Gaussian Process-based Storage Location Assignments with Risk Assessments for Progressive Zone Picking Systems

Abstract

E-commerce warehouses are under constant pressure to adapt their order picking systems and reassign product storage locations to meet fluctuating customer demands. Most existing approaches optimize storage location reassignments based on customer orders and operational configurations to maintain high order picking performance. This paper presents a Gaussian process surrogate model (GPSM) approach to predict the performance metrics for storage location reassignments. The GPSM estimates the expected flow time of orders from the historical data on previous storage location assignments and aids in identifying the new assignments that yield the minimum estimated average flow times. Management can also take advantage of the GPSM's uncertainty quantification capability to assess the probability of improvement for a given storage reassignment and its implementation. The proposed model and assignment policy are validated using discrete-event simulations and industrial data. Experimental results demonstrate that the GPSM can improve expected flow time by 7.51% and reduce unnecessary reassignment operations by 43.25%.

Keywords: Facilities planning and design; Zone order picking; Storage location assignment; Gaussian process approach.

1. Introduction

Zone picking (ZP) systems are popular owing to their high-throughput capability, flexibility to handle small and large order volumes, and adaptability to a wide range of product sizes with a variety of order pickers (Gaast et al., 2020). By separating an order picking area into zones and assigning one picker per zone, ZP systems can reduce travel time and congestion among pickers (De Koster et al., 2007).

In ZP systems, performance bottlenecks can still occur due to workload imbalances or other reasons. Therefore ZP comes with order picking planning policies, such as order batching (Pan, Shih, & Wu, 2015), order sequencing (Huang et al., 2018), and storage assignment (Jane, 2000). Order batching reduces order picking variability and mitigates workload gaps between zones by grouping orders so that each zone's workload is evenly distributed between batches. Order sequencing adjusts tote release sequences to improve the distribution of workload balances. Storage assignment balances workloads over zones by assigning products to each zone during a long picking time. Optimizing product assignments is referred to as the storage location assignment problem (SLAP) (Gu et al., 2007).

This paper investigates the SLAP for a progressive bypass ZP flow-rack system with an automated storage/retrieval (S/R) crane. The flow-rack system is popular in warehouse management systems for e-commerce businesses because it handles small and frequent orders and daily demand fluctuations under short flow times. In many warehouses, since assignment capabilities are limited in terms of resources and time, the SLAP problem is optimized by reassigning selected products instead of reallocating all products in an order. As illustrated in Figure 1, prior to order picking, the S/R crane relocates a few storage locations of products between the zones to rebalance the workloads (Roodbergen & Vis, 2009). We study the optimization of the swap operations considering the current assignment of products and estimates of the expected workload over daily orders.

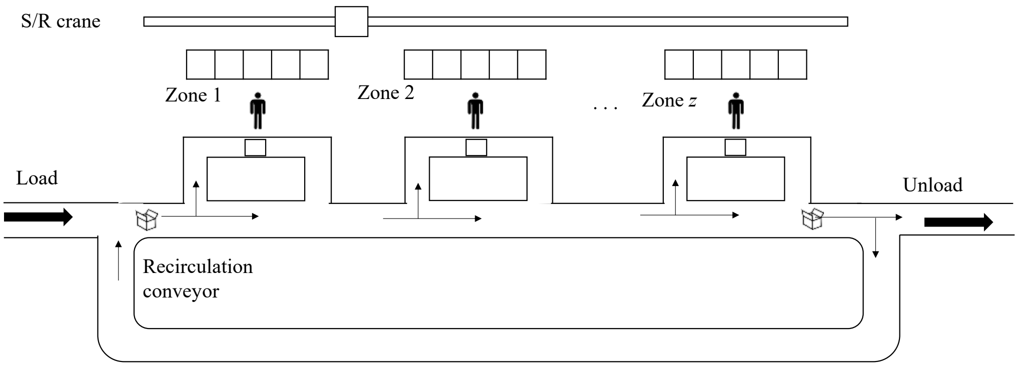

Figure 1. Progressive bypass ZP system with an S/R crane (Kim and Hong (2022)).

Solutions to the SLAP use workload proxies to approximate current workloads and exploit them for a simple rebalancing of the estimated workloads. Popular workload proxies include the number of picks (Jane & Laih, 2005; Jewkes et al., 2004; Kim & Hong, 2022), the expected walk time from a picker's loading depot (Jewkes et al., 2004), picker's skill level (Bartholdi & Eisenstein, 1996), and picker's route distance (De Koster et al., 2012). However, it is still unclear whether the proxies accurately represent workloads, and it is questionable whether using the proxies is effective in optimizing an order picking performance metric such as the average flow time of orders. In this paper, we propose to use a data driven approach to directly estimate the order picking performance metric as a function of current workloads, and we exploit the data-driven estimation to solve SLAP.

To our knowledge, only a few studies have used a data-driven approach to the workload balancing and storage location reassignment problems. Building an analytical model to estimate order picking performance from order picking data has been challenging due to the noise-included uncertainty of complex ZP systems. We develop a Gaussian process surrogate model (GPSM) for a data-driven estimation of the average flow time of orders from the historical data of previous storage location assignments. Surrogate modeling usually uses the Gaussian process, due to good analytical inference, computational flexibility, and straightforward uncertainty quantification (Rasmussen & Williams, 2006). GPSM can measure whether a new storage assignment is likely to improve order picking performance by means of the posterior prediction uncertainty. Based on the GPSM, we develop a storage location reassignment method that sequentially runs one swap operation at a time and suggests the reliable optimal swap operation with respect to the GPSM's average flow time estimate.

Contributions of this study are summarized as follows:

- **Model-driven approach vs. data-driven approach.** We directly estimate order picking performance with a data-driven approach by implementing the GPSM. Since estimating order picking performance is challenging due to warehouse complexity, size, and noise in the data, implicit workload proxies have been used to estimate performance in traditional methods (De Koster et al., 2012; Jane & Laih, 2005; Jane, 2000; Jewkes et al., 2004; Kim & Hong, 2020). Instead of approximating workload proxies, we optimize storage location assignments by using historical order picking data to estimate the average flow time with GPSM.
- **Domain knowledge-based performance improvement.** We improve the GPSM's estimation performance with our domain specific feature generation and training data configuration. As a result, we present an accurate learning model that effectively estimates order picking performance to solve SLAP.
- **Risk assessment using GPSM.** We propose a screening test based on the improvement probability obtained from the GP estimates and the storage reassignment procedure based on the estimation model and screening test The procedure optimizes storage space with minimal adjustments, reduces material handling costs, and enhances customer service in the warehouse.
- **Performance evaluation.** We use discrete event simulation and statistical tests to demonstrate the proposed models' ability to reduce flow time and reassignment effort, thereby improving warehouse management systems and operational efficiency.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 explains the fulfillment system and the proposed GPSM for average flow-time estimation. Section 4 introduces the proposed storage reassignment method using the flow-time estimation. Section 5 describes the experiment validating the model, and the simulations, results, and sensitivity analysis. Section 6 concludes and suggests future research.

2. Literature review

2.1 General storage location assignment

Most studies of the SLAP use travel distance and travel time as the performance measurements. Caron et al. (2010) developed an optimization model that minimizes the travel distance and travel time, and they also developed an estimation model that estimates the two performance measurements. Brynzér and Johansson (1996) suggested that grouping products by characterizing their variant information could reduce the product movements for traveling and picking. Muppani and Adil (2008) used nonlinear integer programming to capture the impact of a class-based storage system on the required space and material handling cost. Several studies have used probabilistic models and Markov chains to estimate travel distance or travel time to evaluate order picking performance and analyze the optimal storage assignment policy (Le-Duc & De Koster, 2005; Pan et al., 2014; Pan & Wu, 2009).

Motivated by a real problem, SLAP studies have also considered controlling the number of reassignment operations. Quintanilla et al. (2014) developed heuristic algorithms to maximize the available storage space by reoptimizing pallet locations in a random storage system. Kübler et al. (2020) suggested the ABC class-based iterative storage reassignment method considering order batching and picker routing together, and evaluated the reassignment effort with the future travel distance to identify promising reassignment. Pazour and Carlo (2015) developed mathematical model formulations for reassignment operating policies using an automated S/R crane and quantified the total loaded and unloaded travel distances while optimizing the reassignment operations.

In progressive zone picking systems, the SLAP has been focused on balancing workloads between zones for operational efficiency. Studies of workload balancing include Jane (2000), who proposed a heuristic algorithm to balance pickers' workloads by adjusting the number of storage zones. Pan, Shih, Wu, et al. (2015) developed a heuristic based on a genetic algorithm to solve the SLAP considering workload balance in a progressive zone picking system. Kim and Hong (2022) used mixed-integer programming to construct the storage location reassignment (SLR) model and applied it to a progressive bypass ZP system with a circulation conveyor and an S/R crane. The authors' model relocated the storage locations of products considering workload balance and recirculation reduction.

The models above, however, require solver tools, experts who can adapt the model to the warehouse environment, and lengthy computation to solve the kinds of large-size problems encountered by e-commerce warehouses. Most model-based optimization assumes a static operational environment, while data-driven optimization allows solving SLAP in an uncertain environment. Thus, we focus on data-driven storage location assignments for more agile and flexible SLAP optimization.

2.2 Data-driven storage location assignments

Various types of data-driven approaches have been developed for SLAP, and several studies have utilized clustering algorithms. Jane and Laih (2005) suggested a clustering algorithm to distribute the frequently requested products into several zones for workload balance in a synchronized zone picking system. Chuang et al. (2012) clustered associated items into groups and determined the sequence of the order groups for SLAP to minimize picking distance.

Chiang et al. (2011) used a data association algorithm to group products into similarity groups by order frequency and other product characteristics and introduced a data-mining based storage assignment (DMSA), which aimed to increase the association index (AIX) between products and their storage locations. Chiang et al. (2014) extended their research using a weighted support count (WSC) to calculate each AIX. The heuristic considered the relationship between a family and a cluster of products.

Pang and Chan (2016) developed a data mining-based assignment using association rules to minimize travel distances by controlling storage locations of correlated items and items near entry points. In a dynamic environment, Li et al. (2016) optimized the storage assignment based on the ABC classification and mutual affinity of products. The product affinity-based heuristic (PABH) identified the relationship between products. The authors used a greedy genetic algorithm because the problem was a quadratic assignment problem.

Regression techniques have been investigated for their use in storage location assignments. Sadiq et al. (1996) built a regression model to analyze the performance of storage location assignments considering order picking time and reassignment. Larco et al. (2016) used linear regression models to estimate worker discomfort factor and order cycle times. Larco et al. (2016) solved SLAP using the estimated values as the parameters of bi-objective optimization. Larco et al. (2016); Sadiq et al. (1996) estimated order picking performance with simple linear models based on factors such as storage location, bin type, product life cycle, and management policies, but neglected workload balance factors, a key influence on the performance of ZP systems.

Two practical issues deserve further investigation. Previous studies of data-driven SLAP consider the allocation of storage locations for entire products, yet when real-world reassignments are both resource- and cost-intensive, many warehouses choose to reallocate a limited range of products due to demand and operational uncertainty. The reassignment benefit should outweigh the reassignment cost while considering variables such as order sequence, workload balance, and rack storage policy. We use the Gaussian process to solve the data-driven SLAP with limited reassignment capacity and to quantify the reassignment risk to screen for reassignment operations with high potential for improvement.

3. Progressive ZP system and GPSM

In this section, we describe the progressive ZP system and explain the average flow time estimation using GPSM. We consider order picking in the concept of *wave management*, where a large set of orders is scheduled to be picked during a time period (*pick wave*) (Bartholdi & Hackman, 2008). We define the set of orders in a *pick wave* as an order list. Prior to the *pick wave*, there is an opportunity to reassign the storage location of items within the picking zones. The flow time of order refers to the time interval between the time a tote enters to the time it exits the ZP system. Flow time is a critical performance metric that needs to be reduced to shorten in-transit inventory and ensure fast delivery (Bartholdi & Hackman, 2008). According to practitioners, when considering throughput time or other indicators, it is challenging to operate a smooth transition to a post-process (e.g., packing), so the average flow time of orders is considered. We propose a learning model that estimates the average flow times of upcoming order lists based on workload balance information from historical order picking data.

3.1 Progressive ZP system characteristics

Our progressive zone picking system consists of an S/R crane and multiple zones. S/R cranes are automated material handling systems that are used in numerous manufacturing and warehousing systems to handle, store and retrieve discrete products (Ghomri & Sari, 2017). The S/R crane transfers each carton of products from a reserve rack to a flow rack, between reserve racks, or from the entrance or exit of the fulfillment center to any racks. The S/R crane can exchange the storage locations of a pair of products at a time by temporarily placing the products to be exchanged in the reserve rack area, which we refer to as a swap operation. Figure 2 shows an S/R crane replenishing a flow rack area with cartons from the reserve rack area.

Each zone has a flow rack and one zone-dedicated picker. Pickers pick products from cells in the flow rack and place them in totes. The totes travel on a conveyor through the zones. Each tote visits only the zones containing the products to be picked. When a zone has no room for a tote to enter, the tote skips over (bypasses) the zone and keeps traveling on the conveyor until a room is available.

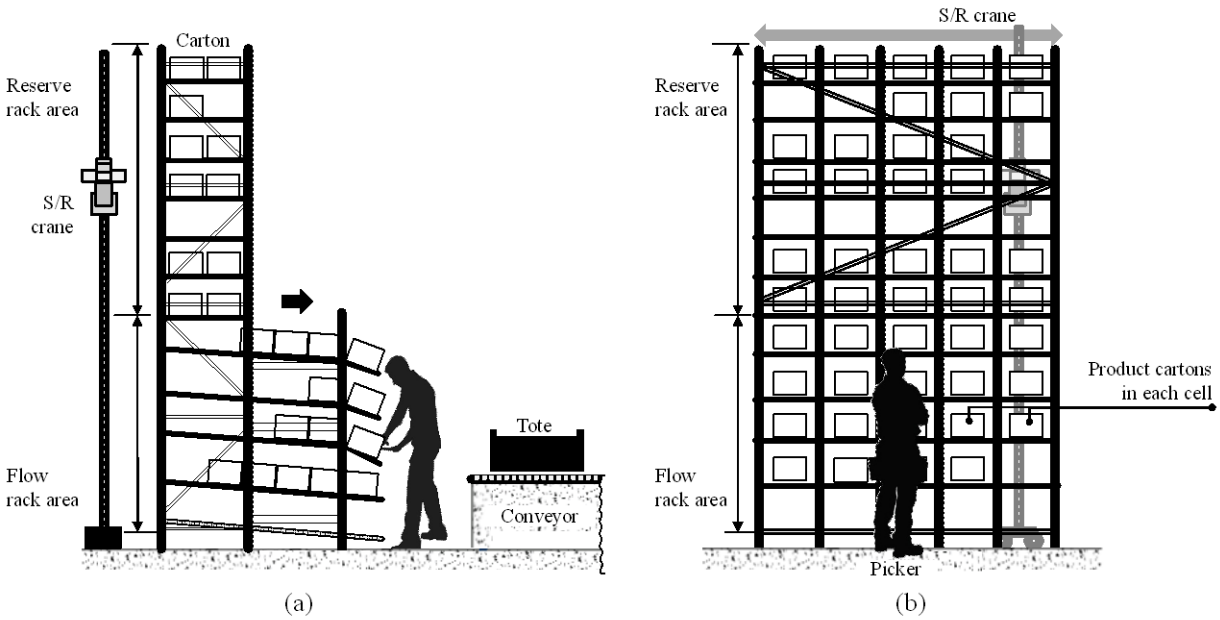

Figure 2. Progressive ZP flow-rack system with S/R crane (a) side view and (b) front view (modified from Cahyo (2017)).

A zone consists of multiple flow racks, and a flow rack consists of multiple cells to store products. We consider a simplified model, i.e., each cell contains only one product type. An order list is the batch of orders handled in the same shift or time window, and an order is the types and amounts of products.

We use the following notations.

Set and indices

O, o	The set of orders, and an order $o \in O$
L, l	The set of order lists, and an order list $l \in L$
Z, z	The set of zones, and a zone z $\in Z$
K, k	The set of cells, and a cell $k \in K$
P, p	The set of products, and a product $p \in P$
R, r	The set of relative rack column positions in a zone, and a rack position $r \in R$

Order and storage location representation

Let **C** be the $|O| \times |P|$ matrix representing products in orders with c_{op} as its (o, p) th element, where c_{op} be a binary variable indicating whether order o contains product p; c_{op} is 1 if order o contains product p, and 0 otherwise. Let **S** denote a $|P| \times |K|$ matrix describing the relationship between the products and cells with its (p, k) th element s_{pk} being a binary variable indicating whether product p is located on cell k. Let **W** be a $|O| \times |K|$ matrix representing the relationship between the orders and cells with its (o, k) th element w_{ok} being a binary variable indicating whether order o requires a product to be picked from cell k. The matrix **W** is related to matrices **C** and **S** through

$$\mathbf{W} = \mathbf{C} \cdot \mathbf{S}$$

$$= \begin{bmatrix} c_{11} & \cdots & c_{1|P|} \\ \vdots & \ddots & \vdots \\ c_{|O|1} & \cdots & c_{|O||P|} \end{bmatrix} \cdot \begin{bmatrix} s_{11} & \cdots & s_{1|K|} \\ \vdots & \ddots & \vdots \\ s_{|P|1} & \cdots & s_{|P||K|} \end{bmatrix}.$$

Abstract representation by order list and Zone/Rack locations

Let **A** be a $|K| \times |L|$ matrix of $a_{lk}'s$ that represents the total number of picks from cell $k \in K$ required to fill order list $l \in L$. Let **B** represent a $|L| \times |Z|$ matrix with b_{lz} as its (l, z) th element which represents the total number of picks from zone $z \in Z$ to fill order list$l \in L$. Let **G** be a $|L| \times |R|$ matrix of $g_{lr}'s$ that represents the total number of picks from flow rack $r \in R$ required to fill order list $l \in L$. Let O_l denote the set of orders in the order list l. Let K_z and K_r denote the set of cells assigned in zone z and the set of cells assigned in rack position r, respectively. We can calculate a_{lk} using the matrix **W** as

$$a_{lk} = \sum_{o \in O_l} w_{ok}.$$

We can calculate b_{lz} as

$$b_{lz} = \sum_{k \in K_z} a_{lk},$$

and calculate g_{lr} as

$$g_{lr} = \sum_{k \in K_r} a_{lk}.$$

Below, we summarize the constant matrices for the problem formulation.

C, c_{op}	Order-Product inclusion matrix and its elements c_{op}, $\forall o \in O$ and $p \in P$
W, w_{ok}	Order-Cell inclusion matrix and its elements w_{or}, $\forall o \in O$ and $k \in K$
S, s_{pk}	Product-Cell inclusion matrix and its elements s_{pr}, $\forall p \in P$ and $k \in K$
A, a_{lk}	Order list-Cell relationship matrix and its values a_{lk}, $\forall l \in L$ and $k \in K$
B, b_{lz}	Order list-Zone relationship matrix and its values b_{lz}, $\forall l \in L$ and $z \in Z$
G, g_{lh}	Order list-Rack position relationship matrix and its values g_{lh}, $\forall l \in L$ and $h \in H$

3.2 GPSM for Average flow time estimation

Feature Generation

To estimate the average flow time of orders in an order list ($\overline{FT}$), an estimation model requires the key input variables (i.e., features) that are expected relate to $\overline{FT}$. We generate features from historical order picking data that can represent the picking time and workload balance across zones. Picker's travel distance and workload balancing measures, such as standard deviations of workload between zones, and zones' maximum and minimum numbers of workload, are potentially related to $\overline{FT}$ (Huang et al. (2018), Vanheusden et al. (2022)). We develop GPSM using three features of zone workload balance and a feature of rack storage policy to estimate the average flow time $\overline{FT}$.

Let $\mathbf{x}_l$ represent the three input features extracted from order list $l \in L$, and let $\mathbf{X} = \{\mathbf{x}_l, \boldsymbol{l} \in \boldsymbol{L}\}$ We consider three models. The first GPSM, which we refer to as GPSM_{AZ}, is only based on the first input feature, i.e., picker's travel distance. We define a total travel distance to complete order list $l \in L$ as

$$TD_l = \sum_{r \in R} d_r \cdot g_{lr},$$

where d_r is the distance from the picker's loading depot to a rack position $r \in R$.

