. *OR Spectrum* 28: 495–518.

Öztürkoğlu, Ö., Gue, K., Meller, R. (2012) Optimal unit-load warehouse designs for single command operations. *IIE Transactions* 44 (6): 459–475

Robenek, T., Umang, N., Bierlaire, M., Ropke, S. (2014) A branch-and-price algorithm to solve the integrated berth allocation and yard assignment problem in bulk ports. *European Journal of Operational Research* 235(2): 399–411.

Rodrigues, F., Agra, A. (2022) Berth allocation and quay crane assignment/scheduling problem under uncertainty: A survey. *European Journal of Operational Research* 303(2): 501–524.

Salido, M. A., Rodriguez-Molins, M., & Barber, F. (2011) Integrated intelligent techniques for remarshaling and berthing in maritime terminals. *Advanced Engineering Informatics* 25(3): 435–451.

Stahlbock, R., Voss, S. (2008) Operations research at container terminals: a literature update. *OR Spectrum* 30 (1): 1–52.

Tang, S., Jin, J., Lu, C. (2022) Investigation of berth allocation problem in container ports considering the variety of disruption. *Computers & Industrial Engineering* 172: 108564.

Thanos, E., Toffolo, T., Santos, H., Vancroonenburg, W., Vanden Berghe, G. (2021) The tactical berth allocation problem with time-variant specific quay crane assignments. *Computers & Industrial Engineering* 155: 107168.

Wang, K., Zhen, L., Wang, S., Laporte, G. (2018) Column generation for the integrated berth allocation, quay crane assignment, and yard assignment problem. *Transportation Science* 52(4): 812–834.

Wang, W., Lin, S., & Zhen, L. (2023) Flexible storage yard management in container terminals under uncertainty. *Computers & Industrial Engineering* 186: 109753.

Xie, F., Wu, T., Zhang, C. (2019) A branch-and-price algorithm for the integrated berth allocation and quay crane assignment problem. *Transportation Science* 53(5): 1427–1454.

Yang, L., Ng, T., Lee, L. (2022a) A robust approximation for yard template optimization under uncertainty. *Transportation Research Part B* 160: 21–53.

Yang, X., Hu, H., Jin, J., Luo, N. (2022b) Joint optimization of space allocation and yard crane deployment in container terminal under uncertain demand. *Computers & Industrial Engineering* 172: 108425.

Yu, H., Ning, J., Wang, Y., He, J., Tan, C. (2021) Flexible yard management in container terminals for uncertain retrieving sequence. *Ocean & Coastal Management* 212: 105794.

Zaerpour, N., Gharehgozli, A., de Koster, R. (2019) Vertical expansion: a solution for future container terminals. *Transportation Science* 53(5):1235–1251.

Zaerpour, N., Yu, Y., de Koster, R. (2015) Storing Fresh Produce for Fast Retrieval in an Automated Compact Cross-Dock System. *Production and Operations Management* 24(8): 1266–1284.

Zhen, L. (2014) Container yard template planning under uncertain maritime market. *Transportation Research Part E: Logistics and Transportation Review* 69: 199–217.

Zhen, L. (2015) Tactical berth allocation under uncertainty. *European Journal of Operational Research* 247(3): 928–944.

Zhen, L., Chew, E., Lee, L. (2011) An integrated model for berth template and yard template planning in transshipment hubs. *Transportation Science* 45(4): 483–504.

Zhen, L., Lee, L., Chew, E., Chang, D., Xu, Z. (2012) A comparative study on two types of automated container terminal systems. *IEEE Transactions on Automation Science & Engineering* 9(1): 56–69.

Zhen, L., Zhuge, D. Wang, S., Wang, K. (2022) Integrated berth and yard space allocation under uncertainty. *Transportation Research Part B* 162: 1–27.

Zhou, C., Chew, E., Lee, L. (2017) Information-based allocation strategy for GRID-based transshipment automated container terminal. *Transportation Science* 52(3): 707–721.

Zhou, C., Chew, E., Lee, L., Liu, D. (2016) An introduction and performance evaluation of the GRID system for transshipment terminals. *Simulation* 92(3): 277–293.

Zhu, M., Fan, X., Cheng, H., He, Q. (2010) Modeling and simulation of automated container terminal operation. *Journal of Computers* 5(6): 951–957.

## Appendix A: Generation methodology for QC profiles

A pivotal component of our research pertains to the concept of Quay Crane (QC) profiles, which are instrumental in dictating the operational tempo and efficiency at two-story container ports, the generation methodology of QC-profiles is as follows:

**Determining Handling Time (Step 1)**: Each vessel's operational duration within the port is ascertained by referencing the handling time ranges for QC-profiles stipulated in Table 2. This determination, denoted as $h_{ip}$ within our model, signifies the number of operational time steps allocated per vessel.