The second GPSM, which we call GPSM_{MM}, uses three input features: TD_l and the maximum and minimum number of workloads of zones. We calculate the maximum number of picks among zones b_l^{max} and the minimum numbers of picks among zones b_l^{min} as

$$b_l^{max} = \max_{\forall z \in Z} b_{lz},$$

$$b_l^{min} = \min_{\forall z \in Z} b_{lz}.$$

The third GPSM, which we refer to as GPSM$_{SD}$, also uses TD_l and one additional input, the standard deviation of the number of picks per zone. We let b_l^{SD}represent the standard deviation for order list $l \in L$, which can be obtained as

$$b_l^{SD} = \sqrt{\frac{\sum_{z\in Z}(b_{lz} - \frac{1}{|Z|}\sum_{z\in Z}(b_{lz}))^2}{|Z|-1}}.$$

GPSM

Gaussian process (GP) regression, a popular surrogate model for computer and physical experiments (Mackay, 1998; Murphy, 2012), provides a predictive relation between input and response variables. For our problem, the input is an input feature set $\boldsymbol{x}$, and the response variable y would be the average flow time $\overline{FT}$. Suppose that we have training data of input features and associated response values for orders in order list L. We denote the training data by $(\boldsymbol{X}, \boldsymbol{y})$, where $\boldsymbol{X} = (\boldsymbol{x_l}, l \in L)$ and $\boldsymbol{y} = (y_l, l = 1 \in L)$ represent the training inputs and responses. The inputs and responses are related via an unknown regression model $f(\boldsymbol{x})$ as

$$y_l = f(\boldsymbol{x_l}) + \epsilon_l, \epsilon_l \sim N(0, \sigma_n^2).$$

We assume the regression function $f(\boldsymbol{x})$ is a zero-mean Gaussian process with covariance function K($\boldsymbol{x}$, $\boldsymbol{x}'$). Given the training data, we like to predict the posterior distribution of $f(\boldsymbol{x}_*)$ for a test input $\boldsymbol{x}_*$, which follows a Gaussian distribution with mean $\hat{\mu}(\mathbf{x}_*)$ and variance $\hat{\Sigma}(\mathbf{x}_*)$ (Rasmussen & Williams, 2006),

$$\hat{\mu}(\mathbf{x}_*) = K(\mathbf{x}_*,\mathbf{X})[K(\mathbf{X},\mathbf{X}) + \sigma_n^2\boldsymbol{I}]^{-1}y, \tag{1}$$

$$\hat{\Sigma}(\mathbf{x}_*) = K(\mathbf{x}_*,\mathbf{x}_*) - K(\mathbf{x}_*,\mathbf{X})[K(\mathbf{X},\mathbf{X}) + \sigma_n^2\boldsymbol{I}]^{-1}K(\mathbf{x}_*,\mathbf{X}), \tag{2}$$

where $\boldsymbol{I}$ is an $|\boldsymbol{L}| \times |\boldsymbol{L}|$ identity matrix, and $\boldsymbol{K}(\boldsymbol{X_1}, \boldsymbol{X_2})$ is the matrix of the covariance function values evaluated between $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$. We use the covariance function composed of a squared exponential (SE) covariance function and noise variance function,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\lambda^2}\right) + \sigma_n^2\delta(\mathbf{x}_i, \mathbf{x}_j), \tag{3}$$

where variable length scale λ, signal variance σ_f^2, and noise variance σ_b^2 control the correlation between observations $\mathbf{x}_i$ and $\mathbf{x}_j$, and δ is the Kronecker delta function. The kernel-function option depends on the *input* and expected patterns in the data, e.g., Richardson et al. (2017) considered SE and Matérn covariance as the kernel function.

4. Storage location reassignment with risk assessment

In this section we define a storage location reassignment as the sequential determination of swap operations, each of which is optimized to minimize the average flow-time of orders based on the

GPSM's output. We also use the GPSM estimate to evaluate the improvement probability of the swap operations by screening out the less promising ones.

4.1 Swap operation

The storage location reassignment adjusts the locations of products to minimize the average flow-time estimate from the GPSM. We use a series of swap operations (i.e., switch a pair of product cartons) to optimize storage locations with limited handling capacity. For each swap operation, we assume that

- Each picker stays in his/her zone and independently retrieves each product per tote.
- Each product occupies one zone and one rack.
- Each zone has the same number of racks.
- Constant picking time includes search time, pick time, and inspection time.
- Management has sufficient time to reassign locations for new order lists (Sadiq et al., 1996).
- All products are available before an order enters the system.
- A swap operation swaps two cartons of products across flow racks at a time.
- The total number of the swaps is finite and determined dynamically.
- The swap distance as defined by the physical distance between two rack locations, and the number of cartons affect swap time.
- The number of swap operation is not predetermined; after a swap, management determines if another swap is necessary.
- The reserved area has empty racks; therefore, swap operations can be performed without additional handling.
- The swap operation utilizes the reserved area to temporarily store the products to be reassigned. The reserved area has empty racks; therefore, swap operations can be performed without additional handling.

Each swap operation is optimized so that the choice minimizes the average flow time estimated by the pre-trained GPSM described in Section 3.2. A swap operation is represented by a column swap of the original product-pick-face inclusion matrix **S**. For example, suppose there are five products ($P = \{1,2,3,4,5\}$) in each pick-face. If the swap operation occurs between the product 2 and product 4, then it makes a change of the original product-pick-face inclusion matrix **S** to **S`**

$$\mathbf{S}=\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix},\qquad \mathbf{S}` =\begin{bmatrix}1&0&0&0&0\\0&0&0&1&0\\0&0&1&0&0\\0&1&0&0&0\\0&0&0&0&1\end{bmatrix}.$$

After the update to **S`**, we can calculate the corresponding testing input $\boldsymbol{x}_*$ and then get the associate posterior predictive distribution $f(\boldsymbol{x}_*)$ from the GPSM to get the estimated distribution of the average flow time resulting from the swap. We repeat this for a list of *N* feasible swap operations and select the best swap operation that minimizes the expected average flow time, i.e., $E[f(\boldsymbol{x}_*)]$. We consider all possible swap cases among $|P|$ products, which gives N equal to two-out-of-$|P|$ combination caes. We iterate the swap operation to the maximum number of swaps *M* determined by management.

4.2 Feedforward heuristic

Since exact algorithm for an optimal sequence of M swap operations is not suitable due to the combinatorial nature of the problem, we propose Algorithm 1, where the best solutions from the precedence process become the input of the next process. The steps of the feedforward heuristic are as follows.

Algorithm 1. Feedforward heuristic.

Step 1.	Set *M* (based on warehouse experience and considering the number of storage locations that can be reassigned), and set i =1 and set S^1 to be the current product-pick-face inclusion matrix.
Step 2.	Generate *N* candidates $\{\mathbf{S}_1^i, \mathbf{S}_2^i, \dots, \mathbf{S}_N^i\}$ from S^{i-1}. (See 4.1 Swap operation)
Step 3.	Calculate the corresponding input features $\{\mathbf{x}_{*1}^i \mathbf{x}_{*2}^i, \dots, \mathbf{x}_{*N}^i\}$ corresponding to the *N* candidates.
Step 4.	Obtain $\hat{\mu}(\mathbf{x}_*^i)$ from each candidate and find the optimal swap operation based on equation (4).
Step 5.	Update S^i.
Step 6.	If $i < M$, $i = i + 1$ and go to step 2, otherwise STOP.

Let $\mathbf{S}^i$ represent the *i*th storage location matrix for the iteration $i = 1\dots, M$, where *M* controls the maximum number of swap operations set according to the warehouse environment. For every iteration *i*, generate storage location candidates $\{\mathbf{S}_1^i, \mathbf{S}_2^i, \dots, \mathbf{S}_N^i\}$ with swap operations from former storage locations $\mathbf{S}^{i-1}$. Each candidate generates the testing inputs $\{\mathbf{x}_{*1}^i \mathbf{x}_{*2}^i, \dots, \mathbf{x}_{*N}^i\}$ to be evaluated with the proposed average flow-time estimation model. Next, determine the storage locations $\mathbf{S}^i$ according to the best $\mathbf{x}_*^i$ that yields the minimum $\hat{\mu}(\mathbf{x}_*^i)$ using Equation (1). Obtain the expected average flow time of the storage locations $\mathbf{S}^i$ as

$$\mathrm{E}[\overline{FT}(\mathbf{S}^i)] = \min\{\hat{\mu}(\mathbf{x}_*^i) \big| \mathbf{x}_*^i \in \{\mathbf{x}_{*1}^i, \mathbf{x}_{*2}^i, \dots, \mathbf{x}_{*N}^i\}\}. \tag{4}$$

Considering an initial storage location $\mathbf{S}^0$ for an order list l and obtaining the historical data in the form listed, use Algorithm 1 to find the new storage location ($\mathbf{S}^1$) at iteration 1 that minimizes the average flow time for an order list l. Note that the new storage location $\mathbf{S}^1$ from the first stage is considered as the initial storage location $\mathbf{S}^{i-1}$ of the second stage and obtains the new storage location

$\mathbf{S^2}$ that yields the minimum $\overline{FT}$ at the second stage. Run these procedures until iteration i reaches the maximum number of swaps M.

4.3 Screening out less promising swaps

Quantifying uncertainty is critical in managing operational risks. The probability that a system will not improve can be estimated with the estimates of Gaussian process model (Bect et al., 2012). We conduct a screening test using the posterior predictive distribution from GPSM. We use the posterior predictive distribution of the average flow time after a swap to evaluate the probability of improvement in the average flow time, compared to the average flow time before that swap. If the improvement probability is not sufficiently high, we revoke the determined swap operation.

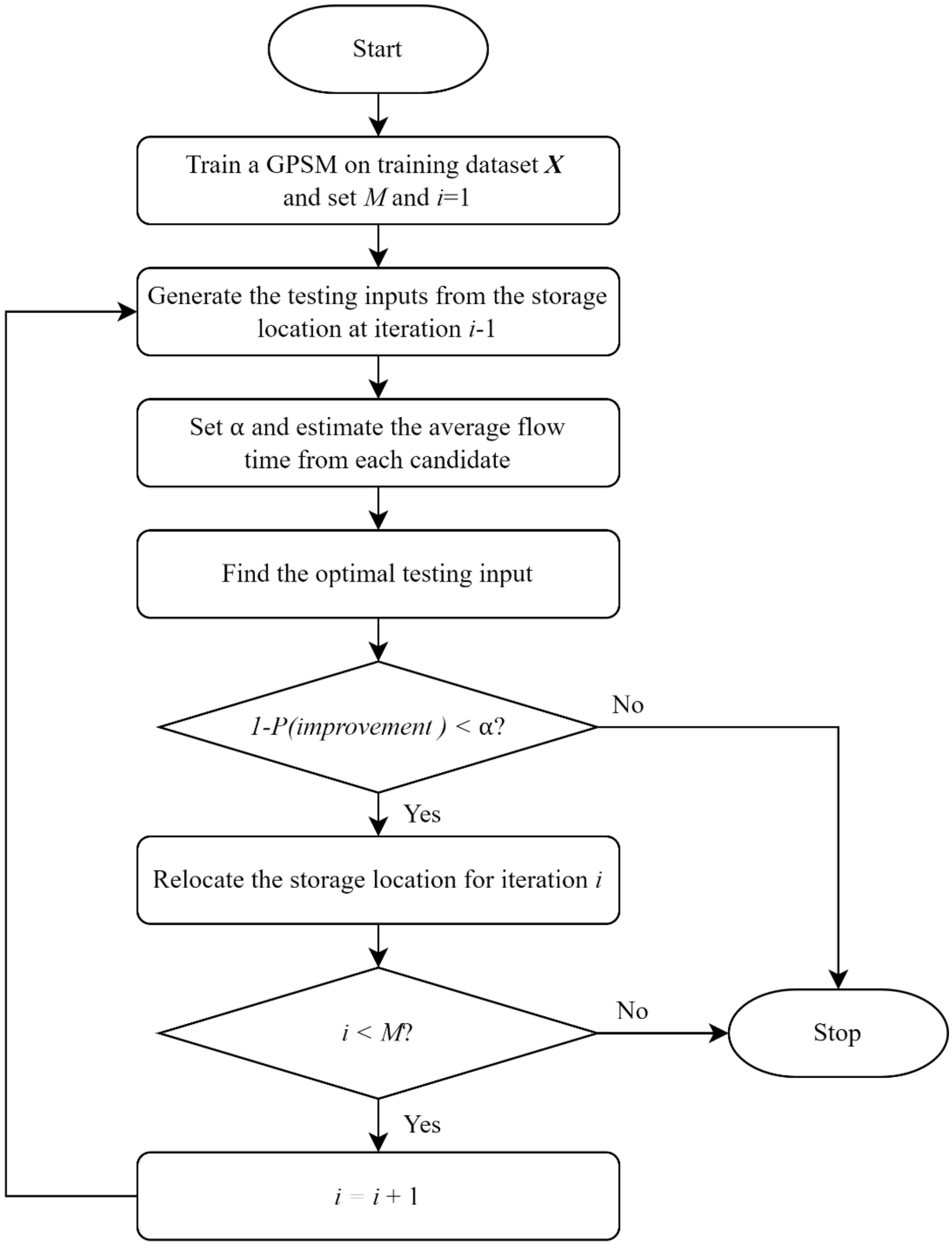

Figure 3. Flowchart of a screening test applied to Algorithm 1.

Given the means and variances of the two mutually independent Gaussian distributions, we use a one-tailed test to statistically compare the difference between the mean values of two distributions. GPSM is used to obtain the mean($\hat{\mu}(\cdot)$) and variance($\hat{\Sigma}(\cdot)$) of $\mathbf{x}_*^{\boldsymbol{i}}$ and $\mathbf{x}_*^{\boldsymbol{i-1}}$, where $\mathbf{x}_*^{\boldsymbol{i}}$ is the testing input expected to be the optimal case from the relocated cases $\{\mathbf{x}_{*\mathbf{1}}^{\boldsymbol{i}}, \mathbf{x}_{*\mathbf{2}}^{\boldsymbol{i}}, \ldots, \mathbf{x}_{*\boldsymbol{N}}^{\boldsymbol{i}}\}$, and $\mathbf{x}_*^{\boldsymbol{i-1}}$ is the testing input of the reassigned order list at iteration $i-1$. The null hypothesis is that the $\overline{FT}$ of $\mathbf{x}_*^{\boldsymbol{i-1}}$ is less than or equal to the $\overline{FT}$ of $\mathbf{x}_*^{\boldsymbol{i}}$. The probability of the null hypothesis (p-value) denotes the probability that the reassignment operation will not improve. We set significance level α for the one-tailed test to classify the case with the risk to increase $\overline{FT}$. We obtain the Z-score of the difference between the two $\overline{FT}$ values and the improvement probability from the Z-score as

$$1 - P(improvement) = P(\overline{FT}(\mathbf{S}^{i-1}) - \overline{FT}(\mathbf{S}^{i}) \leq 0) = P(Z \geq \frac{\hat{\mu}(\mathbf{x}_*^{i-1}) - \hat{\mu}(\mathbf{x}_*^{i})}{\sqrt{\hat{\Sigma}(\mathbf{x}_*^{i-1}) + \hat{\Sigma}(\mathbf{x}_*^{i})}}). \quad (5)$$

We compare the one-tailed p-value with the significance level α. Note that the improvement probability of reassignment cannot be guaranteed if the p-value is greater than the significance level. If the p-value becomes smaller than the significance level, the GPSM expects that the flow time of the GPSM optimal assignment is significantly shorter than the flow time of the original assignment; therefore, it executes the determined swap operation.

5. Industrial application of GPSM

In this section we describe our simulation experiment and the results. Ten scenarios from a real e-commerce warehouse configuration and historical order data were used to evaluate the performance of the proposed model. We analyze the effectiveness of reassignment decisions and conduct sensitivity tests over various warehouse configurations.

5.1 Workload scenarios

To obtain accurate and reliable estimates, we use different workload scenarios to train the proposed models on various storage policy factors. We consider the four zone workload distribution types and two rack storage policies shown in Table 1.

Table 1 Zone workload distribution types and rack storage policies.

Zone workload distribution type	Rack storage policy
Uniform	ABC(5:3:2)
Bottleneck zone	Random
Descending demand	
Irregular demand	

We note that a uniform zone workload distribution type is the most workload-balanced scenario with almost uniformly distributed workloads over zones. We generate Scenarios 1 and 2 with this type to make workload balanced scenarios. The bottleneck zone distribution type refers to the case that one randomly chosen zone is assigned 25% of total orders when all other zones have a uniform workload distribution. The bottleneck zone scenarios can occur in situations when demand for particular products increase dramatically that often arises in e-commerce warehouses. We generate Scenarios 3 and 4 with this type. A descending demand zone distribution type has a descending trend in workload. The first zone has the largest demand volume and the demand gradually decreases toward the last zone. Many warehouse managements set descending demand in progressive zone picking system to avoid blocking delay between zones. We distribute the workload ratio of each zone as the ratio of each zone index to the sum of zone indexes. We generate Scenarios 5 and 6 with this type. An irregular demand-zone distribution type represents the most workload-unbalanced scenario with an uneven distribution of workloads across zones. We demonstrate worst-case workload balancing by randomly shuffling storage locations for descending demand scenarios and generating irregular demand scenarios. We generate Scenarios 7 and 8 with this type.

The rack storage policy defines how products' storage locations are assigned over rack column positions. The ABC (5:3:2) class-based rack storage policy classifies products to classes A, B and C in a ratio (0.5:0.3:0.2) for allocating high-demand products in the rack column positions with the picker's shortest travel distance. The random rack-storage policy refers to the random assignment of products within zones.

We generate the first eight workload scenarios using a full factorial design with four zone workload distribution types and two rack storage policies and two additional scenarios. To train GPSM even with scenarios that cannot be considered by the full factorial design, we generate scenarios 9 and 10 by switching a randomly chosen pair of product locations from Scenario 1. Scenarios 9 and 10 have balanced workloads, but the storage locations of products differ from the preset rack storage policy due to demand fluctuations and a series of previous storage relocations. We classify Scenarios 1, 2, 9, and 10 as workload balanced scenarios and Scenarios 3, 4, 5, 6, 7, and 8 as workload unbalanced scenarios. In Table 2, the average $\overline{FT}$ of 50 order lists per scenario measured by simulation shows that zone workload distribution type and rack storage policy directly influence order picking performance. The workload balanced Scenarios 1, 2, 9, and 10 provide shorter average $\overline{FT}$. We conduct statistical tests to validate the significance of difference between scenarios (Please see Table B1 and Table B2 in Appendix B for detail).

Table 2. Ten workload distribution scenarios.

Scenario	Zone workload distribution type	Rack storage policy	Average $\overline{FT}$
1	Uniform	ABC (5:3:2)	1250.03
2	Uniform	Random	1309.47
3	Bottleneck zone	ABC (5:3:2)	1440.65
4	Bottleneck zone	Random	1523.62

5	Descending demand	ABC (5:3:2)	1513.42
6	Descending demand	Random	1593.70
7	Irregular demand	ABC (5:3:2)	1522.98
8	Irregular demand	Random	1586.75
9	24 times random switch from scenario 1		1284.24
10	100 times random switch from scenario 1		1312.51

5.2 Experimental configuration

To validate our proposed GPSM, we configure a warehouse and its order profiles based on an e-commerce company's progressive bypass ZP system in Korea. We run 200 experiments per configuration, i.e., 20 different order lists per each of the 10 scenarios. We train GPSM on 300 training datasets consisting of 10 scenarios with 30 different order lists. We use Tecnomatix© Plant Simulation 12 to build the simulation model. We generate synthetic historical data based on the order profiles and modified sizes of the ZP system. Since recirculation and bypass disrupt the order sequence, the order release sequence follows the First-Come-First-Serve rule. We use Python 3.7 and scikit-learn toolbox to analyze the data. Table 3 reports the details.

Table 3. Warehouse configuration.

Parameter	Values
Number of swaps (n)	1, 2, 3, 4
Number of zones	8
Number of rack columns per zone	6
Order size	$Uniform$ (3,9)
Number of orders in each order list	100

Note: Default values of each parameter are underscored.

We label the GPSMs as stages 1–4 by setting $M = 4$. GPSM(n) denotes the first stage resulting in the swapping of n pairs of storage locations. We compare the GPSM to Kim and Hong (2022), who proposed an MIP model to reassign a limited number of storage locations for the same order picking system. For simplicity, we call their model SLR(n), where n is stage number (i.e., the number of the swap operations). We also compare the GPSM with the heuristic algorithm proposed by Jane (2000) for workload balancing in a zone picking system. For simplicity, we call their heuristic algorithm the JA. Unlike the GPSM and SLR, the JA reassigns all storage locations without limiting the number of reassignments.

The objective of the GPSM is to minimize the average flow time; the objective of the SLR(n) model is to minimize the maximum number of picks and the maximum number of order visits among all zones; and the objective of the Jane is to distribute the number of picks evenly across all zones.

Since all three models aim to balance the workload in the progressive ZP system to minimize order flow time, we compare the models' performance.

5.3 GPSM effectiveness analysis

Figure 4 illustrates the percentage reductions of average $\overline{FT}$ between the original assignment and three GPSMs, and between the original assignment and SLR(n) when swapping multiple pairs of products. Without the screening test, $GPSM_{SD}(1)$ yields a 3.23% average reduction, $GPSM_{SD}(4)$ yields an 8.87% average reduction, SLR(1) yields a 2.93% average reduction, and Jane yields an 8.21% in $\overline{FT}$. The workload unbalanced scenarios (1, 2, 9, and 10) have larger $\overline{FT}$ reductions than the workload balanced scenarios (3, 4, 5, 6, 7, and 8). For the same zone workload distribution type scenarios in the GPSMs, the random storage policy yields a high reduction percentage of $\overline{FT}$. The results indicate that the workload unbalanced scenarios have scope for further improvement than the workload balanced scenarios.

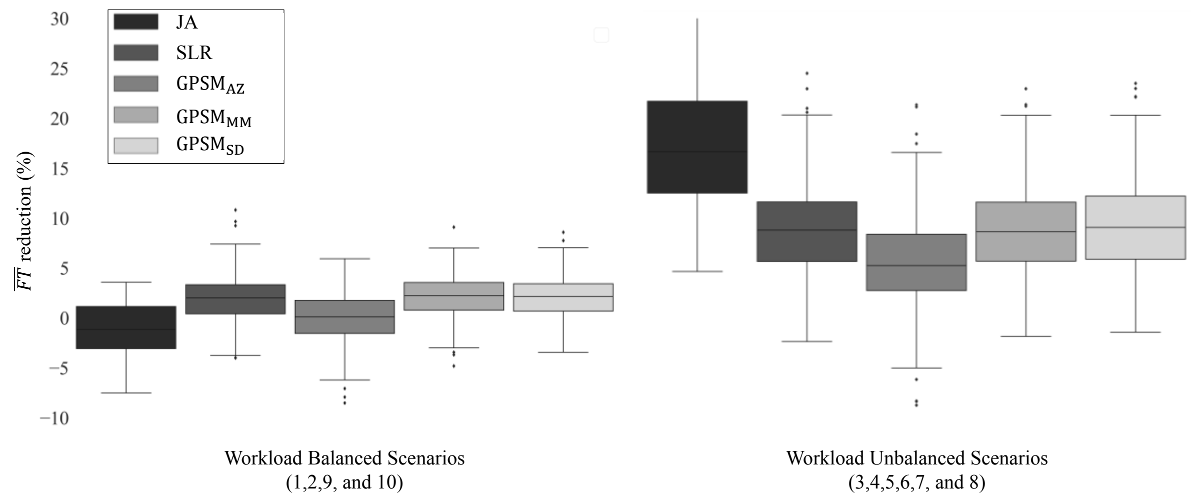

Figure 4. Percentage reductions of average $\overline{FT}$ by workload scenarios without the screening test.

Without the screening test, we observe 24 failures of reassignment that yield a 1.13% increase in the average $\overline{FT}$ compared with the original assignment when $GPSM_{SD}(1)$ runs for 200 instances. For the same instances, Jane obtains 49 failures of reassignment with an increase of 2.96% in the average $\overline{FT}$ over the original assignment, and SLR(1) results in 28 failures of reassignment with an increase of 1.13%. The screening test excludes the reassignment that is expected to improve with low probability.

Table 4 reports the results of the paired t-tests for each scenario and stage to identify statistically significant differences between SLR and $GPSM_{SD}$. The differences are insignificant in the balanced scenarios (1, 2, 9, 10) or low stages (n = 1,2). Where swap operations are performed more than three times, $GPSM_{SD}$ has shorter $\overline{FT}$ than SLR in the unbalanced scenarios (3, 4, 5, 6, 7, 8).

Table 4. Paired t-test results on $\overline{FT}$ of SLR(n) – $GPSM_{SD}$(n)

Stage	Scenario	Paired t-test on the average $\overline{FT}$			
		Mean difference	*t*-value	DF	*P*-value
$n = 1$	1	11.405	2.082	19	0.051
	2	-1.286	-0.227	19	0.823
	3	8.831	1.322	19	0.202
	4	-6.932	-1.246	19	0.228
	5	-3.917	-0.638	19	0.531
	6	3.569	0.451	19	0.657
	7	3.630	0.476	19	0.639
	8	17.554	1.960	19	0.065
	9	-0.853	-0.206	19	0.839
	10	11.577	2.059	19	0.053
$n = 2$	1	16.198	3.604	19	0.002
	2	3.657	0.700	19	0.493
	3	13.045	1.613	19	0.123
	4	-3.310	-0.379	19	0.709
	5	-3.297	-0.608	19	0.551
	6	0.542	0.075	19	0.941
	7	6.574	0.775	19	0.448
	8	7.898	1.730	19	0.100
	9	4.738	0.792	19	0.438
	10	4.738	0.792	19	0.438
$n = 3$	1	-0.846	-0.154	19	0.879
	2	7.804	2.075	19	0.052
	3	15.395	1.676	19	0.110
	4	23.124	4.068	19	0.001
	5	19.973	2.827	19	0.011
	6	20.554	2.211	19	0.040
	7	0.025	0.003	19	0.998
	8	23.433	3.167	19	0.005
	9	3.494	0.973	19	0.343
	10	2.883	0.453	19	0.656
$n = 4$	1	-1.753	-0.277	19	0.785
	2	0.667	0.126	19	0.901
	3	24.278	2.614	19	0.017
	4	18.841	2.228	19	0.038
	5	25.310	3.224	19	0.004
	6	25.168	2.154	19	0.044
	7	16.812	2.512	19	0.021
	8	30.869	4.215	19	0.000
	9	-7.403	-1.706	19	0.104
	10	-5.434	-1.075	19	0.296

Figure 5 illustrates the screening test's effects on productivity for 200 assignments from SLR, SD-GPSM before the screening test, $GPSM_{SD}$ after the screening test, and the original assignments without reassignment. We adopt Youden's J statistic for the significance levels of the screening tests (Youden, 1950). We set swapping time (ST) weight, considering the replenishment time in the real e-commerce warehouse. On the x-axis, we classify experimental cases by different reassignment strategies, stage number, and ST weight. The black bars indicate the time difference between the start and finish of a sequence of an order list (C_{max}; makespan), and the dotted white bars indicate the total swapping time (TST). The y-axis represents the average total elapsed time including C_{max} and TST.

The makespan distinctly decreases as the stages increase. $GPSM_{SD}$ after the screening test yields a shorter TST because the screening test filters out the number of swap operations. When $ST = 2.5$, the total swapping time of $GPSM_{SD}$(2) before the screening test averages 300 seconds, and the total swapping time of $GPSM_{SD}(3)$ after the screening test averages 256.5 seconds. Before the screening test, the makespan of $GPSM_{SD}(2)$ averages 2449.90 seconds, and the makespan of $GPSM_{SD}(3)$ averages 2413.62 seconds. After the screening test, $GPSM_{SD}$ frequently shows a large makespan reduction with less swapping time.

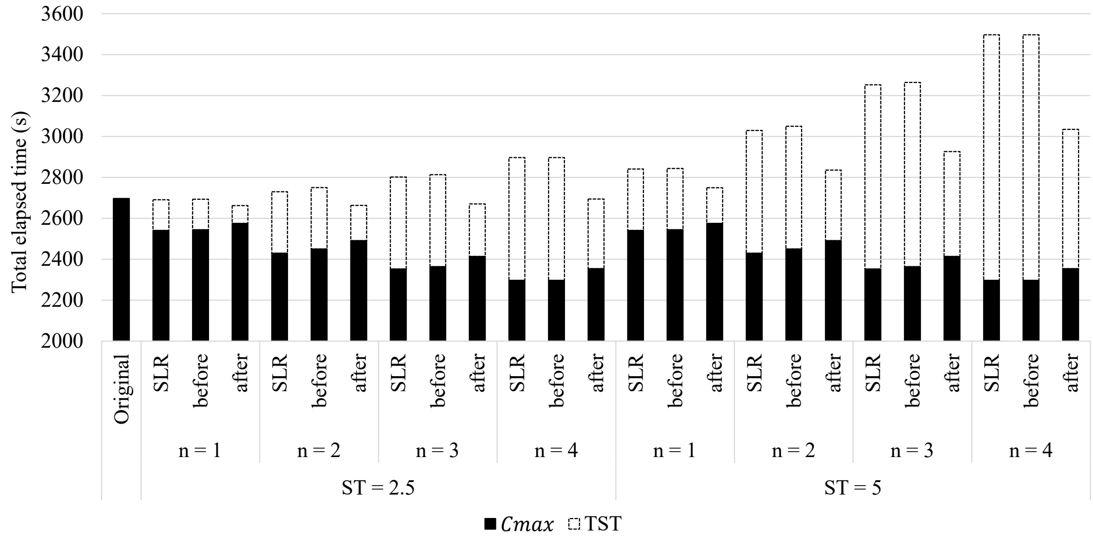

Figure 5. Screening test effectiveness analysis considering swapping times.

5.4 Sensitivity analysis of GPSM with a screening test

We consider different warehouse configurations (i.e., increasing the number of zones, rack columns, orders in an order list, and order sizes). We also consider non-identical picker skill configuration (i.e., two slow-moving pickers in eight zones) (Bartholdi and Eisenstein (1996), Hong (2019)).We use the following measurements:

- Before: The average $\overline{FT}$ reduction before the screening test (%)
- α: The significance level for the screening test
- #swaps: The average number of swaps after the screening test
- Avg: The average $\overline{FT}$ reduction (%)
- Worst: The $\overline{FT}$ reduction of the worst case (%)

Table 5 reports the sensitivity analysis results. In Appendix C, we conduct statistical analyses to identify which methods have statistically significant differences across order picking environments. As a result of the fact that the data variances for each group are not equal, we use Welch's ANOVA test (Table C1) and Games-Howell as post-hoc tests (Table C2, Table C3, Table C4, Table C5, Table C6, and Table C7) for each order picking environment.

Table 5. Results of GPSM sensitivity analysis with screening tests.

Environment	JA		Stage	SLR(n)			$GPSM_{AZ}(n)$					$GPSM_{MM}(n)$					$GPSM_{SD}(n)$				
	Avg.	Worst		#swaps	Avg.	Worst	Before	α	#swaps	Avg	Worst	Before	α	#swaps	Avg	Worst	Before	α	#swaps	Avg	Worst
Default	*8.21*	*-8.13*	$n = 1$	1.00	2.93	-4.16	1.92	0.39	0.27	0.89	-1.76	3.02	0.14	0.55	2.28	-0.82	3.23	0.15	0.59	2.56	-1.90
			$n = 2$	2.00	5.28	-3.98	3.06	0.44	0.50	1.37	-1.76	5.32	0.32	1.00	3.86	-0.10	5.60	0.26	1.15	4.54	0.00
			$n = 3$	3.00	6.82	-3.75	3.90	0.46	0.70	1.78	-1.76	7.26	0.37	1.40	5.06	-0.10	7.59	0.31	1.71	6.23	0.00
			$n = 4$	4.00	8.06	-2.22	4.40	0.48	0.90	2.06	-1.76	8.40	0.37	1.72	5.54	-0.10	8.88	0.37	2.27	7.51	0.00
12 zones	*9.41*	*-8.69*	$n = 1$	1.00	2.66	-6.22	2.00	0.41	0.28	0.94	-2.60	2.71	0.27	0.51	2.14	-3.84	2.40	0.36	0.42	1.69	-3.30
			$n = 2$	2.00	4.59	-6.43	3.00	0.50	0.56	1.47	-2.84	4.62	0.37	0.86	3.44	-3.84	4.20	0.41	0.81	2.98	-3.99
			$n = 3$	3.00	5.98	-4.85	4.38	0.50	0.69	1.62	-0.90	5.82	0.39	1.14	4.32	-3.84	5.64	0.44	1.20	3.88	-3.18
			$n = 4$	4.00	7.31	-5.22	5.32	0.50	0.72	1.66	-0.90	6.96	0.41	1.38	5.07	-3.84	6.44	0.44	1.54	4.50	-3.18
12 rack columns	*7.00*	*-17.59*	$n = 1$	1.00	2.26	-5.68	1.53	0.42	0.25	0.75	-0.37	2.83	0.16	0.52	1.82	-2.83	2.69	0.18	0.51	1.83	-1.39
			$n = 2$	2.00	3.85	-2.85	2.86	0.45	0.45	1.12	-2.37	4.72	0.32	0.99	3.14	-2.83	4.94	0.35	1.02	3.39	-0.47
			$n = 3$	3.00	5.02	-2.41	3.57	0.49	0.63	1.48	-1.11	6.48	0.36	1.45	4.38	-2.83	6.72	0.31	1.51	4.67	0.00
			$n = 4$	4.00	6.04	-2.02	4.48	0.50	0.80	1.70	0.00	8.04	0.40	1.90	5.26	-2.83	8.28	0.35	2.00	5.79	0.00
Unif(10,20) *Order size*	*7.27*	*-8.68*	$n = 1$	1.00	1.91	-2.47	1.41	0.47	0.80	1.38	-7.23	2.43	0.19	0.79	2.46	-1.31	2.53	0.19	0.77	2.54	-1.31
			$n = 2$	2.00	3.51	-2.04	2.50	0.41	1.18	2.20	-7.23	4.12	0.28	1.42	3.95	-1.31	4.36	0.25	1.46	4.20	-0.72
			$n = 3$	3.00	4.87	-2.07	3.33	0.41	1.32	2.52	-7.23	5.54	0.43	2.04	5.21	-1.31	5.82	0.35	2.15	5.56	0.00
			$n = 4$	4.00	6.01	-2.32	3.98	0.46	1.42	2.67	-7.23	6.73	0.39	2.65	6.31	-1.31	7.14	0.38	2.83	6.75	0.00
200 orders in each order list	*13.59*	*-7.60*	$n = 1$	1.00	4.14	-1.19	3.08	0.43	0.53	2.19	-3.60	4.03	0.24	0.86	3.93	-0.42	3.73	0.18	0.38	1.55	-0.42
			$n = 2$	2.00	7.06	-1.05	5.44	0.40	0.79	3.24	-3.60	6.96	0.40	1.67	6.72	0.00	6.76	0.27	0.69	2.64	-0.42
			$n = 3$	3.00	9.22	-1.58	6.84	0.40	0.80	3.27	-3.60	9.21	0.39	2.39	8.80	0.00	9.15	0.33	0.93	3.35	-0.42
			$n = 4$	4.00	10.82	-0.93	7.60	0.50	0.81	3.29	-3.60	10.88	0.42	3.05	10.30	0.00	10.92	0.39	1.15	3.76	-0.42
Non-identical picker skill	*7.74*	*-10.47*	$n = 1$	1.00	2.78	-4.99	3.60	0.41	0.93	3.44	-2.75	2.85	0.37	0.73	2.70	-2.14	2.99	0.33	0.52	2.48	-2.03
			$n = 2$	2.00	4.77	-6.42	6.22	0.39	1.43	4.92	-2.46	5.22	0.43	1.32	4.78	-3.16	5.40	0.40	1.03	4.21	-1.12
			$n = 3$	3.00	6.49	-4.25	8.19	0.47	1.91	5.89	-2.46	7.14	0.43	1.82	6.01	-3.16	7.20	0.43	1.54	5.69	-1.12
			$n = 4$	4.00	7.53	-5.27	9.64	0.48	2.37	6.67	-2.46	8.44	0.44	2.24	6.94	-3.16	8.60	0.43	2.05	6.67	-1.12

Sensitivity analysis over default order picking environment

Table 5 reports only four failures of reassignment with a 1.09% increase in $\overline{FT}$ compared with the original assignment when $GPSM_{SD}(1)$ runs for 200 instances. When the stage increases, the SLR model swaps the number of product pairs equal to the stage number, whereas the GPSM policies swap fewer product pairs because the screening test filters out the swap operation with a low improvement probability. Both $GPSM_{SD}$ and $GPSM_{MM}$ significantly reduce $\overline{FT}$ and swapping time. JA swaps the unlimited number of storage locations and yields 8.21% of the average $\overline{FT}$ reduction. $GPSM_{SD}(4)$ swaps on average 2.27 times to yield 7.51% of the average $\overline{FT}$ reduction, but SLR needs to swap three times to yield 6.82% of the average $\overline{FT}$ reduction. $GPSM_{SD}$ yields 0.00% for the $\overline{FT}$ reduction of the worst cases, whereas SLR(4) and JA yield -2.22% and -8.13%, respectively. Appendix Table C2 shows that $\overline{FT}$ reduction rates of SLR(4) and JA are not significantly different from $GPSM_{SD}(4)$ and $GPSM_{MM}(4)$ before performing screening tests. The results indicate that GPSM is a reliable, cost-effective tool for warehouse management.

Sensitivity analysis over 12 zones

Table 5 also shows that reassignment performance decreases when the number of zones increases. SLR yields 2.66% of the average $\overline{FT}$ reduction with one swap operation and 4.59% of the average $\overline{FT}$ reduction with two swap operations. After the screening test, the GPSM policies skip the low improvement cases and reduce the average number of swap operations. $GPSM_{SD}$ yields 4.50% of the $\overline{FT}$ with average 1.54 swap operations, and $GPSM_{MM}$ yields 5.07% of the $\overline{FT}$ with average 1.38 swap operations. $GPSM_{SD}$ yields -3.18% for the $\overline{FT}$ reduction of the worst cases, whereas JA and SLR(4) yield -8.69% and -5.22%, respectively. Appendix Table C3 shows that $\overline{FT}$ reduction rates of SLR(4) and $GPSM_{SD}(4)$ are not significantly different before performing screening tests. The results indicate that GPSM performs well over large-scale assignments with limited training data resources.

Sensitivity analysis over 12 rack columns per zone

As the number of rack columns in each zone increases, picker's travel distance has a large effect on order picking performance. SLR yields 3.85% for the average $\overline{FT}$ reduction with 2 swap operations, and $GPSM_{SD}$ and $GPSM_{MM}$ yield 5.79% for the average $\overline{FT}$ reduction with average 2.00 swap operations and 5.26%, respectively, with 1.90 swap operations. The results indicate that both $GPSM_{SD}$ and $GPSM_{MM}$ balance the workloads well between zones when travel distance is critical.

Sensitivity analysis over order size = Uniform(10,20)

$GPSM_{SD}$ and $GPSM_{MM}$ consistently reduce the average $\overline{FT}$ from stages 1–4 and provide assignments with a high probability to improve. Intuitively, increasing the order size causes congestion because the totes stay longer in the system. JA yields 7.27% for the average $\overline{FT}$ reduction with unlimited swaps, whereas SLR yields 3.51% when swapping two pairs of products, and 4.87% when swapping three pairs of products, respectively. $GPSM_{SD}$ yields 6.75 % when swapping an average 2.83 pairs of products, whereas $GPSM_{MM}$ yields 6.31% when swapping an average 2.65 pairs of products. Appendix Table C5 shows that $\overline{FT}$ reduction rates of SLR(4), JA, $GPSM_{SD}(4)$, and $GPSM_{MM}(4)$ are not significantly different before performing screening tests. The results confirm well-balanced workloads between zones when the order size is large.

Sensitivity analysis over number of orders = 200

To analyze GPSM performance over a large problem size, we increase the number of orders in each order list. As the number of orders increases, the average $\overline{FT}$ reduction between the original and reassignment increases: $GPSM_{MM}(4)$ yields 10.30% on average and 0.00% in the worst case with the average number of swaps = 3.05, and SLR yields 9.22% on average and −1.58% in the worst case with three swap operations, whereas JA yields 13.59% on average and -7.60% in the worst case with the unlimited number of swap operations.

Sensitivity analysis over non-identical picker skill

$GPSM_{AZ}$ yields the largest $\overline{FT}$ reduction before the screening test, 2.11% and 1.90% higher than SLR(4) and JA, respectively. Despite the poor performance of $GPSM_{AZ}$ in other configurations, it does capture the zones with slow picking speeds and assigns fewer loads to them, which balances the workload in a non-identical picker-skill environment. Appendix Table C7 shows that $\overline{FT}$ reduction rates of $GPSM_{AZ}(4)$ is significantly larger than SLR(4) before performing screening tests. The results indicate that the GPSM policies effectively reassign storage to reduce $\overline{FT}$ when pickers are non-identical in skill and speed.

6. Conclusion

This paper proposed an average flow-time estimation model for estimating travel time and a storage location reassignment with risk assessment in an automated zone picking system. During the storage location reassignment procedure, warehouse management applied the screening test for each reassignment stage to assess the risk of the reassignment decision's failure. Ten scenarios from a real e-commerce warehouse configuration and historical order data were used to evaluate the performance of the proposed model.

The GPSM estimated the average flow time from the historical data with three different types of features: the number of picks in all zones, the min-max number of picks, standard deviation of the number of picks, and the standard deviation of the number of picks. Given storage location assignments, it enabled the GPSM to adjust the storage location by evaluating the minimum estimated average flow time. The GPSM identified new storage location assignments and assessed the risk of a reassignment failure with the probability of improvement.

The simulation experiments and statistical tests validated that the proposed models significantly reduced the average flow time of orders in large-scale order picking and non-identical picker skill environments, and that the screening test limited the number of swap operations.

Future research will investigate constructing the features extracted from an order list based on domain knowledge of the picking system. More analysis should confirm the accuracy of an estimation model that relies on feature extraction from large-scale order picking data with no loss of data integrity. Further investigation of the sequential sampling and kernel functions should indicate improved model performance with limited data.

Appendix A. Numerical validation

We measure the Pearson correlation coefficient (Pearson's r) of the average flow time estimation model. Figure A1 plots the correlation along with the Pearson's r value between the actual and predicted values of true average flow time for 200 different order lists generated using simulation data from industry (see Sections 5.1 and 5.2 for details of the dataset). For the three different GPSMs, Pearson's r values ranging from 0.537 to 0.950 indicate a relatively strong correlation between simulations and predictions.

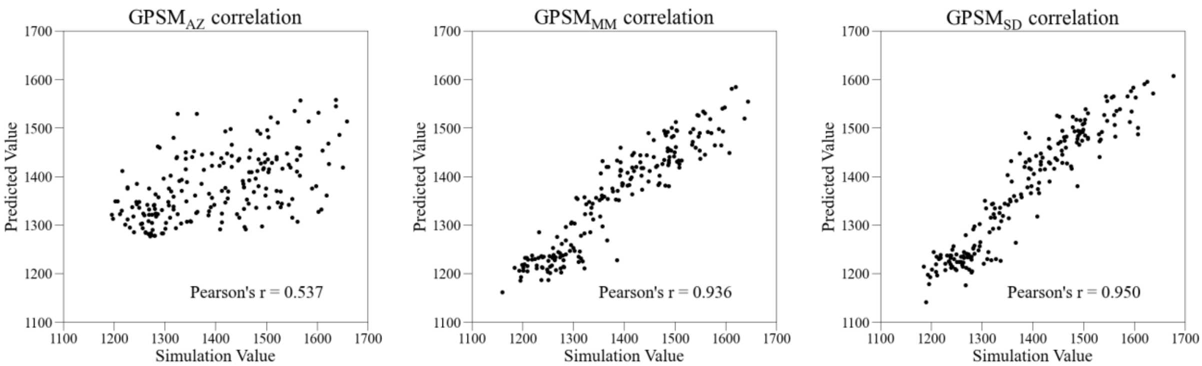

Figure A1. Correlation between the values obtained by simulations and by the three GPSMs.

We also calculate the mean absolute percentage error (MAPE) and R^2 score of the three GPSMs for the dataset. Table A1 reports the values denoting the regression accuracy of the proposed average flow-time estimation models.

Table A1. MAPE and R2 scores of GPSM.

Model	MAPE (%)	R^2
$GPSM_{AZ}$	5.85	0.26
$GPSM_{MM}$	2.74	0.79
$GPSM_{SD}$	2.26	0.89

Appendix B. Statistical analysis between workload scenarios

Table B1. Welch's ANOVA tests of performance between workload scenarios.

Source	Statistic[a]	DF Num	DF Den	P-value
$\overline{FT}$	262.032	9	199.157	< 0.001

[a] Asymptotically F distributed

Table B2. Multiple comparisons of mean differences in the average $\overline{FT}$ between workload scenarios based on Games-Howell post-hoc tests.