**Workload Estimation (Step 2)**: Utilizing $h_{ip}$ values, alongside anticipated load quantities for each vessel, we derive the corresponding workload and QC operational times, ensuring resource adequacy and operational realism.

**Workload Distribution (Step 3)**: Subsequently, the aggregated workload is equitably distributed across each operational time step. This stratagem ensures sustained operational tempo, avoiding QC idleness or, conversely, overburdening.

**QC-profile Segregation (Step 4)**: In the final phase, workloads designated for each time step are bifurcated, though randomly, to service both levels of the two-story structure, guaranteeing workload for each QC stratum. This bifurcation culminates in the formation of distinct upper and lower QC-profiles per vessel.

It is imperative to note that the efficiency of QC-profiles holds direct repercussions for quay crane operational efficacy. Suboptimal profile configurations could precipitate inefficiencies, reflecting areas necessitating future research and optimization endeavors. We acknowledge this aspect as an integral component of our subsequent investigative pursuits, appreciating its profound impact on comprehensive port operational efficiency.

## Appendix B: The construction of two-story port

As depicted in Figure S1, our manuscript highlights the distinctive advantage of double-layered container terminals over traditional single-layered port structures. The key innovation in double-layered docks lies in their vertical space expansion. Unlike conventional ports where containers are stacked in a single horizontal layer, double-layered terminals utilize an additional level above the ground. This design effectively doubles the container storage capacity within the same footprint area.

Therefore, the spatial density of container storage in these double-layered docks is significantly higher compared to traditional single-layer yards. This increased density allows for more efficient use of limited port space, accommodating a larger number of containers without expanding the ground area. This design is particularly beneficial in ports where space is at a premium, enabling them to handle higher volumes of cargo without the need for additional land.

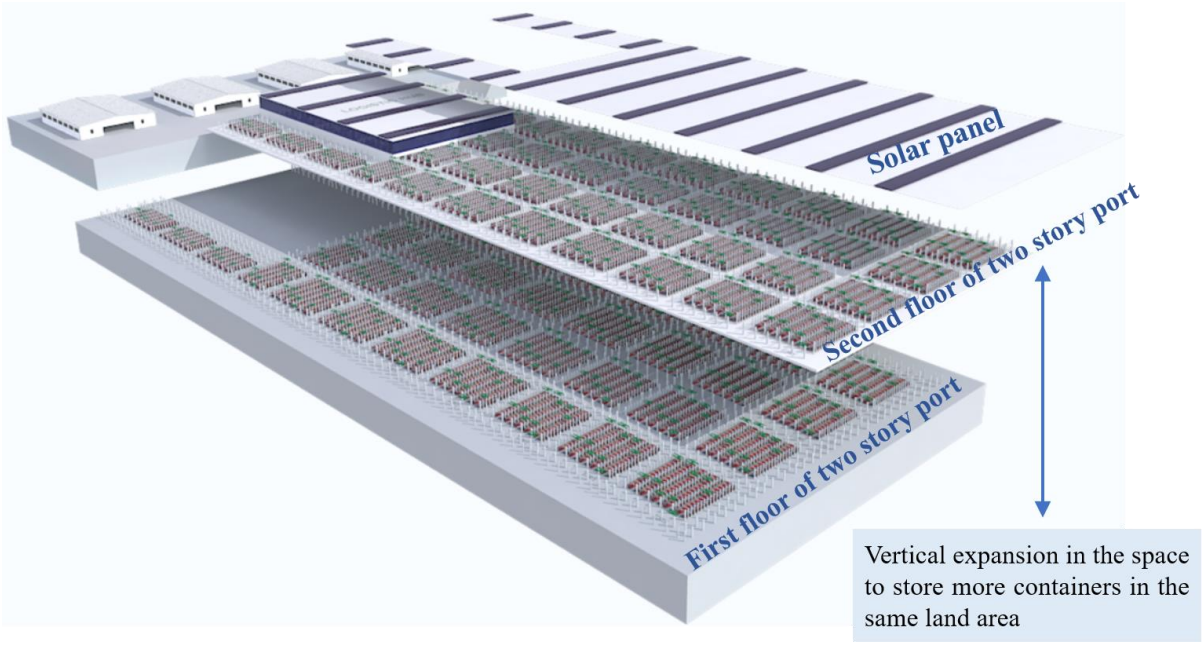


Figure S1: The construction of two-story port

## Appendix C: The concept of "slot" and "lane" in a subblock

In a port, the layout for container stacking within each subblock is illustrated in Figure S2. In this context, a 'lane' refers to the direction parallel to the movement of vehicles, while a 'slot' is perpendicular to the lanes. Therefore, in the subblock shown in Figure S2, one layer can accommodate a total of 6 (lanes) x 8 (slots) = 48 standard 20-foot containers, commonly known as 20-foot equivalent units (TEUs).

Typically, a subblock is designed to stack containers in up to five layers vertically. Therefore, a single subblock has the capacity to hold up to 240 TEUs. This efficient utilization of space, both horizontally and vertically, plays a crucial role in optimizing the storage capacity and operational efficiency of a port.

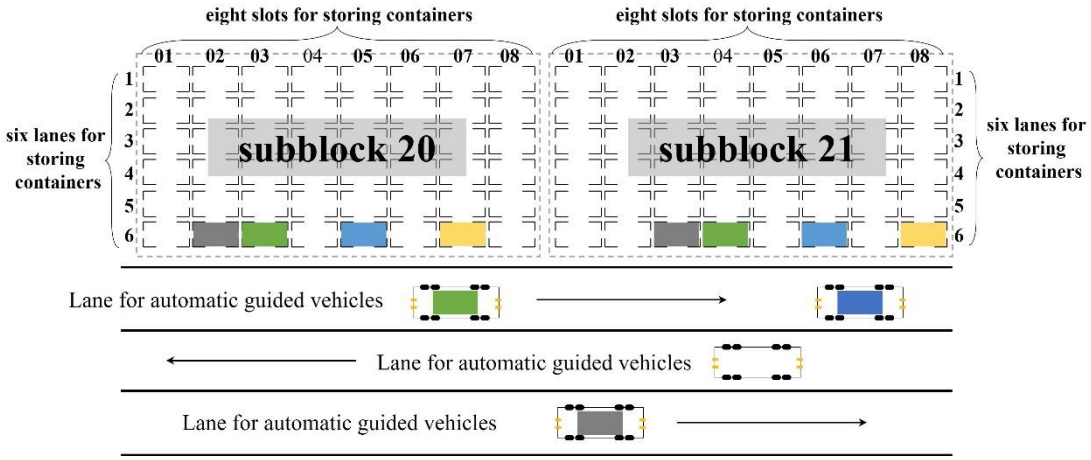


Figure S2: The concept of "slot" and "lane" in a subblock

## Appendix D: Model formulation when applying the plan obtained by method 2 under actual scenarios

The second method in section 6.3, while effective in creating a long-term baseline plan, does not account for uncertainties, which may lead to infeasible scenarios when applying this plan to certain real-world situations. To address this challenge and ensure the plan's performance can be assessed across different scenarios, we propose a modification in the two-stage model. Specifically, we suggest relaxing the constraints (38)–(39) into the objective function. This adjustment entails temporarily disregarding the constraints related to the load limits of triple-trolley quay cranes and the balance between upper and lower layer loads. Such relaxation ensures that each vessel can always complete its container loading and unloading operations within an 'ideal' optimum time window, regardless of unforeseen circumstances. The specific formulation of the model with these adjustments is as follows:

$$[M3]\Theta(\omega, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\alpha}) = \text{Minimize } \sum_{i \in V} \left\{ c_i^D (\tilde{\mu}_{i\omega}^e - \mu_i^e)^+ + d^{LE} \min\{\tilde{l}_{i\omega}, z\sigma_i\} + d^{LS} (\tilde{l}_{i\omega} - z\sigma_i)^+ + d^U \tilde{u}_{i\omega} + c^S \left\lceil \left(\tilde{l}_{i\omega} - z\sigma_i\right)^+ / z \right\rceil \right\} + \varepsilon(\min\{f_t, Q\} - \tilde{\rho}_{t\omega}^1 - \tilde{\rho}_{t\omega}^2)^+ + \varpi(x^{Gap} - |\tilde{\rho}_{t\omega}^2 - \tilde{\rho}_{t\omega}^1|)^+$$

Subject to:

Constraints (27)–(37), (52)–(57), (63)–(66).

In the objective function, the constraints originally defined as (38) and (39) are incorporated as penalized costs. This modification aims to prevent the model from becoming infeasible due to these constraints. $\varepsilon$ and $\varpi$ are introduced as large positive numbers, serving as unit penalty costs for any violations of quay crane (QC) allocation and load distribution constraints in practical scenarios.

The evaluation process under method 2 is as follows:

Initially, we utilize the first-stage model of our original model M1 to generate a baseline plan based on assumed expected data. This baseline plan is subsequently evaluated across different scenarios $w = w'$ to assess the robustness of the baseline plan under varying conditions. Thus, the solution derived from method 2 originates from the evaluation of the baseline plan across multiple deterministic scenarios (i.e., model M3), aiming to calculate the expected value of the objective function under these scenarios. When the baseline plan obtained from M1 is applied to M3 under a specific actual scenario (for example, $w = w''$) and results in no solution, we set the objective function value of M3 to a sufficient large number $M$ to indicate the poor performance of the baseline plan.