Scenario	1	2	3	4	5	6	7	8	9	10
1		-59.445*	-190.629*	-273.590*	-263.397*	-343.676*	-272.955*	-336.721*	-34.219*	-62.487*
2			-131.180*	-214.141*	-203.948*	-284.226*	-213.505*	-277.271*	25.231	-3.037
3				-82.961*	-72.768*	-153.046*	-82.326*	-146.091*	156.410*	128.143*
4					10.193	-70.086*	0.635	-63.130*	239.371*	211.104*
5						-80.279*	-9.558	-73.323*	229.178*	200.911*
6							70.721*	6.955	309.457*	281.189*
7								-63.766*	238.736*	210.468*
8									302.501*	274.234*
9										-28.267
10										

* Mean difference significant at the 0.05 level

Appendix C. Statistical analysis between methods

Table C1. Welch's ANOVA tests of performance between methods.

Environment	Stage	Statistic[a]	DF Num	DF Den	*P*-value
default	$n = 1$	75.511	5	524.212	< 0.001
	$n = 2$	51.969	5	531.754	< 0.001
	$n = 3$	40.901	5	537.173	< 0.001
	$n = 4$	42.488	5	542.384	< 0.001
12 zones	$n = 1$	100.042	5	526.994	< 0.001
	$n = 2$	78.095	5	534.856	< 0.001
	$n = 3$	59.942	5	539.904	< 0.001
	$n = 4$	48.484	5	543.359	< 0.001
12 rack columns	$n = 1$	49.255	5	509.983	< 0.001
	$n = 2$	33.982	5	514.926	< 0.001
	$n = 3$	26.549	5	519.378	< 0.001
	$n = 4$	24.391	5	523.325	< 0.001
Unif(10,20) Order size	$n = 1$	94.673	5	507.882	< 0.001
	$n = 2$	70.097	5	514.426	< 0.001
	$n = 3$	53.633	5	521.611	< 0.001
	$n = 4$	44.641	5	527.161	< 0.001
200 orders in each order list	$n = 1$	178.769	5	495.630	< 0.001
	$n = 2$	137.922	5	503.428	< 0.001
	$n = 3$	105.377	5	510.265	< 0.001
	$n = 4$	86.137	5	517.059	< 0.001
Non-identical picker skill	$n = 1$	47.813	5	510.657	< 0.001
	$n = 2$	28.858	5	518.685	< 0.001
	$n = 3$	20.066	5	527.084	< 0.001
	$n = 4$	20.938	5	535.300	< 0.001

[a] Asymptotically F distributed

Table C2. Multiple comparisons of mean differences in the average $\overline{FT}$ in the default environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(n)	GPSM$_{AZ}$(n)	GPSM$_{MM}$(n)	GPSM$_{SD}$(n)
$n = 1$	Original		129.097*	44.031*	29.906	45.268*	48.389*
	JA			-85.065*	-99.190*	-83.828*	-80.708*
	SLR(n)				-14.125	1.237	4.358
	GPSM$_{AZ}$(n)					15.362	18.483
	GPSM$_{MM}$(n)						3.121
	GPSM$_{SD}$(n)						
$n = 2$	Original		129.097*	79.480*	47.641*	79.993*	83.985*
	JA			-49.617*	-81.456*	-49.104*	-45.112*
	SLR(n)				-31.838*	0.513	4.505
	GPSM$_{AZ}$(n)					32.351*	36.344*
	GPSM$_{MM}$(n)						3.992
	GPSM$_{SD}$(n)						
$n = 3$	Original		129.097*	102.719*	60.490*	109.162*	114.303*
	JA			-26.378*	-68.606*	-19.934	-14.794
	SLR(n)				-42.228*	6.444	11.584
	GPSM$_{AZ}$(n)					48.672*	53.812*
	GPSM$_{MM}$(n)						5.140
	GPSM$_{SD}$(n)						
$n = 4$	Original		129.097*	121.378*	68.870*	126.214*	134.113*
	JA			-7.719	-60.226*	-2.883	5.017
	SLR(n)				-52.507*	4.836	12.735
	GPSM$_{AZ}$(n)					57.344*	65.243*

	GPSM$_{MM}$(*n*)						7.899
	GPSM$_{SD}$(*n*)						

* Mean difference significant at the 0.05 level

Table C3. Multiple comparison of mean differences in the average $\overline{FT}$ in the 12 zones environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(*n*)	GPSM$_{AZ}$(*n*)	GPSM$_{MM}$(*n*)	GPSM$_{SD}$(*n*)
n = 1	Original		142.240*	39.719*	30.324	40.209*	35.453
	JA			-102.521*	-111.916*	-102.031*	-106.787*
	SLR(*n*)				-9.395	0.491	-4.266
	GPSM$_{AZ}$(*n*)					9.885	5.129
	GPSM$_{MM}$(*n*)						-4.756
	GPSM$_{SD}$(*n*)						
n = 2	Original		142.240*	68.487*	45.252*	68.572*	62.562*
	JA			-73.753*	-96.988*	-73.668*	-79.677*
	SLR(*n*)				-23.235	0.085	-5.924
	GPSM$_{AZ}$(*n*)					23.320	17.310
	GPSM$_{MM}$(*n*)						-6.010
	GPSM$_{SD}$(*n*)						
n = 3	Original		142.240*	89.270*	65.932*	86.845*	83.713*
	JA			-52.970*	-76.308*	-55.395*	-58.527*
	SLR(*n*)				-23.338	-2.425	-5.557
	GPSM$_{AZ}$(*n*)					20.913	17.782
	GPSM$_{MM}$(*n*)						-3.132
	GPSM$_{SD}$(*n*)						
n = 4	Original		142.240*	108.714*	80.106*	104.562*	96.093*
	JA			-33.526*	-62.134*	-37.678*	-46.147*
	SLR(*n*)				-28.608*	-4.153	-12.622
	GPSM$_{AZ}$(*n*)					24.455*	15.986
	GPSM$_{MM}$(*n*)						-8.469
	GPSM$_{SD}$(*n*)						

* Mean difference significant at the 0.05 level

Table C4. Multiple comparisons of mean differences in the average $\overline{FT}$ in the 12 rack columns environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(*n*)	GPSM$_{AZ}$(*n*)	GPSM$_{MM}$(*n*)	GPSM$_{SD}$(*n*)
n = 1	Original		144.657*	42.460	29.165	52.112	49.528
	JA			-102.196*	-115.492*	-92.545*	-95.129*
	SLR(*n*)				-13.295	9.652	7.068
	GPSM$_{AZ}$(*n*)					22.947	20.363
	GPSM$_{MM}$(*n*)						-2.584
	GPSM$_{SD}$(*n*)						
n = 2	Original		144.657*	72.334*	54.881*	87.701*	91.614*
	JA			-72.323*	-89.776*	-56.956*	-53.044*
	SLR(*n*)				-17.453	15.367	19.279
	GPSM$_{AZ}$(*n*)					32.821	36.733
	GPSM$_{MM}$(*n*)						3.912
	GPSM$_{SD}$(*n*)						
n = 3	Original		144.657*	94.480*	68.615*	120.754*	125.059*
	JA			-50.177*	-76.042*	-23.904	-19.598
	SLR(*n*)				-25.865	26.274	30.579
	GPSM$_{AZ}$(*n*)					52.138*	56.444*
	GPSM$_{MM}$(*n*)						4.306

Stage	Method	Original	JA	SLR(n)	GPSM$_{AZ}$(n)	GPSM$_{MM}$(n)	GPSM$_{SD}$(n)
	GPSM$_{SD}$(n)						
n = 4	Original		144.657*	113.849*	85.417*	150.018*	154.481*
	JA			-30.808	-59.240*	5.361	9.823
	SLR(n)				-28.432	36.169	40.632
	GPSM$_{AZ}$(n)					64.601*	69.064*
	GPSM$_{MM}$(n)						4.462
	GPSM$_{SD}$(n)						

* Mean difference significant at the 0.05 level

Table C5. Multiple comparisons of mean differences in the average $\overline{FT}$ in the *Unif*(10,20) order size environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(n)	GPSM$_{AZ}$(n)	GPSM$_{MM}$(n)	GPSM$_{SD}$(n)
n = 1	Original		216.892*	54.461	41.753	68.556*	71.486*
	JA			-162.431*	-175.139*	-148.336*	-145.406*
	SLR(n)				-12.708	14.095	17.025
	GPSM$_{AZ}$(n)					26.803	29.733
	GPSM$_{MM}$(n)						2.930
	GPSM$_{SD}$(n)						
n = 2	Original		216.892*	100.288*	73.472*	116.596*	123.383*
	JA			-116.604*	-143.419*	-100.296*	-93.509*
	SLR(n)				-26.815	16.308	23.095
	GPSM$_{AZ}$(n)					43.124	49.910
	GPSM$_{MM}$(n)						6.787
	GPSM$_{SD}$(n)						
n = 3	Original		216.892*	138.916*	98.224*	157.297*	164.920*
	JA			-77.976*	-118.668*	-59.595	-51.971*
	SLR(n)				-40.692	18.381	26.004
	GPSM$_{AZ}$(n)					59.073*	66.696*
	GPSM$_{MM}$(n)						7.623
	GPSM$_{SD}$(n)						
n = 4	Original		216.892*	171.478*	116.889*	191.426*	202.302*
	JA			-45.414*	-100.002*	-25.466	-14.589
	SLR(n)				-54.589*	19.948	30.824
	GPSM$_{AZ}$(n)					74.537*	85.413*
	GPSM$_{MM}$(n)						10.876
	GPSM$_{SD}$(n)						

* Mean difference significant at the 0.05 level

Table C6. Multiple comparisons of mean differences in the average $\overline{FT}$ in 200 orders in the order list environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(n)	GPSM$_{AZ}$(n)	GPSM$_{MM}$(n)	GPSM$_{SD}$(n)
n = 1	Original		235.929*	69.321*	53.654	67.399*	62.195*
	JA			-166.607*	-182.274*	-168.530*	-173.734*
	SLR(n)				-15.667	-1.923	-7.126
	GPSM$_{AZ}$(n)					13.744	8.541
	GPSM$_{MM}$(n)						-5.204
	GPSM$_{SD}$(n)						
n = 2	Original		235.929*	118.812*	94.709*	116.800*	112.905*
	JA			-117.117*	-141.220*	-119.129*	-123.024*
	SLR(n)				-24.103	-2.012	-5.907
	GPSM$_{AZ}$(n)					22.091	18.196
	GPSM$_{MM}$(n)						-3.895
	GPSM$_{SD}$(n)						

n = 3	Original		235.929*	155.636*	118.416*	154.732*	153.536*
	JA			-80.292*	-117.518*	-81.196*	-82.393*
	SLR(n)				-37.220*	-0.904	-2.101
	GPSM$_{AZ}$(n)					36.316*	35.120
	GPSM$_{MM}$(n)						-1.197
	GPSM$_{SD}$(n)						
n = 4	Original		235.929*	182.875*	132.236*	183.524*	184.098*
	JA			-53.054*	-103.693*	-52.405*	-51.830*
	SLR(n)				-50.639*	0.649	1.224
	GPSM$_{AZ}$(n)					51.288*	51.863*
	GPSM$_{MM}$(n)						0.575
	GPSM$_{SD}$(n)						

* Mean difference significant at the 0.05 level

Table C7. Multiple comparisons of mean differences in the average $\overline{FT}$ in the non-identical picker skill environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(n)	GPSM$_{AZ}$(n)	GPSM$_{MM}$(n)	GPSM$_{SD}$(n)
n = 1	Original		147.011*	49.026	61.845*	50.186	52.989
	JA			-97.985*	-85.166*	-96.825*	-94.022*
	SLR(n)				12.819	1.160	3.962
	GPSM$_{AZ}$(n)					-11.659	-8.856
	GPSM$_{MM}$(n)						2.803
	GPSM$_{SD}$(n)						
n = 2	Original		147.011*	84.656*	106.695*	91.732*	94.720*
	JA			-62.356*	-40.316*	-55.279*	-52.291*
	SLR(n)				22.040	7.076	10.064
	GPSM$_{AZ}$(n)					-14.963	-11.975
	GPSM$_{MM}$(n)						2.988
	GPSM$_{SD}$(n)						
n = 3	Original		147.011*	114.430*	141.455*	125.281*	127.133*
	JA			-32.581*	-5.556	-21.730	-19.878
	SLR(n)				27.025	10.851	12.703
	GPSM$_{AZ}$(n)					-16.174	-14.322
	GPSM$_{MM}$(n)						1.852
	GPSM$_{SD}$(n)						
n = 4	Original		147.011*	134.200*	167.025*	148.532*	151.843*
	JA			-12.811	20.014	1.521	4.832
	SLR(n)				32.825*	14.332	17.643
	GPSM$_{AZ}$(n)					-18.494	-15.182
	GPSM$_{MM}$(n)						3.311
	GPSM$_{SD}$(n)						

* Mean difference significant at the 0.05 level

References

Bartholdi, J., & Eisenstein, D. (1996). Bucket brigades: A self-organizing order picking system for a warehouse. *Report, School of Industrial Engineering, Georgia Tech, Atlanta, USA*.

Bartholdi, J. J., & Hackman, S. T. (2008). *Warehouse & Distribution Science: Release 0.89*. Supply Chain and Logistics Institute Atlanta.

Bect, J., Ginsbourger, D., Li, L., Picheny, V., & Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, *22*(3), 773-793. https://doi.org/10.1007/s11222-011-9241-4

Brynzér, H., & Johansson, M. I. (1996). Storage location assignment: Using the product structure to reduce order picking times. *International Journal of Production Economics*, *46-47*, 595-603. https://doi.org/10.1016/0925-5273(94)00091-3

Cahyo, S. D. (2017). *Analysis of Shortage Delay in a Zone Picking System with AS/RS Replenishment* [Master's thesis, Pusan National University]. http://lib.pusan.ac.kr/en/en/resource/catalog/?app=solars&mod=detail&record_id=93752046

Caron, F., Marchet, G., & Perego, A. (2010). Optimal layout in low-level picker-to-part systems. *International Journal of Production Research*, *38*(1), 101-117. https://doi.org/10.1080/002075400189608

Chiang, D. M.-H., Lin, C.-P., & Chen, M.-C. (2011). The adaptive approach for storage assignment by mining data of warehouse management system for distribution centres. *Enterprise Information Systems*, *5*(2), 219-234. https://doi.org/10.1080/17517575.2010.537784

Chiang, D. M.-H., Lin, C.-P., & Chen, M.-C. (2014). Data mining based storage assignment heuristics for travel distance reduction. *Expert Systems*, *31*(1), 81-90. https://doi.org/10.1111/exsy.12006

Chuang, Y.-F., Lee, H.-T., & Lai, Y.-C. (2012). Item-associated cluster assignment model on storage allocation problems. *Computers & Industrial Engineering*, *63*(4), 1171-1177. https://doi.org/10.1016/j.cie.2012.06.021

De Koster, R., Le-Duc, T., & Zaerpour, N. (2012). Determining the number of zones in a pick-and-sort order picking system. *International Journal of Production Research*, *50*(3), 757-771. https://doi.org/10.1080/00207543.2010.543941

De Koster, R. B. M., Le-Duc, T., & Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. *European Journal of Operational Research*, *182*(2), 481-501.

Gaast, J. P. v. d., Koster, R. B. M. d., Adan, I. J. B. F., & Resing, J. A. C. (2020). Capacity Analysis of Sequential Zone Picking Systems. *Operations Research*, *68*(1), 161-179. https://doi.org/10.1287/opre.2019.1885

Ghomri, L., & Sari, Z. (2017). Mathematical modeling of the average retrieval time for flow-rack automated storage and retrieval systems. *Journal of Manufacturing Systems*, *44*, 165-178. https://doi.org/10.1016/j.jmsy.2017.05.002

Gu, J., Goetschalckx, M., & McGinnis, L. F. (2007). Research on warehouse operation: A comprehensive review. *European Journal of Operational Research*, *177*(1), 1-21. https://doi.org/10.1016/j.ejor.2006.02.025

Hong, S. (2019). A performance evaluation of bucket brigade order picking systems: Analytical and simulation approaches. *Computers & Industrial Engineering*, *135*, 120-131. https://doi.org/10.1016/j.cie.2019.05.037

Huang, M., Guo, Q., Liu, J., & Huang, X. (2018). Mixed Model Assembly Line Scheduling Approach to Order Picking Problem in Online Supermarkets. *Sustainability*, *10*(11), 3931. https://www.mdpi.com/2071-1050/10/11/3931

Jane, C.-C., & Laih, Y.-W. (2005). A clustering algorithm for item assignment in a synchronized zone order picking system. *European Journal of Operational Research*, *166*(2), 489-496. https://doi.org/10.1016/j.ejor.2004.01.042

Jane, C. C. (2000). Storage location assignment in a distribution center. *International Journal of Physical Distribution & Logistics Management*, *30*(1), 55-71. https://doi.org/10.1108/09600030010307984

Jewkes, E., Lee, C., & Vickson, R. (2004). Product location, allocation and server home base location for an order picking line with multiple servers. *Computers & Operations Research*, *31*(4), 623-636. https://doi.org/10.1016/S0305-0548(03)00035-2

Kim, J.-h., & Hong, S. (2020). A dynamic storage location assignment model for a progressive bypass zone picking system with an S/R crane. In P. N. University (Ed.).

Kim, J., & Hong, S. (2022). A dynamic storage location assignment model for a progressive bypass zone picking system with an S/R crane. *Journal of the Operational Research Society*, *73*(5), 1155-1166. https://doi.org/10.1080/01605682.2021.1892462

Kübler, P., Glock, C. H., & Bauernhansl, T. (2020). A new iterative method for solving the joint dynamic storage location assignment, order batching and picker routing problem in manual picker-to-parts warehouses. *Computers & Industrial Engineering*, *147*, 106645. https://doi.org/10.1016/j.cie.2020.106645

Larco, J. A., de Koster, R., Roodbergen, K. J., & Dul, J. (2016). Managing warehouse efficiency and worker discomfort through enhanced storage assignment decisions. *International Journal of Production Research*, *55*(21), 6407-6422. https://doi.org/10.1080/00207543.2016.1165880

Le-Duc, T., & De Koster, R. B. M. (2005). Travel distance estimation and storage zone optimization in a 2-block class-based storage strategy warehouse. *International Journal of Production Research*, *43*(17), 3561-3581. https://doi.org/10.1080/00207540500142894

Li, J., Moghaddam, M., & Nof, S. Y. (2016). Dynamic storage assignment with product affinity and ABC classification—a case study. *The International Journal of Advanced Manufacturing Technology*, *84*(9-12), 2179-2194. https://doi.org/10.1007/s00170-015-7806-7

Mackay, D. J. C. (1998). Introduction to Gaussian processes. *NATO ASI series. Series F: computer and system sciences*, 133-165.

Muppani, V. R., & Adil, G. K. (2008). A branch and bound algorithm for class based storage location assignment. *European Journal of Operational Research*, *189*(2), 492-507. https://doi.org/10.1016/j.ejor.2007.05.050

Murphy, K. P. (2012). *Machine learning : a probabilistic perspective*. MIT Press. http://mitpressebooks.mit.edu/product/machine-learning

Pan, J. C.-H., Shih, P.-H., & Wu, M.-H. (2015). Order batching in a pick-and-pass warehousing system with group genetic algorithm. *Omega*, *57*, 238-248. https://doi.org/https://doi.org/10.1016/j.omega.2015.05.004

Pan, J. C.-H., Shih, P.-H., Wu, M.-H., & Lin, J.-H. (2015). A storage assignment heuristic method based on genetic algorithm for a pick-and-pass warehousing system. *Computers & Industrial Engineering*, *81*, 1-13. https://doi.org/10.1016/j.cie.2014.12.010

Pan, J. C.-H., Wu, M.-H., & Chang, W.-L. (2014). A travel time estimation model for a high-level picker-to-part system with class-based storage policies. *European Journal of Operational Research*, *237*(3), 1054-1066. https://doi.org/10.1016/j.ejor.2014.02.037

Pan, J. C. H., & Wu, M. H. (2009). A study of storage assignment problem for an order picking line in a pick-and-pass warehousing system. *Computers & Industrial Engineering*, *57*(1), 261-268. https://doi.org/10.1016/j.cie.2008.11.026

Pang, K.-W., & Chan, H.-L. (2016). Data mining-based algorithm for storage location assignment in a randomised warehouse. *International Journal of Production Research*, *55*(14), 4035-4052. https://doi.org/10.1080/00207543.2016.1244615

Pazour, J. A., & Carlo, H. J. (2015). Warehouse reshuffling: Insights and optimization. *Transportation Research Part E: Logistics and Transportation Review*, *73*, 207-226. https://doi.org/10.1016/j.tre.2014.11.002

Quintanilla, S., Pérez, Á., Ballestín, F., & Lino, P. (2014). Heuristic algorithms for a storage location assignment problem in a chaotic warehouse. *Engineering Optimization*, *47*(10), 1405-1422. https://doi.org/10.1080/0305215x.2014.969727

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press. Table of contents only http://www.loc.gov/catdir/toc/fy0614/2005053433.html

Richardson, R. R., Osborne, M. A., & Howey, D. A. (2017). Gaussian process regression for forecasting battery state of health. *Journal of Power Sources*, *357*, 209-219. https://doi.org/10.1016/j.jpowsour.2017.05.004

Roodbergen, K. J., & Vis, I. F. A. (2009). A survey of literature on automated storage and retrieval systems. *European Journal of Operational Research*, *194*(2), 343-362. https://doi.org/10.1016/j.ejor.2008.01.038

Sadiq, M., Landers, T. L., & Don Taylor, G. (1996). An Assignment Algorithm for Dynamic Picking Systems. *IIE Transactions*, *28*(8), 607-616. https://doi.org/10.1080/15458830.1996.11770706

Vanheusden, S., van Gils, T., Braekers, K., Ramaekers, K., & Caris, A. (2022). Analysing the effectiveness of workload balancing measures in order picking operations. *International Journal of Production Research*, *60*(7), 2126-2150. https://doi.org/10.1080/00207543.2021.1884307

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32-35.

Gaussian Process-based Storage Location Assignments with Risk Assessments for Progressive Zone Picking Systems

Abstract

E-commerce warehouses are under constant pressure to adapt their order picking systems and reassign product storage locations to meet fluctuating customer demands. Most existing approaches optimize storage location reassignments based on customer orders and operational configurations to maintain high order picking performance. This paper presents a Gaussian process surrogate model (GPSM) approach to predict the performance metrics for storage location reassignments. The GPSM estimates the expected flow time of orders from the historical data on previous storage location assignments and aids in identifying the new assignments that yield the minimum estimated average flow times. Management can also take advantage of the GPSM's uncertainty quantification capability to assess the probability of improvement for a given storage reassignment and its implementation. The proposed model and assignment policy are validated using discrete-event simulations and industrial data. Experimental results demonstrate that the GPSM can improve expected flow time by 7.51% and reduce unnecessary reassignment operations by 43.25%.

Keywords: Facilities planning and design; Zone order picking; Storage location assignment; Gaussian process approach.

1. Introduction

Zone picking (ZP) systems are popular owing to their high-throughput capability, flexibility to handle small and large order volumes, and adaptability to a wide range of product sizes with a variety of order pickers (Gaast et al., 2020). By separating an order picking area into zones and assigning one picker per zone, ZP systems can reduce travel time and congestion among pickers (De Koster et al., 2007).

In ZP systems, performance bottlenecks can still occur due to workload imbalances or other reasons. Therefore ZP comes with order picking planning policies, such as order batching (Pan, Shih, & Wu, 2015), order sequencing (Huang et al., 2018), and storage assignment (Jane, 2000). Order batching reduces order picking variability and mitigates workload gaps between zones by grouping orders so that each zone's workload is evenly distributed between batches. Order sequencing adjusts tote release sequences to improve the distribution of workload balances. Storage assignment balances workloads over zones by assigning products to each zone during a long picking time. Optimizing product assignments is referred to as the storage location assignment problem (SLAP) (Gu et al., 2007).

This paper investigates the SLAP for a progressive bypass ZP flow-rack system with an automated storage/retrieval (S/R) crane. The flow-rack system is popular in warehouse management systems for e-commerce businesses because it handles small and frequent orders and daily demand fluctuations under short flow times. In many warehouses, since assignment capabilities are limited in terms of resources and time, the SLAP problem is optimized by reassigning selected products instead of reallocating all products in an order. As illustrated in Figure 1, prior to order picking, the S/R crane relocates a few storage locations of products between the zones to rebalance the workloads (Roodbergen & Vis, 2009). We study the optimization of the swap operations considering the current assignment of products and estimates of the expected workload over daily orders.

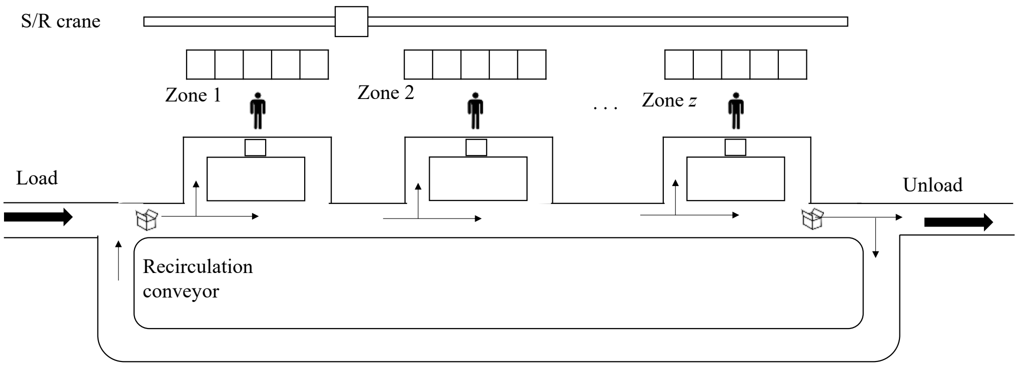

Figure 1. Progressive bypass ZP system with an S/R crane (Kim and Hong (2022)).

Solutions to the SLAP use workload proxies to approximate current workloads and exploit them for a simple rebalancing of the estimated workloads. Popular workload proxies include the number of picks (Jane & Laih, 2005; Jewkes et al., 2004; Kim & Hong, 2022), the expected walk time from a picker's loading depot (Jewkes et al., 2004), picker's skill level (Bartholdi & Eisenstein, 1996), and picker's route distance (De Koster et al., 2012). However, it is still unclear whether the proxies accurately represent workloads, and it is questionable whether using the proxies is effective in optimizing an order picking performance metric such as the average flow time of orders. In this paper, we propose to use a data driven approach to directly estimate the order picking performance metric as a function of current workloads, and we exploit the data-driven estimation to solve SLAP.

To our knowledge, only a few studies have used a data-driven approach to the workload balancing and storage location reassignment problems. Building an analytical model to estimate order picking performance from order picking data has been challenging due to the noise-included uncertainty of complex ZP systems. We develop a Gaussian process surrogate model (GPSM) for a data-driven estimation of the average flow time of orders from the historical data of previous storage location assignments. Surrogate modeling usually uses the Gaussian process, due to good analytical inference, computational flexibility, and straightforward uncertainty quantification (Rasmussen & Williams, 2006). GPSM can measure whether a new storage assignment is likely to improve order picking performance by means of the posterior prediction uncertainty. Based on the GPSM, we develop a storage location reassignment method that sequentially runs one swap operation at a time and suggests the reliable optimal swap operation with respect to the GPSM's average flow time estimate.

Contributions of this study are summarized as follows:

- **Model-driven approach vs. data-driven approach.** We directly estimate order picking performance with a data-driven approach by implementing the GPSM. Since estimating order picking performance is challenging due to warehouse complexity, size, and noise in the data, implicit workload proxies have been used to estimate performance in traditional methods (De Koster et al., 2012; Jane & Laih, 2005; Jane, 2000; Jewkes et al., 2004; Kim & Hong, 2020). Instead of approximating workload proxies, we optimize storage location assignments by using historical order picking data to estimate the average flow time with GPSM.
- **Domain knowledge-based performance improvement.** We improve the GPSM's estimation performance with our domain specific feature generation and training data configuration. As a result, we present an accurate learning model that effectively estimates order picking performance to solve SLAP.
- **Risk assessment using GPSM.** We propose a screening test based on the improvement probability obtained from the GP estimates and the storage reassignment procedure based on the estimation model and screening test The procedure optimizes storage space with minimal adjustments, reduces material handling costs, and enhances customer service in the warehouse.
- **Performance evaluation.** We use discrete event simulation and statistical tests to demonstrate the proposed models' ability to reduce flow time and reassignment effort, thereby improving warehouse management systems and operational efficiency.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 explains the fulfillment system and the proposed GPSM for average flow-time estimation. Section 4 introduces the proposed storage reassignment method using the flow-time estimation. Section 5 describes the experiment validating the model, and the simulations, results, and sensitivity analysis. Section 6 concludes and suggests future research.

2. Literature review

2.1 General storage location assignment

Most studies of the SLAP use travel distance and travel time as the performance measurements. Caron et al. (2010) developed an optimization model that minimizes the travel distance and travel time, and they also developed an estimation model that estimates the two performance measurements. Brynzér and Johansson (1996) suggested that grouping products by characterizing their variant information could reduce the product movements for traveling and picking. Muppani and Adil (2008) used nonlinear integer programming to capture the impact of a class-based storage system on the required space and material handling cost. Several studies have used probabilistic models and Markov chains to estimate travel distance or travel time to evaluate order picking performance and analyze the optimal storage assignment policy (Le-Duc & De Koster, 2005; Pan et al., 2014; Pan & Wu, 2009).

Motivated by a real problem, SLAP studies have also considered controlling the number of reassignment operations. Quintanilla et al. (2014) developed heuristic algorithms to maximize the available storage space by reoptimizing pallet locations in a random storage system. Kübler et al. (2020) suggested the ABC class-based iterative storage reassignment method considering order batching and picker routing together, and evaluated the reassignment effort with the future travel distance to identify promising reassignment. Pazour and Carlo (2015) developed mathematical model formulations for reassignment operating policies using an automated S/R crane and quantified the total loaded and unloaded travel distances while optimizing the reassignment operations.

In progressive zone picking systems, the SLAP has been focused on balancing workloads between zones for operational efficiency. Studies of workload balancing include Jane (2000), who proposed a heuristic algorithm to balance pickers' workloads by adjusting the number of storage zones. Pan, Shih, Wu, et al. (2015) developed a heuristic based on a genetic algorithm to solve the SLAP considering workload balance in a progressive zone picking system. Kim and Hong (2022) used mixed-integer programming to construct the storage location reassignment (SLR) model and applied it to a progressive bypass ZP system with a circulation conveyor and an S/R crane. The authors' model relocated the storage locations of products considering workload balance and recirculation reduction.

The models above, however, require solver tools, experts who can adapt the model to the warehouse environment, and lengthy computation to solve the kinds of large-size problems encountered by e-commerce warehouses. Most model-based optimization assumes a static operational environment, while data-driven optimization allows solving SLAP in an uncertain environment. Thus, we focus on data-driven storage location assignments for more agile and flexible SLAP optimization.

2.2 Data-driven storage location assignments

Various types of data-driven approaches have been developed for SLAP, and several studies have utilized clustering algorithms. Jane and Laih (2005) suggested a clustering algorithm to distribute the frequently requested products into several zones for workload balance in a synchronized zone picking system. Chuang et al. (2012) clustered associated items into groups and determined the sequence of the order groups for SLAP to minimize picking distance.

Chiang et al. (2011) used a data association algorithm to group products into similarity groups by order frequency and other product characteristics and introduced a data-mining based storage assignment (DMSA), which aimed to increase the association index (AIX) between products and their storage locations. Chiang et al. (2014) extended their research using a weighted support count (WSC) to calculate each AIX. The heuristic considered the relationship between a family and a cluster of products.

Pang and Chan (2016) developed a data mining-based assignment using association rules to minimize travel distances by controlling storage locations of correlated items and items near entry points. In a dynamic environment, Li et al. (2016) optimized the storage assignment based on the ABC classification and mutual affinity of products. The product affinity-based heuristic (PABH) identified the relationship between products. The authors used a greedy genetic algorithm because the problem was a quadratic assignment problem.

Regression techniques have been investigated for their use in storage location assignments. Sadiq et al. (1996) built a regression model to analyze the performance of storage location assignments considering order picking time and reassignment. Larco et al. (2016) used linear regression models to estimate worker discomfort factor and order cycle times. Larco et al. (2016) solved SLAP using the estimated values as the parameters of bi-objective optimization. Larco et al. (2016); Sadiq et al. (1996) estimated order picking performance with simple linear models based on factors such as storage location, bin type, product life cycle, and management policies, but neglected workload balance factors, a key influence on the performance of ZP systems.

Two practical issues deserve further investigation. Previous studies of data-driven SLAP consider the allocation of storage locations for entire products, yet when real-world reassignments are both resource- and cost-intensive, many warehouses choose to reallocate a limited range of products due to demand and operational uncertainty. The reassignment benefit should outweigh the reassignment cost while considering variables such as order sequence, workload balance, and rack storage policy. We use the Gaussian process to solve the data-driven SLAP with limited reassignment capacity and to quantify the reassignment risk to screen for reassignment operations with high potential for improvement.

3. Progressive ZP system and GPSM

In this section, we describe the progressive ZP system and explain the average flow time estimation using GPSM. We consider order picking in the concept of *wave management*, where a large set of orders is scheduled to be picked during a time period (*pick wave*) (Bartholdi & Hackman, 2008). We define the set of orders in a *pick wave* as an order list. Prior to the *pick wave*, there is an opportunity to reassign the storage location of items within the picking zones. The flow time of order refers to the time interval between the time a tote enters to the time it exits the ZP system. Flow time is a critical performance metric that needs to be reduced to shorten in-transit inventory and ensure fast delivery (Bartholdi & Hackman, 2008). According to practitioners, when considering throughput time or other indicators, it is challenging to operate a smooth transition to a post-process (e.g., packing), so the average flow time of orders is considered. We propose a learning model that estimates the average flow times of upcoming order lists based on workload balance information from historical order picking data.

3.1 Progressive ZP system characteristics

Our progressive zone picking system consists of an S/R crane and multiple zones. S/R cranes are automated material handling systems that are used in numerous manufacturing and warehousing systems to handle, store and retrieve discrete products (Ghomri & Sari, 2017). The S/R crane transfers each carton of products from a reserve rack to a flow rack, between reserve racks, or from the entrance or exit of the fulfillment center to any racks. The S/R crane can exchange the storage locations of a pair of products at a time by temporarily placing the products to be exchanged in the reserve rack area, which we refer to as a swap operation. Figure 2 shows an S/R crane replenishing a flow rack area with cartons from the reserve rack area.

Each zone has a flow rack and one zone-dedicated picker. Pickers pick products from cells in the flow rack and place them in totes. The totes travel on a conveyor through the zones. Each tote visits only the zones containing the products to be picked. When a zone has no room for a tote to enter, the tote skips over (bypasses) the zone and keeps traveling on the conveyor until a room is available.

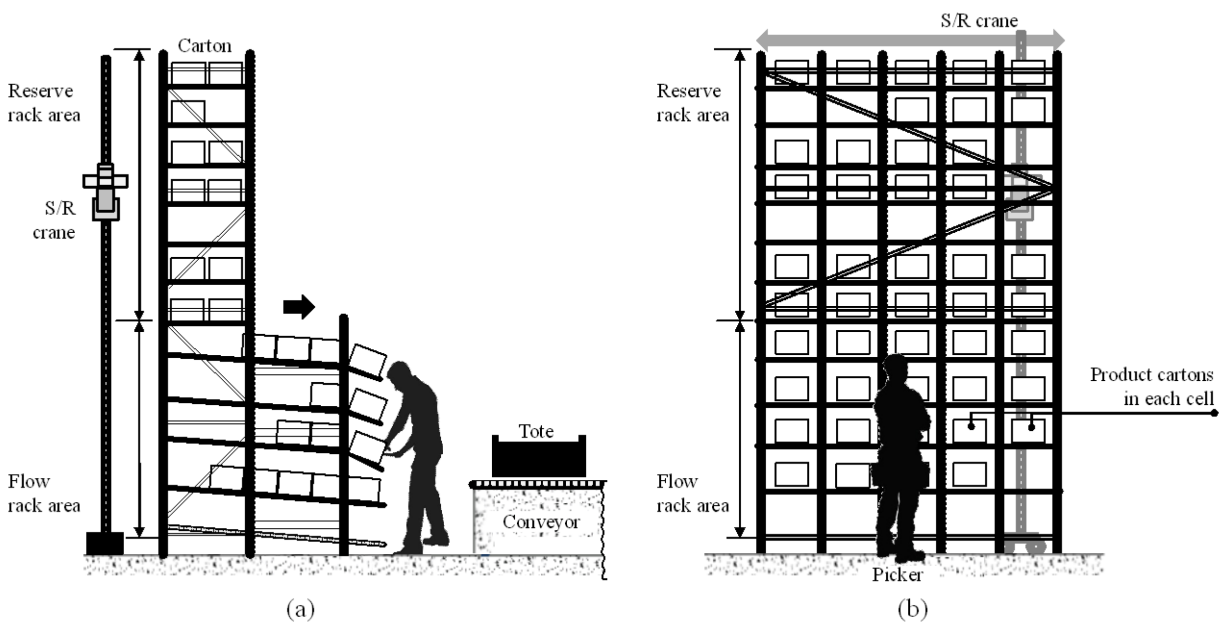

Figure 2. Progressive ZP flow-rack system with S/R crane (a) side view and (b) front view (modified from Cahyo (2017)).

A zone consists of multiple flow racks, and a flow rack consists of multiple cells to store products. We consider a simplified model, i.e., each cell contains only one product type. An order list is the batch of orders handled in the same shift or time window, and an order is the types and amounts of products.

We use the following notations.

Set and indices

O, o	The set of orders, and an order $o \in O$
L, l	The set of order lists, and an order list $l \in L$
Z, z	The set of zones, and a zone z $\in Z$
K, k	The set of cells, and a cell $k \in K$
P, p	The set of products, and a product $p \in P$
R, r	The set of relative rack column positions in a zone, and a rack position $r \in R$

Order and storage location representation

Let **C** be the $|O| \times |P|$ matrix representing products in orders with c_{op} as its (o, p) th element, where c_{op} be a binary variable indicating whether order o contains product p; c_{op} is 1 if order o contains product p, and 0 otherwise. Let **S** denote a $|P| \times |K|$ matrix describing the relationship between the products and cells with its (p, k) th element s_{pk} being a binary variable indicating whether product p is located on cell k. Let **W** be a $|O| \times |K|$ matrix representing the relationship between the orders and cells with its (o, k) th element w_{ok} being a binary variable indicating whether order o requires a product to be picked from cell k. The matrix **W** is related to matrices **C** and **S** through

$$\mathbf{W} = \mathbf{C} \cdot \mathbf{S}$$

$$= \begin{bmatrix} c_{11} & \cdots & c_{1|P|} \\ \vdots & \ddots & \vdots \\ c_{|O|1} & \cdots & c_{|O||P|} \end{bmatrix} \cdot \begin{bmatrix} s_{11} & \cdots & s_{1|K|} \\ \vdots & \ddots & \vdots \\ s_{|P|1} & \cdots & s_{|P||K|} \end{bmatrix}.$$

Abstract representation by order list and Zone/Rack locations

Let **A** be a $|K| \times |L|$ matrix of $a_{lk}'s$ that represents the total number of picks from cell $k \in K$ required to fill order list $l \in L$. Let **B** represent a $|L| \times |Z|$ matrix with b_{lz} as its (l, z) th element which represents the total number of picks from zone $z \in Z$ to fill order list$l \in L$. Let **G** be a $|L| \times |R|$ matrix of $g_{lr}'s$ that represents the total number of picks from flow rack $r \in R$ required to fill order list $l \in L$. Let O_l denote the set of orders in the order list l. Let K_z and K_r denote the set of cells assigned in zone z and the set of cells assigned in rack position r, respectively. We can calculate a_{lk} using the matrix **W** as

$$a_{lk} = \sum_{o \in O_l} w_{ok}.$$

We can calculate b_{lz} as

$$b_{lz} = \sum_{k \in K_z} a_{lk},$$

and calculate g_{lr} as

$$g_{lr} = \sum_{k \in K_r} a_{lk}.$$

Below, we summarize the constant matrices for the problem formulation.

C, c_{op}	Order-Product inclusion matrix and its elements c_{op}, $\forall o \in O$ and $p \in P$
W, w_{ok}	Order-Cell inclusion matrix and its elements w_{or}, $\forall o \in O$ and $k \in K$
S, s_{pk}	Product-Cell inclusion matrix and its elements s_{pr}, $\forall p \in P$ and $k \in K$
A, a_{lk}	Order list-Cell relationship matrix and its values a_{lk}, $\forall l \in L$ and $k \in K$
B, b_{lz}	Order list-Zone relationship matrix and its values b_{lz}, $\forall l \in L$ and $z \in Z$
G, g_{lh}	Order list-Rack position relationship matrix and its values g_{lh}, $\forall l \in L$ and $h \in H$

3.2 GPSM for Average flow time estimation

Feature Generation

To estimate the average flow time of orders in an order list ($\overline{FT}$), an estimation model requires the key input variables (i.e., features) that are expected relate to $\overline{FT}$. We generate features from historical order picking data that can represent the picking time and workload balance across zones. Picker's travel distance and workload balancing measures, such as standard deviations of workload between zones, and zones' maximum and minimum numbers of workload, are potentially related to $\overline{FT}$ (Huang et al. (2018), Vanheusden et al. (2022)). We develop GPSM using three features of zone workload balance and a feature of rack storage policy to estimate the average flow time $\overline{FT}$.

Let $\mathbf{x}_l$ represent the three input features extracted from order list $l \in L$, and let $\mathbf{X} = \{\mathbf{x}_l, \boldsymbol{l} \in \boldsymbol{L}\}$ We consider three models. The first GPSM, which we refer to as GPSM_{AZ}, is only based on the first input feature, i.e., picker's travel distance. We define a total travel distance to complete order list $l \in L$ as

$$TD_l = \sum_{r \in R} d_r \cdot g_{lr},$$

where d_r is the distance from the picker's loading depot to a rack position $r \in R$.

The second GPSM, which we call GPSM_{MM}, uses three input features: TD_l and the maximum and minimum number of workloads of zones. We calculate the maximum number of picks among zones b_l^{max} and the minimum numbers of picks among zones b_l^{min} as

$$b_l^{max} = \max_{\forall z \in Z} b_{lz},$$

$$b_l^{min} = \min_{\forall z \in Z} b_{lz}.$$

The third GPSM, which we refer to as GPSM$_{SD}$, also uses TD_l and one additional input, the standard deviation of the number of picks per zone. We let b_l^{SD} represent the standard deviation for order list $l \in L$, which can be obtained as

$$b_l^{SD} = \sqrt{\frac{\sum_{z \in Z}(b_{lz} - \frac{1}{|Z|}\sum_{z \in Z}(b_{lz}))^2}{|Z|-1}}.$$

GPSM

Gaussian process (GP) regression, a popular surrogate model for computer and physical experiments (Mackay, 1998; Murphy, 2012), provides a predictive relation between input and response variables. For our problem, the input is an input feature set $\boldsymbol{x}$, and the response variable y would be the average flow time $\overline{FT}$. Suppose that we have training data of input features and associated response values for orders in order list L. We denote the training data by $(\boldsymbol{X}, \boldsymbol{y})$, where $\boldsymbol{X} = (\boldsymbol{x_l}, l \in L)$ and $\boldsymbol{y} = (y_l, l = 1 \in L)$ represent the training inputs and responses. The inputs and responses are related via an unknown regression model $f(\boldsymbol{x})$ as

$$y_l = f(\boldsymbol{x_l}) + \epsilon_l, \epsilon_l \sim N(0, \sigma_n^2).$$

We assume the regression function $f(\boldsymbol{x})$ is a zero-mean Gaussian process with covariance function K$(\boldsymbol{x}, \boldsymbol{x}')$. Given the training data, we like to predict the posterior distribution of $f(\boldsymbol{x}_*)$ for a test input $\boldsymbol{x}_*$, which follows a Gaussian distribution with mean $\hat{\mu}(\mathbf{x}_*)$ and variance $\hat{\Sigma}(\mathbf{x}_*)$ (Rasmussen & Williams, 2006),

$$\hat{\mu}(\mathbf{x}_*) = K(\mathbf{x}_*,\mathbf{X})[K(\mathbf{X},\mathbf{X}) + \sigma_n^2 \boldsymbol{I}]^{-1} y, \tag{1}$$

$$\hat{\Sigma}(\mathbf{x}_*) = K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*,\mathbf{X})[K(\mathbf{X},\mathbf{X}) + \sigma_n^2 \boldsymbol{I}]^{-1} K(\mathbf{x}_*,\mathbf{X}), \tag{2}$$

where $\boldsymbol{I}$ is an $|\boldsymbol{L}| \times |\boldsymbol{L}|$ identity matrix, and $\boldsymbol{K}(\boldsymbol{X_1}, \boldsymbol{X_2})$ is the matrix of the covariance function values evaluated between $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$. We use the covariance function composed of a squared exponential (SE) covariance function and noise variance function,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\lambda^2}\right) + \sigma_n^2 \delta(\mathbf{x}_i, \mathbf{x}_j), \tag{3}$$

where variable length scale λ, signal variance σ_f^2, and noise variance σ_b^2 control the correlation between observations $\mathbf{x}_i$ and $\mathbf{x}_j$, and δ is the Kronecker delta function. The kernel-function option depends on the *input* and expected patterns in the data, e.g., Richardson et al. (2017) considered SE and Matérn covariance as the kernel function.

4. Storage location reassignment with risk assessment

In this section we define a storage location reassignment as the sequential determination of swap operations, each of which is optimized to minimize the average flow-time of orders based on the

GPSM's output. We also use the GPSM estimate to evaluate the improvement probability of the swap operations by screening out the less promising ones.

4.1 Swap operation

The storage location reassignment adjusts the locations of products to minimize the average flow-time estimate from the GPSM. We use a series of swap operations (i.e., switch a pair of product cartons) to optimize storage locations with limited handling capacity. For each swap operation, we assume that

- Each picker stays in his/her zone and independently retrieves each product per tote.
- Each product occupies one zone and one rack.
- Each zone has the same number of racks.
- Constant picking time includes search time, pick time, and inspection time.
- Management has sufficient time to reassign locations for new order lists (Sadiq et al., 1996).
- All products are available before an order enters the system.
- A swap operation swaps two cartons of products across flow racks at a time.
- The total number of the swaps is finite and determined dynamically.
- The swap distance as defined by the physical distance between two rack locations, and the number of cartons affect swap time.
- The number of swap operation is not predetermined; after a swap, management determines if another swap is necessary.
- The reserved area has empty racks; therefore, swap operations can be performed without additional handling.
- The swap operation utilizes the reserved area to temporarily store the products to be reassigned. The reserved area has empty racks; therefore, swap operations can be performed without additional handling.

Each swap operation is optimized so that the choice minimizes the average flow time estimated by the pre-trained GPSM described in Section 3.2. A swap operation is represented by a column swap of the original product-pick-face inclusion matrix **S**. For example, suppose there are five products ($P = \{1,2,3,4,5\}$) in each pick-face. If the swap operation occurs between the product 2 and product 4, then it makes a change of the original product-pick-face inclusion matrix **S** to **S`**

$$\mathbf{S}=\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}, \qquad \mathbf{S}` = \begin{bmatrix}1&0&0&0&0\\0&0&0&1&0\\0&0&1&0&0\\0&1&0&0&0\\0&0&0&0&1\end{bmatrix}.$$

After the update to **S**`, we can calculate the corresponding testing input $\boldsymbol{x}_*$ and then get the associate posterior predictive distribution $f(\boldsymbol{x}_*)$ from the GPSM to get the estimated distribution of the average flow time resulting from the swap. We repeat this for a list of N feasible swap operations and select the best swap operation that minimizes the expected average flow time, i.e., $E[f(\boldsymbol{x}_*)]$. We consider all possible swap cases among $|P|$ products, which gives N equal to two-out-of-$|P|$ combination caes. We iterate the swap operation to the maximum number of swaps M determined by management.

4.2 Feedforward heuristic

Since exact algorithm for an optimal sequence of M swap operations is not suitable due to the combinatorial nature of the problem, we propose Algorithm 1, where the best solutions from the precedence process become the input of the next process. The steps of the feedforward heuristic are as follows.

Algorithm 1. Feedforward heuristic.

Step 1.	Set M (based on warehouse experience and considering the number of storage locations that can be reassigned), and set i =1 and set S^1 to be the current product-pick-face inclusion matrix.
Step 2.	Generate N candidates $\{\mathbf{S}_1^i, \mathbf{S}_2^i, \dots, \mathbf{S}_N^i\}$ from S^{i-1}. (See 4.1 Swap operation)
Step 3.	Calculate the corresponding input features $\{\mathbf{x}_{*1}^i \mathbf{x}_{*2}^i, \dots, \mathbf{x}_{*N}^i\}$ corresponding to the N candidates.
Step 4.	Obtain $\hat{\mu}(\mathbf{x}_*^i)$ from each candidate and find the optimal swap operation based on equation (4).
Step 5.	Update S^i.
Step 6.	If $i < M$, $i = i + 1$ and go to step 2, otherwise STOP.

Let $\mathbf{S}^i$ represent the ith storage location matrix for the iteration $i = 1\dots, M$, where M controls the maximum number of swap operations set according to the warehouse environment. For every iteration i, generate storage location candidates $\{\mathbf{S}_1^i, \mathbf{S}_2^i, \dots, \mathbf{S}_N^i\}$ with swap operations from former storage locations $\mathbf{S}^{i-1}$. Each candidate generates the testing inputs $\{\mathbf{x}_{*1}^i \mathbf{x}_{*2}^i, \dots, \mathbf{x}_{*N}^i\}$ to be evaluated with the proposed average flow-time estimation model. Next, determine the storage locations $\mathbf{S}^i$ according to the best $\mathbf{x}_*^i$ that yields the minimum $\hat{\mu}(\mathbf{x}_*^i)$ using Equation (1). Obtain the expected average flow time of the storage locations $\mathbf{S}^i$ as

$$\mathrm{E}[\overline{FT}(\mathbf{S}^i)] = \min\{\hat{\mu}(\mathbf{x}_*^i) \big| \mathbf{x}_*^i \in \{\mathbf{x}_{*1}^i, \mathbf{x}_{*2}^i, \dots, \mathbf{x}_{*N}^i\}\}. \tag{4}$$

Considering an initial storage location $\mathbf{S}^0$ for an order list l and obtaining the historical data in the form listed, use Algorithm 1 to find the new storage location ($\mathbf{S}^1$) at iteration 1 that minimizes the average flow time for an order list l. Note that the new storage location $\mathbf{S}^1$ from the first stage is considered as the initial storage location $\mathbf{S}^{i-1}$ of the second stage and obtains the new storage location

$\mathbf{S}^2$ that yields the minimum $\overline{FT}$ at the second stage. Run these procedures until iteration i reaches the maximum number of swaps M.

4.3 Screening out less promising swaps

Quantifying uncertainty is critical in managing operational risks. The probability that a system will not improve can be estimated with the estimates of Gaussian process model (Bect et al., 2012). We conduct a screening test using the posterior predictive distribution from GPSM. We use the posterior predictive distribution of the average flow time after a swap to evaluate the probability of improvement in the average flow time, compared to the average flow time before that swap. If the improvement probability is not sufficiently high, we revoke the determined swap operation.

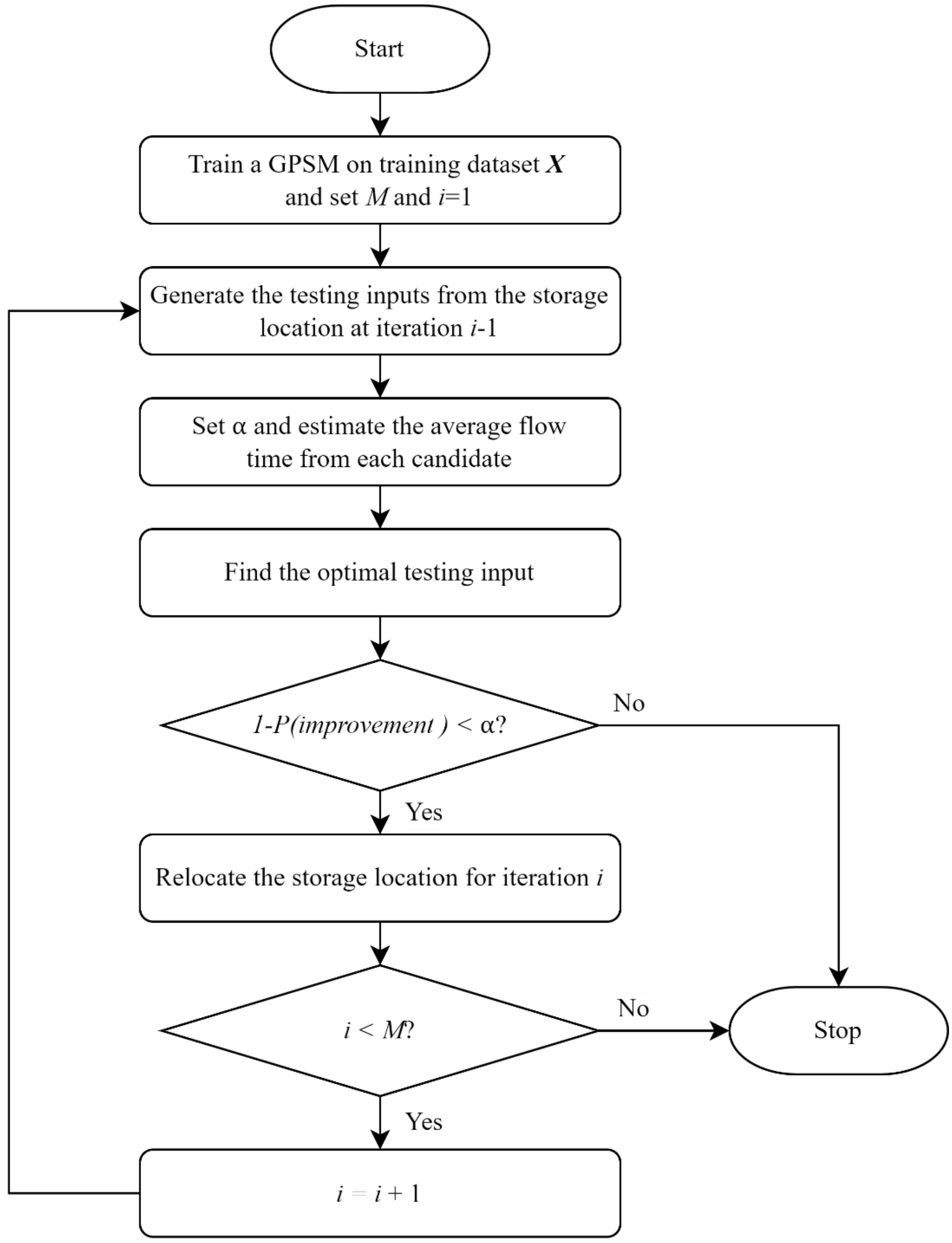

Figure 3. Flowchart of a screening test applied to Algorithm 1.

Given the means and variances of the two mutually independent Gaussian distributions, we use a one-tailed test to statistically compare the difference between the mean values of two distributions. GPSM is used to obtain the mean($\hat{\mu}(\cdot)$) and variance($\hat{\Sigma}(\cdot)$) of $\mathbf{x}_*^{\boldsymbol{i}}$ and $\mathbf{x}_*^{\boldsymbol{i-1}}$, where $\mathbf{x}_*^{\boldsymbol{i}}$ is the testing input expected to be the optimal case from the relocated cases $\{\mathbf{x}_{*\mathbf{1}}^{\boldsymbol{i}}, \mathbf{x}_{*\mathbf{2}}^{\boldsymbol{i}}, \dots, \mathbf{x}_{*\boldsymbol{N}}^{\boldsymbol{i}}\}$, and $\mathbf{x}_*^{\boldsymbol{i-1}}$ is the testing input of the reassigned order list at iteration $i-1$. The null hypothesis is that the $\overline{FT}$ of $\mathbf{x}_*^{\boldsymbol{i-1}}$ is less than or equal to the $\overline{FT}$ of $\mathbf{x}_*^{\boldsymbol{i}}$. The probability of the null hypothesis (p-value) denotes the probability that the reassignment operation will not improve. We set significance level α for the one-tailed test to classify the case with the risk to increase $\overline{FT}$. We obtain the Z-score of the difference between the two $\overline{FT}$ values and the improvement probability from the Z-score as

$$1 - P(improvement) = P(\overline{FT}(\mathbf{S}^{i-1}) - \overline{FT}(\mathbf{S}^{i}) \leq 0) = P\left(Z \geq \frac{\hat{\mu}(\mathbf{x}_*^{i-1}) - \hat{\mu}(\mathbf{x}_*^{i})}{\sqrt{\hat{\Sigma}(\mathbf{x}_*^{i-1}) + \hat{\Sigma}(\mathbf{x}_*^{i})}}\right). \quad (5)$$

We compare the one-tailed p-value with the significance level α. Note that the improvement probability of reassignment cannot be guaranteed if the p-value is greater than the significance level. If the p-value becomes smaller than the significance level, the GPSM expects that the flow time of the GPSM optimal assignment is significantly shorter than the flow time of the original assignment; therefore, it executes the determined swap operation.

5. Industrial application of GPSM

In this section we describe our simulation experiment and the results. Ten scenarios from a real e-commerce warehouse configuration and historical order data were used to evaluate the performance of the proposed model. We analyze the effectiveness of reassignment decisions and conduct sensitivity tests over various warehouse configurations.

5.1 Workload scenarios

To obtain accurate and reliable estimates, we use different workload scenarios to train the proposed models on various storage policy factors. We consider the four zone workload distribution types and two rack storage policies shown in Table 1.

Table 1 Zone workload distribution types and rack storage policies.

Zone workload distribution type	Rack storage policy
Uniform	ABC(5:3:2)
Bottleneck zone	Random
Descending demand	
Irregular demand	

We note that a uniform zone workload distribution type is the most workload-balanced scenario with almost uniformly distributed workloads over zones. We generate Scenarios 1 and 2 with this type to make workload balanced scenarios. The bottleneck zone distribution type refers to the case that one randomly chosen zone is assigned 25% of total orders when all other zones have a uniform workload distribution. The bottleneck zone scenarios can occur in situations when demand for particular products increase dramatically that often arises in e-commerce warehouses. We generate Scenarios 3 and 4 with this type. A descending demand zone distribution type has a descending trend in workload. The first zone has the largest demand volume and the demand gradually decreases toward the last zone. Many warehouse managements set descending demand in progressive zone picking system to avoid blocking delay between zones. We distribute the workload ratio of each zone as the ratio of each zone index to the sum of zone indexes. We generate Scenarios 5 and 6 with this type. An irregular demand-zone distribution type represents the most workload-unbalanced scenario with an uneven distribution of workloads across zones. We demonstrate worst-case workload balancing by randomly shuffling storage locations for descending demand scenarios and generating irregular demand scenarios. We generate Scenarios 7 and 8 with this type.

The rack storage policy defines how products' storage locations are assigned over rack column positions. The ABC (5:3:2) class-based rack storage policy classifies products to classes A, B and C in a ratio (0.5:0.3:0.2) for allocating high-demand products in the rack column positions with the picker's shortest travel distance. The random rack-storage policy refers to the random assignment of products within zones.

We generate the first eight workload scenarios using a full factorial design with four zone workload distribution types and two rack storage policies and two additional scenarios. To train GPSM even with scenarios that cannot be considered by the full factorial design, we generate scenarios 9 and 10 by switching a randomly chosen pair of product locations from Scenario 1. Scenarios 9 and 10 have balanced workloads, but the storage locations of products differ from the preset rack storage policy due to demand fluctuations and a series of previous storage relocations. We classify Scenarios 1, 2, 9, and 10 as workload balanced scenarios and Scenarios 3, 4, 5, 6, 7, and 8 as workload unbalanced scenarios. In Table 2, the average $\overline{FT}$ of 50 order lists per scenario measured by simulation shows that zone workload distribution type and rack storage policy directly influence order picking performance. The workload balanced Scenarios 1, 2, 9, and 10 provide shorter average $\overline{FT}$. We conduct statistical tests to validate the significance of difference between scenarios (Please see Table B1 and Table B2 in Appendix B for detail).

Table 2. Ten workload distribution scenarios.

Scenario	Zone workload distribution type	Rack storage policy	Average $\overline{FT}$
1	Uniform	ABC (5:3:2)	1250.03
2	Uniform	Random	1309.47
3	Bottleneck zone	ABC (5:3:2)	1440.65
4	Bottleneck zone	Random	1523.62

5	Descending demand	ABC (5:3:2)	1513.42
6	Descending demand	Random	1593.70
7	Irregular demand	ABC (5:3:2)	1522.98
8	Irregular demand	Random	1586.75
9	24 times random switch from scenario 1		1284.24
10	100 times random switch from scenario 1		1312.51

5.2 Experimental configuration

To validate our proposed GPSM, we configure a warehouse and its order profiles based on an e-commerce company's progressive bypass ZP system in Korea. We run 200 experiments per configuration, i.e., 20 different order lists per each of the 10 scenarios. We train GPSM on 300 training datasets consisting of 10 scenarios with 30 different order lists. We use Tecnomatix© Plant Simulation 12 to build the simulation model. We generate synthetic historical data based on the order profiles and modified sizes of the ZP system. Since recirculation and bypass disrupt the order sequence, the order release sequence follows the First-Come-First-Serve rule. We use Python 3.7 and scikit-learn toolbox to analyze the data. Table 3 reports the details.

Table 3. Warehouse configuration.

Parameter	Values
Number of swaps (n)	1, 2, 3, 4
Number of zones	8
Number of rack columns per zone	6
Order size	$Uniform$ (3,9)
Number of orders in each order list	100

Note: Default values of each parameter are underscored.

We label the GPSMs as stages 1–4 by setting $M = 4$. GPSM(n) denotes the first stage resulting in the swapping of n pairs of storage locations. We compare the GPSM to Kim and Hong (2022), who proposed an MIP model to reassign a limited number of storage locations for the same order picking system. For simplicity, we call their model SLR(n), where n is stage number (i.e., the number of the swap operations). We also compare the GPSM with the heuristic algorithm proposed by Jane (2000) for workload balancing in a zone picking system. For simplicity, we call their heuristic algorithm the JA. Unlike the GPSM and SLR, the JA reassigns all storage locations without limiting the number of reassignments.

The objective of the GPSM is to minimize the average flow time; the objective of the SLR(n) model is to minimize the maximum number of picks and the maximum number of order visits among all zones; and the objective of the Jane is to distribute the number of picks evenly across all zones.

Since all three models aim to balance the workload in the progressive ZP system to minimize order flow time, we compare the models' performance.

5.3 GPSM effectiveness analysis

Figure 4 illustrates the percentage reductions of average $\overline{FT}$ between the original assignment and three GPSMs, and between the original assignment and SLR(n) when swapping multiple pairs of products. Without the screening test, $GPSM_{SD}$(1) yields a 3.23% average reduction, $GPSM_{SD}$(4) yields an 8.87% average reduction, SLR(1) yields a 2.93% average reduction, and Jane yields an 8.21% in $\overline{FT}$. The workload unbalanced scenarios (1, 2, 9, and 10) have larger $\overline{FT}$ reductions than the workload balanced scenarios (3, 4, 5, 6, 7, and 8). For the same zone workload distribution type scenarios in the GPSMs, the random storage policy yields a high reduction percentage of $\overline{FT}$. The results indicate that the workload unbalanced scenarios have scope for further improvement than the workload balanced scenarios.

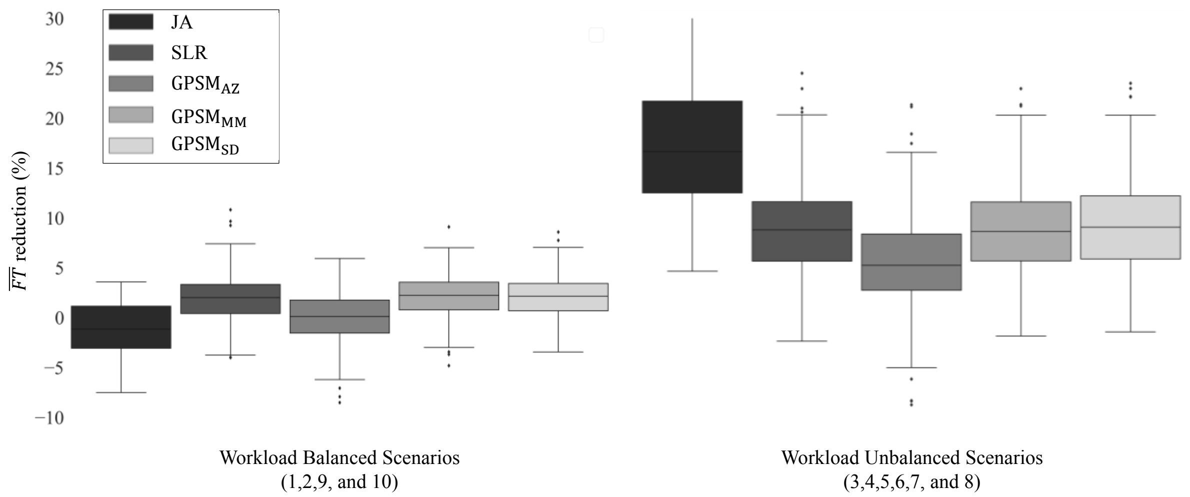

Figure 4. Percentage reductions of average $\overline{FT}$ by workload scenarios without the screening test.

Without the screening test, we observe 24 failures of reassignment that yield a 1.13% increase in the average $\overline{FT}$ compared with the original assignment when $GPSM_{SD}$(1) runs for 200 instances. For the same instances, Jane obtains 49 failures of reassignment with an increase of 2.96% in the average $\overline{FT}$ over the original assignment, and SLR(1) results in 28 failures of reassignment with an increase of 1.13%. The screening test excludes the reassignment that is expected to improve with low probability.

Table 4 reports the results of the paired t-tests for each scenario and stage to identify statistically significant differences between SLR and $GPSM_{SD}$. The differences are insignificant in the balanced scenarios (1, 2, 9, 10) or low stages (n = 1,2). Where swap operations are performed more than three times, $GPSM_{SD}$ has shorter $\overline{FT}$ than SLR in the unbalanced scenarios (3, 4, 5, 6, 7, 8).

Table 4. Paired t-test results on $\overline{FT}$ of SLR(n) – GPSM$_{SD}$(n)

Stage	Scenario	Paired t-test on the average $\overline{FT}$			
		Mean difference	*t*-value	DF	*P*-value
n = 1	1	11.405	2.082	19	0.051
	2	-1.286	-0.227	19	0.823
	3	8.831	1.322	19	0.202
	4	-6.932	-1.246	19	0.228
	5	-3.917	-0.638	19	0.531
	6	3.569	0.451	19	0.657
	7	3.630	0.476	19	0.639
	8	17.554	1.960	19	0.065
	9	-0.853	-0.206	19	0.839
	10	11.577	2.059	19	0.053
n = 2	1	16.198	3.604	19	0.002
	2	3.657	0.700	19	0.493
	3	13.045	1.613	19	0.123
	4	-3.310	-0.379	19	0.709
	5	-3.297	-0.608	19	0.551
	6	0.542	0.075	19	0.941
	7	6.574	0.775	19	0.448
	8	7.898	1.730	19	0.100
	9	4.738	0.792	19	0.438
	10	4.738	0.792	19	0.438
n = 3	1	-0.846	-0.154	19	0.879
	2	7.804	2.075	19	0.052
	3	15.395	1.676	19	0.110
	4	23.124	4.068	19	0.001
	5	19.973	2.827	19	0.011
	6	20.554	2.211	19	0.040
	7	0.025	0.003	19	0.998
	8	23.433	3.167	19	0.005
	9	3.494	0.973	19	0.343
	10	2.883	0.453	19	0.656
n = 4	1	-1.753	-0.277	19	0.785
	2	0.667	0.126	19	0.901
	3	24.278	2.614	19	0.017
	4	18.841	2.228	19	0.038
	5	25.310	3.224	19	0.004
	6	25.168	2.154	19	0.044
	7	16.812	2.512	19	0.021
	8	30.869	4.215	19	0.000
	9	-7.403	-1.706	19	0.104
	10	-5.434	-1.075	19	0.296

Figure 5 illustrates the screening test's effects on productivity for 200 assignments from SLR, SD-GPSM before the screening test, GPSM$_{SD}$ after the screening test, and the original assignments without reassignment. We adopt Youden's J statistic for the significance levels of the screening tests (Youden, 1950). We set swapping time (ST) weight, considering the replenishment time in the real e-commerce warehouse. On the x-axis, we classify experimental cases by different reassignment strategies, stage number, and ST weight. The black bars indicate the time difference between the start and finish of a sequence of an order list (C_{max}; makespan), and the dotted white bars indicate the total swapping time (TST). The y-axis represents the average total elapsed time including C_{max} and TST.

The makespan distinctly decreases as the stages increase. GPSM$_{SD}$ after the screening test yields a shorter TST because the screening test filters out the number of swap operations. When $ST =$ 2.5, the total swapping time of GPSM$_{SD}$(2) before the screening test averages 300 seconds, and the total swapping time of GPSM$_{SD}$(3) after the screening test averages 256.5 seconds. Before the screening test, the makespan of GPSM$_{SD}$(2) averages 2449.90 seconds, and the makespan of GPSM$_{SD}$(3) averages 2413.62 seconds. After the screening test, GPSM$_{SD}$ frequently shows a large makespan reduction with less swapping time.

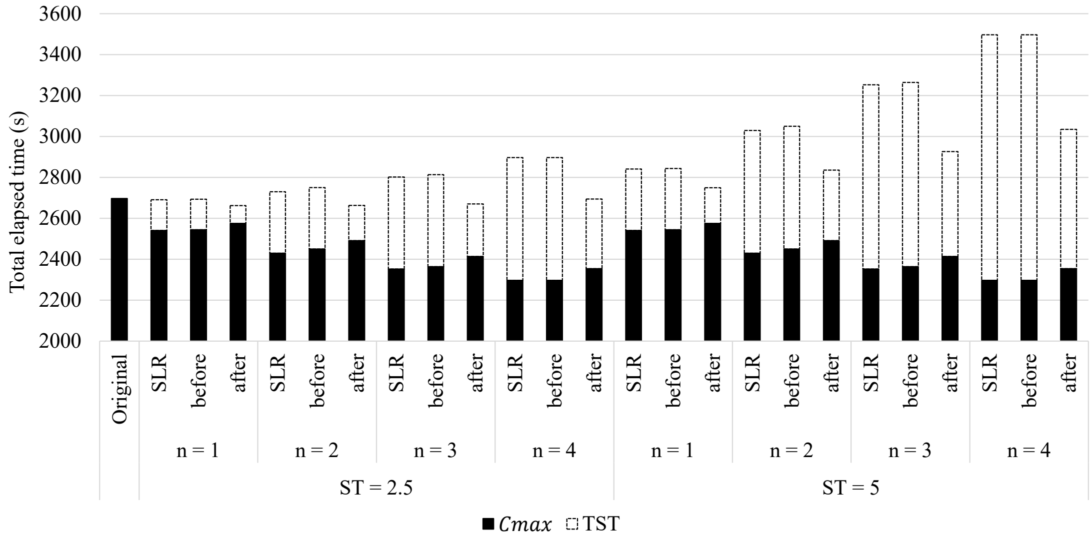

Figure 5. Screening test effectiveness analysis considering swapping times.

5.4 Sensitivity analysis of GPSM with a screening test

We consider different warehouse configurations (i.e., increasing the number of zones, rack columns, orders in an order list, and order sizes). We also consider non-identical picker skill configuration (i.e., two slow-moving pickers in eight zones) (Bartholdi and Eisenstein (1996), Hong (2019)).We use the following measurements:

- Before: The average $\overline{FT}$ reduction before the screening test (%)
- α: The significance level for the screening test
- #swaps: The average number of swaps after the screening test
- Avg: The average $\overline{FT}$ reduction (%)
- Worst: The $\overline{FT}$ reduction of the worst case (%)

Table 5 reports the sensitivity analysis results. In Appendix C, we conduct statistical analyses to identify which methods have statistically significant differences across order picking environments. As a result of the fact that the data variances for each group are not equal, we use Welch's ANOVA test (Table C1) and Games-Howell as post-hoc tests (Table C2, Table C3, Table C4, Table C5, Table C6, and Table C7) for each order picking environment.

Table 5. Results of GPSM sensitivity analysis with screening tests.

Environment	JA		Stage	SLR(n)			$GPSM_{AZ}(n)$					$GPSM_{MM}(n)$					$GPSM_{SD}(n)$				
	Avg.	Worst		#swaps	Avg.	Worst	Before	α	#swaps	Avg	Worst	Before	α	#swaps	Avg	Worst	Before	α	#swaps	Avg	Worst
Default	*8.21*	*-8.13*	$n = 1$	1.00	2.93	-4.16	1.92	0.39	0.27	0.89	-1.76	3.02	0.14	0.55	2.28	-0.82	3.23	0.15	0.59	2.56	-1.90
			$n = 2$	2.00	5.28	-3.98	3.06	0.44	0.50	1.37	-1.76	5.32	0.32	1.00	3.86	-0.10	5.60	0.26	1.15	4.54	0.00
			$n = 3$	3.00	6.82	-3.75	3.90	0.46	0.70	1.78	-1.76	7.26	0.37	1.40	5.06	-0.10	7.59	0.31	1.71	6.23	0.00
			$n = 4$	4.00	8.06	-2.22	4.40	0.48	0.90	2.06	-1.76	8.40	0.37	1.72	5.54	-0.10	8.88	0.37	2.27	7.51	0.00
12 zones	*9.41*	*-8.69*	$n = 1$	1.00	2.66	-6.22	2.00	0.41	0.28	0.94	-2.60	2.71	0.27	0.51	2.14	-3.84	2.40	0.36	0.42	1.69	-3.30
			$n = 2$	2.00	4.59	-6.43	3.00	0.50	0.56	1.47	-2.84	4.62	0.37	0.86	3.44	-3.84	4.20	0.41	0.81	2.98	-3.99
			$n = 3$	3.00	5.98	-4.85	4.38	0.50	0.69	1.62	-0.90	5.82	0.39	1.14	4.32	-3.84	5.64	0.44	1.20	3.88	-3.18
			$n = 4$	4.00	7.31	-5.22	5.32	0.50	0.72	1.66	-0.90	6.96	0.41	1.38	5.07	-3.84	6.44	0.44	1.54	4.50	-3.18
12 rack columns	*7.00*	*-17.59*	$n = 1$	1.00	2.26	-5.68	1.53	0.42	0.25	0.75	-0.37	2.83	0.16	0.52	1.82	-2.83	2.69	0.18	0.51	1.83	-1.39
			$n = 2$	2.00	3.85	-2.85	2.86	0.45	0.45	1.12	-2.37	4.72	0.32	0.99	3.14	-2.83	4.94	0.35	1.02	3.39	-0.47
			$n = 3$	3.00	5.02	-2.41	3.57	0.49	0.63	1.48	-1.11	6.48	0.36	1.45	4.38	-2.83	6.72	0.31	1.51	4.67	0.00
			$n = 4$	4.00	6.04	-2.02	4.48	0.50	0.80	1.70	0.00	8.04	0.40	1.90	5.26	-2.83	8.28	0.35	2.00	5.79	0.00
Unif(10,20) *Order size*	*7.27*	*-8.68*	$n = 1$	1.00	1.91	-2.47	1.41	0.47	0.80	1.38	-7.23	2.43	0.19	0.79	2.46	-1.31	2.53	0.19	0.77	2.54	-1.31
			$n = 2$	2.00	3.51	-2.04	2.50	0.41	1.18	2.20	-7.23	4.12	0.28	1.42	3.95	-1.31	4.36	0.25	1.46	4.20	-0.72
			$n = 3$	3.00	4.87	-2.07	3.33	0.41	1.32	2.52	-7.23	5.54	0.43	2.04	5.21	-1.31	5.82	0.35	2.15	5.56	0.00
			$n = 4$	4.00	6.01	-2.32	3.98	0.46	1.42	2.67	-7.23	6.73	0.39	2.65	6.31	-1.31	7.14	0.38	2.83	6.75	0.00
200 orders in each order list	*13.59*	*-7.60*	$n = 1$	1.00	4.14	-1.19	3.08	0.43	0.53	2.19	-3.60	4.03	0.24	0.86	3.93	-0.42	3.73	0.18	0.38	1.55	-0.42
			$n = 2$	2.00	7.06	-1.05	5.44	0.40	0.79	3.24	-3.60	6.96	0.40	1.67	6.72	0.00	6.76	0.27	0.69	2.64	-0.42
			$n = 3$	3.00	9.22	-1.58	6.84	0.40	0.80	3.27	-3.60	9.21	0.39	2.39	8.80	0.00	9.15	0.33	0.93	3.35	-0.42
			$n = 4$	4.00	10.82	-0.93	7.60	0.50	0.81	3.29	-3.60	10.88	0.42	3.05	10.30	0.00	10.92	0.39	1.15	3.76	-0.42
Non-identical picker skill	*7.74*	*-10.47*	$n = 1$	1.00	2.78	-4.99	3.60	0.41	0.93	3.44	-2.75	2.85	0.37	0.73	2.70	-2.14	2.99	0.33	0.52	2.48	-2.03
			$n = 2$	2.00	4.77	-6.42	6.22	0.39	1.43	4.92	-2.46	5.22	0.43	1.32	4.78	-3.16	5.40	0.40	1.03	4.21	-1.12
			$n = 3$	3.00	6.49	-4.25	8.19	0.47	1.91	5.89	-2.46	7.14	0.43	1.82	6.01	-3.16	7.20	0.43	1.54	5.69	-1.12
			$n = 4$	4.00	7.53	-5.27	9.64	0.48	2.37	6.67	-2.46	8.44	0.44	2.24	6.94	-3.16	8.60	0.43	2.05	6.67	-1.12

Sensitivity analysis over default order picking environment

Table 5 reports only four failures of reassignment with a 1.09% increase in $\overline{FT}$ compared with the original assignment when $GPSM_{SD}(1)$ runs for 200 instances. When the stage increases, the SLR model swaps the number of product pairs equal to the stage number, whereas the GPSM policies swap fewer product pairs because the screening test filters out the swap operation with a low improvement probability. Both $GPSM_{SD}$ and $GPSM_{MM}$ significantly reduce $\overline{FT}$ and swapping time. JA swaps the unlimited number of storage locations and yields 8.21% of the average $\overline{FT}$ reduction. $GPSM_{SD}(4)$ swaps on average 2.27 times to yield 7.51% of the average $\overline{FT}$ reduction, but SLR needs to swap three times to yield 6.82% of the average $\overline{FT}$ reduction. $GPSM_{SD}$ yields 0.00% for the $\overline{FT}$ reduction of the worst cases, whereas SLR(4) and JA yield -2.22% and -8.13%, respectively. Appendix Table C2 shows that $\overline{FT}$ reduction rates of SLR(4) and JA are not significantly different from $GPSM_{SD}(4)$ and $GPSM_{MM}(4)$ before performing screening tests. The results indicate that GPSM is a reliable, cost-effective tool for warehouse management.

Sensitivity analysis over 12 zones

Table 5 also shows that reassignment performance decreases when the number of zones increases. SLR yields 2.66% of the average $\overline{FT}$ reduction with one swap operation and 4.59% of the average $\overline{FT}$ reduction with two swap operations. After the screening test, the GPSM policies skip the low improvement cases and reduce the average number of swap operations. $GPSM_{SD}$ yields 4.50% of the $\overline{FT}$ with average 1.54 swap operations, and $GPSM_{MM}$ yields 5.07% of the $\overline{FT}$ with average 1.38 swap operations. $GPSM_{SD}$ yields -3.18% for the $\overline{FT}$ reduction of the worst cases, whereas JA and SLR(4) yield -8.69% and -5.22%, respectively. Appendix Table C3 shows that $\overline{FT}$ reduction rates of SLR(4) and $GPSM_{SD}(4)$ are not significantly different before performing screening tests. The results indicate that GPSM performs well over large-scale assignments with limited training data resources.

Sensitivity analysis over 12 rack columns per zone

As the number of rack columns in each zone increases, picker's travel distance has a large effect on order picking performance. SLR yields 3.85% for the average $\overline{FT}$ reduction with 2 swap operations, and $GPSM_{SD}$ and $GPSM_{MM}$ yield 5.79% for the average $\overline{FT}$ reduction with average 2.00 swap operations and 5.26%, respectively, with 1.90 swap operations. The results indicate that both $GPSM_{SD}$ and $GPSM_{MM}$ balance the workloads well between zones when travel distance is critical.

Sensitivity analysis over order size = Uniform(10,20)

$GPSM_{SD}$ and $GPSM_{MM}$ consistently reduce the average $\overline{FT}$ from stages 1–4 and provide assignments with a high probability to improve. Intuitively, increasing the order size causes congestion because the totes stay longer in the system. JA yields 7.27% for the average $\overline{FT}$ reduction with unlimited swaps, whereas SLR yields 3.51% when swapping two pairs of products, and 4.87% when swapping three pairs of products, respectively. $GPSM_{SD}$ yields 6.75 % when swapping an average 2.83 pairs of products, whereas GPSM_{MM} yields 6.31% when swapping an average 2.65 pairs of products. Appendix Table C5 shows that $\overline{FT}$ reduction rates of SLR(4), JA, GPSM_{SD}(4), and GPSM_{MM}(4) are not significantly different before performing screening tests. The results confirm well-balanced workloads between zones when the order size is large.

Sensitivity analysis over number of orders = 200

To analyze GPSM performance over a large problem size, we increase the number of orders in each order list. As the number of orders increases, the average $\overline{FT}$ reduction between the original and reassignment increases: GPSM_{MM}(4) yields 10.30% on average and 0.00% in the worst case with the average number of swaps = 3.05, and SLR yields 9.22% on average and −1.58% in the worst case with three swap operations, whereas JA yields 13.59% on average and -7.60% in the worst case with the unlimited number of swap operations.

Sensitivity analysis over non-identical picker skill

GPSM_{AZ} yields the largest $\overline{FT}$ reduction before the screening test, 2.11% and 1.90% higher than SLR(4) and JA, respectively. Despite the poor performance of GPSM_{AZ} in other configurations, it does capture the zones with slow picking speeds and assigns fewer loads to them, which balances the workload in a non-identical picker-skill environment. Appendix Table C7 shows that $\overline{FT}$ reduction rates of GPSM_{AZ}(4) is significantly larger than SLR(4) before performing screening tests. The results indicate that the GPSM policies effectively reassign storage to reduce $\overline{FT}$ when pickers are non-identical in skill and speed.

6. Conclusion

This paper proposed an average flow-time estimation model for estimating travel time and a storage location reassignment with risk assessment in an automated zone picking system. During the storage location reassignment procedure, warehouse management applied the screening test for each reassignment stage to assess the risk of the reassignment decision's failure. Ten scenarios from a real e-commerce warehouse configuration and historical order data were used to evaluate the performance of the proposed model.

The GPSM estimated the average flow time from the historical data with three different types of features: the number of picks in all zones, the min-max number of picks, standard deviation of the number of picks, and the standard deviation of the number of picks. Given storage location assignments, it enabled the GPSM to adjust the storage location by evaluating the minimum estimated average flow time. The GPSM identified new storage location assignments and assessed the risk of a reassignment failure with the probability of improvement.

The simulation experiments and statistical tests validated that the proposed models significantly reduced the average flow time of orders in large-scale order picking and non-identical picker skill environments, and that the screening test limited the number of swap operations.

Future research will investigate constructing the features extracted from an order list based on domain knowledge of the picking system. More analysis should confirm the accuracy of an estimation model that relies on feature extraction from large-scale order picking data with no loss of data integrity. Further investigation of the sequential sampling and kernel functions should indicate improved model performance with limited data.

Appendix A. Numerical validation

We measure the Pearson correlation coefficient (Pearson's r) of the average flow time estimation model. Figure A1 plots the correlation along with the Pearson's r value between the actual and predicted values of true average flow time for 200 different order lists generated using simulation data from industry (see Sections 5.1 and 5.2 for details of the dataset). For the three different GPSMs, Pearson's r values ranging from 0.537 to 0.950 indicate a relatively strong correlation between simulations and predictions.

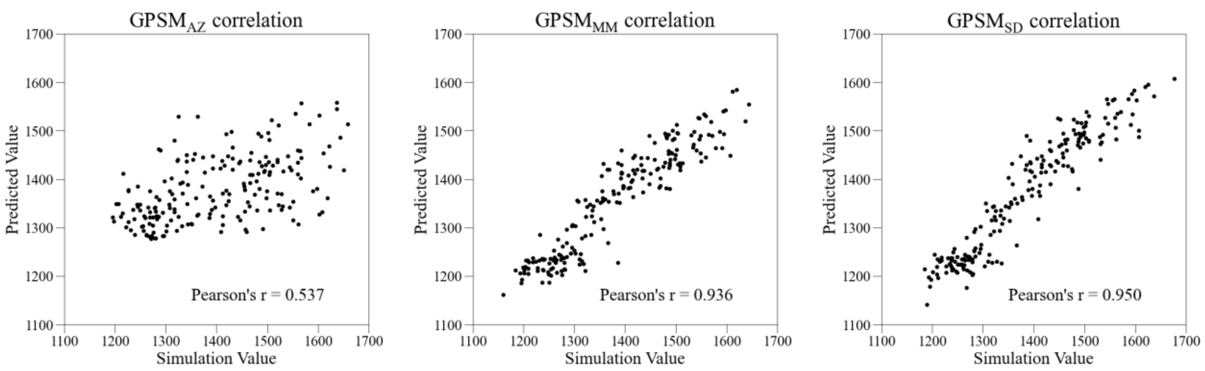

Figure A1. Correlation between the values obtained by simulations and by the three GPSMs.

We also calculate the mean absolute percentage error (MAPE) and R^2 score of the three GPSMs for the dataset. Table A1 reports the values denoting the regression accuracy of the proposed average flow-time estimation models.

Table A1. MAPE and R2 scores of GPSM.

Model	MAPE (%)	R^2
GPSM$_{AZ}$	5.85	0.26
GPSM$_{MM}$	2.74	0.79
GPSM$_{SD}$	2.26	0.89

Appendix B. Statistical analysis between workload scenarios

Table B1. Welch's ANOVA tests of performance between workload scenarios.

Source	Statistic[a]	DF Num	DF Den	*P*-value
$\overline{FT}$	262.032	9	199.157	< 0.001

[a] Asymptotically F distributed

Table B2. Multiple comparisons of mean differences in the average $\overline{FT}$ between workload scenarios based on Games-Howell post-hoc tests.

Scenario	1	2	3	4	5	6	7	8	9	10
1		-59.445*	-190.629*	-273.590*	-263.397*	-343.676*	-272.955*	-336.721*	-34.219*	-62.487*
2			-131.180*	-214.141*	-203.948*	-284.226*	-213.505*	-277.271*	25.231	-3.037
3				-82.961*	-72.768*	-153.046*	-82.326*	-146.091*	156.410*	128.143*
4					10.193	-70.086*	0.635	-63.130*	239.371*	211.104*
5						-80.279*	-9.558	-73.323*	229.178*	200.911*
6							70.721*	6.955	309.457*	281.189*
7								-63.766*	238.736*	210.468*
8									302.501*	274.234*
9										-28.267
10										

* Mean difference significant at the 0.05 level

Appendix C. Statistical analysis between methods

Table C1. Welch's ANOVA tests of performance between methods.

Environment	Stage	Statistic[a]	DF Num	DF Den	*P*-value
default	$n = 1$	75.511	5	524.212	< 0.001
	$n = 2$	51.969	5	531.754	< 0.001
	$n = 3$	40.901	5	537.173	< 0.001
	$n = 4$	42.488	5	542.384	< 0.001
12 zones	$n = 1$	100.042	5	526.994	< 0.001
	$n = 2$	78.095	5	534.856	< 0.001
	$n = 3$	59.942	5	539.904	< 0.001
	$n = 4$	48.484	5	543.359	< 0.001
12 rack columns	$n = 1$	49.255	5	509.983	< 0.001
	$n = 2$	33.982	5	514.926	< 0.001
	$n = 3$	26.549	5	519.378	< 0.001
	$n = 4$	24.391	5	523.325	< 0.001
Unif(10,20) Order size	$n = 1$	94.673	5	507.882	< 0.001
	$n = 2$	70.097	5	514.426	< 0.001
	$n = 3$	53.633	5	521.611	< 0.001
	$n = 4$	44.641	5	527.161	< 0.001
200 orders in each order list	$n = 1$	178.769	5	495.630	< 0.001
	$n = 2$	137.922	5	503.428	< 0.001
	$n = 3$	105.377	5	510.265	< 0.001
	$n = 4$	86.137	5	517.059	< 0.001
Non-identical picker skill	$n = 1$	47.813	5	510.657	< 0.001
	$n = 2$	28.858	5	518.685	< 0.001
	$n = 3$	20.066	5	527.084	< 0.001
	$n = 4$	20.938	5	535.300	< 0.001

[a] Asymptotically F distributed

Table C2. Multiple comparisons of mean differences in the average $\overline{FT}$ in the default environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(n)	GPSM$_{AZ}$(n)	GPSM$_{MM}$(n)	GPSM$_{SD}$(n)
$n = 1$	Original		129.097*	44.031*	29.906	45.268*	48.389*
	JA			-85.065*	-99.190*	-83.828*	-80.708*
	SLR(n)				-14.125	1.237	4.358
	GPSM$_{AZ}$(n)					15.362	18.483
	GPSM$_{MM}$(n)						3.121
	GPSM$_{SD}$(n)						
$n = 2$	Original		129.097*	79.480*	47.641*	79.993*	83.985*
	JA			-49.617*	-81.456*	-49.104*	-45.112*
	SLR(n)				-31.838*	0.513	4.505
	GPSM$_{AZ}$(n)					32.351*	36.344*
	GPSM$_{MM}$(n)						3.992
	GPSM$_{SD}$(n)						
$n = 3$	Original		129.097*	102.719*	60.490*	109.162*	114.303*
	JA			-26.378*	-68.606*	-19.934	-14.794
	SLR(n)				-42.228*	6.444	11.584
	GPSM$_{AZ}$(n)					48.672*	53.812*
	GPSM$_{MM}$(n)						5.140
	GPSM$_{SD}$(n)						
$n = 4$	Original		129.097*	121.378*	68.870*	126.214*	134.113*
	JA			-7.719	-60.226*	-2.883	5.017
	SLR(n)				-52.507*	4.836	12.735
	GPSM$_{AZ}$(n)					57.344*	65.243*

	GPSM$_{MM}$(n)						7.899
	GPSM$_{SD}$(n)						

* Mean difference significant at the 0.05 level

Table C3. Multiple comparison of mean differences in the average $\overline{FT}$ in the 12 zones environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(n)	GPSM$_{AZ}$(n)	GPSM$_{MM}$(n)	GPSM$_{SD}$(n)
$n = 1$	Original		142.240*	39.719*	30.324	40.209*	35.453
	JA			-102.521*	-111.916*	-102.031*	-106.787*
	SLR(n)				-9.395	0.491	-4.266
	GPSM$_{AZ}$(n)					9.885	5.129
	GPSM$_{MM}$(n)						-4.756
	GPSM$_{SD}$(n)						
$n = 2$	Original		142.240*	68.487*	45.252*	68.572*	62.562*
	JA			-73.753*	-96.988*	-73.668*	-79.677*
	SLR(n)				-23.235	0.085	-5.924
	GPSM$_{AZ}$(n)					23.320	17.310
	GPSM$_{MM}$(n)						-6.010
	GPSM$_{SD}$(n)						
$n = 3$	Original		142.240*	89.270*	65.932*	86.845*	83.713*
	JA			-52.970*	-76.308*	-55.395*	-58.527*
	SLR(n)				-23.338	-2.425	-5.557
	GPSM$_{AZ}$(n)					20.913	17.782
	GPSM$_{MM}$(n)						-3.132
	GPSM$_{SD}$(n)						
$n = 4$	Original		142.240*	108.714*	80.106*	104.562*	96.093*
	JA			-33.526*	-62.134*	-37.678*	-46.147*
	SLR(n)				-28.608*	-4.153	-12.622
	GPSM$_{AZ}$(n)					24.455*	15.986
	GPSM$_{MM}$(n)						-8.469
	GPSM$_{SD}$(n)						

* Mean difference significant at the 0.05 level

Table C4. Multiple comparisons of mean differences in the average $\overline{FT}$ in the 12 rack columns environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(n)	GPSM$_{AZ}$(n)	GPSM$_{MM}$(n)	GPSM$_{SD}$(n)
$n = 1$	Original		144.657*	42.460	29.165	52.112	49.528
	JA			-102.196*	-115.492*	-92.545*	-95.129*
	SLR(n)				-13.295	9.652	7.068
	GPSM$_{AZ}$(n)					22.947	20.363
	GPSM$_{MM}$(n)						-2.584
	GPSM$_{SD}$(n)						
$n = 2$	Original		144.657*	72.334*	54.881*	87.701*	91.614*
	JA			-72.323*	-89.776*	-56.956*	-53.044*
	SLR(n)				-17.453	15.367	19.279
	GPSM$_{AZ}$(n)					32.821	36.733
	GPSM$_{MM}$(n)						3.912
	GPSM$_{SD}$(n)						
$n = 3$	Original		144.657*	94.480*	68.615*	120.754*	125.059*
	JA			-50.177*	-76.042*	-23.904	-19.598
	SLR(n)				-25.865	26.274	30.579
	GPSM$_{AZ}$(n)					52.138*	56.444*
	GPSM$_{MM}$(n)						4.306

	GPSM$_{SD}$(*n*)						
n = 4	Original		144.657*	113.849*	85.417*	150.018*	154.481*
	JA			-30.808	-59.240*	5.361	9.823
	SLR(*n*)				-28.432	36.169	40.632
	GPSM$_{AZ}$(*n*)					64.601*	69.064*
	GPSM$_{MM}$(*n*)						4.462
	GPSM$_{SD}$(*n*)						

* Mean difference significant at the 0.05 level

Table C5. Multiple comparisons of mean differences in the average $\overline{FT}$ in the *Unif*(10,20) order size environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(*n*)	GPSM$_{AZ}$(*n*)	GPSM$_{MM}$(*n*)	GPSM$_{SD}$(*n*)
n = 1	Original		216.892*	54.461	41.753	68.556*	71.486*
	JA			-162.431*	-175.139*	-148.336*	-145.406*
	SLR(*n*)				-12.708	14.095	17.025
	GPSM$_{AZ}$(*n*)					26.803	29.733
	GPSM$_{MM}$(*n*)						2.930
	GPSM$_{SD}$(*n*)						
n = 2	Original		216.892*	100.288*	73.472*	116.596*	123.383*
	JA			-116.604*	-143.419*	-100.296*	-93.509*
	SLR(*n*)				-26.815	16.308	23.095
	GPSM$_{AZ}$(*n*)					43.124	49.910
	GPSM$_{MM}$(*n*)						6.787
	GPSM$_{SD}$(*n*)						
n = 3	Original		216.892*	138.916*	98.224*	157.297*	164.920*
	JA			-77.976*	-118.668*	-59.595	-51.971*
	SLR(*n*)				-40.692	18.381	26.004
	GPSM$_{AZ}$(*n*)					59.073*	66.696*
	GPSM$_{MM}$(*n*)						7.623
	GPSM$_{SD}$(*n*)						
n = 4	Original		216.892*	171.478*	116.889*	191.426*	202.302*
	JA			-45.414*	-100.002*	-25.466	-14.589
	SLR(*n*)				-54.589*	19.948	30.824
	GPSM$_{AZ}$(*n*)					74.537*	85.413*
	GPSM$_{MM}$(*n*)						10.876
	GPSM$_{SD}$(*n*)						

* Mean difference significant at the 0.05 level

Table C6. Multiple comparisons of mean differences in the average $\overline{FT}$ in 200 orders in the order list environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(*n*)	GPSM$_{AZ}$(*n*)	GPSM$_{MM}$(*n*)	GPSM$_{SD}$(*n*)
n = 1	Original		235.929*	69.321*	53.654	67.399*	62.195*
	JA			-166.607*	-182.274*	-168.530*	-173.734*
	SLR(*n*)				-15.667	-1.923	-7.126
	GPSM$_{AZ}$(*n*)					13.744	8.541
	GPSM$_{MM}$(*n*)						-5.204
	GPSM$_{SD}$(*n*)						
n = 2	Original		235.929*	118.812*	94.709*	116.800*	112.905*
	JA			-117.117*	-141.220*	-119.129*	-123.024*
	SLR(*n*)				-24.103	-2.012	-5.907
	GPSM$_{AZ}$(*n*)					22.091	18.196
	GPSM$_{MM}$(*n*)						-3.895
	GPSM$_{SD}$(*n*)						

$n = 3$	Original		235.929*	155.636*	118.416*	154.732*	153.536*
	JA			-80.292*	-117.518*	-81.196*	-82.393*
	SLR(n)				-37.220*	-0.904	-2.101
	GPSM$_{AZ}$(n)					36.316*	35.120
	GPSM$_{MM}$(n)						-1.197
	GPSM$_{SD}$(n)						
$n = 4$	Original		235.929*	182.875*	132.236*	183.524*	184.098*
	JA			-53.054*	-103.693*	-52.405*	-51.830*
	SLR(n)				-50.639*	0.649	1.224
	GPSM$_{AZ}$(n)					51.288*	51.863*
	GPSM$_{MM}$(n)						0.575
	GPSM$_{SD}$(n)						

* Mean difference significant at the 0.05 level

Table C7. Multiple comparisons of mean differences in the average $\overline{FT}$ in the non-identical picker skill environment based on Games-Howell post-hoc tests.

Stage	Method	Original	JA	SLR(n)	GPSM$_{AZ}$(n)	GPSM$_{MM}$(n)	GPSM$_{SD}$(n)
$n = 1$	Original		147.011*	49.026	61.845*	50.186	52.989
	JA			-97.985*	-85.166*	-96.825*	-94.022*
	SLR(n)				12.819	1.160	3.962
	GPSM$_{AZ}$(n)					-11.659	-8.856
	GPSM$_{MM}$(n)						2.803
	GPSM$_{SD}$(n)						
$n = 2$	Original		147.011*	84.656*	106.695*	91.732*	94.720*
	JA			-62.356*	-40.316*	-55.279*	-52.291*
	SLR(n)				22.040	7.076	10.064
	GPSM$_{AZ}$(n)					-14.963	-11.975
	GPSM$_{MM}$(n)						2.988
	GPSM$_{SD}$(n)						
$n = 3$	Original		147.011*	114.430*	141.455*	125.281*	127.133*
	JA			-32.581*	-5.556	-21.730	-19.878
	SLR(n)				27.025	10.851	12.703
	GPSM$_{AZ}$(n)					-16.174	-14.322
	GPSM$_{MM}$(n)						1.852
	GPSM$_{SD}$(n)						
$n = 4$	Original		147.011*	134.200*	167.025*	148.532*	151.843*
	JA			-12.811	20.014	1.521	4.832
	SLR(n)				32.825*	14.332	17.643
	GPSM$_{AZ}$(n)					-18.494	-15.182
	GPSM$_{MM}$(n)						3.311
	GPSM$_{SD}$(n)						

* Mean difference significant at the 0.05 level

References

Bartholdi, J., & Eisenstein, D. (1996). Bucket brigades: A self-organizing order picking system for a warehouse. *Report, School of Industrial Engineering, Georgia Tech, Atlanta, USA*.

Bartholdi, J. J., & Hackman, S. T. (2008). *Warehouse & Distribution Science: Release 0.89*. Supply Chain and Logistics Institute Atlanta.

Bect, J., Ginsbourger, D., Li, L., Picheny, V., & Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, *22*(3), 773-793. https://doi.org/10.1007/s11222-011-9241-4

Brynzér, H., & Johansson, M. I. (1996). Storage location assignment: Using the product structure to reduce order picking times. *International Journal of Production Economics*, *46-47*, 595-603. https://doi.org/10.1016/0925-5273(94)00091-3

Cahyo, S. D. (2017). *Analysis of Shortage Delay in a Zone Picking System with AS/RS Replenishment* [Master's thesis, Pusan National University]. http://lib.pusan.ac.kr/en/en/resource/catalog/?app=solars&mod=detail&record_id=93752046

Caron, F., Marchet, G., & Perego, A. (2010). Optimal layout in low-level picker-to-part systems. *International Journal of Production Research*, *38*(1), 101-117. https://doi.org/10.1080/002075400189608

Chiang, D. M.-H., Lin, C.-P., & Chen, M.-C. (2011). The adaptive approach for storage assignment by mining data of warehouse management system for distribution centres. *Enterprise Information Systems*, *5*(2), 219-234. https://doi.org/10.1080/17517575.2010.537784

Chiang, D. M.-H., Lin, C.-P., & Chen, M.-C. (2014). Data mining based storage assignment heuristics for travel distance reduction. *Expert Systems*, *31*(1), 81-90. https://doi.org/10.1111/exsy.12006

Chuang, Y.-F., Lee, H.-T., & Lai, Y.-C. (2012). Item-associated cluster assignment model on storage allocation problems. *Computers & Industrial Engineering*, *63*(4), 1171-1177. https://doi.org/10.1016/j.cie.2012.06.021

De Koster, R., Le-Duc, T., & Zaerpour, N. (2012). Determining the number of zones in a pick-and-sort order picking system. *International Journal of Production Research*, *50*(3), 757-771. https://doi.org/10.1080/00207543.2010.543941

De Koster, R. B. M., Le-Duc, T., & Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. *European Journal of Operational Research*, *182*(2), 481-501.

Gaast, J. P. v. d., Koster, R. B. M. d., Adan, I. J. B. F., & Resing, J. A. C. (2020). Capacity Analysis of Sequential Zone Picking Systems. *Operations Research*, *68*(1), 161-179. https://doi.org/10.1287/opre.2019.1885

Ghomri, L., & Sari, Z. (2017). Mathematical modeling of the average retrieval time for flow-rack automated storage and retrieval systems. *Journal of Manufacturing Systems*, *44*, 165-178. https://doi.org/10.1016/j.jmsy.2017.05.002

Gu, J., Goetschalckx, M., & McGinnis, L. F. (2007). Research on warehouse operation: A comprehensive review. *European Journal of Operational Research*, *177*(1), 1-21. https://doi.org/10.1016/j.ejor.2006.02.025

Hong, S. (2019). A performance evaluation of bucket brigade order picking systems: Analytical and simulation approaches. *Computers & Industrial Engineering*, *135*, 120-131. https://doi.org/10.1016/j.cie.2019.05.037

Huang, M., Guo, Q., Liu, J., & Huang, X. (2018). Mixed Model Assembly Line Scheduling Approach to Order Picking Problem in Online Supermarkets. *Sustainability*, *10*(11), 3931. https://www.mdpi.com/2071-1050/10/11/3931

Jane, C.-C., & Laih, Y.-W. (2005). A clustering algorithm for item assignment in a synchronized zone order picking system. *European Journal of Operational Research*, *166*(2), 489-496. https://doi.org/10.1016/j.ejor.2004.01.042

Jane, C. C. (2000). Storage location assignment in a distribution center. *International Journal of Physical Distribution & Logistics Management*, *30*(1), 55-71. https://doi.org/10.1108/09600030010307984

Jewkes, E., Lee, C., & Vickson, R. (2004). Product location, allocation and server home base location for an order picking line with multiple servers. *Computers & Operations Research*, *31*(4), 623-636. https://doi.org/10.1016/S0305-0548(03)00035-2

Kim, J.-h., & Hong, S. (2020). A dynamic storage location assignment model for a progressive bypass zone picking system with an S/R crane. In P. N. University (Ed.).

Kim, J., & Hong, S. (2022). A dynamic storage location assignment model for a progressive bypass zone picking system with an S/R crane. *Journal of the Operational Research Society*, *73*(5), 1155-1166. https://doi.org/10.1080/01605682.2021.1892462

Kübler, P., Glock, C. H., & Bauernhansl, T. (2020). A new iterative method for solving the joint dynamic storage location assignment, order batching and picker routing problem in manual picker-to-parts warehouses. *Computers & Industrial Engineering*, *147*, 106645. https://doi.org/10.1016/j.cie.2020.106645

Larco, J. A., de Koster, R., Roodbergen, K. J., & Dul, J. (2016). Managing warehouse efficiency and worker discomfort through enhanced storage assignment decisions. *International Journal of Production Research*, *55*(21), 6407-6422. https://doi.org/10.1080/00207543.2016.1165880

Le-Duc, T., & De Koster, R. B. M. (2005). Travel distance estimation and storage zone optimization in a 2-block class-based storage strategy warehouse. *International Journal of Production Research*, *43*(17), 3561-3581. https://doi.org/10.1080/00207540500142894

Li, J., Moghaddam, M., & Nof, S. Y. (2016). Dynamic storage assignment with product affinity and ABC classification—a case study. *The International Journal of Advanced Manufacturing Technology*, *84*(9-12), 2179-2194. https://doi.org/10.1007/s00170-015-7806-7

Mackay, D. J. C. (1998). Introduction to Gaussian processes. *NATO ASI series. Series F: computer and system sciences*, 133-165.

Muppani, V. R., & Adil, G. K. (2008). A branch and bound algorithm for class based storage location assignment. *European Journal of Operational Research*, *189*(2), 492-507. https://doi.org/10.1016/j.ejor.2007.05.050

Murphy, K. P. (2012). *Machine learning : a probabilistic perspective*. MIT Press. http://mitpressebooks.mit.edu/product/machine-learning

Pan, J. C.-H., Shih, P.-H., & Wu, M.-H. (2015). Order batching in a pick-and-pass warehousing system with group genetic algorithm. *Omega*, *57*, 238-248. https://doi.org/https://doi.org/10.1016/j.omega.2015.05.004

Pan, J. C.-H., Shih, P.-H., Wu, M.-H., & Lin, J.-H. (2015). A storage assignment heuristic method based on genetic algorithm for a pick-and-pass warehousing system. *Computers & Industrial Engineering*, *81*, 1-13. https://doi.org/10.1016/j.cie.2014.12.010

Pan, J. C.-H., Wu, M.-H., & Chang, W.-L. (2014). A travel time estimation model for a high-level picker-to-part system with class-based storage policies. *European Journal of Operational Research*, *237*(3), 1054-1066. https://doi.org/10.1016/j.ejor.2014.02.037

Pan, J. C. H., & Wu, M. H. (2009). A study of storage assignment problem for an order picking line in a pick-and-pass warehousing system. *Computers & Industrial Engineering*, *57*(1), 261-268. https://doi.org/10.1016/j.cie.2008.11.026

Pang, K.-W., & Chan, H.-L. (2016). Data mining-based algorithm for storage location assignment in a randomised warehouse. *International Journal of Production Research*, *55*(14), 4035-4052. https://doi.org/10.1080/00207543.2016.1244615

Pazour, J. A., & Carlo, H. J. (2015). Warehouse reshuffling: Insights and optimization. *Transportation Research Part E: Logistics and Transportation Review*, *73*, 207-226. https://doi.org/10.1016/j.tre.2014.11.002

Quintanilla, S., Pérez, Á., Ballestín, F., & Lino, P. (2014). Heuristic algorithms for a storage location assignment problem in a chaotic warehouse. *Engineering Optimization*, *47*(10), 1405-1422. https://doi.org/10.1080/0305215x.2014.969727

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press. Table of contents only http://www.loc.gov/catdir/toc/fy0614/2005053433.html

Richardson, R. R., Osborne, M. A., & Howey, D. A. (2017). Gaussian process regression for forecasting battery state of health. *Journal of Power Sources*, *357*, 209-219. https://doi.org/10.1016/j.jpowsour.2017.05.004

Roodbergen, K. J., & Vis, I. F. A. (2009). A survey of literature on automated storage and retrieval systems. *European Journal of Operational Research*, *194*(2), 343-362. https://doi.org/10.1016/j.ejor.2008.01.038

Sadiq, M., Landers, T. L., & Don Taylor, G. (1996). An Assignment Algorithm for Dynamic Picking Systems. *IIE Transactions*, *28*(8), 607-616. https://doi.org/10.1080/15458830.1996.11770706

Vanheusden, S., van Gils, T., Braekers, K., Ramaekers, K., & Caris, A. (2022). Analysing the effectiveness of workload balancing measures in order picking operations. *International Journal of Production Research*, *60*(7), 2126-2150. https://doi.org/10.1080/00207543.2021.1884307

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32-35.